STAT 801 Lecture 8

Reading for Today's Lecture: ?
Goals of Today's Lecture:

- Develop convergence in distribution tools.
-

Today's notes

Last time we used the Fourier inversion formula to prove the local central limit theorem:

Framework:

- $X_1, \ldots, X_n$ are iid with mean 0 and variance 1
- $T = n^{1/2}\bar{X}$
- $\phi(t) = E(e^{itX_1})$ is the characteristic function of a single X

We concluded

$$E(e^{itT}) = \left[\phi(n^{-1/2}t)\right]^n$$

We differentiated ϕ to obtain

$$\begin{aligned}
\phi(0) &= 1 \\
\phi'(0) &= iE(X_1) = 0 \\
\phi''(0) &= -E(X_1^2) = -1
\end{aligned}$$

It now follows that

$$\begin{aligned}
E(e^{itT}) &\approx [1 - t^2/(2n) + o(1/n)]^n \\
&\to e^{-t^2/2}
\end{aligned}$$

Apply the Fourier inversion formula to deduce

$$f_T(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

which is the standard normal random density.

This proof of the central limit theorem is not terribly general since it requires T to have a bounded continuous density. The usual central limit theorem is a statement about cdfs not densities and is

$$P(T \le t) \to P(Z \le t)$$

Convergence in Distribution

In undergraduate courses we often teach the central limit theorem: if $X_1, \ldots, X_n$ are iid from a population with mean μ and standard deviation σ then $n^{1/2}(\bar{X} - \mu)/\sigma$ has approximately a normal distribution. We also say that a Binomial(n, p) random variable has approximately a $N(np, np(1-p))$ distribution.

To make precise sense of these assertions we need to assign a meaning to statements like "X and Y have approximately the same distribution". The meaning we want to give is that X and Y have nearly the same cdf but even here we need some care. If n is a large number is the $N(0, 1/n)$ distribution close to the distribution of $X \equiv 0$? Is it close to the $N(1/n, 1/n)$ distribution? Is it close to the $N(1/\sqrt{n}, 1/n)$ distribution? If $X_n \equiv 2^{-n}$ is the distribution of X_n close to that of $X \equiv 0$?

The answer to these questions depends in part on how close close needs to be so it's a matter of definition. In practice the usual sort of approximation we want to make is to say that some random variable X, say, has nearly some continuous distribution, like $N(0, 1)$. In this case we must want to calculate probabilities like $P(X > x)$ and know that this is nearly $P(N(0, 1) > x)$. The real difficulty arises in the case of discrete random variables; in this course we will not actually need to approximate a distribution by a discrete distribution.

When mathematicians say two things are close together they either can provide an upper bound on the distance between the two things or they are talking about taking a limit. In this course we do the latter.

Definition: A sequence of random variables X_n converges in distribution to a random variable X if

$$E(g(X_n)) \to E(g(X))$$

for every bounded continuous function g.

Theorem: The following are equivalent:

1. X_n converges in distribution to X.
2. $P(X_n \le x) \to P(X \le x)$ for each x such that $P(X = x) = 0$
3. The characteristic functions of X_n converge to that of X:

   $$E(e^{itX_n}) \to E(e^{itX})$$

   for every real x.

These are all implied by

$$M_{X_n}(t) \to M_X(t) < \infty$$

for all $|t| \le \epsilon$ for some positive ϵ.

Now let's go back to the questions I asked:

- $X_n \sim N(0, 1/n)$ and $X = 0$. Then

$$P(X_n \le x) \to \begin{cases} 1 & x > 0 \\ 0 & x < 0 \\ 1/2 & x = 0 \end{cases}$$

  Now the limit is the cdf of $X = 0$ except for $x = 0$ and the cdf of X is not continuous at $x = 0$ so yes, X_n converges to X in distribution.

- I asked if $X_n \sim N(1/n, 1/n)$ had a distribution close to that of $Y_n \sim N(0, 1/n)$. The definition I gave really requires me to answer by finding a limit X and proving that both X_n and Y_n converge to X in distribution. Take $X = 0$. Then

$$E(e^{tX_n}) = e^{t/n + t^2/(2n)} \to 1 = E(e^{tX})$$

  and

$$E(e^{tY_n}) = e^{t^2/(2n)} \to 1$$

  so that both X_n and Y_n have the same limit in distribution.

- Now multiply both X_n and Y_n by $n^{1/2}$ and let $X \sim N(0, 1)$. Then $\sqrt{n}X_n \sim N(n^{-1/2}, 1)$ and $\sqrt{n}Y_n \sim N(0, 1)$. You can use characteristic functions to prove that both $\sqrt{n}X_n$ and $\sqrt{n}Y_n$ converge to $N(0, 1)$ in distribution.

- If you now let $X_n \sim N(n^{-1/2}, 1/n)$ and $Y_n \sim N(0, 1/n)$ then again both X_n and Y_n converge to 0 in distribution.

- If you multiply these X_n and Y_n by $n^{1/2}$ then $n^{1/2}X_n \sim N(1, 1)$ and $n^{1/2}Y_n \sim N(0, 1)$ so that $n^{1/2}X_n$ and $n^{1/2}Y_n$ are **not** close together in distribution.

- You can check that $2^{-n} \to 0$ in distribution.

Here is the message you are supposed to take away from this discussion. You do distributional approximations by showing that a sequence of random variables X_n converges to some X. The limit distribution should be non-trivial, like say $N(0,1)$. We don't say X_n is approximately $N(1/n, 1/n)$ but that $n^{1/2}X_n$ converges to $N(0,1)$ in distribution.

The Central Limit Theorem

If $X_1, X_2, \cdots$ are iid with mean 0 and variance 1 then $n^{1/2}\bar{X}$ converges in distribution to $N(0,1)$. That is,

$$P(n^{1/2}\bar{X} \le x) \to \frac{1}{2\pi}\int_{-\infty}^{x} e^{-y^2/2}dy\,.$$

Proof: As before

$$E(e^{itn^{1/2}\bar{X}}) \to e^{-t^2/2}$$

This is the characteristic function of a $N(0,1)$ random variable so we are done by our theorem.

Edgeworth expansions

In fact if $\gamma = E(X^3)$ then

$$\phi(t) \approx 1 - t^2/2 - i\gamma t^3/6 + \cdots$$

keeping one more term. Then

$$\log(\phi(t)) = \log(1+u)$$

where

$$u = -t^2/2 - i\gamma t^3/6 + \cdots$$

Use $\log(1+u) = u - u^2/2 + \cdots$ to get

$$\log(\phi(t)) \approx -[t^2/2 - i\gamma t^3/6 + cdots] - [\ldots]^2/2 + \cdots$$

which rearranged is

$$\log(\phi(t)) \approx -t^2/2 + i\gamma t^3/6 + \cdots$$

Now apply this calculation to

$$\log(\phi_T(t)) \approx -t^2/2 + iE(T^3)t^3/6 + \cdots$$

Remember $E(T^3) = \gamma/\sqrt{n}$ and exponentiate to get

$$\phi_T(t) \approx e^{-t^2/2} \exp\{i\gamma t^3/(6\sqrt{n}) + \cdots\}$$

You can do a Taylor expansion of the second exponential around 0 because of the square root of n and get

$$\phi_T(t) \approx e^{-t^2/2}(1 - i\gamma t^3/(6\sqrt{n}))$$

neglecting higher order terms. This approximation to the characteristic function of T can be inverted to get an **Edgeworth** approximation to the density (or distribution) of T which looks like

$$f_T(x) \approx \frac{1}{\sqrt{2\pi}} e^{-x^2/2}[1 - \gamma(x^3 - 3x)/(6\sqrt{n}) + \cdots]$$

Remarks:

1. The error using the central limit theorem to approximate a density or a probability is proportional to $n^{-1/2}$
2. This is improved to n^{-1} for symmetric densities for which $\gamma = 0$.
3. The expansions are **asymptotic**. This means that the series indicated by $\cdots$ usually does **not** converge. When $n = 25$ it may help to take the second term but get worse if you include the third or fourth or more.
4. You can integrate the expansion above for the density to get an approximation for the cdf.

Multivariate convergence in distribution

Definition: $X_n \in R^p$ converges in distribution to $X \in R^p$ if

$$E(g(X_n)) \to E(g(X))$$

for each bounded continuous real valued function g on R^p.

This is equivalent to either of

Cramér Wold Device: $a^t X_n$ converges in distribution to $a^t X$ for each $a \in R^p$

or

Convergence of Characteristic Functions:

$$E(e^{ia^t X_n}) \to E(e^{ia^t X})$$

for each $a \in R^p$.

Extensions of the CLT

1. If $Y_1, Y_2, \cdots$ are iid in R^p with mean μ and variance covariance Σ then $n^{1/2}(\bar{Y} - \mu)$ converges in distribution to $MVN(0, \Sigma)$.

2. If for each n we have a set of independent mean 0 random variables $X_{n1}, \ldots, X_{nn}$ and $E(X_{ni}) = 0$ and $Var(\sum_i X_{ni}) = 1$ and
$$\sum E(|X_{ni}|^3) \to 0$$
then $\sum_i X_{ni}$ converges in distribution to $N(0,1)$. This is the Lyapunov central limit theorem.

3. As in the Lyapunov central CLT but replace the third moment condition with
$$\sum E(X_{ni}^2 1(|X_{ni}| > \epsilon)) \to 0$$
for each $\epsilon > 0$ then again $\sum_i X_{ni}$ converges in distribution to $N(0,1)$. This is the Lindeberg central limit theorem. (Lyapunov's condition implies Lindeberg's.)

4. There are extensions to random variables which are not independent. Examples include the m-dependent central limit theorem, the martingale central limit theorem, the central limit theorem for mixing processes.

5. Many important random variables are not sums of independent random variables. We handle these with Slutsky's theorem and the δ method.

Slutsky's Theorem: If X_n converges in distribution to X and Y_n converges in distribution (or in probability) to c, a constant, then $X_n + Y_n$ converges in distribution to $X + c$.

Warning: the hypothesis that the limit of Y_n be constant is essential.

The delta method: Suppose a sequence Y_n of random variables converges to some y a constant and that if we define $X_n = a_n(Y_n - y)$ then X_n converges in distribution to some random variable X. Suppose that f is a differentiable function on the range of Y_n. Then $a_n(f(Y_n) - f(y))$ converges in distribution to $f'(y)X$. If X_n is in R^p and f maps R^p to R^q then f' is the $q \times p$ matrix of first derivatives of components of f.

Example: Suppose $X_1, \ldots, X_n$ are a sample from a population with mean μ, variance σ^2, and third and fourth central moments μ_3 and μ_4. Then
$$n^{1/2}(s^2 - \sigma^2) \Rightarrow N(0, \mu_4 - \sigma^4)$$

where $\Rightarrow$ is notation for convergence in distribution. For simplicity I define $s^2 = \overline{X^2} - \bar{X}^2$.

We take Y_n to be the vector with components $(\overline{X^2}, \bar{X})$. Then Y_n converges to $y = (\mu^2 + \sigma^2, \mu)$. Take $a_n = n^{1/2}$. Then

$$n^{1/2}(Y_n - y)$$

converges in distribution to $MVN(0, \Sigma)$ with

$$\Sigma = \left[\begin{array}{cc} \mu_4 - \sigma^4 & \mu_3 - \mu(\mu^2 + \sigma^2) \\ \mu_3 - \mu(\mu^2 + \sigma^2) & \sigma^2 \end{array} \right]$$

Define $f(x_1, x_2) = x_1 - x_2^2$. Then $s^2 = f(Y_n)$ and the gradient of f has components $(1, -2x_2)$. This leads to

$$n^{1/2}(s^2 - \sigma^2) \approx n^{1/2}(1, -2\mu) \left[\begin{array}{c} \overline{X^2} - (\mu^2 + \sigma^2) \\ \bar{X} - \mu \end{array} \right]$$

which converges in distribution to the law of $(1, -2\mu)Y$ which is $N(0, a^t \Sigma a)$ where $a = (1, -2\mu)^t$. This boils down to $N(0, \mu_4 - \sigma^2)$.

Remark: In this sort of problem it is best to learn to recognize that the sample variance is unaffected by subtracting μ from each X. Thus there is no loss in assuming $\mu = 0$ which simplifies Σ and a.