The Annals of Statistics
1987, Vol. 15, No. 1, 456–461

IS THE SELECTED POPULATION THE BEST?

BY SAM GUTMANN AND ZAKHAR MAYMIN

Northeastern University

Random variables $X_i \sim N(\theta_i, 1)$, $i = 1, 2, \ldots, k$, are observed. Suppose X_S is the largest observation. If the inference $\theta_S > \max_{i \neq S} \theta_i$ is made whenever $X_S - \max_{i \neq S} X_i > c$, then the probability of a false inference is maximized when two θ_i are equal and the rest are $-\infty$. Equivalently, the inference can be made whenever a two-sample two-sided test for difference of means, based on the largest two observations, would reject the hypothesis of no difference. The result also holds in the case of unknown, estimable, common variance, and in fact for location families with monotone likelihood ratio.

1. Introduction. We observe independent X_i, $i = 1, \ldots, k$, where X_i has density $p(x - \theta_i)$ where p is normal (or, more generally, has monotone likelihood ratio). The best population corresponds to the largest θ_i; the selected population corresponds to the largest X_i. The goal of this paper is to attach the best possible p-value to the inference: the selected population is the best. (This is not a standard hypothesis testing problem, because the hypotheses are set up *after* viewing the data.)

Let S be the random variable such that $X_S = \max_i X_i$. Suppose we infer that the selected population is best, i.e., $\theta_S > \max_{i \neq S} \theta_i$, whenever $(X_1, \ldots, X_k) \in T$. Then the probability of an error is [with $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$]

$$P_\theta\Big((X_1, \ldots, X_k) \in T,\, \theta_S \leq \max_{i \neq S} \theta_i\Big) \tag{1.1}$$

and the probability of correct inference is

$$P_\theta\Big((X_1, \ldots, X_k) \in T,\, \theta_S > \max_{i \neq S} \theta_i\Big). \tag{1.2}$$

Subject to ensuring that (1.1) $\leq \alpha$ for all θ, we want to maximize (1.2).

Let the ordered sample of the X's be denoted

$$X_S = X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(k)}.$$

In Section 2, we choose $T = \{X_S - X_{(2)} > c\}$, and prove in Theorem 1 that the supremum of the probability of error,

$$\sup_\theta P_\theta\Big(X_S - X_{(2)} > c,\, \theta_S \leq \max_{i \neq S} \theta_i\Big), \tag{1.3}$$

is attained when two θ_i are equal, and the rest are $-\infty$. Thus (1.3) reduces to

$$2P(Y_1 - Y_2 > c), \tag{1.4}$$

Received January 1985; revised May 1986.
AMS 1980 *subject classifications*. 62F03, 62F07.
Key words and phrases. Selection, retrospective hypotheses, monotone likelihood ratio.

where $Y_i = X_i - \theta_i$ are independent with density p. This result is independent of k.

Under the hypothesis of logarithmically concave distribution function (weaker than monotone likelihood ratio), Bofinger (1983) constructed $(1 - \alpha)$ confidence intervals of the form

$$\theta_S - \max_{i \neq S} \theta_i > X_S - X_{(2)} - d_k,$$

where d_k satisfies

$$P\left(\bigcup_{i=2}^{k} \{ Y_i - Y_1 > d_k \} \right) = \frac{k-1}{k} \alpha,$$

for i.i.d. $Y_i = X_i - \theta_i$. These intervals allow the inference $\theta_S > \max_{i \neq S} \theta_i$ when $X_S - X_{(2)} > d_k$. Here $d_k > c_0$ (= for $k = 2$) where c_0 is chosen to make (1.4) $= \alpha$. Also note that $d_k \to \infty$ as $k \to \infty$. (On the other hand, Bofinger's intervals provide additional information, e.g., the inference $\theta_S - \max_{i \neq S} \theta_i > -d_k$ is always warranted.)

Fabian's (1962) confidence bounds for $\min(\theta_S - \max_{i \neq S} \theta_i, 0)$ and Hsu's (1981) simultaneous intervals for $\theta_j - \max_i \theta_i$, $j = 1, \ldots, k$, allow the inference $\theta_S \geq \max_{i \neq S} \theta_i$ whenever $X_S - X_{(2)} > b_k$ where b_k satisfies

$$P\left(\bigcup_{i=2}^{k} \{ Y_i - Y_1 \} > b_k \right) = \alpha$$

with $Y_i = X_i - \theta_i$. Again, $b_k \to \infty$ as $k \to \infty$, but for k small, b_k can be smaller than c_0. The reason is the distinction between the inferences $\theta_S > \max_{i \neq S} \theta_i$, and $\theta_S \geq \max_{i \neq S} \theta_i$; a distinction made clear by examining the case $k = 2$. Suppose $X_i \sim N(\theta_i, 1)$ and $\alpha = 0.05$. We can infer $\theta_1 \neq \theta_2$ only if $|X_1 - X_2| > 1.96\sqrt{2}$ (in which case we can infer the stronger $\theta_S > \max_{i \neq S} \theta_i$). But we can infer $\theta_1 \geq \theta_2$ or $\theta_2 \geq \theta_1$ (according as $X_1 > X_2$ or $X_2 > X_1$) if $|X_1 - X_2| > 1.645\sqrt{2}$.

For the normal case with $\alpha = 0.05$, we have $c_0 = 1.96\sqrt{2}$ and $b_3 = 1.92\sqrt{2}$, $b_4 = 2.06\sqrt{2}$, so that the more conservative Fabian–Hsu inference is made less often already when $k = 4$. [The situation is similar for $\alpha = 0.01$; see Table I of Gupta (1963).]

By Theorem 1 below, the inference $\theta_S > \max_{i \neq S} \theta_i$ cannot be made whenever $X_S - X_{(2)} > c$, for any $c < c_0$, if the error probability is to be bounded by α. An additional optimality property is described in Section 3. Among procedures which regard larger values of the gaps $X_S - X_{(i)}$ as stronger evidence for $\theta_S > \max_{i \neq S} \theta_i$, the procedure which infers $\theta_S > \max_{i \neq S} \theta_i$ when $X_S - X_{(2)} > c_0$ makes this inference correctly as often as possible for each θ. This is the content of Theorem 2. The paper concludes with a mention of a possible reinterpretation of the results in terms of conditionally testing hypotheses set up after looking at the data.

2. Main results. The proof of Theorem 1 relies on the following lemma, which follows readily from an old inequality of Chebyshev [see Hardy, Littlewood and Polya (1952)].

LEMMA 1. *Let u, f and g be nonnegative functions defined on some interval of the reals. Assume that u is nondecreasing and f/g is nonincreasing—or vice versa. Then*

$$\frac{\int uf}{\int ug} \leq \frac{\int f}{\int g},$$

where the integrals are taken over the domain of the functions.

In this lemma and everywhere below we interpret $a/b \leq c/d$ as $ad \leq bc$ whenever a, b, c and d are nonnegative numbers.

The next two definitions, and Lemmas 2 and 3, can be found, essentially, in Barlow and Proschan (1975), Chapter 3.

DEFINITION. The density $p(x)$ has monotone likelihood ratio (MLR) if and only if for every $c > 0$ we have $p(x + c)/p(x)$ is nonincreasing in x.

DEFINITION. With $\bar{F}(x) := 1 - F(x)$, any distribution function $F(x)$ is said to have increasing failure rate (IFR) if and only if for every $c > 0$ we have $\bar{F}(x + c)/\bar{F}(x)$ is nonincreasing in x.

As usual we say that a random variable has MLR (or IFR) if its density (or distribution function) does. [Strictly speaking, it is the location family formed from $p(x)$ that has MLR.]

We note that MLR can be defined for integer valued random variables as well, and that the results below will apply in that case almost without change.

LEMMA 2. *If X and Y are independent, X has MLR and Y has IFR, then X + Y has IFR.*

LEMMA 3. *If X has MLR, then X has IFR; if X has MLR, then −X has MLR.*

We now state and prove the main result.

THEOREM 1. *Assume that p has MLR. Then*

$$\sup_{\theta} P_\theta\Big(X_S - X_{(2)} > c,\ \theta_S \leq \max_{i \neq S} \theta_i\Big) = 2P(Y_1 - Y_2 > c),$$

where Y_1 and Y_2 are independent with density $p(y)$.

PROOF. The probability $P_\theta(X_S - X_{(2)} > c,\ \theta_S \leq \max_{i \neq S}\theta_i)$ equals

$$\sum_{s=1}^{k} P_\theta(S = s) I\Big(\theta_s \leq \max_{i \neq s} \theta_i\Big) P_\theta(X_S - X_2 > c | S = s), \tag{2.1}$$

where $I(A) = 1$ if A is true and 0 otherwise. Now

$$(2.1) \le \sum_{s=1}^{k} P_\theta(S = s) \sup_{\substack{\theta_s \le \max \theta_i \\ i \ne s}} P_\theta\big(X_S - X_{(2)} > c | S = s\big)$$

$$= \sup_{\substack{\theta_1 \le \max \theta_i \\ i \ne 1}} P_\theta\big(X_S - X_{(2)} > c | S = 1\big) \tag{2.2}$$

by the symmetry. Now $P_\theta(X_S - X_{(2)} > c|S = 1)$ equals

$$\begin{aligned} &P_\theta(X_1 - X_i > c,\ i = 2,3,\dots,k | X_1 - X_i \ge 0,\ i = 2,3,\dots,k) \\ &\quad = P_\theta(Y_1 - Y_i \ge c + \theta_i - \theta_1, \\ &\qquad i = 2,3,\dots,k | Y_1 - Y_i \ge \theta_i - \theta_1,\ i = 2,3,\dots,k), \end{aligned} \tag{2.3}$$

where $Y_i = X_i - \theta_i$, $i = 1,\dots,k$. Now let $\gamma_i = \theta_i - \theta_1$ for $i = 2,\dots,k$, so that (2.3) equals

$$\begin{aligned} &P(Y_1 - Y_i \ge c + \gamma_i,\ i = 2,3,\dots,k | Y_1 - Y_i \ge \gamma_i,\ i = 2,3,\dots,k) \\ &\quad = \int \prod_2^k F(y - c - \gamma_i) p(y)\, dy \Big/ \int \prod_2^k F(y - \gamma_i) p(y)\, dy, \end{aligned} \tag{2.4}$$

where F is the distribution function of Y_i, that is, the distribution function corresponding to p. The null hypothesis corresponds to the case that at least one γ_i is nonnegative, and we will now evaluate the supremum of the expression in brackets in (2.4) over the null hypothesis. The sup will be attained when a single γ_i is 0, and the rest are $-\infty$.

Without loss of generality, assume $\gamma_2 \ge 0$. Then (2.4) can be rewritten

$$\begin{aligned} &\frac{\int \prod_3^k F(y - \gamma_i) F(y - \gamma_2) p(y + c)\, dy}{\int \prod_3^k F(y - \gamma_i) F(y - \gamma_2) p(y)\, dy} \\ &\quad \le \frac{\int F(y - \gamma_2) p(y + c)\, dy}{\int F(y - \gamma_2) p(y)\, dy} \end{aligned} \tag{2.5}$$

by a direct application of Lemma 1. Now (2.5) equals

$$\begin{aligned} &\int F(y - \gamma_2 - c) p(y)\, dy \Big/ \int F(y - \gamma_2) p(y)\, dy \\ &\quad = P(Y_1 - Y_2 > \gamma_2 + c)/P(Y_1 - Y_2 > \gamma_2). \end{aligned} \tag{2.6}$$

By Lemmas 2 and 3, $Y_1 - Y_2$ has IFR, so, since γ_2 is greater than or equal to 0, it follows that

$$\begin{aligned} (2.6) &\le P(Y_1 - Y_2 > c)/P(Y_1 - Y_2 > 0) \\ &= 2P(Y_1 - Y_2 > c). \end{aligned}$$

Thus we have shown that

$$P_\theta\Big(X_S - X_{(2)} > c,\ \theta_S \le \max_{i \ne S} \theta_i\Big) \le 2P(Y_1 - Y_2 > c). \tag{2.7}$$

To complete the proof, choose $\theta_1 = \theta_2$ and let $\theta_i \to -\infty$ for $i > 2$. Then (2.7) becomes an equality, and the proof is complete. □

REMARK. If the X_i have density $\sigma^{-1}p(\sigma^{-1}(x - \theta_i))$ where σ is unknown, and an independent estimator $\hat{\sigma}$ is available such that $\hat{\sigma}/\sigma$ has known density $q(t)$, then, as in the above proof, we obtain

$$\sup_{\theta, \sigma} P_{\theta, \sigma}\Big(\hat{\sigma}^{-1}\big(X_S - X_{(2)}\big) > c, \theta_S \le \max_{i \ne S} \theta_i\Big) = 2P\big(\hat{\sigma}_1(Y_1 - Y_2) > c\big),$$

where $\hat{\sigma}_1$ has density $q(t)$, and Y_1 and Y_2 are independent with density $p(y)$. Thus the inference that the selected population is best can be made exactly when a two-sample two-sided t-test based on the largest two sample means would reject the hypothesis of no difference (in the normal case).

3. An optimality property. The next result, which follows from Theorem 1, establishes an optimality property (analogous to "uniformly most powerful") for the procedure discussed above. Suppose we make the inference $\theta_S > \max_{i \ne S} \theta_i$ whenever $(X_1, X_2, \dots, X_k) \in T$, where T has a natural monotonicity property, namely: If $(x_1, \dots, x_k) \in T$ and $y_{(1)} - y_{(i)} \ge x_{(1)} - x_{(i)}$ for each $i > 1$, then $(y_1, \dots, y_k) \in T$. Thus larger gaps between the maximum and other observations are interpreted as stronger evidence that $\theta_S > \max_{i \ne S} \theta_i$.

THEOREM 2. *Among T with the above monotonicity property satisfying*

$$\sup_{\theta} P_\theta\Big((X_1, \dots, X_k) \in T, \theta_S \le \max_{i \ne S} \theta_i\Big) \le \alpha,$$

we have

$$P_\theta\Big(X_S - X_{(2)} > c_0, \theta_S > \max_{i \ne S} \theta_i\Big) \ge P_\theta\Big((X_1, \dots, X_k) \in T, \theta_S > \max_{i \ne S} \theta_i\Big), \quad \forall \theta,$$

where c_0 is chosen to satisfy $2P(Y_1 - Y_2 > c_0) = \alpha$.

PROOF. Under the hypothesis of monotonicity, $T = \{x: g(x_{(1)} - x_{(2)}, x_{(1)} - x_{(3)}, \dots, x_{(1)} - x_{(k)}) \ge 0\}$ for some g nondecreasing in each argument. With $g(t) = \sup_{u_3, \dots, u_k} g(t, u_3, \dots, u_k)$, let

$$T^* = \big\{x: g\big(x_{(1)} - x_{(2)}\big) \ge 0\big\}.$$

Then $T \subset T^*$ and $T^* = \{x: x_{(1)} - x_{(2)} > a\}$ (possibly $\ge a$; WLOG take $a \ge 0$). Thus $P_\theta(X \in T, \theta_S \le \max_{i \ne S} \theta_i) \le P_\theta(X \in T^*, \theta_S \le \max_{i \ne S} \theta_i)$. For $\theta = (0, 0, m, m, \dots, m)$ however, as $m \to -\infty$ both $P_\theta(X \in T, \theta_S \le \max_{i \ne S} \theta_i)$ and $P_\theta(X \in T^*, \theta_S \le \max_{i \ne S} \theta_i)$ converge to $2P(Y_1 - Y_2 > a)$. Since by Theorem 1

$$\sup_\theta P_\theta\Big(X \in T^*, \theta_S \le \max_{i \ne S} \theta_i\Big) = 2P(Y_1 - Y_2 > a),$$

it follows that

$$\sup_\theta P_\theta\Big(X \in T, \theta_S \le \max_{i \ne S} \theta_i\Big) = \sup_\theta P_\theta\Big(X \in T^*, \theta_S \le \max_{i \ne S} \theta_i\Big)$$

and hence that $a \ge c_0$. Thus $T \subset \{x: x_{(1)} - x_{(2)} \ge c_0\}$ and the result follows. □

Both Theorems 1 and 2 have conditional analogues. Because of the conditional technique of the proof of Theorem 1, the procedure can be interpreted as a test of a *retrospective* hypothesis,

$$H_0(s)\colon \theta_s \le \max_{i \ne s} \theta_i,$$

formed after observing $S = s$. Theorem 1 calculates the "size," or maximum probability of type I error conditional on $S = s$, for the test which rejects when $X_s - X_{(2)} > c$. The conditional version of Theorem 2 asserts that

$$P_\theta((X_1, \dots, X_k) \in T | S = s) \le P_\theta\Big(X_s - \max_{i \ne s} X_i > c_0 | S = s\Big),$$

for all $\theta \notin H_0(s)$ if T is monotone with

$$\sup_{\theta \in H_0(s)} P_\theta((X_1, \dots, X_k) \in T | S = s) \le \alpha.$$

REFERENCES

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing.* Holt, Rinehart and Winston, New York.

BOFINGER, E. (1983). Multiple comparisons and selection. *Austral. J. Statist.* **25** 198–207.

FABIAN, V. (1962). On multiple decision methods for ranking population means. *Ann. Math. Statist.* **33** 248–254.

GUPTA, S. S. (1963). Probability integrals of the multivariate normal and multivariate *t*. *Ann. Math. Statist.* **34** 792–828.

HARDY, G. H., LITTLEWOOD, J. E. and POLYA, G. (1952). *Inequalities*, 2nd ed. Cambridge Univ. Press.

HSU, J. C. (1981). Simultaneous confidence intervals for all distances from the "best." *Ann. Statist.* **9** 1026–1034.

DEPARTMENT OF MATHEMATICS
NORTHEASTERN UNIVERSITY
360 HUNTINGTON AVENUE
BOSTON, MASSACHUSETTS 02115