مجلّه‌ی پژوهش‌های آماری ایران
J. Statist. Res. Iran **3** (2006): 113-137
دوره‌ی ۳، شماره‌ی ۲، پاییز و زمستان ۱۳۸۵، صص ۱۱۳-۱۳۷
پژوهشکده‌ی آمار

Dynamic Bayesian Information Measures

Nader Ebrahimi[†], S. N. U. A. Kirmani[‡], and Ehsan S. Soofi[⋆,*]

[†] Northern Illinois University
[‡] University of Northern Iowa
[⋆] University of Wisconsin-Milwaukee

Invited Paper

Abstract. This paper introduces measures of information for Bayesian analysis when the support of data distribution is truncated progressively. The focus is on the lifetime distributions where the support is truncated at the current age $t \geqslant 0$. Notions of uncertainty and information are presented and operationalized by Shannon entropy, Kullback-Leibler information, and mutual information. Dynamic updatings of prior distribution of the parameter of lifetime distribution based on observing a survival at age t and observing a failure or the residual lifetime beyond t are presented. Dynamic measures of information provided by the data about the parameter of lifetime distribution, and dynamic predictive information are introduced. These measures are applied to two well-known lifetime models. The paper concludes with some remarks on use of generalized uncertainty and information measures, and some topics for further research.

Keywords. Bayesian information; Entropy; Exponential; Gamma prior; Kullback-Leibler information; Mutual information; Reliability; Residual life; Weibull.

1 Introduction

This paper integrates two lines of information theoretic research: dynamic information measures and Bayesian information measures. In lifetime studies,

* Corresponding author

consideration of the current age truncates the support of lifetime distribution progressively and leads to the past and remaining lifetime distributions where the age becomes a parameter. The information measures of the truncated distributions are functions of time, and thus are dynamic. Several authors have considered information functions that take age into account, see, for example, Ebrahimi (1996), Ebrahimi and Kirmani (1996 a, b), Di Crescenzo and Longobardi (2002), Belzunce et al., (2004), Asadi, Ebrahimi, and Soofi (2005), Asadi et al., (2004, 2005), and Ebrahimi. Kirmani, and Soofi (2007). This line of research has provided some important results for lifetime models.

Bayesian information measures are for an observation y made or to be made from a random variable Y having a distribution $F_{Y|\theta}(y|\theta)$ about the parameter θ when the prior belief can be described by a probability distribution with density $f_\Theta(\theta)$. The objective is to measure information provided by data about the parameter. Bayesian information measures are for the purposes such as design comparison, data evaluation, model comparison, and prior construction; see for example, Lindley (1956), Zellner (1971, 1977), Bernardo (1979), Goel and DeGroot (1979), Polson (1992), Soofi (1994, 2000), Singpurwalla (1997), Yuan and Clarke (1999), Sebastiani and Wynn (2000). We develop dynamic versions of several Bayesian information measures. Given a prior distribution $f_\Theta(\theta)$ for the parameter of the lifetime model $F_{Y|\theta}(y|\theta)$, observing a failure time $y > t$, the age t induces dynamic into the posterior distribution of θ and leads to dynamic Bayesian information measures. Either or both y and θ may be vectors, but to simplify notation, we discuss the case when both are scalar.

Section 2 presents the notion of uncertainty and information functions. Section 3 presents dynamic updating of prior distribution of the parameter of lifetime distribution based on observing a failure beyond age t and dynamic Bayesian information measures in the data about the parameter. Section 4 presents dynamic sequential updating of prior distribution of the parameter of lifetime distribution first based on observing a survival at age t and then observing the residual lifetime beyond t, and the decomposition of total information measure about the parameter into information for the two stages. Section 5 presents dynamic information for the predictive distribution and dynamic version of Zellner's Maximal Data Information Prior (MDIP) criterion. Section 6 presents applications to the exponential and Weibull distributions. Section 7 gives some concluding remarks.

2 Uncertainty and Information Functions

The notion of information refers to the ease of predictability of an unknown prospect having a probability distribution F. The distribution maps uncertainty about all values of the unknown prospect in the range of its possible values $\mathcal{S}$, referred to as the support of the distribution. The overall uncertainty of F is measured by an uncertainty function mapping the difficulty of predicting the unknown prospect using F. The notions of information and uncertainty are relative and involve comparison of F with another distribution, referred to as the reference distribution. We define uncertainty as follows.

The uncertainty associated with a probability distribution F having a density or mass function f is defined by a scaler function $\mathcal{U}(f)$ such that:

(a) $\mathcal{U}(f)$ is concave in f;

(b) $\mathcal{U}(f) \leqslant \mathcal{U}(f^*)$,

where f^* is the uniform density (possibly improper).

This definition is a modification of the uncertainty function defined by Goel and DeGroot (1981) where (a) is the only requirement for $\mathcal{U}(f)$.

The concavity condition (a) implies reduction of uncertainty by averaging. By (b), $\mathcal{U}(f)$ measures the uniformity (lack of concentration) of probabilities under F, where the uniform distribution reflects the most unpredictable situation, when we are unable to forecast in favor or against any values for the unknown prospect. Thus we have no reason to assess if any value is more or less likely than the others and invoke Laplace's "Principle of Insufficient Reason", assigning equal probabilities to all possible values (intervals of equal width in the continuous case). The uniform distribution is the global reference distribution for quantifying uncertainty in terms of unpredictability. This definition is easily extendible to a restricted set such as the set of all distributions having certain moment values where f^* is the density of the maximum uncertainty distribution in the set; the concavity of $\mathcal{U}$ ensures uniqueness of $\mathcal{U}(f^*)$.

The uncertainty function $\mathcal{U}(f)$ includes some well-known measures such as Shannon entropy (Shannon, 1948) and Rényi entropy (Rényi, 1961). The concavity condition (a) includes variance as an uncertainty measure (Goel and DeGroot, 1981). But the uniformity condition (b) excludes variance, in general. First, variance does not necessarily map uniformity of probabilities and hence does not map uncertainty in the sense of difficulty of predicting outcomes; e.g., beta distribution $Be(\alpha, \beta)$ with $\alpha, \beta < 1$, see Ebrahimi, Maasoumi, and Soofi (1999 a, b). Second, uncertainty function provides a scaler measure for the multivariate case, but the natural extension of variance is a variance-covariance matrix, which cannot be summarized uniquely. Third, when the distribution is heavy tail, the variance is not finite and a finite uncertainty measure is need. However, for some distributions like normal and exponential, uncertainty can be measured by variance or a monotone function of it.

Two distributions F_1 and F_2 are compared by an information discrepancy function

$$\mathcal{D}(f_1 : f_2) \geqslant 0,$$

such that $\mathcal{D}(f_1 : f_2)$ is convex in f_1 and $\mathcal{D}(f_1 : f_2) = 0$ if and only if $f_1(x) = f_2(x)$ almost everywhere. An information discrepancy function maps how different are the two distributions. But it does not indicate which of the two distributions is more informative (more concentrated).

A discrepancy function between F and the uniform distribution $\mathcal{D}(f : f^*)$ quantifies the information associated with a probability distribution F. An example of such information measures is the uncertainty difference,

$$\mathcal{D}(f : f^*) = \Delta\mathcal{U}(f : f^*) = \mathcal{U}(f^*) - \mathcal{U}(f) = \mathcal{I}(f) - \mathcal{I}(f^*) \geqslant 0,$$

provided that $\mathcal{U}(f^*) < \infty$. The quantity $\mathcal{I}(f) = -\mathcal{U}(f)$ is referred to as the information function of F.

The uncertainty difference also provides comparison of information for two distributions F_1 and F_2,

$$\Delta\mathcal{D}(f_1 : f_2) = \Delta\mathcal{U}(f_1 : f^*) - \Delta\mathcal{U}(f_2 : f^*) = \mathcal{I}(f_1) - \mathcal{I}(f_2).$$

Note that when $\mathcal{U}(f_j) < \infty,\ j = 1, 2,$

$$\Delta\mathcal{I}(f_1 : f_2) = \mathcal{I}(f_1) - \mathcal{I}(f_2)$$

is well-defined even when $\mathcal{U}(f^*) = \infty$. However, $\Delta\mathcal{I}(f_1 : f_2)$ can be positive or negative depending upon which of the two distributions is more informative (concentrated).

An important question in statistics is to what extent the use of a variable Y reduces uncertainty about predicting the outcomes of another variable X; Retzer, Soofi, and Soyer (2008) provides a comprehensive treatment of this topic. When Y is stochastic with distribution F_Y, information provided by an observation $Y = y$ about predicting outcomes of X can be measured by an information discrepancy function

$$\mathcal{D}(f_{X|y} : f_X) \geqslant 0, \tag{1}$$

where f_X is the marginal distribution of X and $f_{X|y}$ is the conditional distribution of X given y. The information discrepancy function maps how different are the marginal and conditional distributions. It does not indicate which of the two distributions, the conditional density $f_{X|y}$ or the marginal density f_X is more informative.

Alternatively, information provided by an observation $Y = y$ about predicting outcomes of X can be measured by the uncertainty difference

$$\mathcal{I}(X|y) \equiv \Delta\mathcal{I}(f_{X|y} : f_X) = \Delta\mathcal{U}(f_X : f_{X|y}) = \mathcal{U}(f_X) - \mathcal{U}(f_{X|y}),$$

Note that $\Delta\mathcal{I}(f_{X|y} : f_X)$ can be positive (negative) when conditional density $f_{X|y}$ is farther (closer) to uniformity than the marginal density f_X.

The *expected information* provided by outcomes of Y for prediction of X is defined by

$$\mathcal{I}(X|Y) \equiv E_y \left[\mathcal{U}(f_X) - \mathcal{U}(f_{X|y})\right] \geqslant 0, \tag{2}$$

where E_y denotes the expectation with respect to the marginal density f_Y. The inequality is implied by concavity of $\mathcal{U}(\cdot)$ and Jensen's inequality, and the equality holds if and only if X and Y are independent. It is reasonable to require that outcomes of Y, on average, contain some information about prediction of X, and at worst, in the long-run, use of a variable provides no information for predicting outcomes of another variable (DeGroot, 1962).

The *shared information* between a pair of random prospects (X, Y) having joint distribution $F_{X,Y}$ is defined by

$$\mathcal{C}(X,Y) \equiv \mathcal{U}(f_X) + \mathcal{U}(f_Y) - \mathcal{U}(f_{X,Y}). \tag{3}$$

This quantity can be positive, negative, or zero. An uncertainty function $\mathcal{U}(\cdot)$ is said to be *subadditive* if $\mathcal{C}(X,Y) > 0$.

2.1 Examples of Uncertainty and Information Measures

The most well-known example of uncertainty functions is Shannon's entropy

$$H(Y) \equiv H(f_Y) = -\int_S \log f_Y(y)\ dF_Y(y). \tag{4}$$

The entropy of uniform distributions over a finite set and finite interval are finite. In this case

$$\mathcal{D}(f : f^*) = \Delta H(f : f^*) = H(f^*) - H(f).$$

For $\mathcal{U}(f) = H(f)$, we denote the information function $\mathcal{I}(f)$ as $I(f) = -H(f)$ (Lindley, 1956, Zellner, 1971). For two distributions with finite entropy,

$$\Delta I(f_1 : f_2) = I(f_1) - I(f_2)$$

is well-defined even when $H(f^*) = \infty$.

A well-known measure of discrepancy between two probability distributions F_1 and F_2 is the Kullback-Leibler information

$$K(f_1 : f_2) = \int_S \log \frac{f_1(y)}{f_2(y)}\ dF_1(y),$$

given that F_1 is absolutely continuous with respect to F_2.

For an observable random prospect $K(f_1 : f_2)$ assumes an interpretation rooted in the Bayes rule. Given an observation $Y = y$, Bayes' theorem relates the likelihood ratio to the prior and posterior odds in favor of F as follows:

$$\log \frac{f_1(y)}{f_2(y)} = \log \frac{P(F_1|y)}{P(F_2|y)} - \log \frac{P(F_1)}{P(F_2)}, \tag{5}$$

where $P(\cdot)$ and $P(\cdot|y)$ denote the prior and posterior probabilities of the model. As the difference between the posterior and prior log-odds, the logarithm of the likelihood ratio $\log[f_1(y)/f_2(y)]$ quantifies the information in $Y = y$ in favor of F_1 against F_2 (Kullback 1959). When y is not observed specifically and there is no information on y, other than $y \in \mathcal{S}$, then $K(f_1 : f_2)$ gives the mean information per observation y from F_1 against F_2.

For distributions over a finite set and a finite interval we have

$$K(f : f^*) = \Delta H(f : f^*) = \Delta I(f^* : f). \tag{6}$$

This relationship provides $\Delta H(f : f^*)$ the Bayesian interpretation of the Kullback-Leibler information between F and the uniform distribution. In general, $K(f_1 : f_2) \neq \Delta I(f_1 : f_2)$; Soofi, Ebrahimi, and Habibullah (1995) and Ebrahimi, Soofi, and Soyer (2008) provide generalizations of (6).

Information provided by an observation $Y = y$ about a random prospect X can be measured by the entropy difference $\Delta H(f_X : f_{X|y})$ or by the Kullback-Leibler function $K(f_{X|y} : f_X) \geqslant 0$. The expected information, referred to as mutual information (Shannon 1948), is given by:

$$M(X,Y) \equiv E_y[H(X) - H(X|y)] \tag{7}$$
$$= E_y[K(f_{X|y} : f_X)] \geqslant 0. \tag{8}$$

Other representations of mutual information are:

$$M(X,Y) = H(X) - H(X|Y) \tag{9}$$
$$= H(X) + H(Y) - H(X,Y) \tag{10}$$
$$= K(f_{X,Y} : f_X f_Y), \tag{11}$$

where $H(X|Y) = E_y[H(X|y)]$ is referred to as the conditional entropy, and $H(X,Y)$ is the entropy of the joint distribution $F_{X,Y}$. The equalities (7)-(11) follow from the additive decomposition property of Shannon entropy.

Representation (11) is the information discrepancy between the actual joint distributions of two variables and their joint distribution as if they were independent, showing that $M(X,Y) \geqslant 0$ and the equality holds if and only if two variables are independent. Representations (10) and (11) show the symmetry: $M(X,Y) = M(Y,X)$. By (10), Shannon entropy is subadditive. By (7)-(11), the expected information (2) and shared information (3) based on Shannon entropy, and the expected information discrepancy between the conditional $F_{X|y}$ and marginal F_X distributions (1) and information discrepancy between the joint $F_{X,Y}$ and product marginals $F_X F_Y$ based on the Kullback-Leibler function all provide a unique dependence measure $M(X,Y)$.

2.2 Dynamic Information Functions

In reliability and survival analysis problems, Y is a nonnegative random variable representing a lifetime. In such cases, the age can be taken into account when measuring information. For taking the current age into account, let $Y_t \stackrel{d}{=} Y|Y > t$. Then the density function of distribution $F_{Y_t}(y_t) = P(Y \leqslant y_t|Y > t)$ is given by

$$f_{Y_t}(y_t) = -\frac{d}{dy}P(Y > y_t|Y > t) = \frac{f_Y(y_t)}{\bar{F}_Y(t)}, \tag{12}$$

where $\bar{F}_Y(t) = 1 - F_Y(t)$.

Let $R_t \stackrel{d}{=} Y_t - t|Y > t$, i.e., $R_t \stackrel{d}{=} Y_t - t$ is the residual lifetime. Then the residual density (12) can be represented in terms of r_t as

$$f_{R_t}(r_t) = \frac{f_Y(r_t + t)}{\bar{F}_Y(t)}, \qquad r_t \geqslant 0.$$

Since entropy is location-invariant, $H(f;t) = H(f_{Y_t}) = H(f_{R_t})$. In general, the residual entropy is a function of t.

Noting that for the uniform distribution f_{Y_t} is also uniform over $\{y : t < y < b\}$, the residual entropy $H(f;t)$ measures uncertainty due to lack of predictability of the remaining lifetime at age t. It is clear that for $t_0 = \inf\{y : \bar{F}(y) = 1\}$, $H(f;t_0) = H(f)$; without loss of generality, hereafter we let $t_0 = 0$. For each $t \geqslant 0$, $H(f;t)$ possesses all the properties of $H(f)$. If we consider $\mathcal{T} = \{t : t > 0\}$ as an index set, then $H(f;t)$ provides a dynamic entropy ranging over $\mathcal{T}$.

Dynamic discrimination information function between two residual life distributions $F_1(y;t)$ and $F_2(y;t)$ implied by two lifetime distributions $F_1(y)$ and $F_2(y)$ is given by $K(f_1 : f_2;t) = K[f_1(y;t) : f_2(y;t)]$. Dynamic mutual information function $M(X,Y;t_1,t_2)$ is defined when both components in (X,Y) are residual lifetimes. When only one of two components is a residual lifetime $M(X,Y;t)$ gives a dynamic mutual information.

3 Bayesian Information Measures

In Bayesian analysis the focus of interest is a pair of random prospects (Θ, Y) which plays the role of (X, Y) of the preceding sections. The second component, Y, is an observable random variable whose distribution depends on an unknown parameter θ. The first component Θ, represents values that the parameter θ can take according to a prior probability distribution with density $f_\Theta(\theta)$. Thus, Θ can be thought of as an unobservable random prospect.

Suppose that Y is a nonnegative random variable representing a lifetime with distribution $F_{Y|\theta}$. We observe y and update our prior belief about a parameter θ reflected in the prior distribution with density function $f_\Theta(\theta)$. We use the lifetime density $f_{Y|\theta}(y|\theta)$, $y \geqslant 0$, for the likelihood function and obtain the posterior distribution having the density function

$$f_{\Theta|y}(\theta|y) = \frac{f_{Y|\theta}(y|\theta) f_\Theta(\theta)}{f_Y(y)}, \qquad \theta \in \Theta, \ y \geqslant 0. \tag{13}$$

We also obtain a predictive distribution having density function

$$f_Y(y) = \int_\Theta f_{Y|\theta}(y|\theta) f_\Theta(\theta) \, d\theta, \qquad y \geqslant 0.$$

The prior information $I(\Theta) = -H(\Theta)$ and the likelihood information $I(Y|\theta) = -H(Y|\theta)$ are measures of input information. The posterior information $I(\Theta|y) = -H(\Theta|y)$ and predictive information $I(Y) = -H(Y)$ are measures of output information. Three well-known Bayesian measures of information in data y about the parameter θ are based on various combinations of these measures (Lindley, 1956, Zellner, 1977, Abel and Singpurwalla, 1994). The predictive information is studied by (Press 1996).

3.1 Dynamic Updating

Consider updating the prior belief in situations such as the one follows. Suppose that Y is the age at death of an insured person who purchases the policy at age t. The length of time between Y and t, together with the age at which insurance is purchased, is crucial for pricing life insurance products for individuals in various age groups. Since $f_{Y|\theta}(y|\theta)$ is the density of lifetime distribution at birth, various age groups are not distinguished, hence the updating (13) is static and so is the predictive distribution.

The density function (12) takes the age into account and provides the likelihood function

$$f_{Y_t|\theta}(y_t|\theta, t) = \frac{f_{Y|\theta}(y_t|\theta)}{\bar{F}_{Y|\theta}(t|\theta)}, \qquad y_t > t. \tag{14}$$

Given an observation $y_t > t$, the likelihood function (14) is a function of the observation y_t as well as the age t, hence it is dynamic.

Using the dynamic likelihood (14) we update the prior for the lifetime model parameter $f_\Theta(\theta)$ to the posterior distribution having density function

$$f_{\Theta|y_t}(\theta|y_t;t) \propto \frac{f_{Y|\theta}(y_t|\theta)f_\Theta(\theta)}{\bar{F}_{Y|\theta}(t|\theta)}, \qquad y_t > t, \ \theta \in \Theta. \tag{15}$$

In this updating the posterior is a function of the observation y_t as well as the age t at which insurance is purchased, hence it is dynamic. Clearly $f_{\Theta|y_t}(\theta|y_0;0) = f_{\Theta|y}(\theta|y)$, the static posterior (13).

Applications of the dynamic updating extend well beyond life insurance and lifetime distributions. When the subject of duration study is other than lifetime (e.g., search time, unemployment period) the present time point plays the role of "age". More generally, the dynamic measures are applicable to any continuous distribution with a positive support. For example, for the distributions of wage, income, and diminishable natural resources such as petroleum, the minimum wage, poverty line, and amount of oil extracted to date play the role of the current age, respectively.

3.2 Expected Information in Data

The *expected information* in the data about the parameter (Lindley 1956) is given by

$$\begin{aligned}\vartheta(\Theta|Y) &= E_y[H(\Theta) - H(\Theta|y)] \\ &= E_y\left[K\left(f_{\Theta|y} : f_\Theta\right)\right],\end{aligned} \tag{16}$$

where E_y denotes the expectation with respect to the marginal density f_Y. The expected information $\vartheta(\Theta|Y) = M(Y,\Theta)$ is a mutual information, hence it is symmetric in Y and Θ and is invariant under all one-to-one transformations of the data and the parameter. The new notation $\vartheta(\Theta|Y)$ underscores its directional Bayesian interpretation. The second expression in (16) has an expected utility interpretation (Bernardo 1979). The reference priors which is prevalent in Bayesian literature seek to maximize $\vartheta(\Theta|Y)$, approximately, due the fact that a closed form solution is rare.

The *dynamic expected information* in the residual lifetime about the parameter is

$$\begin{aligned}\vartheta(\Theta|Y_t;t) &= E_{y_t}[H(\Theta) - H(\Theta|y_t;t)] \\ &= E_{y_t}\left[K\left(f_{\Theta|y_t} : f_\Theta;t\right)\right],\end{aligned} \tag{17}$$

where E_{y_t} denotes the expectation with respect to the marginal residual density f_{Y_t} and $H(\Theta|y;t)$ is the entropy of (15). Clearly, $\vartheta(\Theta|Y_0;0) = \vartheta(\Theta|Y)$.

Applications of $\vartheta(\Theta|Y;t)$ will take the item's age into account in numerous applications of Lindley's measure such as reference priors and diagnostics for comparison of designs and experiments, model evaluation, information loss due to censoring, collinearity, and multivariate dimension reduction (see, references in Singpurwalla 1997 and Soofi 1994).

3.3 Information in Observed Data

The information provided by an observation y about θ is measured by the entropy difference

$$\vartheta(\Theta|y) = H(\Theta) - H(\Theta|y).$$

Although $\vartheta(\theta|Y) = \Delta I(f_{\Theta|y}, f_\Theta)$ we use a new notation to emphasize its Bayesian interpretation.

In the continuous case, the entropy (4) is not invariant under transformations. For any nonsingular transformation $\lambda = \phi(\theta)$

$$H(\Lambda) = H(\Theta) - E_\Lambda \left[\log \left| \frac{d\phi^{-1}(\Lambda)}{d\Lambda} \right| \right]. \tag{18}$$

Thus, the information measure $\vartheta(\Theta|y)$ is invariant under location and scale transformations of θ. However, unlike the expected information $\vartheta(\Theta|Y)$, the information in an observation $\vartheta(\Theta|y)$ is not invariant under all one-to-one transformations. Abel and Singpurwalla (1994) found the lack of invariance of $\vartheta(\Theta|y)$ useful for comparing the informativeness of outcomes, survival at time $t = t^*$ and failure of an item in a small interval $(t^*, t^* + \Delta)$ about the failure rate and mean lifetime under an exponential model.

The dynamic information in an observation $y_t > t$ about the parameter θ is given by

$$\vartheta(\Theta|y_t;t) = H(\Theta) - H(\Theta|y_t;t), \tag{19}$$

where $H(\Theta|y;t)$ is the entropy of (15). It is clear that $\vartheta(\Theta|y_0;0) = \vartheta(\Theta|y)$.

4 Information Difference Measure

When we observe that a system survival at time $t > 0$, we can update the prior distribution $f_\Theta(\theta)$ for the lifetime model parameter based on the survival event $Y > t$. Let

$$S_t = \begin{cases} 1, & \text{if the system is functioning at time } t \\ 0, & \text{otherwise.} \end{cases}$$

Updating the prior distribution f_Θ for the system functioning at time t we obtain

$$f_{\Theta|S_t=1}(\theta|1,t) \propto \bar{F}_{Y|\theta}(t|\theta) f_\Theta(\theta), \qquad \theta \in \Theta. \tag{20}$$

The information provided by observing the survival event $Y > t$ is given by the difference

$$\vartheta(\Theta|S_t = 1) = H(\Theta) - H(\Theta|S_t = 1),$$

where $H(\Theta|S_t = 1) = H[f_{\Theta|S_t=1}(\theta|1,t)]$ is the entropy of (20).

Suppose that after observing the survival we continue the experiment until the system actually fails at time $y_t > t$. Then the likelihood (14) can be represented in terms of the residual life $r_t = y_t - t$ as

$$f_{R_t|\theta}(r_t|\theta;t) = \frac{f_{Y|\theta}(r_t + t|\theta)}{\bar{F}_{Y|\theta}(t|\theta)}, \qquad r_t \geqslant 0. \tag{21}$$

The posterior (20) updated by the survival event provides a dynamic prior distribution for the model (21) and can be updated again upon observing $R_t = r_t$, i.e., when the system actually fails. We then obtain

$$\begin{aligned} f_\Theta(\theta|r_t,t) &\equiv f_{\Theta|S_t=1,r_t}(\theta|1,r_t) \\ &\propto f_{Y|\theta}(t + r_t|\theta) f_\Theta(\theta) \\ &\quad r_t \geqslant 0,\ \theta \in \Theta \\ &\propto f_{Y|\theta}(y_t|\theta) f_\Theta(\theta) \\ &\quad y_t > t,\ \theta \in \Theta. \end{aligned} \tag{22}$$

The first expression shows the posterior distribution by decomposing the failure time y_t as $t+r_t$ and the second expression shows the dynamic in terms of $y_t > t$. The sequential learning of failure provides the same updating as observing the failure in a single stage. That is, the total information in the two-stage updating remains the same as static updating (13). However, we can compute the information provided by each stage.

For the two-stage updating we have

$$\begin{aligned} \vartheta(\Theta|y_t;t) &= [H(\Theta) - H(\Theta|S_t = 1)] + [H(\Theta|S_t = 1) - H(\Theta|y_t;t)] \\ &= \vartheta(\Theta|S_t = 1) + \vartheta(\Theta_t|r_t;t), \end{aligned} \tag{23}$$

where Θ_t denotes the parameter with the event updated distribution $f_{\Theta|S_t=1}$. The second equality in (23) is obtained from (22).

We therefore have the information difference between observing the survival event $Y > t$ and observing the actual failure $y_t > t$ given by

$$\begin{aligned} \Delta\vartheta(\Theta|S_t = 1, \Theta|y_t;t) &\equiv \vartheta(\Theta|y_t;t) - \vartheta(\Theta|S_t = 1) \\ &= H(\Theta|S_t = 1) - H(\Theta|y_t;t) \\ &= \vartheta(\Theta_t|r_t;t). \end{aligned}$$

5 Predictive Information and MDIP

The information functions presented in the preceding section were defined in terms of difference between the input information about the parameter $I(\Theta)$ and the output information about the parameter $I(\Theta|y)$. This section present dynamic versions of two information measures that involve the predictive information $I(Y)$ and the likelihood information $I(Y|\theta)$.

5.1 Predictive Information

Press (1996) studied the marginal entropy of Y in the Bayesian predictive context. For the dynamic updating (15) the prior predictive density is given by

$$f_{Y_t}(y_t) = \int f_\theta(\theta) f_{Y_t|\theta}(y_t|\theta) d\theta. \tag{24}$$

The dynamic predictive information measure for (24) is $I(Y;t) = -H(Y;t)$. Using (9) and (17), we have

$$I(Y_t;t) = E_\theta[I(Y_t|\theta;t)] - \vartheta(\theta|Y_t;t), \tag{25}$$

where E_θ denotes the expectation with respect to the prior distribution f_Θ. This representation provides an intuitive interpretation of the predictive information as being the difference between the expected information in the residual lifetime likelihood about residual life and the expected information in the residual life data about the parameter.

5.2 MDIP

An information measure proposed by Zellner (1977) for developing priors is defined in terms of difference between the two inputs: prior information about the parameter $I(\Theta)$ and the likelihood information $I(Y|\theta)$. Zellner's measure, referred to as the maximal data information prior (MDIP) criterion, is defined as

$$\begin{aligned} \mathcal{Z}(\Theta) &= E_\theta[H(\Theta) - H(Y|\theta)] \\ &= E_\theta[I(Y|\theta)] - I(\Theta) \\ &= E_{\theta|y}\left[K\left(f_{Y|\theta} : f_\Theta\right)\right], \end{aligned}$$

where $E_{\theta|y}$ denotes the expectation with respect to the posterior distribution $f_{\Theta|y}$. By the virtue of second representation, the MDIP criterion function is interpreted as the *ápriori* expected information in the likelihood from which the information in the prior distribution is subtracted. The third expression is well-defined whenever the sampling distribution $f_{Y|\theta}$ is absolutely continuous with respect to the prior f_Θ. For the case of lifetime distributions, typically parameters are positive and the support of $f_{Y|\theta}$ is a subset of $\Re^+$, hence the absolute continuity is satisfied.

The MDIP criterion can be expressed in terms of expected information about the parameter, the predictive information, and the prior information as

$$\mathcal{Z}(\Theta) = \vartheta(\Theta|Y) + I(Y) - I(\Theta).$$

The MDIP distribution maximizes $\mathcal{Z}(\Theta)$. The MDIP distribution has a closed form density function given by

$$f^*_{MDIP}(\theta) \propto \exp\{-H(Y|\theta)\}.$$

The entropy expression for many well-known families of distribution has closed form. When θ ranges over a finite set or interval, $f^*_{MDIP}(\theta)$ is a proper density functions. Otherwise, it can be improper. The MDIP solution is maximally committal to the data information (minimally committal to non-data information). Note that the MDIP density function $f^*_{MDIP}(\theta)$ is decreasing function of the likelihood entropy.

Interestingly, for $\mathcal{Z}^*(\Theta) = \max_{f(\theta)} \mathcal{Z}(\Theta)$, we have

$$\mathcal{Z}^*(\Theta) - \mathcal{Z}(\Theta) = K[f(\theta) : f^*_{MDIP}(\theta)].$$

Details and applications of the MDIP to various probability and econometrics models can be found in Zellner (1997).

The dynamic MDIP criterion is given by

$$\begin{aligned}\mathcal{Z}(\Theta; t) &= E_\theta[H(\Theta) - H(Y_t|\theta; t)] \\ &= E_{\theta|y_t}\left[K\left(f_{y_t|\theta} : f_\theta; t\right)\right].\end{aligned}$$

The maximum dynamic data information prior is the density $f^*_\theta(\theta)$ that maximizes $\mathcal{Z}(\Theta; t)$. A closed form solution is obtainable whenever the residual entropy has a closed form, however it will be improper for most well known models.

6 Exponential and Weibull Models

This section illustrates applications of the Bayesian dynamic information measures for two lifetime models shown in Table 1. These two models are chosen because: (a) the exponential model is the most fundamental lifetime model and has the unique lack of memory property; and (b) the Weibull model is also

Table 1. Parametric functions of interest for Bayesian information analysis of the exponential and Weibull models.

Lifetime Model	Prior Density	Failure rate	Mean
Exponential			
$f(y\|\theta) = \theta e^{-\theta y}$	$f(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$, $\theta > 0$	$\lambda(y\|\theta) = \theta$	$\mu_{y\|\theta} = \theta^{-1}$
Weibull			
$f(y\|\theta) = c\theta y^{c-1} e^{-\theta y^c}$	$f(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$, $\theta > 0$	$\lambda(y\|\theta) = cy^{c-1}\theta$	$\mu_{y\|\theta} = \Gamma(1+\frac{1}{c})\theta^{-\frac{1}{c}}$

an important model for which information functions provide example of more general features than those for the exponential model. The Weibull shape parameter c is assumed to be known. In each case, for the parameter of interest θ we use the gamma prior $Gam(\alpha, \beta)$. For each model, we discuss the information provided by an observation y_t about the model parameter θ and about two important functions of the parameter, the failure rate $\phi_1(\theta) = \lambda(y|\theta)$ and the mean $\phi_2(\theta) = \mu_{y|\theta}$, shown in Table 1.

For both models, the posterior distributions for the model parameter are also gamma $f_{\Theta|Data} = Gam(\tilde{\alpha}, \tilde{\beta})$, where the updated parameters depend on the updating scenarios. When a failure is observed, $\tilde{\alpha} = \alpha + 1$ and for observing a survival $\tilde{\alpha} = \alpha$. For the exponential parameter θ, posterior (15) is $f_{\Theta|y_t}(\theta|y_t) = Gam(\alpha + 1,\ \beta + y - t) = Gam(\alpha + 1,\ \beta + r_t)$ and posterior (20) is $f_{\Theta|S_t=1}(\theta|1, t) = Gam(\alpha,\ \beta + t)$. For the Weibull parameter θ, posterior (15) is $f_{\Theta|y_t}(\theta|y_t) = Gam(\alpha + 1,\ \beta + y^c - t^c)$ and posterior (20) is $f_{\Theta|S_t=1}(\theta|1, t) = Gam(\alpha,\ \beta + t^c)$.

Figure 1 shows the density functions of the prior and posterior distributions of the exponential failure rate $\lambda(y|\theta) = \theta$ for $\alpha = \beta = 3, t = 3$, and $r_t = 1, 3, 5$. The prior is the same as that used by Abel and Singpurwalla (1994) and the posterior for survival at $t = 3$ is also the same as theirs. The posterior for $r_t = 3$ is equivalent to the static posterior (13) for observing a failure at $y = 3$ used by Abel and Singpurwalla for illustrating that a failure is less informative about the exponential failure rate than a survival at $t = 3$. (The survival posterior for $t = 3$ is more concentrated than the failure posterior for $y = 3$). Figure 1 shows that the posteriors for observing failure become more concentrated as the residual lifetime r_t increases and eventually become more concentrated (e.g., the posterior for $r_t = 5$) than the survival posterior for $t = 3$.

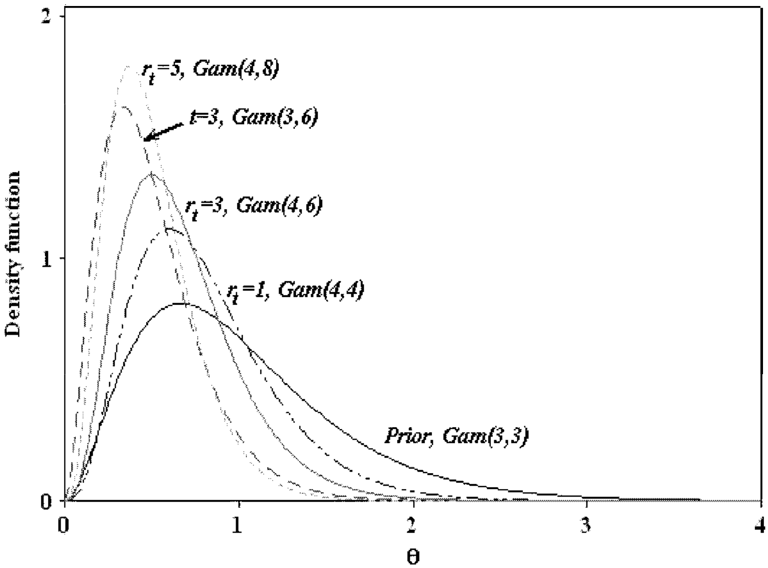

Figure 1. Prior and dynamic posterior distributions of the exponential failure rate θ.

6.1 Expected Information

The entropy of prior distribution $f_\Theta = Gam(\alpha, \beta)$ is given by

$$H(\Theta) = \log \Gamma(\alpha) - \log \beta - (\alpha - 1)\psi(\alpha) + \alpha, \tag{26}$$

where $\psi(\alpha) = d \log \Gamma(\alpha)/d\alpha$ is digamma function. The posterior entropies are given by (26) with the updated parameters.

The exponential and Weibull variables are related by a one-to-one transformation. Due to the invariance of mutual information, the expected information in data about θ is the same for both cases. Since y and t only effect the scale of prior distribution, the expected information is free of t and is given by

$$\vartheta(\Theta|Y) = \frac{1}{\alpha} + \psi(\alpha) - \log \alpha. \tag{27}$$

For each case, $\phi_1(\theta) = \lambda(y|\theta)$ and $\phi_2(\theta) = \mu_{y|\theta}$ are one-to-one transformations, thus the expected information about the failure rate and the mean is also given by (27).

In general, however, dynamic Lindley's measure need not be free of t. For example, consider the case when Y has a uniform distribution over $(\theta, \theta + 1)$ and the prior for θ is uniform over an interval. It can be shown that $\vartheta(\Theta|Y; t)$ depends on t.

6.2 Information in An Observation

For the exponential model, the information in an observation $y_t > t$ about the failure rate is

$$\begin{aligned}\vartheta(\Theta|y_t;t) &= \log\left(1+\frac{y_t-t}{\beta}\right)+\psi(\alpha)-\log\alpha \\ &= A_\lambda(\beta,y_t,t)+B_\lambda(\alpha) \\ &= A^*_\lambda(\beta,r_t)+B_\lambda(\alpha).\end{aligned} \tag{28}$$

The prior and posterior entropies of $\mu = \theta^{-1}$ are obtained by (18), and the information (19) for the mean lifetime is

$$\begin{aligned}\vartheta(\Theta^{-1}|y_t;t) &= -\log\left(1+\frac{y_t-t}{\beta}\right)+\psi(\alpha)-\log\alpha+\frac{2}{\alpha} \\ &= A_\mu(\beta,y_t,t)+B_\mu(\alpha) \\ &= A^*_\mu(\beta,r_t)+B_\mu(\alpha).\end{aligned} \tag{29}$$

The following points are noteworthy.

(a) For $r_t = y_t - t$, $A_\lambda(\beta,y_t,t) = A^*_\lambda(\beta,r_t) = \log\left(1+\frac{y_t-t}{\beta}\right)$; clearly, the function is decreasing in β and t, and increasing in y_t and r_t.

(b) For $r_t = y_t - t$, $A_\mu(\beta,y_t,t) = A^*_\mu(\beta,r_t) = -A^*_\lambda(\beta,r_t)$; clearly, the function is increasing in β and t, and decreasing in y_t and r_t.

(c) As a functions of t, the information measures (28) and (29) are dynamic. However, in terms of the residual life r_t these information measures are memoryless due to the lack of memory of the exponential model.

(d) $B_\lambda(\alpha) = \psi(\alpha) - \log\alpha < 0$; the function is increasing for all α with the upper bound zero.

(e) $B_\mu(\alpha) = \psi(\alpha) - \log\alpha + \frac{2}{\alpha} > 0$; the function is decreasing for all α with a lower bound zero.

(f) The average of (28) and (29) is

$$\frac{1}{2}\left[\vartheta(\Theta|y_t;t)+\vartheta(\Theta^{-1}|y_t;t)\right] = \vartheta(\Theta|Y).$$

That is, the information in an observation $y_t > t$ about the exponential failure rate and about the mean lifetime average to the expected information in data about the exponential parameter $\vartheta(\Theta|Y)$, which is free of t.

For the exponential model, as an item gets older, observation of its lifetime becomes less informative about the failure rate and more informative about the mean lifetime. For $t = 0$ and $y = t^*$, (28) and (29) give the information provided by observing a failure at $y \approx t^*$ about the exponential failure rate and mean studied by Abel and Singpurwalla (1994).

More insights about the information functions can be gained by considering the family of gamma priors that have a given mean, say $E_\theta(\Theta) = 1$ implying $\alpha = \beta$. In this family, the prior entropy is concave in α with maximum at $\alpha = 1$, i.e., when the prior is exponential, but the prior variance is convex and decreasing in α. Figure 2a shows plots of the entropy and variance of this family as functions of α.

It can be shown that the information about the failure rate $\vartheta(\Theta|r_t, \alpha)$ is concave in α and the information about the mean $\vartheta(\Theta^{-1}|r_t, \alpha)$ is convex and decreasing in α. Figure 2b shows examples of information functions for the exponential model when $r_t = 1, 3$. The graph illustrates the pattern of information about the failure rate (dotted lines) is similar to the pattern of prior entropy and the pattern of information about the mean (dashed lines) is similar to the pattern of prior variance. The graph also illustrates that the information about the failure rate (mean) increases (decreases) with r_t. The expected information is also shown which is the average of information about the failure rate and the mean. The expected information is decreasing in α.

For the Weibull model, the information about θ is in the same form as (28) where y and t are replaced with y^c and t^c.

$$\begin{aligned}\vartheta(\Theta|y_t;t) &= \log\left(1 + \frac{y_t^c - t^c}{\beta}\right) + \psi(\alpha) - \log\alpha \\ &= A_\lambda(\beta, y_t^c, t^c) + B_\lambda(\alpha). \qquad (30)\end{aligned}$$

As seen in Table 1, the failure rate is a linear function of θ. Hence, the dynamic information about the failure rate is $\vartheta(\Theta|y_t;t)$. The mean is a scalar multiple of $\theta^{-1/c}$, which has an inverted generalized gamma distribution. Applications of (18) on the prior and posterior entropies of θ gives

$$\begin{aligned}\vartheta\left(\Theta^{-\frac{1}{c}}|y_t;t\right) &= -\frac{1}{c}\log\left(1 + \frac{y_t^c - t^c}{\beta}\right) + \psi(\alpha) - \log\alpha + \left(1 + \frac{1}{c}\right)\frac{1}{\alpha} \\ &= A_{\mu_c}(\beta, y_t, t) + B_{\mu_c}(\alpha). \qquad (31)\end{aligned}$$

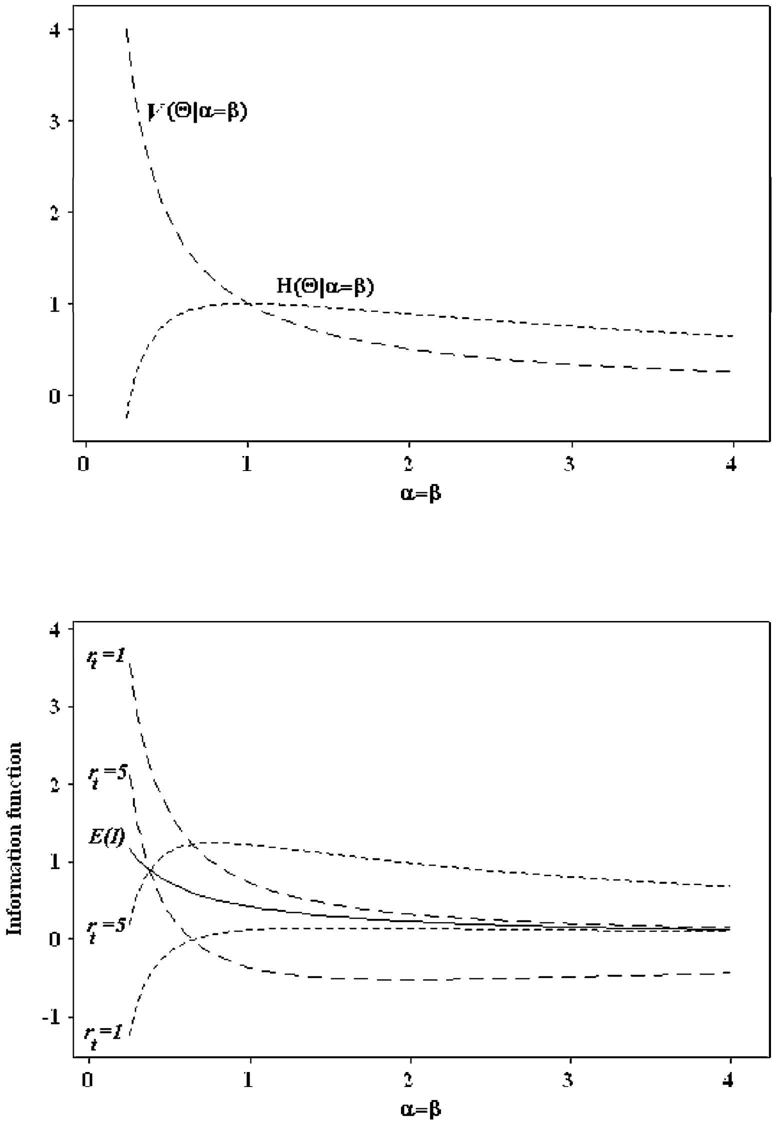

Figure 2. Entropy and variance of gamma prior with parameters $\alpha = \beta$ and information function about the exponential failure rate (dotted), mean (dashed), and the expected information.

The following points are noteworthy.

(a) As functions of α, β, y_t, and t, the dynamic behaviors of the information measures for the Weibull parameter, failure rate, and mean are similar to those for the exponential model.

(b) For $r_t = y_t - t$, $A_{\mu_c}(\beta, y_t^c, t^c) = A_\mu^*(\beta, r_t)$ if and only if $c = 1$; i.e. the model is exponential.

(c) As a functions of t, the information measures (30) and (31) are dynamic. With $y_t = t + r_t$ these information measures are functions of the residual life r_t and t and are memoryless if and only if $c = 1$; i.e. the model is exponential.

(d) $A_{\mu_c}(\beta, y_t^c, t^c) = -A_\lambda(\beta, y_t^c, t^c)$ if and only if $c = 1$; i.e. the model is exponential.

(e) $B_{\mu_c}(\alpha) = \psi(\alpha) - \log\alpha + \left(1 + \frac{1}{c}\right)\frac{1}{\alpha} > 0$; the function is decreasing for all α with a lower bound zero.

(f) The average of (30) and (31) is less than $\vartheta(\Theta|Y)$ for $c < 1$ when the failure rate is decreasing; it is greater than $\vartheta(\Theta|Y)$ for $c > 1$ when the failure rate is increasing; and equal to the expected information $\vartheta(\Theta|Y)$ for $c = 1$ when failure rate is constant (the model is exponential). In general, the average is a function of t.

Figure 3 shows examples of information functions for the Weibull model when $c = 2$ based on the family of gamma priors with parameters $\alpha = \beta$. The plots are for $t = 1, 3$ and $r_t = 1$. The graph illustrates the information about the failure rate (dotted lines) follows similar pattern as the prior entropy and the information about the mean (dashed lines) follows similar pattern as the prior variance. The graph also illustrates that the information about the failure rate (mean) increases (decreases) with t. The expected information is also shown which in this case is not the average of information about the failure rate and the mean. The dashed-dotted lines show plots of averages of information measures for the failure rate and the mean which depend on t; on average, the information decreases as t increases.

6.3 Information Difference Measure

For the exponential parameter θ, the posterior (15) is $Gam(\alpha,\ \beta + t)$. The information about θ provided by observing the survival event $Y > t$ is

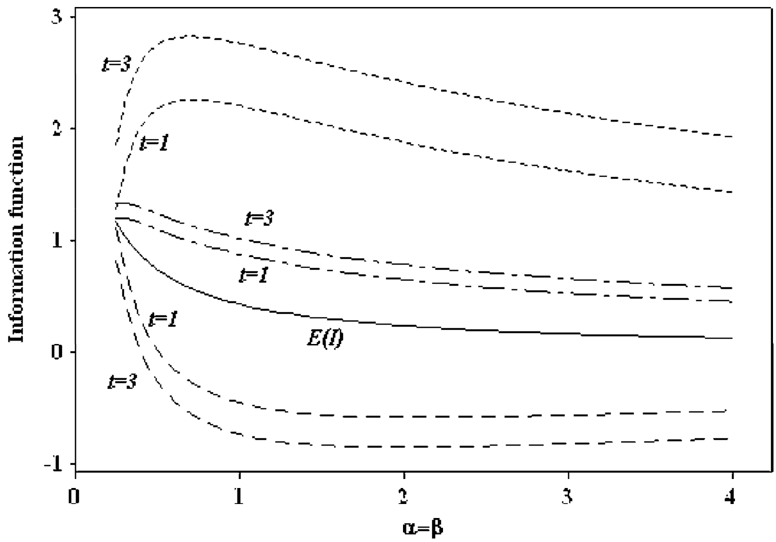

Figure 3. Dynamic information function about the Weibull ($c = 2$) failure rate (dotted), mean (dashed), their average (dotted-dashed) and the expected information when prior parameters are $\alpha = \beta$.

$$\vartheta(\Theta|S_t = 1) = \log\left(1 + \frac{t}{\beta}\right).$$

This gives the information about the exponential failure rate provided by observing a survival at $t = t^*$, which is studied by Abel and Singpurwalla (1994).

The information difference between observing the survival event $Y > t$ and observing an actual failure $y_t > t$ is given by

$$\begin{aligned}\Delta\vartheta(\Theta|S_t = 1, \Theta|y_t) &= \vartheta(\Theta_t|r_t) \\ &= \log\left(1 + \frac{r_t}{\beta + t}\right) + \psi(\alpha) - \log\alpha \\ &= \vartheta(\Theta|y_t; t).\end{aligned}$$

The information difference is nonnegative only when $\frac{r_t}{\beta + t} \geqslant \alpha\exp\{-\psi(\alpha)\} - 1$.

For the Weibull parameter θ, the posterior (15) is $Gam(\alpha,\ \beta + t^c)$. The information about θ provided by observing the survival event $Y > t$ is

$$\vartheta(\Theta|S_t = 1) = \log\left(1 + \frac{t^c}{\beta}\right).$$

The information difference between observing the survival event $Y > t$ and observing an actual failure $y_t > t$ is given by

$$\Delta\vartheta(\Theta|S_t = 1, \Theta|y_t) = \vartheta(\Theta_t|r_t) = \log \frac{\beta + (r_t + t)^c}{\beta + t^c} + \psi(\alpha) - \log \alpha.$$

In this case, the information difference is an increasing function of r_t and decreasing functions of α and β.

6.4 Predictive Information

The dynamic predictive information is obtained by using (27) in (25) and evaluating $E_\theta[I(Y|\theta; t)]$. Since (27) is free of t, the dynamic of the predictive information is induced by the likelihood function. Thus, $I(Y; t)$ is free of t for exponential, but not for Weibull. For the exponential model, the predictive density is Pareto over (β, ∞) and $I(Y; t) = -H(\alpha) - \log \beta$, where

$$H(\alpha) = 1 + \frac{1}{\alpha} - \log \alpha$$

is the entropy of Pareto distribution with parameter α. For Weibull, the predictive entropy is given by $I(Y; t) = -H(\alpha) + A(\alpha, \beta, c, t)$ where $A(\alpha, \beta, c, t)$ is increasing in t for $c > 1$ and decreasing when $c < 1$.

6.5 MDIP Criterion

It is easily seen that $\mathcal{Z}(\Theta; t)$ is free of t if and only if the lifetime is exponentially distributed. For the Weibull case, it can be shown that $\mathcal{Z}(\Theta; t)$ is an increasing function of t for $c < 1$ (the decreasing failure rate case) and is a decreasing function of t for $c > 1$ (the increasing failure rate case). If we wish to develop priors based on MDIP, the priors for θ will be improper for all three cases.

7 Concluding Remarks

We have introduced the concept of Bayesian dynamic information. Four measures that combine prior and residual life data are introduced: dynamic Lindley's expected information in an experiment about a parameter, dynamic information in observed data about a parameter, dynamic information in the predictive distribution, and dynamic Zellner's information in the data about the parameter. These measures are applied to examine information provided by the data beyond an age about the mean and hazard function under the exponential and Weibull models for the lifetime variable, and the respective predictive distributions.

The entropy, discrimination information function, and mutual information for the past lifetime (or down time) of a system are defined similarly and provide dynamic information measures; see Di Crescenzo and Longobardi (2002, 2004). All the dynamic Bayesian measures presented in this paper can simply be adapted using the past lifetime information measures.

Another well-known measure of uncertainty of a distribution is the entropy of order α, (Rényi 1961), which generalizes Shannon entropy. The information divergence of order α (Rényi 1961) generalizes the Kullback-Leibler information. Asadi et al. (2005) and Abraham and Sankaran (2006) have developed dynamic versions of Rényi entropy and information divergence, and Nanda and Paul (2006) have studied other generalization in the dynamic context. These measures can be used in the context of Bayesian information, as well. The choice of an uncertainty function is up to the researcher, taking into consideration the properties and suitability of the measure for a particular study. An advantage of Shannon entropy is its additive decomposition property, which impies the equalities (8)-(11). Use of Rényi entropy and information divergence in (8)-(11) provides four different measures of dependence. The Bayesian interpretation of of Kullback-Leibler information (5) and entropy is another advantage for their applications in the context of Bayesian information analysis.

The concepts and results developed in this paper provide some topics for future research, including implications of the dynamic information about the parameters and dynamic predictive information for system design, experimental design, sampling schemes, and minimal information priors. We explored some properties of dynamic Bayesian information measures for the exponential and Weibull models. Establishing these types of properties for dynamic Bayesian information measures in the context of some broader classes of distributions provides interesting and challenging research topics.

In another route, Mazzuchi et al. (2008) used the quantized entropy for developing a procedure for Bayesian inference about Shannon entropy when $F_{Y|\theta}$ is unknown and the Kullback-Leibler information index of fit of a model $F^*_{Y|\theta}$. The key elements of their procedure is a maximum entropy distribution $F^*_{Y|\theta}$ in a moment class as the first guess in the Dirichlet process prior and estimating the moments using the Dirichlet probabilities such that the Kullback-Leibler information can be estimated properly through difference between two entropies. Extension of their procedure to the dynamic setting requires addressing a few important issues such as what types of maximum entropy models may be considered: the traditional maximum entropy model in a moment class or the maximum dynamic entropy models developed by Asadi et al. (2004)?

References

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

Abraham B.; Sankaran, P.G. (2006). Rényi's entropy for residual lifetime distribution. *Statistical Papers.* **47**, 17-29.

Asadi, M.; Ebrahimi, N.; Soofi, E.S. (2005). Dynamic generalized information measures. *Statistics and Probability Letters.* **71**, 85-98.

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

Asadi, M.; Ebrahimi, N.; Hamedani, G.G.; Soofi, E.S. (2005). Dynamic minimum discrimination information models. *Journal of Applied Probability.* **42**, 643-660.

DeGroot, M.H. (1962). Uncertainty, information, and sequential experiments. *Annals of Mathematical Statistics.* **33**, 404-419.

Belzunce, F.; Navarro, J.; Ruiz, J.M. (2004). Some results on residual entropy functions. *Metrika.* **59**, 147-161.

Bernardo, J.M. (1979). Expected information as expected utility. *The Annals of Statistics.* **7**, 686-690.

Di Crescenzo, A.; Longobardi, M. (2002). Entropy-based measure of uncertainty in past lifetime distributions. *Journal of Applied Probability.* **39**, 434-440.

Di Crescenzo, A.; Longobardi, M. (2004). A Measure of discrimination between past life-time distributions. *Statistics and Probability Letters.* **67**, 173-182.

Ebrahimi, N. (1996). How to measure uncertainty in the residual lifetime distributions. *Sankhya A.* **58**, 48-57.

Ebrahimi, N.; Kirmani, S.N.U.A. (1996 a). A Characterization of the proportional hazards model through a measure of discrimination between two residual life distributions. *Biometrika.* **83**, 233-235.

Ebrahimi, N.; Kirmani, S.N.U.A. (1996 b). Some results on ordering of survival functions through uncertainty. *Statistics and Probability Letters.* **29**, 167-176.

Ebrahimi, N.; Kirmani, S.N.U.A.; Soofi, E.S. (2007). Dynamic multivariate information. *Journal of Multivariate Analysis.* **98**, 328-349.

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

Ebrahimi, N.; Maasoumi, E.; Soofi, E.S. (1999 b). Measuring informativeness of data by entropy and variance. in *Advances in Econometrics: Income Distribution and Methodology of Science, Essays in Honor of Camilo Dagum*, D. Slottje, ed. Physica-Verlag, New York.

Ebrahimi, N.; Soofi, E.S.; Soyer, R. (2008). Multivariate maximum entropy identification, transformation, and dependence. *Journal of Multivariate Analysis.* **99**, 1217-1231.

Goel, P.K.; DeGroot, M.H. (1981). Information about hyperparameters in hierarchical models. *Journal of the American Statistical Association.* **76**, 140-147.

Kullback, S. (1959), *Information Theory and Statistics.* Wiley, New York.

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

Mazzuchi, T.A.; Soofi, E.S.; Soyer, R. (2008). Bayes estimate and inference for entropy and information index of fit. *Econometric Reviews.* **27**, 428-456.

Nanda A.K.; Paul, P. (2006). Some results on generalized residual entropy. *Information Sciences.* **176**, 27-47.

Polson, N.G. (1992). On the expected amount of information from a nonlinear model. *Journal of the Royal Statistical Society B.* **54**, 889-895.

Press, S.J. (1996). The Definetti Transform. in *Maximum Entropy and Bayesian Methods.* Kittanson and R. Silver, eds. Kluwer.

Rényi, A. (1961). On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium.* **1**, 547-561.

Retzer, J.J.; Soofi, E.S.; Soyer, R. (2008). Information importance of predictors: Concepts, measures, Bayesian inference, and applications. *Computational Statistics and Data Analysis*, in press.

Sebastiani, P.; Wynn, H.P. (2000). Maximum entropy sampling and optimal Bayesian experimental design. *Journal of the Royal Statistical Society, Ser. B.* **62**, 145-157.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal.* **27**, 379-423.

Singpurwalla, N.D. (1997). Entropy and information in reliability. in *Bayesian Analysis of Statistics and Econometrics: Essays in Honor of Arnold Zellner*, D. Berry, K. Chaloner, and J. Geweke, eds. Wiley, New York, pp. 459-469.

Soofi, E.S. (2000). Principal information theoretic approaches. *Journal of the American Statistical Association.* **95**, 1349-1353.

Soofi, E.S. (1994). Capturing the intangible concept of information. *Journal of the American Statistical Association.* **89**, 1243-1254.

Soofi, E.S.; Ebrahimi, N.; Habibullah, M. (1995). Information distinguishability with application to analysis of failure data. *Journal of the American Statistical Association.* **90**, 657-668.

Yuan, A.; Clarke, B. (1999). An information criterion for likelihood selection. *IEEE Transactions On Information Theory.* **IT 45**, 562-571.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Zellner, A. (1977). Maximal data information prior distributions. in *New Developments in Applications of Bayesian Methods*, A. Aykac and C, Brumat, eds. North Holland, Amsterdam, pp. 211-232.

Zellner, A. (1997). *Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers*, Cheltenham, UK & Lyme, US: Edward Elgar.

Nader Ebrahimi
Division of Statistics,
Northern Illinois University,
Dekalb, IL 60155.
e-mail: *nader@math.niu.edu*

S. N. U. A. Kirmani
Department of Mathematics,
University of Northern Iowa,
Cedar Falls, IA 50614.
e-mail: *kirmani@math.uni.edu*

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