1. Inequalities

Let us start with an inequality.

**Lemma 13.1.** If $h$ is a nonnegative function, then for all $\lambda > 0$,

$$
P\{h(X) \geq \lambda\} \leq \frac{E[h(X)]}{\lambda}.
$$

**Proof.** We know already that

$$
E[h(X)] = \sum_x h(x)f(x) \geq \sum_{x: h(x) \geq \lambda} h(x)f(x).
$$

If $x$ is such that $h(x) \geq \lambda$, then $h(x)f(x) \geq \lambda f(x)$, obviously. Therefore,

$$
E[h(X)] \geq \lambda \sum_{x: h(x) \geq \lambda} f(x) = \lambda P\{h(X) \geq \lambda\}.
$$

Divide by $\lambda$ to finish. \qed

Thus, for example,

$$
P\{|X| \geq \lambda\} \leq \frac{E(|X|)}{\lambda}
$$

“Markov’s inequality.”

$$
P\{|X - EX| \geq \lambda\} \leq \frac{\text{Var}(X)}{\lambda^2}
$$

“Chebyshev’s inequality.”

To get Markov’s inequality, apply Lemma 13.1 with $h(x) = |x|$. To get Chebyshev’s inequality, first note that $|X - EX| \geq \lambda$ if and only if $|X - EX|^2 \geq \lambda^2$. Then, apply Lemma 13.1 to find that

$$
P\{|X - EX| \geq \lambda\} \leq \frac{E(|X - EX|^2)}{\lambda^2}.
$$

Then, recall that the numerator is $\text{Var}(X)$. 

45
In words:

- If $E(|X|) < \infty$, then the probability that $|X|$ is large is small.
- If $\text{Var}(X)$ is small, then with high probability $X \approx EX$.

2. Conditional distributions

If $X$ is a random variable with mass function $f$, then $\{X = x\}$ is an event. Therefore, if $B$ is also an event, and if $P(B) > 0$, then

$$P(X = x \mid B) = \frac{P(\{X = x\} \cap B)}{P(B)}.$$ 

As we vary the variable $x$, we note that $\{X = x\} \cap B$ are disjoint. Therefore,

$$\sum_x P(X = x \mid B) = \sum_x \frac{P(\{X = x\} \cap B)}{P(B)} = \frac{P(\cup_x \{X = x\} \cap B)}{P(B)} = 1.$$ 

Thus,

$$f(x \mid B) = P(X = x \mid B)$$

defines a mass function also. This is called the conditional mass function of $X$ given $B$.

Example 13.2. Let $X$ be distributed uniformly on $\{1, \ldots, n\}$, where $n$ is a fixed positive integer. Recall that this means that

$$f(x) = \begin{cases} \frac{1}{n} & \text{if } x = 1, \ldots, n, \\ 0 & \text{otherwise}. \end{cases}$$

Choose and fix two integers $a$ and $b$ such that $1 \leq a \leq b \leq n$. Then,

$$P\{a \leq X \leq b\} = \sum_{x=a}^{b} \frac{1}{n} = \frac{b - a + 1}{n}.$$ 

Therefore,

$$f(x \mid a \leq X \leq b) = \begin{cases} \frac{1}{b - a + 1} & \text{if } x = a, \ldots, b, \\ 0 & \text{otherwise}. \end{cases}$$
3. Conditional expectations

Once we have a (conditional) mass function, we have also a conditional expectation at no cost. Thus,

$$E(X \mid B) = \sum_x x f(x \mid B).$$

**Example 13.3** (Example 13.2, continued). In Example 13.2,

$$E(X \mid a \leq X \leq b) = \sum_{k=a}^{b} \frac{k}{b-a+1}.$$ 

Now,

$$\sum_{k=a}^{b} k = \sum_{k=1}^{b} k - \sum_{k=1}^{a-1} k = \frac{b(b+1)}{2} - \frac{(a-1)a}{2} = \frac{b^2 + b - a^2 + a}{2}.$$ 

Write $b^2 - a^2 = (b-a)(b+a)$ and factor $b+a$ to get

$$\sum_{k=a}^{b} k = \frac{b+a}{2}(b-a+1).$$ 

Therefore,

$$E(X \mid a \leq X \leq b) = \frac{b+a}{2}.$$ 

This should not come as a surprise. Example 13.2 actually shows that given $B = \{a \leq X \leq b\}$, the conditional distribution of $X$ given $B$ is uniform on $\{a, \ldots, b\}$. Therefore, the conditional expectation is the expectation of a uniform random variable on $\{a, \ldots, b\}$.

**Theorem 13.4** (Bayes’s formula for conditional expectations). If $P(B) > 0$, then

$$E(X) = E(X \mid B)P(B) + E(X \mid B^c)P(B^c).$$

**Proof.** We know from the ordinary Bayes’s formula that

$$f(x) = f(x \mid B)P(B) + f(x \mid B^c)P(B^c).$$

Multiply both sides by $x$ and add over all $x$ to finish. \qed
Remark 13.5. The more general version of Bayes’s formula works too here: Suppose $B_1, B_2, \ldots$ are disjoint and $\bigcup_{i=1}^{\infty} B_i = \Omega$; i.e., “one of the $B_i$’s happens.” Then,

$$EX = \sum_{i=1}^{\infty} E(X \mid B_i)P(B_i).$$

Example 13.6. Suppose you play a fair game repeatedly. At time 0, before you start playing the game, your fortune is zero. In each play, you win or lose with probability $1/2$. Let $T_1$ be the first time your fortune becomes +1. Compute $E(T_1)$.

More generally, let $T_x$ denote the first time to win $x$ dollars, where $T_0 = 0$.

Let $W$ denote the event that you win the first round. Then, $P(W) = P(W^c) = 1/2$, and so

$$E(T_x) = \frac{1}{2}E(T_x \mid W) + \frac{1}{2}E(T_x \mid W^c).$$

Suppose $x \neq 0$. Given $W$, $T_x$ is one plus the first time to make $x - 1$ more dollars. Given $W^c$, $T_x$ is one plus the first time to make $x + 1$ more dollars. Therefore,

$$E(T_x) = \frac{1}{2} \left[ 1 + E(T_{x-1}) \right] + \frac{1}{2} \left[ 1 + E(T_{x+1}) \right] = 1 + \frac{E(T_{x-1}) + E(T_{x+1})}{2}.$$ 

Also $E(T_0) = 0$.

Let $g(x) = E(T_x)$. This shows that $g(0) = 0$ and

$$g(x) = 1 + \frac{g(x+1) + g(x-1)}{2} \quad \text{for } x = \pm 1, \pm 2, \ldots.$$ 

Because $g(x) = (g(x) + g(x))/2$,

$$g(x) + g(x) = 2 + g(x+1) + g(x-1) \quad \text{for } x = \pm 1, \pm 2, \ldots.$$ 

Solve to find that for all integers $x \geq 1$,

$$g(x+1) - g(x) = -2 + g(x) - g(x-1).$$
Example 14.1 (St.-Petersbourg paradox, continued). We continued with our discussion of the St.-Petersbourg paradox, and note that for all integers \( N \geq 1 \),

\[
g(N) = g(1) + \sum_{k=2}^{N} (g(k) - g(k-1))
\]

\[
= g(1) + \sum_{k=1}^{N-1} (g(k + 1) - g(k))
\]

\[
= g(1) + \sum_{k=1}^{N-1} (-2 + g(k) - g(k-1))
\]

\[
= g(1) - 2(N - 1) + \sum_{k=1}^{N} (g(k) - g(k-1))
\]

\[
= g(1) - 2(N - 1) + g(N).
\]

If \( g(1) < \infty \), then \( g(1) = 2(N - 1) \). But \( N \) is arbitrary. Therefore, \( g(1) \) cannot be finite; i.e.,

\[ E(T_1) = \infty. \]

This shows also that \( E(T_x) = \infty \) for all \( x \geq 1 \), because for example \( T_2 \geq 1 + T_1 \). By symmetry, \( E(T_x) = \infty \) if \( x \) is a negative integer as well.