36-752, Spring 2018
Homework 5

Due Monday, April 23, by 5:00pm in Jisu's mailbox.

1. Recall the Skorohod's representation theorem given in class (see Theorem 6.7 in the book *Weak Convergence in Metric Spaces,* by P. Billingsley, Wiley Series in Probability and Statistics, 1999, second edition).

   Assume that $\{X_n\}$ *and* X *take values in a separable metric space and that* $X_n \overset{D}{\to} X$*. Then, there exist random variables* $\{Y_n\}$ *and* Y*, defined on the same probability space, such that* $X_n \overset{d}{=} Y_n$ *for all* n *and* $X \overset{d}{=} Y$ *and* $Y_n \overset{a.s}{\to} Y$*.*

   (a) Use Skorohod's representation theorem to show that $X_n \overset{D}{\to} X$ if and only if $\lim_n \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]$ for all bounded functions g that are continuous almost everywhere with respect to the distribution of X.

   (b) Use the previous result to give a simple proof of the continuous mapping theorem.

2. Show by example that distribution functions having densities can converge weakly even if the densities do not converge. *Hint: Consider* $f_n(x) = 1 + \cos 2\pi n x$ *on* $[0, 1]$.

3. Let $X_n = (X_n(1), \ldots, X_n(n))$ be a random vector uniformly distributed over $S_{\sqrt{n}} = \{x \in \mathbb{R}^n \colon \|x\| = \sqrt{n}\}$, the n-dimensional sphere of radius $\sqrt{n}$. Show that $X_n(1) \overset{D}{\to} X$, where $X \sim N(0, 1)$. You may use the fact that if the random vector $Z_n = (Z_n(1), \ldots, Z_n(n))$ is comprised of independent standard normals, then the vector $Z_n \frac{\sqrt{n}}{\|Z_n\|}$ is uniformly distributed over $S_{\sqrt{n}}$ (that is, $X_n \overset{d}{=} Z_n \frac{\sqrt{n}}{\|Z_n\|}$).

4. Suppose that the distributions of random variables X_n and X (in $(\mathbb{R}^d, \mathcal{B}^d)$) have densities f_n and f. Show that if $f_n(x) \to f(x)$ for x outside a set of Lebesgue measure 0, then $X_n \overset{D}{\to} X$. *Hint: Use Scheffe's theorem.*

   More, generally, show that convergence in total variation implies convergence in distribution. That is, show that, if $\{\mu_n\}$ and μ are probability measures on $(\mathcal{X}, \mathcal{B})$ (here $\mathcal{X}$ is a metric space and $\mathcal{B}$ the corresponding Borel σ-field), and if

   $$d_{\mathrm{TV}}(\mu_n, \mu) = \sup_{A \in \mathbb{B}} |\mu_n(A) - \mu(A)| \to 0,$$

   then $\mu_n \overset{D}{\to} \mu$.

5. Show that

   $$\rho(F, G) = \inf \{\epsilon > 0 \colon F(x - \epsilon) - \epsilon \le G(x) \le F(x + \epsilon) + \epsilon \text{ for all } x\}$$

   defines a metric on the space of c.d.f.'s and that $\rho(F_n, F) \to 0$ if and only if $X_n \overset{D}{\to} X$, where X_n has c.d.f. F_n for all n and X has c.d.f. F.

6. Assume that $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ is a parametric model over the sample space $(\mathcal{X}, \mathcal{B})$, such that $P_\theta << \mu$ for all $\theta \in \Theta$, for some σ-finite dominating measure μ. Assume also that all the P_θ's have the same support and $\theta \neq \theta'$ implies that $P_\theta \neq P_{\theta'}$. (You may also assume that $K(P_\theta, P_{\theta'}) < \infty$ for all $\theta \neq \theta'$, though this is not necessary.) Let $\mathbb{X}_n = (X_1, \ldots, X_n) \overset{i.i.d.}{\sim} P_{\theta_0}$ for some $\theta_0 \in \Theta$ and write

$$L_n(\theta; \mathbb{X}_n) = \prod_i^n p_\theta(X_i),$$

for the likelihood function at $\theta \in \Theta$, where p_θ is the density of P_θ with respect to μ

Use the law of large numbers to show that, for any $\theta \neq \theta_0$ in Θ,

$$\lim_{n \to \infty} \mathbb{P}\left(L_n(\mathbb{X}_n; \theta_0) > L_n(\mathbb{X}_n; \theta)\right) = 1$$

The previous result offers an asymptotic justification of why in this case the MLE is a sensible choice. *Hint: express the inequality in term of log-likelihood ratio and show that the ratio converges in probability to $K(P_{\theta_0}, P_\theta)$.*

7. Two sequences $\{X_n\}$ and $\{Y_n\}$ of random variables are *asymptotically equivalent* if $X_n - Y_n = o_P(1)$.

   (a) Let $X'_n = (X_n - \mathbb{E}[X_n])/\sqrt{\text{Var}[X_n]}$ and $Y'_n = (Y_n - \mathbb{E}[Y_n])/\sqrt{\text{Var}[Y_n]}$. Show that $\{X'_n\}$ and $\{Y'_n\}$ are *asymptotically equivalent* if $\text{Corr}(X_n, Y_n) \to 1$. Conclude that if $(X_n - \mathbb{E}[X_n])/\sqrt{\text{Var}[X_n]} \overset{D}{\to} X$ and $\text{Corr}(X_n, Y_n) \to 1$ then $(Y_n - \mathbb{E}[Y_n])/\sqrt{\text{Var}[Y_n]} \overset{D}{\to} X$.

   (b) Show that $\text{Corr}(X_n, Y_n) \to 1$ if $\frac{\mathbb{E}(X_n - Y_n)^2}{\text{Var}[X_n]} \to 0$.

8. **The Delta method with higher order expansions.**

   (a) Prove the following: let $\{X_n\}$ and X be a sequence of random vectors and a random vector in $\mathbb{R}^d$ and $\{r_n\}$ a sequence of positive numbers incraesing to ∞ such that $r_n(X_n - \theta) \overset{D}{\to} X$, for some $\theta \in \mathbb{R}^d$. Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be twice differentiable at $\theta \in \mathbb{R}^d$ and with $\nabla f(\theta) = 0$. Show that

   $$r_n^2(f(X_n) - f(\theta) \overset{D}{\to} \frac{1}{2} X^\top H_f(\theta) X,$$

   where $H_f(\theta)$ is the Hessian of f at θ.

   (b) Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Bernoulli}(\theta)$ and let $\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$. We are interested in estimating the variance of the distribution, $\theta(1-\theta)$. Let $f\colon [0,1] \to [0,1]$ be given as $f(x) = x(1-x)$. Consider the estimator $f(\widehat{\theta}) = \widehat{\theta}(1-\widehat{\theta})$. Derive the asymptotic distribution of $f(\widehat{\theta})$, for all $\theta \in (0,1)$. *The limiting distribution will be different depending on whether $\theta = 1/2$ or not.*

9. Let $\{X_n\}$ and X be a sequence of random vectors and a random vector in $\mathbb{R}^d$, respectively, and $\{r_n\}$ a sequence of positive numbers such that $r_n \to \infty$. Suppose that $r_n(X_n - \theta) \xrightarrow{D} X$, for some $\theta \in \mathbb{R}^d$. Show that $X_n = \theta + o_P(1)$.

10. **Records.** Let $Z_1, Z_2, \ldots$ be i.i.d. continuous random variables. We say a record occurs at k if $Z_k > \max_{i<k} Z_i$. Let $R_k = 1$ if a record occurs at k, and let $R_k = 0$ otherwise. Then $R_1, R_2, \ldots$ are independent Bernoulli random variables with $\mathbb{P}(R_k = 1) = 1 - \mathbb{P}(R_k = 0) = 1/k$ for $k = 1, 2, \ldots$. Let $S_n = \sum_{k=1}^n R_k$ denote the number of records in the first n observations. Find $\mathbb{E}[S_n]$ and $\mathrm{Var}[S_n]$, and show that $(S_n - \mathbb{E}[S_n])/\sqrt{\mathrm{Var}[S_n]} \xrightarrow{D} N(0,1)$ (The distribution of S_n is also the distribution of the number of cycles in a random permutation.)

11. **Bonus problem. Written by Jisu.** In Problem 3, we used the fact that if $Z \sim N(0, I_n)$ (multivariate norml with mean 0 and variance identity), then $\frac{Z}{\|Z\|_2}$ is uniformly distributed over $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$. Intuitive way of arguing this is that both the distribution of $\frac{Z}{\|Z\|_2}$ and the uniform distribution on $\mathbb{S}^{n-1}$ are invariant under rotations, and hence they should equal. We formaly argue in the following subproblems. We let $O(n)$ be the set of $n \times n$ orthogonal matrices, i.e. $O(n) = \{A \in \mathbb{R}^{n\times n} : A^\top A = AA^\top = I_n\}$. For $A \in \mathbb{R}^{n\times n}$ and $E \subset \mathbb{R}^n$, we use the notation $AE := \{Ax \in \mathbb{R}^n : x \in E\}$. We also let $\mu_{Z/\|Z\|}$ be the induced measure on $(S, \mathcal{B}_S)$ defined as $\mu_{Z/\|Z\|_2}(E) = P\left(\frac{Z}{\|Z\|_2} \in E\right)$ for all $E \in \mathcal{B}_{\mathbb{S}^{n-1}}$, where $\mathcal{B}_{\mathbb{S}^{n-1}}$ is a Borel set of $\mathbb{S}^{n-1}$.

    We first need to define the uniform distribution on $\mathbb{S}^{n-1}$. It is indeed the $n-1$-dimensional Hausdorff measure. For any subset $U \subset \mathbb{R}^n$, let $diam(U)$ denote its diameter, i.e. $diam(U) = \sup\{\|x - y\|_2 : x, y \in U\}$. Then for any $E \subset \mathbb{R}^n$ and for any $d, \delta > 0$, define

    $$H_\delta^d(E) = \inf\left\{\sum_{i=1}^\infty (diam(U_i))^d : E \subset \bigcup_{i=1}^\infty U_i,\ diam(U_i) < \delta\right\},$$

    and let $H^d(E) = \lim_{\delta\to 0} H_\delta^d(E)$ be the d-dimensional Hausdorff measure. Then the uniform distribution $\lambda_{\mathbb{S}^{n-1}}$ on $\mathbb{S}^{n-1}$ is defined as $\lambda_{\mathbb{S}^{n-1}}(E) = \frac{H^{n-1}(E)}{H^{n-1}(\mathbb{S}^{n-1})}$ for all $E \in \mathcal{B}_{\mathbb{S}^{n-1}}$.

    For $S \subset \mathbb{R}^n$ and $G \subset \mathbb{R}^{n\times n}$, we call that G acts on S if for all $A \in O(n)$, $AS \subset S$. Also, we further say that G acts transitively on S if G acts on S and for any $x, y \in S$, there exists $A \in G$ such that $Ax = y$.

    (a) Show that $O(n)$ acts transitively on $\mathbb{S}^{n-1}$.

    Let $\mathcal{B}_S$ be the Borel set of S and let μ be a finite measure on $(S, \mathcal{B}_S)$, i.e. $\mu(S) < \infty$. We call that μ is a Haar measure on (G, S) if μ is a nonzero and $\mu(AE) = \mu(E)$ for all Borel set $E \in \mathcal{B}_S$ and $A \in G$.

(b) Show that $\mu_{Z/\|Z\|_2}$ is a Haar measure on $(O(n), \mathbb{S}^{n-1})$. You can omit the part that $AE \in \mathcal{B}_{\mathbb{S}^{n-1}}$ for $E \in \mathcal{B}_{\mathbb{S}^{n-1}}$ (which should be a repetition of Homework 2 Problem 9).

(c) Show that $\lambda_{\mathbb{S}^{n-1}}$ is a Haar measure on $(O(n), \mathbb{S}^{n-1})$. You can assume that $0 < H^{n-1}(\mathbb{S}^n) < \infty$.

It is known that Haar measure is unique up to constant, i.e. if both μ and ν are Haar measures on (G, S), then $\mu = \xi\nu$ for some $\xi > 0$ (For reference, see Exercise 11.(i) in `http://terrytao.wordpress.com/2011/09/27/254a-notes-3-haar-measure-and-the-peter-we`
In our case, we have shown that both $\mu_{Z/\|Z\|_2}$ and $\lambda_{\mathbb{S}^{n-1}}$ are Haar measures on $(O(n), \mathbb{S}^{n-1})$. Since both $\mu_{Z/\|Z\|_2}$ and $\lambda_{\mathbb{S}^{n-1}}$ are probability measures, $\mu_{Z/\|Z\|_2} = \lambda_{\mathbb{S}^{n-1}}$ should hold.