

Central limit theorem

Statement

Let x_1, x_2, \dots, x_n are n independent identically distributed random variables with same mean μ and standard deviation σ and if $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, then the variate $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ has a distribution that approaches the standard normal distribution as $n \rightarrow \infty$ provided the MGF of x_i exist.

Proof:

MGF of z about origin is $M_X(t) = E(e^{tz})$

$$= E \left[e^{t \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)} \right]$$

$$= E \left[e^{\frac{\sqrt{nt}}{\sigma} (\bar{x} - \mu)} \right]$$

$$= E \left[e^{\frac{\bar{x}\sqrt{nt}}{\sigma}} e^{-\frac{\mu\sqrt{nt}}{\sigma}} \right]$$

$$= e^{-\frac{\mu\sqrt{nt}}{\sigma}} E \left[e^{\frac{\bar{x}\sqrt{nt}}{\sigma}} \right]$$

$$= e^{-\frac{\mu\sqrt{nt}}{\sigma}} E \left[e^{\frac{\sqrt{nt}}{\sigma} \frac{1}{n} (x_1 + x_2 + \dots + x_n)} \right]$$

$$= e^{-\frac{\mu\sqrt{nt}}{\sigma}} E\left(e^{\frac{tx_1}{\sigma\sqrt{n}}}\right) E\left(e^{\frac{tx_2}{\sigma\sqrt{n}}}\right) \dots E\left(e^{\frac{tx_n}{\sigma\sqrt{n}}}\right)$$

$$= e^{-\frac{\mu\sqrt{nt}}{\sigma}} \left\{M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}^n$$

Taking log on both sides

$$\log M_z(t) = \log e^{-\frac{\mu\sqrt{nt}}{\sigma}} + \log \left\{M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}^n$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + n \log M_X\left(\frac{t}{\sigma\sqrt{n}}\right)$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + n \log E\left(e^{\frac{tx}{\sigma\sqrt{n}}}\right)$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + n \log \left[E\left(1 + \frac{tx}{\sigma\sqrt{n}} + \frac{\left(\frac{tx}{\sigma\sqrt{n}}\right)^2}{2!} + \dots\right) \right] \quad \mu \star$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + n \log \left[E\left(1 + \frac{tx}{\sigma\sqrt{n}} + \frac{1}{2!} \frac{t^2 x^2}{\sigma^2 n} + \dots\right) \right]$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + n \log \left[1 + \frac{tx}{\sigma\sqrt{n}} E(x) + \frac{1}{2!} \frac{t^2 x^2}{\sigma^2 n} E(x^2) + \dots \right]$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + n \log \left[1 + \frac{tx}{\sigma\sqrt{n}} \mu_1' + \frac{1}{2!} \frac{t^2 x^2}{\sigma^2 n} \mu_2' + \dots \right]$$

$$= \frac{-\mu t \sqrt{n}}{\sigma} + n \left[\left(\frac{tx}{\sigma \sqrt{n}} \mu_1' + \frac{1}{2!} \frac{t^2 x^2}{\sigma^2 n} \mu_2' + \dots \right) - \frac{1}{2} \left(\frac{tx}{\sigma \sqrt{n}} \mu_1' + \frac{1}{2!} \frac{t^2 x^2}{\sigma^2 n} \mu_2' + \dots \right)^2 + \dots \right]$$

$$= \frac{-\mu t \sqrt{n}}{\sigma} + \frac{\mu_1' t \sqrt{n}}{\sigma} + \frac{\mu_2' t^2}{2! \sigma} + \dots - \frac{(\mu_1')^2 t^2}{2\sigma^2} + \text{terms containing "n" in the denominator}$$

Put $\mu = \mu_1'$

$$= \frac{-\mu_1' t \sqrt{n}}{\sigma} + \frac{\mu_1' t \sqrt{n}}{\sigma} + \frac{\mu_2' t^2}{2! \sigma} + \dots - \frac{(\mu_1')^2 t^2}{2\sigma^2} + \text{terms containing "n" in the denominator}$$

$$= \frac{t^2}{2\sigma^2} (\mu_2' - (\mu_1')^2) + \text{terms containing "n" in the denominator}$$

$$= \frac{t^2}{2\sigma^2} \sigma^2 + \text{terms containing "n" in the denominator}$$

$$\log M_z(t) = \frac{t^2}{2} + \text{terms containing "n" in the denominator}$$

Letting $n \rightarrow \infty$, $\log M_z(t) = \frac{t^2}{2}$

$$\Rightarrow M_z(t) = e^{\frac{t^2}{2}} = \text{MGF of } N(0, 1)$$

Hence z follows standard normal distribution as $n \rightarrow \infty$

Standard Normal Distribution

Let $z = \frac{X-\mu}{\sigma}$, z follows normal distribution with mean 0 and variance 1, then z follows standard normal distribution.

Problems on Central limit theorem

1. If X_1, X_2, \dots, X_n are Poisson variables with parameter $\lambda = 2$, use central limit theorem to estimate $P(120 < S_n < 160)$ where $S_n = X_1 + X_2 + \dots + X_n$ and $n = 75$

Solution:

To find mean and variance

Given mean = 2

Variance = 2

(For Poisson distribution Mean = variance = λ)

To find $n\mu$ and $n\sigma^2$

$$n\mu = 75 \times 2 = 150$$

$$n\sigma^2 = 75 \times 2 = 150$$

$$\sigma\sqrt{n} = \sqrt{150}$$

Application of central limit theorem

$$S_n \sim N(n\mu, \sigma\sqrt{n})$$

$$\sim N(150, \sqrt{150})$$

To find $P(120 < S_n < 160)$

$$\begin{aligned} \text{Let } z &= \frac{S_n - n\mu}{\sigma\sqrt{n}} \\ &= \frac{S_n - 150}{\sqrt{150}} \end{aligned}$$

If $S_n = 120$

$$z = \frac{120 - 150}{\sqrt{150}} = -2.45$$

If $S_n = 160$

$$z = \frac{160 - 150}{\sqrt{150}} = 0.85$$

$$P(120 < S_n < 160) = P\left(\frac{S_n - 150}{\sqrt{150}} \leq z \leq \frac{S_n + 150}{\sqrt{150}}\right)$$

$$= P(-2.45 \leq z \leq 0.85)$$

$$= P(-2.45 \leq z \leq 0) + P(0 \leq z \leq 0.85)$$

$$= 0.4927 + 0.2939 = 0.7866$$

2. Let X_1, X_2, \dots, X_n be independent identically distributed random variable variables with mean = 2 and variance = $\frac{1}{4}$. Find $P(192 < X_1 + X_2 + \dots + X_n < 210)$

Solution:

To find mean and variance

Given mean = 2

Variance = $\frac{1}{4}$, $n = 4$

To find $n\mu$ and $n\sigma^2$

$$n\mu = 100 \times 2 = 200$$

$$n\sigma^2 = 100 \times \frac{1}{4} = 25$$

$$\sigma\sqrt{n} = 5$$

Application of central limit theorem

$$S_n \sim N(n\mu, \sigma\sqrt{n}) \sim N(200, 5)$$

To find $P(192 < S_n < 210)$

$$\begin{aligned} \text{Let } z &= \frac{S_n - n\mu}{\sigma\sqrt{n}} \\ &= \frac{S_n - 200}{5} \end{aligned}$$

If $S_n = 192$

$$z = \frac{192 - 200}{5} = -1.6$$

If $S_n = 210$

$$z = \frac{210 - 200}{5} = 2$$

$$\begin{aligned} P(192 < S_n < 210) &= P\left(\frac{S_n - 200}{5} \leq z \leq \frac{S_n + 200}{5}\right) \\ &= P(-1.6 \leq z \leq 2) \\ &= P(-1.6 \leq z \leq 0) + P(0 \leq z \leq 2) \end{aligned}$$

$$= 0.4452 + 0.4772 = 0.9224$$

3. The resistors r_1, r_2, r_3 and r_4 are independent random variables and is uniform in the interval (450, 550). Using the central limit theorem, find $P(1900 < r_1 + r_2 + r_3 + r_4 < 2100)$

Solution:

To find mean and variance

A random variable X is said to have uniform distribution on the interval (a, b) if its probability density function is given by

$$f(x) = \frac{1}{b-a}, a < x < b$$

$$\text{Mean} = \frac{a+b}{2}, \text{Variance} = \frac{(b-a)^2}{12}$$

$$\text{Mean} = \frac{450+550}{2} = 500$$

$$\text{Variance} = \frac{(550-450)^2}{12} = 833.33, n = 4$$

To find $n\mu$ and $n\sigma^2$

$$n\mu = 4 \times 500 = 2000$$

$$n\sigma^2 = 4 \times 833.33 = 3333.32$$

$$\sigma\sqrt{n} = 2\sqrt{833.33} = 57.73$$

Application of central limit theorem

$$S_n \sim N(n\mu, \sigma\sqrt{n})$$

$$\sim N(2000, 57.73)$$

To find $P(1900 < S_n < 2100)$

$$\begin{aligned} \text{Let } z &= \frac{S_n - n\mu}{\sigma\sqrt{n}} \\ &= \frac{S_n - 2000}{57.73} \end{aligned}$$

If $S_n = 1900$

$$z = \frac{1900 - 2000}{57.73} = -1.73$$

If $S_n = 2100$

$$z = \frac{2100 - 2000}{57.73} = 1.73$$

$$\begin{aligned} P(1900 < S_n < 2100) &= P\left(\frac{S_n - 2000}{57.73} \leq z \leq \frac{S_n + 2000}{57.73}\right) \\ &= P(-1.73 \leq z \leq 1.73) \end{aligned}$$

$$= P(-1.73 \leq z \leq 0) + P(0 \leq z \leq 1.73)$$

$$= 2 \times P(0 \leq z \leq 1.73)$$

$$= 2 \times 0.4582 = 0.9164$$

4. If $x_i, i = 1, 2, \dots, 50$ are independent random variables each having a Poisson distribution with parameter $\lambda = 0.03$ and $S_n = X_1 + X_2 + \dots + X_n$ evaluate

$$P(S_n \geq 3)$$

Solution:

To find mean and variance

Given mean = 0.03

Variance = 0.03, $n = 4$

To find $n\mu$ and $n\sigma^2$

$$n\mu = 50 \times 0.03 = 1.5$$

$$n\sigma^2 = 50 \times 0.03 = 1.5$$

$$\sigma\sqrt{n} = \sqrt{1.5}$$

Application of central limit theorem

$$S_n \sim N(n\mu, \sigma\sqrt{n})$$

$$\sim N(1.5, \sqrt{1.5})$$

To find $P(S_n \geq 3)$

$$\begin{aligned} \text{Let } z &= \frac{S_n - n\mu}{\sigma\sqrt{n}} \\ &= \frac{S_n - 1.5}{\sqrt{1.5}} \end{aligned}$$

If $S_n = 3$

$$z = \frac{3 - 1.5}{\sqrt{1.5}} = \sqrt{1.5}$$

$$\begin{aligned} P(S_n \geq 3) &= P(z \geq \sqrt{1.5}) \\ &= P(z \geq 1.23) \\ &= 0.5 - P(z < 1.23) \\ &= 0.1112 \end{aligned}$$

5. A coin is tossed 300 times. What is the probability that heads will appear more than 140 times and less than 150 times.

Solution:

To find mean and variance

Let P be the probability of getting head in a single trial.

$$p = \frac{1}{2}, q = 1 - \frac{1}{2} = \frac{1}{2}$$

Here $n = 300$

To find np and npq

$$\text{mean} = np = 300 \times \frac{1}{2} = 150$$

$$\text{Variance} = npq = 300 \times \frac{1}{2} \times \frac{1}{2} = 75$$

To find $P(140 < S_n < 150)$

$$\begin{aligned} \text{Let } z &= \frac{X - \mu}{\sigma} \\ &= \frac{X - 150}{\sqrt{75}} \end{aligned}$$

If $X = 140$

$$z = \frac{140 - 150}{\sqrt{75}} = -1.15$$

If $X = 150$

$$z = \frac{150 - 150}{\sqrt{75}} = 0$$

$$\begin{aligned} P(140 < X < 150) &= P\left(\frac{X - 150}{\sqrt{75}} \leq z \leq \frac{X - 150}{\sqrt{75}}\right) \\ &= P(-1.15 \leq z \leq 0) \end{aligned}$$

$$= P(0 \leq z \leq 1.15)$$

$$= 0.3749$$

