36-752, Spring 2018
Homework 5 Solution

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

Points: 100+10 pts total for the assignment.

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.

   Points: 12 pts = 7 + 5.

   Solution.

   (a)

   ($\Longrightarrow$) Since $X_n \overset{D}{\to} X$, there exists random variables Y_n and Y such that $X_n \overset{d}{=} Y_n$ for all n, $X \overset{d}{=} Y$, and $Y_n \to Y$ a.s.. Also, let $C_g := \{x : g \text{ is continuous at } x\}$, then $P(Y \in C_g) = 1$. Then for all $\omega \in \{Y_n(\omega) \to Y(\omega)\} \cap \{Y(\omega) \in G\}$, $g(Y_n(\omega)) \to g(Y(\omega))$ as well. And hence

$$
\begin{aligned}
P(g(Y_n(\omega)) \to g(Y(\omega))) &\geq P(\{Y_n(\omega) \to Y(\omega)\} \cap \{Y(\omega) \in C_g\}) \\
&= 1 - P(\{Y_n(\omega) \nrightarrow Y(\omega)\} \cup \{Y(\omega) \notin C_g\}) \\
&\geq 1 - P(Y_n(\omega) \nrightarrow Y(\omega)) + P(Y(\omega) \notin C_g) \\
&= 1.
\end{aligned}
$$

   Hence $g(Y_n) \to g(Y)$ a.s. as well. Then since g is bounded, by dominated convergence theorem (or bounded convergence theorem),

$$
\lim_{n\to\infty} \mathbb{E}[g(Y_n)] = \mathbb{E}[g(Y)].
$$

   ($\Longleftarrow$) From condition, $\lim_n \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ for all bounded continuous function f. Hence $X_n \overset{D}{\to} X$ by definition.

   (b)

Suppose random variables $\{X_n\}_{n\in\mathbb{N}}$, X, and a function g satisfy $X_n \xrightarrow{D} X$ and $P(X \in C_g) = 1$, where $C_g := \{x : g \text{ is continuous at } x\}$. Then for any bounded continuous function f and $x \in C_g$, for all $\epsilon > 0$, there exists $\epsilon' > 0$ with $\|g(x) - z\| < \epsilon'$ implying $\|f(g(x)) - f(z)\| < \epsilon$. Then from $x \in C_g$, there exists $\delta > 0$ with $\|x - y\| < \delta$ implying $\|g(x) - g(y)\| < \epsilon'$. Then $\|f(g(x)) - f(g(y))\| < \epsilon$ as well, so $f \circ g$ is continuous at x, i.e.

$$x \in C_{f\circ g}.$$

Also, since $f \circ g$ is bounded, hence from (a), $\lim_{n\to\infty} \mathbb{E}[f(g(Y_n))] = \mathbb{E}[f(g(Y))]$. Since this holds for any bounded continuous function f,

$$g(Y_n) \xrightarrow{P} g(Y).$$

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

Points: 8 pts.

Solution.

Consider measures on $([0, 1], \mathcal{B}([0, 1]))$ having densities $f_n(x) = 1 + \cos 2\pi nx$ on $[0, 1]$. Then corresponding distribution functions are for each $n \in \mathbb{N}$,

$$F_n(x) = \int_0^x f_n(x)dx = x + \frac{1}{2\pi n} \sin 2\pi nx.$$

Then for all $x \in [0, 1]$, $F_n(x) = x + \frac{1}{2\pi n} \sin 2\pi nx \to x$, so

$$F_n \xrightarrow{D} F \quad \text{with} \quad F(x) = \begin{cases} 0, & x \le 0, \\ x & -1 \le x < 1, \\ 1, & x \ge 1. \end{cases}$$

However, $f_n(x) = 1 + \cos 2\pi nx$ does not converge for any $x \in [0, 1]$.

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 $X_n(1) \xrightarrow{D} X$, 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, $X_n \overset{d}{=} Z_n \frac{\sqrt{n}}{\|Z_n\|}$).

Points: 8 pts.

Solution.
Let $Z_n \sim N(0, I_n)$, then $Z_n \frac{\sqrt{n}}{\|Z_n\|}$ is uniformly distributed over $S_{\sqrt{n}}$ from Problem 11. Hence,

$$X_n(1) \stackrel{d}{=} \frac{\sqrt{n}Z_n(1)}{\|Z_n\|} = \frac{Z_n(1)}{\sqrt{\frac{1}{n}\sum_{k=1}^n Z_n(k)^2}}.$$

Then since $\mathbb{E}[Z_n(k)^2] = 1$, from weak law of large numbers, $\frac{1}{n}\sum_{k=1}^n Z_n(k)^2 \xrightarrow{P} 1$. And $Z_n(1) \stackrel{d}{=} X$ with $X \sim N(0,1)$, hence from Slutsky Theorem,

$$X_n(1) \xrightarrow{D} \frac{X}{\sqrt{1}} = X.$$

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 $X_n \xrightarrow{D} X$. *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 $\mu_n \xrightarrow{D} \mu$.

Points: 12 pts.

Solution.
Let λ be the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}^d)$. Since $f_n \to f$ a.e. and $\int |f_n| d\lambda = \int |f| d\lambda = 1$, Scheffé's theorem implies that $\int |f_n - f| d\lambda \to 0$ as $n \to \infty$. Then for any bounded countinuous function g,

$$\begin{aligned}
|\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| &= \left| \int g(x) f_n(x) d\lambda - \int g(x) f(x) d\lambda \right| \\
&\leq \int g(x) |f_n(x) - f(x)| d\lambda \\
&\leq \sup_{x \in \mathcal{X}} |g(x)| \int |f_n(x) - f(x)| d\lambda \to 0 \text{ as } n \to \infty,
\end{aligned}$$

And hence $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ as $n \to \infty$ for any bounded continuous g, i.e. $X_n \xrightarrow{D} X$ holds.
More generally, note that $d_{\mathrm{TV}}(\mu_n, \mu) \to 0$ implies that for all $A \in \mathcal{B}$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$. Then from Portmanteau theorem, $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ holds. And hence $X_n \xrightarrow{D} X$ holds.

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 $X_n \xrightarrow{D} X$, where X_n has c.d.f. F_n for all n and X has c.d.f. F.

Points: 12 pts.

Solution.

We first check wheather $\rho(F,G)$ is a metric. Note that if $F(x-\epsilon)-\epsilon \le G(x) \le F(x+\epsilon)+\epsilon$ holds for all x, then $G(x-\epsilon) \le F((x-\epsilon)+\epsilon)+\epsilon = F(x)+\epsilon$ and $F(x)-\epsilon = F((x+\epsilon)-\epsilon)-\epsilon \le G(x+\epsilon)$, and hence $G(x-\epsilon)-\epsilon \le F(x) \le G(x+\epsilon)+\epsilon$ holds for all x, and hence $\rho(F,G) = \rho(G,F)$ holds. Also from definition, $\rho(F,G) \ge 0$. And $\rho(F,G) = 0$ implies that for all x, $F(x) \le \inf_{\epsilon>0}\{G(x+\epsilon)+\epsilon\} = G(x)$ and correspondingly $\rho(G,F) = 0$ implies $G(x) \le \inf_{\epsilon>0}\{F(x+\epsilon)+\epsilon\} = F(x)$ holds, and hence $F = G$ holds. And $F = G$ trivially implies $\rho(F,G) = 0$, so $\rho(F,G) = 0$ if and only if $F = G$. Also, for F, G, H, for any $\epsilon > 0$, and for any $x \in \mathbb{R}$,

$$\begin{aligned}
H(x) &\ge G(x-\rho(G,H)-\epsilon)-\rho(G,H)-\epsilon \\
&\ge F(x-\rho(F,G)-\rho(G,H)-2\epsilon)-\rho(F,G)-\rho(G,H)-2\epsilon, \\
H(x) &\le G(x+\rho(G,H)+\epsilon)+\rho(G,H)+\epsilon \\
&\le F(x+\rho(F,G)+\rho(G,H)+2\epsilon)+\rho(F,G)+\rho(G,H)+2\epsilon.
\end{aligned}$$

Hence $\rho(F,H) \le \rho(F,G)+\rho(G,H)+2\epsilon$ for any $\epsilon > 0$, and hence $\rho(F,G) \le \rho(F,G)+\rho(G,H)$. Hence ρ defines a metric on the space of c.d.f.'s.

Next, we show $\rho(F_n,F) \to 0 \iff X_n \xrightarrow{D} X$.

($\Longrightarrow$) For any continuous point x of F, let $\epsilon_n := 2\rho(F_n,F)$ and then $F(x-\epsilon_n)-\epsilon_n \le F_n(x) \le F(x+\epsilon_n)+\epsilon_n$. Then $\rho(F_n,F) \to 0$ implies that

$$F(x) = \lim_{n\to\infty}(F(x-\epsilon_n)-\epsilon_n) \le \lim_{n\to\infty} F_n(x) \le \lim_{n\to\infty}(F(x+\epsilon_n)+\epsilon_n) = F(x),$$

i.e. $\lim_{n\to\infty} F_n(x) = F(x)$. And hence $X_n \xrightarrow{D} X$ from Portmanteau theorem.

($\Longleftarrow$) Fix $\epsilon > 0$, and choose $x_1 < \cdots < x_k$ be such that $F(x_1) \le \frac{\epsilon}{2}$, $F(x_k) \ge 1-\frac{\epsilon}{2}$, $|x_{i+1}-x_i| \le \epsilon$, and all x_i's are continuous point of F. Choose N large enough so that for all $n \ge N$ and for any $i = 1,\ldots,k$, $|F_n(x_i)-F(x_i)| \le \frac{\epsilon}{2}$ holds.

Now, for any $x \in [x_1, x_k]$, there exists some $i \in [1, k-1]$ such that $x \in [x_i, x_{i+1}]$. Then

$$\begin{aligned}
F(x) &\ge F(x_i) \ge F_n(x_i) - \frac{\epsilon}{2} > F_n(x-\epsilon)-\epsilon, \\
F(x) &\le F(x_{i+1}) \le F_n(x_{i+1}) + \frac{\epsilon}{2} < F_n(x+\epsilon)+\epsilon.
\end{aligned}$$

And hence $F_n(x-\epsilon)-\epsilon \le F(x) \le F_n(x+\epsilon)+\epsilon$ holds.

For any $x < x_1$, note that $F_n(x_1) \le F(x_1) + |F_n(x_1) - F(x_1)| \le \epsilon$ holds, and hence

$$F(x) \ge 0 \ge F_n(x_1) - \epsilon \ge F_n(x-\epsilon) - \epsilon,$$
$$F(x) \le F(x_1) \le \frac{\epsilon}{2} < F_n(x+\epsilon) + \epsilon.$$

And hence $F_n(x-\epsilon)-\epsilon \le F(x) \le F_n(x+\epsilon)+\epsilon$ holds.

For any $x > x_k$, note that $F_n(x_k) \ge F(x_k) - |F_n(x_k) - F(x_k)| \ge 1-\epsilon$ holds, and hence

$$F(x) \ge F(x_k) \ge 1 - \frac{\epsilon}{2} > F_n(x-\epsilon) - \epsilon,$$
$$F(x) \le 1 \le F_n(x_k) + \epsilon \le F_n(x+\epsilon) + \epsilon.$$

And hence $F_n(x-\epsilon)-\epsilon \le F(x) \le F_n(x+\epsilon)+\epsilon$ holds.

Hence in any case, $F_n(x-\epsilon)-\epsilon \le F(x) \le F_n(x+\epsilon)+\epsilon$ holds, so $\rho(F_n, F) \le \epsilon$ for $n \ge N$. Therefore, $\rho(F_n, F) \to 0$ as $n \to \infty$.

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 $\theta \ne \theta'$ implies that $P_\theta \ne P_{\theta'}$. (You may also assume that $K(P_\theta, P_{\theta'}) < \infty$ for all $\theta \ne \theta'$, 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 $\theta \ne \theta_0$ 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 $K(P_{\theta_0}, P_\theta)$.*

Points: 8 pts.

Solution.

Note that condition $L_n(\mathbb{X}_n;\theta_0) > L_n(\mathbb{X}_n;\theta)$ can be equivalently written as

$$\begin{aligned} L_n(\mathbb{X}_n;\theta_0) > L_n(\mathbb{X}_n;\theta) &\iff \log L_n(\mathbb{X}_n;\theta_0) > \log L_n(\mathbb{X}_n;\theta) \\ &\iff \sum_{i=1}^n \log p_{\theta_0}(X_i) > \sum_{i=1}^n \log p_\theta(X_i) \\ &\iff \frac{1}{n}\sum_{i=1}^n \log \frac{p_{\theta_0}(X_i)}{p_\theta(X_i)} > 0 \\ &\iff \frac{1}{n}\sum_{i=1}^n Y_i > 0, \end{aligned}$$

where $Y_i = \log \frac{p_{\theta_0}(X_i)}{p_\theta(X_i)}$ for $i = 1, \cdots, n$. Then since all P_θ has same support, $P_{\theta_0} \ll P_\theta$ holds, so expectation of Y_i under P_{θ_0} can be computed as

$$\mathbb{E}_{\theta_0}[Y_i] = \mathbb{E}_{\theta_0}\left[\log \frac{p_{\theta_0}(X_i)}{p_\theta(X_i)}\right] = K(P_{\theta_0}, P_\theta) \in (0,\infty).$$

Hence by weak law of large number,

$$\frac{1}{n}\sum_{i=1}^n Y_i \xrightarrow{P} K(P_{\theta_0}, P_\theta) > 0,$$

and hence $\frac{1}{n}\sum_{i=1}^n Y_i \xrightarrow{D} K(P_{\theta_0}, P_\theta)$ as well. since 0 is a continuous point of the cdf of constant random variable $K(P_{\theta_0}, P_\theta)$, so

$$\lim_{n\to\infty} P\left(\frac{1}{n}\sum_{i=1}^n Y_i > 0\right) = \lim_{n\to\infty} P\left(K(P_{\theta_0}, P_\theta) > 0\right) = 1.$$

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

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

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

Points: 12 pts = 5 + 7.

Solution.

(a)

From definition, $\mathbb{E}[X_n'^2] = \mathbb{E}[Y_n'^2] = 1$ and $\mathbb{E}[X_n'Y_n'] = Corr(X_n, Y_n)$ holds. Hence the L_2 difference of X_n' and Y_n' can be bounded as

$$\begin{aligned}\mathbb{E}\left[(X_n' - Y_n')^2\right] &= \mathbb{E}[X_n'^2] + \mathbb{E}[Y_n'^2] - 2\mathbb{E}[X_n'Y_n'] \\ &= 2(1 - Corr(X_n, Y_n)) \to 0 \text{ as } n \to \infty.\end{aligned}$$

And hence $X_n' - Y_n' \to 0$ in L_2, which implies $X_n' - Y_n' \xrightarrow{P} 0$. Then by applying Slutsky's theorem,

$$\frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{Var[Y_n]}} = Y_n' = X_n' - (X_n' - Y_n') \xrightarrow{P} X - 0 = X.$$

(b)

Since $Var[X_n - Y_n] \leq \mathbb{E}(X_n - Y_n)^2$, $\frac{\mathbb{E}(X_n - Y_n)^2}{\mathrm{Var}[X_n]} \to 0$ implies $a_n := \frac{Var[X_n - Y_n]}{\mathrm{Var}[X_n]} \to 0$ as well. Then a_n can be expanded as

$$\begin{aligned}a_n &= \frac{Var[X_n] + Var[Y_n] - 2Cov(X_n, Y_n)}{\mathrm{Var}[X_n]} \\ &= \sqrt{\frac{Var[Y_n]}{Var[X_n]}}\left(\sqrt{\frac{Var[X_n]}{Var[Y_n]}} + \sqrt{\frac{Var[Y_n]}{Var[X_n]}} - 2Corr(X_n, Y_n)\right).\end{aligned}$$

Hence rearranging term gives

$$Corr(X_n, Y_n) = \frac{1}{2}(1 + a_n)\sqrt{\frac{Var[X_n]}{Var[Y_n]}} + \frac{1}{2}\sqrt{\frac{Var[Y_n]}{Var[X_n]}}.$$

Then, further applying AM-GM inequality gives

$$\begin{aligned}Corr(X_n, Y_n) &\geq \frac{1}{2}(1 - |a_n|)\left(\sqrt{\frac{Var[X_n]}{Var[Y_n]}} + \sqrt{\frac{Var[Y_n]}{Var[X_n]}}\right) \\ &\geq \frac{1}{2}(1 - |a_n|) \times 2 = 1 - |a_n|,\end{aligned}$$

and hence

$$1 - |a_n| \leq Corr(X_n, Y_n) \leq 1.$$

Then from $a_n \to 0$ as $n \to \infty$, $Corr(X_n, Y_n) \to 1$ as $n \to \infty$.

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 $r_n(X_n - \theta) \xrightarrow{D} X$, 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

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

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

(b) Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim}$ Bernoulli(θ) 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.*

Points: 12 pts = 7 + 5.

Solution.

(a)

From f being twice differentiable at θ, $f(X_n)$ can be Taylor expanded at θ as

$$f(X_n) = f(\theta) + \nabla f(\theta)^\top (X_n - \theta) + \frac{1}{2}(X_n - \theta)^\top H_f(\theta)(X_n - \theta) + R(X_n - \theta),$$

where R satisfies $\lim_{h\to 0} \frac{R(h)}{\|h\|_2^2} = 0$. Hence from $\nabla f(\theta) = 0$,

$$r_n^2(f(X_n) - f(\theta)) = \frac{1}{2}(r_n(X_n - \theta))^\top H_f(\theta)(f_n(X_n - \theta)) + \|r_n(X_n - \theta)\|_2^2 \frac{R(\|X_n - \theta\|)}{\|X_n - \theta\|_2^2}.$$

Then applying continuous mapping theorem on $r_n(X_n - \theta) \xrightarrow{D} X$ imply that

$$(r_n(X_n - \theta))^\top H_f(\theta)(f_n(X_n - \theta)) \xrightarrow{D} \frac{1}{2}(r_n(X_n - \theta))^\top H_f(\theta)(f_n(X_n - \theta)),$$
$$\|r_n(X_n - \theta)\|_2^2 \xrightarrow{D} \|X\|_2^2.$$

Now we consider the remainder term $\frac{R(\|X_n - \theta\|)}{\|X_n - \theta\|_2^2}$. For $\epsilon > 0$, there exists $\delta > 0$ such that $\|h\| \leq \delta$ implies $\frac{|R(h)|}{\|h\|_2^2} \leq \epsilon$. Hence $\|X_n - \theta\| \leq \delta$ implies $\frac{|R(\|X_n - \theta\|)|}{\|X_n - \theta\|_2^2} \leq \epsilon$, i.e.

$$P\left(\frac{|R(\|X_n - \theta\|)|}{\|X_n - \theta\|_2^2} > \epsilon\right) \leq P\left(\|X_n - \theta\| > \delta\right).$$

Also, note that from Slutsky theorem, $X_n - \theta = \frac{1}{r_n}(r_n(X_n - \theta)) \xrightarrow{D} 0 \times X = 0$, and hence $X_n - \theta \xrightarrow{P} 0$. Hence, $P(\|X_n - \theta\| > \delta) \to 0$ as $n \to \infty$, which implies

$$P\left(\frac{|R(\|X_n - \theta\|)|}{\|X_n - \theta\|_2^2} > \epsilon\right) \to 0 \text{ as } n \to \infty,$$

i.e. $\frac{R(\|X_n - \theta\|)}{\|X_n - \theta\|_2^2} \xrightarrow{P} 0$. Then from Slutsky theorem,

$$\|r_n(X_n - \theta)\|_2^2 \frac{R(\|X_n - \theta\|)}{\|X_n - \theta\|_2^2} \xrightarrow{D} \|X\|_2^2 \times 0 = 0, \text{ and hence } \|r_n(X_n - \theta)\|_2^2 \frac{R(\|X_n - \theta\|)}{\|X_n - \theta\|_2^2} \xrightarrow{P} 0.$$

And then again from Slutsky theorem,

$$\begin{aligned} r_n^2(f(X_n) - f(\theta)) &= \frac{1}{2}(r_n(X_n - \theta))^\top H_f(\theta)(f_n(X_n - \theta)) + \|r_n(X_n - \theta)\|_2^2 \frac{R(\|X_n - \theta\|)}{\|X_n - \theta\|_2^2} \\ &\xrightarrow{D} \frac{1}{2} X^\top H_f(\theta) X + 0 = \frac{1}{2} X^\top H_f(\theta) X. \end{aligned}$$

(b)
Since $\mathbb{E}[X_i] = \theta$ and $Var[X_i] = \theta(1-\theta)$, we have the asymptotic normality of $\hat{\theta}$ as

$$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_i]) \xrightarrow{D} N(0, Var[X_i]) = N(0, \theta(1-\theta)).$$

Note that $f'(x) = 1 - 2x$ and $f''(x) = -2$.
When $\theta \neq \frac{1}{2}$, $f'(\theta) \neq 0$. Hence from the usual delta method, we have the asymptotic distribution of $f(\widehat{\theta})$ as

$$\sqrt{n}(f(\hat{\theta}) - f(\theta)) \xrightarrow{D} N(0, f'(\theta)^2\theta(1-\theta)) = N(0, \theta(1-\theta)(1-2\theta)^2).$$

When $\theta = \frac{1}{2}$, $f'(\theta) = 0$ and $f''(\theta) = -2$, and hence from (a), we have the asymptotic normality of $\hat{\theta}$ as

$$n(f(\hat{\theta}) - f(\theta)) \xrightarrow{D} \frac{1}{2} f''(\theta)(N(0, \theta(1-\theta)))^2 \overset{d}{=} -\frac{1}{4}\chi_1^2,$$

where χ_1^2 is the chi-square distribution with degree of freedom 1.

Remark.

In (a), the Taylor remainder term $R(X_n - \theta) = o_P(\|X_n - \theta\|_2^2)$ when $\|X_n - \theta\|_2 = o_P(1)$, but otherwise not necessarily. For example, let $X_n = X$ be some nonconstant random variable, and $f(x) = x^3$ with $\theta = 0$. Then the remainder term $R(X_n - \theta) = X^3$is not $o_P(|X|^2)$.

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 $X_n = \theta + o_P(1)$.

Points: 8 pts.

Solution.
From Slutsky theorem,

$$X_n - \theta = \frac{1}{r_n}\left(r_n(X_n - \theta)\right) \xrightarrow{D} 0 \times X = 0.$$

Then $X_n - \theta \xrightarrow{P} 0$ as well, i.e. $X_n - \theta = o_P(1)$. Hence,

$$X_n = \theta + (X_n - \theta) = \theta + o_P(1).$$

10. **Records.** Let $Z_1, Z_2, \dots$ 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 $R_1, R_2, \dots$ are independent Bernoulli random variables with $\mathbb{P}(R_k = 1) = 1 - \mathbb{P}(R_k = 0) = 1/k$ for $k = 1, 2, \dots$. 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 $\text{Var}[S_n]$, and show that $(S_n - \mathbb{E}[S_n])/\sqrt{\text{Var}[S_n]} \xrightarrow{D} N(0,1)$ (The distribution of S_n is also the distribution of the number of cycles in a random permutation.)

Points: 8 pts.

Solution.
Note that $\mathbb{E}[R_k] = \frac{1}{k}$ and $Var[R_k] = \frac{1}{k}(1 - \frac{1}{k}) = \frac{k-1}{k^2}$. Let $X_{n,1}, \dots, X_{n,n}$ be $X_{n,k} = R_k - \mathbb{E}[R_k]$, then $X_{n,k}$ is mean 0 and variance $\frac{k-1}{k^2}$.Also, $S_n - \mathbb{E}[S_n] = \sum_{k=1}^n X_{n,k}$ and

$$\sigma_n := \sqrt{Var[S_n]} = \sqrt{\sum_{k=1}^n \frac{k-1}{k^2}}.$$

Note that $\sigma_n^2 > \sum_{k=2}^n \frac{1}{2k} \geq \int_2^{n+1} \frac{1}{2x} dx = \frac{1}{2}\log\left(\frac{n+1}{2}\right) \to \infty$ as $n \to \infty$. Hence for any $\epsilon > 0$, there exists N such that for all $n \geq N$, $\sigma_n \geq \frac{1}{\epsilon}$. Since $|X_{n,k}| < 1$, the Lindeberg-Feller condition is satisfied as

$$\frac{1}{\sigma_n^2}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 I(|X_{n,k}| \geq \epsilon\sigma_n)\right] \leq \frac{1}{\sigma_n^2}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 I(|X_{n,k}| \geq 1)\right] = 0,$$

and hence $\lim_{n\to\infty} \frac{1}{\sigma_n^2}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 I(|X_{n,k}| \geq \epsilon\sigma_n)\right] = 0$. Hence from Lindeberg-Feller central limit theorem, $\frac{1}{\sigma_n}\sum_{k=1}^n X_{n,k}$ converges to $N(0,1)$, i.e., $(S_n - \mathbb{E}[S_n])/\sqrt{\text{Var}[S_n]} \xrightarrow{D} N(0,1)$.

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

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

and let $H^d(E) = \lim_{\delta\to 0} H^d_\delta(E)$ 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.

Points: 10 pts = 3 + 4 + 3.

Solution.

(a)

First, we show that for all $A \in O(n)$, $A\mathbb{S}^{n-1} \subset \mathbb{S}^{n-1}$. $x \in \mathbb{S}^{n-1}$ is equivalent to $\|x\|_2^2 = 1$. Then for any $A \in O(n)$, $x \in \mathbb{S}^{n-1}$ implies $\|Ax\|_2^2 = x^\top A^\top Ax = x^\top x = \|x\|_2^2 = 1$, and hence $Ax \in \mathbb{S}^{n-1}$. Therefore, $A\mathbb{S}^{n-1} \subset \mathbb{S}^{n-1}$, and $O(n)$ acts on $\mathbb{S}^{n-1}$.

Second, we show that for any $x, y \in \mathbb{S}^{n-1}$, there exists $Ax = y$. Let $\{x_1, \ldots, x_n\}$, $\{y_1, \ldots, y_n\}$ be two orthonormal basis of $\mathbb{R}^n$ with $x_1 = x$ and $y_1 = y$. Let $X = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}$ and $Y = \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix}$, then from orthonormality, $X, Y \in O(n)$. Let $A = YX^\top$, then $AA^\top = YX^\top XY^\top = YY^\top = I_n$ and $A^\top A = XY^\top YX^\top = XX^\top = I_n$, and hence $A \in O(n)$, also, $AX = YX^\top X = Y$, hence in particular, $Ax_1 = y_1$ holds, i.e. $Ax = y$. Hence $O(n)$ acts transitively on $\mathbb{S}^{n-1}$.

(b)

We need to show that $\mu_{Z/\|Z\|_2}(AE) = \mu_{Z/\|Z\|_2}(E)$ for all Borel set $E \in \mathcal{B}_{\mathbb{S}^{n-1}}$ and $A \in O(n)$. Note that $\mu_{Z/\|Z\|_2}(AE)$ can be expanded as

$$\mu_{Z/\|Z\|_2}(AE) = P\left(\frac{Z}{\|Z\|_2} \in AE\right) = P\left(\frac{A^{-1}Z}{\|Z\|_2} \in E\right).$$

Then, note that $\|Z\|_2^2 = Z^\top Z = Z^\top AA^\top Z = \|A^{-1}Z\|_2$, and hence $\frac{A^{-1}Z}{\|Z\|_2} = \frac{A^{-1}Z}{\|A^{-1}Z\|_2}$ and

$$P\left(\frac{A^{-1}Z}{\|Z\|_2} \in E\right) = P\left(\frac{A^{-1}Z}{\|A^{-1}Z\|_2} \in E\right).$$

Also note that $A^{-1}Z = A^\top Z$ is distributed as $N(0, A^\top I_n A) = N(0, I_d)$, and hence $\frac{A^{-1}Z}{\|A^{-1}Z\|_2}$ has the same distribution as $\frac{Z}{\|Z\|_2}$. Hence

$$P\left(\frac{A^{-1}Z}{\|A^{-1}Z\|_2} \in E\right) = P\left(\frac{Z}{\|Z\|_2} \in E\right) = \mu_{Z/\|Z\|_2}(E).$$

Hence combining above gives $\mu_{Z/\|Z\|_2}(AE) = \mu_{Z/\|Z\|_2}(E)$, and therefore $\mu_{Z/\|Z\|_2}$ is a Haar measure.

(c)

We need to show that $\lambda_{\mathbb{S}^{n-1}}(AE) = \lambda_{\mathbb{S}^{n-1}}(E)$ for all Borel set $E \in \mathcal{B}_{\mathbb{S}^{n-1}}$ and $A \in O(n)$. Since $\lambda_{\mathbb{S}^{n-1}}(E) = \frac{H^{n-1}(E)}{H^{n-1}(\mathbb{S}^{n-1})}$, it is equivalent to showing that $H^{n-1}(AE) = H^{n-1}(E)$. Note that $A \in O(n)$ prserves norm, i.e. $\|Ax\|_2 = \sqrt{x^\top A^\top Ax} = \sqrt{x^\top x} = \|x\|_2$, hence $diam(AU) = diam(U)$ for any subset $U \subset \mathbb{R}^n$. Hence for fixed $\delta > 0$ and for any $\{U_i\}$ with $E \subset \bigcup_{i=1}^\infty U_i$, $diam(U_i) < \delta$, $AE \subset \bigcup_{i=1}^\infty AU_i$ and $diam(AU_i) < \delta$

with $\sum_{i=1}^{\infty}(diam(AU_i))^n = \sum_{i=1}^{\infty}(diam(U_i))^n$. Hence $H_\delta^n(AE) \leq H_\delta^n(E)$, and by $E = A^{-1}(AE)$, $H_\delta^n(E) \leq H_\delta^n(AE)$ as well, and hence

$$H_\delta^n(AE) = H_\delta^n(E).$$

Hence, $H^n(AE) = \lim_{\delta\to 0} H_\delta^n(AE) = \lim_{\delta\to 0} H_\delta^n(E) = H^n(E)$, and correspondingly $\lambda_{\mathbb{S}^{n-1}}(AE) = \lambda_{\mathbb{S}^{n-1}}(E)$. Therfore, $\lambda_{\mathbb{S}^{n-1}}$ is a Haar measure.