Disclaimer: *These notes have not been subjected to the usual scrutiny for formal publications. They are to be used only for the class.*
Outline:

1. Conditional Probability, Conditional Distribution and Conditional Expectation
2. Stable Marriage Problem (Part 1)

1 Conditional Probability, Conditional Distribution and Conditional Expectation

Recall the simple example of a joint distribution we had in recitation 2:

Prob of X,Y	x_1	x_2
y_1	0.1	0.2
y_2	0.4	0.3

There we define *conditional probability* as the ratio of the *joint probability* over the *marginal probability*. For instance,

$$\begin{aligned}
\Pr[Y = y_1 | X = x_1] &= \frac{\Pr[X = x_1, Y = y_1]}{\Pr[X = x_1]} \\
&= \frac{\Pr[X = x_1, Y = y_1]}{\Pr[X = x_1, Y = y_1] + \Pr[X = x_1, Y = y_2]} \\
&= \frac{0.1}{0.1 + 0.4} = 0.2
\end{aligned}$$

Similarly, we can deduce that

$$\Pr[Y = y_2 | X = x_1] = \frac{0.4}{0.1 + 0.4} = 0.8$$

Notice the two numbers above add up to 1. This is not a coincidence. Generally, when we have a joint distribution and we fix one of the random variables, say $X = x_1$, then the other random variable follows a *conditional distribution* which is denoted as $Y \mid X = x_1$. (Hence the probability masses add up to 1). This lets us focus on the distribution of Y given the fact that $X = x_1$. Of course, when X changes, say it changes to $X = x_2$, the conditional distribution changes, too. Often you will see the short-hand $Y \mid X$, which is in fact a random variable whose distribution depends on the particular value of X.

Think of $Y \mid X = x_1$ as simply another random variable. Based on its distribution (the conditional distribution), we can talk about its expectation (and variance and so on), if they exist.

The expected value of $Y \mid X = x$ is called the *conditional expectation* and is defined as

$$E[Y \mid X = x] = \sum_y y \cdot \Pr[Y = y | X = x]$$

(by convention, capital letters denote random variables, which are essentially functions. And lower-case letters denote usual variables.) Because the conditional distribution depends on the particular value of X, so does the conditional expectation. You may even think of it as a function of x as:

$$E[Y \mid X = x] \equiv u(x)$$

Let's go through a simple example of conditional expectation based on the example above:

$$\begin{aligned} E[Y \mid X = x_1] \equiv u(x_1) &= y_1 \cdot \Pr[Y = y_1 \mid X = x_1] + y_2 \cdot \Pr[Y = y_2 \mid X = x_1] \\ &= y_1 \cdot 0.2 + y_2 \cdot 0.8 \end{aligned}$$

Here is a lemma that's useful in proving numerous facts about conditional probability:

$$E[r(X)E(Y \mid X)] = E[r(X)Y]$$

As we said before, you can think of $E(Y \mid X)$ as a function $u(X)$, then the left hand side is really just $E[r(X)u(X)]$ (that is to say the outer expectation doesn't involve the distribution of Y).

To prove this, invoke the definitions of conditional probability, and use the fact known as "change of variable for computing expectation": to compute $E[f(X)]$, one does not have to figure out the distribution of the random variable $f(X)$, but only to plug the values of X into function f:

$$E[f(X)] = \sum_x f(x) \Pr[X = x]$$

The proof then follows:

$$\begin{aligned} E[r(X)E(Y \mid X)] &= \sum_x r(x)E(Y \mid X = x) \Pr[X = x] \text{ by change of variable for computing the expectation} \\ &= \sum_x r(x) \sum_y y \Pr[Y = y \mid X = x] \Pr[X = x] \text{ by definition of conditional expectation} \\ &= \sum_{x,y} r(x) y \Pr[Y = y \mid X = x] \Pr[X = x] \\ &\quad \text{changing the order of summation in any way we want, which typically is OK} \\ &= \sum_{x,y} r(x) y \Pr[Y = y, X = x] \text{ by definition of conditional probability} \\ &= E[r(X)Y] \end{aligned}$$

An immediate corollary is the double expectation theorem:

$$E[E[Y \mid X]] = E[Y] \text{ i.e. } \sum_x$$

Another corollary: if X and Y are independent (the information from X doesn't tell anything about Y),

$$E[Y \mid X] = E[Y]$$

Also, it turns out (in the context of statistics) that $u(X) = E[Y \mid X]$ is the best *predictor* of Y given X (in the situation where you can only observe X but you have to guess what Y is). Then the lemma tells us that the *prediction error* $\varepsilon = Y - u(X) = Y - E[Y \mid X]$ is un-correlated with any

function of X. (Recall that two random variables X, Y are un-correlated iff $E[X] \cdot E[Y] = E[XY]$.) It also follows easily from the lemma.

Recall that with conditional probability, we had

$$\Pr[B \mid A] = \Pr[B \cap A \mid A]$$

(that is, you can assume that the event A always happens when everything is conditioned on A). For instance, when computing $\Pr[A \mid X = x]$, you can assume $X = x$ everywhere in the event A. This leads to the following *substitution rule* for conditional expectation, which formalizes the same kind of intuition:

$$E[s(X,Y) \mid X = x] = E[s(x,Y) \mid X = x]$$

Finally, similar to the usual expectation, conditional expectation is also linear:

$$E[Y + Z \mid X] = E[Y \mid X] + E[Z \mid X] \text{ and } E[cY \mid X] = cE[Y \mid X]$$

2 Stable Marriage Problem (Part 1)

Please refer to the accompanying lecture notes for the description and an example of the problem. We will only cover up to the example part. (the section "analysis" and everything afterwards will be dealt later.)