Lecture 10

0.1 Expectation of a Function of a Random Variable

Let $X : \Omega \to S$ be a random variable and let $f : S \to \mathbb{R}$ be a function. Then $f(X)$ is also a random variable, and we could ask ourselves what is the value of $\mathbb{E}(f(X))$ (e.g., $\mathbb{E}(1/X)$, $\mathbb{E}(X^2)$ etc.). It follows by the linearity of expectation that $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ for every $a, b \in \mathbb{R}$. That is, $\mathbb{E}(f(X)) = f(\mathbb{E}(X))$ whenever f is a linear function. One could imagine that such an equality holds for every function f but, as illustrated by the following example, this is not the case.

Example 1: Let X be a random variable satisfying $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ and let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Then $\mathbb{E}(X) = 1 \cdot 1/2 + (-1) \cdot 1/2 = 0$ implying that $f(\mathbb{E}(X)) = 0$. On the other hand, $f(X) = X^2 = 1$, i.e., $\mathbb{P}(f(X) = 1) = 1$. Hence $\mathbb{E}(f(X)) = 1$. We conclude that $\mathbb{E}(f(X)) \neq f(\mathbb{E}(X))$ in this case.

Denoting $f(X)$ by Y, one could of course calculate $\mathbb{E}(f(X))$ using the formula $\mathbb{E}(Y) = \sum_y y \cdot \mathbb{P}(Y = y)$. However, this requires calculating the distribution of Y and does not make a real use in the fact that Y is a function of X. The following claim suggests a better course of action.

Claim 0.1. *If f is a non-negative function, namely $f : S \to [0, \infty)$, then*

$$\mathbb{E}(f(X)) = \sum_x f(x) \cdot \mathbb{P}(X = x).$$

Moreover, for a general function f, $\mathbb{E}(f(X))$ is finite if and only if the above series absolutely converges, in which case the series $\sum_x f(x) \cdot \mathbb{P}(X = x)$ converges to $\mathbb{E}(f(X))$.

Proof. We have

$$\begin{aligned}
\sum_x f(x) \cdot \mathbb{P}(X = x) &= \sum_{y \in [0,\infty)} \sum_{x \in S : f(x) = y} f(x) \cdot \mathbb{P}(X = x) \\
&= \sum_{y \in [0,\infty)} y \sum_{x \in S : f(x) = y} \mathbb{P}(\{w \in \Omega : X(\omega) = x\}) \\
&= \sum_{y \in [0,\infty)} y \cdot \mathbb{P}\left(\bigcup_{x \in S : f(x) = y} \{w \in \Omega : X(\omega) = x\} \right) \\
&= \sum_{y \in [0,\infty)} y \cdot \mathbb{P}(\{\omega \in \Omega : \exists x \in S \text{ such that } X(\omega) = x \text{ and } f(x) = y\}) \\
&= \sum_{y \in [0,\infty)} y \cdot \mathbb{P}(\{\omega \in \Omega : f(X(\omega)) = y\}) \\
&= \sum_{y \in [0,\infty)} y \cdot \mathbb{P}(f(X) = y) \\
&= \mathbb{E}(f(X)),
\end{aligned}$$

where the third equality holds since $\{w \in \Omega : X(\omega) = x\} \cap \{w \in \Omega : X(\omega) = x'\} = \emptyset$ whenever $x \neq x'$. □

Example 2: Let $X \sim \text{Geom}(1/2)$ and let $Y = 2^X$. Then the support of μ_X is $\mathbb{N}$ and for every $k \in \mathbb{N}$ we have $\mathbb{P}(X = k) = \frac{1}{2^k}$. Moreover, it is evident that the support of μ_Y is $\{2^k : k \in \mathbb{N}\}$ and that $\mathbb{P}\left(Y = 2^k\right) = \mathbb{P}(X = k) = \frac{1}{2^k}$ for every $k \in \mathbb{N}$, where the first equality holds since $f(x) = 2^x$ is a bijection between $\mathbb{N}$ and $\{2^k : k \in \mathbb{N}\}$. We now calculate $\mathbb{E}(Y)$ using two methods. The first method uses the distribution of Y and the second uses Claim 0.1. Using the distribution of Y we get

$$\mathbb{E}(Y) = \sum_{y \in \{2^k : k \in \mathbb{N}\}} y \cdot \mathbb{P}(Y = y) = \sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = \infty.$$

Using Claim 0.1 we get

$$\mathbb{E}(Y) = \sum_{k \in \mathbb{N}} 2^k \cdot \mathbb{P}(X = k) = \sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = \infty.$$

As we can see, both methods yield the same result.

1 Variance

Definition 1.1. *Let X be a random variable with finite expectation. The* variance *of X is defined to be*

$$\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

The standard deviation *of X is*

$$\sigma = \sigma_X = \sqrt{\operatorname{Var}(X)}.$$

Remark 1.2. *The variance is defined for every random variable with finite expectation. Note, however, that it could be infinite.*

1.1 Basic properties of the variance

In the following, let X be a random variable with finite expectation μ.

Claim 1.3. $\operatorname{Var}(X) \geq 0$. *Moreover,* $\operatorname{Var}(X) = 0$ *if and only if* $\mathbb{P}(X = \mu) = 1$.

Proof. Let $f(x) = (x - \mu)^2$. Clearly, f is a non-negative function. Hence, it follows by the Monotonicity of Expectation (See Claim 1.15 in Lecture 9) that $\operatorname{Var}(X) = \mathbb{E}(f(X)) \geq \mathbb{E}(0) = 0$, and, moreover, equality is attained if and only if $\mathbb{P}(X = \mu) = \mathbb{P}(f(X) = 0) = 1$. □

Remark 1.4. *This shows, in particular, that the standard deviation is a non-negative real number.*

Claim 1.5. *For every $a \in \mathbb{R}$ it holds that* $\operatorname{Var}(aX) = a^2 \operatorname{Var}(X)$.

Proof. We have

$$\operatorname{Var}(aX) = \mathbb{E}\left((aX - \mathbb{E}(aX))^2\right) = \mathbb{E}\left((aX - a\mathbb{E}(X))^2\right) = a^2\mathbb{E}\left((X - \mathbb{E}(X))^2\right) = a^2 \operatorname{Var}(X),$$

where the second and third equalities hold by the linearity of expectation. □

Claim 1.6. *For every $b \in \mathbb{R}$ it holds that* $\operatorname{Var}(X + b) = \operatorname{Var}(X)$.

Proof. We have

$$\operatorname{Var}(X + b) = \mathbb{E}\left((X + b - \mathbb{E}(X + b))^2\right) = \mathbb{E}\left((X + b - \mathbb{E}(X) - b)^2\right) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \operatorname{Var}(X),$$

where the second equality holds by the linearity of expectation. □

The following claim presents a useful way of calculating variance.

Claim 1.7. $\operatorname{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Proof. We have

$$\begin{aligned}
\operatorname{Var}(X) &= \mathbb{E}\left((X - \mu)^2\right) \\
&= \mathbb{E}\left(X^2 - 2\mu \cdot X + \mu^2\right) \\
&= \mathbb{E}\left(X^2\right) - 2\mu\mathbb{E}(X) + \mu^2 \\
&= \mathbb{E}\left(X^2\right) - 2\mu^2 + \mu^2 \\
&= \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2,
\end{aligned}$$

where the third equality holds by the linearity of expectation. □

1.2 Variance of Common Distributions

1.2.1 Uniform Distribution

Recall the definition of the Uniform distribution: $X \sim \mathrm{U}(S)$ if

$$\mathbb{P}(X = s) = \begin{cases} \frac{1}{|S|} & s \in S \\ 0 & s \notin S \end{cases}$$

Claim 1.8. *If $S = \{a, a+1, \ldots, b\}$ for some $a, b \in \mathbb{N}$, then $\mathrm{Var}(X) = \frac{(b-a+1)^2-1}{12}$.*

Proof. Let $Y = X - a + 1$ and let $n = b - a + 1$. Then $Var(X) = Var(Y + a - 1) = Var(Y)$. Hence, it suffices to show that $\mathrm{Var}(Y) = \frac{n^2-1}{12}$. Observe that $Y \sim \mathrm{U}(1, \ldots, n)$ and thus, in particular, $\mathbb{E}(Y) = \frac{n+1}{2}$. We next calculate $\mathbb{E}(Y^2)$.

$$\begin{aligned} \mathbb{E}\left(Y^2\right) &= \sum_{y=1}^{n} y^2 \cdot \mathbb{P}(Y = y) \\ &= \sum_{y=1}^{n} y^2 \cdot \frac{1}{n} \\ &= \frac{1}{n} \cdot \sum_{y=1}^{n} y^2 \\ &= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} \\ &= \frac{(n+1)(2n+1)}{6}, \end{aligned}$$

where the first equality holds by Claim 0.1 and the fourth equality holds by the identity

$$\sum_{i=1}^{m} i^2 = \frac{m(m+1)(2m+1)}{6}.$$

We conclude that

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}(Y) \\ &= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\ &= \frac{n+1}{12} \cdot (4n + 2 - 3n - 3) \\ &= \frac{(n+1)(n-1)}{12} \\ &= \frac{n^2-1}{12} \end{aligned}$$

$\square$

1.2.2 Bernoulli Distribution

Recall the definition of the Bernoulli distribution: $X \sim \operatorname{Ber}(p)$ if

$$X \sim \begin{cases} 1 & p \\ 0 & 1-p \end{cases}$$

Claim 1.9. $\operatorname{Var}(X) = p(1-p)$.

Proof. As we have seen $\mathbb{E}(X) = p$. Therefore

$$\begin{aligned} \operatorname{Var}(X) &= \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2 \\ &= 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 \\ &= p(1-p). \end{aligned}$$

□

1.2.3 Binomial Distribution

Recall the definition of the Binomial distribution: $X \sim \operatorname{Bin}(n, p)$ if

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

for every integer $k \in \{0, 1, \ldots, n\}$.

Claim 1.10. $\operatorname{Var}(X) = np(1-p)$.

The proof is left as an exercise.