Stat310: Test 2

> 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.
> You have **80 minutes** and you may use one double-sided page of notes. You may not use your text book or a computer (calculators are fine). Please pledge on the last page.

Name: Hadley Wickham

1. Let P be a random variable that is 1 if someone smokes pot and 0 otherwise. $P \sim \text{Bernoulli}(p)$

   (a) What does $P \sim \text{Bernoulli}(p)$ mean in this context? A person either smokes pot or doesn't, so where does the randomness come from? (1)

   Random selection of the person ✓

   (b) We can estimate the true value of p by performing a survey and asking each participant whether or not they smoke pot. Let Q_i be 1 if person i smokes pot and 0 otherwise. Use the plug-in principle to write an estimate for p if $i = 1, \ldots, n$. (2)

   $$\hat{p} = \sum_{i=1}^{n} \frac{Q_i}{n}$$ ✓ ✓

   (c) Compute the variance of that estimator. (Hint: think about the distribution of the sum) (3)

   $$\sum_{i=1}^{n} Q_i \sim \text{Binomial}(n, p) ✓$$

   $$\Rightarrow Var(\Sigma Q_i) = np(1-p) ✓$$

   $$\Rightarrow Var\left(\frac{\Sigma Q_i}{n}\right) = \frac{p(1-p)}{p} ✓$$

   OR

   $$Var\left(\frac{\Sigma Q_i}{n}\right) = \frac{1}{n^2} Var(\Sigma Q_i) ✓$$

   $$= \frac{1}{n^2} \Sigma Var(Q_i) ✓ \quad Q_i \text{ indep}$$

   $$= \frac{1}{n^2} n p(1-p)$$

   $$= \frac{p(1-p)}{n} ✓$$

Another way to estimate p is to use a randomised response question like we did in class. Instead of asking the participant directly, we can ask them to flip a coin and say yes if it's heads, and to answer truthfully otherwise. Let H be 1 for heads and 0 for tails, and let R be 1 if they answer yes to this question and 0 otherwise.

(d) Fill in the blanks and then use that information to derive the joint distribution of H and R, and then the marginal distribution of R. (3)

$P(H = 0) =$ 0.5 $\quad P(H = 1) =$ 0.5 $\qquad$ All correct = 1 p

$P(R = 1|H = 0) =$ p $\quad P(R = 1|H = 1) =$ 1

This question confused many people. Give the benefit of the doubt

	H = 0	H = 1
R = 0	$(1-p)/2$	0
R = 1	$p/2$	$1/2$

Checking:

$\frac{p}{2} + \frac{1-p}{2} + \frac{1}{2} = 1$ ✓

All correct = 1 pt

If early mistake made give points for correctly carrying-through

$P(R = 0) =$ $(1-p)/2$ $\quad P(R = 1) =$ $p/2 + 1/2$ $\qquad$ Both correct = 1 pt

(e) Give a transformation of P and H that produces R. (1)

$$R = H + (1 - H)\,P$$ ✓

1 if heads $\qquad$ 1 if tails $\qquad$ whether or not they smoke pot

(f) If $\bar{R}_n = \sum_i^n R_i/n$, then we can estimate p with $2\bar{R}_n - 1$. Show that the expected value of this statistic is p. (2)

$$E(2\bar{R}_n - 1) = 2E(\bar{R}_n) - 1$$

$$= 2E\left(\frac{\Sigma R_i}{n}\right) - 1$$

$$= \frac{2}{n}\Sigma E(R_i) - 1$$

1 pt

$$E(R_i) = 0 \cdot P(R=0) + 1\,P(R=1)$$
$$= p/2 + 1/2$$

$$= \frac{\cancel{2}}{\cancel{n}}\,\cancel{n}\,(p/\cancel{2} + 1/\cancel{2}) - 1$$

1 pt

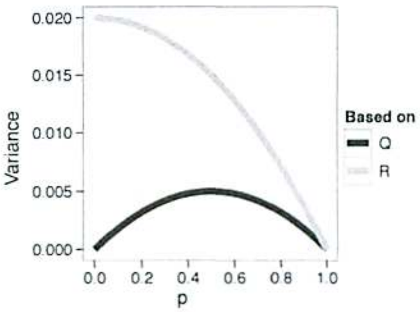
$$= p + 1 - 1$$

$$= p \quad \square$$

(g) The following figure shows the variance of the two estimators for different values of true p. What do you notice? Is there any reason you would choose an estimator based on R instead of Q? (3)

The variance of the estimator based on R is uniformly higher that that based on Q. ✓

We'd use R if we expected that it would lead more people to tell the truth ie. $E(\bar{Q}_n) < p$ ✓✓



2. Let X_i be independent Poisson(λ_i). (Therefore the mgf of X_i is $M_{x_i}(t) = e^{\lambda_i(e^t-1)}$).

(a) Let $S_n = \sum_i^n X_i$. What is M_{S_n}, the mgf of S_n? (3)

$$M_{S_n}(t) = E(e^{S_n t}) = E(e^{\sum X_i t})$$

✓ i.e prod mgf = sum of r.v.

$$= \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} e^{\lambda_i (e^t - 1)}$$ ✓

$$= e^{\sum_{i=1}^{n} \lambda_i (e^t-1)}$$ ✓

(b) What is the distribution of S_n? ✓ (2)

$$S_n \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right)$$ ✓

(c) If $\lambda_i = \lambda$ and n is large, give a distribution (with parameters) that will closely approximate the distribution of S_n. (3)

$E(X_i) = \lambda \quad Var(X_i) = \lambda$ ✓

By the CLT

$$\frac{S_n}{n} \dot\sim \text{Normal}\left(\lambda, \frac{\lambda}{n}\right)$$ ✓ (i.e. invoked the CLT)

$$\Rightarrow \quad S_n \dot\sim \text{Normal}(n\lambda, n\lambda)$$ ✓

(d) Suggest two estimators for λ using the plug-in principle. (2)

$$\frac{E(S_n)}{n} \quad \checkmark \qquad \frac{Var(S_n)}{n} \quad \checkmark$$

3. The joint distribution of X and Y is given by the pdf: $f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)}$, $x \in (-\infty, \infty)$, $y \in (-\infty, \infty)$.

(a) Are X and Y independent? Why/why not? (2)

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2} \frac{1}{\sqrt{2\pi}} e^{-y^2} \quad \checkmark \Rightarrow$$ independent because can right as product of two marginals ✓

(or similar)

If they calculated the marginals they would verify $f(x,y) \neq f(x)f(y)$ which also deserves credit

(b) Let $A = X + Y$, $B = X - Y$. Find $f(a, b)$ (6)

$$X = \frac{A+B}{2} \quad \checkmark \qquad Y = \frac{A-B}{2} \quad \checkmark \qquad \Rightarrow -\infty < a < \infty \quad \checkmark$$
$$-\infty < b < \infty$$

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} \checkmark = -\frac{1}{2} \quad \checkmark$$

$$f(a,b) = f(h_1(a,b), h_2(a,b))\, |J| \quad \checkmark$$

$$= \frac{1}{2\pi} e^{-\left(\left(\frac{a+b}{2}\right)^2 + \left(\frac{a-b}{2}\right)^2\right)} \cdot \frac{1}{2} \quad \checkmark$$

$$\frac{1}{4}\left(a^2 + \cancel{2ab} + b^2 + a^2 - \cancel{2ab} + b^2\right)$$
$$= \frac{1}{2}(a^2 + b^2)$$

$$= \frac{1}{4\pi} e^{-\frac{1}{2}(a^2+b^2)} \quad \checkmark$$

(c) Are A and B independent? Why/why not? (1)

$$f(a,b) = \frac{1}{2\sqrt{\pi}} e^{-a^2/2} \frac{1}{2\sqrt{\pi}} e^{-b^2/2} \Rightarrow \text{independent}$$

✓

(d) What is the distribution of A? (1)

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

$\Rightarrow$ $A \sim \text{Normal}(0, ?)$

✓ as long as they mention normal

Hadley the pdf of a standard normal is $e^{-x^2/2}$ so you screwed this question up!

Pledge:

Question	Points	Score
1	15	
2	10	
3	10	
Total:	35	