STA258H5

Al Nosedal
and Alison Weir

Winter 2017

THE CENTRAL LIMIT THEOREM

Theorem 7.5

Let Y and $Y_1, Y_2, Y_3, \ldots$ be random variables with moment-generating functions $M(t)$ and $M_1(t), M_2(t), \ldots$, respectively. If

$$\lim_{n \to \infty} M_n(t) = M(t) \qquad \text{for all real } t,$$

then the distribution function of Y_n converges to the distribution function of Y as $n \to \infty$.

Maclaurin Series

A Maclaurin series is a Taylor series expansion of a function about 0,

$$f(x) = f(0) + f^{'}(0)x + \frac{f^{''}(0)x^2}{2!} + \frac{f^3(0)x^3}{3!} + ... + \frac{f^n(0)x^n}{n!} + ...$$

Maclaurin series are named after the Scottish mathematician Colin Maclaurin.

Useful properties of MGFs

- $M_Y(0) = E(e^{0Y}) = E(1) = 1$.
- $M'_Y(0) = E(Y)$.
- $M''_Y(0) = E(Y^2)$.
- $M_{aY}(t) = E(e^{t(aY)}) = E(e^{(at)Y}) = M_Y(at)$, where a is a constant.

Let $Y_1, Y_2, ..., Y_n$ be independent and identically distributed random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2 < \infty$. Define

$$U_n = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.
Then the distribution function of U_n converges to the standard Normal distribution function as $n \to \infty$. That is,

$$\lim_{n\to\infty} P(U_n \leq u) = \int_{\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \text{ for all u.}$$

Proof

Let $Z_i = \frac{X_i - \mu}{\sigma}$. Note that $E(Z_i) = 0$ and $V(Z_i) = 1$. Let us rewrite U_n

$$\sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) = \sqrt{n}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{n\sigma}\right) = \frac{1}{\sqrt{n}}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma}\right)$$

$$U_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i.$$

Since the mfg of the sum of independent random variables is the product of their individual mfgs, if $M_{Z_i}(t)$ denotes the mgf of each random variable Z_i

$$M_{\sum Z_i}(t) = [M_{Z_1}(t)]^n$$

and

$$M_{U_n} = M_{\sum Z_i}(t/\sqrt{n}) = \left[M_{Z_1}(t/\sqrt{n})\right]^n.$$

Recall that $M_{Z_i}(0) = 1$, $M'_{Z_i}(0) = E(Z_i) = 0$, and $M''_{Z_i}(0) = E(Z_i^2) = V(Z_i^2) = 1$.

Now, let us write the Taylor's series of $M_{Z_i}(t)$ at 0

$$M_{Z_i}(t) = M_{Z_i}(0) + tM'_{Z_i}(0) + \frac{t^2}{2!}M''_{Z_i}(0) + \frac{t^3}{3!}M'''_{Z_i}(0) + ...$$

$$M_{Z_i}(t) = 1 + \frac{t^2}{2} + \frac{t^3}{3!}M'''_{Z_i}(0) + ...$$

$$M_{U_n}(t) = \left[M_{Z_1}(t/\sqrt{n})\right]^n = \left[1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{3/2}}M'''_{Z_i}(0) + ...\right]^n$$

Recall that if

$$\lim_{n \to \infty} b_n = b \qquad \lim_{n \to \infty} \left(1 + \frac{b_n}{n}\right)^n = e^b$$

But

$$\lim_{n \to \infty} \left[\frac{t^2}{2} + \frac{t^3}{3! n^{1/2}} M'''_{Z_i}(0) + ...\right] = \frac{t^2}{2}$$

Therefore,

$$\lim_{n \to \infty} M_{U_n}(t) = exp\left(\frac{t^2}{2}\right)$$

which is the moment-generating function for a standard Normal random variable. Applying Theorem 7.5 we conclude that U_n has a distribution function that converges to the distribution function of the standard Normal random variable.

Using the CLT

Two approximate distributions:

1. $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
2. $T = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$

Example

The number of accidents per week at a hazardous intersection varies with mean 2.2 and standard deviation 1.4. This distribution takes only whole-number values, so it is certainly not Normal.

a) Let $\bar{x}$ be the mean number of accidents per week at the intersection during a year (52 weeks). What is the approximate distribution of $\bar{x}$ according to the Central Limit Theorem?

b) What is the approximate probability that $\bar{x}$ is less than 2?

c) What is the approximate probability that there are fewer than 100 accidents at the intersection in a year?

Solution

a) By the Central Limit Theorem, $\bar{X}$ is roughly Normal with mean $\mu^* = 2.2$ and standard deviation $\sigma^* = \sigma/\sqrt{n} = 1.4/\sqrt{52} = 0.1941$.

b) $P(\bar{X} < 2) = P\left(\frac{\bar{X}-\mu^*}{\sigma^*} < \frac{2-2.2}{0.1941}\right)$
$= P(Z < -1.0303) = 0.1515$

Solution

Let X_i be the number of accidents during week i.

c) $P(Total < 100) = P\left(\sum_{i=1}^{52} X_i < 100\right)$

$$= P\left(\frac{\sum_{i=1}^{52} X_i}{52} < \frac{100}{52}\right)$$

$$= P\left(\bar{X} < 1.9230\right)$$

$$= P(Z < -1.4270) = 0.0768$$

An insurance company knows that in the entire population of millions of apartment owners, the mean annual loss from damage is $\mu = 75$ and the standard deviation of the loss is $\sigma = 300$. The distribution of losses is strongly right-skewed: most policies have $0 loss, but a few have large losses. If the company sells 10,000 policies, can it safely base its rates on the assumption that its average loss will be no greater than $85?

Solution

The Central Limit Theorem says that, in spite of the skewness of the population distribution, the average loss among 10,000 policies will be approximately $N(75; \frac{300}{\sqrt{10000}}) = N(75; 3)$. Now

$$P(\bar{X} > 85) = P\left(Z > \frac{85 - 75}{3}\right) = P(Z > 3.33) = 1 - 0.996 = 0.0004$$

We can be about 99.96% certain that average losses will not exceed \$85 per policiy.

Example

A freight elevator can transport a maximum of 9800 kg. Suppose a load of cargo containing 49 boxes must be transported via the elevator. Experience has shown that the weight of boxes of this type of cargo follows a distribution with mean 205 kg and standard deviation 15 kg. Based on this information, what is the probability that all 49 boxes can be safely loaded onto the freight elevator and transported?

Example

An insurance company audits health insurance claims in its database. They randomly select 400 items from the database (each item represents a dollar amount). Assume the population values are $\mu = 8$ dollars and $\sigma = 20$ dollars.

a. Find the probability that the sample mean would be less than \$ 6.50.

b. Do you think the population of overpayments has a Normal distribution. Why?

c. Why can we use the Normal distribution to answer part a)?

d. For what value of k does $P(\mu - k < X < \mu + k)$ equal 80%?