Math/Stat 341: Probability: Fall ‘21 (Williams)

Professor Steven J Miller: sjm1@williams.edu

Homepage: https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21

Lecture 1: 10-17-21: https://youtu.be/iltn3Oks-9k

Lecture 17: 10/18/19: Linearity of expectation, variances and covariances, power of linearity of expectation, bernoulli and binomial, convolution: https://youtu.be/WdITkk5zac0

Plan for the day: Lecture 17: October 25, 2021:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Linearity of expectation
- variances and covariances
- power of linearity of expectation
- bernoulli and binomial
- convolution

General items.

- **Path through the algebra….**

Theorem 9.5.1 (Linearity of Expectation) *Let $X_1, \ldots, X_n$ be random variables, let $g_1, \ldots, g_n$ be functions such that $\mathbb{E}[|g_i(X_i)|]$ exists and is finite, and let $a_1, \ldots, a_n$ be any real numbers. Then*

$$\mathbb{E}[a_1 g_1(X_1) + \cdots + a_n g_n(X_n)] \;=\; a_1 \mathbb{E}[g_1(X_1)] + \cdots + a_n \mathbb{E}[g_n(X_n)].$$

Note the random variables are not *assumed to be independent. Also, if $g_i(X_i) = c$ (where c is a fixed number) then $\mathbb{E}[g_i(X_i)] = c$.*

Lemma 9.5.2 *Let X be a random variable with mean μ_X and variance σ_X^2. If a and b are any fixed constants, then for the random variable $Y = aX + b$ we have*

$$\mu_Y \;=\; a\mu_X + b \quad \text{and} \quad \sigma_Y^2 \;=\; a^2 \sigma_X^2.$$

Lemma 9.5.3 *Let X be a random variable. Then*

$$\mathrm{Var}(X) \;=\; \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

$\mu = \int_{-\infty}^{\infty} x f_X(x)\,dx$

$\sigma^2 = E[X^2] - E[X]^2$

or $E[(X-\mu)^2]$

where

$\mu = E[X]$

Theorem 9.6.1 *If X and Y are independent random variables, then*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

A particularly important case is

$$\mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[X - \mu_X]\mathbb{E}[Y - \mu_Y] = 0.$$

Prove: X, Y indep so are $\tilde{X} = X - \mu_X$ and $\tilde{Y} = Y - \mu_Y$

Test when X and Y not independent….

$X = X$, $Y = X$ not indep

$\mathbb{E}[X^2] = \mathbb{E}[X]\mathbb{E}[X]$ take $X \sim Unif(-1,1)$ Then $\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] > 0$

Proof: key idea $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ as indep

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\,dx\,dy$$

$$= \underbrace{\int_{x=-\infty}^{\infty} x\, f_X(x)\,dx}_{\mathbb{E}[X]} \; \underbrace{\int_{y=-\infty}^{\infty} y\, f_Y(y)\,dy}_{\mathbb{E}[Y]} \qquad \blacksquare$$

$\mathbb{E}$[prod of indep RV] = prod of the expected values

Theorem 9.6.2 (Means and Variances of Sums of Random Variables) *Let $X_1, \ldots, X_n$ be random variables with means $\mu_{X_1}, \ldots, \mu_{X_n}$ and variances $\sigma^2_{X_1}, \ldots, \sigma^2_{X_n}$. If $X = X_1 + \cdots + X_n$, then*

$$\mu_X = \mu_{X_1} + \cdots + \mu_{X_n}.$$

If the random variables are independent, then we also have

$$\sigma^2_X = \sigma^2_{X_1} + \cdots + \sigma^2_{X_n} \quad \text{or} \quad \mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

In the special case when the random variables are independent and identically distributed (so all the means equal μ and all the variances equal σ^2), then

$$\mu_X = n\mu \quad \text{and} \quad \sigma^2_X = n\sigma^2.$$

Is this reasonable?

Rescale….

Look at general linear combination….

$$\mathrm{Var}(a_1 X_1 + \cdots + a_n X_n)$$

$$= a_1^2\,\mathrm{Var}(X_1) + \cdots + a_n^2\,\mathrm{Var}(X_n)$$

(Just let $\tilde{X}_k = a_k X_k$)

Standard Deviation: $\sigma_X = \sqrt{n}\,\sigma$

$$Var(X_1 + \cdots + X_n) = E\left[\left((X_1 + \cdots + X_n) - (\mu_{X_1} + \cdots + \mu_{X_n})\right)^2\right]$$

$$\sout{= E\left[X_1^2 + \cdots + X_n^2 + X\text{-cross terms} + X\text{'s times } \mu\text{'s} + \cdots\right]}$$

$$= E\left[(X_1 - \mu_{X_1})^2 + \cdots + (X_n - \mu_{X_n})^2 + \sum_{i \neq j} X_i X_j + \sum_{i \neq j} \mu_{X_i}\mu_{X_j} - 2\sum_{i \neq j} X_i \mu_{X_j}\right]$$

$$= E\left[(X_1 - \mu_{X_1})^2\right] + \cdots + E\left[(X_n - \mu_{X_n})^2\right] + \sum_{i \neq j} E\left[X_i X_j - X_i \mu_{X_j} + X_j \mu_{X_i} + \mu_{X_i}\mu_{X_j}\right]$$

$$= Var(X_1) + \cdots + Var(X_n) + \sum_{i \neq j} E\left[(X_i - \mu_{X_i})(X_j - \mu_{X_j})\right]$$

(COVARIANCE, it's zero if indep)

Covariance. Let X and Y be random variables. The covariance of X and Y, denoted by σ_{XY} or $\mathrm{Cov}(X, Y)$, is

$$\sigma_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)].$$

Note $\mathrm{Cov}(X, X)$ equals the variance of X. Also, if $X_1, \ldots, X_n$ are random variables and $X = X_1 + \cdots + X_n$, then

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j).$$

$$\sum_{i \neq j} \underbrace{\mathrm{Cov}(X_i, X_j)}_{\sigma_{X_i X_j}}$$

The Method of the Cumulative Distribution Function. Let X be a random variable with density f_X whose density is non-zero on some interval I, and let $Y = g(X)$ where $g : I \to \mathbb{R}$ is a differentiable function with inverse h. Assume the derivative of g is either always positive or always negative in I, except at finitely many points where it may vanish. To find the density f_Y:

1. Identify the interval I where the random variable X is defined.

2. Prove the function g has a derivative that is always positive or always negative (except, of course, at potentially finitely many points).

3. Determine the inverse function $h(y)$, where $g(h(y)) = y$ and $h(g(x)) = x$.

4. Determine $h'(y)$, either by directly differentiating h or using the relation $h'(y) = 1/g'(h(y))$.

5. The density of Y is $f_Y(y) = f_X(h(y))|h'(y)|$.

Definition 10.1.1 *The **convolution** of* independent *continuous random variables X and Y on $\mathbb{R}$ with densities f_X and f_Y is denoted $f_X * f_Y$, and is given by*

$$(f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(t) f_Y(z-t) dt.$$

If X and Y are discrete, we have

$$(f_X * f_Y)(z) = \sum_n f_X(x_n) f_Y(z - x_n);$$

note of course that $f_Y(z - x_n)$ is zero unless $z - x_n$ is one of the values where Y has positive probability (i.e., one of the special points y_m).

The convolution of two random variables has many wonderful properties, including the following theorem.

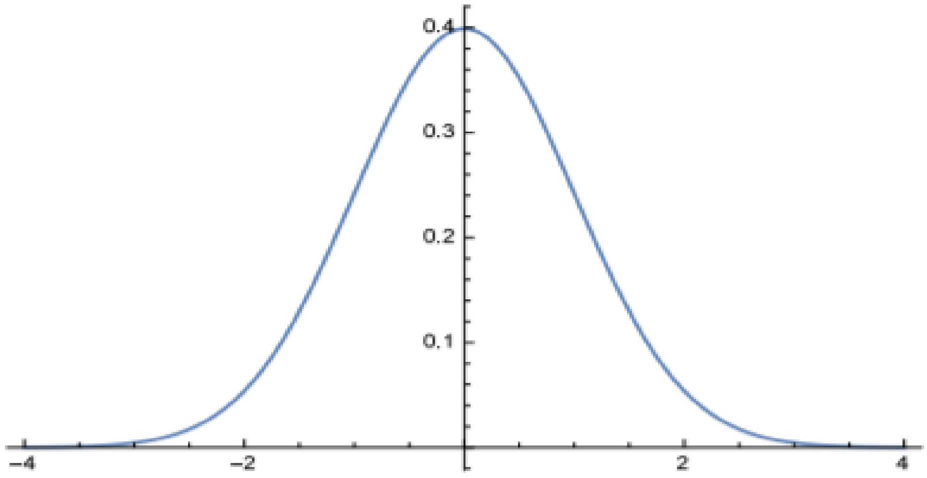
Theorem 10.1.2 *Let X and Y be continuous or discrete* independent *random variables on $\mathbb{R}$ with densities f_X and f_Y. If $Z = X + Y$, then*

$$f_Z(z) = (f_X * f_Y)(z).$$

Further, convolution is commutative: $f_X * f_Y = f_Y * f_X$.

Central Limit Theorem

$$\text{Normal } N(\mu, \sigma^2): \; p(x) = e^{-(x-\mu)^2/2\sigma^2} / \sqrt{2\pi\sigma^2}.$$



Theorem

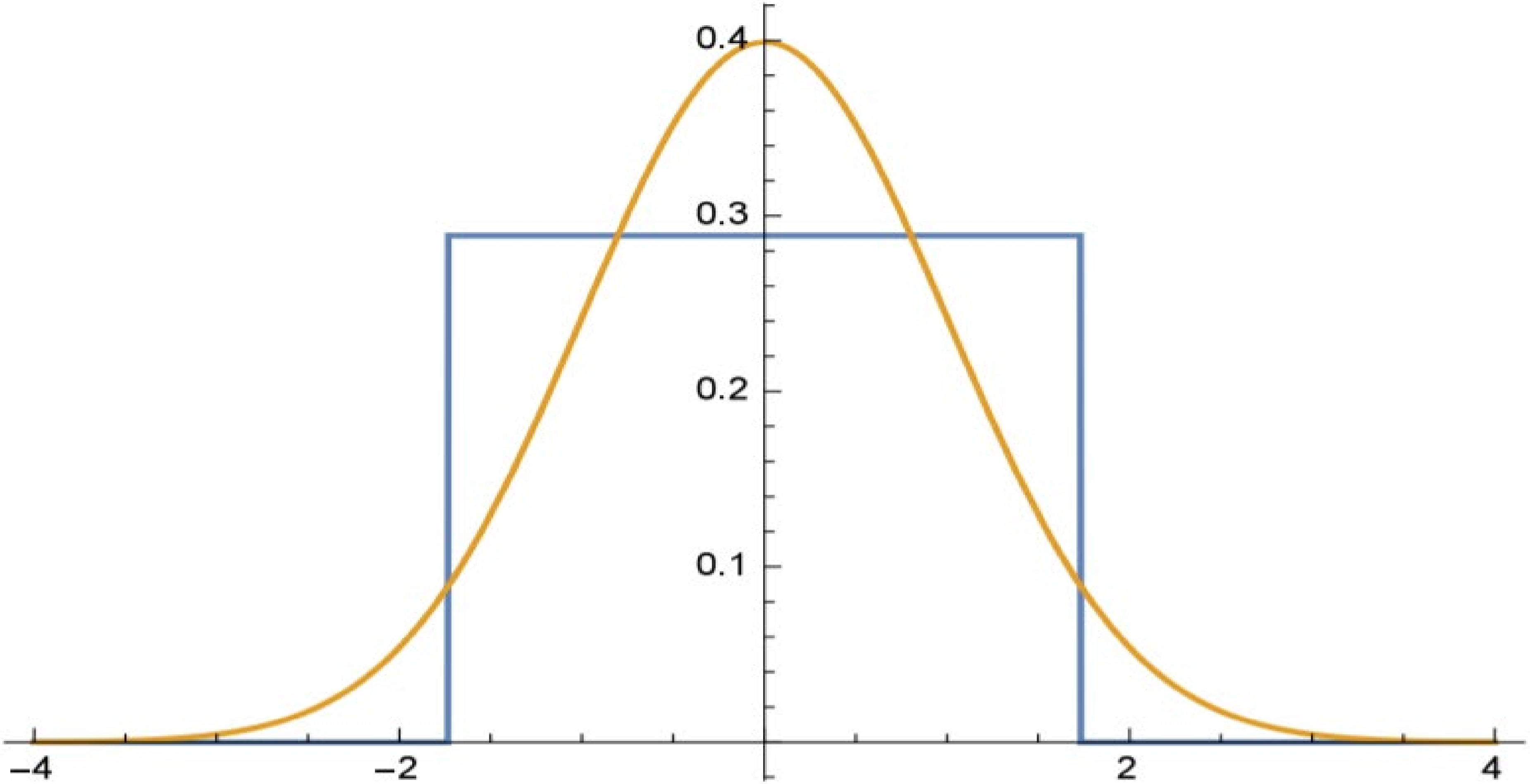
If $X_1, X_2, \ldots$ independent, identically distributed random variables (mean μ, variance σ^2, finite moments) then

mean 0 variance 1

$$S_N := \frac{X_1 + \cdots + X_N - N\mu}{\sigma\sqrt{N}} \text{ converges to } N(0,1).$$

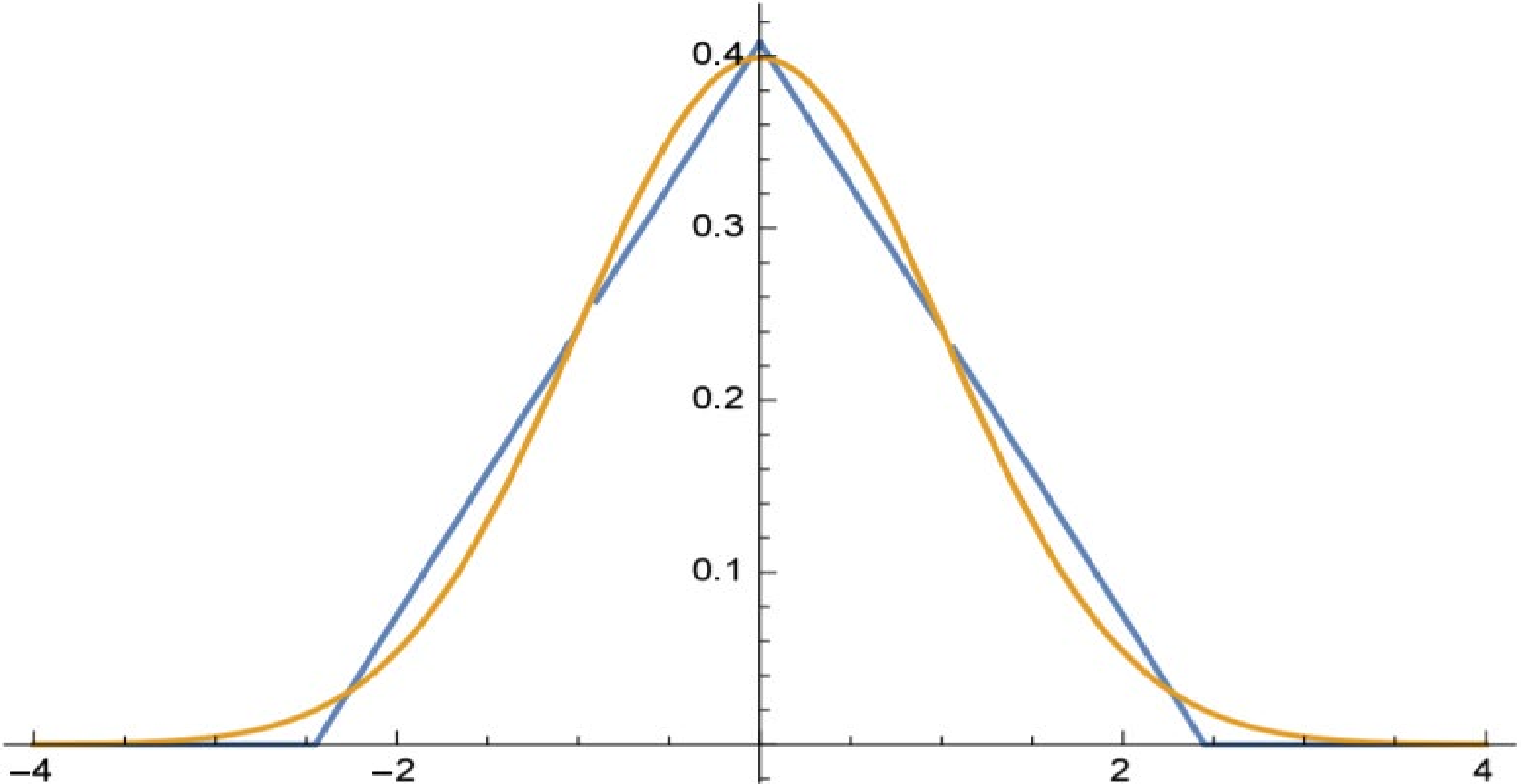
Central Limit Theorem: Sums of Uniform Random Variables
$X_i \sim \mathrm{Unif}(-1/2, 1/2)$

$Y_1 = X_1/\sigma_{X_1}$ vs $N(0,1)$.



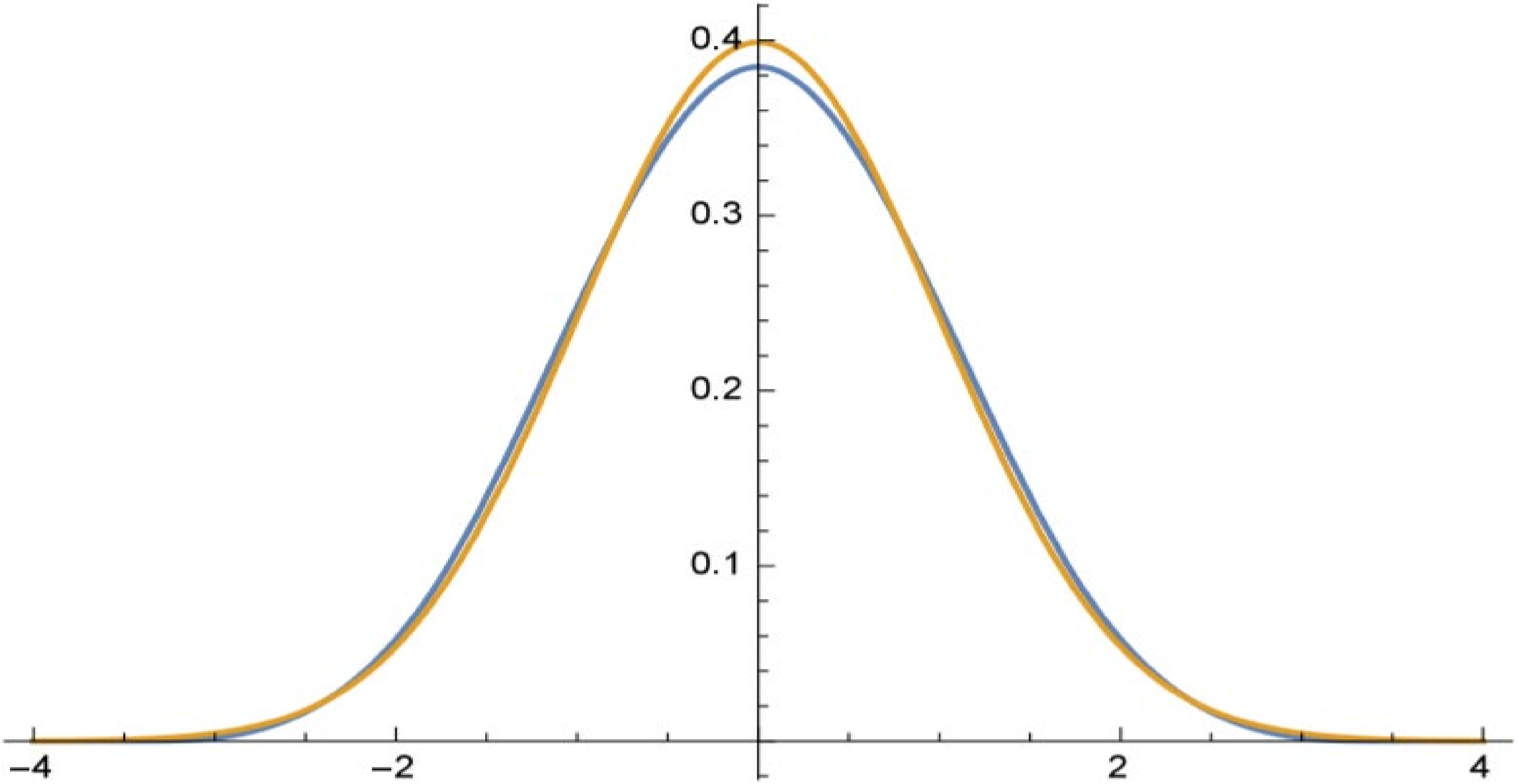
Central Limit Theorem: Sums of Uniform Random Variables
$X_i \sim \mathrm{Unif}(-1/2, 1/2)$

$$Y_2 = (X_1 + X_2)/\sigma_{X_1+X_2} \text{ vs } N(0,1).$$



Central Limit Theorem: Sums of Uniform Random Variables
$X_i \sim \mathrm{Unif}(-1/2, 1/2)$

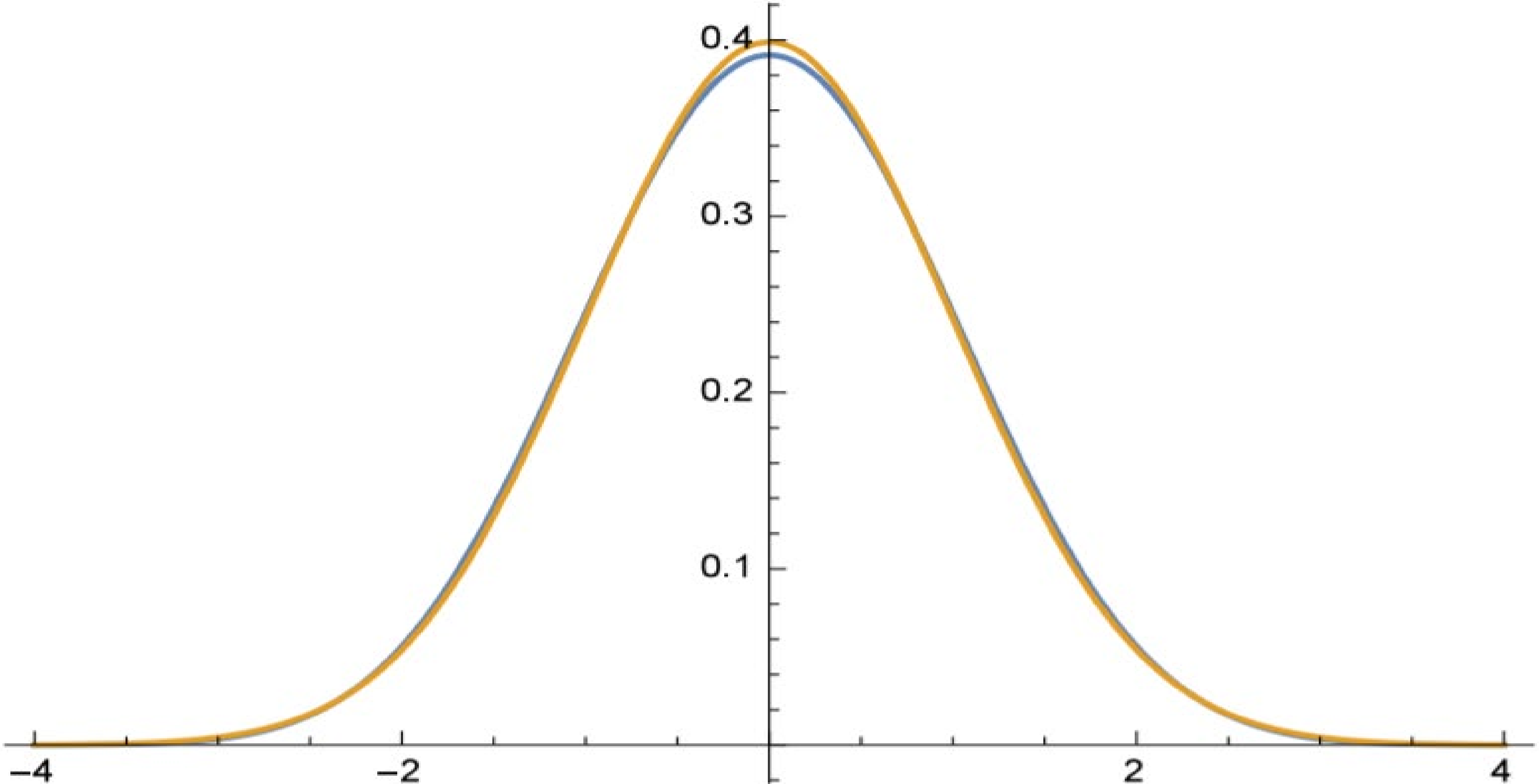
$$Y_4 = ((X_1 + X_2) + (X_3 + X_4))/\sigma_{X_1+X_2+X_3+X_4} \text{ vs } N(0,1).$$



Central Limit Theorem: Sums of Uniform Random Variables

$X_i \sim \mathrm{Unif}(-1/2, 1/2)$

$$Y_8 = (X_1 + \cdots + X_8)/\sigma_{X_1+\cdots+X_8} \text{ vs } N(0,1).$$



Central Limit Theorem: Sums of Uniform Random Variables
$X_i \sim \mathrm{Unif}(-1/2, 1/2)$

Density of $Y_4 = (X_1 + \cdots + X_4)/\sigma_{X_1+\cdots+X_4}$.

$$\frac{\begin{cases} \frac{1}{27}\left(18 + 9\sqrt{3}\,y - \sqrt{3}\,y^3\right) & y == 0 \\ \frac{1}{18}\left(12 - 6\,y^2 - \sqrt{3}\,y^3\right) & -\sqrt{3} < y < 0 \\ \frac{1}{54}\left(72 - 36\sqrt{3}\,y + 18\,y^2 - \sqrt{3}\,y^3\right) & \sqrt{3} < y < 2\sqrt{3} \\ \frac{1}{54}\left(18\sqrt{3}\,y - 18\,y^2 + \sqrt{3}\,y^3\right) & y == \sqrt{3} \\ \frac{1}{18}\left(12 - 6\,y^2 + \sqrt{3}\,y^3\right) & 0 < y < \sqrt{3} \\ \frac{1}{54}\left(72 + 36\sqrt{3}\,y + 18\,y^2 + \sqrt{3}\,y^3\right) & -2\sqrt{3} < y \le -\sqrt{3} \\ 0 & \mathrm{True} \end{cases}}{\sqrt{3}}$$

(Don't even think of asking to see Y_8's!)

The Bernoulli Distribution: X has a **Bernoulli distribution** with parameter $p \in [0,1]$ if $\text{Prob}(X = 1) = p$ and $\text{Prob}(X = 0) = 1 - p$. We view the outcome 1 as a **success**, and 0 as a **failure**. We write $X \sim \text{Bern}(p)$. We also call X a **binary indicator random variable**.

$$E[X] = \sum_n n \,\text{Prob}(X = n)$$

$$= 0 \cdot (1-p) + 1 \cdot p = p$$

$$E[X^2] = \sum_n n^2 \,\text{Prob}(X = n)$$

$$= 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

$$\text{Var}(X) = E[X^2] - E[X]^2 = p - p^2 = p(1-p) \in (0,1)$$

largest at $p = 1/2$: $f(p) = p(1-p) = p - p^2$ so $f'(p) = 1 - 2p$

Theorem: Simplify, Simplify

$f'(p) = 0 \rightarrow p = 1/2$

$f''(p) = -2$ so it's a max

The Binomial Distribution: Let n be a positive integer and let $p \in [0,1]$. Then X has the **binomial distribution** with parameters n and p if

$$\text{Prob}(X = k) \;=\; \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k \in \{0, 1, \ldots, n\} \\ 0 & \text{otherwise.} \end{cases}$$

We write $X \sim \text{Bin}(n, p)$. The mean of X is np and the variance is $np(1-p)$.

$X = X_1 + \cdots + X_n$ each $X_i \sim \text{Bern}(p)$ and indep

$E[X] = E[X_1] + \cdots + E[X_n] = np$

$Var(X) = Var(X_1) + \cdots + Var(X_n) = np(1-p)$

$$\binom{5}{3} = \frac{5!}{3!\,2!} = 10$$

$$\binom{3}{5} = \frac{3!}{5!\,(-2)!} = 0$$

$(-2)!$ is "infinity"

Γ (Gamma function)

$\Gamma(n+1) = n!$ n non-neg int, $\left(-\frac{1}{2}\right)! = \sqrt{\pi}$

$X \sim Bin(n, p)$

$$E[X] = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

$$E[X^2] = \sum_{k=0}^{n} k^2 \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

Know $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$

Does this help?

Differentiating Identities

The multinomial distribution and coefficients. Let n, k be positive integers, and let $p_1, p_2, \ldots, p_n \in [0, 1]$ be such that $p_1 + \cdots + p_n = 1$. Let $x_1, \ldots, x_n \in \{0, 1, \ldots, n\}$ be such that $x_1 + \cdots + x_n = n$. The corresponding **multinomial coefficient** is

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1! x_2! \cdots x_k!},$$

and all other choices of the x_i's evaluate to zero. The **multinomial distribution** with parameters n, k and $p_1, \ldots, p_k$ is non-zero only for such $(x_1, \ldots, x_k)$, where the density is

$$\binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2 x_2 \cdots p_k^{x_k}.$$

We write $X \sim \text{Multinomial}(n, k, p_1, \ldots, p_k)$.