The Austalian National University
School of Computing
Jo Ciucă & Thang Bui

Semester 2, 2023
Tutorial 7

COMP3670/6670: Introduction to Machine Learning

Question 1 — Marginals and Conditionals

Consider two discrete random variables X and Y , with sample spaces $\{1, 2, 3\}$ and $\{1, 2\}$ respectively, and the following joint distribution $p(x, y)$.

$p(x,y)$	$x=1$	$x=2$	$x=3$
$y=1$	$1/16$	$4/16$	$1/16$
$y=2$	$2/16$	$3/16$	$5/16$

1. Compute the marginal distributions $p(x)$ and $p(y)$

   Solution.

$$p(x) = \sum_y p(x, y)$$

$p(x)$	$x=1$	$x=2$	$x=3$
	$3/16$	$7/16$	$6/16$

$$p(y) = \sum_x p(x, y)$$

$p(y)$	$y=1$	$y=2$
	$6/16$	$10/16$

2. Compute the conditional distributions $p(x \mid Y = 1)$ and $p(y \mid X = 2)$.

   Solution.

$$p(x \mid Y = 1) = p(x, Y = 1)/p(Y = 1) = p(x, Y = 1)(16/6)$$

$p(x \mid Y=1)$	$x=1$	$x=2$	$x=3$
	$1/6$	$4/6$	$1/6$

$$p(y \mid X = 2) = p(y, X = 2)/p(X = 2) = p(y, X = 2)(16/7)$$

$p(y \mid X=2)$	$y=1$	$y=2$
	$4/7$	$3/7$

3. Are X and Y statistically independent?

   Solution. No. $p(X = 1, Y = 1) = 1/16 \neq p(X = 1)p(Y = 1) = 3/16 \times 6/16 = 9/256$

4. Compute the expectation values $\mathbb{E}[X]$ and $\mathbb{E}[Y]$. [1]

   Solution.

$$\mathbb{E}[X] = \sum_x xP(X = x) = 1 \cdot \frac{3}{16} + 2 \cdot \frac{7}{16} + 3 \cdot \frac{6}{16} = \frac{35}{16}$$

$$\mathbb{E}[Y] = \sum_y yP(Y = y) = 1 \cdot \frac{6}{16} + 2 \cdot \frac{10}{16} = \frac{26}{16}$$

[1] Note that $\mathbb{E}[X]$ is shorthand for $\mathbb{E}_X[x]$.

Question 2 **Bayes' Rule**

Here is a bag containing three coins: A fair coin (equally likely to land on heads or tails), a two headed coin (always lands on heads) and a two tailed coin (always lands on tails).

1. You select one of the coins uniformly at random, and flip it. The result is heads. What is the probability of the other side being tails?

   Solution. We apply Bayes's rule. Note that if we see heads, asking if the other side is tails, is the same as asking if the coin we just flipped was the fair coin (as no other coin has both heads and tails). Let HH, HT, TT represent the two headed, the fair, and the two tailed coin respectively.

$$
\begin{aligned}
&p(HT \mid \text{ saw heads }) \\
&= \frac{p(\text{ saw heads } \mid HT)p(HT)}{p(\text{ saw heads })} \\
&= \frac{p(\text{ saw heads } \mid HT)p(HT)}{\sum_{\text{coin}} p(\text{ saw heads } \mid \text{coin})p(\text{coin})} \\
&= \frac{p(\text{ saw heads } \mid HT)p(HT)}{p(\text{ saw heads } \mid HH)p(HH) + p(\text{ saw heads } \mid TT)p(TT) + p(\text{ saw heads } \mid HT)p(HT)} \\
&= \frac{1/2 \cdot 1/3}{1 \cdot 1/3 + 0 \cdot 1/3 + 1/2 \cdot 1/3} \\
&= 1/3
\end{aligned}
$$

2. You select one of the coins uniformly at random, and flip the coin N times. The outcome of every trial is that the coin lands on heads. What is the probability of the other side being tails? What does this result tend to as $N \to \infty$? Can you explain the result?

   Solution. Same as previous answer, but the probability of getting N heads in a row with a fair coin is 2^{-N}. With a two headed coin it's a certainty, with a two tailed coin, an impossibility.

$$
\begin{aligned}
&p(HT \mid N \text{ heads }) \\
&= \frac{p(N \text{ heads } \mid HT)p(HT)}{p(N \text{ heads })} \\
&= \frac{p(N \text{ heads } \mid HT)p(HT)}{\sum_{\text{coin}} p(N \text{ heads } \mid \text{coin})p(\text{coin})} \\
&= \frac{p(N \text{ heads } \mid HT)p(HT)}{p(N \text{ heads } \mid HH)p(HH) + p(N \text{ heads } \mid TT)p(TT) + p(N \text{ heads } \mid HT)p(HT)} \\
&= \frac{2^{-N} \cdot 1/3}{1 \cdot 1/3 + 0 \cdot 1/3 + 2^{-N} \cdot 1/3} \\
&= \frac{2^{-N}}{1 + 2^{-N}} = \frac{1}{1 + 2^{N}}
\end{aligned}
$$

   As $N \to \infty, p(HT \mid N \text{ heads }) \to 0$. This means that as we observe longer and longer sequences of heads in a row, we are increasingly less confident that we took the fair coin out of the bag (and increasingly more confident that we have the 2 headed coin.) Note that $p(HT \mid N \text{ heads})$ will never be exactly 0 for any finite number of observations of heads, as it could be the case (however vanishly unlikely) that we selected the fair coin, and got really lucky and got lots of heads in a row.

Question 3 **Expected Value**

A crooked gambler approaches you with the opportunity to play a game.
"I've got a perfectly ordinary deck of 52 cards. It costs \$1 to play! I draw three cards from the deck. If two of them are red, you win \$1. If all three are red, you win five dollars! What do you say?"

1. Should you play this game or not?

   Solution. Solution. We should play the game if the expected payout is positive (that is, the expectation value of the payout is positive). There are three options. Getting three reds results in a net gain of \$4. Two reds results in a net gain \$0 (we break even). Anything else is a net gain of \$-1 (a loss of \$1). Letting X denote the payout of this game

   $$\mathbb{E}[X] = p(3\text{ reds }) \cdot \$4 + p(2\text{ reds}) \cdot \$0 + p(\text{ anything else })\$(-1)$$

   The odds of three reds is given by

   $$\frac{26}{52} \cdot \frac{25}{51} \cdot \frac{24}{50} = \frac{2}{17}$$

   (as the cards are drawn without replacement). There are three sequences with two reds, RRB, RBR, BRR, and all are equally likely. So the odds of two reds is given by

   $$3 \cdot \frac{26}{52} \cdot \frac{25}{51} \cdot \frac{26}{50} = \frac{13}{34}$$

   So,

   $$p(\text{ anything else }) = 1 - p(3\text{ reds }) - p(2\text{ reds }) = 1 - \frac{2}{17} - \frac{13}{34} = 1/2$$

   Hence,

   $$\begin{aligned}\mathbb{E}[X] &= p(3\text{ reds }) \cdot \$4 + p(2\text{ reds }) \cdot \$0 + p(\text{ anything else })\$(-1) \\ &= \frac{2}{17} \cdot \$4 + p(2\text{ reds }) \cdot \$0 + \frac{1}{2} \cdot \$(-1) \\ &= \$\frac{8}{17} - \$\frac{1}{2} = \$\frac{-1}{34}\end{aligned}$$

   So we stand to lose $\$\frac{1}{34}$, (approximately 3 cents), on average per game. We should not play this game.

2. You reply to the gambler "I'll play if we change the rules, and each time you draw a card, we write the result down, and then reshuffle the card back into the deck."
   Should you play this game now?

   Solution. Same answer as before, but now the drawings are without replacement, so

   $$\begin{aligned}p(3\text{ reds}) &= \left(\frac{26}{52}\right)^3 = \frac{1}{8} \\ p(2\text{ reds}) &= 3 \times \left(\frac{26}{52}\right)^3 = \frac{3}{8} \\ p(\text{ anything else }) &= 1 - \frac{1}{8} - \frac{3}{8} = \frac{1}{2}\end{aligned}$$

   Hence,

   $$\begin{aligned}\mathbb{E}[X] &= p(3\text{ reds }) \cdot \$4 + p(2\text{ reds }) \cdot \$0 + p(\text{ anything else })\$(-1) \\ &= \frac{1}{8} \cdot \$4 + p(2\text{ reds }) \cdot \$0 + \frac{1}{2} \cdot \$(-1) \\ &= \$\frac{1}{2} - \$\frac{1}{2} = \$0\end{aligned}$$

   So, on average, we won't win or lose any money, it's a fair game. So we should be indifferent whether we play this game or not.[2]

[2] Or not play it, if you value your time more than nothing. Or do play it, if you enjoy gambling!

Question 4 **Properties of Conditional Distributions**

Let X, Y be random variables, with corresponding probability distribution functions $p(x) : \mathbb{R} \to \mathbb{R}$ and $p(y) : \mathbb{R} \to \mathbb{R}$ respectively. Let $p(x, y) : \mathbb{R}^2 \to \mathbb{R}$ be the joint probability distribution function. We define the expectation value of a function $f : \mathbb{R} \to \mathbb{R}$ of X to be

$$\mathbb{E}_X[f(x)] = \int_{-\infty}^{\infty} f(x)p(x)dx$$

and the expectation value of a function $g : \mathbb{R}^2 \to \mathbb{R}$ with respect to X and Y to be

$$\mathbb{E}_{X,Y}[g(x, y)] := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)p(x, y)dxdy$$

1. Prove that if a binary function $g(x, y) : \mathbb{R}^2 \to \mathbb{R}$ has no dependence on the second argument y (that is, if $g(x, y) = h(x)$ for some function $h : \mathbb{R} \to \mathbb{R}$), then

$$\mathbb{E}_{X,Y}[g(x, y)] = \mathbb{E}_X[h(x)]$$

Solution.

$$\begin{aligned}\mathbb{E}_{X,Y}[g(x, y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)p(x, y)dxdy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x)p(x, y)dxdy\end{aligned}$$

Swap the limits of integration,

$$\begin{aligned}&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x)p(x, y)dydx \\ &= \int_{-\infty}^{\infty} h(x) \left(\int_{-\infty}^{\infty} p(x, y)dy \right) dx\end{aligned}$$

Apply sum rule,

$$= \int_{-\infty}^{\infty} h(x)p(x)dx = \mathbb{E}_X[h(x)]$$

2. It was given in lectures that the covariance between two random variables can be expressed in one of two ways:

$$\mathrm{Cov}_{X,Y}[x, y] := \mathbb{E}_{X,Y}\left[(x - \mathbb{E}_X[x])(y - \mathbb{E}_Y[y])\right]$$

or as the alternate form

$$\mathrm{Cov}_{X,Y}[x, y] = \mathbb{E}_{X,Y}[xy] - \mathbb{E}_X[x]\mathbb{E}_Y[y]$$

Prove these are equivalent.

Solution.

$$\begin{aligned}&\mathrm{Cov}_{X,Y}[x, y] \\ &= \mathbb{E}_{X,Y}\left[(x - \mathbb{E}_X[x])(y - \mathbb{E}_Y[y])\right] \\ &= \mathbb{E}_{X,Y}\left[xy - x\mathbb{E}_Y[y] - y\mathbb{E}_X[x] + \mathbb{E}_X[x]\mathbb{E}_Y[y]\right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(xy - x\mathbb{E}_Y[y] - y\mathbb{E}_X[x] + \mathbb{E}_X[x]\mathbb{E}_Y[y]\right] p(x, y)dxdy\end{aligned}$$

$$\begin{aligned}= &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp(x, y)\, dx\, dy - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\mathbb{E}_Y[y]p(x, y)\, dx\, dy \\ &- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\mathbb{E}_X[x]p(x, y)\, dx\, dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{E}_X[x]\mathbb{E}_Y[y]p(x, y)\, dx\, dy\end{aligned}$$

We evaluate term by term. Note that $\mathbb{E}_X[x]$ and $\mathbb{E}_Y[y]$ are constants, and so can be taken out the front of any integral.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyp(x,y)\;dx\;dy = \mathbb{E}_{X,Y}[xy]$$

$$\begin{aligned}
&-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\mathbb{E}_Y[y]p(x,y)\;dx\;dy \\
&= -\mathbb{E}_Y[y]\int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} p(x,y)dy\right)dx \\
&= -\mathbb{E}_Y[y]\int_{-\infty}^{\infty} xp(x)dx = -\mathbb{E}_Y[y]\mathbb{E}_X[x]
\end{aligned}$$

$$\begin{aligned}
&-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\mathbb{E}_X[x]p(x,y)\;dx\;dy \\
&= -\mathbb{E}_X[x]\int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} p(x,y)\;dx\right)\;dy \\
&= -\mathbb{E}_X[x]\int_{-\infty}^{\infty} yp(y)dy \\
&= -\mathbb{E}_X[x]\mathbb{E}_Y[y]
\end{aligned}$$

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E}_X[x]\mathbb{E}_Y[y]p(x,y)\;dx\;dy \\
&= \mathbb{E}_X[x]\mathbb{E}_Y[y]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\;dx\;dy \\
&= \mathbb{E}_X[x]\mathbb{E}_Y[y]
\end{aligned}$$

Hence, combining the results

$$\begin{aligned}
\mathrm{Cov}_{X,Y}[x,y] &= \mathbb{E}_{X,Y}[xy] - \mathbb{E}_X[x]\mathbb{E}_Y[y] - \mathbb{E}_Y[y]\mathbb{E}_X[x] + \mathbb{E}_X[x]\mathbb{E}_Y[y] \\
&= \mathbb{E}_{X,Y}[xy] - \mathbb{E}_X[x]\mathbb{E}_Y[y]
\end{aligned}$$

as required.

3. It was given in lectures that if X and Y are statistically independent, then $\mathrm{Cov}_{X,Y}[x,y] = 0$ Prove this.

Solution. If X and Y are statistically independent, we have that $p(x,y) = p(x)p(y)$. Hence,

$$\begin{aligned}
\mathbb{E}_{X,Y}[xy] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\;p(x,y)\;dx\;dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\;p(x)p(y)\;dx\;dy \\
&= \int_{-\infty}^{\infty} xp(x)dx\int_{-\infty}^{\infty} y\;p(y)\;dy \\
&= \mathbb{E}_X[x]\mathbb{E}_Y[y]
\end{aligned}$$

Hence, by the previous question

$$\begin{aligned}
\mathrm{Cov}_{X,Y}[x,y] &= \mathbb{E}_{X,Y}[xy] - \mathbb{E}_X[x]\mathbb{E}_Y[y] \\
&= \mathbb{E}_X[x]\mathbb{E}_Y[y] - \mathbb{E}_X[x]\mathbb{E}_Y[y] = 0
\end{aligned}$$

as required.