36710-36752, Fall 2020
Homework 5

Due Friday, December 11.

1. For each n, let $X_n \sim N(0, 1/n)$ and denote its c.d.f. with F_{X_n}. Also, let X be a degenerate random variable such that $\mathbb{P}(X = 0) = 1$. Since $X_n \xrightarrow{p} 0$, $X_n \xrightarrow{D} X$. Now, the c.d.f. of X is
$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$
but the pointise limit of the sequence of c.d.f.'s $\{F_{X_n}\}_n$ is not a c.d.f., since it is not right continuous:
$$\lim_n F_{X_n}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/2 & \text{if } x = 0 \\ 1 & \text{if } x > 0. \end{cases}$$
Is there an issue here?

2. Show by example that distribution functions having densities can converge weakly even if the densities do not converge. *Hint: Consider the sequence of densities* $\{f_n\}$ *where* $f_n(x) = 1 + \cos 2\pi n x$ *on* $[0, 1]$. *Compute the corresponding c.d.f.'s...*

3. Let $X_n = (X_n(1), \ldots, X_n(n))$ be a random vector uniformly distributed over $S_{\sqrt{n}} = \{x \in \mathbb{R}^n : \|x\| = \sqrt{n}\}$, 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, $X_n \stackrel{d}{=} Z_n \frac{\sqrt{n}}{\|Z_n\|}$). This result says that, in very high-dimensions, the marginal distribution of any coordinate of a random vector uniformly distributed over the unit sphere is approximately normal. Another counter-intuitive fact about measures in high-dimensions.

4. **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 increasing to ∞ such that $r_n(X_n - \theta) \xrightarrow{D} X$, for some $\theta \in \mathbb{R}^d$. Let $f : \mathbb{R}^d \to \mathbb{R}$ 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 $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} \text{Bernoulli}(\theta)$ 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 $f : [0, 1] \to [0, 1]$ 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.*

5. Let X and Y be random variables over the probability space $(\Omega, \mathcal{F}, P)$. Assume that the range of Y is a countable subset $\mathcal{Y}$ of $\mathbb{R}$ such that $P\left(Y^{-1}(\{y\})\right) > 0$ for all $y \in \mathcal{Y}$. Show that the conditional expectation of X given Y is the random variable $g(Y)$, where the function $g\colon \mathbb{R} \to \mathbb{R}$ is given by
$$y \mapsto \frac{1}{P\left(Y^{-1}(\{y\})\right)} \int_{Y^{-1}(\{y\})} X dP.$$
In particular, if $Y = 1_A$ for some $A \in \mathcal{F}$ we may speak of the conditional expectation of X given A when referring to $\mathbb{E}[X|Y]$. This is what "conditioning on an event" means.[1]

6. If X and Y are independent random variables with finite expectations on a common probability space $(\Omega, \mathcal{F}, P)$, show that $\mathbb{E}(X|Y) = \mathbb{E}[X]$, a.e. $[P]$.
This can be proved in many ways, some simpler than others. You should try to provide a measure-theoretic proof of the following, more general result: if $\mathcal{C}$ and $\sigma(X)$ are independent σ-fields contained in $\mathcal{F}$, then $\mathbb{E}[X|\mathcal{C}] = \mathbb{E}[X]$, a.e. $[P]$.

7. Let X be a random variable on $(\Omega, \mathcal{F}, P)$ and $\mathcal{C} \subset \mathcal{F}$ a σ-field. Show that, for each $p \geq 1$,
$$\mathbb{E}\left[|\mathbb{E}[X|\mathcal{C}]|^p\right] \leq \mathbb{E}|X|^p.$$
That is, the condition expectation is a contraction on the L_p space of random variables on $(\Omega, \mathcal{F}, P)$ with finite p-th moment. In particular, show that the variance of $\mathbb{E}[X|\mathcal{C}]$ is smaller than the variance of X. This is a way of formalizing the intuition that conditioning (which can be thought of as extra information) reduces uncertainty.

[1] Ale's rant: in the theoretical statistical literature you will often see the following mis-use of the expression "conditioning on an event". In proving that a certain property holds, a general strategy is to define a high-probability good event and to show that the desired property holds in that event. Way too often the authors will then say that *"...conditionally on this good event, the claimed result follows."* In fact, there is no conditioning at all! The argument is instead as follows: let R the event that the result holds and G the good event. Then if $G \subseteq R$ and $P(G)$ is large, we must have that the probability $P(R^c)$ that the result fails is small, smaller than $P(G^c)$. As you can see, there is no conditioning.