Journal of the Korean
Data & Information Science Society
2000, Vol. 11, No.1, pp. 103 ~ 110

Second Order Approximations to the Stopping Time with Fixed Proportional Accuracy [1]

Kiheon Choi [2]

Abstract

Suppose that there is a population of hidden objects of which the total number N is unknown. From such data, we derive second-order approximations to the stopping time with fixed proportional accuracy.

Key Words and Phrases: stopping time, second order approximation.

1. Introduction

Consider a problem which require us to find, observe, or catch some of or all of a group of hidden objects as prey. Examples of such prey are fish in a lake, potential voters in a voter registration drives, donors to charitable organizations, disintegrating atoms in a radioative source, disease carriers, or relics at the site of an archaeological dig. This problem has been considered by several authors, including Starr(1974), Vardi(1980), Dalal and Mallows(1988).

Thus, consider an area containing N prey. Imagine the prey are labelled $1, \cdots, N$; let T_i denote the time at which we would capture the prey labelled i if we are to search indefinitely. We suppose throughout that $T_1, \cdots, T_N$ are independent and identically distributed with a continuous distribution function F for which $F(0) = 0$. The distribution function F may depend on an unknown parameter θ, or not. Let $t_1 \leq \cdots \leq t_N$ denote the order statistics of $T_1, \cdots, T_N$. If the search is continued for t units of times, then the available data consists of the number of objects found and the times at which they were found; in symbols,

$$K_t = \#\{k \leq N : T_k \leq t\}.$$

[1]This work was supported by the Institute of Natural Science Fund, Duksung Women's University, 1999

[2]Associate professor, Department of Statistics, Duksung Women's University, Seoul, Korea, 132-714

Let $\hat{N}_t$ denote an estimator of N. Then, stopping time τ_h are sought for which

$$lim_{N\to\infty} P_N\{|\hat{N}_{\tau_h} - N| \leq h\hat{N}_{\tau_h}\},$$

is computed for fixed $h > 0$ for all values of unknown parameter θ. We derive a second order term to the confidence interval for stopping time with fixed proportional accuracy.

2. Second order term for stopping time

F is assumed to be known, continuous distribution function that is strictly increasing on the interval $(0, b_F)$, where $b_F = sup\{t : F(t) < 1\} \leq \infty$. Then the maximum likelihood estimator of F after t time units of observation is (an integer adjacent to)

$$\hat{N}_t = \frac{K_t}{F(t)},$$

for $0 < t < b_F$. Since K_t has binomial distribution, the mean and variance of $\hat{N}_t$ are $E_N[\hat{N}_t] = N$ and $D_N^2[\hat{N}_t] = N\sigma^2(t)$, where

$$\sigma^2(t) = \frac{1}{F(t)} - 1,$$

and $\hat{N}_t$ is asymptotically normal as $N \to \infty$ for fixed $t > 0$; that is

$$\frac{\hat{N}_t - N}{\sqrt{N\sigma^2(t)}} \Rightarrow Z, \tag{1}$$

where $\Rightarrow$ denotes convergence in distribution and Z denotes the standard normal distribution Φ. In fact, (1) holds for sequences $t = t_N > 0$ for which $N\sigma^2(t_N) \to \infty$ as $N \to \infty$. Using asymptotic normality to set an approximate to confidence interval and imposing the condition that the half width of the interval be at most $h\hat{N}_t$, as in Chow and Robbins(1965), suggests sampling until $\hat{N}_t \geq z^2\sigma^2(t)/h^2$, where $\Phi(z) = (1+\gamma)/2$ (and γ is desired confidence coefficient).

Now $F = F_\theta$ is assumed to be an exponential distribution with unknown failure rate θ, $F_\theta = 1 - e^{-\theta x}$ for $x \geq 0$.

Let h be a fixed length, we have

$$P_{N,\theta}\left(\left|\frac{\hat{N}_t}{N} - 1\right| \leq h\right) \approx 2\Phi\left[\frac{h\sqrt{N}}{\sigma(t\theta)}\right] - 1$$

as $N \to \infty$, and we need

$$\frac{h\sqrt{N}}{\sigma(t\theta)} \geq 2.$$

This suggests that we continue sampling until

$$\frac{\hat{N}_t}{\sigma^2(t\hat{\theta}_t)} \geq \frac{4}{h^2}. \tag{2}$$

Consider small t, say

$$t \downarrow 0.$$

Then

$$1 - e^{-t\theta} \sim t\theta,$$
$$\sigma^2(t\theta) \sim \frac{12}{t^3\theta^3},$$

and

$$\frac{\hat{N}_t}{\sigma^2(t\hat{\theta}_t)} \approx \frac{(t\hat{\theta}_t)^3\hat{N}_t}{12} \approx \frac{(t\hat{\theta}_t)^2 K_t}{12}.$$

Further, recalling that K_t has a binomial distribution, it is then easily seen that the conditional density of $X_j = t_j/t, j = 1, \cdots, k$, given that $K_t = k$, is the same as the distribution of the order statistics of a sample of size k from the density

$$f_w(x) = \frac{we^{-wx}}{1 - e^{-w}}, \quad 0 \leq x \leq 1,$$

where $w = \theta t$. Let $\mu(w)$ denote the mean of f_w. Then, by Taylor expansion,

$$\mu(w) = \frac{1}{2} - \frac{1}{12}w + O(w^2),$$

so that

$$t\theta_t \approx 6\left[1 - 2\mu(t\hat{\theta}_t)\right] = 6\left[1 - 2\frac{S_t}{tK_t}\right],$$

where $S_t = t_1 + \cdots + t_{K_t}$. So,

$$\frac{\hat{N}_t}{\sigma^2(t\hat{\theta}_t)} \approx 3K_t\left[1 - 2\frac{S_t}{tK_t}\right]^2.$$

Using these approximations and letting $x_+ = max(0, x)$, relation (2) may be rewritten: continue sampling until

$$3K_t\left[1 - 2\frac{S_t}{tK_t}\right]_+^2 \geq \frac{4}{h^2}.$$

This in turn suggests the stopping time,

$$\tau_h = \inf\left\{t > 0 : 3[tK_t - 2S_t]_+^2 \geq \frac{4t^2K_t + 1}{h^2}\right\}.$$

where the term $1/h^2$ has been added to discourage early stopping.

For the asymptotic, suppose (without essential loss of generality) that

$$\theta = 1$$

and consider small t, say

$$t = N^{-\frac{1}{3}}s,$$

where $0 < s < \infty$. Then

$$t^2 K_t \xrightarrow{p} s^3,$$

as $N \to \infty$. Let

$$K_N^0(s) = N^{-\frac{1}{3}}[K_t - N(1 - e^{-t})]$$

and

$$A_N(s) = \sqrt{3}[tK_t - 2S_t].$$

Thus,

$$N^{\frac{1}{3}}\tau_h = \inf\left\{s > 0 : A_N(s) > \frac{1}{h}\sqrt{4t^2K_t + 1}\right\}$$

There is a simple relation between K_N and K_N^0,

$$K_N(s) - K_N^0(s) = N^{\frac{2}{3}}[1 - e^{-t} - t] \to \frac{1}{2}s^2$$

uniformly on compacts in $0 \le s < \infty$.

Next let $B(s)$, $0 \le s < \infty$, be a standard Brownian motion, and let

$$A(s) = \sqrt{3}\int_0^s (s - 2u)dB(u) + \frac{\sqrt{3}}{6}s^3.$$

Observe that $A(s)$ is a Gaussian process with mean and covariance functions

$$E[A(s)] = \frac{\sqrt{3}}{6}s^3$$

and

$$\begin{aligned} Cov[A(s_1), A(s_2)] &= 3\int_0^{s_1}(s_1 - 2u)(s_2 - 2u)du \\ &= 3[s_1 s_2 u - (s_1 + s_2)u^2 + \frac{4}{3}u^3]_{u=0}^{s_1} \\ &= s_1^3 \end{aligned}$$

for $0 \le s_1 \le s_2 < \infty$. Thus $\tilde{A}(s) := A(s^{1/3})$ is Brownian motion with drift parameter and unit diffusion parameter. This may be written as the stopping time

$$N^{\frac{1}{3}}\tau_h \Longrightarrow \eta_h = inf\left\{s > 0 : A(s)_+ \ge \frac{1}{h}\sqrt{4s^3 + 1}\right\},$$

as $N \to \infty$.

Now if the procedure is modified slightly, then it is possible to make of the coverage probabilities exceed γ for all sufficiently small $h > 0$. Let W be a Brownian motion with drift θ and let us consider the stopping time

$$T_a = inf\{t > 0 : W(t) + \theta t \geq \psi_a(t)\}$$

with $T_a = \infty$ if the set is empty. We will consider continuously differentiable functions

$$\psi_a(t) = \sqrt{2(at + c)},\ c > 0.$$

Therefore

$$lim_{N\to\infty} P_N\{|\hat{N}_{T_h} - N| \leq h\hat{N}_{T_h}\} = P\left\{\left|\frac{W(T_a)}{\theta T_a} - 1\right| \leq h\right\},$$

where $a = 2/h^2$. Using the second order approximation to the density of Brownian first exit times. Jennen(1985) showed that the density $f_a(t)$ of T_a is asymptotically equivalent to the density of the first exit time over the tangent approximation with second order term

$$f_a(t) = \left[\frac{\Lambda_a(t)}{t^{3/2}} + \frac{t^{3/2}\psi_a''(t)}{2\Lambda_a(t)^2}(1 + o(1))\right] \cdot \frac{1}{\sqrt{2\pi}} exp\left(-\frac{\psi_a(t)^2}{2t}\right),$$

where $\Lambda_a(t) = \psi_a(t) - t\psi_a'(t)$ is the intercept on the vertical axis of the tangent to the curve at t.

Theorem 2.1

$$\begin{aligned} P\left\{\left|\frac{W(T_a)}{\theta T_a} - 1\right| \leq h\right\} &= \int_{a_h}^{b_h} f_a(t)dt \\ &= 2\Phi(2) - 1 + O(h^2), \end{aligned}$$

where $a_h = 2a/\theta^2 + d\sqrt{a}$, $b_h = 2a/\theta^2 - b\sqrt{a}$, $d = \frac{4\sqrt{2}}{\theta^2} - \frac{4}{\theta^2\sqrt{a}}$ *and* $b = \frac{4\sqrt{2}}{\theta^2} + \frac{4}{\theta^2\sqrt{a}}$.

Proof. Now we know

$$f_a(t) \approx \left[\frac{\Lambda_a(t)}{t^{3/2}} + \frac{t^{3/2}\psi_a''(t)}{2\Lambda_a(t)^2}\right] \cdot \frac{1}{\sqrt{2\pi}} exp\left(-\frac{(\sqrt{2(at+c)} - \theta t)^2}{2t}\right) \tag{3}$$

And

$$\frac{\Lambda_a(t)}{\sqrt{t}} \approx \sqrt{\frac{a}{2}} + \frac{3c}{\sqrt{8at}}$$

Let $s = 2a/\theta^2 + x$, then

$$\sqrt{2(as + c)} \approx \frac{2a}{\theta} + \frac{\theta}{2a}(ax + c).$$

So the integral of the first part of right hand side of (3) is

$$\begin{aligned}
\int_{2a/\theta^2-b\sqrt{a}}^{2a/\theta^2+d\sqrt{a}} & \frac{\Lambda_a(s)}{s^{3/2}}\frac{1}{\sqrt{2\pi}}e^{-a-c/s-\theta^2 s/2+\theta\sqrt{2(as+c)}}ds \\
&= e^{-a}\frac{1}{\sqrt{2\pi}}\int\left(\frac{\sqrt{\frac{a}{2}}+3c}{\sqrt{8as}}\right)\frac{1}{s}e^{-c/s-\theta^2 s/2+\theta\sqrt{2(as+c)}}ds \\
&= e^{-a}\frac{1}{\sqrt{2\pi}}\int\sqrt{\frac{a}{2}}\frac{1}{s}e^{-c/s-\theta^2 s/2+\theta\sqrt{2}(as+c)}ds \\
&\quad +e^{-a}\frac{1}{\sqrt{2\pi}}\int\frac{3c}{\sqrt{8as}}\frac{1}{s}e^{-c/s-\theta^2 s/2+\theta\sqrt{2(as+c)}}ds \\
&= (I)\ +\ (II) \qquad (say)
\end{aligned}$$

Now

$$\begin{aligned}
(I) &= e^{-a}\frac{1}{\sqrt{2\pi}}\int_{-b\sqrt{a}}^{d\sqrt{a}}\sqrt{\frac{a}{2}}\frac{1}{s}e^{-c/s-\theta^2 s/2+\theta\sqrt{2(as+c)}}ds \\
&\approx \frac{1}{\sqrt{2\pi}}\int_{-b\sqrt{a}}^{d\sqrt{a}}\frac{\theta^2}{2\sqrt{2}}\frac{1}{\sqrt{a}}\frac{1}{s}\left(1-\frac{\theta^2 x}{4a}\right)e^{-\frac{\theta^4 x^2}{16a}+\frac{\theta^4 cx}{8a^2}}dx
\end{aligned}$$

Applying the change of variables, we get

$$\begin{aligned}
(I) &\approx \frac{1}{\sqrt{2\pi}}\int_{-b}^{d}\frac{\theta^2}{2\sqrt{2}}\left(1-\frac{\theta^2 u}{4a}\right)e^{-\frac{\theta^4 u^2}{16}}du \\
&= \int_{-b}^{d}\frac{1}{\sqrt{2\pi}\left(\frac{2\sqrt{2}}{\theta^2}\right)}e^{-\frac{u^2}{2\left(\frac{2\sqrt{2}}{\theta^2}\right)^2}}du \\
&\quad -\int_{-b}^{d}\frac{1}{\sqrt{2\pi}\left(\frac{2\sqrt{2}}{\theta^2}\right)}\left(\frac{\theta^2 u}{4\sqrt{a}}\right)e^{-\frac{u^2}{2\left(\frac{2\sqrt{2}}{\theta^2}\right)^2}}du \\
&= (I_1)\ +\ (I_2) \qquad (say)
\end{aligned}$$

Then

$$\begin{aligned}
(I_1) &= \int_{-b}^{d}\frac{1}{\sqrt{2\pi}\left(\frac{2\sqrt{2}}{\theta^2}\right)}e^{-\frac{u^2}{2\left(\frac{2\sqrt{2}}{\theta^2}\right)^2}}du \\
&= P\left\{-\frac{4\sqrt{2}}{\theta^2}-\frac{4}{\theta^2\sqrt{a}} < N\left(0,\left(2\frac{\sqrt{2}}{\theta^2}\right)^2\right) < \frac{4\sqrt{2}}{\theta^2}-\frac{4}{\theta^2\sqrt{a}}\right\}
\end{aligned}$$

$$\begin{aligned} &= P\left\{-2-\sqrt{\frac{2}{a}} < Z < 2-\sqrt{\frac{2}{a}}\right\} \\ &= \Phi\left(2-\sqrt{\frac{2}{a}}\right) - \Phi\left(-2-\sqrt{\frac{2}{a}}\right) \\ &\approx \Phi(2) - \Phi(-2) - \frac{2\sqrt{2}}{\sqrt{\pi}} e^{-2} \frac{1}{a}. \end{aligned}$$

Similarly, we get

$$\begin{aligned} (I_2) &= -\int_{-b}^{d} \frac{1}{\sqrt{2\pi}\left(\frac{2\sqrt{2}}{\theta^2}\right)} \left(\frac{\theta^2 u}{4\sqrt{a}}\right) e^{-\frac{u^2}{2\left(\frac{2\sqrt{2}}{\theta^2}\right)^2}} du \\ &\approx \frac{2\sqrt{2}}{\sqrt{\pi}} e^{-2} \frac{1}{a}. \end{aligned}$$

So,

$$\begin{aligned} (I) &= (I_1) + (I_2) \\ &= \Phi(2) - \Phi(-2) + O\left(\frac{1}{a^2}\right). \end{aligned}$$

Also,

$$\begin{aligned} (II) &= e^{-a} \frac{1}{\sqrt{2\pi}} \int \frac{3c}{\sqrt{8as}} \frac{1}{s} e^{-c/s-\theta^2 s/2+\theta\sqrt{2(as+c)}} ds \\ &= O\left(\frac{1}{a^2}\right). \end{aligned}$$

The second part of right hand side of (3) is

$$\begin{aligned} &\int_{2a/\theta^2-b\sqrt{a}}^{2a/\theta^2+d\sqrt{a}} \frac{\psi_a''(s)s^3/2}{2\Lambda_a(s)^2} \frac{1}{\sqrt{2\pi}} e^{-a-c/s-\theta^2 s/2+\theta\sqrt{2(as+c)}} ds \\ &\approx -\frac{1}{a}\left[\Phi(2) - \Phi(-2)\right]. \end{aligned}$$

Thus we prove the theorem.

References

1. Chow, Y. S. and Robbins, H. (1965). Asymptotic theory of fixed width confidence intervals for the mean, *Annals of Mathematical Statistics,* 36: 457-462.

2. Dalal, S.C. and Mallows, C.L. (1988)a. When should on stop testing software, *Journal of the American Statistical Association,* Vol. 83, No. 403, 872-879.

3. Jennen, C. (1985). Second Order Approximations to the Density, Mean and Variance of Brownian First-Exit Times, *The Annals of Probability,* Vol. 13, No. 1, 126-144.

4. Jennen, C. and Lerche, H. R. (1981). Fixed Exit Densities of Brownian Motion Through One-sided Moving Boundaries, *Z. Wahrscheïnlichkeitstheorie verw. Gebiete* 55. 133-148.

5. Lerche, H. R. (1986), *Boundary Crossing of Brownian Motion,* Springer-Verlag.

6. Starr, Norman (1972). Optimal and adaptive stoping based on capture times, *Journal of Applied Probability,* 11, 294-301.

7. Vardi, Yehuda (1980). On a stopping time of Starr and its use in estimating the number of transmission sources, *Journal of Applied Probability,* 17, 235-242.