CSSE-12 February 1, 2001

Parameter Estimation for Grouped and Truncated Data or Truncated Data

Yoshio Komori and Hideo Hirose

Department of Control Engineering and Science
Kyushu Institute of Technology

680-4 Kawazu
Iizuka, 820-8502, Japan
{*komori* & *hirose*}*@ces.kyutech.ac.jp*

Abstract

In the present article we deal with parameter estimation about truncated data, which are of a variety of truncated time—a_1 truncated data are of a truncated time b_1, a_2 truncated data are of a truncated time b_2 and so on. We show some conditions to get the maximum likelihood estimations in the two cases, in one of which the truncated data are given as data values and in the other of which the truncated data are given as grouped data (that is, each expresses the number of the data fall into a subinterval). Here, it is supposed that all the data are subject to an exponential distribution.

1 Introduction

We are concerned with a problem to predict what situation will happen respecting the failure of industrial products after they were shipped. As one of such practical examples, we suppose that inferior goods were shipped being mixed in the group of normal ones and some of those have been returned by now. Then, our purpose is to infer how many inferior goods leave on the market without being returned in order to decide to recall all goods or not. The following is the setting for our problems: Products were shipped at intervals of τ and s times in total, and the investigation of the number of the returned products has been continued for T after the 1st shipment.

In the present article we deal with two cases, one of which is the grouped and truncated data case and the other of which is the truncated data case, and state a condition to obtain a maximum likelihood estimator (MLE) in each case, provided that failure time obeys an exponential distribution.

2 Grouped and Truncated Data Case

Divide the shipping interval τ into g non-overlapping subintervals of length t/g, and denote by $r_j^{(i)}$ the number of the products broken in the subinterval between $(j-1)\tau/g$ and $j\tau/g$ on the ith shipment. Figure 1 indicates the aspect of the occurrence of failure during the observing period.

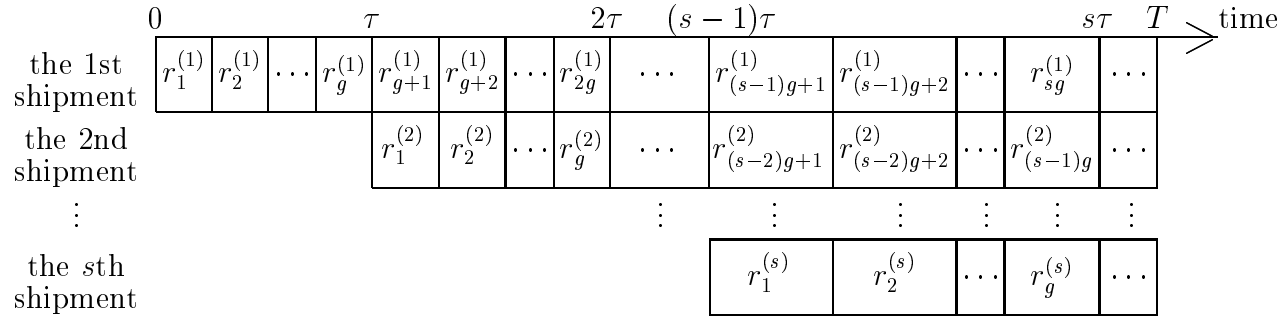

Figure 1: The aspect of the occurrence of failure on each shipment

Next, suppose that T may be expressed as $\eta\tau/g$ with some positive integer η ($\geq (s-1)g+1$), then we obtain Fig. 2 by rearranging the data on Fig. 1 in order of the length of passed time since products were shipped.

We assume that time t, which has passed by a product led to failure since its shipment, obeys a distribution whose density function $f(t;\boldsymbol{\theta})$ depending on $\boldsymbol{\theta}$. Then, the logarithmic likelihood function about data on Fig. 2 is the following:

$$\ln L_{tg}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \sum_{i=1}^{s} \sum_{j=1}^{\eta-(i-1)g} r_j^{(i)} \ln \left\{ \frac{\displaystyle\int_{(j-1)\tau/g}^{j\tau/g} f(t;\boldsymbol{\theta})\mathrm{d}t}{\displaystyle\int_{0}^{T-(i-1)\tau} f(t;\boldsymbol{\theta})\mathrm{d}t} \right\}. \tag{1}$$

Let us consider a case, where $f(t;\boldsymbol{\theta})$ in (1) is the density function of an exponential

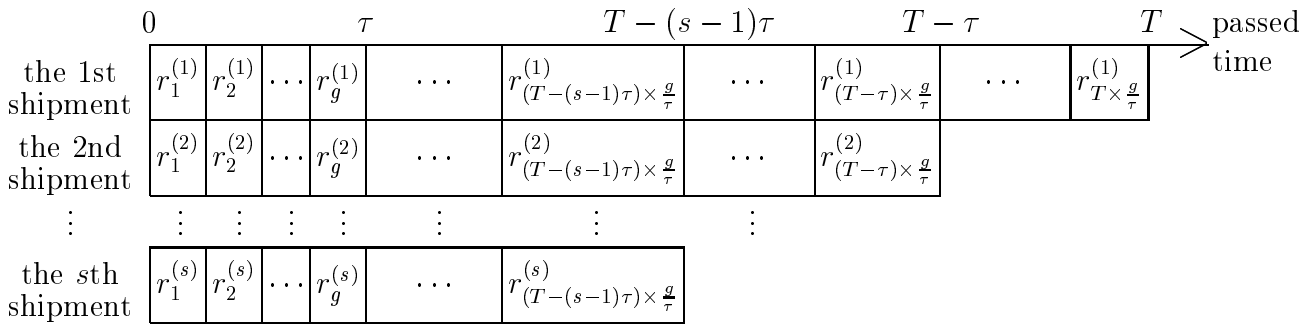

Figure 2: The occurrence of failure versus passed time

distribution, that is, ce^{-ct}. Here, c (>0) is a parameter. Setting

$$\Delta t \stackrel{\text{def}}{=} \tau/g, \quad t_j \stackrel{\text{def}}{=} j\Delta t, \quad \tau_i \stackrel{\text{def}}{=} T-(i-1)\tau,$$

we obtain

$$\ln L_{tg}(c) = \sum_{i=1}^{s} \sum_{j=1}^{\eta-(i-1)g} r_j^{(i)} \left\{ -ct_{j-1} + \ln\left(1-\mathrm{e}^{-c\Delta t}\right) - \ln\left(1-\mathrm{e}^{-c\tau_i}\right) \right\}. \tag{2}$$

In connection with this, the following lemma holds.

Lemma 2.1 *We set*

$$n_i \stackrel{\text{def}}{=} \sum_{j=1}^{\eta-(i-1)g} r_j^{(i)}, \quad N \stackrel{\text{def}}{=} \sum_{i=1}^{s} n_i, \quad \tilde{t}_a \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{s} \sum_{j=1}^{\eta-(i-1)g} r_j^{(i)} \left(\frac{t_{j-1}+t_j}{2} \right).$$

If

$$0 < \tilde{t}_a < \frac{1}{2N} \sum_{i=1}^{s} n_i \tau_i,$$

there exists a solution of $\dfrac{\partial \ln L_{tg}}{\partial c}(c) = 0$ *and it is MLE. If not, MLE does not exist.* ■

Proof. From (2)

$$\frac{\partial \ln L_{tg}}{\partial c}(c) = \sum_{i=1}^{s} \sum_{j=1}^{\eta-(i-1)g} r_j^{(i)} \left\{ -t_{j-1} + \frac{\Delta t}{\mathrm{e}^{c\Delta t}-1} - \frac{\tau_i}{\mathrm{e}^{c\tau_i}-1} \right\}.$$

According to Lemma 2.2, the right hand above is a strictly decreasing function of c (>0) if $\tau_i > \Delta t$. Since

$$\lim_{c\to+0} \left\{ \frac{\Delta t}{\mathrm{e}^{c\Delta t}-1} - \frac{\tau_i}{\mathrm{e}^{c\tau_i}-1} \right\} = -\frac{1}{2}\Delta t + \frac{1}{2}\tau_i,$$

$$\lim_{c\to+0} \frac{\partial \ln L_{tg}}{\partial c}(c) = N \left[\frac{1}{2N} \sum_{i=1}^{s} n_i \tau_i - \tilde{t}_a \right].$$

On the other hand

$$\lim_{c\to+\infty} \frac{\partial \ln L_{tg}}{\partial c}(c) = -\sum_{i=1}^{s} \sum_{j=1}^{\eta-(i-1)g} r_j^{(i)} t_{j-1} < 0.$$

In addition, $\dfrac{\partial \ln L_{tg}}{\partial c}(c)$ is continuous. Consequently, there exists a solution $c = c_{tg0}$ (> 0) satisfying $\dfrac{\partial \ln L_{tg}}{\partial c}(c) = 0$ if $0 < \tilde{t}_a < \dfrac{1}{2N}\sum_{i=1}^{s} n_i\tau_i$, and $\ln L_{tg}(c) \uparrow \sup_{c>0}\{\ln L_{tg}(c)\}$ as $c \downarrow 0$ if $\tilde{t}_a \geq \dfrac{1}{2N}\sum_{i=1}^{s} n_i\tau_i$. □

Lemma 2.2 *The function*

$$g(c) \stackrel{\text{def}}{=} \frac{\Delta t}{\mathrm{e}^{c\Delta t} - 1} - \frac{s}{\mathrm{e}^{cs} - 1} \qquad (0 < \Delta t < s)$$

strictly increases in $(0, \infty)$. ■

Proof. Because that

$$g'(c) = \frac{1}{c^2}\left\{-\frac{(c\Delta t)^2\mathrm{e}^{c\Delta t}}{(\mathrm{e}^{c\Delta t} - 1)^2} + \frac{(cs)^2\mathrm{e}^{cs}}{(\mathrm{e}^{cs} - 1)^2}\right\},$$

it suffices for $g'(c) < 0$ $(c > 0)$ holding that $h(x) \stackrel{\text{def}}{=} \dfrac{x^2\mathrm{e}^x}{(\mathrm{e}^x - 1)^2}$ is strictly increasing in $(0, \infty)$. Since

$$h'(x) = \frac{x\mathrm{e}^x}{(\mathrm{e}^x - 1)^3}\left\{(2 - x)\mathrm{e}^x - (2 + x)\right\},$$

we set $u(x) \stackrel{\text{def}}{=} (2 - x)\mathrm{e}^x - (2 + x)$ and investigate this. By differentiating $u(x)$ up to twice we get

$$u'(x) = (1 - x)\mathrm{e}^x - 1, \qquad u''(x) = -x\mathrm{e}^x.$$

From these equations, $u''(x) < 0$ for $x > 0$ and $u'(0) = 0$, thus $u'(x) < 0$. In analogy, $u'(x) < 0$ for $x > 0$ and $u(0) = 0$, thus $u(x) < 0$. Consequently, $h(x)$ is strictly decreasing in $(0, \infty)$. Therefore, $g(c)$ is a strictly decreasing function. □

3 Truncated Data Case

In Sec. 2 we discussed the case in which given data were only the number $r_j^{(i)}$ of products broken in the subinterval between $(j-1)\tau/g$ and $j\tau/g$ for each j on the ith shipment. In this section we devote ourselves to the case in which failure time is given as data.

Denote by $t_k^{(i,j)}$ $(k = 1, 2, \ldots, r_j^{(i)})$ failure time of $r_j^{(i)}$ products broken in the subinterval between $(j-1)\tau/g$ and $j\tau/g$ on the ith shipment, and set $\boldsymbol{t}_j^{(i)} \stackrel{\text{def}}{=} (t_1^{(i,j)}, t_2^{(i,j)}, \ldots, t_{r_j^{(i)}}^{(i,j)})$. As Fig. 2, Fig. 3 indicates the aspect of the occurrence of failure arranged in order of the length of passed time after each shipping.

The logarithmic likelihood function about data on Fig. 3 is defined by

$$\ln L_t(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \sum_{i=1}^{s}\sum_{j=1}^{\eta-(i-1)g}\sum_{k=1}^{r_j^{(i)}} \ln\left\{\frac{f(t_k^{(i,j)};\boldsymbol{\theta})}{\displaystyle\int_0^{T-(i-1)\tau} f(t;\boldsymbol{\theta})\mathrm{d}t}\right\}. \tag{3}$$

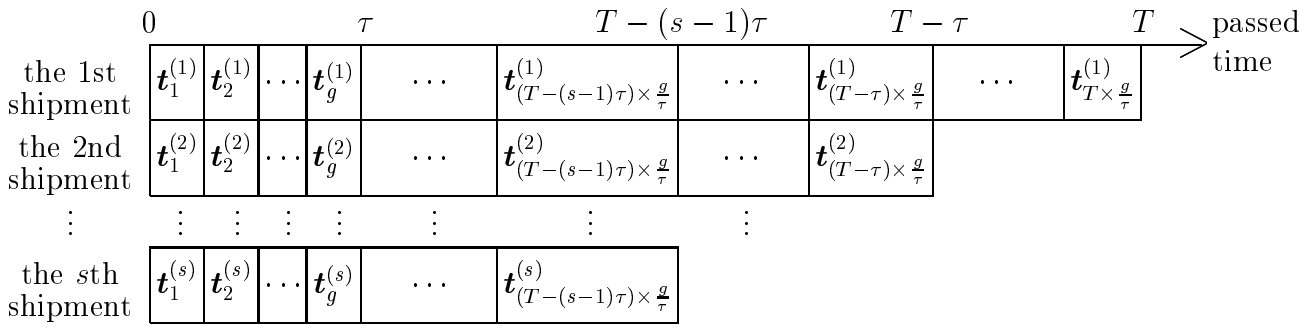

Figure 3: The occurrence of failure versus passed time

Let us consider of a case, where $f(t;\boldsymbol{\theta})$ in (3) is the density function of an exponential distribution. Setting

$$\bar{t}^{(i)} \stackrel{\text{def}}{=} \frac{1}{n_i} \sum_{j=1}^{\eta-(i-1)g} \sum_{k=1}^{r_j^{(i)}} t_k^{(i,j)},$$

we obtain

$$\ln L_t(c) = \sum_{i=1}^{s} n_i \left\{ \ln c - c\bar{t}^{(i)} - \ln\left(1 - \mathrm{e}^{-c\tau_i}\right) \right\}. \tag{4}$$

In connection with this, the following lemma holds.

Lemma 3.1 *We set*

$$\bar{t}_a \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{s} n_i \bar{t}^{(i)}.$$

If

$$0 < \bar{t}_a < \frac{1}{2N} \sum_{i=1}^{s} n_i \tau_i,$$

there exists a solution of $\dfrac{\partial \ln L_t}{\partial c}(c) = 0$ *and it is MLE. If not, MLE does not exist.* ■

Proof. From (4)

$$\frac{\partial \ln L_t}{\partial c}(c) = N \left[\sum_{i=1}^{s} \frac{n_i}{N} \left\{ c^{-1} - \frac{\tau_i}{\mathrm{e}^{c\Delta t} - 1} \right\} - \bar{t}_a \right].$$

According to Lemma 3.2, the function in { } on the right hand above is a strictly decreasing for c (> 0).

Since

$$\lim_{c \to +0} \left\{ c^{-1} - \frac{\tau_i}{\mathrm{e}^{c\tau_i} - 1} \right\} = \frac{1}{2}\tau_i,$$

$$\lim_{c \to +0} \frac{\partial \ln L_t}{\partial c}(c) = N \left[\frac{1}{2N} \sum_{i=1}^{s} n_i \tau_i - \bar{t}_a \right].$$

On the other hand

$$\lim_{c \to +\infty} \frac{\partial \ln L_t}{\partial c}(c) = -N\bar{t}_a < 0.$$

In addition, $\dfrac{\partial \ln L_t}{\partial c}(c)$ is continuous. Consequently, there exists a solution $c = c_{t0}$ (> 0) satisfying $\dfrac{\partial \ln L_t}{\partial c}(c) = 0$ if $0 < \bar{t}_a < \dfrac{1}{2N}\sum_{i=1}^{s} n_i\tau_i$, and $\ln L_t(c) \uparrow \sup_{c>0}\{\ln L_t(c)\}$ as $c \downarrow 0$ if $\bar{t}_a \geq \dfrac{1}{2N}\sum_{i=1}^{s} n_i\tau_i$. □

Lemma 3.2 *The function*

$$v(c) \stackrel{\text{def}}{=} \frac{1}{c} - \frac{s}{\mathrm{e}^{cs} - 1} \qquad (s \neq 0)$$

strictly increases in $(0, \infty)$. ■

Proof. By differentiating $v(c)$ we obtain

$$v'(c) = \frac{\mathrm{e}^{cs}\left\{2 + (cs)^2 - (\mathrm{e}^{cs} + \mathrm{e}^{-cs})\right\}}{c^2\left(\mathrm{e}^{cs} - 1\right)^2}.$$

Using Maclaurin expansion of e^x, we can easily show that $\{\ \}$ part on the right hand is negative. Thus, $v(c)$ is a strictly decreasing function in $(0, \infty)$. □

4 Summary

In the present paper we stated the conditions under which MLEs may be given for grouped and truncated data or truncated data when shipment is performed in several times.

Deemer [1] gave a similar condition in the case where products are shipped only one time: Let T_r be the truncated time, $\bar{t}$ the average failure time of broken products. Then, MLE exists if $0 < \bar{t} < \dfrac{1}{2}T_r$, and MLE does not exist if $\bar{t} \geq \dfrac{1}{2}T_r$.

Comparing this with the results in Sec. 2 or Sec. 3, we explain about them. The expression $\dfrac{1}{N}\sum_{i=1}^{s} n_i\tau_i$ is the average truncated time because that n_i and τ_i are the total number of broken products and the truncated time on the ith shipment, respectively. If $\dfrac{t_{j-1} + t_j}{2}$ is chosen as the representative value of failure time in the subinterval, $\tilde{t}_a$ means an approximate value to the average failure time. On the other hand $\bar{t}_a$ is the average failure time on all the shipments since $\bar{t}^{(i)}$ is the average failure time on the ith shipment. Summarizing the things above, we can say as follows: If we replace the truncated time in Deember's result with the average one , we obtain the result in Sec. 3. Besides the truncated time, if we replace the average failure time with the approximate value, we obtain the result in Sec. 2.

References

[1] W.L. Deemer, Jr. and D.F. Votaw, Jr. Estimation of parameters of truncated or censored exponential distributions. *Ann. Math. Statist.*, **26**:498–504, 1955.