ECE 361 Probability for Engineers (Fall, 2016)
Lecture 4a

§2.5 Joint PMFs of multiple RVs

Functions of multiple RVs

Just as we can discuss functions of a single RV $Y = g(X)$, we can discuss functions of multiple RVs $Z = g(X, Y)$. Such functions are of course themselves RVs and have a PMF given by

$$\boxed{p_Z(z) = \sum_{(x,y):g(x,y)=z} p_{X,Y}(x, y), \ \forall z \in \mathcal{Z}} \tag{1}$$

which is to say, we compute the probability of the pre-image of the function g from z. Just as we compute the expectation $\mathbb{E}[Y]$ for $Y = g(X)$ as $\mathbb{E}[Y] = \sum_{x\in\mathcal{X}} g(x)p_X(x)$, we similarly compute for $Z = g(X, Y)$

$$\mathbb{E}[Z] = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} g(x, y)p_{X,Y}(x, y). \tag{2}$$

Recall that for the special case of a linear function $Y = aX + b$ we had that $\mathbb{E}[Y] = a\mathbb{E}[X] + b$, i.e., the expectation of a linear function is the linear function of the expectation. The similar result holds for a pair of RVs:

$$\boxed{Z = aX + bY + c \ \Rightarrow \ \mathbb{E}[Z] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c.} \tag{3}$$

Example 2.9 is a helpful illustration of the two procedures to find $\mathbb{E}[Z]$ for $Z = X + 2Y$ for the joint PMF $\mathbf{p}_{X,Y}$ in Fig. 2.10.

More than two RVs

For three RVs (X, Y, Z) we have a joint PMF $\mathbf{p}_{X,Y,Z} = (p_{X,Y,Z}(x, y, z), \ x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z})$. From here we can find various marginals, including:

$$p_{X,Y}(x, y) = \sum_{z\in\mathcal{Z}} p_{X,Y,Z}(x, y, z), \ p_X(x) = \sum_{y\in\mathcal{Y},z\in\mathcal{Z}} p_{X,Y,Z}(x, y, z). \tag{4}$$

The simple rule is this: given a joint PMF for a collection of RVs, say $(X_1, \ldots, X_n)$ and a subset of those RVs of interest $X_A \equiv (X_i, i \in A)$ for some $A \subset [n]$, the PMF for X_A is found by summing over RVs not in the set. For three RVs (X, Y, Z) and a function $W = g(X, Y, Z)$ we find its expectation as you would now expect:

$$\mathbb{E}[W] = \sum_{x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}} g(x, y, z)p_{X,Y,Z}(x, y, z). \tag{5}$$

Again, a special form holds for linear functions:

$$W = aX + bY + cZ + d \ \Rightarrow \ \mathbb{E}[W] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c\mathbb{E}[Z] + d. \tag{6}$$

Clearly this linearity of expectation holds more generally: given RVs $X_1, \ldots, X_n$ and scalars $(a_1, \ldots, a_n)$:

$$\boxed{\mathbb{E}\left[\sum_{i=1}^n a_i X_i\right] = \mathbb{E}[a_1X_1 + \cdots + a_nX_n] = a_1\mathbb{E}[X_1] + \cdots + a_n\mathbb{E}[X_n] = \sum_{i=1}^n a_i\mathbb{E}[X_i].} \tag{7}$$

Example. Mean of the binomial. In a class of 300 students where each student earns an A with probability 0.3, what is the expected number of As? Define the Bernoulli RV X_i to take value 1 when student i earns an A and 0 else, thus $\mathbb{P}(X_i = 1) = 0.3 = 1 - \mathbb{P}(X_i = 0)$ for $i = 1, \ldots, 300$. Then the random number of As is $X = X_1 + \cdots + X_{300}$. By linearity of expectation:

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \cdots + X_{300}] = 300\mathbb{E}[X_1] = 300 \times 0.3 = 100. \tag{8}$$

More generally, if we have a sequence of n trials each of which is successful with probability p, then we write $X_i \sim \mathrm{Ber}(p)$ and $X = \sum_{i=1}^n X_i$, and find $\mathbb{E}[X] = n\mathbb{E}[X_1] = np$. We then recognize $X \sim \mathrm{Bin}(n,p)$, and thus we've shown that

$$\boxed{X \sim \mathrm{Bin}(n,p) \;\Rightarrow\; \mathbb{E}[X] = np.} \tag{9}$$

Example. The hat problem. Suppose n people throw their hat in a pile and then each person picks a hat at random. On average how many people pick their own hat? Define $X_i \sim \mathrm{Ber}(1/n)$ where $X_i = 1$ when person i picks their own hat so that $X = X_1 + \cdots + X_n \sim \mathrm{Bin}(n, 1/n)$. Note X is the desired random quantity of the number of individuals that pick their own hat. By the previous result: $\mathbb{E}[X] = n\mathbb{E}[X_1] = n \times 1/n = 1$.

§2.6 Conditioning

Conditioning a RV on an event

Recall the definition of conditional probability: $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ for any event B with $\mathbb{P}(B) > 0$. If we view the elements of a PMF as probabilities of the corresponding event, i.e., $p_X(x) = \mathbb{P}(X = x)$ then the formula for conditioning an RV on an event looks quite natural:

$$\boxed{p_{X|A}(x) = \mathbb{P}(X = x|A) = \frac{\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)}.} \tag{10}$$

Example. Let X be the roll of a fair six-sided die and let A be the event that the roll is even. Find the PMF for X given A:

$$p_{X|A}(x) = \mathbb{P}(X = x|A) = \mathbb{P}(X = x \cap X \text{ even })/\mathbb{P}(X \text{ even}) = \begin{cases} 1/3, & x \in \{2,4,6\} \\ 0, & \text{else} \end{cases} \tag{11}$$

Example. A student passes a test independently at each attempt with probability p, and will retake the exam until she passes, with a maximum of n attempts. Let X be the RV for the number of attempts. View $X \sim \mathrm{Geo}(p)$ and $A = \{X \leq n\}$, so that $X|A$ is the random number of attempts conditioned on the number of attempts being at most n. Then:

$$\mathbb{P}(A) = 1 - \mathbb{P}(\bar{A}) = 1 - \mathbb{P}(X > n) = 1 - (1-p)^n, \tag{12}$$

and

$$p_{X|A}(k) = \begin{cases} \frac{(1-p)^{k-1}p}{1-(1-p)^n}, & k \in [n] \\ 0, & \text{else} \end{cases} \tag{13}$$

See Fig. 2.11 and Fig. 2.12 for a visualization of the effect of conditioning on A on the PMF for X.

Conditioning one RV on another

Given knowledge that $Y = y$ affects the probability that $X = x$, just as knowledge of the event B affects the probability of event A. Again, interpreting $\{Y = y\}$ and $\{X = x\}$ as events these conclusions follow directly from the definitions in Chapter 1. The notation is:

$$p_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y). \tag{14}$$

This quantity is read as the probability the RV $X = x$ given the RV $Y = y$. It is evaluated as:

$$\boxed{p_{X|Y}(x|y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p_{X,Y}(x,y)}{p_Y(y)}.} \tag{15}$$

Note that $\mathbf{p}_{X|Y}(\mathbf{x}|y) = (p_{X|Y}(x|y),\ x \in \mathcal{X})$ is a valid PMF for X for each $y \in \mathcal{Y}$ because

$$\sum_{x \in \mathcal{X}} p_{X|Y}(x|y) = \sum_{x \in \mathcal{X}} \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_Y(y)}{p_Y(y)} = 1. \tag{16}$$

See Fig. 2.13 for an illustration.

One value of a conditional PMF is that it is often easier to compute the joint PMF from the conditional and the marginal:

$$\boxed{p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y) = p_{Y|X}(y|x)p_X(x).} \tag{17}$$

Example. A professor answers questions incorrectly with probability 1/4, and the number of questions she is asked is uniformly likely to be 0, 1, or 2. Let X be the number of questions she is asked and Y the number of questions she answers wrongly. Find the joint PMF for X, Y. The support of the pair is clearly $[2] \times [2]$ and we use the conditional PMF to find the joint:

$$p_{X,Y}(x,y) = p_{Y|X}(y|x)p_X(x). \tag{18}$$

For example, the probability that $(X,Y) = (1,1)$ is $1/4 \times 1/3 = 1/12$. The entire PMF is shown in Fig. 2.14. The joint PMF can then be used to answer questions like: what is the probability she has at least one wrong answer. Here $A = \{(1,1),(2,1),(2,2)\}$ and the joint PMF gives $\mathbb{P}(A) = (4+6+1)/48$.

Note the conditional PMF also gives an equation for computing the marginals:

$$\boxed{p_X(x) = \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y) = \sum_{y \in \mathcal{Y}} p_Y(y) p_{X|Y}(x|y).} \tag{19}$$

References

[1] *Introduction to Probability, 2nd Edition* by Dimitri P. Bertsekas and John N. Tsitsiklis, Athina Scientific Press, 2008.