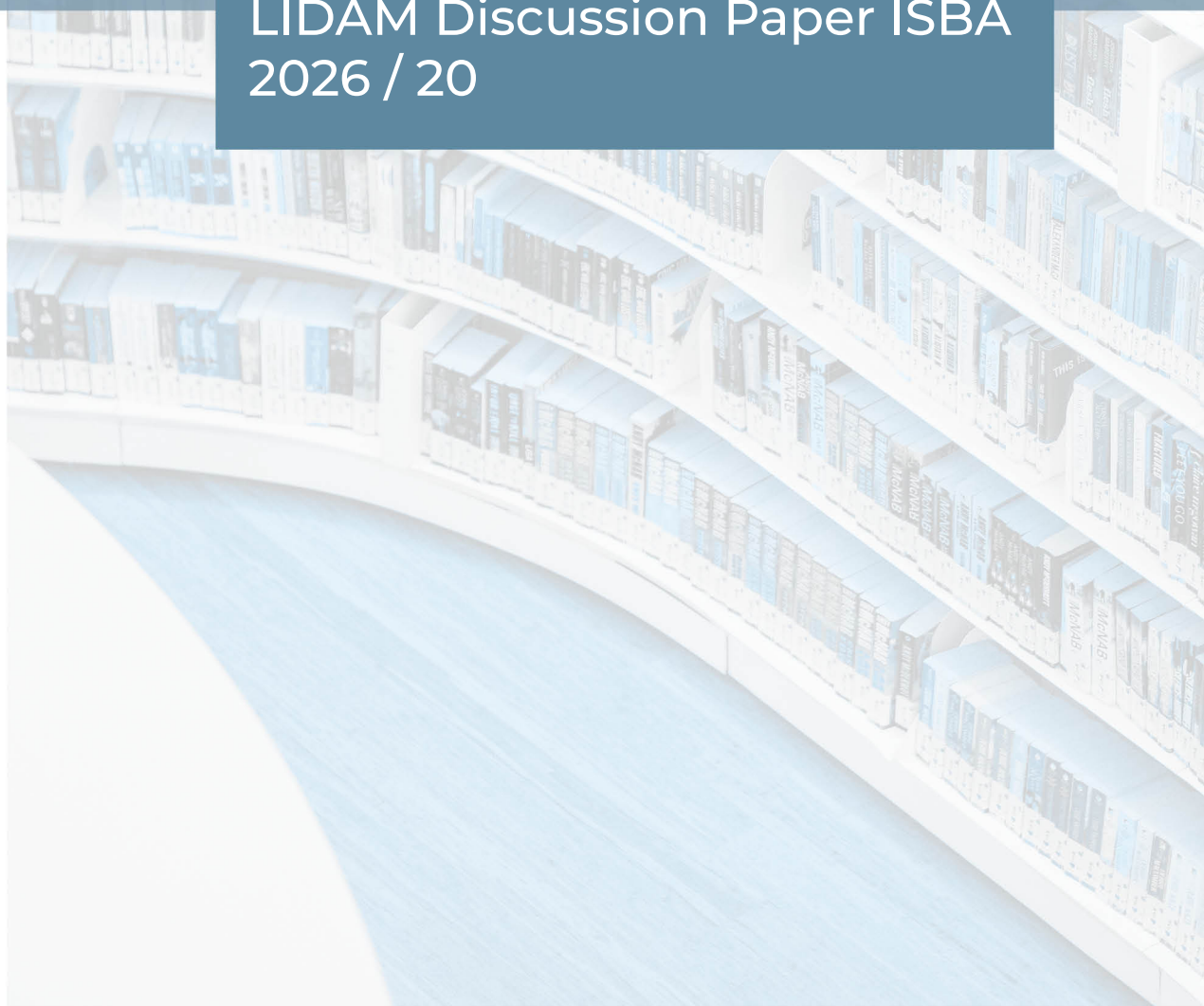
CONVEX ORDER MONOTONICITY OF CONDITIONAL MEAN PREDICTORS IN LATENT MIXTURE MODELS

Valentin Dendoncker, Michel Denuit, Christian Y. Robert

LIDAM Discussion Paper ISBA
2026 / 20



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

CONVEX ORDER MONOTONICITY OF CONDITIONAL MEAN PREDICTORS IN LATENT MIXTURE MODELS

Valentin Dendoncker
Institute of Statistics, Biostatistics and Actuarial Science - ISBA
Louvain Institute of Data Analysis and Modeling - LIDAM
UCLouvain
Louvain-la-Neuve, Belgium
valentin.dendoncker@uclouvain.be
Quantitative and Artificial Intelligence Research Methods at ICHEC - Quaresmi
ICHEC
Bruxelles, Belgium
valentin.dendoncker@ichec.be

Michel Denuit
Institute of Statistics, Biostatistics and Actuarial Science - ISBA
Louvain Institute of Data Analysis and Modeling - LIDAM
UCLouvain
Louvain-la-Neuve, Belgium
michel.denuit@uclouvain.be

Christian Y. Robert
Laboratoire de Sciences Actuarielle et Financière – LSAF
UCBL – Université Claude Bernard Lyon 1, ISFA
Lyon, France
christian.robert@univ-lyon1.fr

June 1, 2026

Abstract

This paper studies the monotonicity, in the convex order, of conditional mean predictors based on aggregate statistics in latent mixture models. For partial sums $S_j = X_1 + \cdots + X_j$, we give general conditions under which $\mathrm{E}[X_i \mid S_j]$ is dominated by $\mathrm{E}[X_i \mid S_{j-1}]$ in the convex order, $i < j$. Such an ordering means that enlarging the conditioning sum reduces the dispersion, and hence the informativeness, of the corresponding conditional mean predictor. The first set of results is based on a projection identity for conditional expectations and includes the classical case where an independent noise term is added to the signal. We then show that the same convex-order reduction follows from a Markov dependence structure. This provides a tractable route for latent mixture models: when the variables are conditionally independent given a common factor and belong to natural exponential families with a common canonical parameter, the sufficiency of partial sums yields the required Markov property. A complementary analysis characterizes situations in which the conditional mean predictor is affine or proportional to the conditioning sum. In particular, for convolution semigroups associated with infinitely divisible distributions, we identify the latent scaling structures leading to proportional predictors. The results apply to a range of standard models, including Poisson, binomial, negative binomial, Gamma, normal, inverse Gaussian, Tweedie, compound Poisson and tempered stable distributions.

Keywords: Convex order; Conditional mean prediction; Latent mixture models; Natural exponential families; Sufficient statistics; Infinitely divisible distributions.

MSC2020 subject classification: 60E15, 60E07, 62P05.

1 Introduction and motivation

Conditional mean predictors play a central role in statistical inference. When a random variable of interest cannot be directly observed, or when only an aggregate statistic is available, the conditional expectation provides the optimal predictor under squared-error loss. More precisely, if X denotes the unobserved quantity and S denotes the available statistic, then $\mathrm{E}[X \mid S]$ is the best L^2-predictor of X among all measurable functions of S. Beyond its optimality property, the distribution of $\mathrm{E}[X \mid S]$ measures the amount of information about X that is contained in S. A highly dispersed conditional mean predictor indicates that the statistic S is informative about X, whereas a nearly constant predictor indicates that little information about X can be recovered from S.

This paper studies such conditional mean predictors when the statistic is an aggregate sum. Specifically, for random variables $X_1, \ldots, X_n$ and partial sums

$$S_j = X_1 + \cdots + X_j, \qquad j = 1, \ldots, n,$$

we investigate conditions under which $\mathrm{E}[X_i \mid S_j]$ is dominated by $\mathrm{E}[X_i \mid S_{j-1}]$ in the convex order for $i < j$. This stochastic order relation is a natural tool for this purpose, since it compares random variables with the same mean in terms of dispersion and is stronger than a variance comparison. In the present setting, the convex order means that, as the conditioning sum incorporates additional components, the conditional mean predictor of X_i becomes less dispersed, and hence less informative about the individual component X_i.

A basic example is the additive-noise model. Assume that an analyst cannot observe a random variable X directly but only a noisy signal $S = X + Y$, where Y represents an error term. The predictor $\mathrm{E}[X \mid S]$ is then the best mean-square reconstruction of X based on S. If further independent noise is added to the signal, the resulting conditional mean predictor becomes less dispersed in the convex order. This formalizes the intuitive idea that accumulating noise reduces the information content of the signal. Related questions have been studied by Ernst et al. (2022) and Huang (2023), while sufficient conditions for convex ordering of conditional expectations were obtained by Mizuno (2006). Denuit (2010) related such inequalities to the bivariate $(2,1)$-increasing convex order introduced by Denuit, Lefevre and Mesfioui (1999). See also Shaked et al. (2012) for global dependence orders based on conditional means and variances.

The same issue arises in regression models with latent or unobserved covariates. Consider, for instance, a linear model

$$S = \alpha_1 Y_1 + \cdots + \alpha_p Y_p + \mathcal{E} \text{ with } \mathrm{E}[\mathcal{E} \mid Y_1, \ldots, Y_p] = 0.$$

Suppose that the contribution $\alpha_1 Y_1$ is not observed and must be imputed from the aggregate response S, or from a statistic derived from it. The natural imputation is then the conditional mean predictor $\mathrm{E}[\alpha_1 Y_1 \mid S]$. This predictor is optimal under squared-error loss, but its usefulness depends on how informative S is about the missing contribution. If the remaining terms $\alpha_2 Y_2, \ldots, \alpha_p Y_p$ and the error $\mathcal{E}$ act as additional noise, the conditional mean predictor of $\alpha_1 Y_1$ is expected to become less variable. Convex-order comparisons provide a distributional way to quantify this loss of information. This interpretation is particularly

relevant when the regressors are not independent but are affected by a common latent environment, so that conditional independence given an unobserved factor is a plausible modeling assumption.

The conditional mean predictors studied in this paper also appear in several other contexts. In risk sharing, if two agents face losses X and Y and share the aggregate loss $S = X + Y$, the quantities $\mathrm{E}[X \mid S]$ and $\mathrm{E}[Y \mid S]$ define the conditional mean risk-sharing rule proposed by Denuit and Dhaene (2012). In this context, participants generally prefer less variable contributions, which makes convex-order comparisons directly meaningful. See Denuit (2019) and Denuit and Robert (2020, 2021, 2023) for related developments. Conditional expectations given sums also occur in selection problems, as in Van Bochove (2011), and in random-walk models, where one may be interested in the conditional expectation of a previous increment given the current position. Related limit results were studied by Zabell (1980, 1993) and Bar-Lev et al. (2013). In economic theory, agents often infer a value from imperfect signals; for example, Ganuza and Penalva (2010) used dispersion-based signal orderings to compare the information conveyed by signals in auctions.

Most classical results focus on the independent case. If X and Y are independent, the addition of noise to a signal reduces the dispersion of the conditional mean predictor. Under dependence, however, the situation is more subtle. Noise need not be detrimental in the same way: if Y is a monotone function of X, then $X+Y$ may still reveal X exactly, so that $\mathrm{E}[X \mid X + Y] = X$. Thus the dependence structure between the variable of interest and the other components of the aggregate statistic is crucial. We refer to Denuit and Robert (2022) for a study of conditional mean decompositions under dependence and conditional independence.

The present paper focuses on latent mixture models, where the variables are typically dependent marginally but become independent conditionally on a common latent factor. Such models are natural in statistical applications involving unobserved heterogeneity, common environments, random effects, systematic risk factors or experimental conditions shared by several observations. Our first contribution is to identify a general projection identity under which conditioning on a larger sum reduces the corresponding conditional mean predictor in the convex order. This identity includes, as a special case, the addition of an independent noise term.

Our second contribution is to connect this projection identity with a Markov dependence structure. Precisely, if $X_i \to S_{j-1} \to S_j$ is a Markov chain then it is established that $\mathrm{E}[X_i \mid S_j]$ is dominated by $\mathrm{E}[X_i \mid S_{j-1}]$ in the convex order. We then show that this Markov property arises naturally in latent mixture models based on natural exponential families with a common canonical parameter. In this case, the partial sum is a sufficient statistic for the common latent factor, and this sufficiency mechanism yields the required convex-order monotonicity.

Our third contribution concerns models in which the conditional mean predictor is affine, or proportional, to the conditioning sum. We provide algebraic conditions ensuring that such affine predictors are ordered in the convex sense. We then specialize the analysis to convolution semigroups associated with infinitely divisible distributions. This gives explicit conditions under which proportional predictors arise in conditionally independent latent scaling models. The results apply to many standard distributions used in probability and statistics, including Poisson, binomial, negative binomial, gamma, normal, inverse Gaussian,

Tweedie, compound Poisson and tempered stable models.

The remainder of the paper is organized as follows. Section 2 recalls the convex order and establishes a general projection criterion ensuring that conditional mean predictors decrease in convex order when the conditioning sum is enlarged. Section 3 relates this criterion to Markov dependence structures and applies the result to latent mixtures of natural exponential families with common canonical parameter. Section 4 studies affine and proportional conditional mean predictors and derives consequences for infinitely divisible distributions and convolution semigroups. Technical arguments are collected in the Appendix.

Throughout the paper, all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$. They are assumed to be integrable and to have finite variance, that is, to belong to $L^2(\Omega, \mathcal{F}, \mathrm{P})$. All equalities and inequalities between random variables are understood to hold almost surely.

2 Convex-order stochastic inequalities for conditional mean predictors given sums

2.1 Convex order

Given two random variables V and W, V is said to be smaller than W in the convex order, henceforth denoted by $V \preceq_{\mathrm{CX}} W$ if the inequality $\mathrm{E}[g(V)] \leq \mathrm{E}[g(W)]$ holds for all convex functions g for which the expectations exist. Since the functions $x \mapsto \pm x$ are both convex, $\preceq_{\mathrm{CX}}$ only applies to random variables with the same expected value. The stochastic inequality $V \preceq_{\mathrm{CX}} W$ intuitively means that V and W have the same "size" (as $\mathrm{E}[V] = \mathrm{E}[W]$ holds) but that W is "more variable" than V. For instance, the variance of W is larger than the variance of V. A general treatment of this order relation can be found in Müller and Stoyan (2002) or Shaked and Shanthikumar (2007), for instance.

In this paper, the convex order is used to compare conditional mean predictors based on aggregate statistics. More precisely, if a random variable of interest is observed through a sum involving additional components, we study whether the conditional mean predictor becomes less dispersed when the conditioning sum is enlarged. Such a decrease in convex order can be interpreted as a loss of informativeness of the corresponding aggregate statistic.

2.2 A projection criterion for convex-order reduction

We first state a general projection criterion ensuring that conditioning on a larger sum reduces the conditional mean predictor in the convex order. Denuit and Robert (2023) derived conditions on a triplet (U_1, U_2, W) of random variables ensuring that $\mathrm{E}[U_1|U_2 + W]$ is smaller than $\mathrm{E}[U_1|U_2]$ in the convex order. Notice that this statement is not always true. For instance, with $W = U_1 - U_2$ we get $\mathrm{E}[U_1|U_2 + W] = U_1$ which dominates $\mathrm{E}[U_1|U_2]$ in the convex order (by Jensen's inequality). The following result provides a sufficient condition ensuring that conditioning on a sum decreases the conditional expectation in the convex order.

Proposition 2.1. *Let* (U_1, U_2, W) *be such that*

$$\mathrm{E}[U_1 \mid U_2 + W] = \mathrm{E}\big[\mathrm{E}[U_1 \mid U_2]\big|U_2 + W\big] \tag{2.1}$$

holds true. Then, $\mathrm{E}[U_1 \mid U_2 + W] \preceq_{\mathrm{CX}} \mathrm{E}[U_1 \mid U_2]$.

Proof. Let $g : \mathbb{R} \to \mathbb{R}$ be a convex function for which the expectations $\mathrm{E}[g\,(\mathrm{E}[U_1 \mid U_2])]$ and $\mathrm{E}[g\,(\mathrm{E}\,[U_1 \mid U_2 + W])]$ are finite. By Jensen's inequality,

$$\begin{aligned}\mathrm{E}[g(\mathrm{E}[U_1 \mid U_2])] &= \mathrm{E}\big[\mathrm{E}[g\,(\mathrm{E}[U_1 \mid U_2])\,|U_2 + W]\big] \\ &\geq \mathrm{E}\big[g\,(\mathrm{E}[\mathrm{E}[U_1 \mid U_2] \mid U_2 + W])\,\big] \\ &= \mathrm{E}[g\,(\mathrm{E}\,[U_1 \mid U_2 + W])],\end{aligned}$$

which ends the proof. □

Remark 2.2. Setting $Z := U_2 + W$, $U := \mathrm{E}[U_1 \mid Z]$ and $V := \mathrm{E}[U_1 \mid U_2]$, condition (2.1) can be rewritten as

$$U = \mathrm{E}[V \mid Z].$$

Since $\sigma(U) \subseteq \sigma(Z)$, the tower property gives

$$\mathrm{E}[V \mid U] = \mathrm{E}\big[\mathrm{E}[V \mid Z]\big|U\big] = U.$$

The joint distribution of (U, V) is thus a martingale coupling between the laws of U and V in the sense of Strassen. Consequently, Theorem 1.2 in Leskelä and Vihola (2017) yields $U \preceq_{\mathrm{CX}} V$, which is precisely the stochastic inequality $\mathrm{E}[U_1 \mid U_2 + W] \preceq_{\mathrm{CX}} \mathrm{E}[U_1 \mid U_2]$ stated under Proposition 2.1.

Example 2.3. *Denuit and Robert (2023) considered* (U_1, U_2, W) *such that*

$$\mathrm{E}[U_1 \mid U_2, W] = \mathrm{E}[U_1 \mid U_2]. \tag{2.2}$$

In this case,

$$\begin{aligned}\mathrm{E}\big[\mathrm{E}[U_1 \mid U_2]\big|U_2 + W\big] &= \mathrm{E}\big[\mathrm{E}[U_1 \mid U_2, W]\big|U_2 + W\big] \\ &= \mathrm{E}\big[\mathrm{E}[U_1 \mid U_2, U_2 + W]\big|U_2 + W\big] \\ &= \mathrm{E}[U_1 \mid U_2 + W]\end{aligned}$$

so that (2.1) *holds true. Proposition 2.1 then shows that* $\mathrm{E}[U_1 \mid U_2 + W] \preceq_{\mathrm{CX}} \mathrm{E}[U_1 \mid U_2]$ *as established by Denuit and Robert (2023) in their Proposition 2.1.*

Remark 2.4. The implication (2.2) ⇒ (2.1) admits an intuitive interpretation in terms of Hilbert spaces, viewing conditional expectations as orthogonal projections in $L^2(\Omega, \mathcal{F}, \mathrm{P})$. In this way, (2.2) is equivalent to requiring that

$$\mathrm{E}[U_1 \mid \sigma(U_2, W)] = \mathrm{E}[U_1 \mid \sigma(U_2)],$$

which means that the projections of U_1 onto $L^2(\sigma(U_2, W))$ and $L^2(\sigma(U_2))$ coincide. Projecting further onto $L^2(\sigma(U_2 + W))$ yields

$$\mathrm{E}\big[\mathrm{E}[U_1 \mid \sigma(U_2)]\big|\sigma(U_2 + W)\big] = \mathrm{E}[U_1 \mid \sigma(U_2 + W)],$$

that is, the residual $U_1 - \mathrm{E}[U_1|\sigma(U_2)]$ is orthogonal to $L^2(\sigma(U_2 + W))$. This provides a geometric explanation of the implication (2.2) ⇒ (2.1) and why the converse fails in general.

Remark 2.5. Clearly, (2.2) is valid when W is independent of (U_1, U_2). Taking $W = Z$ independent of (X, Y), $U_1 = X$ and $U_2 = X + Y$, Example 2.3 gives

$$\mathrm{E}[X|X+Y+Z] \preceq_{\mathrm{CX}} \mathrm{E}[X|X+Y] \Rightarrow \mathrm{Var}[\mathrm{E}[X|X+Y+Z]] \leq \mathrm{Var}[\mathrm{E}[X|X+Y]]. \quad (2.3)$$

Adding an independent term to the conditioning sum thus always reduces the conditional expectation in the convex order as established by Denuit and Robert (2023). The variance inequality in (2.3) corresponds to formula (4.3) in Ernst et al. (2022). It can thus be strengthened as a stochastic inequality in terms of the convex order. Also, (2.3) extends the variance inequality stated under Lemma 3.2 in Huang (2023) where X and Y are independent and Z is Normally distributed and independent of (X, Y).

2.3 Accumulation of independent noise

Let us consider the application to signals discussed in the introduction. If the signal becomes more noisy, in the sense that more errors accumulate, then the conditional expectation given signal becomes smaller in convex order and thus less informative. Precisely, let $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ be independent errors. Define $Y_j := \sum_{k=1}^{j} \mathcal{E}_k$, $j = 1, 2, \ldots, n$. Condition (2.2) is fulfilled with $U_1 = X$, $U_2 = Y_{j-1}$, and $W = \mathcal{E}_j$. Example 2.3 then gives

$$\mathrm{E}[X|X+Y_j] \preceq_{\mathrm{CX}} \mathrm{E}[X|X+Y_{j-1}] \text{ for } j = 2, \ldots, n. \quad (2.4)$$

This expresses the fact that $\mathrm{E}[X|X+Y_j]$ becomes less informative as j increases because the signal $S_j = X + Y_j = X + \sum_{k=1}^{j} \mathcal{E}_k$ is subject to more noise and thus loses its predictive power.

Assume for instance that X is observed up to error $\mathcal{E}_1 \sim \mathrm{Normal}(0, \sigma_1^2)$, where $\mathrm{Normal}(a, b)$ denotes the normal distribution with mean a and variance b and "$\sim$" reads as "is distributed as". Then, increasing the variance to σ_2^2, $\sigma_2^2 > \sigma_1^2$, is equivalent to replace $\mathcal{E}_1$ with $\mathcal{E}_1 + \mathcal{E}_2$ where $\mathcal{E}_2 \sim \mathrm{Normal}(0, \sigma_2^2 - \sigma_1^2)$. Thus, we know from (2.4) that the conditional expectation gets smaller in the convex order when the variance of the Gaussian error increases, so that the signal becomes less informative. The same comment applies with the Binomial error structures considered by Ernst et al. (2022) in their Section 5.1, as well as to Poisson or gamma noise, for instance. These findings are summarized in the following result, where $\mathrm{Poisson}(a)$ denotes the Poisson distribution with mean a, $\mathrm{Binomial}(k, p)$ denotes the binomial distribution with size k and success probability p, and $\mathrm{Gamma}(a, b)$ denotes the gamma distribution with mean a/b and variance a/b^2.

Property 2.6. *Considering independent random variables X, Y_1, and Y_2, the stochastic inequality $\mathrm{E}[X|X+Y_2] \preceq_{\mathrm{CX}} \mathrm{E}[X|X+Y_1]$ holds true if*

(i) $Y_i \sim \mathrm{Normal}(0, \sigma_i^2)$ *with* $\sigma_1^2 \leq \sigma_2^2$.

(ii) $Y_i \sim \mathrm{Poisson}(\lambda_i)$ *with* $\lambda_1 \leq \lambda_2$.

(iii) $Y_i \sim \mathrm{Binomial}(k_i, p)$ *with* $k_1 \leq k_2$, *for any* $p \in (0, 1)$.

(iv) $Y_i \sim \mathrm{Gamma}(a_i, b)$ *with* $a_1 \leq a_2$ *for any* $b > 0$.

Similar results are obtained when Y_i belongs to any parametric family satisfying the semi-group property.

The remainder of this paper studies whether results similar to those stated under Property 2.6 still hold when the terms entering the condition are no more independent but only conditionally independent.

3 Markov dependence structures and natural exponential families

3.1 Markov chain structure

Condition (2.2) is referred to as "conditional mean independence" in econometrics, where this concept is used to decide whether additional features must be included in a regression model. This section considers Markov chains as another conditional independence structure leading to the desired reduction in convex order, as shown by the next result.

Proposition 3.1. *Let X, Y, and Z be real-valued random variables and assume that $X \to Y \to Z$ is a Markov chain, that is, there exists a Markov kernel $K : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0,1]$ such that, for every bounded Borel function $h : \mathbb{R} \to \mathbb{R}$,*

$$\mathrm{E}[h(Z) \mid X, Y] = \int_{\mathbb{R}} h(z)\, K(Y, dz).$$

Then, $\mathrm{E}[X \mid Z] \preceq_{\mathrm{CX}} \mathrm{E}[X \mid Y]$.

Proof. By assumption, X and Z are conditionally independent given Y. Hence,

$$\mathrm{E}[X \mid Y, Z] = \mathrm{E}[X \mid Y] \Rightarrow \mathrm{E}[X \mid Z] = \mathrm{E}\big[\mathrm{E}[X \mid Y, Z]\big|Z\big] = \mathrm{E}\big[\mathrm{E}[X \mid Y]\big|Z\big].$$

Applying Proposition 2.1 with $U_1 := X$, $U_2 := Y$, and $W := Z - Y$ then ends the proof. □

Proposition 3.1 shows that the conditional expectation of the first element in a Markov chain, given future states, decreases in the convex order when the time horizon increases.

Let us now apply Proposition 3.1 to conditional expectations given sums.

Corollary 3.2. *Let $X_1, \ldots, X_n$ be real-valued random variables with partial sums $S_j = X_1 + \ldots + X_j$, $j \in \{1, \ldots, n\}$. Consider indices i and $j \in \{1, \ldots, n\}$ such that $i < j$. If $X_i \to S_{j-1} \to S_j$ is a Markov chain, then* $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$.

Proof. The announced result is a direct consequence of Proposition 3.1 with $X := X_i$, $Y = S_{j-1}$, and $Z = S_j$. □

3.2 Latent mixtures of natural exponential families

We now identify a broad class of latent mixture models for which the Markov property in Corollary 3.2 is automatically satisfied. The key mechanism is sufficiency. Conditionally on the latent factor, the variables belong to natural exponential families with a common

canonical parameter. As a consequence, the partial sum S_{j-1} is sufficient for the latent parameter from the observation of $(X_1, \ldots, X_{j-1})$. This removes the additional information contained in the individual decomposition of S_{j-1}, and yields the required Markov structure.

Proposition 3.3. *Let $X_1, \ldots, X_n$ be real-valued random variables with partial sums $S_j = X_1 + \ldots + X_j$, $j \in \{1, \ldots, n\}$. Let Θ be a random variable such that $\mathrm{P}[\Theta \in D] = 1$ for some Borel set $D \subseteq \mathbb{R}$. Assume that, conditionally on Θ, the random variables $X_1, \ldots, X_n$ are independent and that the conditional distribution of X_k given Θ is as follows:*

$$\mathrm{P}[X_k \in dx \mid \Theta = \theta] = \exp(\theta x - \kappa_k(\theta))\, \rho_k(dx), \qquad \theta \in D,$$

where the σ-finite measure ρ_k on $\mathbb{R}$ is such that

$$0 < \int_{\mathbb{R}} e^{\theta x} \rho_k(dx) < \infty \text{ and } \kappa_k(\theta) := \ln \int_{\mathbb{R}} e^{\theta x} \rho_k(dx), \ \ \theta \in D.$$

Then, $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds for every $1 \leq i < j \leq n$.

Proof. Fix $i < j$ and set $\boldsymbol{Y}_{j-1} := (X_1, \ldots, X_{j-1})$. For $k = 1, \ldots, j-1$, write $q_{k,\theta}(x) := \exp(\theta x - \kappa_k(\theta))$ so that

$$\mathrm{P}[X_k \in dx_k \mid \Theta = \theta] = q_{k,\theta}(x_k)\rho_k(dx_k), \qquad \theta \in D.$$

Let $\mu_{j-1} := \rho_1 \otimes \cdots \otimes \rho_{j-1}$ and $P_\theta^{(j-1)}$ be the conditional law of $\boldsymbol{Y}_{j-1}$ given $\Theta = \theta$. For every non-negative Borel function $\varphi : \mathbb{R}^{j-1} \to \mathbb{R}^+$,

$$\int_{\mathbb{R}^{j-1}} \varphi(x_1, \ldots, x_{j-1}) P_\theta^{(j-1)}(dx_1, \ldots, dx_{j-1}) = \int_{\mathbb{R}^{j-1}} \varphi(x_1, \ldots, x_{j-1}) \prod_{k=1}^{j-1} q_{k,\theta}(x_k) \mu_{j-1}(dx_1, \ldots, dx_{j-1}).$$

Hence, $P_\theta^{(j-1)} \ll \mu_{j-1}$ and

$$\frac{dP_\theta^{(j-1)}}{d\mu_{j-1}}(x_1, \ldots, x_{j-1}) = \prod_{k=1}^{j-1} q_{k,\theta}(x_k) = \exp\left(\theta \sum_{k=1}^{j-1} x_k - \sum_{k=1}^{j-1} \kappa_k(\theta)\right) = c_\theta(s),$$

where

$$s = \sum_{k=1}^{j-1} x_k \text{ and } c_\theta(s) = \exp\left(\theta s - \sum_{k=1}^{j-1} \kappa_k(\theta)\right).$$

By the Fisher–Neyman factorization criterion, there exists a Markov kernel $Q : \mathbb{R} \times \mathcal{B}(\mathbb{R}^{j-1}) \to [0, 1]$, independent of θ, such that $\mathrm{P}[\boldsymbol{Y}_{j-1} \in d\boldsymbol{y} \mid S_{j-1}, \Theta] = Q(S_{j-1}, d\boldsymbol{y})$ almost surely holds. Consequently, for every $A \in \mathcal{B}(\mathbb{R}^{j-1})$,

$$\mathrm{P}[\boldsymbol{Y}_{j-1} \in A \mid S_{j-1}] = \mathrm{E}\big[\mathrm{P}[\boldsymbol{Y}_{j-1} \in A \mid S_{j-1}, \Theta] \big| S_{j-1}\big] = Q(S_{j-1}, A).$$

Hence, for every integrable measurable function m,

$$\mathrm{E}[m(\boldsymbol{Y}_{j-1}) \mid S_{j-1}, \Theta] = \int_{\mathbb{R}^{j-1}} m(\boldsymbol{y}) Q(S_{j-1}, d\boldsymbol{y}) = \mathrm{E}[m(\boldsymbol{Y}_{j-1}) \mid S_{j-1}]. \tag{3.1}$$

In particular, since X_i is a measurable function of $\boldsymbol{Y}_{j-1}$ because $i < j$, we have

$$\mathrm{E}[X_i \mid S_{j-1}, \Theta] = \mathrm{E}[X_i \mid S_{j-1}].$$

We now prove the Markov property ensuring the desired convex-order comparison. Since X_j is conditionally independent of $\boldsymbol{Y}_{j-1}$ given Θ, for every bounded Borel function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathrm{E}[g(X_j) \mid \boldsymbol{Y}_{j-1}, \Theta] = \mathrm{E}[g(X_j) \mid \Theta].$$

Since S_{j-1} is measurable with respect to $\boldsymbol{Y}_{j-1}$, we have $\sigma(\boldsymbol{Y}_{j-1}, S_{j-1}, \Theta) = \sigma(\boldsymbol{Y}_{j-1}, \Theta)$, and therefore

$$\mathrm{E}[g(X_j) \mid \boldsymbol{Y}_{j-1}, S_{j-1}, \Theta] = \mathrm{E}[g(X_j) \mid \Theta].$$

On the other hand, taking conditional expectation with respect to (S_{j-1}, Θ) yields

$$\mathrm{E}[g(X_j) \mid S_{j-1}, \Theta] = \mathrm{E}\big[\mathrm{E}[g(X_j) \mid \boldsymbol{Y}_{j-1}, S_{j-1}, \Theta]\big|S_{j-1}, \Theta\big] = \mathrm{E}[g(X_j) \mid \Theta].$$

Thus,

$$\mathrm{E}[g(X_j) \mid \boldsymbol{Y}_{j-1}, S_{j-1}, \Theta] = \mathrm{E}[g(X_j) \mid S_{j-1}, \Theta],$$

which shows that X_j and $\boldsymbol{Y}_{j-1}$ are conditionally independent given (S_{j-1}, Θ). In particular, X_j and X_i are conditionally independent given (S_{j-1}, Θ).

Let $h, g : \mathbb{R} \to \mathbb{R}$ be bounded Borel functions. Since X_i is a measurable function of $\boldsymbol{Y}_{j-1}$, the preceding conditional independence gives

$$\mathrm{E}[h(X_i)g(X_j) \mid S_{j-1}, \Theta] = \mathrm{E}[h(X_i) \mid S_{j-1}, \Theta]\mathrm{E}[g(X_j) \mid S_{j-1}, \Theta].$$

Using (3.1), we get $\mathrm{E}[h(X_i) \mid S_{j-1}, \Theta] = \mathrm{E}[h(X_i) \mid S_{j-1}]$ so that

$$\begin{aligned}
\mathrm{E}[h(X_i)g(X_j) \mid S_{j-1}] &= \mathrm{E}\big[\mathrm{E}[h(X_i)g(X_j) \mid S_{j-1}, \Theta]\big|S_{j-1}\big] \\
&= \mathrm{E}\big[\mathrm{E}[h(X_i) \mid S_{j-1}, \Theta]\,\mathrm{E}[g(X_j) \mid S_{j-1}, \Theta]\big|S_{j-1}\big] \\
&= \mathrm{E}[h(X_i) \mid S_{j-1}]\,\mathrm{E}\big[\mathrm{E}[g(X_j) \mid S_{j-1}, \Theta]\big|S_{j-1}\big] \\
&= \mathrm{E}[h(X_i) \mid S_{j-1}]\,\mathrm{E}[g(X_j) \mid S_{j-1}].
\end{aligned}$$

Therefore, X_i and X_j are conditionally independent given S_{j-1}.

It remains to check whether the Markov chain dependence structure is present. Let $R : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0, 1]$ be a regular conditional distribution of X_j given S_{j-1}. Since X_i and X_j are conditionally independent given S_{j-1}, the conditional law of X_j given (X_i, S_{j-1}) is also $R(S, dx)$. Thus, for every bounded Borel function $h : \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned}
\mathrm{E}[h(S_j) \mid X_i, S_{j-1}] &= \mathrm{E}[h(S_{j-1} + X_j) \mid X_i, S_{j-1}] \\
&= \int_{\mathbb{R}} h(S_{j-1} + x)R(S_{j-1}, dx).
\end{aligned}$$

Define the Markov kernel $K(s, A) := R(s, A - s)$, $s \in \mathbb{R}$, $A \in \mathcal{B}(\mathbb{R})$, where $A - s := \{x \in \mathbb{R} : s + x \in A\}$. Then,

$$\mathrm{E}[h(S_j) \mid X_i, S_{j-1}] = \int_{\mathbb{R}} h(z)K(S_{j-1}, dz).$$

The conditional law of S_j given (X_i, S_{j-1}) therefore depends only on S_{j-1}. Equivalently, $X_i \to S_{j-1} \to S_j$ is a Markov chain. Applying Corollary 3.2 then ends the proof. □

Remark 3.4. The common canonical parametrization in Proposition 3.3 can be viewed as the form naturally associated with sufficiency of partial sums in product exponential models. To see this, consider for instance a more general model in which, conditionally on $\Theta = \theta$, the density of X_k with respect to Lebesgue measure is of the form $h_k(x)\exp(\eta_k(\theta)x - A_k(\theta))$. If, for some $j \geq 2$, the statistic $S_j = X_1 + \cdots + X_j$ is sufficient for the conditional family $\{\mathcal{L}(X_1, \ldots, X_j \mid \Theta = \theta) : \theta \in D\}$, then the Fisher–Neyman factorization criterion implies that likelihood ratios depend on $(x_1, \ldots, x_j)$ only through $x_1 + \cdots + x_j$. Hence, for any fixed $\theta_0 \in D$, the linear form $\sum_{k=1}^{j}(\eta_k(\theta) - \eta_k(\theta_0))x_k$ must be a function of $x_1 + \cdots + x_j$, and therefore a multiple of this sum. Thus, choosing $\eta(\theta) := \eta_1(\theta)$, one can write $\eta_k(\theta) = \eta(\theta) + c_k$, where $c_k := \eta_k(\theta_0) - \eta_1(\theta_0)$ is independent of θ. Absorbing these constants into the base measures, namely by setting $\rho_k(dx) = e^{c_k x}h_k(x)\,dx$, yields $\mathrm{P}[X_k \in dx \mid \Theta = \theta] = \exp(\eta(\theta)x - \kappa_k(\eta(\theta)))\rho_k(dx)$, with $\kappa_k(u) := \ln \int e^{ux}\rho_k(dx)$. This is the form used in Proposition 3.3, up to the reparametrization $\theta \mapsto \eta(\theta)$.

3.3 Applications

Proposition 3.3 applies to many standard latent mixture models. In each case below, the conditional distributions belong to natural exponential families with a common canonical parameter, while heterogeneity across components is absorbed into the base measures or exposure parameters.

3.3.1 Poisson distribution

Assume that $X_1, \ldots, X_n$ are independent given $\Lambda > 0$, with $X_k \sim \mathrm{Poisson}(w_k\Lambda)$ where $w_k > 0$. Set $\Theta := \ln \Lambda$. Then, for $r \in \mathbb{N}$,

$$\mathrm{P}[X_k = r \mid \Theta] = \exp\bigl(r\Theta - w_k e^{\Theta}\bigr)\,\frac{w_k^r}{r!}.$$

Conditionally on Θ, X_k thus belongs to a natural exponential family with common canonical parameter Θ, with

$$\rho_k(\{r\}) = \frac{w_k^r}{r!}, \qquad r \in \mathbb{N}, \qquad \text{and} \qquad \kappa_k(\theta) = w_k e^{\theta}, \qquad \theta \in \mathbb{R}.$$

By virtue of Proposition 3.3, the stochastic inequality $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds true for every $i < j \leq n$. Thus, given any counting variables conditionally independent and Poisson distributed given Λ, with mean proportional to Λ, the conditional expectation of each count given its sum with an increasing number of other counts decreases in the convex order.

3.3.2 Binomial distribution

Assume that, conditionally on a latent success probability $P \in (0, 1)$, the variables $X_1, \ldots, X_n$ are independent and such that $X_k \sim \mathrm{Binomial}(m_k, P)$, where $m_k \in \mathbb{N} \setminus \{0\}$. Set $\Theta := \ln \frac{P}{1-P}$.

Then, for $r = 0, \ldots, m_k$,

$$\begin{aligned} \mathrm{P}[X_k = r \mid \Theta] &= \binom{m_k}{r} \left(\frac{e^{\Theta}}{1+e^{\Theta}}\right)^r \left(\frac{1}{1+e^{\Theta}}\right)^{m_k - r} \\ &= \exp\bigl(r\Theta - m_k \log(1+e^{\Theta})\bigr) \binom{m_k}{r}. \end{aligned}$$

Conditionally on Θ, X_k thus belongs to a natural exponential family with common canonical parameter Θ, with

$$\rho_k(\{r\}) = \binom{m_k}{r} \mathbf{1}_{\{0,\ldots,m_k\}}(r), \qquad r \in \mathbb{N}, \qquad \text{and} \qquad \kappa_k(\theta) = m_k \log(1+e^{\theta}), \qquad \theta \in \mathbb{R}.$$

Therefore, Proposition 3.3 applies, which implies that $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds for every $i < j \leq n$. Thus, considering a number of successes recorded over m_k repetitions of a Bernoulli experiment, its conditional expectation given its sum with additional numbers of successes that are conditionally independent given the common, unknown success probability, decreases in the convex order.

3.3.3 Negative binomial distribution

Suppose that, conditionally on a latent parameter $P \in (0,1)$ such that $\mathrm{E}\left[\frac{P}{1-P}\right] < \infty$, the variables $X_1, \ldots, X_n$ are independent and follow the negative binomial distribution with parameters (r_k, P), which is denoted as $X_k \sim \mathrm{NegBin}(r_k, P)$, where $r_k > 0$. Set $\Theta := \ln P < 0$. Then,

$$\begin{aligned} \mathrm{P}[X_k = r \mid \Theta] &= \frac{1}{r!} \frac{\Gamma(r_k + r)}{\Gamma(r_k)} \exp\bigl(r\Theta + r_k \log(1 - e^{\Theta})\bigr) \\ &= \exp(r\Theta - \kappa_k(\Theta))\, \rho_k(\{r\}), \end{aligned}$$

where

$$\kappa_k(\theta) = -r_k \log(1 - e^{\theta}), \qquad \theta < 0, \qquad \text{and} \qquad \rho_k(\{r\}) = \frac{1}{r!} \frac{\Gamma(r_k + r)}{\Gamma(r_k)}, \qquad r \in \mathbb{N}.$$

By virtue of Proposition 3.3, the stochastic inequality $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds true for every $i < j \leq n$. This result can be interpreted in terms of number of failures before getting a specified number of successes in Bernoulli trials or as Poisson common mixture model with gamma distributed mean.

3.3.4 Gamma distribution

Assume that, conditionally on a latent factor $B > 0$ satisfying $\mathrm{E}\left[\frac{1}{B}\right] < \infty$, the variables $X_1, \ldots, X_n$ are independent and such that $X_k \sim \mathrm{Gamma}(\alpha_k, B)$, where $\alpha_k > 0$. Set $\Theta := -B < 0$. The conditional probability density function is

$$\begin{aligned} f_k(x \mid \Theta) &= \frac{B^{\alpha_k}}{\Gamma(\alpha_k)} x^{\alpha_k - 1} e^{-Bx}, \qquad x > 0, \\ &= \exp(\Theta x - \kappa_k(\Theta))\, \rho_k(dx), \end{aligned}$$

where

$$\kappa_k(\theta) = -\alpha_k \log(-\theta), \qquad \theta < 0, \qquad \text{and} \qquad \rho_k(dx) = \frac{x^{\alpha_k - 1}}{\Gamma(\alpha_k)} \mathbf{1}_{\{x>0\}}\, dx.$$

Therefore, Proposition 3.3 applies, which implies that $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds for every $i < j \leq n$.

3.3.5 Normal distribution

Let $a_k > 0$ and $b_k \in \mathbb{R}$. Assume that, conditionally on a latent factor Θ satisfying $\mathrm{E}[|\Theta|] < \infty$, the variables $X_1, \ldots, X_n$ are independent and such that $X_k \sim \mathrm{Normal}(b_k + a_k\Theta, a_k)$. Then, the conditional probability density function is

$$f_k(x \mid \Theta) = \exp(\Theta x - \kappa_k(\Theta))\, \rho_k(dx),$$

where

$$\kappa_k(\theta) = b_k\theta + \frac{1}{2}a_k\theta^2 \qquad \text{and} \qquad \rho_k(dx) = \frac{1}{\sqrt{2\pi a_k}} \exp\left(-\frac{(x-b_k)^2}{2a_k}\right) dx.$$

By virtue of Proposition 3.3, the stochastic inequality $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds true for every $i < j \leq n$.

3.3.6 Inverse Gaussian distribution

Assume that, conditionally on a latent factor $B > 0$ satisfying $\mathrm{E}\left[\frac{1}{B}\right] < \infty$, the variables $X_1, \ldots, X_n$ are independent and follow the inverse Gaussian distribution with mean parameter $\frac{w_k}{B}$ and shape parameter w_k^2, which is denoted as $X_k \sim \mathrm{InvGauss}\left(\frac{w_k}{B}, w_k^2\right)$, where $w_k > 0$. Set $\Theta := -\frac{B^2}{2} < 0$. Then,

$$\mathrm{P}[X_k \in dx \mid \Theta = \theta] = \exp(\theta x - \kappa_k(\theta))\, \rho_k(dx), \qquad \theta < 0,$$

where

$$\kappa_k(\theta) = -w_k\sqrt{-2\theta}, \qquad \theta < 0, \qquad \text{and} \qquad \rho_k(dx) = \frac{w_k}{\sqrt{2\pi x^3}} \exp\left(-\frac{w_k^2}{2x}\right) \mathbf{1}_{\{x>0\}}\, dx.$$

Therefore, Proposition 3.3 applies, which implies that $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds for every $i < j \leq n$.

3.3.7 Tweedie distribution

Let $p \in (1, 2)$, $\alpha := \frac{2-p}{p-1} > 0$, and $c_p := \frac{(p-1)^{-\alpha}}{2-p}$. Assume that, conditionally on a latent factor $B > 0$ satisfying $\mathrm{E}\left[B^{-\frac{1}{p-1}}\right] < \infty$, the variables $X_1, \ldots, X_n$ are independent and admit the compound Poisson–gamma representation $X_k \sim \sum_{\ell=1}^{N_k} Z_{k,\ell}$, where, conditionally on B, the variables N_k are independent and such that $N_k \sim \mathrm{Poisson}(w_k c_p B^{-\alpha})$, with $w_k > 0$, while $Z_{k,\ell}$ are independent and satisfy $Z_{k,\ell} \sim \mathrm{Gamma}(\alpha, B)$, where B is the rate parameter. Set

$\Theta := -B < 0$. Conditionally on Θ, X_k thus belongs to the Tweedie natural exponential family with power parameter p and weight w_k. More precisely,

$$\mathrm{P}[X_k \in dx \mid \Theta = \theta] = \exp(\theta x - \kappa_k(\theta))\, \rho_k(dx), \qquad \theta < 0,$$

where

$$\kappa_k(\theta) = w_k c_p (-\theta)^{-\alpha}, \qquad \theta < 0, \qquad \text{and} \qquad \rho_k(dx) = \delta_0(dx) + \sum_{m=1}^{\infty} \frac{(w_k c_p)^m}{m!\,\Gamma(m\alpha)} x^{m\alpha-1} \mathbf{1}_{\{x>0\}}\, dx.$$

By virtue of Proposition 3.3, the stochastic inequality $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds true for every $i < j \leq n$.

4 Affine and proportional conditional mean predictors

This section studies situations in which the conditional mean predictor $\mathrm{E}[X_i \mid S_j]$ is an affine, or more specifically proportional, function of the conditioning sum S_j. In such cases, the convex-order monotonicity studied in the previous sections can be checked through explicit relations between the affine coefficients. This provides a complementary route to the Markov and sufficiency arguments developed in Section 3.

4.1 Independent case

Consider independent, non-negative random variables X and Y. Theorem 3.2 in Furman et al. (2018) characterizes the situation where the conditional expectation $\mathrm{E}[X|S]$ is proportional to $S = X + Y$. Precisely, denoting as $\mathcal{L}_X(t) = \mathrm{E}[\exp(-tX)]$ and $\mathcal{L}_Y(t) = \mathrm{E}[\exp(-tY)]$ the respective Laplace transforms of X and Y,

$$\mathrm{E}[X|S] = \frac{\mathrm{E}[X]}{\mathrm{E}[S]} S \Leftrightarrow t \mapsto \frac{\ln \mathcal{L}_X(t)}{\ln \mathcal{L}_Y(t)} = \frac{\mathrm{E}[X]}{\mathrm{E}[Y]} \text{ for all } t. \tag{4.1}$$

The next result extends (4.1) to real-valued random variables, provided they are independent.

Proposition 4.1. *Let X and Y be two integrable and independent real-valued random variables. Assume that there exists a C^1 function $\Psi : \mathbb{R} \to \mathbb{C}$ such that, for all $z \in \mathbb{R}$,*

$$\mathrm{E}[e^{\mathrm{i}zX}] = e^{a\Psi(z)} \qquad \textit{and} \qquad \mathrm{E}[e^{\mathrm{i}zY}] = e^{b\Psi(z)},$$

for some $a, b \geq 0$ with $a + b > 0$. Let $S := X + Y$. Then,

$$\mathrm{E}[X \mid S] = \frac{a}{a+b} S.$$

Proof. By independence, the joint characteristic function of X and Y is given by

$$\varphi(u, v) := \mathrm{E}[e^{\mathrm{i}(uX+vY)}] = e^{a\Psi(u)+b\Psi(v)},$$

which is continuously differentiable. Taking the partial derivatives gives

$$\frac{\partial}{\partial u}\varphi(u,v) = \mathrm{i}\,\mathrm{E}[Xe^{\mathrm{i}(uX+vY)}] \quad \text{and} \quad \frac{\partial}{\partial v}\varphi(u,v) = \mathrm{i}\,\mathrm{E}[Ye^{\mathrm{i}(uX+vY)}],$$

or, equivalently,

$$\frac{\partial}{\partial u}\varphi(u,v) = a\Psi'(u)e^{a\Psi(u)+b\Psi(v)} \quad \text{and} \quad \frac{\partial}{\partial v}\varphi(u,v) = b\Psi'(v)e^{a\Psi(u)+b\Psi(v)}.$$

Evaluating identities above at $u = v = z$ then yields

$$\mathrm{E}[Xe^{\mathrm{i}zS}] = \frac{a}{a+b}\,\mathrm{E}[Se^{\mathrm{i}zS}].$$

Now, let $g(S) := \mathrm{E}[X \mid S]$. Since

$$\mathrm{E}[|g(S)|] = \mathrm{E}\left[|\mathrm{E}[X \mid S]|\right] \leq \mathrm{E}[|X|] < \infty$$

and S is integrable, the random variable

$$h(S) := g(S) - \frac{a}{a+b}S$$

is integrable, too. Consider the associated finite signed measure μ on $\mathbb{R}$ defined for any Borel set $A \in \mathcal{B}(\mathbb{R})$ by $\mu(A) := \mathrm{E}\left[h(S)\mathbf{1}_{S\in A}\right]$. Its Fourier transform is given by

$$\widehat{\mu}(z) = \int_{\mathbb{R}} e^{\mathrm{i}zs}\mu(ds) = \mathrm{E}\left[h(S)e^{\mathrm{i}zS}\right].$$

Moreover,

$$\mathrm{E}[Xe^{\mathrm{i}zS}] = \mathrm{E}\left[\mathrm{E}[Xe^{\mathrm{i}zS} \mid S]\right] = \mathrm{E}\left[g(S)e^{\mathrm{i}zS}\right],$$

so that, for all $z \in \mathbb{R}$, $\widehat{\mu}(z) = 0$. The uniqueness of the Fourier transform for finite signed measures then implies that $\mu = 0$ (that is, $\mathrm{E}[h(S)\mathbf{1}_{S\in A}] = 0$ for all Borel sets A), which yields $h(S) = 0$ almost surely. This ends the proof. □

4.2 Linearity of the conditional expectation given the sum

The following result provides sufficient conditions based on the algebraic structure of the conditional expectations ensuring that a conditional expectation given a sum is a linear function of the sum.

Proposition 4.2. *Let $X_1, \ldots, X_n$ be real-valued random variables with partial sums $S_j = X_1 + \ldots + X_j$, $j \in \{1, \ldots, n\}$. Assume that, for all $1 \leq i < j \leq n$, there exist deterministic coefficients $(\alpha_{i,j}, c_{i,j})$ such that*

$$\mathrm{E}[X_i \mid S_j] = \alpha_{i,j} + c_{i,j}S_j, \tag{4.2}$$

and satisfying the compatibility relation

$$c_{i,j} = c_{i,j-1} \sum_{k=1}^{j-1} c_{k,j}. \tag{4.3}$$

Then, $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds true for all $1 \leq i < j \leq n$. Moreover, $\alpha_{i,j} = \mathrm{E}[X_i] - c_{i,j}\mathrm{E}[S_j]$. In particular, for all $1 \leq i < j \leq n$ such that $\alpha_{i,j} = 0$ and $\mathrm{E}[S_j] \neq 0$, one has $c_{i,j} = \frac{\mathrm{E}[X_i]}{\mathrm{E}[S_j]}$.

Proof. Fix $i < j \leq n$, and let $U_1 := X_i$, $U_2 := S_{j-1}$, and $W := X_j$, so that $U_2 + W = S_j$. From assumption (4.2) and the law of iterated expectations, we get

$$\mathrm{E}[U_1 \mid U_2] = \mathrm{E}[U_1] + c_{i,j-1}\big(U_2 - \mathrm{E}[U_2]\big).$$

Similarly,

$$\mathrm{E}[U_1 \mid U_2 + W] = \mathrm{E}[U_1] + c_{i,j}\big((U_2 + W) - \mathrm{E}[U_2 + W]\big).$$

Moreover,

$$\mathrm{E}[U_2 - \mathrm{E}[U_2] \mid U_2 + W] = \sum_{k=1}^{j-1} \mathrm{E}[X_k - \mathrm{E}[X_k] \mid S_j] = \left(\sum_{k=1}^{j-1} c_{k,j}\right) \left((U_2 + W) - \mathrm{E}[U_2 + W]\right).$$

With $a = c_{i,j-1}$, $b = c_{i,j}$, and $c = \sum_{k=1}^{j-1} c_{k,j}$, we thus have

$$\mathrm{E}[U_1 \mid U_2] = \mathrm{E}[U_1] + a\big(U_2 - \mathrm{E}[U_2]\big), \tag{4.4}$$

$$\mathrm{E}[U_1 \mid U_2 + W] = \mathrm{E}[U_1] + b\big((U_2 + W) - \mathrm{E}[U_2 + W]\big), \tag{4.5}$$

and

$$\mathrm{E}[U_2 \mid U_2 + W] = \mathrm{E}[U_2] + c\big((U_2 + W) - \mathrm{E}[U_2 + W]\big). \tag{4.6}$$

The compatibility condition (4.3) ensures that $b = ac$. Define the centered variables

$$\widetilde{U}_1 := U_1 - \mathrm{E}[U_1], \qquad \widetilde{U}_2 := U_2 - \mathrm{E}[U_2], \qquad \widetilde{S} := (U_2 + W) - \mathrm{E}[U_2 + W].$$

Then (4.4)–(4.6) can be rewritten as

$$\mathrm{E}[\widetilde{U}_1 \mid U_2] = a\,\widetilde{U}_2, \qquad \mathrm{E}[\widetilde{U}_1 \mid U_2 + W] = b\,\widetilde{S}, \qquad \mathrm{E}[\widetilde{U}_2 \mid U_2 + W] = c\,\widetilde{S}.$$

Hence,

$$\mathrm{E}\Big[\mathrm{E}[\widetilde{U}_1 \mid U_2] \mid U_2 + W\Big] = a\,\mathrm{E}[\widetilde{U}_2 \mid U_2 + W] = ac\,\widetilde{S}.$$

As (4.3) holds true, this yields

$$\mathrm{E}\Big[\mathrm{E}[\widetilde{U}_1 \mid U_2] \mid U_2 + W\Big] = b\,\widetilde{S} = \mathrm{E}[\widetilde{U}_1 \mid U_2 + W].$$

Applying Proposition 2.1, we obtain $\mathrm{E}[\widetilde{U}_1 \mid U_2 + W] \preceq_{\mathrm{CX}} \mathrm{E}[\widetilde{U}_1 \mid U_2]$. Since the convex order $\preceq_{\mathrm{CX}}$ is invariant under translations, this implies $\mathrm{E}[U_1 \mid U_2 + W] \preceq_{\mathrm{CX}} \mathrm{E}[U_1 \mid U_2]$. Therefore, $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds true. Finally, taking expectations in (4.2) yields $\mathrm{E}[X_i] = \alpha_{i,j} + c_{i,j}\mathrm{E}[S_j]$, which ends the proof. □

Remark 4.3. The compatibility condition (4.3) of Proposition 4.2 guarantees that the conditional expectations $(\mathrm{E}[X_i \mid S_j])_j$, viewed as orthogonal projections in L^2-spaces (see Remark 2.4), exhibit the geometric behavior required by Proposition 2.1.

Remark 4.4. It is known from formula (3.A.33) in Shaked and Shanthikumar (2007) that U is smaller than V in the Lorenz order if, and only if,

$$\frac{U}{\mathrm{E}[U]} \preceq_{\mathrm{CX}} \frac{V}{\mathrm{E}[V]}. \tag{4.7}$$

This is a convenient way to extend the convex order to the comparison of random variables with different means. Under the conditions of Proposition 4.2 with $\alpha_{i,j} = 0$, we have

$$\mathrm{E}[X_i|S_j] = \frac{\mathrm{E}[X_i]}{\mathrm{E}[S_j]} S_j \Rightarrow \mathrm{E}\left[\frac{X_i}{\mathrm{E}[X_i]}\middle|\frac{S_j}{\mathrm{E}[S_j]}\right] = \frac{S_j}{\mathrm{E}[S_j]}.$$

By Jensen's inequality, this implies that

$$\frac{S_j}{\mathrm{E}[S_j]} \preceq_{\mathrm{CX}} \frac{X_i}{\mathrm{E}[X_i]}$$

so that the sum S_j is smaller compared to each term X_i in the Lorenz order in this case.

Remark 4.5. It turns out that every distribution discussed in Section 3.3 yields a conditional expectation given the sum of the form (4.2). Hence, if $X_1, \ldots, X_n$ are as in one of the cases considered therein, Proposition 4.2 implies that $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$, $1 \leq i < j \leq n$. This therefore provides an algebraic corroboration of the observations made in Section 3.3 on the basis of Proposition 3.3. However, (4.2) is not necessarily fulfilled under the assumptions of Proposition 3.3. For instance, take $n = 3$ and let X_1, X_2, X_3 be independent conditionally on $\Theta = \theta$ with $X_k \sim \text{Binomial}\left(2, \frac{a_k e^\theta}{1+a_k e^\theta}\right)$, with $a_1 = a_2 = 1$ and $a_3 = 3$. This falls within the scope of Proposition 3.3, since

$$\kappa_k(\theta) = \ln \sum_{r=0}^{2} e^{\theta r} \binom{2}{r} a_k^r = 2\ln(1 + a_k e^\theta), \qquad \rho_k(\{r\}) = \binom{2}{r} a_k^r, \qquad r = 0, 1, 2.$$

Conditionally on $S_3 = s$, we get

$$\mathrm{P}[(X_1, X_2, X_3) = (x_1, x_2, x_3) \mid S_3 = s] \propto \prod_{k=1}^{3} \binom{2}{x_k} a_k^{x_k}, \qquad x_1 + x_2 + x_3 = s.$$

Hence

$$\mathrm{E}[X_1 \mid S_3 = 0] = 0, \qquad \mathrm{E}[X_1 \mid S_3 = 6] = 2,$$

whereas

$$\mathrm{E}[X_1 \mid S_3 = 1] = \frac{2}{2+2+6} = \frac{1}{5}.$$

If $\mathrm{E}[X_1 \mid S_3]$ were affine in S_3, the first two equalities would force $\mathrm{E}[X_1 \mid S_3 = s] = \frac{s}{3}$, contradicting the value obtained at $s = 1$. Thus, $\mathrm{E}[X_1 \mid S_3]$ is not affine in S_3 in this case.

4.3 Latent scaling in infinitely divisible convolution semigroups

Recall from Cont and Tankov (2004, Proposition 3.1) that given a Lévy process $(L_t)_{t\geq 0}$, L_t has an infinitely divisible distribution for all t. Conversely, if F is an infinitely divisible distribution then there exists a Lévy process $(L_t)_{t\geq 0}$ such that the distribution of L_1 is given by F. In other words, any infinitely divisible distribution can be seen as a distribution associated to a Lévy process at a given time (say $t = 1$).

The following result provides some explicit conditions yielding the appropriate linear structure (4.2) for the conditional expectations.

Proposition 4.6. *Let Λ be a random variable, let $X_1, \ldots, X_n$ be real-valued random variables conditionally independent given Λ, with partial sums $S_j = X_1 + \ldots + X_j$, $j \in \{1, \ldots, n\}$, and let $(L_t)_{t\geq 0}$ be a non-deterministic Lévy process such that $\mathrm{E}[|L_t|] < \infty$ for some t. Let Ψ be the characteristic exponent of $(L_t)_{t\geq 0}$, that is,*

$$\Phi_{L_t}(z) := \mathrm{E}\left[e^{\mathrm{i}z.L_t}\right] = e^{t\Psi(z)}, \qquad z \in \mathbb{R}.$$

Let $(L_t^{(1)})_{t\geq 0}, \ldots, (L_t^{(n)})_{t\geq 0}$ be n independent copies of $(L_t)_{t\geq 0}$. Assume that given Λ, $X_j \sim L^{(j)}_{b_j(\Lambda)}$ for some positive measurable function b_j with $\mathrm{E}[b_j(\Lambda)] < \infty$. For every $j \in \{1, \ldots, n\}$, define $B_j(\Lambda) := \sum_{k=1}^{j} b_k(\Lambda)$. Assume moreover that for every $1 \leq i < j \leq n$, there exists a measurable function $\varphi_{i,j}$ such that $b_i(\Lambda) = \varphi_{i,j}(B_j(\Lambda))$.

Then the following statements are equivalent:

(i) There exists $c_{i,j} \in \mathbb{R}$ such that $\mathrm{E}[X_i \mid S_j] = c_{i,j} S_j$ for every $1 \leq i < j \leq n$.

(ii) There exist constants $\alpha_1, \ldots, \alpha_n > 0$ and a random variable $U(\Lambda) > 0$ such that $b_j(\Lambda) = \alpha_j U(\Lambda)$ for every $1 \leq j \leq n$.

Moreover, if either (i) or (ii) holds, then for every $1 \leq i < j \leq n$,

$$c_{i,j} = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k} = \frac{\mathrm{E}[b_i(\Lambda)]}{\mathrm{E}[B_j(\Lambda)]}. \tag{4.8}$$

In particular, $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds true for every $1 \leq i < j \leq n$.

Proof. We first prove that $(ii) \Rightarrow (i)$. Assume that there exist constants $\alpha_1, \ldots, \alpha_n > 0$ and a random variable $U(\Lambda) > 0$ such that $b_j(\Lambda) = \alpha_j U(\Lambda)$ for all $1 \leq j \leq n$. Let us fix i and j such that $1 \leq i < j \leq n$. Conditionally on Λ, the random variables $X_1, \ldots, X_j$ are independent and satisfy $X_i \sim L_{\alpha_i U(\Lambda)}$ and $S_j - X_i \sim L_{\left(\sum_{k=1}^{j} \alpha_k - \alpha_i\right) U(\Lambda)}$ given Λ. Considering Proposition 3.13 in Cont and Tankov (2004), the fact that $\mathrm{E}[|L_t|] < \infty$ for some t is equivalent to the fact that $\int_{|x|\geq 1} |x| \nu(dx) < \infty$, where ν is the Lévy measure associated with $(L_t)_{t\geq 0}$, which in turn implies that Ψ is a C^1 function. Proposition 4.1 then applies conditionally on Λ. In particular, it yields

$$\mathrm{E}[X_i \mid S_j, \Lambda = \lambda] = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k} S_j.$$

Since the coefficient does not depend on λ, we also get

$$\mathrm{E}[X_i \mid S_j] = \mathrm{E}[\mathrm{E}[X_i \mid S_j, \Lambda] \mid S_j] = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k} S_j.$$

Hence, (i) holds with $c_{i,j} = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k}$.

We now prove that $(i) \Rightarrow (ii)$. Assume that for every $1 \leq i < j \leq n$,

$$\mathrm{E}[X_i \mid S_j] = c_{i,j} S_j. \tag{4.9}$$

In particular, for every $z \in \mathbb{R}$,

$$0 = \mathrm{E}\big[(X_i - c_{i,j} S_j)\, e^{\mathrm{i}zS_j}\big]. \tag{4.10}$$

Let us fix $1 \leq i < j \leq n$. Conditioning with respect to Λ in (4.10), we thus get

$$0 = \mathrm{E}\Big[\mathrm{E}\big[\,(X_i - c_{i,j} S_j)\, e^{\mathrm{i}zS_j}\big|\Lambda\big]\Big] = \mathrm{E}\Big[\mathrm{E}\big[\mathrm{E}[(X_i - c_{i,j} S_j)\, e^{\mathrm{i}zS_j} \mid S_j, \Lambda]\big|\Lambda\big]\Big]. \tag{4.11}$$

Applying Proposition 4.1 conditionally on $\Lambda = \lambda$, notice that

$$\mathrm{E}[X_i \mid S_j, \Lambda = \lambda] = \frac{b_i(\lambda)}{B_j(\lambda)} S_j.$$

Moreover, since

$$\mathrm{E}[e^{\mathrm{i}zS_j} \mid \Lambda = \lambda] = e^{B_j(\lambda)\Psi(z)},$$

differentiating with respect to z gives

$$\mathrm{E}[S_j e^{\mathrm{i}zS_j} \mid \Lambda = \lambda] = \frac{B_j(\lambda)\Psi'(z)}{\mathrm{i}}\, e^{B_j(\lambda)\Psi(z)}.$$

Hence, (4.11) yields

$$0 = \frac{\Psi'(z)}{\mathrm{i}}\, \mathrm{E}\big[(b_i(\Lambda) - c_{i,j} B_j(\Lambda))\, e^{B_j(\Lambda)\Psi(z)}\big], \qquad z \in \mathbb{R}. \tag{4.12}$$

By the non-degeneracy assumption on the Lévy process together with Remark A.3 in Appendix, there exists a non-trivial interval $I \subset \mathbb{R}$ such that $\Re(\Psi(z)) < 0$, $z \in I$, and Ψ is not constant on I. Since the underlying Lévy process is integrable, Ψ is C^1. Hence, there exists a non-trivial interval $J \subset I$ such that $\Psi'(z) \neq 0$, $z \in J$. Therefore, (4.12) implies that for every $z \in J$,

$$\mathrm{E}\big[\big(b_i(\Lambda) - c_{i,j} B_j(\Lambda)\big) e^{B_j(\Lambda)\Psi(z)}\big] = 0.$$

Now, consider the signed measure $\mu_{i,j}$ on $\mathbb{R}^+$, defined for any Borel set $A \in \mathcal{B}(\mathbb{R}^+)$ by

$$\mu_{i,j}(A) := \mathrm{E}\big[\big(b_i(\Lambda) - c_{i,j} B_j(\Lambda)\big) \mathbf{1}_{\{B_j(\Lambda) \in A\}}\big].$$

Observe that

$$|\mu_{i,j}|([0, +\infty[) \leq \mathrm{E}[|b_i(\Lambda) - c_{i,j} B_j(\Lambda)|] < \infty,$$

since $\mathrm{E}[b_i(\Lambda)] < \infty$ and $\mathrm{E}[B_j(\Lambda)] < \infty$. Moreover, (4.12) implies that for every $z \in J$,

$$\int_0^\infty e^{x\Psi(z)}\, \mu_{i,j}(dx) = 0.$$

By Lemma A.2 together with Remark A.3 in Appendix, we conclude that $\mu_{i,j} = 0$. Hence,

$$\mathrm{E}[b_i(\Lambda) - c_{i,j}B_j(\Lambda)|B_j(\Lambda)] = 0.$$

Since $b_i(\Lambda)$ is assumed to be $\sigma\left(B_j(\Lambda)\right)$-measurable, the above reasoning implies that $b_i(\Lambda) = c_{i,j}B_j(\Lambda)$ holds for every $1 \leq i < j \leq n$. Setting $U(\Lambda) := B_n(\Lambda)$, we get $b_i(\Lambda) = c_{i,n}U(\Lambda)$ for $i < n$, and $b_n(\Lambda) = \left(1 - \sum_{i=1}^{n-1} c_{i,n}\right) U(\Lambda)$. Hence, (ii) holds with

$$\alpha_i := c_{i,n} \quad (i < n), \qquad \alpha_n := 1 - \sum_{i=1}^{n-1} c_{i,n}.$$

To end the proof, notice that if (ii) holds then $B_j(\Lambda) = \left(\sum_{k=1}^{j} \alpha_k\right) U(\Lambda)$. In particular,

$$\frac{\mathrm{E}[b_i(\Lambda)]}{\mathrm{E}[B_j(\Lambda)]} = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k} = c_{i,j},$$

and (4.8) holds. Conversely, if (i) holds, then $b_i(\Lambda) = c_{i,j}B_j(\Lambda)$ for every $1 \leq i < j \leq n$, which implies in particular that (4.8) holds too. Now, for every $1 \leq i < j \leq n$, notice that

$$c_{i,j-1} \sum_{k=1}^{j-1} c_{k,j} = \frac{\mathrm{E}[b_i(\Lambda)]}{\mathrm{E}[B_{j-1}(\Lambda)]} \sum_{k=1}^{j-1} \frac{\mathrm{E}[b_k(\Lambda)]}{\mathrm{E}[B_j(\Lambda)]} = \frac{\mathrm{E}[b_i(\Lambda)]}{\mathrm{E}[B_j(\Lambda)]} = c_{i,j},$$

which means that (4.3) holds. Applying Proposition 4.2, we conclude that $\mathrm{E}\left[X_i \mid S_j\right] \preceq_{\mathrm{CX}} \mathrm{E}\left[X_i \mid S_{j-1}\right]$ holds for every $1 \leq i < j \leq n$, as announced. This ends the proof. □

4.4 Applications

4.4.1 Convolution semigroups

Proposition 4.6 can be interpreted in terms of convolution semigroups. Indeed, since any non-trivial infinitely divisible probability distribution μ on $\mathbb{R}$ gives rise to a Lévy process $(L_t)_{t\geq 0}$ with characteristic exponent Ψ, μ can also be associated with a convolution semigroup $(\mu_t)_{t\geq 0}$ defined by

$$\widehat{\mu_t}(z) = \mathrm{E}[e^{\mathrm{i}zL_t}] = e^{t\Psi(z)}.$$

For any positive constants $\alpha_1, \ldots, \alpha_n$ and any positive measurable random variable $U(\Lambda)$, the random variables $X_1, \ldots, X_n$ such that $X_i \sim \mu_{\alpha_i U(\Lambda)}$ given Λ, $i = 1, \ldots, n$, fall within the scope of Proposition 4.6. Hence, $\mathrm{E}[X_i \mid S_j] = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k} S_j$.

4.4.2 Poisson distribution

Assume that $X_1, \ldots, X_n$ are conditionally independent given Λ, with $X_i \sim \mathrm{Poisson}(w_i U(\Lambda))$, $1 \leq i \leq n$, for a given function U such that $U(\Lambda) > 0$. The corresponding Laplace transform is then given by

$$\mathcal{L}_{X_i|\Lambda=\lambda}(u) := \mathrm{E}\left[e^{-uX_i} \mid \Lambda = \lambda\right] = e^{-w_i U(\lambda)\psi(u)},$$

with Laplace exponent $\psi : \mathbb{R}^+ \to \mathbb{R}^+ : u \mapsto 1 - e^{-u}$. Denoting by Ψ the underlying characteristic exponent of the random vector $(X_1, \ldots, X_n)$, one has $\psi(u) = -\Psi(\mathrm{i}u)$. Applying Proposition 4.6 then yields $\mathrm{E}[X_i \mid S_j] = \frac{w_i}{\sum_{k=1}^{j} w_k} S_j$, $1 \le i < j \le n$. This expression can also be obtained from the multinomial distribution of Poisson counts given their sum.

Remark 4.7. In the Poisson case above, the Markov property $X_i \to S_{j-1} \to S_j$ (which is required to apply Corollary 3.2) does hold for $i < j$, as already observed in Section 3.3.1.

4.4.3 Normal distribution

Let $(L_t)_{t \ge 0}$ be a Brownian motion with drift, that is, $L_t = \mu t + \sigma W_t$ where $(W_t)_{t \ge 0}$ is a standard Brownian motion, $\mu \in \mathbb{R}$, and $\sigma > 0$. Its characteristic exponent is

$$\Psi(z) = \mathrm{i}\mu z - \frac{1}{2}\sigma^2 z^2.$$

Let $X_1, \ldots, X_n$ be random variables such that $X_1, \ldots, X_n$ are independent conditionally on Λ with $X_i \sim \mathrm{Normal}(\mu\alpha_i\Lambda, \sigma^2\alpha_i\Lambda)$, where $\alpha_i > 0$ for all i. Then, applying Proposition 4.6, we have $\mathrm{E}[X_i \mid S_j] = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k} S_j$, $1 \le i < j \le n$. This expression for $\mathrm{E}[X_i \mid S_j]$ can also be directly obtained in the normal case from the joint distribution of (X_i, S_j).

Remark 4.8. In the normal case above, the Markov property $X_i \to S_{j-1} \to S_j$ (required to apply Corollary 3.2) does not hold in general for $i < j$. Indeed, choose for instance $n = 3$, $\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\mu = 0$, $\sigma = 1$, and let Λ be such that $\mathrm{P}[\Lambda = 1] = \frac{1}{2} = 1 - \mathrm{P}[\Lambda = 4]$. Conditionally on $\Lambda = \lambda$, X_1, X_2 and X_3 are independent with $X_k \sim \mathrm{Normal}(0, \lambda)$. If $X_1 \to S_2 \to S_3$ were a Markov chain, then X_3 would be conditionally independent of X_1 given S_2. However, conditioning on both $S_2 = s$ and $X_1 = x$ gives information on the latent variable Λ, since it also determines $X_2 = s - x$. In particular,

$$\mathrm{P}[\Lambda = 4 \mid S_2 = s, X_1 = x] = \frac{\frac{1}{4}\exp\left(\frac{3}{8}(x^2 + (s-x)^2)\right)}{1 + \frac{1}{4}\exp\left(\frac{3}{8}(x^2 + (s-x)^2)\right)}.$$

Thus,

$$\mathrm{E}[X_3^2 \mid S_2 = s, X_1 = x] = \mathrm{E}[\Lambda \mid S_2 = s, X_1 = x]$$

depends on x, even for fixed s. Hence, given (S_2, X_1), the conditional distribution of X_3, and therefore of S_3, still depends on X_1. Thus $X_1 \to S_2 \to S_3$ is not a Markov chain in this case.

4.4.4 Gamma distribution

Let $X_1, \ldots, X_n$ be conditionally independent given Λ, such that $X_i \sim \mathrm{Gamma}(\alpha_i U(\Lambda), \tau)$, $1 \le i \le n$, for a given function U such that $U(\Lambda) > 0$. The corresponding Laplace transform is then given by

$$\mathcal{L}_{X_i \mid \Lambda = \lambda}(u) = e^{-\alpha_i U(\lambda)\psi(u)},$$

with Laplace exponent $\psi : \mathbb{R}^+ \to \mathbb{R}^+ : u \mapsto \ln\left(1 + \frac{u}{\tau}\right)$. Applying Proposition 4.6 then yields $\mathrm{E}[X_i \mid S_j] = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k} S_j$, $1 \le i < j \le n$. This expression can also be obtained from the Dirichlet distribution of Gamma measurements divided by their sum.

4.4.5 Compound Poisson distribution

Let $(L_t)_{t\geq 0}$ be a compound Poisson Lévy process defined by $L_t = \sum_{k=1}^{N_t} D_k$, where $(N_t)_{t\geq 0}$ is a Poisson process with intensity $\lambda > 0$, and $(D_k)_{k\geq 1}$ is a sequence of independent real-valued random variables, independent of $(N_t)_{t\geq 0}$, with common distribution function F_D. The characteristic exponent of $(L_t)_{t\geq 0}$ is given by

$$\Psi(z) = \lambda \left(\mathrm{E}[e^{\mathrm{i}zD_1}] - 1 \right).$$

In addition, since $\mathrm{E}[D_1^2] < \infty$, one has $\mathrm{E}|L_t| < \infty$ and $\mathrm{Var}[L_t] < \infty$ for all $t \geq 0$. Now, let $X_1, \ldots, X_n$ be conditionally independent given Λ, with $X_i \sim L_{\alpha_i U(\Lambda)}$ given Λ, $1 \leq i \leq n$, for some constants $\alpha_i > 0$ and a random variable $U(\Lambda) > 0$. More precisely, assume that conditionally on $\Lambda = \lambda$, the characteristic function of X_i is given by

$$\mathrm{E}\big[e^{\mathrm{i}zX_i} \mid \Lambda = \lambda\big] = e^{\alpha_i U(\lambda)\Psi(z)},$$

with corresponding Lévy measure $\nu_{D_i}(dx \mid \Lambda = \lambda) = \alpha_i U(\lambda)\, \lambda F_D(dx)$. Proposition 4.6 then gives $\mathrm{E}[X_i \mid S_j] = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k} S_j$, $1 \leq i < j \leq n$.

4.4.6 Tempered stable distribution

Let $(L_t)_{t\geq 0}$ be a generalized tempered stable Lévy process with Lévy triplet $(0, \nu, b)$, with $b \in \mathbb{R}$ and

$$\nu(dx) = \left(C_+ \frac{e^{-Mx}}{x^{1+Y}} \mathbf{1}_{\{x>0\}} + C_- \frac{e^{-G|x|}}{|x|^{1+Y}} \mathbf{1}_{\{x<0\}} \right) dx,$$

for some parameters $C_+, C_- \geq 0$, not both equal to 0, $G, M > 0$, and $Y \in (0, 2)$. Notice that the exponential tempering and the choice $Y < 2$ imply that

$$\int_{|x|>1} |x|\, \nu(dx) < \infty \qquad \text{and} \qquad \int_{\mathbb{R}\setminus\{0\}} x^2\, \nu(dx) < \infty.$$

Therefore, $\mathrm{E}[|L_t|] < \infty$ and $\mathrm{Var}[L_t] < \infty$ for every $t \geq 0$.

Let $X_1, \ldots, X_n$ be conditionally independent given Λ, with $X_i \sim L_{\alpha_i U(\Lambda)}$, $1 \leq i \leq n$, for some constants $\alpha_i > 0$ and a random variable $U(\Lambda) > 0$. More precisely, assume that conditionally on $\Lambda = \lambda$, conditional Lévy triplet of X_i is given by $(0, \alpha_i U(\lambda)\nu, \alpha_i U(\lambda) b)$. Proposition 4.6 then gives $\mathrm{E}[X_i \mid S_j] = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k} S_j$, $1 \leq i < j \leq n$.

4.4.7 Beyond Proposition 4.6

As we have seen, Proposition 4.6 provides us with sufficient conditions for obtaining an expression of the form (4.2) with zero $\alpha_{i,j}$-coefficients, guaranteeing the desired convex order (2.4), when the random variables under consideration belong to the same convolution semigroup associated with a given Lévy process. In this section, we explore situations that fall outside of the scope of Proposition 4.6, but where the convex order (2.4) is nevertheless still satisfied.

Normal distribution Let us start with a Gaussian example where the conditional expectation given the sum is strictly affine with respect to the sum.

Example 4.9. *Let $\Lambda \sim \mathrm{Normal}(\mu, \tau^2)$ be a latent random variable, and let $\mathcal{E}_1, \dots, \mathcal{E}_n$ be independent centered Gaussian random variables such that $\mathcal{E}_k \sim \mathrm{Normal}(0, \sigma_k^2)$ for every $1 \le k \le n$, independent of Λ. Let $d_1, \dots, d_n \in \mathbb{R}$, and define*

$$Z_k := d_k \Lambda + \mathcal{E}_k, \qquad 1 \le k \le n.$$

Here, each Z_k can be considered to be a signal for the underlying Λ.

For $j \le n$, set $T_j := \sum_{k=1}^{j} Z_k$, $D_j := \sum_{k=1}^{j} d_k$, and $V_j := \sum_{k=1}^{j} \sigma_k^2$. Then, $T_j = D_j \Lambda + \sum_{k=1}^{j} \mathcal{E}_k$, so that (Z_i, T_j) is jointly Gaussian for every $i \le j$. Therefore,

$$\mathrm{E}[Z_i \mid T_j] = \alpha_{i,j} + c_{i,j} T_j \text{ where } c_{i,j} = \frac{\mathrm{Cov}[Z_i, T_j]}{\mathrm{Var}[T_j]} = \frac{d_i D_j \tau^2 + \sigma_i^2}{D_j^2 \tau^2 + V_j}, \tag{4.13}$$

and $\alpha_{i,j} = \mathrm{E}[Z_i] - c_{i,j}\mathrm{E}[T_j]$. Since $\mathrm{E}[Z_i] = d_i \mu$ and $\mathrm{E}[T_j] = D_j \mu$, we get

$$\alpha_{i,j} = \mu \frac{d_i V_j - D_j \sigma_i^2}{D_j^2 \tau^2 + V_j}.$$

In particular, when $\mu \neq 0$, the conditional expectations $\mathrm{E}[Z_i \mid T_j]$ are strictly affine with respect to T_j.

Remark 4.10. In the setting of Example 4.9, notice that $Z_i \sim \mathrm{Normal}(d_i \Lambda, \sigma_i^2)$ given Λ, so that the conditional variance does not depend on λ. Therefore, except in degenerate cases, the conditional law of Z_i cannot be written in the form $L_{b_i(\Lambda)}$ for some Lévy process $(L_t)_{t \ge 0}$. Example 4.9 therefore does not fall within the scope of Proposition 4.6.

In order to ensure the desired convex order (2.4), the coefficients appearing in (4.13) must also satisfy the compatibility relation (4.3), which is not true in general. The next result characterizes precisely when convex order still holds in the Gaussian setting.

Property 4.11. *In the framework of Example 4.9, let $i \le j - 1$. Then,*

$$\mathrm{E}[Z_i \mid T_j] \preceq_{\mathrm{CX}} \mathrm{E}[Z_i \mid T_{j-1}] \Leftrightarrow \frac{(d_i D_j \tau^2 + \sigma_i^2)^2}{D_j^2 \tau^2 + V_j} \le \frac{(d_i D_{j-1} \tau^2 + \sigma_i^2)^2}{D_{j-1}^2 \tau^2 + V_{j-1}}.$$

Proof. Both $\mathrm{E}[Z_i \mid T_{j-1}]$ and $\mathrm{E}[Z_i \mid T_j]$ are Gaussian random variables with the same mean. The convex order reduces to the comparison of the respective variances in this case. Hence, it suffices to compare the variances of $\mathrm{E}[Z_i \mid T_{j-1}]$ and $\mathrm{E}[Z_i \mid T_j]$. Note that

$$\mathrm{Var}[T_{j-1}] = D_{j-1}^2 \tau^2 + V_{j-1}, \qquad \mathrm{Cov}[Z_i, T_{j-1}] = d_i D_{j-1} \tau^2 + \sigma_i^2.$$

Therefore,

$$\mathrm{Var}\big[\mathrm{E}[Z_i \mid T_{j-1}]\big] = \frac{(\mathrm{Cov}[Z_i, T_{j-1}])^2}{\mathrm{Var}[T_{j-1}]} = \frac{(d_i D_{j-1} \tau^2 + \sigma_i^2)^2}{D_{j-1}^2 \tau^2 + V_{j-1}}.$$

Similarly,

$$\mathrm{Var}\big[\mathrm{E}[Z_i \mid T_j]\big] = \frac{(d_i D_j \tau^2 + \sigma_i^2)^2}{D_j^2 \tau^2 + V_j}.$$

This proves the announced result. □

In the context of Example 4.9, notice that if there exists j_0 such that $d_{j_0} = 0$ (that is, such that $Z_{j_0} = \mathcal{E}_{j_0}$ is independent of Z_i for all $i < j_0$), then $D_{j_0} = D_{j_0-1}$, while

$$V_{j_0} = V_{j_0-1} + \sigma_{j_0}^2 \geq V_{j_0-1}.$$

Hence, for $i \leq j_0 - 1$,

$$\frac{(d_i D_{j_0}\tau^2 + \sigma_i^2)^2}{D_{j_0}^2\tau^2 + V_{j_0}} \leq \frac{(d_i D_{j_0-1}\tau^2 + \sigma_i^2)^2}{D_{j_0-1}^2\tau^2 + V_{j_0-1}}.$$

By virtue of Property 4.11, the above inequality implies that $\mathrm{E}[Z_i \mid T_{j_0}] \preceq_{\mathrm{CX}} \mathrm{E}[Z_i \mid T_{j_0-1}]$. Moreover, if $d_j = 0$ for all $j \geq j_0$, the same conclusion holds for every $j \geq j_0$. This observation is of course a consequence of Remark 2.5 about the addition of independent noise.

The following example highlights the role played by the order of arrival of the signal components in the convex ordering, within the conditionally independent Gaussian framework of Example 4.9.

Example 4.12. *Consider the Gaussian affine framework of Example 4.9 with $m = 4$, and assume for simplicity that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$. More precisely, suppose that*

$$Z_k := d_k\Lambda + \mathcal{E}_k, \qquad 1 \leq k \leq 4,$$

where $\Lambda \sim \mathrm{Normal}(\mu, \tau^2)$ is a latent random variable, and where $\mathcal{E}_k \sim \mathrm{Normal}(0, 1)$ for every $1 \leq k \leq 4$. Let us now consider two different patterns for the coefficients d_1, d_2, d_3, d_4.

Case 1: $(d_1, d_2, d_3, d_4) = (1, 1, 0, 0)$. Then $D_2 = D_3 = D_4 = 2$, while $V_2 = 2$, $V_3 = 3$, $V_4 = 4$. For $i \leq j - 1$,

$$Q_{i,j} := \frac{(d_i D_j\tau^2 + \sigma_i^2)^2}{D_j^2\tau^2 + V_j}$$

has constant numerator for $j \geq 2$, while the denominator is increasing in j. Hence, $Q_{i,j} \leq Q_{i,j-1}$, $j = 3, 4$, so that $\mathrm{E}[Z_i \mid T_j] \preceq_{\mathrm{CX}} \mathrm{E}[Z_i \mid T_{j-1}]$ holds true by Property 4.11, for all admissible i, j.

Case 2: $(d_1, d_2, d_3, d_4) = (1, 0, 0, 1)$. Then $D_2 = D_3 = 1$ and $D_4 = 2$, while $V_2 = 2$, $V_3 = 3$, $V_4 = 4$. As above, $Q_{i,3} \leq Q_{i,2}$ for all $i \leq 2$. However, for $i = 1$,

$$Q_{1,3} = \frac{(\tau^2 + 1)^2}{\tau^2 + 3}, \qquad Q_{1,4} = \frac{(2\tau^2 + 1)^2}{4(\tau^2 + 1)}.$$

Hence, a direct computation shows that $Q_{1,4} > Q_{1,3}$ as soon as $\tau > \sqrt{\frac{\sqrt{17}-1}{8}}$. For such a value of τ, Property 4.11 then implies that $\mathrm{E}[Z_1 \mid T_4]$ and $\mathrm{E}[Z_1 \mid T_3]$ cannot be ordered in the $\preceq_{\mathrm{CX}}$-sense.

Example 4.12 thus shows that the desired convex order of the sequence may not be satisfied, depending on when the components of the signal of interest reappear (even though the conditional expectations remain affine).

Poisson distribution The next example considers a situation where conditional expectations are not all proportional to sums, in the conditionally independent Poisson case. It thus falls outside of the scope of Proposition 4.6. The convex-order reduction is nevertheless still present because of the Markov structure.

Example 4.13. *Let* $\Lambda \sim \text{Gamma}(2,1)$*, and assume that given* $\Lambda = \lambda$*,* $Z_1 \sim \text{Poisson}(\lambda)$*,* $Z_2 \sim \text{Poisson}(2\lambda)$*,* $Z_3 \sim \text{Poisson}(3\lambda)$*, and* $Z_4 \sim \text{Poisson}(\lambda^2)$*, with* Z_1*,* Z_2*,* Z_3*, and* Z_4 *conditionally independent given* Λ*. For* $j = 2,3,4$*, set* $T_j := Z_1 + \cdots + Z_j$*. The conditional expectations are given by*

$$\mathrm{E}[Z_1 \mid T_2] = \frac{T_2}{3}, \qquad \mathrm{E}[Z_2 \mid T_2] = \frac{2T_2}{3},$$

$$\mathrm{E}[Z_1 \mid T_3] = \frac{T_3}{6}, \qquad \mathrm{E}[Z_2 \mid T_3] = \frac{T_3}{3}, \qquad \mathrm{E}[Z_3 \mid T_3] = \frac{T_3}{2},$$

and, for every $t \in \mathbb{N}$*,*

$$\mathrm{E}[Z_1 \mid T_4 = t] = t\, r_t, \qquad \mathrm{E}[Z_2 \mid T_4 = t] = 2t\, r_t, \qquad \mathrm{E}[Z_3 \mid T_4 = t] = 3t\, r_t,$$

where

$$r_t := \mathrm{E}\left[\frac{1}{6+\Lambda} \;\middle|\; T_4 = t\right] = \frac{\displaystyle\int_0^\infty \lambda^{t+1}(6+\lambda)^{t-1} e^{-(7\lambda+\lambda^2)}\, d\lambda}{\displaystyle\int_0^\infty \lambda^{t+1}(6+\lambda)^{t} e^{-(7\lambda+\lambda^2)}\, d\lambda}.$$

Equivalently,

$$\mathrm{E}[Z_1 \mid T_4] = T_4\, \mathrm{E}\left[\frac{1}{6+\Lambda} \;\middle|\; T_4\right], \qquad \mathrm{E}[Z_2 \mid T_4] = 2T_4\, \mathrm{E}\left[\frac{1}{6+\Lambda} \;\middle|\; T_4\right],$$

and

$$\mathrm{E}[Z_3 \mid T_4] = 3T_4\, \mathrm{E}\left[\frac{1}{6+\Lambda} \;\middle|\; T_4\right].$$

For $j = 2$*, since* $T_1 = Z_1$*, one has* $\mathrm{E}[Z_1 \mid T_2] \preceq_{\mathrm{CX}} \mathrm{E}[Z_1 \mid T_1] = Z_1$*. For* $j = 3$*, the vector* (Z_1, Z_2) *is conditionally independent given* $\Lambda = \lambda$*, with intensities proportional to* λ*. Hence, using the same argument as in Proposition 3.3 with* $Z_i \to T_2 \to T_3$*,* $i = 1,2$*, we conclude that* $\mathrm{E}[Z_i \mid T_3] \preceq_{\mathrm{CX}} \mathrm{E}[Z_i \mid T_2]$*,* $i = 1,2$*. For* $j = 4$*, the vector* (Z_1, Z_2, Z_3) *is conditionally independent given* $\Lambda = \lambda$*, with intensities proportional to* λ*. Hence, using again the same argument as in Proposition 3.3 with* $Z_i \to T_3 \to T_4$*,* $i = 1,2,3$*, we get* $\mathrm{E}[Z_i \mid T_4] \preceq_{\mathrm{CX}} \mathrm{E}[Z_i \mid T_3]$*,* $i = 1,2,3$*. This shows that* $\mathrm{E}[Z_i \mid T_j] \preceq_{\mathrm{CX}} \mathrm{E}[Z_i \mid T_{j-1}]$ *holds for every* $i < j$*.*

The following example illustrates another situation which falls outside of the scope of Proposition 4.6.

Example 4.14. *Let* U *be a positive latent random variable, and assume that* Z_1*,* Z_2*, and* Z_3 *are independent conditionally on* U*, with* $Z_1 \sim \text{Poisson}(\alpha_1 U)$*,* $Z_2 \sim \text{Poisson}(\alpha_2 U)$*, and* $Z_3 \sim \text{Poisson}(\alpha_3 U)$ *given* U*, where* $\alpha_1, \alpha_2, \alpha_3 > 0$*. Also, consider* $Z_4 \sim \text{Poisson}(\lambda)$*,* $\lambda > 0$*,*

independent of (U, Z_1, Z_2, Z_3). *Set* $A_2 := \alpha_1 + \alpha_2$ *and* $A_3 := \alpha_1 + \alpha_2 + \alpha_3$. *Denote the partial sums as* $T_j = Z_1 + \ldots + Z_j$, $j = 1, \ldots, 4$.

Since the Poisson laws form a convolution semigroup, the random vector (Z_1, Z_2, Z_3) *falls within the scope of Proposition 4.6. Hence,*

$$\mathrm{E}[Z_1 \mid T_2] = \frac{\alpha_1}{A_2} T_2, \;\; \mathrm{E}[Z_1 \mid T_3] = \frac{\alpha_1}{A_3} T_3 \text{ and } \mathrm{E}[Z_1 \mid T_3] \preceq_{\mathrm{CX}} \mathrm{E}[Z_1 \mid T_2].$$

Moreover, since Z_4 *is independent of* (Z_1, T_3),

$$\mathrm{E}[Z_1 \mid T_3, Z_4] = \mathrm{E}[Z_1 \mid T_3] \Rightarrow \mathrm{E}[Z_1 \mid T_4] \preceq_{\mathrm{CX}} \mathrm{E}[Z_1 \mid T_3]$$

by Example 2.3. Hence, the sequence $\left(\mathrm{E}[Z_1 \mid T_j]\right)_{j=2,3,4}$ *is non-increasing with respect to the convex order.*

Remark 4.15. In Example 4.14, observe that unlike (Z_1, Z_2, Z_3), the vector (Z_1, Z_2, Z_3, Z_4) does not fall within the scope of Proposition 4.6. In particular, there is then no guarantee that $\mathrm{E}[Z_1 \mid T_4]$ is linear with respect to T_4. Indeed, the fact that $T_3 \sim \mathrm{Poisson}(A_3 U)$ given U and $Z_4 \sim \mathrm{Poisson}(\lambda)$ yields

$$\mathrm{E}[T_3 \mid T_4, U] = T_4 \frac{A_3 U}{A_3 U + \lambda}.$$

Moreover, using the fact that $\mathrm{E}[Z_1 \mid T_3, U] = \frac{\alpha_1}{A_3} T_3$, we get

$$\mathrm{E}[Z_1 \mid T_4] = \alpha_1 T_4 \mathrm{E}\left[\frac{U}{A_3 U + \lambda} \,\middle|\, T_4 \right].$$

Choosing for instance U such that $\mathrm{P}[U = u_1] = p = 1 - \mathrm{P}[U = u_2]$ yields

$$\mathrm{E}[Z_1 \mid T_4 = t] = \alpha_1 t \left(\frac{u_1}{A_3 u_1 + \lambda} \pi_t + \frac{u_2}{A_3 u_2 + \lambda} (1 - \pi_t) \right),$$

where

$$\pi_t := \mathrm{P}[U = u_1 \mid T_4 = t] = \frac{p\, e^{-(A_3 u_1 + \lambda)} (A_3 u_1 + \lambda)^t}{(1-p) e^{-(A_3 u_2 + \lambda)} (A_3 u_2 + \lambda)^t + p\, e^{-(A_3 u_1 + \lambda)} (A_3 u_1 + \lambda)^t}.$$

In such a case, the map $t \mapsto \mathrm{E}[Z_1 \mid T_4 = t]$ is thus nonlinear in t.

As a consequence of both Example 2.3 and Proposition 4.6, Example 4.14 thus shows that a sequence of conditional expectations may first satisfy the linear structure of Proposition 4.6, then become nonlinear after the addition of a pure noise, while still preserving the desired convex ordering.

Example 4.16. *Consider the same variables as in Example 4.14, but change their order. Let* $\widetilde{Z}_1, \widetilde{Z}_2, \widetilde{Z}_4$ *be conditionally independent given* U, *with*

$$\widetilde{Z}_1 \sim \mathrm{Poisson}(\alpha_1 U), \qquad \widetilde{Z}_2 \sim \mathrm{Poisson}(\alpha_2 U), \qquad \widetilde{Z}_4 \sim \mathrm{Poisson}(\alpha_3 U),$$

and let

$$\widetilde{Z}_3 \sim \text{Poisson}(\lambda)$$

be independent of $(U, \widetilde{Z}_1, \widetilde{Z}_2, \widetilde{Z}_4)$*. Define*

$$\widetilde{T}_j = \widetilde{Z}_1 + \cdots + \widetilde{Z}_j, \qquad j = 1, \ldots, 4.$$

Since $\widetilde{Z}_3$ *is independent of* $(U, \widetilde{Z}_1, \widetilde{Z}_2)$*, Example 2.3 gives*

$$\mathrm{E}[\widetilde{Z}_1 \mid \widetilde{T}_3] \preceq_{\mathrm{CX}} \mathrm{E}[\widetilde{Z}_1 \mid \widetilde{T}_2].$$

However, when moving from $\widetilde{T}_3$ *to* $\widetilde{T}_4$*, the added variable* $\widetilde{Z}_4$ *carries information about the latent factor* U*. Conditionally on* $U = u$*,* $\widetilde{Z}_1$*,* $\widetilde{Z}_2$*, and* $\widetilde{Z}_3$ *are independent Poisson variables with parameters* $\alpha_1 u, \alpha_2 u, \lambda$*. Hence*

$$\widetilde{Z}_1 \mid \widetilde{T}_3 = t, U = u \sim \text{Binomial}\left(t, \frac{\alpha_1 u}{A_2 u + \lambda}\right),$$

and therefore

$$\mathrm{E}[\widetilde{Z}_1 \mid \widetilde{T}_3] = \alpha_1 \widetilde{T}_3 \mathrm{E}\left[\frac{U}{A_2 U + \lambda} \,\middle|\, \widetilde{T}_3\right].$$

Similarly,

$$\mathrm{E}[\widetilde{Z}_1 \mid \widetilde{T}_4] = \alpha_1 \widetilde{T}_4 \mathrm{E}\left[\frac{U}{A_3 U + \lambda} \,\middle|\, \widetilde{T}_4\right].$$

Thus adding $\widetilde{Z}_4$ *can increase the information about* U*, and there is no general reason for the desired convex-order monotonicity to hold. For instance, take*

$$\alpha_1 = \alpha_2 = \alpha_3 = 1, \qquad \lambda = 2, \qquad \mathrm{P}[U = 1] = \mathrm{P}[U = 4] = \frac{1}{2}.$$

Set

$$M_3 := \mathrm{E}[\widetilde{Z}_1 \mid \widetilde{T}_3], \qquad M_4 := \mathrm{E}[\widetilde{Z}_1 \mid \widetilde{T}_4].$$

A direct numerical computation gives

$$\mathrm{Var}[M_3] \approx 3.09, \qquad \mathrm{Var}[M_4] \approx 2.89.$$

This variance comparison alone does not disprove the convex-order relation $M_4 \preceq_{\mathrm{CX}} M_3$*. However, the stop-loss transforms cross. Indeed,*

$$\mathrm{E}[(M_4 - 1.84)_+] \approx 1.0779 > 1.0712 \approx \mathrm{E}[(M_3 - 1.84)_+],$$

whereas

$$\mathrm{E}[(M_4 - 4.28)_+] \approx 0.1517 < 0.1898 \approx \mathrm{E}[(M_3 - 4.28)_+].$$

Thus, neither $M_4 \preceq_{\mathrm{CX}} M_3$ *nor* $M_3 \preceq_{\mathrm{CX}} M_4$ *holds in this example. The two conditional mean predictors are not comparable in the convex order.*

5 Conclusion

This paper has studied the convex-order monotonicity of conditional mean predictors based on aggregate statistics. For partial sums $S_j = X_1 + \cdots + X_j$, $j = 1, \ldots, n$, we have investigated conditions under which the stochastic inequality $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ is valid, with $i < j$. Such an ordering means that, as the conditioning sum incorporates additional components, the corresponding conditional mean predictor becomes less dispersed. In the statistical interpretation adopted in this paper, this provides a distributional measure of information loss: the aggregate statistic S_j is less informative about the individual component X_i than the smaller statistic S_{j-1}.

The first contribution of the paper is a general projection criterion for convex-order reduction. We showed that the identity $\mathrm{E}[U_1 \mid U_2 + W] = \mathrm{E}\big[\mathrm{E}[U_1 \mid U_2] \mid U_2 + W\big]$ is sufficient to imply $\mathrm{E}[U_1 \mid U_2 + W] \preceq_{\mathrm{CX}} \mathrm{E}[U_1 \mid U_2]$. This criterion includes the classical case where an independent noise term is added to a signal. It therefore strengthens variance-based comparisons by showing that the whole distribution of the conditional mean predictor is reduced in the convex order.

The second contribution is to identify a Markov mechanism leading to this projection identity. If $X_i \to S_{j-1} \to S_j$ is a Markov chain, then the desired convex-order monotonicity follows immediately. We proved that this Markov structure arises naturally in latent mixture models based on natural exponential families with a common canonical parameter. In such models, the variables are conditionally independent given a common latent factor, and the partial sum S_{j-1} is sufficient for this factor from the observation of $(X_1, \ldots, X_{j-1})$. The sufficiency of partial sums is therefore the key statistical mechanism behind the convex-order reduction. This result covers a broad range of standard models, including Poisson, binomial, negative binomial, gamma, normal, inverse Gaussian and Tweedie distributions.

The third contribution is an algebraic analysis of affine and proportional conditional mean predictors. We derived a compatibility condition on the affine coefficients ensuring that $\mathrm{E}[X_i \mid S_j] \preceq_{\mathrm{CX}} \mathrm{E}[X_i \mid S_{j-1}]$ holds true. This provides an alternative route to convex-order monotonicity when $\mathrm{E}[X_i \mid S_j]$ is an affine function of S_j. We then specialized this analysis to convolution semigroups associated with infinitely divisible distributions. In latent scaling models of the form $X_i \mid \Lambda \sim L_{\alpha_i U(\Lambda)}$, where L is a common Lévy process, the conditional mean predictor is proportional to the conditioning sum:

$$\mathrm{E}[X_i \mid S_j] = \frac{\alpha_i}{\sum_{k=1}^{j} \alpha_k} S_j.$$

This gives explicit convex-order monotonicity results for Poisson, gamma, normal, compound Poisson and tempered stable models, among others.

The examples developed in the paper also show that the different mechanisms leading to convex-order monotonicity should not be confused. In some models, the desired ordering follows from a Markov and sufficiency structure. In others, it follows from an affine or proportional convolution-semigroup structure, even when the Markov property fails. Conversely, when a new component added to the aggregate statistic carries additional information about the latent factor, convex-order monotonicity may fail. Thus, the order in which signal-bearing and noise components enter the aggregate statistic can be decisive.

The results have a natural interpretation in statistical prediction and imputation. If an individual component is unobserved and must be predicted from an aggregate response, the conditional expectation is the optimal predictor under squared-error loss. The convex-order comparisons obtained in this paper quantify how the informativeness of this predictor is affected by aggregation, noise accumulation and latent dependence. This perspective is relevant, for example, in regression models with unobserved covariates, common-factor models, risk allocation, signal extraction and random effects models.

Several extensions could be considered as future research avenues. A first direction would be to replace scalar sums by more general aggregate statistics, including vector-valued statistics or linear transformations. A second direction would be to study estimation procedures for the conditional mean predictors and to investigate whether the convex-order monotonicity is preserved under plug-in or non-parametric estimation. A third direction would be to extend the analysis to other stochastic order relations. See, e.g., Ganuza and Penalva (2010) for comparisons of conditional expectations given signals in the dispersive order. These questions would further connect stochastic-order theory with statistical inference under aggregation and latent dependence.

Acknowledgements

Valentin Dendoncker and Michel Denuit gratefully acknowledge funding from the FWO and F.R.S.-FNRS under the Excellence of Science (EOS) programme, project ASTeRISK (40007517).

References

- Bar-Lev, S.K., Schulte-Geers, E., Stadje, W. (2013). Conditional limit theorems for the terms of a random walk revisited. Journal of Applied Probability 50, 871-882.

- Cont, R., Tankov, P. (2004). Financial Modelling with Jump Processes. Chapman & Hall/CRC Financial Mathematics Series.

- Denuit, M. (2010). Positive dependence of signals. Journal of Applied Probability 47, 893-897.

- Denuit, M. (2019). Size-biased transform and conditional mean risk sharing, with application to P2P insurance and tontines. ASTIN Bulletin 49, 591-617.

- Denuit, M., Dhaene, J. (2012). Convex order and comonotonic conditional mean risk sharing. Insurance: Mathematics and Economics 51, 265-270.

- Denuit, M., Lefevre, Cl., Mesfioui, M. (1999). A class of bivariate stochastic orderings with applications in actuarial sciences. Insurance: Mathematics and Economics 24, 31-50.

- Denuit, M., Robert, C.Y. (2020). Large-loss behavior of conditional mean risk sharing. ASTIN Bulletin 50, 1093-1122.

- Denuit, M., Robert, C.Y. (2021). Efron's asymptotic monotonicity property in the Gaussian stable domain of attraction. Journal of Multivariate Analysis 186, Article 104803.

- Denuit, M., Robert, C.Y. (2022). Conditional tail expectation decomposition and conditional mean risk sharing for dependent and conditionally independent risks. Methodology and Computing in Applied Probability 24, 1953-1985.

- Denuit, M., Robert, C.Y. (2023). From risk reduction to risk elimination by conditional mean risk sharing of independent losses. Insurance: Mathematics and Economics 108, 46-59.

- Ernst, P. A., Kagan, A. M., Rogers, L.C.G. (2022). The least favorable noise. Electronic Communications in Probability 27, Article 28.

- Furman, E., Kuznetsov, A., Zitikis, R. (2018). Weighted risk capital allocations in the presence of systematic risk. Insurance: Mathematics and Economics 79, 75-81.

- Ganuza, J.-J., Penalva, J.S. (2010). Signal orderings based on dispersion and the supply of private information in auctions. Econometrica 78, 1007-1030.

- Huang, D. (2023). The existence of the least favorable noise. Electronic Communications in Probability 28, Article 25.

- Leskelä, L., Vihola, M. (2017). Conditional convex orders and measurable martingale couplings. Bernoulli 23, 2784–2807.

- Mizuno, T. (2006). A relation between positive dependence of signal and variability of conditional expectation given signal. Journal of Applied Probability 43, 1181-1185.

- Müller, A., Stoyan, D. (2002). Comparison Methods for Stochastic Models and Risks. Wiley.

- Shaked, M., Sordo, M. A., Suarez-Llorens, A. (2012). Global dependence stochastic orders. Methodology and Computing in Applied Probability 14, 617-648.

- Shaked, M., Shanthikumar, J.G. (2007). Stochastic Orders. Springer, New York.

- Van Bochove, C.A. (2011). Expectation of a uniform random variable with uniform observation errors after selection of the highest observations. Statistical Papers 52, 971-977.

- Zabell, S.L. (1980). Rates of convergence for conditional expectations. The Annals of Probability 8, 928-941.

- Zabell, S.L. (1993). A limit theorem for expectations conditional on a sum. Journal of Theoretical Probability 6, 267-283.

Appendix

A Technical material for Section 4

Definition A.1. *Let $\mathcal{M}$ be a class of measures on a subset $E \subset \mathbb{R}$, and let $\{f_\theta\}_{\theta\in\Theta}$ be a family of $\mathbb{C}$-valued measurable functions, integrable with respect to every $\mu \in \mathcal{M}$. For $\mu \in \mathcal{M}$, let $T_\theta(\mu)$ be defined by*

$$T_\theta(\mu) := \int_E f_\theta(x)\mu(dx).$$

The family of functions $\{f_\theta\}_{\theta\in\Theta}$ is a separating family for the class of measures $\mathcal{M}$ if

$$T_\theta(\mu) = T_\theta(\tilde{\mu}),\ \forall\theta \in \Theta \Rightarrow \mu = \tilde{\mu}.$$

Lemma A.2. *Let $\Psi : \mathbb{R} \to \mathbb{C}$ be a continuous function. Assume that there exists a non-trivial interval $I \subset \mathbb{R}$ such that $\Re(\Psi(z)) < 0$, $z \in I$, and that Ψ is not constant on I. Then the family of functions*

$$\mathcal{F}_\Psi := \{f_z : \mathbb{R}_+ \to \mathbb{C},\ x \mapsto e^{x\Psi(z)}\}_{z\in\mathbb{R}}$$

separates finite signed measures on $\mathbb{R}_+$. More precisely, if μ is a finite signed measure on $[0,\infty[$ such that

$$\int_0^\infty e^{x\Psi(z)}\,\mu(dx) = 0, \qquad \forall z \in I,$$

then $\mu = 0$.

Proof. Let μ be a finite signed measure on $[0,\infty[$ such that

$$\int_0^\infty e^{x\Psi(z)}\,\mu(dx) = 0, \qquad \forall z \in I.$$

Define its Laplace transform by

$$\mathcal{L}_\mu(s) := \int_0^\infty e^{-sx}\,\mu(dx), \qquad \Re(s) > 0.$$

Since μ is finite, $\mathcal{L}_\mu$ is well-defined and holomorphic on the open half-plane $\{\Re(s) > 0\}$ by dominated convergence, since $|e^{-sx}| \le 1$ for $\Re(s) > 0$. Note now that for every $z \in I$, the condition $\Re(\Psi(z)) < 0$ is equivalent to $\Re(-\Psi(z)) > 0$, so that $-\Psi(z)$ belongs to the domain of holomorphy of $\mathcal{L}_\mu$. Moreover,

$$\mathcal{L}_\mu(-\Psi(z)) = \int_0^\infty e^{x\Psi(z)}\,\mu(dx) = 0, \qquad \forall z \in I.$$

Thus, $\mathcal{L}_\mu$ vanishes on the set $-\Psi(I)$, which has an accumulation point in $\{\Re(s) > 0\}$ (since Ψ is continuous and non-constant on a non-trivial interval I). By the identity theorem for holomorphic functions, $\mathcal{L}_\mu \equiv 0$ on $\{\Re(s) > 0\}$. The uniqueness of the Laplace transform for finite signed measures on $[0,\infty[$ then implies that $\mu = 0$. This ends the proof. □

Remark A.3. The assumptions in Lemma A.2 become very natural when Ψ is the characteristic exponent of an integrable real-valued Lévy process. Indeed, denoting by $(L_t)_{t\geq 0}$ such a Lévy process with characteristic triplet (A, ν, γ) and assuming that $\mathrm{E}|L_t| < \infty$ for some $t > 0$, then

$$\Psi(z) = -\frac{1}{2}Az^2 + i\gamma z + \int_{\mathbb{R}} \left(e^{\mathrm{i}zx} - 1 - \mathrm{i}zx\,\mathbf{1}_{\{|x|\leq 1\}}\right) \nu(dx).$$

Taking real parts yields

$$\Re(\Psi(z)) = -\frac{1}{2}Az^2 - \int_{\mathbb{R}} (1 - \cos(zx))\,\nu(dx) \leq 0, \qquad z \in \mathbb{R}.$$

Moreover, if $(L_t)_{t\geq 0}$ is non-degenerate (that is, if $A > 0$ or $\nu \neq 0$), then $\Re(\Psi)$ is not identically zero and there exists a non-trivial interval $I \subset \mathbb{R}$ such that $\Re(\Psi(z)) < 0$ for all $z \in I$. Hence, Lemma A.2 applies to the family $\{x \mapsto f_z(x) := e^{x\Psi(z)}\}_{z\in\mathbb{R}}$.