
Analytically deriving Partial Information Decomposition for affine systems of stable and convolution-closed distributions

Chaitanya Goswami
Electrical & Computer Engineering,
Carnegie Mellon University,
Pittsburgh, PA-15213.
cgoswami@andrew.cmu.edu

Amanda Merkley
Electrical & Computer Engineering,
Carnegie Mellon University,
Pittsburgh, PA-15213.
amerkley@andrew.cmu.edu

Abstract

Bivariate partial information decomposition (PID) has emerged as a promising tool for analyzing interactions in complex systems, particularly in neuroscience. PID achieves this by decomposing the information that two sources (e.g., different brain regions) have about a target (e.g., a stimulus) into unique, redundant, and synergistic terms. However, the computation of PID remains a challenging problem, often involving optimization over distributions. While several works have been proposed to compute PID terms *numerically*, there is a surprising dearth of work on computing PID terms *analytically*. The only known analytical PID result is for jointly Gaussian distributions. In this work, we present two theoretical advances that enable analytical calculation of the PID terms for numerous well-known distributions, including distributions relevant to neuroscience, such as Poisson, Cauchy, and binomial. Our first result generalizes the analytical Gaussian PID result to the much larger class of stable distributions. We also discover a theoretical link between PID and the emerging fields of data thinning and data fission. Our second result utilizes this link to derive analytical PID terms for two more classes of distributions: convolution-closed distributions and a sub-class of the exponential family. Furthermore, we provide an analytical upper bound for approximately computing PID for convolution-closed distributions, whose tightness we demonstrate in simulation.

1 Introduction

Bivariate partial information decomposition¹ (PID) is an information-theoretic framework developed for answering a central inquiry in many neuroscientific and machine learning studies: *how do two sources, X and Y , jointly process information about a target M ?* PID answers this question by quantifying the M -specific information contained in different interactions between X and Y . Specifically, it decomposes the total information X and Y have about M into *four* components: (i) the information about M contained uniquely in X , (ii) the information about M contained uniquely in Y , (iii) the redundant information about M contained in both X and Y , and (iv) the synergistic information about M which arises from the interaction between X and Y .

The PID terms offer novel insights for understanding interactions within complex systems, particularly in neuroscience. For instance, PID has been used to understand the firing patterns of grid cells [1, 2] and to study the flow of information in the visual cortex [3]. The following list of works [1, 3, 4, 5, 6, 7, 8] demonstrate the application of PID in studying diverse neuroscientific questions.

¹Throughout this work we refer to bivariate PID as PID.

Beyond neuroscience, PID has found applications in multimodal learning for interpreting model predictions [9, 10], fair machine learning for defining and quantifying bias [11, 12, 13, 14], and understanding financial markets [15].

However, the primary constraint hindering broader adoption of PID is the difficulty of computing and estimating the PID terms. BROJA-PID [16], a widely applied PID framework, requires solving a constrained minimization problem over a set of probability distributions (see Sec. 2). This minimization problem can pose a considerable challenge, particularly when the underlying distributions are continuous. As a result, several recent works have been dedicated to providing *numerical algorithms* that solve the aforementioned minimization problem exactly or approximately [3, 9, 17, 18, 19, 20], with more emphasis on the case of discrete distributions.

Despite significant progress in *numerically* calculating PID (typically through discrete approximations), very few works exist on *analytically*² calculating PID. As of this writing, the only known analytical PID expressions exist for jointly Gaussian M , X , and Y [21, 22]. A fundamental property of the Gaussian system is that one of the unique information (UI) terms in its PID is guaranteed to be *zero*. This property greatly simplifies the computation of PID for Gaussian systems, as the rest of the PID terms can be easily derived by solving desirable linear equations specified in many PID frameworks (see Sec 2), bypassing the need for optimizing over distributions (see Sec. 3).

In this work, we show that numerous systems of random variables M , X , and Y expressing a particular “affine dependence structure” also exhibit the same fundamental property that at least one of the UI terms is zero. Consequently, we expand significantly on the existing Gaussian PID result by using this fundamental property to analytically compute PID for various systems of M , X , and Y employing well-known distributions such as Poisson, exponential, gamma, beta, negative binomial, multinomial, Cauchy, Lévy-stable and more. The main contributions of this work are:

1. We extend the Gaussian PID result to a much larger class of distributions, known as the stable distribution family [23, 24] in Sec. 4. These results provide the first known analytical PID for fat-tailed distributions.
2. We highlight a theoretical link between PID calculation and the fields of data thinning [25] and data fission [26] in Sec. 5. We utilize this link to derive analytical PID terms for two more distribution families: convolution-closed distributions [25] and certain exponential family distributions [26].
3. For convolution-closed distributions, we further derive an analytical upper bound on the objective of the minimization used for computing BROJA-PID. We use this upper bound to approximately compute PID for systems of M , X , and Y having a non-affine dependence structure. We show the goodness of our approximation by a simulation study in Sec. 6.

2 Background

Notation: We denote the set of all natural numbers, real numbers, and positive real numbers as \mathbb{N} , \mathbb{R} , and \mathbb{R}^+ , respectively. Vectors are denoted by bold-faced font and an arrow, and matrices are denoted by bold-faced font. Define $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Let \mathbf{I}_d be the identity matrix of size $d \times d$ and $[d] = \{1, \dots, d\} \forall d \in \mathbb{N}$. We denote $\vec{\mathbf{1}}_d$ and $\vec{\mathbf{0}}_d$ as d -dimensional vectors having all elements as 1 and 0, respectively. For brevity, the probability notations of the form $P(A|B)$ and $P(A)$ are always understood to be as $P(A = a|B = b)$ and $P(A = a)$, respectively. The general term ‘distribution’ is used to refer to both probability density function (p.d.f.) and probability mass function (p.m.f.) (whichever is appropriate depending upon the context). The symbol $A \perp\!\!\!\perp B|C$ denotes that A and B are conditionally independent given C , and similarly, $A \perp\!\!\!\perp B$ implies A and B are independent. The L_p -norm is denoted by $\|(\cdot)\|_p \forall p \in [0, \infty)$, and $|\cdot|$ denotes the absolute function.

PID Background: Suppose M, X, Y are random variables with joint distribution $P(M, X, Y)$. According to [16, 27], three desirable equalities for a bivariate PID are as follows:

$$\begin{aligned} I_P(M; [X, Y]) &= UI(M; X \setminus Y) + UI(M; Y \setminus X) + RI(M; X, Y) + SI(M; X, Y), \\ I_P(M; X) &= UI(M; X \setminus Y) + RI(M; X, Y), \\ I_P(M; Y) &= UI(M; Y \setminus X) + RI(M; X, Y). \end{aligned} \quad (1)$$

²Section 2 precisely defines the notion of “analytical calculating PID” adopted in this work.

Here, $I_P(M; X)$ and $I_P(M; Y)$ denote the mutual information [28] between M and X , and M and Y , respectively, under the distribution $P(M, X, Y)$. Similarly, $I_P(M; [X, Y])$ denotes the joint mutual information between M and $[X, Y]$ under the distribution $P(M, X, Y)$. $UI(M; X \setminus Y)$ and $UI(M; Y \setminus X)$ are the unique information about M in X and Y , respectively. $SI(M; X, Y)$ and $RI(M; X, Y)$ denote the respective synergistic and redundant information shared between X and Y about M . The linear system defined in (1) contains four variables in three equations. Hence, only one of UI , RI , or SI need be defined to evaluate all four PID terms. Proposing a suitable definition of PID is the focus of much research [16, 27, 29, 30, 31, 32, 33, 34, 35] and all our theoretical results are applicable for any Blackwellian PID [22, 36] or a PID definition satisfying the assumption (*) of Bertschinger et al. [16] (see Appx. B for a formal justification). For example, our results are applicable for the PID definitions proposed in [16, 22, 27, 29, 37]. In this work, we discuss our results in the context of the BROJA-PID [16, 38], which is the definition of PID adopted in previous works [3, 9, 20]:

$$UI(M; X \setminus Y) = \min_{Q \in \Delta_P} I_Q(M; X|Y) \text{ or equivalently } UI(M; Y \setminus X) = \min_{Q \in \Delta_P} I_Q(M; Y|X), \quad (2)$$

where $\Delta_P = \{Q(M, X, Y) : Q(M, X) = P(M, X), Q(M, Y) = P(M, Y)\}$ and $I_Q(M; X|Y)$ is the conditional mutual information under the distribution $Q(M, X, Y)$. Note that the minimizing distributions for both problems shown in (2) are the same [16], and consequently, both minimization problems are equivalent. We refer the reader to [39] for a review on PID. In this work, we define “analytically calculating PID” as analytically solving (2) by providing an explicit construction of the minimizing distribution $Q^*(M, X, Y)$ that minimizes (2). Note that only the distributions $Q^*(M, X, Y)$ and $P(M, X, Y)$ are needed to compute the BROJA-PID terms.

We briefly describe two distribution families used in Sec. 4 and Sec. 5 (see Appx. N for more details).

Stable distribution family: Stable distributions are a family of distributions that naturally arise in the context of generalized central limit theorems. Some well-known members of this family are the Gaussian, Cauchy, Poisson, and Lévy distributions. A defining feature of stable distributions is that the sum of two independent copies of a random variable X , denoted as X_1 and X_2 , has the same distribution as a translated and scaled version of X [23, 24, 40]. In this work, we consider five sub-classes of stable distributions:

(i) *Continuous stable distributions* are parameterized through four parameters: stability parameter $\alpha \in (0, 2]$, skewness parameter $\beta \in [-1, 1]$, scale parameter $\gamma \in (0, \infty)$, and location parameter $\mu \in \mathbb{R}$. We denote its p.d.f. as $p_{CS}(\alpha, \beta, \gamma, \mu)$. Note that all continuous stable distribution (except Gaussian) are fat-tailed.

(ii) *Independent component multivariate stable distributions* describe the distribution of \vec{X} consisting of d independent random variables $\{X_j\}_{j=1}^d$, such that each $X_j \sim p_{CS}(\alpha, \beta_j, \gamma_j, \mu_j)$. We denote its p.d.f. as $p_{CS-IC}(\alpha, \vec{\beta}, \vec{\gamma}, \vec{\mu})$.

(iii) *Elliptically-contoured multivariate stable distributions* are the distributions of continuous stable random vectors whose p.d.f. has elliptical contours, e.g., the multivariate Gaussian distribution. We denote its p.d.f. as $p_{CS-EC}(\alpha, \Sigma, \vec{\mu})$. Here, Σ is a positive definite matrix, $\vec{\mu} \in \mathbb{R}^d$, and $\alpha \in (0, 2]$.

(iv) *Discrete stable distributions* are the discrete analogues of the continuous stable distributions. The p.m.f. of discrete stable distributions, denoted as $P_{DS}(\nu, \tau)$, are parameterized through two parameters: rate parameter $\tau > 0$ and exponent $0 < \nu \leq 1$. Discrete stable distributions do not have well-known multivariate generalizations except the Poisson distribution.

(v) *Multivariate Poisson distribution:* We use the multivariate Poisson distribution proposed in [41, 42, 43, 44], denoted as $\text{Poisson}(d, d', \vec{\Lambda})$, which represents each random variable in the random vector \vec{N} as a sum of independent Poisson random variables. Formally, $\vec{N} \sim \text{Poisson}(d, d', \vec{\Lambda})$ if $\vec{N} = \mathbf{A}\vec{N}^g$, where $\mathbf{A} = [\mathbf{A}_1 \dots \mathbf{A}_{d'}]$. Here, \mathbf{A}_i denotes a $d \times \binom{d}{i}$ submatrix having no duplicate columns, where each of its columns contain exactly i ones and $(d-i)$ zeros [44]. The vector

$$\vec{N}^g = [N_1^g \dots N_d^g \ N_{12}^g \dots \ N_{d-(d'-1)\dots d}^g]^T$$

with mutually independent $N_{i_1 \dots i_j}^g \sim \text{Poisson}(\lambda_{i_1 \dots i_j}) \ \forall \ (i_1, \dots, i_j) \in \mathbb{A}_j^d, j \in [d']$, where $\mathbb{A}_i^d = \{(j_1, \dots, j_i) : j_1 < j_2 < \dots < j_i, \text{ and } j_1, \dots, j_i \in [d]\}$, e.g., $\mathbb{A}_2^3 = \{(1, 2), (1, 3), (2, 3)\}$. Note that $d' \leq d$, and $\vec{\Lambda} = [\lambda_1 \dots \lambda_{d-(d'-1)\dots d}]^T$.

We refer the reader to [23, 24, 40, 42, 44, 45, 46, 47] for more details on stable distributions.

Convolution-closed distributions: Convolution-closed distributions form a large class of distributions that are closed under convolution in some parameter δ . Formally, we define the convolution-closed distribution as follows: Let $\mathcal{F}_{\mathcal{D}}$ denote a family of distributions, where each member distribution $f(\delta) \in \mathcal{F}_{\mathcal{D}}$ is indexed by a parameter $\delta \in \mathcal{D}$. Consider $X_1 \sim f(\delta_1)$, $X_2 \sim f(\delta_2)$, and $X_1 \perp\!\!\!\perp X_2$ for some $\delta_1, \delta_2 \in \mathcal{D}$. Then, $\mathcal{F}_{\mathcal{D}}$ is convolution-closed in the parameter δ if

$$X_1 + X_2 \sim f(\delta_1) * f(\delta_2) = f(\delta_1 + \delta_2) \quad \forall \delta_1, \delta_2 \in \mathcal{D} \text{ such that } \delta_1 + \delta_2 \in \mathcal{D}, \quad (3)$$

where $*$ denotes the convolution operator. Many well-known distributions can be considered convolution-closed in some parameter [25], such as the Poisson distribution, Gaussian distribution, gamma distribution, etc. Table 1 in [25] lists various examples of convolution-closed distributions.

3 A sufficient condition for computing PID terms analytically

The main focus of this work is to analyze the cases where (2) is analytically solvable. A sufficient condition for solving (2) is to show that Δ_P contains a distribution $Q_{MC}(M, X, Y)$ with the Markovian structure $M \rightarrow X \rightarrow Y$ or $M \rightarrow Y \rightarrow X$. We briefly discuss the argument justifying this sufficient condition. Consider the case where $Q_{MC}(M, X, Y) \in \Delta_P$ and has the Markovian structure $M \rightarrow X \rightarrow Y$. First, we note that $UI(M; Y|X) = \min_{Q \in \Delta_P} I_Q(M; Y|X) \geq 0$ due to the non-negativity of conditional mutual information [28]. Second, $Y \perp\!\!\!\perp M|X$ due to the Markovian structure of $Q_{MC}(M, X, Y)$, which implies that $I_{Q_{MC}}(M; Y|X) = 0$ (conditionally independent random variables have zero conditional mutual information [28]). Hence, $Q_{MC}(M, X, Y)$ achieves the lower bound of zero for the minimization problem $\min_{Q \in \Delta_P} I_Q(M; Y|X)$, showing that $Q_{MC}(M, X, Y)$ indeed minimizes (2). A similar argument can be made for the case when $Q_{MC}(M, X, Y)$ has the Markovian structure $M \rightarrow Y \rightarrow X$. Proposition 1 in Appx. A formalizes the above argument.

The above sufficient condition provides an easy way to analytically calculate the PID terms, as it ensures one of the unique information terms is always zero. Consequently, the remaining PID terms can be calculated by substituting zero for the appropriate unique information term in (1) and solving the resultant linear system. Hence, if applicable, the above sufficient condition considerably simplifies the calculation of PID terms by circumventing the need for optimizing over a set of distributions. Surprisingly, many well-known distribution families allow intuitive constructions of $P(M, X, Y)$, for which the above sufficient condition is applicable. In the following sections, we provide theorems that specify sufficient conditions under which the existence of these Markov chains in Δ_P can be guaranteed for these $P(M, X, Y)$.

4 Computing PID for stable distributions

In this section, we extend existing results for analytical PID computation of jointly Gaussian M , X , and Y [21, 22] to the much larger class of stable distributions. Our results utilize two key observations to provide the generalization of the Gaussian results: (i) We first identify that the analytical computation of PID for jointly Gaussian systems is due to their particular “affine dependence” structure; (ii) We show that these particular affine dependence structures are not unique to Gaussian systems, but rather extend to many members of the stable distribution family.

For jointly Gaussian M , X , and Y , the conditional distributions $P(X|M) = \mathcal{N}(aM + b, \sigma_X^2)$ and $P(Y|M) = \mathcal{N}(cM + d, \sigma_Y^2)$ are also Gaussian distributions, where their means are an affine function of M and their variances are fixed with respect to M . This particular affine dependence of X and Y on M is the key to the analytical calculation of their PID, as it guarantees existence of a Markov chain $Q_{MC}(M, X, Y) \in \Delta_P$. Thus, we can apply the sufficient condition described in Sec. 3 to compute the PID terms. We illustrate through an example: consider $P(X|M) = \mathcal{N}(M, \sigma_X^2)$ and $P(Y|M) = \mathcal{N}(M, \sigma_Y^2)$ where $\sigma_X^2 < \sigma_Y^2$, and $M \sim P(M)$ for some appropriate $P(M)$. Then, we can explicitly construct $Q_{MC}(M, X, Y) \in \Delta_P$ with the Markovian structure $M \rightarrow X \rightarrow Y$ as follows: Choose $Q_{MC}(M, X, Y) = P(M)P(X|M)Q_{MC}(Y|X)$ where $Q_{MC}(Y|X)$ is specified through the addition of independent Gaussian noise, i.e., $Y = X + \epsilon$. Here, $\epsilon \sim \mathcal{N}(0, \sigma_Y^2 - \sigma_X^2)$ and $\epsilon \perp\!\!\!\perp (M, X)$. It is easy to verify that $Q_{MC}(Y|M) = \mathcal{N}(M, \sigma_Y^2)$, which implies that $Q_{MC}(M, Y) = P(M, Y)$. By construction, $Q_{MC}(M, X) = P(M, X)$, and hence, $Q_{MC}(M, X, Y) \in \Delta_P$.

The above example is a special case of a well-known result where, for a scalar M , one can always construct a lower signal-to-noise ratio (SNR) “Gaussian channel”, i.e. $P(Y|M)$ in our example, by adding independent Gaussian noise to a higher SNR “Gaussian channel” [22, 48], i.e. $P(X|M)$ in our example. Surprisingly, the technique of adding independent noise to construct Markov chains contained in Δ_P can be extended to $P(X|M)$ and $P(Y|M)$ as members of stable distributions. Theorems 1 and 4 generalize the above construction for $P(X|M)$ and $P(Y|M)$ as members of univariate continuous and univariate discrete stable distributions, respectively. Theorems 2, 3, and 5 consider the case of multivariate stable distributions (see Sec. 2). The key technique for proving these theorems is that a Markov chain $Q_{MC}(M, X, Y) \in \Delta_P$ can always be constructed by adding appropriate independent noise to a higher SNR $P(X|M)$ to obtain a lower SNR $P(Y|M)$, similar to our above example.

4.1 PID of univariate affine continuous stable system

Theorem 1 can be viewed as a direct generalization of Barrett’s Gaussian PID result [21] to stable distributions, showing one of the UI terms is always zero. We begin by formally describing the univariate affine continuous stable system that generalizes the Gaussian system of Barrett’s [21]. Let the joint distribution of M, X , and Y , denoted as $P(M, X, Y)$, satisfy the following properties: (i) $M \sim P(M)$ with support set $\mathcal{M} \subseteq \mathbb{R}$; (ii) $P(X|M)$ and $P(Y|M)$ are univariate continuous stable distributions with an affine dependence on M , i.e., $P(X|M) = p_{CS}(\alpha, \beta_X, \gamma_X, aM + b)$ and $P(Y|M) = p_{CS}(\alpha, \beta_Y \text{sgn}(ac), \gamma_Y, cM + d)$, where $a, b, c, d \in \mathbb{R}$.

Theorem 1. *Let M, X , and Y be random variables whose joint distribution $P(M, X, Y)$ is described by a univariate affine continuous stable system. Without the loss of generality, assume $|a|/\gamma_X \geq |c|/\gamma_Y$. If $1 - \beta_Y \geq (\gamma_X |c|/\gamma_Y |a|)^\alpha (1 - \beta_X)$ and $1 + \beta_Y \geq (\gamma_X |c|/\gamma_Y |a|)^\alpha (1 + \beta_X)$, then Δ_P contains a Markov chain of the form $M \rightarrow X \rightarrow Y$ and $UI(M; Y \setminus X) = 0$.*

Proof. See Appx. F for the proof. Here, $|a|/\gamma_X$ and $|c|/\gamma_Y$ are the SNR analogues. \square

4.2 PID of multivariate affine continuous stable systems

We analyze two multivariate generalizations of Theorem 1 in Theorems 2 and 3. Namely, we consider independent component multivariate stable distribution (where all the components of the random vector are independent) and elliptically-contoured multivariate stable distribution (where the p.d.f. is elliptically contoured, similar to multivariate Gaussian distributions). We construct two systems employing these two sub-classes of multivariate continuous stable distributions and show that one of the UI terms is always zero for these systems. For both cases, denote the joint distribution of M, \vec{X} , and \vec{Y} as $P(M, \vec{X}, \vec{Y})$, where the support set of M is $\mathcal{M} \subseteq \mathbb{R}$. The dimensions of M, \vec{X} , and \vec{Y} are 1, d_X , and d_Y , respectively.

System 1: The random vectors \vec{X} and \vec{Y} satisfy the following equations:

$$\vec{X} = \vec{H}_X M + \mathbf{A}_X \vec{Z}_X + \vec{b}_X \text{ and } \vec{Y} = \vec{H}_Y M + \mathbf{A}_Y \vec{Z}_Y + \vec{b}_Y, \quad (4)$$

where $\vec{Z}_X \sim p_{CS-IC}(\alpha, \vec{0}_{d_X}, \vec{1}_{d_X}, \vec{0}_{d_X})$, $\vec{Z}_Y \sim p_{CS-IC}(\alpha, \vec{0}_{d_Y}, \vec{1}_{d_Y}, \vec{0}_{d_Y})$, \mathbf{A}_X and \mathbf{A}_Y are invertible matrices, $\vec{H}_X, \vec{b}_X \in \mathbb{R}^{d_X}$, and $\vec{H}_Y, \vec{b}_Y \in \mathbb{R}^{d_Y}$.

Theorem 2. *Let System 1 describe the joint distribution $P(M, \vec{X}, \vec{Y})$ of M, \vec{X} , and \vec{Y} . Without the loss of generality, assume $\|\mathbf{A}_Y^{-1} \vec{H}_Y\|_\kappa \leq \|\mathbf{A}_X^{-1} \vec{H}_X\|_\kappa$, where $\kappa = \alpha/\alpha - 1 \forall \alpha \in (1, 2]$ and $\kappa = \infty \forall \alpha \in (0, 1]$. Then, Δ_P contains a Markov chain of the form $M \rightarrow \vec{X} \rightarrow \vec{Y}$ and $UI(M; \vec{Y} \setminus \vec{X}) = 0$.*

Proof. See Appx. G for the proof. Here, $\|\mathbf{A}_Y^{-1} \vec{H}_Y\|_\kappa$ and $\|\mathbf{A}_X^{-1} \vec{H}_X\|_\kappa$ are the SNR analogues. \square

System 2: The conditional distribution of \vec{X} and \vec{Y} conditioned on M are as follows: $P(\vec{X}|M) = p_{CS-EC}(\alpha, \Sigma_X, \vec{H}_X M + \vec{b}_X)$ and $P(\vec{Y}|M) = p_{CS-EC}(\alpha, \Sigma_Y, \vec{H}_Y M + \vec{b}_Y)$, where Σ_X and Σ_Y are positive definite matrices, $\vec{H}_X, \vec{b}_X \in \mathbb{R}^{d_X}$, and $\vec{H}_Y, \vec{b}_Y \in \mathbb{R}^{d_Y}$.

Theorem 3. Let System 2 describe the joint distribution $P(M, \vec{X}, \vec{Y})$ of M , \vec{X} , and \vec{Y} . Define $\Sigma_X^{-1/2}$ and $\Sigma_Y^{-1/2}$ as the respective inverses of the matrices $\Sigma_X^{1/2}$ and $\Sigma_Y^{1/2}$ which satisfy: $(\Sigma_X^{1/2})^T \Sigma_X^{1/2} = \Sigma_X$, and $(\Sigma_Y^{1/2})^T \Sigma_Y^{1/2} = \Sigma_Y$. Without the loss of generality, assume $\|\Sigma_Y^{-1/2} \vec{H}_Y\|_2 \leq \|\Sigma_X^{-1/2} \vec{H}_X\|_2$. Then, Δ_P contains a Markov chain of the form $M \rightarrow \vec{X} \rightarrow \vec{Y}$ and $UI(M; \vec{Y} \setminus \vec{X}) = 0$.

Proof. See Appx. H for the proof. Here, $\|\Sigma_Y^{-1/2} \vec{H}_Y\|_2$ and $\|\Sigma_X^{-1/2} \vec{H}_X\|_2$ are the SNR analogues. \square

4.3 PID of univariate affine discrete stable system

The univariate affine discrete stable system is the discrete counterpart of the univariate affine continuous stable system described in Sec. 4.1. The formal description of the univariate affine discrete stable system is as follows. Let the joint distribution of M , X , and Y , denoted as $P(M, X, Y)$, satisfy the following properties: (i) $M \sim P(M)$ with support set $\mathcal{M} \subseteq (0, \infty)$; (ii) $P(X|M)$ and $P(Y|M)$ are univariate discrete stable distributions with an affine dependence on M , i.e., $P(X|M=m) = P_{DS}(\nu, am + b)$ and $P(Y|M=m) = P_{DS}(\nu, cm + d)$, where $a, b, c, d \in (0, \infty)$.

Theorem 4. Let M , X , and Y be random variables whose joint distribution $P(M, X, Y)$ is described by a univariate affine discrete stable system. Without the loss of generality, assume $a \geq c$. If $a/b \geq c/d$, then Δ_P contains a Markov chain of the form $M \rightarrow X \rightarrow Y$ and $UI(M; Y \setminus X) = 0$.

Proof. See Appx. I for the proof. Here, a and c are the SNR analogues. \square

4.4 PID of multivariate linear Poisson system

The Poisson distribution is the only the discrete stable distribution with a well-known multivariate extension. Hence, we analyze vector-generalizations of Theorem 4 only for the Poisson distribution. We now describe the multivariate linear Poisson system: Let the joint distribution $P(M, \vec{X}, \vec{Y})$ of M , \vec{X} , and \vec{Y} satisfy the following properties: (i) $M \sim P(M)$ with support set $\mathcal{M} \subseteq (0, \infty)$; (ii) $P(\vec{X}|M) = \text{Poisson}(d_X, d'_X, \vec{\Lambda}_X)$ and $P(\vec{Y}|M) = \text{Poisson}(d_Y, d'_Y, \vec{\Lambda}_Y)$, with:

$$\begin{aligned} \vec{\Lambda}_X &= \left[\lambda_1^X \cdots \lambda_{d_X - (d'_X - 1) \dots d_X}^X \right]^T, \lambda_{i_1 \dots i_j}^X = \gamma_{i_1 \dots i_j}^X M^j \forall j \in [d_X] \text{ and } (i_1, \dots, i_j) \in \mathbb{A}_j^{d_X}, \\ \vec{\Lambda}_Y &= \left[\lambda_1^Y \cdots \lambda_{d_Y - (d'_Y - 1) \dots d_Y}^Y \right]^T, \lambda_{i_1 \dots i_j}^Y = \gamma_{i_1 \dots i_j}^Y M^j \forall j \in [d_Y] \text{ and } (i_1, \dots, i_j) \in \mathbb{A}_j^{d_Y}. \end{aligned} \quad (5)$$

Theorem 5. Let the joint distribution $P(M, \vec{X}, \vec{Y})$ of M , \vec{X} , and \vec{Y} be described by the multivariate linear Poisson system defined above. Without the loss of generality, assume $d'_X \geq d'_Y$. If $\sum_{(i_1, \dots, i_j) \in \mathbb{A}_j^{d_X}} \gamma_{i_1 \dots i_j}^X \geq \sum_{(i_1, \dots, i_j) \in \mathbb{A}_j^{d_Y}} \gamma_{i_1 \dots i_j}^Y \forall j \in [d'_Y]$, then Δ_P contains a Markov chain of the form $M \rightarrow \vec{X} \rightarrow \vec{Y}$ and $UI(M; \vec{Y} \setminus \vec{X}) = 0$.

Proof. See Appx. J for the proof. Here, $\gamma_1^X, \dots, \gamma_{d_X - (d'_X - 1) \dots d_X}^X, \gamma_1^Y, \dots, \gamma_{d_Y - (d'_Y - 1) \dots d_Y}^Y \in \mathbb{R}^+$ are the SNR analogues. \square

5 Computing PID using data thinning and data fission strategies

Data thinning and data fission are emerging fields in machine learning and statistics dedicated to studying the procedures of splitting a random variable X into N different component random variables. These splitting procedures provide an attractive alternative to the standard splitting of datasets into training, validation, and test splits for performing cross-validation to select statistical model parameters, as they enable users to perform cross-validation even for the extreme case of a dataset containing a *single* datapoint. We refer the reader to the recent works of Neufeld et al. [25] and Leiner et al. [26] for a more comprehensive discussion on data thinning and fission, respectively.

The fields of data thinning and data fission share an important theoretical link with analytically calculating PID terms. Specifically, the tools developed for splitting random variables for data thinning and data fission can be readily used to calculate PID terms analytically. To give an intuition, suppose X contains more information about M than Y . Then, we can employ data fission and thinning strategies to decompose X into two components, $f_1(X)$ and $f_2(X)$, such that $f_2(X)$ follows the same distribution as Y . Then, $f_2(X)$ and Y convey the same information about M as they

are identically distributed and represent the redundant component of the PID terms. Similarly, $f_1(X)$ contains the information uniquely contained in X about M . Thus, $f_1(X)$ represents the unique information term. Theorems 6 and 7 utilize the data thinning and data fission proposed in [25, 26] to construct a Markov chain $Q_{MC}(M, X, Y) \in \Delta_P$ for several systems of random variables M , X , and Y . This allows us to use the sufficient condition discussed in Sec. 3 to compute their PID.

5.1 PID for convolution-closed distribution based on data thinning strategies of Neufeld et al.

Neufeld et al. [25] introduces data thinning for a large family of distributions known as convolution-closed distributions (see Sec. 2). An attractive property of convolution-closed distributions is that they provide a natural way to define a dilation/thinning operation. Formally, let $X \sim f(\delta)$, then we define X_ϵ as the ϵ -dilated version of X if $X_\epsilon \sim f(\epsilon\delta)$ for some $\epsilon \in (0, 1)$ such that $\epsilon\delta \in \mathcal{D}$. Furthermore, if we assume $(1 - \epsilon)\delta \in \mathcal{D}$, then $P(X_\epsilon|X) = P(X_\epsilon|X_\epsilon + X_{1-\epsilon})$, where $X_{1-\epsilon} \sim f((1 - \epsilon)\delta)$ and $X_\epsilon \perp\!\!\!\perp X_{1-\epsilon}$ (see lemma 19 for a formal justification). We denote $P(X_\epsilon|X = x) = P(X_\epsilon|X_\epsilon + X_{1-\epsilon} = x)$ as $G(\epsilon\delta, (1 - \epsilon)\delta, x)$. This dilation operation forms the basis of data thinning, as it enables X to be split into its dilated components X_ϵ and $X_{1-\epsilon}$, such that $X = X_\epsilon + X_{1-\epsilon}$, where $X_\epsilon \perp\!\!\!\perp X_{1-\epsilon}$. We utilize this dilation operation for analytically calculating the PID terms of the following linear convolution-closed system:

Linear convolution-closed system: Let \mathcal{F}_D be a convolution-closed distribution family as described in Sec. 2. The joint distribution $P(M, X, Y)$ of the random variables M , X , and Y describes a *linear convolution-closed system* if the distributions $P(X|M)$ and $P(Y|M)$ are defined as follows:

$$P(X|M=m) = f(\delta_m^X) \text{ and } P(Y|M=m) = f(\delta_m^Y) \text{ such that } \delta_m^X, \delta_m^Y \in \mathcal{D} \forall m \in \mathcal{M}, \quad (6)$$

where \mathcal{M} is the support of M . Furthermore, we assume $\delta_m^X = \gamma\delta_m^Y \forall m \in \mathcal{M}$ for some $\gamma \in \mathbb{R}^+$.

Theorem 6. Let the joint density $P(M, X, Y)$ of random variables M , X , and Y be described by a linear convolution-closed system. Without the loss of generality, assume $\gamma \leq 1$. If

- (a) $(1 - \gamma)\delta_m^Y \in \mathcal{D} \forall m \in \mathcal{M}$,
- (b) $P(X_\gamma|X_\gamma + X_{1-\gamma} = x, M=m) = G(\gamma\delta_m^X, (1 - \gamma)\delta_m^X, x)$ does not depend on m , where $P(X_\gamma|M) = f(\gamma\delta_m^X)$, $P(X_{1-\gamma}|M) = f((1 - \gamma)\delta_m^X)$ and $X_\gamma \perp\!\!\!\perp X_{1-\gamma}|M$,

then Δ_P contains a Markov chain of the form $M \rightarrow X \rightarrow Y$ and $UI(M; Y \setminus X) = 0$.

The proof of Theorem 6 is provided in Appx. K. Theorem 6 enables PID calculation for several well-known distributions such as gamma, Poisson, beta etc. (see Appx. C for a (non-exhaustive) list).

5.2 PID for distributions based on data fission strategies of Leiner et al.

Leiner et al. [26] propose a “conjugate-prior reversal” strategy for splitting a random variable X into two components, $f(X)$ and $g(X)$, for certain exponential family distributions. The distributions proposed for performing conjugate-prior reversal provide natural descriptions of $P(M, X, Y)$ (specified in theorem 7) for which PID can be calculated analytically. We briefly describe the distributions used in the conjugate-prior reversal strategy. Let $X \sim p_{exp1}(X)$, where

$$p_{exp1}(X = x; \theta_1, \theta_2) = H(\theta_1, \theta_2) \exp(\theta_1^T x - \theta_2^T A(x)), \quad (7)$$

for some appropriately defined $H(\cdot, \cdot)$, $A(\cdot)$, θ_1 and θ_2 . Furthermore, define a random variable Y through its conditional density $p(Y|X = x)$:

$$p(Y = y|X = x; \theta_3) = h(y) \exp(x^T T(y) - \theta_3^T A(x)), \quad (8)$$

for some $h(\cdot)$, $T(\cdot)$, and θ_3 , such that $p(Y = y|X = x; \theta_1, \theta_2)$ is a well-defined distribution. Then, the decomposition terms $f(X)$ and $g(X)$ are Y and X , respectively, in the conjugate-prior reversal strategy. Furthermore, the marginal distribution of Y is expressed as:

$$p_{exp2}(Y = y; \theta_1, \theta_2, \theta_3) = h(y)H(\theta_1, \theta_2)/H(\theta_1 + T(y), \theta_2 + \theta_3) \text{ (see proof of Theorem 1 in [26])}.$$

Theorem 7. Let M , X , and Y be random variables having the joint distribution $P(M, X, Y)$. Furthermore, the conditional distribution of X and Y conditioned on M are as follows: $P(X|M=m) = p_{exp1}(X; \theta_1(m), \theta_2(m))$ and $P(Y|M=m) = p_{exp2}(Y; \theta_1(m), \theta_2(m), \theta_3)$. Then, Δ_P contains a Markov chain of the form $M \rightarrow X \rightarrow Y$ and $UI(M; Y \setminus X) = 0$.

The proof of Theorem 7 can be found in Appx. L. The proof essentially stems from observing that the joint distribution $Q_{MC}(M, X, Y) = P(M)P(X|M)Q_{MC}(Y|X)$ lies in Δ_P , where $Q_{MC}(Y|X) = h(y) \exp(x^T T(y) - \theta_3^T A(x))$. A (non-exhaustive) list of well-known distributions for which theorem 7 is applicable is provided in Appx. D. Leiner et al. [26] also discuss some more strategies for performing data fission that do not follow the conjugate-prior reversal strategy. We provide the corresponding results for computing PID for these remaining data fission strategies and for additional miscellaneous distributions in Appx. E.

6 Upper bound for convolution-closed distributions

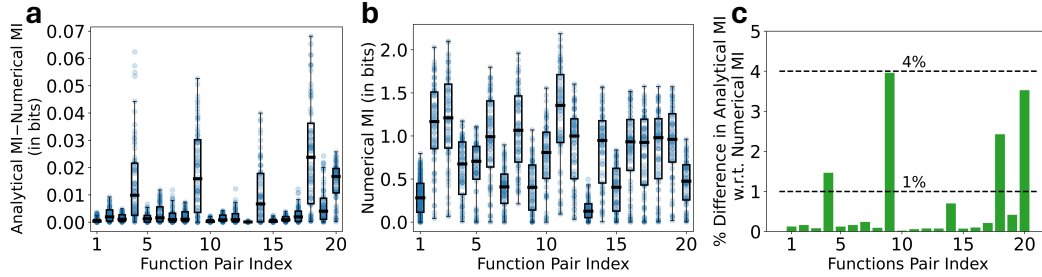


Figure 1: **a** and **b**, respectively, show the box plot of the difference $I_{Q_A}(M, X, Y) - I_{Q_N}(M, X, Y)$ and the corresponding values of $I_{Q_N}(M, X, Y)$ for the 20 different function pairs across the 75 different $P(M)$ distributions. The light-blue dots show the corresponding data points used for making the box plots. **c** shows the ratio of the median difference $I_{Q_A}(M, X, Y) - I_{Q_N}(M, X, Y)$ and the median value of $I_{Q_N}(M, X, Y)$ in percentage, for each function pair.

Several numerical methods, such as [3, 9, 20], approximately solve (2) by considering a smaller constraint set or minimizing an appropriate upper bound for calculating the PID terms. These numerical methods employ general approximations that rely on weak assumptions on the underlying distributions for solving (2), as their goal is to estimate PID for a large class of distributions. However, for specific applications, it is possible to make stronger assumptions on the underlying distributions (e.g., assuming the Poisson distribution for modeling neural spikes). This section illustrates how our theoretical analysis can benefit these numerical algorithms by providing more refined approximations for solving (2) that harness these stronger assumptions. Specifically, we construct an upper bound for the objective of (2) for convolution-closed distributions and show that, under certain assumptions, the upper bound can be *analytically minimized* over Δ_P in Sec. 6.1. Note that these upper bounds are applicable for more general cases than our theoretical results, as they do not require the sources X and Y to have an affine dependence on M . Consequently, our upper bound is also applicable in cases where both of the UI terms in the PID are non-zero, unlike our results in Sec. 4 and 5. We demonstrate the tightness of our upper bound through a simulation study in Sec. 6.2.

Notations and assumptions: We consider the minimization problem $\min_{Q \in \Delta_P} I_Q(M; [X, Y])$ for deriving the upper bound, as it was shown to be equivalent to (2) in [16]. The distribution $P(M, X, Y)$ for which we will construct our upper bound is specified as follows. The random variable M has support over \mathcal{M} , the conditional distributions $P(X|M)$ and $P(Y|M)$ are members of some convolution-closed distribution family $\mathcal{F}_{\mathcal{D}}$, and there exists some $\delta_{bias}^X, \delta_{bias}^Y \in \mathcal{D}$ such that

$$\begin{aligned} P(X|M=m) &= f(\delta_m^X) \text{ and } P(Y|M=m) = f(\delta_m^Y) \text{ where } \delta_m^X, \delta_m^Y \in \mathcal{D} \forall m \in \mathcal{M}, \text{ with} \\ (\delta_m^Y - \delta_{bias}^Y), (\delta_m^X - \delta_{bias}^X), (\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)) &\in \mathcal{D} \text{ and } \delta_m^X - \delta_{bias}^X = \epsilon_m^{(1)}(\delta_m^Y - \delta_{bias}^Y), \\ \delta_m^Y &= \epsilon_m^{(2)}(\delta_m^X - \delta_{bias}^X), \text{ and } \delta_m^X = \epsilon_m^{(3)}(\delta_m^X - \delta_{bias}^X) \text{ for } \epsilon_m^{(1)}, \epsilon_m^{(2)}, \epsilon_m^{(3)} \in [0, 1] \forall m \in \mathcal{M}. \end{aligned} \quad (9)$$

The above assumptions ensure that X and Y can always be decomposed into new random variables X', Y'', n_X, Y' , and n_Y (see (10)). We utilize these decomposed random variables to construct our upper bound. Hence, our upper bound is only applicable for systems satisfying (9). The above assumptions describe a large class of systems. For example, in the Poisson case, our upper bound is applicable for any $P(M, X, Y)$ having $P(X|M) = \text{Poisson}(f_1(M))$ and $P(Y|M) = \text{Poisson}(f_2(M))$, as long as $f_1(M) \geq f_2(M)$ over \mathcal{M} . In Appx. C, we provide numerous examples of systems for which (9) holds to illustrate the assumptions of (9).

6.1 Upper bound construction

First, we consider an arbitrary distribution $Q(M, X, Y) \in \Delta_P$. Therefore, we know $Q(X|M=m) = P(X|M=m) = f(\delta_m^X)$ and $Q(Y|M=m) = P(Y|M=m) = f(\delta_m^Y)$. We use the dilation properties of convolution-closed distributions (Sec. 5.1) to decompose X and Y into their respective dilated versions: (X', Y'', n_X) and (Y', n_Y) . From the results of Appx. M.1, we know that $X = X' + Y'' + n_X$ and $Y = Y' + n_Y$. Furthermore, (X', Y'', n_X) are mutually conditionally independent given M and $Y' \perp\!\!\!\perp n_Y | M$. Hence, we can construct the following Markov chain for any arbitrary $Q(M, X, Y) \in \Delta_P$:

$$M \rightarrow [X' \ Y'' \ Y' \ n_X \ n_Y]^T \rightarrow [X' + Y'' + n_X \ Y' + n_Y]^T = [X \ Y]^T. \quad (10)$$

We denote the joint distribution of $(M, X', Y'', Y', n_X, n_Y)$ as $\bar{Q}(M, X', Y'', Y', n_X, n_Y)$. We appropriately choose the dilation amounts for X and Y such that the respective conditional distributions of (X', Y'', Y', n_X, n_Y) are as follows:

$$\begin{aligned} \bar{Q}(X'|M=m) &= f(\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)), \quad \bar{Q}(Y''|M=m) = f(\delta_m^Y - \delta_{bias}^Y), \\ \bar{Q}(Y'|M=m) &= f(\delta_m^Y - \delta_{bias}^Y), \quad \bar{Q}(n_X|M) = f(\delta_{bias}^X), \text{ and } \bar{Q}(n_Y|M) = f(\delta_{bias}^Y). \end{aligned} \quad (11)$$

The distributions in (11) are well-defined due to (9). Appx. M.1 formally shows that a \bar{Q} , defined as above, exists for each $Q \in \Delta_P$. We use data-processing inequality and (10) to conclude:

$$I_Q(M; [X', Y'', n_X, Y', n_Y]) \geq I_Q(M; [X, Y]) \quad \forall Q(M, X, Y) \in \Delta_P. \quad (12)$$

Hence, (12) provides us the desired upper bound for our objective $I_Q(M; [X, Y])$. Furthermore, the minimization problem $\min_{Q \in \Delta_P} I_{\bar{Q}}(M; [X', Y'', n_X, Y', n_Y])$ is analytically solvable, and the minimizing distribution \bar{Q}^* has the structure:

$$Y' = Y'' \text{ and } (n_X, n_Y) \perp\!\!\!\perp (M, X', Y').$$

The corresponding distribution of (M, X, Y) , denoted as $Q_A(M, X, Y)$, can be found by appropriately manipulating \bar{Q}^* , as $X = X' + Y'' + n_X$ and $Y = Y' + n_Y$. The distribution Q_A serves as an approximate solution for the problem $\min_{Q \in \Delta_P} I_Q(M; [X, Y])$ and, consequently, (2). Note that if multiple δ_{bias}^X and δ_{bias}^Y exist satisfying (9), we optimize over the pairs $(\delta_{bias}^X, \delta_{bias}^Y)$ to further refine our approximate solution of (2).

6.2 Simulation study for numerically validating the upper bound

We illustrate the tightness of our upper bound through a simulation study on the Poisson distribution (a convolution-closed distribution). The reason for choosing the Poisson distribution is two-fold: (i) Poisson distribution is easily approximated as a discrete distribution over finite support, enabling calculation of ground-truth PID terms through numerical solvers such as [10, 17, 18], which are not readily available for continuous distributions; (ii) Many practical applications of PID have been in neuroscience, and the Poisson distribution is frequently used for modeling neural spikes in neuroscientific studies [49].

We compare the performance of our analytical estimate $Q_A(M, X, Y)$ with the numerical ground-truth estimate $Q_N(M, X, Y)$ for the Poisson distribution. The simulation setup is as follows: We choose a $P(M, X, Y)$ such that $P(X|M) = \text{Poisson}(f_1(M))$ and $P(Y|M) = \text{Poisson}(f_2(M))$. We chose 20 different pairs of non-linear $f_1(\cdot)$ and $f_2(\cdot)$ (enumerated in Table 1 in Appx. M), such that the assumptions in (9) are satisfied. The distribution $P(M)$ is chosen to be a discrete distribution. For each function pair, we compare $Q_A(M, X, Y)$ and $Q_N(M, X, Y)$ over 75 different distributions of $P(M)$ as shown in Fig. 1a. The 75 distributions are further sub-divided into three groups of 25 based on the number of the outcomes of M , which are 2, 4, and 8. For each of the 75 distributions, the values of M are randomly sampled from $[1, 4]$ and the values of $P(M)$ are randomly sampled from a simplex of appropriate dimensions. We compare the mutual information $I_{Q_A}(M; [X, Y])$ calculated using Q_A with the numerical ground-truth $I_{Q_N}(M; [X, Y])$ calculated using numerical solvers (Fig. 1b). For calculating $I_{Q_N}(M; [X, Y])$, we approximate the Poisson distribution as a finite discrete distribution by terminating its support at the smallest integer where its cumulative distribution is greater than 0.99. We use the code provided in [50] (under Apache 2.0-license) to numerically solve (2) using this discrete approximation of the Poisson distribution. Fig. 1c demonstrates that the analytical upper bound provided by Q_A is very tight for the tested function pairs (within $< 1\%$ of

the numerical ground-truth for 16 function pairs) and serves as a good approximation for solving (2). Furthermore, the tightness of our analytical upper bound suggests that Q_A might be an analytical solution of (2) for a larger class of systems of M , X , and Y having non-affine dependence on M . Similar results demonstrating the tightness of upper-bound for negative-binomial and binomial distributions are presented in Appx. M.4.

7 Discussion and Limitations

In this work, we analytically compute PID for large classes of distributions, greatly expanding upon the analytical result for the Gaussian system. We provide the first known analytical PID for systems employing Poisson, gamma, exponential, Cauchy, beta, Dirichlet, Lévy-stable, binomial, multinomial, negative binomial, and uniform distributions. Furthermore, we generalize the previous Gaussian PID result [21] in an additional way by showing the target M need not be Gaussian. Our stable distribution results provide the first known analytical computation of PID for fat-tailed distributions (all continuous stable distributions have infinite variance except the Gaussian distribution). A practical utility of our analytical results is that they provide a large test bed in which the performance of numerical PID estimators can be compared and evaluated. This test bed may benefit future works on numerical PID estimation by enabling more comprehensive testing of PID estimators.

Our results on the Poisson, Cauchy, and binomial are of particular relevance in the neuroscientific context. Poisson and Cauchy distributions are widely used to model neural spikes [49] and network dynamics in the brain [51]. Binomial thinning is a frequently-used operator in neuroscience [49]. Our generalization of the Gaussian result could be helpful in refining approximations already used in computing PID for neural data [3], e.g., by relaxing the assumption of joint Gaussianity. As continuous stable distributions have been shown to better model Magnetic Resonance Imaging data [52, 53], our results may also be helpful in computing PID in this application.

We also connect the fields of data thinning and data fission with PID by using their decomposition strategies for analytically computing the PID of systems based on convolution-closed distributions (see Sec. 5). Conversely, our PID results also suggest decomposition strategies for data thinning/fission purposes, e.g., our stable distribution results suggest that stable random variables can be decomposed by adding independent noise (similar to the Gaussian case discussed in [26]). Convolution-closed distributions are particularly promising for studying PID as they allow intuitive construction of upper bounds that can be analytically minimized (see Sec. 6). These upper bounds can complement the work on numerical estimation of PID by providing more refined approximations. Another promising avenue is to combine our upper bounds with lower bounds (e.g., from [54]) to create branch and bound algorithms [55] for solving (2). Overall, our analytical results greatly facilitate the computation of PID, either by directly using the analytical expressions or by providing refined approximations for numerical methods.

Limitations and Future Work: We study PID for univariate M , X , and Y and provide some vector extensions. More vector extensions of our results are a promising direction for subsequent works. Most of our analytical results require $P(X|M)$ and $P(Y|M)$ to depend on some affine functions of M . The existence of analytical solutions for the cases where $P(X|M)$ and $P(Y|M)$ depend on non-affine functions of M remains an open question. The upper bound discussed in Sec. 6 can be further refined with more careful analysis, and more rigorous testing of these upper bounds is required to understand in which cases the upper bound is a good approximation for solving (2). We defer the testing and refinement of these upper bounds for subsequent works, as our primary goal in this work is to study analytical solutions of PID. Niu & Quinn [56] also propose a duality result between the synergistic and redundant components in the Gaussian broadcast and multiple-access channels utilizing the analytical PID expressions of the Gaussian system. It may be possible to derive similar duality results for appropriately defined broadcast and multiple-access channels employing other distributions (e.g., Poisson or Cauchy) with our theoretical results.

Broader Impact: Due to their theoretical nature, our results' negative impact primarily depends on how they are used and interpreted. Our theoretical results are applicable only under the specific assumptions outlined in this work, and using these results without ensuring that the theorem's assumptions are satisfied can lead to incorrect scientific conclusions. We also caution reader against naively using PID to draw causal inferences, as PID is inherently a correlational quantity.

Acknowledgments and Disclosure of Funding

Authors would like to thank Prof. Pulkit Grover, Dr. Praveen Venkatesh, Prof. Alireza Chamanzar, and Neil Ashim Mehta for helpful discussions. This work was supported by National Science Foundation (NSF). Authors declare no competing interests.

References

- [1] Praveen Venkatesh and Pulkit Grover. Understanding encoding and redundancy in grid cells using partial information decomposition. *Comput. Syst. Neurosci.(Cosyne)*, 2020.
- [2] Ariel K Feldman, Praveen Venkatesh, Douglas J Weber, and Pulkit Grover. Information-theoretic tools to understand distributed source coding in neuroscience. *IEEE Journal on Selected Areas in Information Theory*, 2024.
- [3] Praveen Venkatesh, Corbett Bennett, Sam Gale, Tamina Ramirez, Gregory Heller, Severine Durand, Shawn Olsen, and Stefan Mihalas. Gaussian partial information decomposition: Bias correction and application to high-dimensional data. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 74602–74635. Curran Associates, Inc., 2023.
- [4] Nigel Colenbier, Frederik Van de Steen, Lucina Q Uddin, Russell A Poldrack, Vince D Calhoun, and Daniele Marinazzo. Disambiguating the role of blood flow and global signal with partial information decomposition. *NeuroImage*, 213:116699, 2020.
- [5] Tjeerd W Boonstra, Luca Faes, Jennifer N Kerkman, and Daniele Marinazzo. Information decomposition of multichannel EMG to map functional interactions in the distributed motor system. *NeuroImage*, 202:116093, 2019.
- [6] Jana Krohova, Luca Faes, Barbora Czipelova, Zuzana Turianikova, Nikoleta Mazgutova, Riccardo Pernice, Alessandro Busacca, Daniele Marinazzo, Sebastiano Stramaglia, and Michal Javorka. Multiscale information decomposition dissects control mechanisms of heart rate variability at rest and during physiological stress. *Entropy*, 21(5):526, 2019.
- [7] Giuseppe Pica, Eugenio Piasini, Houman Safaai, Caroline Runyan, Christopher Harvey, Mathew Diamond, Christoph Kayser, Tommaso Fellin, and Stefano Panzeri. Quantifying how much sensory information in a neural code is relevant for behavior. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [8] Praveen Venkatesh, Sanghamitra Dutta, and Pulkit Grover. Information flow in computational systems. *IEEE Transactions on Information Theory*, 66(9):5456–5491, 2020.
- [9] Paul Pu Liang, Yun Cheng, Xiang Fan, Chun Kai Ling, Suzanne Nie, Richard Chen, Zihao Deng, Nicholas Allen, Randy Auerbach, Faisal Mahmood, Russ R Salakhutdinov, and Louis-Philippe Morency. Quantifying & modeling multimodal interactions: An information decomposition framework. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 27351–27393. Curran Associates, Inc., 2023.
- [10] Paul Pu Liang, Yun Cheng, Ruslan Salakhutdinov, and Louis-Philippe Morency. Multimodal fusion interactions: A study of human and automatic quantification. In *Proceedings of the 25th International Conference on Multimodal Interaction*, pages 425–435, 2023.
- [11] Sanghamitra Dutta, Praveen Venkatesh, Piotr Mardziel, Anupam Datta, and Pulkit Grover. An information-theoretic quantification of discrimination with exempt features. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3825–3833, 2020.
- [12] Sanghamitra Dutta and Faisal Hamman. A review of partial information decomposition in algorithmic fairness and explainability. *Entropy*, 25(5):795, 2023.

- [13] Faisal Hamman and Sanghamitra Dutta. Demystifying local & global fairness trade-offs in federated learning using partial information decomposition. In *The Twelfth International Conference on Learning Representations*, 2024.
- [14] Faisal Hamman and Sanghamitra Dutta. A unified view of group fairness tradeoffs using partial information decomposition. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 214–219, 2024.
- [15] Tomas Scagliarini, Luca Faes, Daniele Marinazzo, Sebastiano Stramaglia, and Rosario N Mantegna. Synergistic information transfer in the global system of financial markets. *Entropy*, 22(9):1000, 2020.
- [16] Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014.
- [17] Abdullah Makkeh, Dirk Oliver Theis, and Raul Vicente. Bivariate partial information decomposition: The optimization perspective. *Entropy*, 19(10):530, 2017.
- [18] Pradeep Kr Banerjee, Johannes Rauh, and Guido Montúfar. Computing the unique information. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 141–145. IEEE, 2018.
- [19] Abdullah Makkeh, Dirk Oliver Theis, and Raul Vicente. BROJA-2PID: A robust estimator for bivariate partial information decomposition. *Entropy*, 20(4):271, 2018.
- [20] Ari Pakman, Amin Nejatbakhsh, Dar Gilboa, Abdullah Makkeh, Luca Mazzucato, Michael Wibral, and Elad Schneidman. Estimating the unique information of continuous variables. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 20295–20307. Curran Associates, Inc., 2021.
- [21] Adam B Barrett. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Physical Review E*, 91(5):052802, 2015.
- [22] Praveen Venkatesh and Gabriel Schamberg. Partial information decomposition via deficiency for multivariate Gaussians. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 2892–2897. IEEE, 2022.
- [23] John P Nolan. Univariate stable distributions. *Springer Series in Operations Research and Financial Engineering*, 10:978–3, 2020.
- [24] Fred W Steutel and Klaas van Harn. Discrete analogues of self-decomposability and stability. *The Annals of Probability*, pages 893–899, 1979.
- [25] Anna Neufeld, Ameer Dharamshi, Lucy L. Gao, and Daniela Witten. Data thinning for convolution-closed distributions. *Journal of Machine Learning Research*, 25(57):1–35, 2024.
- [26] James Leiner, Boyan Duan, Larry Wasserman, and Aaditya Ramdas. Data fission: Splitting a single data point. *Journal of the American Statistical Association*, 0(0):1–12, 2023.
- [27] Paul L Williams and Randall D Beer. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*, 2010.
- [28] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006.
- [29] Malte Harder, Christoph Salge, and Daniel Polani. Bivariate measure of redundant information. *Physical Review E*, 87(1):012130, 2013.
- [30] Pradeep Kr Banerjee, Eckehard Olbrich, Jürgen Jost, and Johannes Rauh. Unique informations and deficiencies. In *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 32–38. IEEE, 2018.
- [31] Jim W Kay and Robin AA Ince. Exact partial information decompositions for Gaussian systems based on dependency constraints. *Entropy*, 20(4):240, 2018.

- [32] Conor Finn and Joseph T Lizier. Pointwise partial information decomposition using the specificity and ambiguity lattices. *Entropy*, 20(4):297, 2018.
- [33] Xueyan Niu and Christopher J Quinn. A measure of synergy, redundancy, and unique information using information geometry. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 3127–3131. IEEE, 2019.
- [34] Fernando E Rosas, Pedro AM Mediano, Borzoo Rassouli, and Adam B Barrett. An operational information decomposition via synergistic disclosure. *Journal of Physics A: Mathematical and Theoretical*, 53(48):485001, 2020.
- [35] Keerthana Gurushankar, Praveen Venkatesh, and Pulkit Grover. Extracting unique information through Markov relations. In *2022 58th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1–6. IEEE, 2022.
- [36] David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, pages 265–272, 1953.
- [37] Praveen Venkatesh, Keerthana Gurushankar, and Gabriel Schamberg. Capturing and interpreting unique information. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 2631–2636. IEEE, 2023.
- [38] Virgil Griffith and Christof Koch. Quantifying synergistic mutual information. *Guided self-organization: inception*, pages 159–190, 2014.
- [39] Joseph T Lizier, Nils Bertschinger, Jürgen Jost, and Michael Wibral. Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work. *Entropy*, 20(4):307, 2018.
- [40] John P Nolan. *Stable distributions*. 2012.
- [41] Jeremie Fish, Jie Sun, and Erik Bollt. Interaction networks from discrete event data by Poisson multivariate mutual information estimation and information flow with applications from gene expression data. *Applied Network Science*, 7(1):1–22, 2022.
- [42] AS Krishnamoorthy. Multivariate binomial and Poisson distributions. *Sankhyā: The Indian Journal of Statistics*, pages 117–124, 1951.
- [43] DM Mahamunulu. A note on regression in the multivariate Poisson distribution. *Journal of the American Statistical Association*, 62(317):251–258, 1967.
- [44] Dimitris Karlis and Loukia Meligkotsidou. Multivariate Poisson regression with covariance structure. *Statistics and Computing*, 15(4):255–265, 2005.
- [45] Murad S Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*. Chapman & Hall, 1994.
- [46] S James Press. Multivariate stable distributions. *Journal of Multivariate Analysis*, 2(4):444–462, 1972.
- [47] Norman Lloyd Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Discrete multivariate distributions*, volume 165. Wiley New York, 1997.
- [48] Xiaohu Shang and H Vincent Poor. Noisy-interference sum-rate capacity for vector Gaussian interference channels. *IEEE Transactions on Information Theory*, 59(1):132–153, 2012.
- [49] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press, 2005.
- [50] Abdullah Makkeh, Dirk Oliver Theis, and Raul Vicente. Broja-2pid: A robust estimator for bertschinger et al.’s bivariate partial information decomposition. *Entropy*, 20(4):271, 2018.
- [51] Viktoras Pyragas and Kestutis Pyragas. Mean-field equations for neural populations with q-Gaussian heterogeneities. *Physical Review E*, 105(4):044402, 2022.

- [52] Diego Salas-González, JM Górriz, Javier Ramírez, Matthias Schloegl, Elmar Wolfgang Lang, and Andrés Ortiz. Parameterization of the distribution of white and grey matter in MRI using the α -stable distribution. *Computers in biology and medicine*, 43(5):559–567, 2013.
- [53] Diego Salas-Gonzalez, Juan M Górriz, Javier Ramírez, and Elmar W Lang. Why using the alpha-stable distribution in NeuroImage? In *2014 International Conference on Signal Processing and Multimedia Applications (SIGMAP)*, pages 297–301. IEEE, 2014.
- [54] Seiya Tokui and Issei Sato. Disentanglement analysis with partial information decomposition. *arXiv preprint arXiv:2108.13753*, 2021.
- [55] Narendra and Fukunaga. A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers*, 100(9):917–922, 1977.
- [56] Xueyan Niu and Christopher J Quinn. Synergy and redundancy duality between Gaussian multiple access and broadcast channels. In *2020 International Symposium on Information Theory and Its Applications (ISITA)*, pages 6–10. IEEE, 2020.
- [57] Robert B Ash and Catherine A Doléans-Dade. *Probability and measure theory*. Academic press, 2000.
- [58] Gilbert Strang. *Introduction to linear algebra*. SIAM, 2022.
- [59] Alan V Oppenheim, Alan S Willsky, Syed Hamid Nawab, and Jian-Jiun Ding. *Signals and systems*, volume 2. Prentice hall Upper Saddle River, NJ, 1997.
- [60] Jagadeesh Chandramohan and Lung-Kuang Liang. Bernoulli, multinomial and Markov chain thinning of some point processes and some results about the superposition of dependent renewal processes. *Journal of Applied Probability*, 22(4):828–835, 1985.
- [61] Sergio Verdú. Poisson communication theory. *International Technion Communication Day in Honor of Israel Bar-David*, 66, 1999.
- [62] Christian Walck et al. *Hand-book on statistical distributions for experimentalists*. Stockholms universitet, 1996.
- [63] Kalimuthu Krishnamoorthy. *Handbook of statistical distributions with applications*. Chapman and Hall/CRC, 2006.
- [64] Lenka Slámová and Lev B Klebanov. Generalized definitions of discrete stability. *arXiv preprint arXiv:1502.02588*, 2015.

A Formal proof of the sufficient condition for solving (2) discussed in Sec. 3

Proposition 1. *For the minimization problem defined in (2), if there exists a joint density $Q_{MC}(M, X, Y) \in \Delta_P$, having the Markovian structure of the form $M \rightarrow X \rightarrow Y$ or $M \rightarrow Y \rightarrow X$, then $Q_{MC}(M, X, Y)$ minimizes (2). Furthermore,*

1. *If $Q_{MC}(M, X, Y)$ has the Markovian structure $M \rightarrow X \rightarrow Y$, then $UI(M; Y \setminus X) = I_{Q_{MC}}(M; Y|X) = 0$.*
2. *If $Q_{MC}(M, X, Y)$ has the Markovian structure $M \rightarrow Y \rightarrow X$, then $UI(M; X \setminus Y) = I_{Q_{MC}}(M, X|Y) = 0$.*

Proof. Case 1: $Q_{MC}(M, X, Y) \in \Delta_P$ having the Markovian structure $M \rightarrow X \rightarrow Y$.

By the definition of the Markov chain, we know that M and Y are conditionally independent given X . Denote $Q_{MC}(M, X, Y)$ as the joint distribution of the Markov chain $M \rightarrow X \rightarrow Y$. Then,

$$I_{Q_{MC}}(M; Y|X) = 0, \quad (13)$$

where $I_{Q_{MC}}(M; Y|X)$ is the conditional mutual information between M and Y given X under the distribution $Q_{MC}(M, X, Y)$. Equation (13) follows from a well-known property of conditional mutual information which states that $I(A; B|C) = 0$ iff $A \perp\!\!\!\perp B|C$ [28].

Since $I_Q(M; Y|X) \geq 0 \forall Q \in \Delta_P$, due to the non-negativity of conditional mutual information [28], we can conclude that $\min_{Q \in \Delta_P} I_Q(M; Y|X) \geq 0$. The joint distribution $Q_{MC}(M, X, Y)$ having the Markovian structures $M \rightarrow X \rightarrow Y$ achieves the lower bound of 0 for the minimization problem $\min_{Q \in \Delta_P} I_Q(M; Y|X)$ implying $Q_{MC}(M, X, Y)$ minimizes (2).

Case 2: $Q_{MC}(M, X, Y) \in \Delta_P$ having the Markovian structure $M \rightarrow Y \rightarrow X$.

A similar argument as case 1, with the random variables X and Y switched, would allow us to conclude that the joint distribution $Q_{MC}(M, X, Y)$ specifying the Markov chain $M \rightarrow Y \rightarrow X$ minimizes (2). \square

B Other PID definitions for which theorems 1-7 are valid

Our theoretical results (Theorems 1-7) hold for a wide range of PID measures. Particularly, our results hold for two broad classes of PID definitions: PID definitions satisfying assumption (*) in Bertschinger et al. [16] and Blackwellian PIDs (see Appendix B-C of Venkatesh & Schamberg [22] for a deeper discussion on differences between these two families of PID definition). Some examples of PID definitions satisfying assumption (*) in Bertschinger et al. [16] are the Williams & Beer's PID definition [27], the PID definition proposed in Harder et al. [29], and I-PID proposed in Venkatesh et al. [37]. Similarly, an example of a Blackwellian PID is the δ -PID discussed in Venkatesh & Schamberg [22]. Section III-F of Venkatesh et al. [37] further discusses a family of Blackwellian PID definitions for which our theorems are also applicable. We now provide formal arguments showing that our theoretical results are applicable for PID definitions that either satisfy assumption (*) of Bertschinger et al. [16] or are Blackwellian

PID definitions satisfying assumption (*): We invoke the argument presented in the proof discussed in Section 4.2 of Barrett [21] to show that our results hold for any PID definition satisfying assumption (*) in Bertschinger et al. [16]. Barrett's argument relies on the key observation that the UI calculated using BROJA-PID upper bounds the UI calculated using any other PID definition satisfying assumption (*) in Bertschinger et al. Entropy'14. Consequently, if one of the UI atoms of the BROJA-PID is zero then the corresponding UI atom of the other PID definitions must also be zero due to BROJA PID upper-bounding them and the non-negativity of the PID atoms. Since the remaining PID atoms are calculated using (1), it must be the case that the PID atoms calculated using any PID definition satisfying assumption (*) must be the same as the PID atoms calculated using BROJA-PID whenever one of the UI terms of BROJA-PID is zero.

Blackwellian PID: The defining feature of a Blackwellian PID is that the UI atom is zero iff it is possible to construct the Markovian structure between the random variables (i.e., $M \rightarrow X \rightarrow Y$ or $M \rightarrow Y \rightarrow X$), while preserving their pairwise marginals (see Venkatesh et al. [22] for more details). In all our theoretical results, we show that the random variables in the analyzed system admit

a Markovian structure, while preserving the pairwise marginal structure. Hence, the UI atom for the systems analyzed in Theorems 1-7 is zero for any Blackwellian PID. Consequently, as the rest of the PID terms are calculated using (1), it must be the case that the PID atoms calculated using any Blackwellian PID must be the same as the PID atoms calculated using BROJA-PID in our results.

C Lists of Systems that can solved using Theorem 6

In this section, we provide examples of systems of random variables (M, X, Y) employing well-known distributions whose PID terms can be obtained through theorem 6. The following list is modified from Table 2 of [25]. Note that for the joint distribution $P(M, X, Y)$ of the system of random variable (M, X, Y) , we only specify $P(M)$, $P(X|M)$, and $P(Y|M)$, as any $P(M, X, Y)$ having these marginals can be solved through theorem 6. We also discuss some possible extensions of the corresponding systems for which the upper bound discussed in Sec. 6 is applicable. Note that the extension of the systems provided are not exhaustive, and there exists more systems for which the upper bound is applicable. In the following examples, we assume $a > b$.

1. **System:** The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support $(0, \infty)$.
- (b) $P(X|M) = \text{Poisson}(aM)$.
- (c) $P(Y|M) = \text{Poisson}(bM)$.

The corresponding $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, x)$ specified in theorem 6 is $\text{Binomial}(x, \gamma)$ with $\gamma = b/a$, and $a, b > 0$.

Result: The Markov chain $M \rightarrow X \rightarrow Y \in \Delta_P$ and $UI(M; Y \setminus X) = 0$.

An extension of the system for which the upper bound of Sec. 6 in applicable: The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support \mathcal{M} .
- (b) $P(X|M) = \text{Poisson}(f_1(M))$, for $f_1(M) > 0 \forall m \in \mathcal{M}$.
- (c) $P(Y|M) = \text{Poisson}(f_2(M))$, for $f_2(M) > 0 \forall m \in \mathcal{M}$.
- (d) $f_1(m) \geq f_2(m) \forall m \in \mathcal{M}$ and all the underlying distributions are well-defined.

2. **System:** The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support $(-\infty, \infty)$.
- (b) $P(X|M) = \mathcal{N}(aM, a\sigma^2)$.
- (c) $P(Y|M) = \mathcal{N}(bM, b\sigma^2)$.

The corresponding $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, x)$ specified in theorem 6 is $\mathcal{N}(\gamma x, \gamma(1-\gamma)\sigma^2)$ with $\gamma = b/a$, and $a, b > 0$.

Result: The Markov chain $M \rightarrow X \rightarrow Y \in \Delta_P$ and $UI(M; Y \setminus X) = 0$.

An extension of the system for which the upper bound of Sec. 6 in applicable: The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support \mathcal{M} .
- (b) $P(X|M) = \text{Gaussian}(f_1(M), f_1(M))$, for $f_1(M) > 0 \forall m \in \mathcal{M}$.
- (c) $P(Y|M) = \text{Gaussian}(f_2(M), f_2(M))$, for $f_2(M) > 0 \forall m \in \mathcal{M}$.
- (d) $f_1(m) \geq f_2(m) \forall m \in \mathcal{M}$ and all the underlying distributions are well-defined.

3. **System:** The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support $(0, 1]$.
- (b) $P(X|M) = \text{Negative Binomial}(a, M)$.
- (c) $P(Y|M) = \text{Negative Binomial}(b, M)$.

The corresponding $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, x)$ specified in theorem 6 is $\text{Beta-Binomial}(x, \gamma a, (1-\gamma)a)$ with $\gamma = b/a$, and $a, b \in \mathbb{N}$.

Result: The Markov chain $M \rightarrow X \rightarrow Y \in \Delta_P$, and $UI(M; Y \setminus X) = 0$.

An extension of the system for which the upper bound of Sec. 6 in applicable: The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support \mathcal{M} .

- (b) $P(X|M) = \text{Negative Binomial}(f_1(M), p)$, for $f_1(M) \in \mathbb{N}_0 \forall m \in \mathcal{M}$ and $p \in [0, 1]$.
- (c) $P(Y|M) = \text{Negative Binomial}(f_2(M), p)$, for $f_2(M) \in \mathbb{N}_0 \forall m \in \mathcal{M}$.
- (d) $f_1(m) \geq f_2(m) \forall m \in \mathcal{M}$ and all the underlying distributions are well-defined.

4. **System:** The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support $(0, \infty)$.
- (b) $P(X|M) = \text{Gamma}(a, M)$.
- (c) $P(Y|M) = \text{Gamma}(b, M)$.

The corresponding $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, x)$ specified in theorem 6 is obtained as follows: $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, x) = xZ$, $Z \sim \text{Beta}(b, (1-b))$ with $\gamma = b/a$, and $a, b \in (0, \infty)$. Note that we are following the $\text{Gamma}(\alpha, \beta)$ notation.

Result: The Markov chain $M \rightarrow X \rightarrow Y \in \Delta_P$ and $UI(M; Y \setminus X) = 0$.

An extension of the system for which the upper bound of Sec. 6 is applicable: The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support \mathcal{M} .
- (b) $P(X|M) = \text{Gamma}(f_1(M), \beta)$, for $f_1(M) > 0 \forall m \in \mathcal{M}$ and $\beta > 0$.
- (c) $P(Y|M) = \text{Gamma}(f_2(M), \beta)$, for $f_2(M) > 0 \forall m \in \mathcal{M}$.
- (d) $f_1(m) \geq f_2(m) \forall m \in \mathcal{M}$ and all the underlying distributions are well-defined.

5. **System:** The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support $(0, \infty)$.
- (b) $P(X|M) = \text{Exponential}(M)$.
- (c) $P(Y|M) = \text{Gamma}(b, M)$.

The corresponding $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, x)$ specified in theorem 6 is obtained as follows: $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, x) = xZ$, $Z \sim \text{Beta}(b, (1-b))$, and $0 < b \leq 1$. Note that we are following the $\text{Gamma}(\alpha, \beta)$ notation.

Result: The Markov chain $M \rightarrow X \rightarrow Y \in \Delta_P$ and $UI(M; Y \setminus X) = 0$.

An extension of the system for which the upper bound of Sec. 6 is applicable: The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support \mathcal{M} .
- (b) $P(X|M) = \text{Exponential}(f_1(M))$, for $f_1(M) > 0 \forall m \in \mathcal{M}$.
- (c) $P(Y|M) = \text{Gamma}(f_2(M), f_1(M))$, for $f_2(M) > 0 \forall m \in \mathcal{M}$.
- (d) $f_2(m) \leq 1 \forall m \in \mathcal{M}$ and all the underlying distributions are well-defined.

6. **System:** The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support $[0, 1]$.
- (b) $P(X|M) = \text{Binomial}(a, M)$.
- (c) $P(Y|M) = \text{Binomial}(b, M)$.

The corresponding $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, x)$ specified in theorem 6 is Hypergeometric($x, \gamma a, (1-\gamma)a$) with $\gamma = b/a$, and $a, b \in \mathbb{N}$.

Result: The Markov chain $M \rightarrow X \rightarrow Y \in \Delta_P$ and $UI(M; Y \setminus X) = 0$.

An extension of the system for which the upper bound of Sec. 6 is applicable: The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support \mathcal{M} .
- (b) $P(X|M) = \text{Binomial}(f_1(M), p)$, for $f_1(M) \in \mathbb{N}_0 \forall m \in \mathcal{M}$ and $p \in [0, 1]$.
- (c) $P(Y|M) = \text{Binomial}(f_2(M), p)$, for $f_2(M) \in \mathbb{N}_0 \forall m \in \mathcal{M}$.
- (d) $f_1(m) \geq f_2(m) \forall m \in \mathcal{M}$ and all the underlying distributions are well-defined.

7. **System:** The distributions of $P(\vec{M})$, $P(\vec{X}|\vec{M})$, and $P(\vec{Y}|\vec{M})$ are as follows:

- (a) $\vec{M} \sim P(\vec{M})$ having support over a k -dimensional simplex.
- (b) $P(\vec{X}|\vec{M}) = \text{Multinomial}_k(a, \vec{M})$.
- (c) $P(\vec{Y}|\vec{M}) = \text{Multinomial}_k(b, \vec{M})$.

The corresponding $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, x)$ specified in theorem 6 is Multivariate Hypergeometric (\vec{x}, γ_Y) with $\gamma = b/a$, and $a, b \in \mathbb{N}$.

Result: The Markov chain $\vec{M} \rightarrow \vec{X} \rightarrow \vec{Y} \in \Delta_P$ and $UI(\vec{M}; \vec{Y} \setminus \vec{X}) = 0$.

An extension of the system for which the upper bound of Sec. 6 is applicable: The distributions of $P(M)$, $P(\vec{X}|M)$, and $P(\vec{Y}|M)$ are as follows:

- (a) $M \sim P(M)$ having support \mathcal{M} .
- (b) $P(\vec{X}|M) = \text{Multinomial}_k(f_1(M), \vec{p})$, for $f_1(M) \in \mathbb{N}_0 \forall m \in \mathcal{M}$.
- (c) $P(\vec{Y}|M) = \text{Multinomial}(f_2(M), \vec{p})$, for $f_2(M) \in \mathbb{N}_0 \forall m \in \mathcal{M}$.
- (d) $f_1(m) \geq f_2(m) \forall m \in \mathcal{M}$, \vec{p} is a valid probability vector, and all the underlying distributions are well-defined.

D Lists of Systems that can solved using Theorem 7

In this section, we provide examples of systems of random variables (M, X, Y) employing well-known distributions whose PID terms can be obtained through theorem 7. The following list is appropriately modified from Appendix A of [26]. Similar to Appx. C, we only specify $P(M)$, $P(X|M)$, and $P(Y|M)$ of the joint distribution $P(M, X, Y)$, as any $P(M, X, Y)$ having the same marginals can be solved through theorem 7. In the following examples, we assume $a > b$.

1. **System:** The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support $(0, \infty)$.
- (b) $P(X|M) = \text{Exponential}(M)$.
- (c) $P(Y|M) = \text{Geometric}(M/(\tau+M))$, where $\tau \in (0, \infty)$.

Results: The Markov chain $M \rightarrow X \rightarrow Y \in \Delta_P$, and $UI(M; Y \setminus X) = 0$. The corresponding $Q_{MC}(Y|X)$ required for constructing the Markov chain is $\text{Poisson}(\tau X)$.

2. **System:** The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support \mathcal{M} .
- (b) $P(X|M) = \text{Gamma}(f_1(M), f_2(M))$, where $f_1(m), f_2(m) > 0 \forall m \in \mathcal{M}$.
- (c) $P(Y|M) = \text{Negative Binomial}(f_1(M), f_2(M)/(f_2(M)+\tau))$, where $\tau \in (0, \infty)$

Results: The Markov chain $M \rightarrow X \rightarrow Y \in \Delta_P$, and $UI(M; Y \setminus X) = 0$. The corresponding $Q_{MC}(Y|X)$ required for constructing the Markov chain is $\text{Poisson}(\tau X)$.

3. **System:** The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support $(0, \infty)$.
- (b) $P(X|M) = \text{Beta}(M, 1)$.
- (c) $P(Y|M)$ is a discrete distribution with over support set $\{0, 1, \dots, N\}$, with

$$P(Y = y|M = m) = m \frac{\Gamma(y+m)N!}{\Gamma(N+1+m)y!},$$

where $\Gamma(\cdot)$ is the Gamma function.

Results: The Markov chain $M \rightarrow X \rightarrow Y \in \Delta_P$, and $UI(M; Y \setminus X) = 0$. The corresponding $Q_{MC}(Y|X)$ required for constructing the Markov chain is $\text{Binomial}(N, X)$.

4. **System:** The distributions of $P(M)$, $P(X|M)$, and $P(Y|M)$ are as follows:

- (a) $M \sim P(M)$ having support $(0, \infty)$.
- (b) $P(X|M) = \text{Beta}(1, M)$.
- (c) $P(Y|M)$ is a discrete distribution with over support set $\{0, 1, \dots, N\}$, with

$$P(Y = y|M = m) = m \frac{\Gamma(y+m)N!}{\Gamma(N+1+m)y!},$$

where $\Gamma(\cdot)$ is the Gamma function.

Results: The Markov chain $M \rightarrow X \rightarrow Y \in \Delta_P$, and $UI(M; Y \setminus X) = 0$. The corresponding $Q_{MC}(Y|X)$ required for constructing the Markov chain is $\text{Binomial}(N, 1-X)$.

5. **System:** The distributions of $P(\vec{M})$, $P(\vec{X}|\vec{M})$, and $P(\vec{Y}|\vec{M})$ are as follows:

- (a) $\vec{M} \sim P(\vec{M})$ having support $(0, \infty)^d$ for some dimension d .
- (b) $P(\vec{X}|\vec{M}) = \text{Dirichlet}(M_1, \dots, M_d)$.
- (c) $P(\vec{Y}|\vec{M})$ is a discrete distribution having support $\{y_i \in \{1, \dots, N\} \mid \sum_{i=1}^d y_i = N\}$, and

$$P(\vec{Y}|\vec{M}) = \frac{\Gamma\left(\sum_{i=1}^d M_i\right) N!}{\Gamma\left(\sum_{i=1}^d M_i + N\right)} \prod_{i=1}^d \frac{\Gamma(M_i + Y_i)}{\Gamma(M_i) Y_i!},$$

where $\Gamma(\cdot)$ is the Gamma function.

Results: The Markov chain $\vec{M} \rightarrow \vec{X} \rightarrow \vec{Y} \in \Delta_P$, and $UI(\vec{M}; \vec{Y} \setminus \vec{X}) = 0$. The corresponding $Q_{MC}(\vec{Y}|\vec{X})$ required for constructing the Markov chain is Multinomial $_d(N, \vec{X})$.

E Miscellaneous PID Results

This section explores the analytical calculation of PID terms for five more systems of random variables (M, X, Y) with joint distribution $P(M, X, Y)$. These five systems' $P(M, X, Y)$ employ well-known distributions, but their respective analytical PID calculation does not result from Theorem 1- Theorem 7. Of note are the results of $P(M, X, Y)$ employing exponential and uniform distribution, which states that these systems will always have both unique information terms zero.

E.1 PID for the miscellaneous system M1.

Miscellaneous system M1: Let (M, X, Y) be a system of random variables with joint distribution $P(M, X, Y)$. The joint distribution $P(M, X, Y)$ describes the miscellaneous system **M1** if it satisfies the following two properties:

- 1. $M \sim P(M)$, and has support $(0, \infty)$.
- 2. $P(X|M) = \text{Exponential}(\gamma_X M)$, and $P(Y|M) = \text{Exponential}(\gamma_Y M)$, where $\gamma_X, \gamma_Y \in (0, \infty)$.

Proposition 2. Let M, X , and Y be random variables whose joint distribution $P(M, X, Y)$ describes the miscellaneous system **M1** defined above. Then, Δ_P contains a distribution $Q_{MC}(M, X, Y)$, which admits both Markovian structures, i.e., $M \rightarrow X \rightarrow Y$ and $M \rightarrow Y \rightarrow X$. Consequently, both $UI(M; Y \setminus X) = UI(M; X \setminus Y) = 0$.

Proof. We prove proposition 2 by explicitly constructing a joint distribution $Q_{MC}(M, X, Y) \in \Delta_P$ that admits both Markovian structures, namely, $M \rightarrow X \rightarrow Y$ and $M \rightarrow Y \rightarrow X$. A necessary and sufficient condition for the distribution $Q_{MC}(M, X, Y)$ to admit both Markovian structures is that it should satisfy (14).

$$Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(X|M)Q_{MC}(Y|X) = Q_{MC}(M)Q_{MC}(Y|M)Q_{MC}(X|Y). \quad (14)$$

We now briefly outline the structure of the proof. The rest of the proof can be divided into three parts:

- 1. In the first part, we explicitly construct $Q_{MC}(M, X, Y)$.
- 2. In the second part, we show that the $Q_{MC}(M, X, Y)$ constructed in the first part lies in Δ_P .
- 3. In the third part, we conclude our proof by showing that $Q_{MC}(M, X, Y)$ satisfies (14) and, consequently, admits both Markovian structures.

Part 1: Specifying $Q_{MC}(M, X, Y)$

We specify $Q_{MC}(M, X, Y)$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(X|M)$, and $Q_{MC}(Y|X)$. We choose $Q_{MC}(M)$ and $Q_{MC}(X|M)$ as follows:

$$Q_{MC}(M) = P(M), \text{ and } Q_{MC}(X|M) = P(X|M). \quad (15)$$

The distribution $Q_{MC}(Y|X)$ is specified through the following deterministic transformation:

$$Y = \frac{\gamma_X}{\gamma_Y} X. \quad (16)$$

Part 2: Showing $Q_{MC}(M, X, Y) \in \Delta_P$

For showing $Q_{MC}(M, X, Y) \in \Delta_P$, we first need to calculate $Q_{MC}(Y|M)$:

Calculating $Q_{MC}(Y|M)$: By lemma 1, if $X \sim \text{Exp}(\lambda)$, then $aX \sim \text{Exponential}(\lambda/a)$. Since $Q_{MC}(X|M) = \text{Exponential}(\gamma_X M)$, we can use the result of lemma 1 to calculate $Q_{MC}(Y|M)$ as follows:

$$\begin{aligned} Q_{MC}(Y|M) &= Q_{MC}\left(\frac{\gamma_X}{\gamma_Y} X \middle| M\right) = \text{Exponential}\left(\frac{\gamma_Y}{\gamma_X} \gamma_X M\right) \\ &= \text{Exponential}(\gamma_Y M) = P(Y|M). \end{aligned} \quad (17)$$

Then, from (15) and (17), we can conclude:

$$Q_{MC}(M, X) = Q_{MC}(M)Q_{MC}(X|M) = P(M)P(X|M) = P(M, X), \quad (18)$$

$$Q_{MC}(M, Y) = Q_{MC}(M)Q_{MC}(Y|M) = P(M)P(Y|M) = P(M, Y). \quad (19)$$

Hence, $Q_{MC}(M, X, Y) \in \Delta_P$.

Part 3: Showing $Q_{MC}(M, X, Y)$ satisfies (14)

It is trivial to see that by the construction of $Q_{MC}(M, X, Y)$ given in Part 1, we have:

$$Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(X|M)Q_{MC}(Y|X). \quad (20)$$

By the chain rule, we have:

$$Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(Y|M)Q_{MC}(X|Y, M). \quad (21)$$

Hence, for showing that $Q_{MC}(M, X, Y)$ satisfies (14), we only need to show that $Q_{MC}(X|Y, M) = Q_{MC}(X|Y)$.

Showing $Q_{MC}(X|Y, M) = Q_{MC}(X|Y)$: Observe that $Y = \gamma_X/\gamma_Y X$ by construction. Therefore, we can equivalently express $X = \gamma_Y/\gamma_X Y$. Consequently,

$$Q_{MC}(X|Y, M) = \mathbb{I}[X = \gamma_Y/\gamma_X Y] \text{ as } X = \gamma_Y/\gamma_X Y, \quad (22)$$

where $\mathbb{I}[\cdot]$ is the indicator function. Since R.H.S of the above equation does not depend on M , we have:

$$Q_{MC}(X|Y, M) = \mathbb{I}[X = \gamma_Y/\gamma_X Y] = Q_{MC}(X|Y). \quad (23)$$

Combining (21) and (23), we have:

$$Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(Y|M)Q_{MC}(X|Y). \quad (24)$$

Combining (20) and (24), we can show that $Q_{MC}(M, X, Y)$ satisfies (14).

Conclusion: Combining the results of parts 2 and 3, we can conclude that the $Q_{MC}(M, X, Y) \in \Delta_P$ and contains both Markovian structure $M \rightarrow X \rightarrow Y$ and $M \rightarrow Y \rightarrow X$. A unique implication of $Q_{MC}(M, X, Y)$ admitting both Markovian structures is that both $I_{Q_{MC}}(M; X|Y) = I_{Q_{MC}}(M; Y|X) = 0$. Consequently, both minimization problems presented in (2) are minimized by $Q_{MC}(M, X, Y)$ with $UI(M; X \setminus Y) = \min_{Q \in \Delta_P} I_Q(M; X|Y) = 0$ and $UI(M; Y \setminus X) = \min_{Q \in \Delta_P} I_Q(M; Y|X) = 0$, completing our proof. \square

Lemma 1. If $X \sim \text{Exponential}(\lambda)$ with $\lambda \in (0, \infty)$, then $\gamma X \sim \text{Exponential}(\lambda/\gamma)$ for $\gamma \in (0, \infty)$.

Proof. We denote the p.d.f. of the random variable X as $p_X(\cdot)$ and, by lemma statement, we know:

$$p_X(X = x) = \text{Exponential}(\lambda) = \begin{cases} \lambda e^{-x\lambda} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}. \quad (25)$$

Denote the p.d.f. of the function $f(X)$ of the random variable X as $p_{f(X)}(\cdot)$. We want to derive $p_{f(X)}(\cdot)$ for $f(X) = \gamma X$. Since, $f(X) = \gamma X$ is an invertible function, we can use the change-of-variable technique for deriving $p_{f(X)}(\cdot)$:

$$p_{f(X)}(f(X) = y) = p_X(X = f^{-1}(y)) \left| \frac{df^{-1}(y)}{dy} \right| = \begin{cases} \frac{\lambda}{\gamma} e^{-y\lambda/\gamma} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$= \text{Exponential}(\lambda/\gamma). \quad (26)$$

□

E.2 PID for the miscellaneous system M2.

Miscellaneous System M2: Let (M, X, Y) be a system of random variables with joint distribution $P(M, X, Y)$. The joint distribution $P(M, X, Y)$ describes the miscellaneous system **M2** if it satisfies the following two properties:

1. $M \sim P(M)$ and has support $(0, \infty)$.
2. $P(X|M) = \text{Negative Binomial}(M, p_X)$, and $P(Y|M) = \text{Negative Binomial}(M, p_Y)$.

This system is adapted from the negative binomial decomposition listed in Appendix A [26].

Proposition 3. Let M, X , and Y be random variables whose joint distribution $P(M, X, Y)$ describes the miscellaneous system **M2** defined above.

1. If $p_X \leq p_Y$, then Δ_P contains a Markov chain of the form $M \rightarrow X \rightarrow Y$, and $UI(M; Y|X) = 0$.
2. If $p_X \geq p_Y$, then Δ_P contains a Markov chain of the form $M \rightarrow Y \rightarrow X$, and $UI(M; X|Y) = 0$.

Proof. We only provide an explicit proof of condition 1. The proof of condition 2 follows the same steps as the proof of condition 1 with parameters of X and Y switched.

Proof of condition 1: We briefly outline the proof structure. The proof can be divided into two parts:

1. In the first part, we explicitly construct a joint distribution $Q_{MC}(M, X, Y)$ having the Markovian structure $M \rightarrow X \rightarrow Y$.
2. In the second part, we conclude our proof by showing the $Q_{MC}(M, X, Y)$ constructed in the first part, lies in Δ_P .

Part 1: Specifying $Q_{MC}(M, X, Y)$

Since $Q_{MC}(M, X, Y)$ is the joint distribution of a Markov chain $M \rightarrow X \rightarrow Y$, we have $Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(X|M)Q_{MC}(Y|X)$. Taking advantage of this special structure of $Q_{MC}(M, X, Y)$, we specify $Q_{MC}(M, X, Y)$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(X|M)$, and $Q_{MC}(Y|X)$. We choose $Q_{MC}(M)$, $Q_{MC}(X|M)$ and $Q_{MC}(Y|X)$ as follows:

$$Q_{MC}(M) = P(M), Q_{MC}(X|M) = P(X|M) \text{ and } Q_{MC}(Y|X) = \text{Binomial}(X, p),$$

$$\text{where } p = \frac{p_X(1 - p_Y)}{p_Y(1 - p_X)}. \quad (27)$$

Part 2: Showing $Q_{MC}(M, X, Y) \in \Delta_P$

For showing $Q_{MC}(M, X, Y) \in \Delta_P$, we first need to calculate $Q_{MC}(Y|M)$:

Calculating $Q_{MC}(Y|M)$: Since $Q_{MC}(X|M) = \text{Negative Binomial}(M, p_X)$ and $Q_{MC}(Y|X) = \text{Binomial}(X, p)$, we can use lemma 2 for calculating $Q(Y|M)$:

$$Q_{MC}(Y|M) = \text{Negative Binomial}(M, p_Y) = P(Y|M). \quad (28)$$

Then, from (27) and (28), we can conclude:

$$Q_{MC}(M, X) = Q_{MC}(M)Q_{MC}(X|M) = P(M)P(X|M) = P(M, X), \quad (29)$$

$$Q_{MC}(M, Y) = Q_{MC}(M)Q_{MC}(Y|M) = P(M)P(Y|M) = P(M, Y). \quad (30)$$

Hence, $Q_{MC}(M, X, Y) \in \Delta_P$ and consequently, by proposition 1, $UI(M; Y \setminus X) = 0$. This concludes our proof. \square

Lemma 2. Let $X \sim \text{Negative Binomial}(N, p_X)$ and $P(Y|X) = \text{Binomial}(X, p)$ with $p = \frac{p_X(1-p_Y)}{p_Y(1-p_X)}$, $p_Y \geq p_X$, $p_Y, p_X \in [0, 1]$, and $N \in \mathbb{N}_0$. Then, $Y \sim \text{Negative Binomial}(N, p_Y)$.

Proof. First, in order for $P(Y|X)$ to be a legitimate distribution, we need to ensure that $p \in [0, 1]$, i.e., we need to show that:

$$0 \leq p \leq 1. \quad (31)$$

Since $p_X, (1 - p_Y), (1 - p_X), p_Y \geq 0$ as $p_X, p_Y \in [0, 1]$, we can easily verify the first inequality:

$$p = \frac{p_X(1 - p_Y)}{(1 - p_X)p_Y} \geq 0. \quad (32)$$

For verifying $p \leq 1$, we will use the assumption that $p_Y \geq p_X$:

$$p_X \leq p_Y \Rightarrow \frac{1}{p_Y} \leq \frac{1}{p_X}. \quad (33)$$

Subtracting 1 from both sides of the above inequality.

$$\frac{1}{p_Y} - 1 \leq \frac{1}{p_X} - 1 \Rightarrow \frac{1 - p_Y}{p_Y} \leq \frac{1 - p_X}{p_X}. \quad (34)$$

Rearranging terms, we obtain:

$$\frac{p_X(1 - p_Y)}{(1 - p_X)p_Y} \leq 1 \Rightarrow p \leq 1 \left(\text{as } p = \frac{p_X(1 - p_Y)}{(1 - p_X)p_Y} \right). \quad (35)$$

For calculating $P(Y)$, we borrow the result of proposition 13 in [26], which states that if $X \sim \text{Negative Binomial}(M, p_X)$ and $P(Y|X) = \text{Binomial}(X, p)$, then $P(Y) = \text{Negative Binomial}\left(N, \frac{p_X}{p_X + p - pp_X}\right)$. Hence,

$$\begin{aligned} P(Y) &= \text{Negative Binomial}\left(N, \frac{p_X}{p_X + p - pp_X}\right) = \text{Negative Binomial}\left(N, \frac{p_X}{p_X + p(1 - p_X)}\right) \\ &= \text{Negative Binomial}\left(N, \frac{p_X}{p_X + \frac{p_X(1-p_Y)}{(1-p_X)p_Y}(1-p_X)}\right) \\ &= \text{Negative Binomial}\left(N, \frac{p_X}{p_X + \frac{p_X(1-p_Y)}{p_Y}}\right) \\ &= \text{Negative Binomial}\left(N, \frac{p_X p_Y}{p_X p_Y + p_X(1 - p_Y)}\right) \\ &= \text{Negative Binomial}\left(N, \frac{p_X p_Y}{p_X}\right) = \text{Negative Binomial}(N, p_Y). \end{aligned} \quad (36)$$

\square

E.3 PID for the miscellaneous system M3.

Miscellaneous System M3: Let (M, X, Y) be a system of random variables with joint distribution $P(M, X, Y)$. The joint distribution $P(M, X, Y)$ describes the miscellaneous system **M3** if it satisfies the following two properties:

1. $M \sim P(M)$ and has support $(0, \infty)$.
2. $P(X|M) = \text{Uniform}(0, \gamma_X M)$ and $P(Y|M) = \text{Uniform}(0, \gamma_Y M)$, where $\gamma_X, \gamma_Y \in \mathbb{R} \setminus \{0\}$.

Proposition 4. Let M, X , and Y be random variables whose joint distribution $P(M, X, Y)$ describes the miscellaneous system **M3** defined above. Then, Δ_P contains a distribution $Q_{MC}(M, X, Y)$, which admits both Markovian structures, i.e., $M \rightarrow X \rightarrow Y$ and $M \rightarrow Y \rightarrow X$. Consequently, both $UI(M; Y \setminus X) = UI(M; X \setminus Y) = 0$.

Proof. We prove proposition 4 by explicitly constructing a joint distribution $Q_{MC}(M, X, Y) \in \Delta_P$ that admits both Markovian structures, namely, $M \rightarrow X \rightarrow Y$ and $M \rightarrow Y \rightarrow X$. A necessary and sufficient condition for the distribution $Q_{MC}(M, X, Y)$ to admit both Markovian structures is that it should satisfy (37):

$$Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(X|M)Q_{MC}(Y|X) = Q_{MC}(M)Q_{MC}(Y|M)Q_{MC}(X|Y). \quad (37)$$

We now briefly outline the structure of the proof. The rest of the proof can be divided into three parts:

1. In the first part, we explicitly construct $Q_{MC}(M, X, Y)$.
2. In the second part, we show that the $Q_{MC}(M, X, Y)$ constructed in the first part, lies in Δ_P .
3. In the third part, we conclude our proof by showing that $Q_{MC}(M, X, Y)$ satisfies (37) and, consequently, admits both Markovian structures.

Part 1: Specifying $Q_{MC}(M, X, Y)$

We specify $Q_{MC}(M, X, Y)$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(X|M)$, and $Q_{MC}(Y|X)$. We choose $Q_{MC}(M)$ and $Q_{MC}(X|M)$ as follows:

$$Q_{MC}(M) = P(M), \text{ and } Q_{MC}(X|M) = P(X|M). \quad (38)$$

The distribution $Q_{MC}(Y|X)$ is specified through the following deterministic transformation:

$$Y = \frac{\gamma_Y}{\gamma_X} X. \quad (39)$$

Part 2: Showing $Q_{MC}(M, X, Y) \in \Delta_P$

For showing $Q_{MC}(M, X, Y) \in \Delta_P$, we first need to calculate $Q_{MC}(Y|M)$:

Calculating $Q_{MC}(Y|M)$: Since $Q_{MC}(X|M) = \text{Uniform}(0, \gamma_X M)$, and $Y = \gamma_Y/\gamma_X X$, we can use the result of lemma 3 to calculate $Q_{MC}(Y|M)$ as follows:

$$\begin{aligned} Q_{MC}(Y|M) &= Q_{MC}\left(\frac{\gamma_Y}{\gamma_X} X \middle| M\right) = \text{Uniform}\left(0, \frac{\gamma_Y}{\gamma_X} \gamma_X M\right) = \text{Uniform}(0, \gamma_Y M) \\ &= P(Y|M). \end{aligned} \quad (40)$$

Then, from (38) and (40), we can conclude:

$$Q_{MC}(M, X) = Q_{MC}(M)Q_{MC}(X|M) = P(M)P(X|M) = P(M, X), \quad (41)$$

$$Q_{MC}(M, Y) = Q_{MC}(M)Q_{MC}(Y|M) = P(M)P(Y|M) = P(M, Y). \quad (42)$$

Hence, $Q_{MC}(M, X, Y) \in \Delta_P$.

Part 3: Showing $Q_{MC}(M, X, Y)$ satisfies (37)

It is trivial to see that by the construction of $Q_{MC}(M, X, Y)$ given in Part 1, we have:

$$Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(X|M)Q_{MC}(Y|X). \quad (43)$$

By the chain rule, we have:

$$Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(Y|M)Q_{MC}(X|Y, M). \quad (44)$$

Hence, for showing that $Q_{MC}(M, X, Y)$ satisfies (37), we only need to show that $Q_{MC}(X|Y, M) = Q_{MC}(X|Y)$.

Showing $Q_{MC}(X|Y, M) = Q_{MC}(X|Y)$: Observe that $Y = \gamma_Y/\gamma_X X$ by construction. Therefore, we can equivalently express $X = \gamma_X/\gamma_Y Y$. Consequently,

$$Q_{MC}(X|Y, M) = \mathbb{I}[X = \gamma_X/\gamma_Y Y] \text{ as } X = \gamma_X/\gamma_Y Y, \quad (45)$$

where $\mathbb{I}[\cdot]$ is the indicator function. Since R.H.S of the above equation does not depend on M , we have:

$$Q_{MC}(X|Y, M) = \mathbb{I}[X = \gamma_X/\gamma_Y Y] = Q_{MC}(X|Y). \quad (46)$$

Combining (44) and (46), we have:

$$Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(Y|M)Q_{MC}(X|Y). \quad (47)$$

Combining (43) and (47), we can show that $Q_{MC}(M, X, Y)$ satisfies (37).

Conclusion: Combining the results of parts 2 and 3, we can conclude that $Q_{MC}(M, X, Y) \in \Delta_P$ and contains both Markovian structures: $M \rightarrow X \rightarrow Y$ and $M \rightarrow Y \rightarrow X$. A unique implication of $Q_{MC}(M, X, Y)$ admitting both Markovian structures is that both $I_{Q_{MC}}(M; X|Y) = I_{Q_{MC}}(M; Y|X) = 0$. Consequently, both minimization problems presented in (2) are minimized by $Q_{MC}(M, X, Y)$ with $UI(M; X|Y) = \min_{Q \in \Delta_P} I_Q(M; X|Y) = 0$ and $UI(M; Y|X) = \min_{Q \in \Delta_P} I_Q(M; Y|X) = 0$, completing our proof. \square

Lemma 3. Let $X \sim \text{Uniform}(0, a)$ for some $a \in \mathbb{R}$, then $\gamma X \sim \text{Uniform}(0, \gamma a)$ for $\gamma \in \mathbb{R}$ and $a, \gamma \neq 0$.

Proof. We denote the p.d.f. of the random variable X as $p_X(\cdot)$. By lemma statement, we know:

$$p_X(X = x) = \text{Uniform}(0, a) = \begin{cases} \frac{1}{a} & x \in [0, a] \\ 0 & \text{otherwise} \end{cases}. \quad (48)$$

Denote the p.d.f. of the function $f(X)$ of the random variable X as $p_{f(X)}(\cdot)$. We want to derive $p_{f(X)}(\cdot)$ for $f(X) = \gamma X$. Since, $f(X) = \gamma X$ is an invertible function, we can use the change-of-variable technique for deriving $p_{f(X)}(\cdot)$:

$$p_{f(X)}(f(X) = y) = p_X(X = f^{-1}(y)) \left| \frac{df^{-1}(y)}{dy} \right| = \begin{cases} 1/|\gamma a| & y \in [0, \gamma a] \\ 0 & \text{otherwise} \end{cases} = \text{Uniform}(0, \gamma a).$$

\square

E.4 PID for the miscellaneous system M4.

Miscellaneous System M4: Let (M, X, Y) be a system of random variables with joint distribution $P(M, X, Y)$. The joint distribution $P(M, X, Y)$ describes the miscellaneous system **M4** if it satisfies the following two properties:

1. $M \sim P(M)$ and has support $[0, 1]$.
2. $P(X|M) = \text{Bernoulli}(M)$, and $P(Y|M) = \text{Bernoulli}(M + p - 2pM)$, where $p \in [0, 1]$.

This system is adapted from the Bernoulli decomposition listed in Appendix A [26].

Proposition 5. Let M, X , and Y be random variables whose joint distribution $P(M, X, Y)$ describes the miscellaneous system **M4** defined above. Then, Δ_P contains a Markov chain of the form $M \rightarrow X \rightarrow Y$ and $UI(M; Y|X) = 0$.

Proof. We briefly outline the proof structure. The proof can be divided into two parts:

1. In the first part, we explicitly construct a joint distribution $Q_{MC}(M, X, Y)$ having the Markovian structure $M \rightarrow X \rightarrow Y$.
2. In the second part, we conclude our proof by showing the $Q_{MC}(M, X, Y)$ constructed in the first part, lies in Δ_P .

Part 1: Specifying $Q_{MC}(M, X, Y)$

Since $Q_{MC}(M, X, Y)$ is the joint distribution of a Markov chain $M \rightarrow X \rightarrow Y$, we have $Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(X|M)Q_{MC}(Y|X)$. Leveraging the special structure of $Q_{MC}(M, X, Y)$, we specify $Q_{MC}(M, X, Y)$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(X|M)$, and $Q_{MC}(Y|X)$. We choose $Q_{MC}(M)$, $Q_{MC}(X|M)$ as follows:

$$Q_{MC}(M) = P(M), \text{ and } Q_{MC}(X|M) = P(X|M). \quad (49)$$

The distribution $Q_{MC}(Y|X)$ is specified through the following stochastic transformation:

$$Y = X(1 - Z) + Z(1 - X), \quad (50)$$

where $Z \sim \text{Bernoulli}(p)$ and $Z \perp\!\!\!\perp (X, M)$.

Part 2: Showing $Q_{MC}(M, X, Y) \in \Delta_P$

For showing $Q_{MC}(M, X, Y) \in \Delta_P$, we first need to calculate $Q_{MC}(Y|M)$:

Calculating $Q_{MC}(Y|M)$: For calculating $Q(Y|M)$, we use the result of proposition 10 in [26], which states that if $X' \sim \text{Bernoulli}(\theta)$, $Z' \sim \text{Bernoulli}(p')$, $Z' \perp\!\!\!\perp X'$, and $Y' = X'(1 - Z') + (1 - X')Z'$, then $Y \sim \text{Bernoulli}(\theta + p' - 2p'\theta)$.

Noting that $Q_{MC}(X|M) = \text{Bernoulli}(M)$ and $Y = X(1 - Z) + Z(1 - X)$, where $Z \sim \text{Bernoulli}(p)$ and $Z \perp\!\!\!\perp X|M$, we can calculate $Q_{MC}(Y|M)$ as follows:

$$Q_{MC}(Y|M) = \text{Bernoulli}(M + p - 2pM) = P(Y|M). \quad (51)$$

Then, from (49) and (51), we can conclude:

$$Q_{MC}(M, X) = Q_{MC}(M)Q_{MC}(X|M) = P(M)P(X|M) = P(M, X), \quad (52)$$

$$Q_{MC}(M, Y) = Q_{MC}(M)Q_{MC}(Y|M) = P(M)P(Y|M) = P(M, Y). \quad (53)$$

Hence, $Q_{MC}(M, X, Y) \in \Delta_P$ and consequently, by proposition 1, $UI(M; Y \setminus X) = 0$. This concludes our proof. \square

E.5 PID for the miscellaneous system M5.

Miscellaneous System M5: Let (\vec{M}, X, Y) be a system of random variables with joint distribution $P(\vec{M}, X, Y)$. The joint distribution $P(\vec{M}, X, Y)$ describes the miscellaneous system **M5** if it satisfies the following two properties:

1. $\vec{M} \sim P(\vec{M})$ and has support over simplex of dimension d .
2. $P(X|\vec{M}) = \text{Categorical}(M_1, \dots, M_d)$, and $P(Y|\vec{M}) = \text{Categorical}(\phi_1, \dots, \phi_d)$, where $\phi_i = (1 - p)M_i + p/d$.

This system is adapted from the categorical decomposition listed in Appendix A [26].

Proposition 6. Let \vec{M} , X , and Y be random variables whose joint distribution $P(\vec{M}, X, Y)$ describes the miscellaneous system **M5** defined above. Then, Δ_P contains a Markov chain of the form $\vec{M} \rightarrow X \rightarrow Y$ and $UI(\vec{M}; Y \setminus X) = 0$.

Proof. We briefly outline the proof structure. The proof can be divided into two parts:

1. In the first part, we explicitly construct a joint distribution $Q_{MC}(\vec{M}, X, Y)$ having the Markovian structure $\vec{M} \rightarrow X \rightarrow Y$.

2. In the second part, we conclude our proof by showing the $Q_{MC}(\vec{M}, X, Y)$ constructed in the first part, lies in Δ_P .

Part 1: Specifying $Q_{MC}(\vec{M}, X, Y)$

Since $Q_{MC}(\vec{M}, X, Y)$ is the joint distribution of a Markov chain $\vec{M} \rightarrow X \rightarrow Y$, we have $Q_{MC}(\vec{M}, X, Y) = Q_{MC}(\vec{M})Q_{MC}(X|\vec{M})Q_{MC}(Y|X)$. Taking advantage of this special structure of $Q_{MC}(\vec{M}, X, Y)$, we specify $Q_{MC}(\vec{M}, X, Y)$ by individually specifying $Q_{MC}(\vec{M})$, $Q_{MC}(X|\vec{M})$, and $Q_{MC}(Y|X)$. We choose $Q_{MC}(\vec{M})$, $Q_{MC}(X|\vec{M})$ as follows:

$$Q_{MC}(\vec{M}) = P(\vec{M}), \text{ and } Q_{MC}(X|\vec{M}) = P(X|\vec{M}). \quad (54)$$

The distribution $Q_{MC}(Y|X)$ is specified through the following transformation:

$$Y = X(1 - Z) + ZD, \quad (55)$$

where $Z \sim \text{Bernoulli}(p)$, $D \sim \text{Categorical}(\frac{1}{d}, \dots, \frac{1}{d})$, and D , Z , and (X, \vec{M}) are jointly independent.

Part 2: Showing $Q_{MC}(\vec{M}, X, Y) \in \Delta_P$

For showing $Q_{MC}(\vec{M}, X, Y) \in \Delta_P$, we first need to calculate $Q_{MC}(Y|\vec{M})$:

Calculating $Q_{MC}(Y|\vec{M})$: For calculating $Q_{MC}(Y|\vec{M})$, we use the result of proposition 11 in [26], which states that:

If $X' \sim \text{Categorical}(\theta_1, \dots, \theta_d)$, $Z' \sim \text{Bernoulli}(p')$, $D' \sim \text{Categorical}(\frac{1}{d}, \dots, \frac{1}{d})$, and $Y' = X'(1 - Z') + Z'D'$ with X' , Z' , and D' jointly independent, then $Y' \sim \text{Categorical}(\phi'_1, \dots, \phi'_d)$ with $\phi'_i = (1 - p')\theta_i + p'/d$.

Noting that $Q_{MC}(X|\vec{M}) = \text{Categorical}(M_1, \dots, M_d)$ and $Y = X(1 - Z) + ZD$, with X, D , and Z jointly conditionally independent conditioned on M , we have:

$$Q_{MC}(Y|\vec{M}) = \text{Categorical}(\phi_1, \dots, \phi_d), \text{ with } \phi_i = M_i(1 - p) + \frac{p}{d} = P(Y|\vec{M}). \quad (56)$$

From (54) and (56), we can conclude:

$$Q_{MC}(\vec{M}, X) = Q_{MC}(\vec{M})Q_{MC}(X|\vec{M}) = P(M)P(X|\vec{M}) = P(\vec{M}, X), \quad (57)$$

$$Q_{MC}(\vec{M}, Y) = Q_{MC}(\vec{M})Q_{MC}(Y|\vec{M}) = P(M)P(Y|\vec{M}) = P(\vec{M}, Y). \quad (58)$$

Hence, $Q_{MC}(\vec{M}, X, Y) \in \Delta_P$ and consequently, by proposition 1, $UI(\vec{M}; Y \setminus X) = 0$. This concludes our proof. \square

E.6 PID for the miscellaneous system M6.

Miscellaneous System M6: Let (M, \vec{X}, \vec{Y}) be a system of random variables with joint distribution $P(M, \vec{X}, \vec{Y})$. The joint distribution $P(M, \vec{X}, \vec{Y})$ describes the miscellaneous system **M6** if it satisfies the following two properties:

1. $M \sim P(M)$ and has support \mathbb{N} .
2. $P(\vec{X}|M) = \text{Multinomial}_{d_X}(\vec{X}; M, \vec{p}_X)$, and $P(\vec{Y}|M) = \text{Multinomial}_{d_Y}(\vec{Y}; M, \vec{p}_Y)$.

Proposition 7. Let M be a random variable. Furthermore, assume \vec{X} and \vec{Y} are random vectors of size d_X and d_Y , respectively, such that $P(M, \vec{X}, \vec{Y})$ describes miscellaneous system **M6**.

1. If, $\min_{i \in [d_Y]} p_i^Y \geq \min_{i \in [d_X]} p_i^X$, then there exists a Markov chain of the form $M \rightarrow \vec{X} \rightarrow \vec{Y}$ in Δ_P , and $UI(M; \vec{Y} \setminus \vec{X}) = 0$.
2. If, $\min_{i \in [d_Y]} p_i^Y \leq \min_{i \in [d_X]} p_i^X$, then there exists a Markov chain of the form $M \rightarrow \vec{Y} \rightarrow \vec{X}$ in Δ_P , and $UI(M; \vec{X} \setminus \vec{Y}) = 0$.

Proof. We only provide an explicit proof of condition 1. The proof of condition 2 follows the same steps as the proof of condition 1 with parameters of \vec{X} and \vec{Y} switched.

We briefly outline the proof structure. The proof can be divided into two major parts:

1. In the first part, we explicitly construct a joint distribution $Q_{MC}(M, \vec{X}, \vec{Y})$ having the Markovian structure $M \rightarrow \vec{X} \rightarrow \vec{Y}$.
2. In the second part, we conclude our proof by showing the $Q_{MC}(M, \vec{X}, \vec{Y})$ constructed in the first part, lies in Δ_P .

Part 1: Specifying $Q_{MC}(M, \vec{X}, \vec{Y})$

Since $Q_{MC}(M, \vec{X}, \vec{Y})$ is the joint distribution of a Markov chain $M \rightarrow \vec{X} \rightarrow \vec{Y}$, we have $Q_{MC}(M, \vec{X}, \vec{Y}) = Q_{MC}(M)Q_{MC}(\vec{X}|M)Q_{MC}(\vec{Y}|\vec{X})$. Taking advantage of this special structure of $Q_{MC}(M, \vec{X}, \vec{Y})$, we specify $Q_{MC}(M, \vec{X}, \vec{Y})$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(\vec{X}|M)$, and $Q_{MC}(\vec{Y}|\vec{X})$. We choose $Q_{MC}(M)$, $Q_{MC}(\vec{X}|M)$ as follows:

$$Q_{MC}(M) = P(M), \text{ and } Q_{MC}(\vec{X}|M) = P(\vec{X}|M). \quad (59)$$

Specifying $Q(\vec{Y}|\vec{X})$: Let $k_X^* = \arg \min_{i \in [d_X]} p_i^X$, and $k_Y^* = \arg \min_{i \in [d_Y]} p_i^Y$. Then, $Q(\vec{Y}|\vec{X}) = \text{Multinomial}_{d_Y}(\vec{Y}; \sum_{i \in [d_X] \setminus k_X^*} x_i, \mathbf{p}_Y^*)$, where

$$\mathbf{p}_Y^* = \left[\frac{p_1^Y}{\sum_{i \in [d_X] \setminus k_X^*} p_i^X} \cdots \frac{p_{d_Y-1}^Y}{\sum_{i \in [d_X] \setminus k_X^*} p_i^X} 1 - \frac{\sum_{i=1}^{d_Y-1} p_i^Y}{\sum_{i \in [d_X] \setminus k_X^*} p_i^X} \right]^T. \quad (60)$$

We first show the proposed $Q_{MC}(\vec{Y}|\vec{X})$ is a valid distribution. By condition 1:

$$\min_{i \in [d_Y]} p_i^Y \geq \min_{i \in d_X} p_i^X \quad (61)$$

$$\Rightarrow \sum_{i \in [d_X] \setminus k_X^*} p_i^X \geq \sum_{i \in [d_Y] \setminus k_Y^*} p_i^Y \geq p_j^Y, \forall j \in [d_Y] \quad (62)$$

$$\Rightarrow 1 \geq \frac{p_j^Y}{\sum_{i \in [d_X] \setminus k_X^*} p_i^X} \geq 0, \forall j \in [d_Y], \quad (63)$$

where the last inequality is due to \vec{p}_X and \vec{p}_Y forming valid multinomial distributions.

Part 2: Showing $Q_{MC}(M, \vec{X}, \vec{Y}) \in \Delta_P$

For showing $Q_{MC}(M, \vec{X}, \vec{Y}) \in \Delta_P$, we first need to calculate $Q_{MC}(\vec{Y}|M)$:

Calculating $Q_{MC}(\vec{Y}|M)$: We use lemma 4 to calculate $Q(\vec{Y}|M)$. Note that $Q_{MC}(\vec{X}|M) = \text{Multinomial}_{d_X}(M, \vec{p}_X)$, hence we have:

$$\begin{aligned} Q_{MC}(Y|M) &= \text{Multinomial}_{d_Y}(M, \vec{p}_Y'), \text{ where,} \\ \vec{p}_Y' &= \left[\frac{p_1^Y \sum_{i \in [d_X] \setminus k_X^*} p_i^X}{\sum_{i \in [d_X] \setminus k_X^*} p_i^X} \cdots \frac{p_{d_Y-1}^Y \sum_{i \in [d_X] \setminus k_X^*} p_i^X}{\sum_{i \in [d_X] \setminus k_X^*} p_i^X} 1 - \frac{\sum_{i \in [d_X] \setminus k_X^*} p_i^X \sum_{i=1}^{d_Y-1} p_i^Y}{\sum_{i \in [d_X] \setminus k_X^*} p_i^X} \right] \\ \Rightarrow \vec{p}_Y' &= [p_1^Y \cdots p_{d_Y-1}^Y \ 1 - \sum_{i \in [d_Y]} p_i^Y] = \vec{p}_Y \\ \Rightarrow Q_{MC}(Y|M) &= \text{Multinomial}_{d_Y}(M, \vec{p}_Y) = P(Y|M). \end{aligned} \quad (64)$$

Then, from (59) and (64), we can conclude:

$$Q_{MC}(M, \vec{X}) = Q_{MC}(M)Q_{MC}(\vec{X}|M) = P(M)P(\vec{X}|M) = P(M, \vec{X}), \quad (65)$$

$$Q_{MC}(M, \vec{Y}) = Q_{MC}(M)Q_{MC}(\vec{Y}|M) = P(M)P(\vec{Y}|M) = P(M, \vec{Y}). \quad (66)$$

Hence, $Q_{MC}(M, \vec{X}, \vec{Y}) \in \Delta_P$ and consequently, by proposition 1, $UI(M; \vec{Y} \setminus \vec{X}) = 0$. This concludes our proof. \square

Lemma 4. Suppose $\vec{X} \sim \text{Multinomial}_{d_X}(n, \vec{p}_X)$ and $\vec{Y}|\vec{X} \sim \text{Multinomial}_{d_Y}\left(\sum_{j \in \mathcal{I}} X_j, \vec{q}\right)$, where $\mathcal{I} \subset \{1, \dots, d_X\}$. Then $\vec{Y} \sim \text{Multinomial}_{d_Y}(n, \vec{q}^*)$, where

$$q_i^* = \begin{cases} \left[\sum_{j \in \mathcal{I}} p_j\right] \cdot q_i, & 1 \leq i < d_Y \\ 1 - \left[\sum_{j \in \mathcal{I}} p_j\right] \cdot \left[\sum_{j=1}^{d_Y-1} q_j\right], & i = d_Y \end{cases}. \quad (67)$$

Proof. Without loss of generality, reshuffle the class indices of \vec{Y} such that $\mathcal{I} = \{i: 1 \leq i \leq w\}$, where $|\mathcal{I}| = w$. Let $C = \{\vec{x}' \in \mathbb{N}_0^{d_X-1} : \sum_{i=1}^{d_X-1} x_i \leq n\}$. By the law of total probability,

$$P(\vec{Y} = \vec{y}) = \sum_{\vec{x} \in C} P(\vec{X} = \vec{x}) P(\vec{Y} = \vec{y} | \vec{X} = \vec{x}) \quad (68)$$

$$= \sum_{\vec{x} \in C} \frac{n!}{\prod_{i=1}^{d_X} x_i!} \prod_{i=1}^{d_X} p_i^{x_i} \cdot \frac{\left(\sum_{j \in \mathcal{I}} x_j\right)!}{\prod_{i=1}^{d_Y} y_i!} \prod_{i=1}^{d_Y} q_i^{y_i}. \quad (69)$$

Remove from the sum all terms that do not depend on \vec{x} . Note that $y_{d_Y} = \sum_{j \in \mathcal{I}} x_j - \sum_{i=1}^{d_Y-1} y_i$ and $q_{d_Y}^{y_{d_Y}} = \left(1 - \sum_{i=1}^{d_Y-1} q_i\right)^{y_{d_Y}}$ are both functions of \vec{y} . Let $y_s = \sum_{i=1}^{d_Y-1} y_i$ and $q_s = \sum_{i=1}^{d_Y-1} q_i$.

$$P(\vec{Y} = \vec{y}) = \left(\frac{n!}{\prod_{i=1}^{d_Y-1} y_i!} \prod_{i=1}^{d_Y-1} q_i^{y_i}\right) \sum_{\vec{x} \in C} \frac{\prod_{i=1}^{d_X} p_i^{x_i} \cdot \left(\sum_{j \in \mathcal{I}} x_j\right)!}{\prod_{i=1}^{d_X} x_i! \left(\sum_{j \in \mathcal{I}} x_j - y_s\right)!} (1 - q_s)^{\sum_{j \in \mathcal{I}} x_j - y_s}.$$

Form a multinomial coefficient outside the sum:

$$P(\vec{Y} = \vec{y}) = \left(\frac{n!}{\prod_{i=1}^{d_Y-1} y_i! \cdot (n - y_s)!} \prod_{i=1}^{d_Y-1} q_i^{y_i}\right) \quad (70)$$

$$\times \sum_{\vec{x} \in C} \frac{(n - y_s)! \prod_{i=1}^{d_X} p_i^{x_i} \cdot \left(\sum_{j \in \mathcal{I}} x_j\right)!}{\prod_{i=1}^{d_X} x_i! \left(\sum_{j \in \mathcal{I}} x_j - y_s\right)!} (1 - q_s)^{\sum_{j \in \mathcal{I}} x_j - y_s}. \quad (71)$$

Let $\mathcal{I}^0 = \{i: w < i \leq d_X\}$ be the set of indices that are not taken in the number of trials for \vec{Y} . Consider the sum over $\vec{x} \in C$ alone in (71). Separating terms that belong to \mathcal{I} and \mathcal{I}^0 ,

$$\sum_{\vec{x} \in C} \frac{(n - y_s)!}{\prod_{j \in \mathcal{I}^0} x_j! \left(\sum_{j \in \mathcal{I}} x_j - y_s\right)!} \prod_{j \in \mathcal{I}^0} p_j^{x_j} (1 - q_s)^{\sum_{j \in \mathcal{I}} x_j - y_s} \cdot \frac{\left(\sum_{j \in \mathcal{I}} x_j\right)!}{\prod_{j \in \mathcal{I}} x_j!} \prod_{j \in \mathcal{I}} p_j^{x_j}.$$

Let $C' = \{\vec{x}' \in \mathbb{N}_0^{d_X-1} : \sum_{i=1}^{d_X-1} x_i - \sum_{i=1}^{d_Y-1} y_i \leq n - \sum_{i=1}^{d_Y-1} y_i\}$. Note $\vec{x} \in C \Leftrightarrow \vec{x} \in C'$, so we can equivalently sum over elements in C' . Perform a change of variable with $u = \sum_{j \in \mathcal{I}} x_j - y_s$ and define $B = \{x_j \in \vec{x}: j \in \mathcal{I}, \sum_{j \in \mathcal{I}} x_j = u + y_s\}$. Then the sum becomes

$$= \sum_{\vec{x} \in C'} \frac{(n - y_s)!}{\prod_{j \in \mathcal{I}^0} x_j! u!} \prod_{j \in \mathcal{I}^0} p_j^{x_j} (1 - q_s)^u \sum_{x_j \in B} \binom{u + y_s}{x_{j_1}, \dots, x_{j_m}} \prod_{j \in \mathcal{I}} p_j^{x_j}. \quad (72)$$

Note that the inner sum simplifies by the multinomial theorem:

$$= \sum_{\vec{x} \in C'} \frac{(n - y_s)!}{\prod_{j \in \mathcal{I}^0} x_j! u!} \prod_{j \in \mathcal{I}^0} p_j^{x_j} (1 - q_s)^u \left(\sum_{j \in \mathcal{I}} p_j\right)^{u + y_s} \quad (73)$$

$$= \left(\sum_{j \in \mathcal{I}} p_j\right)^{y_s} \sum_{\vec{x} \in C'} \frac{(n - y_s)!}{\prod_{j \in \mathcal{I}^0} x_j! \cdot u!} \prod_{j \in \mathcal{I}^0} p_j^{x_j} \left(\sum_{j \in \mathcal{I}} p_j - q_s \sum_{j \in \mathcal{I}} p_j\right)^u. \quad (74)$$

Reapply the multinomial theorem to the sum of \vec{x} over C' :

$$= \left(\sum_{j \in \mathcal{I}} p_j \right)^{y_s} \left(\sum_{j \in \mathcal{I}^0} p_j + \sum_{j \in \mathcal{I}} p_j - q_s \sum_{j \in \mathcal{I}} p_j \right)^{n-y_s} \quad (75)$$

$$= \left(\sum_{j \in \mathcal{I}} p_j \right)^{y_s} \left(1 - q_s \sum_{j \in \mathcal{I}} p_j \right)^{n-y_s}. \quad (76)$$

Inserting (76) into (71), the marginal becomes

$$P(\vec{Y} = \vec{y}) = \frac{n! \prod_{i=1}^{d_Y-1} q_i^{y_i}}{\prod_{i=1}^{d_Y-1} y_i! \cdot (n - y_s)!} \left(\sum_{j \in \mathcal{I}} p_j \right)^{y_s} \left(1 - q_s \sum_{j \in \mathcal{I}} p_j \right)^{n-y_s} \quad (77)$$

$$= \frac{n! \prod_{i=1}^{d_Y-1} \left(\sum_{j \in \mathcal{I}} p_j \cdot q_i \right)^{y_i}}{\prod_{i=1}^{d_Y-1} y_i! \cdot (n - y_s)!} \left(1 - q_s \sum_{j \in \mathcal{I}} p_j \right)^{n-y_s} \quad (78)$$

$$= \text{Multinomial}_{d_Y}(n, \vec{q}^*) \quad (79)$$

where \vec{q}^* is defined as in (67). \square

F Proof of Theorem 1

In this section, we provide the proof of theorem 1. Since theorem 1 provides the analytical PID terms for the univariate affine continuous stable system defined in Sec. 4, we briefly restate certain key properties of the univariate stable distributions and the corresponding univariate affine continuous stable system for convenience.

F.1 Univariate continuous stable distribution

Univariate continuous stable distribution are a large class of distributions that naturally arise in the context of generalized central limit theorems. We refer the reader to Appx. N.2 for more details on univariate continuous stable distributions. We now list certain key properties of univariate continuous stable distributions that we make use of in the proof of theorem 1:

1. If X is distributed according to a univariate continuous stable distribution, then the sum of two independent copies of X , denoted as X_1 and X_2 , follows the same univariate continuous stable distribution upto a scaling and translation operation, i.e., $aX_1 + bX_2 \stackrel{d}{=} cX + d$ for $a, b, c > 0$ and $d \in \mathbb{R}$.
2. The p.d.f. of univariate continuous stable distributions is characterized by four parameters: stability parameter denoted as $\alpha \in (0, 2]$, skewness parameter denoted as $\beta \in [-1, 1]$, scale parameters denoted as $\gamma \in (0, \infty)$, and location parameter denoted as $\mu \in \mathbb{R}$. We denote the p.d.f. of a univariate continuous stable distribution as $p_{CS}(\alpha, \beta, \gamma, \mu)$.

F.2 Definition of univariate affine continuous stable system

Let (M, X, Y) be a system of random variables with joint distribution $P(M, X, Y)$. The joint distribution $P(M, X, Y)$ describes the univariate affine continuous stable system if it satisfies the following two properties:

1. $M \sim P(M)$ having some support set $\mathcal{M} \subseteq \mathbb{R}$.
2. The conditional distributions of random variables X and Y conditioned on M can be expressed through univariate continuous stable distributions with an affine dependence on M , i.e., $P(X|M) = p_{CS}(\alpha, \beta_X, \gamma_X, aM + b)$ and $P(Y|M) = p_{CS}(\alpha, \beta_Y \text{sgn}(ac), \gamma_Y, cM + d)$, where $a, b, c, d \in \mathbb{R}$.

F.3 Formal proof of Theorem 1

We first discuss and provide the proofs of lemma 5 and lemma 6 that we will use for proving theorem 1. In the following proofs, $\text{sgn}(x)$ denotes the sign function:

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} . \quad (80)$$

Lemma 5. *If $X \sim p_{CS}(\alpha, \beta, \gamma, \mu)$, then for any $\eta \neq 0$ and $\kappa \in \mathbb{R}$, we have*

$$\eta X + \kappa \sim \begin{cases} p_{CS}(\alpha, \text{sgn}(\eta)\beta, |\eta|\gamma, \eta\mu + \kappa) & \alpha \neq 1 \\ p_{CS}(\alpha, \text{sgn}(\eta)\beta, |\eta|\gamma, \eta\mu + \kappa - 2/\pi\beta\gamma\eta \log(|\eta|)) & \alpha = 1 \end{cases} .$$

Proof. See the proof of proposition 1.4 part (b) in [23]. \square

Lemma 6. *If $X_1 \sim p_{CS}(\alpha, \beta_1, \gamma_1, \mu_1)$ and $X_2 \sim p_{CS}(\alpha, \beta_2, \gamma_2, \mu_2)$ are independent, then $X_1 + X_2 \sim p_{CS}(\alpha, \beta, \gamma, \mu)$ where*

$$\beta = \frac{\beta_1\gamma_1^\alpha + \beta_2\gamma_2^\alpha}{\gamma_1^\alpha + \gamma_2^\alpha}, \quad \gamma^\alpha = \gamma_1^\alpha + \gamma_2^\alpha, \quad \text{and } \mu = \mu_1 + \mu_2.$$

Proof. See the proof of proposition 1.4 part (c) in [23]. \square

Theorem 1. *Let M, X , and Y be random variables whose joint distribution $P(M, X, Y)$ describes a univariate affine continuous stable system. Without the loss of generality, assume $|a|/\gamma_X \geq |c|/\gamma_Y$. If $1 - \beta_Y \geq (\gamma_X|c|/\gamma_Y|a|)^\alpha (1 - \beta_X)$, and $1 + \beta_Y \geq (\gamma_X|c|/\gamma_Y|a|)^\alpha (1 + \beta_X)$, then Δ_P contains a Markov chain of the form $M \rightarrow X \rightarrow Y$ and $UI(M; Y \setminus X) = 0$.*

Proof. We first note that we can always assume $|a|/\gamma_X \geq |c|/\gamma_Y$ without the loss of generality because if $|c|/\gamma_Y \geq |a|/\gamma_X$, then we can always switch our nomenclature to refer to Y as X , and X as Y .

We now briefly outline the proof structure. We divide the proof into two cases $|a|/\gamma_X > |c|/\gamma_Y$ and $|a|/\gamma_X = |c|/\gamma_Y$. The proof for both cases follows essentially the same structure consisting of two major parts:

1. In the first part, we explicitly construct a joint distribution $Q_{MC}(M, X, Y)$ having the Markovian structure $M \rightarrow X \rightarrow Y$.
2. In the second part, we show that the $Q_{MC}(M, X, Y)$ constructed in the first part, lies in Δ_P . Therefore, we can then apply the result of proposition 1 to conclude $UI(M; Y \setminus X) = 0$.

Case 1: $|a|/\gamma_X > |c|/\gamma_Y$

Part 1: Specifying $Q_{MC}(M, X, Y)$

We explicitly construct the desired Markov chain $M \rightarrow X \rightarrow Y$ by constructing a larger Markov chain $M \rightarrow X \rightarrow X' \rightarrow Y' \rightarrow Y$ and then marginalizing the larger Markov chain to obtain the desired Markov chain $M \rightarrow X \rightarrow Y$.

Denote the joint distribution of the Markov chain $M \rightarrow X \rightarrow X' \rightarrow Y' \rightarrow Y$ as $Q_{MC}(M, X, X', Y', Y)$. We can decompose $Q_{MC}(M, X, X', Y', Y)$ by utilizing its Markovian structure as follows:

$$Q_{MC}(M, X, X', Y', Y) = Q_{MC}(M)Q_{MC}(X|M)Q_{MC}(X'|X)Q_{MC}(Y'|X')Q_{MC}(Y|Y').$$

Consequently, we can specify/construct the distribution $Q_{MC}(M, X, X', Y', Y)$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(X|M)$, $Q_{MC}(X'|X)$, $Q_{MC}(Y'|X')$, and $Q_{MC}(Y|Y')$.

Specifying $Q_{MC}(M)$ and $Q_{MC}(X|M)$: We choose $Q_{MC}(M)$ and $Q_{MC}(X|M)$ as follows:

$$Q_{MC}(M) = P(M) \text{ and } Q_{MC}(X|M) = P(X|M), \quad (81)$$

where $P(M)$ and $P(X|M)$ are marginal distributions derived from the original joint distribution $P(M, X, Y)$ (discussed in the theorem statement) over which the bivariate PID is being calculated.

Specifying $Q_{MC}(X'|X)$: The distribution $Q_{MC}(X'|X)$ is specified by the following deterministic transformation:

$$X' = \begin{cases} \frac{1}{a}X & \alpha \neq 1 \\ \frac{1}{a}X + (2\beta_X\gamma_X \log(|1/a|))/(\pi a) & \alpha = 1 \end{cases} \quad (82)$$

We will also derive the distribution $Q(X'|M)$ before proceeding with our construction, as we will need it later in the proof to show that the constructed $Q_{MC}(M, X, Y) \in \Delta_P$.

Deriving $Q_{MC}(X'|M)$: Since $Q_{MC}(X|M = m) = p_{CS}(\alpha, \beta_X, \gamma_X, am + b)$ for a fixed m and X' is scaled and translated version of X , we can use lemma 5 to derive $Q_{MC}(X'|M)$. The exact expression of $Q_{MC}(X'|M = m)$ is provided in (83).

$$\begin{aligned} Q_{MC}(X'|M = m) &= p_{CS}(\alpha, \beta_{X'}, \gamma_{X'}, \mu_{X'}), \text{ where} \\ \beta_{X'} &= \begin{cases} \text{sgn}(1/a)\beta_X & \alpha \neq 1 \\ \text{sgn}(1/a)\beta_X & \alpha = 1 \end{cases} = \text{sgn}(1/a)\beta_X = \text{sgn}(a)\beta_X, \\ \gamma_{X'} &= \begin{cases} |1/a|\gamma_X & \alpha \neq 1 \\ |1/a|\gamma_X & \alpha = 1 \end{cases} = \frac{\gamma_X}{|a|}, \\ \mu_{X'} &= \begin{cases} \frac{1}{a}(am + b) & \alpha \neq 1 \\ \frac{1}{a}(am + b) + 2\beta_X\gamma_X \log(|1/a|)/\pi a - 2\beta_X\gamma_X \log(|1/a|)/\pi a & \alpha = 1 \end{cases} = m + \frac{b}{a}, \\ \Rightarrow Q_{MC}(X'|M = m) &= p_{CS}(\alpha, \beta_X \text{sgn}(a), \gamma_X/|a|, m + b/a). \end{aligned} \quad (83)$$

Specifying $Q_{MC}(Y'|X')$: In order to define $Q_{MC}(Y'|X')$, we need to define an auxiliary variable ϵ . The random variable $\epsilon \sim p_{CS}(\epsilon; \alpha, \beta', \gamma', \mu')$, where

$$\beta' = \text{sgn}(a) \frac{(\gamma_Y/|c|)^\alpha \beta_Y - (\gamma_X/|a|)^\alpha \beta_X}{(\gamma_Y/|c|)^\alpha - (\gamma_X/|a|)^\alpha}, \quad \gamma' = \left(\left(\frac{\gamma_Y}{|c|} \right)^\alpha - \left(\frac{\gamma_X}{|a|} \right)^\alpha \right)^{\frac{1}{\alpha}}, \quad \mu' = \frac{d}{c} - \frac{b}{a}. \quad (84)$$

In order for ϵ to have a legitimate continuous stable distribution, we need to ensure that β' , γ' , and μ' lie within their appropriate bounds as specified in Appx. F.1. It is trivial to see that $\mu' \in \mathbb{R}$.

Showing $\gamma' \in (0, \infty)$. Under the assumption of the case 1, we know that:

$$\frac{|a|}{\gamma_X} > \frac{|c|}{\gamma_Y}.$$

Manipulating the above inequality, we obtain:

$$\begin{aligned} \frac{\gamma_Y}{|c|} > \frac{\gamma_X}{|a|} &\Rightarrow \left(\frac{\gamma_Y}{|c|} \right)^\alpha > \left(\frac{\gamma_X}{|a|} \right)^\alpha \Rightarrow \left(\frac{\gamma_Y}{|c|} \right)^\alpha - \left(\frac{\gamma_X}{|a|} \right)^\alpha > 0 \\ &\Rightarrow \left(\left(\frac{\gamma_Y}{|c|} \right)^\alpha - \left(\frac{\gamma_X}{|a|} \right)^\alpha \right)^{\frac{1}{\alpha}} = \gamma' > 0. \end{aligned} \quad (85)$$

Hence, (85) shows that $\gamma' \in (0, \infty)$.

Lastly, showing that $\beta' \in [-1, 1]$. Note that the condition $\beta' \in [-1, 1]$ can be equivalently expressed as $|\beta'| \leq 1$. For showing that $\beta' \in [-1, 1]$, we will show that the inequality $|\beta'| \leq 1$ is equivalent to the inequalities: $(\gamma_X/|a|)^\alpha (1 - \beta_X) \leq 1 - \beta_Y$ and $(1 + \beta_X) (\gamma_X/|a|)^\alpha \leq 1 + \beta_Y$, where the last two inequalities hold by the theorem assumptions:

$$\begin{aligned} |\beta'| \leq 1 &\Rightarrow \left| \text{sgn}(a) \frac{(\gamma_Y/|c|)^\alpha \beta_Y - (\gamma_X/|a|)^\alpha \beta_X}{(\gamma_Y/|c|)^\alpha - (\gamma_X/|a|)^\alpha} \right| \leq 1 \Rightarrow \left| \frac{(\gamma_Y/|c|)^\alpha \beta_Y - (\gamma_X/|a|)^\alpha \beta_X}{(\gamma_Y/|c|)^\alpha - (\gamma_X/|a|)^\alpha} \right| \leq 1, \\ &\Rightarrow -1 \leq \frac{(\gamma_Y/|c|)^\alpha \beta_Y - (\gamma_X/|a|)^\alpha \beta_X}{(\gamma_Y/|c|)^\alpha - (\gamma_X/|a|)^\alpha} \leq 1. \end{aligned} \quad (86)$$

Multiplying $(\gamma_Y/|c|)^\alpha - (\gamma_X/|a|)^\alpha$ on both sides of the inequalities in (86).

$$(\gamma_X/|a|)^\alpha - (\gamma_Y/|c|)^\alpha \leq (\gamma_Y/|c|)^\alpha \beta_Y - (\gamma_X/|a|)^\alpha \beta_X \leq (\gamma_Y/|c|)^\alpha - (\gamma_X/|a|)^\alpha. \quad (87)$$

Adding $(\gamma_X/|a|)^\alpha \beta_X$ on both sides of the inequalities in (87).

$$\begin{aligned} & (\gamma_X/|a|)^\alpha + (\gamma_X/|a|)^\alpha \beta_X - (\gamma_Y/|c|)^\alpha \leq (\gamma_Y/|c|)^\alpha \beta_Y \leq (\gamma_Y/|c|)^\alpha - (\gamma_X/|a|)^\alpha + (\gamma_X/|a|)^\alpha \beta_X, \\ \Rightarrow & (1 + \beta_X) (\gamma_X/|a|)^\alpha - (\gamma_Y/|c|)^\alpha \leq (\gamma_Y/|c|)^\alpha \beta_Y \leq (\gamma_Y/|c|)^\alpha + (\gamma_X/|a|)^\alpha (\beta_X - 1). \end{aligned} \quad (88)$$

Dividing $(\gamma_Y/|c|)^\alpha$ on both sides of the inequalities in (88).

$$(1 + \beta_X) (\gamma_X|c|/|a|\gamma_Y)^\alpha - 1 \leq \beta_Y \leq 1 + (\gamma_X|c|/|a|\gamma_Y)^\alpha (\beta_X - 1). \quad (89)$$

Simplifying the inequalities in (89).

$$(1 + \beta_X) (\gamma_X|c|/|a|\gamma_Y)^\alpha \leq 1 + \beta_Y, \quad (90)$$

$$\beta_Y - 1 \leq (\gamma_X|c|/|a|\gamma_Y)^\alpha (\beta_X - 1). \quad (91)$$

Multiplying -1 on both sides in (91):

$$(\gamma_X|c|/|a|\gamma_Y)^\alpha (1 - \beta_X) \leq 1 - \beta_Y. \quad (92)$$

Hence, by above analysis we can conclude that the inequality $|\beta'| \leq 1$ is equivalent to the inequalities: $(\gamma_X|c|/|a|\gamma_Y)^\alpha (1 - \beta_X) \leq 1 - \beta_Y$ and $(1 + \beta_X) (\gamma_X|c|/|a|\gamma_Y)^\alpha \leq 1 + \beta_Y$, which ensures $\beta' \in [-1, 1]$.

As $\beta' \in [-1, 1]$, $\gamma' \in (0, \infty)$ and $\mu' \in \mathbb{R}$, we can conclude ϵ follows a legitimate univariate continuous stable distribution. Furthermore, we define $\epsilon \perp\!\!\!\perp (M, X, X')$, i.e., $\Pr(\epsilon|M, X, X') = \Pr(\epsilon)$ and $\Pr(M, X, X'|\epsilon) = \Pr(M, X, X')$.

We now use ϵ to specify $Q_{MC}(Y'|X')$ as follows:

$$Y' = X' + \epsilon. \quad (93)$$

We also derive the distribution $Q(Y'|M)$ before proceeding with our construction, as we will need it later in the proof to show that the constructed $Q_{MC}(M, X, Y) \in \Delta_P$.

Deriving $Q_{MC}(Y'|M)$: We will use the distributions $Q_{MC}(X'|M)$ and $Q_{MC}(\epsilon)$ to derive the distribution of $Q_{MC}(Y'|M)$:

$$\begin{aligned} Q_{MC}(Y'=y'|M=m) &= Q_{MC}(X'+\epsilon=y'|M=m), \\ &= \int_{k=-\infty}^{\infty} Q_{MC}(X'=k, \epsilon=y'-k|M=m)dk, \\ &= \int_{k=-\infty}^{\infty} Q_{MC}(\epsilon=y'-k|M=m)Q_{MC}(X'=k|\epsilon=y'-k, M=m)dk. \end{aligned}$$

Using the fact that $\epsilon \perp\!\!\!\perp M$ and $\epsilon \perp\!\!\!\perp X'|M$, we have:

$$Q_{MC}(Y'=y'|M=m) = \int_{k=-\infty}^{\infty} Q_{MC}(\epsilon=y'-k)Q_{MC}(X'=k|M=m)dk. \quad (94)$$

Since $Q_{MC}(\epsilon)$ and $Q_{MC}(X'|M=m)$ are univariate continuous stable distributions for a fixed m (see (84) and (83)), equation (94) describes a convolution of two univariate stable distributions for a fixed m . Convolution of two univariate continuous stable distribution is akin to summing two independent random variable having the same univariate continuous stable distributions, and

consequently $Q_{MC}(Y'|M)$ can be derived using lemma 6. Hence,

$$\begin{aligned}
Q_{MC}(Y'|M = m) &= p_{CS}(\alpha, \tilde{\beta}, \tilde{\gamma}, \tilde{\mu}), \text{ where} \\
\tilde{\mu} &= m + \frac{b}{a} + \mu' = m + \frac{b}{a} + \frac{d}{c} - \frac{b}{a} = m + \frac{d}{c}, \\
\tilde{\gamma} &= \left(\left(\frac{\gamma_X}{|a|} \right)^\alpha + \gamma'^\alpha \right)^{\frac{1}{\alpha}} = \left(\left(\frac{\gamma_X}{|a|} \right)^\alpha + \left(\frac{\gamma_Y}{|c|} \right)^\alpha - \left(\frac{\gamma_X}{|a|} \right)^\alpha \right)^{\frac{1}{\alpha}} = \frac{\gamma_Y}{|c|}, \\
\tilde{\beta} &= \frac{(\gamma_X/|a|)^\alpha \text{sgn}(a)\beta_X + \gamma'^\alpha \beta'}{(\gamma_X/|a|)^\alpha + \gamma'^\alpha} \\
&= \frac{(\gamma_X/|a|)^\alpha \text{sgn}(a)\beta_X + ((\gamma_Y/|c|)^\alpha - (\gamma_X/|a|)^\alpha) \left(\text{sgn}(a) \frac{(\gamma_Y/|c|)^\alpha \beta_Y - (\gamma_X/|a|)^\alpha \beta_X}{(\gamma_Y/|c|)^\alpha - (\gamma_X/|a|)^\alpha} \right)}{(\gamma_X/|a|)^\alpha + (\gamma_Y/|c|)^\alpha - (\gamma_X/|a|)^\alpha} \\
&= \text{sgn}(a) \frac{(\gamma_X/|a|)^\alpha \beta_X + (\gamma_Y/|c|)^\alpha \beta_Y - (\gamma_X/|a|)^\alpha \beta_X}{(\gamma_Y/|c|)^\alpha} = \text{sgn}(a) \frac{(\gamma_Y/|c|)^\alpha \beta_Y}{(\gamma_Y/|c|)^\alpha} \\
&= \text{sgn}(a)\beta_Y. \tag{95}
\end{aligned}$$

Specifying $Q_{MC}(Y|Y')$: $Q_{MC}(Y|Y')$ is specified by the following deterministic transformation:

$$Y = \begin{cases} cY' & \alpha \neq 1 \\ cY' + (2\text{sgn}(a)c\beta_Y\gamma_Y \log(|c|))/\pi & \alpha = 1 \end{cases}. \tag{96}$$

Finally, we construct the desired $Q_{MC}(M, X, Y)$ from $Q_{MC}(M, X, X', Y', Y)$ by marginalizing X' and Y' .

Part 2: Showing $Q_{MC}(M, X, Y) \in \Delta_P$

For showing that $Q_{MC}(M, X, Y) \in \Delta_P$, we first need to calculate $Q_{MC}(Y|M)$.

Calculating $Q(Y|M)$: We use lemma 5 to calculate $Q(Y|M)$ as from (95), we know that $Q_{MC}(Y'|M) = p_{CS}(Y'; \alpha, \text{sgn}(a)\beta_Y, \gamma_Y/|c|, M + d/c)$ is a continuous stable distribution for a fixed m , and Y a translated and scaled version of Y' . The exact expression of $Q_{MC}(Y|M = m)$ is provided in (97).

$$\begin{aligned}
Q_{MC}(Y|M = m) &= p_{CS}(\alpha, \tilde{\beta}_Y, \tilde{\gamma}_Y, \tilde{\mu}_Y), \text{ where} \\
\tilde{\beta}_Y &= \begin{cases} \text{sgn}(c)\text{sgn}(c)\beta_Y & \alpha \neq 1 \\ \text{sgn}(c)\text{sgn}(a)\beta_Y & \alpha = 1 \end{cases} = \text{sgn}(a)\text{sgn}(c)\beta_Y = \text{sgn}(ac)\beta_Y, \\
\tilde{\gamma}_Y &= \begin{cases} |c| \frac{\gamma_Y}{|c|} & \alpha \neq 1 \\ |c| \frac{\gamma_Y}{|c|} & \alpha = 1 \end{cases} = \gamma_Y, \\
\tilde{\mu}_Y &= \begin{cases} c(m + \frac{d}{c}) & \alpha \neq 1 \\ c(m + \frac{d}{c}) + 2\text{sgn}(a)c\beta_Y\gamma_Y \log(|c|)/\pi - 2\text{sgn}(a)c\beta_Y\gamma_Y \log(|c|)/\pi & \alpha = 1 \end{cases} = cm + d, \\
\Rightarrow Q_{MC}(Y|M = m) &= p_{CS}(\alpha, \beta_Y \text{sgn}(ac), \gamma_Y, cm + d) = P(Y|M). \tag{97}
\end{aligned}$$

From (81) and (97), we can conclude:

$$Q_{MC}(M, X) = Q_{MC}(M)Q_{MC}(X|M) = P(M)P(X|M) = P(M, X), \tag{98}$$

$$Q_{MC}(M, Y) = Q_{MC}(M)Q_{MC}(Y|M) = P(M)P(Y|M) = P(M, Y). \tag{99}$$

Hence, $Q_{MC}(M, X, Y) \in \Delta_P$ and consequently, by proposition 1, $UI(M; Y|X) = 0$, concluding our proof for case 1.

Case 2: $|a|/\gamma_X = |c|/\gamma_Y$

The proof of case 2 is extremely similar to case 1. For case 2, we are able to directly construct the distribution $Q_{MC}(M, X, Y)$ without the need of specifying a larger Markov chain.

Part 1: Specifying $Q_{MC}(M, X, Y)$

Similarly to case 1, we specify $Q_{MC}(M, X, Y)$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(X|M)$ and $Q_{MC}(Y|X)$. The distributions $Q_{MC}(M)$ and $Q_{MC}(X|M)$ are as follows:

$$Q_{MC}(M) = P(M) \text{ and } Q_{MC}(X|M) = P(X|M). \tag{100}$$

The distribution $Q_{MC}(Y|X)$ is defined using the following deterministic transformation:

$$Y = \begin{cases} \frac{c}{a}X - \frac{cb}{a} + d & \alpha \neq 1 \\ \frac{c}{a}X + \frac{2c}{a\pi}\beta\gamma_X \log(|c/a|) - \frac{cb}{a} + d & \alpha = 1 \end{cases} \quad (101)$$

Part 2: Showing $Q_{MC}(M, X, Y) \in \Delta_P$

For showing that $Q_{MC}(M, X, Y) \in \Delta_P$, we first need to calculate $Q_{MC}(Y|M)$.

Calculating $Q(Y|M)$: We use lemma 5 to calculate $Q(Y|M)$ as $Q_{MC}(X|M=m)$ is a continuous stable distribution for a fixed m , and Y a translated and scaled version of X . The exact expression of $Q_{MC}(Y|M=m)$ is provided in (102).

$$\begin{aligned} Q_{MC}(Y|M=m) &= p_{CS}(y; \alpha, \tilde{\beta}, \tilde{\gamma}, \tilde{\mu}), \\ \tilde{\mu} &= \begin{cases} \frac{ca}{a}m + \frac{cb}{a} - \frac{bc}{a} + d & \alpha \neq 1 \\ \frac{ca}{a}m + \frac{cb}{a} - \frac{bc}{a} + d - \frac{2c}{a\pi}\beta\gamma_X \log(|c/a|) + \frac{2c}{a\pi}\beta\gamma_X \log(|c/a|) & \alpha = 1 \end{cases} \\ &= cm + d, \\ \tilde{\gamma} &= |c|^{\gamma_X/|a|} = |c|^{\gamma_Y/|c|} = \gamma_Y, \text{ (using the fact that } \gamma_X/|a| = \gamma_Y/|c| \text{ by assumption)} \\ \tilde{\beta} &= \beta \text{sgn}(c/a) = \beta \text{sgn}(ac), \\ \Rightarrow Q_{MC}(Y|M=m) &= p_{CS}(\alpha, \text{sgn}(ac)\beta_Y, \gamma_Y, cm + d) = P(Y|M). \end{aligned} \quad (102)$$

From (100) and (102), we can conclude:

$$Q_{MC}(M, X) = Q_{MC}(M)Q_{MC}(X|M) = P(M)P(X|M) = P(M, X), \quad (103)$$

$$Q_{MC}(M, Y) = Q_{MC}(M)Q_{MC}(Y|M) = P(M)P(Y|M) = P(M, Y). \quad (104)$$

Hence, $Q_{MC}(M, X, Y) \in \Delta_P$ and consequently, by proposition 1, $UI(M; Y \setminus X) = 0$, concluding our proof for case 2 and theorem 1. \square

G Proof of Theorem 2

In this section, we provide the proof of theorem 2. Since theorem 2 provides the analytical PID terms for system 1 of the multivariate affine continuous stable system defined in Sec. 4, we briefly restate certain key properties of the corresponding independent component multivariate stable distributions and the corresponding multivariate affine continuous stable system for convenience.

G.1 Independent component multivariate stable distribution

Independent component multivariate stable distributions are a specific multivariate generalization of univariate continuous stable distributions describing a collection of independent random variables, with each random variable distributed according to a univariate continuous stable distribution. We refer the reader to Appx. N.3 for more details on multivariate continuous stable distributions. We now list certain key properties of independent component multivariate stable distributions that we make use of in the proof of theorem 2:

1. If \vec{X} is distributed according to an independent component multivariate stable distribution, then the sum of two independent copies of \vec{X} , denoted as \vec{X}_1 and \vec{X}_2 , follows an independent component multivariate stable distribution up to a scaling and translation operation, i.e., $a\vec{X}_1 + b\vec{X}_2 \stackrel{d}{=} c\vec{X} + \vec{d}$ for $a, b, c > 0$ and $\vec{d} \in \mathbb{R}$.
2. We denote the p.d.f. of the independent component multivariate stable continuous distribution as $p_{CS-IC}(\alpha, \vec{\beta}, \vec{\gamma}, \vec{\mu})$, where

$$\vec{\beta} = [\beta_1 \quad \dots \quad \beta_d]^T, \text{ with } \beta_j \in [-1, 1], \vec{\gamma} = [\gamma_1 \quad \dots \quad \gamma_d]^T \text{ with } \gamma_j \in (0, \infty),$$

$$\vec{\mu} = [\mu_1 \quad \dots \quad \mu_d]^T \in \mathbb{R}^d, \text{ and } \alpha \in (0, 2].$$

In general, the p.d.f. of independent component multivariate stable distribution do not have a closed-form analytical expression, and are expressed through their characteristic function.

3. The characteristic function of a d -dimensional random vector \vec{X} having independent component multivariate distribution $p_{CS-IC}(\alpha, \vec{\beta}, \vec{\gamma}, \vec{\mu})$ is expressed as follows:

$$\mathbb{E} \left[e^{i\vec{t}^T \vec{X}} \right] = \exp \left(- \sum_{j=1}^d |\gamma_j t_j|^\alpha (1 - i\beta_j \text{sgn}(t_j) \Phi(\alpha)) + i\vec{t}^T \vec{\mu} \right) \quad \forall \vec{t} \in \mathbb{R}^d, \quad (105)$$

$$\text{where } \Phi(\alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ \frac{-2}{\pi} \log(|t|) & \alpha = 1 \end{cases}, \quad \text{sgn}(t) = \begin{cases} -1 & t < 0 \\ 0 & t = 0 \\ 1 & t > 0 \end{cases}.$$

4. The random vector $\vec{X} = [X_1 \dots X_d]^T$ having the distribution $p_{CS-IC}(\alpha, \vec{\beta}, \vec{\gamma}, \vec{\mu})$ is essentially a collection of independent random variables $\{X_j\}_{j=1}^d$, where $X_j \sim p_{CS}(\alpha, \beta_j, \gamma_j, \mu_j)$.

G.2 Definition of the system 1 of the multivariate affine continuous stable system

Let the random variable $M \sim P(M)$ with support $\mathcal{M} \subseteq \mathbb{R}$. \vec{X} and \vec{Y} are d_X -dimensional and d_Y -dimensional random vectors, respectively. The joint distribution $P(M, \vec{X}, \vec{Y})$ describes system 1 of the multivariate affine continuous stable system if it satisfies the following two properties

1. $M \sim P(M)$ having some support set $\mathcal{M} \subseteq \mathbb{R}$.
2. The conditional distributions of random variables \vec{X} and \vec{Y} conditioned on M are expressed as follows:

$$\vec{X} = \vec{H}_X M + \mathbf{A}_X \vec{Z}_X + \vec{b}_X \text{ and } \vec{Y} = \vec{H}_Y M + \mathbf{A}_Y \vec{Z}_Y + \vec{b}_Y, \quad (106)$$

where $\vec{Z}_X \sim p_{CS-IC}(\alpha, \vec{0}_{d_X}, \vec{1}_{d_X}, \vec{0}_{d_X})$, $\vec{Z}_Y \sim p_{CS-IC}(\alpha, \vec{0}_{d_Y}, \vec{1}_{d_Y}, \vec{0}_{d_Y})$, \mathbf{A}_X and \mathbf{A}_Y are invertible matrices, $\vec{H}_X, \vec{b}_X \in \mathbb{R}^{d_X}$, and $\vec{H}_Y, \vec{b}_Y \in \mathbb{R}^{d_Y}$.

G.3 Formal proof of Theorem 2

Theorem 2. Let the joint distribution $P(M, \vec{X}, \vec{Y})$ of M , \vec{X} , and \vec{Y} describe the system 1 of the multivariate affine continuous stable system. Without the loss of generality, assume $\|\mathbf{A}_Y^{-1} \vec{H}_Y\|_\kappa \leq \|\mathbf{A}_X^{-1} \vec{H}_X\|_\kappa$, where $\kappa = \alpha/\alpha-1 \forall \alpha \in (1, 2]$ and $\kappa = \infty \forall \alpha \in (0, 1]$. Then, Δ_P contains a Markov chain of the form $M \rightarrow \vec{X} \rightarrow \vec{Y}$ and $UI(M; \vec{Y} \setminus \vec{X}) = 0$.

Proof. We first note that we can always assume $\|\mathbf{A}_Y^{-1} \vec{H}_Y\|_\kappa \leq \|\mathbf{A}_X^{-1} \vec{H}_X\|_\kappa$ without the loss of generality because if $\|\mathbf{A}_Y^{-1} \vec{H}_Y\|_\kappa \geq \|\mathbf{A}_X^{-1} \vec{H}_X\|_\kappa$, then we can always switch our nomenclature to refer to \vec{Y} as \vec{X} , and \vec{X} as \vec{Y} .

The proof of above theorem relies on the result of lemma 7 and the fact that the linear system described in (106) can always be reduced to the special case of the linear system used in lemma 7. We briefly outline the proof structure.

1. In the first part, we explicitly construct a joint distribution $Q_{MC}(M, \vec{X}, \vec{Y})$ having the Markovian structure $M \rightarrow \vec{X} \rightarrow \vec{Y}$.
2. In the second part, we show that the $Q_{MC}(M, \vec{X}, \vec{Y})$ constructed in the first part, lies in Δ_P . Therefore, we can then apply the result of proposition 1 to conclude $UI(M; \vec{Y} \setminus \vec{X}) = 0$.

Part 1: Specifying $Q_{MC}(M, \vec{X}, \vec{Y})$

We explicitly construct the desired Markov chain $M \rightarrow \vec{X} \rightarrow \vec{Y}$ by constructing a larger Markov chain $M \rightarrow \vec{X} \rightarrow \vec{X}' \rightarrow \vec{Y}' \rightarrow \vec{Y}$ and then marginalizing the larger Markov chain to obtain the desired Markov chain $M \rightarrow \vec{X} \rightarrow \vec{Y}$.

Denote the joint distribution of the Markov chain $M \rightarrow \vec{X} \rightarrow \vec{X}' \rightarrow \vec{Y}' \rightarrow \vec{Y}$ as $Q_{MC}(M, \vec{X}, \vec{X}', \vec{Y}', \vec{Y})$. We can decompose $Q_{MC}(M, \vec{X}, \vec{X}', \vec{Y}', \vec{Y})$ by utilizing its Markovian structure as follows:

$$Q_{MC}(M, \vec{X}, \vec{X}', \vec{Y}', \vec{Y}) = Q_{MC}(M)Q_{MC}(\vec{X}|M)Q_{MC}(\vec{X}'|\vec{X})Q_{MC}(\vec{Y}'|\vec{X}')Q_{MC}(\vec{Y}|\vec{Y}').$$

Consequently, we can specify/construct the distribution $Q_{MC}(M, \vec{X}, \vec{X}', \vec{Y}', \vec{Y})$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(\vec{X}|M)$, $Q_{MC}(\vec{X}'|\vec{X})$, $Q_{MC}(\vec{Y}'|\vec{X}')$, and $Q_{MC}(\vec{Y}|\vec{Y}')$. We specify $Q_{MC}(M, \vec{X}) = P(M, \vec{X})$ and $Q(\vec{X}'|\vec{X})$ using the following deterministic transformation:

$$\vec{X}' = \mathbf{A}_X^{-1} (\vec{X} - \vec{b}_X). \quad (107)$$

Substituting (106) in the above equation:

$$\vec{X}' = \mathbf{A}_X^{-1} (\vec{H}_X M + \mathbf{A}_X \vec{Z}_X + \vec{b}_X - \vec{b}_X) = \mathbf{A}_X^{-1} \vec{H}_X M + \vec{Z}_X. \quad (108)$$

Specifying $Q_{MC}(\vec{Y}'|\vec{X}')$: We define \vec{Y}' according to the following affine system:

$$\vec{Y}' = \mathbf{A}_Y^{-1} \vec{H}_Y M + \vec{Z}_Y. \quad (109)$$

Observe that both \vec{X}' and \vec{Y}' can be described as an affine function of M satisfying the special structure described in lemma 7 (see (115)). Furthermore, $Q_{MC}(M) = P(M)$ by construction. Consequently, $Q_{MC}(M)$ satisfies the properties required by the distribution of M outlined in lemma 7, as $P(M)$ in both lemma 7 and theorem 2 follow the same properties. Since $\|\mathbf{A}_Y^{-1} \vec{H}_Y\|_\kappa \leq \|\mathbf{A}_X^{-1} \vec{H}_X\|_\kappa$ by the assumption in the theorem, we can apply the result of lemma 7 to construct $Q_{MC}(\vec{Y}'|\vec{X}')$, where \vec{H}_X and \vec{H}_Y would be replaced by $\mathbf{A}_X^{-1} \vec{H}_X$ and $\mathbf{A}_Y^{-1} \vec{H}_Y$, respectively.

Lastly, we specify $Q_{MC}(\vec{Y}|\vec{Y}')$ through the following linear system:

$$\vec{Y} = \mathbf{A}_Y \vec{Y}' + \vec{b}_Y. \quad (110)$$

Finally, we construct the desired $Q_{MC}(M, \vec{X}, \vec{Y})$ from $Q_{MC}(M, \vec{X}, \vec{X}', \vec{Y}', \vec{Y})$ by marginalizing \vec{X}' and \vec{Y}' .

Part 2: Showing $Q_{MC}(M, \vec{X}, \vec{Y}) \in \Delta_P$

For showing that $Q_{MC}(M, \vec{X}, \vec{Y}) \in \Delta_P$, we first need to derive $Q_{MC}(\vec{Y}|M)$.

Deriving $Q(\vec{Y}|M)$: Substituting (109) in (110) we obtain:

$$\vec{Y} = \mathbf{A}_Y (\mathbf{A}_Y^{-1} \vec{H}_Y M + \vec{Z}_Y) + \vec{b}_Y = \vec{H}_Y M + \mathbf{A}_Y \vec{Z}_Y + \vec{b}_Y. \quad (111)$$

Inspecting the above equation, we can conclude:

$$Q_{MC}(\vec{Y}|M) = P(\vec{Y}|M). \quad (112)$$

From (112), we can conclude:

$$Q_{MC}(M, \vec{Y}) = Q_{MC}(M)Q_{MC}(\vec{Y}|M) = P(M)P(\vec{Y}|M) = P(M, \vec{Y}). \quad (113)$$

We know that by construction:

$$Q_{MC}(M, \vec{X}) = P(M, \vec{X}). \quad (114)$$

Hence, $Q_{MC}(M, \vec{X}, \vec{Y}) \in \Delta_P$ and consequently, by proposition 1, $UI(M; \vec{Y} \setminus \vec{X}) = 0$, concluding our proof. \square

Lemma 7. Let the joint distribution of M, \vec{X} and \vec{Y} , denoted as $P(M, \vec{X}, \vec{Y})$, describe the system 1 of the multivariate affine continuous stable system. Furthermore, assume $\mathbf{A}_X = \mathbf{I}_{d_X}$, $\mathbf{A}_Y = \mathbf{I}_{d_Y}$, $\vec{b}_X = \vec{0}_{d_X}$ and $\vec{b}_Y = \vec{0}_{d_Y}$, where \mathbf{I}_d is a $d \times d$ identity matrix and $\vec{0}_d$ is d -dimensional vector of zeros. Then,

1. If $\|\vec{\mathbf{H}}_Y\|_\kappa \leq \|\vec{\mathbf{H}}_X\|_\kappa$, where $\kappa = \alpha/\alpha-1 \forall \alpha \in (1, 2]$ and $\kappa = \infty \forall \alpha \in (0, 1]$, then Δ_P contains a Markov chain of the form $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{Y}}$ and $UI(M; \vec{\mathbf{Y}}|\vec{\mathbf{X}}) = 0$.
2. If $\|\vec{\mathbf{H}}_X\|_\kappa \leq \|\vec{\mathbf{H}}_Y\|_\kappa$, where $\kappa = \alpha/\alpha-1 \forall \alpha \in (1, 2]$ and $\kappa = \infty \forall \alpha \in (0, 1]$, then Δ_P contains a Markov chain of the form $M \rightarrow \vec{\mathbf{Y}} \rightarrow \vec{\mathbf{X}}$ and $UI(M; \vec{\mathbf{X}}|\vec{\mathbf{Y}}) = 0$.

Proof. We only provide an explicit proof of condition 1. Condition 2 is essentially the same as condition 1, with the parameters of X and Y switched. Consequently, the proof of condition 2 follows the same steps as the proof of condition 1, with parameters about X and Y switched.

Proof of condition 1:

We briefly outline the proof structure.

1. In the first part, we explicitly construct a joint distribution $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$ having the Markovian structure $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{Y}}$.
2. In the second part, we show that the $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$ constructed in the first part, lies in Δ_P . Therefore, we can then apply the result of proposition 1 to conclude $UI(M; \vec{\mathbf{Y}}|\vec{\mathbf{X}}) = 0$.

Part 1: Specifying $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$

We explicitly construct the desired Markov chain $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{Y}}$ by constructing a larger Markov chain $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{X}}' \rightarrow \vec{\mathbf{Y}}$ and then marginalizing the larger Markov chain to obtain the desired Markov chain $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{Y}}$.

Denote the joint distribution of the Markov chain $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{X}}' \rightarrow \vec{\mathbf{Y}}$ as $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{X}}', \vec{\mathbf{Y}})$. We can decompose $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{X}}', \vec{\mathbf{Y}})$ by utilizing its Markovian structure as follows:

$$Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{X}}', \vec{\mathbf{Y}}) = Q_{MC}(M)Q_{MC}(\vec{\mathbf{X}}|M)Q_{MC}(\vec{\mathbf{X}}'|\vec{\mathbf{X}})Q_{MC}(\vec{\mathbf{Y}}|\vec{\mathbf{X}}').$$

Consequently, we can specify/construct the distribution $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{X}}', \vec{\mathbf{Y}})$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(\vec{\mathbf{X}}|M)$, $Q_{MC}(\vec{\mathbf{X}}'|\vec{\mathbf{X}})$, and $Q_{MC}(\vec{\mathbf{Y}}|\vec{\mathbf{X}}')$.

Under the assumptions of lemma 7, we can simplify the linear system describing $P(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$ shown in (106) as follows:

$$\vec{\mathbf{X}} = \vec{\mathbf{H}}_X M + \vec{\mathbf{Z}}_X \text{ and } \vec{\mathbf{Y}} = \vec{\mathbf{H}}_Y M + \vec{\mathbf{Z}}_Y. \quad (115)$$

Correspondingly, we can write the characteristic functions of conditional distributions $P(\vec{\mathbf{X}}|M = m)$ and $P(\vec{\mathbf{Y}}|M = m)$ for a fixed m by employing lemma 9:

$$\begin{aligned} \mathbb{E}[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{X}}}|M=m] &= \exp \left(- \sum_{j=1}^n |t_j|^\alpha + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_X m \right), \\ \mathbb{E}[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{Y}}}|M=m] &= \exp \left(- \sum_{j=1}^n |t_j|^\alpha + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_Y m \right), \end{aligned} \quad (116)$$

where the characteristic function can be derived by realizing that for a given m , $\vec{\mathbf{H}}_X m$ and $\vec{\mathbf{H}}_Y m$ are constants being added to the random vectors $\vec{\mathbf{Z}}_X \sim p_{CS-IC}(\alpha, \vec{\mathbf{0}}_{d_x}, \vec{\mathbf{1}}_{d_x}, \vec{\mathbf{0}}_{d_x})$ and $\vec{\mathbf{Z}}_Y \sim p_{CS-IC}(\alpha, \vec{\mathbf{0}}_{d_y}, \vec{\mathbf{1}}_{d_y}, \vec{\mathbf{0}}_{d_y})$, respectively, and then using the result of lemma 9.

Specifying $Q_{MC}(M)$ and $Q_{MC}(\vec{\mathbf{X}}|M)$: We specify $Q_{MC}(M)$ and $Q_{MC}(\vec{\mathbf{X}}|M)$ as follows:

$$Q_{MC}(M) = P(M) \text{ and } Q_{MC}(\vec{\mathbf{X}}|M) = P(\vec{\mathbf{X}}|M), \quad (117)$$

where $P(M)$ and $P(\vec{\mathbf{X}}|M)$ are marginal distributions derived from the original joint distribution $P(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$ (discussed in the lemma statement) over which the bivariate PID is being calculated.

Specifying $Q_{MC}(\vec{\mathbf{X}}'|\vec{\mathbf{X}})$: We specify $Q_{MC}(\vec{\mathbf{X}}'|\vec{\mathbf{X}})$ through the following deterministic transformation:

$$\vec{\mathbf{X}}' = \vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T \vec{\mathbf{X}}, \quad (118)$$

where $\vec{\mathbf{H}}_X^k$ is as defined in lemma 10, and $k \in [0, \infty]$.

We will also derive the distribution $Q(\vec{\mathbf{X}}'|M)$ before proceeding with our construction, as we will need it later in the proof to show that the constructed $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}}) \in \Delta_P$.

Deriving $Q_{MC}(\vec{\mathbf{X}}'|M)$: To derive $Q_{MC}(\vec{\mathbf{X}}'|M)$, we represent $\vec{\mathbf{X}}'$ as function of M and $\vec{\mathbf{Z}}_X$ in a similar manner to $\vec{\mathbf{X}}$ in (115), and then use the result of lemma 9 to derive its conditional characteristic function $\mathbb{E}[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{X}}'}|M]$, and correspondingly the desired $Q_{MC}(\vec{\mathbf{X}}'|M)$.

We multiply $\vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T$ on both sides of equality in (115) to obtain the following linear equation:

$$\vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T \vec{\mathbf{X}} = \vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T (\vec{\mathbf{H}}_X M + \vec{\mathbf{Z}}_X) = \vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T \vec{\mathbf{H}}_X M + \vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T \vec{\mathbf{Z}}_X.$$

Substituting (118) in the above equation, we obtain:

$$\vec{\mathbf{X}}' = \vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T \vec{\mathbf{X}} = \vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T \vec{\mathbf{H}}_X M + \vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T \vec{\mathbf{Z}}_X. \quad (119)$$

Using the result of lemma 10, i.e., $\vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T \vec{\mathbf{H}}_X = \vec{\mathbf{H}}_Y$, in the above equation we obtain:

$$\vec{\mathbf{X}}' = \vec{\mathbf{H}}_Y M + \vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T \vec{\mathbf{Z}}_X. \quad (120)$$

From (120), we can observe that $\vec{\mathbf{X}}'$ is an affine function of $\vec{\mathbf{Z}}_X$ for a fixed m . Hence, we can employ the result of lemma 9 to derive the distribution of $Q_{MC}(\vec{\mathbf{X}}'|M)$ (by deriving its corresponding characteristic function).

Note that the j -th column of $\vec{\mathbf{H}}_Y(\vec{\mathbf{H}}_X^k)^T$ can be expressed as $h'_j \vec{\mathbf{H}}_Y$, where h'_j is the j -th component of $\vec{\mathbf{H}}_X^k$. Consequently:

$$\begin{aligned} \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{X}}'} \middle| M = m \right] &= \exp \left(- \sum_{j=1}^{d_X} |h'_j \vec{\mathbf{H}}_Y^T \vec{\mathbf{t}}|^\alpha + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_Y m \right) \\ &= \exp \left(- \sum_{j=1}^{d_X} |h'_j|^\alpha |\vec{\mathbf{H}}_Y^T \vec{\mathbf{t}}|^\alpha + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_Y m \right) \\ &= \exp \left(- |\vec{\mathbf{H}}_Y^T \vec{\mathbf{t}}|^\alpha \left(\sum_{j=1}^{d_X} |h'_j|^\alpha \right) + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_Y m \right). \end{aligned} \quad (121)$$

We substitute the value h'_j in (121) using the definition of $\vec{\mathbf{H}}_X^k$ given in lemma 10. We will divide the substitution into two cases:

Case 1: k is finite, i.e., $k \in [0, \infty)$. Then, we have $h'_j = |h_j^X|^k / \|\vec{\mathbf{H}}_X\|_{1+k}^{1+k}$, which implies:

$$\begin{aligned} \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{X}}'} \middle| M = m \right] &= \exp \left(- |\vec{\mathbf{H}}_Y^T \vec{\mathbf{t}}|^\alpha \sum_{j=1}^{d_X} \frac{|h_j^X|^{k\alpha}}{\|\vec{\mathbf{H}}_X\|_{1+k}^{\alpha+k\alpha}} + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_Y m \right) \\ &= \exp \left(- \frac{|\vec{\mathbf{H}}_Y^T \vec{\mathbf{t}}|^\alpha}{\|\vec{\mathbf{H}}_X\|_{1+k}^{\alpha+k\alpha}} \sum_{j=1}^{d_X} |h_j^X|^{k\alpha} + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_Y m \right) \\ &= \exp \left(- \frac{|\vec{\mathbf{H}}_Y^T \vec{\mathbf{t}}|^\alpha \|\vec{\mathbf{H}}_X\|_{k\alpha}^{k\alpha}}{\|\vec{\mathbf{H}}_X\|_{1+k}^{\alpha+k\alpha}} + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_Y m \right). \end{aligned} \quad (122)$$

Case 2: k is infinite, i.e., $k = \infty$. Then, we have $h'_j = 0 \forall j \neq j^*$, and $h'_{j^*} = 1/h_1^X$, which implies:

$$\begin{aligned}\mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{X}}'} \middle| M = m \right] &= \exp \left(- \left| \vec{\mathbf{H}}_Y^T \vec{\mathbf{t}} \right|^\alpha \left| \frac{1}{h_{j^*}^X} \right|^\alpha + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_Y m \right) \\ &= \exp \left(- \frac{\left| \vec{\mathbf{H}}_Y^T \vec{\mathbf{t}} \right|^\alpha}{\left\| \vec{\mathbf{H}}_X \right\|_\infty^\alpha} + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_Y m \right).\end{aligned}\quad (123)$$

Combining (122) and (123), we can re-write $\mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{X}}'} \middle| M = m \right]$ as:

$$\begin{aligned}\mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{X}}'} \middle| M = m \right] &= \exp \left(-f(k) \left| \vec{\mathbf{H}}_Y^T \vec{\mathbf{t}} \right|^\alpha + i\vec{\mathbf{t}}^T \vec{\mathbf{H}}_Y m \right), \text{ where} \\ f(k) &= \begin{cases} \left\| \vec{\mathbf{H}}_X \right\|_{k\alpha}^{k\alpha} / \left\| \vec{\mathbf{H}}_X \right\|_{1+k}^{\alpha+\alpha k} & k \text{ finite} \\ 1/\left\| \vec{\mathbf{H}}_X \right\|_\infty^\alpha & k \text{ infinite} \end{cases}.\end{aligned}\quad (124)$$

Specifying $Q_{MC}(\vec{\mathbf{Y}}|\vec{\mathbf{X}}')$: We specify $Q_{MC}(\vec{\mathbf{Y}}|\vec{\mathbf{X}}')$ through the following stochastic transformation:

$$\vec{\mathbf{Y}} = \vec{\mathbf{X}}' + \vec{\epsilon}, \quad (125)$$

where $\vec{\epsilon}$ follows a multivariate stable distribution. Furthermore, we assume that $\vec{\epsilon}$ is jointly independent from $(\vec{\mathbf{X}}, \vec{\mathbf{X}}', M)$, i.e., $\vec{\epsilon} \perp\!\!\!\perp (\vec{\mathbf{X}}, \vec{\mathbf{X}}', M)$. The characteristic function of $\vec{\epsilon}$ is defined as follows:

$$\mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\epsilon}} \right] = \exp \left(- \sum_{j=1}^{d_Y} |t_j|^\alpha + f(k) \left| \vec{\mathbf{H}}_Y^T \vec{\mathbf{t}} \right|^\alpha \right). \quad (126)$$

In order for $Q_{MC}(\vec{\mathbf{Y}}|\vec{\mathbf{X}}')$ to be a valid distribution, we need to ensure that the distribution of $\vec{\epsilon}$ defined in (126) is a legitimate multivariate stable distribution. We will utilize property 1 for showing that $\vec{\epsilon}$ follows a stable multivariate stable distribution. Property 1 states a random vector is distributed according to an independent component multivariate stable distribution iff every 1-dimensional projection of this random vector follows a univariate stable continuous distribution. Hence, to show that $\vec{\epsilon}$ is distributed according to a legitimate multivariate stable distribution, we will show that every 1-dimensional projection of $\vec{\epsilon}$ follows a univariate stable continuous distribution.

Let $\vec{\mathbf{t}} \in \mathbb{R}^{d_Y}$, then the characteristic function of the 1-dimensional projection of $\vec{\epsilon}$ along $\vec{\mathbf{t}}$, i.e., $\vec{\mathbf{t}}^T \vec{\epsilon}$ can be trivially deduced using (126).

$$\begin{aligned}\mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\epsilon}} \right] &= \exp \left(- \sum_{j=1}^{d_Y} |t_j|^\alpha + f(k) \left| \vec{\mathbf{H}}_Y^T \vec{\mathbf{t}} \right|^\alpha \right) \\ &= \exp \left(- \left\| \vec{\mathbf{t}} \right\|_\alpha^\alpha + f(k) \left| \vec{\mathbf{H}}_Y^T \vec{\mathbf{t}} \right|^\alpha \right).\end{aligned}\quad (127)$$

Comparing the characteristic function of $\vec{\mathbf{t}}^T \vec{\epsilon}$ described in (127) with the standard characteristic function of a univariate stable characteristic function defined in (323), we can conclude $\vec{\mathbf{t}}^T \vec{\epsilon} \sim p_{CS}(\alpha, \beta(\vec{\mathbf{t}}), \gamma(\vec{\mathbf{t}}), \mu(\vec{\mathbf{t}}))$, where:

$$\gamma(\vec{\mathbf{t}}) = \left\| \vec{\mathbf{t}} \right\|_\alpha^\alpha - f(k) \left| \vec{\mathbf{H}}_Y^T \vec{\mathbf{t}} \right|^\alpha, \mu(\vec{\mathbf{t}}) = 0, \text{ and } \beta(\vec{\mathbf{t}}) = 0. \quad (128)$$

In order for $\vec{\mathbf{t}}^T \vec{\epsilon}$ to have a legitimate univariate stable distribution, we just need to show that $\gamma(\vec{\mathbf{t}}) \geq 0 \forall \vec{\mathbf{t}} \in \mathbb{R}^{d_Y}$. Note that $\gamma(\vec{\mathbf{t}}) = 0$ would correspond to the case where all the mass of the distribution is centered at 0, which does satisfy definition 2, and is an example of a degenerate univariate stable distribution. Let us analyze the function $\gamma(\vec{\mathbf{t}})$:

$$\gamma(\vec{\mathbf{t}}) = \left\| \vec{\mathbf{t}} \right\|_\alpha^\alpha - f(k) \left| \vec{\mathbf{H}}_Y^T \vec{\mathbf{t}} \right|^\alpha, \quad (129)$$

Applying lemma 8 on the term $\|\vec{\mathbf{H}}_Y^T \vec{\mathbf{t}}\|$, we obtain:

$$\|\vec{\mathbf{H}}_Y^T \vec{\mathbf{t}}\| \leq \begin{cases} \|\vec{\mathbf{t}}\|_\alpha^\alpha \|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha & \alpha \in (1, 2] \\ \|\vec{\mathbf{t}}\|_1^\alpha \|\vec{\mathbf{H}}_Y\|_\infty^\alpha & \alpha \in (0, 1] \end{cases}. \quad (130)$$

Applying (130) in (129), we obtain:

$$\gamma(\vec{\mathbf{t}}) \geq \begin{cases} \|\vec{\mathbf{t}}\|_\alpha^\alpha - f(k) \|\vec{\mathbf{t}}\|_\alpha^\alpha \|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha & \alpha \in (1, 2] \\ \|\vec{\mathbf{t}}\|_\alpha^\alpha - f(k) \|\vec{\mathbf{t}}\|_1^\alpha \|\vec{\mathbf{H}}_Y\|_\infty^\alpha & \alpha \in (0, 1] \end{cases}. \quad (131)$$

Denote $l(\vec{\mathbf{t}}, \alpha)$ as follows:

$$l(\vec{\mathbf{t}}, \alpha) = \begin{cases} \|\vec{\mathbf{t}}\|_\alpha^\alpha - f(k) \|\vec{\mathbf{t}}\|_\alpha^\alpha \|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha & \alpha \in (1, 2] \\ \|\vec{\mathbf{t}}\|_\alpha^\alpha - f(k) \|\vec{\mathbf{t}}\|_1^\alpha \|\vec{\mathbf{H}}_Y\|_\infty^\alpha & \alpha \in (0, 1] \end{cases}. \quad (132)$$

In order to show $\gamma(\vec{\mathbf{t}}) \geq 0 \forall \vec{\mathbf{t}} \in \mathbb{R}^{d_Y}$, it suffices to show that $l(\vec{\mathbf{t}}, \alpha) \geq 0 \forall \vec{\mathbf{t}} \in \mathbb{R}^{d_Y}$ and $\alpha \in (0, 2]$.

Case 1: $\alpha \in (1, 2]$.

For $\alpha \in (1, 2]$, we have:

$$l(\vec{\mathbf{t}}, \alpha) = \|\vec{\mathbf{t}}\|_\alpha^\alpha - f(k) \|\vec{\mathbf{t}}\|_\alpha^\alpha \|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha = \|\vec{\mathbf{t}}\|_\alpha^\alpha \left(1 - f(k) \|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha\right). \quad (133)$$

Since, the above equation holds for any $k \in [0, \infty]$, we choose $k = 1/(\alpha-1)$. Substituting the form $f(k)$ from (124) in the above equation, we obtain:

$$\begin{aligned} l(\vec{\mathbf{t}}, \alpha) &= \|\vec{\mathbf{t}}\|_\alpha^\alpha \left(1 - \frac{\|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha \|\vec{\mathbf{H}}_X\|_{\alpha/\alpha-1}^{\alpha/\alpha-1}}{\|\vec{\mathbf{H}}_X\|_{1+\frac{1}{\alpha-1}}^{\alpha(1+\frac{1}{\alpha-1})}}\right) = \|\vec{\mathbf{t}}\|_\alpha^\alpha \left(1 - \frac{\|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha \|\vec{\mathbf{H}}_X\|_{\alpha/\alpha-1}^{\alpha/\alpha-1}}{\|\vec{\mathbf{H}}_X\|_{\alpha/\alpha-1}^{\alpha(\frac{\alpha}{\alpha-1})}}\right) \\ &= \|\vec{\mathbf{t}}\|_\alpha^\alpha \left(1 - \frac{\|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha}{\|\vec{\mathbf{H}}_X\|_{\alpha/\alpha-1}^{\alpha(\frac{\alpha}{\alpha-1}) - \frac{\alpha}{\alpha-1}}}\right) = \|\vec{\mathbf{t}}\|_\alpha^\alpha \left(1 - \frac{\|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha}{\|\vec{\mathbf{H}}_X\|_{\alpha/\alpha-1}^\alpha}\right). \end{aligned} \quad (134)$$

Note that $\|\vec{\mathbf{t}}\|_\alpha^\alpha \geq 0$, and by condition 1 we know that $\|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1} \leq \|\vec{\mathbf{H}}_X\|_{\alpha/\alpha-1} \Rightarrow \|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha \leq \|\vec{\mathbf{H}}_X\|_{\alpha/\alpha-1}^\alpha$. Hence, using these two previous facts and the above equation, we can conclude:

$$l(\vec{\mathbf{t}}, \alpha) = \underbrace{\|\vec{\mathbf{t}}\|_\alpha^\alpha}_{\geq 0} \underbrace{\left(1 - \frac{\|\vec{\mathbf{H}}_Y\|_{\alpha/\alpha-1}^\alpha}{\|\vec{\mathbf{H}}_X\|_{\alpha/\alpha-1}^\alpha}\right)}_{\geq 0} \Rightarrow l(\vec{\mathbf{t}}, \alpha) \geq 0. \quad (135)$$

Case 2: $\alpha \in (0, 1]$.

$$l(\vec{\mathbf{t}}, \alpha) = \|\vec{\mathbf{t}}\|_\alpha^\alpha - f(k) \|\vec{\mathbf{t}}\|_1^\alpha \|\vec{\mathbf{H}}_Y\|_\infty^\alpha \stackrel{(a)}{\geq} \|\vec{\mathbf{t}}\|_\alpha^\alpha - f(k) \|\vec{\mathbf{t}}\|_\alpha^\alpha \|\vec{\mathbf{H}}_Y\|_\infty^\alpha, \quad (136)$$

where (a) is derived using the monotonicity of L_p norms in p , see [57]. More precisely, we use the inequality $\|\vec{\mathbf{t}}\|_\alpha \geq \|\vec{\mathbf{t}}\|_1 \forall \alpha \in (0, 1]$. Simplifying the above equation, we have:

$$l(\vec{\mathbf{t}}, \alpha) \geq \|\vec{\mathbf{t}}\|_\alpha^\alpha \left(1 - f(k) \|\vec{\mathbf{H}}_Y\|_\infty^\alpha\right). \quad (137)$$

Since, the above equation holds for any $k \in [0, \infty]$, we choose $k = \infty$. Substituting the form $f(k)$ from (124) in the above equation, we obtain:

$$l(\vec{t}, \alpha) \geq \left\| \vec{t} \right\|_{\alpha}^{\alpha} \left(1 - \frac{\|\vec{H}_Y\|_{\infty}^{\alpha}}{\|\vec{H}_X\|_{\infty}^{\alpha}} \right). \quad (138)$$

Note that $\left\| \vec{t} \right\|_{\alpha}^{\alpha} \geq 0$, and by condition 1 we know that $\|\vec{H}_Y\|_{\infty} \leq \|\vec{H}_X\|_{\infty} \Rightarrow \|\vec{H}_Y\|_{\infty}^{\alpha} \leq \|\vec{H}_X\|_{\infty}^{\alpha}$. Hence, using these two facts and the above equation, we can conclude:

$$l(\vec{t}, \alpha) \geq \underbrace{\left\| \vec{t} \right\|_{\alpha}^{\alpha}}_{\geq 0} \underbrace{\left(1 - \frac{\|\vec{H}_Y\|_{\infty}^{\alpha}}{\|\vec{H}_X\|_{\infty}^{\alpha}} \right)}_{\geq 0} \Rightarrow l(\vec{t}, \alpha) \geq 0. \quad (139)$$

From (135) and (139), we can conclude that $l(\vec{t}, \alpha) \geq 0 \forall \vec{t} \in \mathbb{R}^{d_Y}$ and $\alpha \in (0, 2]$, which implies $\gamma(\vec{t}) \geq 0 \forall \vec{t} \in \mathbb{R}^{d_Y}$. Hence, $\vec{\epsilon}$ follows a legitimate multivariate stable distribution.

Finally, we construct the desired $Q_{MC}(M, \vec{X}, \vec{Y})$ from $Q_{MC}(M, \vec{X}, \vec{X}', \vec{Y})$ by marginalizing \vec{X}' .

Part 2: Showing $Q_{MC}(M, \vec{X}, \vec{Y}) \in \Delta_P$

For showing that $Q_{MC}(M, \vec{X}, \vec{Y}) \in \Delta_P$, we first need to derive $Q_{MC}(\vec{Y}|M)$.

Deriving $Q(\vec{Y}|M)$: We will derive $Q_{MC}(\vec{Y}|M)$ by using the fact that $\vec{\epsilon} \perp\!\!\!\perp \vec{X}'|M$, hence their conditional characteristic functions would just result in multiplication.

$$\begin{aligned} \mathbb{E} \left[e^{i\vec{t}^T \vec{Y}} \middle| M = m \right] &= \mathbb{E} \left[e^{i\vec{t}^T (\vec{X}' + \vec{\epsilon})} \middle| M = m \right] = \mathbb{E} \left[e^{i\vec{t}^T \vec{X}' + i\vec{t}^T \vec{\epsilon}} \middle| M = m \right] \\ &= \mathbb{E} \left[e^{i\vec{t}^T \vec{X}'} e^{i\vec{t}^T \vec{\epsilon}} \middle| M = m \right] \stackrel{(a)}{=} \mathbb{E} \left[e^{i\vec{t}^T \vec{X}'} \middle| M = m \right] \mathbb{E} \left[e^{i\vec{t}^T \vec{\epsilon}} \middle| M = m \right], \end{aligned} \quad (140)$$

where (a) is due to $\vec{\epsilon} \perp\!\!\!\perp \vec{X}'|M$. Substituting the characteristic functions of \vec{X}' and $\vec{\epsilon}$ from (124) and (126), respectively, in (140):

$$\begin{aligned} &= \exp \left(- \sum_{j=1}^{d_Y} |t_j|^{\alpha} + f(k) \left| \vec{H}_Y^T \vec{t} \right|^{\alpha} \right) \exp \left(-f(k) \left| \vec{H}_Y^T \vec{t} \right|^{\alpha} + i\vec{t}^T \vec{H}_Y m \right) \\ &= \exp \left(- \sum_{j=1}^{d_Y} |t_j|^{\alpha} + f(k) \left| \vec{H}_Y^T \vec{t} \right|^{\alpha} - f(k) \left| \vec{H}_Y^T \vec{t} \right|^{\alpha} + i\vec{t}^T \vec{H}_Y m \right) \\ &= \exp \left(- \sum_{j=1}^{d_Y} |t_j|^{\alpha} + i\vec{t}^T \vec{H}_Y m \right). \end{aligned} \quad (141)$$

By comparing the characteristic function shown in (141) with the characteristic function derived using $P(\vec{Y}|M)$ shown in (116), we can conclude:

$$Q_{MC}(\vec{Y}|M) = P(\vec{Y}|M). \quad (142)$$

From (142) and (117), we can conclude:

$$Q_{MC}(M, \vec{Y}) = Q_{MC}(M)Q_{MC}(\vec{Y}|M) = P(M)P(\vec{Y}|M) = P(M, \vec{Y}). \quad (143)$$

We know that by construction:

$$Q_{MC}(M, \vec{X}) = P(M, \vec{X}). \quad (144)$$

Hence, $Q_{MC}(M, \vec{X}, \vec{Y}) \in \Delta_P$ and consequently, by proposition 1, $UI(M; \vec{Y} \setminus \vec{X}) = 0$, concluding our proof. \square

Lemma 8. Let $\vec{\mathbf{a}} = [a_1 \ \dots \ a_d] \in \mathbb{R}^d$ and $\vec{\mathbf{b}} = [b_1 \ \dots \ b_d] \in \mathbb{R}^d$, then we have:

$$\left| \vec{\mathbf{a}}^T \vec{\mathbf{b}} \right|^\alpha \leq \begin{cases} \|\vec{\mathbf{a}}\|_\alpha^\alpha \|\vec{\mathbf{b}}\|_{\alpha/(\alpha-1)}^\alpha & \alpha \in (1, 2] \\ \|\vec{\mathbf{a}}\|_1^\alpha \|\vec{\mathbf{b}}\|_\infty^\alpha & \alpha \in (0, 1] \end{cases}, \quad (145)$$

where $\|(\cdot)\|_p$ is the standard L_p norm.

Proof. **Case 1:** $\alpha \in (1, 2]$

Analysing the LHS of (145)

$$\left| \vec{\mathbf{a}}^T \vec{\mathbf{b}} \right| = \left| \sum_{i=1}^d a_i b_i \right| \stackrel{(a)}{\leq} \sum_{i=1}^d |a_i b_i| \stackrel{(b)}{\leq} \left(\sum_{i=1}^d |a_i|^\alpha \right)^{\frac{1}{\alpha}} \left(\sum_{i=1}^d |b_i|^{\frac{\alpha}{\alpha-1}} \right)^{\frac{(\alpha-1)}{\alpha}} = \|\vec{\mathbf{a}}\|_\alpha \|\vec{\mathbf{b}}\|_{\alpha/(\alpha-1)}, \quad (146)$$

where (a) is due to the sub-additivity of $|\cdot|$ operator, and (b) is due to Hölder's inequality [57]. Raising both sides of the inequality to α in (146) gives us the desired inequality. Since, $\alpha > 0$ raising both sides of the inequalities to power α does not change the direction of inequality.

$$\left| \vec{\mathbf{a}}^T \vec{\mathbf{b}} \right|^\alpha \leq \|\vec{\mathbf{a}}\|_\alpha^\alpha \|\vec{\mathbf{b}}\|_{\alpha/(\alpha-1)}^\alpha. \quad (147)$$

Case 2: $\alpha \in (0, 1]$

Analysing the LHS of (145)

$$\left| \vec{\mathbf{a}}^T \vec{\mathbf{b}} \right| = \left| \sum_{i=1}^d a_i b_i \right| \stackrel{(a)}{\leq} \sum_{i=1}^d |a_i b_i| \stackrel{(b)}{\leq} \left(\sum_{i=1}^d |a_i| \right) \max_{i \in \{1, \dots, d\}} |b_i| = \|\vec{\mathbf{a}}\|_1 \|\vec{\mathbf{b}}\|_\infty, \quad (148)$$

where (a) is again due to the sub-additivity of $|\cdot|$ operator, and (b) is due to the fact that $|a_i b_i| \leq |a_i| \max_{i \in \{1, \dots, d\}} |b_i| \ \forall \ i \in \{1, \dots, d\}$. Raising both sides of the inequality to α in (148) gives us the desired inequality. Since, $\alpha > 0$ raising both sides of the inequalities to power α does not change the direction of inequality.

$$\left| \vec{\mathbf{a}}^T \vec{\mathbf{b}} \right|^\alpha \leq \|\vec{\mathbf{a}}\|_1^\alpha \|\vec{\mathbf{b}}\|_\infty^\alpha. \quad (149)$$

Combining (147) and (149), we obtain the desired inequality. \square

Lemma 9. Let $\vec{\mathbf{Z}} \sim p_{CS-IC}(\alpha, \vec{\mathbf{0}}_d, \vec{\mathbf{1}}_d, \vec{\mathbf{0}}_d)$ be a d -dimensional random vector; $\mathbf{A} = [a_{ij}]_{i,j=1}^{n,d} \in \mathbb{R}^{n \times d}$ be a $n \times d$ matrix, and $\vec{\mathbf{b}} \in \mathbb{R}^d$. Then, the characteristic function of $\mathbf{A}\vec{\mathbf{Z}} + \vec{\mathbf{b}}$ is given by (150)

$$\mathbb{E} \left[e^{i\vec{\mathbf{t}}^T (\mathbf{A}\vec{\mathbf{Z}} + \vec{\mathbf{b}})} \right] = \exp \left(- \sum_{j=1}^d \left| \vec{\mathbf{a}}_j^T \vec{\mathbf{t}} \right|^\alpha + i\vec{\mathbf{t}}^T \vec{\mathbf{b}} \right) \ \forall \ \vec{\mathbf{t}} \in \mathbb{R}^d, \quad (150)$$

where $\vec{\mathbf{a}}_i$ is the i -th column of \mathbf{A} .

Proof. Analyzing L.H.S of (150)

$$\mathbb{E} \left[e^{i\vec{\mathbf{t}}^T (\mathbf{A}\vec{\mathbf{Z}} + \vec{\mathbf{b}})} \right] = \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \mathbf{A}\vec{\mathbf{Z}} + i\vec{\mathbf{t}}^T \vec{\mathbf{b}}} \right] = \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \mathbf{A}\vec{\mathbf{Z}}} e^{i\vec{\mathbf{t}}^T \vec{\mathbf{b}}} \right] \quad (151)$$

Using the fact that $e^{i\vec{\mathbf{t}}^T \vec{\mathbf{b}}}$ is constant w.r.t expectation over $\vec{\mathbf{Z}}$, and linearity of expectation, we have:

$$\begin{aligned} \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T (\mathbf{A}\vec{\mathbf{Z}} + \vec{\mathbf{b}})} \right] &= e^{i\vec{\mathbf{t}}^T \vec{\mathbf{b}}} \mathbb{E} \left[\exp \left(i\vec{\mathbf{t}}^T \mathbf{A}\vec{\mathbf{Z}} \right) \right], \\ &= e^{i\vec{\mathbf{t}}^T \vec{\mathbf{b}}} \mathbb{E} \left[\exp \left(i\vec{\mathbf{t}}^T [\vec{\mathbf{a}}_1 \ \dots \ \vec{\mathbf{a}}_d] \vec{\mathbf{Z}} \right) \right], \\ &= e^{i\vec{\mathbf{t}}^T \vec{\mathbf{b}}} \mathbb{E} \left[\exp \left(i [\vec{\mathbf{t}}^T \vec{\mathbf{a}}_1 \ \dots \ \vec{\mathbf{t}}^T \vec{\mathbf{a}}_d] \vec{\mathbf{Z}} \right) \right], \end{aligned}$$

Simplifying the term $[\vec{\mathbf{t}}^T \vec{\mathbf{a}}_1 \quad \dots \quad \vec{\mathbf{t}}^T \vec{\mathbf{a}}_d] \vec{\mathbf{Z}}$ by substituting $\mathbf{Z} = [Z_1 \quad \dots \quad Z_d]^T$, we obtain:

$$= e^{i\vec{\mathbf{t}}^T \vec{\mathbf{b}}} \mathbb{E} \left[\exp \left(i \sum_{j=1}^d (\vec{\mathbf{t}}^T \vec{\mathbf{a}}_j) Z_j \right) \right],$$

Using properties of exponentials, we obtain:

$$= e^{i\vec{\mathbf{t}}^T \vec{\mathbf{b}}} \mathbb{E} \left[\prod_{j=1}^d \exp \left(i\vec{\mathbf{t}}^T \vec{\mathbf{a}}_j Z_j \right) \right],$$

Using the fact $\{Z_j\}_{j=1}^d$ are jointly independent we obtain:

$$= e^{i\vec{\mathbf{t}}^T \vec{\mathbf{b}}} \prod_{j=1}^d \mathbb{E} \left[\exp \left(i\vec{\mathbf{t}}^T \vec{\mathbf{a}}_j Z_j \right) \right],$$

Using the fact that $Z_j \sim p_{CS}(\alpha, 0, 1, 0)$ and the result of lemma 5, we know that $\vec{\mathbf{t}}^T \vec{\mathbf{a}}_j Z_j \sim p_{CS}(\alpha, 0, |\vec{\mathbf{t}}^T \vec{\mathbf{a}}_j|, 0)$. Substituting the corresponding characteristic function of $\vec{\mathbf{t}}^T \vec{\mathbf{a}}_j Z_j$ using (323)

$$\begin{aligned} &= e^{i\vec{\mathbf{t}}^T \vec{\mathbf{b}}} \prod_{j=1}^d \exp \left(-|\vec{\mathbf{t}}^T \vec{\mathbf{a}}_j|^\alpha \right) = e^{i\vec{\mathbf{t}}^T \vec{\mathbf{b}}} \exp \left(-\sum_{j=1}^d |\vec{\mathbf{t}}^T \vec{\mathbf{a}}_j|^\alpha \right) \\ &= \exp \left(-\sum_{j=1}^d |\vec{\mathbf{t}}^T \vec{\mathbf{a}}_j|^\alpha + i\vec{\mathbf{t}}^T \vec{\mathbf{b}} \right). \end{aligned} \quad (152)$$

□

Lemma 10. Let $\vec{\mathbf{H}}_1 = [h_1^{(1)} \quad \dots \quad h_{d_1}^{(1)}]^T \in \mathbb{R}^{d_1}$ and $\vec{\mathbf{H}}_2 = [h_1^{(2)} \quad \dots \quad h_{d_2}^{(2)}]^T \in \mathbb{R}^{d_2}$. Define $\vec{\mathbf{H}}_1^k$ as follows:

$$\vec{\mathbf{H}}_1^k = \frac{1}{\|\vec{\mathbf{H}}_1\|_{1+k}^{1+k}} \left[\text{sgn}(h_1^{(1)}) |h_1^{(1)}|^k \quad \dots \quad \text{sgn}(h_{d_1}^{(1)}) |h_{d_1}^{(1)}|^k \right]^T \quad \forall k \in [0, \infty).$$

Furthermore, define $\vec{\mathbf{H}}_1^\infty$ as follows:

$$\begin{aligned} j^* &= \arg \max_{j \in \{1, \dots, d_1\}} |h_j^{(1)}|, \\ \vec{\mathbf{H}}_1^\infty &= [h_1^\infty \quad \dots \quad h_{d_1}^\infty]^T, \text{ with } h_j^\infty = 0 \quad \forall j \in \{1, \dots, d_1\} \setminus \{j^*\}, \text{ and } h_{j^*}^\infty = \frac{1}{h_{j^*}^{(1)}}. \end{aligned} \quad (153)$$

Then, we have:

$$\vec{\mathbf{H}}_2 (\vec{\mathbf{H}}_1^k)^T \vec{\mathbf{H}}_1 = \vec{\mathbf{H}}_2 \quad \forall k \in [0, \infty) \text{ and } \vec{\mathbf{H}}_2 (\vec{\mathbf{H}}_1^\infty)^T \vec{\mathbf{H}}_1 = \vec{\mathbf{H}}_2, \quad (154)$$

where $\|(\cdot)\|_{1+k}$ is the standard L_{1+k} norm [57]. We can alternatively state (154) in a more compact notation:

$$\vec{\mathbf{H}}_2 (\vec{\mathbf{H}}_1^k)^T \vec{\mathbf{H}}_1 = \vec{\mathbf{H}}_2 \quad \forall k \in [0, \infty], \quad (155)$$

where the case of $k = \infty$ is to be understood as $\vec{\mathbf{H}}_1^k = \vec{\mathbf{H}}_1^\infty$.

Proof. **Case 1:** $k \in [0, \infty)$, i.e., k is finite,

$$\begin{aligned} \vec{\mathbf{H}}_2 (\vec{\mathbf{H}}_1^k)^T \vec{\mathbf{H}}_1 &= \frac{1}{\|\vec{\mathbf{H}}_1\|_{1+k}^{1+k}} \vec{\mathbf{H}}_2 \left[\text{sgn}(h_1^{(1)}) |h_1^{(1)}|^k \quad \dots \quad \text{sgn}(h_{d_1}^{(1)}) |h_{d_1}^{(1)}|^k \right] \begin{bmatrix} h_1^{(1)} \\ \vdots \\ h_{d_1}^{(1)} \end{bmatrix} \\ &= \frac{1}{\|\vec{\mathbf{H}}_1\|_{1+k}^{1+k}} \underbrace{\left(\sum_{j=1}^d |h_j^{(1)}|^{1+k} \right)}_{=\|\vec{\mathbf{H}}_1\|_{1+k}^{1+k}} \vec{\mathbf{H}}_2 = \vec{\mathbf{H}}_2. \end{aligned}$$

Case 2: $k = \infty$,

$$\vec{\mathbf{H}}_2(\vec{\mathbf{H}}_1^\infty)^T \vec{\mathbf{H}}_1 = \vec{\mathbf{H}}_2 \left(\sum_{j=1}^{d_1} h_j^\infty h_j^{(1)} \right) \stackrel{(a)}{=} \vec{\mathbf{H}}_2 \left(h_{j^*}^\infty h_{j^*}^{(1)} \right) \stackrel{(b)}{=} \vec{\mathbf{H}}_2 \frac{1}{h_{j^*}^{(1)}} h_{j^*}^{(1)} = \vec{\mathbf{H}}_2,$$

where (a) is due to the fact that all elements of $\vec{\mathbf{H}}_1^\infty$ are 0 except the j^* element, and (b) is due to the fact that $h_{j^*}^\infty = 1/h_{j^*}^{(1)}$. \square

H Proof of Theorem 3

In this section, we provide the proof of theorem 3. Since theorem 3 provides the analytical PID terms for the system 2 of the multivariate affine continuous stable system defined in Sec. 4, we briefly restate certain key properties of the corresponding elliptically-contoured multivariate stable distributions and the corresponding multivariate affine continuous stable system for convenience.

H.1 Elliptically-contoured multivariate stable distribution

Elliptically-contoured multivariate stable distributions are another example of a multivariate generalization of univariate continuous stable distributions. As the name suggests, the defining features of these distributions is that the corresponding p.d.f. has elliptical contours, similar to the multivariate Gaussian distribution. We refer the reader to Appx. N.3 for more details on multivariate continuous stable distributions. We now list certain key properties of elliptically-contoured multivariate stable distributions that we make use of in the proof of theorem 3:

1. We denote the p.d.f. of the elliptically-contoured continuous stable distribution as $p_{CS-EC}(\alpha, \Sigma, \vec{\mu})$, where Σ is a positive definite matrix, $\vec{\mu} \in \mathbb{R}^d$, and $\alpha \in (0, 2]$. In general, the p.d.f. of Elliptically-contoured multivariate stable distribution do not have a closed-form analytical expression, and are expressed through their characteristic function.
2. The characteristic function of a d -dimensional random vector $\vec{\mathbf{X}}$ having elliptically-contoured multivariate distribution $p_{CS-EC}(\alpha, \Sigma, \vec{\mu})$ is expressed as follows:

$$\mathbb{E} \left[e^{i\vec{t}^T \vec{\mathbf{X}}} \right] = \exp \left(- \left(\vec{t}^T \Sigma \vec{t} \right)^{\alpha/2} + i\vec{t}^T \vec{\mu} \right) \quad \forall \vec{t} \in \mathbb{R}^d. \quad (156)$$

H.2 Definition of the system 2 of the multivariate affine continuous stable system

Let the random variable $M \sim P(M)$ with support $\mathcal{M} \subseteq \mathbb{R}$. $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ are d_X -dimensional and d_Y -dimensional random vectors, respectively. The joint distribution $P(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$ describes the second system of the multivariate affine continuous stable distribution if it satisfies the following two properties:

1. $M \sim P(M)$ having some support set $\mathcal{M} \subseteq \mathbb{R}$.
2. The conditional distributions of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ are expressed as follows:

$$\begin{aligned} P(\vec{\mathbf{X}}|M) &= p_{CS-EC}(\alpha, \Sigma_X, \vec{\mathbf{H}}_X M + \vec{\mathbf{b}}_X), \text{ and} \\ P(\vec{\mathbf{Y}}|M) &= p_{CS-EC}(\alpha, \Sigma_Y, \vec{\mathbf{H}}_Y M + \vec{\mathbf{b}}_Y), \end{aligned} \quad (157)$$

where Σ_X and Σ_Y are positive definite matrices, $\vec{\mathbf{H}}_X, \vec{\mathbf{b}}_X \in \mathbb{R}^{d_X}$, and $\vec{\mathbf{H}}_Y, \vec{\mathbf{b}}_Y \in \mathbb{R}^{d_Y}$.

H.3 Formal proof of theorem 3

Theorem 3. Let the joint distribution $P(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$ of M , $\vec{\mathbf{X}}$, and $\vec{\mathbf{Y}}$ describe the system 2 of the multivariate affine continuous stable system. Define $\Sigma_X^{-1/2}$ and $\Sigma_Y^{-1/2}$ as the respective inverses of the matrices $\Sigma_X^{1/2}$ and $\Sigma_Y^{1/2}$ which satisfy: $(\Sigma_X^{1/2})^T \Sigma_X^{1/2} = \Sigma_X$, and $(\Sigma_Y^{1/2})^T \Sigma_Y^{1/2} = \Sigma_Y$. Without the loss of generality, assume $\|\Sigma_Y^{-1/2} \vec{\mathbf{H}}_Y\|_2 \leq \|\Sigma_X^{-1/2} \vec{\mathbf{H}}_X\|_2$. Then, Δ_P contains a Markov chain of the form $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{Y}}$ and $UI(M; \vec{\mathbf{Y}} \setminus \vec{\mathbf{X}}) = 0$.

Proof. We first note two important observations:

1. We can always assume $\|\Sigma_Y^{-1/2}\vec{H}_Y\|_2 \leq \|\Sigma_X^{-1/2}\vec{H}_X\|_2$ without the loss generality because if $\|\Sigma_Y^{-1/2}\vec{H}_Y\|_2 \geq \|\Sigma_X^{-1/2}\vec{H}_X\|_2$, then we can always switch our nomenclature to refer to \vec{Y} as \vec{X} , and \vec{X} as \vec{Y} .
2. Properties of positive definite matrices guarantee the existence of an invertible $\Sigma_X^{1/2}$ and $\Sigma_Y^{1/2}$ [58].

The proof of this theorem is a generalization of the proof used for the deriving a similar result for the special case of the multivariate Gaussian distribution discussed in [22], which borrows known results from Gaussian interference channels discussed in [48]. This proof follows the same structure as our earlier proofs.

1. In the first part, we explicitly construct a joint distribution $Q_{MC}(M, \vec{X}, \vec{Y})$ having the Markovian structure $M \rightarrow \vec{X} \rightarrow \vec{Y}$.
2. In the second part, we show that the $Q_{MC}(M, \vec{X}, \vec{Y})$ constructed in the first part, lies in Δ_P . Therefore, we can then apply the result of proposition 1 to conclude $UI(M; \vec{Y} \setminus \vec{X}) = 0$.

Part 1: Specifying $Q_{MC}(M, \vec{X}, \vec{Y})$

We explicitly construct the desired Markov chain $M \rightarrow \vec{X} \rightarrow \vec{Y}$ by constructing a larger Markov chain $M \rightarrow \vec{X} \rightarrow \vec{X}' \rightarrow \vec{X}'' \rightarrow \vec{Y}' \rightarrow \vec{Y}$, and then marginalizing the larger Markov chain to obtain the desired Markov chain $M \rightarrow \vec{X} \rightarrow \vec{Y}$.

Denote the joint distribution of the Markov chain $M \rightarrow \vec{X} \rightarrow \vec{X}' \rightarrow \vec{X}'' \rightarrow \vec{Y}' \rightarrow \vec{Y}$ as $Q_{MC}(M, \vec{X}, \vec{X}', \vec{X}'', \vec{Y}', \vec{Y})$. We can decompose $Q_{MC}(M, \vec{X}, \vec{X}', \vec{X}'', \vec{Y}', \vec{Y})$ by utilizing its Markovian structure as follows: $Q_{MC}(M, \vec{X}, \vec{X}', \vec{X}'', \vec{Y}', \vec{Y}) = Q_{MC}(M)Q_{MC}(\vec{X}|M)Q_{MC}(\vec{X}'|\vec{X})Q_{MC}(\vec{X}''|\vec{X}')Q_{MC}(\vec{Y}'|\vec{X}'')Q_{MC}(\vec{Y}|\vec{Y}')$. Consequently, we can specify/construct the distribution $Q_{MC}(M, \vec{X}, \vec{X}', \vec{X}'', \vec{Y}', \vec{Y})$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(\vec{X}|M)$, $Q_{MC}(\vec{X}'|\vec{X})$, $Q_{MC}(\vec{X}''|\vec{X}')$, $Q_{MC}(\vec{Y}'|\vec{X}'')$, and $Q_{MC}(\vec{Y}|\vec{Y}')$.

Specifying $Q_{MC}(M)$ and $Q_{MC}(\vec{X}|M)$: We choose $Q_{MC}(M)$ and $Q_{MC}(\vec{X}|M)$ as follows:

$$Q_{MC}(M) = P(M) \text{ and } Q_{MC}(\vec{X}|M) = P(\vec{X}|M). \quad (158)$$

Specifying $Q_{MC}(\vec{X}'|\vec{X})$: $Q_{MC}(\vec{X}'|\vec{X})$ is specified through the following deterministic transformation:

$$\vec{X}' = (\Sigma_X^{-1/2})^T (\vec{X} - \vec{b}_X). \quad (159)$$

We will also derive the distribution $Q(\vec{X}'|M)$ before proceeding with our construction, as we will need it later in the proof to show that the constructed $Q_{MC}(M, \vec{X}, \vec{Y}) \in \Delta_P$.

Deriving $Q_{MC}(\vec{X}'|M)$: We use lemma 11 to derive $Q_{MC}(\vec{X}'|M = m)$, as $Q_{MC}(\vec{X}|M = m)$ is an elliptically contoured multivariate stable distribution for a fixed m , and (159) defines a scaling and

translation operation.

$$\begin{aligned}
Q_{MC}(\vec{\mathbf{X}}'|M=m) &= p_{CS-EC} \left(\alpha, (\Sigma_X^{-1/2})^T \Sigma_X \Sigma_X^{-1/2}, (\Sigma_X^{-1/2})^T \vec{\mathbf{H}}_X m + (\Sigma_X^{-1/2})^T (\vec{\mathbf{b}}_X - \vec{\mathbf{b}}_X) \right), \\
\Rightarrow Q_{MC}(\vec{\mathbf{X}}'|M=m) &= p_{CS-EC} \left(\alpha, (\Sigma_X^{-1/2})^T \Sigma_X \Sigma_X^{-1/2}, (\Sigma_X^{-1/2})^T \vec{\mathbf{H}}_X m \right), \\
\Rightarrow Q_{MC}(\vec{\mathbf{X}}'|M=m) &= p_{CS-EC} \left(\alpha, (\Sigma_X^{-1/2})^T (\Sigma_X^{1/2})^T \Sigma_X^{1/2} \Sigma_X^{-1/2}, (\Sigma_X^{-1/2})^T \vec{\mathbf{H}}_X m \right), \\
\Rightarrow Q_{MC}(\vec{\mathbf{X}}'|M=m) &= p_{CS-EC} \left(\alpha, (\Sigma_X^{1/2} \Sigma_X^{-1/2})^T \Sigma_X^{1/2} \Sigma_X^{-1/2}, (\Sigma_X^{-1/2})^T \vec{\mathbf{H}}_X m \right), \\
\Rightarrow Q_{MC}(\vec{\mathbf{X}}'|M=m) &= p_{CS-EC} \left(\alpha, \mathbf{I}_{d_X}, (\Sigma_X^{-1/2})^T \vec{\mathbf{H}}_X m \right), \\
\Rightarrow Q_{MC}(\vec{\mathbf{X}}'|M=m) &= p_{CS-EC} \left(\alpha, \mathbf{I}_{d_X}, \vec{\mathbf{K}}_X m \right), \tag{160}
\end{aligned}$$

where \mathbf{I}_{d_X} is a $d_X \times d_X$ identity matrix, and $\vec{\mathbf{K}}_X = (\Sigma_X^{-1/2})^T \vec{\mathbf{H}}_X$.

Specifying $Q_{MC}(\vec{\mathbf{X}}''|\vec{\mathbf{X}}')$: $Q_{MC}(\vec{\mathbf{X}}''|\vec{\mathbf{X}}')$ is again specified through a deterministic transformation:

$$\vec{\mathbf{X}}'' = \vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T \vec{\mathbf{X}}', \tag{161}$$

where $\vec{\mathbf{K}}_Y = (\Sigma_Y^{-1/2})^T \vec{\mathbf{H}}_Y$ and $\vec{\mathbf{K}}_X^1 = 1/\|\vec{\mathbf{K}}_X\|_2^2 \vec{\mathbf{K}}_X$. We also derive the distribution $Q(\vec{\mathbf{X}}'|M)$ which will be needed for showing that the constructed $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}}) \in \Delta_P$.

Deriving $Q_{MC}(\vec{\mathbf{X}}''|M=m)$: We again use lemma 11 to calculate $Q_{MC}(\vec{\mathbf{X}}''|M=m)$, as $Q_{MC}(\vec{\mathbf{X}}'|M=m)$ is an elliptically contoured multivariate stable distribution (see (160)) for a fixed m , and (159) defines a scaling and translation operation.

$$\begin{aligned}
Q_{MC}(\vec{\mathbf{X}}'|M=m) &= p_{CS-EC} \left(\alpha, (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) \mathbf{I}_{d_X} (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T, \vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T \vec{\mathbf{K}}_X m \right), \\
&= p_{CS-EC} \left(\alpha, (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T, \vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T \vec{\mathbf{K}}_X m \right),
\end{aligned}$$

From lemma 10, we know that $\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T \vec{\mathbf{K}}_X = \vec{\mathbf{K}}_Y$, which implies

$$Q_{MC}(\vec{\mathbf{X}}'|M=m) = p_{CS-EC} \left(\alpha, (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T, \vec{\mathbf{K}}_Y m \right). \tag{162}$$

Specifying $Q_{MC}(\vec{\mathbf{Y}}'|\vec{\mathbf{X}}'')$: We specify $Q_{MC}(\vec{\mathbf{Y}}'|\vec{\mathbf{X}}'')$ through the following stochastic transformation:

$$\vec{\mathbf{Y}}' = \vec{\mathbf{X}}'' + \vec{\epsilon}, \tag{163}$$

where $\vec{\epsilon}$ follows a multivariate stable distribution. Furthermore, we assume that $\vec{\epsilon}$ is jointly independent from $(\vec{\mathbf{X}}, \vec{\mathbf{X}}', \vec{\mathbf{X}}'', M)$, i.e., $\vec{\epsilon} \perp\!\!\!\perp (\vec{\mathbf{X}}, \vec{\mathbf{X}}', \vec{\mathbf{X}}'', M)$. The characteristic function of $\vec{\epsilon}$ is defined as follows:

$$\mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\epsilon}} \right] = \exp \left(- \left(\vec{\mathbf{t}}^T \mathbf{I}_{d_Y} \vec{\mathbf{t}} \right)^{\alpha/2} + \left(\vec{\mathbf{t}}^T (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T \vec{\mathbf{t}} \right)^{\alpha/2} \right). \tag{164}$$

In order for $Q_{MC}(\vec{\mathbf{Y}}'|\vec{\mathbf{X}}'')$ to be a valid distribution, we need to ensure that the distribution of $\vec{\epsilon}$ defined in (164) is a legitimate multivariate stable distribution. Similarly to the proof of theorem 2, we will utilize property 1 for showing that $\vec{\epsilon}$ follows a stable multivariate stable distribution. Property 1 states a random vector is distributed according to a multivariate stable distribution iff every 1-dimensional projection of this random vector follows a univariate stable continuous distribution. Hence, to show that $\vec{\epsilon}$ is distributed according to a legitimate multivariate stable distribution, we will show that every 1-dimensional projection of $\vec{\epsilon}$ follows a univariate stable continuous distribution.

Let $\vec{\mathbf{t}} \in \mathbb{R}^{d_Y}$, then the characteristic function of the 1-dimensional projection of $\vec{\epsilon}$ along $\vec{\mathbf{t}}$, i.e., $\vec{\mathbf{t}}^T \vec{\epsilon}$ can be trivially derived from (164).

$$\mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\epsilon}} \right] = \exp \left(- \left(\vec{\mathbf{t}}^T \mathbf{I}_{d_Y} \vec{\mathbf{t}} \right)^{\alpha/2} + \left(\vec{\mathbf{t}}^T (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T \vec{\mathbf{t}} \right)^{\alpha/2} \right). \tag{165}$$

Comparing the characteristic function of $\vec{\mathbf{t}}^T \vec{\epsilon}$ described in (127) with the standard characteristic function of a univariate stable characteristic function defined in (323), we can conclude $\vec{\mathbf{t}}^T \vec{\epsilon} \sim p_{CS}(\alpha, \beta(\vec{\mathbf{t}}), \gamma(\vec{\mathbf{t}}), \mu(\vec{\mathbf{t}}))$, where:

$$\gamma(\vec{\mathbf{t}}) = \left(\vec{\mathbf{t}}^T \mathbf{I}_{d_Y} \vec{\mathbf{t}} \right)^{\alpha/2} - \left(\vec{\mathbf{t}}^T (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T \vec{\mathbf{t}} \right)^{\alpha/2}, \mu(\vec{\mathbf{t}}) = 0, \text{ and } \beta(\vec{\mathbf{t}}) = 0. \quad (166)$$

In order for $\vec{\mathbf{t}}^T \vec{\epsilon}$ to have a legitimate univariate stable distribution, we just need to show that $\gamma(\vec{\mathbf{t}}) \geq 0 \forall \vec{\mathbf{t}} \in \mathbb{R}^{d_Y}$. Note that $\gamma(\vec{\mathbf{t}}) = 0$ would correspond to the case where all the mass of the distribution is centered at 0, which does satisfy definition 2, and is an example of a degenerate univariate stable distribution.

By the assumption in theorem, $\|\vec{\mathbf{K}}_Y\|_2 = \|\Sigma_Y^{-1/2} \vec{\mathbf{H}}_Y\|_2 \leq \|\vec{\mathbf{K}}_X\|_2 = \|\Sigma_X^{-1/2} \vec{\mathbf{H}}_X\|_2$. Hence, we can use the result of lemma 12, which shows that $(\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T \preceq \mathbf{I}_{d_Y}$. Using the fact that $(\mathbf{I}_{d_Y} - (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T)$ is a positive semi-definite matrix, we have:

$$\begin{aligned} & \vec{\mathbf{t}}^T \left(\mathbf{I}_{d_Y} - (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T \right) \vec{\mathbf{t}} \geq 0 \forall \vec{\mathbf{t}} \in \mathbb{R}^{d_Y}, \\ \Rightarrow & \vec{\mathbf{t}}^T \mathbf{I}_{d_Y} \vec{\mathbf{t}} \geq \vec{\mathbf{t}}^T (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T \vec{\mathbf{t}} \forall \vec{\mathbf{t}} \in \mathbb{R}^{d_Y}, \\ \Rightarrow & \left(\vec{\mathbf{t}}^T \mathbf{I}_{d_Y} \vec{\mathbf{t}} \right)^{\alpha/2} \geq \left(\vec{\mathbf{t}}^T (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T \vec{\mathbf{t}} \right)^{\alpha/2} \forall \vec{\mathbf{t}} \in \mathbb{R}^{d_Y}, \\ \Rightarrow & \gamma(\vec{\mathbf{t}}) = \left(\vec{\mathbf{t}}^T \mathbf{I}_{d_Y} \vec{\mathbf{t}} \right)^{\alpha/2} - \left(\vec{\mathbf{t}}^T (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T \vec{\mathbf{t}} \right)^{\alpha/2} \geq 0 \forall \vec{\mathbf{t}} \in \mathbb{R}^{d_Y}. \end{aligned} \quad (167)$$

Hence, we know that $\vec{\epsilon}$ follows a legitimate multivariate stable distribution. We also derive $Q_{MC}(\vec{\mathbf{Y}}'|M)$ which is needed for showing $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}}) \in \Delta_P$.

Deriving $Q_{MC}(\vec{\mathbf{Y}}'|M)$: We will derive $Q_{MC}(\vec{\mathbf{Y}}'|M)$ using the fact that $\vec{\epsilon} \perp\!\!\!\perp \vec{\mathbf{X}}''|M = m$, hence their conditional characteristic functions would just result in multiplication.

$$\begin{aligned} \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{Y}}'} \middle| M = m \right] &= \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T (\vec{\mathbf{X}}'' + \vec{\epsilon})} \middle| M = m \right] = \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{X}}'' + i\vec{\mathbf{t}}^T \vec{\epsilon}} \middle| M = m \right], \\ &= \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{X}}''} e^{i\vec{\mathbf{t}}^T \vec{\epsilon}} \middle| M = m \right] \stackrel{(a)}{=} \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\mathbf{X}}''} \middle| M = m \right] \mathbb{E} \left[e^{i\vec{\mathbf{t}}^T \vec{\epsilon}} \middle| M = m \right], \end{aligned} \quad (168)$$

where (a) is due to $\vec{\epsilon} \perp\!\!\!\perp \vec{\mathbf{X}}''|M$. Substituting the characteristic functions of $\vec{\mathbf{X}}''$ and $\vec{\epsilon}$ from (162) and (164), respectively, in (168)

$$\begin{aligned} &= \exp \left(- \left(\vec{\mathbf{t}}^T (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T \vec{\mathbf{t}} \right)^{\alpha/2} + i\vec{\mathbf{t}}^T \vec{\mathbf{K}}_Y m \right) \times \\ & \quad \exp \left(- \left(\vec{\mathbf{t}}^T \mathbf{I}_{d_Y} \vec{\mathbf{t}} \right)^{\alpha/2} + \left(\vec{\mathbf{t}}^T (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T) (\vec{\mathbf{K}}_Y (\vec{\mathbf{K}}_X^1)^T)^T \vec{\mathbf{t}} \right)^{\alpha/2} \right), \\ &= \exp \left(- \left(\vec{\mathbf{t}}^T \mathbf{I}_{d_Y} \vec{\mathbf{t}} \right)^{\alpha/2} + i\vec{\mathbf{t}}^T \vec{\mathbf{K}}_Y m \right). \end{aligned} \quad (169)$$

By inspecting the characteristic function shown in (169), we can conclude:

$$Q_{MC}(\vec{\mathbf{Y}}'|M = m) = p_{CS-EC}(\alpha, \mathbf{I}_{d_Y}, \vec{\mathbf{K}}_Y m). \quad (170)$$

Specifying $Q_{MC}(\vec{\mathbf{Y}}|\vec{\mathbf{Y}}')$: Lastly, we specify $Q_{MC}(\vec{\mathbf{Y}}|\vec{\mathbf{Y}}')$ through the following deterministic transformation:

$$\vec{\mathbf{Y}} = (\Sigma_Y^{1/2})^T \vec{\mathbf{Y}}' + \vec{\mathbf{b}}_Y. \quad (171)$$

Finally, we construct the desired $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$ from $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{X}}', \vec{\mathbf{X}}'', \vec{\mathbf{Y}}', \vec{\mathbf{Y}})$ by marginalizing $\vec{\mathbf{X}}', \vec{\mathbf{X}}''$, and $\vec{\mathbf{Y}}'$.

Part 2: Showing $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}}) \in \Delta_P$

For showing that $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}}) \in \Delta_P$, we first need to derive $Q_{MC}(\vec{\mathbf{Y}}|M)$.

Deriving $Q(\vec{\mathbf{Y}}|M)$: We use lemma 11 to calculate $Q_{MC}(\vec{\mathbf{Y}}|M = m)$, as $Q_{MC}(\vec{\mathbf{X}}|M = m)$ is an elliptically contoured multivariate continuous stable distribution for a fixed m (see (170)), and (171) defines a scaling and translation operation.

$$\begin{aligned} Q_{MC}(\vec{\mathbf{X}}'|M = m) &= p_{CS-EC} \left(\alpha, (\boldsymbol{\Sigma}_Y^{1/2})^T \boldsymbol{\Sigma}_Y^{1/2}, (\boldsymbol{\Sigma}_Y^{1/2})^T \vec{\mathbf{K}}_Y m + \vec{\mathbf{b}}_Y \right), \\ &\stackrel{(a)}{=} p_{CS-EC} \left(\alpha, \boldsymbol{\Sigma}_Y, (\boldsymbol{\Sigma}_Y^{1/2})^T \vec{\mathbf{K}}_Y m + \vec{\mathbf{b}}_Y \right), \\ &\stackrel{(b)}{=} p_{CS-EC} \left(\alpha, \boldsymbol{\Sigma}_Y, (\boldsymbol{\Sigma}_Y^{1/2})^T (\boldsymbol{\Sigma}_Y^{-1/2})^T \vec{\mathbf{H}}_Y m + \vec{\mathbf{b}}_Y \right), \end{aligned} \quad (172)$$

$$= p_{CS-EC} \left(\alpha, \boldsymbol{\Sigma}_Y, \vec{\mathbf{H}}_Y m + \vec{\mathbf{b}}_Y \right), \quad (173)$$

where (a) is due to the fact that $(\boldsymbol{\Sigma}_Y^{1/2})^T \boldsymbol{\Sigma}_Y^{1/2} = \boldsymbol{\Sigma}_Y$, and (b) is due to the fact that $\vec{\mathbf{K}}_Y = \boldsymbol{\Sigma}_Y^{-1/2} \vec{\mathbf{H}}_Y$. Comparing (173) with $P(\vec{\mathbf{Y}}|M = m)$ shown in Appx. H.2, we can conclude:

$$Q_{MC}(\vec{\mathbf{Y}}|M) = P(\vec{\mathbf{Y}}|M). \quad (174)$$

From (174) and (158), we can conclude:

$$Q_{MC}(M, \vec{\mathbf{Y}}) = Q_{MC}(M) Q_{MC}(\vec{\mathbf{Y}}|M) = P(M) P(\vec{\mathbf{Y}}|M) = P(M, \vec{\mathbf{Y}}). \quad (175)$$

We know that by construction:

$$Q_{MC}(M, \vec{\mathbf{X}}) = P(M, \vec{\mathbf{X}}). \quad (176)$$

Hence, $Q_{MC}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}}) \in \Delta_P$ and consequently, by proposition 1, $UI(M; \vec{\mathbf{Y}} \setminus \vec{\mathbf{X}}) = 0$, concluding our proof. \square

Lemma 11. Let $\vec{\mathbf{X}} \sim p_{CS-EC}(\alpha, \boldsymbol{\Sigma}, \vec{\boldsymbol{\mu}})$ be a d -dimensional random vector, where $\boldsymbol{\Sigma}$ is a $d \times d$ positive definite matrix, $\vec{\boldsymbol{\mu}} \in \mathbb{R}^d$, and $\alpha \in (0, 2]$. Furthermore, let $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $\vec{\mathbf{b}} \in \mathbb{R}^d$. Then,

$$\mathbf{A} \vec{\mathbf{X}} + \vec{\mathbf{b}} \sim p_{CS-EC}(\alpha, \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^T, \mathbf{A} \vec{\boldsymbol{\mu}} + \vec{\mathbf{b}}).$$

Proof. Calculating the characteristic function of $\mathbf{A} \vec{\mathbf{X}} + \vec{\mathbf{b}}$:

$$\mathbb{E} \left[\exp \left(i \vec{\mathbf{t}}^T (\mathbf{A} \vec{\mathbf{X}} + \vec{\mathbf{b}}) \right) \right] = \mathbb{E} \left[\exp \left(i \vec{\mathbf{t}}^T \mathbf{A} \vec{\mathbf{X}} + i \vec{\mathbf{t}}^T \vec{\mathbf{b}} \right) \right] = \mathbb{E} \left[\exp \left(i \vec{\mathbf{t}}^T \mathbf{A} \vec{\mathbf{X}} \right) \exp \left(i \vec{\mathbf{t}}^T \vec{\mathbf{b}} \right) \right],$$

Employing linearity of expectation to pull the term $\exp \left(i \vec{\mathbf{t}}^T \vec{\mathbf{b}} \right)$ out of the expectation, as it is constant with respect to $\vec{\mathbf{X}}$:

$$\mathbb{E} \left[\exp \left(i \vec{\mathbf{t}}^T (\mathbf{A} \vec{\mathbf{X}} + \vec{\mathbf{b}}) \right) \right] = \exp \left(i \vec{\mathbf{t}}^T \vec{\mathbf{b}} \right) \mathbb{E} \left[\exp \left(i \vec{\mathbf{t}}^T \mathbf{A} \vec{\mathbf{X}} \right) \right]. \quad (177)$$

Substituting $\vec{\mathbf{t}}' = \vec{\mathbf{t}}^T \mathbf{A}$ in (177):

$$\mathbb{E} \left[\exp \left(i \vec{\mathbf{t}}^T (\mathbf{A} \vec{\mathbf{X}} + \vec{\mathbf{b}}) \right) \right] = \exp \left(i \vec{\mathbf{t}}^T \vec{\mathbf{b}} \right) \mathbb{E} \left[\exp \left(i (\vec{\mathbf{t}}')^T \vec{\mathbf{X}} \right) \right].$$

Substituting the formula of the characteristic function of an Elliptically-contoured multivariate stable distribution (given in (156)) for calculating $\mathbb{E} \left[\exp \left(i (\vec{\mathbf{t}}')^T \vec{\mathbf{X}} \right) \right]$ as $\vec{\mathbf{X}} \sim p_{CS-EC}(\alpha, \boldsymbol{\Sigma}, \vec{\boldsymbol{\mu}})$:

$$\mathbb{E} \left[\exp \left(i \vec{\mathbf{t}}^T (\mathbf{A} \vec{\mathbf{X}} + \vec{\mathbf{b}}) \right) \right] = \exp \left(i \vec{\mathbf{t}}^T \vec{\mathbf{b}} \right) \exp \left(- \left((\vec{\mathbf{t}}')^T \boldsymbol{\Sigma} \vec{\mathbf{t}}' \right)^{\alpha/2} + i (\vec{\mathbf{t}}')^T \vec{\boldsymbol{\mu}} \right).$$

Re-substituting $\vec{\mathbf{t}}' = \vec{\mathbf{t}}^T \mathbf{A}$ in the above equation

$$\begin{aligned} &= \exp \left(i \vec{\mathbf{t}}^T \vec{\mathbf{b}} \right) \exp \left(- \left(\vec{\mathbf{t}}^T \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^T \vec{\mathbf{t}} \right)^{\alpha/2} + i \vec{\mathbf{t}}^T \mathbf{A} \vec{\boldsymbol{\mu}} \right), \\ &= \exp \left(- \left(\vec{\mathbf{t}}^T \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^T \vec{\mathbf{t}} \right)^{\alpha/2} + i \vec{\mathbf{t}}^T (\mathbf{A} \vec{\boldsymbol{\mu}} + \vec{\mathbf{b}}) \right). \end{aligned} \quad (178)$$

By inspecting the characteristic function described in (178), we can determine that $\mathbf{A}\vec{\mathbf{X}} + \vec{\mathbf{b}}$ is distributed according to an elliptically-contoured multivariate stable distribution, i.e., $\mathbf{A}\vec{\mathbf{X}} + \vec{\mathbf{b}} \sim p_{CS-EC}(\alpha, \mathbf{A}\Sigma\mathbf{A}^T, \mathbf{A}\vec{\mu} + \vec{\mathbf{b}})$. \square

Lemma 12. Let $\vec{\mathbf{H}}_1 \in \mathbb{R}^{d_1}$ and $\vec{\mathbf{H}}_2 \in \mathbb{R}^{d_2}$. Then, $\|\vec{\mathbf{H}}_1\|_2 \geq \|\vec{\mathbf{H}}_2\|_2$ if and only if \exists a matrix $\mathbf{A} \in \mathbb{R}^{d_2 \times d_1}$ such that $\vec{\mathbf{H}}_2 = \mathbf{A}\vec{\mathbf{H}}_1$, and $\mathbf{A}\mathbf{A}^T \preceq \mathbf{I}_{d_2}$, with $\mathbf{A} = \vec{\mathbf{H}}_2(\vec{\mathbf{H}}_1^1)^T$, where $\mathbf{B} \preceq \mathbf{C}$ means $\mathbf{B} - \mathbf{C}$ is a positive semi-definite matrix, $\vec{\mathbf{H}}_1^1$ is as defined in lemma 10, and \mathbf{I}_{d_2} is $d_2 \times d_2$ identity matrix.

Proof. See the proof of lemma 5 in [48] for the special case $t = 1$. \square

I Proof of Theorem 4

In this section, we provide the proof of theorem 4. Since theorem 4 provides the analytical PID terms for the univariate affine discrete stable system defined in Sec. 4, we briefly restate certain key properties of the univariate discrete stable distributions and the corresponding univariate affine discrete stable system for convenience.

I.1 Univariate discrete stable distribution

Univariate discrete stable distributions are the discrete analogues of univariate continuous stable distributions. We refer the reader to Appx. N.4 for more details on univariate discrete stable distributions. We now list certain key properties of univariate continuous stable distributions that we make use of in the proof of theorem 4:

1. If X is distributed according to a univariate discrete stable distribution, then $X \stackrel{d}{=} \gamma \circ X_1 + (1 - \gamma) \circ X_2$, where X_1 and X_2 are two independent copies of X , $\nu \in (0, 1]$, $\gamma \in [0, 1]$, and \circ denotes the binomial thinning operation.
2. The p.m.f. of discrete stable distributions are characterized by two parameters: exponent $\nu \in (0, 1]$, and rate parameter $\tau \in (0, \infty)$. We denote the p.m.f. of a univariate discrete stable distribution as $P_{DS}(\nu, \tau)$. In general, discrete stable distributions do not have a “nice” analytical form consisting of well-known elementary functions. Consequently, the univariate discrete stable distribution are typically expressed through their probability generating function.
3. The probability generating function of a discrete random variable N having a stable discrete distribution is given by (179):

$$\mathbb{P}_N(z) = \exp(-\tau(1 - z)^\gamma). \quad (179)$$

I.2 Definition of univariate affine discrete stable system

Let M, X , and Y be a system of random variables with the joint distribution $P(M, X, Y)$. The joint distribution $P(M, X, Y)$ describes the univariate affine discrete stable system if it satisfies the following two properties:

1. $M \sim P(M)$ having some support set $\mathcal{M} \subseteq (0, \infty)$.
2. The conditional distributions of random variables X and Y conditioned on M can be expressed through discrete stable family distributions with an affine dependence on M . Formally, $P(X|M=m) = P_{DS}(\nu, am + b)$ and $P(Y|M=m) = P_{DS}(\nu, cm + d)$, where $a, b, c, d \in (0, \infty)$.

I.3 Formal proof of Theorem 4

Theorem 4. Let M, X , and Y be random variables whose joint distribution $P(M, X, Y)$ describes a univariate affine discrete stable system. Without the loss of generality, assume $a \geq c$. If $a/b \geq c/d$, then Δ_P contains a Markov chain of the form $M \rightarrow X \rightarrow Y$ and $UI(M; Y \setminus X) = 0$.

Proof. We first note that we can always assume $a \geq c$ without the loss of generality because if $a \leq c$, then we can always switch our nomenclature to refer to Y as X , and X as Y .

We briefly outline the proof structure.

1. In the first part, we explicitly construct a joint distribution $Q_{MC}(M, X, Y)$ having the Markovian structure $M \rightarrow X \rightarrow Y$.
2. In the second part, we show that the $Q_{MC}(M, X, Y)$ constructed in the first part lies in Δ_P . Therefore, we can then apply the result of proposition 1 to conclude $UI(M; Y \setminus X) = 0$.

Part 1: Specifying $Q_{MC}(M, X, Y)$

We explicitly construct the desired Markov chain $M \rightarrow X \rightarrow Y$ by constructing a larger Markov chain $M \rightarrow X \rightarrow X' \rightarrow Y$ and then marginalizing the larger Markov chain to obtain the desired Markov chain $M \rightarrow X \rightarrow Y$.

Denote the joint distribution of the Markov chain $M \rightarrow X \rightarrow X' \rightarrow Y$ as $Q_{MC}(M, X, X', Y)$. We can decompose $Q_{MC}(M, X, X', Y)$ by utilizing its Markovian structure as follows:

$$Q_{MC}(M, X, X', Y) = Q_{MC}(M)Q_{MC}(X|M)Q_{MC}(X'|X)Q_{MC}(Y|X').$$

Consequently, we can specify/construct the distribution $Q_{MC}(M, X, X', Y)$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(X|M)$, $Q_{MC}(X'|X)$, and $Q_{MC}(Y|X')$.

Specifying $Q_{MC}(M)$ and $Q_{MC}(X|M)$: We choose $Q_{MC}(M)$ and $Q_{MC}(X|M)$ as follows:

$$Q_{MC}(M) = P(M) \text{ and } Q_{MC}(X|M) = P(X|M), \quad (180)$$

where $P(M)$ and $P(X|M)$ are marginal distributions derived from the original joint distribution $P(M, X, Y)$ (discussed in the theorem statement) over which the bivariate PID is being calculated.

Specifying $Q_{MC}(X'|X)$: $Q_{MC}(X'|X)$ is specified through a binomial thinning operation, specifically:

$$X' = X \circ (c/a)^{1/\nu}. \quad (181)$$

Note that the above binomial thinning operation is valid as $c \leq a$ by the assumption in the theorem statement. We also derive $Q_{MC}(X'|M)$ as it is needed for showing $Q_{MC}(M, X, Y) \in \Delta_P$.

Deriving $Q_{MC}(X'|M)$: We use lemma 13 to derive $Q_{MC}(X'|M = m)$ as $Q_{MC}(X|M = m)$ describes a univariate discrete stable distribution for a fixed m , and X' is a binomially-thinned version of X . The exact expression of $Q_{MC}(X'|M = m)$ is provided in (182).

$$\begin{aligned} Q_{MC}(X'|M = m) &= P_{DS}(\nu, (c/a)^{\nu/\nu}(am + b/a)), \\ &= P_{DS}(\nu, cm + cb/a). \end{aligned} \quad (182)$$

Specifying $Q_{MC}(Y|X')$: $Q_{MC}(Y|X')$ is specified through the following stochastic transformation:

$$Y = X' + \epsilon, \quad (183)$$

where $\epsilon \sim P_{DS}(\nu, d - cb/a)$. For $Q_{MC}(Y|X')$ to be a valid distribution, we need to ensure that ϵ is distributed according to a legitimate discrete stable distribution. In order to show ϵ is distributed according to a legitimate discrete stable distribution, we need to ensure that ν and $d - cb/a$ lie within their appropriate bounds as specified in Sec. I.1. It is trivial to see $\nu \in (0, 1]$.

From the assumptions in the theorem statement, we know that

$$\frac{a}{b} \geq \frac{c}{d} \Rightarrow d \geq \frac{bc}{a} \Rightarrow d - \frac{bc}{a} \geq 0. \quad (184)$$

Hence, ϵ follows a legitimate discrete stable distribution. Furthermore, we choose $\epsilon \perp (M, X, X')$.

Finally, we construct the desired $Q_{MC}(M, X, Y)$ from $Q_{MC}(M, X, X', Y)$ by marginalizing X' .

Part 2: Showing $Q_{MC}(M, X, Y) \in \Delta_P$

For showing that $Q_{MC}(M, X, Y) \in \Delta_P$, we first need to derive $Q_{MC}(Y|M)$.

Deriving $Q(Y|M)$: Note that ϵ follows discrete stable distribution $P_{DS}(\nu, d - bc/a)$ and $Q_{MC}(X'|M = m)$ is also a discrete stable distribution $P_{DS}(X'; \nu, cm + bc/a)$ for a fixed m (see (182)). Furthermore, $\epsilon \perp\!\!\!\perp X'|M$. Hence, using lemma 14, we have

$$Q_{MC}(Y|M = m) = P_{DS}(\nu, cm + bc/a - bc/a + d) = P_{DS}(\nu, cm + d). \quad (185)$$

As $P(Y|M) = P_{DS}(\nu, cm + d)$ (see Appx. I.2), we have:

$$Q(Y|M) = P(Y|M). \quad (186)$$

From (180) and (186), we can conclude:

$$Q_{MC}(M, X) = Q_{MC}(M)Q_{MC}(X|M) = P(M)P(X|M) = P(M, X), \quad (187)$$

$$Q_{MC}(M, Y) = Q_{MC}(M)Q_{MC}(Y|M) = P(M)P(Y|M) = P(M, Y). \quad (188)$$

Hence, $Q_{MC}(M, X, Y) \in \Delta_P$ and consequently, by proposition 1, $UI(M; Y \setminus X) = 0$, concluding our proof. \square

Lemma 13. Let $N \sim P_{DS}(\nu, \tau)$. Then, $N_1 = \gamma \circ N \sim P_{DS}(\nu, \gamma^\nu \tau) \forall \gamma \in [0, 1]$, where \circ is the binomial thinning operation.

Proof. The probability generating function of N , denoted as $\mathbb{P}_N(z)$, can be described using (179):

$$\mathbb{P}_N(z) = \exp(-\tau(1-z)^\nu). \quad (189)$$

For calculating the probability generating function of N_1 , we use the following relationship between the probability generating function of a random variable R and its binomially-thinned version $R \circ p$, with $p \in [0, 1]$:

$$\mathbb{P}_{R \circ p}(z) = \mathbb{P}_R(1 - p(1 - z)) \text{ (see equation 3.3 in [24])}. \quad (190)$$

From (189) and (190), we can calculate the probability generating function of N_1 as follows:

$$\begin{aligned} \mathbb{P}_{N_1}(z) &= \mathbb{P}_{N \circ \gamma}(z) = \mathbb{P}_N(1 - \gamma(1 - z)) = \exp(-\tau(1 - (1 - \gamma(1 - z)))^\nu) \\ &= \exp(-\tau(\gamma(1 - z))^\nu) = \exp(-\tau\gamma^\nu(1 - z)^\nu). \end{aligned} \quad (191)$$

Inspecting the probability generating function of N_1 in (191), we can conclude that $N_1 \sim P_{DS}(\nu, \gamma^\nu \tau)$. \square

Lemma 14. Let $N_1 \sim P_{DS}(\nu, \tau_1)$, $N_2 \sim P_{DS}(\nu, \tau_2)$, and $N_1 \perp\!\!\!\perp N_2$. Then, $N_1 + N_2 = N \sim P_{DS}(\nu, \tau_1 + \tau_2)$.

Proof. The probability generating function of N_1 and N_2 , denoted as $\mathbb{P}_{N_1}(z)$ and $\mathbb{P}_{N_2}(z)$, can be described using (179):

$$\mathbb{P}_{N_1}(z) = \exp(-\tau_1(1-z)^\nu) \text{ and } \mathbb{P}_{N_2}(z) = \exp(-\tau_2(1-z)^\nu). \quad (192)$$

Calculating the probability generating function of $N = N_1 + N_2$,

$$\begin{aligned} \mathbb{P}_N(z) &= \mathbb{P}_{N_1+N_2}(z) \stackrel{(a)}{=} \mathbb{P}_{N_1}(z)\mathbb{P}_{N_2}(z) \stackrel{(b)}{=} \exp(-\tau_1(1-z)^\nu) \exp(-\tau_2(1-z)^\nu), \\ &= \exp(-(\tau_1 + \tau_2)(1-z)^\nu) \end{aligned} \quad (193)$$

where (a) is due to the fact that $N_1 \perp\!\!\!\perp N_2$, and (b) is achieved by substituting the explicit forms of $\mathbb{P}_{N_1}(z)$ and $\mathbb{P}_{N_2}(z)$ from (192). Inspecting the probability generating function of $N_1 + N_2$ in (193), we can conclude that $N_1 + N_2 = N \sim P_{DS}(\nu, \tau_1 + \tau_2)$. \square

J Proof of Theorem 5

In this section, we provide the proof of theorem 5. Since theorem 5 provides the analytical PID terms for the multivariate linear Poisson system defined in Sec. 4, we briefly restate certain key properties of multivariate Poisson distributions and the multivariate linear Poisson system for convenience.

J.1 Multivariate Poisson Distribution

We use the multivariate Poisson distribution proposed in [42]. For more details, see Appx. N.5. We list certain key properties of multivariate Poisson distributions that we make use of in the proof of theorem 5:

1. Under the definition of the multivariate Poisson distribution proposed in [42], each component of the random vector is a sum of independent Poisson random, i.e.,

$$\vec{N} = \mathbf{A}\vec{N}^g, \quad (194)$$

where \mathbf{A} is an appropriate matrix of 0's and 1's. We can decompose $\mathbf{A} = [\mathbf{A}_1 \dots \mathbf{A}_{d'}]$, where \mathbf{A}_i is a $d \times \binom{d}{i}$ submatrix having no duplicate columns, and each of its columns containing exactly i ones and $(d-i)$ zeros [44], and $\vec{N}^g = [N_1^g \dots N_d^g N_{12}^g \dots N_{(d-1)d}^g \dots N_{d-(d'-1)\dots d}^g]^T$, with $N_{i_1 \dots i_j}^g \sim \text{Poisson}(\lambda_{i_1 \dots i_j}) \forall (i_1, \dots, i_j) \in \mathbb{A}_j^d, j \in [d']$. Furthermore, the random variables $\{N_1, \dots, N_{d-(d'-1)\dots d}\}$ are mutually independent.

2. We denote the p.m.f. of the multivariate Poisson distribution as $\text{Poisson}(d, d', \vec{\Lambda})$, where $d \geq d'$, and

$$\vec{\Lambda} = [\lambda_1 \dots \lambda_d \lambda_{12} \dots \lambda_{d-(d'-1)\dots d}]^T. \quad (195)$$

3. For $d' = 1$, we have that \vec{N} is a collection of independent Poisson random variables, and when both $d = d' = 1$, we recover the scalar Poisson distribution: $\Pr(N = n) = \frac{e^{-\lambda} \lambda^n}{n!}, \forall n \in \mathbb{N}_0$.
4. Let $\vec{N} \sim \text{Poisson}(d, d', \vec{\Lambda})$, where \vec{N} is a d -dimensional random vector. Then the corresponding p.m.f. of \vec{N} is described below:

Let $\vec{n}' = [n_{12} \dots n_{(d-1)d} \dots n_{d-(d'-1)\dots d}]^T$, and $d_{\vec{n}'}$ be the dimension of \vec{n}' , then:

$$P(\vec{N}=\vec{n}) = e^{-\vec{1}^T \vec{\Lambda}} \prod_{i=1}^d \lambda_i^{n_i} \sum_{\vec{n}' \in C} \left(\prod_{(i_1, i_2) \in \mathbb{A}_2^d} \left(\frac{\lambda_{i_1 i_2}}{\lambda_{i_1} \lambda_{i_2}} \right)^{n_{i_1 i_2}} \times \dots \times \prod_{(i_1, \dots, i_{d'}) \in \mathbb{A}_{d'}^d} \left(\frac{\lambda_{i_1 \dots i_{d'}}}{\prod_{j=1}^{d'} \lambda_{i_j}} \right)^{n_{i_1 \dots i_{d'}}} \times Q(\vec{n}, \vec{n}') \right), \quad (196)$$

where $C = \{\vec{n}' \in \mathbb{N}_0^{d_{\vec{n}'}} : (\vec{a}_i')^T \vec{n}' \leq n_i \forall i \in [d]\}$, and

$$Q(\vec{n}, \vec{n}') = \prod_{i=1}^d \frac{1}{(n_i - \vec{a}_i'^T \vec{n}')!} \prod_{j=2}^{d'} \prod_{(i_1, \dots, i_j) \in \mathbb{A}_j^d} \frac{1}{n_{i_1 \dots i_j}^g!},$$

with \vec{a}_i' being the i -th row of the matrix $\mathbf{A}' = [\mathbf{A}_2 \dots \mathbf{A}_{d'}]$.

J.2 Definition of the multivariate linear Poisson system

Let the random variable $M \sim P(M)$. \vec{X} and \vec{Y} are d_X -dimensional and d_Y -dimensional random vectors, respectively. The joint distribution $P(M, \vec{X}, \vec{Y})$ describes the multivariate linear Poisson system if it satisfies the following two properties:

1. $M \sim P(M)$ having some support set $\mathcal{M} \subseteq (0, \infty)$.
2. The conditional distributions of random vectors \vec{X} and \vec{Y} are multivariate Poisson distributions, i.e., $P(\vec{X}|M) = \text{Poisson}(d_X, d'_X, \vec{\Lambda}_X)$ and $P(\vec{Y}|M) = \text{Poisson}(d_Y, d'_Y, \vec{\Lambda}_Y)$, with:

$$\vec{\Lambda}_X = [\lambda_1^X \dots \lambda_{d_X-(d'_X-1)\dots d_X}^X]^T, \lambda_{i_1 \dots i_j}^X = \gamma_{i_1 \dots i_j}^X M^j \forall j \in [d_X] \text{ and } (i_1, \dots, i_j) \in \mathbb{A}_j^{d_X},$$

$$\vec{\Lambda}_Y = [\lambda_1^Y \dots \lambda_{d_Y-(d'_Y-1)\dots d_Y}^Y]^T, \lambda_{i_1 \dots i_j}^Y = \gamma_{i_1 \dots i_j}^Y M^j \forall j \in [d_Y] \text{ and } (i_1, \dots, i_j) \in \mathbb{A}_j^{d_Y}.$$

Here, $\gamma_1^X, \dots, \gamma_{d_X-(d'_X-1)\dots d_X}^X, \gamma_1^Y, \dots, \gamma_{d_Y-(d'_Y-1)\dots d_Y}^Y \in \mathbb{R}^+$.

Furthermore, let \mathbf{A}_X , and \mathbf{A}_Y be the corresponding \mathbf{A} -matrices (defined in (194)) associated with $P(\vec{\mathbf{X}}|M)$ and $P(\vec{\mathbf{Y}}|M)$, respectively.

J.3 Formal proof of Theorem 5

Theorem 5. *Let the joint distribution $P(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$ of M , $\vec{\mathbf{X}}$, and $\vec{\mathbf{Y}}$ describe the multivariate linear Poisson system. Without the loss of generality, assume $d'_X \geq d'_Y$. If $\sum_{(i_1, \dots, i_j) \in \mathbb{A}_j^{d_X}} \gamma_{i_1 \dots i_j}^X \geq \sum_{(i_1, \dots, i_j) \in \mathbb{A}_j^{d_Y}} \gamma_{i_1 \dots i_j}^Y \forall j \in [d'_Y]$, then Δ_P contains a Markov chain of the form $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{Y}}$ and $UI(M; \vec{\mathbf{Y}} \setminus \vec{\mathbf{X}}) = 0$.*

Proof. We first note that we can always assume $d'_X \geq d'_Y$ without the loss of generality because if $d'_X \leq d'_Y$, then we can always switch our nomenclature to refer to $\vec{\mathbf{Y}}$ as $\vec{\mathbf{X}}$, and $\vec{\mathbf{X}}$ as $\vec{\mathbf{Y}}$.

We provide an explicit construction of the Markov chain $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{X}}^g \rightarrow \vec{\mathbf{Y}}^g \rightarrow \vec{\mathbf{Y}}$ having the marginals $P(M, \vec{\mathbf{X}})$, and $P(M, \vec{\mathbf{Y}})$. Denote the joint density of $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{X}}^g \rightarrow \vec{\mathbf{Y}}^g \rightarrow \vec{\mathbf{Y}}$ as $Q(M, \vec{\mathbf{X}}, \vec{\mathbf{X}}^g, \vec{\mathbf{Y}}^g, \vec{\mathbf{Y}})$. For $Q(M)$ and $Q(\vec{\mathbf{X}}|M)$, we choose them to be equal to $P(M)$ and $P(\vec{\mathbf{X}}|M)$, i.e. $Q(M) = P(M)$, and $Q(\vec{\mathbf{X}}|M) = P(\vec{\mathbf{X}}|M)$. Note that due to this construction, $P(M, \vec{\mathbf{X}}) = Q(M, \vec{\mathbf{X}})$ holds trivially.

For $Q(\vec{\mathbf{X}}^g|\vec{\mathbf{X}})$, we use the result described in lemma 17. Note that the construction for $Q(\vec{\mathbf{X}}^g|\vec{\mathbf{X}})$ is not explicit but rather implicit. Let $d_{\vec{\mathbf{A}}_X}$ be the dimension of $\vec{\mathbf{A}}_X$, then we explicitly choose: $Q(\vec{\mathbf{X}}|\vec{\mathbf{X}}^g, M) = \delta_K(\vec{\mathbf{X}} = \mathbf{A}_X \vec{\mathbf{X}}^g)$, and $Q(\vec{\mathbf{Y}}^g|M) = \text{Poisson}(d_{\vec{\mathbf{A}}_X}, 1, \vec{\mathbf{A}}_Y)$. Note that $Q(\vec{\mathbf{X}}|M) = \text{Poisson}(d_X, d'_X, \vec{\mathbf{A}}_X)$ by construction, and we derive $Q(\vec{\mathbf{X}}^g|\vec{\mathbf{X}}, M)$ through the Bayes theorem. Here, $\delta_K(\cdot)$ is the Kronecker delta function [59]. By lemma 17, we know that $Q(\vec{\mathbf{X}}^g|\vec{\mathbf{X}}, M) = Q(\vec{\mathbf{X}}^g|\vec{\mathbf{X}})$, and hence we have the Markov chain $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{X}}^g$.

For choosing $Q(\vec{\mathbf{Y}}^g|\vec{\mathbf{X}}^g)$, we rely on the result of lemma 18. Let us define the random vectors:

$$\vec{\mathbf{X}}_j = \left[X_{i_1 \dots i_j}^g \right]_{(i_1, \dots, i_j) \in \mathbb{A}_j^{d_1}}^T, \text{ and } \vec{\mathbf{Y}}_j = \left[Y_{i_1 \dots i_j}^g \right]_{(i_1, \dots, i_j) \in \mathbb{A}_j^{d_2}}^T,$$

i.e., $\vec{\mathbf{X}}_j$ and $\vec{\mathbf{Y}}_j$ are random vectors containing all terms of the form X_{i_1, \dots, i_j}^g , and Y_{k_1, \dots, k_j}^g , where $(i_1, \dots, i_j) \in \mathbb{A}_j^{d_1}$ and $(k_1, \dots, k_j) \in \mathbb{A}_j^{d_2}$, respectively. Note that we can write $\vec{\mathbf{X}}^g = \left[\vec{\mathbf{X}}_1^T \dots \vec{\mathbf{X}}_{d'_X}^T \right]^T$, and we define $\vec{\mathbf{Y}}^g = \left[\vec{\mathbf{Y}}_1^T \dots \vec{\mathbf{Y}}_{d'_Y}^T \right]^T$. Then, we construct $Q(\vec{\mathbf{Y}}^g|\vec{\mathbf{X}}^g)$ as a product of d'_Y multinomial distributions, described below:

$$Q(\vec{\mathbf{Y}}^g|\vec{\mathbf{X}}^g) = \prod_{j=1}^{d'_Y} Q(\vec{\mathbf{Y}}_j|\vec{\mathbf{X}}_j),$$

where $Q(\vec{\mathbf{Y}}_j|\vec{\mathbf{X}}_j) = \text{Multinomial}(\mathbf{k}_j; N_j, \mathbf{p}_i)$, and

$$\begin{aligned} N_j &= \vec{\mathbf{1}}^T \vec{\mathbf{x}}_j, \quad \vec{\mathbf{k}}_j = \left[\vec{\mathbf{y}}_j^T \quad \vec{\mathbf{1}}^T \vec{\mathbf{x}}_j - \vec{\mathbf{1}}^T \vec{\mathbf{y}}_j \right]^T, \\ \vec{\mathbf{p}}_i &= \left[p_{1 \dots j} \dots p_{d_2 - (j-1) \dots d_2} \quad 1 - \sum_{(i_1, \dots, i_j) \in \mathbb{A}_j^{d_2}} p_{i_1 \dots i_j} \right]^T, \\ p_{i_1 \dots i_j} &= \frac{\gamma_{i_1 \dots i_j}^Y}{\sum_{(i_1, \dots, i_j) \in \mathbb{A}_j^{d_1}} \gamma_{i_1 \dots i_j}^X} \quad \forall (i_1, \dots, i_j) \in \mathbb{A}_j^{d_2}. \end{aligned}$$

By construction $\vec{\mathbf{X}}^g$ consists of mutually conditionally independent Poisson random variables (conditioned on M), and the condition of lemma 18 (specified in (219)) is satisfied by condition 1. Therefore, using the result of lemma 18 on the Markov chain $M \rightarrow \vec{\mathbf{X}}^g \rightarrow \vec{\mathbf{Y}}^g$ (obtained by marginalizing $\vec{\mathbf{X}}$), we conclude that $Q(\vec{\mathbf{Y}}^g|M) = \text{Poisson}(1, d_{\vec{\mathbf{A}}_Y}, \vec{\mathbf{A}}_Y)$, where $d_{\vec{\mathbf{A}}_Y}$ is the dimension of $\vec{\mathbf{A}}_Y$.

We specify $Q(\vec{\mathbf{Y}}|\vec{\mathbf{Y}}^g)$ through the following deterministic transformation $\vec{\mathbf{Y}} = \mathbf{A}_Y \vec{\mathbf{Y}}^g$ to obtain the Markov chain $M \rightarrow \vec{\mathbf{X}} \rightarrow \vec{\mathbf{X}}^g \rightarrow \vec{\mathbf{Y}}^g \rightarrow \vec{\mathbf{Y}}$. Marginalizing $\vec{\mathbf{X}}$ and $\vec{\mathbf{X}}^g$ in the above Markov chain,

we get the following Markov chain: $M \rightarrow \vec{Y}^g \rightarrow \vec{Y}$. Now, since $\vec{Y} = \mathbf{A}_Y \vec{Y}^g$ and $P(\vec{Y}^g|M) = \text{Poisson}(d_{\vec{A}_Y}, 1, \vec{A}_Y)$, $P(\vec{Y}|M) = \text{Poisson}(d_Y, d_Y', \Lambda_Y)$ (by definition of the multivariate Poisson discussed in Sec. J.1). Since we have $Q(M) = P(M)$ and $Q(\vec{Y}|M) = P(\vec{Y}|M)$, we also have $Q(M, \vec{Y}) = P(M, \vec{Y})$. Marginalizing \vec{X}^g and \vec{Y}^g from the Markov chain $M \rightarrow \vec{X} \rightarrow \vec{X}^g \rightarrow \vec{Y}^g \rightarrow \vec{Y}$ results in the desired Markov chain $M \rightarrow \vec{X} \rightarrow \vec{Y}$ having $Q(M, \vec{X}) = P(M, \vec{X})$ and $Q(M, \vec{Y}) = P(M, \vec{Y})$, concluding our proof. \square

Lemma 15. *If $N_j \sim \text{Poisson}(\lambda_j) \forall j \in \{1, \dots, d\}$, and all N_j 's are mutually independent, then $P(N_1, \dots, N_j | \sum_{j=1}^d N_j = N) = \text{Multinomial}(N_1, \dots, N_j; N, [\lambda_1 / \sum_{i=1}^d \lambda_i \dots \lambda_d / \sum_{i=1}^d \lambda_i])$.*

Proof. See Section 3 of [60]. \square

Lemma 16. *Let $\vec{N}_1 = [N_1^{(1)} \dots N_{d_1}^{(1)}]^T$, where $N_j \sim \text{Poisson}(\lambda_j^{(1)}) \forall j \in \{1, \dots, d_1\}$ and all N_j 's are mutually independent from each other. Define $P(\vec{N}_2|\vec{N}_1) = \text{Multinomial}(\vec{k}; N, \vec{p})$, where*

$$N = \sum_{j=1}^{d_1} N_j^{(1)}, \quad \vec{k} = [N_1^{(2)} \dots N_{d_2}^{(2)} \sum_{j=1}^{d_1} N_j^{(1)} - \sum_{j=1}^{d_2} N_j]{}^T,$$

$$\vec{p} = \left[\frac{\lambda_1^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}} \dots \frac{\lambda_{d_2}^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}} 1 - \frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}} \right]{}^T,$$

where $\{\lambda_j^{(2)}\}_{j=1}^{d_2}$ are some positive constants, i.e., $\lambda_j^{(2)} \in (0, \infty)$, such that $\sum_{i=1}^{d_2} \lambda_j^{(2)} \leq \sum_{j=1}^{d_1} \lambda_j^{(1)}$. Then, $\vec{N}_2 = [N_1^{(2)} \dots N_{d_2}^{(2)}]^T$, where $N_j^{(2)} \sim \text{Poisson}(\lambda_j^{(2)})$, and all $N_j^{(2)}$'s are mutually independent from each other.

Proof. Lemma 16 is a straight-forward generalization of lemma 15. Denote the probability of \vec{N}_2 shown in the lemma statement as $P(\vec{N}_2)$. We will prove the above lemma by constructing a Markov chain $\vec{N}_1 \rightarrow N'_1 \rightarrow N'_2 \rightarrow \vec{N}_2$ with joint density $Q(\vec{N}_1, N'_1, N'_2, \vec{N}_2)$, and showing that $Q(\vec{N}_2) = P(\vec{N}_2)$ and $Q(\vec{N}_2|\vec{N}_1) = P(\vec{N}_2|\vec{N}_1)$.

Specifying $Q(N'_1|\vec{N}_1)$: We specify $Q(N'_1|\vec{N}_1)$ through the following deterministic transformation:

$$N'_1 = \vec{1}^T \vec{N}_1 = \sum_{j=1}^{d_1} N_j^{(1)}. \quad (197)$$

Calculating $Q(N'_1)$: We use the property that the sum of independent Poisson random variables is a Poisson random variable with its rate parameter being the sum of the rate parameters of its summands [61]. Hence, $Q(N'_1) = \text{Poisson}(\sum_{j=1}^{d_1} \lambda_j^{(1)})$.

Specifying $Q(N'_2|N'_1)$: We specify $Q(N'_2|N'_1 = n'_1)$ as follows:

$$Q(N'_2|N'_1 = n'_1) = \text{Binomial}\left(N'_2; n'_1, \sum_{j=1}^{d_2} \lambda_j^{(2)} / \sum_{j=1}^{d_1} \lambda_j^{(1)}\right). \quad (198)$$

Calculating $Q(N'_2)$: We use lemma 15 to calculate $Q(N'_2)$. Note that $Q(N'_1) = \text{Poisson}(\sum_{j=1}^{d_1} \lambda_j^{(1)})$, and $Q(N'_2|N'_1) = \text{Binomial}\left(N'_2; n'_1, \sum_{j=1}^{d_2} \lambda_j^{(2)} / \sum_{j=1}^{d_1} \lambda_j^{(1)}\right)$. Hence, by lemma 15, we have:

$$Q(N'_2) = \text{Poisson}\left(\sum_{j=1}^{d_1} \lambda_j^{(1)} \times \frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}}\right) = \text{Poisson}\left(\sum_{j=1}^{d_2} \lambda_j^{(2)}\right). \quad (199)$$

Specifying $Q(N'_2|N'_1)$: We specify $Q(\vec{N}_2|N'_2 = n'_2)$ as follows:

$$Q(\vec{N}_2|N'_2 = n'_2) = \text{Multinomial} \left(\vec{N}'_2; n'_2, \left[\frac{\lambda_1^{(2)}}{\sum_{j=1}^{d_2} \lambda_j^{(2)}} \quad \dots \quad \frac{\lambda_{d_2}^{(2)}}{\sum_{j=1}^{d_2} \lambda_j^{(2)}} \right]^T \right). \quad (200)$$

Calculating $Q(\vec{N}_2)$: We again use lemma 15 to calculate $Q(\vec{N}_2)$. Note that,

$$Q(N'_2) = \text{Poisson} \left(\sum_{j=1}^{d_2} \lambda_j^{(2)} \right), \text{ and}$$

$$Q(\vec{N}_2|N'_2 = n'_2) = \text{Multinomial} \left(\vec{N}'_2; n'_2, \left[\frac{\lambda_1^{(2)}}{\sum_{j=1}^{d_2} \lambda_j^{(2)}} \quad \dots \quad \frac{\lambda_{d_2}^{(2)}}{\sum_{j=1}^{d_2} \lambda_j^{(2)}} \right]^T \right). \quad (201)$$

Hence, by lemma 15, we have:

$$\vec{N}_2 = \left[N_1^{(2)} \quad \dots \quad N_{d_2}^{(2)} \right]^T, \text{ where } N_j^{(2)} \sim \text{Poisson}(\lambda_j^{(2)}) \quad \forall j \in [d_2]. \quad (202)$$

Hence, by (202), we know that $Q(\vec{N}_2) = P(\vec{N}_2)$.

Calculating $Q(\vec{N}_2|\vec{N}_1)$:

$$Q(\vec{N}_2 = \vec{n}_2|\vec{N}_1 = \vec{n}_1) = \sum_{n'_1, n'_2=0}^{\infty} Q(N'_1 = n'_1, N'_2 = n'_2, \vec{N}_2 = \vec{n}_2|\vec{N}_1 = \vec{n}_1),$$

$$\stackrel{(a)}{=} \sum_{n'_1, n'_2=0}^{\infty} Q(n'_1|\vec{n}_2)Q(n'_2|n'_1)Q(\vec{n}_2|n'_2),$$

where (a) is due to the Markovian nature of $Q(\vec{N}_1, N'_1, N'_2, \vec{N}_2)$. We drop the random variable notation for brevity. Let $\vec{n}_1 = \left[n_1^{(1)} \quad \dots \quad n_{d_1}^{(1)} \right]^T$, and $\vec{n}_2 = \left[n_1^{(2)} \quad \dots \quad n_{d_2}^{(2)} \right]^T$. Then, substituting the exact form of $Q(n'_1|\vec{n}_1), Q(n'_2|n'_1)$, and $Q(\vec{n}_2|n'_2)$ in the above equation:

$$= \sum_{n'_1, n'_2=0}^{\infty} \mathbb{I} \left[n'_1 = \sum_{j=1}^{d_1} n_j^{(1)} \right] \frac{(n'_1)!}{(n'_1 - n'_2)!(n'_2)!} \left(\frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}} \right)^{n'_2} \left(1 - \frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}} \right)^{n'_1 - n'_2} \times,$$

$$\mathbb{I} \left[\sum_{j=1}^{d_2} n_j^{(2)} = n'_2 \right] \frac{(n'_2)!}{\prod_{j=1}^{d_2} (n_j^{(2)})!} \prod_{j=1}^{d_2} \left(\frac{\lambda_j^{(2)}}{\sum_{j=1}^{d_2} \lambda_j^{(2)}} \right)^{n_j^{(2)}},$$

where $\mathbb{I}[\cdot]$ is an indicator function. Simplifying the above equation by using the fact that only non-zero term in the above summation is $n'_1 = \sum_{j=1}^{d_1} n_j^{(1)}$ and $n'_2 = \sum_{j=1}^{d_2} n_j^{(2)}$.

$$= \frac{\left(\sum_{j=1}^{d_1} n_j^{(1)} \right)!}{\left(\sum_{j=1}^{d_1} n_j^{(1)} - \sum_{j=1}^{d_2} n_j^{(2)} \right)! \left(\sum_{j=1}^{d_2} n_j^{(2)} \right)!} \left(\frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}} \right)^{\sum_{j=1}^{d_2} n_j^{(2)}} \times$$

$$\left(1 - \frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}} \right)^{\sum_{j=1}^{d_1} n_j^{(1)} - \sum_{j=1}^{d_2} n_j^{(2)}} \frac{\left(\sum_{j=1}^{d_2} n_j^{(2)} \right)!}{\prod_{j=1}^{d_2} (n_j^{(2)})!} \prod_{j=1}^{d_2} \left(\frac{\lambda_j^{(2)}}{\sum_{j=1}^{d_2} \lambda_j^{(2)}} \right)^{n_j^{(2)}},$$

Collecting all factorial terms together, and rearranging some terms:

$$= \frac{\left(\sum_{j=1}^{d_1} n_j^{(1)} \right)!}{\left(\sum_{j=1}^{d_1} n_j^{(1)} - \sum_{j=1}^{d_2} n_j^{(2)} \right)! \prod_{j=1}^{d_2} (n_j^{(2)})!} \left(1 - \frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}} \right)^{\sum_{j=1}^{d_1} n_j^{(1)} - \sum_{j=1}^{d_2} n_j^{(2)}} \times$$

$$\left(\frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}} \right)^{\sum_{j=1}^{d_2} n_j^{(2)}} \prod_{j=1}^{d_2} \left(\frac{\lambda_j^{(2)}}{\sum_{j=1}^{d_2} \lambda_j^{(2)}} \right)^{n_j^{(2)}}.$$

Substituting $\left(\sum_{j=1}^{d_2} \lambda_j^{(2)} / \sum_{j=1}^{d_1} \lambda_j^{(1)}\right)^{\sum_{j=1}^{d_2} n_j^{(2)}} = \prod_{j=1}^{d_2} \left(\sum_{j=1}^{d_2} \lambda_j^{(2)} / \sum_{j=1}^{d_1} \lambda_j^{(1)}\right)^{n_j^{(2)}}$ in the above equation:

$$\begin{aligned}
&= \frac{\left(\sum_{j=1}^{d_1} n_j^{(1)}\right)!}{\left(\sum_{j=1}^{d_1} n_j^{(1)} - \sum_{j=1}^{d_2} n_j^{(2)}\right)! \prod_{j=1}^{d_2} (n_j^{(2)})!} \left(1 - \frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}}\right)^{\sum_{j=1}^{d_1} n_j^{(1)} - \sum_{j=1}^{d_2} n_j^{(2)}} \times \\
&\quad \prod_{j=1}^{d_1} \left(\frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}}\right)^{n_j^{(2)}} \prod_{j=1}^{d_2} \left(\frac{\lambda_j^{(2)}}{\sum_{j=1}^{d_2} \lambda_j^{(2)}}\right)^{n_j^{(2)}} \\
&= \frac{\left(\sum_{j=1}^{d_1} n_j^{(1)}\right)!}{\left(\sum_{j=1}^{d_1} n_j^{(1)} - \sum_{j=1}^{d_2} n_j^{(2)}\right)! \prod_{j=1}^{d_2} (n_j^{(2)})!} \left(1 - \frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}}\right)^{\sum_{j=1}^{d_1} n_j^{(1)} - \sum_{j=1}^{d_2} n_j^{(2)}} \times \\
&\quad \prod_{j=1}^{d_1} \left(\frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}}\right)^{n_j^{(2)}} \left(\frac{\lambda_j^{(2)}}{\sum_{j=1}^{d_2} \lambda_j^{(2)}}\right)^{n_j^{(2)}} \\
&= \frac{\left(\sum_{j=1}^{d_1} n_j^{(1)}\right)!}{\left(\sum_{j=1}^{d_1} n_j^{(1)} - \sum_{j=1}^{d_2} n_j^{(2)}\right)! \prod_{j=1}^{d_2} (n_j^{(2)})!} \left(1 - \frac{\sum_{j=1}^{d_2} \lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}}\right)^{\sum_{j=1}^{d_1} n_j^{(1)} - \sum_{j=1}^{d_2} n_j^{(2)}} \times \\
&\quad \prod_{j=1}^{d_1} \left(\frac{\lambda_j^{(2)}}{\sum_{j=1}^{d_1} \lambda_j^{(1)}}\right)^{n_j^{(2)}}. \tag{203}
\end{aligned}$$

By inspecting (203), we can see that $Q(\vec{N}_2 | \vec{N}_1) = \text{Multinomial}(\vec{k}; N, \vec{p})$ with,

$$N = \sum_{j=1}^{d_1} N_j^{(1)}, \quad \vec{k} = \left[N_1^{(2)} \dots N_{d_2}^{(2)} \sum_{j=1}^{d_1} N_j^{(1)} - \sum_{j=1}^{d_2} N_j \right]^T, \tag{204}$$

$$\vec{p} = \left[\lambda_1^{(2)} / \sum_{j=1}^{d_1} \lambda_j^{(1)} \dots \lambda_{d_2}^{(2)} / \sum_{j=1}^{d_1} \lambda_j^{(1)} \quad 1 - \sum_{j=1}^{d_2} \lambda_j^{(2)} / \sum_{j=1}^{d_1} \lambda_j^{(1)} \right]^T, \tag{205}$$

Hence, by (205), we can conclude that $Q(\vec{N}_2 | \vec{N}_1) = P(\vec{N}_2 | \vec{N}_1)$ finishing our proof. \square

Lemma 17. Let \vec{N} be a d -dimensional vector and M be a positive random variable, i.e. $M \sim P(M), P(M \leq 0) = 0$. Let $P(\vec{N} | M) = \text{Poisson}(d, d', \vec{\Lambda})$, where:

$$\begin{aligned}
\vec{\Lambda} &= [\lambda_1 \dots \lambda_d \quad \lambda_{12} \dots \lambda_{d-(d'-1)\dots d}]^T, \\
\lambda_{i_1 \dots i_j} &= \gamma_{i_1 \dots i_j} M^j \quad \forall j \in [d'] \text{ and } (i_1, \dots, i_{d'}) \in \mathbb{A}_j^d. \tag{206}
\end{aligned}$$

Let $\vec{N}^g = [N_1^g \dots N_d^g \quad N_{12}^g \dots N_{(d-1)d}^g \dots N_{d-(d'-1)\dots d}^g]^T$, where $P(\vec{N}^g | M) = \text{Poisson}(d_{\vec{\Lambda}}, 1, \vec{\Lambda})$ and $\vec{N} = \mathbf{A} \vec{N}^g$, where $d_{\vec{\Lambda}}$ is the dimension of $\vec{\Lambda}$, and $\mathbf{A} = [\mathbf{A}_1 \dots \mathbf{A}_{d'}]$ as defined in Sec. J.1. Then $P(\vec{N}^g | \vec{N}, M) = P(\vec{N}^g | \vec{N})$, i.e. M, \vec{N}^g and \vec{N} form the following Markov chain $M \rightarrow \vec{N} \rightarrow \vec{N}^g$.

Proof. To prove lemma 17, all we need to show is that $P(\vec{N}^g | \vec{N}, M)$ does not depend on M . Let us to calculate $P(\vec{N}^g | \vec{N}, M)$. By Bayes Theorem, we know:

$$P(\vec{N}^g | \vec{N}, M) = \frac{P(\vec{N}^g | M) P(\vec{N} | \vec{N}^g, M)}{P(\vec{N} | M)}. \tag{207}$$

Let us now write the expression for $P(\vec{N} | M)$ using (196):

$$P(\vec{N} = \vec{n} | M) = Q(\vec{n}, \vec{n}') e^{-\vec{1}^T \vec{\Lambda}} \prod_{i=1}^d (\gamma_i M)^{n_i} \left(\sum_{\vec{n}' \in C} \prod_{j=2}^{d'} \prod_{(i_1, \dots, i_j) \in \mathbb{A}_j^d} \left(\frac{\gamma_{i_1 \dots i_j} M^j}{\prod_{l=1}^j \gamma_{i_l} M} \right)^{n'_{i_1 \dots i_j}} \right). \tag{208}$$

Canceling M in the above equation for the terms inside the summation, we get:

$$P(\vec{\mathbf{N}}=\vec{\mathbf{n}}|M)=Q(\vec{\mathbf{n}},\vec{\mathbf{n}}')e^{-\vec{\mathbf{I}}^T\vec{\Lambda}}\prod_{i=1}^d(\gamma_i M)^{n_i}\left(\sum_{\vec{\mathbf{n}}'\in C}\prod_{j=2}^{d'}\prod_{(i_1,\dots,i_j)\in\mathbb{A}_j^d}\left(\frac{\gamma_{i_1\dots i_j}}{\prod_{l=1}^j\gamma_{i_l}}\right)^{n'_{i_1\dots i_j}}\right). \quad (209)$$

Absorbing all terms that do not depend upon M into B , we have

$$B=\sum_{\vec{\mathbf{n}}'\in C}\prod_{j=2}^{d'}\prod_{(i_1,\dots,i_j)\in\mathbb{A}_j^d}\left(\frac{\gamma_{i_1\dots i_j}}{\prod_{l=1}^j\gamma_{i_l}}\right)^{n'_{i_1\dots i_j}}, \quad (210)$$

Then, we can rewrite (209) as:

$$\begin{aligned} P(\vec{\mathbf{N}}=\vec{\mathbf{n}}|M) &= Be^{-\vec{\mathbf{I}}^T\vec{\Lambda}}\prod_{i=1}^d(\gamma_i M)^{n_i} = Be^{-\vec{\mathbf{I}}^T\vec{\Lambda}}M^{\sum_{i=1}^d n_i}\prod_{i=1}^d(\gamma_i)^{n_i} \\ &\stackrel{(a)}{=} Be^{-\vec{\mathbf{I}}^T\vec{\Lambda}}M^{\sum_{i=1}^d n_i} \stackrel{(b)}{=} Be^{-\vec{\mathbf{I}}^T\vec{\Lambda}}M^{\vec{\mathbf{I}}^T\vec{\mathbf{n}}}, \end{aligned} \quad (211)$$

where, in (a) we further absorb $\prod_{i=1}^d(\gamma_i)^{n_i}$ into B since it does not depend upon M , and in (b) we substitute $\sum_{i=1}^d n_i = \vec{\mathbf{I}}^T\vec{\mathbf{n}}$. Similarly, let us write out the expression for $P(\vec{\mathbf{N}}^g|\vec{\mathbf{n}}^g)$:

$$P(\vec{\mathbf{N}}^g=\vec{\mathbf{n}}^g|M) = e^{-\vec{\mathbf{I}}^T\vec{\Lambda}}\prod_{j=1}^{d'}\prod_{(i_1,\dots,i_j)\in\mathbb{A}_j^d}(\gamma_{i_1\dots i_j}M^j)^{n^g_{i_1\dots i_j}}. \quad (212)$$

Collecting all the M terms, and absorbing all the terms that do not depend upon M into D , we obtain:

$$P(\vec{\mathbf{N}}^g=\vec{\mathbf{n}}^g|M) = De^{-\vec{\mathbf{I}}^T\vec{\Lambda}}M^{\sum_{j=1}^{d'}\sum_{(i_1,\dots,i_j)\in\mathbb{A}_j^d}(jn^g_{i_1\dots i_j})}. \quad (213)$$

Now let us analyze the term $\vec{\mathbf{I}}^T\mathbf{A}\vec{\mathbf{n}}^g$:

$$\vec{\mathbf{I}}^T\mathbf{A}\vec{\mathbf{n}}^g \stackrel{(a)}{=} (\vec{\mathbf{n}}^g)^T\mathbf{A}^T\vec{\mathbf{I}} \stackrel{(b)}{=} (\vec{\mathbf{n}}^g)^T\begin{bmatrix}\mathbf{A}_1^T\vec{\mathbf{I}} \\ \vdots \\ \mathbf{A}_{d'}^T\vec{\mathbf{I}}\end{bmatrix} \stackrel{(c)}{=} (\vec{\mathbf{n}}^g)^T\begin{bmatrix}\vec{\mathbf{I}} \\ 2\vec{\mathbf{I}} \\ \vdots \\ d'\vec{\mathbf{I}}\end{bmatrix},$$

where (a) uses the fact that $\vec{\mathbf{I}}^T\mathbf{A}\vec{\mathbf{n}}^g$ is a scalar and hence is equal to its transpose, (b) uses the fact $\mathbf{A}=[\mathbf{A}_1\ \dots\ \mathbf{A}_{d'}]$, and (c) follows from the special structure of \mathbf{A}_i , i.e., that each column only contains i ones and $d-i$ zeros, and the fact that $\mathbf{A}_1^T\vec{\mathbf{I}}$ is akin to summing up the columns, hence $\mathbf{A}_i^T\vec{\mathbf{I}} = i\vec{\mathbf{I}}$. Equivalently we can rewrite the above equation as:

$$\vec{\mathbf{I}}^T\mathbf{A}\vec{\mathbf{n}}^g = \sum_{j=1}^{d'}\sum_{(i_1,\dots,i_j)\in\mathbb{A}_j^d}(jn^g_{i_1\dots i_j}). \quad (214)$$

Substituting (214) into (213):

$$P(\vec{\mathbf{N}}^g|M) = De^{-\vec{\mathbf{I}}^T\vec{\Lambda}}M^{\vec{\mathbf{I}}^T\mathbf{A}\vec{\mathbf{n}}^g}. \quad (215)$$

Now, let us write out the expression for $P(\vec{\mathbf{N}}|\vec{\mathbf{N}}^g,M)$. Since $\vec{\mathbf{N}} = \mathbf{A}\vec{\mathbf{N}}^g$, $P(\vec{\mathbf{N}}|\vec{\mathbf{N}}^g)$ can be represented as a Kronecker delta function with the condition $\vec{\mathbf{N}} = \mathbf{A}\vec{\mathbf{N}}^g$, i.e.,

$$P(\vec{\mathbf{N}}=\vec{\mathbf{n}}|\vec{\mathbf{N}}^g=\vec{\mathbf{n}}^g,M) = P(\vec{\mathbf{N}}=\vec{\mathbf{n}}|\vec{\mathbf{N}}^g=\vec{\mathbf{n}}^g) = \delta_K(\vec{\mathbf{n}}=\mathbf{A}\vec{\mathbf{n}}^g), \quad (216)$$

where $\delta_K(\cdot)$ is the Kronecker delta function. Substituting (211), (215) and (216) in (207), we get:

$$\begin{aligned} P(\vec{\mathbf{N}}^g|\vec{\mathbf{N}},M) &= \frac{De^{-\vec{\mathbf{I}}^T\vec{\Lambda}}M^{\vec{\mathbf{I}}^T\mathbf{A}\vec{\mathbf{n}}^g}\delta_K(\vec{\mathbf{n}}=\mathbf{A}\vec{\mathbf{n}}^g)}{Be^{-\vec{\mathbf{I}}^T\vec{\Lambda}}M^{\vec{\mathbf{I}}^T\vec{\mathbf{n}}}} \stackrel{(a)}{=} \frac{DM^{\vec{\mathbf{I}}^T\mathbf{A}\vec{\mathbf{n}}^g}\delta_K(\vec{\mathbf{n}}=\mathbf{A}\vec{\mathbf{n}}^g)}{BM^{\vec{\mathbf{I}}^T\vec{\mathbf{n}}}} \\ &\stackrel{(b)}{=} \frac{DM^{\vec{\mathbf{I}}^T\mathbf{A}\vec{\mathbf{n}}^g}\delta_K(\vec{\mathbf{n}}=\mathbf{A}\vec{\mathbf{n}}^g)}{BM^{\vec{\mathbf{I}}^T\mathbf{A}\vec{\mathbf{n}}^g}} \stackrel{(c)}{=} \frac{D\delta_K(\vec{\mathbf{n}}=\mathbf{A}\vec{\mathbf{n}}^g)}{B}, \end{aligned} \quad (217)$$

where we obtain (a) by canceling the term $e^{-\vec{1}^T \vec{\Lambda}}$, (b) by using the fact $\vec{n} = \mathbf{A} \vec{n}^g$ due to the delta function, and (c) by canceling the term $M^{\vec{1}^T \mathbf{A} \vec{n}^g}$. Since the terms D, B and $\delta_K(\vec{n} = \mathbf{A} \vec{n}^g)$ do not depend upon M , we can conclude $P(\vec{N}^g | \vec{N}, M)$ also does not depend upon M , i.e. $P(\vec{N}^g | \vec{N}, M) = P(\vec{N}^g | \vec{N})$, or $M \rightarrow \vec{N} \rightarrow \vec{N}^g$. \square

Lemma 18. Let $M \sim P(M)$ have support over $(0, \infty)$. Define, $\vec{\mathbf{X}} = [\vec{\mathbf{X}}_1^T \dots \vec{\mathbf{X}}_{d_1}^T]^T$ and $\vec{\mathbf{Y}} = [\vec{\mathbf{Y}}_1 \dots \vec{\mathbf{Y}}_{d_2}^T]^T$, where $\vec{\mathbf{X}}_i$ and $\vec{\mathbf{Y}}_j$ are random vectors of size $q_i \in \mathbb{N}$ and $r_j \in \mathbb{N}$, respectively, $\forall i \in [d_1]$, and $j \in [d_2]$. Let the conditional distribution of $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$ conditioned on M be specified as follows:

$$\begin{aligned} P(\vec{\mathbf{X}} | M) &= \prod_{i=1}^{d_1} P(\vec{\mathbf{X}}_i | M), \text{ and } P(\vec{\mathbf{Y}} | M) = \prod_{i=1}^{d_2} P(\vec{\mathbf{Y}}_i | M), \text{ where} \\ P(\vec{\mathbf{X}}_i | M) &= \text{Poisson}(q_i, 1, \vec{\Lambda}_i^X), \vec{\Lambda}_i^X = [\gamma_{i1}^X M^{i_1} \dots \gamma_{iq_i}^X M^{i_{q_i}}]^T \forall i \in [d_1], \text{ and} \\ P(\vec{\mathbf{Y}}_j | M) &= \text{Poisson}(r_j, 1, \vec{\Lambda}_j^Y), \vec{\Lambda}_j^Y = [\gamma_{j1}^Y M^{j_1} \dots \gamma_{jr_j}^Y M^{j_{r_j}}]^T \forall j \in [d_2]. \end{aligned} \quad (218)$$

If $d_1 \geq d_2$ and (219) hold

$$\sum_{j=1}^{q_i} \gamma_{ij}^X \geq \sum_{j=1}^{r_i} \gamma_{ij}^Y \forall i \in [d_2] \quad (219)$$

then the distribution $\tilde{Q}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$ defined in (220) lies in Δ_P .

$$\tilde{Q}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}}) = P(M) P(\vec{\mathbf{X}} | M) \tilde{Q}(\vec{\mathbf{Y}} | \vec{\mathbf{X}}), \quad (220)$$

where $\tilde{Q}(\vec{\mathbf{Y}} | \vec{\mathbf{X}})$ is a product of d_2 multinomial distributions, i.e. $\tilde{Q}(\vec{\mathbf{Y}} | \vec{\mathbf{X}}) = \prod_{i=1}^{d_2} \tilde{Q}(\vec{\mathbf{Y}}_i | \vec{\mathbf{X}}_i)$, where $\tilde{Q}(\vec{\mathbf{Y}}_i | \vec{\mathbf{X}}_i) = \text{Multinomial}(\vec{\mathbf{k}}_i; N_i, \vec{\mathbf{p}}_i)$ and

$$\begin{aligned} N_i &= \sum_{j=1}^{n_i} x_{ij}, \vec{\mathbf{k}}_i = [y_{i1} \dots y_{im_i} \sum_{j=1}^{n_i} x_{ij} - \sum_{j=1}^{m_i} y_{ij}]^T, \\ \vec{\mathbf{p}}_i &= \left[\frac{\gamma_{i1}^Y}{\sum_{j=1}^{n_i} \gamma_{ij}^X} \dots \frac{\gamma_{im_i}^Y}{\sum_{j=1}^{n_i} \gamma_{ij}^X} 1 - \frac{\sum_{j=1}^{m_i} \gamma_{ij}^Y}{\sum_{j=1}^{n_i} \gamma_{ij}^X} \right]^T. \end{aligned}$$

Proof. Lemma 18 follows from a straightforward application of lemma 16. For showing $\tilde{Q}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$ specified in (220) lies in Δ_P , we need to show that $\tilde{Q}(M, \vec{\mathbf{X}}) = P(M, \vec{\mathbf{X}})$ and $\tilde{Q}(M, \vec{\mathbf{Y}}) = P(M, \vec{\mathbf{Y}})$. $\tilde{Q}(M, \vec{\mathbf{X}}) = P(M, \vec{\mathbf{X}})$ is satisfied trivially by construction of $\tilde{Q}(M, \vec{\mathbf{X}}, \vec{\mathbf{Y}})$.

Proving $\tilde{Q}(\vec{\mathbf{Y}}, M) = P(\vec{\mathbf{Y}}, M)$:

For showing the second condition, we will explicitly show $\tilde{Q}(\vec{\mathbf{Y}} | M) = P(\vec{\mathbf{Y}} | M)$ which, combined with the fact that $\tilde{Q}(M) = P(M)$ (by construction), will show $\tilde{Q}(\vec{\mathbf{Y}}, M) = P(\vec{\mathbf{Y}}, M)$ completing our proof.

Calculating $\tilde{Q}(\vec{Y}|\vec{X})$: From the law of total probability, we have:

$$\begin{aligned}
\tilde{Q}(\vec{Y}|M) &= \sum_{\vec{x}} P(\vec{X} = \vec{x}|M) \tilde{Q}(\vec{Y}|\vec{X} = \vec{x}) \\
&\stackrel{(a)}{=} \sum_{\vec{x}} \prod_{i=1}^{d_1} P(\vec{X}_i = \vec{x}_i|M) \prod_{i=1}^{d_2} \tilde{Q}(\vec{Y}_i|\vec{X}_i = \vec{x}_i) \\
&\stackrel{(b)}{=} \sum_{\vec{x}} \prod_{i=1}^{d_2} P(\vec{X}_i = \vec{x}_i|M) \tilde{Q}(\vec{Y}_i|\vec{X}_i = \vec{x}_i) \prod_{i=d_2+1}^{d_1} P(\vec{X}_i = \vec{x}_i|M) \\
&\stackrel{(c)}{=} \prod_{i=1}^{d_2} \left(\sum_{\vec{x}_i} P(\vec{X}_i = \vec{x}_i|M) \tilde{Q}(\vec{Y}_i|\vec{X}_i = \vec{x}_i) \right) \prod_{i=d_2+1}^{d_1} \underbrace{\left(\sum_{\vec{x}_i} P(\vec{X}_i = \vec{x}_i|M) \right)}_{=1} \\
&\stackrel{(d)}{=} \prod_{i=1}^{d_2} \underbrace{\left(\sum_{\vec{x}_i} P(\vec{X}_i = \vec{x}_i|M) \tilde{Q}(\vec{Y}_i|\vec{X}_i = \vec{x}_i) \right)}_{=\tilde{Q}(\vec{Y}_i|M)} = \prod_{i=1}^{d_2} \tilde{Q}(\vec{Y}_i|M), \tag{221}
\end{aligned}$$

where (a) is due to the particular structure of $P(\vec{X}|M)$ and $\tilde{Q}(\vec{Y}|\vec{X})$ specified in (218) and (220), respectively, (b) is just collecting the terms of $P(\vec{X}_i|M)$ and $\tilde{Q}(\vec{Y}_i|\vec{X}_i)$ into a single product, (c) is due to spreading the summation of the components of \vec{x} over their corresponding distribution components, and (d) is due to the fact that sum over all probabilities of all possible \vec{x}_i is 1.

Note that for a fixed m , $P(\vec{X}_i|M = m) \sim \text{Poisson}(n_i, 1, [\gamma_{i1}m^i \dots \gamma_{im_i}m^i]^T)$, and $Q(\vec{Y}_i|\vec{X}_i) = \text{Multinomial}(\vec{k}_i; N_i, \vec{p}_i)$, where

$$\begin{aligned}
N_i &= \sum_{j=1}^{n_i} x_{ij}, \quad \vec{k}_i = [y_{i1} \dots y_{im_i} \sum_{j=1}^{n_i} x_{ij} - \sum_{j=1}^{m_i} y_{ij}]^T, \\
\vec{p}_i &= \left[\frac{\gamma_{i1}^Y}{\sum_{j=1}^{n_i} \gamma_{ij}^X} \dots \frac{\gamma_{im_i}^Y}{\sum_{j=1}^{n_i} \gamma_{ij}^X} 1 - \frac{\sum_{j=1}^{m_i} \gamma_{ij}^Y}{\sum_{j=1}^{n_i} \gamma_{ij}^X} \right]^T.
\end{aligned}$$

Note that we can equivalently write \vec{p}_i as follows,

$$\vec{p}_i = \left[\frac{\gamma_{i1}^Y M^i}{\sum_{j=1}^{n_i} \gamma_{ij}^X M^i} \dots \frac{\gamma_{im_i}^Y M^i}{\sum_{j=1}^{n_i} \gamma_{ij}^X M^i} 1 - \frac{\sum_{j=1}^{m_i} \gamma_{ij}^Y M^i}{\sum_{j=1}^{n_i} \gamma_{ij}^X M^i} \right]^T. \tag{222}$$

In the setup of lemma 16, choose $d_1 = q_i$ and $d_2 = r_i$, $\lambda_j^{(2)} = \gamma_{ij}^Y M^i \forall j \in [r_i]$ and $\lambda_j^{(1)} = \gamma_{ij}^X M^i \forall j \in [q_i]$. From (219), we know the condition $\sum_{j=1}^{d_2} \lambda_j^{(2)} \leq \sum_{j=1}^{d_2} \lambda_j^{(2)}$ is satisfied by assumption. Hence, we use the result of lemma 15 to conclude $\tilde{Q}(\vec{Y}_i|M) = \text{Poisson}(r_i, 1, \vec{\Lambda}')$, with

$$\vec{\Lambda}' = [\lambda_{i1} \dots \lambda_{ir_i}]^T, \text{ with } \lambda_{ij} = \gamma_{ij}^Y M^i. \tag{223}$$

Comparing $\tilde{Q}(\vec{Y}_i|M)$ specified in (223) with $P(\vec{Y}_i|M)$ specified in (218), we can conclude $Q(\vec{Y}_i|M) = P(\vec{Y}_i|M)$. Substituting $Q(\vec{Y}_i|M) = P(\vec{Y}_i|M)$ in (221), we obtain:

$$\tilde{Q}(\vec{Y}|M) = \prod_{i=1}^{d_2} P(\vec{Y}_i|M). \tag{224}$$

Comparing (224) with (218), we can conclude $\tilde{Q}(\vec{Y}|M) = P(\vec{Y}|M)$, finishing our proof. \square

K Proof of Theorem 6

In this section, we provide the proof of theorem 6. Since theorem 6 provides the analytical PID terms for the linear convolution-closed system defined in Sec. 5.1, we briefly restate certain key properties of convolution-closed distributions and the linear convolution-closed system for convenience.

K.1 Convolution-closed distributions

Convolution-closed distributions are a large class of distributions that are defined as follows:

Definition 1. Let $\mathcal{F}_{\mathcal{D}}$ denote a family of distributions, where each member distribution, $f(\delta) \in \mathcal{F}_{\mathcal{D}}$, is indexed by a parameter δ . Consider $X_1 \sim f(\delta_1)$, $X_2 \sim f(\delta_2)$, and $X_1 \perp\!\!\!\perp X_2$ for some $\delta_1, \delta_2 \in \mathcal{D}$. Then, $\mathcal{F}_{\mathcal{D}}$ is convolution-closed in the parameter δ , if

$$X_1 + X_2 \sim f(\delta_1) * f(\delta_2) = f(\delta_1 + \delta_2) \quad \forall \delta_1, \delta_2 \in \mathcal{D} \text{ such that } \delta_1 + \delta_2 \in \mathcal{D}, \quad (225)$$

where $*$ denotes the convolution operator.

Convolution-closed distributions define a natural dilation/thinning operator. Formally, let $X \sim f(\delta)$. Then, we define X_ϵ as the ϵ -dilated version of X if $X \sim f(\epsilon\delta)$ for some $\epsilon \in (0, 1)$ such that $\epsilon\delta \in \mathcal{D}$. Furthermore, if we assume $(1 - \epsilon)\delta \in \mathcal{D}$, then $P(X_\epsilon|X)$ can be defined as $P(X_\epsilon|X_\epsilon + X_{(1-\epsilon)})$, where $X_{1-\epsilon} \sim f((1 - \epsilon)\delta)$ and $X_\epsilon \perp\!\!\!\perp X_{1-\epsilon}$. Denote, $P(X_\epsilon|X = x) = P(X_\epsilon|X_\epsilon + X_{(1-\epsilon)} = x) = G(\epsilon\delta, (1 - \epsilon)\delta, x)$.

K.2 Definition of linear convolution-closed system

Let $\mathcal{F}_{\mathcal{D}}$ be a convolution-closed distribution family in parameter $\delta \in \mathcal{D}$, and M be a target/message random variable having distribution $P(M)$ over some support set \mathcal{M} . Define the conditional distribution of random variables X and Y conditioned on M as follows:

$$P(X|M=m) = f(\delta_m^X) \text{ and } P(Y|M=m) = f(\delta_m^Y) \text{ such that } \delta_m^X, \delta_m^Y \in \mathcal{D} \quad \forall m \in \mathcal{M}. \quad (226)$$

Furthermore, we assume that there exists $\gamma \in (0, \infty)$ such that: $\delta_m^X = \gamma\delta_m^Y \quad \forall m \in \mathcal{M}$. We denote such a system of random variables (M, X, Y) , having the joint distribution $P(M, X, Y)$, a *linear convolution-closed system*.

K.3 Formal proof of theorem 6

Theorem 6. Let the joint density $P(M, X, Y)$ of random variables M , X , and Y describe a linear convolution-closed system. Without the loss of generality, assume $\gamma \leq 1$. If

- (a) $(1 - \gamma)\delta_m^Y \in \mathcal{D} \quad \forall m \in \mathcal{M}$,
- (b) $P(X_\gamma|X_\gamma + X_{1-\gamma}=x, M=m) = G(\gamma\delta_m^X, (1 - \gamma)\delta_m^X, x)$ does not depend on m , where $P(X_\gamma|M) = f(\gamma\delta_m^X)$, $P(X_{1-\gamma}|M) = f((1 - \gamma)\delta_m^X)$ and $X_\gamma \perp\!\!\!\perp X_{1-\gamma}|M$,

then Δ_P contains a Markov chain of the form $M \rightarrow X \rightarrow Y$ and $UI(M; Y \setminus X) = 0$.

Proof. We first note that we can always assume $\gamma \leq 1$ without the loss of generality because if $\gamma \geq 1$, then we can always switch our nomenclature to refer to Y as X , and X as Y .

We briefly outline the proof structure.

1. In the first part, we explicitly construct a joint distribution $Q_{MC}(M, X, Y)$ having the Markovian structure $M \rightarrow X \rightarrow Y$.
2. In the second part, we show that the $Q_{MC}(M, X, Y)$ constructed in the first part, lies in Δ_P . Therefore, we can then apply the result of proposition 1 to conclude $UI(M; Y \setminus X) = 0$.

Part 1: Specifying $Q_{MC}(M, X, Y)$

Denote the joint distribution of the Markov chain $M \rightarrow X \rightarrow Y$ as $Q_{MC}(M, X, Y)$. We can decompose $Q_{MC}(M, X, Y)$ by utilizing its Markovian structure as follows:

$$Q_{MC}(M, X, Y) = Q_{MC}(M)Q_{MC}(X|M)Q_{MC}(Y|X).$$

Consequently, we can construct the distribution $Q_{MC}(M, X, Y)$ by individually specifying $Q_{MC}(M)$, $Q_{MC}(X|M)$, and $Q_{MC}(Y|X)$.

Specifying $Q_{MC}(M)$ and $Q_{MC}(X|M)$: We specify $Q_{MC}(M)$ and $Q_{MC}(X|M)$ as follows:

$$Q_{MC}(M) = P(M) \text{ and } Q_{MC}(X|M) = P(X|M). \quad (227)$$

Specifying $Q_{MC}(Y|X)$: By assumption (b) in the theorem, we know that $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, x)$ does not depend on m , and consequently, does not depend on δ_m^X . Hence, we can simplify notation of $G(\gamma\delta_m^X, (1-\gamma)\delta_m^X, m)$ as $G(\gamma, (1-\gamma), x)$. We specify $Q(Y|X)$ as follows:

$$Q(Y|X = x) = G(\gamma, 1 - \gamma, x). \quad (228)$$

Part 2: Showing $Q_{MC}(M, X, Y) \in \Delta_P$

For showing that $Q_{MC}(M, X, Y) \in \Delta_P$, we first need to derive $Q_{MC}(Y|M)$.

Deriving $Q_{MC}(Y|M)$: We use the result of lemma 19 to derive $Q_{MC}(Y|M)$. As $Q_{MC}(X|M = m) = f(\delta_m^X)$ is a convolution-closed distribution for a fixed m , and $Q_{MC}(Y = y|X = x) = G(\gamma, 1 - \gamma, x)$, the result of lemma 19 shows that $Q_{MC}(Y|M)$ can be expressed as:

$$Q_{MC}(Y|M = m) = f(\gamma \times \delta_m^X) = f(\gamma\delta_m^X) \stackrel{(a)}{=} f(\delta_m^Y) = P(Y|M = m), \quad (229)$$

where (a) is due to the assumption $\delta_m^X = \gamma\delta_m^Y$ (see Appx. K.2).

From (227) and (229), we can conclude:

$$Q_{MC}(M, X) = Q_{MC}(M)Q_{MC}(X|M) = P(M)P(X|M) = P(M, X), \quad (230)$$

$$Q_{MC}(M, Y) = Q_{MC}(M)Q_{MC}(Y|M) = P(M)P(Y|M) = P(M, Y). \quad (231)$$

Hence, $Q_{MC}(M, X, Y) \in \Delta_P$ and consequently, by proposition 1, $UI(M; Y \setminus X) = 0$, concluding our proof for case 1. □

Lemma 19. Let $\mathcal{F}_{\mathcal{D}}$ be a convolution-closed distribution family convolution-closed in the parameter δ . For a $\delta \in \mathcal{D}$, assume there exists $\epsilon \in (0, 1)$ such that $\epsilon\delta, (1-\epsilon)\delta \in \mathcal{D}$. Consider the following random variables:

$$X'_\epsilon \sim f(\epsilon\delta) \in \mathcal{F}_{\mathcal{D}}, \quad X'_{1-\epsilon} \sim f((1-\epsilon)\delta) \in \mathcal{F}_{\mathcal{D}}, \quad \text{and } X \sim f(\delta) \in \mathcal{F}_{\mathcal{D}},$$

such that $X'_\epsilon \perp\!\!\!\perp X'_{1-\epsilon}$. If $P(X_\epsilon = x_\epsilon | X = x) = P(X'_\epsilon = x_\epsilon | X'_\epsilon + X'_{1-\epsilon} = x)$, then we have $P(X_\epsilon = x_\epsilon) = P(X'_\epsilon = x_\epsilon) = f(x_\epsilon; \epsilon\delta)$ or $X_\epsilon \stackrel{d}{=} X'_\epsilon$, where $\stackrel{d}{=}$ implies equality in distribution [57].

Proof. First let us derive the distribution of $X' = X'_{1-\epsilon} + X'_\epsilon$. Since, $X'_\epsilon \perp\!\!\!\perp X'_{1-\epsilon}$, we can write $P(X')$ as follows:

$$P(X') = f(\epsilon\delta) * f((1-\epsilon)\delta) \stackrel{(a)}{=} f(\epsilon\delta + (1-\epsilon)\delta) = f(\delta) \stackrel{(b)}{=} P(X), \quad (232)$$

where $*$ is the convolution operator, (a) is due to the properties of convolution-closed distribution (see Appx. K.1), and (b) is due to the definition of X in the lemma statement. From (232), we can conclude $P(X = x) = P(X' = x)$, and by assumption we have $P(X_\epsilon = x_\epsilon | X = x) = P(X'_\epsilon = x_\epsilon | X' = x) = P(X'_\epsilon = x_\epsilon | X'_\epsilon + X'_{1-\epsilon} = x)$. Hence, we have:

$$\begin{aligned} P(X = x, X_\epsilon = x_\epsilon) &= P(X = x)P(X_\epsilon = x_\epsilon | X = x) = P(X' = x)P(X'_\epsilon = x_\epsilon | X' = x), \\ &= P(X' = x)P(X'_\epsilon = x_\epsilon | X'_\epsilon + X'_{1-\epsilon} = x) \\ &= P(X'_\epsilon + X'_{1-\epsilon} = x, X'_\epsilon = x_\epsilon). \end{aligned} \quad (233)$$

From (233), we can conclude that $P(X_\epsilon) = P(X'_\epsilon) = f(\epsilon\delta)$. □

L Proof of Theorem 7

In this section, we provide the proof of theorem 7.

L.1 Definition of p_{exp1} and p_{exp2}

We restate the definitions of p_{exp1} and p_{exp2} defined in Sec. 5.2, which is used in the proof of theorem 7. Let $X \sim p_{exp1}(X)$, where

$$p_{exp1}(X = x; \theta_1, \theta_2) = H(\theta_1, \theta_2) \exp(\theta_1^T x - \theta_2^T A(x)), \quad (234)$$

for some appropriately defined $H(\cdot, \cdot)$, $A(\cdot)$, θ_1 and θ_2 . Furthermore, define a random variable Y through its conditional density $p(Y|X = x)$:

$$p(Y = y|X = x; \theta_1, \theta_2) = h(y) \exp(x^T T(y) - \theta_3^T A(x)), \quad (235)$$

for some $h(\cdot)$, $T(\cdot)$, and θ_3 , such that $p(Y = y|X = x; \theta_1, \theta_2)$ is a well-defined distribution. Furthermore, the marginal distribution of Y is expressed as:

$$p_{exp2}(Y = y; \theta_1, \theta_2, \theta_3) = h(y) \frac{H(\theta_1, \theta_2)}{H(\theta_1 + T(y), \theta_2 + \theta_3)} \text{ (see proof of theorem 1 in [26])}. \quad (236)$$

L.2 Formal proof of theorem 7

Theorem 7. *Let M, X , and Y be random variables having the joint distribution $P(M, X, Y)$. Furthermore, the conditional distribution of X and Y conditioned on M are as follows: $P(X|M=m) = p_{exp1}(X; \theta_1(m), \theta_2(m))$ and $P(Y|M=m) = p_{exp2}(Y; \theta_1(m), \theta_2(m), \theta_3)$. Then, Δ_P contains a Markov chain of the form $M \rightarrow X \rightarrow Y$ and $UI(M; Y \setminus X) = 0$.*

Proof. We prove theorem 7 by explicitly constructing a Markov chain $M \rightarrow X \rightarrow Y$, having joint distribution $Q_{MC}(M, X, Y)$, and showing that $Q_{MC}(M, X, Y) \in \Delta_P$.

Specifying $Q_{MC}(M)$ and $Q_{MC}(X|M)$: We specify $Q_{MC}(M)$ and $Q_{MC}(X|M)$ as follows:

$$Q_{MC}(M) = P(M), \text{ and } Q_{MC}(X|M) = P(X|M). \quad (237)$$

Specifying $Q_{MC}(Y|X)$: We specify $Q_{MC}(Y|X) = h(y) \exp(x^T T(y) - \theta_3^T A(x))$. Then, by (236), we know that

$$Q(Y = y|M = m) = p_{exp2}(y; \theta_1(m), \theta_2(m), \theta_3) = \frac{h(y)H(\theta_1(m), \theta_2(m))}{H(\theta_1(m) + T(y), \theta_2(m) + \theta_3(m))}. \quad (238)$$

From (237) and (238), we can conclude:

$$Q_{MC}(M, Y) = Q_{MC}(Y|M)Q_{MC}(M) = P(Y|M)P(M) = P(M, Y), \quad (239)$$

$$Q_{MC}(M, X) = Q_{MC}(X|M)Q_{MC}(M) = P(X|M)P(M) = P(M, X). \quad (240)$$

Hence, $Q_{MC}(M, X, Y) \in \Delta_P$ and consequently, by proposition 1, $UI(M; Y \setminus X) = 0$, concluding our proof. \square

M Additional proofs and details for Sec. 6

In this section, we provide the proofs of certain statements made in Sec. 6, as well as additional details regarding the simulation study presented in the same section.

M.1 Generating \bar{Q} for every $Q \in \Delta_P$

Our goal in this section is to show that there exists a \bar{Q} (as described in Sec. 6) for every $Q \in \Delta_P$, where the corresponding $P(M, X, Y)$ is as defined in Sec. 6. We first briefly restate the assumptions on $P(M, X, Y)$ described in Sec. 6. The distribution $P(M, X, Y)$ has the following properties: the random variable M has support over \mathcal{M} , the conditional distributions $P(X|M)$ and $P(Y|M)$ are

members of some convolution-closed distribution family $\mathcal{F}_{\mathcal{D}}$, and there exists some $\delta_{bias}^X, \delta_{bias}^Y \in \mathcal{D}$ such that:

$$\begin{aligned} P(X|M=m) &= f(\delta_m^X) \text{ and } P(Y|M=m) = f(\delta_m^Y) \text{ where } \delta_m^X, \delta_m^Y \in \mathcal{D} \forall m \in \mathcal{M}, \text{ with} \\ \delta_m^X - \delta_{bias}^X &= \epsilon_m^{(1)}(\delta_m^Y - \delta_{bias}^Y) \text{ for some } \epsilon_m^{(1)} \in [0, 1] \forall m \in \mathcal{M}, \\ \delta_m^Y &= \epsilon_m^{(2)}(\delta_m^X - \delta_{bias}^X) \text{ for some } \epsilon_m^{(2)} \in [0, 1] \forall m \in \mathcal{M}, \\ \delta_m^X &= \epsilon_m^{(3)}(\delta_m^X - \delta_{bias}^X) \text{ for some } \epsilon_m^{(3)} \in [0, 1] \forall m \in \mathcal{M}, \text{ and} \\ (\delta_m^Y - \delta_{bias}^Y), (\delta_m^X - \delta_{bias}^X), (\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)) &\in \mathcal{D}. \end{aligned} \quad (241)$$

As $Q(M, X) = P(M, X)$ and $Q(M, Y) = P(M, Y)$ due to $Q \in \Delta_P$, we have:

$$Q(X|M=m) = f(\delta_m^X), \text{ and } Q(Y|M=m) = f(\delta_m^Y) \text{ where } \delta_m^X, \delta_m^Y \in \mathcal{D} \forall m \in \mathcal{M}. \quad (242)$$

M.1.1 Explicitly constructing a \bar{Q} for every $Q \in \Delta_P$

We show existence of a \bar{Q} for every $Q \in \Delta_P$ by explicitly constructing a \bar{Q} for each $Q \in \Delta_P$.

We define two vectors to simplify our notations:

$$\vec{\mathbf{A}} = [X \quad Y]^T, \quad \vec{\mathbf{B}} = [X' \quad Y'' \quad n_X \quad Y' \quad n_Y]^T, \quad (243)$$

where X and Y are the random variables defined above. (X', Y'', n_X) , and (Y', n_Y) are the corresponding dilated versions of X and Y as defined in Sec. 6. Note that the joint distribution of $(M, X', Y'', n_X, Y', n_Y)$ is $\bar{Q}(M, X', Y'', n_X, Y', n_Y)$. We reiterate the two main properties of $\bar{Q}(X', Y'', n_X, Y', n_Y)$, mentioned in Sec. 6, that we want to prove:

1. The marginal conditional distributions of (X', Y'', n_X, Y', n_Y) given M are as follows:

$$\begin{aligned} \bar{Q}(X'|M=m) &= f(\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)), \quad \bar{Q}(Y''|M=m) = f(\delta_m^Y - \delta_{bias}^Y), \\ \bar{Q}(Y'|M=m) &= f(\delta_m^Y - \delta_{bias}^Y), \quad \bar{Q}(n_X|M) = f(\delta_{bias}^X), \quad \bar{Q}(n_Y|M) = f(\delta_{bias}^Y). \end{aligned}$$

2. The random variables (X', Y'', n_X) are jointly conditionally independent given M . Similarly, the random variables (Y', n_Y) are also conditionally independent given M .

We construct $\vec{\mathbf{B}}$ from $\vec{\mathbf{A}}$ by specifying the conditional distribution $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$. Note that given the values of $\vec{\mathbf{A}}$ and M , we can always use $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$ to construct the random variables $\vec{\mathbf{B}}$. The goal of this remaining section is to show that the corresponding distribution $\bar{Q}(\vec{\mathbf{B}}, M)$ (specified through $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$) satisfies the above two properties. We first explicitly state the structure of $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$.

Specifying $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$:

We first impose the following structure on $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$ as follows:

$$\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M) = \bar{Q}(X', Y'', n_X|X, M) \bar{Q}(Y', n_Y|Y, M), \quad (244)$$

where $\bar{Q}(X', Y'', n_X|X, M)$ and $\bar{Q}(Y', n_Y|Y, M)$ are the conditional distributions through which we specify $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$. Since $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$ is a product of two well-defined distributions, it is a valid distribution. Let us now analyze the joint distribution $\bar{Q}(M, \vec{\mathbf{A}}, \vec{\mathbf{B}})$. By chain rule, we have:

$$\bar{Q}(M, \vec{\mathbf{A}}, \vec{\mathbf{B}}) = \bar{Q}(M, \vec{\mathbf{A}}) \bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M). \quad (245)$$

Substituting the fact that $(M, \vec{\mathbf{A}}) = (M, X, Y)$ and using the structure of $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$ specified in (244), we have:

$$\bar{Q}(M, \vec{\mathbf{A}}, \vec{\mathbf{B}}) = \bar{Q}(M) \bar{Q}(X, Y|M) \bar{Q}(X', Y'', n_X|X, M) \bar{Q}(Y', n_Y|Y, M). \quad (246)$$

We now specify the two components of $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$.

Specifying $\bar{Q}(Y', n_Y|Y, M)$:

Using the chain rule on $\bar{Q}(Y', n_Y|Y, M)$, we obtain

$$\bar{Q}(Y', n_Y|Y, M) = \bar{Q}(Y'|Y, M)\bar{Q}(n_Y|Y, Y', M). \quad (247)$$

Choose:

$$\begin{aligned} \bar{Q}(Y'|Y=y, M=m) &= G(\epsilon_m^{(2)}\delta_m^Y, (1-\epsilon_m^{(2)})\delta_m^Y, y), \text{ where } \delta_m^Y = \epsilon_m^{(2)}(\delta_m^Y - \delta_{bias}^Y) \forall m \in \mathcal{M}, \text{ and} \\ \bar{Q}(n_Y|Y=y, Y'=y', M=m) &= \delta_K(n_Y = y - y'). \end{aligned} \quad (248)$$

Here, $\delta_K(\cdot)$ denotes the Kronecker delta function [59]. Marginalizing (X, X', Y'', n_X) from (246), we obtain:

$$\bar{Q}(M, Y, Y', n_Y) = \bar{Q}(M)\bar{Q}(Y|M)\bar{Q}(Y', n_Y|M, Y) = \bar{Q}(M)\bar{Q}(Y, Y', n_Y|M). \quad (249)$$

Consider the distribution $\bar{Q}(Y, Y', n_Y|M)$. Then, we know:

$$\begin{aligned} \bar{Q}(Y|M=m) &= f(\delta_m^Y), \\ \bar{Q}(n_Y|Y=y, Y'=y', M=m) &= \delta_K(n_Y = y - y'), \text{ and} \\ \bar{Q}(Y'=y'|Y=y, M=m) &= G(\epsilon_m^{(2)}\delta_m^Y, (1-\epsilon_m^{(2)})\delta_m^Y, y). \end{aligned} \quad (250)$$

Comparing the above equation with the structure of $\bar{Q}(Y, Y', n_Y|M)$ specified in proposition 8, we can conclude that (Y', n_Y) are just dilated versions of Y . Consequently, using the result of proposition 8, we have:

$$\bar{Q}(Y'|M=m) = f(\delta_m^Y - \delta_{bias}^Y), \bar{Q}(n_Y|M) = f(\delta_{bias}^Y), \text{ and } Y' \perp\!\!\!\perp n_Y|M. \quad (251)$$

Specifying $\bar{Q}(X', Y'', n_X|X, M)$:

We similarly decompose $\bar{Q}(X', Y'', n_X|X, M)$ using the chain rule as follows:

$$\bar{Q}(X', Y'', n_X|X, M) = \bar{Q}(n_X|X, M)\bar{Q}(X'|X, n_X, M)\bar{Q}(Y''|X, n_X, X', M). \quad (252)$$

We choose:

$$\begin{aligned} \bar{Q}(n_Y|X=x, M=m) &= G((1-\epsilon_m^{(3)})\delta_m^X, \epsilon_m^{(3)}\delta_m^X, x), \text{ where } \delta_m^X = \epsilon_m^{(3)}(\delta_m^X - \delta_{bias}^X) \forall m \in \mathcal{M}, \\ \bar{Q}(X'|X=x, n_X=x_{n_X}, M=m) &= G((1-\epsilon_m^{(1)})\delta_m^X - \delta_{bias}^X, \epsilon_m^{(1)}(\delta_m^X - \delta_{bias}^X), x - x_{n_X}), \text{ where} \\ \delta_m^X - \delta_{bias}^X &= \epsilon_m^{(1)}(\delta_m^Y - \delta_{bias}^Y) \forall m \in \mathcal{M}, \text{ and} \\ \bar{Q}(Y''|X=x, n_X=x_{n_X}, X'=x', M=m) &= \delta_K(n_X = x - x' - x_{n_X}). \end{aligned} \quad (253)$$

Here $\delta_K(\cdot)$ denotes the Kronecker delta function [59]. Marginalizing (Y, Y', n_Y) from (246), we obtain:

$$\bar{Q}(M, X, X', Y'', n_X) = \bar{Q}(M)\bar{Q}(X|M)\bar{Q}(X', n_X, Y''|M, X) = \bar{Q}(M)\bar{Q}(X, X', Y'', n_X|M).$$

Consider the distribution $\bar{Q}(X, X', n_X, Y''|M)$. Then, we know:

$$\begin{aligned} \bar{Q}(X|M=m) &= f(\delta_m^X) \text{ and} \\ \bar{Q}(n_Y|X, M), \bar{Q}(X'|X, n_X, M), \text{ and } \bar{Q}(Y''|X, n_X, X', M) &\text{ are as defined in (253).} \end{aligned} \quad (254)$$

Comparing the above equation with the structure of $\bar{Q}(X, X', n_X, Y''|M)$ specified in proposition 9, we can conclude that (X', Y'', n_X) are just dilated versions of X . Consequently, using the result of proposition 9, we have:

$$\begin{aligned} \bar{Q}(X'|M=m) &= f(\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)), \\ \bar{Q}(n_X|M) &= f(\delta_{bias}^X), \\ \bar{Q}(Y''|M) &= f(\delta_m^Y - \delta_{bias}^Y), \text{ and} \\ (X', Y'', n_X) &\text{ are jointly conditionally independent conditioned on } M. \end{aligned} \quad (255)$$

This concludes our construction of $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$. We obtain our desired $\bar{Q}(M, X', Y'', n_X, Y', n_Y)$ by marginalizing $\vec{\mathbf{A}}$ out of $\bar{Q}(M, \vec{\mathbf{A}}, \vec{\mathbf{B}})$. From (251) and (255) we know that $\bar{Q}(M, X', Y'', n_X, Y', n_Y)$ satisfies the two main properties specified earlier. Since we did not make any specific assumptions on Q aside from the fact that it lies in Δ_P , the above construction can always be used to construct (X', n_X, Y'', Y', n_Y) from (M, X, Y) using the $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$. Note that due to the particular construction of $\bar{Q}(\vec{\mathbf{B}}|\vec{\mathbf{A}}, M)$, we have:

$$X = X' + n_X + Y'' \text{ and } Y = Y' + n_Y. \quad (256)$$

M.1.2 Propositions required for constructing \bar{Q}

We describe two important propositions that we use to show the existence of a \bar{Q} for every $Q \in \Delta_P$.

Proposition 8. Suppose the random variables M and Y have the joint distribution $Q(M, Y)$ such that

$$Q(M) = P(M) \text{ and } Q(Y|M=m) = f(\delta_m^Y) \forall m \in \mathcal{M}, \quad (257)$$

where \mathcal{M} denotes the support of M and $f(\delta_m^Y) \in \mathcal{F}_D$ for some convolution-closed distribution family \mathcal{F}_D . Furthermore, there also exists a $\delta_{bias}^Y \in \mathcal{D}$ such that

$$\delta_m^Y = \epsilon_m^{(2)}(\delta_m^Y - \delta_{bias}^Y) \text{ for some } \epsilon_m^{(2)} \in [0, 1] \text{ and } (\delta_m^Y - \delta_{bias}^Y) \in \mathcal{D} \forall m \in \mathcal{M}. \quad (258)$$

Consider two dilated versions of Y , denoted as Y' and n_Y , defined as follows:

$$\bar{Q}(Y'|Y=y, M=m) = G(\epsilon_m^{(2)}\delta_m^Y, (1 - \epsilon_m^{(2)})\delta_m^Y, y), \text{ and } n_Y = Y - Y'. \quad (259)$$

Then, we have

$$1. Y' \perp\!\!\!\perp n_Y|M$$

2. The conditional distributions $\bar{Q}(Y'|M)$ and $\bar{Q}(n_Y|M)$ are specified as follows:

$$\begin{aligned} \bar{Q}(Y'|M=m) &= f(\epsilon_m^{(2)}\delta_m^Y) = f(\delta_m^Y - \delta_{bias}^Y) \text{ and} \\ \bar{Q}(n_Y|M=m) &= f((1 - \epsilon_m^{(2)})\delta_m^Y) = f(\delta_{bias}^Y) \forall m \in \mathcal{M}. \end{aligned} \quad (260)$$

Proof. We essentially modify the argument given in theorem 1 of [25] to prove the above proposition. Consider the conditional distribution of Y , Y' , and n_Y conditioned on M , denoted as $\bar{Q}(Y, Y', n_Y|M)$. Then, by chain rule, we have $\bar{Q}(Y, Y', n_Y|M) = \bar{Q}(Y|M)\bar{Q}(Y'|M, Y)\bar{Q}(n_Y|Y, Y', M)$. By the assumptions in (257) and (259), we know that

$$\begin{aligned} \bar{Q}(Y|M=m) &= f(\delta_m^Y), \bar{Q}(Y'|M=m, Y=y) = G(\epsilon_m^{(2)}\delta_m^Y, (1 - \epsilon_m^{(2)})\delta_m^Y, y), \text{ and} \\ \bar{Q}(n_Y|Y=y, Y'=y', M) &= \delta_K(n_Y = y - y'), \end{aligned} \quad (261)$$

where $\delta_K(\cdot)$ is the Kronecker delta function [59].

Observation 1: Our first key observation is that the distribution $\bar{Q}(Y', Y, n_Y|M=m)$ is completely specified through the distributions $\bar{Q}(Y|M)$ and $\bar{Q}(Y'|Y, M)$, as n_Y is completely determined by Y' and Y and $\bar{Q}(n_Y|Y, Y', M)$ is specified through a Kronecker delta function.

Similarly, consider a new pair of random variables Y'_{new}, n_Y^{new} such that $Y'_{new} \perp\!\!\!\perp n_Y^{new}|M$ and the respective conditional distributions of Y'_{new} and n_Y^{new} are specified as follows:

$$\bar{Q}_{new}(Y'_{new}|M=m) = f(\epsilon_m^{(2)}\delta_m^Y), \text{ and } \bar{Q}_{new}(n_Y^{new}|M=m) = f((1 - \epsilon_m^{(2)})\delta_m^Y) \forall m \in \mathcal{M}. \quad (262)$$

We will now show two key properties for the random variables $(Y'_{new}, n_Y^{new}, Y'_{new} + n_Y^{new})$ that will form the crux of our argument:

Property 1: $\bar{Q}_{new}(Y'_{new} + n_Y^{new} = y|M=m) = \bar{Q}(Y=y|M=m)$.

Let us now calculate the conditional distribution $\bar{Q}(Y'_{new} + n_Y^{new} = y|M)$:

$$\bar{Q}_{new}(Y'_{new} + n_Y^{new} = y|M) = \int_{k=-\infty}^{\infty} \bar{Q}_{new}(Y'_{new} = k, n_Y^{new} = y - k|M). \quad (263)$$

Using conditional dependence of Y'_{new} and n_Y^{new} , we have $\bar{Q}_{new}(Y'_{new} = k, n_Y^{new} = y - k|M) = \bar{Q}_{new}(Y'_{new} = k|M)\bar{Q}_{new}(n_Y^{new} = y - k|M)$:

$$\bar{Q}_{new}(Y'_{new} + n_Y^{new} = y|M) = \int_{k=-\infty}^{\infty} \bar{Q}_{new}(Y'_{new} = k|M)\bar{Q}_{new}(n_Y^{new} = y - k|M). \quad (264)$$

Note that the R.H.S of the above equation defines a convolution of $\bar{Q}_{new}(Y'_{new}|M)$ and $\bar{Q}(n_Y^{new}|M)$, hence:

$$\bar{Q}_{new}(Y'_{new} + n_Y^{new} = y|M) = \bar{Q}_{new}(Y'_{new} = k|M) * \bar{Q}_{new}(n_Y^{new} = y - k|M), \quad (265)$$

where $*$ denotes the convolution operator. Substituting (262) in the above equation:

$$\bar{Q}_{new}(Y'_{new} + n_Y^{new} = y | M = m) = f(\epsilon_m^{(2)} \delta_m^Y) * f((1 - \epsilon_m^{(2)}) \delta_m^Y). \quad (266)$$

Using the properties of convolution-closed distributions:

$$\bar{Q}_{new}(Y'_{new} + n_Y^{new} = y | M = m) = f(\epsilon_m^{(2)} \delta_m^Y) + (1 - \epsilon_m^{(2)}) \delta_m^Y = f(\delta_m^Y). \quad (267)$$

Hence, by comparing above equation with (261) we have:

$$\bar{Q}_{new}(Y'_{new} + n_Y^{new} = y | M = m) = Q(Y = y | M = m). \quad (268)$$

Property 2: $\bar{Q}(Y'_{new} = y' | Y'_{new} + n_Y^{new} = y, M = m) = G(\epsilon_m^{(2)} \delta_m^Y, 1 - \epsilon_m^{(2)} \delta_m^Y, y)$

By definition of $G(\epsilon_m^{(2)} \delta_m^Y, (1 - \epsilon_m^{(2)}) \delta_m^Y, z)$ given in Sec. 5.1, we have:

$$G(\epsilon_m^{(2)} \delta_m^Y, (1 - \epsilon_m^{(2)}) \delta_m^Y, z) = \Pr(Z_{\epsilon_m^{(2)}} | Z_{\epsilon_m^{(2)}} + Z_{1-\epsilon_m^{(2)}} = z), \text{ where} \\ \Pr(Z_{\epsilon_m^{(2)}}) = f(\epsilon_m^{(2)} \delta_m^Y), \Pr(Z_{1-\epsilon_m^{(2)}}) = f((1 - \epsilon_m^{(2)}) \delta_m^Y), \text{ and } Z_{\epsilon_m^{(2)}} \perp\!\!\!\perp Z_{1-\epsilon_m^{(2)}}. \quad (269)$$

In the following steps, we slightly modify the notation of $G(\epsilon_m^{(2)} \delta_m^Y, (1 - \epsilon_m^{(2)}) \delta_m^Y, z)$ to $G(z_{\epsilon_m^{(2)}}; \epsilon_m^{(2)} \delta_m^Y, (1 - \epsilon_m^{(2)}) \delta_m^Y, z)$ to make our notations more explicit as follows:

$$G(z_{\epsilon_m^{(2)}}; \epsilon_m^{(2)} \delta_m^Y, (1 - \epsilon_m^{(2)}) \delta_m^Y, z) = \Pr(Z_{\epsilon_m^{(2)}} = z_{\epsilon_m^{(2)}} | Z_{\epsilon_m^{(2)}} + Z_{1-\epsilon_m^{(2)}} = z). \quad (270)$$

We use the definition of the conditional distribution to express $G(\epsilon_m^{(2)} \delta_m^Y, (1 - \epsilon_m^{(2)}) \delta_m^Y, z)$ as:

$$G(z_{\epsilon_m^{(2)}}; \epsilon_m^{(2)} \delta_m^Y, (1 - \epsilon_m^{(2)}) \delta_m^Y, z) = \frac{\Pr(Z_{\epsilon} = z_{\epsilon_m^{(2)}} | Z_{\epsilon_m^{(2)}} + Z_{1-\epsilon_m^{(2)}} = z)}{\Pr(Z_{\epsilon_m^{(2)}} + Z_{1-\epsilon_m^{(2)}} = z)} \quad (271)$$

$$= \frac{\Pr(Z_{\epsilon} = z_{\epsilon_m^{(2)}}) \Pr(Z_{\epsilon_m^{(2)}} + Z_{1-\epsilon_m^{(2)}} = z | Z_{\epsilon_m^{(2)}} = z_{\epsilon_m^{(2)}})}{\Pr(Z_{\epsilon_m^{(2)}} + Z_{1-\epsilon_m^{(2)}} = z)}. \quad (272)$$

Using the fact that $\Pr(Z_{\epsilon_m^{(2)}} + Z_{1-\epsilon_m^{(2)}} = z | Z_{\epsilon_m^{(2)}} = z_{\epsilon_m^{(2)}}) = \Pr(Z_{1-\epsilon_m^{(2)}} = z - z_{\epsilon_m^{(2)}})$, since $Z_{\epsilon_m^{(2)}} \perp\!\!\!\perp Z_{1-\epsilon_m^{(2)}}$:

$$= \frac{\Pr(Z_{\epsilon} = z_{\epsilon_m^{(2)}}) \Pr(Z_{1-\epsilon_m^{(2)}} = z - z_{\epsilon_m^{(2)}})}{\Pr(Z_{\epsilon_m^{(2)}} + Z_{1-\epsilon_m^{(2)}} = z)}. \quad (273)$$

Note that by properties of convolution-closed distributions, $\Pr(Z_{\epsilon_m^{(2)}} + Z_{1-\epsilon_m^{(2)}} = z) = f(z; \delta_m^Y)$, where $f(z; \delta_m^Y)$ is to be understood as the distribution $f(\delta_m^Y)$ evaluated at z . Substituting (269) in the above equation:

$$= \frac{f(z_{\epsilon_m^{(2)}}; \epsilon_m^{(2)} \delta_m^Y) f(z - z_{\epsilon_m^{(2)}}; (1 - \epsilon_m^{(2)}) \delta_m^Y)}{f(z; \delta_m^Y)}. \quad (274)$$

Now, note that by (262) and (267), we have:

$$\bar{Q}_{new}(Y'_{new} = z_{\epsilon} | M = m) = f(z_{\epsilon_m^{(2)}}; \epsilon_m^{(2)} \delta_m^Y), \\ \bar{Q}_{new}(n_Y^{new} = z - z_{\epsilon} | M = m) = f(z - z_{\epsilon_m^{(2)}}; (1 - \epsilon_m^{(2)}) \delta_m^Y), \text{ and} \\ \bar{Q}_{new}(Y'_{new} + n_Y^{new} = z | M = m) = f(z; \delta_m^Y) \quad (275)$$

Hence, combining the above equation with (274), and renaming $z_{\epsilon_m^{(2)}}$ as y' and z as y , we have:

$$G(\epsilon_m^{(2)} \delta_m^Y, (1 - \epsilon_m^{(2)}) \delta_m^Y, y) = \frac{\bar{Q}_{new}(Y'_{new} = y' | M = m) \bar{Q}_{new}(n_Y^{new} = y - y' | M = m)}{\bar{Q}_{new}(Y'_{new} + n_Y^{new} = y | M = m)}. \quad (276)$$

Noting the fact that $\bar{Q}_{new}(n_Y^{new}=y-y'|M=m) = \bar{Q}_{new}(Y'_{new} + n_Y^{new}=y|M=m, Y'_{new} = y')$ due to $Y'_{new} \perp\!\!\!\perp n_Y^{new}|M$ and abbreviating notation pertaining to y' , we have:

$$G(\epsilon_m^{(2)}\delta_m^Y, (1 - \epsilon_m^{(2)})\delta_m^Y, y) = \frac{\bar{Q}_{new}(Y'_{new}|M=m)\bar{Q}_{new}(Y'_{new} + n_Y^{new}=y|M=m)}{\bar{Q}_{new}(Y'_{new} + n_Y^{new}=y|M=m)}. \quad (277)$$

Using Bayes rule in the above equation provides us with our desired property:

$$G(\epsilon_m^{(2)}\delta_m^Y, (1 - \epsilon_m^{(2)})\delta_m^Y, y) = \bar{Q}_{new}(Y'_{new}|M=m, Y'_{new} + n_Y^{new}=y) \quad (278)$$

Furthermore, by comparing (259) with the above equation, we have:

$$\bar{Q}_{new}(Y'_{new} = y'|Y'_{new} + n_Y^{new} = y, M = m) = \bar{Q}(Y' = y'|Y = y, M = m). \quad (279)$$

Observation 2: Note that the distribution $\bar{Q}(Y'_{new}, n_Y^{new}, Y'_{new} + n_Y^{new}|M)$ is also completely determined by specifying $\bar{Q}(Y'_{new} + n_Y^{new}|M)$ and $\bar{Q}(Y'_{new}|Y'_{new} + n_Y^{new}, M)$ since n_Y^{new} is completely determined by $Y'_{new} + n_Y^{new}$ and Y'_{new} .

By property 1 and (279), we know that

$$\bar{Q}_{new}(Y'_{new} + n_Y^{new}|M) = Q(Y|M), \text{ and } \bar{Q}_{new}(Y'_{new}|Y'_{new} + n_Y^{new}, M) = \bar{Q}(Y'|Y, M). \quad (280)$$

Therefore, combining the above equation with observations 1 and 2, we have:

$$\bar{Q}(Y', Y, n_Y|M) = \bar{Q}(Y'_{new}, Y'_{new} + n_Y^{new}, n_Y^{new}|M). \quad (281)$$

Consequently the distribution $\bar{Q}(Y', n_Y|M) = \bar{Q}_{new}(Y'_{new}, n_Y^{new}|M)$. Since, $Y'_{new} \perp\!\!\!\perp n_Y^{new}|M$, we also have $Y' \perp\!\!\!\perp n_Y|M$. Furthermore, by (262),

$$\begin{aligned} \bar{Q}_{new}(Y'_{new}|M=m) &= \bar{Q}(Y'|M) = f(\epsilon_m^{(2)}\delta_m^Y) \forall m \in \mathcal{M}, \text{ and} \\ \bar{Q}_{new}(n_Y^{new}|M=m) &= \bar{Q}(n_Y|M) = f((1 - \epsilon_m^{(2)})\delta_m^Y) \forall m \in \mathcal{M}. \end{aligned} \quad (282)$$

By (258), we know that $\epsilon_m^{(2)}\delta_m^Y = \delta_m^Y - \delta_{bias}^Y$ and $(1 - \epsilon_m^{(2)})\delta_m^Y = \delta_{bias}^Y$. Hence,

$$\begin{aligned} \bar{Q}(Y'|M) &= f(\epsilon_m^{(2)}\delta_m^Y) = f(\delta_m^Y - \delta_{bias}^Y) \forall m \in \mathcal{M}, \text{ and} \\ \bar{Q}(n_Y|M) &= f((1 - \epsilon_m^{(2)})\delta_m^Y) = f(\delta_{bias}^Y) \forall m \in \mathcal{M}. \end{aligned} \quad (283)$$

□

Proposition 9. Suppose the random variables M and X have the joint distribution $Q(M, X)$, such that

$$Q(M) = P(M) \text{ and } Q(X|M=m) = f(\delta_m^X) \forall m \in \mathcal{M}, \quad (284)$$

where \mathcal{M} denotes the support of M and $f(\delta_m^X) \in \mathcal{F}_{\mathcal{D}}$ for some convolution-closed distribution family $\mathcal{F}_{\mathcal{D}}$. Furthermore, there also exist $\delta_{bias}^X \in \mathcal{D}$, and $\delta_m^Y \in \mathcal{D} \forall m \in \mathcal{M}$ such that

$$\begin{aligned} \delta_m^X - \delta_{bias}^X &= \epsilon_m^{(1)}(\delta_m^Y - \delta_{bias}^Y) \text{ for some } \epsilon_m^{(1)} \in [0, 1], \\ \delta_m^X &= \epsilon_m^{(3)}(\delta_m^X - \delta_{bias}^X) \text{ for some } \epsilon_m^{(3)} \in [0, 1], \text{ and} \\ (\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)), (\delta_m^X - \delta_{bias}^X), (\delta_m^Y - \delta_{bias}^Y) &\in \mathcal{D} \forall m \in \mathcal{M}. \end{aligned} \quad (285)$$

Then consider three dilated versions of X , denoted as X', Y'' , and n_X , defined as follows:

$$\begin{aligned} \bar{Q}(X', Y'', n_X|X, M) &= \bar{Q}(n_X|X, M)\bar{Q}(X'|X, n_X, M)\bar{Q}(Y''|X, n_X, M, X'), \text{ where} \\ \bar{Q}(n_X|X=x, M=m) &= G((1 - \epsilon_m^{(3)})\delta_m^X, \epsilon_m^{(3)}\delta_m^X, x), \\ \bar{Q}(X'|X=x, n_X=x_{n_X}, M=m) &= G((1 - \epsilon_m^{(1)})\delta_m^X - \delta_{bias}^X, \epsilon_m^{(1)}(\delta_m^X - \delta_{bias}^X), x - n_x), \text{ and} \\ Y'' &= X - X - n_X. \end{aligned} \quad (286)$$

Then, we have

1. (n_X, X', Y'') are jointly conditionally independent conditioned on M .

2. The conditional distributions $\bar{Q}(X'|M)$, $\bar{Q}(Y''|M)$, and $\bar{Q}(n_X|M)$ are specified as follows:

$$\begin{aligned}\bar{Q}(X'|M) &= f(\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)), \\ \bar{Q}(Y''|M) &= f(\delta_m^X - \delta_{bias}^Y), \text{ and} \\ \bar{Q}(n_X|M) &= f(\delta_{bias}^X).\end{aligned}\quad (287)$$

Proof. The proof essentially consists of recursively applying the proof of proposition 8. Consider a new auxiliary random variable $Z = X - n_X$. Then, the distribution $\bar{Q}(X, n_X, Z|M)$ has the following properties:

$$\begin{aligned}\bar{Q}(X|M) &\stackrel{(a)}{=} f(\delta_m^X), \bar{Q}(n_X|M=m, X=x) \stackrel{(b)}{=} G((1 - \epsilon_m^{(3)})\delta_m^X, \epsilon_m^{(3)}\delta_m^X, x), \text{ and} \\ \bar{Q}(Z|X=x, n_X=x_{n_X}, M=m) &= \delta_K(Z = x - x_{n_X}),\end{aligned}\quad (288)$$

where $\delta_K(\cdot)$ is the Kronecker delta function, and the equalities (a) and (b) follow from (284) and (285), respectively. Note that the distribution $\bar{Q}(X, n_X, Z|M)$ has exactly the same structure as $\bar{Q}(Y, Y', n_Y|M)$ described in (261), hence Z and n_Y are just dilated versions of X . Consequently, from the result of proposition 8, we know that $Z \perp\!\!\!\perp n_X|M$, and the conditional distributions $\bar{Q}(Z|M)$ and $\bar{Q}(n_Y|M)$ are described as follows:

$$\bar{Q}(n_X|M=m) = f((1 - \epsilon_m^{(3)})\delta_m^X), \text{ and } \bar{Q}(Z|M=m) = f(\epsilon_m^{(3)}\delta_m^X). \quad (289)$$

From (285), we have $\epsilon_m^{(3)}\delta_m^X = \delta_m^X - \delta_{bias}^X$ and $(1 - \epsilon_m^{(3)})\delta_m^X = \delta_{bias}^X$. Consequently,

$$\bar{Q}(n_X|M=m) = f(\delta_{bias}^X), \text{ and } \bar{Q}(Z|M=m) = f(\delta_m^X - \delta_{bias}^X). \quad (290)$$

We alternatively express $\bar{Q}(X'|n_X, X, M)$ as follows:

$$\bar{Q}(X'|n_X=x_{n_X}, X=x, M=m) = G((1 - \epsilon_m^{(1)})(\delta_m^X - \delta_{bias}^X), \epsilon_m^{(1)}(\delta_m^X - \delta_{bias}^X), x - x_{n_X}). \quad (291)$$

Note that the distribution $\bar{Q}(X'|n_X=x_{n_X}, X=x, M=m)$ only depends on X and n_X through their difference $X - n_X$. Hence, substituting $Z = X - n_X$ in the above equation:

$$\begin{aligned}\bar{Q}(X'|n_X=x_{n_X}, X=x, M=m) &= \bar{Q}(X'|Z=x - x_{n_X}, M=m) \\ &= G((1 - \epsilon_m^{(1)})(\delta_m^X - \delta_{bias}^X), \epsilon_m^{(1)}(\delta_m^X - \delta_{bias}^X), x - x_{n_X}),\end{aligned}\quad (292)$$

$$\Rightarrow \bar{Q}(X'|Z=z, M=m) = G((1 - \epsilon_m^{(1)})(\delta_m^X - \delta_{bias}^X), \epsilon_m^{(1)}(\delta_m^X - \delta_{bias}^X), z). \quad (293)$$

Hence, consider the following distribution $\bar{Q}(Z, X', Y''|M)$, where from (290) and (293), we have:

$$\begin{aligned}\bar{Q}(Z|M=m) &= f(\delta_m^X - \delta_{bias}^X), \\ \bar{Q}(X'|Z=z, M=m) &= G((1 - \epsilon_m^{(1)})(\delta_m^X - \delta_{bias}^X), \epsilon_m^{(1)}(\delta_m^X - \delta_{bias}^X), z).\end{aligned}\quad (294)$$

Furthermore, by (286), we have $Y'' = X - n_X - X' \stackrel{(c)}{=} Z - X'$, where (c) is due to $Z = X - n_X$. Hence,

$$\bar{Q}(Y''|Z=z, M, X'=x) = \delta_K(Y'' = z - x). \quad (295)$$

Comparing (295) and (294) with (261), we can again conclude that X'' and Y' are just dilated versions of Z . Hence, again using the result of proposition 8, we have $X' \perp\!\!\!\perp Y''|M$, and

$$\bar{Q}(X'|M=m) = f((1 - \epsilon_m^{(1)})(\delta_m^X - \delta_{bias}^X)), \text{ and } \bar{Q}(Y''|M=m) = f(\epsilon_m^{(1)}(\delta_m^X - \delta_{bias}^X)). \quad (296)$$

From (285), we have $\epsilon_m^{(1)}(\delta_m^X - \delta_{bias}^X) = \delta_m^Y - \delta_{bias}^Y$ and $(1 - \epsilon_m^{(1)})(\delta_m^X - \delta_{bias}^X) = \delta_m^X - \delta_M^Y - (\delta_{bias}^X - \delta_{bias}^Y)$. Consequently,

$$\bar{Q}(X'|M=m) = f(\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)), \text{ and } \bar{Q}(Y''|M=m) = f(\delta_m^Y - \delta_{bias}^Y). \quad (297)$$

Hence, by (290) and (297), we can conclude that the conditional distributions $\bar{Q}(X'|M)$, $\bar{Q}(Y''|M)$, and $\bar{Q}(n_X|M)$ are specified as follows:

$$\begin{aligned}\bar{Q}(X'|M=m) &= f(\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)), \\ \bar{Q}(Y''|M) &= f(\delta_m^X - \delta_{bias}^Y), \text{ and} \\ \bar{Q}(n_X|M) &= f(\delta_{bias}^X).\end{aligned}\tag{298}$$

Showing joint conditional independence of (n_Y, X', Y'') given M :

Consider the distribution $\bar{Q}(n_X, Z, X', Y''|M)$. Using the chain rule, we have:

$$\begin{aligned}\bar{Q}(n_X, Z, X', Y''|M) &= \bar{Q}(n_X, Z|M)Q(X', Y''|M, n_X, Z), \\ &\stackrel{(b)}{=} \bar{Q}(n_X|M)\bar{Q}(Z|M)Q(X', Y''|M, n_X, Z), \\ &\stackrel{(c)}{=} \bar{Q}(n_X|M)\bar{Q}(Z|M)Q(X', Y''|M, Z),\end{aligned}\tag{299}$$

where (b) follows from $n_X \perp\!\!\!\perp Z|M$. For showing (c), observe that $Y'' = Z - X'$, and hence we can write $\bar{Q}(Y''|M, Z, n_X, X') = \bar{Q}(Y''|Z, X')$. Furthermore, by analyzing $\bar{Q}(X'|n_X, Z, M)$, we can conclude:

$$\bar{Q}(X'|n_X, Z, M) = \bar{Q}(X'|n_X, X - n_X, M) = \bar{Q}(X'|n_X, X, M) = \bar{Q}(X'|Z, M).\tag{300}$$

Marginalizing Z out of (299), we obtain:

$$\bar{Q}(n_X, X', Y''|M) = \bar{Q}(n_X|M)\bar{Q}(X', Y''|M) \stackrel{(d)}{=} \bar{Q}(n_X|M)Q(X'|M)\bar{Q}(Y''|M),\tag{301}$$

where (d) follows from $X' \perp\!\!\!\perp Y''|M$. By (301), we know that (n_X, X'', Y') are jointly conditionally independent given M , concluding our proof. \square

M.2 Analytically optimizing the upper bound

Proposition 10. Let $\bar{Q}(M, X', Y'', n_X, Y', n_Y)$ be the distribution derived from a distribution $Q \in \Delta_P$ using the construction scheme specified in Appx. M.1. Then, the minimization problem specified in (302) is analytically solvable with the optimizing \bar{Q}^* given by (303).

$$\min_{Q \in \Delta_P} I_{\bar{Q}}(M; [X', Y'', n_X, Y', n_Y]), \text{ (note that } \bar{Q} \text{ is a function of } Q)\tag{302}$$

$$\begin{aligned}\bar{Q}^*(M, X', Y'', n_X, n_Y, Y') &= \bar{Q}^*(M)\bar{Q}^*(X'|M)\bar{Q}^*(Y''|M)\bar{Q}^*(n_X)\bar{Q}^*(n_Y)\bar{Q}^*(Y'|Y''), \text{ with} \\ \bar{Q}^*(M) &= P(M), \bar{Q}^*(X'|M=m) = f(\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)), \bar{Q}^*(n_X) = f(\delta_{bias}^X), \\ \bar{Q}^*(Y''|M=m) &= f(\delta_m^Y - \delta_{bias}^Y), \bar{Q}^*(n_Y) = f(\delta_{bias}^Y), \text{ and } \bar{Q}^*(Y'|Y'') = \mathbb{I}[Y'' = Y'],\end{aligned}\tag{303}$$

where $\mathbb{I}[\cdot]$ is the indicator function. Note that for the minimizing \bar{Q}^* , we have $(n_X, n_Y, (M, X', Y''))$ are jointly independent and $Y' = Y''$.

Proof. First note that by the construction scheme defined in Appx. M.1, we know that there is a unique \bar{Q} for every $Q \in \Delta_P$. Hence, we can interpret \bar{Q} as a function of Q , and consequently the minimization problem in (302) is well-defined. Using the chain rule of mutual information, we derive the following lower bound for $I_{\bar{Q}}(M; [X', Y'', n_X, Y', n_Y])$:

$$I_{\bar{Q}}(M; [X', Y'', n_X, Y', n_Y]) = I_{\bar{Q}}(M; [X', Y'']) + I_{\bar{Q}}(M; [n_X, Y', n_Y] | [X', Y'])\tag{304}$$

$$\Rightarrow I_{\bar{Q}}(M; [X', Y'', n_X, Y', n_Y]) \stackrel{(a)}{\geq} I_{\bar{Q}}(M; [X', Y'']),\tag{305}$$

where (a) is due to the non-negativity of conditional mutual information. Furthermore, note that by properties of \bar{Q} derived in Appx. M.1, we know that $X' \perp\!\!\!\perp Y''|M \forall Q \in \Delta_P$. Consequently, the term $I_{\bar{Q}}(M; [X', Y''])$ does not vary as we choose different Q from Δ_P as the corresponding $\bar{Q}(M, X', Y'') = \bar{Q}(M)\bar{Q}(X'|M)\bar{Q}(Y''|M)$ and the terms:

$$\begin{aligned}\bar{Q}(M) &= Q(M) = P(M), \bar{Q}(X'|M=m) = f(\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)), \text{ and} \\ \bar{Q}(Y''|M) &= f(\delta_m^Y - \delta_{bias}^Y)\end{aligned}\tag{306}$$

do not vary for all $Q \in \Delta_P$ (see Appx. M.1). Therefore, we have:

$$I(M; [X', Y'']) \leq I_{\bar{Q}}(M; [X', Y'', n_X, Y', n_Y]) \quad \forall Q \in \Delta_P, \quad (307)$$

where we forgo the subscript \bar{Q} for $I_{\bar{Q}}(M; [X'', Y'])$ to indicate that the term $I_{\bar{Q}}(M; [X'', Y'])$ does not vary over Δ_P . Hence, if we show that the distribution $\bar{Q}^*(M, X', Y'', n_X, Y', n_Y)$ specified in (303) achieves the lower bound in (307), then $\bar{Q}^*(M, X', Y'', n_X, Y', n_Y)$ should minimize the value of $I_{\bar{Q}}(M; [X', Y'', n_X, Y', n_Y])$ over the set Δ_P . We now calculate $I_{\bar{Q}^*}(M; [X', Y'', n_X, Y', n_Y])$. By the chain rule of mutual information:

$$I_{\bar{Q}^*}(M; [X', Y'', n_X, Y', n_Y]) = I_{\bar{Q}^*}(M; [n_X, n_Y, X', Y'']) + I_{\bar{Q}^*}(M; Y' | n_X, n_Y, X, Y'').$$

Using the fact that $Y' = Y''$, we know that $I_{\bar{Q}^*}(M; Y' | n_X, n_Y, X, Y'')$ is zero. A simple justification follows from the fact $I_{\bar{Q}^*}(M; Y' | n_X, n_Y, X, Y'') = H(Y' | n_X, n_Y, X, Y'') - H(Y' | n_X, n_Y, X, Y'', M)$, where $H(\cdot | \cdot)$ is the conditional entropy function [28]. Observe that both conditional entropy terms are zero as $Y' = Y''$, and condition entropy of a random variable conditioned on itself is zero. Therefore, we can simplify the above equation as:

$$I_{\bar{Q}^*}(M; [X', Y'', n_X, Y', n_Y]) = I_{\bar{Q}^*}(M; [n_X, n_Y, X', Y'']). \quad (308)$$

Applying the chain rule of mutual information again, we obtain:

$$I_{\bar{Q}^*}(M; [X', Y'', n_X, Y', n_Y]) = I_{\bar{Q}^*}(M; [n_X, n_Y]) + I_{\bar{Q}^*}(M; [X', Y''] | n_X, n_Y). \quad (309)$$

Using the fact that $(n_X, n_Y) \perp\!\!\!\perp (M, X', Y'')$, we know that $I(M; [n_X, n_Y]) = 0$ as mutual information between independent random variables is zero [28]. Consequently,

$$I_{\bar{Q}^*}(M; [X', Y'', n_X, Y', n_Y]) = I_{\bar{Q}^*}(M; [X', Y''] | n_X, n_Y) \stackrel{(b)}{=} I_{\bar{Q}^*}(M; [X', Y'']), \quad (310)$$

where (b) again follows from the fact that $(n_X, n_Y) \perp\!\!\!\perp (M, X', Y'')$, and, consequently, conditioning (M, X', Y'') on (n_X, n_Y) has no effect. From (310), we see the proposed \bar{Q}^* in (303) does indeed achieve the lower bound specified in (307), and consequently minimizes (302).

The final step remaining in our proof is to ensure that the proposed \bar{Q}^* corresponds to a valid $Q_{UB}^*(M, X, Y)$ lying in Δ_P , i.e., there is some $Q_{UB}^*(M, X, Y) \in \Delta_P$ that generates \bar{Q}^* using the construction scheme outlined in Appx. M.1. We know that the relationship between the random variables for $(M, X', Y'', n_X, Y', n_Y)$ having the joint distribution \bar{Q} and (M, X, Y) having the distribution Q can be expressed as follows: $X = X' + Y'' + n_X$ and $Y = Y' + n_Y$. Hence, we use these relationships for calculating the conditional marginals $Q_{UB}^*(X|M)$ and $Q_{UB}^*(Y|M)$ of the corresponding distribution $Q_{UB}^*(M, X, Y)$ for the proposed distribution $\bar{Q}^*(M, X', Y'', n_X, Y', n_Y)$:

$$\begin{aligned} Q_{UB}^*(X|M=m) &= \bar{Q}^*(X' + Y'' + n_X | M=m) \\ &\stackrel{(c)}{=} \bar{Q}^*(X' | M=m) * \bar{Q}^*(Y'' | M=m) * \bar{Q}^*(n_X | M=m) \\ &= f(\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y)) * f(\delta_m^Y - \delta_{bias}^Y) * f(\delta_{bias}^X) \\ &\stackrel{(d)}{=} f(\delta_m^X - \delta_m^Y - (\delta_{bias}^X - \delta_{bias}^Y) + \delta_m^Y - \delta_{bias}^Y + \delta_{bias}^X) \\ &= f(\delta_m^X) = P(X|M=m). \end{aligned} \quad (311)$$

where $*$ denotes the convolution operator, and (c) follows from the fact that (X', Y'', n_X) are jointly conditionally independent given M under \bar{Q}^* , and (d) follows from the properties of convolution-closed distributions. Similarly, calculating $Q_{UB}^*(Y|M=m)$:

$$\begin{aligned} Q_{UB}^*(Y|M=m) &= \bar{Q}^*(Y' + n_Y | M=m) \\ &\stackrel{(e)}{=} \bar{Q}^*(Y'' | M=m) * \bar{Q}^*(n_Y | M=m) \\ &= f(\delta_m^Y - \delta_{bias}^Y) * f(\delta_{bias}^Y) \\ &= f(\delta_m^Y - \delta_{bias}^Y + \delta_{bias}^Y) \\ &= f(\delta_m^Y) = P(Y|M=m), \end{aligned} \quad (312)$$

where (e) follows from the fact that $Y' \perp\!\!\!\perp n_Y | M$ under \bar{Q}^* . It is easy to see that $Q_{UB}^*(M) = \bar{Q}(M) = P(M)$. Hence, we have $Q_{UB}^*(M, X) = P(M)P(X|M) = P(M, X)$ and $Q_{UB}^*(M, Y) = P(M)P(Y|M) = P(M, Y)$, which implies $Q_{UB}^*(M, X, Y) \in \Delta_P$, concluding our proof. \square

M.3 Additional details regarding simulation

The list of different function pairs used in the simulation study are listed in Table 1. The function pairs were chosen with the only requirement being that they satisfied the assumptions in (9). For constructing the analytical estimate $\hat{Q}_A(M, X, Y)$, the corresponding δ_{bias}^X and δ_{bias}^Y were chosen as follows:

1. For each distribution and function pair, we first determine the range of possible values for δ_{bias}^X and δ_{bias}^Y .
2. For δ_{bias}^Y , we choose the range as $[0, \min_{m \in \mathcal{M}} f_2(m))$, where \mathcal{M} is the support set of M for that distribution. Note that we cannot choose $\delta_{bias}^Y > \min_{m \in \mathcal{M}} f_2(m)$ as the corresponding conditional distribution $\bar{Q}^*(Y'|M) = \text{Poisson}(f_2(M) - \delta_{bias}^Y)$ (required for the optimal \bar{Q}^* from which Q_A is constructed, see Appx. M.2) would have a negative rate parameter, which is not allowed. Within this range, we choose 10 different, uniformly spaced values for δ_{bias}^Y .
3. For each chosen value of δ_{bias}^Y , the corresponding range for δ_{bias}^X is then chosen as $[0, \min\{\min_{m \in \mathcal{M}} f_1(m), \min_{m \in \mathcal{M}}(f_1(m) - f_2(m) + \delta_{bias}^Y)\}]$. Note that choosing the upper limit of δ_{bias}^X as $\min\{\min_{m \in \mathcal{M}} f_1(m), \min_{m \in \mathcal{M}}(f_1(m) - f_2(m) + \delta_{bias}^Y)\}$ ensures that the distribution $\bar{Q}^*(X'|M) = \text{Poisson}((f_1(M) - \delta_{bias}^X) - (f_2(M) - \delta_{bias}^Y))$ and $\bar{Q}^*(Z|M) = \text{Poisson}(f_1(M) - \delta_{bias}^X)$ (required for specifying the optimal \bar{Q}^* from which Q_A is constructed, see Appx. M.1 and Appx. M.2) have positive rate parameters. For each value of δ_{bias}^Y , we again choose 10 different values of δ_{bias}^X within the above specified range. In total, we chose 100 different pairs of $(\delta_{bias}^X, \delta_{bias}^Y)$.

For each pair of $(\delta_{bias}^X, \delta_{bias}^Y)$, we calculate the analytical $Q_A(M, X, Y)$ as follows:

First, we construct the optimal \bar{Q}^* distribution using the structure provided in proposition 10. The exact specification of \bar{Q}^* is described below. We simplify the structure of \bar{Q}^* by using the fact that $Y' = Y''$:

$$\begin{aligned}
 \bar{Q}^*(M, X', Y', n_X, n_Y, Y'') &\equiv \bar{Q}^*(M, X', Y', n_X, n_Y), \\
 &= \bar{Q}^*(M) \bar{Q}^*(X'|M) \bar{Q}^*(Y'|M) \bar{Q}^*(n_X) \bar{Q}^*(n_Y), \text{ with} \\
 \bar{Q}^*(M) &= P(M), \\
 \bar{Q}^*(X'|M) &= \text{Poisson}(f_1(M) - f_2(M) - (\delta_{bias}^X - \delta_{bias}^Y)), \\
 \bar{Q}^*(n_X) &= \text{Poisson}(\delta_{bias}^X), \\
 \bar{Q}^*(Y''|M) &= \text{Poisson}(f_2(M) - \delta_{bias}^Y), \\
 \bar{Q}^*(n_Y) &= \text{Poisson}(\delta_{bias}^Y).
 \end{aligned} \tag{313}$$

Then, we use the property that $X = X' + Y' + n_X$, and $Y = Y' + n_Y$, and use that to appropriately modify \bar{Q}^* to obtain Q_A as follows:

- We calculate a new distribution $\bar{Q}^*(M, X' + Y', Y', n_X, n_Y)$:

$$\begin{aligned}
 \bar{Q}^*(M, X' + Y', Y', n_X, n_Y) &= \bar{Q}^*(M) \bar{Q}^*(n_X) \bar{Q}^*(n_Y) \bar{Q}^*(X' + Y'|M) \times \\
 &\quad \bar{Q}^*(Y'|X' + Y', M). \\
 &= \bar{Q}^*(M) \bar{Q}^*(n_X) \bar{Q}^*(n_Y) \bar{Q}^*(Z|M) \bar{Q}^*(Y'|Z, M),
 \end{aligned} \tag{314}$$

where $Z = X' + Y'$ and alternatively also $Z = X - n_X$ (as $X = X' + Y' + n_X$). First note that $\bar{Q}^*(Z|M)$ from Appx. M.2 has the form $f(\delta_m^X - \delta_{bias}^X)$, which for the Poisson case is $\text{Poisson}(f_1(M) - \delta_{bias}^X)$. Furthermore, we can recognize that $\bar{Q}^*(Y' = y|Z = z, M = m) = \bar{Q}^*(Y' = y'|X' + Y' = x' + y', M = m) = G(y'; \epsilon_m^{(1)}, 1 - \epsilon_m^{(1)}, x' + y')$, which for the Poisson case can be written as $\text{Binomial}(Z, p_m)$ where $p_m = f_2(M) - \delta_{bias}^Y / f_1(M) - \delta_{bias}^X$.

- Now define $\mathbb{S}_x^m = \{x_{n_X} \in \mathcal{X}_{n_X}^m : x_{n_X} + z = x\}$, where $\mathcal{X}_{n_X}^m$ and \mathcal{Z}_m are the corresponding support sets of the distributions $\bar{Q}^*(n_X|M = m)$ and $\bar{Q}^*(Z|M = m)$,

respectively. Then, using the fact that $X = Z + n_X$, we have:

$$\bar{Q}^*(M=m, X=x, n_Y, Y') = \bar{Q}^*(M, Z + n_X, n_Y, Y') = \bar{Q}^*(M=m) \bar{Q}^*(n_Y) \times \sum_{(x_{n_X}, z) \in \mathbb{S}_x^m} \bar{Q}^*(n_X = x_{n_X}) \bar{Q}^*(Z = z | M = m) \bar{Q}^*(Y' | Z = z, M = m), \quad (315)$$

where we know that for the Poisson case:

$$\bar{Q}^*(M) = P(M), \bar{Q}^*(n_Y) = \text{Poisson}(\delta_{bias}^Y), \bar{Q}^*(Z|M) = \text{Poisson}(f_1(M) - \delta_{bias}^X), \\ \bar{Q}^*(n_X) = \text{Poisson}(\delta_{bias}^X), \text{ and } \bar{Q}^*(Y'|Z, M) = \text{Binomial}\left(Z, \frac{f_2(M) - \delta_{bias}^Y}{f_1(M) - \delta_{bias}^X}\right). \quad (316)$$

- Using a similar procedure, we derive $Q_A(M, X, Y)$ from $\bar{Q}^*(M, X, Y', n_Y)$ as $Y = n_Y + Y'$. Define $\mathbb{S}_y^m = \{y_{n_Y} \in \mathcal{Y}_{n_X}^m, z \in \mathcal{Y}_m' : y_{n_X} + y' = y\}$, then:

$$Q_A(M, X, Y=y) = \bar{Q}(M, X, n_Y + Y') = \sum_{(x_{n_Y}, y') \in \mathbb{S}_y^m} \bar{Q}^*(M, X, Y' = y', n_Y = x_{n_Y}).$$

- After analytically calculating Q_A for all 100 bias pairs of $(\delta_{bias}^X, \delta_{bias}^Y)$, we chose Q_A corresponding to the pair $(\delta_{bias}^X, \delta_{bias}^Y)$ which results in the smallest value of the corresponding $I_{Q_A}(M; [X, Y])$.

| Index | $f_1(M)$ | $f_2(M)$ | Index | $f_1(M)$ | $f_2(M)$ |
|-------|-----------------------|--------------------------|-------|-------------------|------------------|
| 1 | $3M$ | $3 \log(1 + M)$ | 11 | Me^M | M |
| 2 | $1.5M^3$ | $0.5e^M$ | 12 | $3M^2 - 2M$ | $M + \log(M)$ |
| 3 | $0.5M^4$ | $0.25e^M \sin^2(M\pi/8)$ | 13 | $8 \sin(M\pi/10)$ | $2\sqrt{M}$ |
| 4 | $7M$ | $4M \sin^2(M\pi/8)$ | 14 | $2\cosh(M)$ | $6 \sin(M\pi/8)$ |
| 5 | M^2 | M | 15 | $5M$ | $5\sqrt{M}$ |
| 6 | M^3 | M^2 | 16 | e^M | $\sinh(M)$ |
| 7 | $5M$ | $5 \sin(M\pi/8)$ | 17 | $2\cosh(M)$ | $\sinh(M)$ |
| 8 | $M^2 + M^3$ | $2M$ | 18 | $2.5M^2 + 5M$ | $0.5M^3 + M$ |
| 9 | $7M$ | $M^2 \cos^2(M\pi/16)$ | 19 | $1.5\cosh(M)$ | $0.5M^3$ |
| 10 | $8(M - \cos(M\pi/8))$ | $2 \sin(M\pi/8)$ | 20 | $8M + 3$ | $7M$ |

Table 1: Table 1 lists the 20 different function pairs used in the simulation study described in Sec. 6 for comparing the analytical estimate of the minimizing distribution of (2) proposed in Sec. 6 with the corresponding ground-truth numerical estimate. The index in Table 1 maps the corresponding function pair to the results shown in Fig. 1. To illustrate, the number 1 on the x -axis of all the plots shown in Fig. 1 corresponds to the function pair with index 1 in Table 1.

M.4 Additional results and analysis

We report the tightness of the upper-bound for systems employing binomial and negative-binomial distributions in Fig. 2(a). We find that the results for binomial and negative-binomial remain consistent with the Poisson study performed in Sec. 6 and the upper-bound continues to remain tight. We also performed further analysis to understand how our upper-bound performs when both of the UI terms are non-zero. We sub-selected the trials in the 20 pairs (analyzed in Section 6) that had both $UI_x > 0.01$ bits and $UI_y > 0.01$ bits, and report the corresponding median error of these sub-selected trials in Fig. 2(d).

N Distributions

We provide brief descriptions of the distribution families and the corresponding distributions used in this work. We also provide citations to relevant works for the reader interested in learning more about these distributions.

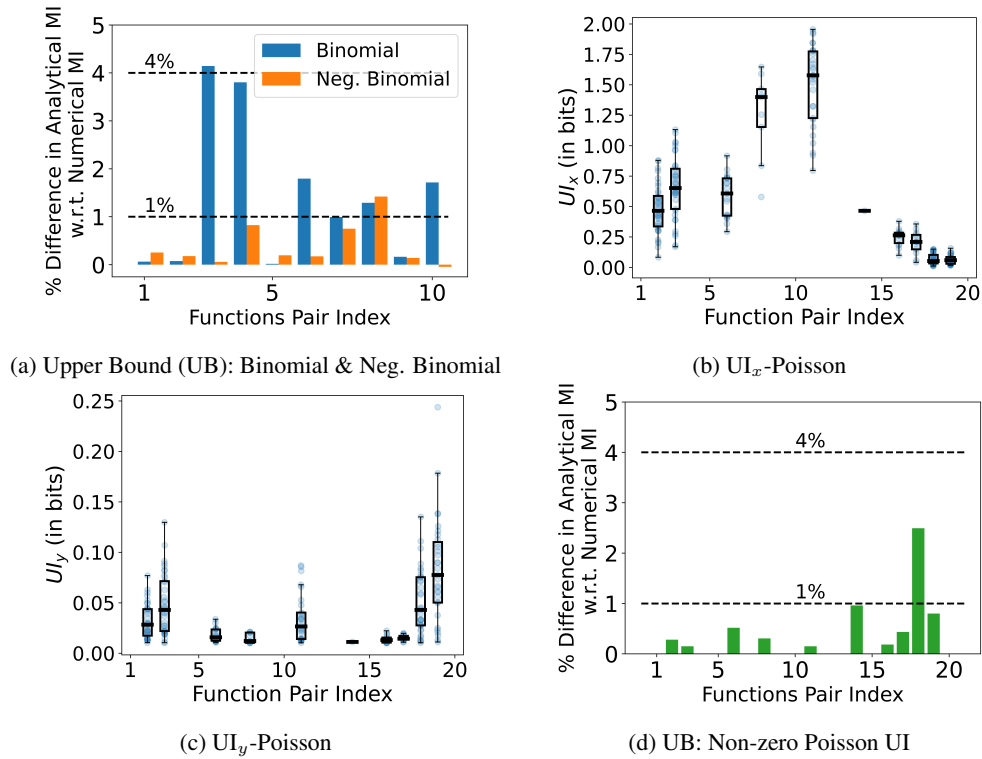


Figure 2: **(a)** shows additional results of simulation study performed in Section 6 of the manuscript for the distributions: Negative Binomial and Binomial. The experimental setup is the same as described in Section 6, with the differences being that only 10 function pairs were tested, the number of outcomes of M was restricted to 2 and 4, and the support of M was appropriately modified for Binomial and Negative Binomial distributions (see caption of Table 2 below). The function pairs tested are provided in Table 2. **(b)** and **(c)** show the unique information atoms of the systems having non-zero unique information in both X and Y tested in the simulation study presented in Section 6 of the manuscript. More specifically, of the 20 pairs analyzed in Section 6, we only selected trials that had both $UI_x > 0.01$ bits and $UI_y > 0.01$ bits. **(d)** provides the corresponding median error for the subselected trials (originally shown in Fig.1c in the manuscript for all trials) for the systems shown in **(b)** and **(c)**.

| Index | Negative Binomial | | Binomial | |
|-------|--------------------|--------------------------------------|---------------------------------------|------------------------------------------------------|
| | $f_1(M)$ | $f_2(M)$ | $f_1(M)$ | $f_2(M)$ |
| 1 | $0.5M^2$ | M | $\lfloor 0.5M^2 \rfloor$ | M |
| 2 | $0.1M^3$ | $0.25M^2$ | $\lfloor 0.04M^3 \rfloor$ | $\lfloor 0.1M^2 \rfloor$ |
| 3 | $4M$ | $M \sin^2(M\pi/16) + 2$ | $4M$ | $\lfloor M \sin^2(M\pi/16) + 2 \rfloor$ |
| 4 | $5M$ | $10 \sin(M\pi/16) + 2$ | $5M$ | $\lfloor 10 \sin(M\pi/16) + 2 \rfloor$ |
| 5 | $0.05(M^2 + M^3)$ | M | $\lfloor 0.05(M^2 + M^3) \rfloor$ | M |
| 6 | $7M$ | $0.5M^2 \cos^2(\frac{M\pi}{36}) + 2$ | $7M$ | $\lfloor 0.5M^2 \cos^2(\frac{M\pi}{36}) + 2 \rfloor$ |
| 7 | $M - \cos(M\pi/8)$ | $2 \sin(M\pi/16) + 2$ | $\lfloor 2(M - \cos(M\pi/8)) \rfloor$ | $\lfloor 12 \sin(M\pi/40) + 2 \rfloor$ |
| 8 | $2M$ | $0.004Me^{M/2} + 9$ | $2M$ | $\lfloor 0.004Me^{M/3} + 5 \rfloor$ |
| 9 | $0.5(3M^2 - 2M)$ | $M + \log(M)$ | $\lfloor 0.1(3M^2 - 2M) \rfloor$ | $\lfloor M + \log(M) \rfloor$ |
| 10 | $18 \sin(M\pi/18)$ | $5\sqrt{M}$ | $\lfloor 19 \sin(M\pi/40) \rfloor$ | $\lfloor 3\sqrt{M} \rfloor$ |

Table 2: For negative binomial the outcomes of M were uniformly randomly sampled from the interval $[5, 10]$, and for the binomial experiment the outcomes of M were uniformly chosen from the set $\{5, 6, \dots, 14\}$. The symbol $\lfloor \cdot \rfloor$ denotes the rounding to closest integer operation. The corresponding system models of binomial and negative binomial can be found in Appendix C (systems 3 and 6, respectively) of the manuscript.

N.1 Common distributions used in this work

We employed various well-known distributions in this work. We refer the reader to [62, 63] for a review of these distributions. We use the most commonly used notations in literature to represent these distributions. Regardless, we provide explicit definitions of our notations for these distributions to remove ambiguity.

Poisson Distribution Poisson distribution is a discrete distribution having support over all non-negative integers. The p.m.f. of the Poisson distribution is described by a single rate parameter $\lambda \in [0, \infty)$. We define the Poisson distribution with the rate parameter $\lambda = 0$ as a degenerate point distribution with all its mass at 0. We use $\text{Poisson}(\lambda)$ to denote the p.m.f. of a random variable X distributed according to the Poisson distribution. The exact definition of the p.m.f. of X is provided below:

$$P(X = x) = \text{Poisson}(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \forall x \in \mathbb{N}_0.$$

Gaussian Distribution Gaussian distribution is a continuous distribution having support over \mathbb{R} . The p.d.f. of the Gaussian distribution is described by its mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in [0, \infty)$. We define the Gaussian distribution with the variance $\sigma^2 = 0$ as a degenerate point distribution with all its mass at μ . We use $\mathcal{N}(\mu, \sigma^2)$ to denote the p.d.f. of a random variable X distributed according to the Gaussian distribution. The exact definition of the p.d.f. of X is provided below:

$$p(X = x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \forall x \in \mathbb{R}.$$

Negative Binomial Distribution Negative binomial distribution is a discrete distribution having support over all non-negative integers. The p.m.f. of the negative binomial distribution is described by two parameters: the number of successes $r \in \mathbb{N}$, and the probability of success $p \in (0, 1]$. We use $\text{Negative Binomial}(r, p)$ to denote the p.m.f. of a random variable X distributed according to the negative binomial distribution. The exact definition of the p.m.f. of X is provided below:

$$P(X = x) = \text{Negative Binomial}(x; r, p) = \binom{x+r-1}{x} (1-p)^x p^r \quad \forall x \in \mathbb{N}_0,$$

where $\binom{x+r-1}{x} = \frac{(x+r-1)!}{x!(r-1)!}.$

Geometric Distribution Geometric distribution is a special case of negative binomial distribution with $r = 1$. The p.m.f. of the geometric distribution is described by a success probability $p \in (0, 1]$. We use $\text{Geometric}(p)$ to denote the p.m.f. of a random variable X distributed according to the geometric distribution. The exact definition of the p.m.f. of X is provided below:

$$P(X = x) = \text{Geometric}(x; p) = p(1-p)^x \quad \forall x \in \mathbb{N}_0.$$

Beta-Binomial Distribution Beta-binomial distribution is a discrete distribution having support over $\{0, \dots, N\}$ for some $N \in \mathbb{N}_0$. The p.m.f. of the beta-binomial distribution is described by three parameters: N specifies the support, $\alpha \in (0, \infty)$ and $\beta \in (0, \infty)$. We use $\text{Beta-Binomial}(N, \alpha, \beta)$ to denote the p.m.f. of a random variable X distributed according to the beta-binomial distribution. The exact definition of the p.m.f. of X is provided below:

$$P(X = x) = \text{Beta-Binomial}(x; N, \alpha, \beta) = \binom{N}{x} \frac{B(x+\alpha, N-x+\beta)}{B(\alpha, \beta)} \quad \forall n \in \{0, \dots, N\},$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the beta function and $\Gamma(\cdot)$ is the Gamma function.

Gamma Distribution Gamma distribution is a continuous distribution having support over $(0, \infty)$. The p.d.f. of the gamma distribution is described by its shape parameter $\alpha > 0$ and rate parameter $\beta > 0$. We use $\text{Gamma}(\alpha, \beta)$ to denote the p.d.f. of a random variable X distributed according to the gamma distribution. The exact definition of the p.d.f. of X is provided below:

$$p(X = x) = \text{Gamma}(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad \forall x \in (0, \infty),$$

where $\Gamma(\cdot)$ is the Gamma function.

Exponential Distribution Exponential distribution is special case of the Gamma distribution with $\alpha = 1$. The p.d.f. of the exponential distribution is described by its rate parameter $\lambda > 0$. We use $\text{Exponential}(\lambda)$ to denote the p.d.f. of a random variable X distributed according to the exponential distribution. The exact definition of the p.d.f. of X is provided below:

$$p(X = x) = \text{Exponential}(x; \lambda) = \lambda e^{-\lambda x} \forall x \in [0, \infty). \quad (317)$$

Beta Distribution Beta distribution is a continuous distribution having support over $[0, 1]$ or $(0, 1)$ depending upon the values of its parameter. The p.d.f. of the beta distribution is described by two shape parameters: $\alpha > 0$ and $\beta > 0$. We use $\text{Beta}(\alpha, \beta)$ to denote the p.d.f. of a random variable X distributed according to the beta distribution. The exact definition of the p.d.f. of X is provided below:

$$p(X = x) = \text{Beta}(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \forall x \in (0, 1),$$

where $\Gamma(\cdot)$ is the Gamma function. The end-point 0 is included in the support of the beta distribution if $\alpha \geq 1$ and the end-point 1 is included in the support of the beta distribution if $\beta \geq 1$.

Dirichlet Distribution Dirichlet distribution is a multivariate generalization of the beta distribution. The p.d.f. of Dirichlet distribution is described by d -shape parameters: $\{\alpha_1, \dots, \alpha_d\}$, where $\alpha_i \in (0, \infty) \forall i \in [d]$. We use $\text{Dirichlet}(\alpha_1, \dots, \alpha_d)$ to denote the p.d.f. of a d -dimensional random vector \vec{X} distributed according to the Dirichlet distribution. The exact definition of the p.d.f. of X is provided below:

$$p(\vec{X} = \vec{x}) = \text{Dirichlet}(\vec{x}; \alpha_1, \dots, \alpha_d) = \frac{\Gamma(\sum_{i=1}^d \alpha_i)}{\prod_{i=1}^d \Gamma(\alpha_i)} \prod_{i=1}^d x_i^{\alpha_i-1},$$

where $\Gamma(\cdot)$ is the Gamma function, and $\vec{x} = [x_1 \dots x_d]^T$, with $x_i \in [0, 1]$ and $\sum_{i=1}^d x_i = 1$.

Binomial Distribution Binomial distribution is a discrete distribution having support over $\{0, \dots, N\}$ for some $N \in \mathbb{N}_0$. The p.m.f. of the binomial distribution is described by two parameters: the total number of trials, N , and the success probability, p . We use $\text{Binomial}(N, p)$ to denote the p.m.f. of a random variable X distributed according to the binomial distribution. The exact definition of the p.m.f. of X is provided below:

$$P(X = x) = \text{Binomial}(x; N, p) = \binom{N}{x} p^x (1-p)^{n-x} \forall n \in \{0, \dots, N\}, \quad (318)$$

where $\binom{N}{x} = \frac{N!}{N!(N-x)!}$.

Multinomial Distribution Multinomial distribution is a multivariate generalization of the binomial distribution. The p.m.f. of the multinomial distribution is described by two parameters: the total number of trials, N , and the success probability vector, \vec{p} . We use $\text{Multinomial}_d(N, \vec{p})$ to denote the p.m.f. of a d -dimensional random vector \vec{X} distributed according to the multinomial distribution. The exact definition of the p.m.f. of \vec{X} is provided below:

$$P(\vec{X} = \vec{x}) = \text{Multinomial}_d(\vec{x}; N, \vec{p}) = \frac{N!}{\prod_{i=1}^d x_i!} \prod_{i=1}^d p_i^{x_i},$$

where, $\vec{p} = [p_1 \dots p_d]^T$ is a d -dimensional probability vector, i.e. $0 \leq p_i \leq 1, \forall 1 \leq i \leq d$ and $\sum_{i=1}^d p_i = 1$, and $\vec{x} = [x_1 \dots x_d]^T$ is a d -dimensional categorical vector, i.e. $x_i \in \mathbb{N}_0, \forall 1 \leq i \leq d$, and $\sum_{i=1}^d x_i = n$.

Hypergeometric Distribution The p.m.f. of the hypergeometric distribution is described by three parameters: the total number of success states in the population $S \in \mathbb{N}_0$, the total number of failure states $F \in \mathbb{N}_0$ and the number of draws $n \in \{0, \dots, S + F\}$. We use $\text{Hypergeometric}(S, K, n)$ to

denote the p.m.f. of a random variable X distributed according to the hypergeometric distribution. The exact definition of the p.m.f. of X is provided below:

$$\begin{aligned} P(X = x) &= \text{Hypergeometric}(x; S, K, n) \\ &= \frac{\binom{S}{x} \binom{F}{n-x}}{\binom{S+F}{n}} \forall x \in \{\max\{0, n-F\}, \dots, \min\{n, S\}\}, \\ \text{where } \binom{A}{B} &= \frac{A!}{B!(A-B)!} \text{ for some } A, B \in \mathbb{N}_0 \text{ and } B \leq A. \end{aligned}$$

Multivariate Hypergeometric Distribution Multivariate hypergeometric distribution is a multivariate generalization of hypergeometric distribution. The p.m.f. of the multivariate hypergeometric distribution is described by the following parameters: the total number of object types $d \in \mathbb{N}_0$, the size of population of each type $\{n_1, \dots, n_d\}$ with $n_i \in \mathbb{N}_0$, and the number of draws $n \in \{0, \dots, \sum_{i=1}^d n_i\}$. Define $\vec{n} = [n_1 \dots n_d]$. Then, we use Multivariate Hypergeometric(\vec{n}, n) to denote the p.m.f. of a d -dimensional random vector \vec{X} distributed according to the multivariate hypergeometric distribution. The exact definition of the p.m.f. of \vec{X} is provided below:

$$\begin{aligned} P(\vec{X} = \vec{x}) &= \text{Multivariate Hypergeometric}(\vec{x}; \vec{n}, n), \\ &= \frac{\prod_{i=1}^d \binom{n_i}{x_i}}{\binom{\sum_{i=1}^d n_i}{n}}, \end{aligned} \quad (319)$$

where $\binom{A}{B} = \frac{A!}{B!(A-B)!}$ for some $A, B \in \mathbb{N}_0$ and $B \leq A$, and $\vec{x} = [x_1 \dots x_d]^T$ with $x_i \in \{0, n_i\}$ and $\sum_{i=1}^d x_i = n$.

Uniform Distribution Uniform distribution is a continuous distribution having support over (a, b) where $a, b \in \mathbb{R}$. We use Uniform(a, b) to denote the p.d.f. of a random variable X distributed according to the uniform distribution. The exact definition of the p.d.f. of X is provided below:

$$p(X = x) = \text{Uniform}(x; a, b) = \frac{1}{|a - b|} \forall x \in [a, b].$$

For uniform distribution, we abuse notation by not ensuring $a < b$ for representing the interval $[a, b]$. In the context of uniform distribution, the notation $[a, b]$ is to be understood as $[a, b]$ if $a \leq b$ and $[b, a]$ if $a \geq b$.

Bernoulli Distribution The p.m.f. of the random variable X distributed according to the Bernoulli distribution is denoted as Bernoulli(p) for $p \in [0, 1]$. The exact definition of the p.d.f. of X is provided below:

$$P(X = 0) = p, \text{ and } P(X = 1) = 1 - p. \quad (320)$$

Categorical Distribution The p.m.f. of a random variable X distributed according to the categorical distribution is denoted as Categorical(p_1, \dots, p_d) for $p_i \in [0, 1] \forall i \in [d]$ and $\sum_{i=1}^d p_i = 1$. The exact definition of the p.d.f. of X is provided below:

$$P(X = i) = p_i \forall i \in [d]. \quad (321)$$

N.2 Univariate Stable Distribution Definition

We employ the definition of a univariate stable distribution discussed in [23]:

Definition 2. A random variable X is said to have a stable continuous distribution if, for two independent copies of X , denoted as X_1 and X_2 , and positive constants $a > 0$ and $b > 0$, (322) holds for some $c > 0$ and $d \in \mathbb{R}$.

$$aX_1 + bX_2 \stackrel{d}{=} cX + d, \text{ where } \stackrel{d}{=} \text{means equality in distribution [57]}. \quad (322)$$

Continuous stable distributions are typically expressed through their characteristic functions, as the analytical form of many continuous stable distributions is not known. The characteristic function of a random variable X having a continuous stable distribution is typically parameterized by four parameters: stability parameter denoted as $\alpha \in (0, 2]$, skewness parameter denoted as $\beta \in [-1, 1]$, scale parameter denoted as $\gamma \in (0, \infty)$, and location parameter denoted as $\mu \in \mathbb{R}$. The exact form of the characteristic function is given in (323).

$$\mathbb{E}[e^{itX}] = \phi(t; \alpha, \beta, \gamma, \mu) = \exp(it\mu - |\gamma t|^\alpha (1 - i\beta \operatorname{sgn}(t)\Phi(\alpha))), \quad (323)$$

$$\text{where } \Phi(\alpha) = \begin{cases} \tan(\frac{\pi\alpha}{2}) & \alpha \neq 1 \\ \frac{-2}{\pi} \log(|t|) & \alpha = 1 \end{cases}, \quad \operatorname{sgn}(t) = \begin{cases} -1 & t < 0 \\ 0 & t = 0 \\ 1 & t > 0 \end{cases}.$$

We further denote the p.d.f. of X having the characteristic function defined in (323) as $p_{CS}(X; \alpha, \beta, \gamma, \mu)$. Some well-known examples of continuous stable distribution are the Gaussian/normal distribution, Cauchy distribution, and Lévy distribution. All univariate continuous stable distributions have infinite variance except the Gaussian distribution, and their tails follow a power law behavior [23]. Univariate stable distributions have been widely used to model many financial and physical phenomena. Chapter 2 of [23] provides a list of applications where stable distributions have been used. For a more in-depth treatment of stable distributions, we refer the reader to [23].

N.3 Multivariate Stable Distribution Definition

A multivariate stable distribution is defined similarly to a univariate stable distribution described in Appx. N.2. We employ the definition of multivariate distribution given in [23].

Definition 3. A non-generate d -dimensional random variable \vec{X} is said to have a multivariate stable continuous distribution if, for two independent copies of \vec{X} , denoted as \vec{X}_1 and \vec{X}_2 , and positive constants $a > 0$ and $b > 0$, (324) holds for some $c > 0$ and $\vec{d} \in \mathbb{R}^d$.

$$a\vec{X}_1 + b\vec{X}_2 \stackrel{d}{=} c\vec{X} + \vec{d}, \text{ where } \stackrel{d}{=} \text{ means equality in distribution [57].} \quad (324)$$

A main source of complexity while analyzing multivariate continuous stable distributions is the wide range of dependence structures this family of distribution admits [23]. We refer the reader [46, 45] for a more general treatment of multivariate stable distributions. In this work, we focus on two special classes of multivariate stable distributions: independent component multivariate stable continuous distribution denoted as $p_{CS-IC}(\alpha, \vec{\beta}, \vec{\gamma}, \vec{\mu})$ and elliptically contoured multivariate stable distribution denoted as $p_{CS-EC}(\alpha, \Sigma, \vec{\mu})$. Similarly to the univariate continuous stable distribution, both independent component and elliptically contoured multivariate stable distributions do not have explicit general analytical forms and are expressed through their characteristic functions.

The characteristic function of a d -dimensional random vector \vec{X} having independent component multivariate distribution $p_{CS-IC}(\alpha, \vec{\beta}, \vec{\gamma}, \vec{\mu})$ is expressed as follows:

$$\mathbb{E}[e^{i\vec{t}^T \vec{X}}] = \exp\left(-\sum_{j=1}^d |\gamma_j t_j|^\alpha (1 - i\beta_j \operatorname{sgn}(t_j)\Phi(\alpha)) + i\vec{t}^T \vec{\mu}\right) \quad \forall \vec{t} \in \mathbb{R}^d, \quad (325)$$

where $\Phi(\alpha)$ and $\operatorname{sgn}(\cdot)$ are as defined in (323), $\vec{\beta} = [\beta_1 \dots \beta_d]^T$ with $\beta_j \in [-1, 1]$, $\vec{\gamma} = [\gamma_1 \dots \gamma_d]^T$ with $\gamma_j \in (0, \infty)$, $\vec{\mu} = [\mu_1 \dots \mu_d]^T \in \mathbb{R}^d$, and $\alpha \in (0, 2]$. One can also alternatively think of the random vector $\vec{X} = [X_1 \dots X_d]^T$ having the distribution $p_{CS-IC}(\alpha, \vec{\beta}, \vec{\gamma}, \vec{\mu})$ as a collection of independent random variables $\{X_j\}_{j=1}^d$, where $X_j \sim p_{CS}(\alpha, \beta_j, \gamma_j, \mu_j)$.

The characteristic function of a d -dimensional random vector \vec{X} having elliptically-contoured multivariate distribution $p_{CS-EC}(\alpha, \Sigma, \vec{\mu})$ is expressed as follows:

$$\mathbb{E}[e^{i\vec{t}^T \vec{X}}] = \exp\left(-\left(\vec{t}^T \Sigma \vec{t}\right)^{\alpha/2} + i\vec{t}^T \vec{\mu}\right) \quad \forall \vec{t} \in \mathbb{R}^d, \quad (326)$$

where Σ is a positive definite matrix, $\vec{\mu} = [\mu_1 \dots \mu_d]^T \in \mathbb{R}^d$, and $\alpha \in (0, 2]$. Note that allowing Σ to be positive semi-definite will violate the fact that \vec{X} should be a non-degenerate d -dimensional random variable. We also list an important property of multivariate continuous stable distributions used in this paper.

Property 1. A d -dimensional random vector \vec{X} has a multivariate stable continuous distribution if and only if every 1-dimensional projection has a univariate continuous stable distribution, i.e., $\vec{t}^T \vec{X} \sim p_{CS}(\alpha, \beta(\vec{t}), \gamma(\vec{t}), \mu(\vec{t})) \forall \vec{t} \in \mathbb{R}^d$.

Proof. See the proof of P.1. in [46]. □

N.4 Univariate Discrete Stable Distributions

Discrete stable distributions form the discrete analogues of the continuous stable distributions. We employ the following definition for a univariate discrete distributions taken from [24]:

Definition 4. A random variable X is said to have a discrete stable distribution with exponent $\nu \in (0, 1]$ if for two independent copies of X , denoted as X_1 and X_2 , (322) holds for all $\gamma \in [0, 1]$.

$$\gamma \circ X_1 + (1 - \gamma^\nu)^{1/\nu} \circ X_2 \stackrel{d}{=} X, \quad (327)$$

where $\stackrel{d}{=}$ means equality in distribution [57], and \circ implies a binomial thinning operation defined as follows: $\gamma \circ X = \sum_{j=1}^X N_j$, where $N_j \sim \text{Bernoulli}(\gamma)$, and all N_j 's are jointly independent.

Similar to univariate continuous stable distributions, univariate discrete stable distribution do not have explicit analytical forms, and are typically expressed through their probability generating functions and/or characteristic functions. The characteristic function of a discrete random variable N having a stable discrete distribution is given by (328):

$$\mathbb{E}[e^{itN}] = \exp(\tau(e^{it} - 1)^\nu) = \phi(t; \nu, \tau), \text{ and } P_N(z) = \exp(-\tau(1 - z)^\nu), \quad (328)$$

where $P_N(z; \nu, \tau)$ is the probability generating function of N , $i = \sqrt{-1}$, $\tau > 0$, and $0 < \nu \leq 1$. We further denote the probability mass function of N as $P_{DS}(N = n; \nu, \tau)$. $P_{DS}(N; 1, \lambda)$ is a standard Poisson distribution with mean λ [24]. All discrete stable distribution asymptotically follow a power law distribution [64].

N.5 Multivariate Poisson Distribution

To our knowledge, there are no well-known multivariate generalizations of univariate discrete stable distributions, except for the special case of the Poisson distribution. In this section, we describe a popular multivariate extension of the Poisson distribution.

An intuitive way to define multivariate Poisson distribution is to represent each random variable in the multivariate Poisson distribution as a sum of independent Poisson random variables [41, 42, 43, 44].

To illustrate, let us construct a bivariate Poisson random vector $\vec{N} = [N_1 \ N_2]^T$, where

$$N_1 = N_1^g + N_{12}^g, \ N_2 = N_2^g + N_{12}^g, \quad (329)$$

and N_1^g, N_2^g , and N_{12}^g are mutually independent Poisson random variables with rates λ_1, λ_2 , and λ_{12} , respectively. Here, the dependence between N_1 and N_2 is expressed through N_{12}^g , with covariance between N_1 and N_2 being equal to λ_{12} [47]. For dimensions $d > 2$, the Poisson random vector:

$$\vec{N} = \mathbf{A}\vec{N}^g, \quad (330)$$

where \mathbf{A} is an appropriate matrix of 0's and 1's. We can decompose $\mathbf{A} = [\mathbf{A}_1 \ \dots \ \mathbf{A}_{d'}]$, where \mathbf{A}_i is a $d \times \binom{d}{i}$ submatrix having no duplicate columns, and each of its columns containing exactly i ones and $(d - i)$ zeros [44], and $\vec{N}^g = [N_1^g \ \dots \ N_d^g \ N_{12}^g \ \dots \ N_{(d-1)d}^g \ \dots \ N_{d-(d-1)\dots d}^g]^T$, with $N_{i_1 \dots i_j}^g \sim \text{Poisson}(\lambda_{i_1 \dots i_j}) \forall (i_1, \dots, i_j) \in \mathbb{A}_j^d, j \in [d']$. Furthermore, the random variables $\{N_1, \dots, N_{d-(d-1)\dots d}\}$ are mutually independent. An intuitive way to think about this multivariate Poisson distribution is to interpret $\{N_i^g\}_{i=1}^d$ as the “main effect”, $\{N_{i_1 i_2}^g\}_{(i_1, i_2) \in \mathbb{A}_2^d}$ as the “two-way covariance effects” and so on until $\{N_{i_1 \dots i_{d'}}^g\}_{(i_1, \dots, i_{d'}) \in \mathbb{A}_{d'}^d}$ as the d' -way covariance effect in an ANOVA-like fashion [44]. For a more detailed discussion, see [43, 44, 47, 42]. Note that $d' \leq d$. In this work, we denote the d -dimensional random vector \vec{N} having a multivariate Poisson distribution with parameters $\vec{\Lambda} = [\lambda_1 \ \dots \ \lambda_d \ \lambda_{12} \ \dots \ \lambda_{d-(d-1)\dots d}]^T$ specifying up to d' -way covariance effects (where $d' \leq d$), as $\vec{N} \sim \text{Poisson}(d, d', \vec{\Lambda})$. Note that for $d' = 1$, we have that K is a collection of independent Poisson random variables, and when both $d = d' = 1$, we recover the scalar Poisson distribution: $\Pr(K = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \forall k \in \mathbb{N}_0$.

Deriving the p.m.f. of the Multivariate Poisson Distribution

Let $\vec{N} \sim \text{Poisson}(d, d', \vec{\Lambda})$, where \vec{N} is a d -dimensional random vector. Then the corresponding p.m.f. of \vec{N} is described below:

Let $\vec{n}' = [n_{12} \dots n_{(d-1)d} \dots n_{d-(d'-1)\dots d}]^T$, and $d_{\vec{n}'}$ be the dimension of \vec{n}' , then:

$$P(\vec{N}=\vec{n}) = e^{-1^T \vec{\Lambda}} \prod_{i=1}^d \lambda_i^{n_i} \sum_{\vec{n}' \in C} \left(\prod_{(i_1, i_2) \in \mathbb{A}_2^d} \left(\frac{\lambda_{i_1 i_2}}{\lambda_{i_1} \lambda_{i_2}} \right)^{n_{i_1 i_2}} \times \dots \times \prod_{(i_1, \dots, i_{d'}) \in \mathbb{A}_{d'}^d} \left(\frac{\lambda_{i_1 \dots i_{d'}}}{\prod_{j=1}^{d'} \lambda_{i_j}} \right)^{n_{i_1 \dots i_{d'}}} \times Q(\vec{n}, \vec{n}') \right), \quad (331)$$

where $C = \{\vec{n}' \in \mathbb{N}_0^{d_{\vec{n}'}} : (\vec{a}'_i)^T \vec{n}' \leq n_i \forall i \in [d]\}$, and

$$Q(\vec{n}, \vec{n}') = \prod_{i=1}^d \frac{1}{(n_i - \vec{a}'_i^T \vec{n}')!} \prod_{j=2}^{d'} \prod_{(i_1, \dots, i_j) \in \mathbb{A}_j^d} \frac{1}{n_{i_1 \dots i_j}^g!},$$

with \vec{a}'_i being the i -th row of the matrix $\mathbf{A}' = [\mathbf{A}_2 \dots \mathbf{A}_{d'}]$.

Proof of $\vec{N} \sim \text{Poisson}(d, d', \vec{\Lambda})$ having the p.m.f. described in (331).

We know that $\vec{N} = \mathbf{A} \vec{N}^g$, where $\mathbf{A} = [\mathbf{A}_1 \dots \mathbf{A}_{d'}]$. Let the dimension of \vec{N}^g be $d_{\vec{N}^g}$ and $\mathcal{S}_{\vec{n}} = \{\vec{n}^g \in \mathbb{N}_0^{d_{\vec{N}^g}} : \mathbf{k} = \mathbf{A} \mathbf{k}^g\}$. Then, using the fact $\vec{N} = \mathbf{A} \vec{N}^g$, the p.m.f. of \vec{N} can be expressed using the p.m.f. of \vec{N}^g in the following manner:

$$P(\vec{N} = \vec{n}) = \sum_{\vec{n}^g \in \mathcal{S}_{\vec{n}}} P(\vec{N}^g = \vec{n}^g). \quad (332)$$

Since \vec{N}^g is just a collection of mutually independent Poisson random variables, we can write $P(\vec{N}^g = \vec{n}^g)$ as a product of scalar Poisson distributions, i.e.,

$$P(\vec{N}^g = \vec{n}^g) = \prod_{j=1}^{d'} \prod_{(i_1, \dots, i_j) \in \mathbb{A}_j^d} e^{\lambda_{i_1 \dots i_j}} \frac{\lambda_{i_1 \dots i_j}^{n_{i_1 \dots i_j}^g}}{n_{i_1 \dots i_j}^g!}. \quad (333)$$

Substituting (333) into (332), and collecting all $e^{\lambda_{i_1 \dots i_j}}$ terms we obtain:

$$P(\vec{N}=\vec{n}) = e^{-1^T \vec{\Lambda}} \sum_{\vec{n}^g \in \mathcal{S}_{\vec{n}}} \prod_{j=1}^{d'} \prod_{(i_1, \dots, i_j) \in \mathbb{A}_j^d} \frac{\lambda_{i_1 \dots i_j}^{n_{i_1 \dots i_j}^g}}{n_{i_1 \dots i_j}^g!}. \quad (334)$$

Decompose the matrix $\mathbf{A} = [\mathbf{A}_1 \mathbf{A}']$, where $\mathbf{A}' = [\mathbf{A}_2 \dots \mathbf{A}_{d'}]$ and $\vec{n}^g = [(\vec{n}_1^g)^T \vec{n}'^T]^T$, where $\vec{n}_1^g = [n_1^g \dots n_d^g]^T$ and \vec{n}' contains the rest of the elements in \vec{n}^g . Using the fact that $\vec{n}^g \in \mathcal{S}_{\vec{n}} \Rightarrow \vec{n} = \mathbf{A} \vec{n}^g$ we have: $\vec{n} = \mathbf{A}_1 \vec{n}_1^g + \mathbf{A}' \vec{n}'$. Combining this with the fact that \mathbf{A}_1 is an identity matrix, we obtain:

$$\vec{n} = \vec{n}_1^g + \mathbf{A}' \vec{n}' \Rightarrow \vec{n}_1^g = \vec{n} - \mathbf{A}' \vec{n}' \Rightarrow n_i^g = n_i - (\vec{a}'_i)^T \vec{n}', \quad (335)$$

where \vec{a}'_i is the i -th row of \mathbf{A}' and $i \in [d]$. Substituting (335) into (334):

$$P(\vec{N}=\vec{n}) = e^{-1^T \vec{\Lambda}} \sum_{\vec{n}' \in \mathcal{S}_{\vec{n}'}} \left[\prod_{i_1 \in \mathbb{A}_1^d} \frac{\lambda_{i_1}^{n_i - (\vec{a}'_i)^T \vec{n}'}}{(n_i - (\vec{a}'_i)^T \vec{n}')!} \prod_{j=2}^{d'} \prod_{(i_1, \dots, i_j) \in \mathbb{A}_j^d} \frac{\lambda_{i_1 \dots i_j}^{n_{i_1 \dots i_j}^g}}{n_{i_1 \dots i_j}^g!} \right] \quad (336)$$

where the summation constraint is transformed to $\mathcal{S}_{\vec{n}'} = \{\vec{n}' \in \mathbb{N}_0^{d_{\vec{n}'}} : (\vec{a}'_i)^T \vec{n}' \leq n_i \forall i \in [d]\}$. The equivalence of the summations over the sets $\mathcal{S}_{\vec{n}'}$ and $\mathcal{S}_{\vec{n}}$ can be derived by considering that

$n_i^g = n_i - (\vec{a}'_i)^T \vec{n}'$, and $n_i^g \geq 0$. Collecting all the factorial terms, we get:

$$Q(\vec{n}, \vec{n}') = \prod_{i=1}^d \frac{1}{(n_i - \vec{a}'_i^T \vec{n}')!} \prod_{j=2}^{d'} \prod_{(i_1, \dots, i_j) \in \mathbb{A}_j^d} \frac{1}{n_{i_1 \dots i_j}^g!}. \quad (337)$$

Substituting (337) into (336), we obtain:

$$P(\vec{N}=\vec{n}) = e^{-\vec{1}^T \vec{\Lambda}} \sum_{\vec{n}' \in \mathcal{S}_{\vec{n}'}} \left[\prod_{i_1 \in \mathbb{A}_1^d} \lambda_{i_1}^{n_i - (\vec{a}'_i)^T \vec{n}'} \prod_{j=2}^{d'} \prod_{(i_1, \dots, i_j) \in \mathbb{A}_j^d} \lambda_{i_1 \dots i_j}^{n_{i_1 \dots i_j}^g} \right] Q(\vec{n}, \vec{n}'). \quad (338)$$

Taking the term $\prod_{i_1 \in \mathbb{A}_1} \lambda_{i_1}^{n_i}$ out of the summation as it does not depend upon \vec{n}' , we obtain:

$$P(\vec{N}=\vec{n}) = e^{-\vec{1}^T \vec{\Lambda}} \prod_{i_1 \in \mathbb{A}_1^d} \lambda_{i_1}^{n_i} \sum_{\vec{n}' \in \mathcal{S}_{\vec{n}'}} \left[\prod_{i_1 \in \mathbb{A}_1^d} \lambda_{i_1}^{-(\vec{a}'_i)^T \vec{n}'} \prod_{j=2}^{d'} \prod_{(i_1, \dots, i_j) \in \mathbb{A}_j^d} \lambda_{i_1 \dots i_j}^{n_{i_1 \dots i_j}^g} \right] Q(\vec{n}, \vec{n}'). \quad (339)$$

Notice that $(\vec{a}'_i)^T \vec{n}'$ contains all the elements of \vec{n}' which have i in their subscript. Hence, expanding the term $\prod_{i_1 \in \mathbb{A}_1^d} \lambda_{i_1}^{-(\vec{a}'_i)^T \vec{n}'}$ and distributing over the product, we obtain the desired form:

$$\begin{aligned} P(\vec{N}=\vec{n}) &= e^{-\vec{1}^T \vec{\Lambda}} \prod_{i=1}^d \lambda_i^{n_i} \sum_{\vec{n}' \in C} \left(\prod_{(i_1, i_2) \in \mathbb{A}_2^d} \left(\frac{\lambda_{i_1 i_2}}{\lambda_{i_1} \lambda_{i_2}} \right)^{n_{i_1 i_2}} \times \right. \\ &\quad \left. \dots \times \prod_{(i_1, \dots, i_{d'}) \in \mathbb{A}_{d'}^d} \left(\frac{\lambda_{i_1 \dots i_{d'}}}{\prod_{j=1}^{d'} \lambda_{i_j}} \right)^{n_{i_1 \dots i_{d'}}} \times Q(\vec{n}, \vec{n}') \right). \end{aligned} \quad (340)$$

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: The three main results of this work are clearly enumerated in Section 1 (Introduction) that demonstrate the main claim of the work which is analytically computing partial information decomposition for numerous systems employing well-known distributions.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: We have provided a discussion on the limitations of this work in Section 7.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory Assumptions and Proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: A formal proof of each theorem is provided in the Appx. F- L, with clearly stated assumptions. A formal justification of the arguments used in Sec. 6 is provided in Appx. M. Section 3 discusses the general proof technique used in proving our main theorems. We also provide informal discussions on the constructions used for proving theorems 1- 7 in their respective sections, namely Sec. 4 and Sec. 5.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental Result Reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The details of our simulation results are provided in Section 6, followed by more details in Appendix M.3. We have also provided the exact code used for running the simulation study.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility.

In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We have provided our code with the supplementary materials.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental Setting/Details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We have provided all simulation details in Section 6 and Appendix L.3.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment Statistical Significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We have provided a box-plot of our results accompanied by the actual data points used for computing the box-plot in Figure 1. The factors of variability are clearly stated in Sec. 6.

Guidelines:

- The answer NA means that the paper does not include experiments.

- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments Compute Resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [No]

Justification: We did not feel the need to provide information on the compute resources required to run our simulation, as our main results are of theoretical nature. Furthermore, all our simulation results can be reproduced on any ordinary personal laptop and do not require any special compute resources.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code Of Ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The research conducted is mainly theoretical in nature and we believe it does not violate the NeurIPS code of ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader Impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: While our research is mainly theoretical, we have tried to anticipate how it might affect the society negatively and positively.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: We believe our work poses no such risk.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We have provided the citation and the corresponding license for the code we used for running our simulation in Section 6.

Guidelines:

- The answer NA means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New Assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release any new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and Research with Human Subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: Our work does not involve crowdsourcing nor human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: Our work does not involve crowdsourcing nor human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.