2.3 Variance

The variance of a random variable X is a nonnegative number that summarizes on average how much X differs from its mean, or expectation. The first expression that comes to mind is

$$X - E(X)$$

i.e. the difference between X and its mean. This itself is a random variable, since even though EX is just a number, X is still random. Hence we would need to take an expectation to turn this expression into the average amount by which X differs from its expected value. This leads us to

$$E(X - EX)$$

This is almost the definition for variance. We require that the variance always be nonnegative, so the expression inside the expectation should always be ≥ 0. Instead of taking the expectation of the difference, we take the expectation of the squared difference.

Definition 2.10. *The* ***variance*** *of X, denoted by $Var(X)$ is defined*

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

Below we give and prove some useful properties of the variance.

Proposition 2.11. *If X is a random variable with mean EX and $c \in \mathbb{R}$ is a real number,*

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

(b) $Var(cX) = c^2 Var(X)$.

(c) $Var(X) = E(X^2) - E(X)$.

(d) If X and Y are independent random variables, then

$$Var(X + Y) = Var(X) + Var(Y)$$

Proof.

(a) Since we always have $(X - EX)^2 \geq 0$, its average is also ≥ 0. Hence,

$$E[(X - EX)^2] \geq 0.$$

(b) Going by the definition, we have

$$\begin{aligned}
\mathrm{Var}(cX) &= E[(cX - E[cX])^2] \\
&= E[(cX - cEX)^2] \\
&= E[c^2(X - EX)^2] \\
&= c^2E[(X - EX)^2] \\
&= c^2\mathrm{Var}(X)
\end{aligned}$$

(c) Expanding out the square in the definition of variance gives

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X - EX)^2] \\
&= E[X^2 - 2XEX + (EX)^2] \\
&= E[X^2] - E(2XEX) + E((EX)^2) \\
&= E[X^2] - 2EXEX + (EX)^2 \\
&= E[X^2] - (EX)^2
\end{aligned}$$

where the third equality comes from linearity of E (Exercise 2.3 (a)) and the fourth equality comes from Exercise 2.3 (b) and the fact that since EX and $(EX)^2$ are constants, their expectations are just EX and $(EX)^2$ respectively.

(d) By the definition of variance,

$$\begin{aligned}
\mathrm{Var}(X + Y) &= E[(X + Y)^2] - (E[X + Y])^2 \\
&= E[X^2 + 2XY + Y^2] - \Big((E[X])^2 + 2E[X]E[Y] + (E[Y])^2\Big) \\
&= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y]^2) + 2E[XY] - 2E[X]E[Y] \\
&= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y]^2) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y)
\end{aligned}$$

where the fourth equality comes from the fact that if X and Y are independent, then $E[XY] = E[X]E[Y]$. Independence of random variables will be discussed in the "Random Variables" section, so don't worry if this proof doesn't make any sense to you yet.

$\square$

Exercise 2.12. Compute the variance of a die roll, i.e. a uniform random variable over the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Solution. Let X denote the outcome of the die roll. By definition, the variance is

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X - EX)]^2 \\
&= E(X^2) - (EX)^2 && \text{(Proposition 2.11 (c))} \\
&= \Big(\sum_{k=1}^{6} k^2 \cdot \frac{1}{6}\Big) - (3.5)^2 && \text{(Definition of Expectation)} \\
&= \frac{1}{6} \cdot (1 + 4 + 9 + 16 + 25 + 36) - 3.5^2 \\
&= \frac{1}{6} \cdot 91 - 3.5^2 \\
&\approx 2.92
\end{aligned}$$

□

Remark 2.13. The square root of the variance is called the **standard deviation**.

2.3.1 Markov's Inequality

Here we introduce an inequality that will be useful to us in Section 5.1 (Point Estimation). Feel free to skip this section and return to it when you read "Chebyschev's inequality" and don't know what's going on.

Markov's inequality is a bound on the probability that a nonnegative random variable X exceeds some number a.

Theorem 2.14 (Markov's inequality)**.** *Suppose X is a nonnegative random variable and $a \in \mathbb{R}$ is a positive constant. Then*

$$P(X \geq a) \leq \frac{EX}{a}$$

Proof. By definition of expectation, we have

$$\begin{aligned}
EX &= \sum_{k \in X(\Omega)} kP(X = k) \\
&= \sum_{k \in X(\Omega) \text{ s.t. } k \geq a} kP(X = k) + \sum_{k \in X(\Omega) \text{ s.t. } k < a} kP(X = k)
\end{aligned}$$

Since X is nonnegative and also since probabilities are nonnegative, each term in the right sum in the above expression is positive. Therefore if we remove this piece from the above expression, we get a smaller number. Hence the above is

$$\begin{aligned}
&\geq \sum_{k \in X(\Omega) \text{ s.t. } k \geq a} kP(X = k) \\
&\geq \sum_{k \in X(\Omega) \text{ s.t. } k \geq a} aP(X = k) \\
&= a \sum_{k \in X(\Omega) \text{ s.t. } k \geq a} P(X = k) \\
&= aP(X \geq a)
\end{aligned}$$

where the second inequality above follows from the fact that $k \geq a$ over the set $\{k \in X(\Omega) \text{ s.t. } k \geq a\}$.

Notation: "s.t." stands for "such that".

Dividing both sides by a, we recover

$$P(X \geq a) \leq \frac{EX}{a}$$

□

Corollary 2.15 (Chebyschev's inequality)**.** *Let X be a random variable. Then*

$$P(|X - EX| > \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

Proof. This is marked as a corollary because we simply apply Markov's inequality to the nonnegative random variable $(X - EX)^2$. We then have

$$\begin{aligned}
P(|X - EX| > \varepsilon) &= P((X - EX)^2 > \varepsilon^2) && \text{(statements are equivalent)} \\
&\leq \frac{E[(X - EX)^2]}{\varepsilon^2} && \text{(Markov's inequality)} \\
&= \frac{\mathrm{Var}(X)}{\varepsilon^2} && \text{(definition of variance)}
\end{aligned}$$

□