Statistics 581, Final Exam Solutions
Wellner; 12/14/2005

1. (40 points) **Define** the following terms. In each case, provide an appropriate (brief) context for your definition.
(a) The *Kullback - Leibler* divergence (or information) between a probability measure P and another (sub-)probability measure Q on the same measurable space $(\mathcal{X}, \mathcal{A})$.
(b) The *Hellinger distance* between two probability measures P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$.
(c) The vector of score functions for a sample of size one in a regular parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d$.
(d) The *information matrix* for a sample of size one in a regular parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d$.
(e) An *asymptotically linear estimator* with influence function ψ.
(f) The efficient influence function $\tilde{\mathbf{l}}_\nu$ for a differentiable parameter $q(\theta) = \nu(P_\theta)$ in a regular parametric model $\mathcal{P}$.

   Solution: See chapters 1-4 of the course notes.

2. (24 points) **State** the following results:
(a) The multiparameter Cramér - Rao inequality (for an unbiased estimator) $T = T(\underline{X})$ of a real-valued parameter $q(\theta)$.
(b) LAN (Local Asymptotic Normality) of the local - log likelihood ratios for a regular parametric model satisfying the Cramér hypotheses (continuous third derivatives with integrable bounds).
(c) The asymptotic behavior of the likelihood ratio statistic $2 \log \lambda_n$ for testing a simple null hypothesis $\theta = \theta_0$ versus $\theta \neq \theta_0$ under a fixed alternative P_θ with $\theta \neq \theta_0$.
(d) A limiting distribution result for $\sqrt{n}(\mathbb{F}_n^{-1}(t_1) - F^{-1}(t_1), \ldots, \mathbb{F}_n^{-1}(t_k) - F^{-1}(t_k))$ for fixed numbers $0 < t_1 < \ldots < t_k < 1$.

   Solution: See chapters 1-4 of the course notes.

3. (40 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. $P_{\theta_0} \in \mathcal{P}$ where $\mathcal{P} = \{P_\theta : \theta = (\alpha, \beta) \in (0, \infty) \times (0, \infty)\}$ is the Weibull family: thus

$$p_\theta(x) = (\beta/\alpha)(x/\alpha)^{\beta-1} \exp(-(x/\alpha)^\beta) 1_{(0,\infty)}(x)$$

   for $\alpha > 0$, $\beta > 0$. Suppose that we are interested in estimating $\nu(P_\theta) = \theta_1 = \alpha$, and that T_n is an asymptotically linear estimator of $\nu(P_\theta)$ with influence function ψ (that is, T_n could be a method of moments or quantile based estimator of

$\theta_1 = \alpha$.
A. What is the efficient influence function $\tilde{\mathbf{l}}_1$ for estimation of θ_1 (in terms of the scores and information quantities)? What is the efficient score $\mathbf{l}_1^*$ function for estimation of θ_1?
B. Draw a picture showing the relationship of the influence function ψ of T_n to $\tilde{\mathbf{l}}_1$ and the tangent space $\dot{\mathcal{P}} = [\dot{\mathbf{l}}]$ of $\mathcal{P}$ (at θ_0).
C. The picture you drew in B shows that

$$\psi - \tilde{\mathbf{l}}_1 \perp \dot{\mathcal{P}} \qquad \text{in } L_2(P_{\theta_0}). \tag{1}$$

Show that (1) is equivalent to the pair of equations

$$E_0(\psi \dot{\mathbf{l}}_1) = 1, \qquad E_0(\psi \dot{\mathbf{l}}_2) = 0. \tag{2}$$

(That is, (1) implies both of the equalities in (2), while (2) implies (1).)
D. Draw a picture relating the *score* function $\dot{\mathbf{l}}_1 = \dot{\mathbf{l}}_\alpha$ to the efficient score function $\mathbf{l}_1^*$ and the "nuisance parameter" score function $\dot{\mathbf{l}}_2 = \dot{\mathbf{l}}_\beta$.
E. Relate the information bound $I_{11\cdot 2}^{-1}$ to the efficient influence function $\tilde{\mathbf{l}}_1$ and the efficient socre function $\mathbf{l}_1^*$ in parts B and D.

Solution:
A. The efficient influence function for θ_1 is $\tilde{\mathbf{l}}_1 = I_{11\cdot 2}^{-1}\mathbf{l}_1^*$ where $\mathbf{l}_1^*$ is the efficient score function given by

$$\mathbf{l}_1^* = \dot{\mathbf{l}}_1 - I_{12}I_{22}^{-1}\dot{\mathbf{l}}_2$$

and $I_{11\cdot 2} = I_{11} - I_{12}I_{22}^{-1}I_{21}$.
B. For T_n with an influence function ψ not in the tangent space $\dot{\mathcal{P}} = [\dot{\mathbf{l}}_\theta]$ of the model (which is the usual case for an ad-hoc estimator), $\psi - \tilde{\mathbf{l}}_1$ is orthogonal to $[\dot{\mathbf{l}}_\theta] = \dot{\mathcal{P}}$; in our present case with $d = 2$ and scores $\dot{\mathbf{l}}_1$ and $\dot{\mathbf{l}}_2$, this means that

$$E\{(\psi - \tilde{\mathbf{l}}_1)\dot{\mathbf{l}}_j\} = 0 \qquad \text{for } j = 1, 2. \tag{3}$$

C. Suppose that $\psi - \tilde{\mathbf{l}}_1 \perp [\dot{\mathbf{l}}_\theta]$ so that the two equalities in (3) hold. Thus, with $j = 1$ it follows that

$$0 \;=\; E\{(\psi - \tilde{\mathbf{l}}_1)\dot{\mathbf{l}}_1\} = E\{\psi\dot{\mathbf{l}}_1 - I_{11\cdot 2}^{-1}E\{\mathbf{l}_1^*\dot{\mathbf{l}}_1\}$$

$$
\begin{aligned}
&= E\{\psi \dot{\mathbf{l}}_1\} - I_{11\cdot 2}^{-1} E\{\mathbf{l}_1^* (\dot{\mathbf{l}}_1 - I_{12} I_{22}^{-1} \dot{\mathbf{l}}_2)\} \qquad \text{since } \ \mathbf{l}_1^* \perp \dot{\mathbf{l}}_2 \\
&= E\{\psi \dot{\mathbf{l}}_1\} - I_{11\cdot 2}^{-1} E\{\mathbf{l}_1^{*2} \\
&= E\{\psi \dot{\mathbf{l}}_1\} - I_{11\cdot 2}^{-1} I_{11\cdot 2} = E\{\psi \dot{\mathbf{l}}_1\} - 1.
\end{aligned}
$$

Thus $E\{\psi \dot{\mathbf{l}}_1\} = 1$. Similarly,

$$
\begin{aligned}
0 &= E\{(\psi - \tilde{\mathbf{l}}_1)\dot{\mathbf{l}}_2\} = E\{\psi \dot{\mathbf{l}}_2 - I_{11\cdot 2}^{-1} E\{\mathbf{l}_1^* \dot{\mathbf{l}}_2\} \\
&= E\{\psi \dot{\mathbf{l}}_2\} - I_{11\cdot 2}^{-1} \cdot 0 \qquad \text{since } \ \mathbf{l}_1^* \perp \dot{\mathbf{l}}_2 \\
&= E\{\psi \dot{\mathbf{l}}_2\}.
\end{aligned}
$$

Thus the two claimed relations hold. Reversing the argument gives the orthogonality relationships (3).
D. The efficient score function $\mathbf{l}_1^*$ is the projection of $\dot{\mathbf{l}}_1$ onto the orthogonal complement of the nuisance parameter tangent space $[\dot{\mathbf{l}}_2]$: that is, $\mathbf{l}_1^* = \Pi(\dot{\mathbf{l}}_1 | [\dot{\mathbf{l}}_2]^\perp) = \dot{\mathbf{l}}_1 - I_{12} I_{22}^{-1} \dot{\mathbf{l}}_2$ where $I_{12} I_{22}^{-1} \dot{\mathbf{l}}_1 = \Pi(\dot{\mathbf{l}}_1 | [\dot{\mathbf{l}}_2])$.

E. The information bound $I_{11\cdot 2}^{-1} = E\{\tilde{\mathbf{l}}_1^2\}$ where (as in A), $\tilde{\mathbf{l}}_1 = I_{11\cdot 2}^{-1} \mathbf{l}_1^*$ and $\mathbf{l}_1^* = \dot{\mathbf{l}}_1 - I_{12} I_{22}^{-1} \dot{\mathbf{l}}_2$ is the efficient score function for $\theta_1 = \alpha$.

4. (40 points). This is a continuation of problem 3 above; all the questions below concern the Weibull model with $\theta = (\alpha, \beta)$ as in problem 3.
A. Consider testing $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$. Describe the test statistics we have available for testing H versus K and give their asymptotic distributions under the null hypothesis H.
B. Describe the limiting distributions of the test statistics in A under local alternatives of the form $\theta_0 + tn^{-1/2}$.
C. Describe the limiting behavior of the test statistics in A under a fixed alternative $\theta \neq \theta_0$.
D. Now consider testing $H : \beta = 1$ versus $K : \beta \neq 1$. Discuss the test statistics we have available for testing H versus K and give their asymptotic distributions under the null hypothesis H.
E. What is the limiting behavior of the test statistics in D under local alternatives of the form $\theta_n = \theta_0 + tn^{-1/2}$ with $\theta_0 = (\alpha, \beta_0) = (\alpha, 1) \in H$?
F. Write out the test statistics in part D as explicitly as possible for the given

Weibull model. Recall that the information matrix for the Weibull model is given by

$$I(\theta) = \begin{pmatrix} \beta^2/\alpha^2 & a/\alpha \\ a/\alpha & b^2/\beta^2 \end{pmatrix}$$

where $a = -(1-\gamma)$ and $b^2 = \pi^2/6 + (1-\gamma)^2$ where

$$\gamma = .577216... = \lim_{m\to\infty} \{\sum_{k=1}^{m} \frac{1}{k} - \log m\}$$

is Euler's constant.

Solution:

A. For testing the simple null hypothesis H versus the alternative K we can use the likelihood ratio, Wald, or Rao (score) statistics given by

$$\begin{aligned} 2\log\lambda &= 2\{l_n(\widehat{\theta}_n) - l_n(\theta_0)\} = 2\log\frac{L_n(\widehat{\theta}_n)}{L_n(\theta_0)}, \\ W_n &= \sqrt{n}(\widehat{\theta}_n - \theta_0)\hat{I}(\widehat{\theta}_n)\sqrt{n}(\widehat{\theta}_n - \theta_0), \\ R_n &= Z_n(\theta_0)^T I(\theta_0)^{-1} Z_n(\theta_0) \end{aligned}$$

where $Z_n(\theta_0) = n^{-1/2}\sum_{i=1}^n \dot{\mathbf{l}}(\theta_0|X_i)$. All three statistics converge in distribution under the null hypothesis H to $D^T I(\theta_0) D \sim \chi_2^2$.

B. Under local alternatives of the form $\theta_n = \theta_0 + tn^{-1/2}$ all three of the test statistics in A converge in distribution to $(D+t)^T I(\theta_0)(D+t) \sim \chi_2^2(\delta)$ (noncentral chi-square with 2 degrees of freedom) and noncentrality parameter $\delta = t^T I(\theta_0) t$.

C. Under a fixed alternative $\theta \neq \theta_0$ (and assuming $E_\theta|\dot{\mathbf{l}}_\theta(\theta_0|X_1)| < \infty$ for the Rao statistic) we have (under the regularity hypotheses A0 - A4 of section 4.2),

$$\begin{aligned} & n^{-1} 2\log\lambda_n \to_p 2K(P_\theta, P_{\theta_0}), \\ & n^{-1} W_n = (\widehat{\theta}_n - \theta_0)^T \hat{I}(\widehat{\theta}_n)(\widehat{\theta}_n - \theta_0) \to_p (\theta - \theta_0)^T I(\theta)(\theta - \theta_0), \\ & n^{-1} R_n \to \{E_\theta \dot{\mathbf{l}}(\theta_0|X_1)\}^T I(\theta_0)^{-1}\{E_\theta \dot{\mathbf{l}}(\theta_0|X_1)\}. \end{aligned}$$

D. For testing $H : \beta = 1$ versus $K : \beta \neq 1$ we can use the likelihood ratio, Wald, or Rao (score) statistics given by

$$\begin{aligned} 2\log\lambda &= 2\{l_n(\widehat{\theta}_n) - l_n(\widehat{\theta}_n^0)\} = 2\log\frac{L_n(\widehat{\theta}_n)}{L_n(\widehat{\theta}_n^0)}, \\ W_n &= \sqrt{n}(\widehat{\beta}_n - 1)\hat{I}_{22\cdot 1}(\widehat{\theta}_n)\sqrt{n}(\widehat{\beta}_n - 1), \\ R_n &= Z_n(\widehat{\theta}_n^0)^T I(\widehat{\theta}_n^0)^{-1} Z_n(\widehat{\theta}_n^0). \end{aligned}$$

where now $\widehat{\theta}_n^0$ is the MLE of θ under H. All three of these statistics converge in distribution to χ_1^2 under the null hypothesis.
E. Under local alternatives of the form $\theta_n = \theta_0 + tn^{-1/2}$, the three statistics in D converge in distribution to $(D_2 + t_2)I_{22\cdot 1}(\theta_0)(D_2 + t_2) \sim \chi_1^2(\delta)$ where $\delta = t_2^2 I_{22\cdot 1}(\theta_0) = t_2^2\pi^2/6$.
F. The log-likelihood for all the data is given by

$$l_n(\theta) = n\log(\beta/\alpha) + \sum_{i=1}^n \{(\beta-1)\log(X_i/\alpha) - (X_i/\alpha)^\beta\}.$$

and the scores for $\theta_1 = \alpha$ and $\theta_2 = \beta$ (for one observation at x) are

$$\begin{aligned} \dot{\mathrm{l}}_\alpha(x) &= (\beta/\alpha)\left\{\left(\frac{x}{\alpha}\right)^\beta - 1\right\}, \\ \dot{\mathrm{l}}_\beta(x) &= \frac{1}{\beta}\{1 - \log(x/\alpha)^\beta((x/\alpha)^\beta - 1)\}. \end{aligned}$$

Thus, using $I_{22\cdot 1}(\theta) = \pi^2/(6\beta^2)$ and $\widehat{\alpha}_n^0 = \overline{X}$,

$$\begin{aligned} 2\log\lambda_n &= 2\{l_n(\widehat{\theta}_n) - l_n(\widehat{\theta}_n^0)\} \\ &= 2\left\{n\log\frac{\widehat{\beta}_n}{1}\frac{\widehat{\alpha}_n^0}{\widehat{\alpha}_n} + \sum_{i=1}^n\{(\widehat{\beta}_n - 1)\log(X_i/\widehat{\alpha}_n) - (X_i/\widehat{\alpha}_n)^{\widehat{\beta}_n}\} + \sum_{i=1}^n (X_i/\widehat{\alpha}_n^0)\}\right\}, \\ W_n &= n(\widehat{\beta}_n - 1)^2\frac{\pi^2/6}{\widehat{\beta}_n^2}, \\ R_n &= Z_{n2}(\widehat{\theta}_n^0)^2\frac{6}{\pi^2} \end{aligned}$$

where

$$Z_{n2}(\widehat{\theta}_n^0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\mathrm{l}}_\beta((\widehat{\alpha}_n^0, 1)|X_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{1 - \log(X_i/\overline{X})\left((X_i/\overline{X}) - 1\right)\right\}.$$

5. (40 points) Suppose that $\mathcal{P} = \{P_\theta :\ \theta \in \Theta \subset \mathbb{R}^d\}$ is a regular parametric model with densities p_θ with respect to a dominating measure μ. Suppose that $\theta_0 \in \Theta$ is fixed.
A. Express $H^2(P_\theta, P_{\theta_0})$ in terms of the densities p_{θ_0} and p_θ.
B. If $g(\theta) = H^2(P_\theta, P_{\theta_0})$, compute

$$\dot{g}(\theta) = \nabla g(\theta) = (\frac{\partial}{\partial\theta_1}g(\theta), \ldots, \frac{\partial}{\partial\theta_d}g(\theta))^T$$

and the $d \times d$ matrix

$$\ddot{g}(\theta) = \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} g(\theta) \right),$$

and $\dot{g}(\theta_0)$ and $\ddot{g}(\theta_0)$ (express the latter in terms of scores and information if possible). You may assume that the necessary interchanges of integration and differentiation are permissible.

C. Suppose that $\tilde{\theta}_n$ is an estimator satisfying $\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$ under P_{θ_0}, and define $\tilde{h}_n^2 \equiv 4ng(\tilde{\theta}_n)$. Find the limiting distribution of $\tilde{h}_n^2$ under P_{θ_0}.

D. Carry out the computations in A, B, and C explicitly when $d = 1$ and $p_\theta(x) = \theta \exp(-\theta x) 1_{(0,\infty)}(x)$ with $\theta > 0$.

Solution: A. The square of the Hellinger distance is given by

$$H^2(P_\theta, P_{\theta_0}) = (1/2) \int \{\sqrt{p_\theta} - \sqrt{p_{\theta_0}}\}^2 d\mu(x) = 1 - \int \sqrt{p_\theta p_{\theta_0}} d\mu \equiv g(\theta).$$

B. Differentiation with respect to θ_j yields

$$\begin{aligned} \frac{\partial}{\partial \theta_j} g(\theta) &= -\int \sqrt{p_{\theta_0}} \frac{\partial}{\partial \theta_j} \sqrt{p_\theta} d\mu \\ &= -\frac{1}{2} \int \sqrt{p_{\theta_0}} \frac{1}{\sqrt{p_\theta}} \frac{\partial p_\theta}{\partial \theta_j} d\mu \\ &= -\frac{1}{2} \int \sqrt{p_\theta p_{\theta_0}} \frac{\partial}{\partial \theta_j} \log p_\theta d\mu \\ &= -\frac{1}{2} \int \sqrt{p_\theta p_{\theta_0}}\ \dot{\mathbf{l}}_j d\mu \end{aligned}$$

so

$$\dot{g}(\theta) = -\frac{1}{2} \int \sqrt{p_\theta p_{\theta_0}}\ \dot{\mathbf{l}}_\theta(\theta|x) d\mu,$$

and $\dot{g}(\theta_0) = -(1/2) E_{\theta_0} \dot{\mathbf{l}}_\theta(\theta_0|X) = 0$. Furthermore, differentiating with respect to θ_i as well yields

$$\begin{aligned} \frac{\partial^2}{\partial \theta_i \partial \theta_j} g(\theta) &= -\frac{1}{2} \int \sqrt{p_{\theta_0}} \left\{ \frac{\partial \sqrt{p_\theta}}{\partial \theta_i} \dot{\mathbf{l}}_j(\theta|x) + \sqrt{p_\theta}\, \ddot{\mathbf{l}}_{ij}(\theta|x) \right\} d\mu \\ &= -\frac{1}{4} \int \sqrt{p_\theta p_{\theta_0}}\, \dot{\mathbf{l}}_i \dot{\mathbf{l}}_j d\mu - \frac{1}{2} \sqrt{p_\theta p_{\theta_0}}\, \ddot{\mathbf{l}}_{ij}(\theta|x) d\mu, \end{aligned}$$

so that

$$\begin{aligned}
\ddot{g}(\theta_0) &= -\frac{1}{4}\int \dot{\mathbf{l}}_\theta(\theta_0|x)\dot{\mathbf{l}}_\theta^T(\theta_0|x)p_{\theta_0}(x)d\mu(x) - \frac{1}{2}\int \ddot{\mathbf{l}}_{\theta,\theta}(\theta_0|x)p_{\theta_0}(x)d\mu(x) \\
&= -\frac{1}{4}I(\theta_0) + \frac{1}{2}I(\theta_0) = \frac{1}{4}I(\theta_0).
\end{aligned}$$

From B it follows that

$$\begin{aligned}
g(\theta) &= g(\theta_0) + \dot{g}^T(\theta_0)(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)^T\ddot{g}(\theta^*)(\theta-\theta_0) \\
&= \frac{1}{2}(\theta-\theta_0)^T\ddot{g}(\theta^*)(\theta-\theta_0) \\
&= \frac{1}{8}(\theta-\theta_0)^T I(\theta^*)(\theta-\theta_0)
\end{aligned}$$

for some θ^* satisfying $|\theta^*-\theta_0| \le |\theta-\theta_0|$. Thus, with $\tilde{\theta}_n^*$ satisfying $|\tilde{\theta}_n-\theta_0| \le |\tilde{\theta}_n-\theta_0| \to_p 0$,

$$\begin{aligned}
8nH^2(P_{\theta_0}, P_{\tilde{\theta}_n}) &= 8ng(\tilde{\theta}_n) \\
&= n(\tilde{\theta}_n-\theta_0)^T I(\tilde{\theta}_n^*)(\tilde{\theta}_n-\theta_0) \\
&\to_d D^T I(\theta_0)D \sim \chi_d^2.
\end{aligned}$$

D. When $p_\theta(x) = \theta e^{-\theta x}1_{[0,\infty)}(x)$ for $\theta > 0$ with μ Lebesgue measure on $\mathbb{R}$,

$$\rho(P_\theta, P_{\theta_0}) = \int_0^\infty \sqrt{\theta\theta_0}\exp(-(\theta+\theta_0)x/2)dx = \frac{\sqrt{\theta\theta_0}}{(\theta+\theta_0)/2},$$

so

$$H^2(P_\theta, P_{\theta_0}) = 1 - \rho(P_\theta, P_{\theta_0}) = 1 - \frac{\sqrt{\theta\theta_0}}{(\theta+\theta_0)/2} \equiv g(\theta).$$

Thus we calculate

$$\dot{g}(\theta) = -\frac{(1/2)\theta^{-1/2}\theta_0^{1/2}}{(\theta+\theta_0)/2} + \frac{2\sqrt{\theta\theta_0}}{(\theta+\theta_0)^2},$$

and

$$\dot{g}(\theta_0) = -\frac{1}{2\theta_0} + \frac{2\theta_0}{(2\theta_0)^2} = 0$$

Before calculating the second derivative, we rewrite the first derivative as

$$
\begin{aligned}
\dot{g}(\theta) &= \frac{\sqrt{\theta\theta_0}}{(\theta+\theta_0)/2}\left\{\frac{1}{\theta+\theta_0}-\frac{1}{2\theta}\right\} \\
&= \frac{2\sqrt{\theta\theta_0}}{\theta+\theta_0}\left\{\frac{2\theta-(\theta+\theta_0)}{2\theta(\theta+\theta_0)}\right\} \\
&= \frac{1}{\theta_0}\frac{1}{\sqrt{\theta/\theta_0}}\frac{(\theta/\theta_0-1)}{(\theta/\theta_0+1)^2}.
\end{aligned}
$$

Thus

$$
\ddot{g}(\theta) = \frac{1}{\theta_0^2}\left\{-\frac{1}{2}(\theta/\theta_0)^{-1/2}\frac{(\theta/\theta_0-1)}{(\theta/\theta_0+1)^2}+\frac{1}{\sqrt{\theta/\theta_0}}\frac{1}{(\theta/\theta_0+1)^2}-\frac{2}{\sqrt{\theta/\theta_0}}\frac{(\theta/\theta_0-1)}{(\theta/\theta_0+1)^3}\right\},
$$

so that

$$
\ddot{g}(\theta_0) = \frac{1}{\theta_0^2}\left\{0+\frac{1}{(1+1)^2}+0\right\} = \frac{1}{4}\frac{1}{\theta_0^2} = \frac{1}{4}I(\theta_0).
$$

6. (40 points) A. Compute $H^2(P_1, P_\theta)$, $d_{TV}(P_1, P_\theta)$, $K(P_1, P_\theta)$, and $K(P_\theta, P_1)$ when $P_\theta = \text{Uniform}[0, \theta]$ with $\theta > 0$. [Hint: consider computing $H^2(P, Q)$ and $d_{TV}(P, Q)$ by way of the corresponding affinities

$$
\rho(P, Q) = \int \sqrt{pq}\, d\mu, \qquad \text{and} \qquad \eta(P, Q) = \int p \wedge q\, d\mu.]
$$

B. If P_θ^n denotes the probability measure (distribution) corresponding to $X_1, \ldots, X_n$ i.i.d. as $P_\theta = \text{Uniform}[0, \theta]$ show that for $\theta_n = 1 + t/n$ with $t \in \mathbb{R}$ fixed, it follows that $H^2(P_{\theta_n}^n, P_1^n) \to$ something and find "something".
C. In Chapter 3, section 5, we established a "basic lower bound inequality" as follows: for any estimator T_n of $\nu(P)$

$$
\begin{aligned}
&\max\{E_{n,1} l(|T_n - \nu(P_1)|),\ E_{n,2} l(|T_n - \nu(P_2)|)\} \\
&\qquad \geq l(\frac{1}{4}|\nu(P_1) - \nu(P_2)|)\{1 - H^2(P_1, P_2)\}^{2n}.
\end{aligned}
$$

Apply this basic lower bound in the context of B above with $P_1 = \text{Uniform}[0, 1]$, $P_2 = \text{Uniform}[0, \theta_n]$ with $\theta_n = 1 + t/n$, $l(x) \equiv x$, and $\nu(P_\theta) = \theta$. Conclude that for any estimators T_n of θ for $\theta_0 = 1$ we have

$$
\liminf_{n\to\infty} \max\{nE_1|T_n - 1|,\ nE_{\theta_n}|T_n - \theta_n|\} \geq LB(t) > 0
$$

and find $LB(t)$ as a function of t.

Solution: A. $H^2(P_1, P_\theta) = 1 - \rho(P_1, P_\theta)$ where

$$\rho(P_1, P_\theta) = \int \sqrt{1_{[0,1]}(x)(1/\theta)1_{[0,\theta]}(x)}dx = \frac{1}{\sqrt{\theta}} \int_0^{\theta\wedge 1} dx = \frac{\theta \wedge 1}{\sqrt{\theta}} = \begin{cases} \sqrt{\theta}, & \theta < 1 \\ 1/\sqrt{\theta}, & \theta \geq 1. \end{cases}$$

Similarly, $d_{TV}(P_1, P_\theta) = 1 - \eta(P_1, P_\theta)$ where

$$\begin{aligned} \eta(P_1, P_\theta) &= \int 1_{[0,1]}(x) \wedge (1/\theta)1_{[0,\theta]}(x)dx = \int_0^{\theta\wedge 1} \frac{1}{\theta \vee 1}dx \\ &= \begin{cases} \theta, & \theta < 1 \\ 1/\theta, & \theta \geq 1. \end{cases} \end{aligned}$$

The Kullback-Leibler divergence $K(P_1, P_\theta)$ is given by

$$K(P_1, P_\theta) = \begin{cases} \int_0^\theta \log\theta dx + \int_\theta^1 \log\infty dx = \infty, & \theta < 1 \\ \int_0^1 \log\theta dx = \log\theta, & \theta \geq 1. \end{cases}$$

and, on the other hand,

$$K(P_\theta, P_1) = \begin{cases} \int_0^\theta \frac{1}{\theta}\log(1/\theta)dx = \log(1/\theta), & \theta < 1 \\ \int_0^1 \frac{1}{\theta}\log\theta dx + \int_1^\infty \frac{1}{\theta}\log(\infty) = \infty, & \theta \geq 1. \end{cases}$$

B. When $\theta = \theta_n = 1 + n^{-1}t$, we have

$$\begin{aligned} H^2(P_{\theta_n}^n, P_1^n) &= 1 - \rho(P_{\theta_n}^n, P_1^n) = 1 - \rho(P_{\theta_n}, P_1)^n \\ &= \left\{ \begin{array}{ll} 1 - (1 + t/n)^{n/2}, & t < 0 \\ 1 - (1+t/n)^{-n/2}, & t > 0 \end{array} \right\} \to \left\{ \begin{array}{ll} 1 - e^{t/2}, & t < 0 \\ 1 - e^{-t/2}, & t > 0 \end{array} \right\} \\ &= 1 - \exp(-|t|). \end{aligned}$$

C. The "basic lower bound proposition" gives

$$\begin{aligned} \max\{E_1|T_n - 1|,\ E_{\theta_n}|T_n - \theta_n|\} &\geq \frac{1}{4}\frac{|t|}{n}\left\{1 - H^2(P_1, P_{\theta_n})\right\}^{2n} \\ &= \frac{1}{4}\frac{|t|}{n}\rho(P_1, P_{\theta_n})^{2n} = \frac{1}{4}\frac{|t|}{n} \begin{cases} (1+t/n)^n, & t < 0 \\ (1+t/n)^{-n}, & t \geq 0. \end{cases} \end{aligned}$$

Thus, multiplying across by n and taking the $\liminf$ across the inequality yields

$$\begin{aligned} &\liminf_{n\to\infty} \max\{nE_1|T_n - 1|,\ nE_{\theta_n}|T_n - \theta_n|\} \\ &\geq \frac{1}{4}|t|\left\{1 - H^2(P_1, P_{\theta_n})\right\}^{2n} \\ &= \frac{1}{4}|t| \liminf_{n\to\infty} \rho(P_1, P_{\theta_n})^{2n} = \frac{1}{4}|t| \liminf_{n\to\infty} \begin{cases} (1+t/n)^n, & t < 0 \\ (1+t/n)^{-n}, & t \geq 0. \end{cases} \\ &= \frac{1}{4}|t|\exp(-|t|) \equiv LB(t). \end{aligned}$$

Note that this is maximized by the choice $|t| = 1$, so we can conclude that for $C \geq 1$ we have

$$\liminf_{n\to\infty} \sup_{|t|\leq C} nE_{\theta_n}|T_n - \theta_n| \geq \frac{1}{4}.$$

Similarly, repeating these calculations with 1 replaced by θ_0 and $\theta_n = \theta_0 + t/n$ yields

$$\liminf_{n\to\infty} \max\{nE_{\theta_0}|T_n - \theta_0||,\ nE_{\theta_n}|T_n - \theta_n|\} \quad = \quad \frac{1}{4}|t|\exp(-|t|/\theta_0)$$

which is maximized by $|t| = \theta_0$, so we can conclude that for $C \geq \theta_0$ we have

$$\liminf_{n\to\infty} \sup_{|t|\leq C} nE_{\theta_n}|T_n - \theta_n| \geq \frac{e^{-1}}{4}\theta_0.$$

(Recall that for $T_n = \max_{1\leq i\leq n} X_i = X_{(n)}$ we have $n(\theta_0 - T_n) \to_d Y \sim \text{Exponential}(1/\theta_0)$ under P_{θ_0}, and furthermore $E_{\theta_0}|n(\theta_0 - T_n)| \to EY = \theta_0$.)