Proving the central limit theorem

Gareth Tribello
July 7, 2017

1 Purpose

In the lectures and exercises we have learnt about the law of large numbers and the central limit theorem. We have largely focussed on how these theorems are used in statistics and have not worried too much about how these theorems are proved. We have done so because for statisticians knowing how to use these ideas is much more important than being able to reproduce these very standard proofs. Proving these theorems is not so difficult, however, and as such the following exercise explains how to go about the proof for those who are interested.

2 Proof of the weak law of large numbers

One reason why expectation values are useful is that they allow us to simplify the information contained in the probability distribution. Two particularly important examples of this in practise are explained below.

2.1 Markov Inequality

Notice that we can write the expectation value of a discrete random variable as:

$$\mathbb{E}(X) = \sum_{x_i} x_i f_X(x_i) = \sum_{x_i < a} x_i f_X(x_i) + \sum_{x_i \geq a} x_i f_X(x_i) \tag{2.1}$$

Alternatively, if the random variable is continuous we can write:

$$\mathbb{E}(X) = \int_{-\infty}^{a} x f_X(x) \mathrm{d}x + \int_{a}^{+\infty} x f_X(x) \mathrm{d}x \tag{2.2}$$

If $X \geq 0$ and $a > 0$ what does this imply about the following:

$$\begin{aligned} \mathbb{E}(X) - \sum_{x_i \geq a} x_i f_X(x_i) &= \\ \mathbb{E}(X) - \int_{a}^{\infty} x f_X(x) \mathrm{d}x &= \end{aligned} \tag{2.3}$$

Given that $x \geq 0$ and the fact that $f_X(x)$ is a probability density what can we immediately state about the expressions on the right had side of the equals signs above:

The relationship we have just derived is known as the Markov inequality.

2.2 Chebyshev Inequality

If $|X| \geq a$ then $X^2 \quad a^2$. What does this allow us to say about:

$$P(|X| \geq a) \qquad P(X^2 \geq a^2) \tag{2.4}$$

Given the Markov inequality derived in the previous section what can we thus say about $\frac{\mathbb{E}(X^2)}{a^2}$?

In the equation we just derived replace X with $X - \mu$ (where μ is $\mathbb{E}(X)$). The resulting equation is the Chebyshev Inequality

2.3 Sums of random variables

At school you will have learnt that the sample mean, μ, is calculated using:

$$\mu = \frac{1}{N} \sum_{n=1}^{N} x_n \tag{2.5}$$

When we take a mean using this equation we assume that we have taken N samples $X_1, X_2, X_3, \dots$ from an underlying probability distribution. What is the expectation value for the quantity we calculate using equation 2.5?

We can derive something similar for the variance (although in this case we will divide it by n for reasons that will become clear in a moment).

$$\begin{aligned} \mathrm{var}\left(\frac{S_n}{n}\right) &= \frac{1}{n^2}\left(\mathbb{E}(S_n^2) - [\mathbb{E}(S_n)]^2\right) \\ &= \frac{1}{n^2}\left(\mathbb{E}[(X_1 + X_2 + X_3 + \cdots + X_n)^2] - [n\mathbb{E}(X)]^2\right) \end{aligned} \tag{2.6}$$

When we expand $(X_1 + X_2 + X_3 + \cdots + X_n)^2$ there are two types of terms. Terms like $\mathbb{E}(X_1^2)$ and terms like $\mathbb{E}(X_1 X_2)$. How many terms of each of these two types are there:

- Terms like $\mathbb{E}(X_1^2)$:
- Terms like $\mathbb{E}(X_1 X_2)$:

How can we rewrite $\mathbb{E}(X_1 X_2)$ if X_1 and X_2 are independent?

Given all this and the fact that the definition of the variance tells us that $\mathbb{E}(X^2) = \text{var}(X) + [\mathbb{E}(X)]^2$ show that:

$$\text{var}\left(\frac{S_n}{n}\right) = \frac{1}{n}\text{var}(X)$$

2.4 The weak law of large numbers

What does Chebyshev Inequality expression tell us above the value of (use what we know about the total variance $\text{var}(S_n)$ to help you:

$$P\left(\left|\frac{S_n}{N} - \mu\right| > \epsilon\right) \tag{2.7}$$

where S_n is the sum over all our samples from the distribution $S_n = X_1 + X_2 + X_3 + \dots$

What then is the value of the following limit:

$$\lim_{N\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \tag{2.8}$$

The expression we have just derived is called the *weak law of large numbers*. What is the significance of this expression?

3 Proof of central limit theorem

3.1 Moment generating function

The moment generating function is given by:

$$M_X(t) = \mathbb{E}(e^{tX}) \tag{3.1}$$

One reason the moment generating function is important because there is a one-to-one relationship between the probability mass/density functions and the moment generating functions. As such whatever we deduce to be true for the moment-generating function is also true for the corresponding probability mass/density function. We can thus prove the central limit theorem by demonstrating that the moment generating for the sample variance is the same as the moment generating function for a normal distribution.

The variance, σ^2, can be calculated using:

$$\sigma^2 = \int_0^1 (X - \mathbf{E}[X])^2 \mathrm{d}F_X(x) \tag{3.2}$$

Much as we did when we proved the law of large numbers we start by proving something about the variance, $\mathrm{var}(S)$, of a random variable that is formed by taking the sum of a series of independent and identically distributed random variables i.e. $S = X_1 + X_2 + X_3 + \cdots + X_n$. In particular we wish to prove that

$$\mathrm{var}(S) = \mathrm{var}(X_1 + X_2 + X_3 + \cdots + X_n) = n\mathrm{var}(X) \tag{3.3}$$

Use the space below to discuss how this proof is performed

3.2 Step 1: Moment generating function for a Gaussian

What integral do we have to perform to calculate the moment generating function for a Gaussian distribution with $\mathbb{E}(X) = 0$ and $\mathrm{var}(X) = 1$?

Expand and simplify the following expression $\frac{1}{2}t^2 - \frac{1}{2}(x-t)^2$

Hence, by using the equality that you have arrived at above and the substitution $y = x - t$, show that the moment generating function for a Gaussian with $\mathbb{E}(X) = 0$ and $\text{var}(X) = 1$ is $M_X(t) = e^{t^2/2}$

3.3 Step 2: Moment generating function for the experiment

Lets introduce the following random variable:

$$Z = \frac{(S_n/n) - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{n\sigma/\sqrt{n}} = \frac{\sum_i X_i - \mu}{\sigma\sqrt{n}} \tag{3.4}$$

where S_n is the sum over all our samples from the distribution $S_n = X_1 + X_2 + X_3 + \dots$ and σ is the sample variance. Show that this random variable has the following useful properties:

$$\mathbb{E}(Z) = \mathbb{E}\left(\frac{(S_n/n) - \mu}{\sigma/\sqrt{n}}\right) = 0$$

Now show that:

$$\text{var}\,[Z] = \frac{\mathbb{E}(S_n^2)}{n\sigma^2} - \frac{2\mathbb{E}(S_n)\mu}{\sigma^2} + \frac{n\mu^2}{\sigma^2} \tag{3.5}$$

Use the fact from earlier that $\mathrm{var}(S_n) = n\sigma^2$ to show that:

$$\frac{\mathbb{E}(S_n^2)}{n\sigma^2} - \frac{2\mathbb{E}(S_n)\mu}{\sigma^2} + \frac{n\mu^2}{\sigma^2} = 1 \tag{3.6}$$

We now note that the moment generating function of Z is:

$$M_z(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\left[\exp\left(\sum_{i=1}^{n} \frac{t}{\sigma\sqrt{n}}(X_i - \mu)\right)\right] \tag{3.7}$$

How can we simplify this expression by using (1) the properties of a exponential of a sum of terms and (2) the fact that all our experiments are independent.

Through the above manipulations you arrive at a function of:

$$\exp\left(\frac{t(X-\mu)}{\sigma\sqrt{n}}\right) \tag{3.8}$$

Make a Maclaurin expansion for the exponential function above

Substitute this expansion into the equation for the expectation and simplify using the linearity of the expectation operator.

Now substitute the sum of exponentials that you arrived at above into the expression that we arrived at for the moment generating function. Note that in the limit as $n \to \infty$ all terms of higher order than t^2 in your Maclaurin expansion are zero. Furthermore, recall the Euler definition of the exponential function:

$$e^x = \lim_{n\to\infty} \left[1 + \frac{x}{n}\right]^n \tag{3.9}$$

Use these results to show that the Moment generating function for the random variable Z is nothing more than $e^{t^2/2}$ as doing so will prove that the probability distribution function for Z is a Normal distribution centred on 0 and with variance 1.