$\mathcal{I}^2\mathcal{SDS}$

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2010-1

January 25, 2010

Information in Survival and Failure about Parameter and Prediction

Nader Ebrahimi
Division of Statistics
Northern Illinois University

Ehsan S. Soofi
Sheldon B. Lubar School of Business
University of Wisconsin-Milwaukee

Refik Soyer
Department of Decision Sciences
The George Washington University

Information in Survival and Failure about Parameter and Prediction

Nader Ebrahimi
Division of Statistics
Northern Illinois University
DeKalb, IL 60155
nader@math.niu.edu

Ehsan S. Soofi
Sheldon B. Lubar School of Business
University of Wisconsin-Milwaukee
P.O.Box 742, Milwaukee, WI 53201
esoofi@uwm.edu

Refik Soyer
Department of Decision Sciences
George Washington University
Washington D.C. 20052
soyer@gwu.edu

Abstract: Lifetime prediction of manufactured items often requires performing a life test and analyzing observed times to failure. A natural question arising in life testing is, "During the conduct of the test do we want to observe more failures or more survivals?" Many engineers probably will say more failures, because intuitively speaking, failures are presumed to provide more information about the parameters of a failure model. It has been pointed out in the literature that this is not always true for the exponential model and that it depends on the particular parameterization that is chosen to model failure times. Our goal is to show that the existing results for the exponential model are general and hold for many lifetime models. We provide results that enable us to compare the information provided by failure and survival times. We give sufficient conditions for observing failures to be more (or less) informative than survivals about the lifetime prediction. Shannon entropy is used as the measure of information.

Index terms: Bayesian predictive distribution, entropy, Lindley's information, prior distribution, posterior distribution.

Notations:

Y	Lifetime of an item
Θ	unknown parameter which may be scalar or vector
θ	a given value of Θ
$f(y\|\theta)$	the conditional probability density function of Y given θ
$f(\theta)$	prior distribution of θ
$H(\Theta)$	the Shannon entropy of $\Theta = -\int f(\theta)\log f(\theta)d\theta$
$S(y\|\theta)$	the conditional survival function of Y given θ
$\lambda(y\|\theta)$	the conditional failure rate of Y given $\theta, = \dfrac{f(y\|\theta)}{S(y\|\theta)}$
$h(\theta)$	$\prod_{i=k+1}^{n} \lambda(y_i\|\theta)$
$\phi(y)$	one-to-one transformation of Y with exponential distribution having the parameter θ
$Y_1, Y_2, \cdots$	a sequence of identical lifetimes and independent conditional on θ
$\boldsymbol{Y}$	$\boldsymbol{Y} = (Y_1, \cdots, Y_n)$ consists of lifetimes of n items to be tested
$\boldsymbol{y}$	$\boldsymbol{y} = (y_1, \cdots, y_n)$ consists of all failure times
$f(\boldsymbol{y}\|\theta)$	the conditional probability density function of $\boldsymbol{Y}$ given θ, $= \int f(\theta) \prod_{i=1}^{n} f(y_i\|\theta)d\theta$
$f(\boldsymbol{y})$	marginal distribution of $\boldsymbol{y}, = \int f(\boldsymbol{y}\|\theta)f(\theta)d\theta$
$f(\theta\|\boldsymbol{y})$	the posterior probability density function of θ given $\boldsymbol{y}$
$H(\Theta\|\boldsymbol{y})$	$-\int f(\theta\|\boldsymbol{y})\log f(\theta\|\boldsymbol{y})d\theta$

t_n	sufficient statistic for θ
$\boldsymbol{y}^*$	$\boldsymbol{y}^* = (y_1, \cdots, y_k, y^*_{k+1}, \cdots, y^*_n$ consist of k failure times $y_1, \cdots, y_k, k < n$ and $n-k$ survival times $y^*_{k+1}, \cdots, y^*_n$ such that $y_i = y^*_i, k = k+1, \cdots n$
$f(\boldsymbol{y}^*\|\theta)$	the conditional probability density function of $\boldsymbol{Y}$ given θ, $= \prod_{i=1}^k f(y_i\|\theta) \prod_{i=k+1}^n S(y_i\|\theta)$
$f(\boldsymbol{y}^*)$	marginal distribution of $\boldsymbol{y}^*$, $= \int f(\theta) f(\boldsymbol{y}^*\|\theta) d\theta$
$f(\theta\|\boldsymbol{y}^*)$	the posterior probability density function of θ given $\boldsymbol{y}^*$
$H(\Theta\|\boldsymbol{y}^*)$	$-\int f(\theta\|\boldsymbol{y}^*) \log f(\theta\|\boldsymbol{y}^*) d\theta$
Y_ν	Lifetime of an untested item
$f(y_\nu)$	prior predictive distribution of $Y_\nu = \int f(y_\nu\|\theta) f(\theta) d\theta$
$f(y_\nu\|\boldsymbol{y})$	posterior predictive distribution of $Y_\nu = \int f(y_\nu\|\theta) f(\theta\|\boldsymbol{y}) d\theta$
$H(Y_\nu)$	the Shannon entropy of $Y_\nu = -\int f(y_\nu) \log f(y_\nu) dy_\nu$
$H(Y_\nu\|\boldsymbol{y})$	$-\int f(y_\nu\|\boldsymbol{y}) \log f(y_\nu\|\boldsymbol{y}) dy_\nu$
$G(\alpha, \beta)$	Gamma distribution with parameters α and β,$= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta)$
$\psi(\alpha)$	digamma function $= \psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$
$H_G(\alpha)$	the entropy of Gamma distribution with parameters α and $\beta = 1$, $= \log \Gamma(\alpha) - (\alpha - 1)\psi(\alpha) + \alpha$
TTE	the time transformed exponential $S(y\|\theta) = \exp(-\theta \log S_0(y))$ and $S_0(y)$ is a survival function
IFR	increasing failure rate
DFR	decreasing failure rate

1 Introduction

Accurate prediction of life times of new items requires performing a life test. This is often accomplished by getting more information about the parameters of a failure model. Information refers to changes in probability distributions of parameters and prediction as a result of data obtained from a life test. The predictive inference is considered a distinguishing feature of Bayesian analysis. But one cannot obtain the predictive distribution without estimation, that is, without obtaining the posterior distribution of the parameters.

Consider a situation in which we are testing n non-repairable items taken randomly from a population. In the typical life-test scenario, we have a fixed time T (time truncation) to run the items to see if they survive or fail. The data obtained are called type I censored data. Another, though much less common, way to test is to decide in advance that we want exactly r failures (item censorship) and then test until they occur. This is called type II censored data. It is clear that both scenarios deprive us the opportunity of observing life history of all n items in the test. In view of this curtailment, many articles quantifying loss of information due to "censorship" have appeared in the literature, see Brooks (1982) and Ebrahimi and Soofi (1990).

In this paper, our focus is not on how much to censor or truncate, but on what we would rather observe, a failure or survival, given a certain amount of fixed or random test time. Clearly, if the answer is the former, then one may choose to induce failures through an accelerated test. Accelerated life-testing involves acceleration of failures with the purpose of assessment of life characteristics of the item at normal use conditions. Of course, many engineers prefer to observe failures over survivals. The intuition for choosing failure over survival is that a failure tells us all we need to know about the item's survival characteristic. Surprisingly, the answer to our question depends on the characteristic life which we wish to learn about, as well as on the model for the lifetime $f(y|\theta)$ that we use and the prior distribution that we have about the parameter $f(\theta)$. In fact, Abel and Singpurwalla (1994) studied the problem for the exponential model for $f(y|\theta)$ and gamma model for prior $f(\theta)$. They compared information about the exponential failure rate and mean failure time $\frac{1}{\theta}$. They concluded that survival is more informative than failure for estimating θ. However, failure is more informative than survival for estimating $\frac{1}{\theta}$. We address this problem at a more general level and present sufficient conditions for failures to be more or less informative than survivals about parameters and prediction of a lifetime.

Lindley (1956) formulated the problem of measuring sample information about the parameters in terms of Shannon entropy and mutual information. Throughout this paper we also use these measures. Evans (1969) discussed usefulness of entropy as a measure of information at a conceptual level. El-Sayyad (1969) used entropy to quantify the amount of information contained in exponential samples about various functions of the exponential parameter. Ebrahimi and Soofi (2004) and Singpurwalla (2006) provide overviews and applications.

This paper is organized as follows. Section 2 presents the measures of information provided by the sample about an unknown quantity. This section also shows comparisons of the information about the model parameter provided by samples of all failures with samples containing both failures and survivals. Section 3 gives results on the comparison of information for prediction of a new lifetime provided by samples of all failures and samples that include both failures and survivals. Section 4 shows applications to several lifetime models. Section 5 gives concluding remarks. Throughout this paper increasing means "nondecreasing" and decreasing means "non-increasing".

2 Information Measures

Let Q denote the unknown quantity of interest with a prior distribution $f(q),\ q \in \mathcal{Q}$. In our case Q can be Θ or Y_ν. Information provided by the life test data D refers to a measure that quantifies changes from $f(q)$ to the posterior distribution $f(q|D)$. We measure information provided by D about Q in terms of Shannon's (1948) entropy. Shannon entropy of the posterior distribution is defined by

$$H(Q|D) = H[f(Q|D)] = -\int_{\mathcal{Q}} f(q|D) \log f(q|D) dq,$$

provided that the integral is finite. If Q is discrete, the integral changes to the sum over all values of Q. The entropy measures closeness of $f(q|D)$ to the uniform distribution over $\mathcal{Q}$. The uniform distribution is the least concentrated distribution, which reflects the most unpredictable (non-informative) situation. As a measure of closeness to uniformity, $H(Q|D)$ is a measure of lack of concentration measuring uncertainty in the sense of unpredictability of the outcomes of Q based on $f(q|D)$.

2.1 Comparison of Survival and Failure

We consider the following two scenarios for D:

1. *Failure scenario*, $D = \boldsymbol{y} = (y_1, \cdots, y_n)$ consist of all failure times.

2. *Survival scenario*, $D = \boldsymbol{y}^* = (y_1, \cdots, y_k, y^*_{k+1}, \cdots, y^*_n)$ consist of k failure times $y_1, \cdots, y_k$, $k < n$ and $n - k$ survival times $y^*_{k+1}, \cdots, y^*_n$, such that $y^*_i = y_i,\ i = k+1, \cdot, n$.

The assumption of each survival time in the second scenario is equal to a failure time in the first scenario, $y^*_i = y_i$, ensures that the comparison is about the types, but not the magnitudes, of the observations. These scenarios depict, for example, two experiments which include two identical sets of $n-1$ failure records, but the nth record $y_n = t$ is a failure for one experiment and is a survival $y^*_n = t$ for the other experiment. Are these two experiments different in information that they provide for inferences about the parameter and prediction?

The sample of all failures $D = \boldsymbol{y}$ and the sample of some failures and some survivals $D = \boldsymbol{y}^*$ provide two different posterior distributions $f(q|\boldsymbol{y})$ and $f(q|\boldsymbol{y}^*)$. Since the prior is common in both cases, the two types of life tests can be compared according to the informativeness of the respective posterior distributions. The sample of all failures $\boldsymbol{y}$ is more informative about Q than the sample of some failures and some survivals $\boldsymbol{y}^*$ whenever

$$H(Q|\boldsymbol{y}^*) > H(Q|\boldsymbol{y}), \tag{1}$$

irrespective of the prior distribution $f(\theta)$. That is, $\boldsymbol{y}$ is more informative if it produces a more concentrated posterior than $\boldsymbol{y}^*$. The comparison (1) is well-defined for improper priors such as Jeffreys' prior, as long as the posterior density functions $f(q|\boldsymbol{y})$ and $f(q|\boldsymbol{y}^*)$ are proper.

The comparison of the two scenarios according to the observed sample information measures can be extended to the comparions according to the expected sample information measures by viewing the entropies in (1) as functions of $\boldsymbol{y}$ and $\boldsymbol{y}^*$ and averaging each posterior entropy with respect to the respective marginal distribution of the data. The average posterior entropy is referred to in general terms as the *conditional entropy.* For the sample of all failures, the conditional entropy is given by

$$\mathcal{H}(Q|\boldsymbol{Y}) = E_{\boldsymbol{y}}\{H(Q|\boldsymbol{y})\} = \int H(Q|\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{y},$$

where $E_{\boldsymbol{y}}$ denotes averaging with respect to $f(\boldsymbol{y})$. The conditional entropy for $\boldsymbol{y}^*$ is defined similarly. On average, $\boldsymbol{y}$ is more informative about Q than $\boldsymbol{y}^*$ if

$$\mathcal{H}(Q|\boldsymbol{Y}^*) > \mathcal{H}(Q|\boldsymbol{Y}). \tag{2}$$

This inequality is also well-defined for improper priors, if in addition to the posterior density functions $f(q|\boldsymbol{y})$ and $f(q|\boldsymbol{y}^*)$, the marginal distributions $f(\boldsymbol{y})$ and $f(\boldsymbol{y}^*)$ are proper.

When $f(\theta)$ is a proper prior distribution, we can compute the observed sample information provided by each sample about Q defined by the change in entropy due to a revision of $f(q)$ to its posterior distribution. The observed information provided by the sample of all failures is

$$\Delta H(\boldsymbol{y}; Q) = H(Q) - H(Q|\boldsymbol{y}).$$

The observed information provided by $\boldsymbol{y}^*$ is computed similarly. Thus, (1) can be equivalently stated as: $\boldsymbol{y}$ is more informative about Q than $\boldsymbol{y}^*$ whenever

$$\Delta H(\boldsymbol{y}; Q) > \Delta H(\boldsymbol{y}^*; Q).$$

With a proper $f(\theta)$, we can also compute the expected entropy difference, known in general terms as the *mutual information.* For the sample of all failures, the mutual information is given by

$$M(\boldsymbol{Y}; Q) = E_{\boldsymbol{y}}\{\Delta H(\boldsymbol{y}; Q)\} = H(Q) - \mathcal{H}(Q|\boldsymbol{Y}) \geq 0, \tag{3}$$

the equality holds if an only if Q and $\boldsymbol{Y}$ are independent. The expected sample information is invariant under one-to-one transformations of Q and $\boldsymbol{Y}$. The expected sample information measure for $\boldsymbol{y}^*$ is defined similarly. Thus, (2) can be equivalently stated as: on average the sample of all failures $\boldsymbol{y}$ is more informative about Q than the sample of some failures and some survivals $\boldsymbol{y}^*$ whenever

$$M(\boldsymbol{Y}; Q) > M(\boldsymbol{Y}^*; Q).$$

The expected sample information about the parameter, $M(\boldsymbol{Y}; \Theta)$ is known as Lindley's measure (Lindley 1956) and is referred to as the parameter information. The expected information in the data for prediction $M(\boldsymbol{Y}; Y_\nu)$ is referred to as the predictive information (San Martini and Spezzaferri 1984, Amaral and Dunsmore 1985).

2.2 Parameter Information

Comparison study of information between survival and failure data was initiated by Abel and Singpurwalla (1994). They considered the exponential model $f(y|\theta) = \theta e^{-\theta y}$ and compared the information provided by an observed data point $D = y$ and $D = y^*$ about $Q = \theta$ and $Q = \mu(\theta) = \theta^{-1}$. To compute $\Delta H(y; \Theta)$, they used the gamma prior $G(\alpha, \beta)$ with density function

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}. \tag{4}$$

This prior induces the inverted gamma prior for $\mu(\theta) = \theta^{-1}$.

We consider models in the exponential family that provide likelihood functions in the form of

$$\mathcal{L}(\theta) \propto \theta^{na} e^{-\theta t_n}, \quad \theta > 0, \tag{5}$$

where $t_n = t_n(\boldsymbol{y})$ is a sufficient statistic for θ and a is a constant assumed to be known. This family is very broad and includes many well-known parametric families. Several members of this family and their sufficient statistics are shown in Table 1. The lognormal model is also in this family. An important class of models whose likelihood functions are in the form of (5) with $a = 1$, is the time-transformed exponential (TTE) models defined by the survival function $S(y|\theta) = \exp\{-\theta \log S_0(y)\}$, $y \geq 0$ where θ is the "proportional hazard" (Barlow and Hsiung 1983). Members of TTE family in Table 1 are the exponential, Pareto, Weibull, and the linear failure rate. The family (5) also includes the gamma distribution and all of its one-to-one transformations such as the Half-normal and generalized gamma models shown in Table 1.

The gamma prior (4) is conjugate for the family (5). The posterior distribution based on the sample of all failures $\boldsymbol{y}$ is $f(\theta|\boldsymbol{y}) = G(\alpha + na, \beta + t_n)$. The limiting case of (4) is the Jeffreys prior $f(\theta) \propto \frac{1}{\theta}$, which is improper. However, the posterior distribution is $f(\theta|\boldsymbol{y}) = G(na, t_n)$ which is proper. The posterior distribution based on $\boldsymbol{y}^*$ is $f(\theta|\boldsymbol{y}^*) = G(\alpha + k, t_n^* + \beta)$, where $t_n^* = t_n$ due to

Table 1. Examples of Models for Applications of the Family (5) and Theorems 1 and 2.

Model	Density and Support	Sufficient statistic for parameter Family (5)	More informative outcome for prediction Theorem 1	Theorem 2
Exponential (θ)	$f(y\|\theta) = \theta e^{-\theta y},\ y \geq 0$	$\sum y_i$	Failure	
Pareto Type I (a, θ)	$f(y\|a,\theta) = \frac{a^\theta}{y^{\theta+1}},\ y \geq a > 0$	$\sum \log(y_i/a)$	Failure	
Pareto Type II (θ)	$f(y\|\theta) = \frac{1}{(1+y)^{\theta+1}},\ y \geq 0$	$\sum \log(1+y_i)$	Failure	
Half-normal (θ)	$f(y\|\theta) = \sqrt{\frac{2\theta}{\pi}} e^{-\frac{\theta}{2}y^2},\ y \geq 0$	$\frac{1}{2}\sum y_i^2$	Failure	
Exponential Power	$f(y\|a,\theta) = a\theta y^{a-1} e^{1+\theta y^a - \exp(\theta y^a)},\ y \geq 0$	N/A	Failure	
Gamma (a, θ)	$f(y\|a,\theta) = \frac{\theta^a}{\Gamma(a)} y^{a-1} e^{-\theta y},\ y \geq 0$	$\sum y_i$	$a \leq 1$, Failure	$a > 1$, Survival
Weibull (b, θ)	$f(y\|b,\theta) = b\theta y^{b-1} e^{-\theta y^b},\ y \geq 0$	$\sum y_i^b$	$b \leq 1$, Failure	$b > 1$, Survival
Generalized gamma (a, b, θ)	$f(y\|a,b,\theta) = \frac{b\theta^a}{\Gamma(a)} y^{ab-1} e^{-\theta y^b},\ y \geq 0$	$\sum y_i^b$	$ab \leq 1$, Failure	$ab > 1$, Survival
Power (θ)	$f(y\|\theta) = \theta y^{\theta-1},\ 0 < y \leq 1$	$\sum \log y_i$	$\theta < 1$, Failure; $\theta > 1$, Survival	
Linear Failure Rate (a, θ)	$f(y\|a,\theta) = 2a\theta(1+ay) e^{-\theta(1+ay)^2},\ y \geq 0$	$\sum (1+ay_i)^2$	$\theta \geq 2a$, Failure	Survival
Extreme Value	$f(y\|a,\theta) = a e^{a\theta} e^{(a/\theta)[1-\exp(\theta y)]},\ y \geq 0$	N/A	$\theta \leq a$, Failure	Survival

the assumption that the survival and failure times are assumed identical under the two scenarios. The posterior entropies for the samples under the two scenarios are given by

$$H(\Theta|t_n) = H_G(\alpha + n_s a) - \log(\beta + t_n), \quad \alpha, \beta \geq 0, \quad n_s = k, n. \tag{6}$$

The difference between the information about the scale parameter provided by the observed sample under the two scenarios is

$$H_G(\alpha + ka) < H_G(\alpha + na), \quad k < n, \quad \alpha \geq 0. \tag{7}$$

Since the gamma entropy H_G is increasingly ordered by the shape parameter (Ebrahimi et al. 1999), observing survival is always more informative than observing failure about the scale parameter for all distributions in the family (5).

Recall that (7) is not invariant under nonlinear transformations, so the observed information ranks the two scenarios differently for different functions of the θ. For example, for comparing information provided by sample in the two scenarios about the mean of the exponential model $\mu(\theta) = \frac{1}{\theta}$ or about the variance of the Half-normal model $\sigma^2(\theta) \propto \frac{1}{\theta}$, the entropies in (6) change to the entropies of the inverse-gamma distributions and the inequality in (7) reverses due to ordering of the inverted gamma distribution shown in Ebrahimi et al. (1999).

We also should note that the proper prior (4) is needed for computing the observed and expected information about the parameter based on the two scenario. The information in the observed samples are given by subtracting (6) from the prior entropy $H_G(\alpha) - \log\beta$. Thus, unlike the comparison of the two scenario which only depends on the sample size n and the number of survivals k, the observed information measure depends on sufficient statistic t_n. Similarly, the expected information in the samples depend on the prior entropy and the expected value of t_n.

3 Predictive Information

In this section we develop results for comparison information provided by $\boldsymbol{y}$ and $\boldsymbol{y}^*$ about prediction of the lifetime of a new item $Q = y_\nu$. We use the following definitions.

Definition 1

(a) The random variable X with survival function $S_1(x)$ is said to be stochastically less than or equal to random variable Y with survival function $S_2(y)$, denoted by $X \overset{ST}{\leq} Y$, if $S_1(v) \leq S_2(v)$ for all v.

(b) The distribution F of a random variable X is said to be increasing (decreasing) failure rate, (IFR (DFR)) if the failure rate $\lambda(x)$ is increasing (decreasing) in $x \geq 0$.

The next result gives a sufficient condition for the entropy ordering under stochastic ordering.

Lemma 1 *Let X and Y be two random variables with stochastic order $X \overset{st}{\leq} Y$. If the density function of Y is decreasing (increasing), then $H(X) \leq (\geq) H(Y)$.*

Proof. We show the proof for the decreasing case. The proof the increasing case is similar. Let F_1 and F_2 be the distribution of X and Y, respectively.

$$\begin{aligned} -H(X) &= \int f_1(u)\log f_1(u)du \\ &\geq \int f_1(u)\log f_2(u)du \end{aligned}$$

$$\geq \int f_2(u) \log f_2(u) du = -H(Y),$$

where f_1 and f_2 are the probability density functions of X and Y respectively. The first inequality is implied by Kullback-Leibler information

$$K(f_1 : f_2) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} dx \geq 0, \tag{8}$$

and the second inequality is implied by the assumption that $f_2(y)$ is a decreasing function.

This result is the univariate version of a result in Asadi et al. (2010). The condition of monotone univariate density function in Lemma 1 replaces a complicated functional assumption about the multivariate density function.

Theorem 1 *Let $y_i, \ i = 1, \cdots, k$ denote data points representing failure and $y_i^*, \ i = k+1, \cdots, n$ denote data points representing either failure or survival. If the predictive density function $f(y_\nu|\boldsymbol{y})$ is decreasing (increasing), then a survival at y_i^* is less (more) informative than a failure about the prediction of Y_ν, if and only if $cov(S(y|\Theta), \ h(\Theta)|\boldsymbol{y}^*) < 0$.*

Proof. Let $\boldsymbol{y}^* = (y_1, \cdots, y_k, y_{k+1}^*, \cdots, y_n^*)$. Then,

$$f(\theta|\boldsymbol{y}^*) = \frac{f(\boldsymbol{y}^*, \theta)}{f(\boldsymbol{y}^*)},$$

where

$$f(\boldsymbol{y}^*, \theta) = f(\theta) \prod_{i=1}^{k} f(y_i|\theta) \prod_{i=k+1}^{n} S(y_i|\theta),$$

and $f(\boldsymbol{y}^*) = \int_\Theta f(\boldsymbol{y}^*, \theta) d\theta$. The posterior distribution based on the sample with all failures $\boldsymbol{y}$ is given by

$$f(\theta|\boldsymbol{y}) = \frac{f(\boldsymbol{y}, \theta)}{f(\boldsymbol{y})} = \frac{f(\boldsymbol{y}^*, \theta) h(\theta)}{f(\boldsymbol{y})}. \tag{9}$$

The second equality in (9) is obtained by noting that $f(y|\theta) = S(y|\theta)\lambda(y|\theta)$, so

$$\begin{aligned} f(\boldsymbol{y}, \theta) &= f(\theta) \prod_{i=1}^{n} (y_i|\theta) \\ &= f(\theta) \prod_{i=1}^{k} f(y_i|\theta) \prod_{i=k+1}^{n} S(y_i|\theta)\lambda(y_i|\theta) \\ &= f(\boldsymbol{y}^*, \theta) \prod_{i=k+1}^{n} \lambda(y_i|\theta). \end{aligned}$$

The difference between the posterior densities under the two scenarios is

$$f(\theta|\boldsymbol{y}) - f(\theta|\boldsymbol{y}^*) = \left[\frac{f(\boldsymbol{y}^*)}{f(\boldsymbol{y})} h(\theta) - 1 \right] f(\theta|\boldsymbol{y}^*). \tag{10}$$

The difference between the two posterior predictive survival functions is

$$
\begin{aligned}
S(y_\nu|\boldsymbol{y}) - S(y_\nu|\boldsymbol{y}^*) &= \int_y^\infty \int_\Theta f(u|\theta)[f(\theta|\boldsymbol{y}) - f(\theta|\boldsymbol{y}^*)]d\theta du \\
&= \int_\Theta S(y_\nu|\theta) \left[\frac{f(\boldsymbol{y}^*)}{f(\boldsymbol{y})} h(\theta) - 1\right] f(\theta|\boldsymbol{y}^*) d\theta \\
&= \frac{f(\boldsymbol{y}^*)}{f(\boldsymbol{y})} \operatorname{cov}\left(S(y_\nu|\Theta),\ h(\Theta)|\boldsymbol{y}^*\right),
\end{aligned}
$$

where the second equality is from changing the order of integration and using (10), and the last equality is due to the linear property of the covariance and by (10), $E_{\theta|\boldsymbol{y}^*}\left[\frac{f(\boldsymbol{y}^*)}{f(\boldsymbol{y})} h(\theta) - 1\right] = 0$. Thus, $S(y_\nu|\boldsymbol{y}) - S(y_\nu|\boldsymbol{y}^*) < 0$ if and only if $\operatorname{cov}\left(S(y_\nu|\Theta), h(\Theta)|\boldsymbol{y}^*\right) < 0$, and we obtain the results by Lemma 1.

We should note that all DFR distributions have decreasing density functions and mixtures of DFR distributions are also DFR. Thus, if the model $f(y|\theta)$ is DFR, then the predictive density is also DFR, and the decreasing condition in Theorem 1 can be replaced with the stronger condition that $f(y|\theta)$ is DFR. However, this cannot be said for the IFR distributions. Next, we give results for IFR predictive distributions.

Lemma 2 *Let X and Y be two random variables with stochastic order $X \overset{st}{\leq} Y$. If Y has an IFR distribution, then $H(X) \geq H(Y)$.*

Proof. Let F_1 and F_2 be the distribution of X and Y, respectively. By (8),

$$
-H(Y) = \int f_1(u) \log f_1(u) du \geq \int f_1(u) \log f_2(u) du.
$$

Now,

$$
\begin{aligned}
\int f_1(u) \log f_2(u) du &= \int f_1(u) \log \lambda_2(u) du + \int f_1(u) \log S_2(u) du \\
&\geq \int f_1(u) \log \lambda_2(u) du + \int f_1(u) \log S_1(u) du \\
&\geq \int f_2(u) \log \lambda_2(u) du + \int f_2(u) \log S_2(u) du = -H(X).
\end{aligned}
$$

This first inequality is from the stochastic order assumption and the last inequality is due to the IFR property and the fact that $\int f_i(u) \log S_i(u) du = -1,\ i = 1, 2$.

Theorem 2 *Let $y_i,\ i = 1, \cdots, k$ denote data points representing failure and $y_i^*,\ i = k+1, \cdots, n$ denote data points representing either failure or survival. If the predictive density function $f(y_\nu|\boldsymbol{y})$ is IFR, then a survival at y_i^* is more informative than a failure about the prediction of Y_ν, if and only if $cov(S(y|\Theta),\ h(\Theta)|\boldsymbol{y}^*) < 0$.*

Proof. . Using Lemma 2 and following the steps in the proof of Theorem 1 give the result.

Some remarks are in order.

(a) It is clear that by Theorem 1 a survival at y_i^* is more (less) informative than a failure about the prediction of Y_ν, if and only if $\text{cov}\,(S(y|\Theta),\ h(\Theta)|\boldsymbol{y}^*) > 0$.

(b) Theorems 1 and 2 give sufficient conditions.

(c) Since Theorems 1 and 2 hold for each outcome of type $\boldsymbol{y}$ and $\boldsymbol{y}*$, the results hold for the comparison according to the expected information 2, provided that the marginal density functions $f(\boldsymbol{y})$ and $f(\boldsymbol{y}^*)$ are proper.

4 Applications

In applications, verification of $\text{cov}\,(S(y|\Theta),\ h(\Theta)|\boldsymbol{y}^*) < \ 0$ can be difficult. A more restrictive, but easy to check condition is as follows. If θ orders the failure rate $\lambda(y|\theta)$ for all y, then it orders $S(y|\theta) = \exp\left\{-\int_0^y \lambda(u|\theta)du\right\}$ in reverse, and hence $\text{cov}\,(S(y|\Theta),\ h(\Theta)|\boldsymbol{y}^*) < \ 0$. Table 1 shows examples of distributions where θ orders the failure rate, thus satisfy the condition $\text{cov}\,(S(y|\Theta),\ h(\Theta)|\boldsymbol{y}^*) < \ 0$.

Among the distributions listed in Table 1, the density functions of exponential, Pareto, Half-normal, and power exponential models are decreasing in y. For $a \leq 1$ ($b \leq 1$), the gamma (Weibull) density function is deceasing in y. The same result holds for the generalized gamma model with $ab \leq 1$. These models satisfy conditions of Theorem 1 without any restriction on θ. Thus, for these models under any prior distribution for θ, the predictive density is decreasing in y, and by Theorem 1, survival is less informative than the failure about prediction.

The density function of the power family is decreasing in y for $0 < \theta \leq 1$ and increasing in y for $\theta \geq 1$. Thus, under any prior distribution such that $P(\Theta \leq 1) = 1$, by Theorem 1, survival is less informative than failure about prediction of the lifetime. However, under any prior distribution such that $P(\Theta \geq 1) = 1$, survival is more informative than failure about prediction of the lifetime. The density functions of the linear failure rate and extreme value models are decreasing in y for $\theta \geq 2a$ and $\theta \geq a$, respectively. Thus, for the linear failure rate model, under any prior distribution such that $P(\Theta \geq 2a) = 1$, by Theorem 1, survival is less informative than failure about prediction of the lifetime. The same result holds for the extreme value model with $P(\Theta \leq a) = 1$. Theorem 1 is also applicable to the gamma and Weibull models when θ is given, but the shape parameters a and b are unknown, and are dustributed according to priors such that $P(A \leq 1) = 1(P(B \leq 1) = 1)$. Since the failure rate is ordered by $a(b)$, we have $\text{cov}\,(S(y|a),\ h(a)|\boldsymbol{y}^*) < 0$ $(\text{cov}\,(S(y|b),\ h(b)|\boldsymbol{y}^*) < 0)$. Thus, under any prior for the shape parameter $a(b)$ such that $P(A \leq 1) = 1(P(B \leq 1) = 1)$, the conditions of Theorem 1 are satisfied and a failure is more informative than a survival about prediction.

The last column of Table 1 indicates application of Theorem 2 to the models whose density functions are not always monotone, so Theorem 1 is not always applicable. For these models we use results from Lynch (1999) who showed that the mixture of IFR distributions with log concave survival functions in the parameter is closed under mixing with distributions that are also IFR or have log concave densities. By this result, Theorem 2 is applicable to the IFR distributions with log concave survival functions in the parameter under any prior distribution which is also IFR or the prior density is log concave. The survival functions of the IFR distributions shown in Table 1 are not log concave in θ. However, Block et al. (2003) showed that the survival functions of the Extreme value (Gompertz) and the IFR Gamma are log concave in $\xi = \frac{1}{\theta}$, and the IFR Weibull and IFR Generalized Gamma survival functions are log concave in $\xi = \frac{1}{\theta^{b-1}}$. Similarly, it can be shown that the the survival functions of the linear failure rate model is log concave in $\xi = \frac{1}{\theta}$. Thus with these reparameterizations, using any prior distribution for ξ which is IFR or the prior density $f(\xi)$ is log concave, Theorem 2 is applicable to the IFR models shown in Table 1. For example, the result is applicable for the uniform prior over the parameter space $\Xi = \{\xi : 0 < \xi \leq \xi_0\}$. Under such priors, we can conclude that for these models, survival is more informative than failure about prediction.

5 Conclusion

Abel and Singpurwalla (1994) posed the interesting and important question of which observation, a failure at a given time point or a survival at the same time point is more informative about a parameter of the lifetime distribution of an item. Their parameters of interest were the mean and failure rate of the exponential model, and they used proper priors for the parameters. We extended their findings to the parameter of a broad class of models in the exponential family and to the samples of n observations, using a formulation that allows proper as well as improper prior for the parameter.

Our main results pertain to the comparison of the informativeness of survival and failure observations about prediction of the lifetime of a new item. We provided solutions for the question of which of types of outcomes, failures or survivals are more informative about prediction of a lifetime, without assuming any specific model for the data or prior distribution for the parameter. A result on entropy ordering under stochastic dominance led to identifying some sufficient conditions in terms of monotonicity of the predictive density function. Depending on the general analytic of the model a failure can be more or less informative than a survival about prediction of a lifetime. If the lifetime model has a decreasing (increasing) density function and the model parameter orders the failure rate, then the failure is more (less) informative than survival about the prediction of the lifetime of a new item. Another result on entropy ordering led to identifying some sufficient conditions in terms of IFR for the predictive distribution and similar condition for the prior density function. For the IFR distributions where the identified conditions hold, survival is more informative than failure about the prediction of the lifetime of a new item.

We illustrated that our results are applicable to many distributions actually used as lifetime models for comparing the informativeness of survival and failure observations about the model parameter and about predicting a the lifetime of a new item.

References

Abel, P.S., and Singpurwalla, N.D. (1994) To survive or to fail: that is the question. *The American Statistician*, 48, 18-21.

Amaral-Turkman, M.A., and Dunsmore, I. (1985) Measures of information in the predictive distribution. In *Bayesian Statistics, 2*, (eds J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith), 603-612.

Asadi, M., Ashrafi, S., Ebrahimi, N., and Soofi, E.S. (2010). Models based on partial information about survival and hazard gradient. *Probability in the Engineering and Informational Sciences*, in press.

Asadi, M., Ebrahimi, N., Hamedani, G.G. and Soofi, E.S. (2004). Maximum dynamic entropy models. *Journal of Applied Probability*, 41, 379-390.

Barlow, R.E. and Hsiung, J.H. (1983). Expected information from a life test experiment. *The Statistician*, 48, 18-21.

Block, H.W., Li, Y., and Savits, T.H. (2003). Preservations of properties under mixture. *Probability in the Engineering and Informational Sciences* 17, 205-212.

Brooks, R. J. (1982). On loss of information through censoring. *Biometrika*, 69, 137-144.

Ebrahimi, N. and Soofi, E. S. (1990). Relative information loss under type II censored exponential data. *Biometrika*, 77, 429-435.

Ebrahimi, N. and Soofi, E. S. (2004). Information measures for reliability. *Mathematical Reliability: An Expository Perspective*, Eds. R. Soyer ,T.A. Mazzuchi, and N. D. Singpurwalla, pp. 127-159. Netherlands: Kluwer.

Ebrahimi, N., Maasoumi, E. and Soofi, E. S. (1999). Ordering univariate distributions by entropy and variance. *Journal of Econometrics*, 90, 317-336.

El-Sayyed, G. M. (1969). Information and sampling from exponential distribution. *Technometrics*, 11, 41-45.

Evans, R.A. (1969). The principle of minimum information, *IEEE Transactions on Reliability*, 87-90.

Kullback, S. (1959). *Information Theory and Statistics*, New York: Wiley (reprinted in 1968 by Dover).

Lindley, D.V. (1956). On a measure of information provided by an experiment. *The Annals of Mathematical Statistics*, 27, 986-1005.

Lynch, J.D. (1999). On conditions for mixtures of increasing failure rate distributions to have an increasing failure rate. *Probability in the Engineering and Informational Sciences* 13, 33-36.

San Martini, A. and Spezzaferri, F. (1984). A predictive model selection criteria. *Journal of the Royal Statistical Society Ser. B*, 46, 296-303.

Singpurwalla, N.D. (2006). *Reliability and Risk: A Bayesian Perspective*, New York: Wiley.