Variance, Covariance, Correlation

1. Variance:

Given a r.v. X, we know that the average value of X is $E(X)$. However, this doesn't tell us how far X tends to be from $E(X)$. The variance is a measure of how spread out the distribution of X is.

Formula: $\mathrm{Var}(X) = E((X - E(X))^2)$.

Alternatively, it can be written as such: $\sigma^2_X = \mathrm{Var}(X) = E((X - \mu_X)^2)$, where $E(X) = \mu_X$.

Properties:

1. $\mathrm{Var}(X) \geq 0 \rightarrow (E(X))^2 \leq E(X^2)$
2. $\mathrm{Var}(X) = E(X^2) - (E(X))^2$
3. $\mathrm{Var}(a+bX) = b^2 \cdot \mathrm{Var}(X) \rightarrow \mathrm{Var}(X+b) = \mathrm{Var}(X)$

   $\rightarrow \mathrm{Var}(bX) = b^2\,\mathrm{Var}(X)$

E.g. 1 Let X have the pmf given by

$$P_X(x) = \begin{cases} \frac{1}{2}, & x=2 \\ \frac{1}{6}, & x=3 \\ \frac{1}{6}, & x=4 \\ \frac{1}{6}, & x=5 \\ 0, & \text{otherwise} \end{cases}$$

Find $\mathrm{Var}(X)$

$$E(X) = 2\left(\tfrac{1}{2}\right) + 3\left(\tfrac{1}{6}\right) + 4\left(\tfrac{1}{6}\right) + 5\left(\tfrac{1}{6}\right) = 3$$

$$(E(X))^2 = 9$$

$$E(X^2) = 4\left(\tfrac{1}{2}\right) + 9\left(\tfrac{1}{6}\right) + 16\left(\tfrac{1}{6}\right) + 25\left(\tfrac{1}{6}\right) = \frac{31}{3}$$

$$Var(x) = E(x^2) - (E(x))^2$$
$$= \frac{31}{3} - 9$$
$$= \frac{31-27}{3}$$
$$= \frac{4}{3}$$

Variance of Various Values:

1. $Var(c) = 0$, c is a constant
2. $X \sim Ber(\theta) \rightarrow Var(x) = \theta(1-\theta)$
3. $X \sim Bin(n,\theta) \rightarrow Var(x) = n\theta(1-\theta)$
4. $X \sim Po(\lambda) \rightarrow Var(x) = \lambda$
5. $X \sim Geo(\theta) \rightarrow Var(x) = \frac{1-\theta}{\theta^2}$
6. $X \sim exp(\lambda) \rightarrow Var(x) = \frac{1}{\lambda^2}$
7. $X \sim N(0,1) \rightarrow Var(x) = 1$
8. $X \sim N(\mu,\sigma^2) \rightarrow Var(x) = \sigma^2$
9. $X \sim Gamma(\alpha,\lambda) \rightarrow Var(x) = \frac{\alpha^2}{\lambda^2}$
10. $X \sim Beta(a,b) \rightarrow Var(x) = \frac{ab}{(a+b)^2(a+b+1)}$

2. Standard Deviation:

$$SD(x) = +\sqrt{Var(x)}$$
$$SD(a+bx) = +\sqrt{Var(a+bx)}$$
$$= \sqrt{b^2 \cdot Var(x)}$$
$$= |b|\sqrt{Var(x)}$$
$$= |b|\, SD(x)$$

3. Covariance

Let x and y be 2 r.v.

$cov(x,y) = E(xy) - E(x) \cdot E(y)$

Properties:

1. $cov(ax+by, z) = a\,cov(x,z) + b\,cov(y,z)$
2. $cov(x, a+by) = b\,cov(x,y)$
3. $cov(x, a+y) = cov(x,y)$
4. $cov(x, by) = b\,cov(x,y)$
5. If $x \perp y$, $cov(x,y) = 0$

E.g. 2

$x \sim$ uniform $[-1,1]$ and $y = x^2$

Find $cov(x,y)$

$cov(x,y) = E(xy) - E(x)E(y)$

$$\begin{aligned} E(xy) &= E(x \cdot x^2) \\ &= E(x^3) \\ &= \int_{-1}^{1} x^3 f_x(x)\, dx \\ &= \int_{-1}^{1} x^3 \left(\tfrac{1}{2}\right) dx \\ &= \tfrac{1}{2} \int_{-1}^{1} x^3\, dx \\ &= \tfrac{1}{2} \left[\tfrac{x^4}{4} \Big|_{-1}^{1} \right] \\ &= 0 \end{aligned}$$

$$E(x) = \int_{-1}^{1} x f_x(x)\,dx$$

$$= \int_{-1}^{1} x\left(\tfrac{1}{2}\right) dx$$

$$= \frac{1}{2}\int_{-1}^{1} x\,dx$$

$$= \frac{1}{2}\left[\frac{x^2}{2}\Big|_{-1}^{1}\right]$$

$$= 0$$

$\therefore \ cov(x,y) = 0$

Note:

1. $var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2cov(X_1, X_2)$

2. $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i}\sum_{\substack{j \\ i<j}} cov(X_i, X_j)$

3. If $X_1, X_2, \ldots X_n$ are independent, then

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

4. $cov(x,x) = Var(x)$

4. Correlation

$$Corr(x,y) = \frac{cov(x,y)}{SD(x) \cdot SD(y)}$$
$$= \frac{cov(x,y)}{\sqrt{var(x) \cdot var(y)}}$$

E.g. 3

$y = a + bx$
Find corr(x,y)

$$corr(x,y) = \frac{cov(x,y)}{\sqrt{var(x) \cdot var(y)}}$$
$$= \frac{cov(x, a+bx)}{\sqrt{var(x) \cdot var(a+bx)}}$$
$$= \frac{b\,cov(x,x)}{\sqrt{var(x) \cdot b^2 \cdot var(x)}}$$
$$= \frac{b\,var(x)}{|b|\,var(x)}$$
$$= \frac{b}{|b|}$$

5. Proofs

1. $var(x+b) = var(x)$, b is a constant

$$var(x+b) = E(x+b)^2 - (E(x+b))^2$$
$$= E(x^2 + 2xb + b^2) - (E(x) + E(b))^2$$
$$= E(x^2) + E(2xb) + E(b^2) - (E(x))^2 - 2E(x)E(b) - (E(b))^2$$
$$= E(x^2) + 2bE(x) + b^2 - (E(x))^2 - 2bE(x) - b^2$$
$$= E(x^2) - (E(x))^2$$
$$= var(X)$$

QED

2. $E((x - E(x))^2) = E(x^2) - (E(x))^2$

$$
\begin{aligned}
E((x - E(x))^2) &= E(x^2 - 2xE(x) + (E(x))^2) \\
&= E(x^2) - 2(E(x))^2 + (E(x))^2 \\
&= E(x^2) - (E(x))^2
\end{aligned}
$$

QED

3. $var(a+bx) = b^2 \, var(x)$

$$
\begin{aligned}
var(a+bx) &= E(a+bx)^2 - (E(a+bx))^2 \\
&= E(a^2 + 2abx + b^2x^2) - (E(a+bx))^2 \\
&= E(a^2) + E(2abx) + E(b^2x^2) - (E(a) + E(bx))^2 \\
&= a^2 + 2ab\,E(x) + b^2\,E(x^2) - (E(a))^2 - 2\,E(a)E(bx) - (E(bx))^2 \\
&= a^2 + 2abE(x) + b^2\,E(x^2) - a^2 - 2abE(x) - b^2(E(x))^2 \\
&= b^2\,E(x^2) - b^2(E(x))^2 \\
&= b^2(E(x^2) - (E(x))^2) \\
&= b^2\,var(x)
\end{aligned}
$$

QED