STAT 450 Lecture 13

Goals of Today's Lecture:

- Introduce the idea of approximate distribution theory.
- State the usual central limit theorem
- Introduce the idea of convergence in distribution.

Today's notes

Approximate Distribution Theory

Up to now we have tried to compute the density or cdf of some transformation, $Y = g(X)$, of the data X exactly. In most cases this is not possible. Instead theoretical statisticians try to find methods to compute f_Y or F_Y approximately. There are really two standard methods:

- **Large sample theory** or **asymptotics**.
- **Monte Carlo** or **simulation** methods.

In this course we focus on large sample theory.

Asymptotics

You already know two large sample theorems:

- **The law of large numbers** or the law of averages: If $X_1, X_2, \cdots$ are independent and identically distributed random variables then the sequence of sample means,:

  $$\bar{X}_n = \frac{1}{n} \sum_1^n X_i \, ,$$

  converges to $\mu = \mathrm{E}(X_1)$.

- **The central limit theorem**: If $X_1, X_2, \cdots$ are independent and identically distributed then the sequence of standardized sample means,

  $$Z_n = n^{1/2} \frac{\bar{X}_n - \mu}{\sigma}$$

  converges in distribution to $N(0,1)$; here $\mu = \mathrm{E}(X_1)$ and $\sigma^2 = \mathrm{Var}(X_1)$. It is a hypothesis of the theorem that $\sigma < \infty$.

I want to make these theorems precise (though I don't intend to prove them convincingly). I want you to have a good intuitive grasp of the meanings of the assertions, however.

The law of large numbers

There are two different senses in which $\bar{X}_n \to \mu$: convergence in probability and almost sure convergence.

Definition: A sequence Y_n of random variables **converges in probability** to Y if, for each $\epsilon > 0$:

$$P(|Y_n - Y| > \epsilon) \to 0$$

Definition: A sequence Y_n of random variables **converges almost surely** (or **strongly**) to Y if

$$P(Y_n \to Y) = 1$$

Notice that the second kind of convergence asks to to calculate a single probability — of an event whose definition is very complicated since it mentions **all** the Y_n at the same time. The first kind of convergence involves computing a sequence of probabilities. The nth probability in the sequence mentions only 2 random variables, Y_n and Y. Typically convergence in probability is easier to prove; it is a theorem that $Y_n \to Y$ almost surely implies $Y_n \to Y$ in probability. (Notice the way we write those assertions.)

Corresponding to the two kinds of convergence are two precise versions of the law of large numbers:

Theorem: the Weak Law of Large Numbers. If $X_1, X_2, \cdots$ are independent and identically distributed random variables such that $E(|X_1|) < \infty$ then the sequence of sample means,

$$\bar{X}_n = \frac{1}{n}\sum_1^n X_i\,,$$

converges to $\mu = E(X_1)$ in probability.

Theorem: the Strong Law of Large Numbers. If $X_1, X_2, \cdots$ are independent and identically distributed random variables such that $E(|X_1|) < \infty$ then the sequence of sample means,

$$\bar{X}_n = \frac{1}{n}\sum_1^n X_i\,,$$

converges to $\mu = E(X_1)$ almost surely.

The SLLN (note the abbreviation) is harder to prove. The WLLN can be deduced, provided $E(X_1^2) < \infty$ from Chebyshev's inequality:

Chebyshev's inequality: If $\mathrm{Var}(Y) < \infty$ then

$$P(|Y - \mu_Y| > t) \le \frac{\mathrm{Var}(Y)}{t^2}$$

for all $t > 0$.

To apply the theorem let $X_1, \ldots, X_n$ be independent and identically distributed (iid) with mean μ and variance σ^2. Then $\bar{X}_n$ has mean μ and $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$. Thus

$$P(|\bar{X}_n - \mu| > \epsilon) \le \frac{\mathrm{Var}(\bar{X}_n)}{\epsilon_2} = \frac{\sigma^2}{n\epsilon^2} \to 0$$

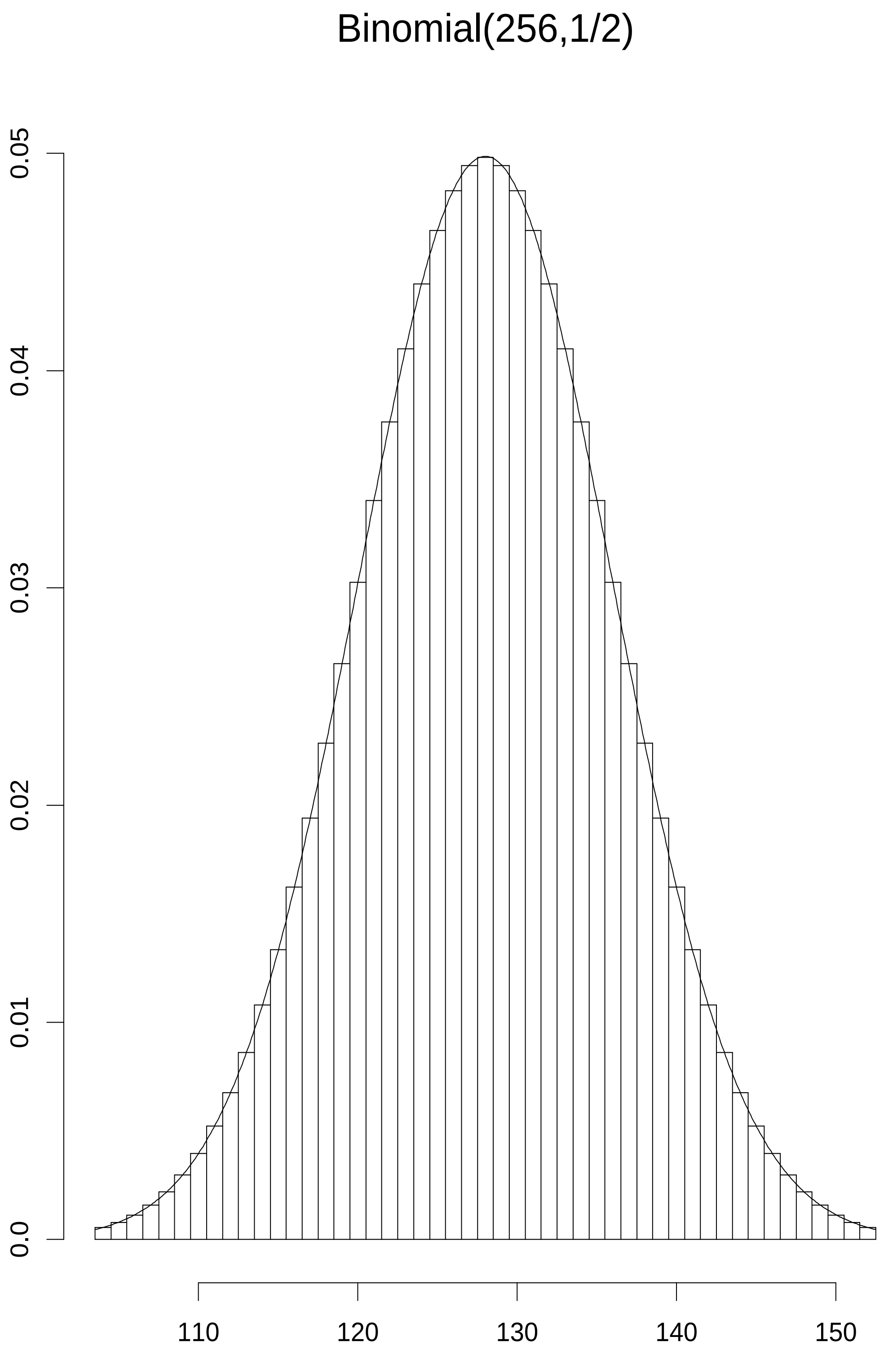
for each $\epsilon > 0$. This proves the weak law of large numbers in the special case $\mathrm{Var}(X_1) < \infty$.

This is a sort of approximate distribution calculation. We say that a certain random variable has almost the same distribution as another, namely, if n is large then the random variable $\bar{X}_n$ has almost the same distribution as the (nonrandom) quantity μ.

In some of the calculations we are about to make that approximation is good enough. In others, however, we want to know just how close to μ the quantity $\bar{X}_n$ is likely to be. The answer is provided by the central limit theorem.

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 $N(0,1)$. In previous textbooks you will have seen pictures like the following:



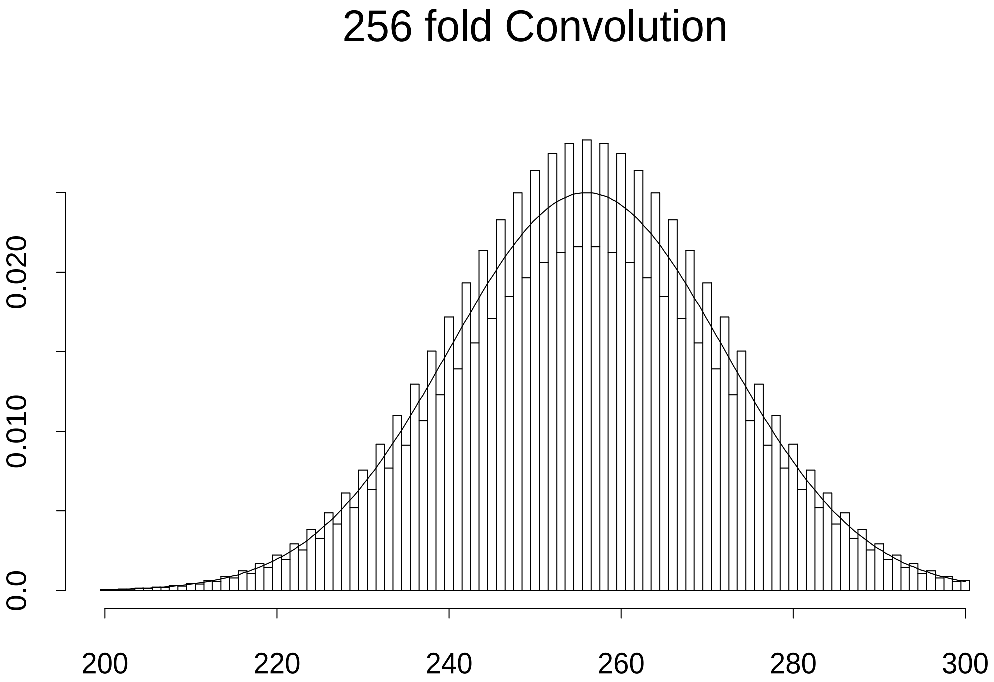
The picture above is an example of one version of the central limit theorem. It shows that the probability mass function of a $Binomial(100, 0.5)$ random variable Y_{100} is close to that of a $N(50, 5)$ random variable. To see that this is the central limit theorem setting let $X_1, X_2, \cdots$ be independent random variables with $P(X_i = 1) = p = 1 - P(X_i = 0)$. (We call the X_i Bernoulli random variables.) Then $Y_{100} = X_1 + \cdots + X_100 \sim Binomial(100, 1/2)$. The central limit theorem asserts that $\sqrt{n}(\bar{X}_n - p)/\sqrt{p(1-p)}$ is nearly $N(0, 1)$ and this quantity is actually $(Y_n - np)/\sqrt{np(1-p)}$.

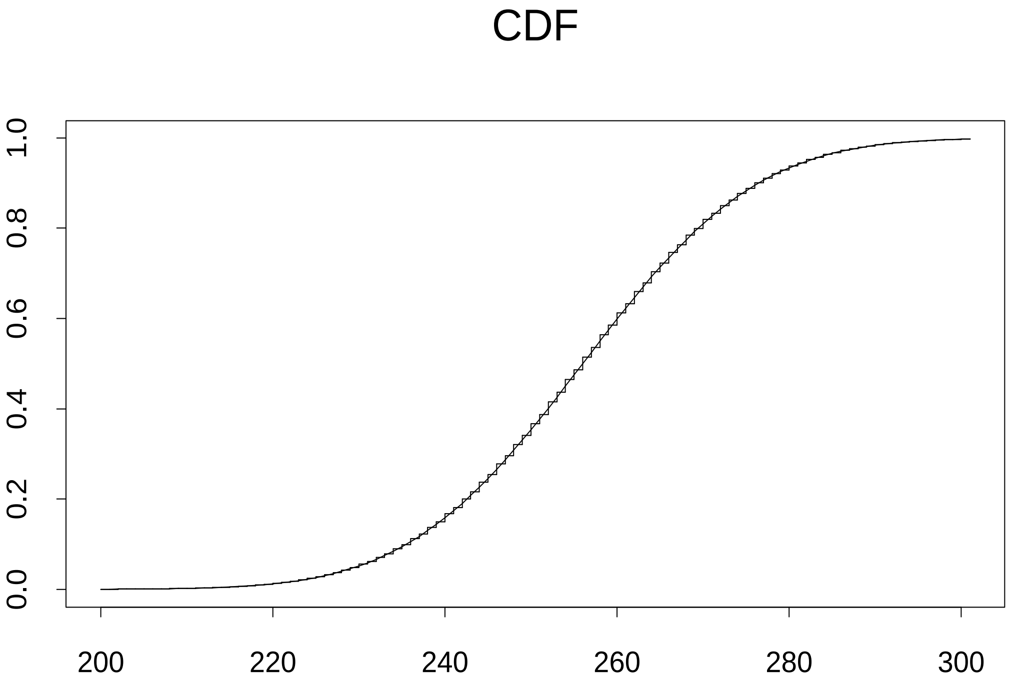
Here is another example of the central limit theorem. Here the X_i are independent and identically distributed random variables with $P(X_i = 0) = P(X_i = 2) = 127.5/256$ and $P(X_i = 1) = 1/256$. The top plot shows a histogram style plot of $P(\sum_1^{256} X_i = k)$ against k. The variable Y has mean 128 and standard deviation $16\sigma \approx 15.97$. You are meant to see that the superimposed normal curve goes between the odd number bars and the even number bars.

The plots illustrate the difference between the **local** central limit theorem, which says that the density of $\bar{X}_n$ is close to the normal density and the **global** central limit theorem which says that the cdf of $\bar{X}_n$ is close to the normal cdf. That is,

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

You should see that for this example the global central limit theorem (generally the word global is omitted and we just call this one "the" central limit theorem) provides a good approximation while the local central limit theorem does not. In general the local CLT requires more hypotheses than the global CLT.





The central limit theorem: If $X_1, X_2, \cdots$ are independent and identically distributed with mean μ and finite standard deviation σ then the sequence of standardized sample means,

$$Z_n = n^{1/2} \frac{\bar{X}_n - \mu}{\sigma}$$

converges in distribution to $N(0,1)$in the sense that

$$\mathrm{P}(Z_n \le z) \to \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$

In this course we will state (but not really prove) a number of theorems with conclusions of this form. To do so we need some mathematical tools.

Convergence in Distribution

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 moment generating 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.