Mean and Variance

Prof. Asim Tewari
IIT Bombay

Concepts of Population and Sample

- Mean
- Variance
- Covariance

- Population and Sample
- Population mean and variance
- Sample mean and variance

Mean and Variance

Expected value of $g(x)$

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

X is a random variable with p.d.f. $f(x)$

Mean value of X is $\mu \equiv \int_{-\infty}^{\infty} x f(x)\, dx = E(X)$

Variance of X is $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$

$$= E\left[(X - E(X))^2\right] = E\left[X^2 - 2X E(X) + [E(X)]^2\right]$$

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

$$\Rightarrow \mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

Variance and Covariance

$$Cov(X,Y) = E\left([X - E(X)][Y - E(Y)]\right)$$

$$= E(XY) - 2E(X)E(Y) + E(X)E(Y)$$

$$= E(XY) - E(X)E(Y)$$

$$\therefore Var(X) = E(X^2) - [E(X)]^2$$

and $Cov(X,Y) = E(X.Y) - E(X)E(Y)$

Variance and Covariance

1.) $Var(X) \geq 0$

2.) $P(X=a)=1 \iff Var(X)=0$

3.) $Var(X+a) = Var(X)$

4.) $Var(aX) = a^2 Var(X)$

5.) $Var(X) = Cov(X,X)$

6.) $Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\, Cov(X,Y)$

Variance of sum of random variables

$$\mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i,j}^{N} \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{N} \mathrm{Var}(X_i) + \sum_{i \neq j}^{N} \mathrm{Cov}(X_i, X_j)$$

$$\mathrm{Var}\left[\sum_{i=1}^{N} a_i X_i\right] = \sum_{i,j}^{N} a_i a_j \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{N} a_i^2 \mathrm{Var}(X_i) + \sum_{i \neq j}^{N} a_i a_j \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{N} a_i^2 \mathrm{Var}(X_i) + \sum_{1 \le i < j \le N} a_i a_j \mathrm{Cov}(X_i, X_j)$$

Variance and Covariance

If $Cov(X_i, X_j) = 0 \quad \forall \; i \neq j$

$\Rightarrow \quad X_i, X_j$ are uncorrelated

For N independent r.v. $X_1, X_2 \cdots\cdots X_N$

$$Var\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} Var(X_i)$$

If all the N r.v. have the same variance σ^2

then $$Var\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} Var(X_i) = N\sigma^2$$

Variance of mean

Mean of n r.v.s is $\frac{1}{n}\sum_{i=1}^{n} x_i$

Variance of mean of n r.v.s would be

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} var(x_i)$$ Assuming x_i's are independent

$$= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$ Assuming x_i's are iid

$$\therefore Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{\sigma^2}{n}$$ if x_i's are iid

Population and sample

Population : Population of size N with value x_i

Population mean : $\mu \equiv \frac{1}{N} \sum_{i=1}^{N} x_i$ $\Big\}$ $\mu = E(X)$

Population variance : $\sigma^2 \equiv \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$ $\Big\}$ $\sigma^2 = E[X - E(X)]^2$

Sample : Take n random values (with replacement) from the population. $y_1, y_2, \ldots \ldots y_n$

Sample mean : $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$

Sample variance : ?

Sample mean

Expected value of Sample mean:

$$E(\bar{y}) = E\left[\frac{1}{n}\sum_{i=1}^{n} y_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(y_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu \qquad = \mu$$

$\therefore \quad E(\bar{y}) = \mu$

Expected value of sample mean is equal to population mean.

$\Rightarrow$ Sample mean is an unbiased estimator of the population mean.

Sample Variance

How to define sample variance so that it is an unbiased estimator of the population variance.

Let squared deviation be defined as

$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

then $E(\sigma_y^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2\right)$

$$= E\left(\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\frac{1}{n}\sum_{j=1}^{n}y_j\right)^2\right)$$

Sample Variance

$$\Rightarrow E(\sigma_y^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{j=1}^{n} y_j\right)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\left(y_i^2 - \frac{2}{n} y_i \sum_{j=1}^{n} y_j + \frac{1}{n^2}\sum_{j=1}^{n} y_j \sum_{k=1}^{n} y_k\right)$$

$j=i$ $\quad j\neq i$ $\quad k\neq j$ $\quad k=j$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{2}{n}\right)E(y_i^2) - \frac{2}{n}\sum_{j\neq i} E(y_i y_j) + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k\neq j} E(y_j y_k) + \frac{1}{n^2}\sum_{j=1}^{n} E(y_j^2)$$

$$\Rightarrow E(\sigma_y^2) = \frac{1}{n}\sum_{i=1}^{n}\left[\left(1-\frac{2}{n}\right)(\sigma^2+\mu^2) - \frac{2}{n}(n-1)\mu^2 + \frac{1}{n^2}n(n-1)\mu^2 + \frac{n}{n^2}(\sigma^2+\mu^2)\right]$$

Sample Variance

$$\Rightarrow E(\sigma_g^2) = \frac{1}{n}\sum\left(\frac{n-1}{n}\right)\sigma^2$$

$$= \frac{1}{n}\, n \left[\frac{n-1}{n}\right]\sigma^2 = \left(\frac{n-1}{n}\right)\sigma^2$$

$$\therefore\ E(\sigma_g^2) = \frac{n-1}{n}\,\sigma^2$$

$\therefore\ \sigma_g^2$ estimates the population variance with a bias of $\frac{n-1}{n}$ factor

Sample Variance

$\therefore$ if we define sample variance $s^2 = \frac{n}{n-1} \sigma_y^2$

then it will be an unbiased estimator of the population variance.

$\therefore$ Sample variance s^2 is defined as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$