**MTH 7241 Fall 2022: Prof. C. King**

**Notes 2: Conditioning**

## Conditional probability

$$\mathbb{P}(B|A) = \text{conditional probability that } B \text{ is true given that } A \text{ is true}$$

This is computed using the formula

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}, \quad \mathbb{P}(A) \neq 0$$

It is important to note that $\mathbb{P}(B|A)$ is defined only if $\mathbb{P}(A) \neq 0$.

**Example 1** *The probability of drawing an Ace from a standard deck of cards is*

$$\mathbb{P}(Ace) = \frac{4}{52} = \frac{1}{13}$$

*Draw two cards in sequence, and let $A_1$, $A_2$ be the events that the first, second cards are Aces respectively, then it is easy to see that*

$$\mathbb{P}(A_1) = \frac{4}{52}, \quad \mathbb{P}(A_2|A_1) = \frac{3}{51}, \quad \mathbb{P}(A_2|A_1^c) = \frac{4}{51}$$

In the above example it is perhaps not immediately obvious how to compute $\mathbb{P}(A_2)$. We can use the formula for total probability, which in this case says that

$$\begin{aligned}
\mathbb{P}(A_2) &= \mathbb{P}(A_2 \text{ and } A_1) + \mathbb{P}(A_2 \text{ and } A_1^c) \\
&= \mathbb{P}(A_2|A_1)\,\mathbb{P}(A_1) + \mathbb{P}(A_2|A_1^c)\,\mathbb{P}(A_1^c) \\
&= \frac{3}{51}\,\frac{4}{52} + \frac{4}{51}\,\frac{48}{52} \\
&= \frac{4}{52}
\end{aligned}$$

The general formula for total probability is this: suppose that there is a collection of events $A_1, A_2, \ldots, A_n$ which are mutually disjoint, so $A_i \text{and} A_j = \emptyset$ for all $i \neq j$, and also exhaustive, meaning they include every outcome in the sample space $S$, so that $A_1 \cup A_2 \cup \cdots \cup A_n = S$. Then for any event $B$,

$$\begin{aligned}
\mathbb{P}(B) &= \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \cdots + \mathbb{P}(B \cap A_n) \\
&= \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \cdots + \mathbb{P}(B|A_n)P(A_n)
\end{aligned}$$

**Example 2** Bayes Rule *is a useful application of conditional probability. The formula is*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\,\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)}$$

*Suppose an insurance application contains the question "Is the applicant a smoker?". Assume that* 30% *of the population smokes, and that* 40% *of smokers will lie about it. Assuming no non-smokers will lie, what percentage of applicants who say they are non-smokers actually are non-smokers?*

If the events $A, B$ are independent then the formula $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$ implies that both of the following are true:

$$\mathbb{P}(A|B) = \mathbb{P}(A) \quad \text{and} \quad \mathbb{P}(B|A) = \mathbb{P}(B)$$

The famous *Monty Hall* gameshow problem. There are 3 doors, and a prize is hidden behind one door. The contestant chooses a door. The host then opens *one of the other doors* to show that it does not conceal a prize. The contestant may now change her choice to the third remaining door. Should she switch her choice?

**Conditional expectation**

If $X$ and $Y$ are discrete r.v.'s then we can compute conditional probabilities as above:

$$\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

There is also the formula for total probability

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x|Y = y)\, \mathbb{P}(Y = y)$$

where the sum runs over all possible values of $Y$. Conditioning is a very useful method for solving problems in probability, because it is often much easier to compute conditional probabilities and then sum over the result to find the 'unconditioned' probability.

**Example 3** *Best prize: $n$ distinct prizes arrive in sequence, all have different values, and one is the best. You must pick a prize or else move on to the next one (no going back to earlier ones). Your knowledge consists of the values of the previous prizes. You want to use a strategy that will maximize the probability of selecting the best prize. The prizes are randomly arranged in sequence.*

*Strategy: reject the first $k$ prizes, then select the first one which is better than all of these previous ones. Let $X$ be the position of the best prize. Use*

$$P_k(best) = \sum_{i=1}^{n} P_k(best|X = i)\, P(X = i)$$

*to deduce*

$$P_k(best) \simeq \frac{k}{n} \log \frac{n}{k}$$

*Find value of $k$ to maximize this.*

We define the conditional expectation of $X$ conditioned on the value $Y = y$ as

$$\mathbb{E}[X|Y = y] = \sum_x x\,\mathbb{P}(X = x|Y = y)$$

This number is defined for each possible value of $Y$. Putting these all together we get the r.v. $\mathbb{E}[X|Y]$ as a function of $Y$. You should think of $\mathbb{E}[X|Y]$ as a random variable which is determined by the random variable $Y$, like $Y^2$ or $e^{tY}$: if you know the value of $Y$, then you know the value of $\mathbb{E}[X|Y]$. There is a very useful relation between the conditional expectation $\mathbb{E}[X|Y]$ and the 'unconditioned' expectation $\mathbb{E}[X]$.

**Theorem 1**

$$\mathbb{E}\Big[\mathbb{E}[X|Y]\Big] = \mathbb{E}[X]$$

Note that on the left side we are first averaging over $X$, with $Y$ fixed, and then we average over $Y$. On the right side we do it all in just one step.

**Example 4** *Let* $N, X_1, X_2, \dots$ *be independent, where* $X_i$ *are IID. Define*

$$Y = \sum_{i=1}^{N} X_i$$

*For example,* $N$ *is the number of insurance claims in a month, and* $X_i$ *is the size of the* $i^{th}$ *claim. Then*

$$\mathbb{E}[Y] = \mathbb{E}[X]\,\mathbb{E}[N]$$

**Example 5** *Rats in a maze.*

**Conditioning with respect to a continuous random variable**

Although we will not define conditioning with respect to a continuous random variable in full detail, it is a very useful notion. Let $X$ be a continuous random variable, then for any event $A$ we have

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x)\, f_X(x) dx$$

It is often convenient to use a shorthand and write this as

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{P}(A \mid X)]$$

where it is understood that the quantity $\mathbb{P}(A \mid X)$ is a random variable which is a function of $X$. Many interesting examples arise when the event $A$ involves another random variable.

**Example 6** *Suppose that $X, Y$ are independent exponentials with mean 1 and we want $\mathbb{P}(X+Y \geq z)$ where $z \geq 0$. Now*

$$\mathbb{P}(X+Y \geq z \,|\, X=x) = \mathbb{P}(Y \geq z-x \,|\, X=x) = \mathbb{P}(Y \geq z-x)$$

*because they are independent. Thus*

$$\mathbb{P}(Y \geq z-x) = \begin{cases} e^{-(z-x)} & \textit{for } z-x \geq 0 \\ 1 & \textit{for } z-x < 0 \end{cases}$$

*and hence*

$$\begin{aligned} \mathbb{P}(X+Y \geq z) &= \int_0^\infty \mathbb{P}(X+Y \geq z \,|\, X=x) e^{-x} dx \\ &= \int_0^\infty \mathbb{P}(Y \geq z-x) e^{-x} dx \\ &= \int_0^z e^{-z} dx + \int_z^\infty e^{-x} dx \\ &= z e^{-z} + e^{-z} \end{aligned}$$

The same technique can be applied even when the random variables are dependent.

**Example 7** *Suppose $X$ is uniform on $[0,1]$ and $Y$ is uniform on $[0,X]$. Calculate $\mathbb{E}[Y]$.*

**Memoryless property of exponential r.v.'s**

Conditioning can have quite unexpected effects on the distributions of random variables. One well-known example is the memoryless property of the exponential random variable. Suppose that $X$ is exponential with rate $\lambda$, so that its pdf is

$$f_X(t) = \lambda e^{-\lambda t} \quad \text{for } t \geq 0$$

Then an easy calculation shows that

$$\mathbb{P}(X > t) = e^{-\lambda t}$$

for all $t > 0$. If we condition on this event we find that

$$\mathbb{P}(X > t + s \,|\, X > s) = e^{-\lambda t}$$

This can be interpreted as a memoryless property by viewing $X$ as the time to failure of a device. Conditioning on the event $\{X > s\}$ means that we condition on the device not having failed up to time $s$. The result above says that given this event, the subsequent lifetime of the device has the same distribution as a fresh lifetime.

**Example 8** *Cars pass a point on a highway. The times between successive cars are independent exponential random variables with the same mean* $m$*. Suppose at a random time you stand at the point on the highway. What is the mean time until the next car passes?*

**Example 9** *Cars pass a point on a highway. The times between successive cars are independent identically distributed random variables. The distribution of the time is binary, and has only two values* $\{L, S\}$ *(long and short). The probabilities are* $P(X = L) = p$ *and* $P(X = S) = 1 - p$*. The mean value is thus* $\mathbb{E}[X] = pL + (1 - p)S$*. Suppose at a random time you stand at the point on the highway. What is the mean time until the next car passes?*