9. Variance and standard deviation

**Problem 9.1** (Best prediction). Consider a random variable $X$ with $\mathbb{E}X^2 < \infty$ and suppose we want the best predictor of a future realization of $X$, in the sense of having the smallest root mean square error (RMSE):

$$\text{RMSE}(c) := \sqrt{\mathbb{E}(X-c)^2}.$$ 

What $c$ should we choose?

*Solution.* Let $\mu := \mathbb{E}X$. Then

$$(X - c)^2 = ((X - \mu) + (\mu - c))^2$$
$$= (X - \mu)^2 + 2(\mu - c)(X - \mu) + (\mu - c)^2.$$ 

and

$$\mathbb{E}(X - c)^2 = \mathbb{E}(X - \mu)^2 + 2(\mu - c)\mathbb{E}(X - \mu) + (\mu - c)^2$$
$$= \mathbb{E}(X - \mu)^2 + (\mu - c)^2.$$ 

Since $\mathbb{E}X^2 < \infty$, the above expression is uniquely minimized at $c = \mu$.

The minimum RMSE of prediction is called the *standard deviation* of $X$, defined as

$$\sigma_X := \text{SD}(X) := \sqrt{\mathbb{E}(X - \mu)^2}.$$ 

The variance of $X$ is given by

$$\text{Var}(X) := \text{SD}^2(X) = \sigma_X^2 = \mathbb{E}(X - \mu)^2.$$ 

9.1. A computing formula for variance. Let $X$ have mean $\mu$ and standard deviation $\sigma$. Putting $c = 0$ in

$$\mathbb{E}(X - c)^2 = \mathbb{E}(X - \mu)^2 + (\mu - c)^2$$

gives

$$\mathbb{E}X^2 = \text{Var}(X) + [\mathbb{E}X]^2,$$

from which we obtain the *computing formula* for variance:

$$\text{Var}(X) = \mathbb{E}X^2 - [\mathbb{E}X]^2,$$

which is often more convenient than the definition (16).

**Example 9.2** (Bernoulli trial). Let $X$ be a Bernoulli random variable with success probability $p$. Then $p_X(0) = 1 - p = 1 - p_X(1)$ and

$$\mathbb{E}X = p \times 1 + (1 - p) \times 0,$$
$$\mathbb{E}X^2 = p \times 1^2 + (1 - p) \times 0^2 = p = \mathbb{E}X,$$
$$\text{Var}(X) = \mathbb{E}X^2 - [\mathbb{E}X]^2 = p - p^2 = p(1 - p),$$

and

$$\sigma = \sqrt{p(1-p)}.$$
9.2. Changes of scale and location. Let $X$ be a random variable and $a, b \in \mathbb{R}$ be real-valued constants. We define $Y := aX + b$, i.e., $Y : \Omega \to \mathbb{R}$ is a random variable satisfying

$$Y(\omega) = aX(\omega) + b, \quad \omega \in \Omega.$$ 

$Y$ is obtained from $X$ by a change in scale of $a$ and a change in location of $b$. By properties of expectation, we observe

$$\mathbb{E}Y = \mathbb{E}(aX + b) = a\mathbb{E}X + b$$

and so

$$(Y - \mathbb{E}Y)^2 = a^2(X - \mathbb{E}X)^2.$$ 

Therefore,

$$\text{Var}(Y) = a^2\text{Var}(X)$$

and

$$\text{SD}(Y) = |a| \text{SD}(X).$$

**Example 9.3 (Standardizing).** Let $X$ have mean $\mu$ and standard deviation $\sigma$ and let

$$Z := \frac{X - \mu}{\sigma}.$$ 

Then $Z$ gives the number of standard deviations $X$ is above or below its mean. Furthermore, $\mathbb{E}Z = 0$ and $\text{SD}(X) = 1$.

9.3. Sums of random variables. Let $X$ and $Y$ be random variables on $\Omega$. Then for $S = X + Y$, we have $\mu_S = \mu_X + \mu_Y$ and

$$S - \mu_S = (X - \mu_X) + (Y - \mu_Y);$$

therefore,

$$\mathbb{E}(Y - \mu_Y)^2 = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y),$$

where

$$\text{Cov}(X,Y) = \mathbb{E}(X - \mu_X)(Y - \mu_Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$$

is called the covariance of $X$ and $Y$.

You should verify the following properties of the covariance function.

(CV1) **Symmetry**: $\text{Cov}(X,Y) = \text{Cov}(Y,X)$.

(CV2) **Scale and location**: $\text{Cov}(aX + b, Y) = a\text{Cov}(X, Y)$.

(CV3) **Additivity**: $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.

(CV4) **Bilinearity**: $\text{Cov}\left(\sum_i a_i X_i, \sum_j b_j Y_j\right) = \sum_{i,j} a_i b_j \text{Cov}(X_i,Y_j)$.

(CV5) **Relation to variance**: $\text{Var}(X) = \text{Cov}(X, X)$.

Properties of the variance function can be derived from (CV1)–(CV4) via (CV5).

**IMPORTANT**: $\text{Cov}(X,Y) = 0$ does **NOT** imply that $X$ and $Y$ are independent, as the next example demonstrates.

**Example 9.4.** Let

<table>
<thead>
<tr>
<th>$X = x$</th>
<th>-2</th>
<th>-1</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_X(x)$</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
</tbody>
</table>

Then, if $Y = X^2$, we have
Here, $E_{XY} = E_X^3 = 0$ and $E_X E_Y = 0$ implies $Cov(X, Y) = 0$, but

$$P[Y = i | X = x] = \begin{cases} 1, & i = x^2 \\ 0, & \text{otherwise,} \end{cases}$$

which does not coincide with $P[Y = i]$. Therefore, $Cov(X, Y) = 0$ but $X$ and $Y$ are not independent.