Statistics 582, Midterm Exam Solutions

Wellner; 2/11/2011

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

   Solution: See Chapter 4 and 5 notes.

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.
   (c) A uniform strong law of large numbers (or "general" Glivenko - Cantelli theorem).
   (d) Wald's theorem on strong consistency of maximum likelihood estimators.

   Solution: See Chapter 4 and 5 notes.

3. (30 points) Suppose that $(X|\theta) \sim \text{Poisson}(\theta)$,

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

   and the prior distribution of θ 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 θ.
   (b) Find the Bayes estimator of θ for squared error loss, $L(\theta, a) = (\theta - a)^2$.
   (c) Find the Bayes estimator for testing $H_0: \theta \in (0, 2]$ versus $H_1: \theta \in (2, \infty)$ with $0 - 1$ loss.
   (d) Find the Bayes estimator of θ for the loss function $L(\theta, a) = (\theta - a)^2/\theta$.
   (e) How do your answers to (a) and (b) change if $X_1, \ldots, X_n$ are i.i.d. Poisson(θ)? (Hint: use sufficiency.)

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 θ 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 $P(\theta \in \Theta_1|X) > P(\theta \in \Theta_0|X) = 1 - P(\theta \in \Theta_1|X)$", or, equivalently if $P(\theta \in \Theta_1|X) > 1/2$. Here $P(\theta \in \Theta_j|X) = \int_{\Theta_j} \lambda(\theta|X)d\theta$, $j = 0, 1$, For example, if $\alpha = 2$, $\beta = 2$, then

$$P(\theta \in \Theta_1|X) = \int_2^\infty \frac{(3)^{X+2}\theta^{X+2-1}}{\Gamma(X+2)}e^{-3\theta}\, d\theta\ ,$$

and the Bayes rule rejects H_0 if $X \geq 5$, 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,2,Infinity}]
TP=Table[{j,post[j,2,2]}, {j,0,16}]
Out[11]=
{{0, 0.0174}, {1, 0.0620}, {2, 0.1512}, {3, 0.2850}, {4, 0.4457},
 {5, 0.6063}, {6, 0.7440}, {7, 0.8472}, {8, 0.9160}, {9, 0.9574},
 {10, 0.9799}, {11, 0.9912}, {12, 0.9964}, {13, 0.9986}, {14, 0.9995},
 {15, 0.9998}, {16, 0.9999}}
```

(d) When the loss function is the weight 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,

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

But

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

Hence the Bayes estimator $d_{wB}(X) = (X+\alpha-1)/(\beta+1)$.
(e) When $X_1,\ldots,X_n$ are i.i.d. Poisson(θ), then by sufficiency it suffices to consider $S = \sum_1^n X_i \sim$ Poisson$(n\theta)$. Then the joint density is

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

so the posterior density of θ is Gamma$(s+\alpha,\beta+n)$ with density

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

The resulting Bayes estimator with respect to squared error loss is

$$
\begin{aligned}
d_B(S) &= \frac{S+\alpha}{n+\beta} = \frac{\beta}{n+\beta}\frac{\alpha}{\beta} + \frac{n}{n+\beta}\frac{S}{n} \\
&= \frac{S+\alpha}{n+\beta} = \frac{\beta}{n+\beta}\frac{\alpha}{\beta} + \frac{n}{n+\beta}\overline{X}_n.
\end{aligned}
$$

4. (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 modeled with the parametric mixture distribution in (a)? What is the natural nonparametric maximum likelihood estimator here?

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

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

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

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

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

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

5. (30 points) Suppose that (X_i, Y_i), $i = 1, \ldots, n$ are independent pairs of random variables with

$$X_i \sim \text{exponential}(\beta_i/\alpha), \qquad Y_i \sim \text{exponential}(1/\beta_i\alpha)$$

independent. Here $\alpha > 0$ and $\beta_i > 0$ for $i = 1, \ldots, n$ are all unknown. Thus the joint density of (X_i, Y_i) is

$$f_{X_i,Y_i}(x_i, y_i) = \alpha^{-2} \exp(-\beta_i x/\alpha) \exp(-y_i/\alpha\beta_i) 1_{[0,\infty)}(x_i) 1_{[0,\infty)}(y_i).$$

(a) Find the maximum likelihood estimator $\widehat{\alpha}_n$ of α.
(b) Do our theorems about consistency and asymptotic normality of maximum likelihood estimators apply to $\widehat{\alpha}_n$? Why or why not? To what (famous) model is the above model analogous?
(c) Compute $E\sqrt{X_i}$, $E\sqrt{Y_i}$, and use these together with independence of X_i and Y_i to compute $E\sqrt{X_i Y_i}$. Also compute $Var(\sqrt{X_i Y_i})$. [Hint: $\Gamma(1/2) = \sqrt{\pi}$.]
(d) use the results of (c) to show that $\hat{\alpha}_n \to_p K\alpha$ for some constant K and identify the constant K.

Solution: (a) The likelihood is

$$L_n(\alpha, \underline{\beta}) = \alpha^{-2n} \exp(-\frac{1}{\alpha}\{\sum_{i=1}^n \beta_i X_i + \sum_{i=1}^n \beta_i^{-1} Y_i\}),$$

so

$$l(\alpha, \underline{\beta}) = -2n \log \alpha - \frac{1}{\alpha}\{\sum_{i=1}^n \beta_i X_i + \sum_{i=1}^n \beta_i^{-1} Y_i\},$$

and

$$\dot{l}_\alpha(\alpha, \beta) = \frac{\partial}{\partial \alpha} l(\alpha, \underline{\beta}) = -\frac{2n}{\alpha} + \frac{1}{\alpha^2}\{\sum_{i=1}^n \beta_i X_i + \sum_{i=1}^n \beta_i^{-1} Y_i\}$$

and

$$\dot{l}_{\beta_i}(\alpha, \beta) = \frac{\partial}{\partial \beta_i} l(\alpha, \underline{\beta}) = -\frac{1}{\alpha}(X_i - \frac{1}{\beta_i^2} Y_i)\,,$$

so $\widehat{\beta}_i = \sqrt{Y_i/X_i}$, $i = 1, \ldots, n$ and

$$\widehat{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} \sqrt{X_i Y_i}$$

(b) No, since the model involves $(n+1)$ parameters – which increases with the number of observations. This model is a close relative of the famous "Neyman - Scott" example, in which the MLE of σ^2 is inconsistent.
(c) We compute, using independence of X_i and Y_i in the third expectation,

$$\begin{aligned}
E\sqrt{X_i} &= \int_0^\infty \sqrt{x} \frac{\beta_i}{\alpha} \exp(-\beta_i x/\alpha) dx \\
&= \int_0^\infty x^{1/2} \left(\frac{\beta_i}{\alpha}\right)^{3/2} \exp(-\beta_i x/\alpha) dx \sqrt{\frac{\alpha}{\beta_i}} \\
&= \int_0^\infty y^{3/2-1} e^{-y} dy \sqrt{\frac{\alpha}{\beta_i}} = \frac{\sqrt{\pi}}{2} \sqrt{\frac{\alpha}{\beta_i}} \\
E\sqrt{Y_i} &= \int \sqrt{x} \frac{1}{\alpha\beta_i} \exp(-x/(\alpha\beta_i)) dx \\
&= \int_0^\infty x^{1/2} \left(\frac{1}{\alpha\beta_i}\right)^{3/2} \exp(-x/(\alpha\beta_i)) dx \sqrt{\alpha\beta_i} \\
&= \int_0^\infty y^{3/2-1} e^{-y} dy \sqrt{\alpha\beta_i} = \frac{\sqrt{\pi}}{2} \sqrt{\alpha\beta_i}, \\
E\sqrt{X_i Y_i} &= E\sqrt{X_i} E\sqrt{Y_i} = \frac{\pi}{4}\alpha.
\end{aligned}$$

[Here we used $\Gamma(3/2) = (1/2)\Gamma(1/2) = (1/2)\sqrt{\pi}$ twice.] Furthermore, again using independence,

$$\begin{aligned}
Var(\sqrt{X_i Y_i}) &= E(X_i Y_i) - [E(\sqrt{X_i})]^2 [E(\sqrt{Y_i})]^2 \\
&= E(X_i) E(Y_i) - \frac{\pi}{4} \frac{\alpha}{\beta_i} \frac{\pi}{4} \alpha\beta_i \\
&= \frac{\alpha}{\beta_i} \alpha\beta_i - \left(\frac{\pi}{4}\right)^2 \alpha^2 = \alpha^2 \left(1 - \frac{\pi^2}{16}\right).
\end{aligned}$$

(d) From (c) we compute $E(\hat{\alpha}_n) = \pi\alpha/4$. [Note that this is $< \alpha$!] By Chebychev's inequality, for $\epsilon > 0$

$$P(|\hat{\alpha}_n - E(\hat{\alpha}_n)| > \epsilon) \leq \frac{Var(\hat{\alpha}_n)}{\epsilon^2} = \frac{\sum_{i=1}^n Var(\sqrt{X_i Y_i})}{n^2 \epsilon^2} = \frac{\alpha^2 (1 - \pi^2/16)}{n\epsilon^2} \to 0$$

as $n \to \infty$. Thus $\hat{\alpha}_n \to_p \pi\alpha/4$. Many of you tried to apply either the weak law of large numbers or the CLT here. This requires checking that the random variables $Z_i \equiv \sqrt{X_i Y_i}$ are i.i.d. This is true, but it is not immediately clear from the fact that the means and variances do not depend on i. Here is one way to see that the Z_i's are i.i.d. Note that $X_i =_d \alpha U_i/\beta_i$ and $Y_i =_d \alpha\beta_i V_i$ where U_i, V_i are all independent standard exponential(1) random variables. Thus $Z_i = \sqrt{X_i Y_i} =_d \sqrt{\alpha^2 U_i V_i} = \alpha\sqrt{U_i V_i}$. It follows that the Z_i's are i.i.d., and we can apply the WLLN, SLLN, and the CLT. But our use of Chebychev's inequality above did not require checking this. What is the distribution of $\sqrt{UV}$ where U and V are independent exponential(1) rv's? Here is a computation:

$$\begin{aligned}
1 - F_{\sqrt{UV}}(t) &\equiv P(\sqrt{UV} > t) \\
&= EP(U > t^2/V|V) \\
&= E\{\exp(-t^2/V)\} = \int_0^\infty \exp(-t^2/v - v)dv
\end{aligned}$$

with density

$$f_{\sqrt{UV}}(t) = 2t\int_0^\infty v^{-1}\exp(-t^2/v - v)dv 1_{(0,\infty)}(t).$$

Here is a plot of the density $f_{\sqrt{UV}}$:

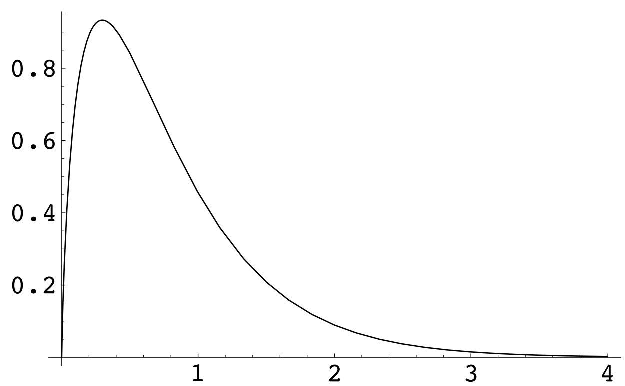

Figure 1: Density of the square root of the product of two exponentials

This is related to a problem of R. A. Fisher called the "problem of the Nile". See e.g. Cassella and Berger, *Statistical Inference, Second Edition*, exercises 6.37 and 7.54, pages 305 and 365 and T. Kariya (1989): "Equivariant estimation in a model with an ancillary statistic", *Annals of Statistics* **17**, 920 - 928.