Lecture 7: Introducing the Normal Distribution

Relevant textbook passages:
Pitman [5]: Sections 1.2, 2.2, 2.4

7.1 Standardized random variables

7.1.1 Definition Given a random variable $X$ with finite mean $\mu$ and variance $\sigma^2$, the standardization of $X$ is the random variable $X^\ast$ defined by

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

Note that $E X^* = 0$, and $\text{Var} X^* = 1$,

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 $X_\ast$ to denote the standardization of $X$.]

A convenient feature of standardization is that they are invariant under multiplication by positive scalars.

7.1.2 Proposition Let $X$ be a random variable with mean $\mu$ and standard deviation $\sigma$, and let $Y = aX$, where $a > 0$. Then

$$
X^* = Y^*.
$$

Proof: The proof follows from Propositions 6.2.1 and 6.5.1, which assert that $E Y = a\mu$ and s.d. $Y = a\sigma$. So

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

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.

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

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

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

The cdf of the standard normal is often denoted by $\Phi$. 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).

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

7.2.2 Proposition

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

The proof is an exercise in integration theory. See for instance, Pitman [5, pp. 358–359].
7.2.3 Proposition The mean of a normal \( N(0,1) \) random variable \( Z \) is 0 and its variance is 1.

Proof of Proposition 7.2.3:

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

by symmetry.

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

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

\[
\text{Var} Z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left( -z \right) \left( -ze^{-z^2/2} \right) \, dz
\]

\[
= \frac{1}{\sqrt{2\pi}} \left( \left. \left( -ze^{-z^2/2} \right) \right|_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} e^{-z^2/2} \, dz \right)
\]

\[
= \frac{1}{\sqrt{2\pi}} \left( -\lim_{z \to -\infty} \left( -ze^{-z^2/2} \right) - \int_{-\infty}^{\infty} e^{-z^2/2} \, dz \right) = 1.
\]

The error function

A function closely related to \( \Phi \) that is popular in error analysis in some of the sciences is the error function, \( \text{erf} \), denoted \( \text{erf} \) defined by

\[
\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2} \, dt, \quad x \geq 0.
\]

It is related to \( \Phi \) by

\[
\Phi(z) = \frac{1}{2} + \frac{1}{2} \text{erf} \left( \frac{z}{\sqrt{2}} \right),
\]

or

\[
\text{erf}(z) = 2\Phi(z\sqrt{2}) - 1.
\]
The “tails” of a standard normal

While we cannot explicitly write down \( \Phi \), we do know something about the “tails” of the cdf. Pollard [6, p. 191] provides the following useful estimate of \( 1 - \Phi(x) \), the probability that standard normal takes on a value at least \( x > 0 \):

7.2.4 Fact

\[
\left( \frac{1}{x} - \frac{1}{x^3} \right) \frac{\exp\left(-\frac{1}{2}x^2\right)}{\sqrt{2\pi}} \leq 1 - \Phi(x) \leq \frac{1}{x} \frac{\exp\left(-\frac{1}{2}x^2\right)}{\sqrt{2\pi}} \quad (x > 0).
\]

So a reasonable simple approximation of \( 1 - \Phi(x) \) is just

\[
1 - \Phi(x) \approx \frac{1}{x} \frac{\exp\left(-\frac{1}{2}x^2\right)}{\sqrt{2\pi}},
\]

which is just \( \frac{1}{x} \) times the density at \( x \). The error is at most \( \frac{1}{x^3} \) of this estimate.

7.3 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 \( \mu \) and \( \sigma \). 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 6.1 and Proposition 6.5.1 that

\[
E X = \mu, \quad \text{Var} X = \sigma^2.
\]

What is the density of \( X \)? Well,

\[
\text{Prob } X \leq x = \text{Prob } \sigma Z + \mu \leq x = \Phi\left( x = \mu \right) / \sigma = \Phi \left( x = \mu / \sigma \right).
\]

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

\[
\frac{d}{dx} \Phi \left( \frac{x - \mu}{\sigma} \right) = \Phi' \left( \frac{x - \mu}{\sigma} \right) \frac{d}{dx} \left[ \frac{x - \mu}{\sigma} \right] = f \left( \frac{x - \mu}{\sigma} \right) \frac{1}{\sigma},
\]

where \( f \) is the derivative of \( \Phi \), that is, the standard normal density. Thus the density of \( X \) at \( x \) is

\[
\frac{1}{\sqrt{2\pi} \sigma} e^{\frac{(x - \mu)^2}{2\sigma^2}}
\]

This density is called the Normal \( N(\mu, \sigma^2) \) density. Its integral is the Normal \( N(\mu, \sigma^2) \) cdf. We say that \( X \) has a Normal \( N(\mu, \sigma^2) \) distribution. As we observed above, the mean of a \( N(\mu, \sigma^2) \) random variable is \( \mu \), and its variance is \( \sigma^2 \).

\(^1\)According to B. L. van der Waerden [7, p. 11], Gauss assumed this density represented the distribution of errors in astronomical data.
We have just proved the following:

**7.3.1 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.

Later on we shall prove the following.

**7.3.2 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. See Pitman [5], page 363.

**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.

### 7.4 The Binomial$(n, p)$ and the Normal$(np, np(1 - p))$

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 $N(np, \sigma^2_n)$ density was a good approximation to $B_n(k)$. In fact, for each $k$,

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

---

---
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

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

(2)
7.5 DeMoivre–Laplace Limit Theorem

The normal approximation to the Binomial can be rephrased as:

**7.5.1 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 $a, b$,

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


**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 \(z = (k - \mu)/\sigma\) is \(f(z) = \frac{1}{\sqrt{2\pi}} z^{-2}/2\), but the area under the density is approximated by the area under a step function, where the steps are centered at the points \((k - \mu)/\sigma\) and have width \(1/\sigma\). Thus each \(z = (k - \mu)/\sigma\) 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 \(f(z) \approx \sigma p_{n,p}(k)\).

**7.5.2 Remark** We can rewrite the standardized random variable \(X^\ast\) in terms of the average frequency of success, \(f = X/n\). Then

\[
X^\ast = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{np(1 - p)}} = \frac{f - p}{\sqrt{p(1 - p)}} \sqrt{n}.
\]

This formulation is sometimes more convenient. For the special case \(p = 1/2\), this reduces to

\[
X^\ast = 2(f - p)\sqrt{n}.
\]

**7.6 Using the Normal approximation**

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 were 11,748 Tails out of 23,552 tosses, or 49.88%. Is this “close enough” to 1/2?

The Binomial(23552, 1/2) random variable has expectation \(\mu = 11,776\) and standard deviation \(\sigma = 76.73\). So assuming the coin is fair the value of the standardized result is \((11748 - 11776)/76.73 = -0.365\). This value called the z-score of the experiment.

We can use the DeMoivre–Laplace Limit Theorem to treat \(X^\ast\) 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 \((-0.365, 0.365)\). 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, $\Phi(-0.365) + 1 - \Phi(0.365) = 0.715$. This value is called the **p-value** of the experiment. It means that if the coins are really “fair,” then $71.5\%$ 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 figure shows the Standard Normal density. For a Standard Normal random variable $Z$, the shaded area is the probability of the event $(|Z| > 0.365)$, which is equal to $\Phi(-0.365) + 1 - \Phi(0.365) = 0.715$. (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 $\Phi(-1.96) + 1 - \Phi(1.96) = 0.05$. Values outside the interval $(-1.96, 1.96)$ are often regarded as unlikely to have occurred “by chance.”
The exact $p$-value

The $p$-value above was derived by treating the standardized Binomial(23552, 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 exact $p$-value which is given by

$$2\sum_{k=0}^{t} \left( \begin{array}{c} 23552 \\ k \end{array} \right) \left( \frac{1}{2} \right)^{23552},$$

where $t$ is the smaller of the number of Tails and Heads. (Why?) Mathematica 10 reports this value to be 0.720.

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.

Four years

What if I take the data from four years of this class? There were 55,515 Tails in 110,848 tosses which gives a $z$-score of 0.547 for a $p$-value of 0.584. The exact $p$-value is 0.587. But the exact value is much more time consuming to compute. For this year’s data, it took Mathematica 10.3 running on my Early-2009 floor-top Mac Pro, just 3.3 seconds, but for four years worth of tosses it took 106 seconds. (Before I optimized my code, it took 288 seconds.)

7.7 Sample size and the Normal approximation

If $X$ is 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 7–9) 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 $\alpha$ is a “low” probability).
- We want to choose $n$ big enough so that

$$P \left( |f - p| > \varepsilon \right) \leq \alpha. \quad (4)$$

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

$$P \left( \frac{|f - p|}{\sqrt{p(1-p)}} \sqrt{n} > \frac{\varepsilon}{\sqrt{p(1-p)}} \sqrt{n} \right) \approx P \left( |Z| > \frac{\varepsilon}{\sqrt{p(1-p)}} \sqrt{n} \right), \quad (5)$$
where $Z \sim N(0, 1)$. Thus

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

Define the function $\zeta(a)$ by

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

or equivalently

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

where $\Phi$ is the Standard Normal cumulative distribution function. See Figure 7.1. This is something you can look up with R or Mathematica’s built-in quantile functions. By symmetry,

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

Figure 7.1. Given $\alpha$, $\zeta(\alpha)$ is chosen so that the shaded region has area $= \alpha$.

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

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

or in other words, find $n$ so that

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

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

$$n = \frac{\zeta^2(\alpha/2)p(1 - p)}{\epsilon^2}.$$
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 \( n \) is large enough, regardless of the value of \( p \), we only need to choose \( n \) so that

\[
n \geq \frac{\zeta^2(\alpha/2)}{4\varepsilon^2},
\]

where

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

Let’s take some examples:

<table>
<thead>
<tr>
<th>( \varepsilon )</th>
<th>( \alpha )</th>
<th>( \zeta(\alpha/2) )</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>1.96</td>
<td>385</td>
</tr>
<tr>
<td>0.03</td>
<td>0.05</td>
<td>1.96</td>
<td>1068</td>
</tr>
<tr>
<td>0.01</td>
<td>0.05</td>
<td>1.96</td>
<td>9604</td>
</tr>
<tr>
<td>0.05</td>
<td>0.01</td>
<td>2.58</td>
<td>664</td>
</tr>
<tr>
<td>0.03</td>
<td>0.01</td>
<td>2.58</td>
<td>1844</td>
</tr>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>2.58</td>
<td>16588</td>
</tr>
</tbody>
</table>

**Bibliography**


