Statistics 582, Midterm Exam Solutions

Wellner; 2/13/2015

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) 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 $(\mathcal{X}, \mathcal{A})$.

   Solutions: See course notes, chapters 4 and 5.

2. (24 points) **State and prove** *any two* of the following results:
   (a) A theorem concerning admissibility of Bayes rules (in the context of finite parameter, action, and sample spaces).
   (b) An inequality satisfied by the Kullback - Leibler information or divergence.
   (c) A theorem connecting least favorable prior distributions and minimax rules.
   (d) A result concerning inadmissibility of the rules $d_{a,b}(X) = aX + b$ for estimation of $\theta = E(X)$ with respect to squared error loss for a random variable with finite variance: $Var(X) = \sigma^2 < \infty$.

   Solutions: See course notes, chapters 4 and 5.

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

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

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

   and the prior distribution of $\boldsymbol{\theta}$ 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 rule for testing $H_0 : \ \theta \in (0, 1]$ versus $H_1 : \ \theta \in (1, \infty)$ 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)?

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

$$P(\theta \in \Theta_1|X) = \int_1^\infty \frac{(3)^{X+2}\theta^{X+2-1}}{\Gamma(X+2)}e^{-3\theta}\, d\theta\ ,$$

and the Bayes rule rejects H_0 if $X \geq 2$, 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,1,Infinity}]
TP=Table[{j,post[j,2,2]}, {j,0,16}]
Out[11]=
{{0, 0.199148}, {1, 0.42319}, {2, 0.647232}, {3, 0.815263}, {4,
  0.916082}, {5, 0.966491}, {6, 0.988095}, {7, 0.996197}, {8,
  0.998898}, {9, 0.999708}, {10, 0.999929}, {11, 0.999984}, {12,
  0.999997}, {13, 0.999999}, {14, 1.}, {15, 1.}, {16, 1.}}
```

(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) Let $\Theta = (0,1)$, $\mathcal{A} = [0,1]$, $L(\theta, a) = (\theta - a)^2$, and suppose that $X \sim \text{Binomial}(n, \theta)$:

$$p_\theta(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x \in \{0, 1, \ldots, n\}.$$

Let the prior distribution of $\boldsymbol{\theta}$ be the Beta(α, β) distribution:

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

(a) Show that the posterior distribution of $\boldsymbol{\theta}$ given X is Beta$(\alpha + X, \beta + n - X)$.
(b) If $\boldsymbol{\theta} \sim \text{Beta}(\alpha, \beta)$, then $E(\boldsymbol{\theta}) = \frac{\alpha}{\alpha+\beta}$ and

$$E(\boldsymbol{\theta}^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1}.$$

Show that the Bayes rule with respect to the prior Λ is $d_\Lambda(X) = (\alpha + X)/(\alpha + \beta + n)$.
(c) Compute the risk and indicate how to compute the Bayes risk of d_Λ.
(d) Now suppose that the loss function is changed to $L(\theta, a) = (\theta - a)^2/\{\theta(1-\theta)\}$. Find the Bayes rule d_Λ with respect to the uniform distribution Λ on $(0,1)$: that is consider the Beta(α, β) prior with $\alpha = 1 = \beta$.
(e) Find the minimax estimator of θ for the weighted loss function $L(\theta, a)$ given in (d).

Solution: (a) Now

$$\begin{aligned} p_\theta(x)\lambda(\theta) &= \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &\propto \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}, \end{aligned}$$

and hence $(\boldsymbol{\theta}|X) \sim \text{Beta}(\alpha + X, \beta + n - X)$.
(b) It follows that for squared error loss the Bayes estimator of θ is given by

$$d_\Lambda(X) = E(\boldsymbol{\theta}|X) = \frac{\alpha + X}{\alpha+\beta+n} = \frac{\alpha+\beta}{\alpha+\beta+n} \cdot \frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n} \cdot \frac{X}{n}.$$

(c) The risk of d_Λ for the squared error loss is given by

$$\begin{aligned} R(\theta, d_\Lambda) &= Var_\theta(d_\Lambda) + (\text{bias}_\theta(d_\Lambda))^2 \\ &= \left(\frac{n}{\alpha+\beta+n}\right)^2 \frac{\theta(1-\theta)}{n} + \left(\frac{\alpha+n\theta}{\alpha+\beta+n} - \theta\right)^2. \end{aligned}$$

The Bayes risk of d_Λ is given by

$$\begin{aligned} \mathcal{R}(\Lambda, d_\Lambda) &= \int_0^1 R(\theta, d_\Lambda)\lambda(\theta)d\theta \\ &= \int_0^1 R(\theta, d_\Lambda) \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta, \end{aligned}$$

and this can be computed explicitly in terms of α and β by use of the first and second moments of $\boldsymbol{\theta}$ under the prior using the formulas given above.
(d) For the weighted squared error loss function given by $L(\theta, a) = (\theta - a)^2/\{\theta(1-\theta)\} \equiv K(\theta)(\theta - a)^2$, the Bayes rule for estimating of θ is given for general α and β by

$$\begin{aligned}
d_\Lambda(X) &= \frac{E\{K(\boldsymbol{\theta})\boldsymbol{\theta}|X\}}{E\{K(\boldsymbol{\theta})|X\}} \\
&= \frac{\int_0^1 \frac{\theta}{\theta(1-\theta)}\theta^{\alpha+X-1}(1-\theta)^{\beta+n-X-1}d\theta}{\int_0^1 \frac{1}{\theta(1-\theta)}\theta^{\alpha+X-1}(1-\theta)^{\beta+n-X-1}d\theta} \\
&= \frac{\int_0^1 \theta^{\alpha+X-1}(1-\theta)^{\beta+n-X-2}d\theta}{\int_0^1 \theta^{\alpha+X-2}(1-\theta)^{\beta+n-X-2}d\theta} \\
&= \frac{\frac{\Gamma(\alpha+X)\Gamma(\beta+n-X-1)}{\Gamma(\alpha+\beta+n-1)}}{\frac{\Gamma(\alpha+X-1)\Gamma(\beta+n-X-1)}{\Gamma(\alpha+\beta+n-2)}} \\
&= \frac{\Gamma(\alpha+\beta+n-2)}{\Gamma(\alpha+\beta+n-1)} \cdot \frac{\Gamma(\alpha+X)}{\Gamma(\alpha+X-1)} \\
&= \frac{1}{\alpha+\beta+n-2} \cdot \frac{\alpha+X-1}{1}
\end{aligned}$$

by using $\Gamma(y+1) = y\Gamma(y)$. When $\alpha = 1 = \beta$, this becomes $d_\Lambda(X) = X/n$.
(e) Note that the Bayes rule we found in (d) has constant risk for weighted squared error loss: $R(\theta, d_\Lambda) = K(\theta)\theta(1-\theta)/n = 1/n$ for all $0 \le \theta \le 1$. Thus the hypothesis of Theorem 6.1 of our course notes is satisfied, and hence d_Λ is minimax for weighted squared error loss.

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

5. (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) Carefully describe a "complete data problem" with a natural connection to the mixture distribution given above via marginalization.
(b) Specify maximum likelihood estimators for the complete data problem in (a).
(c) Use your results in (b) to describe an EM-algorithm for estimation of (λ, μ, θ).
(d) Describe two natural larger models (semiparametric or nonparametric) which contain the two-point Poisson mixture model in the display above. What is the nonparametric maximum likelihood estimator for the largest natural nonparametric model in this case?

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

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

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) A completely 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

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

One possible semiparametric model might be the class $\mathcal{P}_2$ consisting of all possible mixtures of Poisson distributions:

$$p_x \equiv p_{x,G} \equiv \int_0^\infty e^{-\lambda}\frac{\lambda^x}{x!}dG(\lambda)$$

where G is an arbitrary distribution on $(0,\infty)$; see e.g. Tierney and Lambert (1984), *Ann. Statist.* **12** 1380-1387 and 13881399. Another possible semiparametric model might be to consider mixture models of the form

$$\mathcal{P}_3 = \{p_x = \theta p_x^{(0)} + (1-\theta)p_x^{(1)} : \ \theta \in [0,1], p_x^{(0)} \ Poisson(\lambda), \ p_x^{(1)} \ \text{log-concave}\};$$

see e.g. Balabdaoui et al. (2013) J. R. Stat. Soc. Ser. B. Stat. Methodol. **75**, 769-790.

6. (30 points) Suppose that $X \sim \text{Weibull}(\alpha, \beta)$; thus X has density function

$$f(x) = f_{\alpha,\beta}(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}\exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right)1_{(0,\infty)}(x)$$

where $\alpha > 0$, $\beta > 0$.
(a) Find the hazard rate function $\lambda(t) = f(t)/(1-F(t))$ where $f = f_X$ and $F(t) = F_X(t) = P(X \le t)$.
(b) Find the cumulative hazard function $\Lambda(t) = \int_{[0,t]} \lambda(s)ds$. How is Λ defined for a general distribution function F?
(c) State a general formula expressing a survival function $1-F$ to its corresponding cumulative hazard function Λ.
(d) Specialize the general formula in (c) to the particular case in (a) and (b).
(e) If we observe $(Z_1, \Delta_1), \ldots, (Z_n, \Delta_n)$ where $Z_i = X_i \wedge Y_i$, $\Delta_i = 1\{X_i \le Y_i\}$, where $X_1, \ldots, X_n$ are i.i.d. with d.f. F and $Y_1, \ldots, Y_n$ are i.i.d. G, describe the nonparametric maximum likelihood estimator $1 - \hat{F}_n$ of $1-F$.

Solution: (a) The survival function corresponding to the Weibull density is

$$\begin{aligned}1 - F(x) &= F_{\alpha,\beta}(x) = \int_x^\infty f(t)dt \\ &= \int_x^\infty \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{t}{\alpha}\right)^\beta\right) dt \\ &= \int_{(x/\alpha)^\beta}^\infty \exp(-y)dy = \exp(-(x/\alpha)^\beta)\end{aligned}$$

by the change of variables $y = (t/\alpha)^\beta$ so that $dy = (\beta/\alpha)(t/\alpha)^{\beta-1}dt$. Thus the hazard rate function is given by

$$\lambda(x) = \frac{f(x)}{1 - F(x)} = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}.$$

Note that this reduces to just $1/\alpha$ when $\beta = 1$.
(b) The cumulative hazard function is given by

$$\Lambda(x) = \int_0^x \lambda(y)dy = \int_0^x \frac{\beta}{\alpha}\left(\frac{y}{\alpha}\right)^{\beta-1} dy = \left(\frac{x}{\alpha}\right)^\beta.$$

For a general distribution function F, the cumulative hazard is defined by

$$\Lambda(x) = \int_{[0,x]} \frac{1}{1 - F(y-)} dF(y).$$

(c) The general formula expressing the survival function $1 - F$ in terms of Λ is

$$1 - F(t) = \exp\left(-\Lambda_c(t)\right) \prod_{s \le t}(1 - \Delta\Lambda(s))$$

where $\Lambda_c(t) = \Lambda(t) - \sum_{s \le t} \Delta\Lambda(s)$ where $\Delta\Lambda(s) \equiv \Lambda(s) - \Lambda(s-)$.
(d) In case of the Weibull (α, β) distribution in (a) and (b), Λ is continuous, $\Delta\Lambda(s) = 0$ for all $s \ge 0$, and hence $\Lambda_c(t) = \Lambda(t)$ for all $t \ge 0$. Thus the general formula in (c) becomes

$$1 - F(t) = \exp(-\Lambda(t)) = \exp(-(t/\alpha)^\beta),$$

in agreement with the calculations in (a) and (b).
(e) The nonparametric MLE of $1 - F$ based on right-censored data as described is given by the Kaplan-Meier estimator:

$$1 - \widehat{F}_n(t) = \prod_{s \le t}\left(1 - \Delta\widehat{\Lambda}_n(s)\right)$$

where

$$\widehat{\Lambda}_n(t) = \int_{[0,t]} \frac{1}{1 - \mathbb{H}_n(s-)} d\mathbb{H}_n^{uc}(s)$$

is the Nelson-Aalen estimator of Λ based on

$$1 - \mathbb{H}_n(t-) \equiv n^{-1} \sum_{i=1}^{n} 1_{[t,\infty)}(Z_i), \quad \text{and}$$

$$\mathbb{H}_n^{uc}(t) \equiv n^{-1} \sum_{i=1}^{n} \Delta_i 1_{[0,t]}(Z_i).$$