Recitation Session 1 Solutions

1. (10 points) *Dependence.*

   a. One might wonder whether the outcome of the first card has any effect. One can compute considering all possible cases for the first card. Let $B_{1,2}$ and $R_{1,2}$ denote the events that the first and the second card turns black and red, respectively:

$$\begin{aligned} \mathrm{P}(B_2) &= \mathrm{P}(B_2 \ \&\ B_1) + \mathrm{P}(B_2 \ \&\ R_1) && (1)\\ &= \mathrm{P}(B_2|B_1)\,\mathrm{P}(B_1) + \mathrm{P}(B_2|R_1)\,\mathrm{P}(R_1) && (2)\\ &= \frac{25}{51}\frac{26}{52} + \frac{26}{51}\frac{26}{52} && (3)\\ &= \frac{1}{2} && (4) \end{aligned}$$

   Discussion: The result is as if it didn't matter what card I picked for the first one. This makes sense because there is no information on the outcome of the first card. Therefore we could have argued that $P(\text{first card is black}) = P(\text{second card is black})$. In fact in the extreme case think of the following problem: given a deck, pick 52 cards, throw away the first 51 ones that are picked and, what is the probability that the last card is Black. This is the same as picking just one card from the deck!

2. (10 points) *Independence.*

   a. Toss two independent fair coins. Let $H_{1,2}$ be the event that the first (resp. second) coin is heads. And let S be the event that both coins land the same way. Note that they are pairwise independent. But given any two the third one follows hence three events are not independent.

3. (10 points) *Which box?*

   a. Intuitively one would vote for Box 3 since it is the one that has the highest ratio of white balls. Note that this assumes that at the initial stage a box is chosen at random, in other words uniformly with probability 1/3. In many cases we do not know the *priors*.

   b. We know the probabilities of choosing a white ball given the chosen box. These are called the *likelihoods*, that measures how likely that a white ball is chosen given information on the box number. Our goal is to reverse this and find probabilities of the box given a white ball is chosen, i.e. find *posterior* probabilities given an observation. Let $i = 1, 2, 3$ denote the number of the box.

$$\begin{aligned} \mathrm{P}(\text{Box } i|\text{White}) &= \frac{\mathrm{P}(\text{Box } i \ \&\ \text{White})}{\mathrm{P}(\text{White})} && (5)\\ &= \frac{\mathrm{P}(\text{White}|\text{Box } i)\,\mathrm{P}(\text{Box } i)}{\sum_{i=1,2,3} \mathrm{P}(\text{White}|\text{Box } i)\,\mathrm{P}(\text{Box } i)} && (6)\\ &= \frac{\frac{i}{i+1}\frac{1}{3}}{\sum_{i=1,2,3} \frac{i}{i+1}\frac{1}{3}} && (7) \end{aligned}$$

   Posteriors are: $\frac{6}{23}, \frac{8}{23}, \frac{9}{23}$.

c. Now we don't know what the priors are. Assign three variables and put the constraint: $\pi_1 + \pi_2 + \pi_3 = 1$. Likelihood becomes $\pi_i \frac{i}{i+1}$, if you have access to data you could try to estimate the priors and then pick the one that maximisez the likelihoods since the posterior is proportional to the likelihoods.

4. (10 points) *Factorization.*

a. Using the definition: $\mathrm{P}(A|B) = \frac{\mathrm{P}(A \,\&\, B)}{\mathrm{P}(B)}$ and assuming $\mathrm{P}(A \,\&\, C)$ non zero.

$$\begin{aligned} \mathrm{P}(A, B|C) &= \frac{\mathrm{P}(A \,\&\, B \,\&\, C)}{\mathrm{P}(C)} && (8) \\ &= \frac{\mathrm{P}(A \,\&\, B \,\&\, C)}{\mathrm{P}(C)} \frac{\mathrm{P}(A \,\&\, C)}{\mathrm{P}(A \,\&\, C)} && (9) \\ &= \mathrm{P}(A|C)\, \mathrm{P}(B|A, C) && (10) \end{aligned}$$

5. (10 points) *Complement of the conditional.*

a. An equivalent condition for conditional independence of A, B given C is: $\mathrm{P}(A|B, C) = \mathrm{P}(A|C)$. Show this using the above result. By now you should have examples for statements like: 'independence doesn't imply conditional independence', 'conditional independence doesn't imply independence', 'conditional independence doesn't imply conditional independence conditioned on the complement' however independent events are still independent when you consider their complements. This is because statements involving conditional independence actually changes the measure! Remember $\mathrm{P}(\cdot|\text{Event})$ is a different measure than $\mathrm{P}(\cdot)$. For further examples see: `https://en.wikipedia.org/wiki/Conditional_independence` and for a discussion of *explaining away* see `http://people.cs.ubc.ca/~murphyk/Bayes/bnintro.html`

6. (10 points) *Example 4.4 from lecture notes.*

a. Solution is in the lecture notes.

7. (10 points) *Change of variables.*

a. Main idea for such problems is that probabilities measure sizes of sets, hence if the set is unchanged then probabilities are also unchanged. In other words $\mathrm{P}(X \in E) = \mathrm{P}(f(X) \in f(E))$ note the slight abuse of notation here, what $X \in E$ means is $\{\omega \in \Omega | X(\omega) \in E\}$. Again with a slight abuse of notation we can write $\mathrm{P}(X \in dx) = \mathrm{P}(Y \in dy)$ which gives us: $f_Y(y) = f_X(x)/|dy/dx|$. Absolute values are there because P measures a quantity that is always positive.

$$\begin{aligned} \frac{dy}{dx} &= \frac{1}{\lambda x} && (11) \\ f_Y(y) &= \frac{1}{1/\lambda x} = \lambda x && (12) \\ &= \lambda \exp(-\lambda y) && (13) \\ & && (14) \end{aligned}$$

It turns out that Y is an exponential distribution with parameter λ. This is a standard way to simulate exponential distribution using uniform distribution.

8. (10 points) *Sampling with or without replacement.*

a. There are N^n possible sequences, each one is equally likely. There are total of G good ones and B bad ones so that $G + B = N$, probability of success (picking a good one) is then $p = G/N$. Distribution of successes are binomial:

$$\mathrm{P}\left(\text{picking } g \text{ good and } b \text{ bad}\right) = \binom{n}{g} \frac{G^g B^b}{N^n} \tag{15}$$

b. There are N order n many possible sequences: $(N)_n = N(N-1)\ldots(N-n+1)$. In other words each of the $\binom{N}{n}$ unordered samples of size n can be ordered in $n!$ ways.

$$\mathrm{P}\left(g \text{ good and } b \text{ bad}\right) = \binom{n}{g} \frac{G}{N} \frac{G-1}{N-1} \cdots \frac{G-g+1}{N-g+1} \frac{B}{N-g} \frac{B-1}{N-g-1} \cdots \frac{B-b+1}{N-g-b+1} \tag{16}$$

$$= \binom{n}{g} \frac{(G)_g (B)_b}{(N)_n} \tag{17}$$

Discussion: In general there are two strategies one might follow: (1) first identify whether the sequences are equally likely, then count how many such sequences are there, finally calculate the probability of a particular sequence, or (2) again identify types of sequences that are equally likely, count the total number of desired outcomes and divide by the total number of outcomes. Common mistakes occur when one starts with one line of thinking and halfway through the problem switches to the other way of counting.