When is C-R bound attained?

$\delta(X)$ perfectly correlated with $\frac{d}{d\theta} \log f(X|\theta)$

$$\Rightarrow \quad a + b\,\delta(x) = \frac{d}{d\theta} \log f(x|\theta)$$

$$\Rightarrow \quad (a + b\,\delta(x))\,\theta = \log f(x|\theta) + c$$

$$f(x|\theta)\,\exp c = \exp(a\theta + b\delta(x)\theta)$$

$$f(x|\theta) = \exp(\,b\delta(x)\,\theta + a\theta - c)$$

i.e. This must be an exponential family with natural parameter $b\theta$ and natural sufficient statistic $\delta(X)$

Jump now to Ch4 of B+D and hypothesis testing — we've already said some things about it in decision-theoretic terms — to review we have

$$\Theta = \Theta_0 \sqcup \Theta_1$$

and we want to decide between

$$H_0 : \theta \in \Theta_0 \quad \text{and} \quad H_1 : \theta \in \Theta_1$$

a decision-theoretic approach sets

$\mathcal{A} = \{0,1\}$, 0-1 loss is

$$L(\theta, a) = I[\theta \in \Theta_0] I[a=1] + I[\theta \in \Theta_1] I[a=0]$$

$\phi(x)$ a decision rule/"test"

$\phi(x) = 0$ means "accept H_0"

$= 1$ means "reject H_0 in favor of H_1"

$$R(\theta, \phi) = E_\theta L(\theta, \phi(X))$$

$$= I[\theta \in \Theta_0] E_\theta I[\phi(X)=1] + I[\theta \in \Theta_1] E_\theta [\phi(X)=0]$$

$$= I[\theta \in \Theta_0] P_\theta[\phi(X)=1] + I[\theta \in \Theta_1] \left(1 - P_\theta[\phi(X)=1]\right)$$

The function $\pi(\theta) = P_\theta[\phi(X)=1]$ is usually called the "power function" of the test (especially in non-decision-theoretic treatments of testing)

$\{x \mid \phi(x) = 1\}$ is often called the "rejection region" for ϕ and $\{x \mid \phi(x) = 0\}$ is called the "acceptance region" for test ϕ

In choosing a test we want small risk function i.e. we want small risk function, i.e. small power for $\theta \in \Theta_0$ and large power for $\Theta \in \Theta_1$ — how to achieve this?

We've got a start on this in terms of Bayes risk, where we adopt a prior den $g(\Theta)$ for Θ — we know we want to find the posterior probability that $\Theta \in \Theta_0$ and that $\Theta \in \Theta_1$ and simply take

$$\phi(x) = I\left[G(\Theta_1 | x) > G(\Theta_0 | x) \right]$$

to minimize Bayes risk

$$R(G, \phi) = \int R(\Theta, \phi)\, dG(\Theta)$$

What non-Bayes theory is there? Start with simple vs. simple $\Theta_0 = \{0\}$ and $\Theta_1 = \{1\}$ (2-class classification case)

In this case we know that Bayes optimal tests are of the form

$$\phi(x) = I\left[\underbrace{R(x)}_{\frac{f(x|1)}{f(x|0)}} > \underbrace{\frac{g(0)}{g(1)}}_{k} \right]$$

These tests have properties that can be discussed in non-Bayes terms

Call

$$\Pi_\phi(0) = \text{"}\alpha\text{"}$$
$$= \text{"Type I error probability"}$$
$$= \text{the size of the test}$$

$$\Pi_\phi(1) = 1 - \text{Type II error probability}$$
$$= \text{the power of the test}$$

$$\left(\text{"}\beta\text{"} = 1 - \Pi_\phi(1) \quad \text{The type II error probability} \right)$$

Theorem Neyman-Pearson Lemma (Part I - sufficiency) - If

$$\phi(x) = \begin{cases} 1 & \text{if } R(x) > k \\ 0 & \text{if } R(x) < k \end{cases}$$

for real $k > 0$
then ϕ is most powerful of its size
(Bayes simple vs simple tests are most powerful of their size)

Pf
$$\alpha = \Pi_\phi(0) = P_0[\phi(X) = 1]$$
$$= E_0\, \phi(X)$$

Suppose that ϕ' is a competing test with

$$\Pi_{\phi'}(0) \le \alpha$$

and consider

$$E_1\left[\phi(X)-\phi'(X)\right] - k\,E_0\left[\phi(X)-\phi'(X)\right]$$

cont's case
$$= \int \left[\phi(x)-\phi'(x)\right] f(x|1) - k\left[\phi(x)-\phi'(x)\right] f(x|0)\,dx$$

$$= \int \left[\phi(x)-\phi'(x)\right]\left[\frac{f(x|1)}{f(x|0)} - k\right] f(x|0)\,dx$$

when $R(x) = \dfrac{f(x|1)}{f(x|0)} > k \quad \phi(x) = 1$ so

$\phi(x)-\phi'(x) \geq 0$ while when $L(x) < k \;\; \phi(x) = 0$
and so $\phi(x)-\phi'(x) \leq 0$ Thus

$$\left[\phi(x)-\phi'(x)\right]\left[R(x)-k\right] \geq 0 \quad \forall\, x$$

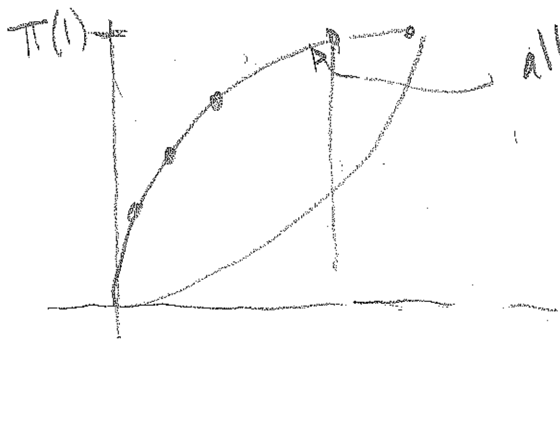
and $\therefore \int$ above $f_1(x)\,dx \geq 0$

So $\quad E_1\left[\phi(X)-\phi'(X)\right] \geq k\,E_0\left[\phi(X)-\phi'(X)\right]$

$$\pi_\phi(1) - \pi_{\phi'}(1) \geq k\left[\pi_\phi(0) - \pi_1(0)\right]$$

≥ 0 by hypothesis

So test of size no more than α has better power than such a ϕ



all $(\pi_\phi(0), \pi_\phi(1))$ points possible — Bayes tests produce upper boundary of the set of such pairs

Alternative (better) proof of N-P Lemma Part I

Suppose ϕ is of the advertised form. Then ϕ is 0-1 loss Bayes for $k = \frac{g(0)}{1-g(0)}$ i.e. for $g(0) = \frac{k}{1+k}$. Then if the Lemma is not true, then $\exists$ a competing test ϕ' with

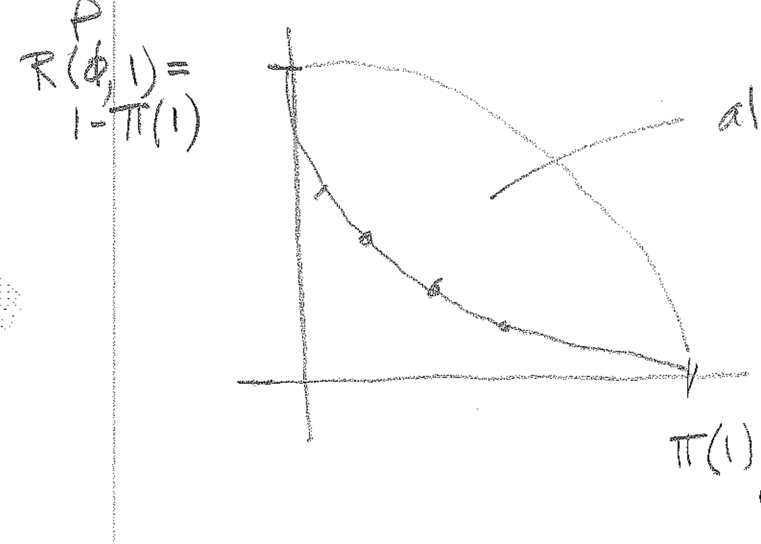
$$\pi_{\phi'}(0) \le \pi_\phi(0) \text{ and } \pi_{\phi'}(1) > \pi_\phi(1)$$

But then

$$\begin{aligned} R(\phi'; g) &= g(0)\, R(0,\phi') + g(1)\, R(1,\phi') \\ &= g(0)\, P_0[\phi'(X)=1] + g(1)\, P_1[\phi'(X)=0] \\ &= g(0)\, \underset{\le \pi_\phi(0)}{\pi_{\phi'}(0)} + g(1)\, \underbrace{(1-\pi_{\phi'}(1))}_{< 1-\pi_\phi(1)} \\ &< R(\phi, g) \end{aligned}$$

and ϕ is not 0-1 loss Bayes. ∎

An alternative version of previous cartoon is



all $(\pi_\phi(0), 1-\pi_\phi(1)) = (\alpha, \beta)$ points possible

Bayes tests are on the lower boundary of the set of such pairs

Example $N(\mu_0, 1)$ vs $N(\mu_1, 1)$ $\mu_0 < \mu_1$

$$L(x) = \exp\left(-\tfrac{1}{2}(x^2 - 2x\mu_1 + \mu_1^2) + \tfrac{1}{2}(x^2 - 2x\mu_0 + \mu_0^2)\right)$$

$$= \exp\left(x(\mu_1 - \mu_0)\right)\exp\left(-\tfrac{1}{2}(\mu_1^2 - \mu_0^2)\right)$$

$\nearrow$ in x

So Bayes test's reject for large x

$$\phi_c(x) = I[x > c]$$

$$\alpha = P_0[X > c] = 1 - \Phi(c - \mu_0)$$

for a given value of this

$$\pi_{\phi_c}(1) = 1 - \Phi(c - \mu_1)$$

you can't beat that

Example Extremely simple/artificial discrete one

Suppose $f(x|\theta)$ for $\theta = 0, 1$ are given as below

	$x=1$	$x=2$	$x=3$
$f(x\|1)$	$1/3$	$1/6$	$1/2$
$f(x\|0)$	$1/3$	$1/3$	$1/3$
$R(x)$	1	$\frac{1}{2}$	$\frac{3}{2}$

$\phi_1(x) = 1$ has size 1 power $\pi_{\phi_1}(1) = 1$

$\phi_2(x) = I[x=3 \text{ or } x=1]$

has size $\frac{2}{3}$ power $\pi_{\phi_2}(1) = \frac{5}{6}$

$\phi_3(x) = I[x=3]$

has size $\frac{1}{3}$ power $\pi_{\phi_3}(1) = \frac{1}{2}$

$\phi_4(x) = 0$

has size 0 power $\pi_{\phi_4} = 0$

These are by N-P guaranteed to be MP of their sizes –

In cases where the statistic $R(X)$ has a dsn conts under $\theta = 0,1$, it's possible to adjust k and get a MP test of any size α of interest – where, as in the discrete X example immediately above, X has a discrete dsn one can get tests of only certain sizes ... unless (mostly for the sake of theoretical completeness) one extends the notion of a test

$$\phi : \mathcal{X} \longrightarrow \{0,1\}$$

to the notion of a "randomized test" –

<u>Def</u> A function $\phi : \mathcal{X} \longrightarrow [0,1]$ is called a randomized test (a randomized decision function)

The interpretation is that if I observe $X = x$ I choose randomly between $a=0$ and $a=1$, with probability $\phi(x)$ assigned to $a=1$

Example simple discrete one

$$\phi_2(x) = I[x=3 \text{ or } x=1] \text{ has size } \tfrac{2}{3}$$

$$\phi_3(x) = I[x=3] \text{ has size } \tfrac{1}{3}$$

Let $\phi(x) = \begin{cases} 1 & \text{if } x=3 \\ .5 & \text{if } x=1 \\ 0 & \text{if } x=2 \end{cases}$

Then $P_0[a=1 \text{ is selected}] = P_0[X=3] + \tfrac{1}{2}P_0[X=1]$

$$= \tfrac{1}{3} + \tfrac{1}{2}(\tfrac{1}{3})$$

$$= E_0\,\phi(X) = \tfrac{1}{2}$$

$$= \pi_\phi(0) = \alpha$$

And this is a test of the form

$$\phi(x) = \begin{cases} 1 & \text{if } R(x) > 1 \\ .5 & \text{if } R(x) = 1 \\ 0 & \text{if } R(x) < 1 \end{cases}$$

i.e. it is a Bayes test (of N-P form) and is therefore MP of its size!

(Bayes form only requires $>$ or $<$ $\frac{g(0)}{g(1)} = k$,

it is silent on what to do if $= k$ — you can do either or even some both without hurting Bayesness!)

In fact, there is a 2nd part of the N-P Lemma that promises that it's always possible to do this kind of thing and get a best size α (potentially randomized) test —

<u>Thm</u> N-P Lemma Part II (Existence)

If $\alpha \in (0,1]$ $\exists$ a test of the form

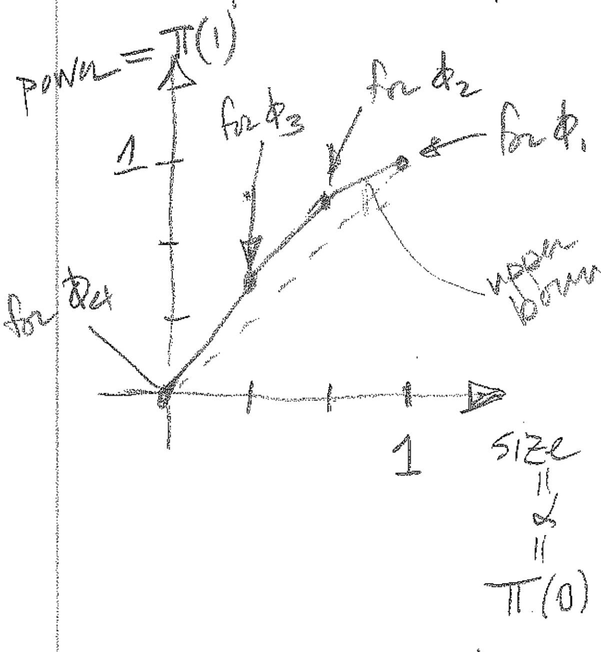
$$\phi(z) = \begin{cases} 1 & \text{if } R(z) > k \\ \gamma & \text{if } \quad = \\ 0 & \quad\quad < \end{cases}$$

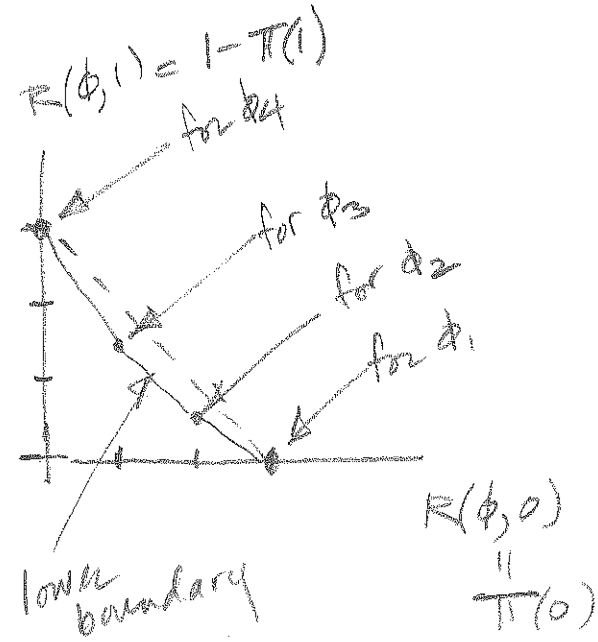
for some $0 < k < \infty$ and $0 \le \gamma \le 1$ that is most powerful of size α for testing $H_0 : \theta = 0$ versus $H_1 : \theta = 1$

(details of proof are not really hard, but are tedious... it's not something to waste time on in Stat ~~542~~)

This is primarily of theoretical importance, as most people would balk at using a randomized test – it does, however, "fill in" the picture of risk functions for discrete problems

<u>Example</u> Simple discrete





The line segments are filled in by the randomized tests

What about testing for problems more complicated than simple vs simple?

Again, <u>Bayes optimal</u> 0-1 loss tests are easy to identify in general terms (if hard to implement in many/most specific contexts) -

For $\theta \sim g(\theta)$ a Bayes optimal test of $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ is

$$\phi(x) = \begin{cases} 1 & \text{if } P[\theta \in \Theta_1 \mid X = x] > \frac{1}{2} \\ 0 & \qquad\qquad\qquad\qquad\quad < \end{cases}$$

$$= \begin{cases} 1 & \text{if } \dfrac{P[\theta \in \Theta_1 \mid X = x]}{P[\theta \in \Theta_0 \mid X = x]} > 1 \\ 0 & \qquad\qquad\qquad\qquad\qquad < \end{cases}$$

Note that in cont$^{\underline{s}}$ cases for example, that ratio is

$$\frac{\int_{\Theta_1} f(x|\theta) g(\theta) d\theta}{\int_{\Theta_0} f(x|\theta) g(\theta) d\theta}$$

and g integrals of $f(x|\theta)$ over Θ_1 and Θ_0 replace $f(x|\theta_1)$ and $f(x|\theta_0)$ that are in numerator and denominator of $R(x)$ for simple vs simple