Assignment 1

Ethan Violette

1/29/2019

2) Hoff Exercise 2.1

Marginal and conditional probability: The social mobility data from Section 2.5 gives a joint probability distribution on (Y__1, Y__2)= (father's occupation, son's occupation). Using this joint distribution, calculate the following distributions:

2.a) the marginal probability distribution of a father's occupation

$Pr(Y_1 = y_1) = Pr(Y_1 = y_1, Y_2 = y_2)$, so the marginal probability distribution of a father's occupation is the sums of the rows of the data, which represents the son's occupation probabilities added up.

	$Y_1 = farms$	$Y_1 = operatives$	$Y_1 = craftsmen$	$Y_1 = sales$	$Y_1 = professional$
$Pr(y_1)$	0.11	0.279	0.277	0.099	0.235

2.b) the marginal probability distribution of a son's occupation

$Pr(Y_2 = y_2) = Pr(Y_1 = y_1, Y_2 = y_2)$, so the marginal probability distribution of a sons's occupation is the sums of the columns of the data, which represents the father's occupation probabilities added up.

	$Y_2 = farms$	$Y_2 = operatives$	$Y_2 = craftsmen$	$Y_2 = sales$	$Y_2 = professional$
$Pr(y_2)$	0.023	0.26	0.24	0.125	0.352

2.c) the conditional distribution of a son's occupation, given that the father is a farmer.

From Bayes' Theorem:

$$Pr(Y_2 = y_2 \mid Y_1 = \text{farmer}) = \frac{Pr(Y_2 = y_2, Y_1 = \text{farmer})}{Pr(Y_1 = \text{farmer})}$$

We know from 2.a) $Pr(Y_1 = \text{farmer}) = .11$:

$$Pr(Y_2 = y_2 \mid Y_1 = \text{farmer}) = \frac{Pr(Y_2 = y_2, Y_1 = \text{farmer})}{.11}$$

For $Pr(Y_2 = y_2, Y_1 = \text{farmer})$, simply select the row from the social mobility data where father's occupation = farm.

$$Pr(Y_2 = y_2 \mid Y_1 = \text{farmer}) = \frac{0.018, 0.035, 0.031, 0.008, 0.018}{.11}$$

	$Y_2 = farms$	$Y_2 = operatives$	$Y_2 = craftsmen$	$Y_2 = sales$	$Y_2 = professional$
$Pr(y_2 \mid Y_1 = farmer)$	0.164	0.318	0.282	0.073	0.164

2.d) the conditional distribution of a father's occupation, given that the son is a farmer.

From Bayes' Theorem:

$$Pr(Y_1 = y_1 \mid Y_2 = \text{farmer}) = \frac{Pr(Y_1 = y_1, Y_2 = \text{farmer})}{Pr(Y_2 = \text{farmer})}$$

We know from 2.b) $Pr(Y_2 = \text{farmer}) = .023$:

$$Pr(Y_1 = y_1 \mid Y_2 = \text{farmer}) = \frac{Pr(Y_1 = y_1, Y_2 = \text{farmer})}{.023}$$

For $Pr(Y_1 = y_1, Y_2 = \text{farmer})$, simply select the column from the social mobility data where son's occupation = farm.

$$Pr(Y_1 = y_1 \mid Y_2 = \text{farmer}) = \frac{0.018, 0.002, 0.001, 0.001, 0.001}{.11}$$

	$Y_1 = farms$	$Y_1 = operatives$	$Y_1 = craftsmen$	$Y_1 = sales$	$Y_1 = professional$
$Pr(y_1 \mid Y_2 = farmer)$	0.783	0.087	0.043	0.043	0.043

3) Hoff Exercise 2.3

Note that f, g and h are just functions. They are not necessarily densities (this is, they don't need to sum/integrate to one). Full conditionals: Let X, Y, Z be random variables with joint density (discrete or continuous) $p(x, y, z) \propto f(x, z)g(y, z)h(z)$. Show that

3.a) p(x|y, z) $\propto$ f(x, z), i.e. p(x|y, z) is a function of x and z

$$\begin{aligned}
p(x \mid y, z) &= \frac{p(x, y, z)}{p(y, z)} \\
&= \frac{p(x, y, z)}{\int p(x, y, z)\ dx} \\
&\propto \frac{f(x, z)g(y, z)h(z)}{\int f(x, z)g(y, z)h(z)\ dx} \\
&\propto \frac{f(x, z)g(y, z)h(z)}{g(y, z)h(z) \int f(x, z)\ dx} \\
&\propto \frac{f(x, z)}{\int f(x, z)\ dx}
\end{aligned}$$

3.b) p(y|x, z) $\propto$ g(y, z), i.e. p(y|x, z) is a function of y and z

$$\begin{aligned}
p(y \mid x, z) &= \frac{p(x, y, z)}{p(x, z)} \\
&\propto \frac{f(x, z)g(y, z)h(z)}{\int p(x, y, z)\ dy} \\
&\propto \frac{f(x, z)g(y, z)h(z)}{\int f(x, z)g(y, z)h(z)\ dy} \\
&\propto \frac{f(x, z)g(y, z)h(z)}{f(x, z)h(z) \int g(y, z)\ dy} \\
&\propto \frac{g(y, z)}{\int g(y, z)\ dy}
\end{aligned}$$

3.c) X and Y are conditionally independent given Z.

$$\begin{aligned}
p(x \mid z) &= \frac{p(x, z)}{p(z)} \\
&\propto \frac{\int p(x, y, z)\ dy}{\int \int p(x, y, z)\ dy\ dx} \\
&\propto \frac{\int f(x, z)g(y, z)h(z)\ dy}{\int \int f(x, z)g(y, z)h(z)\ dy\ dx} \\
&\propto \frac{f(x, z)h(z) \int g(y, z)\ dy}{h(z) \left(\int g(y, z)\ dy\right) \left(\int f(x, z)\ dx\right)} \\
&\propto \frac{f(x, z)}{\int f(x, z)\ dx} \\
&\propto p(x \mid y, z)
\end{aligned}$$

4) (Gelman et al., 2003)

Discuss the following statement. 'The probability of event E is considered "subjective" if two rational persons A and B can assign unequal probabilities to E, $Pr_A(E)$ and $Pr_B(E)$. These probabilities can also be interpreted as "conditional": $Pr_A(E) = Pr(E|I_A)$ and $Pr_B(E) = Pr(E|I_B)$, where I_A and I_B represent the knowledge available to persons A and B, respectively. Apply this idea to the following examples:

4.a) The probability that a '6' appears when a fair die is rolled, where A observes the outcome of the die roll and B does not.

Since A actually sees the outcome of the die roll, A would know ahead of time if the die appeared as 6, or if the number appeared was not six. So A would naturally assign a probability of a '6' appearing as

$$Pr_A(E) = Pr_B(E|I_A) = \begin{cases} 1 & \text{A sees a 6 appear} \\ 0 & \text{A sees a number other than 6 appear} \end{cases}$$

B does not see the outcome of the die roll, so B has no knowledge of the die other than that it is fair. So B would likely assign a probability of a '6' appearing as $Pr_B(E) = Pr_B(E|I_B) = \frac{1}{\text{\# of sides on the die}}$.

Guessing the outcome of a fair die should inherently not be subjective, but the above example results in two rational persons A and B assigning unequal probabilities to the event of interest, which implies that a '6' appearing when a fair die is rolled is subjective. This is because A had more knowledge at their disposal when assigning probability to the event compared to B. A knew the outcome of the roll, so they could accurately assign probability based on this knowledge; unlike B, who did not know the outcome and had to assign probability based on the number of faces on the die.

4.b) The probability that Germany wins the next World Cup, where A is ignorant of soccer and B is a knowledgeable sports fan.

Again I can see a natural disparity in knowledge of the event between A and B. A is ignorant of soccer, so at worst they would give Germany an equal probability of winning the World Cup to any other country they could think of, and at best use information they may know about Germany's large population and proclivity towards soccer to assign a greater probability of Germany winning than, say, Sri Lanka. B is a knowledgeable sports fan, so B would likely know much more about Germany's high ranking history in previous World Cups, the strength of their team, and their various strategies. B would almost certainly assign a different probability to Germany winning the next World Cup compared to A. This implies that $Pr(E|I_A) \neq Pr(E|I_B)$ and $Pr_A(E) \neq Pr_B(E)$, so the probability of Germany winning the next World Cup is subjective as well.