Statistics 582, Midterm Exam Solutions

Wellner; 2/14/2018

This exam is to be taken without any books or notes.

1. (30 points) **Define** *any three* of the following six terms. In each case, provide an appropriate context for your definition.
   (a) The *risk function* of a decision rule d in a decision problem with finite parameter space, action space, sample space, and loss function $L(\theta, a)$.
   (b) An *inadmissible* decision rule.
   (c) A *Bayes rule* with respect to a prior distribution Λ.
   (d) A *minimax decision rule*.
   (e) A *least favorable prior distribution*.
   (f) The *Kullback-Leibler divergence* $K(P, Q)$ between two probability distributions P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$.

2. (24 points) **State and prove** *any two* of the following five 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 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: $Var(X) = \sigma^2 < \infty$.
   (e) Wald's theorem on strong consistency of maximum likelihood estimators.

3. (40 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. with $E|X_1| < \infty$, and let $D_n \equiv n^{-1} \sum_{i=1}^n |X_i - \overline{X}_n|^r$ with $0 < r < 1$ fixed. Use a uniform law of large numbers (or Glivenko-Cantelli theorem) to show that $D_n \rightarrow_{a.s.} d \equiv E\{|X_1 - \mu|^r\}$ with $\mu = E(X_1)$.

   Solution: Let $\delta > 0$. Then $\overline{X}_n \in [\mu - \delta, \mu + \delta]$ for n sufficiently large with probability 1, so we can consider the collection of functions $f(x,t) \equiv f_t(x) \equiv |x - t|^r$ with $t \in [\mu - \delta, \mu + \delta]$, namely:

   $$\mathcal{F} = \{f_t(x) = |x - t|^r : \quad \mu - \delta \le t \le \mu + \delta\}.$$

   Note that: (a) $T = [\mu - \delta, \mu + \delta]$ is compact; (b) $t \mapsto f(x,t) = f_t(x)$ is continuous in t for all x; and (c) there is a function $F(x)$ such that $EF(X) < \infty$ and $|f(x,t)| \le F(x)$ for all $x \in \mathbb{R}$ and $t \in T$: namely, since $|a + b|^r \le |a|^r + |b|^r$

for $r \leq 1$ by the C_r−inequality,

$$\begin{aligned} |f(x,t)| &\leq |x-(\mu+\delta)|^r \vee |x-(\mu-\delta)|^r \\ &\leq (|x|^r + |\mu+\delta|^r) \vee (|x|^r + |\mu-\delta|^r) \\ &\leq 2|x|^r + |\mu+\delta|^r + |\mu-\delta|^r \equiv F(x) \end{aligned}$$

Note that

$$EF(X) = 2E|X|^r + C(\mu,\delta) \leq 2(E|X|)^r + C(\mu,\delta) < \infty$$

by concavity of $x \mapsto x^r$, $x \geq 0$, and $E|X| < \infty$. Thus the uniform strong law of large numbers (or Glivenko-Cantelli theorem) holds for $\mathcal{F}$:

$$\sup_{\mu-\delta \leq t \leq \mu+\delta} |\mathbb{P}_n f_t - P f_t| \rightarrow_{a.s.} 0. \tag{1}$$

But then, for $n > N_\omega$ (so that $\overline{X}_n(\omega) \in [\mu-\delta, \mu+\delta]$),

$$\begin{aligned} |D_n - d| &= |n^{-1}\sum_{i=1}^n |X_i - \overline{X}_n|^r - E|X_1-\mu|^r| \\ &= |\mathbb{P}_n f_{\overline{X}_n} - P f_{\overline{X}_n} + P f_{\overline{X}_n} - P f_\mu| \\ &\leq \sup_{\mu-\delta \leq t \leq \mu+\delta} |\mathbb{P}_n f_t - P f_t| + |P f_{\overline{X}_n} - P f_\mu| \\ &\rightarrow_{a.s.} 0 + 0 = 0 \end{aligned}$$

where the first convergence holds by the uniform strong law of large numbers (1) and the second convergence holds by $\overline{X}_n \rightarrow_{a.s.} \mu$, continuity of $t \mapsto f_t(x)$, and the dominated convergence theorem.

Do **either** problem 4 **or** problem 5.

4. (36 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

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

for some $\theta_0 > 0$ fixed, 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) Is the Bayes estimator (always) 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

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

5. (36 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 Λ of $\boldsymbol{\theta}$ be the Beta(α, β) distribution:

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

where $\alpha > 0$ and $\beta > 0$.

(a) Show that the posterior distribution of $\boldsymbol{\theta}$ given X is Beta$(\alpha + X, \beta + n - X)$.

(b) If $\boldsymbol{\theta} \sim$ Beta(α, β), 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 given by $d_\Lambda(X) = (\alpha + X)/(\alpha + \beta + n)$.

(c) Compute the ordinary risk of d_Λ 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 $(\boldsymbol{\theta}|X) \sim$ 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 6 **or** problem 7.

6. (36 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. P_{θ_0} where $\mathcal{P} = \{P_\theta : \theta \in [1, \infty)\}$ and where P_θ is the Pareto(θ, α) distribution with density

$$p_\theta(x) = \frac{\alpha}{\theta}\left(\frac{\theta}{x}\right)^{\alpha+1} 1_{[\theta,\infty)}(x);$$

here α is assumed to be known.
(a) Find the MLE $\widehat{\theta}_n$ of θ.

(b) Give a proof of consistency of $\hat{\theta}_n$ using Wald's theorem and assuming that $\Theta = [1, M]$ for some $1 < M < \infty$.
(c) Give a direct proof that $\widehat{\theta}_n \rightarrow_p \theta_0$.

Solution. (a) The likelihood for θ of $X_1, \ldots, X_n$ in the given model is

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^n \left(\frac{\alpha}{\theta}\right)\left(\frac{\theta}{X_i}\right)^{\alpha+1} 1_{[\theta,\infty)}(X_i) \\ &= \frac{\alpha^n}{\prod_1^n X_i^{\alpha+1}} \theta^{n\alpha} 1\{\theta \le X_{(1)}\} \end{aligned}$$

where $X_{(1)} = \min_{1\le i\le n} X_i$. This is maximized by $\hat{\theta}_n = X_{(1)}$.
(b) $\Theta = [1, M]$ is compact, and the function $\theta \mapsto p(x;\theta)$ is upper semi-continuous for each fixed x. Furthermore the functions

$$\begin{aligned} f(x,\theta) &= \log \frac{p_\theta(x)}{p_{\theta_0}(x)} \\ &= -\infty \cdot 1\{\theta_0 \le x < \theta\} + \alpha \log(\theta/\theta_0) \cdot 1_{[x\ge\theta]} \\ &\le F(x) \equiv \alpha \log(x/\theta_0) \end{aligned}$$

which satisfies

$$EF(X) = \int_{\theta_0}^\infty \alpha \log(x/\theta_0) \frac{\alpha}{\theta_0}\left(\frac{\theta_0}{x}\right)^{\alpha+1} dx < \infty$$

for every $\alpha > 0$. The function $\sup_{\theta:|\theta-\theta_0|\le\rho} p(x,\theta)$ is measurable for all small ρ by the same argument we used in the Uniform$(0,\theta)$ case. Since the model is identifiable, Wald's theorem yields $\hat{\theta}_n \rightarrow_{a.s.} \theta_0$.
(c) Now $P_{\theta_0}(\hat{\theta}_n < \theta_0) = P_{\theta_0}(X_{(1)} < \theta_0) = 0$ and, for every $\epsilon > 0$,

$$P_{\theta_0}(X_{(1)} > \theta_0 + \epsilon) = P(X_1 > \theta_0 + \epsilon)^n = (1 + \epsilon/\theta_0)^{-\alpha n} \to 0$$

as $n \to \infty$ since $1 - F_{\theta_0}(x) = P_{\theta_0}(X_1 > x) = (x/\theta_0)^{-\alpha}$. Thus $\hat{\theta}_n \rightarrow_p \theta_0$. Since $\sum_1^\infty P_{\theta_0}(X_{(1)} > \theta_0 + \epsilon) < \infty$ for every $\epsilon > 0$, we also have $\hat{\theta}_n \rightarrow_{a.s.} \theta_0$.

7. (36 points) Suppose that $X \sim$ Weibull(α, β); 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 estimators $\widehat{\Lambda}_n$ of Λ and $1-\widehat{F}_n$ 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

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

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