Theory of Probability - Brett Bernstein

Lecture 8

Review Exercises

1. Let $X$ denote the value of rolling a 20-sided die, and let $g : \mathbb{R} \to \mathbb{R}$ be defined as follows:

$$
g(x) = \begin{cases} 
1 & \text{if } x > 12, \\
0 & \text{if } x = 12, \\
-1 & \text{if } x < 12.
\end{cases}
$$

What is $E[g(X)]$?

2. You roll a 4-sided die and flip a fair coin. Let $X$ denote the value of the die, and let $Y$ denote the number of heads from the flip.

(a) What is $P(X + Y = 3)$?

(b) What is $P((X + Y)^2 = 4)$?

(c) What is $E[(X - 2Y)^2]$?

Solutions

1. Here we give a full solution.

$$
E[g(X)] = 1 \cdot p_{g(X)}(1) + (-1)p_{g(X)}(-1)
$$

$$
= 1 \cdot P(g(X) = 1) + (-1)P(g(X) = -1)
$$

$$
= 1 \cdot P(\{s \in S : g(X(s)) = 1\}) + (-1)P(\{s \in S : g(X(s)) = -1\})
$$

$$
= 1 \cdot P(\{s \in S : X(s) > 12\}) + (-1)P(\{s \in S : X(s) < 12\})
$$

$$
= 1 \cdot \frac{8}{20} + (-1) \cdot \frac{11}{20}
$$

$$
= -\frac{3}{20}.
$$

2. (a) Let $S = \{(d, f) : 1 \leq d \leq 4, f \in \{H, T\}\}$ with all outcomes equally likely. Then we have

$$
P(X + Y = 3) = P(\{s \in S : X(s) + Y(s) = 3\}) = P(\{(2, 1), (3, 0)\}) = \frac{2}{8}.
$$
(b) 

\[ P((X + Y)^2 = 4) = P\{s \in S : (X(s) + Y(s))^2 = 4\} \]

\[ = P\{s \in S : X(s) + Y(s) = 2\} \]

\[ = P\{(2, 0), (1, 1)\} \]

\[ = \frac{2}{8}. \]

(c) Let \( Z = X - 2Y \). Then we have

\[ E[(X - 2Y)^2] = 0 \cdot p_{Z^2}(0) + 1 \cdot p_{Z^2}(1) + 4p_{Z^2}(4) + 9p_{Z^2}(9) + 16p_{Z^2}(16) \]

\[ = 1 \cdot P(Z \in \{-1, 1\}) + 4P(Z = 2) + 9P(Z = 3) + 16P(Z = 4) \]

\[ = 1 \cdot \frac{3}{8} + 4 \cdot \frac{2}{8} + 9 \cdot \frac{1}{8} + 16 \cdot \frac{1}{8} \]

\[ = \frac{36}{8}. \]

Expectation of a Function of a Random Variable

Example 1.

Letting \( Y = f(X) \) we compute the following probabilties:

\[ P(Y = 7) = P\{\{a, b\}\} \quad \text{and} \quad P(Y = 8) = P\{\{c, d, e\}\}. \]

Alternatively, we can look at \( f \) as relabeling the labels assigned by \( X \). By unrelabeling we have

\[ P(Y = 7) = P(X = 1) = p_X(1) \quad \text{and} \quad P(Y = 8) = P(X \in \{2, 3\}) = p_X(2) + p_X(3). \]

We can also compute \( E[Y] \) via unrelabeling:

\[ E[Y] = 7 \cdot p_Y(7) + 8 \cdot p_Y(8) = 7 \cdot p_X(1) + 8(p_X(2) + p_X(3)) = f(1)p_X(1) + f(2)p_X(2) + f(3)p_X(3). \]
Using the above example as motivation, we generalize the result below. To help us compute the expected value of functions of random variables we have the LOTUS theorem (which is essentially computing expectations via “unrelabeling”):

**Theorem 2** (Law of the Unconscious Statistician or LOTUS). *Let $X$ be a discrete random variable, and let $f : \mathbb{R} \to \mathbb{R}$ be a function. Then

$$E[f(X)] = \sum_{x : p_X(x) > 0} f(x)p_X(x).$$

To prove this we will use the following which says we can compute the probability $Y = y$ by “unrelabeling”.

**Lemma 3.** *Let $X$ be a discrete random variable and let $Y = f(X)$ for some function $f$. Then we have

$$p_Y(y) = \sum_{x : p_X(x) > 0, f(x) = y} p_X(x).$$

**Proof.** We first partition $S$ by the $X$-labels:

$$S = \bigcup_{x : p_X(x) > 0} \{s : X(s) = x\},$$

a disjoint union that is countable by discreteness. Splitting by this partition we can write $A = \{s : Y(s) = y\}$ as

$$A = \bigcup_{x : p_X(x) > 0} \{s : X(s) = x\} \cap A$$

Thus

$$p_Y(y) = P(A) = \sum_{x : p_X(x) > 0} P(\{s : X(s) = x\} \cap A) = \sum_{x : p_X(x) > 0, f(x) = y} p_X(x).$$

Now we prove the theorem.

**Proof.** Let $Y = f(X)$ so that

$$E[f(X)] = \sum_{y : p_Y(y) > 0} y p_Y(y)$$

$$= \sum_{y : p_Y(y) > 0} y \sum_{x : p_X(x) > 0, f(x) = y} p_X(x)$$

$$= \sum_{y : p_Y(y) > 0} \sum_{x : p_X(x) > 0, f(x) = y} f(x)p_X(x)$$

$$= \sum_{x : p_X(x) > 0} f(x)p_X(x).$$

3
The last equation holds since $\sum p_X(x) = 1$ and these $x$-values can be grouped by common $y$-value thus giving all possible $y$-values.

**Expectation of a Function of an RV Exercises**

1. Suppose $X$ is a random variable that takes on integer values so that
   \[ \sum_{k \in \mathbb{Z}} p_X(k) = 1. \]

   Give an expression for
   \[ E \left[ \sin \left( \frac{\pi}{8} X \right) \right]. \]

2. Let $f(x) = ax + b$ where $a, b \in \mathbb{R}$ are fixed constants. Show that
   \[ E[f(X)] = E[aX + b] = aE[X] + b, \]
   when $X$ is a discrete random variable.

3. Let $X$ be a discrete random variable. Give a formula for $E[X^n]$ where $n$ is a fixed non-negative integer. This is called the $n$th moment of $X$.

4. The variance of a random variable $X$ is defined to be $\text{Var}[X] = E[(X - E[X])^2]$ (also called the second central moment).
   
   (a) Compute the variance of rolling a 6-sided die.
   
   (b) (⋆) Assuming $X$ is discrete, show that the variance is given by $\text{Var}[X] = E[X^2] - E[X]^2$.
   
   (c) Assuming $X$ is discrete, derive a formula for $\text{Var}[aX + b]$ where $a, b \in \mathbb{R}$ are fixed constants.

**Solutions**

1. Using LOTUS we have
   \[ E \left[ \sin \left( \frac{\pi}{8} X \right) \right] = \sum_{k \in \mathbb{Z}} \sin(k\pi/8)p_X(k). \]
2. Using LOTUS we have

\[ E[f(X)] = \sum_{x: p_X(x) > 0} f(x)p_X(x) \]

\[ = \sum_{x: p_X(x) > 0} (ax + b)p_X(x) \]

\[ = a \sum_{x: p_X(x) > 0} xp_X(x) + b \sum_{x: p_X(x) > 0} p_X(x) \]

\[ = aE[X] + b. \]

3. Using LOTUS we have

\[ E[X^n] = \sum_{x: p_X(x) > 0} x^n p_X(x). \]

4. (a) \( E[X] = 3.5 \) so using LOTUS

\[ E[(X - 3.5)^2] = \frac{1}{6} \left( (-2.5)^2 + (-1.5)^2 + (-.5)^2 + (.5)^2 + (1.5)^2 + (2.5)^2 \right) = \frac{35}{12} \]

(b) Using throughout that \( E[X] \) can be treated as a constant we have

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

\[ = E[X^2] - 2E[X]E[X] + E[E[X]^2] \quad \text{(Linearity needed)} \]

\[ = E[X^2] - E[X]^2. \]

The above argument works if we have a property of expectation called linearity. We will prove this in the next section.

(c) Note that

\[ \text{Var}[aX + b] = E[(aX + b - E[aX + b])^2] \]

\[ = E[(aX + b - aE[X] - b)^2] \]

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

\[ = a^2 \text{Var}[X]. \]
Introduction to Multiple Discrete Random Variables

Let $X, Y$ be (arbitrary) random variables defined on the same sample space. We say they are independent if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

for all events $A, B \subset \mathbb{R}$. In words, if $X, Y$ are independent then knowing $X$ is in some $A$ doesn’t change the probability that $Y$ is in some other set $B$. An equivalent statement is that $X, Y$ are independent iff

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) = F_X(x)F_Y(y).$$

The proof is done in a measure theory class.

For discrete random variables, there is a simpler formula. If $X, Y$ are discrete, they are independent iff

$$P(X = x, Y = y) = p_X(x)p_Y(y)$$

for all $x, y \in \mathbb{R}$.

This can be extended to $n$ random variables $X_1, \ldots, X_n$ on the same sample space by requiring

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n)$$

to hold for all sets $A_1, \ldots, A_n \subset \mathbb{R}$. The result for PMFs can be extended to $n$ variables as well:

$$P(X_1 = x_1, \ldots, X_n = x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n).$$

We now illustrate a general theme that will be true for the rest of the course (with the strong law of large numbers being one exception), and in the process prove a very useful theorem as a preview to later material. The theme is that knowing the PMF of a discrete random variable tells you everything you need to know about it. More generally, if we know

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

for a list of $n$ random variables (called the joint PMF), we know everything we need to know about their distributions, and how they interact.

Amazingly, the following result doesn’t require independence.

**Theorem 4 (Linearity of Expectation).** Let $X, Y$ be discrete random variables (defined on the same sample space) whose expectations exist. Then

$$E[X + Y] = E[X] + E[Y].$$

Furthermore, for any $a \in \mathbb{R}$ we have

$$E[aX] = aE[X].$$
This holds for general random variables, as we will see when we learn expectations in a more general context.

We first prove an intuitive and helpful lemma which is similar to the lemma we needed for LOTUS.

**Lemma 5.**

\[ P(X + Y = z) = \sum_{p_X(x) > 0, p_Y(y) > 0, x + y = z} P(X = x, Y = y). \]

**Proof.** Let \( S \) denote our sample space and consider the event

\[ E = \{ s : X(s) + Y(s) = z \}. \]

As \( X \) is discrete, we can express this as a countable disjoint union:

\[ E = \bigcup_{x : p_X(x) > 0} \{ s : X(s) = x, Y(s) = z - x \}. \]

Using countable additivity we have

\[
\begin{align*}
P(X + Y = z) &= P(E) \\
&= \sum_{x : p_X(x) > 0} P\{ s : X(s) = x, Y(s) = z - x \} \\
&= \sum_{p_X(x) > 0, p_Y(y) > 0, x + y = z} P\{ s : X(s) = x, Y(s) = y \} \\
&= \sum_{p_X(x) > 0, p_Y(y) > 0, x + y = z} P\{ X = x, Y = y \}.
\end{align*}
\]

This extends to multiple random variables:

**Corollary 6.** If \( X_1, \ldots, X_n \) are discrete then

\[ P(X_1 + \cdots + X_n = z) = \sum_{x_1 + \cdots + x_n = z} \sum_{p_{X_i}(x_i) > 0} P(X_1 = x_1, \cdots, X_n = x_n). \]

Now we prove the theorem.
Proof. Let $Z = X + Y$. Then

$$E[Z] = \sum_{p_Z(z) > 0} z P(Z = z)$$

$$= \sum_{p_Z(z) > 0} z \sum_{p_X(x) > 0, p_Y(y) > 0, x + y = z} P\{X = x, Y = y\}$$

$$= \sum_{p_Z(z) > 0} \sum_{p_X(x) > 0, p_Y(y) > 0, x + y = z} (x + y) P\{X = x, Y = y\}$$

$$= \sum_{p_X(x) > 0, p_Y(y) > 0} (x + y) P\{X = x, Y = y\}$$

$$= \sum_{p_X(x) > 0, p_Y(y) > 0} x P\{X = x\} + \sum_{p_Y(y) > 0} y P\{Y = y\}.$$ 

The result about $aX$ was proven earlier due to LOTUS.

Applying induction we obtain:

**Corollary 7.** Let $X_1, \ldots, X_n$ be random variables whose expectations exist. Then we have

$$E\left[\sum_{k=1}^{n} X_k\right] = \sum_{k=1}^{n} E[X_k].$$

**Introduction to Multiple Discrete RVs Exercises**

1. Suppose we roll $n$ fair 6-sided dice, and let the random variable $Y_k$ denote the value of the $k$th die. Show that the $Y_k$, for $k = 1, \ldots, n$ are independent.

2. Suppose $X, Y$ are discrete random variables that take values in the integers (i.e., the PMFs only assign non-zero probability to integers).
   
   (a) What is the probability that $X + Y = 3$?
   
   (b) Repeat the above assuming $X, Y$ are independent.

3. Suppose we have an urn with 100 red balls, and 200 black balls. Suppose we draw $k \leq 300$ balls from the urn in sequence, and let $X_i$ be the indicator of whether the $i$th ball drawn is black (i.e., 1 if black, 0 if not).
   
   (a) If the balls are drawn with replacement, what is the expected number of black balls?
(b) If the balls are drawn without replacement, what is the expected number of black balls?
(c) If the balls are drawn without replacement, are $X_1$ and $X_2$ independent?

4. In class of $n$ students, how many pairs of students are expected to have the same birthday (assume the year has 365 days each equally likely to be a birthday)?

Solutions

1. Consider $S = \{(d_1, \ldots, d_n) : 1 \leq d_k \leq 6\}$.

$$P(Y_1 = d_1, \ldots, Y_n = d_n) = P((d_1, \ldots, d_n)) = \frac{1}{6^n} = \prod_{k=1}^{n} P(Y_k = d_k).$$

2. (a) Here are two equivalent ways of writing this:

$$P(X + Y = 3) = \sum_{k \in \mathbb{Z}} P(X = k, Y = 3 - k) = \sum_{k \in \mathbb{Z}} P(X = 3 - k, Y = k).$$

This is the discrete version of an operation called a convolution.

(b) $P(X + Y = 3) = \sum_{k \in \mathbb{Z}} P(X = k, Y = 3 - k) = \sum_{k \in \mathbb{Z}} P(X = k)P(Y = 3 - k)$.

3. (a) The number of black balls drawn is given by $X_1 + X_2 + \cdots + X_k$. Since $E[X_i] = 2/3$ for all $i$ we obtain $2k/3$.

(b) We first show that $E[X_i] = 2/3$ for all $i$. One way to see this is that when counting possible sequences of draws, the problem is symmetric with respect to the positions, so every $X_i$ has the same probability of being 1 as the first. More explicitly, note that

$$P(X_i = 1) = \frac{200}{300} \cdot \frac{299!}{(299-(k-1))!} = \frac{200}{300} \cdot \frac{299!}{(300-k)!}.$$

Thus $E[X_i] = 2/3$ so the full answer is again $2k/3$.

(c) No. Note that

$$P(X_1 = 1, X_2 = 1) = \frac{200 \cdot 199}{300} \cdot \frac{298!}{(298-(k-2))!} = \frac{200 \cdot 199}{300} \cdot \frac{298!}{(300-k)!}.$$

This is not $P(X_1 = 1)P(X_2 = 1) = 4/9$.

4. Each pair of students has the same birthday with probability $1/365$ so the answer is

$$\binom{n}{2} \cdot \frac{1}{365}.$$