Journal of the Korean
Statistical Society
Vol. 23, No. 2, 1994

Comparison of Statistical Experiments and Measures of Information†

KeonTae Sohn[1] and JongWoo Jeon[2]

ABSTRACT

The comparison of statistical experiments with a common parameter and parameter space is discussed using the concept of the Blackwell's sufficiency and the Shannon's entropy. Binomial and censored experiments are considered as applications. The loss of information is studied under the aggregated experiments and truncated experiments, and summerized in some tables which make it possible to indicate the choice of an appropriate experiment.

KEYWORDS: Blackwell's sufficiency, Measures of information, Loss of information, Aggregation, Truncated experiment.

---

†This paper was supported (in part) by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1991
[1]Department of Statistics, Pusan National University, Jangjundong, Kumjunggoo, Pusan, 609-735, Korea.
[2]Department of Computer Science and Statistics, Seoul National University, Sinrimdong, Kwanakgoo, Seoul, 151-742, Korea.

1. INTRODUCTION

Many problems in Statistics begin with a parameter of interest θ, a state of nature, about which we do not have enough information. In order to generate further information about a parameter θ, we plan a statistical experiment which generates the sample x. A *statistical experiment* **E** is defined as a triple $\{S, \Theta, p\}$ where (1) S, the sample space, is the set of all the possible samples x from a performance of **E**, (2) Θ, the parameter space, is the set of all the possible values of parameter θ, and (3) $p = p(x \mid \theta)$ is the probability function defined on $S \times \Theta$. If S is generated by a random variable X, we will notate that $E = \{X, \Theta, P_\theta\}$ where X has a distribution P_θ depending on an unknown parameter $\theta \in \Theta$.

Consider two statistical experiments $E = \{X, \Theta, P_\theta\}$ and $F = \{Y, \Theta, Q_\theta\}$ where P_θ and Q_θ depend on an unknown parameter θ belonging to the common parameter space Θ. If, at the planning stage of the experiment, we are asked to choose one between two experiments, we shall choose one experiment with the large information about the parameter. In order to compare two experiments, statisticains have used various methods as follows:

(1) *decision-theoretic approach*, to compare the risk sets of two experiments for any decision procedure and to choose one experiment with larger risk set.

(2) to find the sufficiency for a parameter between two experiments, and

(3) to propose the measures of information.

Bohnenblust, Shapley and Sherman introduced a method of comparing two experiments in an unpublished Rand Memorandum, according to Blackwell (1951). Their concept is that an experiment **E** is more informative than an another experiment **F**, if, for every possible risk function, any risk attainable with **F** is also attainable with **E**. As a generalization of the Fisher's sufficiency, Blackwell (1951,1953) introduced a concept of sufficiency, called the Blackwell's sufficiency, which needs the existence of an appropriate stochastic kernel and showed that if **E** is sufficient for **F**, **E** is more infomative than **F**. He also gave some equivalent conditions of sufficiency and *being more informative.*

In order to compare the experiments, it is reasonable to consider a simple function which represents the statistical information contained in the experiments. Shannon (1948) introduced an information measure, called *the entropy*, which is a function of distribution measuring the amount of the information in the sense of a communication engineer. He considered the information based on the distribution of the transmitted messages and that of the received messages. Lindley (1956) introduced an information measure which is the expected difference between the entropy of a prior distribution over parameter space and the entropy of the posterior distribution. The measure is considered as the expected gain in information provided by an experiment. Using this measure, he compared the statistical experiments. He also proved that the ordering of the Blackwell's sufficiency preserves the ordering of Lindley information measure.

Bayarri and DeGroot (1987) discussed the selection models. Using the Fisher information number, they give a useful criterion for comparing the selection model with an unrestricted model and consider the general exponential families. The problems of the loss of information on the censored experiments was discussed by Brooks (1980,1982), Turrero (1989) and Ebrahimi and Soofi (1990). Using the Lindley information measure, they consider the Type I censoring, the Type II censoring and the group censoring.

We will give some results for comparison of binomial and censored experiments using the Blackwell's sufficiency. We will show whether an incomplete experiment is less informative than the complete experiment in aggregated experiments and truncated experiments using the expected gain in information. And then we make some tables which indicate an appropriate choice of points of truncation in the normal and exponential experiments.

2. BLACKWELL'S SUFFICIENCY

Consider a statistical experiment $\mathbf{E} = \{X, \Theta, P_\theta\}$ where a random variable X has a distribution P_θ depending on an unknown parameter $\theta \in \Theta$. When the size of the parameter space θ is finite N, we can say that an experiment $\mathbf{E}$ is a set of N probability mesures $P_1, \ldots, P_N$. Let x be a sample point and A be a bounded action set of N dimensional space. The point $a \in A$ is considered as a possible action . The loss from action $a = (a_1, \ldots, a_N)$ is a_i if the actual distribution of X is P_i . A risk vector r is given by

$$r(x) = \Big(\int a_1(x)dP_1, \ldots, \int a_N(x)dP_N\Big).$$

Let $R(E, A)$ be the convex hull of the risk set of all risk vectors in an experiment $\mathbf{E}$.

Let $\mathbf{E} = \{Y, \Theta, Q_\theta\}$ be another statistical experimemt with the common parameter and parameter space Θ and $R(\mathbf{E}, A)$ be the convex hull of the risk set in an experiment $\mathbf{F}$.

Definition 2.1 (Bohnenblust, Shaplay and Sherman). $\mathbf{E}$ is *more informative* than $\mathbf{F}$, written $\mathbf{E} \supset \mathbf{E}$, if we have $R(\mathbf{E}, A) \supset R(\mathbf{E}, A)$ for every A.

Example 2.1. Consider the testing for simple versus simple. Suppose that two different experiments are available for the above testing and these are given by

experiment	random variable	H_0	H_1	MP-α test	power
$\mathbf{E}$	X	f	g	ϕ	$\beta(\alpha)$
$\mathbf{F}$	X'	f'	g'	ϕ'	$\beta'(\alpha)$

Let a_1 (a_2) be actions which represent to accept (reject) the null hypothesis. The risk vector for the experiment E can be given by

$$r = \Big(\int a_i(x)f(x)dx, \int a_j(x)g(x)dx\Big) \quad i, j = 1, 2.$$

Suppose that $\beta'(\alpha) \leq \beta(\alpha)$. We can easily find that $R(\mathbf{E}) \supset R(\mathbf{E})$. By Definition 2.1, $\mathbf{E}$ is more informative than $\mathbf{F}$ if and only if $\beta'(\alpha) \leq \beta(\alpha)$.

Blackwell (1951,1953) introduced a concept which is a generalization of the Fisher's sufficiency as follows.

Definition 2.2 (Blackwell). An experiment $\mathbf{E}$ is said to be *sufficient* for another experiment $\mathbf{F}$, written $\mathbf{E} \succ \mathbf{E}$, if there exists a stochastical kernel $H(\cdot,\cdot)$ such that for any $\theta \in \Theta$,

(1) $H(x, A)$ is a measurable function of x for each set A of reals,

(2) $H(x, \cdot)$ is a probability measure and independent of θ for each x, and

(3) $Q_\theta(Y \in A) = \int H(x, A) dP_\theta$.

Blackwell (1951) showed that $\mathbf{E}$ is more informative than $\mathbf{F}$ if $\mathbf{E}$ is sufficient for $\mathbf{F}$. There is an equivalent definitions to the Blackwell's sufficiency as follows.

Definition 2.3. An experiment $\mathbf{E}$ is said to be *sufficient* for $\mathbf{F}$ if there exists a function $h(x, U)$ where U is a random variable distributed independently of X, with a known distribution, such that the distribution of $h(X, U)$ and Y are identical for any $\theta \in \Theta$.

Let an experiment $\mathbf{E}$ be a combination of n independent experiments $\mathbf{E}_1, \ldots, \mathbf{E}_n$. The probability measure in $\mathbf{E}$ is the product measure generated by the probability measures in $\mathbf{E}_i$'s. Let another experiment $\mathbf{F}$ be a combination of n independent experiments $\mathbf{E}_1, \ldots, \mathbf{E}_n$. Blackwell (1951) showed that the ordering between $\mathbf{E}$ and $\mathbf{F}$ preserves the ordering between $\mathbf{E}_i$ and $\mathbf{E}_i$, $i = 1, \ldots, n$, that is, $\mathbf{E}_i \succ \mathbf{E}_i$, $i = 1, \ldots, n$, imply that $\mathbf{E} \succ \mathbf{E}$.

Example 2.2 (Lehmann, 1986). Consider two normal experiments $\mathbf{E}$ and $\mathbf{F}$ based on $N(\theta, \sigma^2)$ and $N(\theta, a^2\sigma^2), 0 < a < 1$, respectively.

Case 1: σ^2 is known and θ is unknown.

Let Z have the normal distribution $N(0, (1 - a^2)\sigma^2)$ and independent of Y. Since the distribution of $Y + Z$ is identical to X, $\mathbf{F}$ is sufficient for $\mathbf{E}$.

The UMVUE based on **E** and that based on **F** are X and Y respectively. We cannot find an unbiased estimator based on X, whose variance is less than or equal to the variance of Y. Therefore **F** is strictly sufficient for **E**.

Case 2: θ is known and σ^2 is unknown.
Take $\theta = 0$. Then Y/a and X are identically distributed, and aX and Y are also identically distributed. Therefore **F** is as informative as **E**.

Case 3: Both of θ and σ^2 are unknown.
Hansen and Torgersen (1974) showed that **E** and **F** are not comparable.

Example 2.3. Let X and Y be binomial random variables with the proportion θ and $k\theta$ of success, respectively. Then we can show that Y is strictly more informative than X if $k > 1$ and $k\theta < 1$. Since X and Y are the sum of independent and identically distributed Bernoulli random variables, it is sufficient to consider the case that $n = 1$.

(1) Take a loss function $L(\theta, d) = (\theta - d)^2$. Then $E(\theta - X)^2 = var(X) = \theta(1-\theta)$ and $E(\theta - Y/k)^2 = \frac{1}{k^2}var(Y) = \frac{1}{k}\theta(1-k\theta)$. Since $\theta(1-\theta) - \frac{1}{k}\theta(1-k\theta) = (1-k^{-1})\theta > 0$ for any $0 < \theta < 1$, X is not as informative as Y.

(2) Let Z have a uniform distribution $(0,1)$ and be independent of Y. Consider the distribution given by

$$h(Y, Z) = \begin{cases} 1, & \text{if } Y = 1 \text{ and } Z \le 1/k \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} P(h(Y,Z) = 1) &= P(Y=1)P(Z \le 1/k) \\ &= \theta \\ &= P(X = 1). \end{aligned}$$

By Definition 2.3, Y is sufficient for X. Therefore Y is strictly more informative than X.

Example 2.4. Consider a Type I censored experiment. Let X be a lifetime random variable with a distribution function $F(\cdot \mid \theta)$ and a density function $f(\cdot \mid \theta)$, $\theta \in \Theta$. Let c be a fixed censoring time. Then $Y = min\{X, c\}$ is an observable value with a density function given by

$$g(y) = \begin{cases} f(y), & \text{if } x \leq c \\ \bar{F},(y) & \text{if } x > c. \end{cases}$$

Consider two censored experiments, $\mathbf{E} = \{Y_1, \Theta\}$ and $\mathbf{E} = \{Y_2, \Theta\}$ where $Y_1 = min\{X, c_1\}$, $Y_2 = min\{X, c_2\}$ and $c_1 \geq c_2$. Let y_1 be a result from $\mathbf{E}$. Then we can consider a stochastic kernel $H(y_1, \cdot)$ given by

$$H(y_1, A) = \begin{cases} I_A(y_1), & \text{if } A \subset (0, c_2) \\ I_{(c_2,\infty)}(y_1), & \text{if } A = \{c_2\} \\ 0, & \text{otherwise} \end{cases},$$

$$\text{where } I_A(y) = \begin{cases} 1, & \text{if } y \in A \\ 0, & \text{otherwise} \end{cases}.$$

Then $H(y_1, A)$ is a measurable function of y_1 for each set A of reals. $H(y_1, \cdot)$ is a probability measure and independent of θ for each y_1.

If $A \subset (0, c_2)$, then

$$\begin{aligned} P(y_2 \in A \mid \theta) &= \int_A f(x)dx \\ &= \int_A dP_\theta(y_1) \\ &= \int H(y_1, A)dP_\theta(y_1). \end{aligned}$$

If $A = \{c_2\}$, then

$$P_\theta(Y_2 = c_2) = \int_{c_2}^{\infty} f(x)dx$$

$$= \int_{c_2}^{c_1} f(x)dx + \int_{c_1}^{\infty} f(x)dx$$

$$= \int_{c_2}^{c_1} dP_\theta(y_1) + P_\theta(Y_1 = c_1)$$

$$= \int H(y_1, \{c_2\})dP_\theta(y_1).$$

Therefore $P(Y_2 \in A \mid \theta) = \int H(y_1, A)dP_\theta(y_1)$ for any set A of reals. By Definition 2.3, the experiment $\mathbf{E}$ is sufficient for the experiment $\mathbf{F}$.

With taking $c_1 = \infty$, it is clear that an uncensored experiment based on a random variable X is sufficient for any Type I cenored experiment based on the same random variable X.

Consider a Type II censored experiments. Let $(X_1, X_2, \ldots, X_n)$ be the order statistic of a random sample $\{X_1, \ldots, X_n\}$ from $F(\cdot \mid \theta)$ and $f(\cdot \mid \theta)$. Let r be a fixed integer such that $r < n$. Then an observable value in this experiment is

$$Y_i = \begin{cases} X_i, & \text{if } X_i \leq X_{(r)} \\ X_{(r)}, & \text{if } X_i > X_{(r)} \end{cases},$$

where $i = 1, \ldots, n$.

Consider the conditional probability of $Y = (Y_1, \ldots, Y_n)$, given $X = (X_1, \ldots, X_n)$. The value of $P(Y \in A \mid X = x)$ is one or zero and independent of θ. Take $H(x, A) = P(Y \in A \mid X = x)$. Then

$$P(Y \in A \mid \theta) = \int H(x, A)dP_\theta(x).$$

Therefore an uncensored experiment is sufficient for any type II censored experiment.

3. MEASURES OF INFORMATION

In this section, we introduce the Shannon's entropy and the loss of information which we shall use in Section 4.

3.1 Shannon's Entropy.

In the classical framework, the Fisher information number is usually used to measure the information about a parameter θ. The Fisher information number is defined as

$$I_F(\theta) = E\Big(\frac{d}{d\theta}\log p(X \mid \theta)\Big)^2.$$

Consider the statistical experiment from the standpoint of the information theory in communication engineering. Suppose that the knowledge of the state of nature is determined by a finite number of parameters. The knowledge before and after experiment can be viewed as a transmitted and a received message, respectively. The comparison of the knowledge before and after the experiment makes it possible to discuss the amount of information provided by the experiment. Shannon (1948) has introduced two ideas : (1) the first is that information is a statistical concept and (2) the second is that there is a function of the distribution which measures the amount of information. His idea is that the amount of information in a prior distribution can be measured by how much information it is necessary to provide before the value of parameter is known.

Suppose that a parameter space Θ consists of k elements, $\theta_1, \ldots, \theta_k$, and that a prior distribution on θ is $\pi = (p_1, \ldots, p_k)$.

Definition 3.1. The Shannon information measure, called *the entropy*, is defined as

$$H(\pi) = -\sum_{i=1}^{k} p_i \log p_i,$$

where $\pi(\theta) = (p_1, \ldots, p_k)$. If $\pi(\theta)$ is continuous, the entropy is defined as

$$H(\pi) = -\int \pi(\theta)\log \pi(\theta) d\theta.$$

The difference between the information uncertainty before and after an experiment measures the expected gain in information provided by the experiment.

Definition 3.2. *The expected gain in information* provided by the experiment, with a prior $p(\theta)$, is

$$I_{EG}(\mathbf{E}, p(\theta)) = E_X(I(\theta \mid X) - I(\theta))$$

where $I(\theta) = \int p(\theta) \log p(\theta) d\theta$.

This amount of information is the expected difference between the entropy of the prior distribution over θ and the entropy of the posterior distribution with respect to a random variable X such that

$$\begin{aligned} I_{EG}(X; p(\theta)) &= E_X E_{\theta|X}[\log p(\theta \mid X)] - E_\theta[\log p(\theta)] \\ &= E_X E_{\theta|X}\left[\log \frac{p(\theta \mid X)}{p(\theta)}\right] \\ &= E_\theta E_{X|\theta}\left[\log \frac{p(x \mid \theta}{p(X)}\right] \\ &= \int\int p(x,\theta) \log \frac{p(x,\theta)}{p(x)p(\theta)} dx d\theta. \end{aligned}$$

Lindley (1956) discussed the properties of this measure and the results are given below.

(1) $I_{EG}(\mathbf{E}) \geq 0$. $I_{EG}(\mathbf{E}) = 0$ if and only if $p(x \mid \theta)$ does not depend on θ , that is, $p(x \mid \theta) = p(x)$.

(2) $I_{EG}(\mathbf{E}, p(\theta))$ is a concave function of $p(\theta)$.

3.2 Loss of Information.

Consider a complete experiment $\mathbf{E} = \{X, \Theta, p(x \mid \theta)\}$ and an incomplete experiment $\mathbf{E} = \{Y, \Theta, p(y \mid \theta)\}$. It is generally anticipated that the incomplete experiment contains less information than the complete experiment. In this case, we want to measure the loss of information. There are two methods for measuring the loss of information using the Shannon's concepts.

(1) *maximum likelihood approach*: At first we find the maximum likelihood estimator $T = T(X)$ for θ in the experiment $\mathbf{E}$, and $W = W(Y)$ for θ in the experiment $\mathbf{F}$. Using the distribution of T and W, we can calculate the entropy such that

$$I(\mathbf{E}) = \int p(t \mid \theta) \log p(t \mid \theta) dt$$

$$I(\mathbf{E}) = \int p(w \mid \theta) \log p(w \mid \theta) dw.$$

The relation loss of information is defined as

$$L(\mathbf{E}, \mathbf{E}) = \frac{I(\mathbf{E}) - I(\mathbf{E})}{I(\mathbf{E})}.$$

(2) *Bayesian approach*: Let $p(\theta)$ be a given proir density of θ. Using the expected gain in information, we can measure the relative loss of information in the Bayesian sense. Let $p(\theta \mid x)$ be the posterior density of θ, given $X = x$, and $p(x)$ be the marginal density of X. Similarly, $p(\theta \mid y)$ and $p(y)$ are the posterior density of θ, given $Y = y$, and the marginal density of Y respectively. Then the expected gain provided by E and F are

$$I_{EG}(\mathbf{E}) = E_X E_{\theta|X}(\log p(\theta \mid X)) - E_\theta(\log p(\theta))$$

$$I_{EG}(\mathbf{E}) = E_Y E_{\theta|Y}(\log p(\theta \mid Y)) - E_\theta(\log p(\theta)).$$

Then the relative loss of information in a Bayesian sense is

$$L(\mathbf{E}, \mathbf{E}) = \frac{I_{EG}(\mathbf{E}) - I_{EG}(\mathbf{E})}{I_{EG}(\mathbf{E})}.$$

4. AGGREGATED AND TRUNCATED EXPERIMENTS

4.1 Aggregated Data.

In some scientific investigation, the data can be generated from a designed experiment in terms of the some time unit t. However, in some cases, the investigator cannot choose the time interval since controlled experiments may not be possible. That is, the assumed time unit in the experiment may not be the same as the time unit for the available observed data. In this section, suppose that data are available only through aggregation sampling.

Let $Z_t,\ \ t = 1, \ldots, n$, be a sequence of disaggregated observations and Z_t follows an identical normal distribution $N(\theta, \sigma^2)$ independently for each t. Let $X_T,\ \ T = 1, \ldots, l$, be a sequence of aggregated observations such that

$$X_T = \sum_{t=k(T-1)+1}^{kT} Z_t, \quad kl = n,$$

where T is the aggregation-time unit and k is the size of aggregation. Then $X_T \sim N(k\theta, k\sigma^2)$.

Case 1:. *σ^2 is known and θ is unknown.* Let an experiment **E** be the combination of experiments based on $Z_t, t = 1, \ldots, n$. Since Z_t's are independent, the amount of information provided by **E** can be drived as follows.

$$I_F(\mathbf{E}) = -nE_{Z_1}\Big(\frac{\partial^2}{\partial\theta^2}\log p(Z_1 \mid \theta)\Big) = \frac{n}{\sigma^2}.$$

Let another experiment **F** be the combination of experiments under aggregated data.

$$I_F(\mathbf{E}) = l \cdot I_F(X_1) = \frac{n}{k}\frac{k}{\sigma^2} = \frac{n}{\sigma^2}.$$

Therefore **F** is as informative as **E**.

Case 2:. *θ is known and σ^2 is unknown.* Without loss of generality, assume that $\theta = 0$.

$$\begin{aligned} I_F(\mathbf{E}) &= \frac{n}{2\sigma^4} \\ I_F(\mathbf{E}) &= \frac{n}{k} I_F(X_1) \end{aligned}$$

$$= \frac{n}{k}\frac{(2-k)}{2k\sigma^4}$$
$$= \frac{2n-nk}{2k^2\sigma^4}.$$

$$\frac{n}{2\sigma^4} \geq \frac{2n-nk}{2k^2\sigma^4} \Leftrightarrow (k-1)(k+2) \geq 0.$$

Therefore $I_F(\mathbf{E}) \geq I_F(\mathbf{E})$ if and only if $k \geq 1$. It means that aggregated data is always less informative than disaggregated data. The relative loss of information under aggregation with size k is

$$L(\mathbf{E},\mathbf{E}) = \frac{I_F(\mathbf{E}) - I_F(\mathbf{E})}{I_F(\mathbf{E})} = 1 - \frac{2-k}{k^2}.$$

$L(\mathbf{E},\mathbf{E})$ depends only on the size of aggregation but not on the size of total observation. Take $f(k) = 1 - \frac{2-k}{k^2}$. Then $\frac{d}{dk}f(k) = \frac{4-k}{k^3}$. Therefore $f(k)$ has one maximal value at $k = 4$ and $lim_{n\to\infty}f(k) = 1$.

Table 4.1. The Relative Loss of Information under Grouping in Normal Experiments

k	1	2	3	4	5	6	10	1000	∞
$L(E,F)$	0	1	1.111	1.125	1.12	1.111	1.08	1.0098	1

4.2 Truncated Experiments.

4.2.1 Normal Experiments.

Suppose that a random variable X has a normal distribution $N(\theta, 1)$.

Case 1:. $S = \{x \mid a < x < b\}$

$$p(y \mid \theta) = \begin{cases} \frac{\phi(y-\theta)}{\Phi(b-\theta)-\Phi(a-\theta)}, & \text{if } y \in S \\ 0 & \text{otherwise} \end{cases},$$

where $\Phi(x)$ and $\phi(x)$ is the standard normal distribution and density.

$$I_F(\mathbf{E}) \quad = \quad -E_Y\Big(\frac{d^2}{d\theta^2}\log p(Y \mid \theta)\Big)$$

$$
\begin{aligned}
&= 1 + \frac{d^2}{d\theta^2}\log(\Phi(b-\theta) - \Phi(a-\theta)) \\
&= 1 + \frac{(a-\theta)\phi(a-\theta) - (b-\theta)\phi(b-\theta)}{\Phi(b-\theta) - \Phi(a-\theta)} \\
&\quad - \Big(\frac{\phi(b-\theta) - \phi(a-\theta)}{\Phi(b-\theta) - \Phi(a-\theta)}\Big)^2. \\
&\leq 1 \\
&= I_F(\mathbf{E}).
\end{aligned}
$$

After the numerical computation, we can calculate the loss of information. Table 4.2 shows the loss of information from several values of a and b using the expected gain in information. It can be to indicate the choice of truncating point. If we perform the experiments which select the outcomes belonging to the interval $(\mu - \sigma, \mu + \sigma), (\mu - 2\sigma, \mu + 2\sigma)$ and $(\mu - 3\sigma, \mu + 3\sigma)$, the quantile of the information loss are 0.623, 0.194 and 0.018 respectively.

Case 2:. $S = \{x \mid x < a \quad or \quad x > b\}$. In this case, Bayarri and DeGroot (1987) proved that $\mathbf{E}$ and $\mathbf{F}$ are not comparable.

Table 4.2. The Loss of Information under Truncation in Normal Experiments, $\theta \sim N(0,1)$, $X \sim N(\theta,1)$, $a < X < b$.

$b\backslash a$	$-\infty$	-4	-3	-2	-1	0	1	2	3	4	∞
$-\infty$	...	...	...	...	...	...	...	...	...	...	...
-4	100.0	...	...	...	...	...	...	...	...	...	...
-3	62.2	82.5	...	...	...	...	...	...	...	...	...
-2	53.6	57.8	66.6	...	...	...	...	...	...	...	...
-1	73.1	75.1	80.3	82.6	...	...	...	...	...	...	...
0	51.2	52.6	56.1	65.2	83.4	...	...	...	...	...	...
1	29.3	30.4	33.5	42.5	62.3	70.2	...	...	...	...	...
2	6.8	7.8	10.6	19.4	37.7	49.5	63.4	...	...	...	...
3	0.0	0.0	1.8	10.4	28.4	39.8	56.2	43.0	...	...	...
4	0.0	0.0	0.0	8.0	25.8	37.0	52.0	34.1	77.3	...	...
∞	0.0	0.0	0.0	7.3	25.1	36.2	50.9	32.0	68.2	100.0	...

4.2.2 Exponential Experiments.

Consider the comparison between an untruncated experiment **E** and a truncated experiment **F**, using the expected gain in information. Let a random varible X have an exponential distribution with a parameter $\theta(>0)$, that is, the probability density function of X is $p(x \mid \theta) = \theta e^{-\theta x}$. And Y is a truncated random variable which has a probability density function given by

$$p(y \mid \theta) = \begin{cases} \frac{\theta e^{-\theta y}}{P(X \in S \mid \theta)}, & \text{if } y \in S = (a, \infty) \\ 0, & \text{otherwise} \end{cases} .$$

A prior distribution of θ is a gamma with known parameters s and m. The probability density function of θ is

$$p(\theta) = \frac{m^s}{\Gamma(s)} \theta^{s-1} e^{-m\theta}.$$

Theorem 4.1. In the exponential distribution with a gamma prior, the amount of information provided by the left-side truncated experiment is the same as that of the untruncated experiment.

Proof. (1) Amount of information provided by a prior distribution.

$$\begin{aligned} I(\theta) &= \int_0^\infty p(\theta) \log p(\theta) d\theta \\ &= s \log m - \log \Gamma(s) - m \int_0^\infty \theta p(\theta) d\theta + (s-1) \int_0^\infty p(\theta) \log \theta d\theta \\ &= s \log m - \log \Gamma(s) - s + (s-1) E_\theta(\log \theta). \end{aligned}$$

Consider the following equation.

$$\begin{aligned} 0 &= \frac{d}{ds} \int_0^\infty \frac{m^s}{\Gamma(s)} \theta^{s-1} e^{-m\theta} d\theta \\ &= \int_0^\infty \frac{d}{ds} \Big(\frac{m^s}{\Gamma(s)} \theta^{s-1} e^{-m\theta} \Big) d\theta \\ &= \int_0^\infty \frac{m^s \log m \theta^{s-1} e^{-m\theta}}{\Gamma(s)} d\theta + \int_0^\infty \frac{m^s \theta^{s-1} e^{-m\theta} \log \theta}{\Gamma(s)} d\theta \end{aligned}$$

$$
\begin{aligned}
& - \int_0^\infty \frac{m^s \theta^{s-1} e^{-m\theta} d\Gamma(s)/ds}{(\Gamma(s))^2} d\theta \\
&= \log m + E_\theta(\log\theta) - \frac{d}{ds}\log\Gamma(s)
\end{aligned}
$$

$$
E_\theta(\log\theta) = \frac{d}{ds}\log\Gamma(s) - \log m.
$$

Therefore

$$
I(\theta) = (s-1)\frac{d}{ds}\log\Gamma(s) - \log\Gamma(s) - s + \log m.
$$

(2) Expected gain in information provided by an untruncated experiment. To derive the posterior information, find the joint density of X and θ, the marginal density of X and the posterior density of θ, given $X = x$.

$$
\begin{aligned}
p(x,\theta) &= p(x \mid \theta)p(\theta) \\
&= \theta e^{-\theta x}\frac{m^s}{\Gamma(s)}\theta^{s-1}e^{-m\theta} \\
&= \theta^s e^{-\theta(x+m)}\frac{m^s}{\Gamma(s)} \\
p(x) &= \int_0^\infty p(x,\theta)d\theta \\
&= \frac{m^s}{\Gamma(s)}\frac{\Gamma(s+1)}{(x+m)^{s+1}}\int_0^\infty \frac{(x+m)^{s+1}}{\Gamma(s+1)}\theta^s e^{-\theta(x+m)}d\theta \\
&= \frac{sm^s}{(x+m)^{s+1}} \\
p(\theta \mid x) &= \frac{p(x,\theta)}{p(x)} \\
&= \theta^s e^{-\theta(x+m)}\frac{m^s}{\Gamma(s)}\frac{(x+m)^{s+1}}{sm^s} \sim \Gamma(s+1, x+m)
\end{aligned}
$$

$$
\int_0^\infty p(\theta \mid x)\log p(\theta \mid x)d\theta
$$

$$
\begin{aligned}
&= \int_0^\infty \Big\{(s+1)\log(x+m) - \log\Gamma(s+1) + s\log\theta - \theta(x+m)\Big\} p(\theta \mid x) d\theta \\
&= s\frac{d}{ds}\log\Gamma(s+1) - \log\Gamma(s+1) - (s+1) + \log(x+m)
\end{aligned}
$$

$$
\begin{aligned}
\int_0^\infty \int_0^\infty & p(\theta \mid x)\log p(\theta \mid x) d\theta p(x) dx \\
&= s\frac{d}{ds}\log\Gamma(s+1) - \log\Gamma(s+1) - (s+1) \\
&\quad + \int_0^\infty \frac{sm^s}{(x+m)^{s+1}}\log(x+m)dx \\
&= s\frac{d}{ds}\log\Gamma(s+1) - \log\Gamma(s+1) - (s+1) + \log m + \frac{1}{s}
\end{aligned}
$$

$$
\begin{aligned}
I_{EG}(E) &= \int_0^\infty \int_0^\infty p(\theta \mid x)\log p(\theta \mid x) d\theta dx \\
&\quad - \int_0^\infty p(\theta)\log p(\theta) d\theta \\
&= h(s+1) - h(s) + \frac{1}{s},
\end{aligned}
$$

where $h(s) = (s-1)\frac{d}{ds}\log\Gamma(s) - \log\Gamma(s) - s$.

(3) Expected gain in information provided by a truncated experiment. Similarly to (2), we can derive $I_L(F)$.

$$
\begin{aligned}
p(y,\theta) &= \theta^s e^{-\theta(y+m-a)}\frac{m^s}{\Gamma(s)}, \quad y > a \\
p(y) &= \int_0^\infty p(y,\theta)d\theta \\
&= \frac{sm^s}{(y+m-a)^{s+1}}, \quad y > a \\
p(\theta \mid y) &= \frac{p(y,\theta)}{p(y)}
\end{aligned}
$$

$$= \frac{(y+m-a)^{s+1}}{\Gamma(s+1)}\theta^{s}e^{-\theta(y+m-a)} \sim \Gamma(s+1, y+m-a)$$

$$\begin{aligned}
&\int_0^\infty p(\theta \mid y)\log p(\theta \mid y)d\theta \\
&= s\frac{d}{ds}\log\Gamma s+1 - \log\Gamma(s+1) - (s+1) + \log(y+m-a) \\
&= h(s+1) + \log(y+m-a) \\
&\int_a^\infty\int_0^\infty p(\theta \mid y)\log p(\theta \mid y)d\theta p(y)dy \\
&= h(s+1) + sm^s\int_a^\infty (y+m-a)^{-s-1}\log(y+m-a)dy \\
&= h(s+1) + sm^s\int_m^\infty t^{-s-1}\log t dt \\
&= h(s+1) + \log m + \frac{1}{s} \\
I_{EG}(F) &= h(s+1) - h(s) + \frac{1}{s} \\
&= I_{EG}(E).
\end{aligned}$$

The theorem is proved.

Let $\mathbf{E}_a = \{Y, \Theta, p(y \mid \theta)\}$ be a right-side truncated experiment where the probability density of Y depending on θ is given by

$$p(y \mid \theta) = \frac{\theta e^{-\theta y}}{1-e^{-a\theta}}, \quad 0 < y < a.$$

Theorem 4.2. $I_F(\mathbf{E}) \geq I_F(\mathbf{E}_a)$ for any positive a. And $I_F(\mathbf{E}_a)$ is a monotonically increasing function of a.

Proof.

$$I_F(\mathbf{E}_a) = I_F(\mathbf{E}) + \frac{d^2}{d\theta^2}\log(1 - e^{-a\theta})$$

$$\frac{d^2}{d\theta^2}\log(1 - e^{-a\theta}) = \frac{-a^2 e^{-a\theta}}{(1 - e^{-a\theta})^2} < 0 \quad \text{for any } a > 0.$$

Therefore the amount of information provided by an untruncated experiment $\mathbf{E}$ is large than that provided by a truncated experiment $\mathbf{E}_a$ for any positive a.

Take $f(a) = \frac{a^2 e^{-a\theta}}{(1-e^{-a\theta})^2}$. In order to prove the second part of theorem, it is sufficient to show that $f(a)$ is monotonically increasing in a.

$$\frac{d}{da} f(a) \leq 0 \Longleftrightarrow (2 - a\theta) - e^{-a\theta}(2 + a\theta) \leq 0.$$

Take $g(x) = 2 - x - e^{-x}(2 + x), \quad x > 0$. Since $g(0) = 0$ and $g(\infty) = -\infty$, it is sufficient to show that $\frac{d}{dx}g(x)$ is negative for $x > 0$. Since $\log(1 + x) \leq x$, $\frac{d}{dx}g(x) = -1 + e^{-x}(1 + x) \leq 0$. Therefore $I_{\mathbf{E}}(\mathbf{E}_a)$ is monotonically increasing function of a.

After numerical computation, we make Table 4.3 which represents the relative loss of informaton under truncation.

Table 4.3. The Relative Loss of Information under Truncation of Exponential Experiments $\theta \sim \Gamma(1, m), m = 1, 2, 3$, $X \sim$ Truncated Exponential$(\theta), 0 < x < a$.

m=1		m=2		m=3	
a	L	a	L	a	L
1	94.3825	0.5	94.3826	0.333	94.3826
2	86.1591	1.5	78.0472	1.333	67.8836
3	78.0472	2.5	62.9505	2.333	54.3673
4	67.8836	3.5	54.3672	3.333	43.6280
5	62.9505	4.5	47.4183	4.333	36.3320
11	41.1248	10.5	25.2022	10.333	19.4521
21	25.2022	20.5	16.1877	20.333	11.2907
41	16.1876	40.5	9.7686	40.333	7.6424
71	10.5272	70.5	6.9919	70.333	5.7834
101	8.6669	100.5	5.9514	100.333	5.2279
131	7.3076	130.5	5.3808	130.333	5.1496
151	6.6926	150.5	5.2279	150.333	5.1479
201	5.9514	200.5	5.1492	200.333	5.1467

5. CONCLUDING REMARKS

In this paper we considered the loss of information provided by incomplete experiments with aggregation and truncation. Using the expected gain in information based on the Shannon's entropy, we have shown that an incomplete experiment is less informative than the complete experiment. We used a conjugate prior distribution because of easy calculation. Some tables are made in order to indicate an appropriate choice of points of truncation in these experiments. We have proposed some results of binomial experiments and censored experiments using the Blackwell's sufficiency.

We want to find the appropriate measures of information which can be applied on specific experiments with physical meanings. To make the useful criteria for another types of experiments and prior distributions are remained for future works.

REFERENCES

(1) Bayarri, M.J. and DeGroot, M.H. (1987). Information in selection models. *Probability and Bayesian Statistics*, New York, Plenum Press.

(2) Blackwell, D. (1951). Comparision of experiments. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, **93–102**.

(3) Blackwell, D. (1953). Equivalent comparison of experiments. *Annals of Mathematical Statistics*, **24**, **265–272**.

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

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

(6) Goel, P.K. and Degroot, M.H. (1979). Comparison of experiments and information measures. *Annals of Statistics*, **7**, **1066–1077**.

(7) Hansen, O.H. and Torgersen, E.N. (1974). Comparison of linear normal experiments. *Annal of Statistics*, **2**, **367–373**.

(8) Heyer, H. (1982). *Theory of statistical experiments*, New York. Springer-Verlag.

(9) Lehmann, E.L. (1986). Comparing location experiments. *Technical report* No. **75**, University of California, Berkeley.

(10) Lindley, D.V. (1956). On a measure of the information provied by an experiments. *Annals of Mathematical Statistics*, **27**, **986-1005**.

(11) Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, **379–423**, **623–656**.

(12) Turrero, A. (1989). On the relative efficiency of grouped and censored survival data. *Biometrika*, **76**, **125–131**.