Lecture 10: Introducing the Normal Distribution

Relevant textbook passages:

Pitman [5]: Sections 1.2, 2.2, 2.4

10.1 Standardized random variables

Pitman [5]: p. 190

Recall from Lecture 7.2 that given a random variable X with finite mean μ and variance σ^2, the **standardization** of X is the random variable X^* defined by

$$X^* = \frac{X - \mu}{\sigma}.$$

Note that

$$\boldsymbol{E}\, X^* = 0, \quad \text{and} \quad \boldsymbol{Var}\, X^* = 1,$$

and

$$X = \sigma X^* + \mu,$$

so that X^* is just X measured in different units, called **standard units**. Standardized random variables are extremely useful because of the Central Limit Theorem, which will be described in Lecture 12. As Hodges and Lehmann [3, p. 179] put it,

> One of the most remarkable facts in probability theory, and perhaps in all mathematics, is that histograms of a wide variety of distributions are nearly the same when the right units are used on the horizontal axis.

Just what does this standardized histogram look like? It looks like the Standard Normal density.

10.2 The Standard Normal distribution

Pitman [5]: Section 2.2
Larsen–Marx [4]: Section 4.3

10.2.1 Definition *A random variable has the* ***Standard Normal distribution****, denoted* $N(0,1)$*, if it has a density given by*

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad (-\infty < z < \infty).$$

The cdf of the standard normal is often denoted by Φ. That is,

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

See the figures below.

- It is traditional to denote a standard normal random variable by the letter Z.
- There is no closed form expression for the integral $\Phi(x)$ in terms of elementary functions (polynomial, trigonometric, logarithm, exponential).

- However

$$f(z) = 1 - \frac{1}{2}(1 + 0.196854z + 0.115194z^2 + 0.000344z^3 + 0.019527z^4)^{-4}$$

satisfies

$$|\Phi(z) - f(z)| \leqslant .00025 \quad \text{for all } z \geqslant 0.$$

See Pitman [5, p. 95]. Here is a table of selected values as computed by Mathematica 11.

z	$\Phi(z)$	$f(z)$	$f(z) - \Phi(z)$
0.	0.5	0.5	0
0.25	0.598706	0.598671	-3.537×10^{-5}
0.5	0.691462	0.691695	2.324×10^{-4}
0.75	0.773373	0.773381	8.406×10^{-6}
1.	0.841345	0.841124	-2.209×10^{-4}
1.25	0.89435	0.894195	-1.549×10^{-4}
1.5	0.933193	0.93327	7.741×10^{-5}
1.75	0.959941	0.960166	2.255×10^{-4}
2.	0.97725	0.977437	1.871×10^{-4}
2.25	0.987776	0.987805	2.984×10^{-5}
2.5	0.99379	0.993662	-1.285×10^{-4}
2.75	0.99702	0.996803	-2.17×10^{-4}
3.	0.99865	0.998421	-2.293×10^{-4}
3.25	0.999423	0.999229	-1.937×10^{-4}
3.5	0.999767	0.999626	-1.418×10^{-4}
3.75	0.999912	0.999818	-9.378×10^{-5}
4.	0.999968	0.999911	-5.758×10^{-5}
4.25	0.999989	0.999956	-3.349×10^{-5}
4.5	0.999997	0.999978	-1.875×10^{-5}
4.75	0.999999	0.999989	-1.026×10^{-5}
5.	1.	0.999994	-5.545×10^{-6}

- It follows from symmetry of the density around 0 that for $z \geqslant 0$,

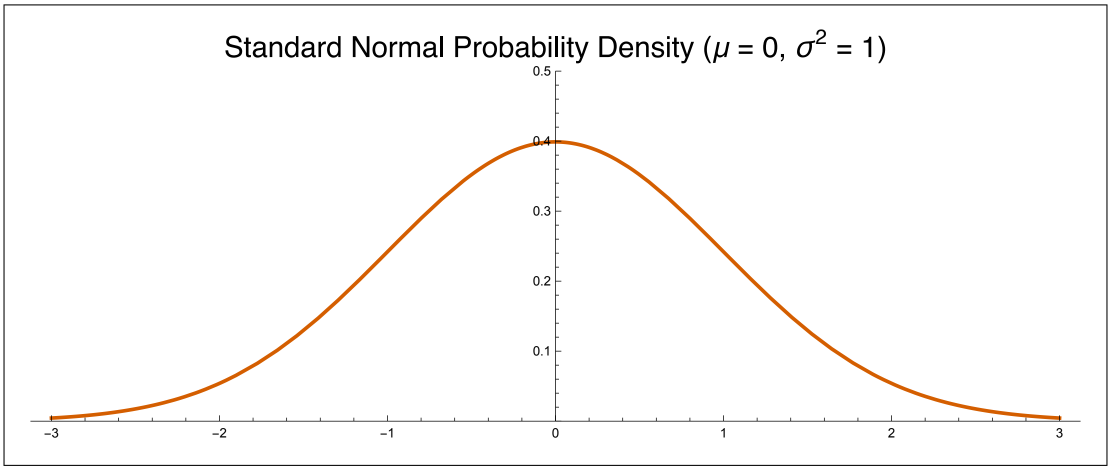
$$\Phi(-z) = 1 - \Phi(z).$$

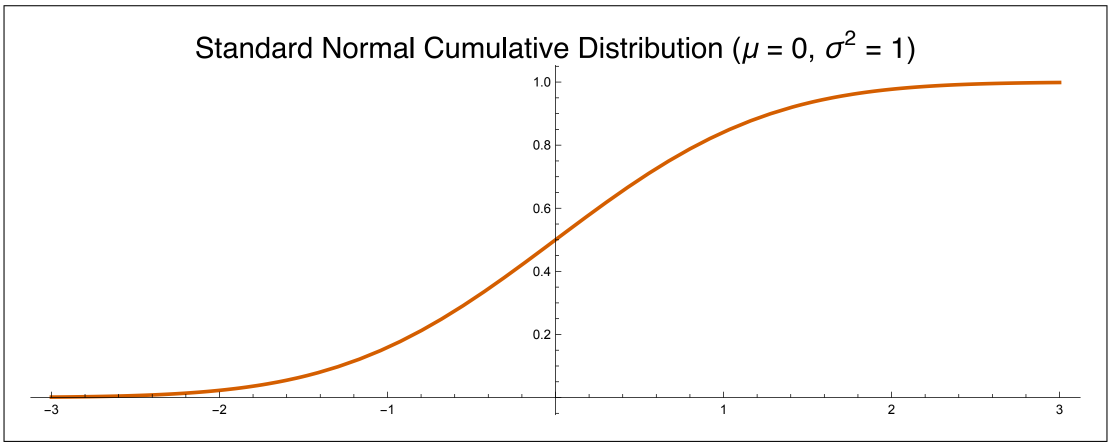
In order for the standard Normal density to be a true density, the following result needs to be true, which it is.

10.2.2 Proposition

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi}.$$

The proof is an exercise in integration theory. See for instance, Pitman [5, pp. 358–359], or Proposition 10.9.1 below.





10.2.3 Proposition *The mean of a normal $N(0,1)$ random variable Z is* 0 *and its variance is* 1.

Proof of Proposition 10.2.3: To show that the expectation exists, we need to show that the integral $\int_0^\infty ze^{-z^2/2}\,dz$ is finite, and then symmetry implies

$$\boldsymbol{E}\,Z = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ze^{-z^2/2}\,dz = 0.$$

Since $z < z^2$ for $z > 1$, it suffices to show that $\int_0^\infty z^2e^{-z^2/2}\,dz$ is finite, which we do next.

Now

$$\boldsymbol{Var}\,Z = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2e^{-z^2/2}\,dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} (-z)\left(-ze^{-z^2/2}\right)\,dz.$$

Integrate by parts: Let $u = e^{-\frac{z^2}{2}}$, $du = -ze^{-\frac{z^2}{2}}dz$, $v = -z$, $dv = -dz$, to get

$$\begin{aligned}
\textit{\textbf{Var}}\, Z &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{(-z)}_{v} \underbrace{\left(-ze^{-\frac{z^2}{2}}\right) dz}_{du} \\
&= \frac{1}{\sqrt{2\pi}} \left(\underbrace{\underbrace{-ze^{-\frac{z^2}{2}}}_{uv} \Bigg|_{-\infty}^{+\infty}}_{=0} \underbrace{- \int_{-\infty}^{\infty} -\underbrace{e^{-\frac{z^2}{2}}}_{u}\, dz}_{\sqrt{2\pi}} \right) \\
&= 1.
\end{aligned}$$

∎

10.2.1 The error function

A function closely related to Φ that is popular in error analysis in some of the sciences is the **error function, erf**, denoted erf defined by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt, \qquad x \geqslant 0.$$

It is related to Φ by

$$\Phi(z) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right),$$

or

$$\operatorname{erf}(z) = 2\Phi(z\sqrt{2}) - 1.$$

10.2.2 The "tails" of a standard normal

While we cannot explicitly write down Φ, we do know something about the "tails" of the cdf. Wasserman [8, Theorem 4.7, p. 65] offers this concentration inequality, known as **Mills inequality**. Compare this to Hoeffding's Inequality 7.3.8.

10.2.4 Proposition (Mills inequality) *If Z is a standard normal random variable, then for $x > 0$,*

$$P(\,|Z| > x) \leqslant \frac{2}{x} \frac{\exp(-\frac{1}{2}x^2)}{\sqrt{2\pi}}.$$

Pollard [6, p. 191] provides the following improvement on this bound.

10.2.5 Proposition *If Z is a standard normal random variable, then for $x > 0$,*

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{\exp(-\frac{1}{2}x^2)}{\sqrt{2\pi}} \leqslant P(Z > x) \leqslant \frac{1}{x} \frac{\exp(-\frac{1}{2}x^2)}{\sqrt{2\pi}} \qquad (x > 0).$$

So a reasonable simple approximation of $1 - \Phi(x)$ for $x \gg 0$, the area under the density to the right of x is just

$$1 - \Phi(x) \approx \frac{1}{x} \frac{\exp(-\frac{1}{2}x^2)}{\sqrt{2\pi}},$$

which is just $\frac{1}{x}$ times the density at x. The error is at most $\frac{1}{x^2}$ of this estimate.

10.3 The Normal Family

There is actually a whole family of **Normal distributions** or **Gaussian distributions**.[1] The family of distributions is characterized by two parameters μ and σ. Because of the Central Limit Theorem, it is one of the most important families in all of probability theory.

Given a standard Normal random variable Z, consider the random variable

$$X = \sigma Z + \mu.$$

It is clear from Section 6.7 and Proposition 6.10.2 that

$$\boldsymbol{E}\, X = \mu, \qquad \boldsymbol{Var}\, X = \sigma^2.$$

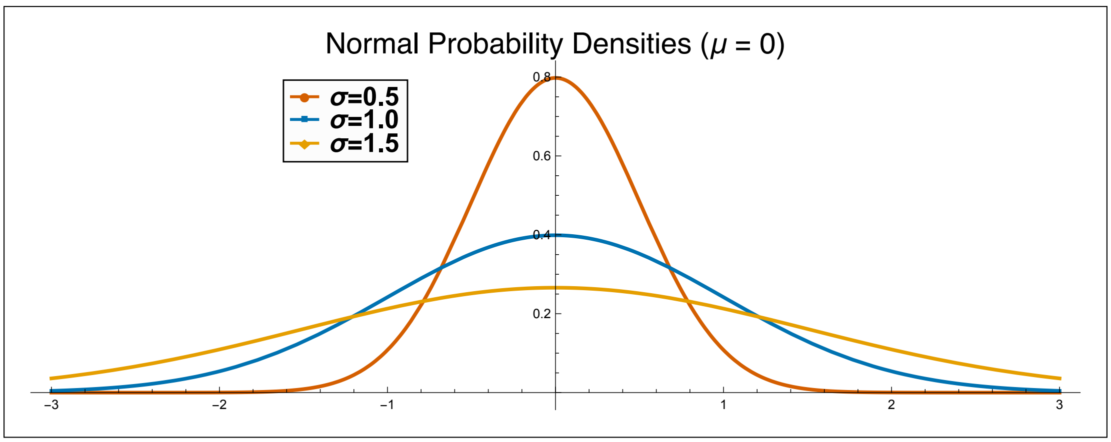
What is the density of X? Well,

$$\operatorname{Prob}(X \leqslant x) = \operatorname{Prob}(\sigma Z + \mu \leqslant x) = \operatorname{Prob}(Z \leqslant (x-\mu)/\sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Now the density of X is the derivative of the above expression with respect to x, which by the chain rule is

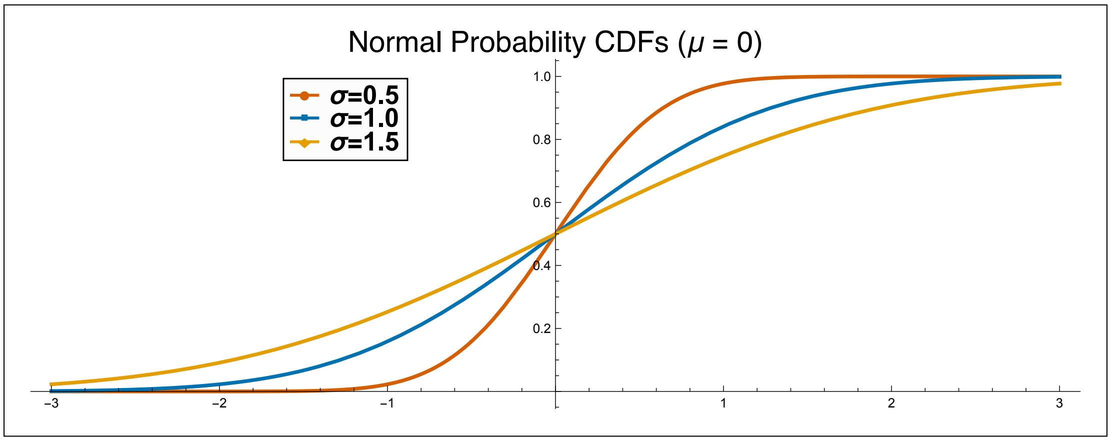
$$\frac{d}{dx}\Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi'\left(\frac{x-\mu}{\sigma}\right)\frac{d}{dx}\left[\frac{x-\mu}{\sigma}\right] = f\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

This density is called the **Normal $\boldsymbol{N(\mu,\sigma^2)}$ density**.[2] Its integral is the **Normal $\boldsymbol{N(\mu,\sigma^2)}$ cdf**, denotes $\Phi_{(\mu,\sigma^2)}$. We say that X has a **Normal $\boldsymbol{N(\mu,\sigma^2)}$ distribution**. As we observed above, the mean of a $N(\mu,\sigma^2)$ random variable is μ, and its variance is σ^2.



[1] According to B. L. van der Waerden [7, p. 11], Gauss assumed this density represented the distribution of errors in astronomical data.

[2] Note that both Mathematica and R parametrize a normal density by its standard deviation σ instead of its variance σ^2.



Pitman [5]: p. 267

We have just proved the following:

10.3.1 Theorem *$Z \sim N(0,1)$ if and only if $X = \sigma Z + \mu \ \sim N(\mu, \sigma^2)$.*

That is, any two normal distributions differ only by **scale** and **location**.

In Proposition 10.8.1 we shall prove the following.

10.3.2 Fact *If $X \sim N[\mu, \sigma^2]$ and $Y \sim N[\lambda, \tau^2]$ are independent normal random variables, then*

$$(X + Y) \sim N[\mu + \lambda, \sigma^2 + \tau^2].$$

The only nontrivial part of this is that $X + Y$ has a normal distribution.

Aside: There are many fascinating properties of the normal family—enough to fill a book, see, e.g., Bryc [1]. Here's one [1, Theorem 3.1.1, p. 39]: If X and Y are independent and identically distributed and X and $(X+Y)/\sqrt{2}$ have the same distribution, then X has a normal distribution.

Or here's another one (Feller [2, Theorem XV.8.1, p. 525]): If X and Y are independent and $X + Y$ has a normal distribution, then both X and Y have a normal distribution.

10.4 The Binomial(n, p) and the Normal$\big(np, np(1-p)\big)$

One of the early reasons for studying the Normal family is that it approximates the Binomial family for large n. We shall see in Lecture 12 that this approximation property is actually much more general.

Fix p and let X be a random variable with a Binomial(n, p) distribution. It has expectation $\mu = np$, and variance $np(1-p)$. Let $\sigma_n = \sqrt{np(1-p)}$ denote the standard deviation of X.

The standardization X^* of X is given by

$$X^* = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{np(1-p)}}.$$

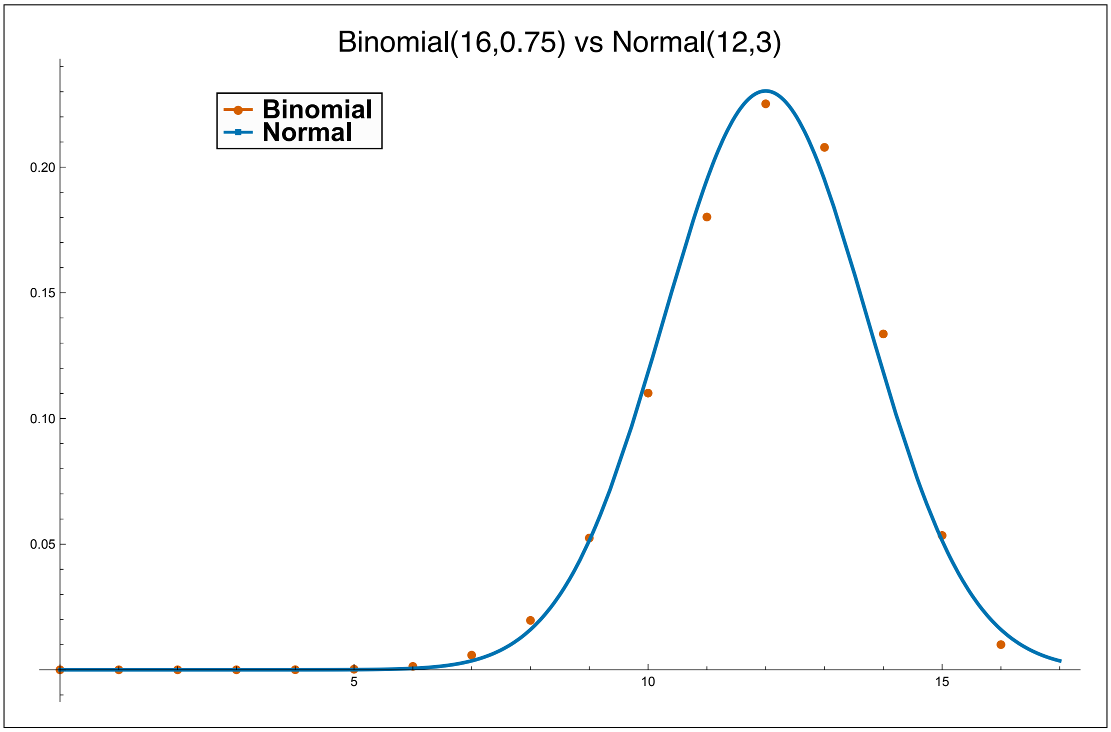
Also, $B_n(k) = P(X = k)$, the probability of k successes, is given by

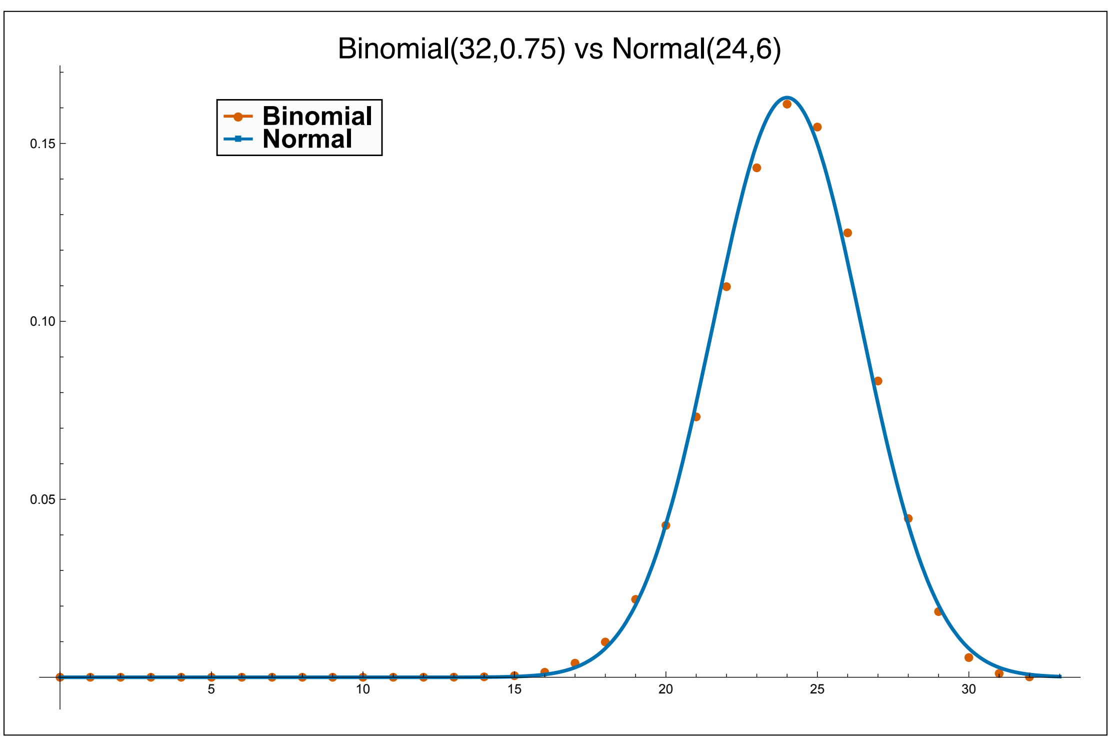
$$B_n(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

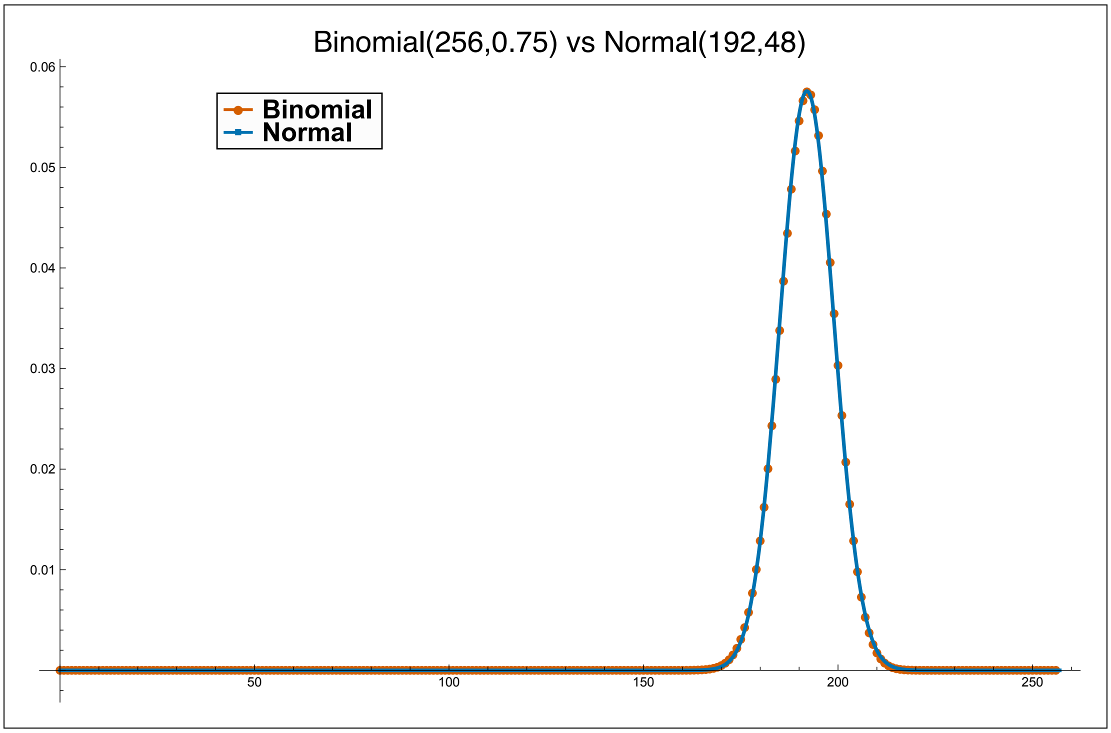
It was realized early on that the normal $N(np, \sigma_n^2)$ density was a good approximation to $B_n(k)$. In fact, for each k,

$$\lim_{n\to\infty} \left| B_n(k) - \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{1}{2}\frac{(k-np)^2}{\sigma_n^2}} \right| = 0. \tag{1}$$

There is a nice proof of this result based on Stirling's approximation on Wikipedia.

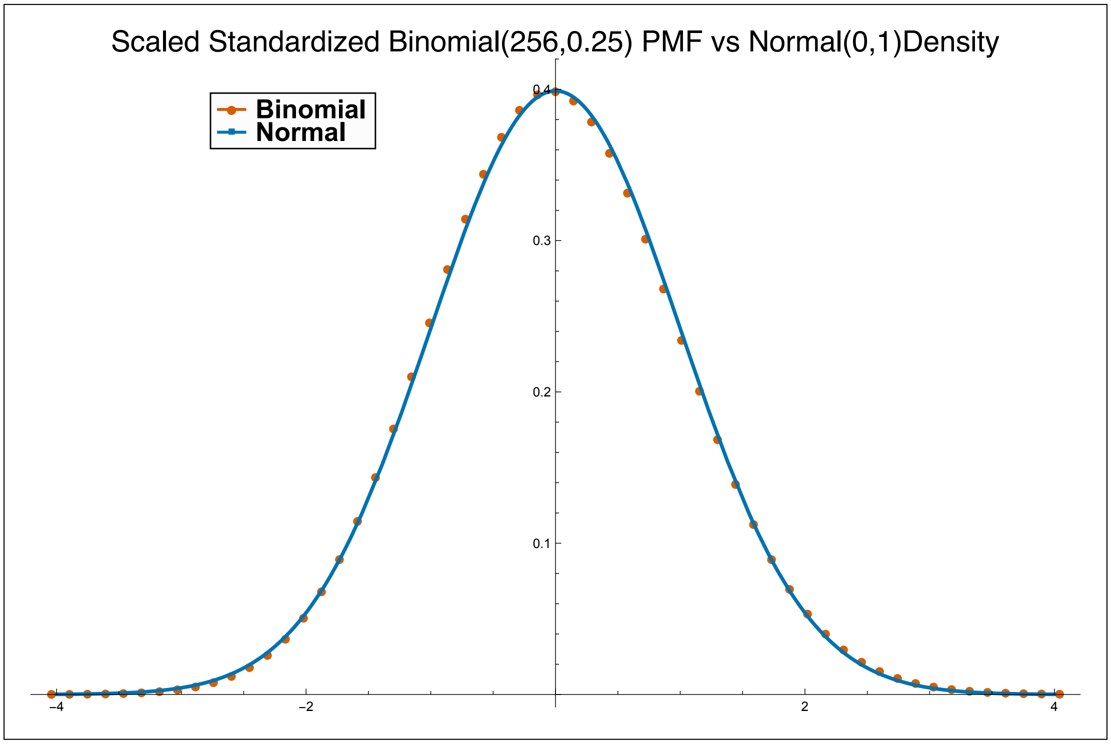





Binomial(256,0.75) vs Normal(192,48)
Binomial
Normal
0.06
0.05
0.04
0.03
0.02
0.01
50
100
150
200
250

In practice, it is simpler to compare the standardized X^* to a standard normal distribution. Defining $\kappa(z) = \sigma_n z + np$, we can rewrite (1) as follows: For each $z = (k - np)/\sigma_n$ we have $\kappa(z) = k$, so

$$\lim_{n\to\infty} \left| \sigma_n B_n\big(\kappa(z)\big) - \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \right| = 0 \tag{2}$$



10.5 DeMoivre–Laplace Limit Theorem

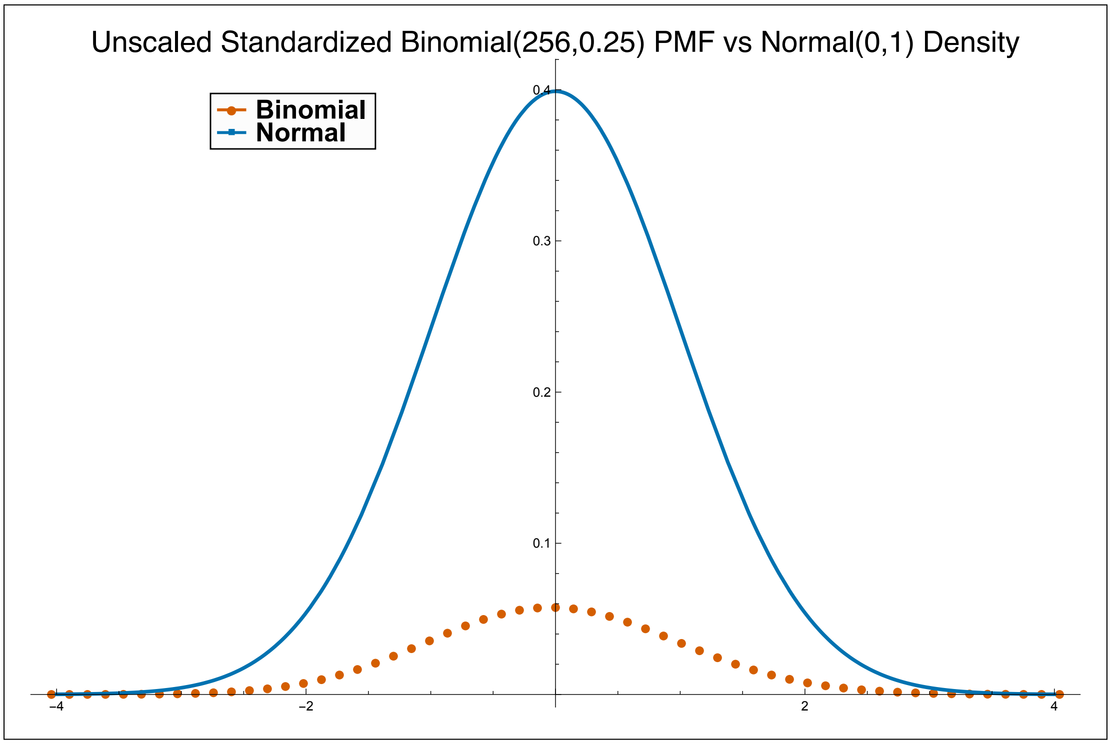
The normal approximation to the Binomial can be rephrased as:

10.5.1 DeMoivre–Laplace Limit Theorem *Let X be Binomial(n, p) random variable. It has mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. Its standardization is $X^* = (X-\mu)/\sigma$. For any real numbers a, b,*

$$\lim_{n\to\infty} P\left(a \leqslant \frac{X - np}{\sqrt{np(1-p)}} \leqslant b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2}\, dx.$$

See Larsen–Marx [4, Theorem 4.3.1, pp. 239–240] or Pitman [5, Section 2.1, esp. p. 99]. For $p \neq 1/2$ it is recommended to use a "skew-normal" approximation, see Pitman [5, p.106]. In practice, [5, p. 103] asserts that the error from using this refinement is "negligible" provided $\sqrt{np(1-p)} \geqslant 3$.

Aside: Note the scaling of B in (2). If we were to plot the probability mass function for X^* against the density for the Standard Normal we would get a plot like this:



The density dwarfs the pmf. Why is this? Well the standardized binomial takes on the value $(k-\mu)/\sigma$ with the binomial pmf $p_{n,p}(k)$. As k varies, these values are spaced $1/\sigma$ apart. The *value* of the density at $z=(k-\mu)$ is $f(z)=\frac{1}{\sqrt{2\pi}}z^{-z^2/2}$, but the *area* under the density is approximated by the area under a step function, where there the steps are centered at the points $(k-\mu)/\sigma$ and have width $1/\sigma$. Thus each $z=(k-\mu)$ contributes $f(z)/\sigma$ to the probability, while the pmf contributes $p_{n,p}(k)$. Thus the DeMoivre–Laplace Theorem says that when $z=(k-\mu)/\sigma)$, then $f(z)/\sigma$ is approximately equal to $p_{n,p}(k)$, or $f(z)\approx\sigma p_{n,p}(k)$.

10.5.2 Remark We can rewrite the standardized random variable X^* in terms of the average frequency of success, $f=X/n$. Then

$$X^*=\frac{X-\mu}{\sigma}=\frac{X-np}{\sqrt{np(1-p)}}=\frac{f-p}{\sqrt{p(1-p)}}\sqrt{n}. \tag{3}$$

This formulation is sometimes more convenient. For the special case $p=1/2$, this reduces to $X^*=(2X-n)/\sqrt{n}$.

10.6 Using the Normal approximation

Update this every year!

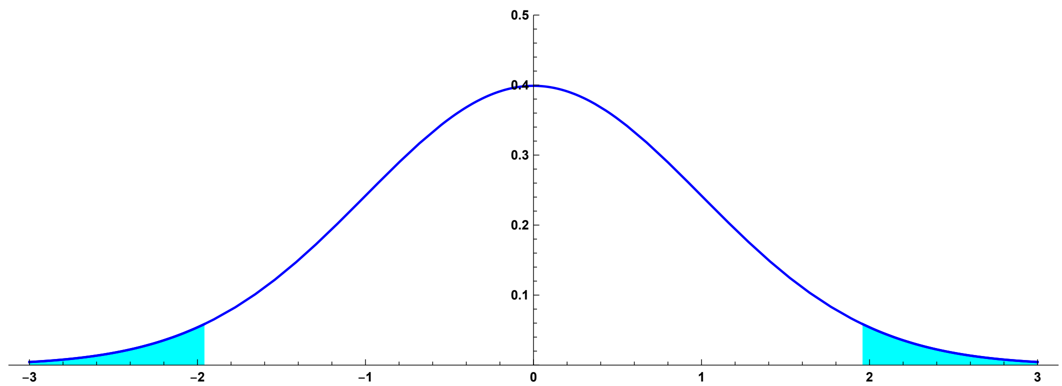
Let's apply the Normal approximation to coin tossing. We will see if the results of your tosses are consistent with $P(\text{Heads})=1/2$. Recall that the results for 2020 were 12,618 Tails out of 25,600 tosses, or 49.289%. This misses the expected number (12,800) by −182. Is this "close enough" to 1/2?

The Binomial$(25600,1/2)$ random variable has expectation $\mu=12,800$ and standard deviation $\sigma=\sqrt{n}/2=80$. So assuming the coin is fair, the value of the standardized result is $(12618-12800)/80=-182/80=-2.28$. This value is called the ***z*-score** of the experiment. It tells us that this year's experiment missed the expectation by −2.28 standard deviations (assuming the coin actually is fair).

We can use the DeMoivre–Laplace Limit Theorem to treat X^* as a Standard Normal (mean $=0$, variance $=1$).

We now ask what the probability is that a Standard Normal Z takes on a value outside the interval $(-2.28, 2.28)$. This gives the probability an absolute deviation of the fraction of Tails from 1/2 of the magnitude we observed could have happened "by chance."

But this is the area under the normal bell curve, or equivalently, $2\Phi(-2.28) = 0.0226$. This value is called the ***p*-value** of the experiment. It means that if the coins are really "fair," then 2.26% of the time we run this experiment we would get a deviation from the expected value at least this large. If the p-value were to be really small, then we would be suspicious of the coin's fairness.



This shaded area is the probability of the event $(|Z| > 1.96)$, which is equal to $2\Phi(-1.96) = 0.05$. Values outside the interval $(-1.96, 1.96)$ are often regarded as unlikely to have occurred "by chance."

The binomial *p*-value

The p-value above was derived by treating the standardized Binomial$(25600, 1/2)$ random variable as if it were a Standard Normal, which it isn't, exactly. But with modern computational tools we can compute the binomial p-value which is given by

$$2\sum_{k=0}^{12618}\binom{25600}{k}(1/2)^{25600},$$

where 12618 is the smaller of the number of Tails (12,618) and Heads (12,982). (Why?) R reports this value to be 0.0233. The efficient way to compute the sum is to use the built-in CDF functions. For R, this is

```
2 * pbinom(12618, 25600, prob = 0.5)
```

and for Mathematica, use

```
2 CDF[BinomialDistribution[25600, 0.5], 12618];
```

Larsen–Marx [4, p. 242] has a section on improving the Normal approximation to deal with integer problems by making a "continuity correction," but it doesn't seem worthwhile in this case.

All years

What if I take the data from all eight years of this class? There were 106,048 Tails in 212,480 tosses, which is This is 49.91%. This gives a z-score of -0.83 for a p-value of 0.4065. The binomial p-value is 0.406.

10.7 Sample size and the Normal approximation

Let X be a Binomial(n, p) random variable, and let $f = X/n$ denote the fraction of trials that are successes. The DeMoivre–Laplace Theorem and equation (3) (on page 10–10) tell us that for large n, we have the following approximation

$$\frac{X - np}{\sqrt{np(1-p)}} = \frac{f - p}{\sqrt{p(1-p)}}\sqrt{n} \sim N(0, 1)$$

We can use this to calculate the probability that f, our experimental result, is "close" to p, the true probability of success.

How large does n have to be for $f = X/n$ to be "close" to be p with "high" probability?

- Let $\varepsilon > 0$ denote the target level of closeness,
- let $1 - \alpha$ designate what we mean by "high" probability (so that α is a "low" probability).
- We want to choose n big enough so that

$$\boxed{P(\,|f - p| > \varepsilon) \leqslant \alpha.} \tag{4}$$

Now

$$|f - p| > \varepsilon \iff \frac{|f - p|}{\sqrt{p(1-p)}}\sqrt{n} > \frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n},$$

and we know from (3) that

$$P\left(\frac{|f - p|}{\sqrt{p(1-p)}}\sqrt{n} > \frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n}\right) \approx P\left(|Z| > \frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n}\right), \tag{5}$$

where $Z \sim N(0, 1)$. Thus

$$P(\,|f - p| > \varepsilon) \approx P\left(|Z| > \frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n}\right).$$

Define the function $\zeta(a)$ by

$$P(Z > \zeta(a)) = a,$$

or equivalently

$$\boxed{\zeta(a) = \Phi^{-1}(1 - a),}$$

where Φ is the Standard Normal cumulative distribution function. See Figure 10.1. This is something you can look up with R or Mathematica's built-in quantile functions. See Figure 10.2. By symmetry,

$$P(\,|Z| > \zeta(a)) = 2a.$$

So by (5) we want to find n such that

$$\zeta(a) = \frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n} \quad \text{where} \quad 2a = \alpha,$$

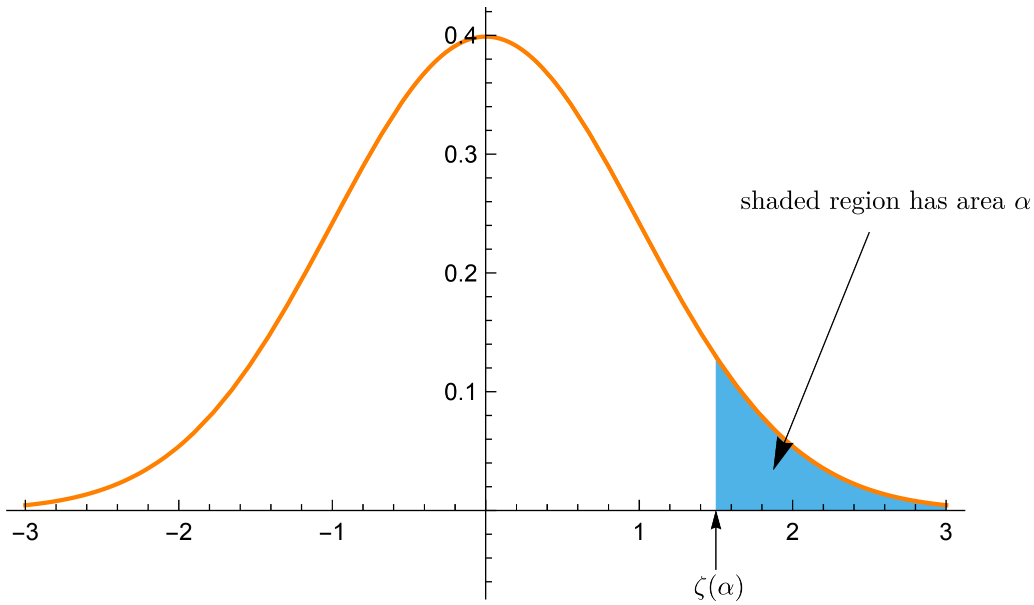


Figure 10.1. Given α, $\zeta(\alpha)$ is chosen so that the shaded region has area $= \alpha$.

or in other words, find n so that

$$\zeta(\alpha/2) = \frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n}$$

$$\zeta^2(\alpha/2) = n\frac{\varepsilon^2}{p(1-p)}$$

$$n = \frac{\zeta^2(\alpha/2)p(1-p)}{\varepsilon^2}.$$

There is a problem with this, namely, it depends on p. But we do have an upper bound on $p(1-p)$, which is maximized at $p = 1/2$ and $p(1-p) = 1/4$.

Thus to be sure that the fraction f of success in n trials is ε-close to p with probability $1-\alpha$, that is,

$$P(\,|f-p| > \varepsilon) \leqslant \alpha.$$

we need to choose n large enough, namely

$$n \geqslant \frac{\zeta^2(\alpha/2)}{4\varepsilon^2},$$

where

$$\zeta(\alpha/2) = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Let's take some examples:

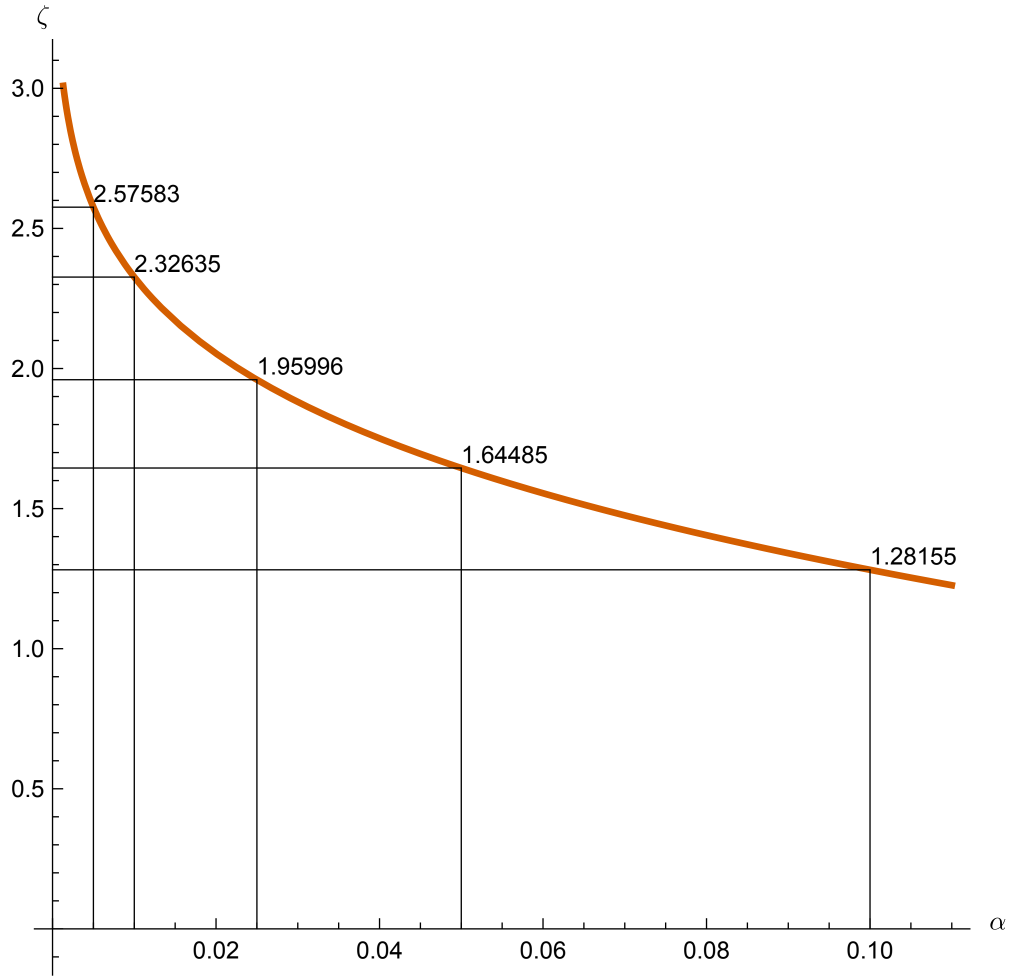


Figure 10.2. The graph of $\zeta(\alpha) = \Phi^{-1}(1-\alpha)$.

ε	α	$\zeta(\alpha/2)$	n
0.05	0.05	1.96	385
0.03	0.05	1.96	1,068
0.01	0.05	1.96	9,604
0.001	0.05	1.96	960,365
0.05	0.01	2.58	664
0.03	0.01	2.58	1,844
0.01	0.01	2.58	16,588
0.001	0.01	2.58	1,658,725
0.05	0.001	3.29	1,083
0.03	0.001	3.29	3,008
0.01	0.001	3.29	27,069
0.001	0.001	3.29	2,706,892

Larsen and Marx [4] have a very nice application to polling and "margin of error" on pages 305–309.

10.8 Sum of Independent Normals

If $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\lambda, \tau^2)$ and X and Y are independent, the joint density is

$$f(x,y) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \frac{1}{\sqrt{2\pi}\tau}\, e^{-\frac{1}{2}\left(\frac{y-\lambda}{\tau}\right)^2}$$

Now we could try to use the convolution formula to find the density of $X+Y$, but that way lies madness. Let's be clever instead.

Pitman [5]: Section 5.3

Start by considering independent standard normals X and Y. The joint density is then

$$\begin{aligned} f(x,y) &= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}} \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}. \end{aligned}$$

The level curves of the density are circles centered at the origin, and the density is invariant under rotation.

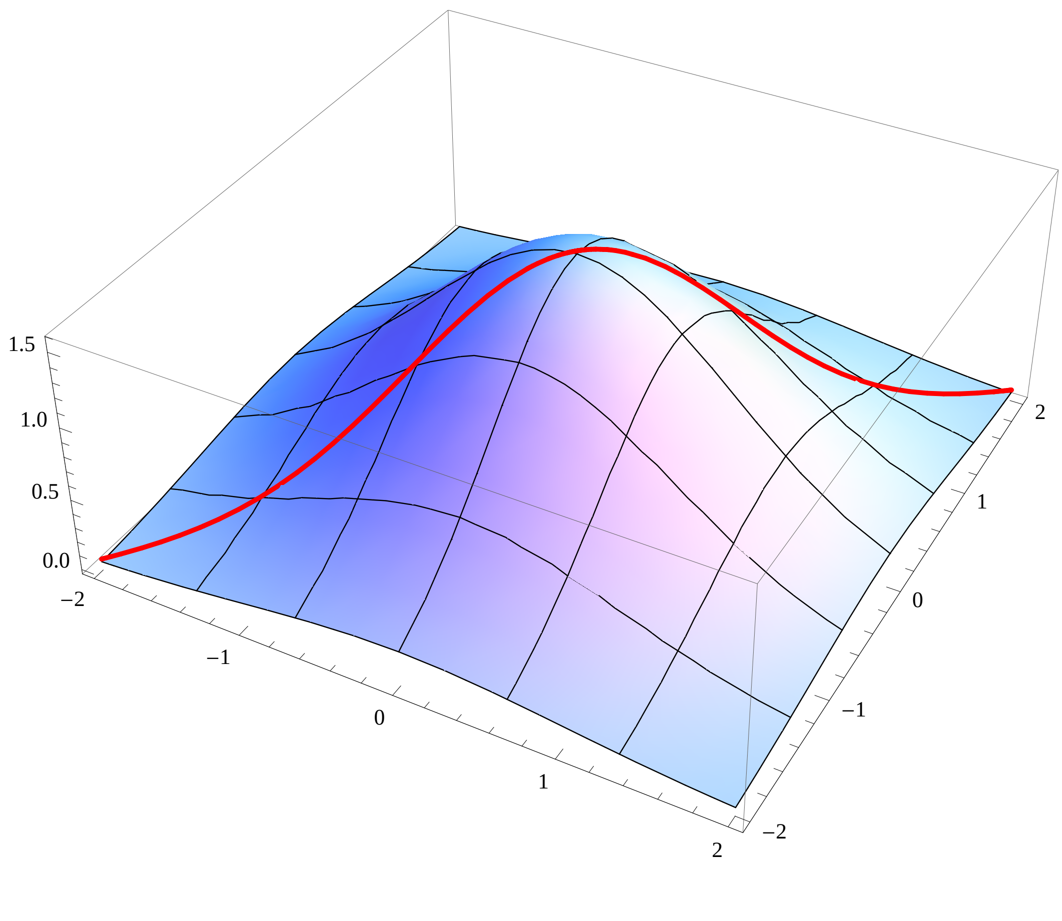


Figure 10.3. The Standard Bivariate Normal density.

Invariance under rotation lets us be clever. Consider rotating the axes by an angle θ. The coordinates of (X,Y) relative to the new axes are (X_θ, Y_θ), given by

$$\begin{bmatrix} X_\theta \\ Y_\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

See Figure 10.4.[3] The first row of the matrix equation gives $X_\theta = X\cos\theta + Y\sin\theta$.

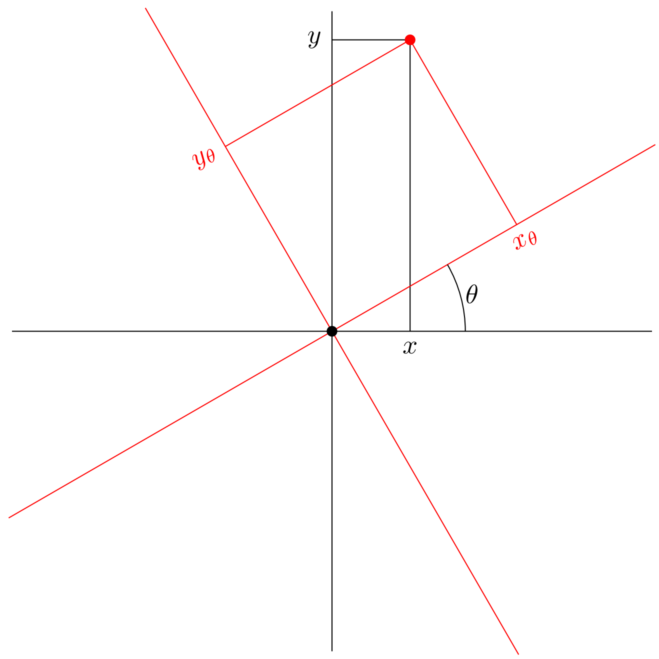


Figure 10.4. Rotated coordinates

By invariance under rotation, we conclude that the marginal density of X_θ and the marginal density of X are the same, namely $N(0,1)$. But any two numbers a and b with $a^2 + b^2 = 1$ correspond to an angle $\theta \in [-\pi, \pi]$ with $\cos\theta = a$ and $\sin\theta = b$. This means that $aX + bY$ is of the form X_θ, and thus has distribution $N(0,1)$. Thus we have proved the following.

The random variable $aX + bY \sim N(0,1)$ whenever $a^2 + b^2 = 1$ and X, Y are independent standard normals.

This fact lets us determine the distribution of the sum of independent normals:

10.8.1 Proposition *If X and Y are independent normals with $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\lambda, \tau^2)$, then*

$$X + Y \sim N(\mu + \lambda, \sigma^2 + \tau^2).$$

[3] To see this, the standard basis vector $(1,0)$ after a rotation through angle θ becomes the new basis vector $(\cos\theta, \sin\theta)$, and the standard basis vector $(0,1)$ becomes the new basis vector $(-\sin\theta, \cos\theta)$, which is orthogonal to the first. The coordinates of the point (x, y) with respect to the new basis are (x_θ, y_θ) where $(x, y) = x_\theta(\cos\theta, \sin\theta) + y_\theta(-\sin\theta, \cos\theta)$. In matrix terms this becomes

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_\theta \\ y_\theta \end{bmatrix}.$$

So

$$\begin{bmatrix} x_\theta \\ y_\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Proof: Standardize X and Y and define W by:

$$X^* = \frac{X-\mu}{\sigma}, \quad Y^* = \frac{Y-\lambda}{\tau}, \quad W = \frac{X+Y-(\lambda+\mu)}{\sqrt{\sigma^2+\tau^2}}.$$

Then $X^* \sim N(0,1)$ and $Y^* \sim N(0,1)$ are independent, and

$$W = \frac{\sigma}{\sqrt{\sigma^2+\tau^2}} X^* + \frac{\tau}{\sqrt{\sigma^2+\tau^2}} Y^*.$$

By the above observation, $W \sim N(0,1)$, which implies

$$X+Y = \sqrt{\sigma^2+\tau^2}\, W + (\lambda+\mu) \sim N(\mu+\lambda, \sigma^2+\tau^2).$$

∎

Of course this generalizes to finite sums of independent normals, not just two.

10.9 ⋆ Appendix: The normal density

10.9.1 Proposition (The Normal Density) *The standard normal density*

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{z^2}{2}\right)}$$

is truly a density. That is, $\int_{\boldsymbol{R}} f(z)\, dz = 1$.

Proof: Set

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

We will show $I^2 = 1$, hence $I = 1$.

Now

$$\begin{aligned} I^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\, dz \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+z^2)}\, dx\, dz. \end{aligned}$$

Let

$$H \colon (r,\theta) \mapsto (r\cos\theta, r\sin\theta)$$

so $H \colon (0,\infty) \times [0,2\pi) \to \boldsymbol{R}^2 \setminus \{0\}$ is one-to-one, and

$$J_H(x,\theta) = \det \begin{bmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{bmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

By the Change of Variables Theorem S4.4.2

$$\begin{aligned} &\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+z^2)}\, dx\, dz \\ &\qquad = \frac{1}{2\pi} \int_{0}^{\infty} \int_{0}^{2\pi} e^{-\frac{1}{2}\left((r\cos\theta)^2+(r\sin\theta)^2\right)}\, |r|\; dr\, d\theta \\ &\qquad = \frac{1}{2\pi} \int_{0}^{\infty} \int_{0}^{2\pi} e^{-\frac{1}{2}r^2} r\, dr\, d\theta. \end{aligned}$$

By Fubini's Theorem S4.3.1 this is equal to

$$
\begin{aligned}
&= \frac{1}{2\pi} \int_0^{2\pi} \left(\int_0^\infty e^{-\frac{1}{2}r^2} r\, dr \right) d\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi} \underbrace{\left[-e^{\frac{-r^2}{z}} \Big|_0^\infty \right]}_{0-(-1)} d\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi} 1\, d\theta = 1.
\end{aligned}
$$

Thus $I^2 = 1$, and since clearly $I > 0$, we must have $I = 1$. ∎

10.9.1 Another calculation of the normal density

This is based on Weld [9, pp. 180–181].

Start by observing that $\frac{d}{dt} e^{-at^2} = -2ate^{-at^2}$ so that for $a > 0$,

$$
\int_0^\infty 2ate^{-at^2} = -e^{-at^2}\Big|_0^\infty = 1. \tag{6}
$$

For $t > 0$, define

$$
I(t) = \int_0^\infty e^{-t^2x^2}\, dx, \qquad \text{and} \qquad I_1(t) = \int_0^\infty te^{-t^2x^2}\, dx, \tag{7}
$$

so that

$$
I(t) = \frac{1}{t} I_1(t).
$$

The first thing to note is that $I_1(t)$ is independent of $t > 0$: Consider the substitution $z = tx$. By the Substitution Theorem S4.4.1 we have

$$
I_1(t) = \int_0^\infty z'(x) e^{-z^2(x)}\, dx = \int_0^\infty e^{-z^2}\, dz = I_1(1). \tag{8}
$$

Let $\bar{I}_1$ denote this common value, which is what we shall now compute. Now

$$
\begin{aligned}
\bar{I}_1 e^{-t^2} &= e^{-t^2} I_1(t) \\
&= \int_0^\infty te^{-t^2} e^{-t^2x^2}\, dx, \\
&= \int_0^\infty te^{-t^2(x^2+1)}\, dx
\end{aligned} \tag{9}
$$

Integrating with respect to t gives

$$
\begin{aligned}
\bar{I}_1 \int_0^\infty e^{-t^2}\, dt &= \int_0^\infty \int_0^\infty te^{-t^2(x^2+1)}\, dx\, dt \\
&= \int_0^\infty \int_0^\infty te^{-t^2(x^2+1)}\, dt\, dx.
\end{aligned} \tag{10}
$$

Now the inner integral satisfies

$$
\int_0^\infty te^{-t^2(x^2+1)}\, dt = \frac{1}{2(x^2+1)} \int_0^\infty 2(x^2+1)te^{-t^2(x^2+1)}\, dt = \frac{1}{2(x^2+1)}, \tag{11}
$$

where the last equality follows from (6) with $a = 2(x^2+1)$. So (10) becomes

$$\bar{I}_1 \int_0^\infty e^{-t^2}\,dt = \frac{1}{2}\int_0^\infty \frac{1}{x^2+1}\,dx. \tag{12}$$

Using the well-known Fact S4.2.1 we have

$$\bar{I}_1 \int_0^\infty e^{-t^2}\,dt = \frac{1}{2}\arctan x\Big|_0^\infty = \frac{\pi}{4}. \tag{13}$$

But $\int_0^\infty e^{-t^2}\,dt = I_1(1) = \bar{I}_1$, so (13) implies

$$\bar{I}_1^2 = \frac{\pi}{4}$$

so

$$\bar{I}_1 = \frac{\sqrt{\pi}}{2}.$$

And the rest follows from simple change of variables.

Bibliography

[1] W. Bryc. 1995. *The normal distribution*. Number 100 in Lecture Notes in Statistics. New York, Berlin, and Heidelberg: Springer–Verlag.

[2] W. Feller. 1971. *An introduction to probability theory and its applications*, 2d. ed., volume 2. New York: Wiley.

[3] J. L. Hodges, Jr. and E. L. Lehmann. 2005. *Basic concepts of probability and statistics*, 2d. ed. Number 48 in Classics in Applied Mathematics. Philadelphia: SIAM.

[4] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[5] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.

[6] D. Pollard. 1984. *Convergence of stochastic processes*. Springer Series in Statistics. Berlin: Springer–Verlag.

[7] B. L. van der Waerden. 1969. *Mathematical statistics*. Number 156 in Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete. New York, Berlin, and Heidelberg: Springer–Verlag. Translated by Virginia Thompson and Ellen Sherman from *Mathematische Statistik*, published by Springer-Verlag in 1965, as volume 87 in the series Grundlerhen der mathematischen Wissenschaften.

[8] L. Wasserman. 2010. *All of statistics: A concise course in statistical inference*. Springer Texts in Statistics. New York: Springer Science+Business Media.

[9] L. D. Weld. 1916. *Theory of errors and least squares*. New York: Macmillan.