Stat310: Final

Name: Hadley Wickham

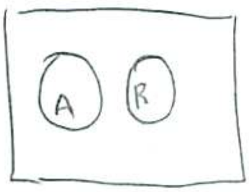
Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page. Do not answer questions on a separate sheet of paper.

You have **three hours** and you may use three double-sided page of notes. You may not use your text book or a computer (calculators are fine), or communicate anything about the contents of this test to anyone else.

Due 5pm May 4.

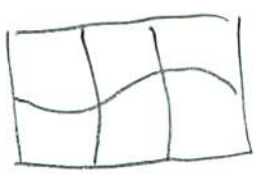
Pledge (including time started and finished):

1. (a) Draw a venn diagram that illustrates the term mutually exclusive. (1)



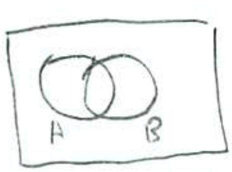
✓ $A \cap B = \emptyset$

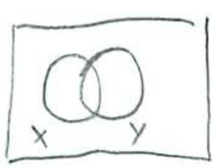
(b) Draw a venn diagram that illustrates the term partition. (1)



✓ mutually exclusive
+ exhaustive

(c) Show that it's hard to illustrate independence in a Venn diagram by creating two events X and Y that are independent, and two events A and B that are not independent, and both sets of events have the same Venn diagram. (2)





✓ 1pt

$P(A \cap B) \neq P(A)P(B)$

$P(X \cap Y) = P(X)P(Y)$ ✓ 1pt

2. Verify that for an event Z with $P(Z) > 0$, the conditional probability function $P(A|Z)$ is a valid probability function.

(a) Show that $P(A|Z) \geq 0$ for all $A \subset S$. (1)

$$P(A|Z) = \frac{P(A \cap Z)^{\,\geq 0}}{P(Z)_{\,\geq 0}} \Rightarrow P(A|Z) \geq 0 \quad \square$$

1pt

(b) Show that $P(S|Z) = 1$ (1)

$$P(S|Z) = \frac{P(S \cap Z)}{P(Z)} = \frac{P(Z)}{P(Z)} = 1 \quad \square$$

1pt

(c) Show that $P(A \cup B|Z) = P(A|Z) + P(B|Z)$ if $A \cap B = \emptyset$ (2)

$$P(A \cup B|Z) = \frac{P((A \cup B) \cap Z)}{P(Z)} = \frac{P((A \cap Z) \cup (B \cap Z))}{P(Z)}$$

$$= \frac{P(A \cap Z) + P(B \cap Z)}{P(Z)} \quad \text{1pt} \quad (A \cap Z) \cap (B \cap Z) = \emptyset$$

$$= P(A|Z) + P(B|Z) \quad \square$$

1pt

Or via any other method if steps correct

3. (a) Using the change of variables technique (or otherwise), find the pdf of X if $X = aZ + b$ and $Z \sim \text{Normal}(0, 1)$. (4)

$$f_z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$x = az + b$$

$$z = \frac{x-b}{a} \qquad v(x) = \frac{x-b}{a} \quad \text{1pt} \qquad v'(x) = \frac{1}{a} \quad \text{1pt}$$

$$f_x(x) = f_z(v(x))\,|v'(x)|$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x-b}{a}\right)^2/2} \left|\frac{1}{a}\right| \quad \text{1pt}$$

$$= \frac{1}{\sqrt{2\pi}\,\underbrace{|a|}} e^{-\frac{(x-b)^2}{2a^2}} = \frac{1}{\sqrt{2\pi a^2}} e^{-\frac{(x-b)^2}{2a^2}}$$

give 1pt
if correct but
missing | |
give 2pts (1 bonus)
if kept | |

(b) What is the distribution of X? (1)

$$X \sim \underline{\text{Normal}}(\mu = \underline{b},\ \sigma^2 = \underline{a^2})$$

- must have all correct to get point
- if pdf wrong, give pt if followed through correctly

4. The waiting time, in minutes, between Lola's barks can be modelled as $W \sim \text{Exponential}(\theta = 10)$.

(a) What's the probability that Lola waits less that 10 minutes (the expected value) before barking next? (1)

$$P(X \leq 10) = 1 - e^{-10/10} = 0.632 \quad 1\text{pt}$$

(b) Given that Lola hasn't barked in the last hour, what's the probability that she barks in the next 5 minutes? (3)

Exponential is memoryless $\Rightarrow$ 1pt

$$P(X \leq 65 \mid X \geq 60) \quad 1\text{pt}$$

$$= P(X \leq 5)$$

$$= 1 - e^{-5/10} = 0.393 \quad 1\text{pt}$$

OR

$$P(X \leq 65 \mid X \geq 60) \quad 1\text{pt}$$

$$= \frac{P(60 \leq X \leq 65)}{P(X \geq 60)} \quad 1\text{pt}$$

$$= \frac{(1 - e^{-65/10}) - (1 - e^{-60/10})}{(1 - e^{-60/10})}$$

$$= \ldots = 0.393 \quad 1\text{pt}$$

(c) How many times do you expect Lola to bark in one day? (1)

1 day = 24×60 minutes

$$\Rightarrow \frac{24 \times 60}{10} = 144 \text{ barks} \quad 1\text{pt}$$

(d) What's the distribution of the number of times Lola barks in a day? (1)

Poisson 1pt

5. Let X and Y be independent random variables with distributions $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(m, q)$. Let $Z = X + Y$

(a) What's the mgf of Z? (Hint: the mgf of X is $(1 - p + pe^t)^n$) (1)

$$M_Z(t) = M_X(t)\, M_Y(t) \quad \text{independent}$$

$$= (1 - p + pe^t)^n (1 - q + qe^t)^m \quad \text{1pt}$$

(b) What's the distribution of Z if $p = q$? (2)

$$\Rightarrow \quad M_Z = (1 - p + pe^t)^n (1 - p + pe^t)^m$$

$$= (1 - p + pe^t)^{n+m} \quad \text{1pt}$$

$$\Rightarrow \quad Z \sim \text{Binomial}(n+m, p) \quad \text{1pt}$$

(c) What's the distribution of Z if $m = n$? (1)

Mgf doesn't simplify $\Rightarrow$ distribution does not have name 1pt

(d) How would you find the distribution of Z if X and Y weren't independent? (1)

Any of :
* change of variables
* $\iint e^{(x+y)t} f(x,y)\, dx\, dy$

$\Rightarrow$ then match to known

6. Given $X_i \overset{iid}{\sim}$ Bernoulli(p), show that the maximum likelihood estimator for p is $\bar{X}_n = \sum_{i=1}^n X_i/n$. (You don't need to verify that the extremum you find is a maximum, but if you do you'll get a bonus point.) (5)

$$f(x_i) = p^{x_i}(1-p)^{1-x_i}$$

$$f(\underset{\sim}{x}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$ 1pt (must use x_i)

$$= p^{\sum x_i}(1-p)^{n-\sum x_i}$$

$$= p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$$

1pt for algebraic simplification either here or when logged (don't need to use $\bar{x}n = \sum x_i$)

$$\ell(\underset{\sim}{x}) = n\bar{x}\ln p + n(1-\bar{x})\ln(1-p)$$ 1 pt

$$\frac{\partial \ell(\underset{\sim}{x})}{\partial p} = \frac{n\bar{x}}{p} - \frac{n(1-\bar{x})}{1-p} = 0$$ 1pt

$$\Rightarrow (1-p)n\bar{x} - np(1-\bar{x}) = 0$$

$$n\bar{x} - \cancel{pn\bar{x}} - np + \cancel{np\bar{x}} = 0$$ 1pt

$$\Rightarrow \hat{p} = \bar{x}$$

1pt (bonus)

$$\frac{\partial^2 \ell(x)}{\partial p^2} = -\frac{n\bar{x}}{p^2} + \frac{np(1-\bar{x})}{(1-p)^2}$$

at $p = \bar{x} \Rightarrow$ $$\frac{-n\cancel{\bar{x}}}{\bar{x}^{\cancel{2}}} + \frac{n\bar{x}(1-\bar{x})}{(1-\bar{x})^2} = \frac{-n(1-\bar{x})^{\cancel{2}} + n\bar{x}\cancel{(1-\bar{x})}}{\bar{x}(1-\bar{x})^{\cancel{2}}} = \frac{\overset{-ve}{-n - \cancel{n\bar{x}} + \cancel{n\bar{x}}}}{\underset{+ve}{\bar{x}(1-\bar{x})}}$$ $\Rightarrow$ maximum

7. Let the event that a stat310 student is in class when I call their name in a your turn be modelled as $C_i \overset{iid}{\sim} Bernoulli(p)$, where p is the probability that a student is in class.

(a) Over the course of the year, I called out 64 names in your turns and 49 people responded. What is the maximum likelihood estimate of p? (1)

$$\hat{p} = \frac{49}{64} = 0.766 \quad \text{1pt}$$

(b) We know that $\hat{p} \sim \text{Normal}(p, p(1-p)/n)$. Use this to construct a 95% confidence interval for p. (Hint: $\Phi(-1.96) = 0.025$) (4)

$$\hat{sd}(\bar{x}_n) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.053 \quad \text{1pt}$$

$$\Rightarrow \frac{\hat{p} - p}{\hat{sd}(\hat{p})} \sim N(0,1) = Z$$

$$P(-a < Z < a) = 0.95 \quad \Rightarrow \quad a = 1.96 \quad \text{1pt}$$

$$\Rightarrow \hat{p} \in (0.766 - 1.96 \cdot 0.053, \quad 0.766 + 1.96 \cdot 0.053) \quad \text{1pt}$$

$$\in (0.66, \quad 0.87) \quad \text{1pt}$$

(c) We could also try to construct a confidence interval using our knowledge that $n\bar{X} \sim Binomial(n, p)$. Why won't this work? (1)

$n\bar{x} \sim \text{Binomial}(n,p) \Rightarrow \bar{X} \sim$ Some Dist with some parameters at least involving p (might expect n to disappear)

can't have p in the dist because we don't know what it is! 1pt

OR Distribution of $\bar{x}$ will be discrete so it may not be possible ~~to~~ to have a confidence interval of the desired length 1pt

8. I collected 15 numbers from an experiment that can be well modelled with a Normal distribution with mean 0 and unknown variance. The mean of the 15 numbers was 0.035 and the variance (S^2) was 0.71. Construct a hypothesis test to see if the numbers came from a distribution with variance 1, or if the real variance was bigger than that. (5)

1 pt
$$\left[\begin{array}{l} H_0 : \sigma^2 = 1 \\ H_1 : \sigma^2 > 1 \end{array}\right.$$

Test statistic : S^2

Under H_0 : $T = \frac{S^2}{\sigma^2}(n-1) \sim \chi^2(n-1)$ 1 pt

$\Rightarrow 14 S^2 \sim \chi^2(14)$ 1 pt

$P(S^2 > 0.71 \mid H_0)$

$= P(T > 14 \times 0.71 = 10)$

$= 1 - 0.24$

$= 0.76$ 1 pt

1 pt wouldn't reject null

CDF of $\chi^2(14)$

x	$P(X \leq x)$
8	0.11
9	0.17
10	0.24
11	0.31
12	0.39
13	0.47
14	0.55
15	0.62

Question	Points	Score
1	4	
2	4	
3	5	
4	6	
5	5	
6	5	
7	6	
8	5	
Total:	40	