COMP2610/COMP6261 — Information Theory: Assignment 0

Sample Solutions

July 30, 2018

1. [8 points] Two fair, six-sided die are rolled. Compute the probability that the sum of the outcomes of the two rolls is:

   (a) (3 pt) Equal to 1

   (b) (3 pt) Equal to 4

   (c) (2 pt) Less than 13

   Solution:

   Let $X_1, X_2 \in \{1, 2, \ldots, 6\}$ be the random variables corresponding to each dice roll. We assume that X_1 is independent of X_2.

   (a) The smallest value X_1 and X_2 each take is 1, therefore $X_1 + X_2 \geq 2$ and so $P(X_1 + X_2 = 1) = 0$.

   (b) The various possibilities for each die that result in a total sum of 4 are $(X_1, X_2) \in \{(1,3), (2,2), (3,1)\}$. Thus,

   $$\begin{aligned}
   &P(X_1 + X_2 = 4) \\
   &= P(X_1 = 1, X_2 = 3) + P(X_1 = 2, X_2 = 2) + P(X_1 = 3, X_2 = 1) \\
   &= P(X_1 = 1)P(X_2 = 3) + P(X_1 = 2)P(X_2 = 2) + P(X_1 = 3)P(X_2 = 1) \\
   &= \left(\frac{1}{6}\right)^2 + \left(\frac{1}{6}\right)^2 + \left(\frac{1}{6}\right)^2 \\
   &= \frac{1}{12}.
   \end{aligned}$$

   (c) The largest value X_1 and X_2 each take is 6. Therefore $X_1 + X_2 \leq 12$ and $P(X_1 + X_2 \leq 13) = 1$.

2. [15 points] Suppose that X, Y are discrete random variables with joint probability

	$Y = -1$	$Y = 0$	$Y = 1$
$X = 0$	0	$\frac{1}{3}$	0
$X = 1$	$\frac{1}{3}$	0	$\frac{1}{3}$

(a) (3 pt) Compute the conditional distributions $p(X|Y=y)$ for all possible values of y.

(b) (2 pt) Compute the conditional distributions $p(Y|X=x)$ for all possible values of x.

(c) (3 pt) Compute $E[X|Y=y]$ for all values of y.

(d) (2 pt) Plot $E[X|Y=y]$ against y.

(e) (5 pt) Are X and Y dependent? If yes, explain why. If no, how do you think X and Y are related?

Solution:

First, we compute the marginal distributions using the sum rule: we have $P(X=x) = \sum_{y\in\{-1,0,1\}} P(X=x,Y=y)$. Thus,

$$\begin{aligned} P(X=1) &= P(X=1,Y=-1)+P(X=1,Y=0)+P(X=1,Y=1) \\ &= \frac{1}{3}+0+\frac{1}{3} \\ &= \frac{2}{3}. \end{aligned}$$

Similarly, $P(Y=y) = \sum_{x\in\{0,1\}} P(X=x,Y=y)$. Thus,

$$\begin{aligned} &P(Y=-1) = P(X=1,Y=-1)+P(X=0,Y=-1) = \frac{1}{3}+0 = \frac{1}{3} \\ &P(Y=0) = P(X=1,Y=0)+P(X=0,Y=0) = 0+\frac{1}{3} = \frac{1}{3} \\ &P(Y=1) = 1-(P(Y=-1)+P(Y=0)) = \frac{1}{3}. \end{aligned}$$

(a) Using the definition of conditional probability:

$$P(X|Y=y) = \frac{P(X,Y=y)}{P(Y=y)}.$$

We plug in each value of y:

- $P(X=1|Y=-1) = (1/3)/(1/3) = 1$
- $P(X=1|Y=0) = 0/(1/3) = 0$
- $P(X=1|Y=1) = (1/3)/(1/3) = 1.$

Note now that $P(X=0|Y=y) = 1-P(X=1|Y=y)$, since X is a binary random variable. The other conditionals may thus be computed trivially.

(b) Using the definition of conditional probability:

$$P(Y=1|X=x) = \frac{P(X=x,Y=1)}{P(X=x)}.$$

We plug in each value of x:

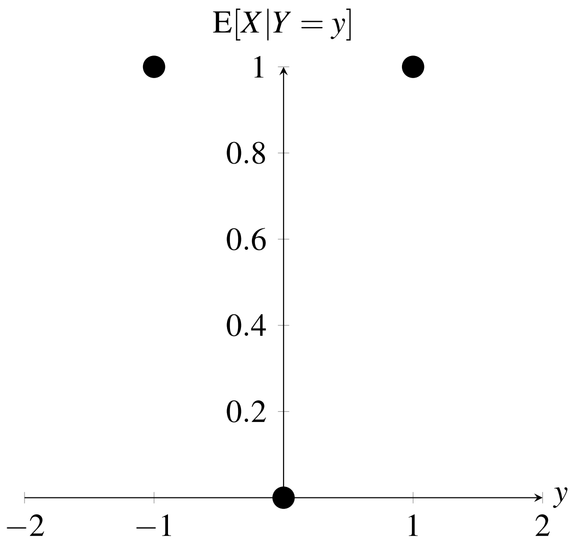
- $P(Y=1|X=0) = 0/(1/3) = 0$, $P(Y=0|X=0) = (1/3)/(1/3) = 1$
- $P(Y=1|X=1) = (\frac{1}{3})/(\frac{2}{3}) = 1/2$, $P(Y=0|X=1) = 0/(1/3) = 0.$

Note now that $P(Y=-1|X=x) = 1-P(Y=1|X=x)-P(Y=0|X=x)$. The other conditionals may thus be computed trivially.

(c) By definition, $\mathrm{E}[X|Y=y] = \sum_{x\in\{0,1\}} x \cdot P(X=x|Y=y)$, which is just $\mathrm{E}[X|Y=y] = 1 \cdot P(X=1|Y=y)$. Since X is Boolean valued, we have:

- $\mathrm{E}[X|Y=-1] = 1 \times P(X=1|Y=-1) = 1$
- $\mathrm{E}[X|Y=0] = 1 \times P(X=1|Y=0) = 0$
- $\mathrm{E}[X|Y=1] = 1 \times P(X=1|Y=0) = 1$.

(d) The plot is given below.



(e) Note that $P(X=0,Y=0) = 1/3$, but $P(X=0)\cdot P(Y=0) = 1/3 \cdot 1/3 = 1/9$, therefore X and Y are not independent. Notably from the plot in the previous part, the dependence is not linear. In fact, we can see that $X = |Y|$, or even $X = Y^2$, is consistent with the observed relationship.

3. [20 points] You have a large jar containing 999 fair coins and one two-headed coin (i.e. a coin that is guaranteed to come up heads). Suppose you pick one coin at random out of the jar and flip it 10 times.

(a) (2 pt) What is the probability you selected a fair coin?

(b) (2 pt) What is the probability of the first flip being heads, assuming you selected a fair coin?

(c) (2 pt) What is the probability of the first flip being heads, assuming you selected the two-headed coin?

(d) (3 pt) What is the overall probability of the first flip being heads?

(e) (3 pt) What is the probability of the first *and* second flip being heads, assuming you selected a fair coin?

(f) (8 pt) What is the probability you selected a fair coin if *all* ten flips turn up heads?

Solution:

Let $N_{\texttt{fair}} = 999$ be the number of fair coins out of a total of $N_{\texttt{total}} = 1000$ coins.

(a) Since we pick a coin at random, $P(\texttt{fair}) = N_{\texttt{fair}}/N_{\texttt{total}} = 999/1000$.

(b) By definition of a fair coin, $P(\texttt{h}|\texttt{fair}) = {}^1\!/_2$.

(c) By definition of a two-headed coin, $P(\texttt{h}|\texttt{biased}) = 1$.

(d) Using the sum rule to find the marginal distribution over h (for heads):

$$\begin{aligned} P(\texttt{h}) &= P(\texttt{h}|\texttt{fair}) \cdot P(\texttt{fair}) + P(\texttt{h}|\texttt{biased}) \cdot P(\texttt{biased}) \\ &= \frac{1}{2} \cdot \frac{999}{1000} + 1 \cdot \left(1 - \frac{999}{1000}\right) \\ &= \frac{1001}{2000} = 0.505. \end{aligned}$$

(e) Since coin flips are independent, and by definition of a fair coin, we have

$$P(\texttt{hh}|\texttt{fair}) = P(\texttt{h}|\texttt{fair}) \cdot P(\texttt{h}|\texttt{fair}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

(f) Like part (d), we can first compute the marginal probability of observing 10 heads by considering the probability under the events of picking the biased and the fair coin:

$$\begin{aligned} P(10\ \texttt{h}s) &= P(10\ \texttt{h}s|\texttt{biased}) \cdot P(\texttt{biased}) + P(10\ \texttt{h}s|\texttt{fair}) \cdot P(\texttt{fair}) \\ &= 1^{10}\left(1 - \frac{999}{1000}\right) + \left(\frac{1}{2}\right)^{10}\left(\frac{999}{1000}\right) \\ &= 0.0019755859375 \end{aligned}$$

Now we use Bayes' rule:

$$\begin{aligned} P(\texttt{fair}|10\ \texttt{h}s) &= \frac{P(10\ \texttt{h}s|\texttt{fair}) \cdot P(\texttt{fair})}{P(10\ \texttt{h}s)} \\ &= \frac{({}^1\!/_2)^{10} \cdot {}^{999}\!/_{1000}}{P(10\ \texttt{h}s)} \\ &= \frac{999}{2023} \\ &= 0.49382105783489866. \end{aligned}$$

4. [32 points] A kidnapping recently occurred in the city of Twin Peaks. Dale, an FBI agent, has been assigned to catch the person who did it. He has determined that the guilty person must be one of the 1,000 members of a club called the White Lodge.

   (a) (2 pt) Leland is a member of the White Lodge. Based on only this information, what is the probability that Leland is guilty?

   (b) (6 pt) Dale has an expensive DNA test that he can use to help his investigation. Suppose the probability of a DNA match given that a person is innocent is 0.1%, and the probability of a DNA match given that a person is guilty is 99%. What is the probability of a DNA match for a randomly chosen member of the White Lodge?

   (c) (9 pt) Dale has found DNA evidence at the scene of the kidnapping. An analysis of the DNA matches a sample that Leland provides. What is the probability that Leland is guilty given the positive outcome of the DNA test? Explain intuitively why this number is higher or lower than that of part (a).

(d) (12 pt) Bob is trying to sell Dale a lie detector machine that he claims is very accurate. He says: *"If my machine starts beeping, you can be 99% sure the person really is guilty"*. What must be the relative fraction of times the machine starts beeping for a guilty person, over the times it starts beeping for an innocent person?

(e) (3 pt) Given just the above information, is it possible to determine the fraction of times that Bob's machine starts beeping for a guilty person? If yes, compute the fraction. If no, explain why not.

Solution:

Let g be the event of Leland being guilty; let m be the event of a DNA match.

(a) If we assume a uniform prior on all 1000 members of the lodge, the prior probability is $P(\mathrm{g}) = {}^{1}\!/_{1000}$.

(b) We are told that the probability of a match of an innocent is 0.1%, so $P(\mathtt{m}|\neg\mathrm{g}) = 0.001$, and the probability of a match for someone who is guilty is $P(\mathtt{m}|\mathrm{g}) = 0.99$. Using the sum and product rules we have:

$$\begin{aligned} P(\mathtt{m}) &= P(\mathtt{m},\mathrm{g}) + P(\mathtt{m},\neg\mathrm{g}) \\ &= P(\mathtt{m}|\mathrm{g})P(\mathrm{g}) + P(\mathtt{m}|\neg\mathrm{g})P(\neg\mathrm{g}) \\ &= 0.99 \cdot \frac{1}{1000} + 0.001 \cdot \frac{999}{1000} \\ &= 0.001989. \end{aligned}$$

(c) Using Bayes' rule we have:

$$P(\mathrm{g}|\mathtt{m}) = \frac{P(\mathtt{m}|\mathrm{g})P(\mathrm{g})}{P(\mathtt{m})} = \frac{0.99 \cdot {}^{1}\!/_{1000}}{0.001989} = 0.497737556561086.$$

We start with some prior beliefs about the guilt of Leland, that is $P(\mathrm{g}) = 0.001$. Upon observing the DNA match we then update our beliefs to take into account the new evidence. The evidence points to Leland being guilty, so naturally our posterior belief of his guilt, $P(\mathrm{g}|\mathtt{m}) \approx 0.498$, is increased relative to the prior.

(d) Let b be the event the machine beeps, g is as before. We're told $P(\mathrm{g}|\mathtt{b}) = 0.99$ and therefore $P(\neg\mathrm{g}|\mathtt{b}) = 0.01$. Using the definition of conditional probability and simplifying, we find

$$\begin{aligned} \frac{P(\mathtt{b}|\mathrm{g})}{P(\mathtt{b}|\neg\mathrm{g})} &= \frac{P(\mathtt{b},\mathrm{g})/P(\mathrm{g})}{P(\mathtt{b},\neg\mathrm{g})/P(\neg\mathrm{g})} \\ &= \frac{P(\mathtt{b},\mathrm{g})P(\neg\mathrm{g})}{P(\mathtt{b},\neg\mathrm{g})P(\mathrm{g})} \\ &= \frac{P(\mathrm{g}|\mathtt{b})P(\mathtt{b})P(\neg\mathrm{g})}{P(\neg\mathrm{g}|\mathtt{b})P(\mathtt{b})P(\mathrm{g})} \\ &= \frac{P(\mathrm{g}|\mathtt{b})P(\neg\mathrm{g})}{P(\neg\mathrm{g}|\mathtt{b})P(\mathrm{g})} \end{aligned} \implies \begin{aligned} \frac{P(\mathtt{b}|\mathrm{g})}{P(\mathtt{b}|\neg\mathrm{g})} &= \frac{P(\mathrm{g}|\mathtt{b})P(\neg\mathrm{g})}{P(\neg\mathrm{g}|\mathtt{b})P(\mathrm{g})} \\ &= \frac{0.99 \cdot 0.999}{0.01 \cdot 0.001} \\ &= 98901. \end{aligned}$$

(e) In the above calculation we were able to compute the relative frequency because the base frequency that the machine beeps, $P(\mathtt{b})$, cancelled out in the step

$$\frac{P(\mathrm{g}|\mathtt{b})P(\mathtt{b})P(\neg\mathrm{g})}{P(\neg\mathrm{g}|\mathtt{b})P(\mathtt{b})P(\mathrm{g})} \longrightarrow \frac{P(\mathrm{g}|\mathtt{b})P(\neg\mathrm{g})}{P(\neg\mathrm{g}|\mathtt{b})P(\mathrm{g})}.$$

This means that regardless of the value of $P(\texttt{b})$, we would end up with the same relative frequency computed above. To compute $P(\texttt{g}|\texttt{b})$ we would need to know $P(\texttt{b})$ since

$$P(\texttt{g}|\texttt{b}) = \frac{P(\texttt{b}|\texttt{g})P(\texttt{g})}{P(\texttt{b})}.$$

So, this is not a number that we can compute with the available information.

5. [25 points] Suppose that X is a random variable with values $\{0,1\}$, and $p(X=1)=\theta$. In the following, express your answer in terms of θ.

(a) (3 pt) Define and calculate the expectation of X.

(b) (4 pt) Define and calculate the variance of X.

(c) (6 pt) Calculate the quantity $\phi(t) := E[e^{tX}]$ for a fixed $t > 0$.

(d) (6 pt) Obtain an expression for the derivative $\phi'(t)$. Explain the relation of $\phi'(0)$ to the result of (a).

(e) (6 pt) Obtain an expression for the second derivative $\phi''(t)$. Explain the relation of $\phi''(0)$ to the result of (b).

Solution:

(a) The expectation of the random variable X is defined as follows:

$$\mathrm{E}[X] := \sum_{x\in\{0,1\}} p(X=x)\cdot x.$$

Therefore

$$\mathrm{E}[X] = 1\cdot p(X=1) + 0\cdot p(X=0) = \theta.$$

(b) One common definition for the variance of X is:

$$\mathrm{var}(X) := \mathrm{E}[X^2] - \mathrm{E}[X]^2.$$

By definition,

$$\mathrm{E}[X^2] = \sum_{x\in\{0,1\}} p(X=x)\cdot x^2 = 1\cdot p(X=1) = \theta.$$

Therefore, by part (a),

$$\mathrm{var}(X) = \theta - \theta^2.$$

(c) By definition,

$$\begin{aligned}
\phi(t) &= \mathrm{E}[e^{tX}] \\
&= \sum_{x\in\{0,1\}} p(X=x)\cdot e^{tx} \\
&= e^{t1}\cdot p(X=1) + e^{t0}\cdot p(X=0) \\
&= \theta e^t + e^0\cdot(1-\theta) \\
&= \theta e^t + 1 - \theta.
\end{aligned}$$

(d) From the expression for $\phi(t)$ in the previous part,

$$\phi'(t) = \frac{d}{dt}\phi(t) = \frac{d}{dt}(\theta e^t + 1 - \theta) = \theta e^t.$$

Note that $\phi'(0) = \theta e^0 = \theta = \mathrm{E}[X]$ by part (a).

(e) From the expression for $\phi'(t)$ in the previous part,

$$\phi''(t) = \frac{d}{dt}\phi'(t) = \frac{d}{dt}\theta e^t = \theta e^t.$$

Revisiting the definition for $\phi(t)$, we equally have

$$\begin{aligned}
\phi''(t) &= \frac{d^2}{dt^2}\left(\mathrm{E}[e^{tX}]\right) \\
&= \mathrm{E}\left[\frac{d^2}{dt^2}e^{tX}\right] \\
&= \mathrm{E}\left[\frac{d}{dt}X \cdot e^{tX}\right] \\
&= \mathrm{E}\left[X^2 \cdot e^{tX}\right].
\end{aligned}$$

Evidently, $\phi''(0) = \mathrm{E}[X^2 \cdot e^0] = \mathrm{E}[X^2] = \theta$. Therefore, $\mathrm{var}[X] = \phi''(0) - \phi'(0)^2$.