The Annals of Statistics,
1983, Vol. 11, No. 4, 1218–1224

ESTIMATING EVENTS

BY MICHAEL EVANS

University of Toronto

The problem of estimating an event, having positive probability content, based on a sample of n observations is considered. A natural metric is shown to exist on the space of possible values for the event. This leads to the definition of optimal estimators. We derive optimal estimators for events which correspond to quantiles for the univariate exponential model. Further optimal estimators are derived for the events bounded by the ellipsoidal contours of the density function in the multivariate normal model.

1. Introduction. The problem we are concerned with here is the estimation of an event which has positive probability content. More precisely we suppose that for statistical model $\mathscr{M} = (S, \mathscr{A}, \{P_\theta \mid \theta \in \Omega\})$ we have the map $Q: \Omega \to \mathscr{A}$ satisfying $P_\theta(Q(\theta)) > 0$ for every $\theta \in \Omega$. The reason for requiring Q to have positive probability content will be explained later. An estimator of Q, based upon a sample of n observations from $\mathscr{M}$, is then a map $C: S^n \to \mathscr{Q} = \{Q(\theta) \mid \theta \in \Omega\}$. We would naturally like C to come "close" to Q, in some sense, in repeated sampling. In Section 2 we define what we mean by "close".

We can consider the problem of estimating the quantiles of a univariate distribution as a special case of our problem. For if $S = \mathbb{R}^1$, $\mathscr{A} = \mathscr{B}^1$ and $q: \Omega \to \mathbb{R}$ is a β-quantile for the model; i.e. $P_\theta((-\infty, q(\theta))) \le \beta \le P_\theta((-\infty, q(\theta)])$ for every $\theta \in \Omega$, then we can speak equivalently of estimating q or $Q: \Omega \to \mathscr{B}^1$ defined by $Q(\theta) = (-\infty, q(\theta)]$. The estimation of quantiles for univariate distributions has been considered by a number of authors; for example Mann (1969); Zidek (1969a); Zidek (1971); Robertson (1977); Dyer, Keating and Hensley (1977); Dyer and Keating (1979); Schafer and Angus (1979); Angus and Schafer (1979) and Reiss (1980). In Section 3 we present optimal estimators for the quantiles of the exponential distributions which are different from those previously obtained.

In Section 4 we consider the problem of estimating the regions bounded by the contours of the density of the multivariate normal distribution. Estimating these sets represents a generalization of the traditional domain of application for estimation theory.

An alternative approach to making inferences about Q arises in the context of the theory of tolerance regions; see for example Fraser and Guttman (1956), Evans and Fraser (1980). The optimality theory of such regions is, however, quite different than the development here which is based on a metric defined on $\mathscr{Q}$.

2. Optimal estimators of events. Suppose that $C: S^n \to \mathscr{Q}$ is an estimator of Q. If θ is true and we have observed $\mathbf{s}$ we want a measure of how close $C(\mathbf{s})$ is to $Q(\theta)$. A purely set-theoretic measure of the difference between two sets A and B is the symmetric difference $A \,\Delta\, B = A \cup B - A \cap B$. Further if μ_θ is a measure on $(S, \mathscr{A})$ then it is easy to prove that $d_\theta: \mathscr{A} \times \mathscr{A} \to \mathbb{R}$ defined by $d_\theta(A, B) = \mu_\theta(A \,\Delta\, B)$ is a semimetric on $\mathscr{A}$ and thus also on $\mathscr{Q}$. If $\theta_1 \ne \theta_2$ implies $\mu_\theta(Q(\theta_1) \,\Delta\, Q(\theta_2)) \ne 0$ then clearly d_θ is a metric on $\mathscr{Q}$. Thus $d_\theta(C(\mathbf{s}), Q(\theta))$ gives a numeric measure of the difference between $C(\mathbf{s})$ and $Q(\theta)$. Taking $d_\theta(\cdot, Q(\theta))$ as the loss function we then have that estimator C is optimal in a class $\mathscr{C}$ of possible estimators if C minimizes $E_\theta[d_\theta(C(\mathbf{s}), Q(\theta))]$ uniformly in θ.

In a given context we must choose the μ_θ. If S is Euclidean we might take $\mu_\theta \equiv \mu$ where μ is Borel measure. Perhaps a more natural choice, and it is the one we will employ here,

Received December 1981; revised April 1983.

AMS 1970 *subject classifications*. Primary, 62F10; secondary, 62H99, 62C99.

Key words and phrases. Random set estimators, equivariance, univariate exponential, multivariate normal.

is to take $\mu_\theta = P_\theta$. Thus it is the expected probability content of the set $C(\mathbf{s}) \Delta Q(\theta)$ which we wish to minimize for each θ. We note that we always have $0 \le E_\theta[d_\theta(C(\mathbf{s}), Q(\theta))] \le 1$. Also if $P_\theta(Q(\theta)) \equiv 0$ then any C satisfying $P_\theta(C(\mathbf{s})) \equiv 0$ would be an optimal estimator of C. For this reason we consider only those Q having positive probability content.

As a direct consequence of Fubini's Theorem, as used in Robbins (1944), we have the following result.

PROPOSITION 1. *For statistical model $\mathscr{M}$, $Q(\theta) \in \mathscr{A}$ and $C: S^n \to \mathscr{A}$ define $\phi: S^{n+1} \to \mathbb{R}^1$ by $\phi(\mathbf{s}, s) = 1$ when $s \in C(\mathbf{s}) \Delta Q(\theta)$ and 0 otherwise. If ϕ is measurable and $P_\theta^n \times \mu_\theta$ integrable then $E_\theta[d_\theta(C(\mathbf{s}), Q(\theta))] = \int_S P_\theta^n(s \in C(\mathbf{s}) \Delta Q(\theta))\, d\mu_\theta(s)$.*

Taking $\mu_\theta = P_\theta$ this result says that the expected probability content of $C(\cdot) \Delta Q(\theta)$ is just the probability that in a sample of size $(n+1)$ from P_θ the $(n+1)$-st value s is covered by $C(\mathbf{s}) \Delta Q(\theta)$. As we will see in Section 4, this proposition can be useful in computations.

As the problem we are discussing here is a decision problem we can speak of unbiased, equivariant, consistent, etc., estimators for Q. In the following sections we will restrict attention to estimators which are equivariant under a group G acting on S which leaves $\{P_\theta \mid \theta \in \Omega\}$ invariant. We note that there does not appear to be an easy analogue of the Rao-Blackwell Theorem in this context and so the class of estimators which are functions of the minimal sufficient statistic may not be essentially complete.

3. Estimating quantiles. We consider the problem of estimating $Q(\theta) = (-\infty, q(\theta)]$ where q is a β-quantile for $(\mathbb{R}^1, \mathscr{B}^1, \{P_\theta \mid \theta \in \Omega\})$. Then the estimator C must be of the form $C(\mathbf{s}) = (-\infty, c(\mathbf{s})]$ and $C(\mathbf{s}) \Delta Q(\theta) = (\min\{c(\mathbf{s}), q(\theta)\}, \max\{c(\mathbf{s}), q(\theta)\}]$. If we take μ_θ to be Borel measure then an optimal estimator c minimizes the expected distance from q. Taking $\mu_\theta = P_\theta$ and denoting the distribution function by F_θ we have that $d_\theta(C(\mathbf{s}), Q(\theta)) = |F_\theta(c(\mathbf{s})) - F_\theta(q(\theta))|$.

In Robertson (1977) estimators of the quantiles of the exponential (θ) distribution are discussed. This paper obtained the estimator in the class of those of the form $c(\mathbf{s}) = \sum_{i=1}^n a_i s_i$ which minimizes $E_\theta[d_\theta^2(C(\mathbf{s}), Q(\theta))]$. We consider this problem using the approach developed here and obtain new estimators.

EXAMPLE. Suppose that $\mathbf{x} = (x_1, \cdots, x_n)'$ is a sample from the exponential (θ) model where $\theta > 0$ is unknown and we wish to estimate the β-quantile $q(\theta) = -\theta \ln(1-\beta)$. This problem is equivariant under the group $G = \mathbb{R}^+$ with the product being ordinary multiplication and with action on $\mathbb{R}^n$ given by $g\mathbf{x} = (gx_1, \cdots, gx_n)'$. Accordingly we restrict our attention to estimators satisfying $c(g\mathbf{x}) = gc(\mathbf{x})$. A maximal invariant statistic under the action of this group is given by $\mathbf{d} = \mathbf{x}/\bar{x}$ where $\bar{x} = (1/n) \sum_{i=1}^n x_i$ and thus $c(\mathbf{x}) = \bar{x}c(\mathbf{d})$. As $\bar{x}$ and $\mathbf{d}$ are statistically independent, see Fraser (1976) page 466, Problem 6, the optimal conditional equivariant estimate given $\mathbf{d}$ will correspond to the optimal equivariant estimator. Hence we write $c(\mathbf{x}) = \bar{x}k$ and find the optimal $k > 0$.

Now $d_\theta(C(\mathbf{x}), Q(\theta)) = |1 - \exp\{-(z/n)k\} - \beta|$ where $z = (n/\theta)\bar{x} \sim$ Gamma (n). Then $E_\theta[d_\theta(C(\mathbf{x}), Q(\theta))] = E[|1 - \exp\{-(z/n)k\} - \beta|] = d(k)$ and since $1 - \exp\{-(z/n)k\} \ge \beta$ if and only if $z \ge -(n/k)\ln(1-\beta)$

$$d(k) = \int_{-(n/k)\ln(1-\beta)}^{\infty} \left(1 - \exp\left\{-\frac{z}{n}k\right\} - \beta\right) h(z)\, dz - \int_0^{-(n/k)\ln(1-\beta)} \left(1 - \exp\left\{-\frac{z}{n}k\right\} - \beta\right) h(z)\, dz \tag{1}$$

where h is the Gamma (n) density. Note that $d(0) = \beta$, $d(\infty) = 1 - \beta$ and thus when $\beta = 0$ the optimal estimator is 0 and when $\beta = 1$ the optimal estimator is ∞.

Differentiating d using Leibnitz's formula, Abramowitz and Stegun (1965), setting

$d'(k) = 0$ and simplifying we obtain that k is a critical point of d if and only if k satisfies $G(-2(1 + n/k)\ln(1 - \beta)) = 1/2$ where G is the chi-squared $(2n + 2)$ distribution function.

Thus k is a critical point of d if and only if $-2(1 + n/k)\ln(1 - \beta)$ is the median of the chi squared $(2n + 2)$ distribution and when this is so we have

$$k = \frac{-n \ln(1 - \beta)}{\ln(1 - \beta) + (1/2)\chi^2_{.5}(2n + 2)}.$$

It is easy to show that $d''(k) > 0$ whenever $d'(k) = 0$ and thus whenever d has a critical point it gives the absolute minimum. If d does not have a critical point then the optimal equivariant estimator is given by $c(\mathbf{x}) = \infty$. We note that a finite k exists satisfying (1) if and only if $\chi^2_{.5}(2n + 2) \geq -2 \ln(1 - \beta)$ and this is always true whenever $\beta \leq .5$. Further, for any β, since $\chi^2_{.5}(n) \to \infty$ as $n \to \infty$, there exists an N_β such that for all $n \geq N_\beta$ the optimal equivariant estimator is finite. We have proved:

PROPOSITION 2. *If* $\mathbf{x} = (x_1, \cdots, x_n)'$ *is a sample from a distribution in the class* $\{$*exponential* $(\theta) \mid \theta > 0\}$ *then a finite optimal equivariant estimator exists for the quantile* $-\theta \ln(1 - \beta)$ *if and only if* $\chi^2_{.5}(2n + 2) > -2 \ln(1 - \beta)$ *and this is always true whenever* $\beta \leq .5$. *When such an estimator exists it is of the form* $\bar{x}k$ *where*

$$k = \frac{-n \ln(1 - \beta)}{\ln(1 - \beta) + (1/2)\chi^2_{.5}(2n + 2)}.$$

We thus have that tables and approximations for the chi-squared distribution can be used to evaluate this estimator. As the group G is transitive and abelian we have, by Lehmann (1959, page 23) that this estimator is unbiased. Further, it is easy to show that the estimator is consistent almost surely for the quantile.

In Dyer, Keating and Hensley (1977) an extensive discussion is given concerning various point estimators for the quantiles of the univariate normal distribution. Following the previous example, optimal equivariant estimators of these quantiles can be developed which differ from those previously considered. Reflecting the symmetry in the normal case, an alternative definition of a β-quantile could be $Q(\mu, \sigma) = [\mu - \sigma z_{(1-\beta)/2}, \mu + \sigma z_{(1-\beta)/2}]$. For a discussion of an applied context where this would be appropriate see Owen (1964). Estimators for such quantiles are obtained as special cases of those derived in the following section. The analysis for two-sided quantiles is substantially different than the one-sided case.

4. Estimating the central events of the multivariate normal. Suppose that $X = (\mathbf{x}_1 \cdots \mathbf{x}_n) \in \mathbb{R}^{p\times n}$ is a sample from the $N_p(\mu, \Sigma)$ distribution where $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p\times p}$ positive definite, are unknown. Let $Q(\mu, \Sigma) = \{\mathbf{y} \mid (\mathbf{y} - \mu)'\Sigma^{-1}(\mathbf{y} - \mu) \leq k_0\}$ where $k_0 = \chi^2_{1-\beta}(p)$ is the point exceeded with probability β by the Chi squared (p) distribution and note that $P_{(\mu,\Sigma)}(Q(\mu, \Sigma)) = \beta$ for every (μ, Σ).

On intuitive grounds we restrict attention to estimators C taking values in $\mathscr{Q}$ and based on the minimal sufficient statistic $(\bar{\mathbf{x}}, S(X)) = ((1/n) \sum_{i=1}^n \mathbf{x}_i, XX' - n\bar{\mathbf{x}}\,\bar{\mathbf{x}}')$. We note that this problem is equivariant under the group $G = \{[\mathbf{a}, B] \mid \mathbf{a} \in \mathbb{R}^p, B \in \mathbb{R}^{p\times p} \det B \neq 0\}$ with product $[\mathbf{a}_1, B_1][\mathbf{a}_2, B_2] = [\mathbf{a}_1 + B_1\mathbf{a}_2, B_1B_2]$ and induced action on the minimal sufficient statistic given by $[\mathbf{a}, B](\bar{\mathbf{x}}, S(X)) = (\mathbf{a} + B\bar{\mathbf{x}}, BS(X)B')$. We then further restrict to those estimators which are also equivariant under G. Thus $C(\bar{\mathbf{x}}, S(X)) = [\bar{\mathbf{x}}, S(X)_T]C(\mathbf{0}, I)$ where $S_T \in \mathbb{R}^{p\times p}$ denotes the unique lower triangular matrix with positive diagonal elements satisfying $S = S_TS_T'$ for positive definite $S \in \mathbb{R}^{p\times p}$. The special case $\bar{\mathbf{x}} = 0$, $S(X) = I$ implies $C(\mathbf{0}, I) \in \mathscr{Q}$; *i.e.* $C(\mathbf{0}, I)$ is an ellipsoid. If $Q \in \mathbb{R}^{p\times p}$ is orthogonal then $[\mathbf{0}, Q]C(\mathbf{0}, I) = C(\mathbf{0}, QQ') = C(\mathbf{0}, I)$ and this implies that $C(\mathbf{0}, I)$ is a sphere in $\mathbb{R}^p$ centered at $\mathbf{0}$. Thus C is of the form $C(\bar{\mathbf{x}}, S(X)) = \{\mathbf{y} \mid (\mathbf{y} - \bar{\mathbf{x}})'S^{-1}(X)(\mathbf{y} - \bar{\mathbf{x}}) \leq k\}$ for some k. We note that, as pointed out by a referee, we could extend the class $\mathscr{Q}$ to include all convex subsets of $\mathbb{R}^p$ and the above argument leads to estimators of the same form. In the following we determine the optimal k. Whether or not this estimator is optimal in the full class of equivariant estimators is undetermined.

Now

$$\begin{aligned} C(\bar{\mathbf{x}}, S(X)) \,\Delta\, Q(\mu, \Sigma) &= \{[\bar{\mathbf{x}}, S(X)_T]C(\mathbf{0}, I)\} \,\Delta\, \{[\mu, \Sigma_T]Q(\mathbf{0}, I)\} \\ &= [\mu, \Sigma_T]\{[\mu, \Sigma_T]^{-1}[\bar{\mathbf{x}}, S(X)_T]C(\mathbf{0}, I) \,\Delta\, Q(\mathbf{0}, I)\} \\ &= [\mu, \Sigma_T]\{C(\bar{\mathbf{z}}, S) \,\Delta\, Q(\mathbf{0}, I)\} \end{aligned}$$

where $\bar{\mathbf{z}} = \Sigma_T^{-1}(\bar{\mathbf{x}} - \mu) \sim N_p(\mathbf{0}, (1/n)I)$ is statistically independent of $S = \{\Sigma_T' S^{-1}(X)\Sigma_T\}^{-1} \sim W_p(I, n-1)$, see Anderson (1958, page 162) and where $W_p(\Sigma, m)$ denotes the Wishart distribution in p dimensions with matrix Σ and m degrees of freedom.

From the above we have that $P_{(\mu,\Sigma)}(C(\bar{\mathbf{x}}, S(X)) \,\Delta\, Q(\mu, \Sigma)) = P_{(\mathbf{0},I)}(C(\bar{\mathbf{z}}, S) \,\Delta\, Q(\mathbf{0}, I))$ where $P_{(\mu,\Sigma)}$ denotes the $N_P(\mu, \Sigma)$ distribution. Thus

$$E_\theta[d_\theta(C(\bar{\mathbf{x}}, S(X)) \,\Delta\, Q(\theta))] = E[P_{(\mathbf{0},I)}(C(\bar{\mathbf{z}}, S) \,\Delta\, Q(\mathbf{0}, I))]$$

where the second expectation is with respect to the joint measure for $(\bar{\mathbf{z}}, S)$ which we will denote by P. By the proposition of Section 2 this expectation equals

$$\int_{\mathbb{R}^p} P(\mathbf{y} \in C(\bar{\mathbf{z}}, S) \,\Delta\, Q(\mathbf{0}, I)) \, dP_{(\mathbf{0},I)}(\mathbf{y}). \tag{2}$$

If $\|\mathbf{y}\|^2 \le k_0$ then $\mathbf{y} \in C(\bar{\mathbf{z}}, S) \,\Delta\, Q(\mathbf{0}, I)$ if and only if $\mathbf{y} \notin C(\bar{\mathbf{z}}, S)$; i.e. if and only if $(\bar{\mathbf{z}} - \mathbf{y})'S^{-1}(\bar{z} - \mathbf{y}) > k$. If $\|\mathbf{y}\|^2 > k_0$ then $\mathbf{y} \in C(\bar{\mathbf{z}}, S) \,\Delta\, Q(\mathbf{0}, I)$ if and only if $\mathbf{y} \in C(\bar{\mathbf{z}}, S)$; i.e. if and only if $(\bar{\mathbf{z}} - \mathbf{y})'S^{-1}(\mathbf{z} - \mathbf{y}) \le k$.

For fixed $\mathbf{y}$ we have $\sqrt{n}(\bar{\mathbf{z}} - \mathbf{y}) \sim N_p(-\sqrt{n}\mathbf{y}, I)$ statistically independent of $S \sim W_p(I, n-1)$ as above. Assuming $n > p$, Anderson (1958, page 106, Theorem 5.22) gives that $((n-p)/p)n(\bar{\mathbf{z}} - \mathbf{y})'S^{-1}(\bar{\mathbf{z}} - \mathbf{y})$ is distributed as a noncentral $F(p, n-p)$ distribution with noncentrality $n\|y\|^2$. If $F(\cdot, \delta, m, n)$ denotes the noncentral $F(m, n)$ distribution function with noncentrality δ and with density $f(\cdot, \delta, m, n)$ then

$$P((\bar{\mathbf{z}} - \mathbf{y})'S^{-1}(\bar{\mathbf{z}} - \mathbf{y}) \le k) = F(((n-p)/p)nk, n\|\mathbf{y}\|^2, p, n-p).$$

Thus putting $x = ((n-p)/p)nk$, (2) is equal to

$$\int_{\|\mathbf{y}\|^2 \le k_0} \{1 - F(x, n\|\mathbf{y}\|^2, p, n-p)\} \, dP_{(\mathbf{0},I)}(\mathbf{y}) + \int_{\|y\|^2 > k_0} F(x, n\|\mathbf{y}\|^2, p, n-p) \, dP_{(\mathbf{0},I)}(\mathbf{y}). \tag{3}$$

Now $\|\mathbf{y}\|^2 \sim \chi^2(p)$, so putting $s = \|y\|^2$, G_p equal to the $\chi^2\,(p)$ distribution function, $g_p = G_p'$, and noting $F(x, ns, p, n-p)$ is a function of s, we have that (3) is equal to

$$d(x) = \int_0^{k_0} \{1 - F(x, ns, p, n-p)\} g_p(s) \, ds + \int_{k_0}^{\infty} F(x, ns, p, n-p) g_p(s) \, ds. \tag{4}$$

Note that $d(0) = G_p(k_0) = \beta$ and thus the optimal estimator of μ is given by $\bar{\mathbf{x}}$.

Differentiating d with respect to x, setting $d'(x) = 0$ and simplifying we obtain that x is a critical point of d if and only if $H(k_0 \mid x) = ½$ where $H(\cdot \mid x)$ is the distribution function of the distribution with density given by

$$h(s \mid x) = f(x, ns, p, n-p) g_p(s) \bigg/ \left\{ \int_0^\infty f(x, ns, p, n-p) \cdot g_p(s) \, ds \right\}.$$

Thus x is a critical point of d if and only if k_0 is a median of $H(\cdot \mid x)$.

We have

$$f(x, ns, p, n-p) = \sum_{m=0}^{\infty} \{2g_{2m+2}(ns)\}\left\{f\left(\frac{p}{p+2m}x, 0, p+2m, n-p\right)\left(\frac{p}{p+2m}\right)\right\}$$

and thus

$$\int_0^{k_0} f(x, ns, p, n-p)g_p(s)\, ds$$

$$= \sum_{m=0}^{\infty}\left\{\int_0^{k_0} 2g_{2m+2}(ns)g_p(s)\, ds\right\}\left\{f\left(\frac{p}{p+2m}x, 0, p+2m, n-p\right)\left(\frac{p}{p+2m}\right)\right\}$$

$$= \sum_{m=0}^{\infty}\left\{\int_0^{k_0}\left(\Gamma\left(\frac{p+2m}{2}\right)\right)^{-1}\left(\frac{n+1}{2}s\right)^{((p+2m)/2)-1} e^{-(n+1)s/2}\,\frac{n+1}{2}\, ds\right\}$$

$$\cdot\left\{\left[\Gamma\left(\frac{p+2m}{2}\right)\Big/\left(\Gamma\left(\frac{2m+2}{2}\right)\Gamma\left(\frac{p}{2}\right)\right)\right]n^m(n+1)^{-((p+2m)/2)}\right.$$

$$\left.\cdot f\left(\frac{p}{p+2m}x, 0, p+2m, n-p\right)\left(\frac{p}{p+2m}\right)\right\}$$

$$= f\left(\frac{x}{n+1}, 0, p, n-p\right)\frac{1}{n+1}\sum_{m=0}^{\infty}\{G_{p+2m}((n+1)k_0)\}\left\{b\left(m, \frac{n}{2}, p(x)\right)\right\}$$

where $b(m, r, p) = \{\Gamma(m+r)/(\Gamma(m+1)\Gamma(r))\}p^r(1-p)^m$ is the negative binomial (r, p) probability function and

$$p(x) = 1 - \left\{\frac{n}{n+1}\right\}\left\{\left(\frac{p}{n-p}x\right)\Big/\left(1+\frac{p}{n-p}x\right)\right\}.$$

Thus we obtain $h(s \mid x) = \sum_{m=0}^{\infty}\{b(m, n/2, p(x))\}\{g_{p+2m}((n+1)s)(n+1)\}$; i.e. an infinite weighted sum of independent chi squareds where the weights are negative binomial probabilities.

Now $d'(x) = \sum_{m=0}^{\infty} a_m(x)$ where

$$a_m(x) = \left\{f\left(\frac{x}{n+1}, 0, p, n-p\right)\left(\frac{1}{n+1}\right)\right\}\{1 - 2G_{p+2m}((n+1)k_0)\}\left\{b\left(m, \frac{n}{2}, p(x)\right)\right\}.$$

Differentiating $a_m(x)$ with respect to x we obtain $a_m'(x) = a_m(x)\{k_1(x) + k_2(x)m\}$ where

$$k_1(x) = \frac{(p-2)(n-p+px) - npx}{2x(n-p+px)} \quad \text{and} \quad k_2(x) = \frac{2n-2p+px}{2x(n-p+px)}.$$

Justifying the differentiation through the summation by the dominated derivative theorem, Fraser (1976, page 551) we have that $d''(x) = k_1(x)d'(x) + k_2(x)\sum_{m=0}^{\infty} ma_m(x)$. Since $\{1 - 2G_{p+2m}((n+1)k_0)\}$ is an increasing function of m, there is an $m^* \in N_0$ such that $a_m(x) \le 0$ for $m \le m^*$ and $a_m(x) > 0$ for $m \ge m^*$. Now $k_2(x) > 0$ and thus whenever $d'(x) = 0$ we have

$$d''(x) = k_2(x)\{\textstyle\sum_{m=0}^{m^*} ma_m(x) + \sum_{m=m^*+1}^{\infty} ma_m(x)\}$$

$$> k_2(x)\{\textstyle\sum_{m=0}^{m^*} m^*a_m(x) + \sum_{m=m^*+1}^{\infty} m^*a_m(x)\} = 0.$$

Therefore any critical point of d is a point where the absolute minimum is achieved. If no critical point exists then as $d(\infty) = 1 - \beta$ we have that d is increasing whenever $\beta \le 1/2$ and thus the estimator is given by $\bar{x}$ and d is decreasing whenever $\beta > 1/2$ and the estimator does not exist.

We note that $G_p((n-1)k_0) = \lim_{k\to 0} H(k_0 \mid ((n-p)/p)nk)$ and

$$\sum_{m=0}^{\infty} b(m, n/2, 1/(n+1))\ G_{p+2m}((n+1)k_0) = \lim_{k\to\infty} H(k_0 \mid ((n-p)/p)nk).$$

Thus a sufficient condition for a critical point to exist is that $G_p((n+1)k_0) > 1/2$ and

$$\sum_{m=0}^{\infty} b(m, n/2, 1/(n+1)) G_{p+2m}((n+1)k_0) < 1/2.$$

It is straightforward to show that these conditions are also necessary for the existence of a critical point. It can also be shown that for a given β there exists an N_β such that for all $n \geq N_\beta$ a finite estimator exists. We have the following

PROPOSITION 3. *If $X = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is a sample from a distribution in $\{N_p(\boldsymbol{\mu}, \Sigma) \mid \boldsymbol{\mu} \in R^p, \Sigma \in \mathbb{R}^{p\times p}$ p.d.$\}$ then an optimal estimator of the form $C(X) = \{\mathbf{y} \mid (\mathbf{y} - \bar{\mathbf{x}})'S^{-1}(x)(\mathbf{y} - \bar{\mathbf{x}}) \leq k\}$ where $k \in [0, \infty)$ exists for $Q(\boldsymbol{\mu}, \Sigma)$ if and only if*

$$G_p((n+1)k_0) > \frac{1}{2} \quad and \quad \sum_{m=0}^{\infty} b\left(m, \frac{n}{2}, \frac{1}{n+1}\right) G_{p+2m}((n+1)k_0) < \frac{1}{2}.$$

When such an estimator exists, k is the solution to $H(k_0 \mid ((n-p)/p)nk) = 1/2$.

We see that the evaluation of the optimal estimators requires the calculation of the k satisfying $1/2 = H(k_0 \mid ((n-p)/p)nk) = \sum_{m=0}^{\infty} b(m, n/2, p(((n-p)/p)nk)) \cdot G_{p+2m}((n+1)k_0)$. If we truncate this series at a value of m such that $G_{p+2m}((n+1)k_0) < \varepsilon$ and iteratively solve for k using a Newton-Raphson routine, then ignoring numerical errors, the true probability differs from $1/2$ by less than ε.

Let C_n be the optimal estimator for $n \geq N_\beta$ and $C_n^*(\bar{\mathbf{x}}, S(X)) = \{\mathbf{y} \mid (\mathbf{y} - \bar{\mathbf{x}})'S{-}1(X)(\mathbf{y} - \bar{\mathbf{x}}) \leq k_0/(n-1)\}$. Then

$$0 \leq E_{(\boldsymbol{\mu},\Sigma)}[d_{(\boldsymbol{\mu},\Sigma)}(C_n(\mathbf{x}, S(X)), Q(\boldsymbol{\mu}, \Sigma))] \leq E_{(\boldsymbol{\mu},\Sigma)}[d_{(\boldsymbol{\mu},\Sigma)}(C_n^*(\mathbf{x}, S(X)), Q(\boldsymbol{\mu}, \Sigma))].$$

As the elements of $\bar{\mathbf{x}}$, $(1/(n-1))S(X)$ converge almost surely to the respective elements of $\boldsymbol{\mu}$ and Σ we must have, for all $\mathbf{y}$, that $(n-1)(\mathbf{y} - \bar{\mathbf{x}})'S^{-1}(X)(\mathbf{y} - \bar{\mathbf{x}})$ converges almost surely to $(\mathbf{y} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})$. Accordingly if $\mathbf{y} \notin Q(\boldsymbol{\mu}, \Sigma)$ then

$$P_{(\boldsymbol{\mu},\Sigma)}^n(\mathbf{y} \in C_n^*(\bar{\mathbf{x}}, S(X)) \,\Delta\, Q(\boldsymbol{\mu}, \Sigma))$$
$$= P((n-1)(\mathbf{y} - \bar{\mathbf{x}})'S^{-1}(X)(\mathbf{y} - \bar{\mathbf{x}}) \leq k_0) \to 0 \quad \text{as} \quad n \to \infty$$

and similarly if $\mathbf{y} \in Q(\boldsymbol{\mu}, \Sigma)$ then $P_{(\boldsymbol{\mu},\Sigma)}^n(\mathbf{y} \in C_n^*(\bar{\mathbf{x}}, S(X)) \,\Delta\, Q(\boldsymbol{\mu}, \Sigma)) \to 0$ as $n \to \infty$. Then by the proposition of Section 2 we must have that $E_{(\boldsymbol{\mu},\Sigma)}[d_{(\boldsymbol{\mu},\Sigma)}(C_n^*(\bar{\mathbf{x}}, S(X)), Q(\boldsymbol{\mu}, \Sigma))] \to 0$ as $n \to \infty$. Therefore we have that C_n converges to $Q(\boldsymbol{\mu}, \Sigma)$ in the mean of the metric.

Events other than the ellipsoidal contours of the density function could also be of interest for the multivariate normal. For example we might be interested in estimating an infinite rectangle with one corner on a particular line in $\mathbb{R}^p$ and containing β of the probability. The approach developed here could be used to obtain optimal estimators of such events.

5. Conclusions. We have been concerned here with obtaining optimal estimators for events with positive probability content. We have applied our approach to univariate problems and also have shown that it leads to estimators in an important class of multivariate problems. Further it can be easily shown that the estimators obtained here agree with the formal Bayes procedures for these problems, with the same restriction on the estimates, following the development of Zidek (1969b).

We note that the optimal estimators obtained here may not exist for certain sample sizes; e.g. in the example of Section 3 we may have ∞ as the value of the estimate. From the examples we considered, it would seem to be the case that the further the set being estimated is from the center of the distribution, the more data we need to avoid this phenomenon. This would seem to be in accord with the intuitive idea that inferences about the tails of a distribution require more data than inference about the centre.

Acknowledgements. The author thanks the associate editor and referees for an extensive set of comments which proved to be very helpful. The research in this paper was supported by grant A3120 of the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. (1965). *Handbook of Mathematical Functions.* Ninth printing. Dover, New York.

ANDERSON, T. W. (1958). *An Introduction to Multivariate Analysis.* Wiley, New York.

ANGUS, J. E. and SCHAFER, R. E. (1979). Estimation of logistic quantiles with minimum error in the predicted distribution function. *Commun. Statist.-Theor. Meth.* **A8**(13) 1271–1284.

DYER, D. D., KEATING, J. P. and HENSLEY, O. L. (1977). Comparison of point estimators of normal percentiles. *Commun. Statist.-Simula. Computa.* **B6**(3) 269–283.

DYER, D. D. and KEATING, J. P. (1979). A further look at the comparison of normal percentile estimators. *Commun. Statist.-Theor. Meth.* **A8**(1), 1–16.

EVANS, M. and FRASER, D. A. S. (1980). An optimum tolerance region for multivariate regression. *J. Multivariate Anal.* **10** 268–272.

FRASER, D. A. S. and GUTTMAN, I. (1956). Tolerance regions. *Ann. Math. Statist.* **27** 162–179.

FRASER, D. A. S. (1976). *Probability and Statistics, Theory and Applications.* Duxbury Press, North Scituate, Mass.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

MANN, N. R. (1969). Optimum estimators for linear functions of location and scale parameters. *Ann. Math. Statist.* **40** 2149–2155.

OWEN, D. B. (1964). Control of percentages in both tails of the normal distribution. *Techn.* **6** 377–387.

REISS, R.-D. (1980). Estimation of quantiles in certain nonparametric models. *Ann. Statist.* **8** 87–105.

ROBBINS, H. E. (1944). On the measure of a random set. *Ann. Math. Statist.* **15** 70–74.

ROBERTSON, C. A. (1977). Estimation of quantiles of exponential distributions with minimum error in predicted distribution functions. *J. Amer. Statist. Assoc.* **72** 162–164.

SCHAFER, R. E. and ANGUS, J. E. (1979). Estimation of Weibull quantiles with minimum error in the distribution function. *Techn.* **21** 367–370.

ZIDEK, J. V. (1969a). Inadmissibility of the best invariant estimator of extreme quantiles of the normal law under squared error loss. *Ann. Math. Statist.* **40** 1801–1808.

ZIDEK, J. V. (1969b). A representation of Bayes invariant procedures in terms of Haar measure. *Ann. Inst. Statist. Math.* **22** 291–308.

ZIDEK, J. F. (1971). Inadmissibility of a class of estimators of a normal quantile. *Ann. Math. Statist.* **42** 1444–1447.

DEPARTMENT OF STATISTICS
UNIVERSITY OF TORONTO
TORONTO, CANADA M5S 1A1