Lab08 Probability and Distributions

Give for a set with the 7 elements $\{A, B, C, D, E, F, G\}$ the ***number of distinct sample outcomes*** given a sample size 5 elements. Show the *professionally typeset equations* that you applied to derive each answer.

(a) Sampling with replacement and distinguishing the order.

$$n^r \Rightarrow 7^5 = 16807$$

(b) Sampling with replacement and irrespectively of the order.

$$C_r^{n+r-1} = \binom{n+r-1}{r} \Rightarrow \frac{11!}{(11-5)! \cdot 5!} = 462$$

(c) Sampling without replacement and distinguishing the order.

$$P_r^n = \frac{n!}{(n-r)!} \Rightarrow \frac{7!}{(7-5)!} = 2520$$

(d) Sampling without replacement and irrespectively of the order.

$$C_r^n = \binom{n}{r} \Rightarrow \binom{7}{5} = \frac{7!}{(7-5)! \cdot 5!} = 21$$

Task 2: Joint, marginal and conditional probabilities (1.0 point)

A speed radar identifies speeders and it also captured the license plates, which helped to distinguish between cars own by local residents, non-residents and commercial vehicles. In a given morning in total 300 cars passed by the speed radar.

(a) Complete table by filling in the missing joint and marginal frequencies. Insert the missing **counts in red**. (0.2 points)

Speed\Plates	Resident	Not Resident	Commercial	Sum
Tolerable	105	120	15	240
Excessive	15	30	15	60
Sum	120	150	30	300

(b) Calculate the ***joint probabilities*** and the ***marginal probabilities*** for the observed table. Report these in a properly formatted table rounded to three decimal places. (0.2 points)

Speed\Plates	Resident	Not Resident	Commercial	Sum
Tolerable	0.350	0.400	0.050	0.800
Excessive	0.050	0.100	0.050	0.200
Sum	0.400	0.500	0.100	1.000

(c) Calculate the joint probabilities of the cells ***assuming independence*** between Origin and Speed. Report these in a properly formatted table rounded to three decimal places. (0.2 points)

$$\Pr(speed \cap plates) = \Pr(speed) \cdot \Pr(plates)$$

Speed\Plates	Resident	Not Resident	Commercial	Sum

Tolerable	0.320	0.400	0.080	0.800
Excessive	0.080	0.100	0.020	0.200
Sum	0.400	0.500	0.100	1.000

(e) Calculate the ***conditional probabilities*** for the speed conditional on who drives the vehicle (local resident, not a resident or commercial driver). Report these in a properly formatted table rounded to three decimal places. (0.2 points)

The condition here is the origin. Therefore, $\Pr(speed|plates) = \frac{\Pr(speed \cap plates)}{\Pr(plates)}$

Speed\Plates	Resident	Not Resident	Commercial
Tolerable	0.875	0.800	0.500
Excessive	0.125	0.200	0.500
Sum	1.000	1.000	1.000

(f) Interpret the conditional probabilities. Do the "*data tell a story*" about the driving habits of the local residents, drivers from outside the neighborhood and commercial drivers? (0.2 points)

Comment:

[a] Local drivers stick to the speed limits in their neighborhood. This is perhaps due to peer-pressure of the neighbors and the realization that their kids may be playing on the street. [b] Outside drivers do not have neighbors peer-pressuring them and it is unlikely that their kids are on these local roads. They, therefore, have a higher tendency ignore the speed limit. [c] Commercial drivers are under severe time pressure and they therefore speed. However, due to the risk of having their license (i.e., their livelihood) suspended, they are a little bit more cautious than the non-residents.

Task 3: Application of Probability Rules (0.4 points)

A group of 3 colleagues fly jointly to a conference. Their "cheapskate" boss only funds no-frill tickets, for which the seat assignment is done randomly by the airline. Their plane has 2×20 no-frill rows of 3 seats each (window, middle, and aisle).

On both flights (to the conference and back) all 3 colleagues end up in the middle seat. Typeset your calculations with the equation editor.

(a) What is the probability that the group of colleagues is jointly unlucky on just one of the flights to end up on a middle seat? (0.1 points)

Comments: See lecture notes Chapt05ProbB.pdf pages 13-14.

The binomial distribution cannot be used because the probability of getting a middle seat changes with each placement, therefore:

$$\Pr(3\text{ middle seats}) = \frac{40}{120} \times \frac{39}{119} \times \frac{38}{118} = 0.0351801$$

(b) What is the probability that the group of colleagues is jointly unlucky on both flights ending up on the middle seats? (0.1 points)

Comment: The seat assignment at the first flight is independent from the seat assignment at the second flight.

$$\begin{aligned}\Pr(\text{middle seats at both flights}) &= \Pr(\text{ first flight}) \cdot \Pr(\text{ second flight}) \\ &= 0.0351801^2 = 0.0012376446\end{aligned}$$

It is highly unlikely in that the colleagues at both flights were assigned just middle seats. Can the colleagues still trust the airline's random seat assignment?

Probabilities based on sets: Let the following relationship between the sets A and B hold: $A \subset B \subset \Omega$

(c) Is the probability of $\Pr(A)$ or $\Pr(B)$ larger? Justify your answer (0.1 points)

Comment: the probability of $\Pr(B)$ is larger because set B has larger sample size.

(d) What is the probability of $\Pr(A|B)$? Justify your answer (0.1 points)

Comment:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)}{\Pr(B)}$$

Task 4: Bayesian Theorem (1 point)

A real estate agent has a client with a specific preference of two school districts. The client is only a short period in town, so that the real estate agent can only show houses in one school district. School district A has 16 available houses whereas school district B has 8 available houses.

(a) Give the ***prior probabilities*** for each school district. Which school district shall the real estate agent take her clients to maximize the likelihood of finding a new home for the client? Make use of standard probability notation and show your equation. (0.2 points)

$$\Pr(\mathrm{A}) = \frac{16}{16+8} = 0.67;\ \Pr(\mathrm{B}) = \frac{8}{16+8} = 0.33$$

According to the higher prior probabilities the real estate agent would visit neighborhood A.

While talking to the client the agent learns that the client prefers a home with mature trees on the property. This information is not part of regular real estate listings. However, from previous visits to districts A and B the agent has seen more trees in district A. She subjectively thinks that a home site in district A is treed with probability $\Pr(\mathrm{Treed}|\mathrm{A}) = 1/4$ whereas for school district B it is $\Pr(\mathrm{Treed}|\mathrm{B}) = 3/4$.

(b) Calculate the total probability $\Pr(\text{Treed})$. Make use of standard probability notation and show your equation. (0.4 points)

$$\begin{aligned}\Pr(\text{Treed}) &= \Pr(\text{Treed}|\text{A}) \cdot \Pr(\text{A}) + \Pr(\text{Treed}|\text{B}) \cdot \Pr(\text{B}) \\ &= \frac{1}{4} \cdot \frac{2}{3} + \frac{3}{4} \cdot \frac{1}{3} \\ &= 0.417\end{aligned}$$

(c) Calculate the posteriori probabilities $\Pr(\text{A}|\text{Treed})$ and $\Pr(\text{B}|\text{Treed})$. Make use of standard probability notation and show your equation. Will the real estate revise her choice of neighborhood in which she will show her client's homes? (0.4 points)

$$\Pr(\text{A}|\text{Treed}) = \frac{\Pr(\text{Treed}|\text{A}) \cdot \Pr(\text{A})}{\Pr(\text{Treed})} = \frac{\frac{1}{4} \cdot \frac{2}{3}}{\frac{5}{12}} = \frac{2}{5} = 0.4$$

$$\Pr(\text{B}|\text{Treed}) = \frac{\Pr(\text{Treed}|\text{B}) \cdot \Pr(\text{B})}{\Pr(\text{Treed})} = \frac{\frac{3}{4} \cdot \frac{1}{3}}{\frac{5}{12}} = \frac{3}{5} = 0.6$$

Taking the preferences for trees into account, the real estate will change to neighborhood B now because the probability of sales success will be higher there.

Task 5: The binomial and Poisson distributions (0.8 points)

Chapter 14 “Probability Distributions” in Lander shows you how to calculate probabilities of events X for the binomial distribution and Poisson distribution.

Calculate the probabilities of X for the binomial distribution with $\pi = 0.4$ and $n = 20$: (0.4 points)
[i] $\Pr(6 \leq X|\pi = 0.4, n = 25) =$ 0.9706378

```
1- pbinom(5, size=25, prob=0.4)
```

[ii] $\Pr(6 < X|\pi = 0.4, n = 25) =$ 0.9264347

```
1- pbinom(6, size=25, prob=0.4)
```

[iii] $\Pr(4 \leq X \leq 8|\pi = 0.4, n = 25) =$ 0.2711647

```
pbinom(8, size=25, prob=0.4) - pbinom(3, size=25, prob=0.4)
```

[iv] $\Pr(X < 4 \cup X > 8|\pi = 0.4, n = 25) =$ 0.7288353

```
1-(pbinom(8, size=25, prob=0.4) - pbinom(3, size=25, prob=0.4))
```

Calculate the probabilities of X for a Poisson distribution with $\lambda = 1$ (0.2 points)
[i] $\Pr(0 < X \leq 6|\lambda = 1)$

```
(ppois(6, lambda = 1)-ppois(0, lambda = 1))
0.6320373
```

[ii] $\Pr(6 < X < \infty | \lambda = 1)$

```
(1-ppois(6, lambda = 1))
8.324115e-05
```

(c) What are the expectation and variance for the binomial and Poisson distributions in tasks 5 (a) and 5(b)? Professionally typeset the used equations. (0.2 points)

[1] Binomial distribution

$$E = n \times p = 25 \times 0.4 = 10$$

$$Var = E \times (1 - p) = 10 \times 0.6 = 6$$

[2] Poisson distribution

$$E = Var = \lambda = 1$$

Task 6: Continuous Density and Distribution Functions (0.4 point)

(a) ***Manually*** draw the density-function and distribution-function of a continuous uniformly distributed random variable $X \sim U(0.5, 1.5)$ into the figures below. (0.2 points)

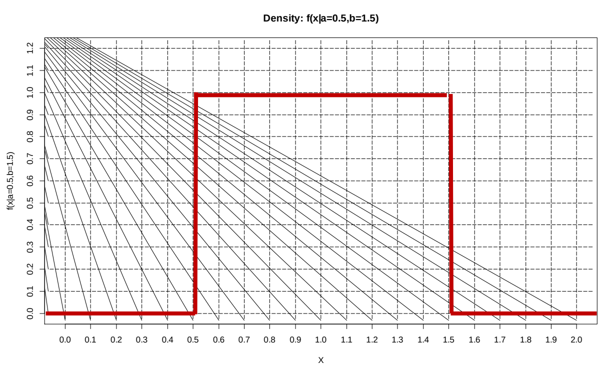

Figure 1: Density Function

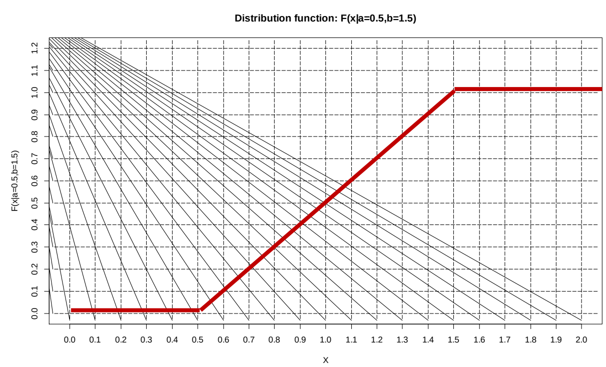

Figure 2: Distribution Function

(b) Manually calculate the following probabilities for $X \sim U(0.5, 1.5)$: (0.2 points)

(i) $\Pr(X \geq 1.0 | a = 0.5, b = 1.5) =$ 0.5

(ii) $\Pr(0.75 \leq X \leq 1.25 | a = 0.5, b = 1.5) =$ 0.5