11. More rules for expectations

So far, we have covered some basic properties of expectations, including linearity, scaling, and conditioning rules. In this chapter, we develop some more intricate theoretical properties of the expectation operator.

**Proposition 11.1.** Suppose \( X \geq 0 \) is a non-negative, discrete random variable. Then

(i) \( \mathbb{E}X \geq 0 \), with equality if and only if \( P\{X = 0\} = 1 \), and

(ii) if \( X > 0 \), then \( \mathbb{E}X = \sum_{k=1}^{\infty} P\{X \geq k\} \).

Also, for general discrete random variables \( X, Y : \Omega \rightarrow \mathbb{R} \) with \( X(\omega) \leq Y(\omega) \) for all \( \omega \in \Omega \), we see that

(iii) \( \mathbb{E}X \leq \mathbb{E}Y \) and

(iv) if \( P\{X = Y\} = 1 \) then \( \mathbb{E}X = \mathbb{E}Y \). Conversely, if \( \mathbb{E}X \) and \( \mathbb{E}Y \) are finite, then \( \mathbb{E}X = \mathbb{E}Y \) implies \( P\{X = Y\} = 1 \).

**Proof.** Item (i) is obvious by the definition of the expectation. To show item (ii), notice

\[
\mathbb{E}X = \sum_{k=1}^{\infty} kP\{X = k\} \\
= \sum_{k=1}^{\infty} P\{X = k\} \sum_{j=1}^{\infty} I_{\{j \leq k\}} \\
= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} P\{X = k\}I_{\{k \geq j\}} \\
= \sum_{j=1}^{\infty} P\{X \geq j\}.
\]

Finally, item (iii) is obvious for \( \mathbb{E}X = -\infty \) or \( \mathbb{E}Y = \infty \). Suppose \( \mathbb{E}X > -\infty \) and \( \mathbb{E}Y < \infty \) and put \( D := Y - X \). Then

\[
0 \leq \mathbb{E}D = \mathbb{E}[Y - X] = \mathbb{E}Y - \mathbb{E}X.
\]

The converse of (iv) is a consequence of item (i).

11.1. Inequalities for expectations.

**Theorem 11.2** (Markov’s inequality). Suppose \( X \geq 0 \) and \( c > 0 \) is a constant. Then

\[
P\{X \geq c\} \leq \frac{1}{c} \mathbb{E}X,
\]

with equality if and only if \( X \) takes values \( 0 \) and \( c \) with probabilities \( p \) and \( 1 - p \), \( 0 < p < 1 \), respectively.

**Proof.** For any non-negative random variables \( X \) and \( c \geq 0 \), we have

\[
cI_{\{X \geq c\}} \leq X \quad \Rightarrow \quad \mathbb{E}(cI_{\{X \geq c\}}) \leq \mathbb{E}X \\
\quad \Rightarrow \quad cP\{X \geq c\} \leq \mathbb{E}X \\
\quad \Rightarrow \quad P\{X \geq c\} \leq \frac{1}{c} \mathbb{E}X.
\]

If \( P\{X = 0\} = p = 1 - P\{X = c\} \), then

\[
\mathbb{E}X = 0 \times p + c \times (1 - p) = cP\{X = c\} = cP\{X \geq c\}.
\]
There are various ways to generalize Markov's inequality \[^{(19)}\]. For example, let \(X \geq 0\) and \(n \in \mathbb{Z}^+\), then \(X^n \geq 0\) and

\[
P[X \geq c] = P[X^n \geq c^n] \leq c^{-n}EX^n.
\]

This approach leads to Chebyshev's inequality.

**Corollary 11.3** (Chebyshev's inequality). Let \(X\) be a random variable with finite mean \(\mu := E\{X\}\) and standard deviation \(SD(X) = \sigma\). Let \(c > 0\). Then

\[
P\{|X - \mu| \geq c\} \leq \frac{\sigma^2}{c^2},
\]

with equality if and only if \(X\) takes only 3 values \(\mu - c, \mu, \mu + c\) such that

\[
P[X = \mu - c] = P[X = \mu + c].
\]

**Proof.** Since \(|X - \mu|\) is a non-negative random variable, we have

\[
P\{|X - \mu| \geq c\} = P\{|X - \mu|^2 \geq c^2\} \leq \frac{E(X - \mu)^2}{c^2} = \frac{\sigma^2}{c^2}.
\]

The conclusion of Chebyshev’s inequality is that it is unlikely for any random variable to take a value several standard deviations from its mean. Notice, however, that, since Chebyshev’s inequality is true for all random variables with finite mean and variance, it gives a very crude upper bound. For example, for \(\sigma^2 = 2\) and \(c = 1\), \[^{(20)}\] asserts

\[
P\{|X - \mu| \geq c\} \leq 2,
\]

which is a tautology.

**Example 11.4** (Estimating a success probability). Suppose \(X\) has the Binomial distribution with parameter \((n, p)\) with \(n\) known and \(p\) unknown. We want to estimate the success probability \(p\) by \(\hat{p} = X/n\). How accurate is \(\hat{p}\)?

- \(E\hat{p} = EX/n = np/n = p\), so \(\hat{p}\) is an unbiased estimator for \(p\).
- Also, \(Var(\hat{p}) = Var(X)/n^2 = npq/n^2 = pq/n \leq 1/(4n)\).

Therefore,

\[
P\{|\hat{p} - p| \geq c\} \leq (4c^2n)^{-1}.
\]