Math 3081 (Probability and Statistics)

Lecture #11 of 27 $\sim$ July 22nd, 2021

Continuous Random Variables (Part 3)

- Independence, Covariance, and Correlation
- Normal Distributions
- Motivation for the Central Limit Theorem

This material represents §2.2.3-2.3.1 from the course notes, and problems 16-20 from WeBWorK 3 + 1-4 from WeBWorK 4.

Recall, I

Definition

If X and Y are continuous random variables, and there is a function $p_{X,Y}(x, y)$, defined on ordered pairs of real numbers (x, y) such that $\iint_R p_{X,Y}(x, y)\, dy\, dx = P[(X, Y) \in R]$ for every plane region R, then we call $p_{X,Y}(x, y)$ the joint probability density function of X and Y.

- To compute the probability that the values of of X and Y together land in a particular planar region R, we integrate the probability density function on the domain R.
- The most difficult part is usually setting up the integral: once it is set up, the rest is just calculation.

Recall, II

As in the case of discrete random variables, we can also recover the individual probability distributions $p_X(x)$ and $p_Y(y)$ for either variable from their joint distribution by integrating over the other variable:

Proposition (Marginal Densities)

If $p_{X,Y}(a,b)$ is the joint probability density function for the continuous random variables X and Y, then for any a and b we may compute the single-variable probability density functions for X and Y as $p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,dy$ and $p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,dx$.

Independence, I

Next, we examine independence.

- As in the discrete case, we consider two continuous random variables X and Y to be independent when knowing the value of one gives no additional information about the value of the other: $P(a < X < b | c < Y < d) = P(a < X < b)$.
- By rearranging, this says $P(a < X < b, c < Y < d) = P(a < X < b) \cdot P(c < Y < d)$, which in terms of probabilities says $\int_a^b \int_c^d p_{X,Y}(x, y)\, dy\, dx = \int_a^b p_X(x)\, dx \cdot \int_c^d p_Y(y)\, dy$.
- But since the right-hand side is also equal to the iterated integral $\int_a^b \int_c^d p_X(x) \cdot p_Y(y)\, dy\, dx$, since both sides are equal on every rectangle $[a, b] \times [c, d]$, we must have $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$ for every x and y.
- This is exactly the same condition as in the discrete case!

Independence, II

Formally, our definition is as follows:

Definition

The continuous random variables X and Y are <u>*independent*</u> *if their joint distribution $p_{X,Y}(x, y)$ is the product of their individual probability distributions $p_X(x)$ and $p_Y(y)$: in other words, when $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$ for all real numbers x, y.*

Independence, II

Formally, our definition is as follows:

Definition

The continuous random variables X and Y are *independent* *if their joint distribution $p_{X,Y}(x,y)$ is the product of their individual probability distributions $p_X(x)$ and $p_Y(y)$: in other words, when $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$ for all real numbers x, y.*

We can also extend this in the natural way to more variables:

Definition

The continuous random variables $X_1, X_2, \ldots, X_n$ are *collectively independent* *if their joint distribution is the product of their individual distributions: in other words, when $p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdot \cdots \cdot p_{X_n}(x_n)$ for all real numbers $x_1, x_2, \ldots, x_n$.*

Independence, III

Example: Determine whether the continuous random variables X and Y with joint probability density function defined by $p_{X,Y}(x, y) = \frac{1}{4}xy$ for $0 \le x \le 2$ and $0 \le y \le 2$ are independent.

Independence, III

Example: Determine whether the continuous random variables X and Y with joint probability density function defined by $p_{X,Y}(x,y) = \frac{1}{4}xy$ for $0 \le x \le 2$ and $0 \le y \le 2$ are independent.

- We just need to compute the marginal distributions and then check the independence condition.
- We see $p_X(x) = \int_0^2 \frac{1}{4}xy\,dy = \frac{1}{2}x$ and also $p_Y(y) = \int_0^2 \frac{1}{4}xy\,dx = \frac{1}{2}y$.
- Then, indeed, $p_{X,Y}(x,y) = \frac{1}{4}xy = \frac{1}{2}x \cdot \frac{1}{2}y = p_X(x) \cdot p_Y(y)$.
- Since the joint distribution function is the product of the individual distributions, X and Y are independent.

Independence, IV

Example: Determine whether the continuous random variables X and Y with joint probability density function defined by $p_{X,Y}(x,y) = \frac{1}{24}(2x+y)$ for $0 \le x \le 3$ and $0 \le y \le 2$ are independent.

- We computed last time that the marginal density functions were $p_X(x) = \dfrac{1}{12}(2x+1)$ and $p_Y(y) = \dfrac{1}{8}(y+3)$.

Independence, IV

<u>Example</u>: Determine whether the continuous random variables X and Y with joint probability density function defined by $p_{X,Y}(x,y) = \frac{1}{24}(2x+y)$ for $0 \le x \le 3$ and $0 \le y \le 2$ are independent.

- We computed last time that the marginal density functions were $p_X(x) = \dfrac{1}{12}(2x+1)$ and $p_Y(y) = \dfrac{1}{8}(y+3)$.
- We see $p_X(x) \cdot p_Y(y) = \dfrac{1}{96}(2x+1)(y+3)$, which is not $p_{X,Y}(x,y)$.
- Since the joint distribution function is not the product of the individual distributions, X and Y are not independent.

Independence, V

In this last example, we did not actually need to have the marginal distribution functions to see that the variables were not independent.

- Specifically, all we needed to notice was that the joint distribution function $p(x, y) = \frac{1}{24}(2x + y)$ cannot be written as the product of a single function of x and a single function of y.

In fact, if we can write $p_{X,Y}(x, y) = q(x) \cdot r(y)$ for some functions $q(x)$ and $r(y)$, the random variables X and Y will be independent.

- This follows because the marginal distribution functions are $p_X(x) = \int_{-\infty}^{\infty} q(x)r(y)\,dy = q(x) \cdot \int_{-\infty}^{\infty} r(y)\,dy$ and $p_Y(y) = \int_{-\infty}^{\infty} q(x)r(y)\,dx = r(y) \cdot \int_{-\infty}^{\infty} q(x)\,dx$, so $p_X(x) \cdot p_Y(y) = q(x)r(y) \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} q(x)r(y)\,dy\,dx = q(x)r(y)$. (The latter double integral is 1 because p is a pdf.)

Independence, VI

Example: If X and Y have the joint probability density function $p_{X,Y}(x,y)$ listed, determine whether X and Y are independent:

1. $p_{X,Y} = 3x^2y^7$ for $0 \le x \le 1$, $0 \le y \le 2$.
2. $p_{X,Y} = 2y^2 + x^2$ for $0 \le x \le 1$, $0 \le y \le 1$.
3. $p_{X,Y} = 2xy + x$ for $0 \le x \le 1$, $0 \le y \le 1$.
4. $p_{X,Y} = \frac{6}{7}(x+y)^2$ for $0 \le x \le 1$, $0 \le y \le 1$.
5. $p_{X,Y} = 2e^{-x-2y}$ for $x, y \ge 0$.

Independence, VI

Example: If X and Y have the joint probability density function $p_{X,Y}(x,y)$ listed, determine whether X and Y are independent:

1. $p_{X,Y} = 3x^2y^7$ for $0 \le x \le 1$, $0 \le y \le 2$.
2. $p_{X,Y} = 2y^2 + x^2$ for $0 \le x \le 1$, $0 \le y \le 1$.
3. $p_{X,Y} = 2xy + x$ for $0 \le x \le 1$, $0 \le y \le 1$.
4. $p_{X,Y} = \frac{6}{7}(x+y)^2$ for $0 \le x \le 1$, $0 \le y \le 1$.
5. $p_{X,Y} = 2e^{-x-2y}$ for $x, y \ge 0$.

We can write some of these as a product as a function of x with a function of y, but not others:

1. $3x^2y^7 = (3x^2) \cdot (y^7)$: independent.
2. $2y^2 + x^2$ cannot be written this way: not independent.
3. $2xy + x = x \cdot (2y+1)$: independent.
4. $\frac{6}{7}(x+y)^2$ cannot be written this way: not independent.
5. $2e^{-x-2y} = (e^{-x}) \cdot (2e^{-2y})$: independent.

Covariance, I

Our last topic is to discuss covariance and correlation. We define covariance the same way as in the discrete case:

Definition

If X and Y are random variables whose expected values exist and are μ_X and μ_Y respectively, then the covariance of X and Y is defined as

$$\mathrm{cov}(X, Y) = E[(X - \mu_X) \cdot (Y - \mu_Y)] = E(XY) - E(X)E(Y).$$

- In order to compute the covariance, we need to know how to compute the expected value of arbitrary functions of X and Y.
- One approach would be to try to write down the pdf of the random variable XY. This can be done, but it is quite an inefficient way to find $E(XY)$. (It's just like that long calculation two lectures ago to find the pdf of X^2.)

Covariance, II... Wait, This Isn't Covariance!

Instead, we extend the principle for finding the expected value of an arbitrary function of X to handle functions of X and Y:

Proposition (Expected Value of Functions of X and Y)

If X and Y are continuous random variables with joint distribution function $p_{X,Y}(x,y)$ and $g(X,Y)$ is any piecewise-continuous function, then $E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \cdot p_{X,Y}(x,y)\, dy\, dx$.

The proof of this proposition is a moderately technical calculation from analysis, so we will omit the details. (The basic idea is simply to show that the result holds for increasingly complex classes of functions g, starting with constant functions.)

Covariance, III... Can We Do Covariance Soon?

As a corollary, we get the additivity property of expected value:

Corollary (Additivity of Expected Value)

If X and Y are continuous random variables, then $E(X+Y) = E(X) + E(Y)$.

Proof: By the previous proposition applied three times, we have

$$\begin{aligned} E(X+Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)\cdot p_{X,Y}(x,y)\,dy\,dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\cdot p_{X,Y}(x,y)\,dy\,dx \\ &\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\cdot p_{X,Y}(x,y)\,dy\,dx \\ &= E(X) + E(Y). \end{aligned}$$

Covariance, IV: Yes, Actually Covariance Now!

Example: If X and Y have joint distribution given by $p_{X,Y}(x,y) = x^2 + 2y^2$ for $0 \leq x \leq 1$, $0 \leq y \leq 1$, find the covariance of X and Y.

Covariance, IV: Yes, Actually Covariance Now!

<u>Example</u>: If X and Y have joint distribution given by $p_{X,Y}(x,y) = x^2 + 2y^2$ for $0 \le x \le 1$, $0 \le y \le 1$, find the covariance of X and Y.

- We compute
  $E(X) = \int_0^1 \int_0^1 x(x^2 + 2y^2)\,dy\,dx = \int_0^1 (x^3 + \frac{2}{3}x)\,dx = 7/12$,
  $E(Y) = \int_0^1 \int_0^1 y(x^2 + 2y^2)\,dy\,dx = \int_0^1 (\frac{1}{2}x^2 + \frac{1}{2})\,dx = 2/3$,
  $E(XY) = \int_0^1 \int_0^1 xy(x^2 + 2y^2)\,dy\,dx = \int_0^1 (\frac{1}{2}x^3 + \frac{1}{2}x)\,dx = 3/8$.
- Thus, the covariance is $\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y)$
  $= (3/8) - (2/3)(7/12) = -1/72$.

Covariance, VI

All of the same properties of variance and covariance hold here:

Proposition (Properties of Variance and Covariance)

If X, Y, Z are continuous random variables whose expected values exist, then we have the following:

1. $\mathrm{cov}(X, X) = \mathrm{var}(X)$.
2. $\mathrm{cov}(Y, X) = \mathrm{cov}(X, Y)$.
3. $\mathrm{cov}(X + Y, Z) = \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$.
4. $\mathrm{cov}(aX + b, Y) = a \cdot \mathrm{cov}(X, Y)$ *for any real* a, b.
5. $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X, Y)$.

Furthermore, if X and Y are independent, then $E(XY) = E(X) \cdot E(Y)$ *and* $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

Proof: These all follow the same way as in the discrete case.

Correlation, I

Perhaps shockingly, we also take the same definition of correlation as in the discrete case:

Definition

If X and Y are discrete random variables whose variances exist and are nonzero, the (Pearson) correlation between X and Y is defined as $\mathrm{corr}(X, Y) = \dfrac{\mathrm{cov}(X, Y)}{\sigma(X)\sigma(Y)}$.

- As before, the correlation describes the strength to which the relationship between X and Y can be captured by a linear model.

Correlation, II

Suppose that the continuous random variables X and Y have joint distribution given by $p_{X,Y}(x, y) = x + y$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Find

1. The covariance of X and Y.
2. The correlation of X and Y.

Correlation, II

Suppose that the continuous random variables X and Y have joint distribution given by $p_{X,Y}(x, y) = x + y$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Find

1. The covariance of X and Y.
2. The correlation of X and Y.

- We just have to set up and evaluate the appropriate integrals to compute $E(X)$, $E(Y)$, $E(XY)$ (for the covariance), and then also $E(X^2)$ and $E(Y^2)$ (to compute $\sigma(X)$ and $\sigma(Y)$ for the correlation).

Correlation, III

Suppose that the continuous random variables X and Y have joint distribution given by $p_{X,Y}(x,y) = x + y$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Find

1. The covariance of X and Y.

- We have
  $E(X) = \int_0^1 \int_0^1 x(x+y)\,dy\,dx = \int_0^1 (x^2 + x/2)\,dx = 7/12$,
  $E(X) = \int_0^1 \int_0^1 y(x+y)\,dy\,dx = \int_0^1 (x/2 + 1/3)\,dx = 7/12$,
  and
  $E(XY) = \int_0^1 \int_0^1 xy(x+y)\,dy\,dx = \int_0^1 (x^2/2 + x/3)\,dx = 1/3$.
- Thus, the covariance is $\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y) = 1/3 - (7/12)(7/12) = -1/144$.

Correlation, IV

Suppose that the continuous random variables X and Y have joint distribution given by $p_{X,Y}(x,y) = x + y$ for $0 \le x \le 1$ and $0 \le y \le 1$. Find

2. The correlation of X and Y.

- We computed $\mathrm{cov}(X,Y) = -1/144$. For the correlation, we also need $\sigma(X)$ and $\sigma(Y)$. First, $E(X) = \int_0^1 \int_0^1 x^2(x+y)\,dy\,dx = \int_0^1 (x^3 + x^2/2)\,dx = 5/12$, so $\sigma(X) = \sqrt{E(X^2) - E(X)^2} = \sqrt{11}/12 \approx 0.2764$.
- Also, $E(Y) = \int_0^1 \int_0^1 y^2(x+y)\,dy\,dx = \int_0^1 (x/3 + 1/4)\,dx = 5/12$, so $\sigma(Y) = \sqrt{E(Y^2) - E(Y)^2} = \sqrt{11}/12 \approx 0.2764$.
- Finally, $\mathrm{corr}(X,Y) = \dfrac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)} = -1/11 \approx -0.091$.

Overview of §2.3: Modeling Applications

We now move into the third portion of this chapter, which (broadly speaking) is about using discrete and continuous random variables to model real-world phenomena.

- We will discuss three important classes of probability distributions: the Gaussian normal distributions, the Poisson distributions, and the exponential distributions.
- Each distribution arises in various practical applications involving phenomena with particular simple properties.
- The normal distribution is by far the most important of these three (so we start with it), but all of them serve as important models for various processes.
- The theme is to describe, in each case, a mathematical property of the distribution that serves as the reason for why it is a good model in appropriate situations.

Normal Distributions, I

We start with the normal distribution:

Definition

A random variable $N_{\mu,\sigma}$ is normally distributed with parameters μ and σ if its probability density function is
$$p_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

A particularly useful case is $\mu = 0$ and $\sigma = 1$:

Definition

The standard normal distribution is $N_{0,1}$, with $\mu = 0$ and $\sigma = 1$.

Normal Distributions, II

It is not trivial to verify that $p_{\mu,\sigma}(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$ actually *is* a probability density function.

- This requires showing $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\, dx = 1$.
- This is not so easy to do, because the integrand does not have an elementary antiderivative, so there is no nice formula for the indefinite integral. (But there are various ways.)
- Computing the expected value is a bit easier (it reduces to a substitution), and the variance can also be evaluated using some manipulations.
- The end result is $E(N_{\mu,\sigma}) = \mu$ and $\mathrm{var}(N_{\mu,\sigma}) = \sigma^2$ so that $\sigma(N_{\mu,\sigma}) = \sigma$.
- This justifies our use of the letters μ and σ for the parameters, because they are just the mean and standard deviation.

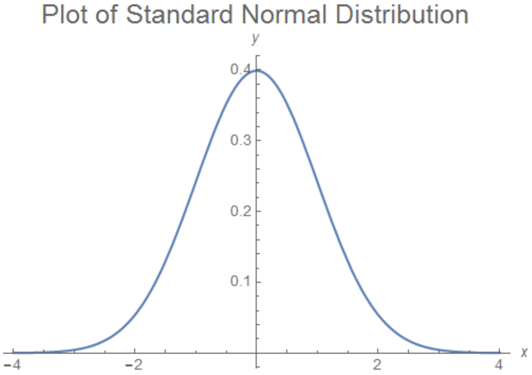
Normal Distributions, III

We will warn that there are many different notations in use for representing normal distributions.

- In particular, it is very common instead to call the parameters μ and σ^2 (for the mean and variance, rather than the two we used, the mean and standard deviation).
- We use somewhat atypical notation $N_{\mu,\sigma}$ for the name of the associated random variable to keep our notation separate.
- It is more common in other sources to write $\mathcal{N}(\mu, \sigma^2)$ as representing the normal distribution with mean μ and variance σ^2.

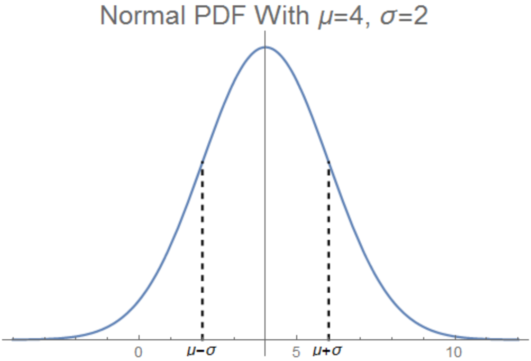
Normal Distributions, IV

The normal distribution is often called the "bell curve" due to the shape of its pdf:



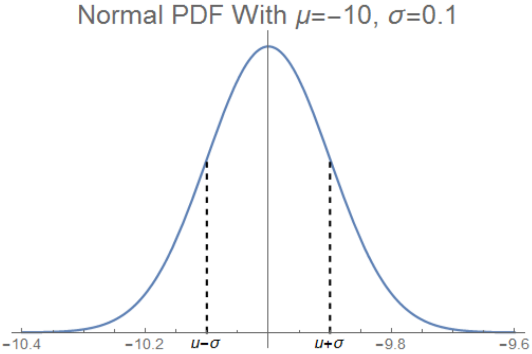
Normal Distributions, V

All normal distributions have the same shape: the mean parameter shifts the location of the peak, while the standard deviation parameter stretches the width.



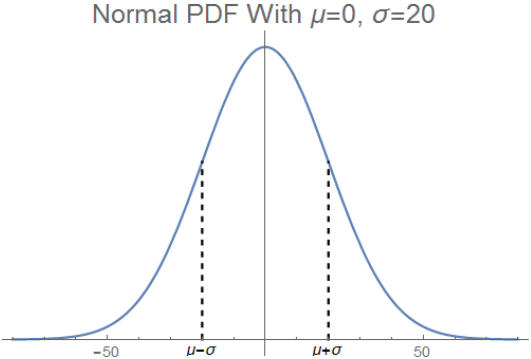
Normal Distributions, VI

All normal distributions have the same shape: the mean parameter shifts the location of the peak, while the standard deviation parameter stretches the width.



Normal Distributions, VII

All normal distributions have the same shape: the mean parameter shifts the location of the peak, while the standard deviation parameter stretches the width.



Normal Distributions, VIII

Example: Suppose $N_{11,2}$ is a normally-distributed random variable with expected value 11 and standard deviation 2. Find:

1. $P(N \leq 11)$.
2. $P(7 \leq N \leq 9)$.
3. $P(N \geq 13)$.

Normal Distributions, VIII

Example: Suppose $N_{11,2}$ is a normally-distributed random variable with expected value 11 and standard deviation 2. Find:

1. $P(N \leq 11)$.
2. $P(7 \leq N \leq 9)$.
3. $P(N \geq 13)$.

- One approach is simply to write down the probability density function and set up the integrals.
- This yields $P(N_{11,2} \leq 11) = \int_{-\infty}^{11} \frac{1}{2\sqrt{2\pi}} e^{-(x-11)^2/8}\, dx$, with $P(7 \leq N_{11,2} \leq 9) = \int_{7}^{9} \frac{1}{2\sqrt{2\pi}} e^{-(x-11)^2/8}\, dx$, and $P(N_{11,2} \geq 13) = \int_{13}^{\infty} \frac{1}{2\sqrt{2\pi}} e^{-(x-11)^2/8}\, dx$.
- Except... now what?

Normal Distributions, IX

Here's the problem: these integrals are essentially impossible to evaluate exactly, because $p_{\mu,\sigma}(x) = e^{-(x-\mu)^2/(2\sigma^2)}$ does not have an elementary antiderivative!

- So in essentially all cases, we must use numerical integration procedures to approximate these integrals.
- On the TI-83 or TI-84, you can do these calculations using the `normalcdf` function, located in the DISTR menu (2nd key $>$ VARS), which has numerous statistical distributions.
- Essentially all computer algebra systems also have an equivalent function built in: MATLAB calls it "normcdf", while Mathematica uses "NormalDistribution" and separate operators for requesting the PDF and CDF.

Normal Distributions, X

<u>Example</u>: Suppose $N_{11,2}$ is a normally-distributed random variable with expected value 11 and standard deviation 2. Find:

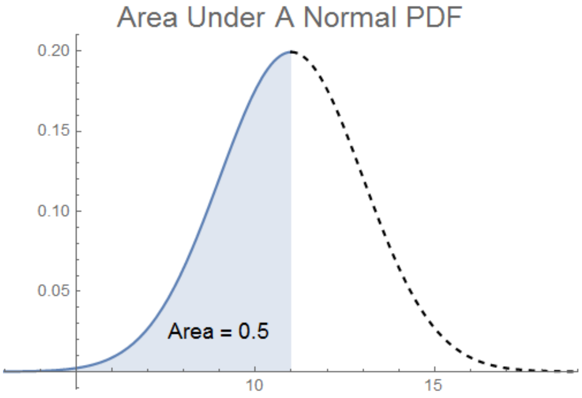
1. $P(N \leq 11)$.
2. $P(7 \leq N \leq 9)$.
3. $P(N \geq 13)$.

- Using whatever form of technology is available, we can evaluate these probabilities, which to four decimal places are as follows:

1. $P(N_{11,2} \leq 11) = 0.5$.
2. $P(7 \leq N_{11,2} \leq 9) \approx 0.1359$.
3. $P(N_{11,2} \geq 13) \approx 0.1587$.

Normal Distributions, XI

We can visualize these calculations as areas (the one given below represents $P(N \leq 11) = 0.5$):



In this case, the area is $1/2$ by symmetry.

Normal Distributions, XII

Another approach to solving problems involving the normal distribution is use a table of computed values for the cumulative distribution function of the standard normal distribution $N_{0,1}$, along with a substitution.

- If we define $z_a = \dfrac{a-\mu}{\sigma}$ and $z_b = \dfrac{b-\mu}{\sigma}$, then it is a fairly straightforward calculation to see that $P(a \le N_{\mu,\sigma} \le b) = P(z_a \le N_{0,1} \le z_b)$.
- This "z-score" simply measures the number of standard deviations away from the mean the associated value a is.
- Because all normal distributions all have the same shape, the area under the curve between two values $x = a$ and $x = b$ will only depend on their position relative to the mean (measured in standard deviations).

Normal Distributions, XIII

Thus, by normalizing to set the mean equal to 0 and the standard deviation equal to 1, we can convert any of these questions to one about the standard normal distribution.

- Once we make this change of variables by computing the z-scores, we can use a table of computed values for the cumulative distribution function of the standard normal $N_{0,1}$ (such as the table given on the next slide) to find the desired probabilities.
- To read the table, add together the row header and the column header to get the value of z. The value in the cell then gives the probability $P(N_{0,1} \leq z)$ to four decimal places.
- Example: In the entry whose column is labeled $+0.1$ and the row labeled -2.0 is given the value $P(N_{0,1} \leq -1.9)$.

Normal Distributions, XIV: Yes, I Made This By Hand

$z\ :\ P(N_{0,1} \leq z)$	$+0.0$	$+0.1$	$+0.2$	$+0.3$	$+0.4$
-3.0	0.0013	0.0019	0.0026	0.0035	0.0047
-2.5	0.0062	0.0082	0.0107	0.0139	0.0179
-2.0	0.0228	0.0287	0.0359	0.0446	0.0548
-1.5	0.0668	0.0808	0.0968	0.1151	0.1357
-1.0	0.1587	0.1841	0.2119	0.2420	0.2743
-0.5	0.3085	0.3446	0.3821	0.4207	0.4602
$+0.0$	0.5	0.5398	0.5793	0.6179	0.6554
$+0.5$	0.6915	0.7257	0.7580	0.7881	0.8159
$+1.0$	0.8413	0.8643	0.8849	0.9032	0.9192
$+1.5$	0.9332	0.9452	0.9554	0.9641	0.9713
$+2.0$	0.9772	0.9821	0.9861	0.9893	0.9918
$+2.5$	0.9938	0.9953	0.9965	0.9974	0.9981
$+3.0$	0.9987	0.9990	0.9993	0.9995	0.9997

Normal Distributions, XV

Historically, most statistical calculations required using a precomputed table like the one on the previous slide.

- Reflect, for a moment, on the amount of work it would have taken to produce all of the data appearing in the previous table in the 19th century.
- We mention this "z-scores" approach not because it is still important to be able to look these things up in a table (it isn't!), but because it is helpful in getting a feeling for how the normal distribution behaves.
- Later, when we are working with applications of the normal distribution in statistics, we will frequently pass back and forth between doing calculations with a normal distribution with mean μ and standard deviation σ and the standard normal. It is important to become comfortable with this procedure now.

Normal Distributions, XVI

Example (again): Suppose $N_{11,2}$ is a normally-distributed random variable with expected value 11 and standard deviation 2. Find:

1. $P(N \leq 11)$.
2. $P(7 \leq N \leq 9)$.
3. $P(N \geq 13)$.

- Instead of using a calculator, solve these by computing z-scores and looking up the results in the table.
- Remember that the entries in the table are values of the standard normal cdf. So you will have to do subtraction to find some of the probabilities above.
- [Stage directions: flip back to the table. Don't say this aloud.]

Normal Distributions, XVII

Example (again): Suppose $N_{11,2}$ is a normally-distributed random variable with expected value 11 and standard deviation 2. Find:

1. $P(N \leq 11)$.
2. $P(7 \leq N \leq 9)$.
3. $P(N \geq 13)$.

- We have $\mu = 11$ and $\sigma = 2$, so $P(N_{11,2} \leq 11) = P(N_{0,1} \leq 0) = \boxed{0.5}$.
- Likewise, $P(7 \leq N_{11,2} \leq 9) = P(-2 \leq N_{0,1} \leq -1) = P(N_{0,1} \leq -1) - P(N_{0,1} \leq -2) = 0.1587 - 0.0228 \approx \boxed{0.1359}$.
- Finally, $P(N_{11,2} \geq 13) = P(N_{0,1} \geq 1) = 1 - P(N_{0,1} \leq 1) = 1 - 0.8413 = \boxed{0.1587}$.

Normal Distributions, XVIII

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the percentage of scores that will be in each range:

1. Between 450 and 550.
2. Less than 600.
3. Greater than 700.

Normal Distributions, XVIII

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the percentage of scores that will be in each range:

1. Between 450 and 550.
2. Less than 600.
3. Greater than 700.

- The scores follow the normal distribution $N_{500,100}$.
- Using the z-score method, we can compute $P(450 < N_{500,100} < 550) = P(-0.5 < N_{0,1} < 0.5) = 0.6915 - 0.3085 \approx 0.3830$.
- Next, we have $P(N_{500,100} < 600) = P(N_{0,1} \leq 1) \approx 0.8413$.
- Lastly, $P(N_{500,100} > 700) = P(N_{0,1} > 2) = 1 - P(N_{0,1} \leq 2) \approx 0.0228$.

Normal Distributions, XIX

In some situations we want to invert our analysis by starting with a probability and finding the corresponding value or range in the distribution.

- Analytically, this corresponds to evaluating the inverse function of the normal cumulative density function (which is usually called the inverse normal cdf for short), which can be done efficiently using a calculator or computer.
- Alternatively, we could look up the needed probabilities in a table to find the associated z-scores.

Normal Distributions, XX (Dos Equis)

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the lowest integer score that is

1. above 80% of the other scores.
2. above 90% of the other scores.
3. above 99% of the other scores.
4. above 99.9% of the other scores.

Normal Distributions, XX (Dos Equis)

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the lowest integer score that is

1. above 80% of the other scores.
2. above 90% of the other scores.
3. above 99% of the other scores.
4. above 99.9% of the other scores.

- We can either use a computer to evaluate the inverse normal cdf, or we can use a reverse-lookup in a table of z-scores. To estimate a value that is between two scores, one can use a linear interpolation, or use a more detailed table, or (the best) simply use a table of the inverse normal cdf directly.

Normal Distributions, XXI

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the lowest integer score that is

1. above 80% of the other scores.

Normal Distributions, XXI

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the lowest integer score that is

1. above 80% of the other scores.

- Using a computer, we can find that $P(N_{500,100} \leq 584) \approx 0.7995$ and $P(N_{500,100} \leq 585) = 0.8023$, so the minimum integer score above 80% of the other scores is 585.
- Alternatively, using a table z-scores (equivalently, the inverse normal cdf for the standard normal), we could find that the value z with $P(N_{0,1} \leq z) = 0.80$ is $z \approx 0.8416$, and so the desired score is $500 + 100z \approx 584.16$.

Normal Distributions, XXII

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the lowest integer score that is

2. above 90% of the other scores.

Normal Distributions, XXII

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the lowest integer score that is

2. above 90% of the other scores.

- Using a computer, we can find that $P(N_{500,100} \leq 628) \approx 0.7995$ and $P(N_{500,100} \leq 629) = 0.9015$, so the minimum integer score above 90% of the other scores is 629.
- Alternatively, using a table of z-scores (equivalently, the inverse normal cdf for the standard normal), we could find that the value z with $P(N_{0,1} \leq z) = 0.90$ is $z \approx 1.2816$, and so the desired score is $500 + 100z \approx 628.16$.

Normal Distributions, XXIII

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the lowest integer score that is

3. above 99% of the other scores.

Normal Distributions, XXIII

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the lowest integer score that is

3. above 99% of the other scores.

- Using a computer, we can find that $P(N_{500,100} \leq 732) \approx 0.9898$ and $P(N_{500,100} \leq 733) = 0.9901$, so the minimum integer score above 99% of the other scores is 733.
- Alternatively, using a table of z-scores (equivalently, the inverse normal cdf for the standard normal), we could find that the value z with $P(N_{0,1} \leq z) = 0.99$ is $z \approx 2.3263$, and so the desired score is $500 + 100z \approx 732.63$.

Normal Distributions, XXIV

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the lowest integer score that is

4. above 99.9% of the other scores.

Normal Distributions, XXIV

Example: A certain standardized test is designed so that its score distribution will be normal with a mean of 500 and a standard deviation of 100. Determine the lowest integer score that is

4. above 99.9% of the other scores.

- Using a computer, we can find that $P(N_{500,100} \leq 809) \approx 0.9989992$ and $P(N_{500,100} \leq 810) = 0.9903240$, so the minimum integer score above 99.9% of the other scores is 810. (Though 809 sure is close!)
- Alternatively, using a table of z-scores (equivalently, the inverse normal cdf for the standard normal), we could find that the value z with $P(N_{0,1} \leq z) = 0.99$ is $z \approx 3.0902$, and so the desired score is $500 + 100z \approx 809.02$.

Normal Distributions, XXV

We mention a few miscellaneous things about the standard normal distribution $N_{0,1}$.

First, the distribution is symmetric about its mean.

- As a consequence, this means $P(N_{0,1} \leq z) = P(N_{0,1} \geq -z) = 1 - P(N_{0,1} \leq -z)$.
- Example: We have $P(N_{0,1} \leq 1.2) = 0.8849$ while $P(N_{0,1} \leq -1.2) = 0.1151$.
- Because of this identity, most tables of z-values only quote results for one half of possible z (either $z > 0$ or $z < 0$), since one may easily calculate the other half of the values as above.

Second, the points that are one standard deviation away from the mean are inflection points of the pdf.

Normal Distributions, XXVI

Third, for doing rough estimates it is useful to remember the approximate proportions of the normal distribution that are within 1, 2, or 3 standard deviations of the mean.

- Observe that $P(-z \leq N_{0,1} \leq z) = 1 - 2 \cdot P(N_{0,1} \leq z)$ (there are two "tails" outside the range $-z \leq N_{0,1} \leq z$).
- So one may calculate $P(-1 \leq N_{0,1} \leq 1) = 0.6827$, $P(-2 \leq N_{0,1} \leq 2) = 0.9545$, $P(-3 \leq N_{0,1} \leq 3) = 0.9973$.
- This is often summarized as the "68-95.5-99.7" rule: these are the percentages of a normal distribution that will fall within 1, 2, and 3 standard deviations of the mean respectively.

Normal Distributions, XXVII

- In particular, compare the "68-95.5-99.7" rule to the worst-case scenario predictions from Chebyshev's inequality (which says that the proportion falling within k standard deviations of the mean must be at least $1 - 1/k^2$): it would give the "0-75-88.8" rule, which is far lower than what occurs with the normal distribution.
- This rule indicates that normally distributed quantities will (the vast majority of the time) be within 3 or fewer standard deviations of the mean.
- In particular, normal models are not generally as appropriate for situations where even 1-2 percent of the distribution represents outliers that are more than 3 standard deviations away from the mean. (Such situations are, of course, still quite possible for arbitrary random variables, per Chebyshev.)

Normal Distributions, XXVIII

The normal distribution arises often as a model for real-world quantities that arise as sums or averages of many small pieces.

Examples:

- Heights of people (typically after sorting by gender)
- Measurement errors in experiments
- Scores on exams or standardized tests
- Diffusion of a particle in a solution (or more generally the position after a random walk)
- Blood pressure readings
- Variation in part sizes made by industrial fabrication

As a matter of practice, most of these quantities are only approximately normally distributed.

Central Limit Theorem, I

The reason for the common appearance of the normal distribution is the following fundamental result:

Theorem (Central Limit Theorem)

Let X_1, X_2, ... , X_n be a sequence of independent, identically-distributed discrete or continuous random variables each with finite expected value μ and standard deviation $\sigma > 0$. Then the distribution of the random variable
$Y_n = \dfrac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$ *will approach the standard normal distribution (of mean 0 and standard deviation 1) as n tends to ∞: explicitly, for any real numbers $a \leq b$ we have*
$P(a \leq Y_n \leq b) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$ *as $n \to \infty$.*

Central Limit Theorem, II

The statement is quite detailed (it took up the entire slide!), so let's unpack it a bit:

- First, if $X_1, \ldots, X_n$ are independent and identically-distributed, we may think of the random variable $X_1 + \cdots + X_n$ as being the sum of the results of independently sampling a random variable X a total of n times.
- One example of this would be flipping a (fair or unfair) coin n times and summing the total number of heads (X is the Bernoulli random variable identifying success=1 or failure=0).
- Another example would be rolling a die n times and summing the outcomes (X is the result of the die roll).

Central Limit Theorem, III

So suppose we have a random variable X, and we sample it (independently) n times, to get values $X_1, X_2, \ldots, X_n$.

- We would like to say something about the sum $X_1 + X_2 + \cdots + X_n$.
- From the linearity of expected value, we can see that $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = n\mu$.
- Also, since $X_1, \ldots, X_n$ are independent, we also have $\mathrm{var}(X_1 + X_2 + \cdots + X_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n) = n\sigma^2$, and so $\sigma(X_1 + \cdots + X_n) = \sqrt{n\sigma^2} = \sigma\sqrt{n}$.

Central Limit Theorem, IV

We just showed that the expected value of $X_1 + \cdots + X_n$ is $n\mu$ and that its standard deviation is $\sigma\sqrt{n}$.

- Therefore, if $Y_n = \dfrac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$, we can see (via expected value properties) that $E(Y_n) = 0$ and $\sigma(Y_n) = 1$.
- We can think of Y_n as a "normalization" of the summed distribution $X_1 + \cdots + X_n$ by translating and rescaling it so that its expected value is 0 and its standard deviation is 1.

Central Limit Theorem, IV

We just showed that the expected value of $X_1 + \cdots + X_n$ is $n\mu$ and that its standard deviation is $\sigma\sqrt{n}$.

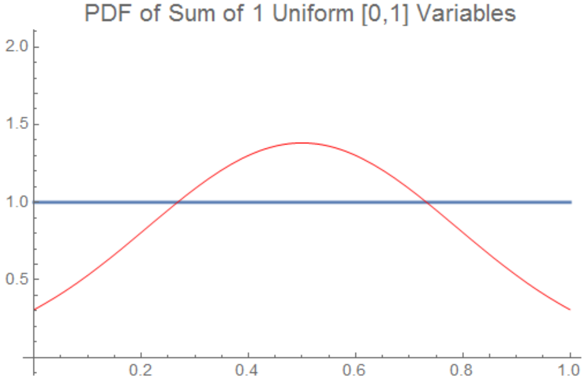
- Therefore, if $Y_n = \dfrac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$, we can see (via expected value properties) that $E(Y_n) = 0$ and $\sigma(Y_n) = 1$.
- We can think of Y_n as a "normalization" of the summed distribution $X_1 + \cdots + X_n$ by translating and rescaling it so that its expected value is 0 and its standard deviation is 1.

The content of the central limit theorem is now that, if we take a sample size n tending to ∞, this normalized average Y_n approaches the standard normal distribution.

- This result is quite powerful, since it says this limiting distribution is the same no matter what the original discrete or continuous random variable X was! (As long as its mean and standard deviation are defined....)

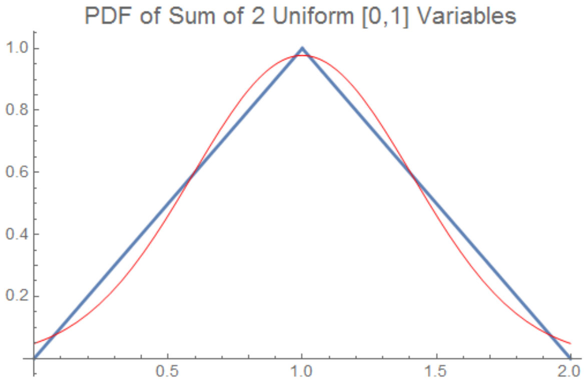
Central Limit Theorem, V

We can illustrate the central limit theorem quite convincingly for X the uniform distribution on $[0, 1]$:



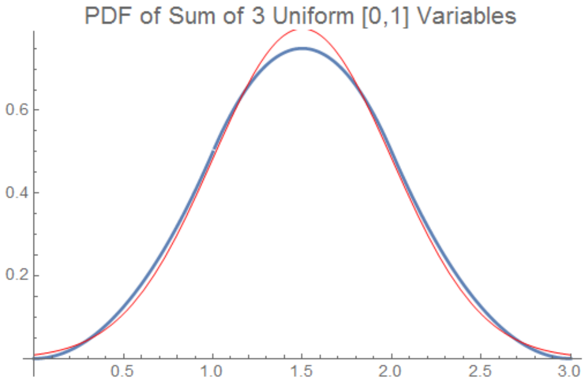
Central Limit Theorem, VI

We can illustrate the central limit theorem quite convincingly for X the uniform distribution on $[0, 1]$ (blue $=$ sum of uniform variables, red $=$ normal distribution):



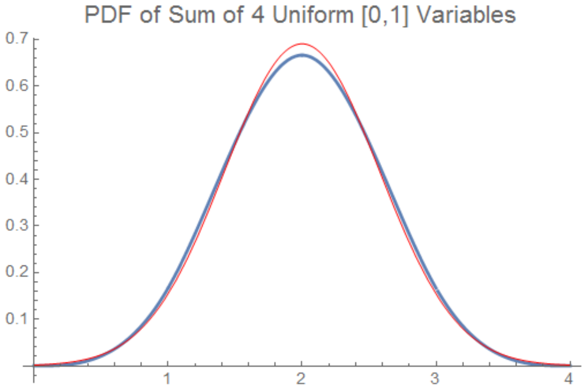
Central Limit Theorem, VII

We can illustrate the central limit theorem quite convincingly for X the uniform distribution on $[0, 1]$ (blue = sum of uniform variables, red = normal distribution):



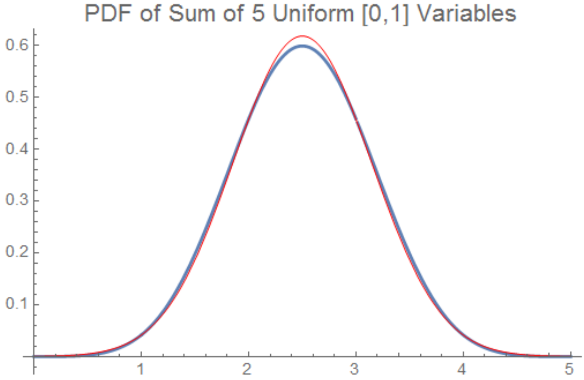
Central Limit Theorem, VIII

We can illustrate the central limit theorem quite convincingly for X the uniform distribution on $[0, 1]$ (blue $=$ sum of uniform variables, red $=$ normal distribution):



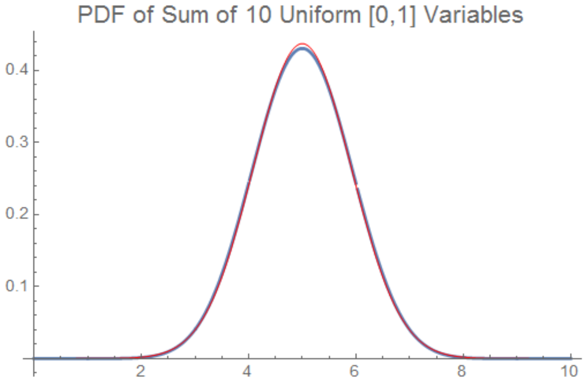
Central Limit Theorem, IX

We can illustrate the central limit theorem quite convincingly for X the uniform distribution on $[0, 1]$ (blue $=$ sum of uniform variables, red $=$ normal distribution):



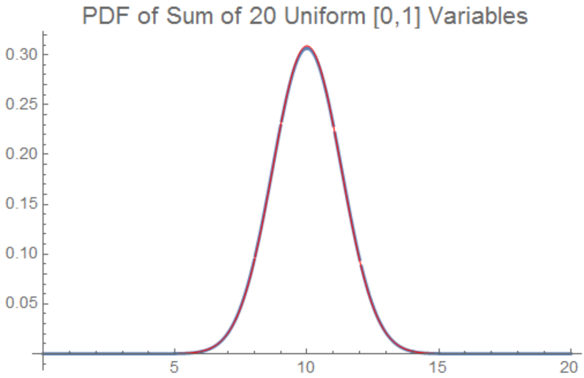
Central Limit Theorem, X

We can illustrate the central limit theorem quite convincingly for X the uniform distribution on $[0, 1]$ (blue = sum of uniform variables, red = normal distribution):



Central Limit Theorem, XI

We can illustrate the central limit theorem quite convincingly for X the uniform distribution on $[0, 1]$ (blue = sum of uniform variables, red = normal distribution):



Summary

We discussed independence, covariance, and correlation for continuous random variables.

We introduced the normal distribution.

We introduced the central limit theorem and gave some analysis of its result.

Next lecture: The central limit theorem and applications