Stats 150, Fall 2019
Lecture 21, Thursday, 11/14/2019

Today we'd like to finish up our discussion on Laplace transforms and move on to discuss Martingales.

1 Convergence of Random Variables and Their Distributions

As a sort of preamble into closing our discussion on Laplace transforms, we want to formalize our discussion of this with more care and precision in our language. Recall that we have a notion of convergence in distribution from before. That is, if X_n is discrete with value in continuous set S, $X_n \xrightarrow{d} X$ means

$$\mathbb{P}(X_n = s) \to \mathbb{P}(X = s), \forall_{s \in S}.$$

The main example for this that we have seen many times before is via the main result of Markov chain theory, which is essentially that $X_n \sim MC(\lambda, P)$ and P **irreducible, periodic, positive recurrent**, so that

$$\mathbb{P}(X_n = s) = \lambda P^n(s) \to \pi(s).$$

Then under these conditions, we have convergence in distribution for a limit X, $\mathbb{P}(X = s) = \pi(s)$. This is for the discrete case; we want to have a version for real (continuous) random variables X_n.
We write that $X_n \xrightarrow{d} X$ means that the CDF of X_n converges to the CDF of X to the extent possible, which means at **all continuity points** of the limit. This is a slightly technical condition, so we will draw pictures and discuss mainly for continuous CDFs as limit. Famous examples are:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \, dz,$$

which is the standard normal. Also, another example is

$$F(x) = \int_{0}^{x} e^{-t} \, dt, \quad x \geq 0,$$

which is the exponential. Convergence means

$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x) \text{ at all continuity points.} \tag{1}$$

From our probability course, we have the famous **Central Limit Theorem** which is to say that for $X_1, X_2, \ldots$ iid, with $\mathbb{E}X_i = \mu, \mathrm{Var}(X_i) = \sigma^2$, then we either add the observations or take their means. This gives:

$$Z_n := \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z.$$

Most statisticians simply call this standard normal variable Z with

$$\mathbb{P}(Z \leq x) = \Phi(x).$$

If we write this out, we have the exact same thing as in (1) above, where we can simply stick anything into X_n in (1).

The question arises in the audience as to why our definition uses the CDF. When we have a convergence in distribution, it has to be that all the point probabilities go to zero.
We have a fact via analysis to check that if $F_n(x)$ is a CDF (which means it is increasing, right-continuous, and so on), and $F_n(x) \to F(x)$ for all x, and F is continuous, then

$$\sup_x \Big| F_n(x) - F(x) \Big| \to 0.$$

Now back to our present setting with the Law of Large Numbers, consider the degenerate case where the limit is discrete. With this same setting, consider:

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{d} \mu = \mathbb{E}X.$$

To see what this says, we can write this out in cases:

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n}{n} \leq x \right) \to \begin{cases} 0, & x < \mu \\ 1, & x > \mu \end{cases}.$$

Now notice that this says nothing at $x = \mu$ because the limit has a jump of 1 at this point. We have to look closer to get anything useful.

Consider the equivalent statement:

$$\mathbb{P}\left(\left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| > \epsilon \right) \to 0.$$

We introduce these concepts only as a unifying language; we have dealt with these in previous courses, and we would like to have convergence in distribution as a way to bring our ideas together.

2 Mean Square

We can have another sort of convergence, namely "Mean square" convergence. We say that $X_n \to X$ in mean square (L^2). This means that

$$\mathbb{E}(X_n - X)^2 \to 0, \quad \text{as } n \to \infty.$$

This implies, via Chebyshev's inequality, that

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \frac{\mathbb{E}(X_n - X)^2}{\epsilon^2} \to 0, \quad \forall_{\epsilon > 0}.$$

This via analysis means $X_n \xrightarrow{d} X$. Now, one of the reasons for these ideas is to bring them into an application.

3 Application of Convergence in Distribution ($\xrightarrow{d}$)

Let $X \geq 0$, $\lambda > 0$, and define $X_\lambda := N_\lambda(X)$, where N_λ is a Poisson process with rate λ independent of X. We have a random variable X with whatever value, and along a timeline, we have a PPP with rate λ. We like to think that λ is under our control, so that we can draw the following diagrams:

From our picture, we can count $N_\lambda(X) = 5$ points under λ before X. Fix arbitrary X in our minds so that we can tweak λ. What happens when $\lambda \uparrow \infty$? First of all, suppose $X = t$ is fixed. Then, $N_\lambda(X) = N_\lambda(t) \sim \text{Poisson}(\lambda t)$, like it always is with this notation. This implies

$$\mathbb{E}N_\lambda(t) = \lambda t, \text{ and } \text{Var}(N_\lambda(t)) = \lambda t.$$

Recall the critical fact about the Poisson distribution that its variance is the same as its mean. Now if we take $N_\lambda(t)$ and divide by λt, taking a Poisson variable and dividing it by its mean, via the Weak Law of Large Numbers (WLLN), we should be getting the rate per unit area in the plane of the PPP. That is,

$$\frac{N_\lambda(t)}{\lambda t} \xrightarrow{d} 1.$$

Essentially, we have this convergence in distribution via convergence in Mean square, from the fact that

$$\mathbb{E}\left(\frac{N_\lambda(t)}{\lambda t} - 1\right)^2 = \frac{\lambda t}{(\lambda t)^2} = \frac{1}{\lambda t} \to 0.$$

This comes to no surprise; these computations are simply for familiarity. However, this was for a fixed time $X = t$. What about a random X?
Let's first notice that

$$\frac{N_\lambda(t)}{\lambda} \xrightarrow{d} t, \text{ as } \lambda \to \infty,$$

which holds for every fixed $t \geq 0$. Now, let's make t random with $t \xmapsto{X}$, where X is independent of $(N_\lambda(t), t \geq 0)$. We may guess, hope, or imagine that

$$\boxed{\frac{N_\lambda(X)}{\lambda} \xrightarrow{d} X}.$$

In fact (and we will check for homework), we have that

$$\frac{N_\lambda(X)}{\lambda} \to X$$

in mean square if $\mathbb{E}X < \infty$. From this, we can get that for any distribution of X,

$$\frac{N_\lambda(X)}{\lambda} \xrightarrow{d} X.$$

We have an assertion that this works for all fixed t, so we should believe that it will work for all t, simultaneously.

Now, consider the exact distribution of $N_\lambda(X)$. In particular, consider that we have a picture previously where we had 5 points. We can ask, what is the probability of getting N points? To get this, we have to condition on X, which is to write down the answer as if we knew what X is. If $X = t$, then the probability is

$$\mathbb{P}(N_\lambda(X) = n) = \frac{e^{-\lambda X}(\lambda X)^n}{n!},$$

which is a trivial formula for fixed X. Now if we do not know the value of X, we take expectations (law of total probability). If we know the conditional probability, we take the expectation. So in the general case (we do not know the value of X), we simply write:

$$\boxed{\mathbb{P}(N_\lambda(X) = n) = \mathbb{E}\left[\frac{e^{-\lambda X}(\lambda X)^n}{n!}\right]} \tag{2}$$

That is, for the discrete and continuous case, we respectively have

$$\begin{aligned} &= \sum_{x=0}^{\infty} \frac{e^{-\lambda x}(\lambda x)^n}{n!} \mathbb{P}(X = x) \\ &= \int_0^{\infty} \frac{e^{-\lambda x}(\lambda x)^n}{n!} f_X(x)\ dx. \end{aligned}$$

This combines

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{E}[\mathbb{1}_A] \\ \mathbb{E}(Y) &= \mathbb{E}\left[\mathbb{E}(Y|X)\right]. \end{aligned}$$

Now, we have another fact:

$$\boxed{\frac{N_\lambda(X)}{\lambda} \xrightarrow{d} X \text{ as } \lambda \to \infty}.$$

Notice that $N_\lambda(X)$ is a discrete random variable with values $\{0, 1, 2, \dots\}$, so that $\frac{N_\lambda(X)}{\lambda}$ is a discrete random variable with values $\{0, \frac{1}{\lambda}, \frac{2}{\lambda}, \dots\}$. Additionally,

$$\mathbb{P}\left(\frac{N_\lambda(X)}{\lambda} \le x\right) \to \mathbb{P}(X \le x)$$

at all continuity points of limit CDF. We may ask why bother with these considerations?
1) First of call, this provides a nice illustration for convergence in distribution.
2) Second it shows that if we know the distribution of $N_\lambda(X)$ for all $\lambda > 0$, we can find (by limit) the distribution of X. We essentially have a formula for working out the possibly continuous object.
That is, we have just shown how to invert a Laplace transform!

4 Review of Laplace Transform

Recall this is simply

$$\varphi_X(\lambda) := \mathbb{E}e^{-\lambda X}.$$

If we know $\phi_X(\lambda)$, we know

$$\mathbb{P}(N_\lambda(X) = 0) = \mathbb{E}e^{-\lambda X}$$

from the case $n := 0$ of the equation (2) above. Let's look at what happens when we differentiate the Laplace transform:

$$\begin{aligned} \frac{d}{dx}\varphi_X(\lambda) &= \frac{d}{d\lambda}\mathbb{E}e^{-\lambda X} \\ &= \mathbb{E}\frac{d}{dx}e^{-\lambda X} \\ &= \mathbb{E}(-X)e^{-\lambda X} \\ &= -\mathbb{E}Xe^{-\lambda X}, \end{aligned}$$

which we can compare against:

$$\begin{aligned}\mathbb{P}(N_\lambda(X) = 1) &= \lambda X \frac{e^{-\lambda X}}{1!} \\ &= \lambda \left(-\frac{d}{dx} \varphi_X(X) \right) / 1!\end{aligned}$$

Then we can continue for the n case:

$$\boxed{\mathbb{P}(N_\lambda(X) = n) = \frac{\lambda^n}{n!} \left(-\frac{d}{d\lambda} \right)^n \varphi_X(\lambda)}.$$

which we can get from repeated differentiation of the Laplace transform! Pitman cites "An Introduction to Probability Theory and Its Applications, Volume II" by Feller and suggests that we look into this if we have a background in analysis. The argument here is extremely clever in the combination of ideas in inverting a Laplace transform, using convergence in distribution and conditioning to bring together small and intuitive steps in a Poisson point process.

5 A First Look at Martingales

Let's paint the setting. Consider some background process $X_0, X_1, X_2, \ldots$ which carries some information that is evolving over time. We think of the stretch as a vector $(X_0, X_1, \ldots, X_n)$ to be a vector of the history up to stage n. We caution that the Xs might not be numerical, as they may be letters, permutations, trees, or anything that can encode the available data up to time n.

We have a numerical process which we acll $M_0, M_1, M_2, \ldots,$ and the idea is that M_n is our accumulated fortune from some gambling game. For a **Martingale**, the game is **fair**. We also have a notion of a "sub martingale" where the game is favorable, or a "super martingale" where the game is unfavorable. All of these definitions are conditionally given the past. Following Durett, we do not assume that M_0 is a function of X_0. That is, we make no assumptions on the starting condition (M_0, X_0) except that $\mathbb{E}|M_0| < \infty$. Now, we can talk about the Martingale property.

Let M_n be a function of $(M_0, X_0, X_1, \ldots, X_n)$ in some general way. The key property is that if we look at the conditional expectation of the next variable, if we know the history, this expectation should be equal to the previous variable (via the fairness condition).

$$\mathbb{E}(M_{n+1} \mid M_0, X_0, X_1, \ldots, X_n) = M_n, \quad \forall_{n=0,1,2,\ldots}$$

This equality is $\leq$ in a super martingale and $\geq$ in a sub martingale. This inequality is the same for every step. To make sure our expectation makes sense, we need to make sure that $\mathbb{E}|M_n| < \infty$. To be pedantic, this is a Martingale with respect to the X_n sequence.

5.1 Examples

There are many old friends who we can revisit here.
Example 1. Take $X_1, X_2, \ldots$ iid with $\mathbb{E}|X_i| < \infty$ and $\mathbb{E}X_i = \mu$. Take $S_n = S_0 + X_1 + \cdots + X_n$, where S_0 is assumed to be independent of $X_1, X_2, \ldots$. Then we can say (S_n) is a Martingale (MG) relative to (X_n) if and only if $\mathbb{E}X = 0$.
Similarly, (S_n) is a super martingale relative to (X_n) if and only if $\mathbb{E}X \leq 0$.
Finally, (S_n) is a sub martingale if and only if $\mathbb{E}X \geq 0$.
Notice that these inequalities are weak (not strict) so that a martingale (fair) is both a super martingale and sub martingale.

Example 2. Consider the same setup but now ask how we may make S_n^2 into a Martingale? Let's write this as:

$$\begin{aligned}
S_{n+1}^2 - S_n^2 &= (S_n + X_{n+1})^2 - S_n^2 \\
&= S_n^2 + 2S_n X_{n+1} + X_{n+1}^2 - S_n^2 \\
&= 2S_n X_{n+1} + X_{n+1}^2
\end{aligned}$$

Then this equality gives that:

$$\begin{aligned}
&\mathbb{E}\left(S_{n+1}^2 - S_n^2 \mid S_0, X_1, \ldots, X_n\right) \\
&= \mathbb{E}(2S_n X_{n+1} + X_{n+1}^2 \mid S_0, X_1, \ldots, X_n) \\
&= 2\mathbb{E}(S_n X_{n+1} \mid S_0, X_1, \ldots, X_n) + \mathbb{E}(X_{n+1}^2 \mid S_0, X_1, \ldots, X_n) \\
&= 2(\mathbb{E}X)S_n + \mathbb{E}X^2 \\
&= \mathbb{E}X^2 \text{ if } \mathbb{E}X = 0,
\end{aligned}$$

where S_n is a function of all the other history. We would like to condition in some way to make S_n go away. One way is to condition $\mathbb{E}X = 0$, so that the above simply equals $\mathbb{E}X^2$.
In conclusion, to get this result, if $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$, then S_n is a Martingale relative to the history of (X_n) and $(S_n^2 - n\mathbb{E}X^2)$ is a Martingale relative to (X_n). This is because

$$\begin{aligned}
&\mathbb{E}(S_{n+1}^2 - (n+1)\mathbb{E}X^2 \mid S_0, X_1, \ldots, X_n) \\
&= \mathbb{E}(S_{n+1}^2 \mid S_0, X_1, \ldots, X_n) - (n+1)\mathbb{E}X^2 \\
&= S_n^2 - n\mathbb{E}X^2.
\end{aligned}$$

That is, $M_n := S_n^2 - n\mathbb{E}X^2$ is a Martingale.

Example 3. Take $X_0, X_1, X_2, \ldots$ to be a Markov chain with parameters (λ, P). Take h to be a real function. Then

$$\mathbb{E}(h(X_{n+1}) \mid X_0, \ldots, X_n) = (Ph)(X_n)$$

Now because

$$(Ph)(x) = \sum_y P(x,y)h(y).$$

So if $Ph = h$ (that is, h is harmonic), then

$$\overbrace{\mathbb{E}(h(X_{n+1} \mid X_0, \ldots, X_n))}^{M_{n+1}} = \overbrace{h(X_n)}^{M_n}.$$

Then $M_n := h(X_n)$ is a Martingale, and the idea of a Martingale is a **generalization** or abstraction of the idea of a harmonic function of a Markov chain. In a way, martingales have been a part of our past discussions. Nearly all the things that have worked out in our past lectures work out due to the Martingale process. We'll see this more in our readings in the text.

Lecture ends here.