Canad. Math. Bull. Vol. 16 (1), 1973

ON SELECTING A SPURIOUS OBSERVATION

BY
K. S. MOUNT AND B. K. KALE(1)

1. Consider a life testing experiment in which $(X_1, X_2, \ldots, X_n)$ are such that $(n-1)$ of them are distributed as $f(x, \sigma)=(1/\sigma)e^{-x/\sigma}$, $x\geq 0$, $\sigma>0$ and one of them is distributed as $f(x, \sigma/\alpha)$, $0<\alpha<1$. A priori each X_i has probability $1/n$ of being a spurious observation distributed as $f(x, \sigma/\alpha)$. For such an experiment Kale and Sinha [2] showed that if u_r denotes the probability that $X_{(r)}$, the r^{th} component of the order statistic, corresponds to the spurious observation, then $u_1<u_2<\cdots<u_n$. Generalizing the above model we assume that $(X_1, \ldots, X_n)$ are such that $(n-1)$ of them are distributed with d.f. $F(x)$, and one of them is distributed with d.f. $G(x)$, where F and G are stochastically ordered, i.e., $G<F$. A priori each X_i has probability $1/n$ of being a spurious observation distributed as G. Then following Kale and Sinha [2],

$$u_r = \binom{n-1}{r-1}\int_{R_1}[F(x)]^{r-1}[1-F(x)]^{n-r}\,dG(x). \tag{1}$$

LEMMA. *Let $dG/dF=\psi(x)$. We show that if $\psi(x)$ is monotone increasing, then $u_1<u_2<\cdots<u_n$.*

Proof.

$$\begin{aligned} u_r &= \binom{n-1}{r-1}\int_{-\infty}^{\infty}[F(x)]^{r-1}[1-F(x)]^{n-r}\psi(x)\,dF(x) \\ &= \binom{n-1}{r-1}\int_0^1 y^{r-1}(1-y)^{n-r}\psi[F^{-1}(y)]\,dy \end{aligned} \tag{2}$$

$$= \frac{1}{n}E[\psi_1(Y_r)] \tag{3}$$

where Y_r denotes a beta r.v. with parameters r and $n-r+1$. Note that $\{Y_r\}_1^n$ is stochastically ordered (increasing) since $[dH(Y_{r+1})/dH(Y_r)]\alpha(y/1-y)$ which is monotone increasing in y for $0\leq y\leq 1$. Further $\psi_1(y)=\psi[F^{-1}(y)]$ is strictly increasing, since $\psi\not\equiv 1$ for otherwise $G\equiv F$. We apply now the results of Lehmann, [3, p. 112, Problem 11] for strictly increasing functions to conclude that $u_1<u_2<\cdots<u_n$.

Some important families of the d.f.'s (F, G) are $G(x)=[F(x)]^k$, $k>1$, i.e., Lehmann alternatives and $G(x)=\sum_{k=1}^{\infty} C_k[F(x)]^k$, $C_k\geq 0$, $\sum C_k=1$, i.e., a convex

Received by the editors June 22, 1971 and, in revised form, September 9, 1971.
(1) Partially supported by a research grant from the National Research Council of Canada.

combination of Lehmann alternatives. The condition $dG/dF=\psi(x)$ where ψ is monotone increasing implies that G and F belong to a monotone likelihood ratio family. A subclass of this is distributions belonging to one parameter exponential class of densities of the form

$$p_\theta(x) = C(\theta)e^{xQ(\theta)}h(x)$$

where $Q(\theta)$ is a monotone increasing function.

Suppose $p_\theta(x)$ is of the form

$$p_\theta(x) = C(\theta)e^{T(x)Q(\theta)}h(x) \tag{4}$$

where $T(x)$ is a real valued function of x. We know the p.d.f. of $Y=T(X)$ is of the form

$$r_\theta(y) = C(\theta)e^{yQ(\theta)}s(y). \tag{5}$$

Let us take a sample of size n say $(y_1, \ldots, y_n)$ with $n-1$ of the observations coming from $f(y)=r_{\theta_0}(y)$ and one observation from $g(y)=r_{\theta_1}(y)$, $\theta_1>\theta_0$. If u_r is the probability that $Y_{(r)}$ corresponds to the spurious observation, then, by our previous remarks, u_r is a monotone increasing function of r. An example of this is the family of distributions $\{N(0, \theta):\theta>0\}$. Here $T(x)=x^2$. Finally, we note that if the ψ in the Lemma is monotone decreasing, then u_r would be a monotone decreasing function of r.

2. **Slippage tests for detecting spurious observations.** We can phrase the problem of detecting spurious observations as a slippage problem. Suppose X_i has d.f. $F_i(x)$, $i=1, \ldots, n$ and the X_i are independent. We wish to test

$$\begin{aligned} &H_0: F_1 = \ldots = F_n = F_0 \qquad F_0\text{—completely specified d.f.} \\ \text{vs.}\quad &H_i: F_1 = \ldots = F_{i-1} = F_{i+1} = \ldots = F_n = F_0 \qquad F_i = G < F_0 \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, \ldots, n \end{aligned} \tag{6}$$

In line with the usual criteria for such tests, [1], we are interested in a test such that:

$$\begin{aligned} &P\{\text{rej. } H_0 \mid H_0 \text{ true}\} = \alpha \\ &P\{\text{acc. } H_i \mid H_i \text{ true}\} \text{ does not depend on } i \\ &P\{\text{acc. } H_i \mid H_i \text{ true}\} \text{ is maximized.} \end{aligned} \tag{7}$$

We assume that if a distribution has slipped, it is equally likely to be any F_i. If the d.f.'s F_0 and G have p.d.f.'s f_0 and g respectively, the joint p.d.f. of $X_1, \ldots, X_n$ is $\prod_{j=1}^n f_0(x_j)$ if H_0 is true and $1/n \sum_{i=1}^n g(x_i) \prod_{j\neq i} f_0(x_j)$ if H_0 is not true. The test satisfying the criteria in display (7) will accept H_0 if

$$\max_j \frac{g(x_j)}{f_0(x_j)} < C_{n,\alpha} \tag{8}$$

and will accept H_i if

$$\frac{g(x_i)}{f_0(x_i)} = \max_j \frac{g(x_j)}{f_0(x_j)} \geq C_{n,\alpha} \tag{9}$$

https://doi.org/10.4153/CMB-1973-015-0 Published online by Cambridge University Press

where the constant $C_{n,\alpha}$ is chosen to satisfy the level α restriction. It is well known [1, p. 307] that if f_0 and g are members of a family which has monotone likelihood ratio in x, then the acceptance regions (8) and (9) become

$$x_{(n)} < C_{n,\alpha} \tag{10}$$

and

$$x_i = x_{(n)} \geq C_{n,\alpha}. \tag{11}$$

This test is often used for detecting spurious observations.

LEMMA. *If $G < F_0$, and has p.d.f. $g(x)$, the test with critical region $x_{(n)} > C_{n,\alpha}$ is unbiased.*

Proof. We know that

$$\begin{aligned}\alpha &= P\{\text{rej. } H_0 \mid H_0\}\\ &= P\{(\text{accept one of the } H_i \mid H_0\}\\ &= \sum_{i=1}^{n} P\{\text{acc. } H_i \mid H_0\}\\ &= nP\{\text{acc. } H_1 \mid H_0\}.\end{aligned}$$

To show this test is unbiased we must show: $P\{\text{acc. } H_i \mid H_i\} \geq P\{\text{acc. } H_i \mid H_j\}$, $i \neq j$. First we show that $P\{\text{acc. } H_i \mid H_i\} = P\{\text{acc. } H_1 \mid H_1\} \geq P\{\text{acc. } H_1 \mid H_0\}$ $(=\alpha/n)$. The point $C_{n,\alpha}$ is chosen so that

$$n\int_{C_{n,\alpha}}^{\infty} [F_0(x)]^{n-1} f_0(x)\,dx = \alpha.$$

We know that $P\{\text{acc. } H_1 \mid H_1\} = \int_{C_{n,\alpha}}^{\infty} [F_0(x)]^{n-1} g(x)\,dx$. Finally, the inequality

$$\int_{C_{n,\alpha}}^{\infty} [F_0(x)]^{n-1} g(x)\,dx \geq \int_{C_{n,\alpha}}^{\infty} [F_0(x)]^{n-1} f_0(x)\,dx \tag{12}$$

can be seen to hold by integrating both sides by parts. For $j > 0$

$$\begin{aligned}P\{\text{acc. } H_i \mid H_j\} &= \int_{C_{n,\alpha}}^{\infty} [F_0(x)]^{n-2} G(x) f_0(x)\,dx\\ &\leq \int_{C_{n,\alpha}}^{\infty} [F_0(x)]^{n-1} f_0(x)\,dx \left(= \frac{\alpha}{n}\right).\end{aligned}$$

Similarly, $P\{\text{acc. } H_0 \mid H_0\} \geq P\{\text{acc. } H_0 \mid H_i\}$, $i = 1, 2, \ldots, n$.

ACKNOWLEDGEMENT. We wish to thank the referee for his helpful comments and suggestions.

https://doi.org/10.4153/CMB-1973-015-0 Published online by Cambridge University Press

REFERENCES

1. T. Ferguson, *Mathematical statistics, a decision theoretic approach*, Academic Press, 1967.

2. B. K. Kale, and S. K. Sinha, *Estimation of expected life in the presence of an outlier observation*. Technometrics, **13** (1971), 755-759.

3. E. L. Lehmann, *Testing statistical hypotheses*, Wiley, New York, 1959.

UNIVERSITY OF MANITOBA,
WINNIPEG, MANITOBA