The Annals of Statistics
1976, Vol. 4, No. 1, 1-10

A CLASS OF UTILITY FUNCTIONS

By D. V. Lindley

The University of Iowa and University College London

In the case of the exponential family of distributions it is well known that the use of a prior distribution belonging to the natural conjugate family substantially simplifies the analysis whilst often being realistic in applications. The present paper explores the related idea of a conjugate family of utilities, and generally the notion of choosing a utility structure that is suitably "matched" to the probability structure and, at the same time, realistic in application.

1. Introduction. In many statistical problems little attention is paid to the precise form of the loss function. In point estimation it is usual to employ quadratic loss, whilst in hypothesis testing a piece-wise, constant function is often preferred. In the present paper we point out that there are other loss functions that can be used without increasing the analytic complexity unduly. We prefer to work in terms of utilities rather than losses, because the former are more readily interpreted in operational terms. We study the form of these utility functions and explore some of the consequences of using them. It is hoped that the availability of a range of amenable utility functions will mean that statisticians will give more consideration to the utility structure of their problems than they have done hitherto.

2. Conjugate utilities. Apart from a few remarks in Section 7, attention is confined to the case of a real variable, x, having a distribution depending on a single real parameter, θ. Until Section 6, this distribution is supposed to be a member of the exponential family. This family can conveniently be described by a density, with respect to some suitable dominating measure, and for a suitable parameterization, proportional to $e^{x\theta}H(x)$ for some nonnegative function $H(x)$. Writing

$$G(\theta)^{-1} = \int e^{x\theta}H(x)\, d\mu(x) \tag{2.1}$$

for all θ for which the integral is finite, the density of x, given θ, is

$$p(x \mid \theta) = e^{x\theta}H(x)G(\theta)\,. \tag{2.2}$$

(In (2.1), $d\mu(x)$ refers to the dominating measure.) The natural conjugate family of densities for θ, Raiffa and Schlaifer (1961) (see also Wetherill (1961)), then has density proportional to $e^{x_0\theta}G(\theta)^{n_0}$ for suitable x_0 and n_0. These last two quantities will be referred to as hyperparameters. Defining

$$K(n_0, x_0)^{-1} = \int e^{x_0\theta}G(\theta)^{n_0}\, d\theta\,, \tag{2.3}$$

Received December 1974; revised May 1975.

AMS 1970 *subject classification.* Primary 62C05.

Key words and phrases. Utility functions, exponential family, conjugate family, hyperparameters, maximum likelihood.

the integral being over the relevant θ-values, the conjugate density of θ, given n_0 and x_0, is

$$p(\theta \mid n_0, x_0) = e^{x_0\theta}G(\theta)^{n_0}K(n_0, x_0)\,. \tag{2.4}$$

The basic result in this theory is that if $(x_1, x_2, \cdots, x_n) = \mathbf{x}$ is a random sample from (2.2), and if the distribution of θ prior to the sample is given by (2.4), then the similar distribution after the sample is to hand is

$$p(\theta \mid n_0, x_0, \mathbf{x}) = e^{\theta \sum x_i}G(\theta)^{n+n_0}K(n + n_0, \textstyle\sum x_i)\,, \tag{2.5}$$

where the summations are from zero (not one) to n. This is the same type of distribution as before, (2.4), but with the hyperparameters changed from x_0 and n_0 to $\sum x_i$ and $n + n_0$. It may therefore be written, $p(\theta \mid n + n_0, \sum x_i)$.

In decision problems it is necessary to go a stage further and, besides a distribution for θ, to introduce a utility function $U(d, \theta)$ for the utility of decision d when the parameter value θ obtains. The optimum decision is that having greatest expected utility, the expectation being with respect to θ, having the distribution (2.5). Just as the posterior analysis is simplified by using distributions for θ that fit nicely with those for x, so the utility calculations can be expedited by employing a convenient utility function. One possibility is to use a *conjugate* utility function defined as a function of the form

$$U(d, \theta) = e^{x(d)\theta}G(\theta)^{n(d)}F(d)\,. \tag{2.6}$$

There $x(d)$, $n(d)$ and $F(d)$ are suitable functions of d, whose form will be discussed below. Since there are no necessary normalizing constraints on a utility function, as there are on a probability density, $F(d)$ does not have to satisfy a result like (2.3): nevertheless we shall find it convenient below to restrict $F(d)$ somewhat, certainly it must be positive.

For hyperparameters x and N (in posterior analysis these will be $x = \sum x_i$ and $N = n + n_0$) the expected utility of d is

$$\begin{aligned} U(d) &= \int e^{[x+x(d)]\theta}G(\theta)^{[N+n(d)]}K(N, x)F(d)\,d\theta \\ &= K(N, x)F(d)/K[N + n(d), x + x(d)] \end{aligned} \tag{2.7}$$

in terms of known functions. Maximization of this over d provides the optimum decision. Notice that we have here a "closure" property analogous to that for conjugate distributions. With the latter we know that whatever data (whatever values of N and x) we obtain, the distribution will always be of the same form, (2.5): with the conjugate utilities the expected value will always have the same form, (2.7). Consequently neither the probability distribution nor the expected utility go outside a closed family.

The possible advantages of the class (2.6) are that it represents a wide class of functions, and therefore provides considerable latitude of choice in a particular application, and that the expected values are available explicitly in terms of known functions, (2.7). We proceed to consider the form of (2.6) in more detail, and then discuss methods of maximizing (2.7).

3. The form of a conjugate utility. To understand (2.6) it is simplest to consider it as a function of θ for fixed d, rather than as is customary with a loss (or utility) function, as a function of d for fixed θ. From this unconventional view with d constant, (2.6) is just the unnormed density (2.2). Typically, therefore, it will be unimodal with tails tending to zero. In other words, d is good for only θ-values around the mode and is unsatisfactory for large and small θ-values. This is the type of utility function that is more useful in what statisticians refer to as point-estimation problems, where the wish is to have d in some sense near to θ. But there are other examples; thus in an inventory problem, where θ is the uncertain demand and d is the stock level, $d = \theta$ is the optimum choice with no unsold stock and no dissatisfied customers. Notice that θ is the "natural" parameter in the exponential family (2.1) and not necessarily the one that is "natural" in applications. An example is given below where θ is the reciprocal of the variance of a normal distribution.

Because of the usual feature of unimodality, consider the maximum of (2.6) for fixed d. Writing $g(\theta) = \log G(\theta)$, the logarithmic derivative of (2.6) vanishes when $x(d) + n(d)g'(\theta) = 0$. It would be natural in most applications for such a maximum to occur at $\theta = d$: that is, for the decision being taken, d, to be the best possible were θ at that value. We shall therefore suppose

$$x(d) = -n(d)g'(d) \tag{3.1}$$

referring to this as condition C_1, and using it to eliminate $x(d)$. At this value of θ the utility (the maximum for that d) is equal to

$$U(d, d) = \exp[n(d)\{g(d) - g'(d)d\}]F(d) .$$

In some applications it would be natural for this maximum to be the same for all d. Roughly this is saying that getting the right answer, $d = \theta$ is equally good whatever that right answer is. (This would not obtain in the inventory example mentioned above, where $U(d, d) = d$ would be more natural, at least for small d.) If we take this common value to be 1, we have $f(d) = \log F(d)$ given by

$$f(d) = n(d)\{g'(d)d - g(d)\} , \tag{3.2}$$

referring to this as condition C_2. Note that typically these utility functions will be bounded, unlike the squared-error usually employed in point-estimation problems. This is a real advantage, since an unbounded utility has always the potentiality of yielding an infinite expected utility when combined with a suitable probability distribution. No decisions are reasonably that good, and it is simplest to suppose that neither "Heaven" nor "Hell" exists, and that expected utilities are always finite.

Under C_1 and C_2 the conjugate utilities are of the form

$$U(d, \theta) = \exp[n(d)\{g(\theta) - g(d) - g'(d)(\theta - d)\}] \tag{3.3}$$

and only $n(d)$ is free to be selected. For θ near d we may expand the expression in braces about d and obtain approximately $\exp[\frac{1}{2}n(d)g''(d)(\theta - d)^2]$, so that near

the θ best for that d the utility behaves like a normal density with spread, expressed in standard deviation terms, $[-n(d)g''(d)]^{-\frac{1}{2}}$. ($g''(d)$ is negative.) The role of $n(d)$ is therefore clear, it measures how near θ has to be to d for the decision to be good: large $n(d)$ says it has to be very near, small $n(d)$ means that it is not critical. A special case would be where the departure was the same for all d; for this to happen we need

$$n(d)^{-1} = -\kappa g''(d) \tag{3.4}$$

for some constant, κ. We refer to this as condition C_3. Under C_1, C_2 and C_3 the only freedom lies in the choice of κ, which is no freedom at all since utility is not affected by a scale change. It should be remembered that C_3 refers to only *local* behavior in θ about $\theta = d$.

Having simplified the conjugate utility structure to the form (3.3), or its special case, (3.4), it is possible to say something about its structure as a function of d for fixed θ, in the usual manner. On taking logarithms of (3.3) and differentiating with respect to d, it is easy to see that it has a maximum at $d = \theta$. A second differentiation shows that the second derivative there is $n(d)g''(d)$, agreeing with that with respect to θ. The above remarks on local behavior therefore apply equally well as a function of d or of θ, with the other variable fixed. The mixed second derivative at $d = \theta$ is $-n(d)g''(d)$ so that around $\theta = d = t$ the utility function has the form

$$\exp\{\tfrac{1}{2}n(t)g''(t)(\theta - d)^2\} .$$

4. Maximization of expected utility. We now turn to considering the maximum value of the expected utility, (2.7). This is in a suitable form for numerical work provided the function $K(n, x)$ is available, but, as it stands, it is difficult to obtain analytic results. Instead we proceed to make some approximations which will help in understanding the form of the optimum decision and, in certain cases, will enable it to be calculated with ease. In applications to random samples of size n, N, in (2.7), is $n + n_0$, where n_0 is a prior hyperparameter. We shall develop approximations for large N which will be useful when either the sample size is large or the prior knowledge substantial. If $N \to \infty$, $x = \sum x_i$ will also increase, so we write $x = N\bar{x}$, thereby defining

$$\bar{x} = (x_0 + \textstyle\sum_{i=1}^{n} x_i)/(n_0 + n) . \tag{4.1}$$

Notice that $\bar{x}$ is only the sample mean when $x_0 = n_0 = 0$, otherwise it is modified in a familiar way by the prior knowledge.

LEMMA. *For large N, $K[N + n(d), N\bar{x} + x(d)]^{-1}$ is asymptotically*

$$\left\{\frac{-2\pi}{Ng''(\theta_0)}\right\}^{\frac{1}{2}} [h(\theta_0) - \tfrac{1}{2}h''(\theta_0)/Ng''(\theta_0)] \exp\{N[\bar{x}\theta_0 + g(\theta_0)]\} . \tag{4.2}$$

There θ_0 is a root of the equation $\bar{x} + g'(\theta) = 0$ and $h(\theta) = \exp[x(d)\theta + n(d)g(\theta)]$.

From (2.3) $K[N + n(d), N\bar{x} + x(d)]^{-1}$ is equal to

$$\int \exp[(N\bar{x} + x(d))\theta]G(\theta)^{N+n(d)}\,d\theta = \int e^{Nf(\theta)}h(\theta)\,d\theta\,,$$

where $f(\theta) = \bar{x}\theta + g(\theta)$. Expand both $f(\theta)$ and $h(\theta)$ in a Taylor series about θ_0, the root of $f'(\theta) = 0$. This gives

$$\int \{h(\theta_0) + (\theta - \theta_0)h'(\theta_0) + \tfrac{1}{2}(\theta - \theta_0)^2h''(\theta_0)\} \exp\{N[f(\theta_0) + \tfrac{1}{2}(\theta - \theta_0)^2f''(\theta_0)]\}\,d\theta\,,$$

retaining only the terms as far as the quadratic. Integration gives the stated answer on recognizing that $f''(\theta_0) = g''(\theta_0)$.

We now apply the lemma to investigate what happens to the expected utility as $N \to \infty$. Expression (2.7) is the ratio of two K-functions, the numerator being a special case of the denominator with $n(d) = x(d) = 0$ and hence $h(\theta) = 1$. Therefore, applying the lemma to both, we easily have that $U(d)$ is asymptotically

$$F(d)[h(\theta_0) - \tfrac{1}{2}h''(\theta_0)/Ng''(\theta_0)]\,. \tag{4.3}$$

Consider first what happens when the term $O(N^{-1})$ is omitted, leaving us with simply $h(\theta_0)F(d)$. Using C_1 and C_2, we have

$$\exp[-n(d)g'(d)\theta_0 + n(d)g(\theta_0) + n(d)\{g'(d)d - g(d)\}]$$

from (3.1) and (3.2). This is equal to

$$\exp[-n(d)\{g'(d)(\theta_0 - d) - g(\theta_0) + g(d)\}]\,,$$

which clearly has its maximum at $\hat{d} = \theta_0$. We therefore have

THEOREM 1. *Under conditions C_1 and C_2 the optimum decision for large N is given by the root, θ_0, of the equation $\bar{x} + g'(\theta) = 0$.*

Note that with $x_0 = n_0 = 0$, θ_0 is just the maximum likelihood estimate of θ; and that in any case the optimum decision does not depend on $n(d)$.

Better results can be obtained at the expense of some complexity by retaining the term $O(N^{-1})$ in (4.3). We prove

THEOREM 2. *Under conditions C_1 and C_2, the optimum decision, to $O(N^{-1})$, is* $\theta_0 - \tfrac{1}{2}n'(\theta_0)/Nn(\theta_0)g''(\theta_0)$.

Simple calculations, principally involving the evaluation of $h''(\theta_0)$, readily show that (4.3) is

$$\exp[-n(d)\{g'(d)(\theta_0 - d) - g(\theta_0) + g(d)\}] \times \left\{1 - \frac{n(d)^2[g'(\theta_0) - g'(d)]^2 + n(d)g''(\theta_0)}{2Ng''(\theta_0)}\right\}.$$

Now we know from Theorem 1 that the maximum must be near $d = \theta_0$, so let us expand terms in d around θ_0, retaining only those as far as the quadratic. The result is

$$\exp[-\tfrac{1}{2}n(\theta_0)g''(\theta_0)(d - \theta_0)^2] \times \left\{1 - \frac{n(\theta_0)^2g''(\theta_0)(d - \theta_0)^2 + n(\theta_0) + (d - \theta_0)n'(\theta_0) + \tfrac{1}{2}(d - \theta_0)^2n''(\theta_0)}{2N}\right\}.$$

If we write $z = d - \theta_0$, this expression is of the form $e^{Az^2}(a + bz + cz^2)$ where b and c are $O(N^{-1})$ but A and a are $O(1)$. This function has a maximum near $z = 0$ at approximately $z_0 = -\frac{1}{2}b/(Aa + c)$ which, to $O(N^{-1})$, is simply $z_0 = -\frac{1}{2}b/A$. On inserting the values for b and A we have the result.

If C_3 is also invoked (equation (3.4)) the optimum is $\theta_0 + \frac{1}{2}g^{(3)}(\theta_0)/N\{g''(\theta_0)\}^2$.

The estimate obtained by invoking C_3 is the same as that obtained by Lindley[1] (1961), equation (2.22), for any distribution, not merely the exponential family, but with quadratic loss. We have already seen above that C_3 essentially gives this form of loss, so the agreement is not surprising.

An interesting question to ask is: when is the modified maximum likelihood estimate, θ_0, correct even to order N^{-1}. From Theorem 2 this occurs only if $n'(\theta_0) = 0$ or $n(d)$ is constant. In that case the utility function (3.3) is locally

$$U(d, \theta) = \exp[g''(d)(\theta - d)^2]\,. \tag{4.4}$$

Rao (1962) has claimed certain second-order properties for maximum likelihood estimates. In the discussion to that paper I pointed out that such claims depend upon an implicit assumption about the loss structure. The same phenomenon is exhibited here and only the (local) form (4.4) for a loss structure will result in second-order optimality properties. This is in no way remarkable; what is remarkable is that to *first* order the loss structure is irrelevant (Theorem 1).

5. Examples. Normal mean, known variance. Here

$$p(x \mid \theta) = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}(x-\theta)^2}\,,$$

assuming the known variance to be one. This is of the form (2.2) with $G(\theta) = e^{-\frac{1}{2}\theta^2}$, so $g(\theta) = -\frac{1}{2}\theta^2$. Theorem 1 gives the optimum decision (under C_1 and C_2) to be $\bar{x}$, and under C_3 this is correct to $O(N^{-1})$ since $g^{(3)}(\theta) = 0$. To illustrate the extra term without C_3 consider the case where $n(d) = e^{-d^2}$. As $g''(\theta)$ is constant, this is a situation where maximum precision is required around $d = 0$, relatively little notice being taken of errors at large d. Simple calculation from Theorem 2 shows that the optimum decision is, to $O(N^{-1})$, $\bar{x}(1 - N^{-1})$, pulling the original value toward zero.

Normal variance, known mean. There, in its usual form with the known mean taken to be zero,

$$p(y \mid \phi) = (2\pi\phi)^{-\frac{1}{2}}e^{-\frac{1}{2}y^2/\phi}\,,$$

ϕ being the variance. To put it into the standard form above write $x = -\frac{1}{2}y^2$, so that $x < 0$, and $\theta = \phi^{-1}$, the precision. Then $G(\theta) = \theta^{\frac{1}{2}}$ and $g(\theta) = \frac{1}{2}\log\theta$, for $\theta > 0$. Theorem 1 gives the large sample decision as $\theta_0^{-1} = -2\bar{x}$. Turning this back into the original data $\{y_i\}$ and the variance ϕ, the decision for ϕ is $(\sum_{i=1}^{n} y_i^2 + y_0^2)/(n + n_0)$, the usual mean-square modified by prior knowledge. With C_3 in addition, Theorem 2 changes the value from θ_0 to $\theta_0(1 + 2/N)$, which,

[1] There are several errors in that paper, all springing from an error on line 4 of (2.18)—the 3! there should be 2. This leads to a 6 in (2.22) which should be replaced by 2.

in terms of the variance gives $(\sum y_i^2 + y_0^2)/(n + n_0 + 2)$. The modification reflects the common discussion about the appropriate divisor for the sum of squares, see, for example, Evans (1964) and references therein. The modification suggested by Theorem 2 without C_3 extends this idea. The "standard deviation" of the utility function $[-n(d)g''(d)]^{-\frac{1}{2}}$ is here $[2d^2/n(d)]^{\frac{1}{2}}$. With $n(d) = d^{-a}$, with $a \geqq -2$, the spread increases with d, from zero to infinity, so that the smaller values of the precision θ are required to greater accuracy. The estimate, to order N^{-1}, is easily seen to be $\theta_0(1 - a/N)$, which, in variance terms is $(\sum y_i^2 + y_0^2)/(n + n_0 - a)$. The restriction to $a \geqq -2$ is needed, since $a < -2$ would mean an increasing spread as $d \to 0$, which would be unreasonable in view of d and θ both being nonnegative. For many purposes the logarithm of the precision is sensibly estimated with constant error. This corresponds to $a = 0$ and leads back to the modified maximum likelihood estimate. So here we have a case where that estimate does have reasonable second-order efficiency properties.

Bernoulli sampling. Here

$$p(x \mid \phi) = \phi^x(1 - \phi)^{1-x}$$

for $x = 0$ or 1, and $0 < \phi < 1$. Writing $\theta = \log\{\phi/(1 - \phi)\}$, the log-odds, this is in the standard form

$$p(x \mid \theta) = e^{x\theta}(1 + e^\theta)^{-1}$$

with $G(\theta) = (1 + e^\theta)^{-1}$ and $g(\theta) = -\log(1 + e^\theta)$. Theorem 1 gives the usual modified maximum likelihood estimate for ϕ (under C_1 and C_2) to be $\bar{x}$. In Theorem 2, using C_3, simple calculations show that to $O(N^{-1})$ the optimum decision for θ is $\theta_0 + \sinh\theta_0/N$, where $\theta_0 = \log\{\bar{x}/(1 - \bar{x})\}$. As an example where increased precision is required for values of θ large in absolute value (ϕ near 0 or 1), take $-n(d)g''(d) = (1 + e^d)^2/e^d$. Simple calculation gives $\theta_0 + 2\sinh\theta_0/N$. Acting in opposite direction, giving more precision around $\phi = \frac{1}{2}$ is $-n(d)g''(d) = e^d/(1 + e^d)^2$. This takes us back to the estimate θ_0.

6. Monotone utility functions. A limitation of the utilities given by (2.6) is that they necessarily have the features of a density function: for example, the utility typically tends to zero at the ends of the range of θ. In many applications utility cannot reasonably have this property: for instance, it may be increasing in θ, the larger θ being, the better the decision. Such a monotone property is possessed by a distribution function and it is therefore tempting to see whether, just as (2.6) imitates the density, (2.4), for θ, we cannot make progress by choosing a utility which imitates the distribution function for θ.

Consider therefore a utility function

$$U(d, \theta) = \int_{-\infty}^{\theta} e^{x(d)t}G(t)^{n(d)}K(x(d), n(d))\,dt \tag{6.1}$$

which, for fixed d, is a distribution function for θ having hyperparameters $x(d)$ and $n(d)$. (The expression could be generalized by not normalizing by K but

using a general function $F(d)$ as in (2.6). However, in most applications $\lim_{\theta\to\infty} U(d, \theta)$ will not depend on d, and (6.1) has this property, the limit always being 1.) Let ψ be a random quantity having density (2.4) with hyperparameters $x(d)$ and $n(d)$. Then (6.1) may be written $U(d, \theta) = p(\tilde{\psi} < \theta)$, the tilde serving to indicate the random quantity. It immediately follows that the expected utility is $p(\tilde{\psi} < \tilde{\theta})$; that is, the probability that one random quantity $\tilde{\psi}$—with hyperparameters $x(d)$ and $n(d)$—is less than an independent random quantity $\tilde{\theta}$ having the same distribution but with hyperparameters, x and N, say. Consequently our expected utility is easily calculable whenever the distribution of the difference or ratio of two independent random quantities from the conjugate family is easily calculable. An obvious case is where the conjugate family is normal, for then the difference is also normal. This is an important, but rather special, case. Fortunately there often exists a transformation which achieves approximate normality, so that the device is of some generality. We therefore consider the normal case in detail. The paper by Berhold (1973) is relevant.

Suppose θ has a posterior distribution which is $N(\mu_1, \sigma_1^2)$, say. This will happen in the first example in Section 5 of sampling from a normal distribution of unknown mean, θ, and known variance. Let $U(d, \theta)$, for a fixed decision d, be a normal distribution function with mean μ_0 and variance σ_0^2—these two quantities will depend on d. Then by the above argument the expected utility involves the difference between two independent normal variables, which is itself normal with mean $\mu_0 - \mu_1$ and variance $\sigma_1^2 + \sigma_0^2$. The expected utility is equal to the probability that this difference is negative, that is to $\Phi[(\mu_1 - \mu_0)/(\sigma_1^2 + \sigma_0^2)^{\frac{1}{2}}]$, where Φ is the standard normal distribution function.

Two applications of this class of utility functions have come to my notice. The first is from the field of education where θ is a measure of the true worth of the subject, what is often called his true score: see Lord and Novick (1968). It is sometimes supposed that the subject is good enough for some task, or for some training if, and only if, θ exceeds some critical threshold level, μ_0, say. A test on the subject gives an observed score and, as a result, θ typically has a posterior distribution which is normal, say $N(\mu_1, \sigma_1^2)$. (μ_1 will depend on the observed score.) Using a utility function which is 1 for $\theta > \mu_0$ and 0 otherwise, the expected utility is $\Phi[(\mu_1 - \mu_0)/\sigma_1]$, and the subject is accepted if this exceeds a critical value; that is, if $\mu_1 > \mu_0 + \lambda\sigma_1$ for some λ.

In criticism of this, it may be argued that to assume a subject with θ just a little larger than μ_0 is satisfactory but one with θ a little less is no good, is unrealistic. Or, to put it differently, to suppose the utility has a discontinuity at μ_0 is inappropriate. The suggestion made here is to replace the discontinuous utility by the normal distribution function with mean μ_0 and variance σ_0^2. (The limit of this as $\sigma_0 \to 0$ is the original function.) The use of this would imply, amongst other things, that subjects with $\theta > \mu_0 + 2\sigma_0$ were almost certainly satisfactory, and those with $\theta < \mu_0 - 2\sigma_0$ were almost certainly not. Subjects with $\theta = \mu_0$ were as likely to be satisfactory as not. By the above argument,

the expected utility for a subject with a true score which is $N(\mu_1, \sigma_1^2)$ is $\Phi[(\mu_1 - \mu_0)/(\sigma_1^2 + \sigma_0^2)^{\frac{1}{2}}]$.

It is of interest to compare the two expected utilities. For fixed μ_0, μ_1, σ_1^2, the use of the new function results in a decrease in expectation whenever $\mu_1 > \mu_0$ and an increase whenever $\mu_1 < \mu_0$. In words, if the cut-off on observed score is high ($\lambda > 0$ above) the student will have to do even better with the new utility: in the contrary case he need not do as well. (Of course, the cut-off value should be determined using another utility function for rejection. A suitable form might be one minus a distribution function: that is, $p(\tilde{\phi} > \theta)$ in the notation used above.) The change required can be easily calculated. Let μ^* be the mean critical value when $\sigma_0^2 = 0$, and $\mu^* + h$ be the value for the new utility function. Then

$$\frac{\mu^* - \mu_0}{\sigma_1} = \frac{\mu^* + h - \mu_0}{(\sigma_1^2 + \sigma_0^2)^{\frac{1}{2}}},$$

so that

$$h = (\mu^* - \mu_0)\{(\sigma_1^2 + \sigma_0^2)^{\frac{1}{2}}/\sigma_1 - 1\}.$$

For small σ_0 this is approximately $(\mu^* - \mu_0)\sigma_0^2/2\sigma_1^2$.

A second application is to the time, θ, available to complete a task. There the expected time available had been used as a criterion, implicitly using a linear utility. This is probably unrealistic. If the time is very short, then the utility is low. It will not reasonably begin to rise until a certain minimum time is available, and will most likely not rise much more after an adequate time is available. If so, the distribution function form can again be used. Of course the results may be quite different from these obtained using a linear utility.

In many situations—even within the exponential family—the distribution of the difference of two random variables is not available explicitly and the above approach can be used only numerically. However, an approximate method is typically available. For example, with Bernoulli sampling, the conjugate family is the Beta family and the log-odds provide a reasonable approximation to normality: $\theta = \log\{\phi/(1 - \phi)\}$, in the notation used in the Bernoulli example above, is approximately normal with mean $\log\{(x + \frac{1}{2})/(N - x + \frac{1}{2})\}$ and variance $(x + 1)^{-1} + (N - x + 1)^{-1}$, when the conjugate family is written in the form $\phi^x(1 - \phi)^{N-x}$. If a normal distribution function is a suitable utility function in terms of θ, then the above methods are available. This amounts to the use of a distribution function in terms of ϕ, the basic parameter.

7. Several parameters. The above ideas extend without any serious difficulty to the case of more than one real parameter. In particular, the use of a bivariate normal distribution function for the utility of a decision that depends on two quantities looks promising. However, conjugate utilities (whether of density or distribution function form) for more than one parameter suffer from the same defect as do conjugate probability distributions: namely, they do not have enough hyperparameters. Generally, with k parameters we have $(k+1)$ hyperparameters.

But to describe only the first- and second-order properties of k quantities we need $k(k+3)/2$ variables—a number greatly in excess of the number of available hyperparameters. It is hoped to explore these ideas further in another paper.

REFERENCES

[1] BERHOLD, M. H. (1973). The use of distribution functions to represent utility functions. *Management Sci.* **19** 825–829.

[2] EVANS, I. G. (1964). Bayesian estimation of the variance of a normal distribution. *J. Roy. Statist. Soc. Ser. B* **26** 63–68.

[3] LINDLEY, D. V. (1961). The use of prior probability distributions in statistical inference and decision. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 453–468. Univ. of California Press.

[4] LORD, F. M. and NOVICK, M. R. (1968). *Statistical Theories of Mental Test Scores.* Addison-Wesley, Reading.

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

[6] RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory.* Division of Research, Harvard Business School, Boston.

[7] WETHERILL, G. B. (1961). Bayesian sequential analysis. *Biometrika* **48** 281–292.

DEPARTMENT OF STATISTICS AND COMPUTER SCIENCE
UNIVERSITY COLLEGE LONDON
GOWER STREET, WC1E 6BT
ENGLAND