I²SDS - Decision Sciences
THE GEORGE WASHINGTON UNIVERSITY
SCHOOL OF BUSINESS

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2014-3
February 24, 2014

Information Theory and Bayesian Reliability Analysis: Recent Advances

Nader Ebrahimi
Division of Statistics
Northern Illinois University, USA

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

Refik Soyer
Department of Decision Sciences
The George Washington University, USA

Information Theory and Bayesian Reliability Analysis: Recent Advances

Nader Ebrahimi, Ehsan S. Soofi and Refik Soyer

1 Introduction and Overview

As noted by Ebrahimi, Soofi and Soyer (2010a) in a recent review, information theory provides measures for handling diverse problems in modelling and data analysis in a unified manner. Information theory statistics have been considered in reliability modelling and life data analysis; see Ebrahimi and Soofi (1998, 2004) for a review of such work. Information theory based work in reliability can be grouped into three main areas as suggested in Ebrahimi and Soofi (2004). These include development of information functions for reliability analysis, information theory-based diagnostics and hypothesis tests for model building and measures that quantify the amount of information for prediction.

Since the seminal work of Lindley (1956), information theory has played an important role in Bayesian statistics. The mutual information which is also known as Lindley's measure has been used by Bernardo (1979a) as the expected utility for the decision problem of reporting a probability distribution. It also has provided the foundation for the reference priors of Bernardo (1979b). Other uses of Lindley's information have been in design problems; see for example Chaloner and Verdinelli (1995) for a comprehensive review. An information processing rule has been defined in Zellner (1988) using information measures and Bayes rule has been shown as the optimal solution.

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

Ehsan S. Soofi
Sheldon B. Lubar School of Business, University of Wisconsin-Milwaukee, Milwaukee, WI e-mail: esoofi@uwm.edu

Refik Soyer
Department of Decision Sciences, The George Washington University, Washington, DC e-mail: soyer@gwu.edu

As noted by Ebrahimi and Soofi (2004) an area of Bayesian reliability analysis where information theory has been often used is the optimal design of life tests; see Chaloner and Verdinelli (1995) and the references therein. Bayesian nonparametric entropy estimation and Bayesian estimation of information indices for lifetime modelling by Mazzuchi et al. (2000, 2008) have been another area of focus of information theoretic work.

In this paper we consider some recent advances in use of information theory in Bayesian reliability analysis. We present a range of information functions for reliability analysis, present their properties and discuss their use in addressing different issues in reliability. Our discussion focuses on use of Bayesian information measures in failure data analysis, prediction, assessment of reliability importance and optimal design of life tests. Below we present some preliminaries associated with Bayesian reliability analysis and introduce notation. Section 2 presents information measures such as mutual information that are used in reliability analysis and their properties. Parameter and predictive information concepts are considered and their properties are discussed in Section 3 with implications on Bayesian designs. Section 4 considers informativeness of observed failures and survivals from life tests and presents some new results for comparison. The notion of information importance is presented is Section 5 as an alternative measure of reliability importance of components of a system. Concluding remarks are given in Section 6.

1.1 Preliminaries

Reliability analysis deals with quantification of uncertainty about certain event(s) and making decisions. Typical issues of interest include: (i) if a component (or a system) performs its mission; (ii) if time to failure of a component (or a system) exceeds a specified (mission) time, and (iii) if mean time to failure exceeds a specified time. For example, if Y denotes lifetime of a component (or a system) then the event of interest is $Y > y$ where y is the specified mission time. The quantity

$$P(Y > y|\theta) = Prob\{Y > y|\theta\} \tag{1}$$

as a function of y is known as the reliability function, where θ is a parameter which can be a scalar or a vector. If $f(y|\theta)$ denotes the the density function of the failure model for Y, then we can write the reliability function as

$$P(Y > y|\theta) = \int_0^y f(x|\theta)dx. \tag{2}$$

Decision problems that involve reliability assessment include design of life tests, system design via reliability optimization, and developing optimal maintenance strategies. In Bayesian reliability analysis, uncertainty about θ is described probabilistically via the prior distribution $f(\theta)$. Prior to observing any data uncertainty statements about a future lifetime Y_v is made using the prior predictive distribution

$$f(y_v) = \int f(y_v|\theta) f(\theta) d\theta. \tag{3}$$

Given n observations, $\mathbf{y} = (y_1, y_2, \cdots, y_n)$ from $f(y|\theta)$, posterior inference for θ is obtained via the posterior distribution

$$f(\theta|\mathbf{y}) \propto f(\theta) f(\mathbf{y}|\theta). \tag{4}$$

The posterior predictive distribution for the future lifetime Y_v, is given by

$$f(y_v|\mathbf{y}) = \int f(y_v|\theta) f(\theta|\mathbf{y}) d\theta. \tag{5}$$

2 Information Functions for Reliability Analysis

Let Q be an unknown quantity of interest which can be a scalar or a vector. Q may be a parameter Θ such as failure rate or may represent a future outcome Y_v such as component lifetime or $Q = (\Theta, Y_v)$. We denote the distribution of Q by F and its probability mass or density function by f.

The unpredictability of Q depends solely on the concentration of its distribution measured by an uncertainty function $\mathscr{U}(f)$. As pointed out by Ebrahimi et al. (2010a), two desirable properties of the uncertainty function are: (i) $\mathscr{U}(\cdot)$ is concave. (ii) $\mathscr{U}(f) \leq \mathscr{U}(f^*)$, where f^* is the pdf of the uniform distribution (the least concentrated model). An uncertainty function with these properties is Shannon entropy

$$H(Q) = H(f) = -\int f(q) \log f(q) dq. \tag{6}$$

The uncertainty about Q is measured by $H(Q)$ and $I(Q) = -H(Q)$ is information about Q; see Lindley (1956).

Information provided by the data $\mathbf{y}$ about Q is measured by the entropy difference

$$\Delta H(\mathbf{y}; Q) = H(Q) - H(Q|\mathbf{y}) \tag{7}$$

where $H(Q|\mathbf{y})$ is obtained using the posterior distribution $f(q|\mathbf{y})$. In (7) $\Delta H(\mathbf{y}; Q)$ is referred to as *observed sample information* about Q and can be positive or negative.

The information discrepancy between $f(q|\mathbf{y})$ and $f(q)$ can also be measured by the Kullback-Leibler divergence

$$K[f(q|\mathbf{y}) : f(q)] = \int f(q|\mathbf{y}) \log \frac{f(q|\mathbf{y})}{f(q)} dq \geq 0, \tag{8}$$

where the equality holds if and only if $f(q|\mathbf{y}) = f(q)$ almost everywhere. The information discrepancy is a *relative entropy* which only detects changes between the prior and the posterior, without indicating which of the two distributions is more informative. It is invariant under all one-to-one transformations of Q.

Expected sample information measures are obtained by viewing the information measures as functions of data $\mathbf{y}$ and averaging them with respect to the distribution of $\mathbf{y}$. Conditional entropy of Q given $\mathbf{y}$ is defined as

$$\mathscr{H}(Q|\mathbf{y}) = E_{\mathbf{y}}\{H(Q|\mathbf{y})\} = \int H(Q|\mathbf{y}) f(\mathbf{y}) d\mathbf{y}. \tag{9}$$

The *conditional information* is then defined as $\mathscr{I}(Q|\mathbf{y}) = -\mathscr{H}(Q|\mathbf{y})$. It follows from the above that the expected sample information is given by

$$E_{\mathbf{y}}[\Delta H(\mathbf{y};Q)] = H(Q) - \mathscr{H}(Q|\mathbf{y}) \geq 0, \tag{10}$$

which is always nonnegative.

The expected entropy difference and expected KL divergence provide the same measure, known as the *mutual information*

$$M(\mathbf{y};Q) = E_{\mathbf{y}}\{\Delta H(\mathbf{y};Q)\} = E_{\mathbf{y}}\left\{K[f(q|\mathbf{y}) : f(q)]\right\}. \tag{11}$$

Another representation of the mutual information, $M(\mathbf{y};Q)$, is given by

$$M(\mathbf{y};Q) = H(Q) - \mathscr{H}(Q|\mathbf{y}) = K[f(q,\mathbf{y}) : f(q)f(\mathbf{y})]. \tag{12}$$

These representations are in terms of the expected uncertainty reduction, and imply that the mutual information is symmetric in Q and $\mathbf{y}$. We note the following properties of mutual information:

1. $M(\mathbf{y};Q) \geq 0$, where the equality holds if an only if Q and $\mathbf{y}$ are independent.
2. We can write $M(\mathbf{y};Q)$ as

$$M(\mathbf{y};Q) = H(Q) + H(\mathbf{y}) - H(Q,\mathbf{y}).$$

3. The conditional mutual information is defined by $\mathscr{M}(\mathbf{y};Q|S) = E_s[M(\mathbf{y};Q|s)] \geq 0$, where the equality holds if an only if Q and $\mathbf{y}$ are conditionally independent.
4. $M(\mathbf{y};Q)$ is invariant under one-to-one transformations of Q and $\mathbf{y}$.

For $Q = \Theta$, the expected sample information about the parameter, $M(\mathbf{y};\Theta)$ is known as Lindley's measure; Lindley (1956). It is also referred to as the *parameter information*. Lindley's measure has been widely used in Bayesian optimal design. It was first considered by Stone (1959) in the context of normal linear models. El-Sayyed (1969) used Lindley's measure for information loss due to censoring in the exponential model and Polson (1993) considered it in design of accelerated life tests.

3 Parameter and Predictive Information and Bayesian Designs

The predictive version of Lindley's measure is referred to as *predictive information*. For $Q = Y_v$, the expected information $M(\mathbf{y};Y_v)$ is referred to as the predictive information; see for example, San Martini and Spezzaferri (1984) and Amaral and Dunsmore (1985). Verdinelli et al. (1993) proposed predictive information for optimal design of accelerated life tests with lognormal lifetimes.

Verdinelli (1992) considered a linear combination of the parameter and predictive information as design criteria

$$U(\mathbf{Y};\Theta,Y_v) = w_1 M(\mathbf{Y};\Theta) + w_2 M(\mathbf{Y};Y_v), \tag{13}$$

where $w_k \geq 0,\ k = 1,2$ reflect the relative importance of the parameter and prediction. As noted by Ebrahimi et al. (2010b), since Θ and Y_v are not independent quantities, $M(\mathbf{Y};\Theta)$ and $M(\mathbf{Y};Y_v)$ are not separable. The weights in the above do not take into account the dependence between the prediction and the parameter.

Taking the dependence between the parameter and prediction into account requires the joint information. Following Ebrahimi et al. (2010b), if we let $Q = (\Theta, Y_v)$ then the observed and expected information measures are given by $\Delta H[\mathbf{y};(\Theta,Y_v)]$ and $M[\mathbf{Y};(\Theta,Y_v)]$. The joint information measures enable us to explore the relationship between $M(\mathbf{Y};\Theta)$ and $M(\mathbf{Y};Y_v)$ as given by the following result.

Theorem 1. *Let $Y_1,Y_2,\cdots$ have distributions $f_{y_i|\theta},\ i = 1,2,\cdots$ which, given θ, are conditionally independent, then*

1. $\Delta H(\mathbf{y};\Theta) = \Delta H[\mathbf{y};(\Theta,Y_v)]$;
2. $M(\mathbf{Y};\Theta) = M[\mathbf{Y};(\Theta,Y_v)]$;
3. $M(\mathbf{Y};Y_v) \leq M(\mathbf{Y};\Theta)$.

Proof of the theorem is given in Ebrahimi et al. (2010b). From Part (1) of the Theorem, we note that information provided by the (observed) sample about the parameter, is the same as joint information about the parameter and prediction. Part (2) of the theorem provides a broader interpretation of Lindley's information, namely expected information provided by the data about the parameter and for the prediction. The inequality in (3) is the Bayesian version of the information processing inequality of information theory. As suggested by Ebrahimi et al. (2010b), it may be referred to as the *Bayesian data processing inequality* mapping the information flow $\mathbf{Y} \to \Theta \to Y_v$. We note that parts (2) and (3) of Theorem 1 are due to the decomposition:

$$M[\mathbf{Y};(\Theta,Y_v)] = M(\mathbf{Y};\Theta) + \mathscr{M}(\mathbf{Y};Y_v|\Theta) = M(\mathbf{Y};Y_v) + \mathscr{M}(\mathbf{Y};\Theta|Y_v). \tag{14}$$

An immediate implication of Theorem 1 is that the design maximizing $M(\mathbf{Y};\Theta)$ also maximizes sample information about the parameter and prediction jointly. However, such optimal design may not be optimal according to $M(\mathbf{Y};Y_v)$. Similarly, the optimal design maximizing $M(\mathbf{Y};Y_v)$ may not be optimal according to $M(\mathbf{Y};\Theta)$.

When $Y_i,\ i = 1,2,\cdots$ are not conditionally independent given θ, the information decomposition is given by

$$M[\mathbf{Y};(\Theta,Y_v)] = M(\mathbf{Y};\Theta) + \mathscr{M}(\mathbf{Y};Y_v|\Theta) \geq M(\mathbf{Y},;\Theta) \tag{15}$$

where $\mathscr{M}(\mathbf{Y};Y_v|\theta) > 0$ is the measure of conditional dependence which reduces to 0 in the conditional independent case. For the conditionally dependent case we have

$$M(\mathbf{Y};Y_v) \leq M(\mathbf{Y};\Theta) \Longleftrightarrow \mathscr{M}(\mathbf{Y};\Theta|\mathbf{Y}_v) \geq \mathscr{M}(\mathbf{Y};Y_v|\Theta) \tag{16}$$

We note that under strong conditional dependence the predictive information $M(\mathbf{Y};Y_v)$ can dominate $M(\mathbf{Y};\Theta)$ the parameter information.

4 Failures versus Survivals

In a probe of the common belief that observing failures in life testing is always more informative than survivals, Abel and Singpurwalla (1994) posed the following question:

> During the conduct of the test, what would you rather observe, a failure or a survival?

The answer to the question has practical implications. For example, if a failure is preferred for inference, then one may wait until a failure occurs or perhaps even induce a failure through an accelerated environment. Abel and Singpurwalla showed that the answer to the question depends on the inferential objective of the life test.

The authors considered an observation $y = y_0$ from the exponential model,

$$f(y|\theta) = \theta e^{-\theta y}$$

and assumed a gamma prior for θ with parameters α and β. They used Shannon entropy for measuring observed information utility

$$H(\Theta|y_0) = -E_{\Theta|y_0}[\log f(\theta|y_0)]$$

As the posterior distribution of Θ is gamma with $(\alpha+1)$ and $(\beta+y_0)$ for the case of a failure at y_0 and with α and $(\beta+y_0)$ for the case of a survival at y_0, they were able to compare the gamma entropies.

The entropy of a gamma distribution with parameters a and b is given by

$$H_G(a) - log(b),$$

where

$$H_G(a) = log\Gamma(a) - (a-1)\psi(a) + a.$$

Since the scale parameter is the same, the comparison of the failure and the survival implies that

$$H_G(\alpha+1) > H_G(\alpha). \tag{17}$$

Thus, for the failure rate Θ survival gives more information than failure.

However, if the objective is to make inference about the mean $\mu = 1/\Theta$, then the failure provides more information than the survival. Having the gamma prior on

Θ implies an inverse gamma prior $1/\Theta$ and it can be shown that in this case the comparison gives

$$H_{IG}(\alpha+1) < H_{IG}(\alpha). \tag{18}$$

It is important to note that the entropy is not invariant under transformations of Θ or the data. Thus, the comparison of information about a parameter depends on the parameterization of the lifetime model.

As pointed out by Ebrahimi et al. (2013), findings by Abel and Singpurwalla raises some questions:

- Is the exponential case a counter example due to the memoriless property ?
- Can the result be generalized to other life models?
- What could possibly explain the preference for failures?
- What would you rather observe, a failure or a survival, for prediction of the lifetime of an untested item?

In order to address the above questions Ebrahimi et al. (2013) considered a more general setup where $D_n = (y_1,\cdots,y_n)$ and $D_k = (y_1,\cdots,y_k,y^*_{k+1},\cdots,y^*_n)$ denote the data provided by the failure and survival scenarios with y_i's and y^*_i's representing failure and survival times, respectively. The setup implies that the sufficient statistic for parameter Θ is the same for both scenarios, that is, $t_n(\mathbf{y}) = t_k(\mathbf{y})$. The corresponding likelihood functions for Θ are given by

$$\mathscr{L}(D_n|\theta) \propto \prod_{i=1}^{n} f(y_i|\theta), \quad \mathscr{L}(D_k|\theta) \propto \prod_{i=1}^{k} f(y_i|\theta) \prod_{i=k+1}^{n} S(y^*_i|\theta),$$

where $S(y|\theta) = P(Y > y|\theta)$ the survival function.

As before, we let Q denote the unknown quantity of interest such as a parameter Θ, a function of the parameter such as $1/\Theta$, or the lifetime of an untested item Y_v with a distribution $f(\cdot)$. Using the Abel and Singpurwalla set up, the sample D_n is said to be more informative than D_k about Q whenever

$$H(Q|D_n) < H(Q|D_k). \tag{19}$$

The comparison is well-defined for improper priors for Θ as long as $f(q|D_n)$ and $f(q|D_k)$ are proper. With proper prior, the above is equivalent to the observed information criteria

$$\Delta H(D_n;Q) = H(Q) - H(Q|D_n) > H(Q) - H(Q|D_k) = \Delta H(D_k;Q). \tag{20}$$

Ebrahimi et al. (2013) considered the class of models with survival function

$$S(y|\theta) = P(Y > y|\theta) = e^{-\theta\phi^{-1}(y)}, \tag{21}$$

where $Y = \phi(X)$, $S(x|\theta) = e^{-\theta x}$, and ϕ is an increasing function such that $\phi(0) = 0 and \phi(\infty) = \infty$. Since the survival function of X is exponential, the class of models is referred to as the time-transformed exponential (TTE) models and θ is referred to as the *proportional hazard* parameter; see Barlow and Hsiung (1983). Examples of lifetime models in the TTE family are given in Table 1. The sufficient statistics for the proportional hazard parameter under the two scenarios are the same

$$t_k(\mathbf{y}) = t_n(\mathbf{y}) = t_n = \sum_{i=1}^{n} \phi^{-1}(y_i).$$

Using the conjugate gamma prior for Θ with parameters α and β, denoted as $G(\alpha,\beta)$, the posterior distributions based on samples from models in the TTE family under both scenarios are gamma

$$f(\theta|D_s) = G(\alpha+n_s, \beta+t_n), n_s = k, n.$$

For the TTE family models, by the observed information criteria, we have:

1. For Θ
$$H_G(\alpha+n) > H_G(\alpha+k),$$
that is, survival is more informative than failure about the proportional hazard parameter Θ;
2. For $1/\Theta$
$$H_{IG}(\alpha+n) < H_{IG}(\alpha+k),$$
that is, failure is more informative than survival about the inverse parameter.

Table 1 Examples of Time Transformed Exponential Family

TTE model	$\phi(x)$
Exponential	x
Weibull	$x^{1/b}$
Linear Failure Rate	$\frac{1}{b}(a^2+bx)^{1/2}$
Pareto Type I	ae^x
Pareto Type II	e^x-1
Pareto Type IV	$(e^x-1)^{1/a}$
Extreme Value	$\log(1+x)$

As noted by Abel and Singpurwalla (1994), ”The aim of life testing is to better predict the life lengths of untested items.” In view of this, Ebrahimi et al. (2013)

investigated if the failure is more informative than the survival to predict future life lengths in the exponential model. Using the gamma prior $G(\alpha,\beta)$ in the exponential model the predictive distributions of Y_V will be Pareto. Thus, the posterior predictive entropies are given by

$$H(Y_V|n_s,t_n) = H_P(\alpha+n_s)+\log(\beta+t_n), \quad \alpha,\beta \geq 0, \quad n_s = k,n,$$

where

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

Since H_P is decreasingly ordered by α, we have

$$H_P(\alpha+n) < H_P(\alpha+k), \quad k<n. \tag{22}$$

In other words, for the exponential model, failure is more informative than survival about prediction of Y_V. Using the conjugate gamma prior for Θ, the result holds for most members of the TTE family for fixed values of other parameters a and b.

As shown by Ebrahimi et al. (2013), a more general result can be obtained in comparing informativeness of failures and survivals about prediction by ordering entropies of predictive distributions. The important quantity for the stochastic ordering of the predictive distributions is

$$\Lambda(\theta) = \prod_{i=k+1}^{n} \lambda(y_i^*|\theta), \tag{23}$$

where $\lambda(y|\theta)$ is the hazard (failure) rate function of Y. The following result provides a comparison of the entropies of predictive distributions.

Theorem 2. *Given the definition of $\Lambda(\Theta)$ in (23)*

1. *If the predictive density function $f(y_V|D_n)$ is decreasing (increasing), then D_n is more (less) informative than D_k about the prediction of Y_V, if and only if*

$$\mathrm{COV}\left[S(y_V|\Theta),\ \Lambda(\Theta)|D_k\right] < 0.$$

2. *If θ orders the survival function $S(y|\theta)$, then* $\mathrm{COV}\left[S(y_V|\Theta),\ \Lambda(\Theta)|D_k\right] < 0$.

The terms $S(y_V|\Theta)$ and $\Lambda(\Theta)$ are functions of Θ and their covariance is obtained under the posterior distribution $f(\theta|D_k)$.

It is important to note that the Theorem 2 enables us to compare the entropies of predictive distributions for many lifetime models without a need to specify any prior distribution. The result is applicable to many of the TTE models. Also, all decreasing failure rate (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 the result is satisfied. For example, the result holds for DFR models such as Pareto Type I, Pareto Type II, and Half-logistic, Weibull with $b \leq 1$, gamma with $a \leq 1$, and generalized-gamma with $ab \leq 1$, but it is not limited to the DFR models. It also applies to the IFR models: linear failure rate, Extreme Value, and models with non-monotone failure rates such as Half-Cauchy. Table 2 gives some examples where θ orders the survival function. When the conditions of the Theorem 2 do not hold, one can do the comparison directly by computing entropy of prediction under both scenarios.

Table 2 Examples where Survival (S) or Failure (F) is more Informative for Prediction

Model (More informative)	Density and Support
Half-normal (F)	$f(y\|\theta) = \sqrt{\frac{2\theta}{\pi}} e^{-\frac{\theta}{2}y^2},\ y \geq 0$
Half-Cauchy (F)	$f(y\|\theta) = \frac{2}{\pi\theta}\left(1+\frac{y^2}{\theta^2}\right)^{-1},\ y \geq 0$
Half-logistic (F)	$f(y\|\theta) = \frac{\theta e^{-\theta y}}{(1+e^{-\theta y})^2}, y \geq 0$
Gamma $(a \leq 1)$ (F)	$f(y\|a,\theta) = \frac{\theta^a}{\Gamma(a)} y^{a-1} e^{-\theta y},\ y \geq 0$
Generalized gamma $(ab \leq 1)$ (F)	$f(y\|a,b,\theta) = \frac{b\theta^a}{\Gamma(a)} y^{ab-1} e^{-\theta y^b}, y \geq 0$
Generalized Pareto $(a > 0, , \theta \neq 1)$ $\theta < 1$ (F) $\theta > 1$ (S)	$f(y\|\theta) = a\left(1-\frac{\theta y}{a}\right)^{1/\theta - 1}, \begin{cases} y \geq 0, \theta < 0 \\ 0 < y \leq a/\theta, \\ \theta > 0 \end{cases}$
Power ($\theta < 1$, F) ($\theta > 1$, S)	$f(y\|\theta) = \theta y^{\theta-1},\ 0 < y \leq 1$
Beta ($\theta < 1$, S) ($\theta > 1$, F)	$f(y\|\theta) = \theta(1-y)^{\theta-1},\ 0 \leq y \leq 1$

Ebrahimi et al. (2013) considered expected sample information as an alternative criterion to the observed information for a plausible explanation for the perception that failures are more informative. The expected information was measured by the conditional entropy $\mathscr{H}(Q|D_s)$ given by

$$\mathscr{H}(Q|D_k) = E_k\{H(Q|D_k)\} = \int H(Q|D_k) f_k(\mathbf{y}) d\mathbf{y},\ \ k = 0, 1, \cdots, n, \tag{24}$$

where E_k denotes averaging with respect to $f_k(\mathbf{y})$. Using the conditional entropy criteria and the conjugate gamma prior for Θ, the authors showed that for all members of the TTE family, failure is more informative than survival about prediction of a new lifetime Y_v, the proportional hazard parameter Θ, and its inverse 1Θ. Furthermore, they showed that unlike the observed information measured by $-H(\Theta|D_k)$, the expected information measured by $-\mathscr{H}(\Theta|D_k)$ is increasing in the number of failures k.

This result may be interpreted as, on average, observing a failure is more informative than observing a survival. As noted by Kruskal (1987), thinking in terms of averages is a tradition in statistics, and this provides a plausible explanation for the perception that failures are more informative.

5 Reliability Importance of System Components

Birnbaum (1969) defined reliability importance of a component i for coherent systems as

$$I_i^B(t) = \frac{\partial \overline{F}(t)}{\partial \overline{F}_i(t)}, \tag{25}$$

where $\overline{F}(t)$ is the system reliability and $\overline{F}_i(t)$ is the component i's reliability at time t. Barlow and Proschan (1975) introduced another measure of relative reliability importance as the conditional probability that the system's failure is caused by component i's failure. It can be shown that

$$I_i^{BP}(t) = \int_0^\infty I_i^B(t)\, dF_i(t)\, dt \tag{26}$$

where $F_i(t)$ is the distribution function for lifetime of component i and $\sum_i I_i^{BP}(t) = 1$. Alternative measures of reliability importance were proposed by Natvig (1979), Natvig (1985), and Armstrong (1995). More recent results are given in Natvig and Gsemyr (2009). All these measures are in terms of contribution of a component to the reliability of the system.

Ebrahimi et al. (2014) noted that reliability importance can be interpreted in terms of how knowledge of status of a component changes our knowledge of the system. In other words, their interpretation is in terms of which component's status knowledge matters most in reducing our uncertainty about systems' status. The authors suggested an alternative notion of component importance in terms of information measures.

Consider a system that consists of n components $C_1, \cdots, C_n$ that collectively determine a random variable of interest Q for the system. Let Q_i be the corresponding random variable associated with $C_i, i = 1, \cdots, n$. More formally, let $Q_i = Q_i(C_i), \mathbf{Q} = (Q_1 \cdots Q_n)$ and define the *system structure function* as

$$Q = \phi(\mathbf{Q}), \phi : \Re^n \to \Re. \tag{27}$$

We note that Q_i can be the indicator variable of the state (failure/survival) or the life length of a component and Q is the respective random variable for the system.

The information notion of importance maps the expected utility of the component variable Q_i for prediction of the system variable Q in terms of the dependence implied by the joint distribution $F(q, q_i)$. More specifically, it is defined as follows.

Definition 1. The importance of component C_i is defined by the expected information utility of C_i for the system measured by the mutual information $M(Q;Q_i)$, provided that $F(q,q_i) \ll F_q(q)F_i(q_i)$. Component C_i is more important than the component C_j for the system if and only if $M(Q;Q_i) \geq M(Q;Q_j)$.

Under the information notion of importance component C_i is more important than the component C_j for the system if and only if

$$M(Q;Q_i) \geq M(Q;Q_j) \iff \mathscr{I}(Q|Q_i) \geq \mathscr{I}(Q|Q_j), \tag{28}$$

where $\mathscr{I}(Q|Q_i) = -\mathscr{H}(Q|Q_i)$ is the conditional information.

Let $T_1, \cdots, T_n$ denote independent random variables representing the life lengths of components $C_1, \cdots, C_n$ and T denote the life length of the system. The survival of system is defined as the survival up to a mission time τ. We define the binary variables for the states of the component and system as

$$Q_i(C_i) = X_i(\tau) = 1(0), \;\; if \;\; T_i > \tau \;\; (T_i \leqslant \tau)$$

and

$$Q = X(\tau) = \phi(X_1(\tau), \ldots, X_n(\tau)) = 1(0), \;\; if \;\; T > \tau \;\; (T \leqslant \tau),$$

where $\phi : \{0,1\}^n \to \{0,1\}$.

The marginal distributions of $X_i(\tau)$ and the conditional distribution of $X(\tau)$ given $X_i(\tau) = x_i$ are Bernoulli with parameters $p_i = p_i(\tau) = \overline{F_i}(\tau)$ and $p_{x|x_i} = p_{x|x_i}(\tau) = P(X(\tau) = x|X_i(\tau) = x_i)$ for $i = 1, \ldots n$, respectively. For each mission time τ, we obtain the conditional information $\mathscr{I}[(X(\tau)|X_i(\tau)]$ as

$$\mathscr{I}[(X(\tau)|X_i(\tau)] = p_i(\tau)I[(X(\tau)|X_i(\tau) = 1] + (1 - p_i(\tau))I[(X(t)|X_i(t) = 0]. \tag{29}$$

Note that this measure ranks the importance of components for a fixed mission time τ.

The following result by Ebrahimi et al. provides a ordering of the components for three types of systems at a fixed time point $t = \tau$.

Theorem 3. *Consider a system with n independent components such that* $P(X_i = 1) = p_i, i = 1, \cdots, n$.

1. *If the system is series, then the component* C_j *is more important than the component* C_i*, for* $p_j < p_i,\ i, j = 1, \cdots, n$.
2. *If the system is parallel, then the component* C_j *is more important than the component* C_i*, for* $p_j > p_i,\ i, j = 1, \cdots, n$.
3. *If the system is k-out-of n and for any i,* $P\left\{S_{(i)}(\boldsymbol{X}(\tau)) \geq k\right\} = p_i(k) \geq \frac{1}{2},\ i = 1, \cdots, n$*, then the component* C_j *is more important than the component* C_i*, where*
$$S_{(i)}(\boldsymbol{v}) = \sum_{k \in \mathscr{N}_i} v_k, \quad \mathscr{N}_i = \{1, \cdots, i-1, i+1, \cdots, n\}.$$
for a vector $\boldsymbol{v} = (v_1, \cdots, v_n)$.

The information measure (28) orders the components in the same way as the reliability importance index of Birnbaum (1969). The information analog of Barlow and Proschan's (1975) measure of component's importance is the expected mutual information $E_i\left[M\Big(X(\tau),X_i(\tau)\Big)\right]$ where the expected value is with respect to the distribution of T_i. Thus, we can order the components by evaluating $E_i\left[\mathscr{I}\Big(X(\tau)|X_i(\tau)\Big)\right]$.

As shown in Ebrahimi et al. (2014), stochastic ordering of life times of the components $T_1 \overset{st}{\leq} \ldots \overset{st}{\leq} T_n$ is sufficient for Theorem 3 to hold. The next example illustrations the implementation of the Theorem.

Example

Consider a system of two components with independent lifetimes.

1. *Bernoulli Distributions*

   We can obtain the conditional information measures for the series and parallel systems as

$$\mathscr{I}\Big(X(\tau)|X_i(\tau)\Big) = p_i(\tau)I(X_j(\tau)) \quad \text{and} \quad \mathscr{I}\Big(X(\tau)|X_i(\tau)\Big) = (1-p_i(\tau))I(X_j(\tau)),$$

   respectively. Thus, C_i is more (less) important for a series system than for a parallel system whenever $\tau > (<)$ median lifetime.

2. *Proportional Hazard Distributions*

   Suppose that the components' lifetimes, $T_i, i = 1,2$, have proportional hazard (PH) distributions

$$\bar{F_i}(\tau) = [\bar{F_0}(\tau)]^{\theta_i},\ \tau \geq 0,\ \theta_i > 0.$$

   For $\theta_1 > \theta_2$, we have $T_1 \overset{st}{\leq} T_2$. Thus for any mission time $\tau,\ p_1(\tau) < p_2(\tau)$ and by Theorem 3 C_1 is more (less) important than C_2 for the series (parallel) system.

3. *TTE Family of Distributions*

   Suppose that the components' lifetimes $T_i, i = 1,2$, having TTE family of distributions as defined in Section 4. Note that the TTE model is a PH model and therefore we can conclude that for $\theta_1 > \theta_2$, $T_1 \overset{st}{\leq} T_2$. Thus, for any mission time $\tau,\ p_1(\tau) < p_2(\tau)$ and C_1 is more (less) important than C_2 for the series (parallel) system.

The information based analog of Barlow-Proschan importance index can also be computed for TTE models. It can be shown that $E_i\left[\mathscr{I}\Big(X(t)|X_i(t)\Big)\right]$ for the series and parallel systems are functions of $\theta = \dfrac{\theta_i}{\theta_j}$, decreasing (increasing) for $\theta < (>) 1$.

The notion of information importance is also applicable to component and system lifetimes as continuous random variables. As shown in Ebrahimi et al. (2014), it is possible to develop results for convolutions and order statistics. For convolutions, results in entropy ordering can be used to obtain component importance ordering. Considering parallel and series systems with continuous lifetimes, the system lifetime being order statistics leads to singular distributions, where the mutual information is not well-defined. In order to alleviate this problem, a modification of the information importance index was used by the authors.

Ebrahimi et al. (2014) also considered an entropy-based importance measure using the Maximal Data Information Prior (MDIP) criterion proposed by Zellner (1977). The MDIP criterion which was originally proposed for developing priors for Bayesian inference, provides the same importance ordering of the components as the mutual information.

6 Concluding Remarks

Information measures play an important role in Bayesian reliability analysis. In this paper our focus was on Bayesian information measures in failure data analysis, prediction, assessment of reliability importance and optimal design of life tests. Other uses of these measures include failure model selection [Mazzuchi et al. (2000, 2008)], characterization of univariate and multivariate failure models and characterization of dependence for system reliability [Ebrahimi et al. (2008)].

Computation of information functions can be quite challenging in many applications but decomposition type results given in Ebrahimi et al. (2008) can be helpful for certain problems. Parametric and nonparametric Bayesian estimation of information functions and related indices [see Dadpay et al. (2007)] are required for multivariate life models. Recent advances in Bayesian computing, especially, those in efficient Markov chain Monte Carlo methods can be exploited in many cases.

References

1. Abel, P. S., Singpurwalla, N. D.: To Survive or to Fail: That is the Question. American Statistician. **48**, 18–21 (1994)
2. Armstrong, M. J.: Joint Reliability-importance of Components. IEEE Transactions on Reliability. **44**, 408–412 (1995)
3. Barlow, R. E., Hsiung, J. H.: Expected Information from a Life Test Experiment. Statistician. **48**, 18–21 (1983)
4. Barlow, R. E., Proschan, F.: Importance of System Components and Fault Tree Event. Stochastic Processes and Their Applications. **3**, 153–172 (1975)
5. Amaral-Turkman, M.A., Dunsmore, I.: Measures of Information in the Predictive Distribution. In: Bernardo, J. M., DeGroot, M. H., Lindley, D. V., Smith, A.F.M. (eds.) Bayesian Statistics, 2, pp. 603–612. Elsevier, Amsterdam (1985)
6. Bernardo, J.M.: Expected Information as Expected Utility. The Annals of Stat. **7**, 686–690 (1979a)

7. Bernardo, J.M.: Reference Posterior Distribution for Bayesian Inference. J. R. Statist. Soc. Ser. B. **41**, 605–647 (1979b)
8. Birnbaum, Z. W.: On the Importance of Different Components in a Multicomponent System. In: Krishnaiah, P. R. (eds.) Multivariate Analysis II, pp. 581–592. Academic Press, New York (1969)
9. Chaloner, K., Verdinelli, I.:Bayesian Experimental Design: A Review. Statistical Science. **10**, 273–304 (1995)
10. Dadpay, A., N., Soofi, E. S., Soyer, R.:Information Measures for Generalized Gamma Family. Journal of Econometrics. **138**, 568–585 (2007)
11. Ebrahimi, N., Soofi, E. S.: Recent Developments In Information Theory and Reliability Analysis. In: Basu, A. P., T. A., Basu, S. K., Mukhopadhyay, S. (eds.) Frontiers in Reliability, pp. 125–132.World Scientific, New Jersey (1998)
12. Ebrahimi, N., Soofi, E. S.: Information Functions for Reliability. In: Soyer, R., Mazzuchi, T. A., Singpurwalla, N. D. (eds.) Mathematical Reliability: An Expository Perspective, pp. 127–159. Kluwer, New York (2004)
13. Ebrahimi, N., Soofi, E. S., Soyer, R.: Multivariate Maximum Entropy Identification, Transformation, and Dependence. Journal of Multivariate Analysis. **99**, 1217–1231 (2008)
14. Ebrahimi, N., Soofi, E. S., Soyer, R.: Information Measures in Perspective. International Statistical Review. **78**, 383–412 (2010a)
15. Ebrahimi, N., Soofi, E. S., Soyer, R.: On the Sample Information about Parameter and Prediction. Statistical Science. **25**, 348–367 (2010b)
16. Ebrahimi, N., Soofi, E. S., Soyer, R.: When are Observed Failures More Informative than Observed Survivals ? Naval Research Logistics. **60**, 102–110 (2013)
17. Ebrahimi, N., Jalali, N. Y., Soofi, E. S., Soyer, R.: Importance of Components for a System. Econometric Reviews. **33**, 395–420 (2014)
18. El-Sayyed, G. M.: Information and Sampling from Exponential Distribution. Technometrics. **11**, 41–45 (1969)
19. Kruskal, W.: Relative Importance by Averaging over Orderings. The American Statistician. **41**, 6–10 (1987)
20. Lindley, D. V.: On a Measure of Information Provided by an Experiment. Annals of Math. Stat. **27**, 986–1005 (1956)
21. Mazzuchi, T. A., Soofi, E. S., Soyer, R.: Computations of Maximum Entropy Dirichlet for Modeling Lifetime Data. Computational Statistics and Data Analysis. **32**, 361–378 (2000)
22. Mazzuchi, T. A., Soofi, E. S., Soyer, R.: Bayes Estimate and Inference for Entropy and Information Index of Fit. Econometric Reviews. **27**, 428–456 (2008)
23. Natvig, B.: A Suggestion of a New Measure of Importance of System Component. Stochastic Processes and Their Applications. **9**, 319–330 (1979)
24. Natvig, B.: New Light on Measures of Importance of System Components. Scandinavian Journal of Statistics. **12**, 43–54 (1985)
25. Natvig, B., Gsemyr, J.: New results on the Barlow-Proschan and Natvig Measures of Component Importance in Nonrepairable and Repairable Systems. Methodology And Computing in Applied Probability. **11**, 603–620 (2009)
26. Polson, N. G.: A Bayesian Perspective on the Design of Accelerated Life Tests. In: Basu, A. P.(eds.) Advances in Reliability, pp. 321–330. North Holland, Amsterdam (1993)
27. San Martini, A., Spezzaferri, F. :A Predictive Model Selection Criteria.. J. R. Statist. Soc. Ser. B. **46**, 296–303 (1984)
28. Stone, M.: Application of a Measure of Information to the Design and Comparison of Regression Experiments. Annals of Math. Stat. **29**, 55–70 (1959)
29. Verdinelli, I.: Advances in Bayesian Experimental Design. In: Berger, J. O., Bernardo, J. M., Dawid, A. P., Smith, A.F.M. (eds.) Bayesian Statistics, 4, pp. 467–481. Oxford University, Oxford (1992)
30. Verdinelli, I., Polson, N.G., Singpurwalla, N.D.: Shannon Information and Bayesian Design for Prediction in Accelerated Life-testing. In: Barlow, R. E., Clarotti, C. A., Spizzichino, F. (eds.) Reliability and Decision Making, pp. 247–256. Chapman Hall, London (1993)

31. Zellner, A.: Maximal Data Information Prior Distributions. In: Aykac, A., Brumat, C. (eds.) New Developments in the Applications of Bayesian Methods., pp. 211–232. North-Holland, Amsterdam (1977)
32. Zellner, A.: Optimal Information Processing and Bayes' Theorem. American Statistician. **42**, 278–284 (1988)