Journal of the American Statistica

ISSN: (Print) (Onlinwww.Jtoaunrdnfaoln lhionmee.pcaogme/:journals/uasa20

Generalized data thinning usin statistics

Ameer Dharamshi, Anna Neufeld, Keshav Motwani Witten & Jacob Bien

Generalized data thinning using sufficient statistics

Ameer Dharamshi[a,*,#], Anna Neufeld[b], Keshav Motwani[a], Lucy L. Gao[c], Daniela Witten[d] and Jacob Bien[e]

[a]Department of Biostatistics, University of Washington
[b]Public Health Sciences Division, Fred Hutchinson Cancer Research Center
[c]Department of Statistics, University of British Columbia
[d]Departments of Biostatistics and Statistics, University of Washington
[e]Department of Data Sciences and Operations, University of Southern, California

*The authors gratefully acknowledge funding from *NIH R01 EB026908, NIH R01 DA047869, ONR N00014-23-1-2589, NSF DMS 2322920, a Simons Investigator Award in Mathematical Modeling of Living Systems, and the Keck Foundation to DW; NIH R01 GM123993 to DW and JB; a Natural Sciences and Engineering Research Council of Canada Discovery Grant to LG; and a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship-Doctoral to AD.*

#adharams@uw.edu

Abstract

Our goal is to develop a general strategy to decompose a random variable *X* into multiple independent random variables, without sacrificing any information about unknown parameters. A recent paper showed that for some well-known natural exponential families, *X* can be *thinned* into independent random variables $X^{(1)}, \ldots, X^{(K)}$, such that $X = \sum_{k=1}^{K} X^{(k)}$. These independent random variables can then be used for various model validation and inference tasks, including in contexts where traditional sample splitting fails. In this paper, we generalize their procedure by relaxing this summation requirement and simply asking that some known function of the independent random variables exactly reconstruct *X*. This generalization of the procedure serves two purposes. First, it greatly expands the families of distributions for which thinning can be performed. Second, it unifies sample splitting and data thinning, which on the surface seem to be very different, as applications of the same principle. This shared principle is sufficiency. We use this insight to perform generalized thinning operations for a diverse set of families.

Keywords: cross-validation, sample splitting, exponential families, selective inference, model validation

1 Introduction

Suppose that we want to *fit* and *validate* a model using a single dataset. Two example scenarios are as follows:

> *Scenario 1.* We want to use the data both to generate and to test a hypothesis.
>
> *Scenario 2.* We want to use the data both to fit a complicated model, and to obtain an accurate estimate of the expected prediction error.

In either case, a naive approach that fits and validates a model on the same data is deeply problematic. In Scenario 1, testing a hypothesis on the same data used to generate it will lead to hypothesis tests that do not control the type 1 error, and to confidence intervals that do not attain the nominal coverage (Fithian et al., 2014). And in Scenario 2, estimating the expected prediction error on the same data used to fit the model will lead to massive downward bias (see Tian, 2020; Oliveira et al., 2021, for recent reviews).

In the case of Scenario 1, recent interest has focused on *selective inference*, a framework that enables a data analyst to generate and test a hypothesis on the same data (see, e.g., Taylor and Tibshirani, 2015). The main idea is as follows: to test a hypothesis generated from the data, we should condition on the event that we selected this particular hypothesis. Despite promising applications of this framework to a number of problems, such as inference after regression (Lee et al., 2016), changepoint detection (Jewell et al., 2022; Hyun et al., 2021), clustering (Gao et al., 2024; Chen and Witten, 2022; Yun and Barber, 2023), and outlier detection (Chen and Bien, 2020), it suffers from some drawbacks:

1. To perform selective inference, the procedure used to generate the null hypothesis must be fully-specified in advance. For instance, if a researcher wishes to cluster the data and then test for a difference in means between the clusters, as in Gao et al. (2024) and Chen and Witten (2022), then they must fully

specify the clustering procedure (e.g., hierarchical clustering with squared Euclidean distance and complete linkage, cut to obtain K clusters) in advance.

2. Finite-sample selective inference typically requires multivariate Gaussianity, though in some cases this can be relaxed to obtain asymptotic results (Taylor and Tibshirani, 2018; Tian and Taylor, 2017; Tibshirani et al., 2018; Tian and Taylor, 2018).

Thus, selective inference is not a flexible, "one-size-fits-all" approach to Scenario 1.

In the case of Scenario 2, proposals to de-bias the "in-sample" estimate of expected prediction error tend to be specialized to simple models, and thus do not provide an all-purpose tool that is broadly applicable (Oliveira et al., 2021).

Sample splitting (Cox, 1975) is an intuitive approach that applies to a variety of settings, including Scenarios 1 and 2; see the left-hand panel of Figure 1. We split a dataset containing n observations into two sets, containing n_1 and n_2 observations (where $n_1 \quad n_2 \quad n$). Then we can generate a hypothesis based on the first set and test it on the second (Scenario 1), or we can fit a model to the first set and estimate its error on the second (Scenario 2). Sample splitting also forms the basis for cross-validation (Hastie et al., 2009).

However, sample splitting suffers from some drawbacks:

1. If the data contain outliers, then each outlier is assigned to a single subsample.
2. If the observations are not independent (for instance, if they correspond to a time series) then the subsamples from sample splitting are not independent, and so sample splitting does not provide a solution to either Scenario 1 or Scenario 2.
3. Sample splitting does not enable conclusions at a per-observation level. For example, if sample splitting is applied to a dataset of the 50 states of the United States, then one can only conduct inference or perform validation on the states not used in fitting.

4. If the model of interest is fit using unsupervised learning, then sample splitting may not provide an adequate solution in either Scenario 1 or 2. The issue relates to #3 above. See Gao et al. (2024); Chen and Witten (2022), and Neufeld et al. (2024b).

In recent work, Neufeld et al. (2024a) proposed *convolution-closed data thinning* to address these drawbacks. They consider splitting, or *thinning*, a random variable X drawn from a convolution-closed family into K independent random variables $X^{(1)}, \ldots, X^{(K)}$ such that $X = \sum_{k=1}^{K} X^{(k)}$, and $X^{(1)}, \ldots, X^{(K)}$ come from the same family of distributions as X (see the right-hand panel of Figure 1). For instance, they show that $X \sim N(\mu, \sigma^2)$ can be thinned into two independent $N(\epsilon\mu, \epsilon\sigma^2)$ and $N((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$ random variables that sum to X. Further, if X is drawn from a Gaussian, Poisson, negative binomial, binomial, multinomial, or gamma distribution, then they can thin it *even when parameters of its distribution are unknown*. Because the thinned random variables are independent, this provides a new approach to tackle Scenarios 1 and 2: After thinning the data into independent parts, we fit a model to one part, and validate it on the rest.

On the surface, it is quite remarkable that one can break up a random variable X into two or more *independent* random variables that sum to X without knowing some (or sometimes any) of the parameters. In this paper, we explain the underlying principles that make this possible. We also show that convolution-closed data thinning can be generalized to increase its flexibility and applicability. The convolution-closed data thinning property $X = \sum_{k=1}^{K} X^{(k)}$ is desirable because it ensures that no information has been lost in the thinning process. However, clearly this would remain true if we were to replace the summation by any other deterministic function. Likewise, the fact that $X^{(1)}, \ldots, X^{(K)}$ are from the same family as X, while convenient, is nonessential.

Our generalization of convolution-closed data thinning is thus a procedure for splitting X into K random variables such that the following two properties hold:

(i) $X \quad T(X^{(1)}, \quad , X^{(K)})$; and (ii) $X^{(1)}, \quad , X^{(K)}$ are mutually independent.

This generalization is broad enough to simultaneously encompass both convolution-closed data thinning and sample splitting. Furthermore, it greatly increases the scope of distributions that can be thinned. In the $K = 2$ case, this generalized goal has been stated before (see Leiner et al., 2023, "P1" property). However, we are the first to develop a widely applicable strategy for achieving this goal. Not only can we thin exponential families that were not previously possible (such as the beta family), but we can even thin outside of the exponential family. For example, generalized thinning enables us to thin $X \sim \text{Unif}(0, \ )$ into $X^{(k)} \overset{\text{iid}}{\sim} \cdot\text{Beta}\left(\frac{1}{K}, 1\right)$, for $k \quad 1, \quad , K$, in such a way that $X \quad \max\{X^{(1)}, \quad , X^{(K)}\}$.

The primary contributions of our paper are as follows:

1. We propose *generalized data thinning*, a general strategy for thinning a single random variable X into two or more independent random variables, $X^{(1)}, \quad , X^{(K)}$, without knowledge of the parameter value(s). Importantly, we show that *sufficiency* is the key property underlying the choice of the function $T(\cdot)$.
2. We apply generalized data thinning to distributions far outside the scope of consideration of Neufeld et al. (2024a): These include the beta, uniform, and shifted exponential, among others. A summary of distributions covered by this work is provided in Table 1. In light of results by Darmois (1935); Koopman (1936), and Pitman (1936), we believe our examples are representative of the full range of cases to which this approach can be applied.
3. We show that sample splitting — which, on its surface, bears little resemblance to convolution-closed data thinning — is in fact based on the same principle:

Both are special cases of generalized data thinning with different choices of the function $T(\cdot)$. In other words, our proposal is a direct *generalization* of sample splitting.

We are not the first to generalize sample splitting. Inspired by Tian and Taylor (2018)'s use of randomized responses, Rasines and Young (2022) introduce the "(U, V)-decomposition", which injects independent noise W to create two independent random variables $U = u(X,W)$ and $V = v(X,W)$ that together are jointly sufficient for the unknown parameters. However, they do not describe how to perform a (U, V)-decomposition other than in the special case of a Gaussian random vector with known covariance. Our generalized thinning framework achieves the goal set out in their paper, providing a concrete recipe for finding such decompositions in a broad set of examples. The "data fission" proposal of Leiner et al. (2023) seeks random variables $f(X)$ and $g(X)$ for which the distributions of $f(X)$ and $g(X) \mid f(X)$ are known and for which $X = h(f(X), g(X))$. When these two random variables are independent (the "P1" property), their proposal aligns with generalized thinning. However, they do not provide a general strategy for performing P1-fission, and the only two examples they provide are the Gaussian vector with known covariance and the Poisson.

The rest of our paper is organized as follows. In Section 2, we define generalized data thinning, present our main theorem, and provide a simple recipe for thinning that is followed throughout the paper. Sections 3–5 demonstrate the utility of our approach in a series of examples organized by the results of Darmois (1935); Koopman (1936), and Pitman (1936): In particular, in Section 3, we consider the case of thinning natural exponential families; this section also revisits the convolution-closed data thinning proposal of Neufeld et al. (2024a) and clarifies the class of distributions that can be thinned using that approach. In Section 4, we apply data thinning to general exponential families. We consider distributions outside of the exponential family in Section 5. Section 6 contains examples of distributions that *cannot* be thinned using the approaches in this paper. Section 7 presents an application of data thinning to

changepoint detection. Finally, we close with a discussion in Section 8; additional technical details are deferred to the supplementary materials.

2 The generalized thinning proposal

We write X to denote a random variable that can be scalar-, vector-, or matrix-valued (and likewise for $X^{(1)}, \ldots, X^{(K)}$). When referring to a random variable or parameter that can only be vector- or matrix-valued, we use bolded symbols.

Definition 1 (Generalized data thinning). *Consider a family of distributions* $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. *Suppose that there exists a distribution* G_t, *not depending on* θ, *and a deterministic function* $T(\cdot)$ *such that when we sample* $(X^{(1)}, \ldots, X^{(K)}) \mid X$ *from* G_X, *for* $X \sim P_\theta$, *the following properties hold:*

1. $X^{(1)}, \ldots, X^{(K)}$ *are mutually independent (with distributions depending on* θ*), and*
2. $X = T(X^{(1)}, \ldots, X^{(K)})$.

Then we say that $\mathcal{P}$ is *thinned* by the function $T(\cdot)$.

When clear from context, sometimes we will say that "P_θ is thinned" or "X is thinned", by which we mean that the corresponding family $\mathcal{P}$ is thinned. Intuitively, we can think of thinning as breaking X up into K independent pieces, but in a very particular way that ensures that none of the information about θ is lost. The fact that no information is lost is evident from the requirement that $X = T(X^{(1)}, \ldots, X^{(K)})$.

Sample splitting (Cox, 1975) can be viewed as a special case of generalized data thinning.

Remark 1 (Sample splitting). *Sample splitting, in which a sample of n independent and identically distributed random variables is partitioned into K subsamples, is a special case of generalized data thinning. Here,* $T(\cdot)$ *is the function that takes in the*

subsamples as arguments, and concatenates and sorts their elements. For more details, see Section 5.2.

Furthermore, Definition 1 is closely related to the proposal of Neufeld et al. (2024a).

Remark 2 (Thinning convolution-closed families of distributions).

Neufeld et al. (2024a) show that some well-known families of convolution-closed distributions, such as the binomial, negative binomial, gamma, Poisson, and Gaussian, can be thinned, in the sense of Definition 1, by addition: $T\left(x^{(1)}, \ldots, x^{(K)}\right) = \sum_{k=1}^{K} x^{(k)}$.

The two examples above do not resemble each other: The first involves a non-parametric family of distributions and applies quite generally, while the second depends on a specific property of the family of distributions. Furthermore, the functions $T(\cdot)$ are quite different from each other. It is natural to ask: How can we find families $\mathcal{P}$ that can be thinned? Is there a unifying principle for the choice of $T(\cdot)$? How can we ensure that there exists a distribution G_t as in Definition 1 that does not depend on θ? The following theorem answers these questions, and indicates that *sufficiency* is the key principle required to ensure that the distribution G_t does not depend on θ.

Theorem 1 (Main theorem). *Suppose* $\mathcal{P}$ *is thinned by a function* $T(\cdot)$ *and, for* $X \sim P$, *let* $Q^{(1)} \otimes \cdots \otimes Q^{(K)}$ *denote the distribution of the mutually independent random variables,* $(X^{(1)}, \ldots, X^{(K)})$, *sampled as in Definition 1. Then, the following hold:*

(a) $T(X^{(1)}, \ldots, X^{(K)})$ *is a sufficient statistic for* θ *based on* $(X^{(1)}, \ldots, X^{(K)})$.

(b) The distribution G_t *in Definition 1 is the conditional distribution*

$$(X^{(1)}, \ldots, X^{(K)}) \,|\, T(X^{(1)}, \ldots, X^{(K)}) = t,$$

where $(X^{(1)}, \ldots, X^{(K)}) \sim Q^{(1)} \otimes \cdots \otimes Q^{(K)}$.

Theorem 1 is proven in Supplement A.1. Further, there is a simple algorithm for finding families of distributions $\mathcal{P}$ and functions $T(\cdot)$ such that $\mathcal{P}$ can be thinned by $T(\cdot)$.

Algorithm 1 (Finding distributions that can be thinned).

1. *Choose K families of distributions,* $\mathcal{Q}^{(k)} \quad \{Q^{(k)} : \quad\}$ *for* $k \quad 1, \quad , K$.
2. *Let* $(X^{(1)}, \quad , X^{(K)}) \sim Q^{(1)} \quad \cdots \quad Q^{(K)}$ *, and let* $T(X^{(1)}, \quad , X^{(K)})$ *denote a sufficient statistic for* θ
3. *Let* P *denote the distribution of* $T(X^{(1)}, \quad , X^{(K)})$.

By construction, the family $\mathcal{P} \quad \{P : \quad\}$ *is thinned by* $T(\cdot)$.

This recipe gives us a very succinct way to describe the distributions that can be thinned: *We can thin the distributions of sufficient statistics*. In particular, the recipe takes as input a joint distribution $Q^{(1)} \quad \cdots \quad Q^{(K)}$, and requires us to choose a sufficient statistic for θ. Then, that statistic's distribution is the P that can be thinned.

3 Thinning natural exponential families

In Section 3.1, we show how to thin a natural exponential family into two or more natural exponential families. In Section 3.2, we show how the convolution-closed thinning proposal of Neufeld et al. (2024a) can be understood in light of natural exponential family thinning. Finally, in Section 3.3, we show how natural exponential families can be thinned into more general (i.e., not necessarily natural) exponential families.

3.1 Thinning natural into natural exponential families

A natural exponential family (Lehmann and Romano, 2005) starts with a known probability distribution *H*, and then forms a family of distributions $\mathcal{P}^H \quad \{P^H : \quad\}$ based on *H*:

$$dP^H(x) \quad e^{x^\top \quad H(\ )} dH(x). \qquad (1)$$

The normalizing constant $e^{A(\theta)}$ ensures that P_θ is a probability distribution, and we take Ω to be the set of θ for which this normalization is possible (i.e. for which $A(\theta) < \infty$).

The next theorem presents a property of H that is necessary and sufficient for the resulting natural exponential family $\mathcal{P}^H$ to be thinned by addition into K natural exponential families. To streamline the statement of the theorem, we start with a definition.

Definition 2 (K-way convolution). *A probability distribution H is the K-way convolution of distributions $H_1, \ldots, H_K$ if $\sum_{k=1}^{K} Y_k \sim H$ for $(Y_1, \ldots, Y_K) \sim H_1 \cdots H_K$.*

Theorem 2 (Thinning natural exponential families by addition). *The natural exponential family $\mathcal{P}^H$ can be thinned by $T(x^{(1)}, \ldots, x^{(K)}) = \sum_{k=1}^{K} x^{(k)}$ into K natural exponential families $\mathcal{P}^{H_1}, \ldots, \mathcal{P}^{H_K}$ if and only if H is the K-way convolution of $H_1, \ldots, H_K$.*

The K natural exponential families in Theorem 2 can be different from each other, but they are all indexed by the same θ that was used in the original family $\mathcal{P}^H$. The proof of Theorem 2 is in Supplement A.2.

Neufeld et al. (2024a) show that it is possible to thin a Gaussian random variable by addition into K independent Gaussians. We now see that this result follows from Theorem 2.

Example 3.1 (Thinning $N_n(\theta, \mathbf{I}_n)$). *Distributions of the form $N_n(\theta, \mathbf{I}_n)$ are a natural exponential family indexed by $\theta \in \mathbb{R}^n$. It can be written in the notation of (1) as $\mathcal{P}^H$, where H represents the $N_n(\mathbf{0}_n, \mathbf{I}_n)$ distribution. Furthermore, H is the K-way convolution of $H_k = N_n(\mathbf{0}_n, \epsilon_k \mathbf{I}_n)$ for $k = 1, 2, \ldots, K$, where $\epsilon_1, \ldots, \epsilon_K \geq 0$ and $\sum_{k=1}^{K} \epsilon_k = 1$. Thus, by Theorem 2, we can thin $\mathcal{P}^H$ by addition into $\mathcal{P}^{H_1}, \ldots, \mathcal{P}^{H_K}$, where $P_\theta^{H_k} = N_n(\epsilon_k \theta, \epsilon_k \mathbf{I}_n)$.*

In Supplement B, we show that Example 3.1 is closely connected to a randomization strategy that has been frequently used in the literature.

Not all natural exponential families satisfy the condition of Theorem 2. We prove in Section 6.1 that the distribution $H \quad \text{Bernoulli}(0.5)$ cannot be written as the sum of two independent, non-constant random variables. Since $\mathcal{P}^{\text{Bernoulli}(0.5)}$ is the $\text{Bernoulli}([1 \quad e \quad]^{1})$ natural exponential family, Theorem 2 implies that Bernoulli random variables cannot be thinned by addition into natural exponential families. In Section 6.1 we will further prove that *no function* $T(\cdot)$ can thin the Bernoulli family.

3.2 Connections to Neufeld et al. (2024a)

Neufeld et al. (2024a) focus on convolution-closed families, i.e., those for which convolving two or more distributions (see Definition 2) in the family produces a distribution that is in the family. They provide a recipe for decomposing a random variable X drawn from a distribution in such a family into independent random variables $X^{(1)}, \quad , X^{(K)}$ that sum to yield X. We now show that their results are encompassed by Theorem 2.

Exponential dispersion families (Jørgensen, 1992; Jørgensen, 1997) are a subclass of convolution-closed families. Given a distribution H with $_{H}( )$ for (as in (1)), we identify the set of distributions H for which $_{H}(\cdot) \quad _{H}(\cdot)$ (i.e., distributions whose cumulant generating function is a multiple of H's cumulant generating function). We define Λ to be the set of λ for which such a distribution H exists. Then, an (additive) *exponential dispersion family* is $\mathcal{P} \quad \bigcup \mathcal{P}^{H}$, where $\mathcal{P}^{H}$ is the natural exponential family generated by H (see (1)). The distributions in $\mathcal{P}$ are indexed over $(\ , \ )$ and take the form $dP^{H}(x) \quad e^{x^{\top}} \quad _{H}( ) dH(x)$.

In words, an exponential dispersion family results from combining a collection of related natural exponential families. For example, starting with $H \quad \text{Bernoulli}(1/2)$, we can take

$\mathbb{Z}$ since for any positive integer λ, H Binomial(,1/2) satisfies the necessary cumulant generating function relationship. Then, $\mathcal{P}^{\text{Binomial}(,1/2)}$ corresponds to the binomial natural exponential family that results from fixing λ Finally, allowing $\lambda$to vary gives the full binomial exponential dispersion family, which is the set of all binomial distributions (varying both of the parameters of the binomial distribution).

By construction, for any $_1$, , $_K$, convolving P^{H_1}, , P^{H_K} gives the distribution P^H, where $_{k\ 1}^{K}{}_k$. The next corollary is an immediate application of Theorem 2 in the context of exponential dispersion families. Notably, the distributions $Q^{(k)}$ themselves still belong to the exponential dispersion family $\mathcal{P}$ to which the distribution of X belongs.

Corollary 1 (Thinning while remaining inside an exponential dispersion family). *Consider an exponential dispersion family* $\mathcal{P} \bigcup \mathcal{P}^H$ *and suppose* $_1$, , $_K$. *Then for* $_{k\ 1}^{K}{}_k$, *we can thin the natural exponential family* $\mathcal{P}^H$ *by* $T(x^{(1)}, ,x^{(K)})$ $_{k\ 1}^{K} x^{(k)}$ *into the natural exponential families* $\mathcal{P}^{H_1}$, , $\mathcal{P}^{H_K}$.

This result corresponds exactly to the data thinning proposal of Neufeld et al. (2024a). We see from Corollary 1 that that proposal thins a natural exponential family, $\mathcal{P}^H$, into a *different* set of natural exponential families, $\mathcal{P}^{H_1}$, , $\mathcal{P}^{H_K}$. However, from the perspective of exponential dispersion families, it thins an exponential dispersion family into the same exponential dispersion family. Continuing the binomial example from above, the corollary tells us that we can thin the binomial family with λas the number of trials into two or more binomial families with smaller numbers of trials, provided that 1.

Neufeld et al. (2024a) focus on convolution-closed families, not exponential dispersion families. However, all convolution-closed families that have moment-generating functions can be written as exponential dispersion families (J ø r g e n s e n a n d

Song, 1998). The Cauchy distribution is convolution-closed, but does not have a moment generating function and thus is not an exponential dispersion family. As we will see in Example 6.1, the $\text{Cauchy}(\theta_1, \theta_2)$ distribution cannot be thinned by addition: Decomposing it using the recipe of Neufeld et al. (2024a) requires knowledge of both unknown parameters. Thus, not all convolution-closed distributions can be thinned by addition in the sense of Definition 1. However, Neufeld et al. (2024a) claim that all convolution-closed distributions *can* be thinned. This apparent discrepancy is due to a slight difference in the definition of thinning between our paper and theirs: Our Definition 1 requires that G_t not depend on θ; however, Neufeld et al. (2024a) have no such requirement. In practice, data thinning is useful only if G_t does not depend on θ, and so there is no meaningful difference between the two definitions.

3.3 Thinning natural into general exponential families

In this section, we apply Algorithm 1 in the case that $\mathcal{Q}^{(k)}$ are (possibly non-natural) exponential families, for which the sufficient statistic need not be the identity. In particular, for $k = 1, \ldots, K$, we let $\mathcal{Q}^{(k)} = \{Q^{(k)}_\theta : \theta \in \Omega\}$ denote an exponential family based on a known distribution H_k and sufficient statistic $T^{(k)}(\cdot)$:

$$dQ^{(k)}_\theta(x) = \exp\{[T^{(k)}(x)]^\top \eta(\theta) - A_k(\eta(\theta))\} dH_k(x). \quad (2)$$

As in Section 3.1, $e^{-A_k(\eta(\theta))}$ is the normalizing constant needed to ensure that $\int dQ^{(k)}_\theta(x) = 1$ and Ω is the set of θ for which $A_k(\eta(\theta)) < \infty$. The function $\eta(\cdot)$ maps θ to the natural parameter. We note that $\sum_{k=1}^{K} T^{(k)}(X^{(k)})$ is a sufficient statistic for θ based on $(X^{(1)}, \ldots, X^{(K)}) \sim Q^{(1)} \times \cdots \times Q^{(K)}$. Then, Algorithm 1 tells us that we can thin the distribution of this sufficient statistic. This leads to the next result.

Proposition 1 (Thinning natural exponential families with more general functions $T(\cdot)$). *Let $X^{(1)}, \ldots, X^{(K)}$ be independent random variables with $X^{(k)} \sim Q^{(k)}$ for $k = 1, \ldots, K$ from any (i.e., possibly non-natural) exponential families $\mathcal{Q}^{(k)}$ as in (2). Let P denote the*

distribution of $\sum_{k=1}^{K} T^{(k)}(X^{(k)})$. *Then,* $\mathcal{P} = \{P : \quad\}$ *is a natural exponential family, and we can thin it into* $X^{(1)}, \quad, X^{(K)}$ *using the function* $T(x^{(1)}, \quad, x^{(K)}) \quad \sum_{k=1}^{K} T^{(k)}(x^{(k)})$.

The fact that $\mathcal{P}$ in this result is a natural exponential family follows from recalling that the sufficient statistic of an exponential family follows a natural exponential family (Lehmann and Romano, 2005, Lemma 2.7.2(i)). Many named exponential families are not natural exponential families, involving non-identity functions $T^{(k)}(\cdot)$, such as the logarithm or polynomials. Therefore, to thin into those families, Proposition 1 will be useful.

Proposition 1 implies that many natural exponential families *can* be thinned by a function of the form $T(x^{(1)}, \quad, x^{(K)}) \quad \sum_{k=1}^{K} T^{(k)}(x^{(k)})$. Theorem 3 shows that if a full-rank natural exponential family can be thinned, then the thinning function *must* take this form.

Theorem 3 (Thinning functions for natural exponential families). *Suppose* $X \sim P$, *where* $\mathcal{P} = \{P : \quad\}$ *is a full-rank natural exponential family with density/mass function* $p(x) = \exp(\quad^{\top} x \quad (\quad)) h(x)$. *If* $\mathcal{P}$ *can be thinned by* $T(\cdot)$ *into* $X^{(1)}, \quad, X^{(K)}$, *then:*

1. *The function* $T(x^{(1)}, \quad, x^{(K)})$ *is of the form* $\sum_{k=1}^{K} T^{(k)}(x^{(k)})$.
2. $X^{(k)} \overset{\text{ind}}{\sim} Q^{(k)}$ *where* $Q^{(k)}$ *is an exponential family with sufficient statistic* $T^{(k)}(X^{(k)})$.

The proof of Theorem 3 is provided in Supplement A.3.

To illustrate the flexibility provided by Proposition 1 and Theorem 3, we demonstrate that a natural exponential family $\mathcal{P}$ can be thinned by different functions $T(\cdot)$, leading to families of distributions $\mathcal{Q}^{(1)}, \quad, \mathcal{Q}^{(K)}$ different from $\mathcal{P}$. Specifically, we consider three

possible *K*-fold thinning strategies for a gamma distribution when the shape, α, is known but the rate[1], θ, is unknown.

Example 3.2 (Thinning $\text{Gamma}(\ ,\ )$ with αknown, approach 1). *Following Algorithm 1, we start with* $X^{(k)} \overset{iid}{\sim} \text{Gamma}\ \frac{}{K},$ *for* $k \quad 1, \quad ,K$ *, and note that* $T(X^{(1)}, \quad ,X^{(K)}) \quad \sum_{k\ 1}^{K} X^{(k)}$ *is sufficient for* θ *Thus, we can thin the distribution of* $\sum_{k\ 1}^{K} X^{(k)}$ *. A well-known property of the gamma distribution tells us that this is a* $\text{Gamma}(\ ,\ )$ *distribution. Sampling from* G_t *as in Theorem 1 corresponds exactly to the multi-fold gamma data thinning recipe of Neufeld et al. (2024a) where* $\epsilon_k \quad \frac{1}{K}$.

Alternatively, when αcan be expressed as half of a natural number, we can apply Proposition 1 to decompose the gamma family into centred normal data.

Example 3.3 (Thinning $\text{Gamma}(\ ,\ )$ with $K/2$ known, approach 2). *Starting with* $X^{(k)} \overset{iid}{\sim} N(0,\frac{1}{2})$ *, notice that* $T^{(k)}(x^{(k)}) \quad (x^{(k)})^2$ *. We thus apply Proposition 1 using* $T(x^{(1)}, \quad ,x^{(K)}) \quad \sum_{k\ 1}^{K} (x^{(k)})^2$ *to thin the sufficient statistic,* $\sum_{k\ 1}^{K} (X^{(k)})^2 \sim \frac{1}{2} \quad {}^{2}_{K} \quad \text{Gamma}\frac{K}{2},$ *, into* $(X^{(1)}, \quad ,X^{(K)})$ *. The function* G_t *from Theorem 1 is the conditional distribution* $(X^{(1)}, \quad ,X^{(K)}) \mid \sum_{k\ 1}^{K} (X^{(k)})^2 \quad t$ *. By rotational symmetry of the* $N_K(0,(2\ )^{1}\mathbf{I}_K)$ *distribution (the joint distribution of* $(X^{(1)}, \quad ,X^{(K)})$ *), G_t is the uniform distribution on the* $(K \quad 1)$ *-sphere of radius* $t^{1/2}$ *. To sample from this conditional distribution, we generate* $\mathbf{Z} \sim N_K(0,\mathbf{I}_K)$ *and then take* $(X^{(1)}, \quad ,X^{(K)})$ *to be* $t^{1/2}\frac{\mathbf{Z}}{\|\mathbf{Z}\|_2}$ *.*

If $\alpha$is a natural number, then applying a similar logic enables us to thin the gamma family with unknown rate into the Weibull family with unknown scale; see Example C.1 in Supplement C.1.1. From a theoretical perspective, when $\alpha$is a natural number, there

is no reason to prefer one of the three gamma thinning strategies over another. However, there may be practical considerations: For instance, the strategy in Example 3.3 may be preferred due to the convenience of working with Gaussian data. In general, if multiple thinning strategies are available, then the choice can be driven by modeling convenience.

4 Indirect thinning of general exponential families

Sometimes rather than thinning X, we may choose to thin a function $S(X)$. When $S(X)$ is sufficient for θ based on X, the next proposition tells us that thinning $S(X)$ rather than X does not result in a loss of information about θ. We emphasize that we are using the concept of sufficiency in two ways here: (i) $S(X)$ is sufficient for θ based on $X \sim P$, and (ii) $T(X^{(1)}, \ldots, X^{(K)})$ is sufficient for θ based on $(X^{(1)}, \ldots, X^{(K)}) \sim Q^{(1)} \times \cdots \times Q^{(K)}$.

Proposition 2 (Thinning a sufficient statistic preserves information). *Suppose $X \sim P \in \mathcal{P}$ has a sufficient statistic S(X) for θ, and we thin S(X) by $T(\cdot)$. That is, conditional on S(X) (and without knowledge of θ) we sample $X^{(1)}, \ldots, X^{(K)}$ that are mutually independent and satisfy $S(X) = T(X^{(1)}, \ldots, X^{(K)})$. Under regularity conditions needed for Fisher information to exist, we have that $I_X(\theta) = \sum_{k=1}^{K} I_{X^{(k)}}(\theta)$.*

This proposition shows that thinning $S(X)$, rather than X, does not result in a loss of information about θ. Its proof (provided in Supplement A.4) follows easily from multiple applications of the fact that sufficient statistics preserve information. Definition 3 formalizes the strategy suggested by Proposition 2.

Definition 3 (Indirect thinning). *Consider $X \sim P \in \mathcal{P}$. Suppose we thin a sufficient statistic S(X) for θ by a function $T(\cdot)$. We say that the family $\mathcal{P}$ is indirectly thinned through $S(\cdot)$ by $T(\cdot)$.*

In light of Proposition 2, indirect thinning does not result in a loss of information.

When $S(\cdot)$ is invertible, then $X \quad S^{-1}(T(X^{(1)}, \quad, X^{(K)}))$, which implies that we can thin X directly by $S^{-1}(T(\cdot))$. It turns out that, regardless of whether we thin X by $S^{-1}(T(\cdot))$ or indirectly thin X through $S(\cdot)$ by $T(\cdot)$, there is little difference between the resulting form of G_t in Theorem 1. In the former case, G_t is the conditional distribution of $(X^{(1)}, \quad, X^{(K)})$ given $S^{-1}(T(X^{(1)}, \quad, X^{(K)})) \quad t$. In the latter case, it is the conditional distribution of $(X^{(1)}, \quad, X^{(K)})$ given $T(X^{(1)}, \quad, X^{(K)}) \quad t$. Since $S(\cdot)$ is invertible, these two conditional distributions are identical following a reparameterization.

We now return to the setting of Proposition 2, where $S(\cdot)$ may or may not be invertible.

Remark 3 (Indirect thinning of general exponential families). *Let* $\mathcal{P} \quad \{P: \quad\}$ *be a full-rank general exponential family. That is,* $dP(x) \quad \exp\{[S(x)]^{\top} \quad ( ) \quad ( )\}dH(x)$, *where* $e^{\quad ( )}$ *is the normalising constant. Since S(X) is sufficient for* θ, *we can indirectly thin X through* $S(\cdot)$ *without a loss of Fisher information (Proposition 2). Furthermore, S(X) belongs to a full-rank natural exponential family (Lehmann and Romano, 2005, Lemma 2.7.2(i)). We can thus indirectly thin X through* $S(\cdot)$ *as follows:*

1. *Provided that the necessary and sufficient condition of Theorem 2 holds for S(X), we can indirectly thin X through* $S(\cdot)$ *by addition into* $X^{(1)}, \quad, X^{(K)}$ *that follow natural exponential families, i.e. (2) where* $T^{(k)}(\cdot)$ *is the identity.*
2. *We now consider* $X^{(1)}, \quad, X^{(K)}$ *that belong to a general exponential family, where* $T^{(k)}(\cdot)$ *in (2) is not necessarily the identity. Suppose further that* $S(X) \overset{D}{\quad} \sum_{k \quad 1}^{K} T^{(k)}(X^{(k)})$. *Then, by Proposition 1, we can indirectly thin X through* $S(\cdot)$ *into* $X^{(1)}, \quad, X^{(K)}$, *by* $T(x^{(1)}, \quad, x^{(K)}) \quad \sum_{k \quad 1}^{K} T^{(k)}(x^{(k)})$.

We see that 1) is a special case of 2).

We now demonstrate indirect thinning with some examples. First, we consider a $\mathrm{Beta}(\quad , \quad)$ random variable, with β a known parameter. This is not a natural exponential

family, and so the results in Section 3 are not directly applicable. The beta family also differs from the other examples that we have seen in the following ways: (i) It is not convolution-closed; (ii) it has finite support; and (iii) the sufficient statistic for an independent and identically distributed sample has an unnamed distribution.

Example 4.1 (Thinning $\mathrm{Beta}(\ ,\ )$ with βknown). *We start with* $X^{(k)} \overset{ind}{\sim} \mathrm{Beta}\left(\frac{1}{K}\quad \frac{k\ \ 1}{K}, \frac{1}{K}\right)$*, for* $k\quad 1,\quad ,K$*; this is a general exponential family (2) with* $T^{(k)}(x^{(k)})\quad \frac{1}{K}\log(x^{(k)})$*. Since* $\sum_{k\ 1}^{K} T^{(k)}(X^{(k)})$ *is sufficient for* θ *based on* $X^{(1)}, ,X^{(K)}$*, we can apply Proposition 1 to thin the distribution of* $\sum_{k\ 1}^{K} T^{(k)}(X^{(k)})$ *by the function*

$$T(x^{(1)},\quad ,x^{(K)})\quad \sum_{k\ 1}^{K} T^{(k)}(x^{(k)})\quad \frac{1}{K}\sum_{k\ 1}^{K}\log(x^{(k)})\quad \log\prod_{k\ 1}^{K} x^{(k)^{1/K}}. \qquad (3)$$

Furthermore, we show in Supplement C.1.2 that $\exp\ T(X^{(1)},\quad ,X^{(K)})\quad \prod_{k\ 1}^{K} X^{(k)^{1/K}}$*, the geometric mean of* $X^{(1)},\quad ,X^{(K)}$*, follows a* $\mathrm{Beta}(\ ,\ )$ *distribution. Therefore, we can indirectly thin a* $\mathrm{Beta}(\ ,\ )$ *random variable through* $S(x)\quad \log(x)$ *by* $T(\cdot)$ *defined in (3). This results in* $X^{(k)} \overset{ind}{\sim} \mathrm{Beta}\left(\frac{1}{K}\quad \frac{k\ \ 1}{K}, \frac{1}{K}\right)$*, for* $k\quad 1,\quad ,K$*.*

Furthermore, since $S(x)\quad \log(x)$ *is invertible, we can directly thin* $X \sim \mathrm{Beta}(\ ,\ )$ *by*

$$T\ (x^{(1)},\quad ,x^{(K)})\quad S^{\ 1}(T(x^{(1)},\quad ,x^{(K)}))\quad \prod_{k\ 1}^{K} x^{(k)^{1/K}}. \qquad (4)$$

To apply either of these thinning strategies, we need to sample from G_t *defined in Theorem 1. This can be done using numerical methods, as detailed in Supplement C.1.2.*

By symmetry of the beta distribution, we can also apply the thinning operations detailed in Example 4.1 to thin a $\text{Beta}(\ ,\ )$ random variable with αknown. In Example C.2 in Supplement C.1.3, we propose an alternative strategy to thin a beta random variable, using a different parametrization. As this example extends naturally to higher dimensions, we derive and prove it for the more general Dirichlet case.

Next, we consider thinning the gamma distribution with unknown shape parameter.

Example 4.2 (Thinning $\text{Gamma}(\ ,\ )$ with βknown). *We start with* $X^{(k)} \overset{ind}{\sim} \text{Gamma}\left(\frac{1}{K}\ \ \frac{k}{K}\ \ 1, \frac{1}{K}\right)$*, for* $k = 1, \ldots, K$*; this is a general exponential family (2) with* $T^{(k)}(X^{(k)}) = \frac{1}{K}\log(x^{(k)})$*. Note that* $T(X^{(1)}, \ldots, X^{(K)}) = \sum_{k=1}^{K} T^{(k)}(X^{(k)})$ *is sufficient for* θ *based on* $X^{(1)}, \ldots, X^{(K)}$*. As* $T^{(k)}(\cdot)$ *is shared with Example 4.1, we can apply Proposition 1 to thin the distribution of* $\sum_{k=1}^{K} T^{(k)}(X^{(k)})$ *by the function defined in (3).*

In Supplement C.1.4 we show that $\exp\left(T(X^{(1)}, \ldots, X^{(K)})\right) = \prod_{k=1}^{K} X^{(k)^{1/K}}$ *follows a* $\text{Gamma}(\ ,\ )$ *distribution. Thus, we can indirectly thin a* $\text{Gamma}(\ ,\ )$ *random variable through* $S(x) = \log(x)$ *by* $T(\cdot)$ *defined in (3). This produces independent random variables* $X^{(k)} \sim \text{Gamma}(\frac{1}{K}\ \ \frac{k}{K}\ \ 1, \frac{1}{K}\ )$ *for* $k = 1, \ldots, K$*. Once again, noting that* $S(\cdot)$ *is invertible, we can instead directly thin* $X \sim \text{Gamma}(\ ,\ )$ *by the function defined in (4).*

To apply either of these thinning strategies to a $\text{Gamma}(\ ,\ )$ *random variable, we must sample from* G_t *as defined in Theorem 1. See Supplement C.1.4.*

Example 4.2 is different from the gamma thinning example from Neufeld et al. (2024a): That involves thinning a $\text{Gamma}(\ ,\ )$ random variable with αknown, whereas here we thin a $\text{Gamma}(\ ,\ )$ random variable with βknown.

Examples 4.1 and 4.2 enable us to thin a random variable into an arbitrary number of independent random variables. However, unlike in the examples in Section 3, the resulting folds are not identically distributed.

In Examples 4.1 and 4.2, the function $S(\cdot)$ through which we indirectly thin X is invertible. Supplement D considers indirect thinning of a sample of n independent and identically distributed normal random variables with both mean and variance unknown. This provides an example of a case in which $S(\cdot)$ is neither invertible, nor scalar-valued.

We close with a list of a few short examples to illustrate the flexibility of indirect thinning.

Example 4.3 (Additional examples of indirect thinning).

1. *Suppose we observe* $X \sim N(\ ,\ )$ *where* μ *is known; here* μ *denotes the mean and* θ *the variance. Then* $S(X) \quad (X \quad)^2 \sim \quad ^2_1 \text{Gamma}\ \frac{1}{2}, \frac{1}{2}$ *. Thus, by applying the Gamma thinning strategy of Neufeld et al. (2024a) discussed in Example 3.2 to S(X), we can indirectly thin a normal distribution with unknown variance through* $S(\cdot)$ *.*
2. *Suppose we observe* $X \sim \text{Weibull}(\ ,\ )$ *where* γ *is known. Then,* $S(X) \quad X \ \sim \text{Exp}$ *. Thus, by applying the Gamma thinning strategy of Example 3.2 or 3.3 to S(X), we can indirectly thin a Weibull distribution with unknown rate through* $S(\cdot)$ *.*
3. *Suppose we observe* $X \sim \text{Pareto}(\ ,\ )$ *where* γ *is known. Then* $S(X) \quad \log \ X / \quad \sim \text{Exp}(\ )$ *. Thus, by applying the Gamma thinning strategy of Example 3.2 or 3.3 to S(X), we can indirectly thin a Pareto distribution with unknown shape through* $S(\cdot)$ *.*

5 Thinning outside of exponential families

In this section, we focus on thinning outside of exponential families. Outside of the exponential family, only certain distributions with domains that vary with the parameter

of interest have sufficient statistics that are bounded as the sample size increases (Darmois, 1935; Koopman, 1936; Pitman, 1936). Thus, we first consider a setting where θ alters the support of the distribution (Section 5.1), and then one where the sufficient statistic's dimension grows as the sample size increases (Section 5.2).

5.1 Thinning distributions with varying support

We consider examples in which the parameter of interest, θ, changes the support of a distribution. In Example 5.1, θ scales the support.

Example 5.1 (Thinning $\mathrm{Unif}(0, \theta)$). *We start with* $X^{(k)} \overset{iid}{\sim} \theta \cdot \mathrm{Beta}\left(\frac{1}{K}, 1\right)$ *for* $k = 1, \ldots, K$*, and note that* $T(X^{(1)}, \ldots, X^{(K)}) = \max(X^{(1)}, \ldots, X^{(K)})$ *is sufficient for* θ*. Furthermore,* $\max(X^{(1)}, \ldots, X^{(K)}) \sim \mathrm{Unif}(0, \theta)$*. Thus, we define* G_t *to be the conditional distribution of* $(X^{(1)}, \ldots, X^{(K)})$ *given* $\max(X^{(1)}, \ldots, X^{(K)}) = t$*. Then, by Theorem 1, we can thin* $X \sim \mathrm{Unif}(0, \theta)$ *by sampling from* G_X*. To do this, we first draw* $\mathbf{C} \sim \mathrm{Categorical}_K(1/K, \ldots, 1/K)$*. Then,* $X^{(k)} = C_k X + (1 - C_k) Z_k$ *where* $Z_k \overset{iid}{\sim} X \cdot \mathrm{Beta}\left(\frac{1}{K}, 1\right)$*.*

This is a special case of Example C.3 in Supplement C.2.1, in which we thin the scale family $\theta \cdot \mathrm{Beta}(\alpha, 1)$ where α is known. Setting $\alpha = 1$ yields Example 5.1.

Similar thinning results can be identified for distributions in which θ shifts the support. In Supplement C.2.2, we show that $X \sim \mathrm{SExp}(\theta, \lambda)$, the location family generated by shifting an exponential random variable by θ, can be thinned by the minimum function.

5.2 Sample splitting as a special case of generalized data thinning

We now consider sample splitting, a well-known approach for splitting a sample of observations into two or more sets (Cox, 1975). We show that sample splitting can be viewed as an instance of generalized data thinning. In this setting, $\mathbf{X} = (X_1, \ldots, X_n)$ is a sample of independent and identically distributed random variables, $X_i \in \mathcal{X}$, each

having distribution $F \in \mathcal{F}$, where $\mathcal{F}$ is some (potentially non-parametric) family of distributions and $\mathcal{X}$ is the set of values that the random variable X_i can take (most commonly $\mathcal{X} = \mathbb{R}^p$). That is, $\mathbf{X} \sim P_F \in \mathcal{P}$, where $\mathcal{P} = \{F^n : F \in \mathcal{F}\}$, and $F^n = F \times \cdots \times F$ denotes the joint distribution of n independent random variables drawn from F.

Example 5.2 (Sample splitting is a special case of generalized data thinning). *We begin with* $\mathbf{X}^{(k)} = (X_1^{(k)}, \ldots, X_{n_k}^{(k)}) \overset{iid}{\sim} F^{n_k}$, *for* $k = 1, \ldots, K$. *Here,* $n_1, \ldots, n_K > 0$, *and* $\sum_{k=1}^{K} n_k = n$. *That is, for* $k = 1, \ldots, K$, $\mathbf{X}^{(k)} \in \mathcal{X}^{n_k}$ *denotes a set of* n_k *independent and identically distributed draws from F.*

Our goal is to thin $S(\mathbf{X})$, *where* $S : \mathcal{X}^n \to \mathcal{X}^n$ *sorts the entries of its input based on their values. We define* $T : \mathcal{X}^{n_1} \times \cdots \times \mathcal{X}^{n_K} \to \mathcal{X}^n$ *as* $T(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}) = S((\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}))$, *the function that concatenates its arguments and then applies* $S(\cdot)$. *Then* $T(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(K)})$ *is a sufficient statistic for F, and furthermore,* $T(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(K)}) \overset{D}{=} S(\mathbf{X})$.

We define $G_{\mathbf{t}}$ *to be the conditional distribution of* $(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(K)})$ *given* $T(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(K)}) = \mathbf{t}$. *Suppose we observe* $\mathbf{X} \sim F^n$. *Then, by Theorem 1, we can indirectly thin* $\mathbf{X}$ *through* $S(\cdot)$ *by* $T(\cdot)$ *by sampling from* $G_{S(\mathbf{X})}$. *This conditional distribution is uniform over all* $\frac{n!}{n_1! \cdots n_K!}$ *assignments of n items to K groups of sizes* $n_1, \ldots, n_K$. *Thus, to sample from* $G_{S(\mathbf{X})}$, *we randomly partition the sample of size n into K groups of sizes* $n_1, \ldots, n_K$. *This is precisely the same as sample splitting.*

We have shown that when one has *n* independent and identically distributed samples from a distribution *F*, then sample splitting is an instance of generalized data thinning. When this assumption holds, it follows from Proposition 2 that sample splitting preserves all information about *F*. In practice, however, sample splitting is often applied in situations where we have *n* random variables that are not independent or not identically distributed. In such a situation, using a valid generalized data thinning

strategy will be advantageous. For example, consider the setting of multivariate Gaussian data with known dense covariance. Since the data are not independent, sample splitting will produce dependent folds whereas multivariate Gaussian data thinning generates independent folds. Next, consider the case of linear regression with a fixed design matrix: The data are independent but not identically distributed. In this setting, Neufeld et al. (2024a) and Rasines and Young (2022) show that Gaussian data thinning is preferable to sample splitting from the standpoint of Fisher information (see Section 4 of Neufeld et al. (2024a) for technical details).

6 Counterexamples

We now present two examples in which thinning strategies do not work. The first involves a natural exponential family that is based on a distribution that *cannot* be written as the convolution of two distributions. In this case, Theorem 2 implies that we cannot thin it by addition. In fact, we will prove a stronger statement: Namely, that there does not exist *any* function $T(\cdot)$ that can thin it. The second example involves a convolution-closed family outside of the natural exponential family in which addition is not sufficient. In this case, taking $T(\cdot)$ to be addition does not enable thinning, as Theorem 1 does not apply.

6.1 The Bernoulli family cannot be thinned

Let P denote the $\text{Bernoulli}(\theta)$ distribution, where θ is the probability of success. Recall that this distribution can be written as a natural exponential family (with natural parameter $\log \frac{\theta}{1-\theta}$). By Theorem 3, if P can be thinned, then the thinning function $T(\cdot)$ must be additive. However, as the next theorem shows, the Bernoulli distribution cannot be written as a convolution of independent, non-constant random variables.

Theorem 4 (The Bernoulli is not a convolution). *If $Z^{(1)}$ and $Z^{(2)}$ are independent, non-constant random variables, then $Z^{(1)} + Z^{(2)}$ cannot be a Bernoulli random variable.*

Theorem 4 is proven in Supplement A.5.

As the Bernoulli distribution cannot be written as a convolution of non-constant random variables, it cannot achieve the two conclusions of Theorem 3 simultaneously. Thus, a contrapositive argument applied to Theorem 3 leads to the next result.

Corollary 2. *The Bernoulli family cannot be thinned by any function* $T(\cdot)$.

This corollary of Theorems 3 and 4 is proven in Supplement A.6. A similar argument reveals that the categorical distribution also cannot be thinned.

The above corollary pertains to a *single* Bernoulli random variable. By contrast, a *vector* of independent and identically distributed Bernoulli random variables can be thinned by sample splitting or by indirect binomial thinning on the sum of the entries.

6.2 The Cauchy family cannot be thinned by addition

Suppose now that our interest lies in a random variable $X = T(X^{(1)}, X^{(2)})$, where $T(X^{(1)}, X^{(2)})$ is *not* sufficient for the parameter θ based on $(X^{(1)}, X^{(2)})$. This means that the conditional distribution of $(X^{(1)}, X^{(2)})$ given $T(X^{(1)}, X^{(2)})$ depends on θ, and thus that we cannot thin X by $T(\cdot)$. We see this in the following example.

Example 6.1 (The trouble with thinning $\text{Cauchy}(\theta_1, \theta_2)$ by addition). *Recall that the Cauchy family,* $\text{Cauchy}(\theta_1, \theta_2)$*, indexed by* $\theta = (\theta_1, \theta_2)$*, is convolution-closed. In particular, if* $X^{(1)}, X^{(2)} \overset{iid}{\sim} \text{Cauchy}\left(\frac{1}{2}\theta_1, \frac{1}{2}\theta_2\right)$*, then* $X^{(1)} + X^{(2)} \sim \text{Cauchy}(\theta_1, \theta_2)$*. It is tempting therefore to try thinning this family by* $T(x^{(1)}, x^{(2)}) = x^{(1)} + x^{(2)}$*. However, the sum* $X^{(1)} + X^{(2)}$ *is not sufficient for either* θ_1 *or* θ_2*, which means that Theorem 1 does not apply. In particular,* G_t*, the conditional distribution of* $(X^{(1)}, X^{(2)})$ *given* $X^{(1)} + X^{(2)} = t$*, is a function of* θ*. Therefore, we cannot thin the Cauchy family with any unknown parameters by addition.*

We can take this result a step further: Given a collection of Cauchy random variables, there is no sufficient statistic for θ that reduces the data beyond the order statistics

(Casella and Berger, 2002, p. 275). Thus, following Algorithm 1 with $\mathcal{Q}^{(k)}$ being $\text{Cauchy}(\ _1, \ _2)$, the only generalized data thinning approach that generates independent Cauchy random variables is sample splitting a vector of independent Cauchy random variables.

7 Changepoint detection in wind speed data

To demonstrate the utility of generalized data thinning, we consider detecting changepoints in the variance of wind speed data. We consider a wind speed dataset (Haslett and Raftery, 1989) collected in the Irish town of Claremorris, available in the R package `gstat` (Pebesma, 2004). Killick and Eckley (2014) took first differences to remove the periodic mean, and then modeled the resulting X_i for $i = 1, \ldots, n$ as independent normal observations with $X_i \sim N(0, \ _i)$. They then estimated changepoints in the variance $\ _1, \ldots, \ _n$. Here, we take their analysis a step further by testing for a difference in variance on either side of each estimated changepoint.

First, we consider a naive approach.

Algorithm 2 (Naive approach for changepoint detection).

1. *Compute* $Z_i := X_i^2$. *Note that* $Z_i \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2\ _i}\right)$.
2. *Estimate changepoints in* $Z_1, \ldots, Z_n$.
3. *Fit a gamma GLM to test for a change in the rate of* Z_i *on either side of each estimated changepoint.*

To carry out Step 2 of Algorithm 2, we use the nonparametric changepoint detection method of Haynes et al. (2017), implemented in the `changepoint.np` R package (Haynes and Killick, 2022), with a BIC penalty and a minimum segment length of 10 days.

However, using the same data to estimate and test changepoints will lead to many false discoveries, as pointed out by Hyun et al. (2021) and Jewell et al. (2022) in a related setting.

A natural alternative is to use *order-preserved sample splitting*, which involves estimating changepoints on a training set composed of odd-indexed observations, and testing those changepoints on a test set composed of even-indexed observations (Zou et al., 2020). Note that order-preserved sample splitting is different from Example 5.2. Since the Z_i are *not* independent and identically distributed, it is *not* a special case of data thinning.

Algorithm 3 (Order-preserved sample splitting approach for changepoint detection).

1. *Compute* $Z_i := X_i^2$. *Note that* $Z_i \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2\lambda_i}\right)$.
2. *Assume n is even. Estimate changepoints in odd observations* $Z_1, Z_3, \ldots, Z_{n-1}$.
3. *Fit a gamma GLM to test for a change in the rate of* Z_i *on either side of each estimated changepoint using even observations* $Z_2, Z_4, \ldots, Z_n$.

In Step 2 of Algorithm 3, we again use the changepoint.np R package with a BIC penalty, but with a minimum segment length of five points (corresponding to 10 days).

Yu (2020) point out that it is important for the findings of a data analysis to be stable across perturbations of the data; a similar argument underlies the stability selection proposal of Meinshausen and Bühlmann (2010). We may wish to assess stability by repeating the splitting procedure many times, and comparing the estimated and rejected changepoints across different splits of the data. However, deterministic approaches like Algorithms 2 and 3 do not lend themselves to repetition.

Generalized data thinning offers a solution to this problem. Each time the procedure is run, sampling from G_t produces a different pair of independent training and test sets. This allows us to assess stability of the procedure across any number of replicates.

Algorithm 4 (Generalized data thinning approach for changepoint detection).

1. *Indirectly thin each X_i through the function $S(x_i) \quad x_i^2$, as in Example 4.3.1 (with μ = 0). This yields $X_1^{(1)}, \quad , X_n^{(1)}$ and $X_1^{(2)}, \quad , X_n^{(2)}$, where $X_i^{(1)}, X_i^{(2)} \sim \text{Gamma} \left(\frac{1}{4}, \frac{1}{2}\right)_i$ and $X_i^{(1)}$ and $X_i^{(2)}$ are independent.*
2. *Estimate changepoints in $X_1^{(1)}, \quad , X_n^{(1)}$.*
3. *Fit a gamma GLM to test whether there is a change in the rate of $X_1^{(2)}, \quad , X_n^{(2)}$ on either side of each estimated changepoint.*

In Step 2 of Algorithm 4, we again apply the nonparametric changepoint detection method, this time with the same 10-point minimum segment length used in Algorithm 2.

We first compare the methods in a simulation study; see Supplement F for details. Figure 2 demonstrates that in the setting where there are no true changepoints, the naive approach fails to control the type 1 error rate. By contrast, both order-preserved sample splitting and generalized data thinning control the type 1 error rate. Figures S2 and S3 of Supplement F overlay the simulated data with the detected changepoints, further illustrating that the naive approach routinely mistakes noise for signal.

Turning back to the wind speed data, the top three panels of Figure 3 show the results of applying the naive, order-preserved sample splitting, and generalized data thinning approaches. To account for the effects of multiple comparisons, when testing changepoints we apply a Bonferroni correction by dividing the standard 0.05 threshold by the number of detected changepoints. We see that the naive method ṡ p-values are below the Bonferroni corrected threshold for over a third of the estimated changepoints. By contrast, the order-preserved sample splitting and generalized data thinning approaches give similar results with no rejections of the null hypothesis. In light of the results in Figure 2 and Supplement F, we believe that most of the changepoints for which we rejected the null hypothesis using the naive approach are false positives.

We now turn to the lower two panels of Figure 3 to see the advantage of the generalized data thinning approach over the order-preserved sample splitting approach. As mentioned previously, the generalized data thinning approach is amenable to a stability analysis whereas the order-preserved sample splitting approach is not. In this spirit, we repeatedly apply Algorithm 4 a total of 100 times and compare results across replicates. The fourth panel of Figure 3 displays, for each 10-day window, the percentage of replicates in which at least one changepoint was estimated using the training set. The fifth panel displays, for each 10-day window, the percentage of replicates for which there was at least one changepoint estimated using the training set *and* that estimated changepoint had a test set p-value below the Bonferroni corrected threshold. As none of the changepoints identified are consistently found to be significant, we are skeptical that they represent true changes in variance. Additional data are likely needed to draw a definitive conclusion.

8 Discussion

Our generalized data thinning proposal encompasses a diverse set of existing approaches for splitting a random variable into independent random variables, from convolution-closed data thinning (Neufeld et al., 2024a) to sample splitting (Cox, 1975). It provides a lens through which these existing approaches follow from the same simple principle — sufficiency — and can be derived through the same simple recipe (Algorithm 1).

The principle of sufficiency is key to generalized data thinning, as it enables a sampling mechanism that does not depend on unknown parameters. When no sufficient statistic that reduces the data is available, as in the non-parametric setting of Section 5.2 and the Cauchy example of Section 6.2, then sample splitting is still possible, provided that the observations are independent and identically distributed. Conversely, in a setting with $n = 1$ or where the elements of $\mathbf{X} \quad (X_1, \quad , X_n)$ are not independent and identically distributed, sample splitting may not be possible, but other generalized thinning approaches may be available.

For example, consider a regression setting with a fixed design, in which each response Y_i has a potentially distinct distribution determined by its corresponding feature vector $\mathbf{x}_i$, for $i = 1, \ldots, n$. It is typical to recast this as random pairs $(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_n, Y_n)$ that are independent and identically distributed from some joint distribution, thereby justifying sample splitting. However, this amounts to viewing the model as arising from a random design, which may not match the reality of how the design matrix was generated, and may not be well-aligned with the goals of the data analysis. For instance, recall the example given in the introduction: Given a dataset consisting of the $n = 50$ states of the United States, it is unrealistic to treat each state as an independent and identically distributed draw, and undesirable to perform inference only on the states that were "held out" of training. In this example, generalized data thinning could provide a more suitable alternative to sample splitting that stays true to the fixed design model underlying the data.

The starting place for any generalized thinning strategy—whether sample splitting or otherwise—is the assumption that the data are drawn from a distribution belonging to a family $\mathcal{P}$. An important topic of future study is the effect of model misspecification. In particular, if we falsely assume that $X \sim P \in \mathcal{P}$, what goes wrong? The random variables $X^{(1)}, \ldots, X^{(K)}$ generated by thinning will still satisfy the property $X = T(X^{(1)}, \ldots, X^{(K)})$; however, $X^{(1)}, \ldots, X^{(K)}$ may not be independent and may no longer have the intended marginals $Q^{(1)}, \ldots, Q^{(K)}$. Can we quantify the effect of the model misspecification? I.e., if the true family is "close" to the assumed family, will $X^{(1)}, \ldots, X^{(K)}$ be only weakly dependent, and will the marginals be close to $Q^{(1)}, \ldots, Q^{(K)}$? Some initial answers to these questions can be found in Neufeld et al. (2024a) and Rasines and Young (2022).

In the introduction, we noted that generalized data thinning with $K = 2$ is a (U, V)-decomposition, as defined in Rasines and Young (2022). We elaborate on that connection here. The (U, V)-decomposition seeks independent random variables $U = u(X, W)$ and $V = v(X, W)$ such that U and V are jointly sufficient for the unknowns,

where W is a random variable possibly depending on X. Suppose we can indirectly thin X through $S(\cdot)$ by $T(\cdot)$. This means we have produced independent random variables $X^{(1)}$ and $X^{(2)}$ for which $S(X) \quad T(X^{(1)}, X^{(2)})$. Since $S(X)$ is sufficient for θ on the basis of X, this implies that $(X^{(1)}, X^{(2)})$ is jointly sufficient for θ It follows that $(X^{(1)}, X^{(2)})$ is a (U, V)-decomposition of X. It is of interest to investigate whether there are (U, V)-decompositions that cannot be achieved through either direct or indirect generalized data thinning.

In Section 6.1, we provided an example of a family for which it is impossible to perform (non-trivial) thinning. In such situations, one may choose to drop the requirement of independence between $X^{(1)}$ and $X^{(2)}$. We expand on this extension in Supplement G.

The data thinning strategies outlined in this paper are implemented in the `datathin` R package, available at https://anna-neufeld.github.io/datathin/. Code to reproduce the simulation study and data analysis results are available at https://github.com/AmeerD/gdt-experiments.

Acknowledgments

We thank Nicholas Irons for identifying a problem with a previous version of the proof about Bernoulli thinning (Section 6.1).

Disclosure Statement

The authors have no relevant financial or non-financial competing interests to report.

SUPPLEMENTARY MATERIAL

Supplementary materials: Contains proofs of technical results, derivations of decompositions, and additional details on the simulation and data analyses. (PDF)

Notes

[1] Although θ is often used in the gamma distribution to denote the scale parameter, here we use it to denote the rate parameter.

References

G. Casella and R.L. Berger. *Statistical Inference*. Thomson Learning, 2002.

Shuxiao Chen and Jacob Bien. Valid inference corrected for outlier removal. *Journal of Computational and Graphical Statistics*, 29(2):323–334, 2020.

Yiqun T Chen and Daniela M Witten. Selective inference for k-means clustering. *arXiv preprint arXiv:2203.15267*, 2022.

David R Cox. A note on data-splitting for the evaluation of significance levels. *Biometrika*, 62 (2):441–444, 1975.

G. Darmois. Sur les lois de probabilité à estimation exhaustive. *Comptes Rendus de l'Académie des Sciences*, 200:1265–1266, 1935.

William Fithian, Dennis Sun, and Jonathan Taylor. Optimal inference after model selection. *arXiv preprint arXiv:1410.2597*, 2014.

Lucy L Gao, Jacob Bien, and Daniela Witten. Selective inference for hierarchical clustering. *Journal of the American Statistical Association*, 119(545):332–342, 2024.

John Haslett and Adrian E Raftery. Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 38(1):1–21, 1989.

Trevor Hastie, Robert Tibshirani, and Jerome H Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, volume 2. Springer, 2009.

Kaylea Haynes and Rebecca Killick. *changepoint.np: Methods for Nonparametric Changepoint Detection*, 2022. R package version 1.0.5.

Kaylea Haynes, Paul Fearnhead, and Idris A Eckley. A computationally efficient nonparametric approach for changepoint detection. *Statistics and computing*, 27:1293–1305, 2017.

Sangwon Hyun, Kevin Z Lin, Max G'Sell, and Ryan J Tibshirani. Post-selection inference for changepoint detection algorithms with application to copy number variation data. *Biometrics*, 77(3):1037–1049, 2021.

Sean Jewell, Paul Fearnhead, and Daniela Witten. Testing for a change in mean after changepoint detection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4): 1082–1104, 2022.

Bent Jørgensen. Exponential dispersion models and extensions: A review. *International Statistical Review/Revue Internationale de Statistique*, pages 5–20, 1992.

Bent Jørgensen and Soo Yeon Kok. Stationary time series models with exponential dispersion model margins. *Journal of Applied Probability*, 35(1):78–92, 1998.

Rebecca Killick and Idris Eckley. changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, 58(3):1–19, 2014.

B. O. Koopman. On distributions admitting a sufficient statistic. *Transactions of the American Mathematical Society*, 39(3):399–409, 1936. ISSN 00029947.

Jason D. Lee, Dennis L. Sun, Yuekai Sun, and Jonathan E. Taylor. Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3):907–927, 2016.

Erich Leo Lehmann and Joseph P Romano. *Testing Statistical Hypotheses: Third Edition*, volume 3. Springer, 2005.

James Leiner, Boyan Duan, Larry Wasserman, and Aaditya Ramdas. Data fission: splitting a single data point. *Journal of the American Statistical Association*, pages 1–12, 2023.

Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

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, 2024a.

Anna Neufeld, Lucy L Gao, Joshua Popp, Alexis Battle, and Daniela Witten. Inference after latent variable estimation for single-cell RNA sequencing data. *Biostatistics*, 25(1):270–287, 2024b.

Natalia L Oliveira, Jing Lei, and Ryan J Tibshirani. Unbiased risk estimation in the normal means problem via coupled bootstrap techniques. *arXiv preprint arXiv:2111.09447*, 2021.

Edzer J Pebesma. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30(7):683–691, 2004.

E. J. G. Pitman. Sufficient statistics and intrinsic accuracy. *Mathematical Proceedings of the Cambridge Philosophical Society*, 32(4):567–579, 1936.

D García Rasines and G A Young. Splitting strategies for post-selection inference. *Biometrika*, 12 2022. ISSN 1464-3510.

Jonathan Taylor and Robert Tibshirani. Post-selection inference for-penalized likelihood models. *Canadian Journal of Statistics*, 46(1):41–61, 2018.

Jonathan Taylor and Robert J Tibshirani. Statistical learning and selective inference. *Proceedings of the National Academy of Sciences*, 112(25):7629–7634, 2015.

Xiaoying Tian. Prediction error after model search. *The Annals of Statistics*, 48(2):763 – 784, 2020.

Xiaoying Tian and Jonathan Taylor. Asymptotics of selective inference. *Scandinavian Journal of Statistics*, 44(2):480–499, 2017.

Xiaoying Tian and Jonathan Taylor. Selective inference with a randomized response. *The Annals of Statistics*, 46(2):679–710, 2018.

Ryan J Tibshirani, Alessandro Rinaldo, Rob Tibshirani, and Larry Wasserman. Uniform asymptotic inference and the bootstrap after model selection. *The Annals of Statistics*, 46(3): 1255–1287, 2018.

Bin Yu. Veridical data science. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 4–5, 2020.

Youngjoo Yun and Rina Foygel Barber. Selective inference for clustering with unknown variance. *arXiv preprint arXiv:2301.12999*, 2023.

Changliang Zou, Guanghui Wang, and Runze Li. Consistent selection of the number of change-points via sample-splitting. *The Annals of Statistics*, 48(1):413 –439, 2020.

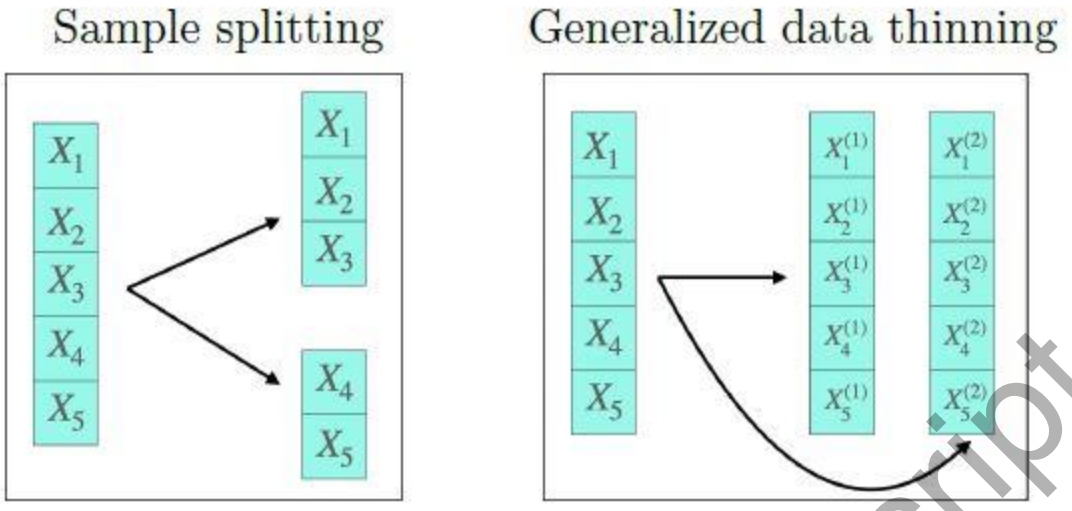

Fig. 1 *Left:* Sample splitting assigns each observation to either a training or a test set. *Right:* Generalized data thinning splits each observation into two parts that are independent and can be used to recover the original observation, i.e. $T(X^{(1)}, X^{(2)}) \quad X$.

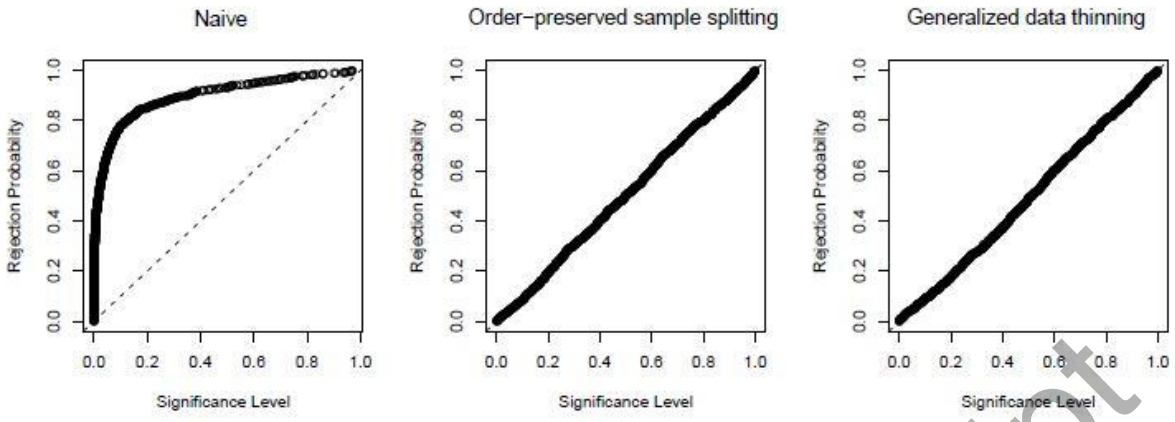

Fig. 2 Type 1 error rate of naive (Algorithm 2), order-preserved sample splitting (Algorithm 3), and generalized data thinning (Algorithm 4) approaches to testing for a change in variance, in a setting where the variance is truly constant.

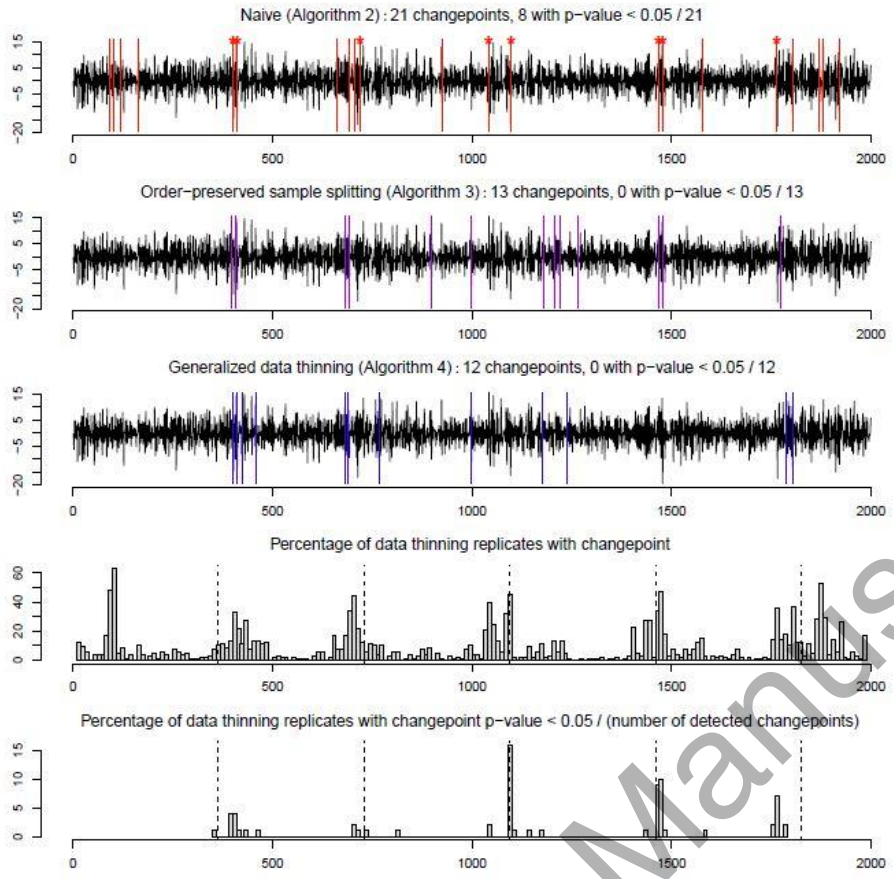

Fig. 3 Results for the wind speed analysis in Section 7. In each panel, the *x*-axis indexes the days. *First three rows:* Wind speed data over time with results of each approach (Algorithms 2, 3, and 4) overlayed: Vertical lines indicate changepoints estimated and asterisks indicate those estimated changepoints for which the computed p-value was below 0.05 divided by the number of detected changepoints. *Fourth row:* We binned the 2,000 days into 10-day windows. For each 10-day window, we display the percentage of replicates of the generalized data thinning approach for which at least one changepoint was estimated on the training set. Dashed lines are drawn every 365 days. *Fifth row:* For each 10-day window, we display the percentage of replicates of the generalized data thinning approach for which at least one changepoint was estimated on the training set *and* that estimated changepoint had a test set p-value below 0.05 divided by the number of detected changepoints.

Table 1 Examples of named families (indexed by an unknown parameter θ) that can be thinned into K components, where K is a positive integer, without knowledge of θ. In cases where they are used, ϵ_k, n_k, and ν are positive tuning parameters to be selected by the analyst, where $\sum_{k=1}^{K} \epsilon_k = 1$ and $n_1, \ldots, n_K$ are integers that sum to n; all other parameters are constrained appropriately. Note that Examples C.1, C.2, C.3, C.4, and D.1 are discussed in the supplementary materials.

Family	Distribution P, where $X \sim P$.	Distribution $Q^{(k)}$, where $X^{(k)} \overset{ind.}{\sim} Q^{(k)}$.	Sufficient statistic T (sufficient for θ)	Reference / notes
Natural exponential family (in parameter θ)	$N(\theta, \sigma^2)$	$N(\epsilon_k \theta, \epsilon_k \sigma^2)$	$\sum_{k=1}^{K} X^{(k)}$	Neufeld et al. (2024a)
	Poisson(θ)	Poisson($\epsilon_k \theta$)		
	NegBin(r, θ)	NegBin($\epsilon_k r$, θ)		
	Binomial(r, θ)	Binomial($\epsilon_k r$, θ)		
	Gamma(θ, β)	Gamma($\epsilon_k \theta$, β)		
	$N_p(\theta, \Sigma)$	$N_p(\epsilon_k \theta, \epsilon_k \Sigma)$	$\sum_{k=1}^{K} \mathbf{X}^{(k)}$	
	Multinomial$_p(r, \theta)$	Multinomial$_p(\epsilon_k r, \theta)$		
	Gamma($K/2$, θ)	$N(0, \frac{1}{2\theta})$	$\sum_{k=1}^{K} X^{(k)\,2}$	Example 3.3
	Gamma(K, θ)	Weibull($\frac{1}{\theta}$, β)	$\sum_{k=1}^{K} X^{(k)}$	Example C.1
General exponential	Beta(α, β)	Beta $\frac{1}{K}$ $\frac{k}{K}\alpha - \frac{1}{K}, \frac{1}{K}\beta$	$\sum_{k=1}^{K} X^{(k)\,1/K}$	Example 4.1

Family	Distribution P,	Distribution $Q^{(k)}$	Sufficient statistic T	Reference / notes
family (in parameter θ)				
	Beta(,)	Beta $\frac{1}{K}$, $\frac{1}{K}$ $\frac{k \ 1}{K}$	$\prod_{k \ 1}^{K} 1 \ X^{(k)}$ $^{1/K}$	Text below Example 4.1
	Gamma(,)	Gamma($\frac{1}{K}$ $\frac{k \ 1}{K}$, $\frac{1}{K}$)	$\prod_{k \ 1}^{K} X^{(k)}$ $^{1/K}$	Example 4.2
	Weibull(,)	Gamma($\frac{1}{K}$,)	$\sum_{k \ 1}^{K} X^{(k)}$ $^{1/}$	Example 4.3
	Pareto(,)	Gamma($\frac{1}{K}$,)	Exp $\sum_{k \ 1}^{K} X^{(k)}$	Example 4.3
	Dirichlet$_K$( )	Gamma($_k$,)	$X^{(1)}$, , $X^{(K)}$ $^{\top}$ / $\sum_{k \ 1}^{K} X^{(k)}$	Example C.2
	N(,)	Gamma($\frac{1}{2K}$, $\frac{1}{2}$)	$(X \quad)^2$ $\sum_{k \ 1}^{K} X^{(k)}$	Indirect only; Example 4.3
	N$_K$($_1\mathbf{1}_K$, $_2\mathbf{I}_K$)	N($_1$, $_2$)	sample mean and variance	Indirect only; Example D.1
Truncated support family	Unif(0,)	·Beta($\frac{1}{K}$,1)	max $X^{(1)}$, , $X^{(K)}$	Example 5.1
	·Beta(,1)	·Beta($\frac{}{K}$,1)		Example C.3
	Exp( )	Exp(/K)	min $X^{(1)}$, , $X^{(K)}$	Example C.4
Non-parametric	F^n	F^{n_k}	See Example 5.2	Example 5.2