Introduction to Probability
Chapter 9
Bivariate Normal Distribution, Central Limit Theorem and Law of Large Numbers

Nitin Gupta
Department of Mathematics
Indian Institute of Technology Kharagpur,
Kharagpur - 721 302, INDIA.

Outline

1. Linear combination of random variable
2. Sum of random variables
3. Bivariate Normal distibution
4. Central Limit Theorem
5. Law of Large Numbers

References

1. Probability and statistics in engineering by Hines et al (2003) Wiley.
2. Mathematical Statistics by Richard J. Rossi (2018) Wiley.
3. Probability and Statistics with reliability, queuing and computer science applications by K. S. Trivedi (1982) Prentice Hall of India Pvt. Ltd.

Linear combination of random variables

Let $X_1, \dots, X_n$ be random variables. Let X_i has mean μ_i and variance σ_i^2, $i = 1, 2, \dots, n$. Let $a_0, a_1, \dots, a_n$ are real valued constants. Then mean and variance of $Y = a_0 + \sum_{i=1}^{n} a_i X_i$, respectively, are

$$E\left(a_0 + \sum_{i=1}^{n} a_i X_i\right) = a_0 + \sum_{i=1}^{n} a_i \mu_i,$$

and if $X_1, \dots, X_n$ be independent random variables, then further

$$Var\left(a_0 + \sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

Example

Example

A patrol pump sells petrol, premium petrol and diesel. Their prices are Rs. 80, 90 and 60 respectively. Let X_1, X_2 and X_3 denote the amount of petrol, premium petrol and diesel purchased on a particular day. Let X_1, X_2 and X_3 are independent with $\mu_1 = 1000$, $\mu_2 = 300$, $\mu_3 = 500$, $\sigma_1 = 100$, $\sigma_2 = 80$ and $\sigma_3 = 50$. The revenue from sales is

$$Y = 80X_1 + 90X_2 + 60X_3.$$

Then

$$E(Y) = 80\mu_1 + 90\mu_2 + 60\mu_3 = 137000$$

and

$$V(Y) = (80)^2\sigma_1^2 + (90)^2\sigma_2^2 + (60)^2\sigma_3^2 = 124840000.$$

Sum of random variables

Let $X_1, \ldots, X_n$ be independent random variables. Then if $S_n = \sum_{i=1}^{n} X_i$, then the MGF of S_n is

$$
\begin{aligned}
M_{S_n}(t) &= E\left(e^{tS_n}\right) \\
&= \prod_{i=1}^{n} M_{X_i}(t)
\end{aligned}
$$

Sum of random variables

Let $X_1, \ldots, X_m$ be independent random variables. If

1. $X_i \sim Bin(n_i, p)$, then $\sum_{i=1}^{m} X_i \sim Bin\left(\sum_{i=1}^{m} n_i, p\right)$
2. $X_i \sim Poiss(\lambda_i)$, then $\sum_{i=1}^{m} X_i \sim Poiss\left(\sum_{i=1}^{m} \lambda_i\right)$
3. $X_i \sim Geo(p)$, then $\sum_{i=1}^{m} X_i \sim NB\left(m, p\right)$
4. $X_i \sim NB\left(n_i, p\right)$, then $\sum_{i=1}^{m} X_i \sim NB\left(\sum_{i=1}^{m} n_i, p\right)$
5. $X_i \sim NB\left(n_i, p\right)$, then $\sum_{i=1}^{m} X_i \sim NB\left(\sum_{i=1}^{m} n_i, p\right)$
6. $X_i \sim Gamma(\alpha_i, \beta)$, then $\sum_{i=1}^{m} X_i \sim Gamma\left(\sum_{i=1}^{m} \alpha_i, \beta\right)$
7. $X_i \sim \chi^2_{r_i}$, then $\sum_{i=1}^{m} X_i \sim \chi^2_{\sum_{i=1}^{m} r_i}$

Sum of random variables

Let $X_1, \ldots, X_m$ be independent random variables. If

1. $X_i \sim N(\mu_i, \sigma_i^2)$, then $\sum_{i=1}^m X_i \sim N\left(\sum_{i=1}^m \mu_i, \sum_{i=1}^m \sigma_i^2\right)$
2. $X_i \sim N(\mu_i, \sigma_i^2)$, then

$$a_0 + \sum_{i=1}^m a_i X_i \sim N\left(a_0 + \sum_{i=1}^m a_i \mu_i, \sum_{i=1}^m a_i^2 \sigma_i^2\right) \quad \text{for } a_i \in \mathbb{R}.$$

Example contd

Example

Total revenue from sales is $Y = 80X_1 + 90X_2 + 60X_3$. Here $\mu_Y = E(Y) = 137000$ and $\sigma_Y^2 = V(Y) = 124840000$. If X_i's are normally distributed, the probability that revenue exceeds 100000 is

$$\begin{aligned} P(Y > 100000) &= P\left(Z > \frac{100000 - 137000}{11173.2}\right) \\ &= P(Z > -3.31) \\ &= \Phi(3.31) \\ &= 0.999533 \end{aligned}$$

Bivariate Normal Distribution (BVN)

$(X, Y) \sim BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ if the joint density is given by

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left(y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)^2},$$

$$-\infty < x < \infty, -\infty < y < \infty, -\infty < \mu_1 < \infty, -\infty < \mu_2 < \infty,$$

$$\sigma_1 \geq 0, \sigma_2 \geq 0, |\rho| < 1.$$

Note that $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$.
Hence $E(X) = \mu_1$, $Var(X) = \sigma_1^2$, $E(Y) = \mu_2$ and $Var(Y) = \sigma_2^2$.

Bivariate Normal Distribution (BVN)

Conditional density of Y given $X = x$ is

$$f_{Y|X=x}(y) = \frac{f_{XY}(x,y)}{f_X(x)}$$
$$= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left(y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)^2}.$$

Hence $[Y|X = x] \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)\right)$.
$E(Y|X = x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$, $Var(Y|X = x) = \sigma_2^2(1-\rho^2)$.
Similarly $[X|Y = y] \sim N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2)\right)$
$E(X|Y = y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)$, $Var(X|Y = y) = \sigma_1^2(1-\rho^2)$.
Also $Cov(X, Y) = \rho\sigma_1\sigma_2$. Hence $\rho = Corr(X, Y)$.

Bivariate Normal Distribution (BVN)

Example

The failure of tube can occur as the result of thermal wear of the internal components. Let X denote the modified life of tube and Y denote the modified thermal wear of the internal components. Let X and Y have a bivariate normal distribution with parameters $\mu_X = 3,\ \mu_Y = 1,\ \sigma_X^2 = 16,\ \sigma_Y^2 = 25$ and $\rho = 3/5$. Determine the probabilities $P(-3 < X < 3)$ and $P(-3 < X < 3 | Y = -4)$.

Solution: (1). $X \sim N(3, 16)$, therefore

$P(-3 < X < 3) = P\left(\frac{-3-3}{4} < Z < 0\right) = P(-1.5 < Z < 0) = 0.433.$

(2). $[X|Y = -4] \sim N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1-\rho^2)\right) \equiv N(0.6, 10.24).$

$$\begin{aligned} P(-3 < X < 3 | Y = -4) &= P\left(\frac{-3-0.6}{\sqrt{10.24}} < Z < \frac{3-0.6}{\sqrt{10.24}}\right) \\ &= P(-1.125 < Z < 0.75) \\ &= 0.64. \end{aligned}$$

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with mean μ and variance σ^2 then $Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ has approximately $N(0,1)$ distribution for n large.

Example

Hard drives are packed 100 to a packet. Drives weights are independent random variable with mean of 0.5 kg and a standard deviation of 0.10 kg. 30 packets are loaded to a box. Suppose we want to find the probability that the drives on a box will exceed 1510 kg in weight. Neglecting both packet and box weight. Let X_i be the weight of ith hard drive $i = 1, 2, \ldots, 3000$. Then total weight is $X = X_1 + \cdots + X_{3000}$. $E(X) = 3000 \times 0.5 = 1500$ and $Var(X) = 3000 \times (0.10)^2 = 30$. Then using CLT, the required solution is

$$\begin{aligned} P(X > 1510) &= P\left(Z > \frac{1510 - 1500}{\sqrt{30}}\right) \\ &= P(Z > 1.83) \\ &= 1 - \Phi(1.83) \\ &= 1 - 0.96637 = 0.03363, \end{aligned}$$

here $Z = \frac{X - 1500}{\sqrt{30}} \sim N(0,1)$.

Law of large numbers

Let $X_1, \ldots, X_n$ be independent random variable with common mean μ and common variance σ^2. Let $S_n = \sum_{i=1}^{n} X_i$. then for any $\epsilon > 0$,

$$P(|\frac{S_n}{n} - \mu| < \epsilon) \to 1,$$

as $n \to \infty$.

Law of large numbers–example

Let $Y = \sum_{i=1}^{n} X_i$, where each X_i takes value 1 with probability p, and 0 with probability $q = 1 - p$, $i = 1, \ldots, n$. Suppose $X_1, \ldots, X_n$ are independent. Let $\hat{p} = Y/n$. Then the law of large number states that

$$P(|\hat{p} - p| < \epsilon) \geq 1 - \frac{p(1-p)}{n\epsilon^2}.$$

This equation may be written as

$$P(|\hat{p} - p| < \epsilon) \geq 1 - \beta,$$

we may determine the value of n satisfying the the above condition for

$$n \geq \frac{p(1-p)}{\epsilon^2 \beta}$$

Summary

In this chapter we presented the sum of random variables, bivariate normal distribution, central limit theorem and law of large numbers.