Fourth Home Assignment
Data Analysis

Nikolaj Dybdahl Rathcke (rfq695)

May 26, 2015

Question 1

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

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

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

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

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.

Question 2

1)

N balls are drawn from a bin with $2N$ balls uniformly and without replacement. The fraction of red balls is ε and $0 < \varepsilon \leq \frac{1}{2}$. The other fraction are green balls $(1-\varepsilon)$. We are asked to show that

$$\mathbb{P}\{N \text{ green balls are pulled in a row}\} \leq e^{-N\varepsilon}$$

The probability of getting N green balls in a row can be defined as

$$\prod_{i=0}^{N-1} \frac{2N(1-\varepsilon)-i}{2N-i}$$

We subtract i since it is without replacement, so the number of balls in the bin decreases.
This can be rewritten:

$$\begin{aligned}\prod_{i=0}^{N-1} \frac{2N(1-\varepsilon)-i}{2N-i} &= \prod_{i=0}^{N-1} \frac{2N-2N\varepsilon-i}{2N-i} \\ &= \prod_{i=0}^{N-1} 1 - \frac{2N\varepsilon}{2N-i}\end{aligned}$$

From the assignment text, we have:

$$1 + x \leq e^x$$

This can be used on the probability of drawing N green balls in a row we found, so we get

$$\begin{aligned}\prod_{i=0}^{N-1} 1 - \frac{2N\varepsilon}{2N-i} &\leq \prod_{i=0}^{N-1} e^{-2N\varepsilon/(2N-i)} \\ &= e^{\sum_{i=0}^{N-1} -2N\varepsilon/(2N-i)}\end{aligned}$$

This neat trick can be used since $e^x \cdot e^x = e^{x+x}$. The constants can be moved outside the sum, so

$$\prod_{i=0}^{N-1} 1 - \frac{2N\varepsilon}{2N-i} \leq e^{-2N\varepsilon \sum_{i=0}^{N-1} 1/(2N-i)}$$

The sum can never exceed 1 and it can never be below $\frac{1}{2}$, which means the factor $-2N\varepsilon$ is at max $-N\varepsilon$ and at minimum $-2N\varepsilon$, i.e. that

$$e^{-2N\varepsilon \sum_{i=0}^{N-1} 1/(2N-i)} \leq e^{-N\varepsilon}$$

If we put the two above inequalities together, we get:

$$\prod_{i=0}^{N-1} 1 - \frac{2N\varepsilon}{2N-i} \leq e^{-2N\varepsilon \sum_{i=0}^{N-1} 1/(2N-i)} \leq e^{-N\varepsilon}$$

Where the left hand side was the probability of drawing N green balls in a row, so we have shown the inequality $\mathbb{P}\{N \text{ green balls are pulled in a row}\} \leq e^{-N\varepsilon}$.

Question 3

This question is about the growth function

1)

We are asked to show that

$$m_{\mathcal{H}}(2N) \leq m_{\mathcal{H}}(N)^2$$

If we split the set of $2N$ into two sets with N points, the most dichotomies we can get is the cross product of the two sets. From "Learning From Data", page 45, we have the upper bound on the growth function

$$m_{\mathcal{H}}(N) \leq 2^N$$

So in each set we have at most 2^N dichotomies. The cross product (and the max number of dichotomies in $m_{\mathcal{H}}(2N)$ is $2^N \cdot 2^N$. So we can insert this upper bound of the left side of the inequality we wanted to show

$$2^N \cdot 2^N \leq m_{\mathcal{H}}(N)^2$$

If the maximum dichotomies is reached each set of N that means we can insert that number on our right side as well:

$$2^N \cdot 2^N \leq \left(2^N\right)^2 \text{ or}$$
$$2^{2N} \leq 2^{2N}$$

Which shows the inequality.

2)

To write a generalization bound that only involves $m_{\mathcal{H}}(N)$, we use equation 2.12 from "Learning From Data":

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}$$

The upper bound that was shown in previous question can easily be substituted into this equation at the cost of a looser bound:

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(N)^2}{\delta}}$$

Question 4

1)

We are asked to prove the following by induction

$$\sum_{i=0}^{d} \binom{N}{i} \leq N^d + 1$$

Basis: The statement should hold for $d = 0$:

$$\sum_{i=0}^{0} \binom{N}{i} \leq N^0 + 1$$
$$1 \leq 2$$

It does.

Inductive step: We assume our first equation holds for d and want to show it holds for $d + 1$ and $d \geq 1$ as well:

$$\sum_{i=0}^{d+1} \binom{N}{i} \leq N^{d+1} + 1$$

We can take out the last element in the sum so we get

$$\binom{N}{d+1} + \sum_{i=0}^{d} \binom{N}{i} \leq N^{d+1} + 1$$
$$= N^d + (N-1)N^d + 1$$

This is where we use the assumption so we insert the upper bound instead of the sum and get:

$$\binom{N}{d+1} + N^d + 1 \leq N^d + (N-1)N^d + 1$$
$$\binom{N}{d+1} \leq (N-1)N^d$$
$$= N^{d+1} - N^d$$

The right hand side grows faster for all N and $d \leq 1$, but not for $d = 0$. However, we have already shown that it holds for $d = 0$.

2)

We can use the above result to derive a bound on $m_{\mathcal{H}}(N)$ by using equation 2.9 from "Learning From Data":

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d} \binom{N}{i}$$

Simply substitute the upper bound in the inequality with the one we found in (1):

$$m_{\mathcal{H}}(N) \leq N^d + 1$$

Again, this bound is a bit looser than equation 2.9.

3)

The equation 2.12 in "Learning From Data" states:

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}$$

Just like before we replace the factor $m_{\mathcal{H}}(N)$ with the upper bound we found in (2):

$$E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4 \cdot ((2N)^d + 1)}{\delta}}$$

For the bound to be meaningful, d should be (alot) less than N.