3 Multiple Discrete Random Variables

3.1 Joint densities

Suppose we have a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and now we have two discrete random variables X and Y on it. They have probability mass functions $f_X(x)$ and $f_Y(y)$. However, knowing these two functions is not enough. We illustrate this with an example.

Example: We flip a fair coin twice and define several RV's: Let X_1 be the number of heads on the first flip. (So $X_1 = 0, 1$.) Let X_2 be the number of heads on the second flip. Let $Y = 1 - X_1$. All three RV's have the same p.m.f. : $f(0) = 1/2$, $f(1) = 1/2$. So if we only look at p.m.f.'s of individual RV's, the pair X_1, X_2 looks just like the pair X_1, Y. But these pairs are very different. For example, with the pair X_1, Y, if we know the value of X_1 then the value of Y is determined. But for the pair X_1, X_2, knowning the value of X_1 tells us nothing about X_2. Later we will define "independence" for RV's and see that X_1, X_2 are an example of a pair of independent RV's.

We need to encode information from $\mathbf{P}$ about what our random variables do together, so we introduce the following object.

Definition 1. *If X and Y are discrete RV's, their joint probability mass function is*

$$f(x, y) = \mathbf{P}(X = x, Y = y)$$

As always, the comma in the event $X = x, Y = y$ means "and". We will also write this as $f_{X,Y}(x, y)$ when we need to indicate the RV's we are talking about. This generalizes to more than two RV's:

Definition 2. *If $X_1, X_2, \cdots, X_n$ are discrete RV's, then their joint probability mass function is*

$$f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = \mathbf{P}(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$$

The joint density for n RV's is a function on $\mathbb{R}^n$. Obviously, it is a non-negative function. It is non-zero only on a finite or countable set of points in $\mathbb{R}^n$. If we sum it over these points we get 1:

$$\sum_{x_1,\cdots,x_n} f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = 1$$

We return to the example we started with.

Example - cont. The joint pmf of X_1, X_2 is

$$f_{X_1,X_2}(0,0) = f_{X_1,X_2}(0,1) = f_{X_1,X_2}(1,0) = f_{X_1,X_2}(1,1) = 1/4$$

and that of X_1, Y is

$$f_{X_1,Y}(0,1) = f_{X_1,Y}(1,0) = 1/2, \quad f_{X_1,Y}(0,0) = f_{X_1,Y}(1,1) = 0$$

As with one RV, the goal of introducing the joint pmf is to extract all the information in the probability measure $\mathbf{P}$ that is relevant to the RV's we are considering. So we should be able to compute the probability of any event defined just in terms of the RV's using only their joint pmf. Consider two RV's X, Y. Here are some events defined in terms of them:

$$\{0 \leq X \leq 10, Y \geq 5\}, \{0 \leq X + Y \leq 10\}, \textit{ or } \{XY \geq 3\}$$

We can write any such event as $\{(X,Y) \in A\}$ where $A \subset \mathbb{R}^2$. For example,

$$\{0 \leq X \leq 10, Y \geq 5\} = \{(X,Y) \in A\},$$

where A is the rectangle $A = \{(x,y) : 0 \leq x \leq 10, y \leq 5\}$. And

$$\{0 \leq X + Y \leq 10\} = \{(X,Y) \in A\}, \quad A = \{(x,y) : 0 \leq x + y \leq 10\}$$

The following proposition tells us how to compute probabilities of events like these.

Proposition 1. *If $X_1, X_2, \cdots, X_n$ are discrete RV's and $A \subset \mathbb{R}^n$, then*

$$\mathbf{P}((X_1, X_2, \cdots, X_n) \in A) = \sum_{(x_1,x_2,\cdots,x_n) \in A} f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n)$$

There are some important special cases of the proposition which we state as corollaries.

Corollary 1. *Let $g(x,y) : \mathbb{R}^2 \to \mathbb{R}$. Let X, Y be discrete RV's. Define a new discrete RV by $Z = g(X,Y)$. Then its pmf is*

$$f_Z(z) = \sum_{(x,y): g(x,y) = z} f_{X,Y}(x,y)$$

When we have the joint pmf of X, Y, we can use it to find the pmf's of X and Y by themselves, i.e., $f_X(x)$ and $f_Y(y)$. These are called "marginal pmf's." The formula for computing them is :

Corollary 2. *Let X, Y be two discrete RV's. Then*

$$
\begin{aligned}
f_X(x) &= \sum_y f_{X,Y}(x,y) \\
f_Y(y) &= \sum_x f_{X,Y}(x,y)
\end{aligned}
$$

Example: Flip a fair coin. Let $X = 0, 1$ be number of heads. If coin is heads roll a four-sided die, if tails a six-sided die. Let Y be the number on the die. Find the joint pmf of X, Y, and the individual pmf's of X and Y.

Example: Roll a die until we get a 6. Let Y be the number of rolls (including the 6). Let X be the number of 1's we got before we got 6. Find $f_{X,Y}, f_X, f_Y$. It is hard to figure out $\mathbf{P}(X = x, Y = y)$ directly. But if we are given the value of Y, say $Y = y$, then X is just a binomial RV with $y - 1$ trials and $p = 1/5$. So we have

$$\mathbf{P}(X = x|Y = y) = \binom{y-1}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{y-1-x}$$

We then use $\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x|Y = y)\mathbf{P}(Y = y)$. The RV Y has a geometric distribution with parameter $1/6$. Note that $\mathbf{P}(X = x|Y = y) = 0$ if $x \geq y$. So we get

$$f_{X,Y}(x,y) = \begin{cases} \binom{y-1}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{y-1-x} \left(\frac{5}{6}\right)^{y-1} \frac{1}{6}, & \text{if } x < y \\ 0, & \text{if } x \geq y \end{cases}$$

3.2 Functions of two RV's

Suppose X and Y are discrete random variables and $g(x, y)$ is a function from $\mathbb{R}^2$ to $\mathbb{R}$. Then $Z = g(X, Y)$ defines a new random variable. We saw in the last section how to compute its pmf from the joint pmf of X and Y:

$$\mathbf{P}(Z = z) = \mathbf{P}(g(X, Y) = z) = \sum_{(x,y):g(x,y)=z} f_{X,Y}(x,y)$$

There is a shortcut for computing its expected value:

Theorem 1. *Let X, Y be discrete RV's, and $g : \mathbb{R}^2 \to \mathbb{R}$. Define $Z = g(X, Y)$. Then*

$$\mathbf{E}[Z] = \sum_{x,y} g(x,y)\, f_{X,Y}(x,y)$$

Proof. **GAP !!!!!!!!!!!!!!!!!!!11** $\square$

Corollary 3. *Let X, Y be discrete RV's and $a, b \in \mathbb{R}$. Then*

$$\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$$

End of September 20 lecture

The theorem and corollary generalize to more than two RV's. The corollary can be used to make some of the calculations we have done for finding the means of certain distributions much easier.

Example: Consider the binomial random variable. So we have an unfair coin with probability p of heads. We flip it n times and let X be the number of heads we get. Now let X_j be the random variable that is 1 if the jth flip is heads, 0 if it is tails. So X_j is the number of heads on the jth flip. Then $X = \sum_{j=1}^n X_j$. So

$$\mathbf{E}[X] = \mathbf{E}[\sum_{j=1}^n X_j] = \sum_{j=1}^n \mathbf{E}[X_j]$$

Note that each X_j is just a Bernoulli trial. $X_j = 1$ with probability p and $X_j = 0$ with probability $1 - p$. So $E[X_j] = p$. Thus we recover our previous result that $E[X] = np$.

Example: We now look at the negative binomial distribution. Again we have an unfair coin with probability p of heads. We flip it until we get heads a total of n times. Then we take X to be the total number of flips including the n heads. So X is at least n. Now define X_1 to be the number of flips it takes to get the first heads, including the flip that gave heads. Let X_2 be the number of flips it takes **after** the first heads to get the second heads. In general, let X_j be the number of flips it takes after the jth heads to get the $(j+1)$th heads. Then $X = \sum_{j=1}^{n} X_j$. So

$$\mathbf{E}[X] = \mathbf{E}[\sum_{j=1}^{n} X_j] = \sum_{j=1}^{n} \mathbf{E}[X_j]$$

What is $\mathbf{E}[X_j]$. We claim that the distribution of X_j is just a geometric distribution with parameter p. So $\mathbf{E}[X_j] = 1/p$. Thus $\mathbf{E}[x] = n/p$. To appreciate how much easier this approach is, the reader is invited to compute the mean of X using the pmf of X.

3.3 Independence of discrete RV's

Recall the definition of independence for events. Two events A and B are independent if $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. We will now define independence of RV's. As we will see, if two RV's X and Y are independent then if we define some event using X and another event using Y, then these two events will be independent. However, the definition is a bit simpler.

Definition 3. *Two discrete RV's X and Y are independent if*

$$\begin{aligned} \mathbf{P}(X = x, Y = y) &= \mathbf{P}(X = x)\mathbf{P}(Y = y), \quad \forall x, y \in \mathbb{R}, i.e., \\ f_{X,Y}(x, y) &= f_X(x)\, f_Y(y), \quad \forall x, y \in \mathbb{R} \end{aligned}$$

Discrete RV's $X_1, X_2, \cdots, X_n$ are independent if

$$f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i), \quad \forall x_1, \cdots, x_n \in \mathbb{R}$$

Remark: In general, knowing the individual pmf's of X and Y, i.e., $f_X(x)$ and $f_Y(y)$, is not enough to determine the joint pmf of X and Y. But if we also know that the two RV's are independent, then $f_X(x)$ and $f_Y(y)$ completely determine the joint pmf. We will often encounter situtations like this: you know X and Y are independent and you know that X has say a Poisson distribution and Y has a binomial distribution. This determines the joint pmf for X and Y.

One of the most important properties of independent RV's is the following.

Theorem 2. *Let X and Y be discrete RV's. Assume that their expected values are defined and the expected value of XY is too. Then*

$$\mathbf{E}[XY] = \mathbf{E}[X]\,\mathbf{E}[Y]$$

Proof.

$$\begin{aligned}
\mathbf{E}[XY] &= \sum_{x,y} xy\, f_{X,Y}(x,y) = \sum_{x,y} xy\, f_X(x)\, f_Y(y) \\
&= \left[\sum_x x\, f_X(x)\right]\left[\sum_y y\, f_Y(y)\right] = \mathbf{E}[X]\,\mathbf{E}[Y]
\end{aligned}$$

$\square$

The next proposition says that if X and Y are independent and we use X to define some event and Y to define some other event, then these two events are independent.

Proposition 2. *Let X and Y be independent discrete RV's. Let $A, B \subset \mathbb{R}$. Then the events $X \in A$ and $Y \in B$ are independent events.*

Proof. **GAP !!!!!!!!!!!!!!!!!!!11** $\square$

Theorem 3. *Let X and Y be independent discrete RV's. Let g, h be functions from $\mathbb{R}$ to $\mathbb{R}$. Then $g(X)$ and $h(Y)$ are independent RV's. Consequently,*

$$\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)]\mathbf{E}[h(Y)]$$

Proof. Did not prove this. $\square$

Suppose we know the joint pmf $f_{X,Y}(x,y)$ but we don't know the individual pmf's $f_X(x)$ and $f_Y(y)$. Can we check if X and Y are independent? We could compute $f_X(x)$ and $f_Y(y)$ from the joint pmf and then check if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. But there is a shortcut:

Proposition 3. *If $f_{X,Y}(x,y)$ can be factored as $g(x)h(y)$ for some functions g and h, then X and Y are independent.*

Caution: Watch out for joint pmf's that look like they factor, but actually don't because of cases involved in the def of the joint pmf. The following example illustrates this.

Example Suppose the joint p.m.f. of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 2p^2\,(1-p)^{x+y-2} & \text{if } x < y \\ p^2\,(1-p)^{2x-2} & \text{if } x = y \\ 0 & \text{if } x > y \end{cases}$$

Are X and Y independent? Note : the marginals of X and Y are geometric if you work them out.

Example: Let X and Y be independent RV's. X has a binomial distribution with n trials and probability p of success. Y also has a binomial distribution with the same probability p of success, but a different number of trials, say m. Argue that $Z = X + Y$ is binomial with probability of success p and $n + m$ trials. Give heuristic explanation and then show what the calculation would be.

Caution: In general it is not true that if we add two independent RV's with the same type of distribution we get back the same distribution.

Example: Let X and Y be independent RV's. They are both geometric with $p = 1/3$. Let $Z = X + Y$. Then Z starts at 2, so it can't possible have a geometric distribuition. Suppose we let $Z = X + Y - 1$. Is Z geometric?

Note that the question is not to find the distribution of Z, but simply to determine if it is geometric. Suppose it is geometric with parameter r. So $\mathbf{P}(Z = z) = (1-r)^{z-1}r$. We will compute $\mathbf{P}(Z = z)$ for a few z and show that we get a contradiction. The pmf of X and Y are

$$f_X(x) = \mathbf{P}(X = x) = \frac{1}{3}\left(\frac{2}{3}\right)^{x-1}, \quad f_Y(y) = \frac{1}{3}\left(\frac{2}{3}\right)^{y-1}$$

Since they are independent the joint pmf is

$$f_{X,Y} = \mathbf{P}(X = x, Y = y) = \frac{1}{27}\left(\frac{2}{3}\right)^{x+y-2}$$

So we have

$$\mathbf{P}(Z = 1) = \mathbf{P}(X + Y = 2) = \mathbf{P}(X = 1, Y = 1) = \frac{1}{27}$$

So we must have $r = \frac{1}{27}$. This would mean that $\mathbf{P}(Z = 2) = \frac{26}{27}\frac{1}{27}$. But

$$\begin{aligned} \mathbf{P}(Z = 2) &= \mathbf{P}(X + Y = 3) = \mathbf{P}(X = 1, Y = 2) + \mathbf{P}(X = 2, Y = 1) \\ &= 2\frac{1}{27}\frac{2}{3} \end{aligned}$$

which is not $26/27^2$. So Z cannot be geometric.

Example: Let X and Y be independent RV's with the geometric distribution with parameters p and q respectively. Let $Z = \min\{X, Y\}$. Find the pmf of Z.

If we try to compute it directly we get stuck:

$$\mathbf{P}(Z = z) = \mathbf{P}(\min\{X, Y\} = z) = \cdots$$

We use a trick. Look at $\mathbf{P}(Z \geq z)$. We have

$$\begin{aligned} \mathbf{P}(Z \geq z) &= \mathbf{P}(\min\{X, Y\} \geq z) = \mathbf{P}(X \geq z, Y \geq z) \\ &= (1-p)^{z-1}(1-q)^{z-1} = [(1-p)(1-q)]^{z-1} \end{aligned}$$

So Z has a geometric distribution. Call the parameter for it r. Then $1 - r = (1-p)(1-q)$, i.e., $r = 1 - (1-p)(1-q)$.

We can understand the above result in a less computational way. We have two biased coins, one for X and one for Y. The coin for X has probability p of heads, the coin for Y probability q. We flip the two coins simultaneously. X is the first flip that the X coin gives heads, Y the first flip that the Y coin gives heads. Since Z is the minimum of X and Y, Z will be the first time that one of the two coins gives heads. We can think of flipping the two coins together as our new trial. Z comes from repeating this new trial until we get success. In the new trial failure means failure for both the X and Y coins and so has probability $(1-p)(1-q)$. So the probability of success in the new trial is $1 - (1-p)(1-q)$.

Recall that for any two random variables $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$, i.e., the mean of their sum is the sum of their means. In general the variance of their sum is not the sum of their variances. But if they are independent it is!

Theorem 4. *If X and Y are independent discrete RV's then $var(X + Y) = var(X) + var(Y)$, provided the variances of X and Y are defined.*

Proof. The proof is a computation that uses two facts. For any RV, $var(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2$. And since X and Y are independent, $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$. So

$$\begin{aligned} var(X+Y) &= \mathbf{E}[(X+Y)^2] - (\mathbf{E}[X+Y])^2 \\ &= \mathbf{E}[X^2 + 2XY + Y^2] - (\mathbf{E}[X] + \mathbf{E}[Y])^2 \end{aligned}$$

$$
\begin{aligned}
&= \mathbf{E}[X^2] + 2\mathbf{E}[XY] + \mathbf{E}[Y^2] - \big(\mathbf{E}[X]^2 + 2\mathbf{E}[X]\mathbf{E}[Y] + \mathbf{E}[Y]^2\big) \\
&= \mathbf{E}[X^2] + 2\mathbf{E}[X]\mathbf{E}[Y] + \mathbf{E}[Y^2] - \big(\mathbf{E}[X]^2 + 2\mathbf{E}[X]\mathbf{E}[Y] + \mathbf{E}[Y]^2\big) \\
&= \mathbf{E}[X^2] - \mathbf{E}[X]^2 + \mathbf{E}[Y^2] - \mathbf{E}[Y]^2 = var(X) + var(Y)
\end{aligned}
$$

$\square$

Remark: The variance of $X - Y$ is **not** the variance of X minus the variance of Y. Think of $X - Y$ as $X + (-Y)$. Then

$$
\begin{aligned}
var(X - Y) &= var(X + (-Y)) = var(X) + var(-Y) \\
&= var(X) + (-1)^2\, var(Y) = var(X) + var(Y)
\end{aligned}
$$

Example: X and Y are independent random variables. They are both Poission. X has mean 1 and Y has mean 3. What is the mean and variance of $2X - 3Y$?

End of September 22 lecture

We are often interested in the sum of independent random variables and in particular in finding the pmf of the sum. The following proposition is about this, but we will see a much more powerful computational method (generating functions) in the next chapter. Note that the formula in the proposition looks like a convolution.

Proposition 4. *Let X and Y be independent RV's. Let $Z = X + Y$. Then*

$$
f_Z(z) = \sum_x f_X(x) f_Y(z - x) = \sum_y f_X(z - y) f_Y(y)
$$

Example: Suppose that X and Y are independent random variables. They both have the geometric distribution with the same parameter p. Let $Z = X + Y$. Find the p.m.f. of Z. After some computation we find

$$
P(Z = z) = (z - 1)p^2(1 - p)^{z-2}
$$

Why do we care about independence? The following paradigm occurs often, expecially in statistics.

Sampling paradigm: We have an experiment with a random variable X. We do the experiment n times. Let $X_1, X_2, \cdots, X_n$ be the resulting values of X. We assume the repetitions of the experiment do not change the experiment and they are independent. So the distribution of each X_j is the same as that of X and $X_1, X_2, \cdots, X_n$ are independent. So the joint pmf is

$$f_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n) = \prod_{j=1}^{n} f_X(x_j)$$

The RV's $X_1, X_2, \cdots, X_n$ are called i.i.d. (independent, identically distributed). We are often interested in the sample mean

$$\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$$

It is a RV. We can find its mean and variance.

The mean of $\bar{X}$ is

$$\mathbf{E}[\bar{X}] = \mathbf{E}[\frac{1}{n} \sum_{j=1}^{n} X_j] = \frac{1}{n} \sum_{j=1}^{n} \mathbf{E}[X_j] = \frac{1}{n} \sum_{j=1}^{n} \mathbf{E}[X] = \mathbf{E}[X]$$

Since the X_j are independent, the variance of $X_1 + X_2 + \cdots X_n$ is the sum of their variances. Since they are identically distributed they have the same variance. In fact, the variance of X_j is the variance of X. So the variance of $X_1 + X_2 + \cdots X_n$ is $n \ var(X)$. Thus

$$var(\bar{X}) = \frac{1}{n^2} n \ var(X) = \frac{1}{n} var(X)$$

So if n is large, the variance of the sample average is much smaller than that of the random variable X.