Applied Stochastic Models in Business and Industry

WILEY

RESEARCH ARTICLE

Information About the Moments or the Likelihood Model Parameters? A Chicken and Egg Problem

Omid M. Ardakani[1] | Majid Asadi[2] | Ehsan S. Soofi[3]

[1]Georgia Southern University, Savannah, Georgia, USA | [2]University of Isfahan, Isfahan, Iran | [3]University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA

Correspondence: Ehsan S. Soofi (esoofi@uwm.edu)

Received: 30 May 2024 | **Revised:** 11 January 2025 | **Accepted:** 13 January 2025

Funding: The authors received no specific funding for this work.

Keywords: entropy | failure rate | Lindley-Singpurwalla model | reparametrization | sample information

ABSTRACT

This article compares the information content of a sample for two competing Bayesian approaches. One approach follows Dennis Lindley's Bayesian standpoint, where one begins by formulating a prior for a parameter related to the problem in question and incorporates a likelihood to transition to a posterior. This contrasts with the usual Bayesian approach, where one starts with a likelihood model, formulates a prior distribution for its parameters, and derives the corresponding posterior. In both cases, the sample information content is measured using the difference between the prior and posterior entropies. We investigate this contrast in the context of learning about the moments of a variable. The maximum entropy principle is used to construct the likelihood model consistent with the given moment parameters. This likelihood model is then combined with the prior information on the parameters to derive the posterior. The model parameters are the Lagrange multipliers for the moment constraints. A prior for the moments induces a prior for the model parameters; however, the data provides differing amounts of information about them. The results obtained for several problems show that the information content using the two formulations can differ significantly. Additional information measures are derived to assess the effects of operating environments on the lifetimes of system components.

1 | Introduction

Irony and Singpurwalla [1] report a written dialogue consisting of 44 questions on non-informative priors with José Bernardo, followed by discussions written by D. R. Cox, A. P. Dawid, J. K. Ghosh, and Dennis Lindley.[12] In his discussion, Lindley [2] described the "Bayesian standpoint" by beginning with a parameter θ related to the problem in question, formulating a prior, incorporating a likelihood, and passing to a posterior. He contrasts this standpoint with the usual Bayesian approach that begins with a likelihood model and specifies a prior for its parameters.

Often the research question is about some moments of a variable. We denote the variable of interest by Y, which varies according to a density function on a support $\mathbb{Y}$. Our quantities of interest are the following moment parameters:

$$\theta_k = E_f\left[T_k(Y|\theta)\right], \quad k = 1, \cdots, K, \tag{1}$$

where $\theta = \left(\theta_1, \ldots, \theta_K\right)^T$ is a vector of unknown finite quantities and T_k's are various *types of moments* of a probability density or mass function (PDF) $f(y|\theta)$. The maximum entropy (ME) principle [3, 4] provides likelihood models parameterized by the Lagrange multipliers for the research moment constraints (Equation (1)). The mapping between the moments and the ME likelihood parameters is invertible, but the information provided by the data about them is not equal. The uncertainty about the unknown θ is described by a PDF $g(\theta)$. We explore the relationship between the utility of learning from the data about the moment parameters and about the likelihood model's parameters. Following Abel and Singpurwalla [5], we measure the utility of data for learning about the parameters in terms of sample information, defined by Lindley [6] as the difference between the entropies of the prior distribution $g(\theta)$ and the posterior distribution $g(\theta|y) \propto f(y|\theta)g(\theta)$.

This paper contributes to the Bayesian information literature by computing and comparing the sample information about the moment parameters with the information about parameters of the likelihood model assumed in the usual Bayesian approach (and other quantities). New contributions include the elaboration of Abel and Singpurwalla [5], applications of Bayesian information measures, and introduction of information measures for the mean number of failures and the failure rate of the geometric distribution, the mean and variance of the normal distribution jointly, and the Lindley and Singpurwalla's [7] distribution of the lifetimes of components in the operation environment. The new information measures extend the Bayesian information literature and provide insights into some subtle issues.

Abel and Singpurwalla [5] addressed the question of "To survive or to fail" assuming the exponential model. They compared the sample information provided by the survival and failure observations at a given time $y = y_0$ about the mean and the failure rate function, respectively. Ebrahimi et al. [8] addressed the same question for several failure time models. The likelihood models to study the question of survival or failure are different; one is based on the survival function for the survival data, and the other is based on the PDF for the failure time data. Our problem is comparison of the information about the moments and the likelihood model parameters (and other parameters such as the failure rate) when all observations are failure times. This is the problem of different parameterizations of the same likelihood model. However, our expressions for the sample information about the mean time to failure and the failure rate shed some light on the conclusions of Abel and Singpurwalla's [5].

Lindley [9] proposed the posterior information rule, defined by the negative entropy, for binomial sampling for learning about the Bernoulli parameter π and $\xi = \log(\pi/(1-\pi))$. Posterior measures include the prior distribution and data information. Only with the uniform prior for π is the posterior information rule equivalent to Lindley's [6] sample information (difference between the prior and posterior entropies). Our analysis illustrates how, with the uniform prior for π, the posterior information differs from the sample information about the log-odds parameter ξ. We also use Jeffreys' prior for π to reveal the effects of the prior on the posterior information rule. We then introduce sample information about the mean number of failures and the failure rate of the geometric distribution.

The sample information about the normal mean, conditional on given variance, is well known. Recently, Ardakani et al. [10] computed the expected prior predictive information measure about the mean and variance jointly which, unlike the sample information, is invariant under reparameterizations. We introduce the sample information measures for the unknown mean and variance jointly, and for the mean and the precision parameter jointly, using the usual dependent normal-gamma prior. We also study the sample information about the normal parameters, individually and jointly, using the independent normal-gamma prior via MCMC and a kernel estimate of entropy developed in Pfughoeft et al. [11].

Lindley and Singpurwalla [7] derived the joint predictive survival function of components of systems operating in an environment, where the failure rates of components are different from those in the test environment. They motivated their model in contrast with the unwieldy distribution of the lifetime of the parallel system. However, the Lindley-Singpurwalla model does not depend on the system structure. We offer a few information measures for assessing the effects of the environment on the distribution of a system's components. Currit and Singpurwalla [12] studied the reliability properties of the distribution of the parallel system in the operating environment and proposed a Bayesian procedure for the parameter of system with two identical components, where the posterior is not in a closed form. They also briefly presented the reliability function of the series system in the operating environment. We also offer a few information measures for assessing the effects of operation environment on the lifetime distribution of the series system.

This paper is organized as follows: Section 2 defines the notations and information measures used in this paper. Section 3 compares sample information measures about the mean time to failure and the failure rate. Section 4 compares sample information measures about the mean of a binary variable and a log-odds parameter. This section also compares the sample information measures about the mean and the failure rate of the geometric distribution. Section 5 presents bivariate sample information measures about the mean and variance of a variable. Section 6 offers information measures for assessing the effects of the operating environment on the lifetimes of components in systems. Section 7 concludes the paper.

2 | Preliminaries

The entropy of a parameter θ (or its distribution g) can be generally written as:

$$H(\theta) = H(g) = -\int_{\Theta} g(\theta) \log g(\theta) d\nu(\theta),$$

where, for the discrete case $d\nu(\theta) = 1$ and the integral becomes a summation, and for the continuous case, g is a Lebesgue PDF with $d\nu(\theta) = d\theta$.

The observed *sample information* is defined by

$$\Delta(\theta; y) = H(\theta) - H(\theta|y). \tag{2}$$

Lindley [6] defined this measure in terms of the information measure of g, given by $I(g) = -H(g)$. This measure can be positive or negative depending on which of the two distributions is more concentrated. $\Delta(\theta; y) < 0$ indicates a "surprising" value of y [6]. The sampling rule of $\Delta(\theta; y)$ to learn about θ is as follows: continue sampling until $\Delta(\theta; y)$ has attained a prescribed value.

Let $\lambda = \xi(\theta)$, where $\xi : \Re^K \to \Re^K$ is one-to-one. The uncertainty about the research parameter θ induces uncertainty about λ and the prior g_θ implies

$$g_\lambda(\lambda) = g_\theta\left(\xi^{-1}(\lambda)\right)\left|\frac{\partial \xi^{-1}(\lambda)}{\partial \lambda}\right|.$$

In the continuous case, the entropy and sample information are not invariant under one-to-one transformations. The entropy transformation formula is

$$H(\lambda) = H(\theta) + E_\theta\left[\log\left|\frac{\partial \xi^{-1}(\lambda)}{\partial \lambda}\right|\right]. \tag{3}$$

where E_θ denotes the expectation with respect to $g(\theta)$. This implies the following information transformation:

$$\Delta(\lambda; y) = \Delta(\theta; y) + E_\theta\left[\log\left|\frac{\partial \xi^{-1}(\lambda)}{\partial \lambda}\right|\right] - E_{\theta|y}\left[\log\left|\frac{\partial \xi^{-1}(\lambda)}{\partial \lambda}\right|\right]. \tag{4}$$

In general, the prior and posterior expected values of the differential terms in Equation (4) are different, so $\Delta(\lambda; y)$ can be more or less than, or equal to $\Delta(\theta; y)$.

The Kullback–Leibler (KL) information divergence between two PDFs on a support $\mathbb{X} \subseteq \Re^K$ is defined by

$$K\left(f_1 : f_2\right) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} dx \geq 0, \tag{5}$$

provided that $f_1(x) = 0$, whenever $f_2(x) = 0$, and $K\left(f_1 : f_2\right) = 0$ if only if $f_1(x) = f_2(x)$ almost everywhere. This measure is invariant under one-to-one transformations, so $K\left(g_{\lambda|y} : g_\lambda\right) = K\left(g_{\theta|y} : g_\theta\right)$. The mutual information between two random variables X and Y with a joint PDF f and marginal distributions f_x and f_y is defined by

$$M(X, Y) = H\left(f_x\right) - E_y\left[H\left(f_{x|y}\right)\right] = E_y\left[K\left(f : f_{x|y}\right)\right] = K\left(f : f_x f_y\right) \geq 0. \tag{6}$$

The KL representation in Equation (6) displays the mutual information as a dependence measure, where $M(X, Y) = 0$ if and only if the two variables are independent.

Lindley [6] defined the expected sample information about θ by the entropy representation in Equation (6) where $M(\theta, Y) = E_y[\Delta(\theta; y)] = E_y\left[K\left(g_{\theta|y} : g_\theta\right)\right]$. This is the prominent Bayesian information measure (see, Ebrahimi et al. [13] and references therein). However, $K\left(g_{\theta|y} : g_\theta\right) \geq 0$ does not indicate which of the two distributions is more informative. Also $M(\theta, Y) = M(\xi(\theta), Y)$. Consequently, $M(\theta, Y)$ is not germane to the parameterization question. We will use Equations (5) and (6) for assessing the effects of environment on lifetimes of components of systems.

The ME principle provides a unique likelihood model consistent with the given moment constraints Equation (1). The ME model that satisfies the K moment constraints Equation (1), if exists, has a PDF in the following form:

$$f_\lambda^*(y|\theta) = C(\lambda) \exp\left\{-\sum_{k=1}^K \lambda_k T_k(y)\right\}, \quad y \in \mathbb{Y}, \tag{7}$$

where $\lambda = \left(\lambda_1, \ldots, \lambda_K\right)^T$ is the vector of Lagrange multipliers for constraints Equation (1), the subscript λ signifies the parameterization of the model, and $C(\lambda)$ is the normalizing factor. The existence of the ME model is determined by $C^{-1}(\lambda) < \infty$. The maximum entropy approach is well studied with a large literature in Statistics, Physics, and many other fields; we refer readers to Foley and Scharfenaker [14] and references therein for the latest developments. By the uniqueness of the ME model, $\lambda = \xi(\theta)$ is a one-to-one function implying that $\theta = \xi^{-1}(\lambda)$ and $f_\lambda^*(y|\theta)$ can be reparameterized and represented simply as $f^*(y|\theta)$.

The exponential family PDFs are in the form of Equation (7) where λ is called the canonical (natural) parameter. The members with finite entropies are ME models [15]. All models for $f(y|\theta)$ considered in this paper are well known members of the exponential family finite entropies. The types of their ME moment constraints can be found in the literature (see, e.g., the Wikipedia page on the maximum entropy probability distributions).

3 | Failure Time of an Item

Consider learning about the expected failure time of an item $E(Y|\mu) = \mu$. We describe uncertainty about μ by the inverse-gamma distribution $g(\mu) = \mathcal{IG}(a, b)$ with shape and scale parameters a and b and PDF

$$g(\mu) = \frac{b^a}{\Gamma(a)} \mu^{-a-1} e^{-b/\mu}, \quad \mu > 0, a, b > 0. \tag{8}$$

The ME model with a single constraint $E(Y|\mu) = \mu$ on the nonnegative support is exponential with PDF $f^*(y|\mu) = (1/\mu)e^{-y/\mu}$, $y \geq 0$. The Lagrange multiplier for the mean constraint is the failure rate $\lambda = 1/\mu$. The implied prior for the failure rate is $g(\lambda) = \mathcal{G}(a, b)$.

A sample of failure times $y_1, \ldots, y_n$ provides the following representations of the likelihood functions in terms of the mean and failure rate parameters:

$$L(\mu) = \frac{1}{\mu^n} e^{-s_n/\mu} \tag{9}$$

$$L(\lambda) = \lambda^n e^{-\lambda s_n}, \tag{10}$$

where $s_n = \sum_{i=1}^n y_i$. The inverse-gamma and gamma priors are conjugate for these likelihood functions, respectively. The posterior parameters are $a_n = a + n$ and $b_n = b + s_n$.

The posterior entropy of the mean is

$$H\left(\mu|s_n\right) = \log\Gamma\left(a_n\right) + \log b_n - \left(a_n + 1\right)\psi\left(a_n\right) + a_n. \tag{11}$$

where $\psi(\alpha) = d\log\Gamma(\alpha)/d\alpha$ is digamma function. The prior entropy $H(\mu)$ is given by Equation (11) with $a_n = a$ and $b_n = b$. The sample information about the mean is

$$\Delta\left(\mu; s_n\right) = -\log\left(1 + \frac{s_n}{b}\right) + \log\frac{\Gamma(a)}{\Gamma\left(a_n\right)} - (a+1)\psi(a) + \left(a_n + 1\right)\psi\left(a_n\right) - n. \tag{12}$$

The sample information for learning about the failure rate requires adjustment of Equation (12) as in Equation (4). The differential term in Equation (3) for $\lambda = 1/\mu$ is $\log|d\mu/d\lambda| = -2\log\lambda$. Using this in Equation (4) we obtain

$$\Delta(\lambda; s_n) = \Delta(\mu; s_n) + 2\log\left(1 + \frac{s_n}{b}\right) - 2\left[\psi(a_n) - \psi(a)\right]. \quad (13)$$

From Equations (12) and (13) we conclude that the failure time data is more (less) information about the mean than about the failure rate if and only if

$$\log\left(1 + \frac{s_n}{b}\right) < (>)\ \psi(a_n) - \psi(a). \quad (14)$$

This inequality has an implication on the conclusions of Abel and Singpurwalla [5]. They considered comparing information about the mean and failure rate when an item survived at time $y = y_0$. Their likelihood functions for this case are based on the survival functions $P(Y > y_0|\mu) = e^{-y_0/\mu}$ and $P(Y > y_0|\lambda) = e^{-\lambda y_0}$, which is different from Equations (9) and (10) with $n = 1$ and $s_n = y_0$. These likelihood functions provide the following sample information measures:

$$\Delta(\mu;\ \text{survival at}\ \ y_0) = -\Delta(\lambda;\ \text{survival at}\ \ y_0) = -\log\left(1 + \frac{y_0}{b}\right).$$

This measure does not include the shape parameter terms in Equation (12) which is derived from likelihood function Equation (9).

Abel and Singpurwalla's [5] likelihood functions for the mean and the failure rate after observing a failure at time $y = y_0$ are the same as Equations (9) and (10) with $n = 1$ and $s_n = y_0$. So, their observed information measures about the mean and failure rate for observing a failure at $y = y_0$ are the same as Equations (12) and (13) with $s_n = y_0$. Letting $s_n = y_0$ and $n = 1$ in Equation (14) and using the recursive $\psi(a+1) = \psi(a) + 1/n$, we can conclude that there is more (less) information about the mean than about the failure rate in a failure observation at $y = y_0$, if and only if

$$\log\left(1 + \frac{y_0}{b}\right) < (>)\ \frac{1}{a}, \quad \text{for all} \quad a > 0.$$

This condition also holds for their comparison between the survival and failure data, which improves on their conclusion where the condition is stated as $a > 1$.

4 | Probability of Failure

Consider a quality control process with outcomes $y \in \{0, 1\}$, denoting the failure and success to pass, respectively. The uncertainty about the probability that the product passes the inspection $\pi = P(Y = 1)$ is described by the Beta prior $\mathcal{B}(a, b)$ with PDF

$$g(\pi) = \frac{1}{B(a, b)}\pi^{a-1}(1-\pi)^{b-1}, \quad 0 \le \pi \le 1, \ \ a, b > 0,$$

where $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is the beta function.

4.1 | Saturated ME Model

The ME model with a single moment constraint $\pi = E(Y)$ on the binary support $y \in \{0, 1\}$ is the Bernoulli distribution with PDF

$$f(y|\pi) = \pi^y(1-\pi)^{1-y}, \quad 0 \le \pi \le 1, \ \ y = 0, 1.$$

This is a saturated ME model where a single constraint in addition to the normalizing constraint exhausts the number of outcomes on the binary support. This distribution can be represented in terms of the general form of the ME model Equation (7) where the Lagrange multiplier is the log-odds parameter in favor of failure, $\lambda = \log[(1-\pi)/\pi]$.

A sample of events $y_1, \ldots, y_n$ provides the following likelihood function for the mean:

$$L(\pi) \propto \pi^{s_n}(1-\pi)^{n-s_n}, \quad 0 \le \pi \le 1, \ \ y = 0, 1,$$

where $s_n = \sum_{i=1}^n y_i$. The posterior distribution of π is $\mathcal{B}(a_n, b_n)$, where $a_n = a + s_n$ and $b_n = b + n - s_n$. The posterior entropy of π is

$$\begin{aligned} H(\pi|s_n) = \log B(a_n, b_n) - (a_n - 1)\left[\psi(a_n) + \psi(a_n + b_n)\right] \\ - (b_n - 1)\left[\psi(b_n) + \psi(a_n + b_n)\right]. \end{aligned} \quad (15)$$

The prior entropy $H(\pi)$ is given by Equation (15) with $a_n = a$ and $b_n = b$.

Lindley [9] proposed continuing binomial sampling until the values of a_n and b_n obtained are such that the posterior information $I(\pi|s_n) = -H(\pi|s_n)$ has attained a prescribed value. In general, the posterior information rule is different from the sample information rule defined based on Equation (2), according to which one continues sampling until the values of a_n and b_n obtained are such that $\Delta(\pi; s_n)$ has attained a prescribed value. An exception is the case of uniform prior for the mean for learning about π, where the prior entropy $H(\pi) = 0$, implying that $\Delta(\pi; s_n) = -H(g_{\pi|s_n})$. The equivalence between Lindley's [6] sample information and Lindley's [9] posterior information rule does not hold for other than uniform prior for the mean to learn about the mean and for a uniform prior to learn about a transformation of the mean.

For example, Lindley [9] also used the posterior information for binomial sampling to learn about the log-odds against failure, $\xi(\pi) = \log(\pi/(1-\pi)) = -\lambda$. The beta $\mathcal{B}(a, b)$ prior for π implies the Generalized Logistic $\mathcal{GL}(a, b)$ prior for the log-odds λ:

$$g(\lambda) = \frac{e^{-b\lambda}}{B(b, a)\left(1 + e^{-\lambda}\right)^{a+b}}, \quad \lambda \in \Re, \ \ a, b > 0.$$

This distribution is also called generalized logistic Type IV and is a conjugate prior for $\mathcal{L}(\lambda)$.

Nadarajah and Zografos [16] gives the following expression for the entropy of $\mathcal{GL}(a, b)$:

$$H(\lambda|s_n) = \log B(b_n, a_n) - a_n\psi(a_n) - b_n\psi(b_n) + (a_n + b_n)\psi(a_n + b_n). \quad (16)$$

The prior entropy $H(\lambda)$ is given by Equation (16) with $a_n = a$ and $b_n = b$. Lindley [9] gives the posterior information $I[\mathcal{GL}(b, a)] = -H(\lambda|s_n)$. However, the uniform prior for π implies the logistic distribution $\mathcal{GL}(1, 1)$ the log-odds.

In comparing the sample information Equation (2) for π and for its transformations, the respective priors also play important roles.

$$\Delta(\pi; s_n) = H[\mathcal{B}(a, b)] - H\left[B(a_n, b_n)\right]; \quad (17)$$

TABLE 1 | The number of trials for attaining one nit (1.44 bits) sample information about the mean and the log-odds parameter.

Prior for π	n	s_n	$\max_{s_n} \Delta(\pi; s_n)$	$\min_{s_n} \Delta(\lambda; s_n)$	s_n	$\min_{s_n} \Delta(\pi; s_n)$	$\max_{s_n} \Delta(\lambda; s_n)$
Uniform	6	0	1.089	0.353	3	0.384	0.865
		6	1.089	0.353			
	9	0	1.403	0.374	4	0.528	1.037
		9	1.403	0.374	5	0.528	1.037
Jeffreys	3						
		0	1.004	0.404	1	−0.053	0.947
		3	1.004	0.404	2	−0.053	0.947
	4	0	1.235	0.421	2	−0.042	1.125
		4	1.235	0.421			

$$\Delta(\lambda; s_n) = H[\mathcal{GL}(a, b)] - H\left[\mathcal{GL}(a_n, b_n)\right]. \tag{18}$$

Table 1 gives the number of trials needed for obtaining one nit (natural log unit equivalent to 1.44 bits) of sample information about the mean with the uniform and Jeffreys priors for π and the induced prior for λ; R codes are available for computing Equations (17) and (18) upon request. The following points are evident:

a. With the uniform prior a larger number of trials is needed to attain a specified amount of information than with the Jeffreys prior.

b. For a given n, $\max_{s_n} \Delta(\pi; s_n)$ and $\min_{s_n} \Delta(\lambda; s_n)$ occur for outcomes $s_n = 0, n$, and $\min_{s_n} \Delta(\mu; s_n)$ and $\max_{s_n} \Delta(\lambda; s_n)$ occur for the median outcome(s).

c. With the uniform prior, the outcome of six (nine) trials attain one nit of information about the mean (log-odds).

d. With the Jeffreys prior only three (four) trials are needed to attain one nit of information about the mean (log-odds).

Figure 1 shows plots of the sample information about the mean (solid red) and the log-odds in favor of failure (dashed blue) when 20 trials (left panels) and 30 trials (right panels) based on the uniform prior (upper panels) and Jeffreys prior (lower panels) for π. The horizontal (dashed-dotted) lines show the sample outcomes s_n required for attaining one nit of sample information. The upper left panel shows that when 20 trials are considered $s_n \le 4$ and $s_n \ge 16$ provide the amount of required information about the mean π, while $2 \le s_n \le 18$ provide the required amount of information about the log-odds parameter λ. The upper right panel shows that when 30 trials are considered, all outcomes attain the required amount of information about π and outcomes $2 \le s_n \le 28$ provide the required amount of information about λ.

The lower panels of Figure 1 show plots of the sample information about π using Jeffreys prior $\mathcal{B}(1/2, 1/2)$ and the implied $\mathcal{GL}(1/2, 1/2)$ prior for λ. When 30 trials are considered, with the Jeffreys prior for π, outcomes $s_n \le 5$ and $s_n \ge 25$ provide the required amount of information about the mean, and with the implied $\mathcal{GL}(1/2, 1/2)$ prior for λ, outcomes $1 \le s_n \le 29$ provide the required amount of information about the log-odds parameter.

4.2 | Number of Failures Before a Success

Consider learning about the expected number of defective items inspected $y = 0,1,2, \ldots$ before finding the first non-defective one. We use inverted beta $\mathcal{IB}(a, b)$ prior for $E(Y) = \mu$ with PDF

$$g(\mu) = \frac{\mu^{a-1}}{B(a, b)(1+\mu)^{a+b}}, \quad \mu > 0, \ \ a, b > 0.$$

This distribution is also called Beta prime and is a special case of the compound gamma distribution [17]. The prior mean $E(\mu) = a/(b-1), a \ne 1, b > 1$. For $a = 1$, the prior for μ reduces to the Pareto Type II distribution $\mathcal{P}_{II}(1, b)$ and the mean is not defined.

The ME model with constraint $E(Y) = \mu$ on the support $\mathbb{Y} = \{0, 1, 2, \ldots\}$ is the geometric distribution with PDF

$$f(y|\mu) = \frac{1}{\mu+1}\left(\frac{\mu}{\mu+1}\right)^y, \quad y = 0, 1, 2, \ldots. \tag{19}$$

The probability of an item being defective in each inspection is $\pi = 1/(\mu + 1)$. This gives the more familiar form of the geometric PDF:

$$f(y|\pi) = \pi(1-\pi)^y, \quad y = 0, 1, 2, \ldots, \quad 0 < \pi < 1.$$

Notice that $\mu = (1-\pi)/\pi$ is odds in favor of failure. The form of the general ME model Equation (7) with Lagrange multiplier $\lambda = \log(1 + 1/\mu) = -\log(1-\pi) > 0$.

A quantity of interest is the failure rate of the process. Sarhan et al. [18] presented two failure rate functions of the geometric distribution based on two related definitions of the failure rate functions for discrete distributions. We adapt the definition given in their Equation (3) for Equation (19) and obtain

$$r = \Pr(Y = y|Y \ge y) = \frac{\pi(1-\pi)^y}{\sum_{i=y}^{\infty} \pi(1-\pi)^i} = \pi = \frac{1}{\mu+1}.$$

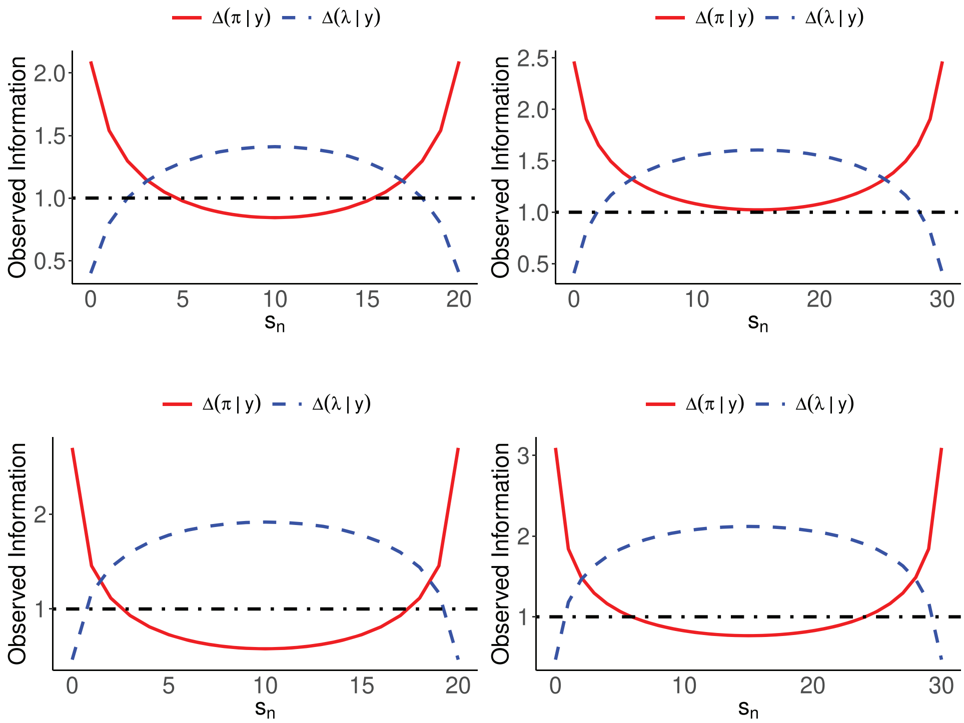

FIGURE 1 | Information about the mean and the log-odds parameter for binomial sampling with the uniform prior (upper panels) and Jeffreys prior (lower panels) for the mean; 20 trials (left panels) and 30 trials (right panels).

A sample of the number of failures before observing a success, $y_1, \ldots, y_n$, provides the following representations of the likelihood function:

$$L(\mu) \propto \frac{\mu^{s_n}}{(\mu+1)^{n+s_n}},$$

where $s_n = \sum_{i=1}^n y_i$. The inverted beta $\mathcal{IB}(a,b)$ prior for μ is conjugate for $L(\mu)$ and the posterior distribution is $\mathcal{IB}(a_n, b_n)$ with updated parameters $a_n = a + s_n$ and $b_n = b + n$.

Using the expression for the entropy of the compound gamma distribution [16] we obtain the sample information $\Delta(\mu; s_n)$ by Equation (2) with

$$H(\mu|s_n) = \log B(a_n, b_n) - (a_n - 1)\psi(a_n) - (b_n + 1)\psi(b_n) + (a_n + b_n)\psi(a_n + b_n).$$

The inverted beta $\mathcal{IB}(a,b)$ prior for $\mu = (1-\pi)/\pi$ implies beta $\mathcal{B}(b,a)$ prior for π. For the failure rate the differential terms in Equation (4) is $\log |\, d\mu/dr \,| = -2\log r$ and the two expectations in Equation (4) give

$$\Delta(r; s_n) = \Delta(\mu; s_n) + 2[\psi(a_n) - \psi(b_n)] - 2[\psi(a) - \psi(b)].$$

As in the case of time to failure, the difference between sample information about the mean and failure rate depends on the hyper-parameters, the sample size and the data. Thus, $\Delta(r; s_n) \leq \Delta(\mu; s_n)$ if and only if

$$\psi(a+n) - \psi(b+s_n) \geq \psi(a) - \psi(b). \tag{20}$$

The digamma function is strictly monotone increasing, so for $a = b$, the inequality in (20) becomes equation when $s_n = n$. It is easy to show that the inequality in (20) holds for $s_n < n$ and is reversed for $s_n > n$.

The sample information about the ME Lagrange multiplier can be computed with $\log |\, d\pi/d\lambda \,| = \log(1-\pi)$ for the differential terms in Equation (4), however, unlike the other ME models considered, λ is not a parameter of interest.

Remark 1. The geometric distribution is also represented as the discrete counterpart of the exponential model. Let $X = Y + 1$. Then $f(x|\pi) = \pi(1-\pi)^{x-1}, \ x = 1, 2, \ldots, \ 0 < \pi < 1$ and $E(X|\theta) = \theta = 1/\pi = 1/r$. The sample information about the mean and failure rate are as shown above, due to $\theta = \mu + 1$ and $\log |\, d\theta/dr \,| = -2\log r$.

5 | Mean and Variance of a Measurement

The problem of learning about the unknown mean μ and variance σ^2 of a variable of interest Y with a PDF on the support $\mathbb{Y} = \Re$ is of paramount interest in many fields. The uncertainty about the mean is described by the normal prior $g(\mu) = \mathcal{N}(\mu_0, \omega_0^2)$ and the uncertainty about the variance prior for σ^2 is described by the inverse gamma prior Equation (8) or the $\mathcal{G}(a,b)$ prior for the precision parameter $\tau = 1/\sigma^2$. When the research question is about the mean and variance jointly, the uncertainty is described by a normal-gamma prior $g_{\mu,\tau}$.

The ME model for the conditional distribution of Y with $E(Y|\mu,\sigma^2) = \mu$ and $E[(Y-\mu)^2|\mu,\sigma^2] = \sigma^2$ is $f^*(y|\mu,\sigma^2) = \mathcal{N}(\mu,\sigma^2), \ y \in \Re$. The ME derivation of the likelihood model gives the Lagrange multipliers for the first two moments as

$$\lambda_1 = \xi_1(\mu) = \frac{\mu}{\sigma^2}, \quad \lambda_2 = \xi_2(\sigma^2) = \frac{1}{2\sigma^2} = \frac{1}{2}\tau.$$

5.1 | One Unknown Moment

When the variance is assumed to be known, the normal prior for the mean is conjugate prior for the conditional normal likelihood function $L(\mu|\sigma^2)$. The posterior distribution is $g(\mu|\overline{y}) = \mathcal{N}\left(\mu_n, \omega_n^2\right)$, where $\overline{y} = \sum_{i=1}^n y_i/n$,

$$\mu_n = \frac{\sigma^2\mu_0 + n\omega_0^2\overline{y}}{\sigma^2 + n\omega_0^2}, \quad \omega_n^2 = \frac{\sigma^2\omega_0^2}{\sigma^2 + n\omega_0^2}. \tag{21}$$

The joint distribution of $(\mu, \overline{y})$ is bivariate normal with the squared correlation

$$\rho_{\overline{y},\mu}^2 = \frac{\omega_0^2}{\sigma^2/n + \omega_0^2}.$$

Using the entropy expression for the normal distribution in Equation (2) gives $\Delta(\mu; y, \sigma^2) = -\frac{1}{2}\log\left(1 - \rho_{\overline{y},\mu}^2\right)$. Because the sample information is determined only by the variance parameters, we have $M(Y, \mu) = \Delta(\mu; y)$ which is the well-known bivariate normal mutual information.

For the sample information about λ_1, the differential terms in Equation (4) are not functions of μ, thus $\Delta\left(\lambda_1; \overline{y}\right) = \Delta(\mu; \overline{y})$.

Consider the case when the variance $\sigma^2 = E\left[(Y-\mu)^2|\mu\right]$ about the given mean is the unknown moment parameter of interest for inference. The normal ME model provides the following two forms of the likelihood function in terms of σ^2 and the precision parameter τ:

$$L\left(\sigma^2\right) = \frac{1}{\left(\sigma^2\right)^{n/2}} e^{-v_n/\sigma^2}$$
$$L(\tau) = \tau^{n/2} e^{-v_n\tau},$$

where

$$v_n = \frac{1}{2}\sum_{i=1}^n \left(y_i - \mu\right)^2. \tag{22}$$

These parameterizations are commonly used in the Bayesian analysis of normal variance model with the inverse-gamma conjugate prior Equation (8) for $L\left(\sigma^2\right)$ and gamma prior for $L(\tau)$ which provide two different sample information measures for the variance and the precision parameter. The conditional posterior distributions $g\left(\sigma^2|v_n, \mu\right)$ and $g\left(\tau|v_n, \mu\right)$ are as in those for the exponential model with $n/2$ and v_n in the places of n and s_n. The sample information about the variance and the precision parameter are given by Equations (12) and (13), adjusted similarly.

5.2 | Both Moments Unknown

We present the information provided by the data about (μ, τ) using two normal-gamma prior distributions used in the Bayesian literature.

5.2.1 | Dependent Prior Moments

For obtaining posterior distributions in a closed form, the normal-gamma prior is constructed as $g_{\mu,\tau}(\mu, \tau) = g_{\mu|\tau}(\mu|\tau)g_\tau(\tau)$, where the prior for μ is linked to τ as follows: $g_{\mu|\tau} = \mathcal{N}\left(\mu_0, \omega_0^2\right)$ with the ratio of prior to model precisions $\tau_0/\tau = n_0$, that is, $\omega_0^2/\sigma^2 = 1/n_0$ and gamma prior $g_\tau = \mathcal{G}(a, b)$ for the precision. Then the posterior distribution of the mean conditional on variance is $g_{\mu|\overline{y},\tau} = \mathcal{N}\left(\mu_{n,n_0}, \omega_{n,n_0}^2\right)$, where the posterior parameters are as in Equation (21) with

$$\mu_{n,n_0} = \frac{n\overline{y} + n_0\mu_0}{n + n_0}, \quad \omega_{n,n_0}^2 = \frac{\sigma^2}{n + n_0} = \frac{1}{\left(n + n_0\right)\tau}.$$

The sample information about the mean, given the variance, is as Equation (12) with $\omega_n^2 = \omega_{n,n_0}^2$.

The posterior distribution of τ is $\mathcal{G}\left(a_n, b_{n,n_0}\right)$, where $a_n = a + n/2$,

$$b_{n,n_0} = b + \frac{n}{2}\left[\hat{\sigma}_n^2 + \frac{n_0}{n + n_0}\left(\overline{y} - \mu_0\right)^2\right],$$

and $\hat{\sigma}_n^2 = \sum_{i=1}^n (y - \overline{y})^2/n$ is the maximum likelihood estimate of the variance.

To compute the sample information about the mean and variance we use the following additive property of Shannon entropy:

$$H(\mu, \tau) = H(\tau) + E_\tau[H(\mu|\tau)],$$

where $H(\tau) = H[\mathcal{G}(a, b)]$ and $E_\tau[H(\mu|\tau)]$ is the conditional entropy given by

$$\begin{aligned} E_\tau[H(\mu|\tau)] &= E_\tau\left\{.5 + .5\log(2\pi) + .5\log\left[\left(n + n_0\right)\tau\right]\right\} \\ &= .5\left[1 + \log(2\pi) + \log\left(n + n_0\right) + \psi(a) - \log b\right]. \end{aligned} \tag{23}$$

Similarly, we obtain the joint posterior entropy, $H(\mu, \tau|y)$, with $H(\tau|y) = H\left[\mathcal{G}\left(a_n, b_n\right)\right]$ and $E_{\tau|y}[H(\mu|\tau, y)]$ by Equation (23) with $a = a_n$ and $b = b_{n,n_0}$. Using the prior and posterior joint entropies we obtain the sample information about the mean and precision parameters as follows:

$$\Delta((\mu\tau); y) = \Delta(\tau; y) + E_{\tau|y}[\Delta(\mu; (\tau y))],$$

where $\Delta(\tau; y)$ is given by Equation (13) with $n/2$ and v_n in the places of n and s_n defined in Equation (22).

The sample information about the mean and variance can be found similarly by noting that $E_{\sigma^2|y}\left[H\left(\mu|\sigma^2, y\right)\right] = E_{\tau|y}[\Delta(\mu; (\tau y))]$ and $\Delta\left(\sigma^2; s_n\right)$ is given by the first term of Equation (12) with v_n in the places of n and s_n. defined in Equation (22).

5.2.2 | Independent Prior Moments

The independent normal-gamma prior is constructed as $g_{\mu,\tau}(\mu, \tau) = g_\mu(\mu)g_\tau(\tau)$, where $g_\mu = \mathcal{N}\left(\mu_0, \omega_0^2\right)$ and $g_\tau = \mathcal{G}(a, b)$. With this prior closed form posterior distributions are analytically intractable, so the sample information about these parameters can not be computed. We pursue using MCMC to generate a large sample from the posterior distributions. Then the information measures are computed by applying the bivariate kernel method developed by Pflughoeft et al. [11] to the posterior samples.

TABLE 2 | Prior and posterior results of 10,000 MCMC samples for the mean, precision parameter, and variance.

Parameter	Mean	SD	Median	2.5 percentile	97.5 percentile	Entropy	Observed information
Mean							
Prior	−0.011	0.998	−0.023	−1.945	1.964	1.429	
Posterior	5.016	0.179	5.013	4.660	5.370	−0.288	1.717
Precision							
Prior	3.006	1.729	2.679	0.621	7.245	1.876	
Posterior	0.310	0.043	0.309	0.232	0.399	−1.731	3.607
Variance							
Prior	0.498	0.453	0.375	0.140	1.632	0.484	
Posterior	3.190	0.316	3.168	2.630	3.875	0.268	0.216

The MCMC samples are generated as follows:

$$y \mid \mu, \tau \sim \mathcal{N}(\mu, \tau),$$
$$\mu \sim \mathcal{N}(0, 1),$$
$$\tau \sim \mathcal{G}(3, 1), \quad \sigma^2 = \frac{1}{\tau}.$$

A sample of $n = 100$ from $\mathcal{N}(5, 4)$ provided $\overline{y} = 5.18$, $\hat{\sigma}_n^2 = 3.333$, and $\hat{\tau} = 0.300$. We obtained 10,000 prior and posterior samples for the three parameters based on generating 101,000 points with a burn-in of 1000 and a thinning interval of 10.

Table 2 presents the prior and posterior MCMC results for μ, τ, and σ^2. These univariate measures are computed using the marginals of the bivariate kernel estimates of PDFs of (μ, τ) and (μ, σ^2). The first five columns give the standard univariate sample summaries. The right two columns give the sample entropy and information measure. The data yields more than twice as much informative about the precision parameter than about the mean. The sample information about the variance is substantially less than information about the mean.

The prior and posterior joint entropies are $H(\mu, \tau) = 3.298$ and $H(\mu, \tau|y) = -2.042$. The prior and posterior joint entropies are $H(\mu, \sigma^2) = 1.631$ and $H(\mu, \sigma^2|y) = -.045$. These measures give the joint sample information measures $\Delta[(\mu\tau); y] = 5.339$ and $\Delta[(\mu\sigma^2); y] = 1.676$, respectively. The amount of information gain about the pair (μ, τ) is substantial and so is its difference with the information gain about (μ, σ^2).

6 | Test and Operating Environments

Systems usually operate in environments that are different from their components were tested. Lindley and Singpurwalla [7] and Currit and Singpurwalla [12] studied failure times of systems whose components have exponential distributions. They assumed the failure rates of components in the test environment are r_k and in the operating environment $r_k^e = \eta r_k$, where the factor $\eta > 0$ captures the effects of environment on the failure rate of the components (and thus the failure of the system). These authors described the uncertainty about the environment effect parameter η by a Gamma distribution. We use the following slightly modified prior $g(\eta) = \mathcal{G}(\alpha, \beta)$, $\alpha, \beta > 0$, which is used by Ebrahimi et al. [19].

The Bayesian predictive distributions of the lifetimes of the components in the operating environment are the Pareto Type II distribution with PDFs

$$f_k^{op}(y_k|r_k, \alpha, \beta) = \frac{r_k \alpha \beta^\alpha}{(r_k y_k + \beta)^{\alpha+1}}, \quad y_k \geq 0, \; r_k > 0, \; \alpha, \beta > 0, k = 1, 2. \tag{24}$$

According to Theorem 2 of Ebrahimi et al. [15], the PDF in Equation (24) is the ME model consistent with the moment constraint $E[\log(r_k Y_k + \beta)] = \theta$, and Lagrange multiplier $\lambda^e = \alpha + 1$.

For a more interpretable moment constraint, let $Z_k = r_k Y_k + \beta$. Then

$$f^{op}(z_k|r_k, \alpha) = \frac{\alpha}{z_k^{\alpha+1}}, \quad z_k \geq 1, \; \alpha > 0.$$

The ME moment parameter is $E[\log Z_k) = \theta^e = 1/\alpha$ and the failure rate function of this distribution is $r_k^e(z_k) = \alpha/z_k$. Thus, the information about λ^e, r^e and θ^e can be obtained via α (the hyper-parameter of prior for η). The conjugate prior for α is gamma $g(\alpha) = \mathcal{G}(c, d)$ [20]. Given a sample of failure time of a component in the operating environment $z_{k1}, \ldots, z_{kn}$, the posterior distribution is $g(\alpha|r, s_{kn}^e) = \mathcal{G}(c + n, d + s_{kn}^e)$, where $s_{kn}^e = \sum_{i=1}^n \log z_{ik}$. Thus,

$$\Delta(\lambda^e; s_{kn}^e) = \Delta(\alpha; s_{kn}^e) = \log\left(1 + \frac{s_{kn}^e}{d}\right) + H(\mathcal{G}_c) - H(\mathcal{G}_{c+n}),$$

where $H(\mathcal{G}_c) = H[\mathcal{G}(c, 1)]$.

6.1 | Lindley-Singpurwalla Distribution

In the testing environment the components are structurally independent and the joint distribution of their lifetimes is the independent exponential. For a system with two components, the joint PDF is

$$f^{test}(y_1, y_2|r_1, r_2) = r_1 r_2 e^{-(r_1 y_1 + r_2 y_2)}. \tag{25}$$

Lindley and Singpurwalla [7] derived the following bivariate distribution for the lifetimes of components in the operating environment:

$$f^{op}(y_1, y_2|r_1, r_2, \alpha, \beta) = \frac{\alpha(\alpha+1)r_1 r_2 \beta^\alpha}{(r_1 y_1 + r_2 y_2 + \beta)^{\alpha+2}}, \quad y_k \geq 0, \ r_k > 0, \quad \alpha, \beta > 0, k = 1, 2. \tag{26}$$

In the literature, Equation (26) is known as the *Lindley-Singpurwalla* bivariate distribution. Factoring out β in the denominator of this PDF gives the bivariate Pareto Type II or Lomax PDF. The Lindley-Singpurwalla model in Equation (26) establishes that in the operating environment, the lifetimes of components, distributed as the univariate Pareto Type II Equation (24), become dependent.

By Theorem 2 of Ebrahimi et al. [15], the transformed PDF is the ME model in with the following moment: $E[\log(r_1 Y_1 + r_2 Y_2 + \beta)] = \theta$, where the Lagrange multiplier is $\lambda = \alpha + 2$. This ME moment constraint is not intuitively interpretable.

The following information measures assess the effects of operating environment on the joint distribution of the components:

$$M^{op}(Y_1, Y_2|r_1, r_2, \alpha, \beta) = \log\left(1 + \frac{1}{\alpha}\right) - \frac{1}{\alpha+1} > M^{test}(Y_1, Y_2|r_1, r_2) = 0;$$

$$H(f^{op}) - H(f^{test}) = \log \frac{\beta^2}{\alpha(\alpha+1)} + \frac{3\alpha+2}{\alpha(\alpha+1)}; \tag{27}$$

$$K(f^{\text{op}} : f^{\text{test}}) = \log \frac{\alpha(\alpha+1)}{\beta^2} - \frac{3\alpha+2}{\alpha(\alpha+1)} + \frac{2\beta}{\alpha-1} > 0, \quad \alpha > 1. \tag{28}$$

The mutual information of the bivariate Pareto distribution is well known. This measure is useful for evaluating the amount of dependence induced on the lifetimes of components in the operating environment in terms of the prior shape parameter α. It is decreasing and approaches zero as $\alpha \to \infty$. The entropy difference compares the extent of concentrations of the joint distributions for the two environments. This measure is decreasing sharply in α and increases slowly in β, It is positive for $\alpha, \beta > 1$ and changes sign at a point whose coordinates increase in both parameters. The KL information divergence decreases sharply and then increases slowly. The first coordinate of the minimal point increases in α and decreases in β.

Remark 2. Lindley and Singpurwalla [7] motivated Equation (26) in contrast with the unwieldy distribution of the lifetime of parallel system with two components, which is given by $T = \max\{Y_1, Y_2\}$. The conditional PDF of $T \mid \eta, r_1, r_2$ in the operating environment is

$$f^P(t|\eta, r_1, r_2) = \sum_{k=1}^{2} \eta r_k e^{-\eta r_k t} - \eta(r_1 + r_2)e^{-\eta(r_1+r_2)t}, \quad t \geq 0, \eta, r_1, r_2 > 0. \tag{29}$$

This PDF and the prior $g(\eta) = \mathcal{G}(\alpha, \beta)$ give the following predictive PDF of the parallel system lifetime T in the operating environment:

$$f^{P.op}(t|r_1, r_2, \alpha, \beta) = \sum_{k=1}^{2} \frac{r_k \alpha \beta^\alpha}{(r_k t + \beta)^{\alpha+1}} - \frac{(r_1 + r_2)\alpha\beta^\alpha}{[(r_1 + r_2)t + \beta]^{\alpha+1}}, \quad t \geq 0, r_1, r_2, \alpha, \beta > 0. \tag{30}$$

Information measures that involve this predictive distribution cannot be obtained in closed form. Furthermore, Equation (30) cannot be represented in the general form of the ME model Equation (7).

6.2 | Series Systems

Currit and Singpurwalla [12] presented the survival function and failure rate of the lifetime of the series system with two components, which is given by $T = \min\{Y_1, Y_2\}$. The conditional PDF of $T \mid \eta, r_1, r_2$ in the operating environment is

$$f^S(t|\eta, r_1, r_2) = \eta(r_1 + r_2)e^{-\eta(r_1+r_2)t}, \quad t \geq 0, \ \eta, r_1, r_2 > 0. \tag{31}$$

This exponential distribution is the ME model, given that the mean lifetime of the system is $E(Y|\mu^e) = \mu^e = 1/[\eta(r_1 + r_2)]$, where the Lagrange multiplier for the constraint $\lambda \mid \eta, r_1, r_2 = \eta(r_1 + r_2)$.

The PDF in Equation (31) and the prior $g(\eta) = \mathcal{G}(\alpha, \beta)$ give the prior predictive distribution of T in the operating environment as the Pareto type II distribution with PDF

$$f^{S.op}(t|r_1, r_2, \alpha, \beta) = \frac{\alpha(r_1 + r_2)\beta^\alpha}{[(r_1 + r_2)t + \beta]^{\alpha+1}}, \quad t \geq 0, \ \alpha, \beta, r_1, r_2 > 0. \tag{32}$$

The ME moment condition for Equation (32) is similar to that given for Equation (24) with r_k replaced by $r_1 + r_2$.

Unlike the case of Equation (30), assortments of information measures can be computed with Equation (32). Given a sample of failure times of the series system in the operating environment $t_1^e, \ldots, t_n^e$, the sample information measures about α, Lagrange multiplier, and the failure rate function can be found similar to those obtained for Equation (24). The distribution of the lifetime of series system in test environment is $f^{S.test}(t|\lambda_1, \lambda_2) = (r_1 + r_2)e^{-(r_1+r_2)t}, t \geq 0, \ \eta, r_1, r_2 > 0$. In addition, the following information measures assess the effects of environment on the lifetime of the series system:

$$H(f^{S.op}) - H(f^{S.test}) = \log \frac{\beta}{\alpha} + \frac{1}{\alpha};$$

$$K(f^{S.op} : f^{S.test}) = \log \frac{\alpha}{\beta} - \frac{1}{\alpha} + \frac{\beta}{\alpha - 1}, \quad \alpha > 1.$$

The analytic behaviors of these measures are similar to Equations (27 and 28).

Some remarks are in order.

Remark 3.

1. Because of the unwieldiness of the PDF in Equation (30), information measures for comparison with Equation (32) cannot be obtained in closed form. However, the information importance of components for failure times of series and parallel systems in the test and operating environment can be investigated under the assumptions of Theorem 1 and Example 2 of Ebrahimi et al. [19].

2. The literature on the information measures of order statistics offers distribution-free measures for comparison of the lifetimes of parallel and series systems, with identical independent (or exchangeable) components. These measures do not distinguish between the test and operating environments. For example, the KL information divergence Equation (5) and the mutual information Equation (6) between consecutive order statistics are given in Ebrahimi et al. [21]. For the parallel and series systems, $K\left(f_{\min\{Y_1,Y_2\}} : f_{\max\{Y_1,Y_2\}}\right) = K\left(f_{\max\{Y_1,Y_2\}} : f_{\min\{Y_1,Y_2\}}\right) = 1$ and $M(\min\{Y_1, Y_2\})$, $\max\{Y_1, Y_2\} = 1 - \log 2$. Accordingly, the extent of divergence and dependence between the lifetimes of the parallel and series systems are equal irrespective of the environment irrespective of the environment. The same also holds for the Jensen-Shannon divergence of system with lifetime T, given by $JS(F_T; f_{V_1} f_{V_2}) = \sum_{k=1}^{2} p_k K(f_T : f_{V_k})$, where $p_k > 0$, $\sum_{k=1}^{2} p_k = 1$ are called the system *signatures* and f_{v_k}, $k = 1, 2$ are PDFs of the order statistics V_1, V_2 of Y_1, Y_2. Each p_k is in fact the probability that the kth component failure causes the system failure; see Samaniego [22] for the notion of system signatures. This measure gives JS (Parallel) = JS (Series) = 0; for details on the Jensen-Shannon divergence of systems see Asadi et al. [23].

7 | Concluding Remarks

The aim of this paper was to show the following contrast between two Bayesian approaches articulated by Lindley [2]: (a) The Bayesian standpoint defined by formulating a prior for the parameter that represents a research question and incorporating a likelihood and passing to a posterior. The likelihood model is a "part of Bayesian *techniques*" (the emphasis is his). (b) The usual Bayesian approach which begins with a likelihood model and specifies a prior for the likelihood model's parameter and passing to a posterior. We showed the consequences of these two approaches on the information provided by data about the parameter that represents a research question and about the likelihood model parameter. Following Able and Singpurwalla [5] we used the observed sample information defined by the difference between the prior and posterior entropies.

We studied the sample information about the mean and some of their well-known functions for variables with distributions on the nonnegative, binary, discrete, and real line supports. Our findings shed light on the conclusions of Able and Singpurwalla [5] and on Lindley's [9] posterior information rule for binomial sampling. We introduced sample information measures for learning about the mean and failure rate of the geometric distribution. We also introduced joint samples information about the mean and variance of the normal distribution under the dependent normal-gamma prior. For the independent normal-gamma prior case, we estimated the sample information measures about the unknown mean and variance of the normal distribution, utilizing the bivariate kernel estimate of the distribution. We offered several information measures for evaluating the effect of operation environment in the context of the Singpurwalla-Lindley model.

We also considered assessing the effects of environment on the lifetimes of components of a system. Lindley and Singpurwalla showed that the lifetimes of independent exponentially distributed components become dependent when their failure rates are proportional between the two environment, when uncertainty about the proportionality factor is described by a gamma distribution. We offered a few information measures for the Lindley-Singpurwalla model that assess the effects of environment. Developing information measures for assessing environment effects provides a high potential area for future research. One way to pursue is to model the effect of environment when the failure rates of components in the test environment are monotonic. Another way is to investigate representing different environmental conditions by different prior distributions, similar to Corollary 1 of Ebrahimi et al. [24].

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Endnotes

[1] An early version of this paper was presented at the symposium New Frontiers in Reliability and Risk Analysis: A Tribute to Nozer D. Singpurwalla, October 13–14.

[2] We thank Refik Soyer for suggesting this dialogue.

References

1. T. Z. Irony and N. D. Singpurwalla, "Non-Informative Priors Do Not Exist: A Dialogue With José M. Bernardo," *Journal of Statistical Planning and Inference* 65 (1997): 159–189.

2. D. V. Lindley, "Some Comments on "Non-informative Priors Do Not Exist"," *Journal of Statistical Planning and Inference* 65 (1997): 182–184.

3. E. T. Jaynes, "Prior probabilities," *IEEE Transactions on Systems Science and Cybernetics* sec-4(3) (1968): 227–241.

4. E. T. Jaynes, "Information Theory and Statistical Mechanics," *Physical Review* 106, no. 4 (1957): 620–630.

5. P. S. Abel and N. D. Singpurwalla, "To Survive or to Fail: That Is the Question," *American Statistician* 48 (1994): 18–21.

6. D. V. Lindley, "On a Measure of the Information Provided by an Experiment," *Annals of Mathematical Statistics* 27, no. 4 (1956): 986–1005.

7. D. V. Lindley and N. D. Singpurwalla, "Multivariate Distributions for the Life Lengths of Components of a System Sharing a Common Environment," *Journal of Applied Probability* 23 (1986): 418–431.

8. N. Ebrahimi, E. S. Soofi, and R. Soyer, "When Are Sample Failures More Informative Than Observed Survivals?," *Naval Research Logistics* 60 (2013): 102–110.

9. D. V. Lindley, "Binomial Sampling Schemes and the Concept of Information," *Biometrika* 44 (1957): 179–186.

10. O. M. Ardakani, R. F. Bordley, and E. S. Soofi, "Expected Information of Noisy Attribute Forecasts for Probabilistic Forecasts," *European Journal of Operational Research* (2025) forthcoming.

11. K. A. Pflughoeft, E. S. Soofi, and R. Soyer, "Information Architecture for Data Disclosure," *Entropy* 24 (2022): 670.

12. A. Currit and N. D. Singpurwalla, "On the Reliability Function of a System of Components Sharing a Common Environment," *Journal of Applied Probability* 25 (1988): 763–771.

13. N. Ebrahimi, E. S. Soofi, and R. Soyer, “On the Sample Information About Parameter and Prediction,” *Statistical Science* 25 (2010): 348–367.

14. D. K. Foley and E. Scharfenaker, “Bayesian Inference and the Principle of Maximum Entropy,” 2024 arXiv:2407.13029.

15. N. Ebrahimi, E. S. Soofi, and R. Soyer, “Multivariate Maximum Entropy Identification, Transformation, and Dependence,” *Journal of Multivariate Analysis* 99 (2008): 1217–1231.

16. S. Nadarajah and K. Zografos, “Formulas for Rényi Information and Related Measures for Univariate Distributions,” *Information Sciences* 155 (2003): 119–138.

17. S. D. Dubey, “Compound Gamma, Beta and F Distributions,” *Metrika* 16 (1970): 27–31.

18. A. M. Sarhan, F. M. Guess, and J. S. Usher, “Estimators for Reliability Measures in Geometric Distribution Model Using Dependent Masked System Life Test Data,” *IEEE Transactions on Reliability* 56, no. 2 (2007): 312–320.

19. N. Ebrahimi, N. Y. Jalali, E. S. Soofi, and R. Soyer, “Importance of Components for a System,” *Econometric Reviews* 33 (2014): 395–420.

20. B. C. Arnold and S. J. Press, “Bayesian Estimation and Prediction for Pareto Data,” *Journal of the American Statistical Association* 84 (1989): 1079–1084.

21. N. Ebrahimi, E. S. Soofi, and H. Zahedi, “Information Properties of Order Statistics and Spacings,” *IEEE Transactions on Information Theory* 50 (2004): 177–183.

22. F. J. Samaniego, *System Signatures and Their Applications in Engineering Reliability* (New York, NY: Springer Science + Business Media, 2007).

23. M. Asadi, N. Ebrahimi, E. S. Soofi, and Y. Zohrevand, “Jensen-Shannon Information of the Coherent System Lifetime,” *Reliability Engineering & System Safety* 156 (2016): 244–255.

24. N. Ebrahimi, S. N. U. A. Kirmani, and E. S. Soofi, “Predictability of Operational Processes,” *Naval Research Logistics* 58 (2011): 531–545.