Notes 22

1 Functions of Random Variables

Recall: A *Random Variable*, R, is a functions $R : S \to \mathbb{R}$, where S denotes the sample space.

Given several random variables, we can combine them into more complicated random variables.

Example: Given 2 six-sided dice (fair, independent), let R_1 be the result of 1st die and R_2 be the result of the 2nd die.

We can define $T = R_1 R_2$. Here, T is a product of two random variables. Note that T is also a random variable. Other combinations of R_1 and R_2 also form random variable. For example, $W = R_1^2 + R_2^2$ is also a random variable.

We often want to compute the distribution of these new random variables. The distribution of $T(= R_1 R_2)$ is

$$Pr(T = x) = \sum_{\{(x_1,x_2):x_1x_2=x\}} Pr(R_1 = x_1 \wedge R_2 = x_2)$$

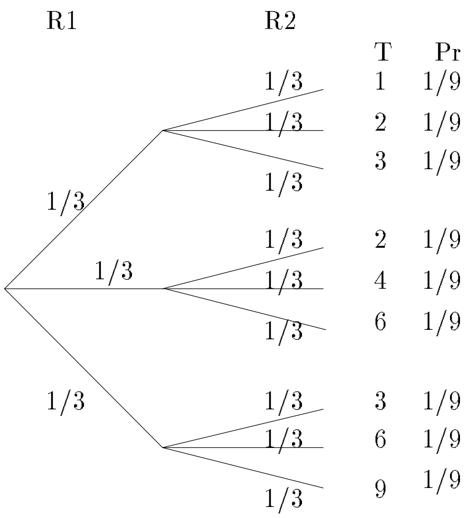
How is this related to our original rule for determining the distribution for a random variable? The original rule says we are supposed to sum over all points in the sample space, but instead we are summing only over values of the original random variables. Why does this work? It follows from the *theorem of total probability*: since at any given sample space point R_1 has a particular value, the events $[R_1 = x_1]$ partition the sample space as required. Therefore,

$$\begin{aligned} \Pr[T = x] &= \sum_{x_1} \Pr[T = x \wedge R_1 = x_1] \\ &= \sum_{x_1} \Pr[R_1 = x_1 \wedge X_2 = x - x_1] \end{aligned}$$

which evaluates to the above sum.

Conveniently, we don't need to consider the original sample space at all—only the values taken on by the underlying random variables.

Let's compute $Pr(T = X)$. For simplicity, let's consider 3-sided dice. The sample space is



$Pr(T = 1) = \frac{1}{9}$

$Pr(T = 2) = \frac{2}{9}$

$Pr(T = 3) = \frac{2}{9}$

$Pr(T = 4) = \frac{1}{9}$

$Pr(T = 5) = 0$

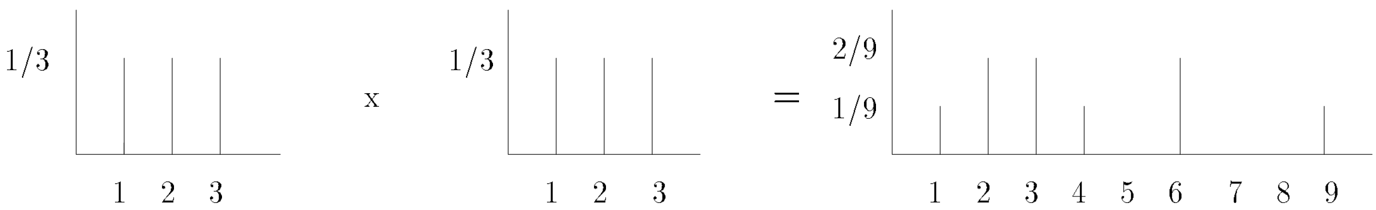
$Pr(T = 6) = \frac{2}{9}$

$Pr(T = 7) = 0$

$Pr(T = 8) = 0$

$Pr(T = 9) = \frac{1}{9}$

This distribution looks like



Although R_1 and R_2 are independent and uniform, the product is not uniform or symmetric. We can apply this technique to any function of random variables. This works even if the variables are not independent. As another example, let $U = R_1 + R_2$, then

$Pr(U = 2) = \frac{1}{9}$

$Pr(U=3)=\frac{2}{9}$

$Pr(U=4)=\frac{3}{9}$

$Pr(U=5)=\frac{2}{9}$

$Pr(U=6)=\frac{1}{9}$

As a sum of independent random variables, U has an interesting probability distribution:

$$\Pr[U=n]=\sum_k \Pr[R_1=k]\Pr[R_2=n-k].$$

This can be interpreted as follows: the distribution for U is the *convolution* (remember generating functions?) of the distributions for R_1 and R_2.

2 Expectation

Expectation compresses a whole distribution down to one number (obviously losing some info on the way).

Definition 2.1 *The* **expected value** *or* **mean** *of random variable R is*

$$E[R]=\sum_x x\cdot\Pr(R=x)$$

Intuitively, the expectation is the weighted average of all possible values. For continuous random variables, the sum is replaced by an integral. Other terms for expectation are the mean or average. Note: the mean is not the same as the median.

Definition 2.2 *The* **median** *of R is $x\in Range(R)$ such that* $\Pr(R<x)\leq 1/2$ *and* $\Pr(R>x)<1/2$. *(sometimes the $\leq$ and $<$ are switched)*

Example 1 Roll 1 six-sided die. the mean is

$$E[R]=\sum_{i=1}^{6} i\left(\frac{1}{6}\right)=\frac{1}{6}\frac{6\cdot 7}{2}=3.5$$

The median is defined to be 4 (sometimes 3).

Example 2 If I is an indicator variable for an event A, then

$$\begin{aligned} E[I] &= 1\cdot\Pr[A]+0\cdot\Pr[\overline{A}] \\ &= \Pr[A] \end{aligned}$$

The median and mean can be very far apart and have different advantages.

Warning: expectation does not tell us a lot about the outcome of a single trial: a variable can diverge significantly from its expectation. But it tells us a lot about the outcome of repeated trials:

Theorem 2.3 (Law of Large Numbers) *If X is a random variable denoting the outcome of an experiment, then for a sufficiently large number of independent experiments, the probability that the* average *outcome diverges significantly from $E[X]$ is tiny.*

In one die roll, there is no particular reason to expect an outcome near 3.5. But over many die rolls, the outcomes will almost certainly average to around 3.5.

Completely different distributions can have the same mean.

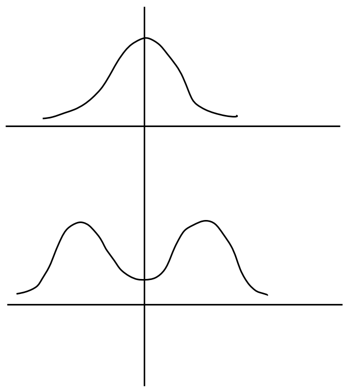


Figure 1: Two distributions with the same mean

2.1 Computing the Mean—a Shortcut

If the range of a random variable is the natural numbers, we can compute the expectation in an interesting manner.

Theorem 2.4 *If $R : S \to \mathbb{N}$, then*

$$E[R] = \sum_{i=0}^{\infty} \Pr(R > i).$$

Proof.

$$
\begin{array}{rcllllll}
\sum_{i=0}^{\infty} \Pr(R > i) & = & \Pr(R > 0)+ & \Pr(R > 1)+ & \Pr(R > 2)+ & \Pr(R > 3)+ & \Pr(R > 4) + \dots \\
 & = & & \Pr(R = 1)+ & \Pr(R = 2)+ & \Pr(R = 3)+ & \Pr(R = 4) + \dots \\
 & & & & \Pr(R = 2)+ & \Pr(R = 3)+ & \Pr(R = 4) + \dots \\
 & & & & & \Pr(R = 3)+ & \Pr(R = 4) + \dots \\
 & & & & & & \Pr(R = 4) + \dots \\
 & = & \sum_{i=1}^{\infty} i \Pr(R = i) & & & & \\
 & = & E[R] & & & &
\end{array}
$$

■

2.2 Application: the Negative Binomial Distribution

BUG Watch: this section was done wrong and corrected later in class. Hopefully the notes are right.

Example: Suppose a system fails independently with probability p at each step (the first event occurs at time 0). Let R denote the time to the first failure. R is distributed according to the *negative binomial distribution*:

$$\Pr[R = k] = p(1-p)^k,$$

since this is the probability that the first k trials are successes and the $(k+1)^{st}$ is a failure.

What is our sample space? Infinite sequences of successes and failures. It is therefore uncountable. But this won't bother us, since we'll never actually sum over the whole sample space.

What is the mean time to failure (MTF)? The event $R > i$ occurs when the first $i+1$ events (trials 0 through i) are all successes. In other words:

$$\Pr(R > i) = (1-p)^{i+1}$$

Therefore, using the shortcut sum,

$$E[R] = (1-p) + (1-p)^2 + (1-p)^3 + \cdots = \frac{1-p}{1-(1-p)} = \frac{1-p}{p}$$

It follows that the time of the first failure has expected time $\frac{1-p}{p}$, meaning (since the first trial is at time 0) that the failure occurs at trial number $1 + \frac{1-p}{p} = \frac{1}{p}$.

Example 3 Consider a couple that wants to have a boy. Assume girls and boys are equally likely. What is the expected number of girls before the first boy?

Here, $p = 1/2$. The boy is analagous to a system failure (not to be taken literally). From the above analysis, the expected birth-number of the boy is

$$E[R] = \frac{1-p}{p} = \frac{1-1/2}{1/2} = 1$$

Thus, you expect only one girl before the first boy. So the average family size is 2! This might seem counterintuitive, since you would expect the attempts to have a boy to lead to sequences of girls; the thing to remember is that these sequences are counterbalanced by the chance of having a boy on the first try.

3 Linearity of Expectation

Theorem 3.1 *Expectation is a* linear function: *for any random variables X and Y, and any scalars a and b, we have*

$$E[aX + bY] = a \cdot E[X] + b \cdot E[Y].$$

Note, we did not need to assume independence of the random variables. LINEARITY OF EXPECTATION does NOT require INDEPENDENCE.

Example 4 Roll two sided dice. Let R_1 and R_2 be the outcome of the 1st and 2nd die. Also, let $R = R_1 + R_2$ be the sum. Then

$$E[R] = E[R_1] + E[R_2] = 3.5 + 3.5 = 7$$

We prove the theorem in two pieces:

Lemma 3.2 *For any random variable R and scalar a*

$$E[aR] = a \cdot E[R].$$

Proof.

$$\begin{aligned}
E[aR] &= \sum_x x \Pr(aR = x) \\
&= \sum_x x \Pr(R = x/a) \\
&= \sum_{x'} ax' \Pr(R = x') \qquad \text{(substitute } x' = x/a\text{)} \\
&= a \sum_{x'} x' \Pr(R = x') \\
&= aE[R]
\end{aligned}$$

■

Lemma 3.3 *For any random variables R_1 and R_2*

$$E[R_1 + R_2] = E[R_1] + E[R_2]$$

We extend this to many random variables by induction.

Corollary 3.4 $E[R_1 + R_2 + R_3 + \cdots + R_N] = E[R_1] + E[R_2] + E[R_3] + \ldots E[R_N]$

To prove the lemma, recall the theorem of total probability.

Theorem 3.5 *If $B_1, B_2, B_3, \ldots$ are disjoint events such that $\bigcup_{i=1}^{\infty} B_i = S$ (S = sample space), then for all $A \subseteq S$*

$$\Pr(A) = \sum_{i=1}^{\infty} \Pr(A \wedge B_i).$$

We can now prove linearity of expectation:

Proof.

$$\begin{aligned}
E[R_1 + R_2] &= \sum_x x \Pr(R_1 + R_2 = x) \\
&= \sum_{x_1, x_2} (x_1 + x_2) \Pr(R_1 = x_1 \wedge R_2 = x_2) \\
&= \sum_{x_1} \sum_{x_2} x_1 \Pr(R_1 = x_1 \wedge R_2 = x_2) + \sum_{x_2} \sum_{x_1} x_2 \Pr(R_1 = x_1 \wedge R_2 = x_2) \\
&= \sum_{x_1} x_1 \sum_{x_2} \Pr(R_1 = x_1 \wedge R_2 = x_2) + \sum_{x_2} x_2 \sum_{x_1} \Pr(R_1 = x_1 \wedge R_2 = x_2) \\
&= \sum_{x_1} x_1 \Pr(R_1 = x_1) + \sum_{x_2} x_2 \Pr(R_2 = x_2) \\
&= E[R_1] + E[R_2]
\end{aligned}$$

■

3.1 The Hat Check Problem

N men check their hat in at dinner. After dinner, each man gets a random hat back. What is the expected number of men to get their own hat?

To calculate directly, the problem is difficult. Let R denote the number of men to get their hat back. $\Pr(R = k)$ is difficult to quantify. However, if we define R_i as

$$R_i = \begin{cases} 1 & \text{if the ith man gets his hat} \\ 0 & \text{otherwise} \end{cases},$$

then $R = R_1 + R_2 + R_3 + \cdots + R_N$. Clearly, the R_i are not independent. For example, if $R_1 = R_2 = \ldots R_{N-1} = 1$, then $R_N = 1$. Nevertheless, we can apply linearity of expectation.

Note, $E[R_i] = 1 \cdot \Pr(R_i = 1) + 0 \cdot \Pr(R_i = 0) = 1/N$.

$$E[R] = E[R_1] + E[R_2] + \cdots + E[R_N] = N(1/N) = 1$$

So, you expect one man to get his hat back.

Example 5 The Chinese appetizer problem: N people at a circular table. N different appetizers on a Lazy Susan. The Lazy Susan is randomly rotated. What is the expected number of people who get the same appetizer?

Answer: 1. Same reasoning as above. The expectation for each individual person is 1/N. Therefore, the expectation for N people is 1.

Note that the probability space is very different from the hat check problem. There are N distinct outcomes, instead of $N!$ in the hat check problem. Either everyone gets the same appetizer back (probability $1/N$) or no one does. Nevertheless, we can solve both problems rather easily by using linearity of expectation. We didn't need to know much about the actual distribution. Also, we don't need independence.

3.2 Combinations of Events

Theorem 3.6 *Given any collection of N events: $A_1, A_2, A_3, \ldots A_N$, the expected number of events that occur is*

$$\sum_i Pr(A_i).$$

Proof. Let R_i be the indicator variable for A_i. That is,

$$R_i = \begin{cases} 1 & \text{if event } A_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Then, the expected number of events is

$$R = R_1 + R_2 + \cdots + R_N.$$

$$\begin{aligned} E[R] &= \sum_i E[R_i] \\ &= \sum_i \Pr(R_i) \\ &= \sum_i \Pr(\text{event } A_i \text{ occurs}) \\ &= \sum_i \Pr(A_i) \end{aligned}$$

■

Example 6 Suppose I flip N coins. The expected number of heads, $E[R]$, is

$$E[R] = E[R_1] + \cdots + E[R_N] = \frac{1}{2} + \ldots \frac{1}{2} = N/2$$

To calculate directly, we could have done

$$E[R] = \sum_{i=0}^{N} i \Pr(R = i) = \sum_{i=0}^{N} i \binom{N}{i} 2^{-N}$$

Using the fact that,

$$\sum_{i=0}^{N} i \binom{N}{i} = N2^{N-1}$$

we would arrive at the same result: $E[R] = N/2$ (alternatively, we can see this as a probabilistic proof of the combinatorial identity: we know two sums compute the same expectation, so they must be equal).

Besides being easier, the expectation argument is also more powerful. It doesn't assume independent distributions on the coins. Suppose I make one random choice, to either use an all heads coin or an all tails coin in my coin flips. Now there are two sample points, one with all heads in the sequence and one with all tails. But the expected value of any one flip is still 1/2. so the expectation argument applies.

Note that with independent coins, the expected value is not a particularly likely value: although we expect the number of heads to $N/2$, the probability of exactly $N/2$ heads is fairly small ($\Theta(\frac{1}{\sqrt{N}})$).

One more little fact about expectation:

Corollary 3.7 *For any random variable, R, and scalar a,*

$$E[R + a] = E[R] + a$$

Proof. The scalar a can be interpreted as a random variable that always evaluates to a. Thus, the theorem follows by linearity of expectation. ■

3.3 Dangers in Expectations

Let's consider the following carnival game.

- Player picks $A \in [1,6]$.
- Roll 3 six-sided dice. (Assume fair and mutually independent).
- If the player matches any of the dice, he wins \$1. Otherwise, he loses \$1.

Let R denote the player's winning. Then, $E[R] = 1 \cdot \Pr(win) - 1 \cdot \Pr(lose)$.

One initial approach may be as follows. The probability of die 1 matching is 1/6. Similary, the probability of die 2 and die 3 matching is 1/6. Therefore, the probability of getting one match is 3/6. Thus, the game is fair, and $E[R] = 0$.

This analysis is wrong. $Pr(win) \neq 1/2$. What was the problem? Well, we imagined an indicator variable for each die: I_k says die k is a winner. But then we claimed that $I = I_1 + I_2 + I_3$ is an indicator variable for whether the player wins. This is wrong, since I isn't an indicator for anything: its value may be 2 or 3. $\sum I_k$ actually measures the *expected number of matching dice*, a different quantity. Indicator variables need to be $0, 1$.

The correct method is to apply inclusion/exclusion.

$$\begin{aligned}
Pr(R = 1) &= \Pr(D_1 = A \vee D_2 = A \vee D_3 = A) \\
&= Pr(D_1 = A) + \Pr(D_2 = A) + \Pr(D_3 = A) - \Pr(D_1 = A \wedge D_2 = A) \\
&\quad - Pr(D_1 = A \wedge D_3 = A) - \Pr(D_2 = A \wedge D_3 = A) + \Pr(D_1 = A \wedge D_2 = A \wedge D_3 = A) \\
&= 1/6 + 1/6 + 1/6 - 1/36 - 1/36 - 1/36 + 1/216 \\
&= .421
\end{aligned}$$

Another way to calculate this probability is to find the probability of a loss. $Pr(lose) = (5/6)^3 = .579 \implies \Pr(win) = 1 - .579 = .421$.

Thus, the game is an expected loss. Another variant of the game is to pay \$2 if two dice match. Again, this is a losing game.