Stat 330 - Tutorial 9

This week's tutorial is about convergence of random variables. With numbers, it's pretty easy to describe convergence: the sequence $\frac{1}{n}$ converges to 0 as $n\to\infty$ because $|\frac{1}{n}-0|\to 0$ as $n\to\infty$. With random variables, it's harder to say what convergence means since random variables can take many different values.

For example, let $X_1, X_2, X_3, \ldots$ be random variables. If we want to be able to say $X_n \to 0$ as $n\to\infty$, we probably want X_n to take values close to 0 with higher probability as n gets larger. Maybe we want to say that the probability X_n takes values other than 0 gets small as $n\to\infty$. It's probably not too hard to imagine how we might define convergence to a constant.

But what about convergence to another R.V.? If Y is a new R.V., Y might take infinitely many values. How should we define convergence of the sequence $\{X_n\}$ to the R.V. Y when the value of Y is random? It turns out that the answer to this question is subtle. In fact, there are many ways to define convergence of R.V.s. In this course, we will focus on two of the most common: convergence in probability and convergence in distribution. Let's start by defining these.

Let $X_1, X_2, \ldots$ be a sequence of R.V.s and Y be another R.V. We say that $\{X_n\}$ converges to Y in probability if, for any $\varepsilon>0$,

$$p(|X_n - Y| > \varepsilon) \to 0 \text{ as } n\to\infty$$

Let F_n be the CDF of X_n and G be the CDF of Y. We say that $\{X_n\}$ converges in distribution to Y if, for every x with $p(Y=x)=0$,

$$F_n(x)\to G(x)$$

We write $X_n \xrightarrow{P} Y$ if $\{X_n\}$ converges to Y in probability, and $X_n \xrightarrow{D} Y$ if $\{X_n\}$ converges to Y in distribution.

Now that we've got these definitions, let's look at some examples.

e.g. I Let $X_n \sim Ber(\frac{1}{n})$. Let $Y \sim Deg(0)$. That is, $p(Y=0)=1$; this is called the degenerate distribution. Then $X_n \xrightarrow{P} Y$ and $X_n \xrightarrow{D} Y$.

We start with convergence in probability. Let $\varepsilon > 0$.

$$p(|X_n - Y| > \varepsilon) = p(|X_n - 0| > \varepsilon) = p(X_n = 1) = \frac{1}{n} \to 0 \text{ as } n \to \infty$$

Therefore, $X_n \xrightarrow{P} Y$.

For convergence in distribution, there is more for us to check. Let F_n be the CDF of X_n and G be the CDF of Y. Then

$$F_n(x) = P(X_n \le x) = \begin{cases} 0 & \text{if } x<0 \\ 1-\frac{1}{n} & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1 \end{cases} \qquad G(x) = p(Y \le x) = \begin{cases} 0 & \text{if } x<0 \\ 1 & \text{if } x \ge 1 \end{cases}$$

Next, notice that $p(Y=x) = \begin{cases} 1 & \text{if } x=0 \\ 0 & \text{if } x \ne 0 \end{cases}$. Therefore, we must check that

$F_n(x) \to G(x)$ for all $x \ne 0$. If $x<0$ then $F_n(x) = 0 = G(x)$ for all n, so $F_n(x) \to G(x)$. Similarly, if $x \ge 1$ then $F_n(x) = 1 = G(x)$ for all n, so $F_n(x) \to G(x)$. If $0<x<1$ then $F_n(x) = 1 - \frac{1}{n} \to 0 = G(x)$, so $F_n(x) \to G(x)$. We have now shown that $F_n(x) \to G(x)$

~~for all x with $p(Y=x)=0$, so $X_n \xrightarrow{D} Y$.~~

for all x with $p(Y=x)=0$, so $X_n \xrightarrow{D} Y$.

$\square$

Part of the reason we went to all the trouble of defining these modes of convergence was so that we can discuss convergence of R.V.s to other (non-degenerate) R.V.s.

e.g. 2 Let $Y = \text{Unif}(0,1)$. Let $X_n = Y + \frac{1}{n}$. Then $X_n \xrightarrow{P} Y$ and $X_n \xrightarrow{D} Y$.

notice: ($X_n \sim \text{Unif}(\frac{1}{n}, 1+\frac{1}{n})$)

Convergence in probability is the easier one to show here. Let $\varepsilon > 0$.

$$p(|X_n - Y| > \varepsilon) = p(|Y + \tfrac{1}{n} - Y| > \varepsilon) = p(\tfrac{1}{n} > \varepsilon) = 1 \text{ for all } n > \lceil 1/\varepsilon \rceil.$$

Therefore, $p(|X_n - Y| > \varepsilon) \to 0$ as $n \to \infty$, so $X_n \xrightarrow{P} Y$.

Showing convergence in distribution is a bit more involved. To start, let

$$F_n(x) = \begin{cases} 0 & \text{if } x \le \frac{1}{n} \\ x - \frac{1}{n} & \text{if } \frac{1}{n} < x \le 1 + 1/n \\ 1 & \text{if } x > 1 + \frac{1}{n} \end{cases} \quad \text{and} \quad G(x) = \begin{cases} 0 & \text{if } x \le 0 \\ x & \text{if } 0 < x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$

be the CDFs of X_n and Y respectively. Notice that $p(Y = x) = 0$ for all x, so we need $F_n(x) \to G(x)$ for all x. It's best to approach this by looking at different cases.

Case 1: $x \le 0$. Here $x \le \frac{1}{n}$ for all n, so $F_n(x) = 0 = G(x)$, so $F_n(x) \to G(x)$.

Case 2: $0 < x \le 1$. Since $x > 0$, there is an $N > 0$ such that $\frac{1}{n} < x$ for all $n \ge N$ (e.g. $N = \lceil 1/x \rceil$). This means that, for all $n > N$, $F_n(x) = x - \frac{1}{n}$. Next, notice that $x - \frac{1}{n} \to x = G(x)$ as $n \to \infty$, so $F_n(x) \to G(x)$.

Case 3: $x > 1$. Similarly to Case 2, there is an $N' > 0$ such that $1 + \frac{1}{n} < x$ for all $n \ge N'$ (e.g. $\lceil \frac{1}{x-1} \rceil$). Then, for all $n \ge N'$, $F_n(x) = 1 = G(x)$, so $F_n(x) \to G(x)$

We have now shown that $F_n(x) \to G(x)$ for all x, so $X_n \xrightarrow{D} Y$.

$\diamond$

~~So far, every example~~

So far, in every example we have seen, convergence in probability and distribution have occurred simultaneously. At this point, you might be wondering whether they must always occur together. The answer to this turns out to be no. Specifically, convergence in distribution can happen without convergence in probability, but not the other way around. Said differently, convergence in probability implies convergence in distribution, but the converse is false.

e.g.3 Consider flipping a fair coin. Let $Y=1$ if the coin is heads and -1 if the coin is tails (the distribution of Y is called Rademacher(1/2)). Let $X_n = Y$ if n is even and $X_n = -Y$ if n is odd. Then $X_n \xrightarrow{D} Y$ but $X_n \not\xrightarrow{P} Y$.

First, we will show that $X_n \not\xrightarrow{P} Y$. Let $\varepsilon = 1/2$ (we only have to show that convergence fails for a single $\varepsilon > 0$)

$$p(|X_n - Y| > \varepsilon) = p(|X_n - Y| > 0.5)$$

$$= \begin{cases} p(|Y-Y|>0.5) & \text{if } n \text{ is even} \\ p(|-Y-Y|>0.5) & \text{if } n \text{ is odd} \end{cases}$$

$$= \begin{cases} p(0>0.5) & \text{if } n \text{ is even} \\ p(2|Y|>0.5) & \text{if } n \text{ is odd} \end{cases}$$

$$= \begin{cases} 0 & \text{if } n \text{ is even} \\ 1 & \text{if } n \text{ is odd} \end{cases}$$

Therefore, the sequence $p(|X_n - Y| > \varepsilon)$ does not converge to 0 (in fact, it does not converge to anything), so $X_n \not\xrightarrow{P} Y$.

For convergence in distribution, notice that, regardless of whether n is even or odd, X_n takes the values ± 1 with equal probability. Therefore, the CDF of X_n is $F_n(x) = \begin{cases} 0 & \text{if } x < -1 \\ 0.5 & \text{if } -1 \le x < 1 \\ 1 & \text{if } x \ge 1 \end{cases}$

This is also the CDF of Y, so $F_n(x) = G(x)$ for all x, where $G(x)$ is the CDF of Y. This also means that $F_n(x) \to G(x)$ for all x, so $X_n \xrightarrow{D} Y$.

◇

The proof that convergence in probability implies convergence in distribution is a bit long and technical, but you can find it in section 4.7 of the textbook (you are not required to know this proof).

Let's finish off our discussion of convergence by discussing some more general properties.

Let $X_1, X_2, \dots$ and Y be R.V.s

- If $X_n \xrightarrow{D} Y$ and $p(Y=a)=1$ for some $a \in \mathbb{R}$, then $X_n \xrightarrow{P} Y$. That is, if Y is a constant (degenerate) ~~rand~~ R.V., then convergence in distribution to Y implies convergence in probability to Y. This implication does not have to hold if Y is not constant!

- Assume that the MGF of Y, $m_Y(t) = E(e^{tY})$, is finite in the interval $(-t_0, t_0)$ for some $t_0 > 0$. Assume also that the MGF of each X_n, $m_{X_n}(t)$, is also finite in $(-t_0, t_0)$. If $m_{X_n}(t) \to m_Y(t)$ for all $t \in (-t_0, t_0)$ then $X_n \xrightarrow{D} Y$. This gives us an alternative way to check for convergence in distribution. This property is also used to prove a version of the ~~central limit~~ Central Limit Theorem.

- <u>The Central Limit Theorem</u>: Assume that ~~we~~ $X_1, X_2, \dots$ are iid and that $E(X_n) = \mu < \infty$ & $V(X_n) = \sigma^2 < \infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1)$ (we haven't defined this exact notation yet, but it just means that the left hand side converges to a R.V. with a $N(0,1)$ distribution).

- <u>The Weak Law of Large Numbers</u>: Assume that $X_1, X_2, \dots$ are iid and that $E(X_n) = \mu < \infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then $\bar{X}_n \xrightarrow{P} \mu$.

There is a lot more to say about convergence of random variables, but let's stop here. The application of these ideas to statistics is called asymptotics or large sample theory.

The Central Limit Theorem is one of the most important results in all of statistics. One consequence of this is that the normal distribution shows up a lot. Let's discuss some properties of the normal distribution that will be useful in this course and in the rest of your statistics studies.

Closure Under Affine Transformations: Let $X_i \sim N(\mu_i, \sigma_i^2)$ for $i=1,\dots,n$ be independent. Let $a_1,\dots,a_n, b \in \mathbb{R}$. ~~Then~~ Let $Y = b + \sum_{i=1}^{n} a_i X_i$. Then $Y \sim N\left(b + \sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$.

Chi-Squared Distribution: Let $X_1,\dots,X_n \overset{iid}{\sim} N(0,1)$. Let $Y = \sum_{i=1}^{n} X_i^2$. ~~The distribution Then~~ Then Y has a chi-squared distribution with n degrees of freedom. We write $Y \sim \chi^2(n)$ or $Y \sim \chi_n^2$. It can be shown that $E(Y)=n$ and $V(Y)=2n$. The density of Y is $f_Y(y) = \frac{1}{\Gamma(n/2)\, 2^{n/2}} y^{(n/2)-1} e^{-y/2}$ for $y>0$. Notice that this is equivalent to the Gamma$(\frac{n}{2}, 2)$ distribution (i.e. the chi-squared distribution is just a special case of the Gamma distribution).

Sample Variance: Let $X_1,\dots,X_n \overset{iid}{\sim} N(\mu,\sigma^2)$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Then $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ and $\bar{X}, S^2$ are independent. Note that if $X_1,\dots,X_n$ are not iid normal then neither of these properties are necessarily true.

T Distribution: Let $X \sim N(0,1)$ and $Y \sim \chi^2(n)$ such that X, Y are independent. Let $W = X/\sqrt{Y/n}$. Then W has a T distribution with n degrees of freedom. We write $W \sim T(n)$ or $W \sim T_n$. The density of W is

$$f_W(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi}\,\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \cdot \frac{1}{\sqrt{n}} \quad \text{for } x \in \mathbb{R}$$

Further, if $W_n \sim T(n)$, then $W_n \xrightarrow{D} N(0,1)$ as $n \to \infty$.

~~F~~ F Distribution: Let $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$ such that X, Y are independent.

Let $Z = \dfrac{X/m}{Y/n}$. Then Z has an F distribution with m numerator degrees of freedom and n denominator degrees of freedom. We write $Z \sim F(m,n)$ or $Z \sim F_{m,n}$. The density of Z is

$$f_Z(z) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n} z\right)^{(m/2)-1} \left(1 + \frac{n}{n} z\right)^{-(m+n)/2} \cdot \frac{m}{n} \quad \text{for } z > 0$$

~~Practically speaking~~ The F distribution arises in regression and ANOVA.
We end with a few properties of the F distribution.

- Let $Z \sim F(m,n)$. Then $\frac{1}{Z} \sim F(n,m)$
- Let $Z_n \sim F(m,n)$. Then $mZ_n \xrightarrow{D} \chi^2(m)$ as $n \to \infty$
- Let $W \sim T(n)$. Then $W^2 \sim F(1,n)$