Data Analysis 2014-2015
Home Assignment 4

Tobias Hallundbæk Petersen - xtv657

November 20, 2015

1 Classification

2 Probability theory refreshment

We han an urn that contains five red, three orange, and one blue ball. We now select two balls at random

1. **What is the sample space of this experiment**
   $\{\{R,R\},\{R,O\},\{R,B\},\{O,O\},\{O,R\},\{O,B\},\{B,R\},\{B,O\}\}$

2. **What is the probability of each point in the sample space**
   If we use the same ordering as above, $\{5/18, 5/24, 5/72, 1/12, 5/24, 1/24, 5/72, 1/24\}$, the sum of these probabilities sum up to one, meaning that we have a complete sample space.

3. **Let X represent the number of orange balls selected. What are the possible values of X?**
   The possible values of X are $\{0, 1, 2\}$.

4. **Calculate $\mathbb{P}\{X = 0\}$**
   $5/18 + 5/72 + 5/72 = 5/12$

5. **Calculate $\mathbb{E}[X]$**
   $5/24 + 1/12 \cdot 2 + 5/24 + 1/24 + 1/24 = 2/3$

3 Probability theory refreshment

From probability theory we have the following definitions and properties:

$$(a)\ p_X(x) = \sum_{y \in \mathcal{Y}} p_{XY}(x,y)$$

$$(b)\text{ If } X \text{ and } Y \text{ are independent, then } P_{XY}(x,y) = p_X(x)p_Y(y)$$

$$(c)\ \mathbb{E}[X] = \sum_{x \in \mathcal{X}} x p_X(x)$$

X and Y are discrete random variables that take values from in $\mathcal{X}$ and $\mathcal{Y}$. p_X is the distribution of X, p_Y the distribution of Y and p_{XY} the distribution of X and Y.

1.

We prove the following identity:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

By (c), the expected value of X is given by:

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x p_X(x)$$

Therefore the expected value of $X + Y$ would be

$$\begin{aligned}
\mathbb{E}[X + Y] &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} (x + y) p_{XY}(x,y) \\
&= \sum_{x \in \mathcal{X}} x \sum_{y \in \mathcal{Y}} p_{XY}(x,y) + \sum_{y \in \mathcal{Y}} y \sum_{\in \mathcal{X}} p_{XY}(x,y) \\
&= \sum_{x \in \mathcal{X}} x p(x) + \sum_{y \in \mathcal{Y}} y p(y) \\
&= \mathbb{E}[X] + \mathbb{E}[Y]
\end{aligned}$$

In the last step we use the definition for the expected value of a random variable. We have now shown that $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

2.

To prove the following identity, we use that the random variables X and Y are independent.

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

We can write $\mathbb{E}[XY]$ as

$$\mathbb{E}[XY] = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} xyp_{XY}(x,y)$$

This is where we use that X and Y are independent - using property (b):

$$\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} xyp_X(x)p_Y(y)$$

This can be reduced to prove our identity

$$\begin{aligned}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} xyp_X(x)p_Y(y) &= \sum_{x\in\mathcal{X}} xp_X(x)\sum_{y\in\mathcal{Y}} yp_Y(y)\\ &= \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

This proves the identity $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

3.

A bag has 2 red apples and 2 green apples. There is taken 2 apples from the bag without putting them back into the bag. Let X be the first apple and let Y be the second apple. The joint distribution table of X and Y is seen below:

X / Y	Red	Green
Red	$\frac{1}{6}$	$\frac{2}{6}$
Green	$\frac{2}{6}$	$\frac{1}{6}$

The probability of apple X being red is:

$$\mathbb{E}[X = \text{Red}] = \frac{1}{2}$$

Which is the same probability for apple Y being red. We have that

$$\mathbb{E}[X = \text{Red} \land Y = \text{Red}] = \frac{1}{6}$$

Since $\frac{1}{2}\frac{1}{2} = \frac{1}{4} \neq \frac{1}{6}$ then

$$\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$$

in this example.

4.

The identity to be proved:

$$\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$$

We know that $\mathbb{E}[X] = k$ and that $\mathbb{E}[k] = k$. That means taking the expected value of an expected value will just return the constant you already found. This can be done more than 2 times and it will always be the constant k that is your result.

5.

We want to show that $\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

$$\begin{aligned}
\mathbb{E}[(X - \mathbb{E}[X])^2] &= \mathbb{E}[X^2 - 2 \cdot X \cdot \mathbb{E}[X] + (\mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2] - 2 \cdot \mathbb{E}[X] \cdot \mathbb{E}[X] + (\mathbb{E}[X])^2 \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}$$

4 Markov's inequality vs. Hoeffding's inequality vs. binomial bound

5 Hoeffding's inequality