Essentials of Data Science With R Software - 1

Probability Theory and Statistical Inference

Univariate Random Variables

:::

Lecture 26

Moments and Variance

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Mathematical Expectation of a Continuous Random Variable:

Let *X* be a continuous random variable having the probability density function *f* (*x*).

Suppose *g*(*X*) is a real valued function of *X*.

Obviously *g*(*X*) will also be a random variable.

Then expectation of *g*(*X*) is defined as is defined as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

provided with $\int |g(x)| f(x)\,d\mathrm{x} < \infty$**.**

Mathematical Expectation of a Discrete Random Variable:

Let *X* be a discrete random variable having the probability mass function $P(X = x_i) = p_i$.

Suppose $g(X)$ is a real valued function of *X*.

Obviously $g(X)$ will also be a random variable.

Thus *X* takes the values $x_1, x_2, \ldots, x_k, \ldots$, with respective probabilities $p_1, p_2, \ldots, p_k, \ldots$

Then expectation of $g(X)$ exists and is defined as

$$E[g(X)] = \sum_{i=1}^{\infty} g(x_i) P(X = x_i) = \sum_{i=1}^{\infty} g(x_i) p_i$$

provided $\sum_{i=1}^{\infty} |g(x_i)| p_i < \infty$.

Special cases of Expectation of a Random Variable: Moments

Moments are used to describe different characteristics and features of a probability distribution, viz., central tendency, disperson, symmetry and peakedness (hump) of probability curve.

Special cases of Expectation of a Random Variable: Moments

1. $g(X) = X^r$ where r is nonnegative integer,

then $E[g(X)] = E(X^r) = \mu_r'$

μ_r' is called as r^{th} moment of X about origin.

2. $g(X) = (X - A)^r$ where r is nonnegative integer,

then $E[g(X)] = E(X - A)^r$

is called as r^{th} moment of X about the point "A".

Special cases of Expectation of a Random Variable: Moments

2. $E[g(X)] = E(X - A)^r$

- **If $A = E(X)$: Mean**

  then $E[(X - A)^r] = E[X - E(X)]^r = \mu_r$

  μ_r is called as r^{th} central moment of X.

- **For $n = 2$, $E[X - E(X)]^2 = \mu_2 = \sigma^2$**

  is called the variance of X.

Gives idea about variation in X relative to mean value..

Special cases of Expectation of a Random Variable: Variance

- **The variance of a continuous random variable *X* is defined as**

$$\boldsymbol{Var(X) = E[X-\mu]^2 = \int_{-\infty}^{\infty} [x-\mu]^2 f(x)\, dx}$$

where $\boldsymbol{\mu = E(x)}$.

- **For a discrete random variable *X*, the variance of *X* is defined as**

$$\boldsymbol{Var(X) = E[X-\mu]^2 = \sum_{i=1}^{n} [x_i - \mu]^2 P(X = x_i)}$$
$$\boldsymbol{= \sum_{i=1}^{n} [x-\mu]^2 p_i}$$

Special cases of Expectation of a Random Variable: Variance

$$E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

If *a* and *b* are any real constants, then

$$Var(a) = 0$$

$$Var[aX + b] = a^2 Var[X]$$

Expectation and Variance of a Random Variable

Numbers are used to summarize a probability distribution for a random variable *X*.

- **The mean is a measure of the center or middle of the probability distribution, and**
- **the variance is a measure of the dispersion, or variability in the probability distribution.**

Expectation and Variance of a Random Variable

These two measures do not uniquely identify a probability distribution but are useful in summarizing the probability distribution of *X*.

Two different distributions can have the same mean and variance.

Still, these measures are simple, useful summaries of the probability distribution of *X*.

Mean and variance are particular "moments".

Special cases of Expectation of a Random Variable: Variance

Example 1

Consider the continuous random variable "waiting time for the train". Suppose that a train arrives every 20 min. Therefore, the waiting time of a particular person is random and can be any time contained in the interval [0, 20]. The required probability density function is

$$\boldsymbol{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{20}} & \textbf{for } \mathbf{0} \le \boldsymbol{x} \le \mathbf{20} \\ \mathbf{0} & \textbf{otherwise}. \end{cases}$$

Special cases of Expectation of a Random Variable: Variance

Example 1

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{20} x\frac{1}{20}dx = 10$$

$$Var\,(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x)dx$$

$$= \int_{-\infty}^{0} (x - 10)^2 f(x)dx + \int_{0}^{20} (x - 10)^2 f(x)dx$$

$$+ \int_{20}^{\infty} (x - 10)^2 f(x)dx$$

$$= 0 + \int_{0}^{20} (x - 10)^2 \frac{1}{20} dx + 0 = \frac{100}{3}$$

Special cases of Expectation of a Random Variable: Variance

Example 2

Suppose we roll a dice and following is the scheme for award based on outcomes –

Point (x)	1	2	3	4	5	6
Reward (INR)	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$

$$E(X^2) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} = 91$$

$$Var(X) = E(X^2) - [E(X)]^2 = 91 - 3.5*3.5 = 78.75$$

Special cases of Expectation of a Random Variable: Variance

Standard Deviation:

s^2 **: (Sample) Variance**

s **: Positive square root of** s^2 **is called as <u>(sample) standard error (se)</u>.**

σ^2 **: (Population) Variance.**

σ **: (Population) standard deviation.**

More popular notation among practitioners

Special cases of Expectation of a Random Variable: Variance

Standard Deviation:

Standard deviation (or standard error) has an advantage that it has the same units as of data, so easy to compare.

For example, if x is in meter, then s^2 is in meter^2 which is not so convenient to interpret.

On the other hand, if x is in meter, then s is in meter which is more convenient to interpret.

Special cases of Expectation of a Random Variable: Variance

Variance (or standard deviation) measures how much the observations vary or how the data is concentrated around the arithmetic mean.

Decision Making

Lower value of variance (or standard deviation, standard error) indicates that the data is highly concentrated or less scattered around the mean.

Higher value of variance (or standard deviation, standard error) indicates that the data is less concentrated or highly scattered around the mean.

Special cases of Expectation of a Random Variable: Variance

Decision Making

The data set having higher value of variance (or standard deviation) has more variability.

The data set with lower value of variance (or standard deviation) is preferable.

If we have two data sets and suppose their variances are Var_1 and Var_2.

If $Var_1 > Var_2$ then the data in Var_1 is said to have more variability (or less concentration around mean) than the data in Var_2.