THE POWER OF THE LIKELIHOOD RATIO TEST

By Bradley Efron

Stanford University and the RAND Corporation

1. Introduction. Suppose we are given n independent and identically distributed observations $x_1, x_2, \cdots, x_n$ of a random variable X having density function $f(x)$ with respect to some measure $\mu(x)$ on a measurable space Ω, and are asked to test the simple hypothesis $f(x) \equiv f_0(x)$ versus the simple alternative $f(x) \equiv f_1(x)$ at a significance level α, $0 < \alpha < 1$. It is well-known that the most powerful test, which rejects for large values of the likelihood ratio

$$\prod_{i=1}^{n} (f_1(x_i)/f_0(x_i)),$$

has an "error probability of the second kind" (probability of mistakenly accepting the null hypothesis) $\beta_n(\alpha)$ satisfying

$$\lim_{n\to\infty} (\log \beta_n(\alpha)/n) = -I, \tag{1}$$

where I is the Kullback-Leibler information number

$$I = E_0(\log (f_0(X)/f_1(x))) = \int_\Omega (\log (f_0(x)/f_1(x)))f_0(x)\, d\mu(x). \tag{2}$$

A nice proof of (1), which requires no additional assumptions, can be found in Section 4 of [4].

Here it is shown that if we make the additional assumption that

$$E_0(|\log (f_0(X)/f_1(X))|^3) < \infty,$$

(E_0 always indicating expectation under the null hypothesis), a better limiting expression for $\beta_n(\alpha)$ can be derived which is sensitive enough to allow power comparisons between different levels of α. In Section 3 the usefulness of similar expressions for simple numerical approximation of the function $\beta_n(\alpha)$ in small samples is illustrated.

In addition to the information number I defined above, let

$$J = E_0(\log (f_0(X)/f_1(X)) - I)^2 \tag{3}$$

and

$$K = E_0(\log (f_0(X)/f_1(X)) - I)^3, \tag{4}$$

which are both finite by the previous assumption. Then we have the following:

Theorem. *If* $\log (f_0(X)/f_1(X))$ *is not a lattice random variable under the null hypothesis, then*

$$\beta_n(\alpha) = \exp \{-[nI - (nJ)^{\frac{1}{2}}z_\alpha + (K/6J)(1 - z_\alpha^2) + \tfrac{1}{2}z_\alpha^2]\} \cdot (2\pi nJ)^{-\frac{1}{2}}(1 + o_n(1)) \tag{5}$$

Received 15 September 1966.

where z_α is the upper α-point of the normal distribution,

(6) $$\Phi(z_\alpha) \equiv \int_{-\infty}^{z_\alpha} e^{-\frac{1}{2}t^2}/(2\pi)^{\frac{1}{2}}\, dt = 1 - \alpha.$$

In the lattice case,

(7) $$\beta_n(\alpha) = (\exp\{-[nI - (nJ)^{\frac{1}{2}} z_\alpha]\} \cdot n^{\frac{1}{2}}) e^{O_n(1)}.$$

(Here $o_n(1)$ approaches zero uniformly for α in any closed interval $[\alpha_l, \alpha_u]$, $0 < \alpha_l < \alpha_u < 1$. The sequence $|O_n(1)|$ is uniformly bounded for values of α in such an interval.)

If α and α' are different values of α, the theorem gives the following estimate for the error ratio in the non-lattice case,

(8) $$\beta_n(\alpha)/\beta_n(\alpha') = \exp[(nJ)^{\frac{1}{2}}(z_\alpha - z_{\alpha'}) + (K/6J - \tfrac{1}{2})(z_\alpha^2 - z_{\alpha'}^2)] \cdot (1 + o_n(1)),$$

and a cruder expression in the lattice case.

2. Proof of Theorem. Let T_n be the random variable

(9) $$T_n = (\textstyle\sum_1^n \log(f_1(X_i)/f_0(X_i)) + nI)/(nJ)^{\frac{1}{2}},$$

which has mean 0 and variance 1 under the null hypothesis. T_n is a monotone function of the likelihood ratio, and therefore a sufficient statistic. Letting $H_n(t)$ and $H_n'(t)$ be the induced cdf's of T_n on the real line under the null and alternative hypotheses respectively, we have by sufficiency

(10) $$dH_n'(t_n)/dH_n(t_n) = \textstyle\prod_1^n f_1(x_i)/\prod_1^n f_0(x_i) = \exp\{(nJ)^{\frac{1}{2}} t_n - nI.\}$$

Consider the non-lattice case. Letting $z_{\alpha,n}$ be the (smallest) upper α point of H_n,

(11) $$H_n(z_{\alpha,n}) = 1 - \alpha,$$

the error probability of the likelihood ratio test[1] is seen to be

(12) $$\beta_n(\alpha) = \int_{-\infty}^{z_{\alpha,n}} \exp\{(nJ)^{\frac{1}{2}} t - nI\}\, dH_n(t).$$

By Theorem 2 of [2], page 210,

(13) $$H_n(t) = \Phi(t) + n^{-\frac{1}{2}} D(t) + o(n^{-\frac{1}{2}}) \quad \text{uniformly in} \quad t,$$

where

(14) $$D(t) = (e^{-t^2/2}/(2\pi)^{\frac{1}{2}})(K/6J^{\frac{3}{2}})(1 - t^2).$$

Expanding the smooth functions Φ and D in a Taylor's series about z_α yields

(15) $$z_{\alpha,n} - z_\alpha = -\{K/6J^{\frac{3}{2}}\}(1 - z_\alpha^2) n^{-\frac{1}{2}} + o(n^{-\frac{1}{2}}).$$

---

[1] If the X_I are discrete random variables, no randomized likelihood ratio test may have exact size α. Equation (13) shows that sizes $\alpha_{1n} \leqq \alpha \leqq \alpha_{2n}$ satisfying $\alpha_{2n} - \alpha_{1n} = o(n^{-\frac{1}{2}})$ are always possible though, and it then follows from the proof above that (5) holds even for those values of α requiring randomization.

If equation (12) is rewritten as

$$(16)\quad \beta_n(\alpha) = \exp\{-[nI - (nJ)^{\frac{1}{2}} z_{\alpha,n}]\}/(nJ)^{\frac{1}{2}} \cdot \int_{-\infty}^{z_{\alpha,n}} (nJ)^{\frac{1}{2}} \exp\{(nJ)^{\frac{1}{2}}(t - z_{\alpha,n})\}\, dH_n(t),$$

equation (15) gives

$$(17)\quad \beta_n(\alpha) = \exp\{-[nI - (nJ)^{\frac{1}{2}} z_\alpha + (K/6J)(1 - z_\alpha^2)]\}/(nJ)^{\frac{1}{2}} \cdot \int_{-\infty}^{z_{\alpha,n}} (nJ)^{\frac{1}{2}} \exp\{(nJ)^{\frac{1}{2}}(t - z_{\alpha,n})\}\, dH_n(t)(1 + o_n(1))$$

uniformly for α in compact intervals excluding 0 and 1. It remains to show that the integral uniformly approaches

$$\varphi(z_\alpha) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2} z_\alpha^2}$$

as n approaches infinity.

Let $c_n \equiv \log n/(nJ)^{\frac{1}{2}}$. Notice that

$$(18)\qquad \int_{-\infty}^{z_{\alpha,n}-c_n} (nJ)^{\frac{1}{2}} \exp\{(nJ)^{\frac{1}{2}}(t - z_{\alpha,n})\}\, dH_n(t) \leqq (J/n)^{\frac{1}{2}}.$$

For any s and any δ, $0 < \delta < 1$, we have

$$(19)\quad e^{-\delta}\left[\frac{H_n\left(s + \frac{\delta}{(nJ)^{\frac{1}{2}}}\right) - H_n(s)}{\Phi\left(s + \frac{\delta}{(nJ)^{\frac{1}{2}}}\right) - \Phi(s)}\right] \leqq \frac{\int_s^{s+\frac{\delta}{(nJ)^{\frac{1}{2}}}} (nJ)^{\frac{1}{2}} \exp\{(nJ)^{\frac{1}{2}}(t - z_{\alpha,n})\}\, dH_n(t)}{\int_s^{s+\frac{\delta}{(nJ)^{\frac{1}{2}}}} (nJ)^{\frac{1}{2}} \exp\{(nJ)^{\frac{1}{2}}(t - z_{\alpha,n})\}\, d\Phi(t)} \leqq e^{\delta}\left[\frac{H_n\left(s + \frac{\delta}{(nJ)^{\frac{1}{2}}}\right) - H_n(s)}{\Phi\left(s + \frac{\delta}{(nJ)^{\frac{1}{2}}}\right) - \Phi(s)}\right].$$

Equation (13) yields

$$(20)\quad [H_n(s + \delta/(nJ)^{\frac{1}{2}}) - H_n(s)]/[\Phi(s + \delta/(nJ)^{\frac{1}{2}}) - \Phi(s)] = 1 + o_n(1)/\delta$$

uniformly for $0 < \delta < 1$ and s in any compact interval. This implies

$$(21)\quad e^{-\delta}[1 + (o_n(1)/\delta)] \leqq \int_{z_{\alpha,n}-c_n}^{z_{\alpha,n}} (nJ)^{\frac{1}{2}} \exp\{(nJ)^{\frac{1}{2}}(t - z_{\alpha,n})\}\, dH_n(t) \cdot \left(\int_{z_{\alpha,n}-c_n}^{z_{\alpha,n}} (nJ)^{\frac{1}{2}} \exp\{(nJ)^{\frac{1}{2}}(t - z_{\alpha,n})\}\, d\Phi(t)\right)^{-1} \leqq e^{\delta}[1 + (o_n(1)/\delta)].$$

As n approaches infinity, the denominator approaches

$$\varphi(z_\alpha) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2} z_\alpha^2}$$

uniformly for values of α bounded away from 0 and 1. Therefore, making use of (18),

$$(22)\quad \varphi(z_\alpha) e^{-\delta}[1 + (o_n(1)/\delta)] \leqq \int_{-\infty}^{z_{\alpha,n}} (nJ)^{\frac{1}{2}} \exp\{(nJ)^{\frac{1}{2}}(t - z_{\alpha,n})\}\, dH_n(t) \leqq \varphi(z_\alpha) e^{\delta}[1 + o_n(1)/\delta].$$

Choosing $\delta_n = (|o_n(1)|)^{\frac{1}{2}}$ yields the desired result

$$\lim_{n\to\infty} \int_{-\infty}^{z_{\alpha,n}} (nJ)^{\frac{1}{2}} \exp \{(nJ)^{\frac{1}{2}}(t - z_{\alpha,n})\} \, dH_n(t) = \varphi(z_\alpha). \tag{23}$$

The proof for the lattice case follows as above from the standard Berry-Esseen theorem. Actually, a somewhat more precise result than the one stated can be obtained from Theorem 1, page 213 of [2].

It should be noted[2] that the proof given above is similar to that given by Bahadur and Rao [1] in their study of the probability of large deviations. As a matter of fact, under slightly stronger conditions than those assumed here our theorem follows as a special case of their results, as extended by Petrov [3]. Letting $L = \log (f_0(X)/f_1(X))$, the condition $E_1 e^{tL} < \infty$ for some $t > 1$ is needed to apply their results (E_1 indicating expectation under the alternative hypothesis) as opposed to the condition $E_1 |L|^3 e^L < \infty$ we have assumed. Conversely, the proof above can be used to weaken slightly the conditions needed for some general large deviation results similar to those given in [1].

The theorem given here can be improved by making stronger assumptions about the distribution of the log likelihood ratio, (primarily, the existence of higher moments) and using more terms of the Cramér expansion of $H_n(t)$ in place of (13) in the proof above. (See page 220 of [2] for this expansion.) We then get approximations of $\beta_n(\alpha)$ accurate up to a factor of $1 + o_n(1/n^p)$ where p is an arbitrarily high power, involving complicated functions of the moments of the log likelihood ratio.

3. Numerical approximation. If we simply substitute Φ for H_n in formula (12), we get the approximation

$$\beta_n(\alpha) \cong e^{-nI} \int_{-\infty}^{z_{\alpha,n}} \exp \{(nJ)^{\frac{1}{2}}t - \tfrac{1}{2}t^2\})/(2\pi)^{\frac{1}{2}} \, dt \tag{24}$$

or

$$\beta_n(\alpha) \cong \exp \{-n(I - J/2)\}\Phi(z_{\alpha,n} - (nJ)^{\frac{1}{2}}) \tag{25}$$

$$\cong \exp \{-n(I - J/2)\}\Phi(z_\alpha - (nJ)^{\frac{1}{2}}). \tag{26}$$

[Note: For "close" alternatives, expansion of $\log (f_1(x)/f_0(x))$ gives $J \cong 2I$ and $I - J/2 \cong \frac{5}{6}E_0[(f_1(x) - f_0(x))/f_0(x))]^3$.] For two different values α and α', we have the approximations to the error ratio

$$\beta_n(\alpha)/\beta_n(\alpha') \cong \Phi(z_{\alpha,n} - (nJ)^{\frac{1}{2}})/\Phi(z_{\alpha',n} - (nJ)^{\frac{1}{2}}) \tag{27}$$

$$\cong \Phi(z_\alpha - (nJ)^{\frac{1}{2}})/\Phi(z_{\alpha'} - (nJ)^{\frac{1}{2}}). \tag{28}$$

(In plotting the error curve, it is also possible to make use of the symmetric approximation

$$\alpha_n(\beta) \cong \exp \{-n(\bar{I} - \bar{J}/2)\}\Phi(z_{\beta,n} - (n\bar{J})^{\frac{1}{2}}), \tag{29}$$

---

[2] I am very grateful to the referee for pointing out the connection between the papers [1] and [3] and the theorem given here.

where

$$\bar{I} = \int_\Omega (\log (f_1(x)/f_0(x)))f_1(x)\, d\mu(x), \quad \text{etc.}).$$

As an example of the efficacy of these approximations consider the case of exponential random variables, $f_0(x) = \lambda_0 e^{-\lambda_0 x}$, $f_1(x) = \lambda_1 e^{-\lambda_1 x}$ for $x > 0$, $\mu(x)$ ordinary Lebesgue measure. Table 1 was computed for $\alpha = .05$, $\alpha' = .10$, and $\lambda_1/\lambda_0 = .6$.

TABLE 1

n	$\alpha = .05$			$\alpha' = .10$			$\beta_n(\alpha)/\beta_n(\alpha')$		
	$\beta_n(\alpha)$	Approx. I (25)	Approx. II (26)	$\beta_n(\alpha')$	Approx. I (25)	Approx. II (26)	Actual	Approx. I (27)	Approx. II (28)
10	.468	.518	.476	.350	.386	.372	1.34	1.34	1.28
15	.338	.372	.339	.235	.259	.248	1.44	1.44	1.37
20	.242	.264	.239	.157	.172	.165	1.54	1.53	1.45
25	.171	.189	.167	.105	.114	.109	1.66	1.66	1.53
30	.120	.130	.116	.0692	.075	.0720	1.73	1.72	1.61
45	.0395	.0422	.0374	.0196	.0212	.0201	2.02	2.00	1.86
50	.0270	.0287	.0253	.0128	.0138	.0130	2.11	2.09	1.94

REFERENCES

[1] Bahadur, R. R., and Rao, R. R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[2] Gnedenko, B. V., and Kolmogorov, A. N. (1954). *Limit distributions for sums of independent random variables.* Addison-Wesley, Cambridge, Mass.

[3] Petrov, V. V. (1965). On the probabilities of large deviations for sums of independent random variables. *Theory of Prob. and its Appl.* **10** 287–298.

[4] Rao, C. R. (1962). Efficient estimates and optimum inference procedures in large samples. *J. Roy. Statist Soc.* **24** (B) 46–72.