Chapter 7 (section 7.4):
Random Variables, Expectation, and
Variance

Kousha Etessami

U. of Edinburgh, UK
Expected Value (Expectation) of a Random Variable

Recall: A **random variable** (r.v.), is a function $X : \Omega \rightarrow \mathbb{R}$, that assigns a real value to each outcome in a sample space $\Omega$.

The **expected value**, or **expectation**, or **mean**, of a random variable $X : \Omega \rightarrow \mathbb{R}$, denoted by $E(X)$, is defined by:

$$E(X) = \sum_{s \in \Omega} P(s)X(s)$$

Here $P : \Omega \rightarrow [0, 1]$ is the underlying probability distribution on $\Omega$.

**Question:** Let $X$ be the r.v. outputing the number that comes up when a fair die is rolled. What is the expected value, $E(X)$, of $X$?
Expected Value (Expectation) of a Random Variable

Recall: A random variable (r.v.), is a function $X : \Omega \rightarrow \mathbb{R}$, that assigns a real value to each outcome in a sample space $\Omega$.

The expected value, or expectation, or mean, of a random variable $X : \Omega \rightarrow \mathbb{R}$, denoted by $E(X)$, is defined by:

$$E(X) = \sum_{s \in \Omega} P(s)X(s)$$

Here $P : \Omega \rightarrow [0, 1]$ is the underlying probability distribution on $\Omega$.

Question: Let $X$ be the r.v. outputing the number that comes up when a fair die is rolled. What is the expected value, $E(X)$, of $X$?

Answer:

$$E(X) = \sum_{i=1}^{6} \frac{1}{6} \cdot i = \frac{21}{6} = \frac{7}{2}.$$
A bad way to calculate expectation

The definition of expectation, \( E(X) = \sum_{s \in \Omega} P(s)X(s) \), can be used directly to calculate \( E(X) \). But sometimes this is horribly inefficient.

**Example:** Suppose that a biased coin, which comes up heads with probability \( p \) each time, is flipped 11 times consecutively.

**Question:** What is the expected # of heads?
A bad way to calculate expectation

The definition of expectation, \( E(X) = \sum_{s \in \Omega} P(s)X(s) \), can be used directly to calculate \( E(X) \). But sometimes this is horribly inefficient.

Example: Suppose that a biased coin, which comes up heads with probability \( p \) each time, is flipped 11 times consecutively.

Question: What is the expected # of heads?

Bad way to answer this: Let’s try to use the definition of \( E(X) \) directly, with \( \Omega = \{ H, T \}^{11} \). Note that \( |\Omega| = 2^{11} = 2048 \).

So, the sum \( \sum_{s \in \Omega} P(s)X(s) \) has 2048 terms!

This is clearly not a practical way to compute \( E(X) \).

Is there a better way? Yes.
Better expression for the expectation

Recall $P(X = r)$ denotes the probability $P(\{s \in \Omega \mid X(s) = r\})$. Recall that for a function $X : \Omega \rightarrow \mathbb{R}$,

$$\text{range}(X) = \{ r \in \mathbb{R} \mid \exists s \in \Omega \text{ such that } X(s) = r \}$$

**Theorem:** For a random variable $X : \Omega \rightarrow \mathbb{R}$,

$$E(X) = \sum_{r \in \text{range}(X)} P(X = r) \cdot r$$

**Proof:** $E(X) = \sum_{s \in \Omega} P(s)X(s)$, but for each $r \in \text{range}(X)$, if we sum all terms $P(s)X(s)$ such that $X(s) = r$, we get $P(X = r) \cdot r$ as their sum. So, summing over all $r \in \text{range}(X)$ we get $E(X) = \sum_{r \in \text{range}(X)} P(X = r) \cdot r$.

So, if $|\text{range}(X)|$ is small, and if we can compute $P(X = r)$, then we need to sum a lot fewer terms to calculate $E(X)$. 
Expected # of successes in $n$ Bernoulli trials

**Theorem:** The expected # of successes in $n$ (independent) Bernoulli trials, with probability $p$ of success in each, is $np$.

**Note:** We’ll see later that we do not need independence for this.

**First, a proof which uses mutual independence:** For $\Omega = \{ H, T \}^n$, let $X : \Omega \rightarrow \mathbb{N}$ count the number of successes in $n$ Bernoulli trials. Let $q = (1 - p)$. Then...

$$E(X) = \sum_{k=0}^{n} P(X = k) \cdot k$$

$$= \sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} \cdot k$$

The second equality holds because, assuming mutual independence, $P(X = k)$ is the binomial distribution $b(k; n, p)$. 
We will soon see this was an unnecessarily complicated proof.

$$E(X) = \sum_{k=0}^{n} P(X = k) \cdot k = \sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} \cdot k =$$

$$= \sum_{k=1}^{n} \frac{n!}{k!(n-k)!} p^k q^{n-k} \cdot k = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k}$$

$$= \sum_{k=1}^{n} n \cdot \frac{(n-1)!}{(k-1)!(n-k)!} p^k q^{n-k} = n \sum_{k=1}^{n} \binom{n-1}{k-1} p^k q^{n-k}$$

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j}$$

$$= np(p + q)^{n-1} = np.$$

We will soon see this was an unnecessarily complicated proof.
Expectation of a geometrically distributed r.v.

**Question:** A coin comes up heads with probability $p > 0$ each time it is flipped. The coin is flipped repeatedly until it comes up heads. What is the expected number of times it is flipped?

**Note:** This simply asks: “What is the expected value $E(X)$ of a geometrically distributed random variable with parameter $p$?”

**Answer:** $\Omega = \{H, TH, TTH, \ldots\}$, and $P(T^{k-1}H) = (1 - p)^{k-1}p$. And clearly $X(T^{k-1}H) = k$. Thus $E(X) = \sum_{s \in \Omega} P(s)X(s) =

$$E(X) = \sum_{k=1}^{\infty} (1 - p)^{k-1}p \cdot k = p \sum_{k=1}^{\infty} k(1 - p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

This is because: $\sum_{k=1}^{\infty} k \cdot x^{k-1} = \frac{1}{(1-x)^2}$, for $|x| < 1$.

**Example:** If $p = 1/4$, then the expected number of coin tosses before we see Heads for the first time is 4.
**Linearity of Expectation (VERY IMPORTANT)**

**Theorem (Linearity of Expectation):** For any random variables \( X, X_1, \ldots, X_n \) on \( \Omega \),
\[
E(X_1 + X_2 + \ldots + X_n) = E(X_1) + \ldots + E(X_n).
\]

Furthermore, for any \( a, b \in \mathbb{R} \),
\[
E(aX + b) = aE(X) + b.
\]
(In other words, the expectation function is a **linear function**.)

**Proof:**
\[
E\left( \sum_{i=1}^{n} X_i \right) = \sum_{s \in \Omega} P(s) \sum_{i=1}^{n} X_i(s) = \sum_{i=1}^{n} \sum_{s \in \Omega} P(s)X_i(s) = \sum_{i=1}^{n} E(X_i).
\]
\[
E(aX + b) = \sum_{s \in \Omega} P(s)(aX(s) + b) = (a \sum_{s \in \Omega} P(s)X(s)) + b \sum_{s \in \Omega} P(s)
\]
\[
= aE(X) + b.
\]
Using linearity of expectation

**Theorem:** The expected # of successes in \( n \) (not necessarily independent) Bernoulli trials, with probability \( p \) of success in each trial, is \( np \).

**Easy proof, via linearity of expectation:** For \( \Omega = \{H, T\}^n \), let \( X \) be the r.v. counting the expected number of successes, and for each \( i \), let \( X_i : \Omega \to \mathbb{R} \) be the binary r.v. defined by:

\[
X_i((s_1, \ldots, s_n)) = \begin{cases} 
1 & \text{if } s_i = H \\
0 & \text{if } s_i = T
\end{cases}
\]

Note that \( E(X_i) = p \cdot 1 + (1 - p) \cdot 0 = p \), for all \( i \in \{1, \ldots, n\} \).

Also, clearly, \( X = X_1 + X_2 + \ldots + X_n \), so:

\[
E(X) = E(X_1 + \ldots + X_n) = \sum_{i=1}^{n} E(X_i) = np. \quad \square
\]

Note: this holds even if the \( n \) coin tosses are totally correlated.
Using linearity of expectation, continued

**Hatcheck problem:** At a restaurant, the hat-check person forgets to put claim numbers on hats. $n$ customers check their hats in, and they each get a random hat back when they leave the restaurant. What is the expected number, $E(X)$, of people who get their correct hat back?

**Answer:** Let $X_i$ be the r.v. that is 1 if the $i$’th customer gets their hat back, and 0 otherwise. Clearly, $E(X) = E(\sum_i X_i)$. Furthermore, $E(X_i) = P(i$’th person gets its hat back$) = 1/n$. Thus, $E(X) = n \cdot (1/n) = 1$.

This would be much harder to prove without using the linearity of expectation.

**Note:** $E(X)$ doesn’t even depend on $n$ in this case.
Independence of Random Variables

**Definition:** Two random variables, $X$ and $Y$, are called independent if for all $r_1, r_2 \in \mathbb{R}$:

$$P(X = r_1 \text{ and } Y = r_2) = P(X = r_1) \cdot P(Y = r_2)$$

**Example:** Two die are rolled. Let $X_1$ be the number that comes up on die 1, and let $X_2$ be the number that comes up on die 2. Then $X_1$ and $X_2$ are independent r.v.’s.

**Theorem:** If $X$ and $Y$ are independent random variables on the same space $\Omega$. Then

$$E(XY) = E(X)E(Y)$$

We will not prove this in class. (The proof is a simple re-arrangement of the sums in the definition of expectation. See Rosen’s book for a proof.)
Variance

The “variance” and “standard deviation” of a r.v., \( X \), give us ways to measure (roughly) “on average, how far off the value of the r.v. is from its expectation”.

**Variance and Standard Deviation**

**Definition:** For a random variable \( X \) on a sample space \( \Omega \), the variance of \( X \), denoted by \( V(X) \), is defined by:

\[
V(X) = E((X - E(X))^2) = \sum_{s \in \Omega} (X(s) - E(X))^2 P(s)
\]

The standard deviation of \( X \), denoted \( \sigma(X) \), is defined by

\[
\sigma(X) = \sqrt{V(X)}
\]
Example, and a useful identity for variance

Example: Consider the r.v., \( X \), such that \( P(X = 0) = 1 \), and the r.v. \( Y \), such that \( P(Y = -10) = P(Y = 10) = 1/2 \). Then \( E(X) = E(Y) = 0 \), but \( V(X) = 0 = \sigma(X) \), whereas \( V(Y) = 100 \) and \( \sigma(Y) = 10 \).

Theorem: For any random variable \( X \),

\[
V(X) = E(X^2) - E(X)^2
\]

Proof:

\[
V(X) = E((X - E(X))^2) \\
= E(X^2 - 2XE(X) + E(X)^2) \\
= E(X^2) - 2E(X)E(X) + E(X)^2 \\
= E(X^2) - E(X)^2.
\]