The Second International Conference "Problems of Cybernetics and Informatics"
September 10-12, 2008, Baku, Azerbaijan. The plenary paper
www.pci2008.science.az/4/08.pdf

SOME APPROXIMATIONS FORMULAS FOR CHARACTERISTICS OF TESTS WITH LINEAR AND CURVED STOPPING BOUNDARIES

Asaf Hajiyev[1], Fada Rahimov[2]

[1]Institute of Cybernetics of ANAS, Baku, Azerbaijan, *asaf@baku-az.net*
[2]Baku State Universitety, Baku, Azerbaijan, *ragimovf@rambler.ru*

The ordinary sequential probability ratio test is defined by the crossing of linear boundaries by random walk. The linear boundaries a rise from sequential probability ratio tests of simple hypotheses against simple alternatives. For problems involving several parameters or composite hypotheses we'll consider curved stopping boundaries, which have a complicated structure and so their investigations meet some difficulties [1,2].

The asymptotic formulas for approximation of a significance level, power and expected sample size for tests with linear and curved stopping boundaries will be discussed in this paper and it will be compared with results of [3].

In [3] the following boundary problem was investigated. Let ξ_n, $n \geq 1$ be a sequence of independent and identically distributed random variable with finite mean value $V = E\xi_1$ and it is assumed that the Borel function $\Delta(x)$, $x \in (-\infty, \infty)$ is given. Additionally assume

$$S_n = \sum_{k=1}^{n} \xi_k , \; \overline{S}_n = \frac{1}{n} S_n, \; T_n = n\Delta(\overline{S}_n), \; \tau_a = \inf\{n \geq 1 : T_n \geq f_a(n)\},$$

where $f_a(t)$, $a > 0$, $t > 0$ is a some family of nonlinear boundaries and $\inf\{\varnothing\} = \infty$.

Note that the series of first passage time in theory of boundary crossing problems for random walks have the form τ_a. For example, if $\Delta(x) \equiv x$, then we obtain the following first passage time

$$t_a = \inf\{n \geq 1 : \; S_n \geq \; f_a(n)\},$$

which was investigated in [3,4].

For $f_a(t) \equiv a$ we have the following form of the first passage time

$$v_a = \inf\{n \geq 1 : \; \Delta(\overline{S}_n) \geq a\}$$

see [1,4].

In sequential analysis the statistics in the form $T_n = n\Delta(\overline{S}_n)$ are widely used.

Let F_θ, $\theta \in \Theta$ be a one-parameter exponential family with natural parameter space Θ, that is

$$F_\theta(dx) = \exp\{\theta x - \psi(\theta)\} \lambda(dx), \; -\infty < x < \infty, \; \theta \in \Omega$$

where λ is a non-degenerated, sigma-finite measure on $(-\infty, \infty)$ and Θ consists all θ for which $\exp\{\theta x\}$ is integrable function with respect to λ; that is

$$e^{\psi(\theta)} = \int e^{\theta x} \lambda(dx) < \infty$$

for $\theta \in \Theta$.

Recall that the log-likelihood function, given $\xi_1,...,\xi_n$ which common distribution F_θ is

$$L_n(\theta) = n[\theta \overline{S}_n - \psi(\theta)],\ \theta \in \Theta.$$

Consider testing of the hypothesis $\theta = \theta_0$ versus $\theta \neq \theta_0$.

Let $\Delta(x) = \sup_\theta \{(\theta - \theta_0)x - [\psi(\theta) - \psi(\theta_0)]\}$, $x \in (-\infty, \infty)$.

Then $T_n = n\Delta(\frac{S_n}{n})$ is the log-likelihood ratio statistic for testing $\theta = \theta_0$ versus $\theta \neq \theta_0$ on the basis of $\xi_1,...,\xi_n$, $n > 1$.

As shown in [] the function $\Delta(x)$ may be infinite for some values of x, but $P(T_n < \infty) = 1$. The function $\Delta(x)$ in special case is straightly forward to compute:

1) If F_θ is the normal distribution with mean θ, $-\infty < \theta < \infty$ and has unit variance, then $\theta = \psi'(\theta)$ and $\psi(\theta) = \frac{\theta^2}{2}$. If the null hypothesis is that $H_0 : \theta = 0$, then it easily follows that

$$\Delta(x) = \frac{x^2}{2}.$$

Consider a problem of testing the null hypothesis $H_0 : \theta = 0$. If a sample has (nonrandom) size *n*, then out comes for which the absolute value of $S_n = \xi_1 + ... + \xi_n$ exceeds $3\sqrt{n}$ would be regarded as strong evidence against the null hypothesis H_0, according to classical statistical theory. If data arrives sequentially and S_n is computed for each $n \geq 1$ them $|S_n|$ exceedes $3\sqrt{n}$ for some *n*, even if H_0 is true. The low of an iterated logarithm asserts that

$$P\left(\limsup_{n\to\infty} \frac{S_n - n\theta}{\sqrt{2n \log\log n}} = 1\right) = 1.$$

In this case we have sample of the size

$$v = v_a = \inf\{n \geq 1 : |S_n| \geq a\sqrt{n}\},$$

where $a \geq 3$ and reject H_0 if $|S_v| > 3\sqrt{v}$.

2) Let $\xi_1, \xi_2, ...$ be independent random variables taking the values 1 and 0 with probabilities θ and $1-\theta$ respectively. Let $S_n = \xi_1 + ... + \xi_n$ and $\Delta(x) = x\log x + (1-x)\log(1-x) + \log 2$. To test $H_0 : \theta = \frac{1}{2}$ against $H_1 : \theta \neq \frac{1}{2}$. Let $1 \leq m_0 \leq m$ and

$$v_a = \inf\left\{n \geq m_0 : n\Delta\left(\frac{S_n}{n}\right) \geq a\right\}.$$

Stop sampling at $\min(v_a, m)$ and reject H_0 if $T \le m$ or $T > m$ and $T_m = m\Delta(S_m/m) \ge d$ ($d \le a$).

We'll assume that the function $\Delta(x)$ is positive, twice continuous-differentiable on $x \in (-\infty, \infty)$, moreover $\mu = \Delta(v) > 0$ and $\Delta'(v) \ne 0$.

For the boundary $f_a(t)$ we'll assume that it satisfies to the following conditions:

1) for each *a* the function $f_a(t)$ increases monotonically, is continuously differentiable for $t > 0$, and $f_a(t) \uparrow \infty$, $a \to \infty$.
2) $n = n(\alpha) \to \infty$, $a \to \infty$. Thus $\frac{1}{n} f_a(n) \to \mu$ and $f_a(n) \to \theta$ for some $\theta \in [0, \mu)$.
3) For each *a* the function $f_a'(t)$ weakly oscillates at infinity, i.e.

$$\frac{f_a'(n)}{f_a'(m)} \to 1 \text{ at } \frac{n}{m} \to 1,\ n \to \infty.$$

Denote $N_a = N_a(\mu)$ a solution of the equation $f_a(n) = n\mu$ which exists for sufficiently large *a* [3]. Also denote $\Phi(x)$ a standard normal distribution.

Theorem. Let $\xi_n, n \ge 1$ be a sequence of independent and identically distributed random variables with $\sigma^2 = D\xi_2 < \infty$, $v = E\xi_1$ and let above mentioned conditions are satisfied for function $\Delta(x)$ and boundary $f_a(t)$.

Then

$$\lim_{a\to\infty} P\left(\tau_a - N_a \le \frac{r\,x}{\lambda}\sqrt{N_a}\right) = \Phi(x),\ r = \left|\Delta'(v)\right|\sigma,$$

where $\lambda = \mu - \theta$.

Corollary. Let the conditions of the theorem are true and $n = n(\alpha) \to \infty$ as $a \to \infty$ such that

$$c_n = \frac{f_a(n) - n\mu}{r\sqrt{n}} = O(1).$$

Then

$$\lim_{a\to\infty} [P(\tau_a \le n) - \Phi(-c_n)] = 0.$$

Theorem and corollary proved in [3].

We present example, which is especially instructive (see [1]). Let $\xi_1, \xi_2, \ldots$ be independent and normally distributed random variables with mean μ and unit variance. It is testing $H_0: \mu = \mu_0$ against $H_1: \mu = \mu_1$ (say $\mu_0 < \mu_1$).

The likelihood ratio is

$$L_n = \prod_{k=1}^{n} \frac{\varphi(\xi_k - \mu_1)}{\varphi(\xi_k - \mu_0)} = e^{(\mu_1 - \mu_0)S_n - \frac{n}{2}(\mu_1^2 - \mu_0^2)},$$

where $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and $S_n=\sum\limits_{k=1}^{n}\xi_k$.

The stopping rule of sequentially probability ratio test can be written

$$\tau=\inf\{n\ge 1: S_n-\tfrac{n}{2}(\mu_1+\mu_0)\notin(a,b)\}, \qquad (1)$$

where $a=\log\dfrac{A}{\mu_1-\mu_0}$, $b=\log\dfrac{B}{\mu_1-\mu_0}$, ($A<1<B$) are constants.

If $\tau<\infty$ the sequential probability ratio test rejects H_0 if and only if

$$S_N\ge b+\frac{\tau}{2}(\mu_1+\mu_0).$$

A simple special case is the symmetric one $\mu_1=-\mu_0$, $b=-a$, for which (1) becomes

$$\tau_b=\inf\{n\ge 1: |S_n|\ge b\}.$$

The main results of [3] implies approximation of the distribution of the sample size τ_b:

$$P_\mu(\tau_b\le n)\approx\Phi\left(\frac{n\mu-b}{\sqrt{n}}\right),\ \mu\ne 0.$$

We also study the approximation of the significance level and power of stopping rule t_a by the results of work [3].

References

1. D. Siegmund. Sequential analysis. Tests and confidence intervals – New York, etc. Springer – Verlag, 1985, 272 p.
2. Hajiev A., Rahimov F.Applcation repeated significance tests comparing more than two treatments in clinical experiment(in Russian). The international conference "Problems of cybernetics and informatics", v.1, pp. 190-193.
3. Hajiev A., Rahimov F. On generalization of one class of the first passage time of random walk for the linear boundary. Transactions of NAS of Azerbaijan 2006. № V. XXVI, pp. 59-66.
4. Woodroof M., Nonlinear renewal theory in sequential analysis. SIAM, 1982, 119 p.