A BAYESIAN APPROACH TO SOME BEST POPULATION PROBLEMS[1]

By Irwin Guttman and George C. Tiao

University of Wisconsin

1. Introduction and summary. There have been several papers recently in the literature devoted to the subject of selecting a "best" population—see for example Gupta and Sobel (1962), Guttman (1961) and others. In these papers, the problem was analyzed from the sampling theory point of view. Except for some simple cases, this approach frequently leads to the problem of eliminating nuisance parameters. And, unless rather strong assumptions about certain of the parameters involved are made, the problem usually becomes intractable. In this paper, we consider certain best population problems adopting a Bayesian approach. As is well known, the modern Bayesian approach provides a logical framework for decision making under uncertainty—Savage (1954), Luce and Raiffa (1957), etc. Indeed, the adoption of this approach in the best population problem leads to intuitively satisfactory decision procedures in the presence of nuisance parameters.

We consider a collection of k populations $\Pi_1, \cdots, \Pi_i, \cdots, \Pi_k$, where Π_i is distributed with probability density function $f(y \mid \theta_i)$ and θ_i may be vector-valued. Associated with the populations is a utility function $U(\theta_i)$. The *best population* is defined to be the one with the largest value among the $U(\theta_i)$. The Bayesian statistical decision procedure for choosing the best population is of course dictated by the principle of maximizing expected utility.

Now it usually is the case that the experimenter's interest focuses on a specific *criterion* $h_i = g(\theta_i)$ where g is known. For example, $g(\theta_i)$ might simply be the mean or the reciprocal of the variance of the ith population, $i = 1, \cdots, k$. It would then be natural for the experimenter to regard his utility function U as a function of h_i, i.e., $U = U(h_i)$.

In this paper, we will be mainly concerned with the case in which the criterion $g(\theta)$ is defined as

$$h = g(\theta) = \int_{a_1}^{a_2} f(y \mid \theta)\, dy,$$

for specified a_1 and a_2. The interval (a_1, a_2) is sometimes referred to as a tolerance interval and the quantity h is called the coverage of this interval. Considerations of tolerance intervals and their coverages frequently arise in engineering application [see for example Bowker and Lieberman (1959), and Ostle (1963)]. For instance, in the manufacture of a certain type of rivet, the product will only be of use if the diameter of the head of the rivet measures

Received 19 March 1963; revised 8 December 1963.

[1] Work supported in part by the Office of Naval Research.

between certain specified limits. Or in the assembling of "stable" amplifiers, certain of the electronic tubes used in the amplifier must have transconductances that lie within specified limits. Hence, it is important for the manufacturer of these products to know what percentage of the items produced meet these required specifications. Suppose the manufacturer can choose among k different processes to produce his product, and suppose further that his utility of the ith population $U(h_i) = h_i$ (that is, other considerations such as cost can be ignored). Naturally, he will wish to choose that process which gives the largest coverage to the specification interval. Of course, we are including the case that one process be capable of k independent modifications. As an example, in the manufacture of precision items on a lathe, there may be a choice of different cutting tools, different coolants, and various speeds of rotation.

Our interest in this type of problem is also motivated by the fact that the coverage criterion h as defined above is usually a complicated function of the parameter(s) θ. One might thus be led to suspect that it would in general be rather difficult to obtain optimal decision procedures for choosing the best population. In fact, one of the present authors—Guttman (1961)—has considered this problem within the sampling theory framework for the cases that f takes the form of a normal density and that of an exponential density. It is evident from this earlier work that one would indeed encounter considerable mathematical difficulties unless certain specific assumptions about the scale parameters of the populations were made. We shall demonstrate, however, that no such difficulty exists in the Bayesian formulation and our analysis leads to results which seem intuitively satisfactory.

In adopting a Bayesian approach, suppose we denote the prior distribution of the parameters of the k populations under consideration as $p(\theta_1, \cdots, \theta_k)$. Then, for given independent samples from these populations, say $(\mathbf{y}_1, \cdots, \mathbf{y}_k)$, we can obtain the posterior distribution of these parameters. As we are interested in $(h_1, \cdots, h_k)$, which are themselves functions of the parameters, we may, in principle at least, determine their joint posterior distribution. This joint posterior distribution summarizes all the relevant information about $(h_1, \cdots, h_k)$.

We assume throughout this paper that the population parameters $(\theta_1, \cdots, \theta_k)$ are locally independent a priori. This means that in the region in which the likelihood is appreciable, the joint distribution $p(\theta_1, \cdots, \theta_k)$ can be written approximately as the product $p_1(\theta_1) \cdots p_k(\theta_k)$. Such an assumption will be appropriate in situations where the prior distribution of the parameters is diffuse and gently changing over a wide region—see for example the discussion in Savage et al. (1962). It is clear that this independence assumption together with the assumption about the independence of samples implies that the h's are locally independent a posteriori and hence can be analyzed individually.

In Sections 2 and 3, we assume that the utility function $U(h_i)$ is h_i for all i, and therefore analyze the problem by comparing the expected values of the posterior distribution of the coverages. The populations involved are assumed to be (1) normal and (2) exponential. The posterior distributions and the moments of the coverages for these populations are derived in Sections 4 and 5 where other types of utility functions are discussed. In Section 6 we briefly discuss some different types of best population problems.

2. The case of the normal density. In this and the following section, the coverage of a given interval (a_1, a_2) is taken as the utility function of the experimenter for a particular population. The decision procedure is then simply to choose as the best population the one with the largest posterior expectation of the coverage.

In this section, we assume that the k populations are normally distributed. The coverage criterion h is then

$$h = g(\mu, \sigma) = \int_{a_1}^{a_2} \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} dy. \tag{2.0}$$

Consider a sample $\mathbf{y} = (y_1, \cdots, y_n)$ taken from one of the populations. The use of n here denotes the size of a sample from a particular population; but we hasten to point out that we do not assume that the sample sizes are common. The likelihood function is

$$\begin{aligned} l(\mu, \sigma \mid \mathbf{y}) &\propto \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(y_j-\mu)^2\right\} \\ &\propto \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\{ns^2 + n(\bar{y}-\mu)^2\}\right\} \end{aligned} \tag{2.1}$$

where $\bar{y}$ and s^2 are the maximum likelihood estimators of μ and σ^2 respectively.

Assume that we are in a situation where little is known a priori about the values of (μ, σ). In other words, we are saying that the information we have concerning (μ, σ) comes primarily from the sample. We may then adopt the approach used by Jeffreys (1948), Savage (1962), and Box and Tiao (1962), and assume that μ and $\log \sigma$ are independent and locally uniformly distributed a priori. That is,

$$p(\mu) \propto k_1, \qquad p(\log \sigma) \propto k_2 \quad \text{or} \quad p(\sigma) \propto \sigma^{-1}. \tag{2.2}$$

Using (2.2), the joint posterior distribution of (μ, σ) is:

$$p(\mu, \sigma \mid \mathbf{y}) = p_1(\sigma \mid s)p_2(\mu \mid \sigma, \bar{y}) \tag{2.3}$$

where

$$p_1(\sigma \mid s) = 2\{\Gamma(\tfrac{1}{2}(n-1))\}^{-1}\{(\tfrac{1}{2}ns^2)^{\frac{1}{2}(n-1)}\}\sigma^{-n} \exp\{-ns^2/2\sigma^2\} \tag{2.4a}$$

and

$$p_2(\mu \mid \sigma, \bar{y}) = \{n/2\pi\sigma^2\}^{\frac{1}{2}} \exp\{(-n/2\sigma^2)(\bar{y}-\mu)^2\}, \tag{2.4b}$$

a result first given by Jeffreys (1948) (see page 121). We note from (2.3) that the adoption of the prior distributions (2.2) amounts to saying that the joint posterior distribution of μ and $\log \sigma$ is proportional to the likelihood function (2.1), and therefore, can be regarded as "being approximated" by this likelihood.

We now derive the posterior mean of h for a given population. From (2.3), the a posteriori expectation of h is:

$$\begin{aligned} E(h \mid \mathbf{y}) &= \int_{-\infty}^{\infty}\int_{0}^{\infty} hp(\mu, \sigma \mid \mathbf{y})\, d\sigma\, d\mu \\ &= \int_{-\infty}^{\infty}\int_{0}^{\infty} \Pr\{a_1 < X < a_2 \mid \mu, \sigma\} p(\mu, \sigma \mid \mathbf{y})\, d\sigma\, d\mu \end{aligned} \tag{2.5}$$

where we have expressed h in the integrand as the conditional probability that an observation X from a normal population falls in (a_1, a_2) for given (μ, σ), while $p(\mu, \sigma \mid \mathbf{y})$ is the posterior distribution of (μ, σ) in (2.3). It is clear that this expectation may be written as

$$E(h \mid \mathbf{y}) = \int_{a_1}^{a_2} p(x \mid \mathbf{y})\, dx, \tag{2.6}$$

where

$$p(x \mid \mathbf{y}) = \int_{-\infty}^{\infty}\int_{0}^{\infty} f(x \mid \mu, \sigma) p(\mu, \sigma \mid \mathbf{y})\, d\sigma\, d\mu \tag{2.7}$$

is the marginal posterior distribution of a "future" observation X given an initial sample $\mathbf{y}$. For the normal case then, (2.7) is:

$$p(x \mid \mathbf{y}) = \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}(n-1))}[(n+1)s^2]^{-\frac{1}{2}}\left\{1 + \frac{(\bar{y} - x)^2}{(n+1)s^2}\right\}^{-\frac{1}{2}n} \tag{2.8}$$

which is seen to be related to the Student-t distribution, a result first derived by Jeffreys (1948) (see page 126). Thus, we have the following remarkably simple result for the posterior expectation:

$$E(h \mid \mathbf{y}) = F_{n-1}(t_2) - F_{n-1}(t_1) \tag{2.9}$$

where F is the cumulative distribution function of the Student-t variable with $(n-1)$ degrees of freedom and

$$t_1 = \left(\frac{n-1}{n+1}\right)^{\frac{1}{2}}\left(\frac{a_1 - \bar{y}}{s}\right), \qquad t_2 = \left(\frac{n-1}{n+1}\right)^{\frac{1}{2}}\left(\frac{a_2 - \bar{y}}{s}\right). \tag{2.10}$$

Similarly, if $a_1 = -\infty$, we find that

$$E(h \mid \mathbf{y}) = F_{n-1}(t_2). \tag{2.11}$$

We recall that here we are regarding the utility as the coverage. Thus, once having computed these posterior expectations for each of the populations, the experimenter arrives at a decision in accordance with the principle of maximizing expected utility by simply choosing as the best population the one with the largest value of these expectations.

In the case where $a_1 = -\infty$ (this will be called the one-sided case), we note that if the sample sizes are the same for all samples, then this selection procedure is equivalent to deciding that the best population is the one with the largest value among the k values of $(a_2 - \bar{y})/s$.

This is an intuitively pleasing result for the following reason. We are interested in the best population, that is, for this case, the population which has largest value of the coverage

$$h = g(\mu, \sigma) = \int_{-\infty}^{a_2} \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx$$
$$= \int_{-\infty}^{(a_2-\mu)/\sigma} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left\{-\frac{t^2}{2}\right\} dt.$$

Since this is a monotone increasing function of the upper limit, say $\tau = (a_2 - \mu)/\sigma$, the best population is then the one with the largest value of the $k\tau$'s. It is intuitively evident that an estimate of τ should be based on $(a_2 - \bar{y})/s$ and hence, that the largest value of $(a_2 - \bar{y})/s$ should be indicative of the best population.

We now turn to sampling from the exponential distribution, employing similar analysis discussed in this section.

3. The case of the exponential density. The probability density function of an exponential population with arbitrary origin, say η, has the form

$$(3.1)\qquad \begin{aligned} f(y \mid \eta, \sigma) &= \sigma^{-1} \exp\{-(y-\eta)/\sigma\}, && y \geqq \eta \\ &= 0 && \text{otherwise,} \end{aligned}$$

where η can be regarded as a location parameter and σ is, of course, a scale parameter. The coverage h is

$$(3.2)\qquad h = g(\eta, \sigma) = \int_{a_1}^{a_2} f(y \mid \eta, \sigma)\, dy.$$

Let a sample $\mathbf{y}$ of size n be taken from each of the k populations and denote the smallest observation in a particular sample by $y*$. The likelihood function is given by:

$$(3.3)\qquad \begin{aligned} l(\eta, \sigma \mid \mathbf{y}) &= \sigma^{-n} \exp\{-\sigma^{-1}\textstyle\sum (y_i - \eta)\} \\ &= \sigma^{-n} \exp\{-\sigma^{-1}[(n-1)w + n(y* - \eta)]\} \end{aligned}$$

where $w = \sum (y_i - y*)/(n-1)$. It is to be noted that the likelihood function (3.3) is a monotonic increasing function of η in the interval $(-\infty, y*)$ and vanishes outside this interval. As in Section 2, we assume that the prior distributions for the location parameter η and $\log \sigma$ are locally uniform. The joint posterior distribution of η and σ then takes the form:

$$(3.4)\qquad p(\eta, \sigma \mid \mathbf{y}) = p_3(\sigma \mid w) p_4(\eta \mid \sigma, y*)$$

where

$$p_3(\sigma \mid w) = \frac{\{(n-1)w\}^{(n-1)}}{\Gamma(n-1)} \sigma^{-n} \exp\left\{-\frac{(n-1)w}{\sigma}\right\}, \qquad \sigma > 0 \tag{3.5a}$$

and

$$p_4(\eta \mid \sigma, y*) = (n/\sigma) \exp\{-(n/\sigma)(y* - \eta)\}, \qquad \eta < y* . \tag{3.5b}$$

As in the preceding section, we base our decision procedure on the posterior expectation of h,

$$E(h \mid \mathbf{y}) = \int_R hp(\eta, \sigma \mid \mathbf{y})\, d\eta\, d\sigma, \tag{3.6}$$

where R denotes the appropriate range of integration. Proceeding as in Section 2, we first evaluate the marginal distribution of a future observation X. Using (3.1) and (3.4), the joint distribution of (X, η, σ) is:

$$\begin{aligned} p(x, \eta, \sigma \mid \mathbf{y}) &= f(x \mid \eta, \sigma)p(\eta, \sigma \mid \mathbf{y}) \\ &= \frac{n[(n-1)w]^{(n-1)}}{\Gamma(n-1)} \sigma^{-(n+2)} \\ &\quad \cdot \exp\left\{-\frac{1}{\sigma}[(n-1)w + n(y* - \eta) + (x - \eta)]\right\} \end{aligned} \tag{3.7}$$

if $\sigma > 0$ and $\eta < \min(y*, x)$, and vanishes otherwise. Integrating out η and σ, we have that

$$\begin{aligned} p(x \mid \mathbf{y}) &= [n/(n+1)w][1 + n(y* - x)/(n-1)w]^{-n} && \text{if } x < y* \\ &= [n/(n+1)w][1 + (x - y*)/(n-1)w]^{-n} && \text{if } x > y* . \end{aligned} \tag{3.8}$$

Thus, the posterior expectation of h is:

$$E(h \mid \mathbf{y}) = \int_{a_1}^{a_2} p(x \mid \mathbf{y})\, dx. \tag{3.9}$$

For the one-sided case, that is $a_1 = -\infty$ and $a_2 = a$, we find that

$$E(h \mid \mathbf{y}) = (n+1)^{-1}[1 + n(y* - a)/(n-1)w]^{-(n-1)} \quad \text{if } y* > a \tag{3.10a}$$

and

$$E(h \mid \mathbf{y}) = 1 - (n/(n+1))[1 + (a - y*)/(n-1)w]^{-(n-1)} \quad \text{if } y* < a. \tag{3.10b}$$

We note that the above results are again intuitively pleasing. To begin with, the coverage h for the one-sided case is easily seen to be a monotone nondecreasing function of $(a - \eta)/\sigma$. Hence, the best population is that with the largest value of $(a - \eta)/\sigma$. It seems natural, therefore, that a selection procedure for the best population should be based on $(a - y*)/w$. In the special case that the sample sizes are common, this leads us to choose that population which yields the sample with the largest value of $(a - y*)/w$. The above reasoning is borne out in expressions (3.10a) and (3.10b) since both are monotonic increasing functions of $(a - y*)/w$.

In addition, we note that if $y* > a$, that is, no sample elements fall in the interval of interest $(-\infty, a)$, then this is a clear indication that the coverage should be of a small magnitude. And in fact, examination of (3.10a) shows that the posterior expectation is of smaller order than n^{-1}.

We remark that the two-sided case may be analysed in a similar way to the above upon using expression (3.8).

4. Posterior distribution of the coverage for normal populations. In this and the following section, we discuss the more general situation in which the experimenter's utility is some function of h, say $U(h)$. Such a situation may arise for example when the experimenter wishes to take into consideration other factors such as cost of sampling, amount of time needed in adjustment of equipment, etc., in addition to h itself. When the experimenter is in such a situation, knowledge of the posterior distribution of h may considerably simplify his subsequent analysis in deciding which is the best population. For example, suppose that $U(h)$ is nearly linear in a certain region where the posterior distribution say $p(h \mid \mathbf{y})$ is sharply concentrated. Then, even though $U(h)$ is not linear for all h in [0, 1], still the posterior expectation of $U(h)$ can, to a good degree of approximation, be determined simply by knowing the posterior expectation of h.

Or, it may be that in the region in which the posterior distribution of h is appreciable, $U(h)$ can be very well represented by a polynomial function of h. So in this case, the expectation of the utility can be evaluated using moments of h. The distribution of h may also be of interest in its own right. For instance, referring to the rivet example of Section 1, suppose that the break even point of a process is accomplished if and only if the proportion of acceptable rivets is greater than a known percentage, say $100\gamma\%$. Suppose further that the manufacturer has already decided which is the best process, perhaps by the method of Sections 2 and 3. He may still want to know for this "best" process the probability of the coverage h exceeding γ. In fact, if this probability is too low, the manufacturer might wish to consider other alternatives. Similar examples could be cited.

Now it is well known that the distribution of a bounded variable such as h is completely determined by its moments. For a normal population, the rth moment of h is

$$E(h^r \mid \mathbf{y}) = \int_{-\infty}^{\infty}\int_0^{\infty} h^r p(\mu, \sigma \mid \mathbf{y})\, d\sigma\, d\mu \tag{4.1}$$

where h and $p(\mu, \sigma \mid \mathbf{y})$ are given respectively in (2.0) and (2.3).

Following the argument of Section 2, (4.1) may be written as

$$E(h^r \mid \mathbf{y}) = \int_{a_1}^{a_2} \cdots \int_{a_1}^{a_2} p(x_1, \cdots, x_r \mid \mathbf{y})\, dx_1 \cdots dx_r \tag{4.2}$$

where

$$p(x_1, \cdots, x_r \mid \mathbf{y}) = \int_{-\infty}^{\infty}\int_{0}^{\infty} \prod_{i=1}^{r} \frac{1}{(2\pi\sigma)^{\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\} p(\mu, \sigma \mid \mathbf{y})\, d\sigma\, d\mu, \tag{4.3}$$

which is the distribution of r future observations, given an initial sample $\mathbf{y}$. This distribution was derived by Raiffa and Schlaifer (1961), p. 345 and takes the form:

$$p(x_1, \cdots, x_r \mid \mathbf{y}) = k(ns^2)^{\frac{1}{2}(n-2)}[ns^2 + \sum_{i,j} c_{ij}(x_i - \bar{y})(x_j - \bar{y})]^{-\frac{1}{2}(n+r-1)} \tag{4.4}$$

with

$$k = \frac{\Gamma(\frac{1}{2}(n + r - 1))}{\Gamma(\frac{1}{2}(n - 1))(\pi^{\frac{1}{2}})^r}\left(\frac{n}{n+r}\right)^{\frac{1}{2}}, \quad \text{and} \quad c_{ii} = 1 - \frac{1}{n+r},$$

$$c_{ij} = -\frac{1}{n+r}, \quad i \neq j.$$

Making the transformation

$$\gamma_i = \left(\frac{n-1}{n+1}\right)^{\frac{1}{2}} \frac{x_i - \bar{y}}{s},$$

we can write (4.2) as

$$E(h^r \mid \mathbf{y}) = k\left[\frac{n+1}{n(n-1)}\right]^{\frac{1}{2}r} \cdot \int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} \left[1 + \frac{\sum_{ij} d_{ij}\gamma_i\gamma_j}{n-1}\right]^{-\frac{1}{2}(n+r-1)} d\gamma_1 \cdots d\gamma_r \tag{4.5}$$

where t_1 and t_2 are given in (2.10) and $d_{ij} = ((n + 1)/n)c_{ij}$.

The rth moment (4.5) is seen to be a multivariate-t integral with $(n - 1)$ degrees of freedom and variance-covariance matrix $\{d_{ij}\}^{-1}$, see Dunnett and Sobel (1954).

To determine the posterior distribution of h for a normal population, one may proceed by making a suitable transformation in $p(\mu, \sigma \mid \mathbf{y})$. In what follows we illustrate the determination of the posterior distribution of h for the one-sided case.

For a normal population, the joint posterior distribution of μ and σ is given by (2.3). We now make the transformation

$$h = g(\mu, \sigma) = \int_{\infty}^{(a_2-\mu)/\sigma} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left\{-\frac{t^2}{2}\right\} dt = \Phi\left(\frac{a_2 - \mu}{\sigma}\right) \tag{4.6}$$

$$v = \sigma.$$

The absolute value of the Jacobian of the transformation is

$$|J| = (2\pi)^{\frac{1}{2}}\sigma \exp\{(a_2 - \mu)^2/2\sigma^2\}. \tag{4.7}$$

Using (4.6) and (4.7), and upon integrating over the range of v, we find after some simplification that the (marginal) posterior distribution of h is:

$$p(h \mid \mathbf{y}) = k \int_c^\infty \frac{1}{\Gamma(n-1)} (t - c)^{n-2} e^{-\frac{1}{2}t^2}\, dt \tag{4.8}$$

where

$$k = \frac{n^{\frac{1}{2}}\Gamma(n-1) \exp\{-\frac{1}{2}m^2[n/(s^2 + b^2) - 1]\}}{2^{\frac{1}{2}n}\,\Gamma(\frac{1}{2}(n-1))(1 + b^2/s^2)^{\frac{1}{2}(n-1)}}$$

$$m = \Phi^{-1}(h),\ b = \bar{y} - a_2 \quad \text{and} \quad c = n^{\frac{1}{2}}bm/(s^2 + b^2)^{\frac{1}{2}}.$$

The integral in (4.8) is known as the $Hh_{n-2}(c)$ function, the properties and uses of which are discussed in an extensive introduction by R. A. Fisher to the 1931 printing of the reports of the British Association for the Advancement of Science (see Greenwood and Hartley, (1962) p. 38). Thus, making use of both a table of the $Hh_{n-2}(c)$ function, and those standard normal tables which would permit accurate inverse interpolation to obtain the value of m, one would be able to calculate the density $p(h \mid \mathbf{y})$.

5. Posterior distribution of the coverage for exponential populations. In this section we determine the moments and the posterior distribution of the coverage h when sampling from the exponential population and when:

(i) $h = \int_{-\infty}^{a} f(y \mid \eta, \sigma)\, dy$ where f is given in (3.1), and

(ii) $y* < a$ with $y*$ denoting the smallest observation.

Paralleling the development in the previous section, and using the posterior distribution of (η, σ) in (3.4), the rth moment of h is

$$E(h^r \mid \mathbf{y}) = \sum_{l=0}^{r} \binom{r}{l} (-1)^l \left(\frac{n}{n+l}\right) \left\{1 + \frac{l(a - y*)}{(n-1)w}\right\}^{-(n-1)}. \tag{5.1}$$

For the posterior distribution of h, we first note that when $y*$ is less than "a", the coverage h is $1 - \exp(-(a - \eta)/\sigma)$. We now make the transformation

$$h = g(\eta, \sigma) = 1 - \exp\{-(a - \eta)/\sigma\}, \qquad v = \sigma. \tag{5.2}$$

The absolute value of the Jacobian of the transformation is $|J| = v/(1 - h)$. We find that the (marginal) posterior distribution of h is:

$$p(h \mid \mathbf{y}) = \frac{n\{(n-1)w\}^{n-1}}{\Gamma(n-1)} (1 - h)^{n-1} \int_{\alpha(h)}^\infty v^{-n} \exp\left\{-\frac{n}{v}\{\bar{y} - a\}\right\} dv \tag{5.3}$$

where $\alpha(h) = -(a - y*)/\log(1 - h)$.

We note that for $\bar{y} - a > 0$, the integral part of (5.3) is of course related to the incomplete gamma integral, as may be seen by making the transformation $x = 1/v$. For $\bar{y} - a < 0$, the integral is no longer related to the incomplete gamma integral. However, if we again make the transformation $x = 1/v$, the resulting form for the integral together with the moderating factor $(1 - h)^{n-1}$ will insure the convergence of the density $p(h \mid \mathbf{y})$ for all values of h.

6. Some other best population problems. In previous sections, we have discussed the case where the criterion function h is the coverage of a certain given interval. However, as explained in Section 1, the criterion function $h = g(\theta)$ may be chosen to suit the particular interest of the experimenter. In fact, as we have seen, the coverage criterion is a rather complicated function of the population parameters. In this section, we turn to some other best population problems, where the definition of "best" used, involves criterions which are simple functions of either a location or a scale parameter of the population distribution. Also, we restrict ourselves to the case where the utility can be taken as the criterion itself.

We will first discuss the case where the k populations are normally distributed and the best population is defined to be the one with the largest mean. This problem has been considered from the sampling theory point of view by Bechhofer (1954). We will then discuss the same situation but where the best population is now taken to be the one with the smallest variance. Here again, for consideration of this problem from the sampling theory point of view, we refer the reader to Gupta and Sobel (1962). We will then turn to the exponential population and discuss this problem using similar criterions.

For normal populations, first suppose that $h = g(\mu, \sigma) = \mu$. From the joint posterior distribution of (μ, σ) given in (2.3), it is known (Jeffreys, 1948) that the marginal posterior distribution of μ is

$$(6.1) \qquad p(\mu \mid \bar{y}, s) = \frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}(n-1))} s^{-1} \left\{1 + \left(\frac{\mu - \bar{y}}{s}\right)^2\right\}^{-\frac{1}{2}n}$$

where we again note that $\bar{y}$ and s are the maximum likelihood estimators of μ and σ, respectively. From (6.1), it is easy to see that the quantity $t_{n-1} = (n - 1)^{\frac{1}{2}}(\mu - \bar{y})/s$ has the Student-t distribution with $n - 1$ degrees of freedom.

Clearly, the mean of the distribution in (6.1) is $\bar{y}$. Since we have assumed that the experimenter's utility function is h itself, he will make the decision that the best population is the one with the largest sample mean.

Now consider the normal case with $h = g(\mu, \sigma) = 1/\sigma$. Again from expression (2.3), we find that the posterior distribution of the quantity $(ns)^{\frac{1}{2}}/\sigma$ follows the χ_{n-1} distribution. It is then easy to verify that the posterior expectation of h is:

$$(6.2) \qquad E(h \mid \mathbf{y}) = 2^{\frac{1}{2}}\Gamma(\tfrac{1}{2}n)/n^{\frac{1}{2}}\Gamma(\tfrac{1}{2}(n-1))s.$$

Turning now to exponential populations—see expression (3.1), we now discuss the situation when $h = g(\eta, \sigma) = \eta$. This situation may arise, for example, in the manufacture of certain electronic components. In some instances, the life times of these parts minus a required minimum standard life time follow an exponential population, with a location parameter η and a scale parameter σ.

When k such processes are available, then clearly the one with the largest (minimum life time) η is desired. On the usual assumptions for the prior distributions of η and σ—see Section 3, it is easy to see that the joint posterior distribution of η and σ takes the form given in expression (3.4). Integrating out σ, the (marginal) posterior distribution of η is

$$p(\eta \mid \mathbf{y}) = nw^{-1}\{1 + n(y* - \eta)/(n-1)w\}^{-n}, \qquad \eta < y* \tag{6.3}$$

where $y*$ and w are as defined in Section 3. Thus, we find that the posterior mean is

$$E(h \mid \mathbf{y}) = y* - (n-1)w/n(n-2). \tag{6.4}$$

We note that the quantity $V = n(y* - \eta)/(n-1)w$ is distributed as a beta variable of the second kind with parameters $(2, n-2)$.

Lastly, for exponential populations, suppose now $h = g(\eta, \sigma) = 1/\sigma$. We have from (3.5a) that the posterior distribution of $(n-1)w/\sigma$ is a gamma variable with parameter $(n-1)$. Hence the posterior mean of h is

$$E(h \mid \mathbf{y}) = 1/w. \tag{6.5}$$

7. Acknowledgment. The authors would like to express their appreciation to Professor L. J. Savage, for his many valued comments and suggestions on various aspects of this paper.

REFERENCES

BECHHOFER, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25** 16–29.

BECHHOFER, R. E. and SOBEL, M. (1954). A single-sample multiple decision procedure for ranking variances of normal populations. *Ann. Math. Statist.* **25** 273–289.

BOWKER, A. H. and LIEBERMAN, G. J. (1959). *Engineering Statistics*. Prentice Hall.

BOX, G. E. P. and TIAO, G. C. (1962). A further look at robustness via Bayes' theorem. *Biometrika* **49** 419–432.

DUNNETT, C. W. and SOBEL, M. (1954). A bivariate generalization of Student's t-distribution, with tables for certain special cases. *Biometrika* **41** 153–169.

GREENWOOD, J. A. and HARTLEY, H. O. (1962). *Guide to Tables in Mathematical Statistics*. Princeton Univ. Press.

GUPTA, S. S. and SOBEL, M. (1962). On selecting a subset containing the population with the smallest variance. *Biometrika* **49** 495–507.

GUTTMAN, IRWIN (1961). Best populations and tolerance regions. *Ann. Inst. Statist. Math.* **13** 9–26.

JEFFREYS, H. (1948). *Theory of Probability*, (2nd edition). Clarendon, Oxford.

LUCE, R. D. and RAIFFA, H. (1957). *Games and Decisions*. Wiley, New York.

OSTLE, BERNARD (1963). *Statistics in Research*, (2nd edition). Iowa Press.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Harvard Univ. Press.

SAVAGE, L. J. (1954). *The Foundations of Statistics*. Wiley, New York.

SAVAGE, L. J., et al. (1962). *The Foundations of Statistical Inference*. Methuen Monograph.