ROBUSTNESS OF BAYES CLASSIFICATION REGION

GIITIRO SUZUKI

(Received Jan. 16, 1973; revised Jan. 30, 1974)

1. Introduction and summary

Let X be a random variable on a sample space $\mathcal{X}$ having density function $f_\lambda(x;\theta)$ for each θ in Ω, the parameter space and for each λ in Λ, the set of kernel indices*. We shall consider the two-way classification problem with the following loss structure:

action \ nature	ω_0	ω_1
a_0	0	α_0
a_1	α_1	0

$\alpha_0, \alpha_1>0$; $\omega_0\cap\omega_1=\phi$, $\omega_0\cup\omega_1=\Omega$

For every measurable subset A_0 of $\mathcal{X}$, define the non-randomized classification rule $\delta=\delta_{A_0}$ as $\delta(x)=a_i$ if $x\in A_i$, $i=0, 1$ (A_1 denoting the complement of A_0). Then the risk of such a rule δ is given by

$$R_\lambda(\delta, \theta)=\alpha_i\int_{A_i} f_\lambda(x;\theta)dx\ , \qquad \text{for } \theta\in\omega_{1-i},\ i=0, 1\ .$$

If we assume that θ is also a random variable having *a priori* probability measure η on Ω, the expected risk of δ is

$$\begin{aligned}R_\lambda(\delta, \eta)&=\int_\Omega R_\lambda(\delta, \theta)\eta(d\theta)\\&=\sum_{i=0}^{1}\alpha_i\int_{\omega_{1-i}}\eta(d\theta)\int_{A_i} f_\lambda(x;\theta)dx\\&=\alpha_1\int_{\omega_0}\eta(d\theta)+\int_{A_0} h_\lambda(x;\eta)dx\ ,\end{aligned}$$

* In a usual classification (decision) problem, we are given a fixed family $\mathcal{P}_\lambda=\{f_\lambda(\cdot\,;\theta):\theta\in\Omega\}$ of conditional densities for some $\lambda\in\Lambda$. In practical cases, however, we rarely have the complete knowledge of $\mathcal{P}_\lambda$. Rather we only know a set of possible families $\{\mathcal{P}_\lambda:\lambda\in\Lambda\}$, where λ is an element of some known index set Λ (which we call the set of kernel indicies), cf. Suzuki, G. (1969), "Bayes procedure with kernel index" *Res. Memo.* No. 25, Inst. Statist. Math.

where

$$h_\lambda(x;\eta)=\sum_{i=0}^{1}(-1)^i\alpha_i\int_{\omega_{1-i}}f_\lambda(x;\theta)\eta(d\theta)\ .$$

Then our main concern is to find a δ^* by which we incur the minimum attainable risk

$$R_\lambda^*(\eta)=\inf_\delta R_\lambda(\delta,\eta)=R_\lambda(\delta^*,\eta)$$

for some fixed η and λ, which may be unknown to us. Writing

$$A^*=A_\lambda^*(\eta)=\{x: h_\lambda(x:\eta)<0\}\ , \tag{1}$$

put $\delta^*=\delta_{A^*}$. Then it is the minimizing rule. $R_\lambda^*(\eta)$ and δ^* are called Bayes risk and Bayes decision rule. Therefore we shall call the set A^* given by (1) *Bayes classification region*.

For any subset A, define the regret function by

$$D_\lambda(A,\eta)=R_\lambda(\delta_A,\eta)-R_\lambda^*(\eta)\ .$$

Then it is easily seen that

$$D_\lambda(A;\eta)=\int_{A^*\triangle A}|h_\lambda(x;\eta)|dx\ ,$$

where $A^*\triangle A$ denotes the symmetric difference of A^* and A. When we can only use an approximate kernel index $\tilde\lambda$ and an approximate *a priori* measure $\tilde\eta$ instead of true λ and true η, we would take the decision $\tilde\delta=\delta_{\tilde A}^*$ $(\tilde A^*=A_{\tilde\lambda}^*(\tilde\eta))$, from which we have the regret value $D_\lambda(\tilde A^*;\eta)$. From the standpoint of practical applications, it would be a most important subject to find how the regret value is related to the "closeness" of some available value $(\tilde\lambda,\tilde\eta)$ and the true values (λ,η). In the next section we state some useful properties of Bayes classification region $A_\lambda^*(\eta)$. Then we discuss robustness of Bayes procedure in some parametric cases (Section 3).

2. Some properties of Bayes classification region

We first state some preliminary lemmas on a strong unimodal density. We say that f is a strongly unimodal symmetric density on its positive carrier $(-M, M)$ and denote by $f\in S_M$ if $r(t)=-\log f(t)$ is strictly convex i.e. $r''(t)>0$ and $r(-t)=r(t)$ on $(-M,M)$.

LEMMA 1. *Let* $f\in S_M$. *For any* $\rho>0$, $\beta_\rho(t)=\log f(t)/f(t+\rho)$ *is strictly increasing for* $-M<t<M-\rho$. *Furthermore when* M *is finite,* $\lim_{t\to M-\rho}\beta_\rho(t)=\infty$ *and* $\lim_{t\to -M}\beta_\rho(t)=-\infty$.

PROOF. The first assertion is the direct consequence of the strong unimodality of f. Next if M is finite, for any $K>0$, there exists t_0 ($<M-\rho$) such that

$$f(t_0+\rho)<f(M-\rho)e^{-K}$$

i.e.

$$r(t_0+\rho)>r(M-\rho)+K$$

consequently we have

$$\begin{aligned}\beta_\rho(t_0)&=r(t_0+\rho)-r(t_0)\\&=r(t_0+\rho)-r(M-\rho)+r(M-\rho)-r(t_0)>K\ ,\end{aligned}$$

which shows $\lim_{t\to M-\rho}\beta_\rho(t)=\infty$. Further $\lim_{t\to -M}\beta_\rho(t)=-\lim_{t'\to M-\rho}\beta_\rho(t')=-\infty$.

Remark. When $M=\infty$, there exists an $f\in S_\infty$ such that $\lim_{t\to\infty}\beta_\rho(t)<\infty$. For example, put $f(x)=e^{-x}/(1+e^{-x})^2$, i.e. logistic density function. Then $f\in S_\infty$ and

$$\beta_\rho(t)=\rho-2\log(1+e^{-t})/(1+e^{-t-\rho})\to\rho \qquad \text{as } t\to\infty\ .$$

Then we shall define the subclass S_M^* of S_M as $f\in S_M^*$ iff $f\in S_M$ and there exists a $\rho_0>0$ for which $\lim_{t\to M-\rho_0}\beta_{\rho_0}(t)=\infty$. From Lemma 1 $S_M^*=S_M$ for finite M. When M is infinite we have

LEMMA 2. *For $f\in S_\infty^*$ the second assertion of Lemma 1 is true. Furthermore if $\lim_{x\to\infty}r'(x)=\lim_{x\to\infty}\{-f'(x)/f(x)\}=\infty$, then $f\in S_\infty^*$.*

PROOF. The first assertion is easily shown using the following two inequalities: For any positive integer k

$$\beta_\rho(t)\leqq k\beta_{\rho/k}\left(t+\frac{k-1}{k}\rho\right)\ , \qquad \text{for } \rho,\ t>0\ .$$

If $0<\rho<\rho'$

$$\beta_\rho(t)<\beta_{\rho'}(t)\ , \qquad \text{for } t>0\ .$$

Next noting that

$$\beta_\rho(t)=r(t+\rho)-r(t)=\int_t^{t+\rho}r'(x)dx\geqq r'(t)\cdot\rho\to\infty \qquad (\text{as } x\to\infty)\ ,$$

we can show the second assertion of Lemma 2.

LEMMA 3. *Let $f\in S_M$. For any s, t ($|s|,|t|<M$) define*

$$\alpha(\rho)=\log\frac{f(\rho s)}{f(\rho t)} \qquad \textit{for } \frac{1}{M}<\rho<\frac{M}{\max(|s|,|t|)}\ .$$

Then $\alpha'(\rho)$ *has the same sign as* $|t|-|s|$.

PROOF. We first note that the function $-xr'(x)$ is symmetric and decreasing for $x>0$. For, because of the strong unimodality of f we have

$$x\frac{d}{dx}[-xr'(x)]=-x^2r''(x)-xr'(x)<0 .$$

Therefore for $0<s<t$

$$\rho\alpha'(\rho)=-\rho sr'(\rho s)+\rho tr'(\rho t)>0 .$$

When $-t<s<0$, by writing

$$\alpha(\rho)=\log\frac{f(\rho s)}{f(\rho t)}=\log\frac{f(\rho(-s))}{f(\rho t)}$$

we have $\alpha'(\rho)>0$ because $0<-s<t$. Consequently we have $\alpha'(\rho)>0$ for $t-s>0$ and $t+s>0$. Along the same lines we can prove other cases.

COROLLARY. *Let* $f\in S_M$. *For* $0<\lambda<\lambda'$ *define*

$$\varphi(s, t)=f\left(\frac{s}{\lambda}\right)f\left(\frac{t}{\lambda'}\right)-f\left(\frac{s}{\lambda'}\right)f\left(\frac{t}{\lambda}\right) .$$

Then $\varphi(s, t)$ *has the same sign as* $|t|-|s|$.

We shall now give some properties of Bayes classification region $A_\lambda^*(\eta)$. With slight loss of generality, we restrict ourselves to the case $\mathcal{X}=(-M_\lambda, M_\lambda)$, $\omega_0=(0, \infty)$, $w_1=(-\infty, 0]$, $\Omega=\omega_0\cup\omega_1$ and $f_\lambda(x;\theta)=f_\lambda(x-\theta)$.

THEOREM 1. *Let* $f_\lambda\in S_{M_\lambda}$ *for each fixed* $\lambda\in\Lambda$. *For any* η *such that* $\int_\Omega r_\lambda'(x-\theta)\eta(d\theta)>-\infty$ *for each* $x\in\mathcal{X}$, *put*

$$\Delta_\lambda(x;\eta)=-r_\lambda(x;\eta_0)+r_\lambda(x;\eta_1) , \tag{2}$$

where

$$r_\lambda(x;\eta_i)=-\log\int_{\omega_i}f_\lambda(x-\theta)\eta(d\theta) , \qquad i=0, 1 .$$

Then $\Delta_\lambda(x;\eta)$ *is a strictly increasing function of* x. *Furthermore, if* $f_\lambda\in S_{M_\lambda}^*$,

$$\Delta_\lambda(x;\eta)\to\pm\infty \qquad (\text{as } x\to\pm M_\lambda)$$

and the Bayes classification region is given by $A_\lambda^*(\eta)=(x^*, M_\lambda)$, *where* $x^*=x_\lambda^*(\eta)$ *is the unique solution of the equation*

$$\Delta_\lambda(x;\eta)=\alpha \qquad (\alpha=\log\alpha_0/\alpha_1) .$$

PROOF. Since

$$\frac{\partial}{\partial x}\Delta_\lambda(x;\eta)=\sum_{i=0}^{1}(-1)^i\left(\frac{\partial}{\partial x}\int_{\omega_i} f_\lambda(x-\theta)\eta(d\theta)\Big/\int_{\omega_i} f_\lambda(x-\theta)\eta(d\theta)\right)$$

$$=\sum_{i=0}^{1}(-1)^i\left(\int_{\omega_i} f_\lambda(x-\theta)\eta(d\theta)\Big/\int_{\omega_i} f_\lambda(x-\theta)\eta(d\theta)\right)$$

$$=\int_{s\leqq x\leqq t}\{r_\lambda'(t)-r_\lambda'(s)\}f_\lambda(s)f_\lambda(t)\eta(ds)\eta(dt)\Big/$$

$$\left(\int_{s\leqq x} f_\lambda(s)ds\int_{t\geqq x} f_\lambda(t)dt\right),$$

then we have $(\partial/\partial x)\Delta_\lambda(x;\eta)>0$ from the convexity of $r_\lambda(t)$.

Next for any $K>0$ there exists a $\rho>0$ such that

$$2(K+1)\int_{-\rho}^{0}\eta(d\theta)<\eta(\omega_0)\ .$$

For $f_\lambda\in S_{M_\lambda}^*$, from Lemma 2, there exists an $M>0$ such that

$$f_\lambda(M)/f_\lambda(M+\rho)>2(K+1)\eta(\omega_1)/\eta(\omega_0)$$

and

$$\int_0^M\eta(d\theta)>\frac{K}{K+1}\eta(\omega_0)\ .$$

Using these inequalities we have

$$\int_0^\infty f_\lambda(M-\theta)\eta(d\theta)\Big/\int_{-\infty}^{0} f_\lambda(M-\theta)\eta(d\theta)$$

$$\geqq\int_0^M f_\lambda(M-\theta)\eta(d\theta)\Big/\left(\int_{-\rho}^{0} f_\lambda(M-\theta)\eta(d\theta)+\int_{-\infty}^{-\rho} f_\lambda(M-\theta)\eta(d\theta)\right)$$

$$\geqq f_\lambda(M)\int_0^M\eta(d\theta)\Big/\left(f_\lambda(M)\int_{-\rho}^{0}\eta(d\theta)+f_\lambda(M+\rho)\eta(\omega_1)\right)>K\ ,$$

which shows that $\lim_{x\to M_\lambda}\Delta_\lambda(x;\eta)=\infty$. In the same way $\lim_{x\to -M_\lambda}\Delta_\lambda(x;\eta)=-\infty$. Noting that

$$A_\lambda^*(\eta)=\{x: h_\lambda(x;\eta)<0\}=\{x:\Delta_\lambda(x;\eta)>\alpha\}$$

the second assertion is proved.

Hereafter, we further restrict ourselves to the following parametric case. Let $f\in S_M^*$, $\Lambda=(1/M,\infty)$ and $f_\lambda(x-\theta)=(1/\lambda)f((x-\theta)/\lambda)$ i.e. λ a scale parameter. Clearly, $f_\lambda\in S_{M_\lambda}^*$, $M_\lambda=\lambda M$. Furthermore we assume that the *a priori* measure η is dominated by a σ-finite measure ν and its derivative $d\eta/d\nu=g_p$ is expressed as

(3) $$g_p(\theta)=[(1-i)p+i(1-p)]g_i(\theta)\ ,\qquad \theta\in\omega_i,\ i=0,1\ ,$$

where p is an unknown constant ($0 \leqq p \leqq 1$) and g_i are known functions such that $\int_{\omega_i} g_i(\theta)d\nu(\theta)=1$.

In this case, the function (2) can be expressed as

$$\Delta_\lambda(x\,;\eta)=\Delta_\lambda(x)-\log(1-p)/p\ ,$$

where

$$(4)\qquad \begin{aligned} &\Delta_\lambda(x)=\log f_\lambda(x\,;g_0)/f_\lambda(x\,;g_1)\\ &f_\lambda(x\,;g_i)=\int_{\omega_i} f\left(\frac{x-\theta}{\lambda}\right)g_i(\theta)d\theta\ , \qquad i=0, 1\ . \end{aligned}$$

Then

$$(5)\qquad \begin{aligned} &h_\lambda(x\,;\eta)\equiv h_\lambda(x\,;p)=\alpha_0(1-p)f_\lambda(x\,;g_1)-\alpha_1 p f_\lambda(x\,;g_0)\\ &A_\lambda^*(\eta)\equiv A_\lambda(p)=(x_\lambda(p),\ \lambda M)\ , \end{aligned}$$

where $x_\lambda(p)$ is the solution of the equation

$$(6)\qquad \Delta_\lambda(x)=\log(1-p)/p+\alpha \qquad (\alpha=\log \alpha_0/\alpha_1)\ .$$

From Theorem 1 we can see that for any fixed λ, $x_\lambda(p)$ is well-defined and strictly decreasing for $0<p<1$.

With slight restriction on the pair (g_0, g_1) we can also show the monotonicity of $x_\lambda(p)$ with respect to λ. For any $\rho>0$, define $\beta_\rho^+(\theta)=\log g_1(-\theta)/g_0(\theta+\rho)$ and $\beta_\rho^-(\theta)=\log g_0(\theta)/g_1(-\theta-\rho)$. Denote by $(g_0, g_1)\in S^p(\nu)$ iff both $\beta_\rho^+(\theta)$ and $\beta_\rho^-(\theta)$ are non-decreasing for $\theta>0$ and strictly increasing on a set of positive ν-measure.

LEMMA 4. *Let $\varphi(s, t)$ have the same sign as $|t|-|s|$ and $\varphi(s, t)=-\varphi(-t, -s)$. For any $(g_0, g_1)\in S^p(\nu)$ the integral*

$$I(x)=\int_0^\infty\int_{-\infty}^0 \varphi(x-s,\ x-t)g_1(t)d\nu(t)g_0(s)d\nu(s)$$

has the same sign as x.

PROOF. When $x>0$ we write

$$I(x)=I_1(x)+I_2(x)+I_3(x)\ ,$$

where

$$I_1(x)=\int_{2x}^\infty\int_{2x-s}^0 \varphi(x-s,\ x-t)g_1(t)d\nu(t)g_0(s)d\nu(s)$$

$$I_2(x)=\int_{2x}^\infty\int_{-\infty}^{2x-s} \varphi(x-s,\ x-t)g_1(t)d\nu(t)g_0(s)d\nu(s)$$

$$I_3(x)=\int_0^{2x}\int_{-\infty}^0 \varphi(x-s,\ x-t)g_1(t)d\nu(t)g_0(s)d\nu(s)\ .$$

We first note $I_3(x)\geqq 0$ from the assumption of φ and non-negativity of g_i. Next, by writing

$$\begin{aligned}I_1(x)+I_2(x)=&\int_0^\infty\int_0^s \varphi(-x-s,\ x+t)g_1(-t)d\nu(t)g_0(s+2x)d\nu(s)\\&+\int_0^\infty\int_0^t \varphi(-x-s,\ x+t)g_0(s+2x)d\nu(s)g_1(-t)d\nu(t)\\=&\int_0^\infty\int_0^s \varphi(-x-s,\ x+t)\{g_0(s+2x)g_1(-t)\\&-g_0(t+2x)g_1(-s)\}d\nu(t)d\nu(s)\ ,\end{aligned}$$

we have $I_1(x)+I_2(x)>0$ from the assumption on β_ρ^+ and φ and consequently $I(x)>0$. Along the same lines we can prove other cases.

Using the above lemma and corollary of Lemma 3, when $\lambda<\lambda'$ we have

$$f_\lambda(x\,;g_0)f_{\lambda'}(x\,;g_1)>f_{\lambda'}(x\,;g_0)f_\lambda(x\,;g_1)\ ,\qquad \text{for } x>0\ ,$$

which implies $\Delta_\lambda(x)>\Delta_{\lambda'}(x)$. Similarly, $\Delta_\lambda(x)<\Delta_{\lambda'}(x)$ for $x<0$. Consequently $\Delta_\lambda(0)=\Delta$ is independent of λ because of the continuity of $\Delta_\lambda(\cdot)$. Putting

$$p^*=1/(1+e^{\Delta-\alpha})\ , \tag{7}$$

we conclude that $x_\lambda(p)$ is a strictly increasing function of λ for each fixed p $(0<p<p^*)$ and if $p^*<p<1$, $x_\lambda(p)$ is a strictly decreasing function of λ.

Finally we shall show that the Bayes risk $R_\lambda^*(\eta)=R_\lambda(p)$ is a concave function of p for each fixed $\lambda\in\Lambda$ and is a strictly increasing function of λ for each fixed p. Since

$$R_\lambda(p)=\alpha_1 p+\int_{x_\lambda(p)}^{\lambda M} h_\lambda(x\,;p)dx\ , \tag{8}$$

then noting $h_\lambda(x_\lambda(p)\,;p)=0$, we have

$$\begin{aligned}\frac{\partial}{\partial p}R_\lambda(p)&=\alpha_1+\int_{x_\lambda(p)}^{\lambda M}\frac{\partial}{\partial p}h_\lambda(x\,;p)dx-\left[\frac{\partial}{\partial p}x_\lambda(p)\right]h_\lambda(x_\lambda(p)\,;p)\\&=\alpha_1-\sum_{i=0}^{1}\alpha_{i-1}\int_{x_\lambda(p)}^{\lambda M} f_\lambda(x\,;g_i)dx\ .\end{aligned}$$

From the monotonicity of $x_\lambda(\cdot)$

$$1-\int_{x_\lambda(p)}^{\lambda M} f_\lambda(x\,;g_i)dx=\int_{\omega_i}F\left(\frac{x_\lambda(p)-\theta}{\lambda}\right)g_i(\theta)d\nu(\theta)$$

is also strictly decreasing for p, where $F(t)=\int_{-M}^t f(x)dx$. This shows the concavity of $R_\lambda(\cdot)$. In a similar way we have

$$\frac{\partial}{\partial\lambda}R_\lambda(p)=\int_{x_\lambda(p)}^{\lambda M}\frac{\partial}{\partial\lambda}h_\lambda(x;p)dx=-\frac{1}{\lambda}\int_{x_\lambda(p)}^{\lambda M}h_\lambda(x;p)dx>0$$

because of the definitions of $x_\lambda(p)$ and $h_\lambda(x;p)$. Summarizing these we have

THEOREM 2. *Let $f_\lambda(x-\theta)=(1/\lambda)f((x-\theta)/\lambda)$, where $f\in S_M^*$ is a known function. Furthermore a priori density g_p is given by* (3), *where $(g_0, g_1)\in S^p(\nu)$ is known to us. Then the Bayes classification region is given by $A_\lambda^*(\eta)\equiv A_\lambda(p)=(x_\lambda(p), \lambda M)$ where $x_\lambda(p)$ is the solution of the equation* (6). *For any fixed λ $x_\lambda(p)$ is strictly decreasing for $0<p<1$ and for any fixed p $(\neq p^*,\ 0<p<1)$ $(p^*-p)x_\lambda(p)$ is strictly increasing for $\lambda>1/M$, where p^* is given by* (7). *Further, Bayes risk function $R_\lambda(p)$ is given by* (8) *and it is concave for p and is increasing for λ.*

3. Robustness of Bayes classification rule

First, we wish to investigate the normal distribution model. Let φ and Φ be the density function and the cummulative distribution function of the standard normal distribution.

LEMMA 5. (i) *For every $\rho>0$, define*

$$\gamma(\rho)=\begin{cases}\rho e^{(1-\rho^2)/2}, & 0<\rho\leqq 1,\\ \rho, & \rho\geqq 1.\end{cases} \tag{9}$$

Then

$$|\Phi(\rho\Phi^{-1}(s))-\Phi(\rho\Phi^{-1}(t))|\leqq\gamma(\rho)|s-t|.$$

(ii) *For any real number r,*

$$|\Phi(tr)-\Phi(r)|\leqq\frac{1}{\sqrt{2\pi e}}|\log t|, \qquad \textit{for } t>0.$$

PROOF. (i) Put $\alpha(s)=\Phi(\rho\Phi^{-1}(s))-r(\rho)s$. Then $\alpha'(s)=\rho\varphi(\rho\Phi^{-1}(s))/\varphi(\Phi^{-1}(s))-\gamma(\rho)=\rho e^{\Phi^{-1}(s^2)(1-\rho^2)/2}-\gamma(\rho)\leqq 0$. Therefore, for $s>t$ $\alpha(s)\leqq\alpha(t)$, namely

$$0<\Phi(\rho\Phi^{-1}(s))-\Phi(\rho\Phi^{-1}(t))\leqq r(\rho)(s-t).$$

(ii) When $r\geqq 0$, for $t\geqq 1$ we put $\beta(t)=\Phi(tr)-(1/\sqrt{2\pi e})\log t$. Then $\beta'(t)=r\varphi(tr)-(1/\sqrt{2\pi e})(1/t)=(1/t)\{tr\varphi(tr)-\varphi(1)\}\leqq 0$, i.e. $\beta(t)$ is decreasing for $t\geqq 1$. This shows

$$\Phi(tr)-\frac{1}{\sqrt{2\pi e}}\log t\leqq\beta(1)=\Phi(r)\ .$$

For $0<t<1$ we also have the inequality

$$0<\Phi(r)-\Phi(tr)=\Phi\Big(\frac{1}{t}(tr)\Big)-\Phi(tr)\leqq\frac{1}{\sqrt{2\pi e}}(-\log t)\ .$$

When $r<0$, we can also show the inequality in a similar way.

LEMMA 6. *For any positive ρ we have*

$$\left|\log\frac{t(t+\rho)}{1+\rho}\right|\leqq\begin{cases}2(1/t-1)\ , & 0<t<1\ ,\\ 2(t-1)\ , & t\geqq 1\ .\end{cases}$$

PROOF. For $t\geqq 1$, put $\xi(t)=\log t(t+\rho)-2t$. Then

$$\xi'(t)=\frac{2t+\rho}{t(t+\rho)}-2=\Big(\frac{2}{t+\rho}-\frac{2}{1+\rho}\Big)+\Big(\frac{\rho}{t(t+\rho)}-\frac{\rho}{1+\rho}\Big)-\frac{\rho}{1+\rho}<0\ .$$

Consequently $\varphi(t)\leqq\varphi(1)$ i.e. $\log(t(t+\rho)/(1+\rho))\leqq 2(t-1)$. For $0<t<1$, putting $s=1/t$, $\rho'=s\rho$, we have

$$\log\frac{1+\rho}{t(t+\rho)}=\log\frac{s(s+\rho')}{1+\rho'}\leqq 2(s-1)=2\Big(\frac{1}{t}-1\Big)\ .$$

We now put $f=\varphi$, $g_i(\theta)=2(1/\sigma)\varphi(\theta/\sigma)$, $\theta\in\omega_i$ (ν being Lebesgue measure). Clearly, $f\in S^*_\infty$ and $(g_0, g_1)\in S^p(\nu)$ and all conditions of Theorem 2 are fulfilled. Then the functions (4), (5) are

$$\varDelta_\lambda(x)=\log\Phi\Big(x\sqrt{\frac{\sigma^2}{\lambda^2(\lambda^2+\sigma^2)}}\Big/\Phi\Big(-x\sqrt{\frac{\sigma^2}{\lambda^2(\lambda^2+\sigma^2)}}\Big)$$

$$h_\lambda(x;p)=2[\alpha_0(1-p)+\alpha_1 p]\frac{1}{\sqrt{\lambda^2+\sigma^2}}\varphi\Big(\frac{x}{\sqrt{\lambda^2+\sigma^2}}\Big)$$
$$\cdot\Big\{\tau(p)-\Phi\Big(x\sqrt{\frac{\sigma^2}{\lambda^2(\lambda^2+\sigma^2)}}\Big)\Big\}\ ,$$

where

$$\tau(p)=\alpha_0(1-p)/[\alpha_0(1-p)+\alpha_1 p]\ .$$

Therefore the solution $x_\lambda(p)$ of (6) is given by

$$x_\lambda(p)=\sqrt{\frac{\lambda^2(\lambda^2+\sigma^2)}{\sigma^2}}\Phi^{-1}(\tau(p))\ . \tag{10}$$

Let $\tilde{p}<p$. Then $x_\lambda(p)<x_\lambda(\tilde{p})$ and $A_\lambda(p)\triangle A_\lambda(\tilde{p})=[x_\lambda(p), x_\lambda(\tilde{p}))$. Since $h_\lambda(x;p)<0$ and

$$0<\Phi\left(x\sqrt{\frac{\sigma^2}{\lambda^2(\lambda^2+\sigma^2)}}\right)-\tau(p)\leqq\tau(\tilde{p})-\tau(p) \qquad \text{for } x\in A_\lambda(p)\triangle A_\lambda(\tilde{p}) ,$$

then

$$\int_{A_\lambda(p)\triangle A_\lambda(\tilde{p})}|h_\lambda(x;p)|dx$$
$$=\int_{x_\lambda(p)}^{x_\lambda(\tilde{p})}-h_\lambda(x;p)dx$$
$$\leqq 2[\alpha_0(1-p)+\alpha_1 p][\tau(\tilde{p})-\tau(p)]\int_{x_\lambda(p)}^{x_\lambda(\tilde{p})}\frac{1}{\sqrt{\lambda^2+\sigma^2}}\varphi\left(\frac{x}{\sqrt{\lambda^2+\sigma^2}}\right)dx$$
$$=2[\alpha_0(1-p)+\alpha_1 p][\tau(\tilde{p})-\tau(p)]\left[\Phi\left\{\frac{\lambda}{\sigma}\Phi^{-1}(\tau(\tilde{p}))\right\}-\Phi\left\{\frac{\lambda}{\sigma}\Phi^{-1}(\tau(p))\right\}\right] .$$

Using Lemma 5 (i) we have

$$\int_{A_\lambda(p)\triangle A_\lambda(p')}|h_\lambda(x;p)|dx\leqq 2[\alpha_0(1-p)+\alpha_1 p]\gamma\left(\frac{\lambda}{\sigma}\right)[\tau(p)-\tau(\tilde{p})]^2$$
$$\leqq[\max(\alpha_0,\alpha_1)]^2\frac{2\gamma(\lambda/\sigma)}{\alpha_0(1-p)+\alpha_1(p)}(\tilde{p}-p)^2 .$$

For $\tilde{p}>p$, we have the same inequality.

Next we consider the sensitivity with respect to the scale parameter λ. Let $0<p<p^*=\alpha_0/(\alpha_0+\alpha_1)$ and $\lambda<\tilde{\lambda}$. Then $0<x_\lambda(p)<x_{\tilde{\lambda}}(p)$ and $A_\lambda(p)\triangle A_{\tilde{\lambda}}(p)=[x_\lambda(p), x_{\tilde{\lambda}}(p))$. Furthermore for $x\in[x_\lambda(p), x_{\tilde{\lambda}}(p))$ $h_\lambda(x;p)<0$ and

$$0<\Phi\left(x\sqrt{\frac{\sigma^2}{\lambda^2(\lambda^2+\sigma^2)}}\right)-\tau(p)\leqq\Phi\left(\frac{x_{\tilde{\lambda}}(p)}{x_\lambda(p)}\Phi^{-1}(\tau p)\right)-\tau(p) .$$

Putting

$$A=\Phi^{-1}(\tau(p)) , \qquad B=x_\lambda(p)/\sqrt{\lambda^2+\sigma^2} ,$$

we have

$$\frac{1}{2[\alpha_0(1-p)+\alpha_1 p]}\int_{A_\lambda(p)\triangle A_{\tilde{\lambda}}(p)}|h_\lambda(x;p)|dx$$
$$=\int_{x_\lambda(p)}^{x_{\tilde{\lambda}}(p)}\left[\Phi\left(x\sqrt{\frac{\sigma^2}{\lambda^2(\lambda^2+\sigma^2)}}\right)-\tau(p)\right]\frac{1}{\sqrt{\lambda^2+\sigma^2}}\varphi\left(\frac{x}{\sqrt{\lambda^2+\sigma^2}}\right)dx$$
$$\leqq\left[\Phi\left(\frac{x_{\tilde{\lambda}}(p)}{x_\lambda(p)}A\right)-\Phi(A)\right]\left[\Phi\left(\frac{x_{\tilde{\lambda}}(p)}{x_\lambda(p)}B\right)-\Phi(B)\right] .$$

Using Lemma 5 (ii) and Lemma 6 we have

$$\int_{A_\lambda(p)\triangle A_{\tilde\lambda}(p)} |h_\lambda(x;p)|\,dx \leqq 2[\alpha_0(1-p)+\alpha_1 p]\frac{1}{2\pi e}\left(\log\frac{x_{\tilde\lambda}(p)}{x_\lambda(p)}\right)^2$$
$$=\frac{\alpha_0(1-p)+\alpha_1 p}{\pi e}\left(\frac{1}{2}\log\frac{\tilde\lambda^2(\tilde\lambda^2+\sigma^2)}{\lambda^2(\lambda^2+\sigma^2)}\right)^2$$
$$\leqq\frac{\alpha_0(1-p)+\alpha_1 p}{\pi e\lambda^4}(\tilde\lambda^2-\lambda^2)^2 .$$

In a similar way we can prove that the same inequality holds in other cases.

Noting that

$$A_\lambda(p)\triangle A_{\tilde\lambda}(\tilde p)\subset[A_\lambda(p)\triangle A_\lambda(\tilde p)]\cup[A_\lambda(\tilde p)\triangle A_{\tilde\lambda}(\tilde p)]$$

we also have

$$\int_{A_\lambda(p)\triangle A_{\tilde\lambda}(\tilde p)} |h_\lambda(x;p)|\,dx$$
$$\leqq[\max(\alpha_0,\alpha_1)]^2\frac{2\gamma(\lambda/\sigma)}{\alpha_0(1-p)+\alpha_1 p}(\tilde p-p)^2+\frac{\alpha_0(1-\tilde p)+\alpha_1\tilde p}{\pi e\lambda^4}(\tilde\lambda^2-\lambda^2)^2$$
$$\leqq[\max(\alpha_0,\alpha_1)]^2\frac{2\gamma(\lambda/\sigma)}{\alpha_0(1-p)+\alpha_1 p}(\tilde p-p)^2+\frac{\max(\alpha_0,\alpha_1)}{\pi e\lambda^4}(\tilde\lambda^2-\lambda^2)^2 .$$

Summarizing these results we have

THEOREM 3. *In the normal distribution model stated above, when we can only use $\tilde p$ or/and $\tilde\lambda$ instead of p or/and λ, putting $\tilde A=A_\lambda(\tilde p)$ $(A_{\tilde\lambda}(p), A_{\tilde\lambda}(\tilde p))$ we take decision $\tilde\delta=\delta_{\tilde A}$. Then the regret values are given*

$$D_\lambda(A_\lambda(\tilde p);p)\leqq k_\lambda(p)(\tilde p-p)^2$$
$$D_\lambda(A_{\tilde\lambda}(p);p)\leqq l_\lambda(p)(\tilde\lambda^2-\lambda^2)^2$$
$$D_\lambda(A_{\tilde\lambda}(\tilde p);p)\leqq k_\lambda(p)(\tilde p-p)^2+l_\lambda(\tilde\lambda^2-\lambda^2)^2 ,$$

where

$$[\alpha_0(1-p)+\alpha_1 p]k_\lambda(p)=\begin{cases}[\max(\alpha_0,\alpha_1)]^2\dfrac{2\lambda}{\sigma}e^{(\sigma^2-\lambda^2)/2\sigma^2} , & 0<\lambda\leqq\sigma\\[2ex] [\max(\alpha_0,\alpha_1)]^2\dfrac{2\lambda}{\sigma} , & \lambda>\sigma\end{cases}$$

$$l_\lambda(p)=[\alpha_0(1-p)+\alpha_1 p]/\pi e\lambda^4 \qquad l_\lambda=\max(\alpha_0,\alpha_1)/\pi e\lambda^4 .$$

We next consider the dichotomy case. Let ν be the counting measure and $g_0(1)=g_1(-1)=1$. It is easily seen that $(g_0, g_1)\in S^p(\nu)$. For any $f\in S_M^*$ the functions (4), (5) are given by

$$\Delta_\lambda(x)=\log f\left(\frac{x-1}{\lambda}\right)\Big/ f\left(\frac{x+1}{\lambda}\right),$$

$$h_\lambda(x;p)=\alpha_0(1-p)\frac{1}{\lambda}f\left(\frac{x+1}{\lambda}\right)-\alpha_1 p\frac{1}{\lambda}f\left(\frac{x-1}{\lambda}\right)$$

and $x_\lambda(p)$ is well-defined for $0<p<1$, $\lambda>1/M$. Furthermore the Bayes risk and the regret value of some non-Bayes rule are

$$R_\lambda(p)=pF\left(\frac{x_\lambda(p)-1}{\lambda}\right)+(1-p)\left[1-F\left(\frac{x_\lambda(p)+1}{\lambda}\right)\right]$$

$$D_\lambda(A_{\tilde\lambda}(\tilde p);p)=H_\lambda(x_\lambda(p);p)-H_\lambda(x_{\tilde\lambda}(\tilde p);p),$$

where

$$F(t)=\int^t f(x)dx$$

$$H_\lambda(t;p)=\int^t h_\lambda(x;p)dx=\alpha_0(1-p)F\left(\frac{t+1}{\lambda}\right)-\alpha_1 pF\left(\frac{t-1}{\lambda}\right).$$

Using Taylor expansion (noting $h_\lambda(x_\lambda(p);p)=0$), we have the following inequalities.

$$D_\lambda(A_{\tilde\lambda}(\tilde p);p)\leqq c_1(\lambda)\,|x_{\tilde\lambda}(\tilde p)-x_\lambda(p)|$$

$$D_\lambda(A_{\tilde\lambda}(\tilde p);p)\leqq c_2(\lambda)\,(x_{\tilde\lambda}(\tilde p)-x_\lambda(p))^2,$$

where

$$c_1(\lambda)=\max(\alpha_0,\alpha_1)f(0)/\lambda$$

$$c_2(\lambda)=\max(\alpha_0,\alpha_1)\cdot\max_t|f'(t)|\,2\lambda^2.$$

For example, if $f(x)=(1/\sqrt{2\pi})e^{-x^2/2}$ then $x_\lambda(p)=(\lambda^2/2)\log((1-p)/p)$. If $f(x)=1-|x|$ (for $|x|<1$) we have $x_\lambda(p)=(1-2p)(\lambda-1)$ and $R_\lambda(p)=2(1-1/\lambda)^2p(1-p)$ for $1<\lambda<2$. When $\lambda>2$, putting $\lambda^*=(\lambda-2)/2(\lambda-1)$ $(0<\lambda^*<1/2)$ we have

$$x_\lambda(p)=\begin{cases}\lambda-\dfrac{1}{1-2p}, & 0<p<\lambda^*,\\ (1-2p)(\lambda-1), & \lambda^*<p<1-\lambda^*,\\ \dfrac{1}{2p-1}-\lambda, & 1-\lambda^*<p<1,\end{cases}$$

$$R_\lambda^*(p)=\begin{cases} p+\dfrac{2p(1-p)}{2p-1}\dfrac{1}{\lambda^2}\,, & 0<p<\lambda^*\,, \\ 2p(1-p)\left(1-\dfrac{1}{\lambda}\right)^2, & \lambda^*<p<1-\lambda^*\,, \\ 1-p-\dfrac{2p(1-p)}{2p-1}\dfrac{1}{\lambda^2}\,, & 1-\lambda^*<p<1\,. \end{cases}$$

INSTITUTE OF STATISTICAL MATHEMATICS