**Statistics 582, Problem Set 8 Solutions**

Wellner; 3/4/2010

1. Lehmann and Casella, problem 5.4, page 289 (with a modification to the posterior variance formula):
Albert and Gupta (1985) investigate theory and applications of the hierarchical model

$$\begin{aligned} X_i|\boldsymbol{\theta}_i &\sim \text{Binomial}(n, \boldsymbol{\theta}_i), \quad i = 1, \ldots, p, \quad \text{independent} \\ \boldsymbol{\theta}_i|\eta &\sim \text{Beta}(k\eta, k(1-\eta)), \quad k \text{ known}, \\ \eta &\sim \text{Uniform}(0,1). \end{aligned}$$

(a) Show that

$$E(\boldsymbol{\theta}_i|\underline{X}) = \left(\frac{n}{n+k}\right)\frac{X_i}{n} + \frac{k}{n+k}E(\eta|\underline{X}),$$

$$Var(\boldsymbol{\theta}_i|\underline{X}) = \frac{k^2}{(n+k)(n+k+1)}Var(\eta|\underline{X}) + \frac{E(\boldsymbol{\theta}_i|\underline{X})(1-E(\boldsymbol{\theta}_i|\underline{X}))}{n+k+1}.$$

(b) Show that the $\boldsymbol{\theta}_i$'s have posterior covariance

$$Cov(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j|\underline{X}) = \left(\frac{k}{n+k}\right)^2 Var(\eta|\underline{X}), \quad i \neq j.$$

(c) Ignoring the prior distribution of $\eta$, show how to construct an empirical Bayes estimator of $\boldsymbol{\theta}_i$. (This might not be expressible in simple closed form.)

**Solution:** (a) Conditionally on $\eta$, the posterior distribution of the $\boldsymbol{\theta}_i$'s is that of independent $\text{Beta}(X_i + k\eta, n - X_i + k(1-\eta))$ random variables. Since the mean and variance of $Y \sim \text{Beta}(\alpha, \beta)$ are given by $E(Y) = \alpha/(\alpha+\beta)$ and $Var(Y) = \alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$, it follows that

$$\begin{aligned} E\{\boldsymbol{\theta}_i|\underline{X}\} &= EE\{\boldsymbol{\theta}_i|\eta, \underline{X}\} = E\left(\frac{X_i + k\eta}{n+k}|\underline{X}\right) \\ &= \frac{n}{n+k}\frac{X_i}{n} + \frac{k}{n+k}E(\eta|\underline{X}), \end{aligned}$$

and

$$\begin{aligned} Var(\boldsymbol{\theta}_i|\underline{X}) &= E\left\{Var(\boldsymbol{\theta}_i|\eta, \underline{X})|\underline{X}\right\} + Var\{E(\boldsymbol{\theta}_i|\eta, \underline{X})\} \\ &= E\left\{\frac{(X_i + k\eta)(n - X_i + k(1-\eta))}{(n+k)^2(n+k+1)}\Big|\underline{X}\right\} + Var\left\{\frac{X_i + k\eta}{n+k}|\underline{X}\right\} \\ &= \frac{1}{n+k+1}E\{\hat{a}_i(1-\hat{a}_i)|\underline{X}\} + \left(\frac{k}{n+k}\right)^2 Var(\eta|\underline{X}) \qquad (1) \end{aligned}$$

with $\hat{a}_i \equiv (X_i + k\eta)/(n+k)$. Now $E(\hat{a}_i|\underline{X}) = E\{\boldsymbol{\theta}_i|\underline{X}\}$, and we can write

$$\hat{a}_i = \hat{a}_i - E(\hat{a}_i|\underline{X}) + E(\hat{a}_i|\underline{X}) \equiv c + d$$

with $c \equiv \hat{a}_i - E(\hat{a}_i|\underline{X})$ and $d \equiv E(\hat{a}_i|\underline{X})$. Then

$$\hat{a}_i(1-\hat{a}_i) = (c+d)(1-(c+d)) = c - 2cd + d(1-d) - c^2$$

where $E(c|\underline{X}) = E\{E(\hat{a}_i - E(\hat{a}_i|\underline{X}))|\underline{X}\} = 0$ a.s. and

$$E(dc|\underline{X}) = E\{E(\hat{a}_i|\underline{X})(\hat{a}_i - E(\hat{a}_i|\underline{X}))|\underline{X}\} = 0 \quad \text{a.s.}$$

Thus it follows that

$$\begin{aligned} E\{\hat{a}_i(1-\hat{a}_i)|\underline{X}\} &= E\{d(1-d)|\underline{X}\} - E\{c^2|\underline{X}\} \\ &= E(\boldsymbol{\theta}_i|\underline{X})(1-E(\boldsymbol{\theta}_i|\underline{X})) - Var(\hat{a}_i|\underline{X}) \\ &= E(\boldsymbol{\theta}_i|\underline{X})(1-E(\{\boldsymbol{\theta}_i|\underline{X})) - \frac{k^2}{(n+k)^2} Var(\eta|\underline{X}). \qquad (2) \end{aligned}$$

Combining (1) and (2) yields

$$\begin{aligned} Var(\boldsymbol{\theta}_i|\underline{X}) &= \frac{1}{n+k+1}\left\{E(\boldsymbol{\theta}_i|\underline{X})(1-E(\boldsymbol{\theta}_i|\underline{X})) - \frac{k^2}{(n+k)^2}Var(\eta|\underline{X})\right\} \\ &\quad + \left(\frac{k}{n+k}\right)^2 Var(\eta|\underline{X}) \\ &= \frac{1}{n+k+1}E(\boldsymbol{\theta}_i|\underline{X})(1-E(\boldsymbol{\theta}_i|\underline{X})) + \frac{k^2}{(n+k)(n+k+1)}Var(\eta|\underline{X}). \end{aligned}$$

(b) Similarly,

$$\begin{aligned} Cov(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j|\underline{X}) &= E\{Cov(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j|\eta, \underline{X})|\underline{X}\} + Cov[E(\boldsymbol{\theta}_i|\eta,\underline{X}), E(\boldsymbol{\theta}_j|\eta,\underline{X})|\underline{X}] \\ &= E\{0|\underline{X}\} + Cov\left[\frac{X_i+k\eta}{n+k}, \frac{X_j+k\eta}{n+k}\Big|\underline{X}\right] \\ &= \left(\frac{k}{n+k}\right)^2 Var(\eta|\underline{X}). \end{aligned}$$

(c) The joint distribution of $\underline{X} = (X_1, \ldots, X_p)$ and $\underline{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)$ is given by

$$p(\underline{x}, \underline{\theta}; \eta, k) = \prod_{i=1}^{p}\left\{\binom{n}{x_i}\theta_i^{x_i+k\eta-1}(1-\theta_i)^{n-x_i+k(1-\eta)-1}\frac{\Gamma(k)}{\Gamma(k\eta)\Gamma(k(1-\eta))}\right\}.$$

As we know, the Bayes estimators of $\theta_i$ for known $\eta$ and $k$ are given by

$$d_i(X_i) = E(\boldsymbol{\theta}_i|X_i) = (X_i + k\eta)/(n+k), \quad i = 1, \ldots, p. \qquad (3)$$

When $\eta$ (and / or $k$) are unknown, the empirical Bayes approach is to estimate these by maximum likelihood from the marginal distribution of $\underline{X}$. In the present case, the marginal distribution of $\underline{X}$ is obtained by integrating over $\underline{\theta}$ in the joint density given in the last display: thus

$$\begin{aligned} p(\underline{x};\eta,k) &= \int_0^1 \cdots \int_0^1 \prod_{i=1}^p \binom{n}{x_i} \frac{\Gamma(k)}{\Gamma(k\eta)\Gamma(k(1-\eta))} \theta_i^{x_i+k\eta-1}(1-\theta_i)^{n-x_i+k(1-\eta)-1} d\underline{\theta} \\ &= \prod_{i=1}^p \binom{n}{x_i} \frac{\Gamma(k)}{\Gamma(k\eta)\Gamma(k(1-\eta))} \int_0^1 \theta_i^{x_i+k\eta-1}(1-\theta_i)^{n-x_i+k(1-\eta)-1} d\theta_i \\ &= \prod_{i=1}^p \binom{n}{x_i} \frac{\Gamma(k)}{\Gamma(k\eta)\Gamma(k(1-\eta))} \frac{\Gamma(x_i+k\eta)\Gamma(n-x_i+k(1-\eta))}{\Gamma(n+k)}. \end{aligned}$$

If we regards $k$ as fixed and known, then the empirical Bayes estimators $\hat{d}_i(X_i)$ of $\theta_i$ are obtained by replacing $\eta$ in (3) by $\hat{\eta}$, the MLE of $\eta$ based on the marginal distribution $p(\underline{x};\eta,k)$ in the last display. This estimator can be calculated numerically by maximizing $\log p(\underline{X};\eta,k)$ as a function of $\eta$, or by solving the score equation

$$0 = \dot{l}_\eta(\underline{X}) = \sum_{i=1}^p k\{\psi(X_i+k\eta) - \psi(n-X_i+k(1-\eta)) - \psi(k\eta) + \psi(k(1-\eta))\}$$

where $\psi = \Gamma'/\Gamma$ is the digamma function.

2. Suppose that $X_i \sim \text{Binomial}(n,\theta_i)$ are independent for $i = 1,\ldots,k$ conditionally on $\boldsymbol{\theta}_i \sim \text{Beta}(a,b)$. Let $L(\underline{\theta},\underline{a}) = \sum_{i=1}^k(\theta_i - a_i)^2$.
(a) Find the Bayes rules $d(\underline{X}) \equiv (d_i(X_i) = d_i(X_i;a,b),\ i = 1,\ldots,k)$ for estimation of $\boldsymbol{\theta}_i$ for $i = 1,\ldots,k$, assuming that $a,b > 0$ are known.
(b) Find the marginal distribution (or mass function) of $\underline{X} = (X_1,\ldots,X_k)$.
(c) Consider estimation of the parameters $a,b$ by maximizing the marginal likelihood resulting from part (b). Denote the resulting estimators of $a,b$ by $\hat{a},\ \hat{b}$. (These will not be expressible in closed form in general.)
(d) Now consider the *empirical Bayes* rules

$$\hat{d}(\underline{X},\hat{a},\hat{b}) \equiv (d_i(X_i,\hat{a},\hat{b}),\ i = 1,\ldots,k).$$

(e) Compare the Bayes risks of $d(\underline{X})$, $\hat{d}(\underline{X},\hat{a},\hat{b})$, and $\underline{X}/n$ when $a = b = 2$, $n = 20$, and $k = 10$. [Hint: see Lehmann and Casella, pages 263-266, including Table 6.1 and Theorem 6.3 on page 265, and (6.16) on page 266.]

**Solution:** This is closely related to problem 5.6.5 on page 295 of Lehmann and Casella.

(a) As in problem 1, the Bayes estimators of $\theta_i$ for fixed $a, b$ are given by

$$d_i(X_i) \equiv d_i(X_i; a, b) = \frac{a + X_i}{a+b+n}, \quad i = 1, \ldots, k.$$

(b) The marginal distribution of $\underline{X}$ is given by

$$\begin{aligned} p(\underline{x}; a, b) &= \int_0^1 \cdots \int_0^1 \prod_{i=1}^p \binom{n}{x_i} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_i^{x_i+a-1}(1-\theta_i)^{n-x_i+b-1} d\underline{\theta} \\ &= \prod_{i=1}^p \binom{n}{x_i} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(x_i+a)\Gamma(n-x_i+b)}{\Gamma(n+a+b)}. \end{aligned}$$

(c) The MLE's $\hat{a}, \hat{b}$ of $a, b$ maximize $l(a, b|\underline{X}) = \log p(\underline{X}; a, b)$.
(d) The empirical Bayes estimators of $\theta_i$ are given by

$$\hat{d}_i(X_i) \equiv d_i(X_i; \hat{a}, \hat{b}) = \frac{\hat{a} + X_i}{\hat{a}+\hat{b}+n}, \quad i = 1, \ldots, k$$

where $\hat{a}, \hat{b}$ are the MLE's based on the marginal likelihood as in (b), (c).
(e) Now the ordinary risks of the Bayes rules are given by

$$\begin{aligned} R_i(\theta_i, d_i) &= E_{\theta_i}(\theta_i - d_i(X_i))^2 = Var_{\theta_i}(d_i(X_i)) + \{\text{bias}_{\theta_i}(d_i(X_i))\}^2 \\ &= \left(\frac{n}{a+b+n}\right)^2 \frac{\theta_i(1-\theta_i)}{n} + \left(\frac{a+b}{a+b+n}\right)^2 \left(\theta_i - \frac{a}{a+b}\right)^2. \end{aligned}$$

Thus the Bayes risk is given by

$$\begin{aligned} \mathcal{R}(\Lambda, d_\Lambda) &= \sum_{i=1}^k \int_0^1 R_{\theta_i}(\theta_i, d_\Lambda) d\Lambda(\theta_i) \\ &= \sum_{i=1}^k \left\{ \left(\frac{n}{a+b+n}\right)^2 \frac{E\boldsymbol{\theta}_i(1-\boldsymbol{\theta}_i)}{n} + \left(\frac{a+b}{a+b+n}\right)^2 Var(\boldsymbol{\theta}_i) \right\} \\ &= k \left\{ \frac{n}{(a+b+n)^2} \frac{ab}{(a+b)^2(a+b+1)} + \left(\frac{a+b}{a+b+n}\right)^2 \frac{ab}{(a+b)(a+b+1)} \right\} \\ &= k \frac{ab}{(a+b+n)^2(a+b+1)} \left\{ \frac{n}{(a+b)} + 1 \right\} \\ &= k \frac{ab}{(a+b+n)(a+b)(a+b+1)}. \end{aligned}$$

The ordinary risks of the unbiased estimator $\underline{X}/n$ are given by

$$R_i(\theta_i, X_i/n) = \frac{\theta_i(1-\theta_i)}{n},$$

so the Bayes risk under the Beta$(a, b)$ prior on all the $\theta_i$'s is

$$\mathcal{R}(\Lambda, \underline{X}/n) = \sum_{i=1}^{k} \frac{1}{n} E\boldsymbol{\theta}_i(1-\boldsymbol{\theta}_i) = \frac{k}{n} \frac{ab}{(a+b)(a+b+1)}.$$

Note that

$$\frac{\mathcal{R}(\Lambda, d_\Lambda)}{\mathcal{R}(\Lambda, \underline{X}/n)} = \frac{k\frac{ab}{(a+b+n)(a+b)(a+b+1)}}{\frac{k}{n}\frac{ab}{(a+b)(a+b+1)}} = \frac{n}{a+b+n}.$$

The Bayes risk of the empirical Bayes estimator $\hat{\underline{d}}$ is more difficult, as far as I know. As noted in (6.37) of Lehmann and Casella, page 272, for any estimator $d$ and squared error loss,

$$\begin{aligned}
\mathcal{R}(\Lambda, d) &= E\|\underline{\boldsymbol{\theta}} - \underline{d}(X)\|^2 \\
&= E\|\underline{\boldsymbol{\theta}} - \underline{d}_\Lambda(X) + \underline{d}_\Lambda(X) - \underline{d}(X)\|^2 \\
&= E\|\underline{\boldsymbol{\theta}} - \underline{d}_\Lambda(X)\|^2 + E\|\underline{d}_\Lambda(X) - \underline{d}(X)\|^2 \\
&\qquad \text{by using } \ d_\Lambda(X) = E(\boldsymbol{\theta}|X) \\
&\qquad \text{and computing conditionally on } \ \boldsymbol{\theta} \\
&= \mathcal{R}(\Lambda, d_\Lambda) + E\|\underline{d}_\Lambda(X) - \underline{d}(X)\|^2.
\end{aligned}$$

Applying this to $\underline{d} = \hat{\underline{d}}$, the empirical Bayes estimator, yields

$$\begin{aligned}
\mathcal{R}(\Lambda, \hat{d}) &= \mathcal{R}(\Lambda, d_\Lambda) + E\|d_\Lambda(X) - \hat{d}_\Lambda(X)\|^2 \\
&= k\frac{ab}{(a+b+n)(a+b)(a+b+1)} + \sum_{i=1}^{k} E\left|\frac{a+X_i}{a+b+n} - \frac{\hat{a}+X_i}{\hat{a}+\hat{b}+n}\right|^2
\end{aligned}$$

where $\hat{a} = \hat{a}(\underline{X})$ and $\hat{b} = \hat{b}(\underline{X})$ are the MLE's of $a$ and $b$ based on the marginal distribution of $\underline{X} = (X_1, \ldots, X_n)$ and the expectation is with respect to this same marginal distribution with the "true" values of the parameters $a$ and $b$. I do not know how to calculate this exactly, so I resorted to Monte-carlo estimation. Here is a table of (exact) values of $\mathcal{R}(\Lambda, d_\Lambda)$, $\mathcal{R}(\Lambda, \underline{X}/n)$ together with estimated values of $\mathcal{R}(\Lambda, \hat{d})$ based on 500 monte-carlo replications with the same parameter values for $n$, $k$, $a$, and $b$ as in the table on page 265 of Lehmann and Casella. For comparison, I have included the column $\delta^{\hat{\pi}}$ (corresponding to my $\mathcal{R}(\Lambda, \hat{d}_\Lambda)$) in column 4.

| $(a,b)$ | $\mathcal{R}(\Lambda, d_\Lambda)$ | $\mathcal{R}(\Lambda, X/n)$ | $\mathcal{R}(\Lambda, \hat{d}_\Lambda)$ | $\widehat{\mathcal{R}}(\Lambda, \hat{d}_\Lambda)$ | $\hat{E}\|d_\lambda - \hat{d}\|^2$ | Est. SD $\times 10^3$ |
|---|---|---|---|---|---|---|
| $(2,2)$ | 0.0833 | 0.1000 | .0850 | 0.0901 | 0.0068 | 0.467 |
| $(6,6)$ | 0.0721 | 0.1154 | .0726 | 0.0867 | 0.0145 | 0.616 |
| $(20,20)$ | 0.0407 | 0.1220 | .0407 | 0.0564 | 0.0157 | 0.615 |
| $(3,1)$ | 0.0625 | 0.0750 | .0641 | 0.0677 | 0.0052 | 0.380 |
| $(9,3)$ | 0.0541 | 0.0865 | .0565 | 0.0657 | 0.0116 | 0.491 |
| $(30,10)$ | 0.0305 | 0.0915 | .0326 | 0.0438 | 0.0132 | 0.553 |

My estimates of the Bayes risk of the empirical Bayes rule $\hat{d}_\Lambda$ are larger than the Bayes risks of the empirical Bayes rule as given the Table 6.1 of Lehmann and Casella. But the general pattern does hold up: my estimated Bayes risks of the empirical Bayes rule $\hat{d}_\Lambda$ are indeed between the Bayes risks $\mathcal{R}(\Lambda, d_\Lambda)$ of the Bayes rule based on knowing $a$ and $b$, and the Bayes risks of the unbiased estimator $\underline{X}/n$.

3. Suppose that $X_1, \ldots, X_n$ are i.i.d. $N(\theta, \sigma^2)$.
(a) Suppose that $\sigma = \sigma_0$ is known. Consider testing $H: \theta = \theta_0 = 0$ versus $K: \theta = \theta_1 = 1$. In the spirit of chapter 5, plot $(R(\theta_0, \phi), R(\theta_1, \phi))$ for your favorite family of tests $\phi$. Find the entire risk body and plot it.
(b) What happens to the risk body as $n$ grows or as $\sigma_0 \to 0$?
(c) What happens to the risk body as $\theta_1$ decreases toward $\theta_0 = 0$?
(d) What happens to the risk bodies $\{(R(\theta_0, \phi), R(\theta_{1,n}, \phi)) : \ n \ge 1\}$ when $\theta_1 \equiv \theta_{1,n} \equiv \theta_0 + cn^{-1/2}$?

**Solution:** (a) If $X_1, \ldots, X_n$ are i.i.d. $N(\theta, \sigma_0)$, to find optimal tests $\phi$ we can reduce (by sufficiency) to consideration of $\overline{X} \sim N(\theta, \sigma_0^2/n)$. My favorite family of tests (in fact the most powerful tests) of $H$ versus $K$ are the tests $\phi_c(\underline{X}) = 1\{\overline{X} > c\}$.For these tests

$$
\begin{aligned}
R(0, \phi_c) &= E_0 \phi_c(\underline{X}) = P_0(\overline{X} > c) \\
&= P_0(\sqrt{n}(\overline{X} - 0)/\sigma_0 > \sqrt{n}c/\sigma_0) \\
&= 1 - \Phi(\sqrt{n}c/\sigma_0)
\end{aligned}
$$

and

$$
\begin{aligned}
R(1, \phi_c) &= E_1(1 - \phi_c(\underline{X})) \\
&= P_1(\overline{X} \le c) = P_1(\sqrt{n}(\overline{X} - 1) \le \sqrt{n}(c - 1)) \\
&= \Phi(\sqrt{n}(c - 1)/\sigma_0).
\end{aligned}
$$

Since these tests are MP for testing $H$ versus $K$, there are no points with risks below the curve given by $\{(R(0, \phi_c), R(1, \phi_c)) : \ c \in \mathbb{R}\}$; this is the lower boundary of the risk body. Note that the tests $\phi_{\text{ignore}}(\underline{X}) \equiv \alpha$ have risks $R(0, \phi_{\text{ignore}}) = \alpha$, $R(1, \phi_{\text{ignore}}) = 1 - \alpha$. Thus the line $\{(\alpha, 1-\alpha) : \ \alpha \in [0,1]\}$ is in the risk body. Furthermore, note that the tests $\phi_c'(\underline{X}) \equiv 1 - \phi_c(\underline{X}) = 1\{\overline{X} \leq c\}$ are MP for testing $H : \theta = 0$ versus $K' : \theta = \theta_1 < 0$, and by the Karlin - Rubin theorem these tests minimize the power function at points $\theta = \theta_1$ in the class of all tests with fixed power function (say at $\alpha$) at $\theta = \theta_0$. Since

$$\text{Power}_{\phi_c'}(\theta) \ = \ E_\theta \phi_c' = 1 - R(\theta, \phi_c),$$

this says that the tests $\phi_c'$ maximize $R(1, \phi_c)$ over tests $\phi$ with $R(0, \phi) = \alpha$. Hence there are no points in the risk body with risks above the curve given by $\{(1 - R(0, \phi_c), 1 - R(1, \phi_c)) : \ c \in \mathbb{R}\}$.
(b) As $n$ grows or $\sigma_0 \to 0$ the risk body expands out toward the boundary of the square $[0,1]^2$; see the plots below.
(c) As $\theta_1 \to \theta_0 = 0$, the risk body contracts toward the diagonal line $(\alpha, 1-\alpha)$ – since the testing problem becomes harder. See the plots below.

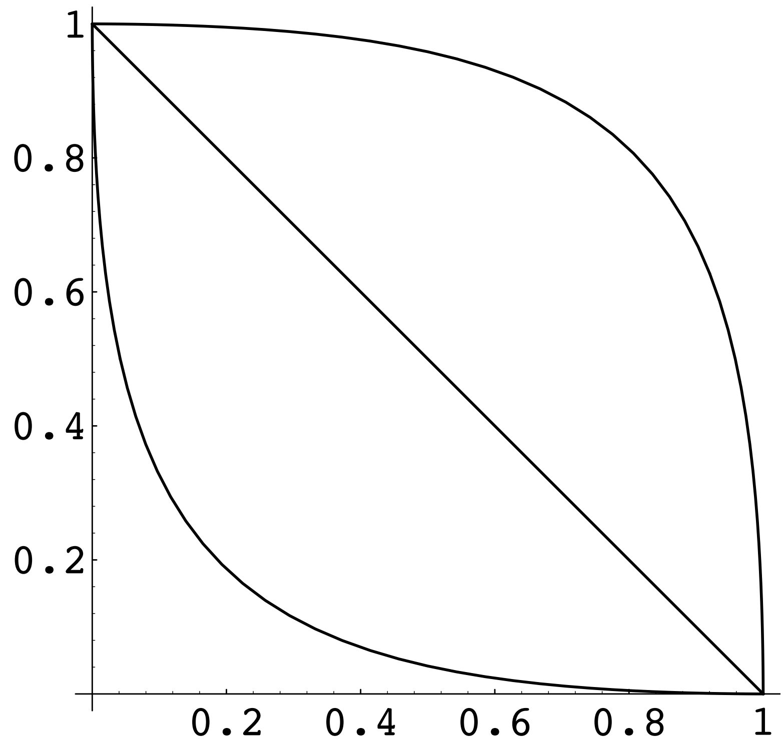


Figure 1: Risks for normal mean test, $n = 3$, $\sigma_0 = 1$, $\theta_1 = 1$

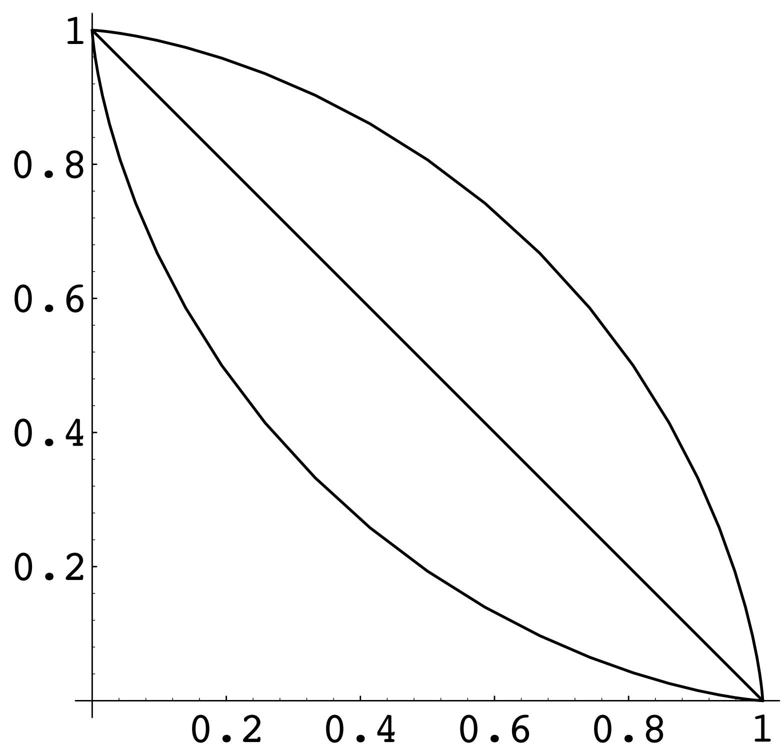


Figure 2: Risks for normal mean test, $n = 3$, $\sigma_0 = 1$, $\theta_1 = .5$

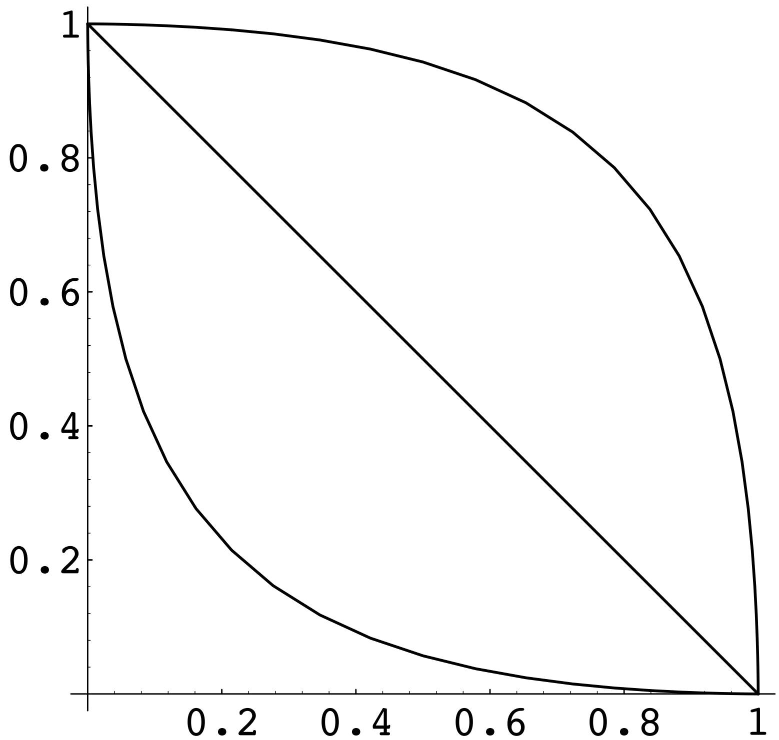


Figure 3: Risks for normal mean test, $n = 10$, $\sigma_0 = 1$, $\theta_1 = .5$

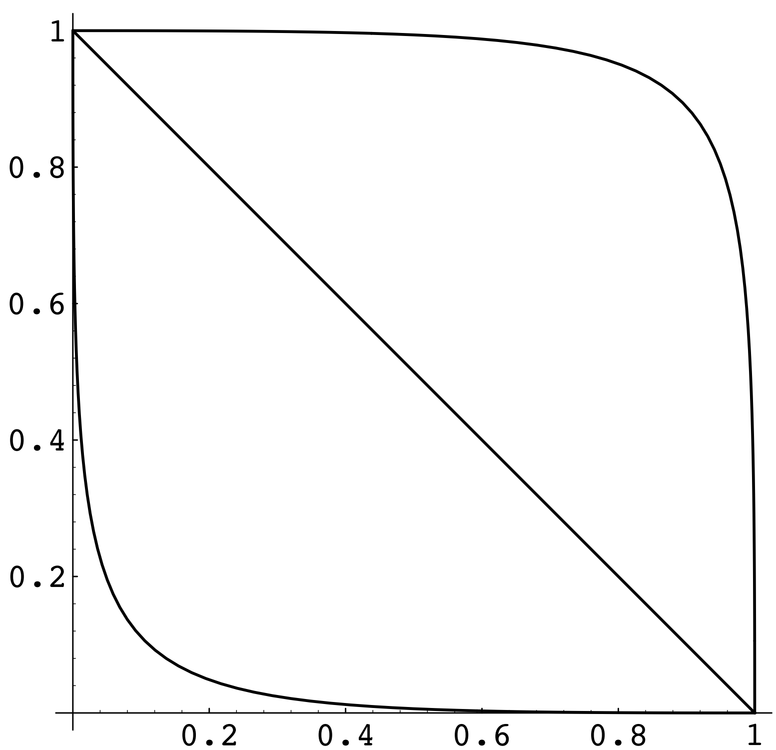


Figure 4: Risks for normal mean test, $n = 25$, $\sigma_0 = 1$, $\theta_1 = .5$

4. A random variable $X$ takes on the values $1,\ 2,\ 3,\ 4$ with probability distribution $p_0(x)$ or $p_1(x)$ as follows:

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $p_0(x)$ | .18 | .06 | .36 | .40 |
| $p_1(x)$ | .36 | .18 | .24 | .22 |

(a) Find a most powerful test of size $\alpha = .2$ for testing $p_0$ versus $p_1$ and determine its power.
(b) Find a Bayes rule $\phi$ for testing $H_0 : p = p_0$ versus $H_1 : p = p_1$ when the prior distribution is given by $(\lambda, 1-\lambda) = (1/2, 1/2)$. What is the relationship between the Bayes rule and the rule which minimizes the sum of risks $a+b$ where $a = E_0\phi$ and $b = E_1(1-\phi)$?

**Solution:** (a) Now $p_1(x)/p_0(x) = 2,\ 3,\ 2/3,\ 22/40 = 11/20$, according as $x = 1,\ 2,\ 3,\ 4$, so a MP test of size $\alpha = .2$ is given by

$$\phi(x) = \begin{cases} 1, & \text{if } x = 2 \\ .14/.18, & \text{if } x = 1 \\ 0, & \text{if } x = 3, 4. \end{cases}$$

Then

$$E_0\phi(X) = P_0(X=2) + \frac{.14}{.18}P_0(X=1) = .06 + \frac{.14}{.18}.18 = .2,$$

while

$$\text{Power} = E_1\phi(X) = P_1(X=2) + \frac{.14}{.18}P_1(X=1) = .18 + \frac{.14}{.18}.36 = .18 + .28 = .46.$$

(b) A Bayes rule (test) $\phi_\lambda$ with respect to the prior $\underline{\lambda} = (1/2, 1/2)$ is given by

$$\begin{aligned} \phi_\lambda(x) &= \begin{cases} 1, & \text{if } p_1(x) \geq p_0(x) \\ 0, & \text{if } p_1(x) < p_0(x) \end{cases} \\ &= \begin{cases} 1, & \text{if } x = 1,\ 2 \\ 0, & \text{if } x = 3, 4\ . \end{cases} \end{aligned}$$

Then $R(0, \phi_\lambda = E_0\phi_\lambda(X) = .24$, $R(1, \phi_\lambda) = E_1(1 - \phi_\lambda(X)) = .46$, and hence the Bayes risk is $\mathcal{R}(\Lambda, \phi_\lambda) = (.24 + .46)/2 = .35$ Note that the sum of the two types of error is

$$R(0, \phi_\lambda) + R(1, \phi_\lambda) = .7 = \int p_0 \wedge p_1 d\mu = 1 - d_{TV}(P_0, P_1),$$

so this result agrees with the solution of a previous problem Also note that for the Neyman Pearson test with $\alpha = .2$, the sum of the two types of error is $.2 + .54 = .74 > .7$.