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

---

§2.6 Conditioning

Conditioning one RV on another

Example. A transmitter sends a message over a computer network. Define X as the travel time of the message and Y as the length of the message. We are given the PMF of the travel time given the message length and the PMF for the message length, and we want to find the PMF for the travel time. That is, we are given $\mathbf{p}_{X|Y}$ and $\mathbf{p}_Y$ and we want to find $\mathbf{p}_X$. Here:

$$p_Y(y) = \begin{cases} 5/6, & y = 10^2 \\ 1/6, & y = 10^4 \end{cases} \tag{1}$$

and

$$p_{X|Y}(x|10^2) = \begin{cases} 1/2, & x = 10^{-2} \\ 1/3, & x = 10^{-1} \\ 1/6, & x = 1 \end{cases}, \quad p_{X|Y}(x|10^4) = \begin{cases} 1/2, & x = 1 \\ 1/3, & x = 10 \\ 1/6, & x = 100 \end{cases} \tag{2}$$

so we find

$$\begin{array}{c|ccccc} x & 10^{-2} & 10^{-1} & 1 & 10 & 100 \\ \hline p_X(x) & \frac{5}{6}\frac{1}{2} & \frac{5}{6}\frac{1}{3} & \frac{5}{6}\frac{1}{6} + \frac{1}{6}\frac{1}{2} & \frac{1}{6}\frac{1}{3} & \frac{1}{6}\frac{1}{6} \end{array} \tag{3}$$

Conditional expectation

Just as conditioned events A affect the probabilities of events B, i.e., $\mathbb{P}(B|A)$, they also affect the expectations of random variables, denoted $\mathbb{E}[X|A]$, the conditional expectation of the RV X given the event A. The corresponding formula are:

Summary of facts about conditional expectations. Let X, Y be RVs.

- The conditional expectation of X given event A with $\mathbb{P}(A) > 0$ is $\mathbb{E}[X|A] = \sum_{x \in \mathcal{X}} x p_{X|A}(x)$.
- For a function $g(X)$: $\mathbb{E}[g(X)|A] = \sum_{x \in \mathcal{X}} g(x) p_{X|A}(x)$.
- The conditional expectation of X given a value y for Y is: $\mathbb{E}[X|Y = y] = \sum_{x \in \mathcal{X}} x p_{X|Y}(x|y)$.
- If $A_1, \ldots, A_n$ are disjoint events that partition the sample space Ω with $\mathbb{P}(A_i) > 0$ for all i then $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X|A_i]\mathbb{P}(A_i)$.
- Furthermore, for any event B: $\mathbb{E}[X|B] = \sum_{i=1}^n \mathbb{E}[X|A_i \cap B]\mathbb{P}(A_i|B)$.
- We have: $\mathbb{E}[X] = \sum_{y \in \mathcal{Y}} p_Y(y)\mathbb{E}[X|Y = y]$.

The last three above are the **Total expectation theorem**, which is the analogue of the **Total probability theorem** in Chapter 1. Here is a proof of one of them:

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x p_X(x) = \sum_{x \in \mathcal{X}} x \left(\sum_{i=1}^n \mathbb{P}(A_i)\mathbb{P}(X = x|A_i) \right) = \sum_{i=1}^n \mathbb{P}(A_i) \sum_{x \in \mathcal{X}} x\mathbb{P}(X = x|A_i) = \sum_{i=1}^n \mathbb{P}(A_i)\mathbb{E}[X|A_i]. \tag{4}$$

Example. Mean and variance of the geometric distribution. Let $X \sim \mathrm{Geo}(p)$ so that $\mathcal{X} = \mathbb{N}$ and $p_X(k) = (1-p)^{k-1}p$ for $k \in \mathbb{N}$. The mean and variance are computable via sums:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k \in \mathbb{N}} k(1-p)^{k-1}p \\ \mathbb{E}[X^2] &= \sum_{k \in \mathbb{N}} k^2(1-p)^{k-1}p \\ \mathrm{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \end{aligned} \tag{5}$$

But this problem outlines an easier way via conditioning. Define $A_1 = \{X = 1\}$ so that $A_1^c = \{X > 1\}$. Then conditioning on A gives:

$$\mathbb{E}[X] = \mathbb{E}[X|X=1]\mathbb{P}(X=1) + \mathbb{E}[X|X>1]\mathbb{P}(X>1). \tag{6}$$

Note $\mathbb{E}[X|X=1] = 1$ and $\mathbb{P}(X=1) = p$ and $\mathbb{P}(X>1) = 1-p$ and $\mathbb{E}[X|X>1] = 1 + \mathbb{E}[X]$. Thus:

$$\mathbb{E}[X] = 1 \cdot p + (1 + \mathbb{E}[X])(1-p), \;\Rightarrow\; \mathbb{E}[X] = \frac{1}{p}. \tag{7}$$

Similarly,

$$\mathbb{E}[X^2] = \mathbb{E}[X^2|X=1]\mathbb{P}(X=1) + \mathbb{E}[X^2|X>1]\mathbb{P}(X>1), \tag{8}$$

and again $\mathbb{E}[X^2|X=1] = 1$, the new calculation is $\mathbb{E}[X^2|X>1] = \mathbb{E}[(1+X)^2] = 1 + 2\mathbb{E}[X] + \mathbb{E}[X^2]$. This gives:

$$\mathbb{E}[X^2] = 1p + (1 + 2\mathbb{E}[X] + \mathbb{E}[X^2])(1-p) \;\Rightarrow\; \mathbb{E}[X^2] = \frac{2}{p^2} - \frac{1}{p}. \tag{9}$$

Combining gives

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \left(\frac{2}{p^2} - \frac{1}{p}\right) - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}. \tag{10}$$

The statement $\mathbb{E}[X|X>1] = 1+\mathbb{E}[X]$ should be intuitive: conditioned on the first trial being a failure, the total expected number of trials until the first success equals one (the failed trial) plus $\mathbb{E}[X]$, the (unconditional) average number of trials until the first success. That is, the random experiment (here, the number of trials until the first success) probabilistically restarts itself. Said differently, the remaining random number of trials until the first success conditioned on the first success being a failure is independent of the outcome of the first trial. If this is not intuitive, here is formal development using the total expectation theorem (TET). First, by definition:

$$\mathbb{E}[X|X>1] = \sum_{x=1}^{\infty} x\mathbb{P}(X=x|X>1) \tag{11}$$

Now focus on $p_{X|X>1}(x) = \mathbb{P}(X=x|X>1)$. For $x=1$ we have $p_{X|X>1}(1) = 0$. For $x > 1$ we have

$$p_{X|X>1}(x) = \frac{\mathbb{P}(X=x, X>1)}{\mathbb{P}(X>1)} = \frac{\mathbb{P}(X=x)}{1 - \mathbb{P}(X=1)} = \frac{p_X(x)}{1-p}, \tag{12}$$

where $p_X(x) = (1-p)^{x-1}p$ is the PMF for X. Substituting this into the TET expression above:

$$\begin{aligned}
\mathbb{E}[X|X>1] &= \sum_{x=2}^{\infty} x \frac{p_X(x)}{1-p} \\
&= \frac{1}{1-p} \sum_{x=2}^{\infty} x p_X(x) \\
&= \frac{1}{1-p} \left(-1 p_X(1) + \sum_{x=1}^{\infty} x p_X(x)\right) \\
&= \frac{\mathbb{E}[X] - p}{1-p}
\end{aligned} \tag{13}$$

Note that $\mathbb{E}[X|X>1] = 1 + \mathbb{E}[X]$ and $\mathbb{E}[X|X>1] = \frac{\mathbb{E}[X]-p}{1-p}$ agree for $\mathbb{E}[X] = 1/p$. A good exercise is to see if you can mimic the argument above to formally establish $\mathbb{E}[X^2|X>1] = \mathbb{E}[(1+X)^2]$, which we used above.

§2.7 Independence

Recall that A is independent of B (and vice-versa) if knowledge of A doesn't affect the probability of B, i.e., $\mathbb{P}(A|B) = \mathbb{P}(A)$, and thus $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ (independence means multiply). The definition for random variables is analogous, as will be made clear below.

Independence of a RV from an event

The key definition here is: the RV X is independent of the event A if each event $\{X = x\}$ is independent of A for $x \in \mathcal{X}$:

$$\mathbb{P}(X = x \cap A) = p_X(x)\mathbb{P}(A), \ \forall x \in \mathcal{X}. \tag{14}$$

Equivalently,

$$p_{X|A}(x) = p_X(x), \ \forall x \in \mathcal{X}. \tag{15}$$

Example. Consider two independent tosses of a fair coin. Let X be the number of heads and A be the event that the number of heads is even. The PMF for X is $p_X(0) = 1/4, p_X(1) = 1/2, p_X(2) = 1/4$ and $\mathbb{P}(A) = 1/2$. The conditional PMF for X given A is $p_{X|A}(0) = 1/2, p_{X|A}(1) = 0, p_{X|A}(2) = 1/2$. Note X and A are not independent.

Independence of RVs

The key definition here is: RVs X, Y are independent if

$$p_{X,Y}(x, y) = p_X(x)p_Y(y), \ \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{16}$$

Equivalently, the events $\{X = x\}$ and $\{Y = y\}$ must be independent events for each $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Just as we say A, B are conditionally independent given C if $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$, we have an analogous notion for RVs. Say X, Y are conditionally independent given event A if

$$\mathbb{P}(X = x, Y = y|A) = \mathbb{P}(X = x|A)\mathbb{P}(Y = y|A), \ \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{17}$$

This is equivalent to $p_{X|Y,A}(x|y) = p_{X|A}(x)$ for all x, y such that $p_{Y|A}(y) > 0$. As with events, conditional independence and independence do not imply one another.

Example. See Fig. 2.15.

If X, Y are independent RVs then the expectation of their product is the product of their expectations:

$$\boxed{X, Y \text{ independent } \Rightarrow \ \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].} \tag{18}$$

To see this:

$$\mathbb{E}[XY] = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} xy p_{X,Y}(x, y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} xy p_X(x)p_Y(y) = \sum_{x \in \mathcal{X}} x p_X(x) \sum_{y \in \mathcal{Y}} y p_Y(y) = \mathbb{E}[X]\mathbb{E}[Y]. \tag{19}$$

More generally, if g, h are functions then

$$\boxed{\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)].} \tag{20}$$

If X, Y are independent RVs then the variance of their sum is the sum of their variances:

$$\boxed{\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y).} \tag{21}$$

Here is the proof:

$$\begin{aligned} \mathrm{var}(X + Y) &= \mathrm{var}(\tilde{X} + \tilde{Y}), \ \tilde{X} = X - \mathbb{E}[X], \ \tilde{Y} = Y - \mathbb{E}[Y] \\ &= \mathbb{E}[(\tilde{X} + \tilde{Y})^2] = \mathbb{E}[\tilde{X}^2 + 2\tilde{X}\tilde{Y} + \tilde{Y}^2] = \mathbb{E}[\tilde{X}^2] + 2\mathbb{E}[\tilde{X}]\mathbb{E}[\tilde{Y}] + \mathbb{E}[\tilde{Y}^2] \\ &= \mathbb{E}[\tilde{X}^2] + \mathbb{E}[\tilde{Y}^2] = \mathrm{var}(\tilde{X}) + \mathrm{var}(\tilde{Y}) = \mathrm{var}(X) + \mathrm{var}(Y) \end{aligned} \tag{22}$$

References

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