Ma 3/103 KC Border
Introduction to Probability and Statistics Winter 2015

Lecture 7: Introducing the Normal Distribution

Relevant textbook passages:

Pitman [6]: Sections 1.2, 2.2, 2.4

7.1 The Inclusion/Exclusion Principle

7.1.1 A multinomial formula

Let us refer to the symbols $x_1, \dots, x_n$ as **letters**. We can use these letters to form symbolic sums and products (**polynomials** or **multinomials**) such x_1x_3, $(1 + x_1)$, $(1 + x_1)(1 + 3x_2)$, etc. Terms in a multinomial are scalar multiples of symbolic products of letters. The **degree** of a term is the sum of the exponents of letters in the term. By convention the product of zero letters is 1 and has degree zero.

The next identity is easy to prove by induction on n.

7.1.1 Proposition (A multinomial identity)

$$(1 + x_1)(1 + x_2) \cdots (1 + x_n) = \sum_{k=0}^{n} \sum_{i_1 < \cdots < i_k} x_{i_1} \cdots x_{i_k}. \tag{1}$$

The somewhat unusual notation for the second sum means to sum the product $x_{i_1} \cdots x_{i_k}$ over all sorted lists of letters having length k. The sorting guarantees that we take each set of k distinct letters exactly once. There are $\binom{n}{k}$ such products. The case $k = 0$ corresponds to taking no letters, and the product of zero letters is the symbol 1. For example,

$$(1 + x_1)(1 + x_2) = \underbrace{1}_{k=0} + \underbrace{(x_1 + x_2)}_{k=1} + \underbrace{x_1x_2}_{k=2},$$

$$(1 + x_1)(1 + x_2)(1 + x_3) = \underbrace{1}_{k=0} + \underbrace{(x_1 + x_2 + x_3)}_{k=1} + \underbrace{(x_1x_2 + x_1x_3 + x_2x_3)}_{k=2} + \underbrace{x_1x_2x_3}_{k=3}.$$

Replacing x_i by $-x_i$ we have the following.

7.1.2 Corollary (Another multinomial identity)

$$(1 - x_1)(1 - x_2) \cdots (1 - x_n) = \sum_{k=0}^{n} \sum_{i_1 < \cdots < i_k} (-1)^k x_{i_1} \cdots x_{i_k}. \tag{2}$$

So, for instance,

$$(1-x_1)(1-x_2) = \underbrace{1}_{k=0} - \underbrace{(x_1+x_2)}_{k=1} + \underbrace{x_1x_2}_{k=2},$$

$$(1-x_1)(1-x_2)(1-x_3) = \underbrace{1}_{k=0} - \underbrace{(x_1+x_2+x_3)}_{k=1} + \underbrace{(x_1x_2+x_1x_3+x_2x_3)}_{k=2} - \underbrace{x_1x_2x_3}_{k=3}.$$

7.1.2 Indicator functions

Recall that for an event A, the indicator function $\mathbf{1}_A$ is defined by

$$\mathbf{1}_A(s) = \begin{cases} 1 & \text{if } s \in A, \\ 0 & \text{if } s \notin A. \end{cases}$$

For an event A,

$$\boldsymbol{E}\,\mathbf{1}_A = P(A).$$

The following identities are immediate consequences of the definition of the indicator function. (Recall that AB denotes the intersection of A and B.)

7.1.3 Proposition *For events A and B,*

$$\mathbf{1}_{AB} = \mathbf{1}_A \cdot \mathbf{1}_B$$
$$\mathbf{1}_{A^c} = 1 - \mathbf{1}_A.$$

7.1.3 The Inclusion/Exclusion Principle

We now harness the expectation operator, and the important fact that expectation is a positive linear operator, to prove the mysterious Inclusion/Exclusion Principle (Pitman [6, p. 31]). Kaplansky [4] refers to it as Poincaré's formula, but it is often attributed to de Moivre.

7.1.4 Inclusion/Exclusion Principle *Let $A_1, \dots, A_n$ be an indexed family of events, not necessarily disjoint, nor even distinct apart from the indexing. Then*

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i_1<i_2} P(A_{i_1}A_{i_2}) + \cdots + (-1)^{n+1}P(A_1 \cdots A_n).$$

Proof: This is the proof outlined in Pitman [6, Exercise 21, p. 184]. Start by observing that

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - P\left((\bigcup_{i=1}^n A_i)^c\right) = 1 - P\left(\cap_{i=1}^n A_i^c\right)$$

Now use Proposition 7.1.3 to rewrite this in the language of indicator functions.

$$\boldsymbol{E}\,\mathbf{1}_{\bigcup_{i=1}^n A_i} = 1 - \boldsymbol{E}\,\mathbf{1}_{\cap_{i=1}^n A_i^c} = 1 - \boldsymbol{E}\prod_{i=1}^n (1 - \mathbf{1}_{A_i}). \tag{3}$$

By the second multinomial identity (2),

$$\prod_{i=1}^{n}(1-\mathbf{1}_{A_i}) = \sum_{k=0}^{n} \sum_{i_1<\cdots<i_k} (-1)^k \mathbf{1}_{A_{i_1}} \cdots \mathbf{1}_{A_{i_k}} = \sum_{k=0}^{n} \sum_{i_1<\cdots<i_k} (-1)^k \mathbf{1}_{A_{i_1}\cdots A_{i_k}}.$$

We now use the fact that **expectation is a positive linear operator** to conclude

$$\boldsymbol{E} \prod_{i=1}^{n}(1-\mathbf{1}_{A_i}) = \sum_{k=0}^{n} \sum_{i_1<\cdots<i_k} (-1)^k \boldsymbol{E}\, \mathbf{1}_{A_{i_1}\cdots A_{i_k}} = \sum_{k=0}^{n} \sum_{i_1<\cdots<i_k} (-1)^k P(A_{i_1}\cdots A_{i_k}).$$

Substituting this back into (3) proves the result. ∎

7.2 Variance of an average vs. variance of a sum

Let $X_1, \dots, X_n$ be *independent* random variables, and let

$$S_n = X_1 + \cdots + X_n, \quad \text{and} \quad A_n = (X_1 + \cdots + X_n)/n.$$

Then

$$\boldsymbol{Var}\, S_n = \sum_{i=1}^{n} \boldsymbol{Var}\, X_i, \quad \text{and} \quad \boldsymbol{Var}\, A_n = \frac{\sum_{i=1}^{n} \boldsymbol{Var}\, X_i}{n^2}.$$

When $X_1, \dots, X_n$ are independent and identically distributed with expectation μ and variance σ^2, then

$$\boldsymbol{Var}\, S_n = n\sigma^2 \quad \text{and} \quad \boldsymbol{Var}\, A_n = \frac{\sigma^2}{n},$$

and the standard deviations satisfy

$$\text{s. d. } S_n = \sqrt{n}\sigma \quad \text{and} \quad \text{s. d. } A_n = \frac{\sigma}{\sqrt{n}}.$$

7.3 Standardized random variables

Pitman [6]: p. 190

7.3.1 Definition *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

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

and

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

so that X^* is just X measured in different units, called ***standard units****.*

[Note: Pitman uses both X^* and later X_* to denote the standardization of X.]

Standardized random variables are extremely useful because of the Central Limit Theorem, which will be described in Lecture 10. 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.

7.4 Binomial

The Binomial(n, p) is the distribution of the number X of successes in n independent Bernoulli(p) trials.

The probability mass function is

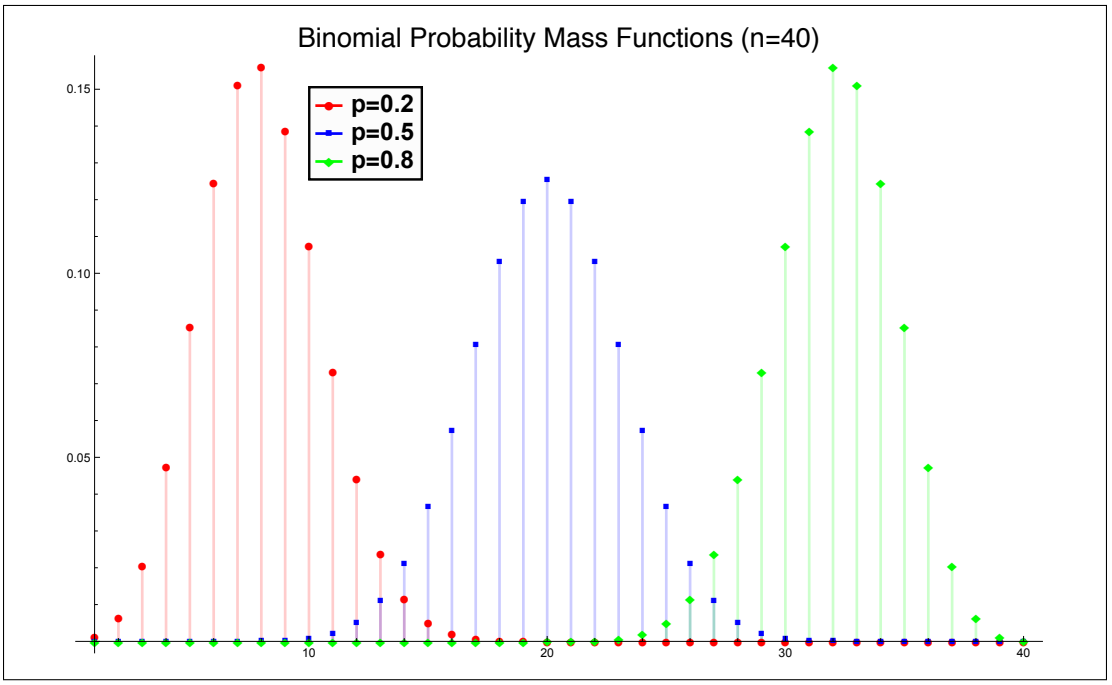
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Since **expectation is a linear operator**, the expectation of a Binomial is the sum of the expectations of the underlying Bernoullis, so if X has a Binomial(n, p) distribution,

$$\boldsymbol{E}\, X = np,$$

and since the variance of the sum of independent random variables is the sum of the variances,

$$\boldsymbol{Var}\, X = np(1-p).$$



7.5 The Normal distribution

Pitman [6]: Section 2.2
Larsen–Marx [5]: Section 4.3

The **Normal distribution** or **Gaussian distribution**[1] is actually a whole family of distributions 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. There is no closed form expression for the cdf of the normal $N[\mu, \sigma^2]$ distribution, but its density is

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

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The case $\mu = 0$ and $\sigma^2 = 1$ is called the **Standard Normal distribution**. See the figures below.

7.5.1 Proposition

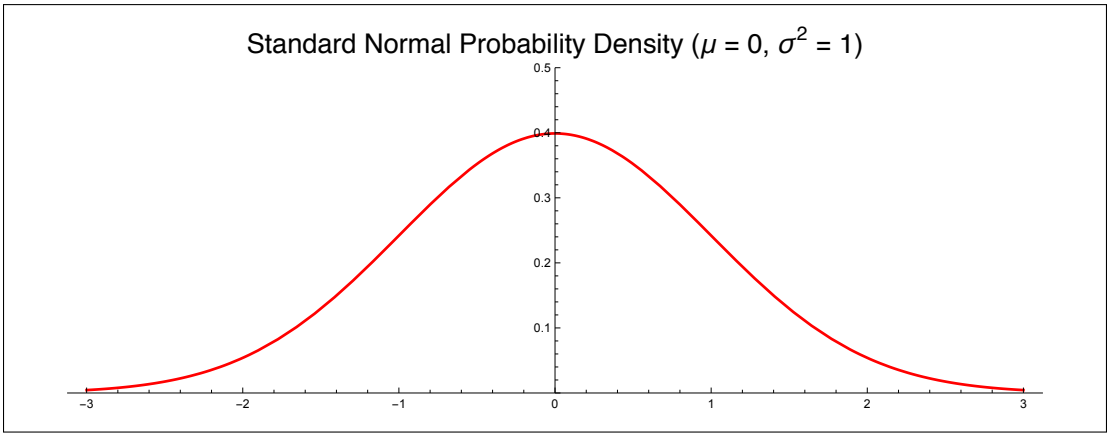
$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx = \sqrt{2\pi}.$$

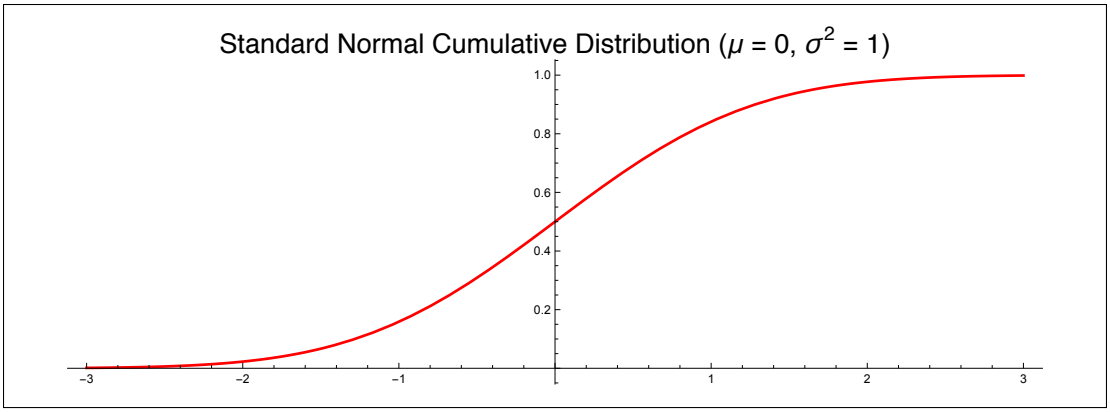
The proof is an exercise in integration theory. See for instance, Pitman [6, pp. 358–359].

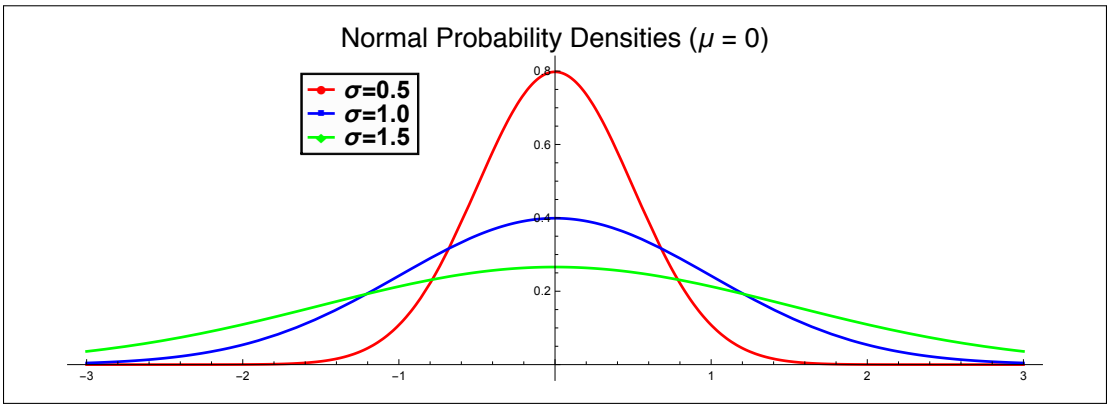
The cdf of the standard normal is often denoted by Φ. That is,

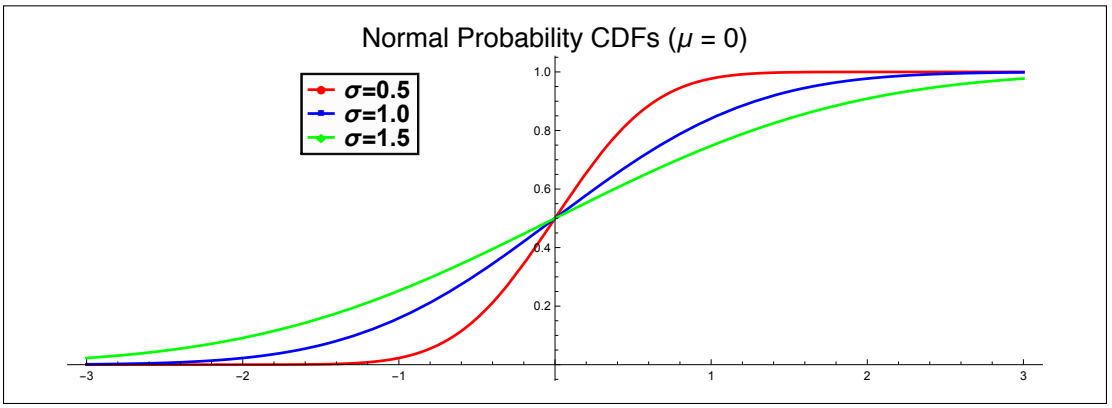
$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2}\, dx.$$

There is no closed form solution to this integral in terms of elementary functions (polynomial, trigonometric, logarithm, exponential).









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

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

It is related to Φ by

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

or

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

7.6 The Normal Family

It is traditional to denote a standard normal random variable by the letter Z. **Pitman [6]:** p. 267

7.6.1 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**.

Proof: Let Z be a standard normal $N(0,1)$ random variable, and let $X = \sigma Z + \mu$, where $\sigma > 0$. Now

$$P(X \leqslant t) = P\left(Z \leqslant \frac{t-\mu}{\sigma}\right) = \Phi\left(\frac{t-\mu}{\sigma}\right),$$

so the density of X is the derivative (wrt t) of this, so

$$\begin{aligned} f_X(t) &= \frac{d}{dt}\Phi\left(\frac{t-\mu}{\sigma}\right) \\ &= \Phi'\left(\frac{t-\mu}{\sigma}\right)\frac{1}{\sigma} \\ &= f_Z\left(\frac{t-\mu}{\sigma}\right)\frac{1}{\sigma} \\ &= \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}, \end{aligned}$$

which is the $N(\mu, \sigma^2)$ density. ∎

Later on we shall prove the following.

7.6.2 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. See Pitman [6], page 363.

7.7 Mean and variance of a Normal

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

7.7.2 Corollary *The mean of a normal $N(\mu, \sigma^2)$ random variable X is μ and its variance is σ^2.*

Proof of Proposition 7.7.1:

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

by symmetry.

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

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

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

∎

The Corollary follows from Theorem 7.6.1 and the change of variables formula for integration.

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.

7.8 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 10 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

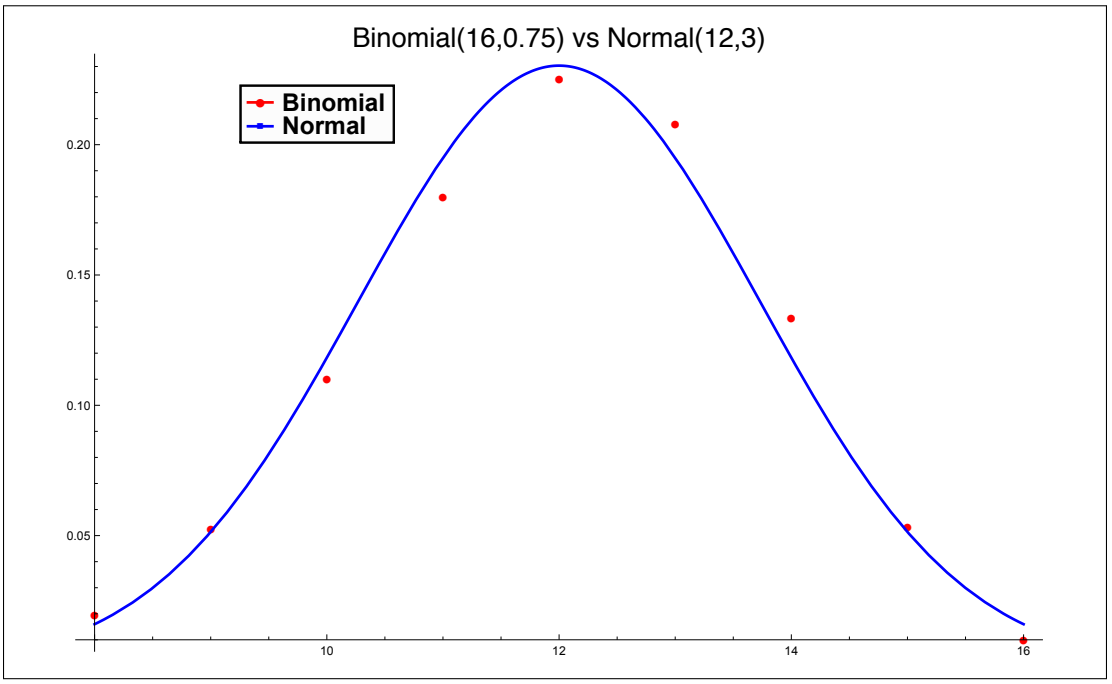
$$X^* = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{np(1-p)}}.$$

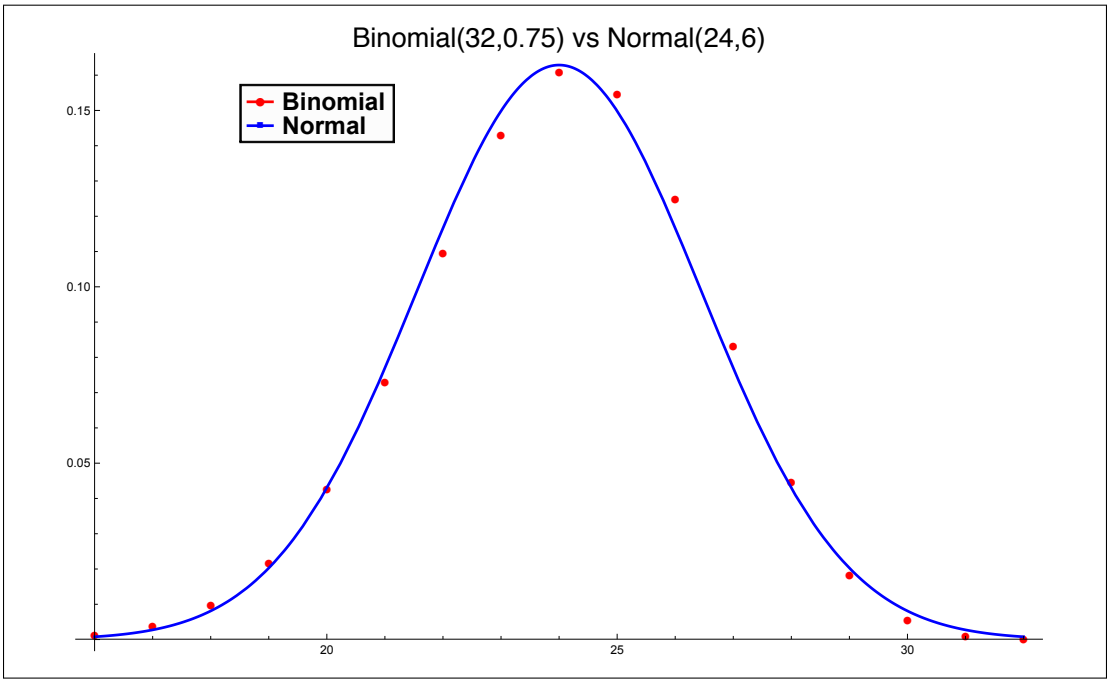
Also, $B_n(k) = P(X = k)$, the probability of k successes, is given by

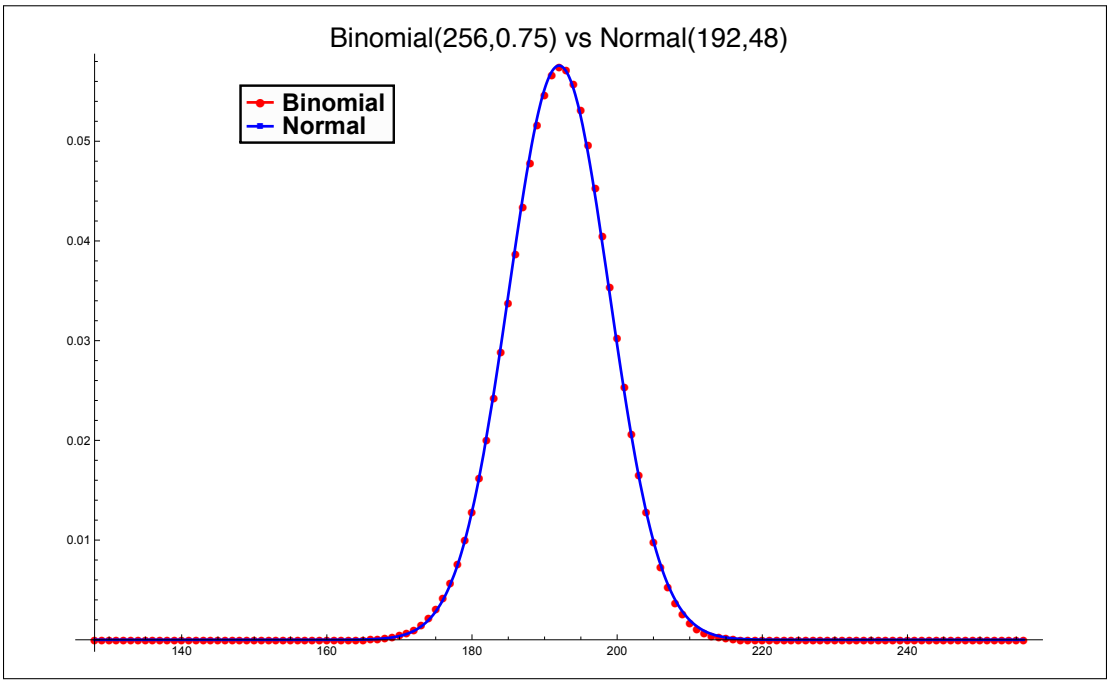
$$B_n(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

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,

$$\lim_{n\to\infty} \left| B_n(k) - \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(k-np)^2}{2\sigma_n^2}} \right| = 0. \tag{4}$$

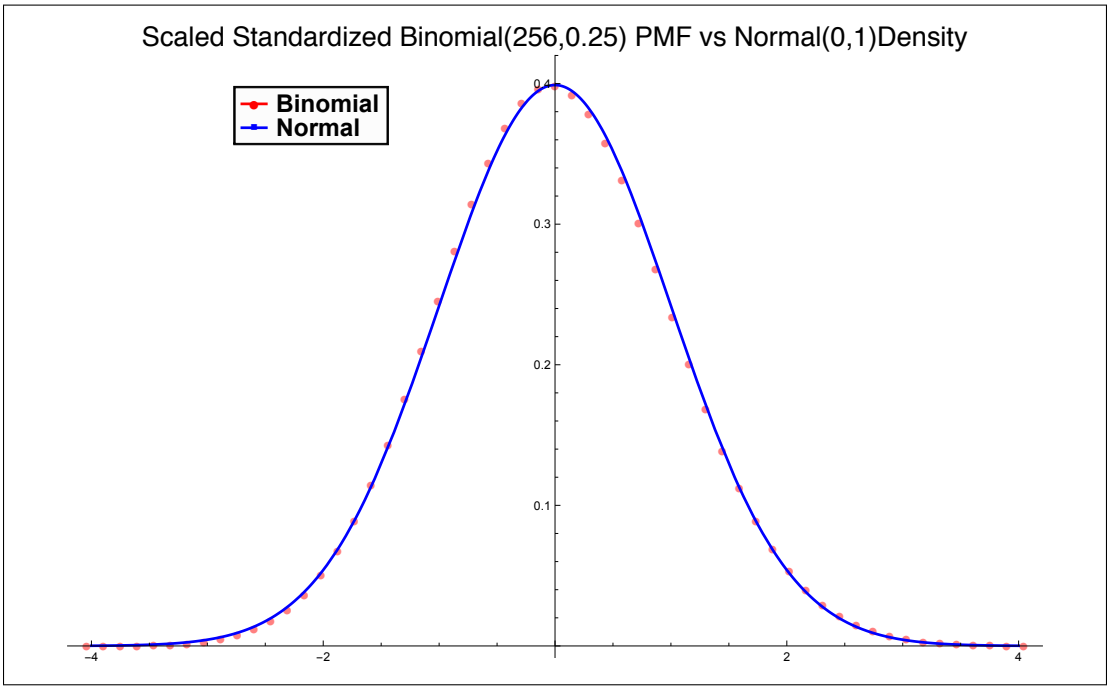

Binomial(16,0.75) vs Normal(12,3)
Binomial
Normal
0.20
0.15
0.10
0.05
10
12
14
16


Binomial(32,0.75) vs Normal(24,6)
Binomial
Normal
0.15
0.10
0.05
20
25
30



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 (4) as follows: For each $z = (k - np)/\sigma_n$ we have $\kappa(z) = k$, so

$$\lim_{n\to\infty} \left| \sigma_n B_n\Big(\kappa(z)\Big) - \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \right| = 0 \tag{5}$$



7.9 DeMoivre–Laplace Limit Theorem

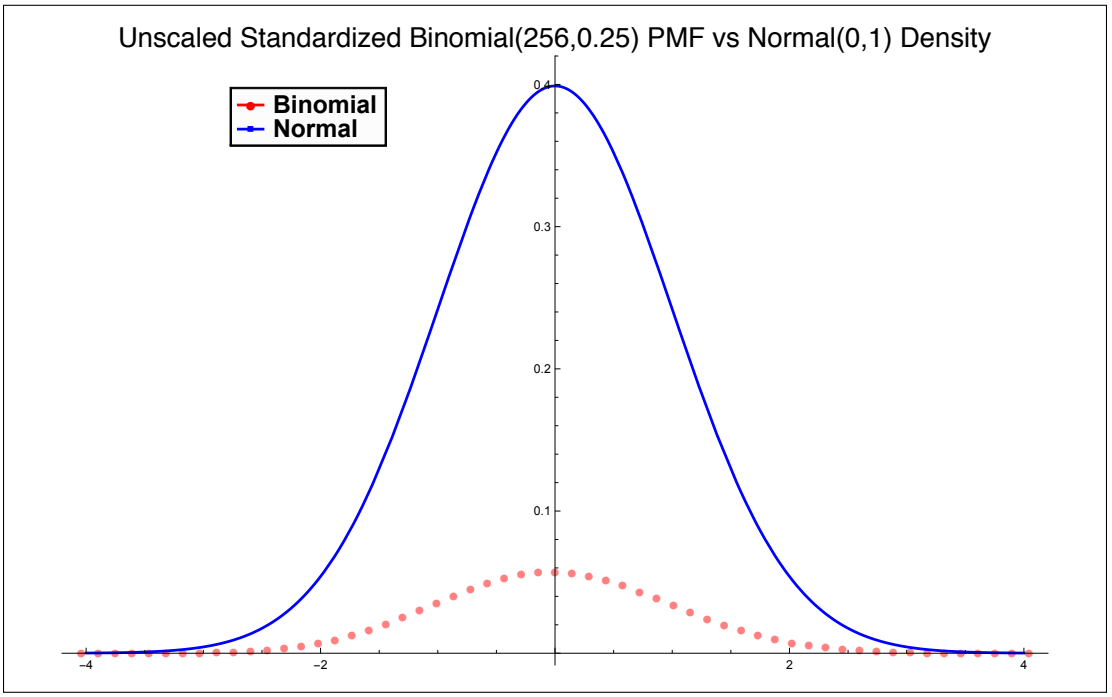
The normal approximation to the Binomial can be rephrased as:

7.9.1 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 [5, Theorem 4.3.1, pp. 239–240] or Pitman [6, Section 2.1, esp. p. 99]. In practice, this approximation requires that $9 \geqslant \max\{p/(1-p)p, (1-p)/p\}$.

Aside: Note the scaling of B in (5). 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 $f(z) \approx \sigma p_{n,p}(k)$.

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

$$X^* = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{np(1-p)}} = \frac{f - p}{\sqrt{p(1-p)}}\sqrt{n}. \tag{6}$$

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

7.10 Using the Normal approximation

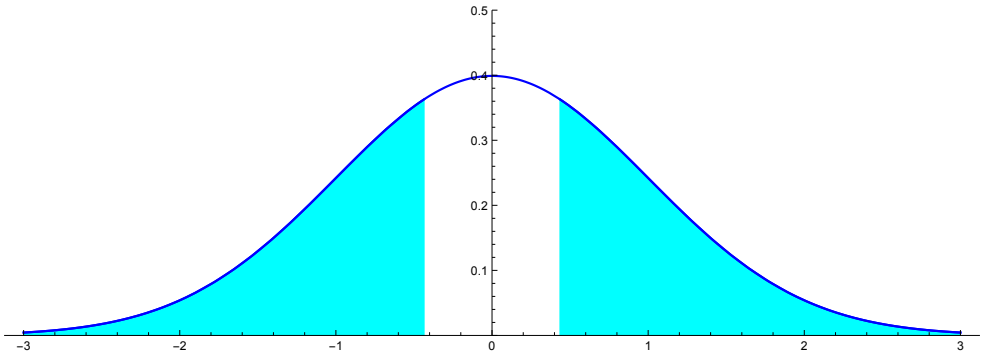
The homework asks you to look at the results of the coin tossing experiment to see if the results are consistent with $P(\text{Heads}) = 1/2$. As of the writing of the assignment, the results were 13,660 Tails out of 27,392 tosses, or 49.8686%. (New submissions have since arrived.) Is this "close" to 1/2?

The Binomial$(27392, 1/2)$ random variable has expectation $\mu = 13696$ and standard deviation $\sigma = 89.74$. So assuming the coin is fair the value of the standardized result is $(13660 - 13696)/89.74 = -0.435$.

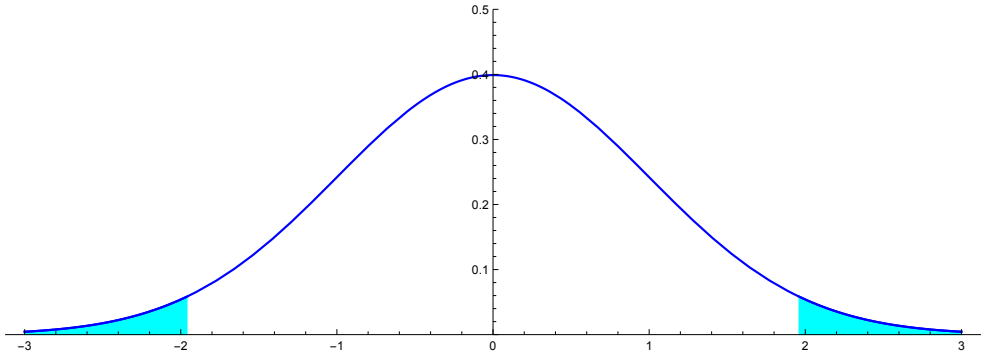
We can use the DeMoivre–Laplace Limit Theorem to treat X^* as a Standard Normal (mean = 0, variance = 1).

We now ask what the probability is that a Standard Normal Z takes on a value outside the interval $(-0.435, 0.435)$. 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, $\Phi(-0.435) + 1 - \Phi(0.435)$.



This figure shows the Standard Normal density. For a Standard Normal random variable Z, the shaded area is the probability of the event $(|Z| > 0.435)$, which is equal to $\Phi(-0.435) + 1 - \Phi(0.435)$. (In the homework, you are asked to compute this numerically.) The smaller this probability, the less likely it is that the coins are fair, and the result was gotten "by chance."



This shaded area is the probability of the event $(|Z| > 1.96)$, which is equal to $\Phi(-1.96) + 1 - \Phi(1.96)$. (In the homework, you are also asked to compute this numerically.) Values outside the interval $(-1.96, 1.96)$ are often regarded as unlikely to have occurred "by chance."

7.11 Sample size and the Normal approximation

If X is a Binomial(n, p) random variable, and let $f = X/n$. The DeMoivre–Laplace Theorem and equation (6) tell us that for large n, we have the following approximation

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

We can use this to calculate the probability that f is "close" to p.

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 closeness,
- let $1 - \alpha$ designate what we mean by "high" probability
- We want to choose n big enough so that

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

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 that

$$P\left(\frac{|f - p|}{\sqrt{p(1-p)}}\sqrt{n} > \frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n}\right) \approx P\left(|Z| > \frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n}\right), \tag{7}$$

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

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

Define the function $\zeta(t)$ by

$$P\left(Z > \zeta(t)\right) = t,$$

or equivalently

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

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

$$P\left(|Z| > \zeta(t)\right) = 2t.$$

So by (7) we want to find n such that

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

or in other words, find n so that

$$\begin{aligned}\zeta(\alpha/2) &= \frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n}\\ \zeta^2(\alpha/2) &= n\frac{\varepsilon^2}{p(1-p)}\\ n &= \frac{\zeta^2(\alpha/2)p(1-p)}{\varepsilon^2}.\end{aligned}$$

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 n is large enough, regardless of the value of p, we only need to choose n so that

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

where

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

Let's take some examples:

ε	α	$\zeta(\alpha/2)$	n
0.05	0.05	1.96	385
0.03	0.05	1.96	1068
0.01	0.05	1.96	9604
0.05	0.01	2.58	664
0.03	0.01	2.58	1844
0.01	0.01	2.58	16588

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] I. Kaplansky. 1944. Symbolic solution of certain problems in permutations. *Bulletin of the American Mathematical Society* 50(12):906–914.
http://projecteuclid.org/euclid.bams/1183506627

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

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

[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.