A NOTE ON THE POWER OF THE BEST CRITICAL REGION FOR INCREASING SAMPLE SIZE

By D. BASU

Statistical Laboratory, Calcutta

1. Let $\{X_i\}$ ($i = 1,2,\ldots$ad inf) be a sequence of random variables and let H_0 be a simple hypothesis about the sequence to be tested against a simple alternative H_1 on the evidence of n random observations $x_1, x_2, \ldots, x_n$ on the first n variables $X_1, X_2 \ldots X_n$. The classical procedure is to control the first kind of error at a fixed level α and then to choose that critical region w_n, in the n-dimensional sample space R_n which has the maximum power β_n. The question whether by making n sufficiently large we can also control the second kind of error $1-\beta_n$ seems not to have been sufficiently investigated. It is easily demonstrated that β_n is a non-decreasing function n, but whether, or under what conditions, $\beta_n \to 1$ as $n \to \infty$ is not known. In the important case when the X's are independently and identically distributed it is here shown that $\beta_n \to 1$. It is however not difficult to construct examples where β_n tends to a constant less than unity and a simple example to this effect has been considered in §3.

2. Let p_{in} ($n=1,2,\ldots$ad inf) be the joint frequency density function of $X_1, X_2, \ldots, X_n$ under hypothesis $H_i (i = 0, 1)$ and let $x_{(n)}$ stand for the sample vector $(x_1, x_2, \ldots, x_n)$. Let w_n be the best critical region of size α in R_n and let w'_{n+1} be that region in R_{n+1} which is constructed by taking w_n in R_n and then giving x_{n+1} all possible values, *i.e.* w'_{n+1} is a cylinder with base w_n.

Clearly, $$\int_{w'_{n+1}} p_{i\,n+1} dx_{(n+1)} = \int_{w_n} dx_{(n)} \int_{-\infty}^{\infty} p_{i\,n+1} dx_{n+1} = \int_{w_n} p_{in} dx_{(n)} \quad (i = 0, 1)$$

Thus w'_{n+1} is a region in R_{n+1} of size α and power β_n. But β_{n+1} is the power of the best critical region in R_{n+1}.

$$\therefore \quad \beta_n \leqslant \beta_{n+1} \text{ for every } n.$$

The monotonic character of β_n was noted by Rao (1948). We now prove that $\beta_n \to 1$ under the assumption that the X's are all independently and identically distributed. In this case the density function p_{in} can be factorised into

$$p_{in} = p_i(x_1)\, p_i(x_2) \ldots p_i(x_n) \qquad (i = 0,1)$$

where p_i is the density function of each of the X's under hypothesis H_i.

Since w_n is the best critical region in R_n we have $p_{1n} \geqslant k_n p_{0n}$ for every $x_{(n)}$ in w_n and $p_{1n} \leqslant k_n p_{0n}$ for every $x_{(n)}$ outside w_n where the constant k_n is so adjusted that w_n is of size α.

Let

$$z_m = \log \frac{p_1(x_m)}{p_0(x_m)} \quad (m = 1, 2, \ldots, \text{ad inf.})$$

The z's are independently and identically distributed and we assume that $E(z)$ exists under both H_0 and H_1. Let $\mu_i = E(z|H_i)$ $(i = 0,1)$.

$$\text{Clearly} \quad \mu_1 - \mu_0 = \int_{-\infty}^{\infty} \log \frac{p_1}{p_0} p_1 dx - \int_{-\infty}^{\infty} \log \frac{p_1}{p_0} . p_0 dx = \int_{-\infty}^{\infty} (\log p_1 - \log p_0)(p_1 - p_0)$$

$$\geqslant 0$$

as the two factors in the integrand are always of the same sign.

Hence $\mu_0 \leqslant \mu_1$, the sign of equality holding only when $p_0(x) = p_1(x)$ almost everywhere in x, and in this trivial case the question of testing H_0 against H_1 does not arise. The more general inequality namely $\mu_0 < 0 < \mu_1$, is due to Wald and is of fundamental importance in sequential analysis.

Now from the definition of w_n we have

$$\left.\begin{array}{l} P(z_1+z_2+\ldots+z_n \geqslant \lambda_n | H_0) = \alpha \\ P(z_1+z_2+\ldots+z_n \geqslant \lambda_n | H_1) = \beta_n \end{array}\right\} \text{ for every } n,$$

where $\lambda_n = \log k_n$.

Now by Khintchine's theorem $\frac{1}{n}(z_1+z_2+\ldots+z_n)$ converges in probability to μ_i under hypothesis H_i $(i = 0, 1)$.

$$\text{But} \qquad P\left(\frac{z_1+\ldots+z_n}{n} \geqslant \frac{\lambda_n}{n} | H_0\right) = \alpha \qquad \text{for every } n.$$

Hence it follows that $\frac{\lambda_n}{n} \to \mu_0$ as $n \to \infty$.

$$\text{But} \qquad \beta_n = P\left(\frac{z_1+\ldots+z_n}{n} \geqslant \frac{\lambda_n}{n} | H_1\right)$$

and $\frac{1}{n}(z_1+\ldots+z_n) \to \mu_1$ (in probability sense) and $\frac{\lambda_n}{n} \to \mu_0$ where $\mu_0 < \mu_1$.

$$\therefore \qquad \beta_n \to 1 \text{ as } n \to \infty.$$

The above result is inherent in the literature but we have not come across any explicit proof. The result was known to the late Prof. Wald. The above simple proof is the result of a little discussion with him.

3. We now demonstrate that when $X_1, X_2, \ldots$ are not independently and identically distributed the power of the best critical region may tend to a constant less than unity.

Let $X_1, X_2, \ldots$ be independent of one another and let H_0 be the simple hypothesis that all the X's are standard normal variates and let H_1 be the alternative hypothesis that X_i is distributed as a normal variable with mean μ_i and s.d. unity ($i = 1,2,\ldots$ad inf)

Thus $$p_0(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}x_i^2\right)$$

and $$p_1(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}(x_i-\mu_i)^2\right)$$

$$\therefore \quad z_i = \log \frac{p_1(x_i)}{p_0(x_i)} = \mu_i x_i - \tfrac{1}{2}\mu_i^2$$

i.e. z_i is distributed as $(-\frac{1}{2}\mu_i^2, |\mu_i|)$ under H_0 and as $(\frac{1}{2}\mu_i^2, |\mu_i|)$ under H_1.
In precisely the same way as before we have

and $$\left.\begin{aligned} P(z_1+z_2+\ldots+z_n \geqslant \lambda_n | H_0) &= \alpha \\ P(z_1+z_2+\ldots+z_n \geqslant \lambda_n | H_1) &= \beta_n \end{aligned}\right\} \quad \text{for every } n.$$

Now $Z_n = z_1+z_2+\ldots+z_n$ is distributed as $(-\frac{1}{2}b_n^2, b_n)$ under H_0 and as $(\frac{1}{2}b_n^2, b_n)$ under H_1, where $b_n^2 = \mu_1^2+\mu_2^2+\ldots+\mu_n^2$.

As $$\alpha = P(Z_n \geqslant \lambda_n \mid H_0) = P\left(\frac{Z_n+\frac{1}{2}b_n^2}{b_n} \geqslant \frac{\lambda_n+\frac{1}{2}b_n^2}{b_n} \Big| H_0\right),$$

$$\therefore \quad \frac{\lambda_n+\frac{1}{2}b_n^2}{b_n} = \gamma$$ where γ is the upper $100\alpha\%$ value of a standard normal variate.

But $$\beta_n = P\,(Z_n \geqslant \lambda_n | H_1) = P\left(\frac{Z_n-\frac{1}{2}b_n^2}{b_n} \geqslant \frac{\lambda_n-\frac{1}{2}b_n^2}{b_n} \Big| H_1\right)$$

and $$\frac{\lambda_n-\frac{1}{2}b_n^2}{b_n} = -b_n + \frac{\lambda_n+\frac{1}{2}b_n^2}{b_n} = -b_n+\gamma.$$

$$\therefore \quad \beta_n \to 1 \text{ if and only if } b_n \to \infty.$$

Hence a necessary and sufficient condition in order that $\beta_n \to 1$ is that $\sum_1^\infty \mu_i^2$ is divergent.

I wish to thank Dr. C. R. Rao for his kind interest in my work.

References

Cramér, H. (1946): *Mathematical Methods of Statistics*, 253. Princeton University Press.

Rao, C. R. (1948): The utilization of multiple measurements in problems of biological classifications. *J. Roy. Stat. Soc.*, Series B, Vol. X, No. 2.