Statistics 581, Final Exam Solutions

Wellner; 12/12/2018

1. (40 points) **Define** *any four* of the following eight terms. In each case, provide an appropriate (brief) context for your definition.
   (a) *Differentiability in quadratic mean* (or Hellinger differentiability) of a family of densities $\{p_\theta : \ \theta \in \Theta \subset \mathbb{R}^d\}$.
   (b) LAN (*Local Asymptotic Normality*) of the local log-likelihood ratios.
   (c) A *locally regular estimator* of a parameter $\nu(P_\theta) = q(\theta)$ in a parametric model $\mathcal{P} = \{P_\theta : \ \theta \in \Theta\}$.
   (d) An *asymptotically linear estimator*.
   (e) A *uniformly integrable sequence of random variables*.
   (f) A *Brownian bridge process* on $[0, 1]$.
   (g) A *Brownian motion process* on $[0, \infty)$.
   (h) Convergence in distribution, $X_n \rightarrow_d X$, in a (separable) metric space (M, d).

   Solution: See Course Notes, Chapters 1-4.

2. (40 points) **State** *four* of the following eight results, providing an appropriate (brief) context for your statements:
   (a) The Lindeberg - Feller central limit theorem.
   (b) The Glivenko-Cantelli theorem for $X_1, \ldots, X_n$ i.i.d. F on $\mathbb{R}$.
   (c) Donsker's theorem for $X_1, \ldots, X_n$ i.i.d. F on $\mathbb{R}$.
   (d) Hájek's convolution theorem.
   (e) Scheffé's theorem.
   (f) A Glivenko-Cantelli theorem for the empirical measure $\mathbb{P}_n$ of $X_1, \ldots, X_n$ i.i.d. P on $(\mathcal{X}, \mathcal{A})$ indexed by a family $\mathcal{F}$ of measurable functions f from $\mathcal{X}$ to $\mathbb{R}$.
   (g) A result connecting a Brownian motion process to a Brownian bridge process.
   (h) A formula connecting $\dot{q}$, the inverse of the Fisher information matrix, and the vector of score functions in a regular parametric model.

   Solution: See Course Notes, Chapters 1-4.

Do either problem 3 or problem 4:

3. (48 points).
Suppose that $X_1, \ldots, X_n$ are independent and identically distributed real-valued random variables with distribution function F and density f.
(a) Consider the sample median $\mathbb{F}_n^{-1}(1/2)$ and the sample mean $\overline{X}_n$. Give conditions under which the sample median $\mathbb{F}_n^{-1}(1/2)$ is an asymptotically linear estimator of the population median $F^{-1}(1/2)$. Identify the influence function $\psi(x)$ and the limiting distribution of $\sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2))$.
(b) Give conditions under which the sample median $\mathbb{F}_n^{-1}(1/2)$ and the sample mean $\overline{X}_n$ have a joint limiting distribution; i.e. conditions which imply that the random vector

$$\begin{pmatrix} \sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2)) \\ \sqrt{n}(\overline{X}_n - \mu) \end{pmatrix}$$

converges in distribution where $\mu = \mu_F = E_F(X_1)$. Find the limiting distribution explicitly.
(c) A simple test for asymmetry of a distribution function is based on the difference of the mean and median: $\gamma(F) \equiv \mu_F - F^{-1}(1/2)$. Note that $\gamma_F = 0$ if F is symmetric about some point, while $\gamma(F)$ is positive for F skewed to the right, and negative for F skewed to the left. Use the results of (b) to find the limiting distribution of

$$\sqrt{n}(\gamma(\mathbb{F}_n) - \gamma(F)) = \sqrt{n}(\overline{X}_n - \mathbb{F}_n^{-1}(1/2) - (\mu_F - F^{-1}(1/2))).$$

Compute the limiting variance in terms of expectations of functions of F. Is $\gamma(\mathbb{F}_n)$ asymptotically linear?

Solution: (a) Assuming that F has a positive density f at $F^{-1}(1/2)$, we know that with $Q(t) \equiv F^{-1}(t)$, so $Q'(t) = 1/f(F^{-1}(t))$,

$$\begin{aligned}\sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2)) &\to_d -Q'(1/2)\mathbb{U}(1/2) \\ &= -\frac{1}{f(F^{-1}(1/2))}\mathbb{U}(1/2) \sim N(0, \frac{1/4}{f^2(F^{-1}(1/2))}).\end{aligned}$$

Moreover,

$$\begin{aligned}\sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{-1}{f(F^{-1}(1/2))}(1_{(-\infty, F^{-1}(1/2)]}(X_i) - 1/2) + o_p(1) \\ &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(X_i) + o_p(1)\end{aligned}$$

where $\psi(x) = -(1_{(-\infty, F^{-1}(1/2)]}(x) - 1/2)/f(F^{-1}(1/2))$ has $E\psi(X) = 0$ and $E\psi^2(X) = (1/4)/f^2(F^{-1}(1/2))$.
(b) If $E(X_1^2) < \infty$, then from the asymptotically linear representation of $\mathbb{F}_n^{-1}(1/2)$ in (a) together with the multivariate central limit theorem it follows that

$$\begin{aligned}\sqrt{n}\begin{pmatrix} \mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2) \\ \overline{X}_n - \mu_F \end{pmatrix} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} \psi(X_i) \\ X_i - \mu_F \end{pmatrix} + o_p(1) \\ &\to_d \underline{V} \sim N_2(0, \Sigma)\end{aligned}$$

where $\Sigma_{11} = E\psi^2(X_1) = (1/4)/f^2(F^{-1}(1/2))$, $\Sigma_{22} = \mathrm{Var}(X_1)$, and

$$\Sigma_{12} = \Sigma_{21} = E\psi(X_1)(X_1 - \mu) = -\frac{E\{1_{(-\infty, F^{-1}(1/2)]}(X_1)(X_1 - \mu)\}}{f(F^{-1}(1/2))}.$$

(c) From the joint convergence result in (b) it follows by the continuous mapping theorem that

$$\begin{aligned}
\sqrt{n}(\gamma(\mathbb{F}_n) - \gamma(F)) &= \sqrt{n}(\overline{X}_n - \mathbb{F}_n^{-1}(1/2) - (\mu_F - F^{-1}(1/2))) \\
&= \sqrt{n}(\overline{X}_n - \mu_F) - \sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2)) \\
&\to_d V_2 - V_1 \sim N(0, \tau^2)
\end{aligned}$$

where

$$\begin{aligned}
\tau^2 &\equiv \tau^2(F) \qquad (0.1) \\
&= \frac{1/4}{f^2(F^{-1}(1/2))} + Var_F(X) \\
&\quad + 2E_F\left\{(X - \mu_F)(1_{(-\infty, F^{-1}(1/2)]}(X) - 1/2)\right\}/f(F^{-1}(1/2)) \\
&= Var_F(X) + \frac{1/4}{f^2(F^{-1}(1/2))} \\
&\quad + 2E_F\left\{(X - \mu_F)1_{(-\infty, F^{-1}(1/2)]}(X)\right\}/f(F^{-1}(1/2)) \\
&= Var_F(X) + \frac{1/4}{f^2(F^{-1}(1/2))} \\
&\quad + 2(E_F\left\{X1_{(-\infty, F^{-1}(1/2)]}(X)\right\} - \mu_F)/f(F^{-1}(1/2)). \qquad (0.2)
\end{aligned}$$

Indeed, $\gamma(\mathbb{F}_n)$ is asymptotically linear as well: from the asymptotic linearity of $\mathbb{F}_n^{-1}(1/2)$ given in (a) and (b)

$$\sqrt{n}(\gamma(\mathbb{F}_n) - \gamma(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{X_i - \mu_F - \psi(X_i)\} + o_p(1)$$

where $\psi(x) = -(1_{(-\infty, F^{-1}(1/2)]}(x) - 1/2)/f(F^{-1}(1/2))$ as in (a)

Remarks: This statistic for testing symmetry was suggested by Edgeworth (1887). See the discussion on pages 105 and 106 of Stigler (1999), which also indicates that the joint asymptotic distribution of the mean and median was known to Laplace in the early 1800's.

4. (48 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. Uniform$(0, \theta)$. Let $X_{(1)} = \min_{1 \le i \le n} X_i$ and $X_{(n)} = \max_{1 \le i \le n} X_i$.

(a) Show that $(nX_{(1)}, n(\theta - X_{(n)})) \to_d \theta(U, V)$ where U and V are independent exponential(1) random variables. Hint: begin by computing $P(X_{(1)} > x,\ X_{(n)} \le y)$; then use this to study the limit of $P(nX_{(1)}/\theta > x,\ n(1 - X_{(n)}/\theta) \ge y)$.

(b) Show that $S_n \equiv (n+1)X_{(n)}/n$ and $T_n \equiv X_{(1)} + X_{(n)}$ are both unbiased estimators of θ.

(c) Find the joint limiting distribution of $(n(S_n - \theta), n(T_n - \theta))$.

(d) Which of the two estimators would you prefer?
[Hint: compute $\lim_n E\{[n(S_n - \theta)]^2\}$ and $\lim_n E\{[n(T_n - \theta)]^2\}$.]

(e) Does the joint density of $(U_n, V_n) \equiv \big(nX_{(1)}/\theta, n(1 - X_{(n)}/\theta\big)$ converge pointwise to a limit density? If so what does this imply about convergence in TV distance?

Solution: (a) First note that $X_1/\theta, \ldots, X_n/\theta$ are i.i.d. Uniform$(0, 1)$. Thus we compute, for $0 \le x \le y \le 1$

$$
\begin{aligned}
P(X_{(1)} > x, X_{(n)} \le y) &= P(X_{(1)}/\theta > x/\theta,\ X_{(n)}/\theta \le y/\theta) \\
&= P(x/\theta < X_i/\theta \le y/\theta \text{ for all } 1 \le i \le n) \\
&= P(x/\theta < X_1/\theta \le y/\theta)^n = (y - x)^n/\theta^n.
\end{aligned}
$$

This implies that for all $x, y > 0$ we have, for n so large that $(x + y)/n \le 1$,

$$
\begin{aligned}
P\left(nX_{(1)}/\theta > x,\ n(1 - X_{(n)}/\theta) \ge y\right) &= P\left(X_{(1)} > x\theta/n,\ X_{(n)} \le 1 - y\theta/n\right) \\
&= \left(1 - \frac{y}{n} - \frac{x}{n}\right)^n \to \exp(-(x + y)) \quad (0.3) \\
&= \exp(-x)\exp(-y).
\end{aligned}
$$

It follows that $(nX_{(1)}, n(\theta - X_{(n)})) \to_d \theta(U, V)$ where U, V are independent exponential(1) random variables.

(b) First, $E_\theta(X_{(1)}/\theta) = 1/(n + 1)$ and $E_\theta(X_{(n)}/\theta) = n/(n + 1)$. Thus it follows that

$$
E_\theta(S_n) = E_\theta\left(\frac{n+1}{n} X_{(n)}\right) = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta,
$$

and

$$
E_\theta(T_n) = E_\theta\left(X_{(n)} + X_{(1)}\right) = \frac{n}{n+1}\theta + \frac{1}{n+1}\theta = \theta.
$$

(c) We see from (a) that

$$
\begin{aligned}
n(S_n - \theta) &= \frac{n+1}{n} n(X_{(n)} - \theta) + \left(\frac{n+1}{n} n - n\right)\theta \\
&= n(X_{(n)} - \theta) + \theta + o_p(1) \\
&\to_d -\theta V + \theta = \theta(1 - V),
\end{aligned}
$$

while

$$
n(T_n - \theta) = n(X_{(n)} - \theta) + nX_{(1)} \to -\theta V + \theta U = \theta(U - V).
$$

Since $n(T_n - \theta)$ and $n(S_n - \theta)$ are both linear transformations of $nX_{(1)}$ and $n(\theta - X_{(n)})$ up to $o_p(1)$, the joint convergence follows from Slutsky's theorem and the continuous mapping theorem (or Mann-Wald theorem) with $g(u, v) = (-v + \theta, u - v)$.
(d) Now from (c) together with uniform square integrability of $\{n(S_n - \theta)\}$ and $\{n(T_n - \theta)\}$ we see that since $Var(U) = Var(V) = 1$

$$E_\theta\{n^2(S_n - \theta)^2\} \to E\{\theta^2(1 - V)^2\} = \theta^2 \cdot 1$$

while

$$E_\theta\{n^2(T_n - \theta)^2\} \to E\{\theta^2(U - V)^2\} = \theta^2 \cdot 2.$$

Thus the asymptotic mean square error of T_n is twice that of S_n; hence I would prefer S_n.
(e) It follows by differentiating (twice) the joint survival function of (U_n, V_n) given in (0.3) that the joint density of (U_n, V_n) is given by

$$f_{U_n,V_n}(x, y) = \frac{n-1}{n}\left(1 - \frac{x+y}{n}\right)^{n-2} 1_{[0,n]\times[0,n]}(x, y).$$

This yields

$$f_{U_n,V_n}(u, v) \to \exp(-(u + v)) = f_{U,V}(u, v) = f_U(u)f_V(v) \quad \text{for each} \quad (u, v) \in [0, \infty)^2,$$

the density of two independent exp(1) random variables. It follows from Scheffé's theorem that

$$d_{TV}(P_{U_n,V_n}, P_{U,V}| = \frac{1}{2}\int_{\mathbb{R}^2} |f_{U_n,V_n}(u, v) - f_{U,V}(u, v)|dudv \to 0.$$

Do either problem 5 or problem 6:

5. (48 points; based on problem 2 of HW # 10:) Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. as (X, Y) where $X \sim$ exponential(μ) and $Y \sim$ exponential(ν) are independent Thus the joint density of each $(X_i, Y_)$ is

$$p_\theta(x, y) = \mu e^{-\mu x} \nu e^{-\nu y} 1_{(0,\infty)\times(0,\infty)}(x, y)$$

where $\theta = (\mu, \nu) \in (0, \infty), (0, \infty) \equiv \Theta$. (a) Find the score functions $\dot{\mathbf{l}}_\mu(x, y)$ and $\dot{\mathbf{l}}_\nu(x, y)$ for $\theta = (\mu, \nu)$ for a sample of size $n = 1$. Find the information matrix $I(\theta)$ and the inverse information matrix $I^{-1}(\theta)$ for $\theta = (\mu, \nu)$ for a sample of size $n = 1$.
(b) Find the score equations and MLE's $\hat{\theta}_n = (\hat{\mu}_n, \hat{\nu}_n)$ based on all the data.
(c) What is the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$?
(d) What is the information bound for estimation of

$$q(\theta) = q(\mu, \nu) = 1/\mu - 1/\nu = E_\theta(X) - E_\theta(Y)?$$

Suggest an estimator that achieves the information bound asymptotically.
(e) What is the information bound for estimation of $q(\theta) = \nu/\mu$? Suggest an estimator that achieves the information bound asymptotically.
(f) Now consider testing $H : \nu/\mu = c$ versus $K : \nu/\mu \neq c$ where c is a known positive constant. [In problem 2 of HW # 10 we reparametrized this problem and considered a likelihood ratio test.] First, draw a picture showing the subset Θ_0 of Θ which defines H. Then propose a Wald type test statistic W_n for testing H versus K by noting that the function q in (e) is constant on Θ_0: $q(\theta) = \nu/\mu = c$ for $\theta \in \Theta_0$. What is the limiting distribution of your test statistic under H? What is the limiting behavior of $n^{-1}W_n$ under a fixed $\theta \in K$? What can you say about the limiting behavior of W_n under local alternatives θ_n of the form $q(\theta_n) = \nu_n/\mu_n = c + n^{-1/2}t$?

Solution: (a) The log of the joint density of each pair (X_i, Y_i) is given by

$$\log p_\theta(x, y) = \log \mu + \log \nu - \mu x - \nu y,$$

and hence the score functions $\dot{\mathbf{l}}_\mu$ and $\dot{\mathbf{l}}_\nu$ and the second derivatives, are given by

$$\begin{aligned} \dot{\mathbf{l}}_\mu(x, y) &= \frac{1}{\mu} - x, \qquad & \ddot{\mathbf{l}}_{\mu\mu} &= -\frac{1}{\mu^2}, \\ \dot{\mathbf{l}}_\nu(x, y) &= \frac{1}{\nu} - y, \qquad & \ddot{\mathbf{l}}_{\nu\nu} &= -\frac{1}{\nu^2}. \end{aligned}$$

These lead to the information matrix $I(\theta$ for $\theta = (\mu, \nu)$ being given by

$$I(\theta) = \begin{pmatrix} \mu^{-2} & 0 \\ 0 & \nu^{-2} \end{pmatrix}.$$

Thus

$$I^{-1}(\theta) = \begin{pmatrix} \mu^{2} & 0 \\ 0 & \nu^{2} \end{pmatrix}.$$

(b) The score equations are simply

$$
\begin{aligned}
0 &= \sum_{i=1}^{n} \dot{\mathrm{l}}_\mu(X_i) = \frac{n}{\mu} - \sum_{1}^{n} X_i = n\left(\frac{1}{\mu} - \overline{X}_n\right), \quad \text{and} \\
0 &= \sum_{i=1}^{n} \dot{\mathrm{l}}_\nu(Y_i) = \frac{n}{\nu} - \sum_{1}^{n} Y_i = n\left(\frac{1}{\nu} - \overline{y}_n\right),
\end{aligned}
$$

so it follows that $\hat{\mu}_n = 1/\overline{X}_n$ and $\hat{\nu}_n = 1/\overline{Y}_n$.
(c) From our general theory, the limiting distribution of the MLE's is

$$
\begin{aligned}
\sqrt{n}(\widehat{\theta}_n - \theta) &= \sqrt{n}\begin{pmatrix} \hat{\mu}_n - \mu \\ \hat{\nu}_n - \nu \end{pmatrix} \\
&\to_d \underline{D} \sim N_2(\underline{0}, I^{-1}(\theta)).
\end{aligned}
$$

(d) For estimation of $q(\theta) = \mu^{-1} - \nu^{-1} = E_\theta X - E_\theta Y$, the information bound is given by $\dot{q}(\theta)^T I^{-1}(\theta)\dot{q}(\theta)$. We calculate $\dot{q}(\theta) = (-1/\mu^2, 1/\nu^2)^T$, and hence

$$
\dot{q}(\mu,\nu)^T I^{-1}(\theta)\dot{q}(\mu,\nu) = \frac{1}{\mu^2} + \frac{1}{\nu^2}.
$$

It follows from our general theory that the plug-in estimator $q(\widehat{\theta}_n) = 1/\hat{\mu}_n - 1/\hat{\nu}_n = \overline{X}_n - \overline{Y}_n$ satisfies

$$
\sqrt{n}(q(\widehat{\theta}_n) - q(\theta))^T = \sqrt{n}(\overline{X}_n - \mu^{-1}, \overline{Y}_n - \nu^{-1})^T \to_d N(0, \mu^{-2} + \nu^{-2}).
$$

Thus it is asymptotically efficient.
(e) The information bound for $q(\theta) = \nu/\mu$ is completely similar: we first calculate $\dot{q}(\theta) = (-\nu/\mu^2, 1/\mu)^T = \mu^{-1}(-\nu/\mu, 1)^T$, and the information bound becomes

$$
I_q^{-1} \equiv \dot{q}^T I^{-1}(\theta)\dot{q} = 2\nu^2/\mu^2.
$$

Here we have

$$
\begin{aligned}
\sqrt{n}(q(\widehat{\theta}_n) - q(\theta)) &= \sqrt{n}\left(\frac{\hat{\nu}_n}{\hat{\mu}_n} - \frac{\nu}{\mu}\right) \\
&= \sqrt{n}\left(\frac{\overline{X}_n}{\overline{Y}_n} - \frac{\nu}{\mu}\right) \to_d N(0, 2\nu^2/\mu^2).
\end{aligned}
$$

(f) The null hypothesis H is given by $\Theta_0 = \{(\mu, c\mu) : \ \mu > 0\}$; this is just the line with slope c in the upper right orthant of the plane. For the function $q(\theta) = \nu/\mu$ we see that $q(\theta_0) = c$ for all $\theta_0 \in \Theta_0$. Thus a natural Wald-type statistic for testing H versus K is given by

$$
W_n \equiv \sqrt{n}(q(\widehat{\theta}_n) - q(\theta_0))\hat{I}_q\sqrt{n}(q(\widehat{\theta}_n) - q(\theta_0))
$$

where $\hat{I}_q^{-1}$ is an estimator of the information bound we found in (e). One reasonable choice for $\hat{I}_q^{-1}$ is just $1/(2\overline{X}_n^2/\overline{Y}_n^2)$ which converges in probability to $2\nu^2/\mu^2 = 2c^2$ under the null hypothesis. Thus we see that under the null hypothesis we have

$$
W_n \to_d N(0, 2\nu^2/\mu^2)^2/(2\nu^2/\mu^2) \stackrel{d}{=} Z^2 \stackrel{d}{=} \chi_1^2.
$$

Under $\theta = (\mu, \nu) \in \Theta_0^c$ fixed, we have $\nu/\mu \neq c$, and we have

$$\begin{aligned} n^{-1} W_n &= (q(\hat{\theta}_n) - q(\theta_0))^2 \hat{I}_q^{-1} \\ &\to_p (q(\theta) - q(\theta_0))^2/(2\nu^2/\mu^2) = \frac{\left(\frac{\nu}{\mu} - c\right)^2}{2\nu^2/\mu^2}. \end{aligned}$$

Under local alternatives of the form $q(\theta_n) = c + tn^{-1/2}$

$$W_n \to_d N(t, 2\nu^2/\mu^2)^2(\mu^2/(2\nu^2) \sim \chi_1^2(\delta)$$

where $\delta = t^2\mu^2/(2\nu^2)$.

Remark: It is instructive to consider the Rao type test statistic for H versus K. To implement this we first need to find the MLE of $(\mu, \nu) = (\mu, c\mu) \in \Theta_0$; i.e. for the smaller model $\mathcal{P}_0$ specified by the null hypothesis. Note that the log-density for one (X, Y) pair is

$$\log p_{\theta^0}(x, y) = \log \mu + \log(c\mu) - \mu(x + cy).$$

This yields the score function for μ in the model $\mathcal{P}_0$,

$$\dot{\mathbf{l}}_\mu(x, y; \mathcal{P}_0) = \frac{2}{\mu} - (x + cy).$$

Thus the likelihood equation for the MLE $\hat{\theta}_n^0$ in $\mathcal{P}_0$ is

$$0 = \sum_{i=1}^n \dot{\mathbf{l}}_\mu(\mu | X_i) = \frac{2n}{\mu} - (n\overline{X}_n + cn\overline{Y}_n),$$

and hence $\hat{\mu}_n^0 = 2/(\overline{X}_n + c\overline{Y}_n)$, which yields, in turn, $(\hat{\mu}_n^0, c\hat{\mu}_n^0)$. Now the Rao statistic is based on

$$\begin{aligned} \underline{Z}_n(\hat{\theta}_n^0) &= n^{-1/2} \sum_{i=1}^n \dot{\mathbf{l}}_\theta(X_i; \hat{\theta}_n^0) = n^{-1/2} \sum_{i=1}^n \begin{pmatrix} 1/\hat{\mu}_n^0 - X_i \\ 1/(c\hat{\mu}_n^0) - Y_i \end{pmatrix} \\ &= \sqrt{n} \begin{pmatrix} (\overline{X}_n + c\overline{Y}_n)/2 - \overline{X}_n \\ (\overline{X}_n + c\overline{Y}_n)/(2c) - \overline{Y}_n \end{pmatrix} \\ &= \frac{\sqrt{n}}{2} \begin{pmatrix} c\overline{Y} - \overline{X} \\ -(c\overline{Y} - \overline{X})/c \end{pmatrix}. \end{aligned}$$

Thus the Rao statistic R_n for testing H versus K is

$$R_n = \underline{Z}_n^T(\hat{\theta}_n^0) I^{-1}(\hat{\theta}_n^0) \underline{Z}_n(\hat{\theta}_n^0).$$

After a bit of algebra this becomes

$$\begin{aligned} n(q(\hat{\theta}_n - c)^2 \cdot \frac{2\overline{Y}_n^2}{(\overline{X} + c\overline{Y}_n)^2} &\to_d N(0, 2c^2)^2 \cdot \frac{2/(c^2\mu^2)}{(1/\mu + 1/\mu)^2} \\ &\stackrel{d}{=} N(0, 1)^2 \sim \chi_1^2 \end{aligned}$$

under $\theta \in \Theta_0$. Note that the Wald and Rao statistics are equal up to an estimator of the information matrix.

6. (48 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. $N(\theta, c\theta)$ where $\theta \in (0, \infty) \equiv \Theta$ and $c > 0$ is a known positive constant.
(a) What is the density $p_\theta(x)$ of each X_i?
(b) Find the score function for a sample of size $n = 1$ and compute the information for θ. Check to make sure that your score function satisfies $E_\theta \dot{\mathbf{l}}_\theta(X_1) = 0$.
(c) Find the MLE $\hat{\theta}_n$ of θ and show that it is consistent.
(d) Show that the sequence of MLE's is asymptotically normal and find the asymptotic variance.
(e) Suggest two alternative inefficient estimators of θ based on the usual $N(\mu, \sigma^2)$ model and compare their asymptotic variances to the variance of the MLE you computed in (e).

Solution: (a) the density p_θ is given by

$$p_\theta(x) = \frac{1}{\sqrt{2\pi c\theta}} \exp\left(-\frac{(x-\theta)^2}{2c\theta}\right).$$

(b) Now

$$\log p_\theta(x) = -(1/2)\log(2\pi c\theta) - \frac{(x-\theta)^2}{2c\theta} = -(1/2)\log\theta - \frac{1}{2c}\left(\frac{x^2}{\theta} - 2x + \theta\right),$$

and hence

$$\begin{aligned} \dot{\mathbf{l}}_\theta(x) &= -\frac{1}{2\theta} + \frac{1}{2c}\left(\frac{x^2}{\theta^2} - 1\right), \\ \ddot{\mathbf{l}}_{\theta\theta}(x) &= \frac{1}{2\theta^2} - \frac{x^2}{c\theta^3}, \end{aligned}$$

and it follows that

$$\begin{aligned} I(\theta) &= -E_\theta \ddot{\mathbf{l}}_{\theta\theta}(X) = \frac{E_\theta X^2}{c\theta^3} - \frac{1}{2\theta^2} \\ &= \frac{c\theta + \theta^2}{c\theta^3} - \frac{1}{2\theta^2} = \frac{2\theta + c}{2c\theta^2}. \end{aligned}$$

Note that

$$E_\theta \dot{\mathbf{l}}_\theta(X_1) = -\frac{1}{2\theta} + \frac{c\theta + \theta^2 - \theta^2}{2c\theta^2} = -\frac{1}{2\theta} + \frac{1}{2\theta} = 0.$$

(c) The score equation for θ is

$$0 = \dot{\mathbf{l}}_\theta(\theta|\underline{X}) = -\frac{n}{2}\theta + \frac{1}{2c}\left(\frac{\sum_1^n X_i^2}{\theta^2} - n\right),$$

or, equivalently,

$$\hat{\theta}_n^2 + c\hat{\theta}_n = \overline{X^2}_n, \quad \text{or} \quad (\hat{\theta}_n^2 + c/2)^2 = \overline{X^2}_n + c^2/4.$$

Thus

$$\hat{\theta}_n = \sqrt{c^2/4 + \overline{X^2}_n} - c/2 \equiv g(\overline{X^2}_n)$$

where $g(v) \equiv \sqrt{c^2/4 + v} - c/2$ is continuous. Note that $\overline{X^2}_n \rightarrow_p c\theta + \theta^2$, and hence by the continuous mapping theorem

$$g(\overline{X^2}_n) \rightarrow g(c\theta + \theta^2) = \sqrt{c^2/4 + c\theta + \theta^2} - c/2 = \theta.$$

(d) From our general theory,

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_d N(0, I^{-1}(\theta)) = N(0, 2c\theta^2/(2\theta + c)).$$

(e) One alternative estimator is $\overline{X}_n$. In this case we know that $\sqrt{n}(\overline{X}_n - \theta) \rightarrow_d N(0, c\theta)$. A second alternative estimator is $S_n^2/c = n^{-1}\sum_1^n (X_i - \overline{X}_n)^2/c \rightarrow_p c\theta/c = \theta$. In this case we have

$$\begin{aligned} \sqrt{n}(c^{-1}S_n^2 - \theta) = \sqrt{n}(S_n^2 - c\theta)/c &\rightarrow_d c^{-1}N(0, Var(X - \theta)^2) \\ &= N(0, c^{-2}Var((X - \theta)^2) = N(0, 2\theta^2). \end{aligned}$$

Note that $\theta > 2c\theta^2/(2\theta + c)$ and $2\theta^2 > 2c\theta^2/(2\theta + c)$, and the asymptotic relative efficiencies of $\overline{X}_n$ and S_n^2/c relative to the MLE are

$$\text{Rel-Eff}(\overline{X}_n, MLE)(\theta) = \frac{2c\theta^2/(2\theta + c)}{c\theta} = \frac{2\theta}{2\theta + c} \leq 1,$$

and

$$\text{Rel-Eff}(S_n^2/c, MLE)(\theta) = \frac{2c\theta^2/(2\theta + c)}{2\theta^2} = \frac{c}{2\theta + c} \leq 1.$$

Note that the relative efficiency of $\overline{X}_n$ decreases as c increases, while the relative efficiency of S_n^2/c increases as c increases; and this pattern reverses as a function of θ.

Do either problem 7 or problem 8:

7. (40points).
(a) State Jensen's inequality including conditions under which equality holds.
(b) For two probability measures P, Q on a measurable space $(\mathcal{X}, \mathcal{A})$, define the Kullback - Leibler divergence $K(P, Q)$.
(c) Show that $K(P, Q) \ge 0$ for all probability measures P and Q with equality if and only if $P = Q$.
(d) Show either $K(P, Q) \ge 2H^2(P, Q)$ or some refinement thereof.

Solution: (a) Jensen's inequality says that $Eg(X) \ge g(EX)$ if g is convex. If g is strictly convex, then equality occurs in Jensen's inequality if and only if $P(X = E(X)) = 1$.
(b) $K(P, Q) = E_P(\log(p/q)) = E_P \log(dP/dQ)$ if $P \prec Q$, $K(P, Q) = +\infty$ otherwise.
(c) Now since $-\log$ is strictly convex,

$$\begin{aligned} K(P, Q) &= -E_P \log dQ/P) = -E_P\{1_{p>0} \log(dQ/dP)\} \\ &\ge -\log E_P(dQ/dP)1\{p > 0\} = -\log Q(p > 0) \ge 0 \ \text{ since } \ Q(p > 0) \le 1. \end{aligned}$$

equality occurs if and only if $dQ/dP = 1$ with P probability 1; i.e. if and only if $Q = P$.
(d) From bonus problem 3(a) on HW # 10:

$$\begin{aligned} K(P, Q) &= -2 \int \log \left(\frac{q}{p}\right)^{1/2} d\mu = -2 \int p \log \left(1 + \sqrt{\frac{p}{q}} - 1\right) d\mu \\ &\ge -2 \int p \left(\sqrt{\frac{q}{p}} - 1\right) d\mu \quad \text{since} \quad -\log(1 + x) \ge -x \\ &= 2\left(1 - \int \sqrt{pq} d\mu\right) = 2H^2(P, Q). \end{aligned}$$

8. (40 points).
(a) Define the Hellinger and total variation distances $H(PQ)$ and $d_{TV}(P, Q)$ between two probability measures P and Q on the same measurable space $(\mathcal{X}, \mathcal{A})$.
(b) Show that $d_{TV}(P, Q) = (1/2) \int |p - q| d\mu$.
(c) Define the Hellinger affinity $\rho(P, Q)$ and the total variation affinity $\eta(P, Q)$ and relate them to the distances $H(P, Q)$ and $d_{TV}(P, Q)$ defined in (a).
(d) State two inequalities relating the Hellinger and total variation metrics defined in (a). Give a proof of at least one of the two inequalities.

Solution: (a) The Hellinger and total variation metrics $H(P, Q)$ and $d_{TV}(P, Q)$ are given by

$$\begin{aligned} H^2(P, Q) &= \frac{1}{2} \int (\sqrt{p} - \sqrt{q})^2 d\mu, \\ d_{TV}(P, Q) &= \sup_{A \in \mathcal{A}} |P(A) - Q(A)|. \end{aligned}$$

(b) Let $r \equiv p - q$ and $B = \{p \ge q\} = \{r \ge 0\}$. Then

$$0 = \int (p - q) d\mu = \int r d\mu = \int (r^+ - r^-) d\mu = \int r^+ d\mu - \int r^- d\mu,$$

so $\int r^- d\mu = \int r^+ d\mu$ and

$$\int |p-q| d\mu = \int |r| d\mu = \int (r^+ + r^-) d\mu = 2 \int r^+ d\mu.$$

Then for any measurable set A,

$$\begin{aligned} P(A) - Q(A) &= \int_A p d\mu - \int_A q d\mu = \int_A (p-q) d\mu \\ &= \int_{A\cap B} (p-q) d\mu + \int_{A\cap B^c} (p-q) d\mu \\ &\le \int r^+ d\mu \le \int_r^+ d\mu = \frac{1}{2} \int |p-q| d\mu. \end{aligned} \tag{0.4}$$

By a symmetric argument

$$Q(A) - P(A) \le \int_A r^- d\mu \le \int r^- d\mu = \frac{1}{2} \int |p-q| d\mu. \tag{0.5}$$

Combining (0.4) and (0.5) yields

$$|P(A) - Q(A)| \le \frac{1}{2} \int |p-q| d\mu.$$

One the other hand,

$$|P(B) - Q(B)| = \Big| \int_B (p-q) d\mu| = \int r^+ d\mu = \frac{1}{2} \int |p-q| d\mu.$$

(c) The Hellinger affinity is $\rho(P,Q) = \int \sqrt{pq} d\mu$ and the total variation affinity is $\eta(P,Q) \equiv \int p \wedge q d\mu$. We have

$$H^2(P,Q) = 1 - \rho(P,Q), \qquad \text{and} \qquad d_{TV}(P,Q) = 1 - \eta(P,Q).$$

(d) Two basic inequalities are:

$$H^2(P,Q) \le d_{TV}(P,Q) \le \sqrt{2} H(P,Q).$$

Let $B \equiv 1\{p \ge q\}$ as in (b). The first inequality follows from the two connection formulas in (c) together with

$$\begin{aligned} \eta(P,Q) &= \int p \wedge q d\mu = \int_B p \wedge q d\mu + \int_{B^c} p \wedge q d\mu \\ &= \int_B q d\mu + \int_{B^c} p d\mu \le \int_B \sqrt{q}\sqrt{p} d\mu + \int_{B^c} \sqrt{p}\sqrt{q} d\mu \\ &= \int \sqrt{pq} d\mu = \rho(P,Q). \end{aligned}$$

Do either problem 9 or problem 10:

9. (40 points) Let $X_1, \ldots, X_n$ be i.i.d. $P_\theta = \text{Normal}(\theta, 1)$.
   (a) Give the Hodges superefficient estimator T_n of θ (with superefficiency at $\theta = 0$).
   (b) What is the limiting distribution of $\sqrt{n}(T_n - \theta)$ as a function of θ?
   (c) What is the limiting distribution of $\sqrt{n}(T_n - \theta_n)$ when sampling from $\theta = \theta_n$ when $\theta_n = cn^{-1/2}$?
   (d) Does the limit distribution in (c) depend on c? Is T_n a locally regular estimator of θ at $\theta = 0$?
   (e) What is the limit of $E_{\theta_n}\{[\sqrt{n}(T_n - \theta_n)]^2\}$ when $\theta_n = cn^{-1/2}$ as in (c)? For what values of c does the limiting risk of T_n exceed the (limiting) risk of $\overline{X}_n$?

   Solution: See chapter 3 notes.

10. (40 points)
    Suppose that $\underline{X}, \underline{X}_1, \ldots, \underline{X}_n$ are i.i.d. $\text{Mult}_k(1, \underline{p})$, so that $\underline{N}_n \equiv \sum_{i=1}^n \underline{X}_i \sim \text{Mult}_k(n, \underline{p})$. Thus

$$P_{\underline{p}}(\underline{X} = \underline{x}) = \prod_{j=1}^k p_j^{x_j} \quad \text{for } x_i \in \{0,1\}, \ \sum_1^k x_i = 1 \, ,$$

$$P_{\underline{p},n}(\underline{N}_n = \underline{m}) = \frac{n!}{\prod_{j=1}^k m_j!} \prod_{j=1}^k p_j^{m_j} \quad \text{for } m_i \geq 0, \text{ integers } \sum_{j=1}^k m_j = n \, .$$

   (a) Compute $K(P_{\underline{q}}, P_{\underline{p}}) \equiv K(\underline{q}, \underline{p})$ for vectors $\underline{q}$, $\underline{p}$ with $\sum p_j = \sum q_j = 1$.
   (b) Evaluate $K(\underline{\widehat{p}}, \underline{p})$ where $\underline{\widehat{p}} = n^{-1}\underline{N}_n$. Relate this to the log-likelihood $\log L_n(\underline{p}|\underline{N}_n)$.
   (c) Use the result of (b) to show, without using any calculus, that the MLE of $\underline{p}$ is $\underline{\widehat{p}} = \underline{N}/n$.

   Solution: (a) First,

$$\log \frac{p_{\underline{q}}(\underline{x})}{p_{\underline{p}}(\underline{x})} = \log \prod_{j=1}^k \frac{q_j^{x_j}}{p_j^{x_j}} = \sum_{j=1}^k x_j \log(\frac{q_j}{p_j}) \, .$$

   Thus

$$K(\underline{q}, \underline{p}) = \sum_{j=1}^k q_j \log \frac{q_j}{p_j} \, .$$

   (b) From (a) it follows that

$$K(\underline{\widehat{p}}, \underline{p}) = \sum_{j=1}^k \widehat{p}_j \log \frac{\widehat{p}_j}{p_j} = -\sum_{j=1}^k \widehat{p}_j \log \frac{p_j}{\widehat{p}_j} \, .$$

   Now

$$\log L_n(\underline{p}|\underline{N}_n) = \sum_{j=1}^k N_j \log p_j + \log\left(\frac{n!}{\prod N_j!}\right)$$

$$
\begin{aligned}
&= n\sum_{j=1}^{k}\widehat{p}_j \log p_j + \log\left(\frac{n!}{\prod N_j!}\right) \\
&= n\sum_{j=1}^{k}\widehat{p}_j \log\left(\frac{p_j}{\widehat{p}_j}\right) + n\sum_{j=1}^{k}\widehat{p}_j \log \widehat{p}_j + \log\left(\frac{n!}{\prod N_j!}\right) \\
&= -nK(\underline{\widehat{p}},\underline{p}) + \text{terms constant in } \underline{p}\,.
\end{aligned}
$$

Even more neatly, as several of you noted,

$$
\log\frac{L_n(\underline{\widehat{p}}|\underline{N}_n)}{L_n(\underline{p}|\underline{N}_n)} \;=\; n\sum_{j=1}^{k}\{\widehat{p}_j \log \widehat{p}_j - \widehat{p}_j \log p_j\} = nK(\underline{\widehat{p}},\underline{p})\,.
$$

(c) Since $K(\underline{\widehat{p}},\underline{p}) \geq 0$ with equality if and only if $\underline{p} = \underline{\widehat{p}}$, we see from the identity in (b) that $L_n(\underline{p}|\underline{N}_n)$ is maximized by $\underline{p} = \underline{\widehat{p}}$.