Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Discussion of "Data Fission: Splitting a Single Data Point"

Anna Neufeld, Ameer Dharamshi, Lucy L. Gao, Daniela Witten & Jacob Bien

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, VOL. 120, NO. 549, 151–157: Theory and Methods Discussion
https://doi.org/10.1080/01621459.2024.2421998

Discussion of "Data Fission: Splitting a Single Data Point"

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

[a]Department of Mathematics and Statistics, Williams College, Williamstown, MA; [b]Department of Biostatistics, University of Washington, Seattle, WA; [c]Department of Statistics, University of British Columbia, Vancouver, Canada; [d]Departments of Statistics and Biostatistics, University of Washington, Seattle, WA; [e]Department of Data Sciences and Operations, University of Southern California, Los Angeles, CA

This comment was presented as part of the JSM 2024 *JASA* Theory and Methods invited session Data Fission: Splitting a Single Data Point.

1. Introduction

Sample splitting, in which some of the observations in a data sample are used as a training set for model fitting or model selection, and the remaining observations are used as a test set for validation or inference, is a fundamental tool in data analysis. We applaud Leiner et al. (2025) for their important paper, which introduces *data fission*, a broad generalization of sample splitting.

Data fission refers to any operation that splits a random variable $X \sim P_\theta$ into two pieces, which we denote here as $X^{(1)}$ and $X^{(2)}$, such that three properties hold: (a) there exists a deterministic function $T(\cdot)$ such that $X = T\left(X^{(1)}, X^{(2)}\right)$; (b) one cannot reconstruct X from $X^{(1)}$ alone; and (c) the marginal distribution of $X^{(1)}$ and the conditional distribution of $X^{(2)} \mid X^{(1)}$ are known, up to the unknown parameter θ. Leiner et al. (2025) use the term *P1 fission* for the special case in which $X^{(1)}$ and $X^{(2)}$ are independent, and *P2 fission* for when independence does not necessarily hold.[1] If X is a vector of independent observations, then sample splitting is an instance of P1 fission. However, data fission extends far beyond sample splitting, even to settings where we observe only a single scalar random variable X that we wish to partition into two pieces. Such a generalization is useful in settings where sample splitting is unsatisfying or inapplicable, such as those discussed in Leiner et al. (2025), Neufeld et al. (2024a), and Dharamshi et al. (2024b).

While P1 fission is extremely useful and easy to use, Leiner et al. (2025) provide P1 fission operations only for the Gaussian and the Poisson distributions. They provide little guidance on how to apply P2 fission operations in practice, leaving the reader unsure of how to apply data fission outside of the Gaussian and Poisson settings. The main contributions of our discussion are as follows:

1. In Section 2, we argue that P1 fission is preferable to P2 fission when both are available, both due to its simplicity but also due to a new information inequality. We also discuss how our own work (Dharamshi et al. 2024b) shows that P1 fission is available for many more distributions than the two noted by Leiner et al. (2025), and offer insight into when P1 fission is possible.
2. In Section 3, we provide guidance on how to actually apply P2 fission in practice by revisiting logistic regression. While this is a setting where P1 fission is impossible, we provide a major improvement on Leiner et al.'s (2025) treatment of logistic regression using P2 fission: we conduct inference on the parameters of interest rather than on targets of convenience.
3. In Section 4, we show that P2 fission can sometimes be interpreted as P1 fission under model misspecification. This suggests room for improvement in the way that previous authors have handled P1 fission under misspecification, and suggests an avenue for developing new P2 fission operations.

2. The Case for P1 Fission over P2 Fission

While Leiner et al. (2025) provide a very large number of P2 fission examples, they provide only two examples of P1 fission: the Gaussian location family and the Poisson. Furthermore, for these two examples, they also provide P2 fission alternatives.

This may give the reader the impression that P1 fission is rarely available, and that P2 fission is as effective as P1 fission when both are possible. In Section 2.1, we argue that P1 fission is preferable to P2 fission when both are available. In Section 2.2, we describe our recent work that establishes that P1 fission is in fact widely available.

2.1. P1 Fission Is Preferable over P2 Fission When Both Are Available

P1 fission decomposes X into independent parts, while P2 fission yields dependent parts. At an intuitive level, having independent parts may seem preferable in the name of simplicity. Indeed, in the two examples where Leiner et al. (2025) have both P1 and

CONTACT Anna Neufeld acn2@williams.edu Department of Mathematics and Statistics, Williams College, 18 Hoxsey Street, Williamstown, MA 01267.

[1]For clarity, we will use "P2 fission" to refer specifically to fission that is not P1 fission, that is the case in which $X^{(1)}$ and $X^{(2)}$ are not independent. Otherwise, "fission" and "P2 fission" would be synonymous.

P2 strategies available, they choose to only work with the P1 strategies for numerical experiments (and a P2 strategy that they carry out in the supplement involves more technical machinery such as quasi-likelihood). However, putting aside convenience, one may wonder whether there is a *statistical* advantage to the independence offered by P1 fission. The following proposition answers this question in the affirmative. For ease of presentation, we focus on a scalar-valued θ, though a similar result holds for general θ.

Proposition 1 (Fisher information allocation of P1 vs. P2). Suppose $X \sim P_\theta$, where P_θ belongs to a family $\mathcal{P}$, parameterized by $\theta \in \mathbb{R}$, for which both P1 and P2 fission operations are available. Let $(X^{(1)}, X^{(2)})$ and $(\tilde{X}^{(1)}, \tilde{X}^{(2)})$ denote the results of P1 fission and P2 fission, respectively. Suppose that the Fisher informations $I_X(\theta)$, $I_{X^{(1)}}(\theta)$, $I_{X^{(2)}}(\theta)$, and $I_{\tilde{X}^{(1)}}(\theta)$ all exist, as well as $I_{\tilde{X}^{(2)}|\tilde{X}^{(1)}}(\theta)$, which denotes the Fisher information contained in $\tilde{X}^{(2)}$ conditional on $\tilde{X}^{(1)}$ (this is a random quantity whose value depends on $\tilde{X}^{(1)}$). If the two fission strategies allocate equal information to the training set, that is $I_{X^{(1)}}(\theta) = I_{\tilde{X}^{(1)}}(\theta)$, then on the test sets,

$$\mathbb{E}_{\tilde{X}^{(1)}}\left[\left(I_{\tilde{X}^{(2)}|\tilde{X}^{(1)}}(\theta)\right)^{-1}\right] \geq \left(I_{X^{(2)}}(\theta)\right)^{-1}. \quad (1)$$

This result, which we prove in Appendix A, is directly inspired by Rasines and Young's (2023) Proposition 1 (restated in Leiner et al. (2025) as "Fact 1").[2] They use this to argue that the deterministic allocation of Fisher information of P1 fission is more efficient than sample splitting. Our result, which applies the same logic, establishes that P1 fission is at least as efficient as any P2 fission strategy. The next example illustrates that the efficiency loss can be extreme.

Remark 1 (Information allocation for Poisson). Let $X \sim \text{Poisson}(\theta)$. If we apply the P1 fission operation from Leiner et al. (2025) to X with tuning parameter ϵ, we have that $X^{(2)} \sim \text{Poisson}((1-\epsilon)\theta)$ and $\left(I_{X^{(2)}}(\theta)\right)^{-1} = \frac{\theta}{1-\epsilon}$. If we apply the P2 fission operation from the supplementary materials of Leiner et al. (2025) to X with tuning parameter τ, we have that $\tilde{X}^{(2)} \mid \tilde{X}^{(1)} \sim \text{Binomial}\left(\tilde{X}^{(1)}, \frac{\theta}{\theta+\tau}\right)$, so that $I_{\tilde{X}^{(2)}|\tilde{X}^{(1)}}(\theta) = \frac{\tau\tilde{X}^{(1)}}{\theta(\theta+\tau)^2}$. Since $\tilde{X}^{(1)} \sim \text{Poisson}(\theta+\tau)$, there is a nonzero probability that $\tilde{X}^{(1)} = 0$. Thus, $\mathbb{E}_{\tilde{X}^{(1)}}\left[\left(I_{\tilde{X}^{(2)}|\tilde{X}^{(1)}}(\theta)\right)^{-1}\right] = \infty$, that is confidence intervals for θ based on P2 fission would have infinite expected width while $\left(I_{X^{(2)}}(\theta)\right)^{-1} = \frac{\theta}{1-\epsilon}$ is finite.

Thus, in addition to the simplicity that comes with independence, there is also a statistical justification for preferring P1 fission when it is available.

2.2. A Systematic Recipe for P1 Fission

In light of the advantages of P1 fission, it is natural to wonder whether P1 fission is possible beyond the normal-location and Poisson families, and what are the underlying principles that determine whether P1 fission is possible in a particular family.

In our own work, we have answered these questions, showing that P1 fission is in fact widely available (Neufeld et al. 2024a; Dharamshi et al. 2024b). In Table 1, we show that in many of the families for which Leiner et al. (2025) provide P2 fission operations (either through their conjugate prior strategy or otherwise), P1 fission operations are also available.

Furthermore, our work elucidates the general principles for when P1 fission is possible, and how a P1 fission operation might be constructed. We first define *data thinning*, which is a K-fold generalization of P1 fission (which itself is equivalent to the "(U, V)-decomposition" defined by Rasines and Young 2023).

Definition 1 (Data thinning, Dharamshi et al. (2024b)). Consider a family of distributions $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. We say that this family is *thinned* by a function $T(\cdot)$ if there exists a distribution G_t, not depending on θ, such that when we draw $\left(X^{(1)}, X^{(2)}, \ldots, X^{(K)}\right) \mid X \sim G_X$, it holds that:

1. $X = T\left(X^{(1)}, \ldots, X^{(K)}\right)$.
2. $\left(X^{(1)}, \ldots, X^{(K)}\right)$ are mutually independent with distributions that depend on θ.

With this definition in place, Dharamshi et al.'s (2024b) Theorem 1 shows that sufficiency is a key ingredient in making data thinning possible. As seen in Table 1, this framework allows us to define P1 fission operations (equivalently, thinning operations) for a wide variety of settings, both within exponential families and beyond.

In light of Section 2.1, Table 1 shows that, for many of the distributions considered in Leiner et al. (2025), P1 fission operations are available and preferable. This may leave one to wonder whether P2 fission should ever be used in practice. As highlighted in the table, Dharamshi et al. (2024b) and Dharamshi et al. (2024a) establish that there are situations in which P1 fission is unavailable or impossible; in these cases, P2 fission is indeed important in practice. The remainder of our discussion focuses on these cases.

3. Improving P2 Fission for Logistic Regression

In their main text, Leiner et al. (2025) do not offer concrete guidance on applying P2 fission in practical contexts. However, Section E.4 of their supplementary materials does provide one case study of applying P2 fission to a concrete problem, namely logistic regression. This is an important example for P2 fission since, as noted in Table 1, Dharamshi et al. (2024b) prove that P1 fission is impossible in the Bernoulli family.

In Section 3.1, we review their proposal and point out a major shortcoming: they are not able to conduct inference on the parameters of interest. Then, in Section 3.2, we show that we can improve their example, establishing more persuasively the practical usefulness of P2 fission.

[2] As noted in Rasines and Young (2023), "optimality of the Fisher information is commonly measured through summary statistics of its inverse."

Table 1. The parameter θ (likewise θ_1 and θ_2) is always unknown (i.e., the fission operation must not rely on it) while all other parameters are known.

	Distribution P_θ	P1	P2
	$N(\theta, \sigma^2)$	Rasines and Young (2023)	
	$N_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$	Rasines and Young (2023)	
	Poisson(θ)	Leiner et al. (2025)	
	NegBin(r, θ)	⋮	⋮
P1 and P2 available, P1 preferable.	Binomial(r, θ)		Leiner et al. (2025)
	Exp(θ)	Neufeld et al. (2024a)	
	Gamma(α, θ)		⋮
		⋮	
	Gamma(θ, β)		
	Dirichlet(θ, ϕ)	⋮	
	Exponential family	Dharamshi et al. (2024b)	
		⋮	
	Multinomial$_p(r, \boldsymbol{\theta})$	Neufeld et al. (2024a)	
	Beta(θ, β)		
	Beta(α, θ)		
	Weibull(θ, γ)	⋮	⋮
P1 available, P2 not yet considered.	Pareto(γ, θ)		
	N(μ, θ)	Dharamshi et al. (2024b)	Not yet considered
	Unif($0, \theta$)	⋮	⋮
	$\theta \cdot$ Beta(α, 1)		
	θ + Exp(λ)		
	Bernoulli(θ)	Impossible (Dharamshi et al. (2024b))	
	Categorical(θ)	Impossible (Dharamshi et al. (2024b))	⋮
P2 available, P1 impossible or not yet considered.	$N_1(\theta_1, \theta_2)$	Impossible (Dharamshi et al. (2024a))	Leiner et al. (2025)
	Gamma(θ_1, θ_2)	Not yet considered	
	NegBin(θ_1, θ_2)	Not yet considered	⋮
	Binomial(θ_1, θ_2)	Not yet considered	

NOTE: For each family, we include a citation to the authors who, to the best of our knowledge, first proposed the decomposition as a general alternative to sample splitting. We have omitted citations to authors who proposed the decompositions for specific tasks. For example, the P1 decomposition of the Poisson follows from a much-used thinning property (see e.g., Chen et al. (2021) and Sarkar and Stephens's (2021) use for specific tasks related to model validation, or Neufeld et al.'s (2024b) use for a specific task related to inference). Similarly, Tian and Taylor (2018) and Tian (2020) use the $N(\theta, \sigma^2)$ decomposition for specific tasks. Finally, Joe (1996) use the natural exponential family P1 decompositions as generative models for time series. Finally, we note that this table is not exhaustive; some of the distributions considered in Dharamshi et al. (2024b) are omitted here for brevity.

3.1. Leiner et al.'s (2025) Proposal for Logistic Regression

In Section A of their supplementary materials, Leiner et al. (2025) describe a P2 fission of the Bernoulli distribution, which for reference we restate here.

Example 1 (P2 fission of a Bernoulli (Leiner et al. 2025)). We observe $Y \sim \text{Bernoulli}(\theta)$. For a tuning parameter ϵ, sample $Z \sim \text{Bernoulli}(\epsilon)$ and then let $Y^{(1)} = (1-Z)Y + Z(1-Y)$ and $Y^{(2)} = Y$, which yields

$$Y^{(1)} \sim \text{Bernoulli}(\theta + \epsilon - 2\theta\epsilon),$$
$$Y^{(2)} \mid Y^{(1)} \sim \text{Bernoulli}\left(\frac{\theta}{\theta + (1-\theta)(\frac{\epsilon}{1-\epsilon})^{2Y^{(1)}-1}}\right). \quad (2)$$

In Section E.4 of the supplementary materials, Leiner et al. (2025) apply this decomposition to logistic regression. We briefly describe this example below.

Example 2 (Inference for logistic regression using P2 fission of a Bernoulli). Let $X \in \mathbb{R}^{n \times p}$ be a (fixed) matrix of covariates with rows denoted x_i, and let $Y \in \{0,1\}^n$ be a vector of binary responses. Assume that

$$Y_i \sim \text{Bernoulli}(\theta_i), \text{ where } \theta_i = \frac{\exp(\beta^T x_i)}{1 + \exp(\beta^T x_i)}, \quad (3)$$

and $\beta \in \mathbb{R}^p$ is sparse. Leiner et al. (2025) suggest the following workflow for inference on the nonzero coefficients.

1. For $i = 1, \ldots, n$, apply the decomposition in Example 1 to split Y_i into $Y_i^{(1)}$ and $Y_i^{(2)}$.
2. Select variables $\mathcal{S} \subseteq \{1, \ldots, p\}$ by running standard logistic lasso software (with cross-validation to select the penalty parameter) on $\left\{(x_1, Y_1^{(1)}), \ldots, (x_n, Y_n^{(1)})\right\}$.
3. Fit a standard logistic GLM on $\left\{(x_{1S}, Y_1^{(2)}), \ldots, (x_{nS}, Y_n^{(2)})\right\}$, where x_{iS} denotes the ith observation subset to the variables in S. This is a misspecified model, since (2) implies that the log odds of the mean of $Y^{(2)}$ is not linear in the covariates. Thus, use sandwich standard errors for inference to obtain valid confidence intervals for the parameters that minimize the KL-divergence between the "working model" (assumed by the use of a standard logistic GLM) and the conditional distribution in (2).

We note that the parameters that are targeted in Step 3 of their approach are *not* the same parameters that appear in the original model (3). This is a serious practical limitation of the approach, as we demonstrate here. We generate 2000 datasets with $n = 500$ and $p = 50$ where $\beta_0 = 0.6$ and $\beta_j = 0$ for $j = 1, \ldots, p$. Note that $Y_i \sim \text{Bernoulli}(0.6)$ for $i = 1, \ldots, n$, and so none of the covariates contribute to the data generating mechanism. For each dataset, we carry out the three-step process described in Example 2, which follows Section E.4 of Leiner et al. (2025). Figure 1 shows that the p-values for the selected variables do not follow a uniform distribution, even though $\beta_1 = \cdots = \beta_p = 0$. Thus, Type 1 error of the selected variables is not controlled.

Table 2. Coverage, selection probability, and conditional power for the method in Example 3, for the simulation described in Section 3.2.

j	Coverage of β_j	Proportion of datasets w/ X_j selected	Proportion of datasets w/ X_j selected where $H_0 : \beta_j = 0$ is rejected
1	0.95	0.32	0.88
2	0.94	0.94	1.00
3	0.94	0.74	0.99
Average over all others	0.95	0.02	0.05

NOTE: The proportion of datasets with X_j selected (for the non-null variables) is a function of how much information is in the training set, whereas the proportion of these datasets for which we reject the null hypothesis that $\beta_j = 0$ is a function of the information in the test set. Thus, we can trade off the performance in these columns by changing the tuning parameter ϵ in Example 1: in these simulations, we use $\epsilon = 0.8$.

As this is the only example of P2 fission given in Leiner et al. (2025), it might appear that P2 fission never lends itself to inference on the parameters of interest. However, we show in the next section that this is not the case. With care, we are able to apply this same decomposition of the Bernoulli for inference on the parameter of interest β.

3.2. A Better Way To Do P2 Fission for Logistic Regression

The problem in Example 2 does not lie in the Bernoulli P2 fission operation in Step 1 (see Example 1): instead, the problem lies in the fact that Step 3 uses a likelihood derived from the marginal distribution of $Y^{(2)}$ rather than the conditional distribution of $Y^{(2)} \mid Y^{(1)}$.

To address this, we re-consider Step 3 of Example 2. It turns out that (2) implies that

$$\log\left(\frac{\Pr\left(Y^{(2)} = 1 \mid Y^{(1)}\right)}{\Pr\left(Y^{(2)} = 0 \mid Y^{(1)}\right)}\right) = \begin{cases} \log\left(\frac{1-\epsilon}{\epsilon}\right) + \beta_0 + \beta^T x_i & \text{if } Y^{(1)} = 1, \\ \log\left(\frac{\epsilon}{1-\epsilon}\right) + \beta_0 + \beta^T x_i & \text{if } Y^{(1)} = 0. \end{cases}$$

Thus, we can improve on Step 3, as shown in the following example.

Example 3 (Improved inference for logistic regression using P2 fission of a Bernoulli). In the setting of Example 2, conduct Steps 1 and 2. Replace Step 3 with the following:

3. Fit a standard logistic GLM to $\left\{(x_{1S}, Y_1^{(2)}), \ldots, (x_{nS}, Y_n^{(2)})\right\}$, with an offset that equals $\log\left(\frac{\epsilon}{1-\epsilon}\right)$ if $Y_i^{(1)} = 0$, and $\log\left(\frac{1-\epsilon}{\epsilon}\right)$ otherwise.

This method makes use of the conditional distribution of $Y^{(2)} \mid Y^{(1)}$, and therefore yields valid confidence intervals for the original parameters.

Figure 1 shows that this leads to uniform p-values for the selected coefficients.

We next generate 2000 datasets with $\beta_0 = 0.6$, $\beta_1 = -0.9$, $\beta_2 = 2.1$, $\beta_3 = -1.5$, and $\beta_4 = \cdots = \beta_p = 0$. Table 2 shows that the standard (non-sandwich) logistic regression confidence intervals from Example 3 achieve nominal coverage for each coefficient whenever it is selected. Table 2 also shows that the method from Example 3 tends to select the truly important variables for the model, and has high power to determine that these variables are significant when they are selected.

Thus, the conditional distribution of $Y^{(2)} \mid Y^{(1)}$ under the Bernoulli P2 fission operation in Example 1 is quite tractable for logistic regression, and allows for valid post-selective inference using standard software. This provides a proof of concept that P2 fission is a powerful tool that may be useful in practice. We believe that Example 3 makes this case more persuasively than the example in Leiner et al.'s (2025) supplement (re-stated as our Example 2), as in the latter example the parameters of interest are not the targets of inference. In the next section, we explore additional possible uses for P2 fission.

Scripts to reproduce the numerical results in this section can be found at *https://github.com/anna-neufeld/fission_logistic/*.

4. P2 Fission as P1 Fission under Misspecification

In this section, we argue that directly working with the conditional distribution of $X^{(2)} \mid X^{(1)}$ provides room for improvement in the way that previous authors have treated the setting of P1 fission under model misspecification.

4.1. Misspecified P1 Fission of the Gaussian Distribution

We first recall the P1 fission operation for the Gaussian distribution with known variance from Neufeld et al. (2024a). This is

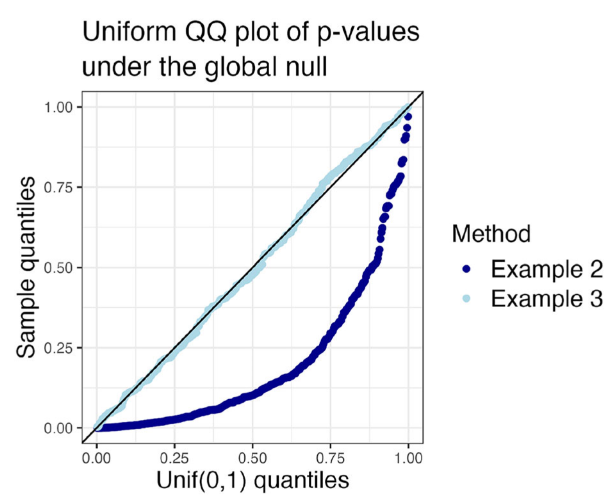

Figure 1. Uniform QQ-plot of p-values for all selected variables across all datasets under the global null, for the simulation described in Section 3. While the method of Example 2 (taken from Leiner et al. (2025)) does not control the Type 1 error, our proposal in Example 3 achieves uniformly distributed p-values by directly working with the conditional distribution.

equivalent (up to a constant scaling) to the operations of Leiner et al. (2025) and Rasines and Young (2023).

Example 4 (P1 fission of the Gaussian with known variance). We observe $X \sim \mathrm{N}\left(\theta, \sigma^2\right)$. For a tuning parameter ϵ, we let

$$\begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \Bigg| X = x \sim \mathrm{N}\left(\begin{bmatrix} \epsilon x \\ (1-\epsilon)x \end{bmatrix}, \begin{bmatrix} \epsilon(1-\epsilon)\sigma^2 & -\epsilon(1-\epsilon)\sigma^2 \\ -\epsilon(1-\epsilon)\sigma^2 & \epsilon(1-\epsilon)\sigma^2 \end{bmatrix}\right), \tag{4}$$

which yields

$$\begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \epsilon\theta \\ (1-\epsilon)\theta \end{bmatrix}, \begin{bmatrix} \epsilon\sigma^2 & 0 \\ 0 & (1-\epsilon)\sigma^2 \end{bmatrix}\right). \tag{5}$$

If σ^2 is unknown, then we cannot draw from the distribution in (4). The question of what to do if σ^2 is unknown has been considered by numerous authors. Both Rasines and Young (2023) and Leiner et al. (2025) propose applying (4) with an estimate $\hat{\sigma}^2$, and then treating the resulting folds $X^{(1)}$ and $X^{(2)}$ as though they were independent; both sets of authors justify this approach using an asymptotic argument. We refer to such an approach as "approximate P1 fission". However, since data fission is most useful when the number of observations n is small (or even $n = 1$), estimating σ^2 from the data is unlikely to be fruitful in practical settings, and the asymptotic arguments of Rasines and Young (2023) and Leiner et al. (2025) are unlikely to apply. Thus, we feel that "approximate P1 fission" leaves something to be desired.

Proposition 10 of Neufeld et al. (2024a) provides the following finite sample result, which quantifies the magnitude of the correlation between $X^{(2)}$ and $X^{(1)}$ when we apply (4) with a "guess" of σ^2.

Example 5 (Gaussian P1 fission under misspecification). Suppose that we observe $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, with both μ and σ^2 unknown. For a $\tilde{\sigma}^2$ that is not a function of X, sampling

$$\begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \Bigg| X = x \sim \mathrm{N}\left(\begin{bmatrix} \epsilon x \\ (1-\epsilon)x \end{bmatrix}, \begin{bmatrix} \epsilon(1-\epsilon)\tilde{\sigma}^2 & -\epsilon(1-\epsilon)\tilde{\sigma}^2 \\ -\epsilon(1-\epsilon)\tilde{\sigma}^2 & \epsilon(1-\epsilon)\tilde{\sigma}^2 \end{bmatrix}\right) \tag{6}$$

yields

$$\begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \epsilon\mu \\ (1-\epsilon)\mu \end{bmatrix}, \begin{bmatrix} \epsilon^2\sigma^2 + \epsilon(1-\epsilon)\tilde{\sigma}^2 & \epsilon(1-\epsilon)(\sigma^2 - \tilde{\sigma}^2) \\ \epsilon(1-\epsilon)(\sigma^2 - \tilde{\sigma}^2) & (1-\epsilon)^2\sigma^2 + \epsilon(1-\epsilon)\tilde{\sigma}^2 \end{bmatrix}\right). \tag{7}$$

On the basis of this result, Neufeld et al. (2024a) argue that as long as σ^2 and $\tilde{\sigma}^2$ are "close", the dependence between folds can be ignored. Of course, in practice, it is not clear how an analyst might obtain an accurate "guess" for the error variance, without using the data!

Notably, up to small changes in scaling and notation, Example 5 is also presented in Section A of the supplementary materials of Leiner et al. (2025) as a Gaussian P2 fission operation for unknown σ^2. However, they do not elaborate on its use.

In Dharamshi et al. (2024a), we show that interpreting (7) as P2 fission provides a valuable toolset for post-selection inference or model validation in the setting of a Gaussian with unknown variance, especially when n is small or $n = 1$. In particular, we work with the conditional distribution of $X^{(2)} \mid X^{(1)}$, rather than ignoring this dependence, as was done by previous authors (Leiner et al. 2025; Rasines and Young 2023; Neufeld et al. 2024a). This conditional distribution poses a number of statistical and computational challenges, which we address. We further explore topics such as the expected allocation of Fisher information between $X^{(1)}$ and $X^{(2)} \mid X^{(1)}$, which depends both on ϵ and the degree of misspecification. As in Section 3.2, we show that Gaussian P2 fission enables inference on parameters of interest using relatively standard techniques.

As far as we know, Dharamshi et al. (2024a) are the first to propose using P2 fission as a remedy to P1 fission under misspecification. In the remainder of this section, we show that this idea can be fruitfully applied far beyond the Gaussian setting.

4.2. Misspecified P1 Fission of the Negative Binomial Distribution

We first restate the P1 fission operation for the negative binomial distribution with known overdispersion from Neufeld et al. (2023) and Neufeld et al. (2024a).

Example 6 (The negative binomial with known overdispersion). We observe $X \sim \mathrm{NB}(r, \theta)$. If r is known, then Neufeld et al. (2024a) propose to sample $X^{(1)}$ and $X^{(2)}$ as

$$\left(X^{(1)}, X^{(2)}\right) \Big| X = x \sim \mathrm{DirichletMultinomial}\left(x, \epsilon r, (1-\epsilon) r\right), \tag{8}$$

which yields $X^{(1)} \sim \mathrm{NB}(\epsilon r, \theta)$, $X^{(2)} \sim \mathrm{NB}((1-\epsilon)r, \theta)$, and $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

If the overdispersion parameter r is unknown, then we cannot draw from the distribution in (8). Neufeld et al. (2023) suggest that one can apply (8) with an estimate of r and treat the resulting folds as independent, and Neufeld et al. (2024a) provide a finite sample result analogous to Example 5 that quantifies the covariance between the folds when a non-data-driven "guess" $\tilde{r}$ is used.

In the setting of Example 6, suppose that we plug in a very particular "guess" $\tilde{r}$ in place of r: we take $\tilde{r} \to \infty$. That is, we perform P1 fission for the incorrect family $\tilde{\mathcal{P}} = \{\lim_{r\to\infty} \mathrm{NB}(r, \theta) : \theta \in (0,1), r\frac{1-\theta}{\theta} = \mu\} = \{\mathrm{Poisson}(\mu) : \mu \in (0, \infty)\}$.

Example 7 (Applying Poisson P1 fission under misspecification). We observe $X \sim NB(r, \theta)$, where both r and θ are unknown. Apply Poisson P1 fission to X: that is, sample

$$\left(X^{(1)}, X^{(2)}\right) \Big| X = x \sim \mathrm{Multinomial}\left(x, \epsilon, (1-\epsilon)\right).$$

This yields $X^{(1)} \sim \mathrm{NB}\left(r, \frac{\theta}{\theta + \epsilon - \epsilon\theta}\right)$ and $X^{(2)} \mid X^{(1)} \sim \mathrm{NB}\left(r + X^{(1)}, \theta + \epsilon - \epsilon\theta\right)$. This is identical to the P2 fission operation for the negative binomial given in Appendix A of Leiner et al. (2025).

Thus, the negative binomial P2 fission operation of Leiner et al. (2025) can be viewed as applying Poisson P1 fission to a negative binomial random variable. It can therefore be interpreted as misspecified P1 fission.

Neufeld et al. (2024b) and Neufeld et al. (2023) explore the possibility of applying Poisson P1 fission to data that follows a negative binomial distribution, but they ignore the resulting dependence between folds of data, and suggest that this is appropriate if the amount of overdispersion in the data is thought to be mild. Negative binomial P2 fission offers an alternative path. However, the conditional distribution of $X^{(2)} \mid X^{(1)}$ takes a complicated form, which may pose a challenge in its application to post-selection inference or model validation. Understanding how to work with this conditional distribution in practice is a topic for future work.

It is clear from Table 1 that there are many distributions beyond the Gaussian and the negative binomial for which the available P1 fission operation requires knowledge of some of the parameters of the model. Using the strategies presented in this section, we can see that it is always possible to come up with a P2 fission operation that does not require knowledge of any parameters: we can just take the P1 recipe and plug in a "guess." This strategy will always result in dependent folds, and thus can always be seen as P2 fission. Thus, in this section, we have seen both a new way to deal with P1 fission under misspecification, as well as a possible future avenue for developing new P2 fission decompositions. For example, one could apply the P1 recipe for $\text{Beta}(\theta, \beta)$ referred to in Table 1 with a guess for β to develop a valid P2 fission strategy for $\{\text{Beta}(\theta_1, \theta_2) : \theta_1, \theta_2 > 0\}$, a family that does not yet appear in the table.

5. Discussion

We applaud Leiner et al. (2025) once again for their important work. Decompositions of a single random variable have already proved useful in a variety of applications far beyond those mentioned in Leiner et al. (2025) (see e.g., Tian 2020; Chen et al. 2021; Oliveira, Lei, and Tibshirani 2022; Neufeld et al. 2023, 2024b), and we have no doubt that the contributions of Leiner et al. (2025) will further increase the scope of application.

In Section 3, we saw that in the setting of logistic regression, operating on the conditional distribution of $Y^{(2)} \mid Y^{(1)}$ arising from P2 fission is actually quite tractable. This raises the following question: while many of the conditional distributions arising from the P2 fission operations proposed in Leiner et al. (2025) appear to be difficult to work with, might some of them be simpler than they seem? For instance, Perry et al. (2024) consider a Gaussian-Laplace setting in which $Y^{(2)} \mid Y^{(1)}$ is complicated, but conditioning on additional information (beyond $Y^{(1)}$) leads to tractable inference. Similar strategies may prove fruitful for other P2 fission decompositions.

Overall, while we would not choose P2 fission over P1 fission in a setting where both are available, we believe that P2 fission is a valuable tool whose potential is far broader than the examples considered in Leiner et al. (2025). We look forward to seeing additional applications of P2 fission in the future.

Appendix A: Proof of Proposition 1

As noted in Section 2.3 of Leiner et al. (2025), $I_X(\theta) = I_{\tilde{X}^{(1)}}(\theta) + \mathbb{E}_{\tilde{X}^{(1)}}[I_{\tilde{X}^{(2)}|\tilde{X}^{(1)}}(\theta)]$. For P1-fission, the same logic applies, but with $X^{(1)}$ and $X^{(2)}$ independent, this reduces to $I_X(\theta) = I_{X^{(1)}}(\theta) + I_{X^{(2)}}(\theta)$ (recovering Proposition 2 of Dharamshi et al. (2024b)). Setting these two decompositions of $I_X(\theta)$ equal to each other and recalling the assumption that the training sets have equal information, $I_{X^{(1)}}(\theta) = I_{\tilde{X}^{(1)}}(\theta)$, we get that

$$\mathbb{E}_{\tilde{X}^{(1)}}[I_{\tilde{X}^{(2)}|\tilde{X}^{(1)}}(\theta)] = I_{X^{(2)}}(\theta).$$

This equality together with Jensen's inequality then implies the result:

$$\mathbb{E}_{\tilde{X}^{(1)}}[\{I_{\tilde{X}^{(2)}|\tilde{X}^{(1)}}(\theta)\}^{-1}] \geq \{E_{\tilde{X}^{(1)}}[I_{\tilde{X}^{(2)}|\tilde{X}^{(1)}}(\theta)]\}^{-1} = I_{X^{(2)}}(\theta)^{-1}.$$

Acknowledgments

We thank Daniel Kessler, Ethan Ancell, and Ronan Perry for helpful conversations that informed some of the ideas in this discussion.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

DW was partially supported by ONR N00014-23-1-2589, NSF DMS 2322920, a Simons Investigator Award in Mathematical Modeling of Living Systems, and NIH 5P30DA048736. LG was supported by an NSERC Discovery Grant. AD was supported by an NSERC Postgraduate Scholarships-Doctoral.

ORCID

Ameer Dharamshi http://orcid.org/0000-0002-5505-4765

References

Chen, F., Roch, S., Rohe, K., and Yu, S. (2021), "Estimating Graph Dimension with Cross-Validated Eigenvalues," arXiv preprint arXiv:2108.03336. [153,156]

Dharamshi, A., Neufeld, A., Gao, L. L., Bien, J., and Witten, D. (2024), "Decomposing Gaussians with Unknown Covariance," arXiv preprint arXiv:2409.11497. [152,153,155]

Dharamshi, A., Neufeld, A., Motwani, K., Gao, L. L., Witten, D., and Bien, J. (2024b), "Generalized Data Thinning Using Sufficient Statistics," *Journal of the American Statistical Association*, 1–26, just-accepted. [151,152,153,156]

Joe, H. (1996), "Time Series Models with Univariate Margins in the Convolution-Closed Infinitely Divisible Class," *Journal of Applied Probability*, 33, 664–677. [153]

Leiner, J., Duan, B., Wasserman, L., and Ramdas, A. (2025), "Data Fission: Splitting a Single Data Point," *Journal of the American Statistical Association*, this issue, DOI: 10.1080/01621459.2023.2270748. [151,152,153,154,155,156]

Neufeld, A., Popp, J., Gao, L. L., Battle, A., and Witten, D. (2023), "Negative Binomial Count Splitting for Single-Cell RNA Sequencing Data," arXiv preprint arXiv:2307.12985. [155,156]

Neufeld, A., Dharamshi, A., Gao, L. L., and Witten, D. (2024), "Data Thinning for Convolution-Closed Distributions," *Journal of Machine Learning Research*, 25, 1–35. [151,152,153,154,155]

Neufeld, A., Gao, L. L., Popp, J., Battle, A., and Witten, D. (2024b), "Inference After Latent Variable Estimation for Single-Cell RNA Sequencing Data," *Biostatistics*, 25, 270–287. [153,156]

Oliveira, N. L., Lei, J., and Tibshirani, R. J. (2022), "Unbiased Test Error Estimation in the Poisson Means Problem via Coupled Bootstrap Techniques," arXiv preprint arXiv:2212.01943. [156]

Perry, R., Xu, Z., McGough, O., and Witten, D. (2024), "Infer-and-Widen Versus Split-and-Condition: Two Tales of Selective Inference," arXiv preprint arXiv:2408.06323. [156]

Rasines, D. G., and Young, G. A. (2023), "Splitting Strategies for Post-Selection Inference," *Biometrika*, 110, 597–614. [152,153,155]

Sarkar, A., and Stephens, M. (2021), "Separating Measurement and Expression Models Clarifies Confusion in Single-Cell RNA Sequencing Analysis," *Nature Genetics*, 53, 770–777. [153]

Tian, X. (2020), "Prediction Error After Model Search," *The Annals of Statistics*, 48, 763–784. [153,156]

Tian, X., and Taylor, J. (2018), "Selective Inference with a Randomized Response," *The Annals of Statistics*, 46, 679–710. [153]