# Math Stats

Instructor: Jonathan Niles-Weed

**Homework 3**
**Due: Sunday October 2, 11:59pm via NYU Gradescope**

Collaborated with Jonah Potzcobutt, and Andre Chen.

1. This question investigates different modes of convergence.

   (a) Show that if $\mathbb{E}|X_n - X| \to 0$ as $n \to \infty$, then $X_n \xrightarrow{p} X$. Show that the conclusion is unchanged if $\mathbb{E}|X_n - X| \to 0$ is replaced by $\mathbb{E}|X_n - X|^r \to 0$ for any $r > 0$.

   From Markov's inequality we know that:

   $$\mathbb{P}\{|X_n - X| \geq \epsilon\} \leq \frac{\mathbb{E}|X_n - X|}{\epsilon} \tag{1}$$

   Evaluating the RHS as $n \to \infty$:

   $$\mathbb{P}\{|X_n - X| \geq \epsilon\} \leq \frac{\mathbb{E}(0)}{\epsilon}$$

   Therefore, the RHS is 0, and since $\epsilon > 0$ and the LHS is less than or equal to the right, the quantity $|X_n - X|$ must be approaching 0 as $n \to \infty$ as well. We can conclude that $X_n \xrightarrow{p} X$.

   For the second part of the question, we can chose a PSD monotic transfer function: $\phi : \phi(x) = x^r \quad \forall r > 0$. We know from lecture 1 that:

   $$\mathbb{P}\{|X_n - X| \geq \epsilon\} \leq \frac{\mathbb{E}\phi(|X_n - X|)}{\phi(\epsilon)}$$

   Yielding:

   $$\mathbb{P}\{|X_n - X| \geq \epsilon\} \leq \frac{\mathbb{E}|X_n - X|^r}{(\epsilon)^r}$$

   And we arrive at the same conclusion, with the same logic as the first part of this question.

   (b) Let $X_n \sim Bern(\lambda_n)$, for some sequence $\lambda_n$ of numbers in $(0, 1)$. Show that if $\lambda_n \to 0$ as $n \to \infty$, then $X_n \xrightarrow{p} 0$.

   By definition of a Bernoulli distribution is $\mathbb{E}Bern(\lambda_n) = \lambda_n$. Defining $X_n$ as a series of coin flips, we can use an indicator function as follows:

   $$X_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X=1}$$

Where $X = 1$ is a positive event, (like flipping a coin and getting heads) and $X = 0$ is a negative event (like flipping tails on a coin). We can see that $X_n$ is our empirical estimator for the average of our bernoulli distribution, and therefore our empirical estimator for $\lambda_n$. We then consider the following inequality:

$$\mathbb{P}\{|X_n| \geq \epsilon\} \leq \frac{\mathbb{E}|X_n|}{\epsilon} = \frac{\mathbb{E}X_n}{\epsilon} \quad \textit{for } X_n \in [0, 1] \tag{2}$$

For some $\epsilon > 0$ We know that:

$$\mathbb{P}\{|X_n - \lambda_n| \geq \epsilon\} \xrightarrow[n\to\infty]{} 0 \quad \forall\epsilon > 0$$

Therefore, as long as $X_1, \ldots, X_n$ are i.i.d. then $X_n \xrightarrow{p} \mathbb{E}\lambda_n$ where $\mathbb{E}\lambda_n = \lambda_n$ and $n \to \infty$. Therefore, if $\lambda_n \to 0$ then $X_n \xrightarrow{p} 0$ as $n \to \infty$.

(c) Let $Y_n = \lambda_n^{-1} X_n$, with $X_n$ as in part (b). Show that if $\lambda_n \to 0$ as $n \to \infty$ then $Y_n \xrightarrow{p} 0$ What do you conclude about part (a)?

Due to the relationship between $\lambda_n^{-1}$ and $X_n$ we can evaluate the following bound:

$$\mathbb{P}\{|\lambda_n^{-1} X_n| \geq \epsilon\} = \mathbb{P}\{X_n \geq \lambda_n \epsilon\} \quad \textit{as } \lambda_n, X_n \in [0, 1] \tag{3}$$

We can break the inequality into 2 cases:

$$\begin{cases} \lambda_n & if \quad \lambda_n \epsilon < 1 \\ 0 & otherwise \end{cases}$$

From the problem statement, we know that if $\lambda_n \to 0$ as $n \to \infty$ then therefore $\mathbb{P}\{|\lambda_n^{-1} X_n| \geq \epsilon\} \xrightarrow{p} 0$. As $\epsilon > 0$, we conclude that $\lambda_n^{-1} X_n \xrightarrow{p} 0$ which then $Y_n \xrightarrow{p} 0$ follows.

(d) Let $X_n \sim Bern(1/2 + 1/n)$ be independent, and let $X \sim Bern(1/2)$. Does $X_n \xrightarrow{p} X$? Does $X_n \xrightarrow{d} X$?

For sake of notation, we define $\lambda_n = \frac{1}{2} + \frac{1}{n}$, as the parameter that defines the Bernoulli distribution which $X_n$ is generated, and $\lambda = \frac{1}{2}$, representing the parameter $\lambda$ that defines the Bernoulli distribution which $X$ is generated from.

To evaluate whether or not the two quantities are equivalent when $n \to \infty$, we look at their defined CDFs for all $t \in \mathbb{R}$. We have:

$$\mathbb{P}\{X_n \leq t\} = \begin{cases} 0 & t < 0 \\ 1 - \lambda_n = .5 - \frac{1}{n} & t \in [0, 1]] \\ 1 & t \geq 1 \end{cases}$$

Doing the same for $X$:

$$\mathbb{P}\{X \leq t\} = \begin{cases} 0 & t < 0 \\ 1 - \lambda = .5 & t \in [0, 1]] \\ 1 & t \geq 1 \end{cases}$$

When we consider when $n \to \infty$ then the term $\frac{1}{n} \to 0$, and the two CDFs are equivalent. Therefore, $X_n \xrightarrow{d} X$ is true. However, since $X_n$ and $X$ are not defined on the same probability disribution, they do not necessarily converge, and we can't show $X_n \xrightarrow{p} X$.

2. Show that Lemma 3.3 is false if the phrase "at which $F_T$ is continuous" is removed from the first part. In other words, construct $T_n$ and $T$ such that $\mathbb{E}f(T_n) \to \mathbb{E}f(t)$ for any bounded, continuous $f : \mathbb{R} \to \mathbb{R}$ but $F_{Tn}$ fails to converge to $F_T$ for some $t \in \mathbb{R}$. (Hint: consider $T_n$ which are uniform on the interval $[-1/n, 1/n]$.)

   Using the hint, we take $T_n$ which are uniform on the interval $[-1/n, 1/n]$ and $T \xrightarrow{d} T$.

   Using the definition of a CDF: $F_T(0) = \mathbb{P}_T(x \leq 0) = 1$. We would imagine that our empirical CDF, $F_{T_n}(0) \xrightarrow{d} F_T(0)$ would also converge, however this is not the case.

   For any finite $n$, we have $F_{T_n}(0) = \mathbb{P}_{T_n}(x \leq .0) = 0.5$ as the defined PDF of $T_n$ is symmetric around 0. Therefore, we conclude that even though $T_n \xrightarrow{d} T$ $F_{T_n}$ does not converge to $F_T$ at $t = 0$.

3. This question proves Proposition 3.5 for convergence in probability, under the additional assumption that $(T_n)_{n \geq 1}$ and $T$ all lie in a compact set.

   (a) Assume that $g$ is Lipschitz, that is, that there exists an $L \in \mathbb{R}$ such that $|g(x) - g(y)| \leq L|x - y|$ for all $x, y \in \mathbb{R}$. Show that for any $\epsilon > 0$:

   $$\mathbb{P}\{|g(T_n) - g(T) \geq \epsilon\} \leq \mathbb{P}\{|T_n - T| \geq \epsilon/L\}$$

   Conclude that $g(T_n) \xrightarrow{p} g(T)$

   I'm assuming that since the wording of "... under the addition assumption..." implies that we can assume $T_n \to T$. Therefore, we can start with the definition of $g$ being Lipscitz, and set $x = T_n$ and $y = T$

   $$\begin{aligned} |g(x) - g(y)| &\leq L|x - y| \\ |g(T_n) - g(T)| &\leq L|T_n - T| \\ \mathbb{P}\{|g(T_n) - g(T) \geq \epsilon\} &\leq \mathbb{P}\{L|T_n - T| \geq \epsilon\} \\ \mathbb{P}\{|g(T_n) - g(T) \geq \epsilon\} &\leq \mathbb{P}\{|T_n - T| \geq \epsilon/L\} \end{aligned} \tag{4}$$

   As $n \to \infty$ then $|T_n - T| \to 0$. As $\epsilon > 0$, the RHS is positive semi definite. Since our LHS is less than or equal to our RHS, our LHS must be approaching 0 as well. Evaluating our expression in the LHS, we arrive at the conclusion: $|g(T_n) - g(T)| \to 0$ as $n \to \infty$, therefore $g(T_n) \xrightarrow{p} g(T)$

(b) Let $g$ be a continuous function on a compact set $K$. It is a fact from real analysis that for any such function, there exists a continuous, non-decreasing function $\omega : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ called a modulus of continuity such that $\omega(0) = 0$ and

$$|g(x) - g(y)| \leq \omega(|x - y|) \quad \forall x, y \in K$$

Mimicking the proof in the first part, show that $g(T_n) \xrightarrow{p} g(T)$

As instructed, we can prove this by recreating the last proof:

$$\begin{aligned} |g(x) - g(y)| &\leq \omega(|x - y|) \\ |g(T_n) - g(T)| &\leq \omega(|T_n - T|) \\ \mathbb{P}\{|g(T_n) - g(T) \geq \epsilon\} &\leq \mathbb{P}\{\omega(|T_n - T|) \geq \epsilon\} \end{aligned} \tag{5}$$

As $n \to \infty$ then $|T_n - T| \to 0$. Its given that $\omega(0) = 0$, and since $\omega : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ and is non-decreasing, and continuous, we can conclude that as $|T_n - T| \to 0$ then $\omega(|T_n - T|) \to 0$ as well. Therefore, our RHS is approaching 0 as $n \to \infty$, and as $\epsilon > 0$, the RHS is positive semi definite. Since our LHS is less than or equal to our right, our LHS must be approaching 0 as well. Evaluating our expression in the LHS, we arrive at the conclusion: $|g(T_n) - g(T)| \to 0$ as $n \to \infty$, therefore $g(T_n) \xrightarrow{p} g(T)$

4. Show that if $\sqrt{n}(T_n - \mu)$ is asymptotically normal, then $T_n - \mu \xrightarrow{p} 0$. (Hint: For any random variable $X$ we have

$$\mathbb{P}\{|X| \leq \epsilon\} \leq \mathbb{P}\{X \leq -\epsilon\} + (1 - \mathbb{P}\{X \leq \epsilon/2\}),$$

for any $\epsilon > 0$.)

Plugging in $T_n - \mu$ as our random variable we have:

$$\mathbb{P}\{|T_n - \mu| \leq \epsilon\} \leq \mathbb{P}\{T_n - \mu \leq -\epsilon\} + (1 - \mathbb{P}\{T_n - \mu \leq \epsilon/2\}),$$

Within each probability, we can scale by $\sqrt{n}$ to yield a standard normal CDF:

$$\begin{aligned} &\mathbb{P}\{\sqrt{n}(|T_n - \mu|) \leq \sqrt{n}\epsilon\} \leq \mathbb{P}\{\sqrt{n}(T_n - \mu) \leq -\sqrt{n}\epsilon\} + (1 - \mathbb{P}\{\sqrt{n}(T_n - \mu) \leq \sqrt{n}\epsilon/2\}) \\ &\mathbb{P}\{\sqrt{n}(|T_n - \mu|) \leq \sqrt{n}\epsilon\} \leq F_{\mathcal{N}(0,\sigma^2)}(-\epsilon\sqrt{n}) + 1 - F_{\mathcal{N}(0,\sigma^2)}(\epsilon\sqrt{n}/2) \\ &\mathbb{P}\{\sqrt{n}(|T_n - \mu|) \leq \sqrt{n}\epsilon\} \leq 0 \end{aligned} \tag{6}$$

The RHS approaches 0 as $n \to \infty$ because each term diverges to positive or negative infinity ($-\epsilon\sqrt{n} \to -\infty$ and $\epsilon\sqrt{n}/2 \to \infty$). Plugging into the CDF of the standard normal gaussian:

$$F_{\mathcal{N}(0,\sigma^2)}(-\infty) = 0 \quad F_{\mathcal{N}(0,\sigma^2)}(\infty) = 1$$

Therefore our RHS is $0 + 1 - 1 = 0$. Therefore, for our inequality to hold, the LHS must be $\leq 0$, and since $n, \epsilon > 0$, then $|T_n - \mu| \to 0$ as $n \to \infty$.

5. In this exercise, we will prove the bound $\mathbb{E}|Z|^3 \leq 3\sqrt{2\pi}\sigma^3 \leq 8\sigma^3$ if $Z \sim \mathcal{N}(0, \sigma^2)$. In fact, we will prove something stronger: that the claim holds as long as $Z$ is centered and $\sigma^2$-subgaussian.

   (a) Show that it suffices to prove the claim for $\sigma = 1$

   Using the defined inequality in the problem statement we have:

$$\begin{aligned} \mathbb{E}|Z|^3 &\leq 8\sigma^3 \\ \mathbb{E}|Z/\sigma|^3 &\leq 8 \end{aligned} \tag{7}$$

   To prove subgaussianity we need to show:

$$\mathbb{E}e^{\lambda(Z - \mathbb{E}Z)} \leq e^{\frac{\lambda^2\sigma^2}{2}}$$

   Choosing $\sigma = 1$:

$$\begin{aligned} \mathbb{E}e^{\lambda\frac{Z}{\sigma} - \mathbb{E}\frac{Z}{\sigma}} &\leq e^{\lambda^2\sigma^2/2} \\ \mathbb{E}e^{\lambda Z - \mathbb{E}Z} &\leq e^{\lambda^2/2} \end{aligned} \tag{8}$$

   Which holds true, showing $\frac{Z}{\sigma}$ is 1-subgassian, which shows $\mathbb{E}|\frac{Z}{\sigma}|^3 \leq 8$ is true and that $\mathbb{E}|Z|^3 \leq 8\sigma^3$ follows.

   (b) Show that

$$\mathbb{E}|Z|^3 = \int_0^\infty 3t^2 \mathbb{P}\{|Z| \geq t\} dt$$

   Using the hint, we can rewrite the probability as the expectation of an indicator function. Changing the integration bounds to the relevant values and moving the expectation out, we can simply evaluate the integral and arrive at our desired statement:

$$\begin{aligned} \mathbb{E}|Z|^3 = \int_0^\infty 3t^2 \mathbb{P}\{|Z| \geq t\} dt &= \int_0^\infty \mathbb{E}3t^2 \mathbb{1}_{|Z| \geq t} dt \\ &= \mathbb{E}3 \int_0^{|Z|} t^2 dt \\ &= \mathbb{E}\frac{3t^3}{3}\Big|_0^{|Z|} \\ &= \mathbb{E}|Z|^3 \end{aligned} \tag{9}$$

   (c) Conclude via Collary 1.8

   We showed that $Z$ is 1-subgaussian, therefore:

$$\begin{aligned}\mathbb{E}|Z|^3 = \int_0^\infty 3t^2\mathbb{P}\{|Z| \geq t\}dt &\leq \int_0^\infty 3t^2 \left(2e^{\frac{-t^2}{2}}\right) dt \\ &\leq 6\int_0^\infty t^2 e^{\frac{-t^2}{2}} dt\end{aligned} \tag{10}$$

We can integrate by parts taking $u = t$ and $v = te^{-t^2/2}$. Doing so we get:

$$\begin{aligned}6\left(-te^{-t^2/2}\right) + \int_0^\infty 1e^{-t^2/2}dt &= 6\left(\sqrt{\frac{\pi}{2}}\right) \\ &= 3\sqrt{2\pi}\end{aligned} \tag{11}$$

Thus we show that:

$$\mathbb{E}|Z|^3 \leq 3\sqrt{2\pi}$$

6. In statistical practice, it is often the case that a sequence of random variables Tn satisfies a central limit theorem with unknown $\mu$. Moreover, in some situations, the limiting variance $\sigma^2$ can depend on $\mu$, which poses a challenge when attempting to use asymptotic normality for inference. The delta method provides a trick for avoiding this problem.

   (a) Let $\mu$ be an unknown constant and suppose that a $T_n$ satisfies a central limit theorem:

   $$\sqrt{n}(T_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\mu))$$

   where $\sigma : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is a known function. Suppose that $g$ is a function such that $g'(\mu) = \frac{1}{\sigma(\mu)}$. Show that $\sqrt{n}(g(T_n) - g(\mu))$ is asymptotically normal with variance 1, no matter what $\mu$ is. Such a g is called variance-stabilizing transformation.

   Applying the delta method to the statement given in the problem definition:

   $$\sqrt{n}(g(T_n) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, g'(\mu)^2\sigma^2(\mu))$$

   Using the definition of the derivative of $g(\mu)$ :

   $$g'(\mu)^2\sigma^2(\mu) = (\frac{1}{\sigma(\mu)})^2 \times \sigma^2(\mu) = 1$$

   We can see that $\sqrt{n}(g(T_n - g(\mu))) \xrightarrow{d} \mathcal{N}(0,1)$, that is, $\sqrt{n}(g(T_n - g(\mu)))$ is asymptotically normal with variance 1, no matter what $\mu$ is.

   (b) If $Z \sim \mathcal{N}(0, \sigma^2)$, then $\mathbb{E}Z^2 = \sigma^2$ and $\mathbb{E}Z^4 = 3\sigma^4$. If $Z_1, \ldots, Z_n \sim \mathcal{N}(0, \sigma^2)$ are i.i.d., show that:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}(0, 2\sigma^4)$$

Examining the expectation of $Z_i^2 - \sigma^2$

$$\mathbb{E}(\hat{Z}^2) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - \sigma^2\right) = \left(\mathbb{E}Z^2 - \sigma^2\right) = (\sigma^2 - \sigma^2) = 0$$

Doing the same for our second moment:

$$\begin{aligned} Var(\hat{Z}^2) = \mathbb{E}\left(\frac{1}{n}(\sum_{i=1}^{n} Z_i^2 - \sigma^2)^2\right) &= \mathbb{E}(Z^2 - \sigma^2)^2 \\ &= \mathbb{E}Z^4 + \sigma^4 - 2\mathbb{E}Z^2\sigma^2 \\ &= 3\sigma^4 + \sigma^4 - 2\mathbb{E}Z^2\sigma^2 \\ &= 4\sigma^4 - 2\sigma^2\mathbb{E}Z^2 \\ &= 4\sigma^4 - 2\sigma^4 \\ &= 2\sigma^4 \end{aligned} \tag{12}$$

Therefore we can see that as $n \to \infty$ then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}(0, 2\sigma^4)$$

(c) In the same setting as the previous item, show that

$$\sqrt{n}\left(log\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2\right) - log(\sigma^2)\right) \xrightarrow{d} \mathcal{N}(0, 2)$$

We can use the Delta method with $g(x) = log(x)$, therefore we have

$$\sqrt{n}\left(log\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2\right) - log(\sigma^2)\right) = \sqrt{n}\left(g\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2\right) - g(\sigma^2)\right)$$

$$\sqrt{n}(g(T_n) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, g'(\mu)^2\sigma^2)$$

The derivative of $g$ is:

$$g'(x) = \frac{1}{x}$$

Then we have:

$$\sqrt{n}(g(T_n) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, \frac{1}{\sigma^2}^2 2\sigma^4) = \mathcal{N}(0, 2)$$

And we have shown our desired quality.