Lecture 15: Normal Distributions and Central Limit Theorem

Contents

1 Normal distribution: Review from last lecture

Last time we discussed the Normal distribution.

Definition 1 *A continuous r.v. X is said to be* **Normal(μ, σ^2)** *or* **Gaussian(μ, σ^2)** *if it has p.d.f. $f_X(x)$ of the form:*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < x < \infty$$

where $\sigma > 0$. The parameter μ is called the **mean** *and the parameter σ is called the* **standard deviation**.

Definition 2 *A Normal$(0, 1)$ r.v. Y is said to be a* **standard Normal**. *Its c.d.f. is denoted by*

$$\Phi(y) = F_Y(y) = \mathbf{P}\{Y \le y\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-t^2/2} dt$$

The Normal(μ, σ^2) p.d.f. has a "bell" shape, and is clearly symmetric around μ, as shown in Figure 1.

We proved the following theorem, which explains the significance of the parameters of the Normal distribution:

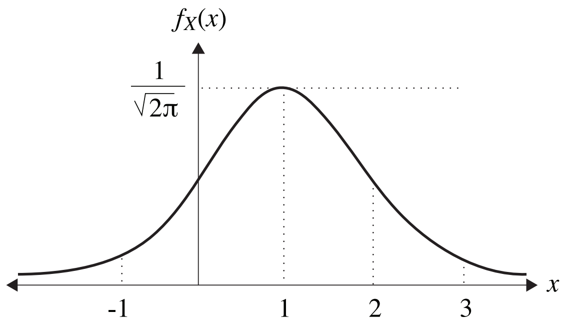

Figure 1: *Normal p.d.f. with mean* $\mu = 1$ *and variance* $\sigma^2 = 1$.

Theorem 3 *Let* $X \sim$ *Normal*(μ, σ^2), *then* $\mathbf{E}\{X\} = \mu$ *and* $\mathbf{Var}(X) = \sigma^2$.

We also proved the Linear transformation property for the Normal distribution, which says that you can multiply a Normal by a scalar and/or shift a Normal left or right and still end up with a Normal distribution.

Theorem 4 *Let* $X \sim$ *Normal*(μ, σ^2). *Let* $Y = aX + b$, *where* $a > 0$ *and* b *are scalars. Then* $Y \sim$ *Normal*$(a\mu + b, a^2\sigma^2)$.

The Linear transformation property is important because it allows us to compute properties of any Normal(μ, σ^2) distribution by first translating the distribution into the (standard) Normal$(0, 1)$ distribution and computing properties of that, and then translating back the results.

Question: Suppose that $X \sim$ Normal(μ, σ^2). How can I define Y in terms of X so that $Y \sim$ Normal$(0, 1)$?

Answer: Let $Y = \frac{X-\mu}{\sigma}$.

We will discuss many applications of this, but let us first see what we know for a standard Normal.

2 The standard Normal

Consider the standard Normal$(0, 1)$ distribution with p.d.f. $f(t)$ and c.d.f. $\Phi(y)$, where

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

and where

$$\Phi(y) = \int_{-\infty}^{y} f(t)dt$$

Unfortunately, $\Phi(y)$ is not known in closed form. That is, we don't know how to integrate $f(t)$ from 0 to y symbolically. We must therefore use a table of numerically integrated results for $\Phi(y)$, such as that given in [2][1]. A subset of the table is given below for reference:

y	$\Phi(y)$
0.5	0.6915
1.0	0.8413
1.5	0.9332
2.0	0.9772
2.5	0.9938
3.0	0.9987

Question: Looking at the table you will see, for example, that $\Phi(1) = 0.8413$. What does this tell us about the probability that the standard Normal is within one standard deviation of its mean?

Answer: We are given that $\mathbf{P}\{Y < 1\} = 0.84$. We want to know $\mathbf{P}\{-1 < Y < 1\}$.

$$\begin{aligned}
\mathbf{P}\{-1 < Y < 1\} &= \mathbf{P}\{Y < 1\} - \mathbf{P}\{Y < -1\} \\
&= \mathbf{P}\{Y < 1\} - \mathbf{P}\{Y > 1\} \quad \text{(by symmetry)} \\
&= \mathbf{P}\{Y < 1\} - (1 - \mathbf{P}\{Y < 1\}) \\
&= 2\mathbf{P}\{Y < 1\} - 1 \\
&= 2\Phi(1) - 1 \\
&= 2 \cdot 0.84 - 1 \\
&= 0.68
\end{aligned}$$

So with probability 68%, we are within one standard deviation of the mean.

Likewise, we can use the same argument to show that with probability 95%, we are within two standard deviations of the mean, and with probability 99.7%, we are within three standard deviations of the mean, etc.

[1] In practice no one ever goes to the table anymore, since there are approximations that allow you to compute the values in the table to within 7 decimal places, see for example [1].

3 Converting between Normal(μ, σ^2) and standard Normal

Question: The results of the last section were expressed for a standard Normal. What if we don't have a standard Normal?

Answer: We can convert a non-standard Normal into a standard Normal using the Linear Transformation Property. Here's how it works:

$$X \sim \text{Normal}(\mu, \sigma^2) \iff Y = \frac{X - \mu}{\sigma} \sim \text{Normal}(0, 1)$$

So

$$\mathbf{P}\{X < k\} = \mathbf{P}\left\{\frac{X - \mu}{\sigma} < \frac{k - \mu}{\sigma}\right\} = \mathbf{P}\left\{Y < \frac{k - \mu}{\sigma}\right\} = \Phi\left(\frac{k - \mu}{\sigma}\right)$$

Theorem 5 *If $X \sim$ Normal(μ, σ^2), then the probability that X deviates from its mean by less than k standard deviations is the same as the probability that the standard Normal deviates from its mean by less than k.*

Proof: Let $Y \sim \text{Normal}(0, 1)$. Then:

$$\mathbf{P}\{-k\sigma < X - \mu < k\sigma\} = \mathbf{P}\left\{-k < \frac{X - \mu}{\sigma} < k\right\} = \mathbf{P}\{-k < Y < k\}$$

■

The above theorem illustrates why it's often easier to think in terms of standard deviations rather than absolute values.

Question: Proponents of IQ testing will tell you that human intelligence (IQ) has been shown to be Normally distributed with mean 100 and standard deviation 15. What fraction of people have IQ greater than 130 ("the gifted cutoff"), given that $\Phi(2) = 0.9772$?

Answer: We are looking for the fraction of people whose IQ is more than two standard deviations *above* the mean. This is the same as probability that the standard normal exceeds its mean by more than two standard deviations, which is $1 - \Phi(2) = 0.023$. Thus only about 2% of people have an IQ above 130.

4 Central limit theorem

Consider sampling the heights of all the individuals in the state and take that average. The Central limit theorem (CLT), which we will define soon, says that this average will tend to be Normally distributed. This would be true even if we took the average of a large number of i.i.d. random variables, where the random variables come from a distribution which is decidedly non-Normal, say a Uniform distribution. It is this property that makes the Normal distribution so important!

We now state this more formally. Let $X_1, X_2, X_3, \ldots, X_n$ be independent and identically distributed r.v.'s with some mean μ and variance σ^2. Note: We're *not* assuming that these are Normally-distributed r.v.'s. In fact we're not even assuming that these are necessarily continuous r.v.'s – they may be discrete r.v.'s.

Let

$$S_n = X_1 + X_2 + \cdots + X_n \tag{1}$$

Question: What is the mean and standard deviation of S_n?

Answer: $\mathbf{E}\{S_n\} = n\mu$ and $\mathbf{Var}(S_n) = n\sigma^2$. Thus the standard deviation is $\sigma\sqrt{n}$.

Let

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Question: What is the mean and standard deviation of Z_n?

Answer: Z_n has mean 0 and standard deviation 1.

Theorem 6 (Central Limit Theorem (CLT)) *Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. r.v.'s with common mean μ and finite variance σ^2, and define:*

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

Then the c.d.f. of Z_n converges to the standard normal c.d.f., i.e.,

$$\lim_{n\to\infty} \mathbf{P}\{Z_n \le z\} = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

for every z.

Proof: Most proofs are pretty tough. We have a very nice proof which makes use of Laplace transforms, which we unfortunately haven't covered yet – Laplace transforms are a transform which applies to continuous distributions in a similar way to the z-transform which applied to discrete distributions. We will discuss the proof of CLT after we cover Laplace transforms. ∎

Question: What is the distribution of S_n in (1)?

Answer: By the Linear Transformation Property, $S_n \sim \text{Normal}(n\mu, n\sigma^2)$.

The Central Limit Theorem is extremely general and explains many natural phenomena that result in Normal distributions. The fact that CLT applies to any sum of i.i.d. r.v.'s allows us to prove that the Binomial(n, p) distribution, which is a sum of i.i.d. Bernoulli(p) r.v.'s, converges to a Normal distribution when n is high.

We now illustrate the use of the Normal distribution in approximating the distribution of a complicated sum.

Example: Normal approximation of a sum

Imagine that we are trying to transmit a signal. During the transmission, there are a hundred sources independently making low noise. Each source produces an amount of noise which is Uniformly distributed between $a = -1$ and $b = 1$. If the total amount of noise is greater than 10 or less than -10, then it corrupts the signal. However if the absolute value of the total amount of noise is under 10, then it's not a problem.

Question: What is the approximate probability that the absolute value of the total amount of noise from the 100 signals is less than 10?

Answer: Let X_i be the noise from source i. Observe that $\mu_{X_i} = 0$. Observe that $\sigma^2_{X_i} = \frac{(b-a)^2}{12} = \frac{1}{3}$ and $\sigma_{X_i} = \frac{1}{\sqrt{3}}$. Let $S_{100} = X_1 + X_2 + \cdots + X_{100}$.

$$\begin{aligned} \mathbf{P}\{-10 < S_{100} < 10\} &= \mathbf{P}\left\{\frac{-10}{\sqrt{100/3}} < \frac{S_{100} - 0}{\sqrt{100/3}} < \frac{10}{\sqrt{100/3}}\right\} \\ &\approx 2\Phi\left(\sqrt{3}\right) - 1 \\ &= 2(0.9572) - 1 \\ &= 0.9144 \end{aligned}$$

Hence the approximate probability of the signal getting corrupted is less than 10%. In practice, this approximation is excellent.

References

[1] Normal Table. Computing the Normal. http://www.math.ucla.edu/~cbm/aands//page_299.htm.

[2] Daniel Zwillinger. *CRC Standard Mathematical Tables and Formulae*. Chapman & Hall, 31st edition, 2003.