Statistics 582, Midterm Exam Solutions

Wellner; 2/117/2009

1. (24 points) **Define** *any three* of the following terms. In each case, provide an appropriate context for your definition.
   (a) An *admissible* decision rule.
   (b) A *Bayes rule* with respect to a prior distribution Λ.
   (c) A *minimax decision rule*.
   (d) A *least favorable prior distribution*.
   (e) The *risk function* of a decision rule d in a decision problem with finite parameter space, action space, sample space, and loss function $L(\theta, a)$.
   (f) The *Kullback-Leibler information* $K(P, Q)$ between two probability distributions P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$.

   Solution: See Chapters 4 and 5.

2. (24 points) **State** *any two* of the following results:
   (a) A theorem relating Bayes rules to minimax rules and least favorable prior distributions.
   (b) Any theorem / result about nonparametric maximum likelihood estimation.
   (d) A uniform strong law of large numbers (or Glivenko - Cantelli theorem).
   (e) Wald's theorem on strong consistency of maximum likelihood estimators.

   Solution: See Chapters 4 and 5.

   Do **either** problem 3 **or** problem 4.

3. (32 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. Uniform$(0, \theta)$ conditional on $\boldsymbol{\theta} = \theta$, and suppose that $\boldsymbol{\theta}$ has the Pareto (θ_0, α) prior with density

$$\lambda(\theta) = \left(\frac{\alpha}{\theta_0}\right)\left(\frac{\theta_0}{\theta}\right)^{\alpha+1} 1_{(\theta_0,\infty)}(\theta),$$

   with prior mean

$$E(\boldsymbol{\theta}) = \frac{\alpha}{\alpha-1}\theta_0 \qquad \text{if} \;\; \alpha > 1.$$

   (a) Find the posterior distribution of $\boldsymbol{\theta}$.
   (b) Find the Bayes estimator of $\boldsymbol{\theta}$ for squared-error loss.
   (c) For what values of θ is the Bayes estimator consistent?
   (d) Compute the risk $R(\theta, d)$ of the usual unbiased estimator of θ, $d_{ub}(\underline{X}) = (n+1)X_{(n)}/n$.

Solution: (a) First, the joint density of the X_i's is

$$p_\theta(\underline{x}) = \prod_{i=1}^n \frac{1}{\theta} 1_{[0,\theta]}(X_i) = \frac{1}{\theta^n} 1_{[0,\theta]}(X_{(n)}),$$

so the posterior distribution of $\boldsymbol{\theta}$ is given by

$$\begin{aligned}
\lambda(\theta|\underline{X}) &= \frac{1}{\theta^n} 1_{[0,\theta]}(X_{(n)}) \cdot \left(\frac{\alpha}{\theta_0}\right)\left(\frac{\theta_0}{\theta}\right)^{\alpha+1} 1_{(\theta_0,\infty)}(\theta) \\
&\propto \frac{1}{\theta^{\alpha+n+1}} 1_{(\theta_0 \vee X_{(n)},\infty)}(\theta),
\end{aligned}$$

and thus we see that

$$\lambda(\theta|\underline{X}) = \frac{\alpha+n}{\theta_0 \vee X_{(n)}} \left(\frac{\theta_0 \vee X_{(n)}}{\theta}\right)^{\alpha+n+1} 1_{(\theta_0 \vee X_{(n)},\infty)}(\theta);$$

i.e. the posterior is Pareto$(\theta_0 \vee X_{(n)}, \alpha+n)$.
(b) Since $E(\boldsymbol{\theta}) = \frac{\alpha}{\alpha-1}\theta_0$ if $\boldsymbol{\theta} \sim$ Pareto(θ_0,α), it follows from (a) that

$$E(\boldsymbol{\theta}|\underline{X}) = \frac{\alpha+n}{\alpha+n-1}(\theta_0 \vee X_{(n)}).$$

(c) since $X_{(n)} \to_p \theta$ when the X_i's are i.i.d. P_θ and $(\alpha+n)/(\alpha+n-1) \to 1$, it follows that

$$E(\boldsymbol{\theta}|\underline{X}) \to_p 1 \cdot (\theta_0 \vee \theta) = \begin{cases} \theta & \text{if } \theta \ge \theta_0, \\ \theta_0 & \text{if } \theta < \theta_0. \end{cases}$$

Thus the Bayes estimator is consistent if $\theta \ge \theta_0$, but inconsistent if $\theta < \theta_0$.
(d) The usual unbiased estimator of θ is $(n+1)X_{(n)}/n$. For this estimator we compute

$$\begin{aligned}
R(\theta, \frac{n+1}{n}X_{(n)}) &= \left(\frac{n+1}{n}\right)^2 Var_\theta(X_{(n)}) + 0^2 \\
&= \left(\frac{n+1}{n}\right)^2 \frac{n}{(n+2)(n+1)^2}\theta^2 \\
&= \frac{1}{n(n+2)}\theta^2
\end{aligned}$$

since

$$\begin{aligned}
E_\theta X_{(n)} &= \frac{n}{n+1}\theta, \\
E_\theta X_{(n)}^2 &= \int_0^\theta x^2 \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} dx = \frac{n}{n+2}\theta^2, \\
Var_\theta(X_{(n)}) &= \left\{\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right\}\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2.
\end{aligned}$$

4. (32 points) Suppose that $(X|\boldsymbol{\theta}) \sim \text{Poisson}(\boldsymbol{\theta})$; i.e.

$$p(x|\theta) = e^{-\theta}\frac{\theta^x}{x!}, \qquad x \in \{0, 1, 2, \ldots\},$$

and the prior distribution of $\boldsymbol{\theta}$ is Gamma(α, β), i.e.

$$\lambda(\theta) = \frac{\beta^\alpha \theta^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta\theta) 1_{(0,\infty)}(\theta).$$

(a) Find the posterior distribution of $\boldsymbol{\theta}$.
(b) Find the Bayes estimator of θ for squared error loss, $L(\theta, a) = (\theta - a)^2$.
(c) Find the Bayes estimator for testing $H_0: \ \theta \in (0, 3]$ versus $H_1: \ \theta \in (3, \infty)$.
(d) Find the Bayes estimator of θ for the loss function $L(\theta, a) = (\theta - a)^2/\theta$.

Solution: (a) The joint distribution of X and θ is given by

$$\begin{aligned} p(x|\theta)\lambda(\theta) &= e^{-\theta}\frac{\theta^x}{x!}\frac{\beta^\alpha\theta^{\alpha-1}}{\Gamma(\alpha)}e^{-\beta\theta} \\ &\propto \theta^{x+\alpha-1}e^{-(\beta+1)x}, \end{aligned}$$

so the posterior density of $\boldsymbol{\theta}$ is Gamma$(x + \alpha, \beta + 1)$ with density

$$\lambda(\theta|x) = \frac{(\beta+1)^{x+\alpha}\theta^{x+\alpha-1}}{\Gamma(x+\alpha)}e^{-(\beta+1)\theta}1_{(0,\infty)}(\theta).$$

(b) The Bayes estimator with respect to squared error loss and the given prior is the posterior mean $d_B(X) = (X + \alpha)/(1 + \beta)$.
(c) The Bayes rule for testing H_0 versus H_1 is "reject H_0 if $P(\boldsymbol{\theta} \in \Theta_1|X) > P(\boldsymbol{\theta} \in \Theta_0|X) = 1 - P(\boldsymbol{\theta} \in \Theta_1|X)$", or, equivalently if $P(\boldsymbol{\theta} \in \Theta_1|X) > 1/2$. Here $P(\boldsymbol{\theta} \in \Theta_j|X) = \int_{\Theta_j} \lambda(\theta|X)d\theta$, $j = 0, 1$, For example, if $\alpha = 2$, $\beta = 2$, then

$$P(\boldsymbol{\theta} \in \Theta_1|X) = \int_3^\infty \frac{(3)^{X+2}\theta^{X+2-1}}{\Gamma(X+2)}e^{-3\theta}\ d\theta\ ,$$

and the Bayes rule rejects H_0 if $X \geq 8$, as can be seen by computing the posterior probabilities as a function of X; see the Mathematica code below:

```
f[j_,t_,a_,b_]:= b^(j+a)*t^(j+a-1) *Exp[-(b+1)*t]/Gamma[j+a]
post[j_,a_,b_]:=NIntegrate[f[j,t,a,b],{t,3,Infinity}]
TP=Table[{j,post[j,2,2]}, {j,0,16}]
Out[11]=
{0,0.00123},{1,0.006232},{2,0.021226},{3,0.05496},{4,0.115691},{5,0.20678},
{6,0.323897},{7,0.45565},{8,0.587408},{9,0.70598},{10,0.80300},{11,0.87577},
{12,0.9261},{13,0.95853},{14,0.97796},{15,0.98889},{16,0.9946}
```

(d) When the loss function is the weighted squared error loss function $L(\theta, a) = (\theta - a)^2/\theta$, the Bayes estimator of θ is, since $K(\theta) = 1/\theta$ in the context of our corollary 5.5.1,

$$d_{wB}(X) = \frac{E\{K(\boldsymbol{\theta})\boldsymbol{\theta}|X\}}{E\{K(\boldsymbol{\theta})|X\}} = \frac{1}{E\{\boldsymbol{\theta}^{-1}|X\}}.$$

But

$$\begin{aligned}
E\{\boldsymbol{\theta}^{-1}|X\} &= \int_0^\infty \theta^{-1}\lambda(\theta|x)d\theta \\
&= \int_0^\infty \theta^{-1}\frac{(\beta+1)^{x+\alpha}\theta^{x+\alpha-1}}{\Gamma(x+\alpha)}e^{-(\beta+1)\theta}d\theta \\
&= \int_0^\infty \frac{(\beta+1)^{x+\alpha}\theta^{x+\alpha-1-1}}{\Gamma(x+\alpha)}e^{-(\beta+1)\theta}d\theta \\
&= \frac{\Gamma(x+\alpha-1)}{\Gamma(x+\alpha)}(\beta+1)\int_0^\infty \frac{(\beta+1)^{x+\alpha-1}\theta^{x+\alpha-1-1}}{\Gamma(x+\alpha-1)}e^{-(\beta+1)\theta}d\theta \\
&= \frac{\beta+1}{X+\alpha-1}.
\end{aligned}$$

Hence the Bayes estimator $d_{wB}(X) = (X+\alpha-1)/(\beta+1)$.

Do **either** problem 5 **or** problem 6.

5. (30 points) Suppose that $\Theta = \{0,1\} = \mathcal{A}$ with losses $L(0,0) = L(1,1) = 0$, $L(0,1) = L(1,0) = 1$. Suppose that P_0 and P_1 have densities p_0 and p_1 with respect to a dominating measure μ, and suppose that P_0 and P_1 have prior probabilities $1-\lambda$ and λ respectively for some $\lambda \in [0,1]$.
(a) Show that the Bayes rules d_Λ (with $d_\Lambda(x)$ =probability of action 1 given that $X = x$ is observed) have the form

$$d_\Lambda(x) = \begin{cases} 1 & \text{if } \lambda p_1(x) > (1-\lambda)p_0(x), \\ \gamma(x) & \text{if } \lambda p_1(x) = (1-\lambda)p_0(x), \\ 0 & \text{if } \lambda p_1(x) < (1-\lambda)p_0(x). \end{cases}$$

(b) Give an expression for the Bayes risk for the prior λ.
(c) Show that when $\lambda = 1/2$ the Bayes risk is related to the total variation distance between P_0 and P_1.

Solution: (a) Let $\phi \equiv \phi(X)$ be a general decision rule (or test function), so that $\phi(X)$ is the probability of action 1 when X is observed. Then the risks are

$$\begin{aligned}
R(0,\phi) &= E_0\phi(X) = \int \phi(x)p_0(x)d\mu(x), \\
R(1,\phi) &= E_1(1-\phi(X)) = \int (1-\phi(x))p_1(x)d\mu(x).
\end{aligned}$$

Thus the Bayes risk of the rule ϕ for the prior distribution $(1-\lambda,\lambda)$ is

$$\begin{aligned}\mathcal{R}(\lambda,\phi) &= \lambda R(1,\phi)+(1-\lambda)R(0,\phi)\\ &= \lambda E_1(1-\phi)+(1-\lambda)E_0\phi\\ &= \lambda+\int \phi\{(1-\lambda)p_0-\lambda p_1\}d\mu,\end{aligned}$$

and it is clear that this is minimized if ϕ is of the form given:

$$\phi(X)\equiv d_\Lambda(X)=\begin{cases}1 & \text{if } \lambda p_1(x)>(1-\lambda)p_0(x),\\ \gamma(x) & \text{if } \lambda p_1(x)=(1-\lambda)p_0(x),\\ 0 & \text{if } \lambda p_1(x)<(1-\lambda)p_0(x).\end{cases}$$

(b) It follows that the Bayes risk is given by

$$\mathcal{R}(\Lambda,d_\Lambda) = \lambda+\int 1\{\lambda p_1(x)>(1-\lambda)p_0(x)\}\{(1-\lambda)p_0-\lambda p_1\}d\mu(x)$$

(c) When $\lambda=1/2$, this becomes

$$\begin{aligned}\mathcal{R}((1/2,1/2),d_{(1/2,1/2)}) &= \frac{1}{2}+\frac{1}{2}\int_{[x:\ p_1(x)>p_0(x)]}\{p_0(x)-p_1(x)\}d\mu(x)\\ &= \frac{1}{2}\left\{\int p_1d\mu-\int_{[p_1>p_0]}p_1d\mu+\int_{[p_1>p_0]}p_0d\mu\right\}\\ &= \frac{1}{2}\left\{\int_{[p_1\le p_0]}p_1d\mu+\int_{[p_1>p_0]}p_0d\mu\right\}\\ &= \frac{1}{2}\int p_0\wedge p_1d\mu=\frac{1}{2}\left\{1-d_{TV}(P_0,P_1)\right\}.\end{aligned}$$

Note that this expression makes imminently good sense: if $d_{TV}(P_0,P_1)=1$ (so that P_0 and P_1 are essentially concentrated on different sets, then $\mathcal{R}((1/2,1/2),d_{(1/2,1/2)}=0$; i.e. we can carry out a test which makes no error. On the other hand, if $d_{TV}(P_0,P_1)$ is zero (so $P_0=P_1$), then the Bayes risk is $1/2$; i.e. we can not carry out a test which does better than tossing a fair coin.

6. (30 points) Consider testing $H_0:\ X\sim U(0,1)$ versus $H_1: X\sim U(1/2,3/2)$ with zero - one loss.
(a) Find the risk set $\mathcal{R}$.
(b) Find the minimax rule.
(c) Find the least favorable prior distribution Λ.

Solution: Let $\phi(x)$ be the probability of action 1 when $X=x$ is observed. To begin, we find all the Bayes rules. Suppose that the prior on P_0, P_1 is given by

$(1-\lambda, \lambda)$ where $0 \leq \lambda \leq 1$. The Bayes rules with respect to the prior determined by λ are all of the form

$$\phi_\lambda(x) = \begin{cases} 1 & \text{if } \lambda p_1(x) > (1-\lambda) p_0(x), \\ \gamma(x) & \text{if } \lambda p_1(x) = (1-\lambda) p_0(x), \\ 0 & \text{if } \lambda p_1(x) < (1-\lambda) p_0(x). \end{cases}$$

Here $p_1(x) = 1_{[1/2,3/2]}(x)$, $p_0(x) = 1_{[0,1]}(x)$, and $\gamma(z) \in [0,1]$ for each x; one simple choice is $\gamma(x) = \gamma$ with $\gamma \in [0,1]$.
Suppose that $\lambda < 1/2$. Then it is easily seen that $\phi_\lambda(x) = 1_{[1,3/2]}(x)$, with corresponding risks

$$R(0, \phi_\lambda) = E_0 \phi_\lambda(X) = 0; \quad R(1, \phi_\lambda) = E_1(1 - \phi_\lambda(X)) = 1/2.$$

Suppose that $\lambda > 1/2$. Then it is easily seen that $\phi_\lambda(x) = 1_{[1/2,3/2]}(x)$, with corresponding risks

$$R(0, \phi_\lambda) = E_0 \phi_\lambda(X) = 1/2; \quad R(1, \phi_\lambda) = E_1(1 - \phi_\lambda) = 0.$$

Finally, when $\lambda = 1/2$, the test $\phi_\lambda(x) = 1_{[1,3/2]}(x) + \gamma(x) 1_{[1/2,1]}(x)$. Taking $\gamma(x) = \gamma$, a constant, the corresponding risks are

$$R(0, \phi_\lambda) = E_0 \phi_\lambda(X) = \gamma/2; \quad R(1, \phi_\lambda) = E_1(1 - \phi_\lambda(X)) = (1-\gamma)/2.$$

Note that if ϕ is a rule, then $1 - \phi$ is also a rule, and the risk points are given by

$$R(0, 1-\phi) = E_0(1-\phi) = 1 - R(0, \phi); \quad R(1, 1-\phi) = E_1 \phi = 1 - R(1, \phi).$$

Also note that $\phi(x) \equiv \gamma$ is a rule for $\gamma \in [0,1]$ with risk points $(\gamma, 1-\gamma)$. Plotting all these resulting risk points gives the following risk body:

As we have seen above, all the points on the line $\{(\gamma/2, (1-\gamma)/2) : \ 0 \leq \gamma \leq 1\}$ are risk points of Bayes rules with respect to the prior $(1/2, 1/2)$. Among these rules there is one with equal risks: the rule

$$\phi_{1/2}(x) = 1_{[1,3/2]}(x) + (1/2) 1_{[1/2,1]}(x),$$

and the corresponding risk point is $(1/4, 1/4)$. Note that another rule with the same risks is $\phi(x) = 1_{[3/4,3/2]}(x)$; and indeed any rule of the form $\phi(x) = 1_{(1,3/2]}(x) + 1_A(x)$ with $A \subset [1/2, 1]$ with $\mu(A) = 1/4$ is a minimax rule (where μ is Lebesgue measure). Thus the least favorable prior distribution is given by $\lambda = 1/2$.

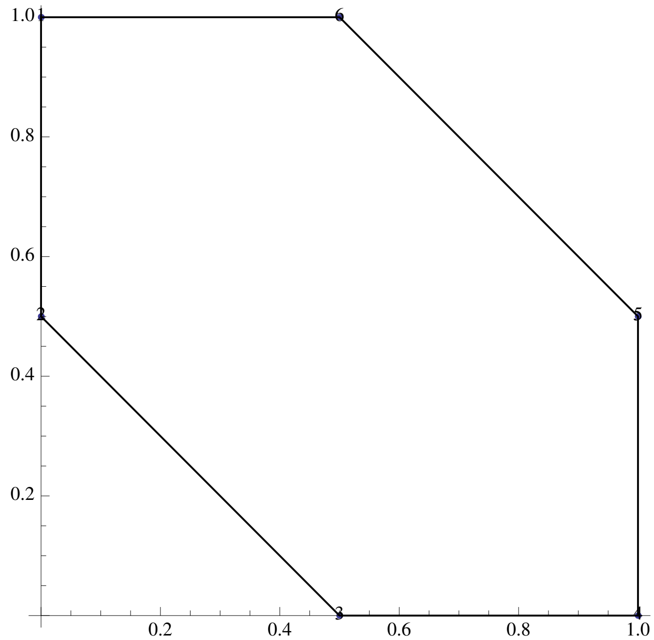

Figure 1: Risk body, problem 6

7. (30 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. with mixture density (mass function)

$$p(x; \lambda, \mu, \theta) = \theta \frac{\lambda^x}{x!} e^{-\lambda} + (1-\theta) \frac{\mu^x}{x!} e^{-\mu}, x = 0, 1, \ldots,$$

where $0 < \theta < 1$, $0 < \lambda \neq \mu < \infty$; in other words, p is the mixture of two Poisson distributions with parameters λ and μ respectively.
(a) Describe an EM - algorithm for estimation of (λ, μ, θ).
(b) What is the natural corresponding nonparametric model for the data which were modelled with the parametric mixture distribution in (a)? What is the natural nonparametric maximum likelihood estimator for the nonparametric model?

Solution: (a) Here it is natural to let the "complete data" $\underline{X}$ be $(X_1, \delta_1), \ldots, (X_n, \delta_n)$ where $\delta_i \in \{0, 1\}$ and (X_i, δ_i) are i.i.d. with density

$$p(x, \delta; \theta, \lambda, \mu) = (\theta \frac{\lambda^x}{x!} e^{-\lambda})^\delta ((1-\theta) \frac{\mu^x}{x!} e^{-\mu})^{1-\delta}$$

for $(x, \delta) \in \{0, 1, \ldots\} \times \{0, 1\}$. Then the incomplete $\underline{Y}$ is $X_1, \ldots, X_n$, which are iid with the mixture distribution

$$p(x; \lambda, \mu, \theta) = \theta \frac{\lambda^x}{x!} e^{-\lambda} + (1-\theta) \frac{\mu^x}{x!} e^{-\mu}.$$

It follows that conditional on $X = x$, δ is Bernoulli$(p(x))$ where

$$p(x) \equiv p(x; \theta, \lambda, \mu) = \frac{\theta \lambda^x e^{-\lambda}/x!}{\theta \frac{\lambda^x}{x!} e^{-\lambda} + (1-\theta) \frac{\mu^x}{x!} e^{-\mu}}. \quad (1)$$

Hence $E(\delta|X) = p(X)$; this is the basis of the E - step of an EM algorithm.
To find the M - step, note that

$$\begin{aligned} l(\theta,\lambda,\mu|X,\delta) &= \delta\{\log\theta + X\log\lambda - \lambda\} + (1-\delta)\{\log(1-\theta) + X\log\mu - \mu\} \\ &\quad +\text{constant}\,, \end{aligned}$$

so that the scores (for a sample of size one) are

$$\dot{l}_\theta(X,\delta) = \frac{\delta}{\theta} - \frac{1-\delta}{1-\theta}\,,$$

$$\dot{l}_\lambda(X,\delta) = \delta\left\{\frac{X}{\lambda} - 1\right\},$$

$$\dot{l}_\mu(X,\delta) = (1-\delta)\left\{\frac{X}{\mu} - 1\right\}.$$

Thus the score equations are solved by

$$\widehat{\lambda}_n = \frac{\sum \delta_i X_i}{\sum \delta_i}, \qquad \widehat{\mu}_n = \frac{\sum (1-\delta_i) X_i}{\sum (1-\delta_i)}\,, \qquad \widehat{\theta}_n = \frac{\sum \delta_i}{n}\,.$$

This is the basis of an M - step.
Set $\theta^{(0)} = 1/2$, $\widehat{\lambda}^{(0)} = \widehat{\mu}^{(0)} = \overline{X}$. Then, for $m = 0, 1, \ldots$, define

$$\widehat{\delta}_i^{(m)} \equiv p(X_i; \widehat{\theta}^{(m)}, \widehat{\lambda}^{(m)}, \widehat{\mu}^{(m)}) \tag{2}$$

where $p(x;\theta,\lambda,\mu)$ is given by (1), and

$$\widehat{\lambda}^{(m+1)} = \frac{\sum \widehat{\delta}_i^{(m)} X_i}{\sum \widehat{\delta}_i^{(m)}}\,, \tag{3}$$

$$\widehat{\mu}^{(m+1)} = \frac{\sum (1-\widehat{\delta}_i^{(m)}) X_i}{\sum (1-\widehat{\delta}_i^{(m)})}\,, \tag{4}$$

$$\widehat{\theta}^{(m+1)} = \frac{\sum \widehat{\delta}_i^{(m)}}{n}\,. \tag{5}$$

Iteration of (2) and (3,4,5) yields an EM algorithm for estimation of (θ,λ,μ).

(b) The natural nonparametric model for this data would be $\mathcal{P} = \{\underline{p} = (p_0, p_1, p_2, \ldots) : \sum_{x=0}^\infty p_x = 1\}$. The nonparametric maximum likelihood estimator is just $\hat{p}_n = (\hat{p}_n(0), \hat{p}_n(1), \ldots)$ where

$$\hat{p}_n(x) \equiv \mathbb{P}_n(\{x\}) = \frac{\#\{i \le n : \; X_i = x\}}{n}\,.$$