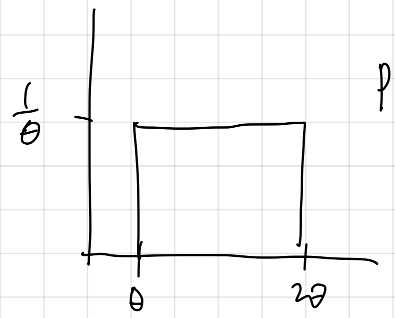
Problem 1

Let $X_1, \ldots, X_n$ be uniformly distributed over the open interval $(\theta, 2\theta)$ where $\theta > 0$ is the unknown parameter of interest.

(a) Show that a sufficient statistic for θ is $T(X) = (X_{(1)},\ X_{(n)})$. Recall that

$$X_{(1)} = \min_i X_i$$
$$X_{(n)} = \max_i X_i$$



pdf $f(x) = \frac{1}{\theta}$

$$f_{X_1^n}(x_1^n) = \prod_{i=1}^{n} \left(\frac{1}{2\theta-\theta}\right) \mathbb{1}\{\theta \le x_i \le 2\theta\} = \left(\frac{1}{\theta}\right)^n \mathbb{1}\{\theta \le x_i \le 2\theta;\ \forall i = 1, \ldots, n\}.$$

$$= \left(\frac{1}{\theta}\right)^n \mathbb{1}\{\theta \le \min_{1\le i\le n} x_i\}\, \mathbb{1}\{\max_{1\le i\le n} x_i \le 2\theta\}.$$

Let $h(x_1^n) = 1$ and $g(\theta, 2\theta)(x_1^n) = \left(\frac{1}{\theta}\right)^n \mathbb{1}\{\theta \le \min_{1\le i\le n} x_i\}\, \mathbb{1}\{\max x_i \le 2\theta\}$.

$\downarrow$ doesn't depend on θ.

$g(\theta, 2\theta)$ depends only on x_1^n through the function $T(X_1^n) = \left(\min_{1\le i\le n} X_i,\ \max_{1\le i\le n} X_i\right)$

so, $T(X) = (X_{(1)}, X_{(n)})$.

(b) Show that $\hat{\theta} = \frac{2}{3}X_1$ is an unbiased estimator for θ.

$$E(\hat{\theta}) = E\left(\tfrac{2}{3}X_1\right) = \tfrac{2}{3}E(X_1) = \tfrac{\cancel{2}}{3}\left(\tfrac{1}{\cancel{2}}(\theta + 2\theta)\right) = \theta$$

since $E(\hat{\theta}) = \theta$, $\hat{\theta}$ is an unbiased estimator of θ.

(c) Compute $\hat{\theta}_{MLE}$.
Hint: Think critically. No need for differentiation here.

$$\text{lik}(\theta) = \frac{1}{\theta^n} \mathbb{1}\left(\max_i x_i \le 2\theta,\ \min_i x_i \ge \theta\right)$$

$\text{lik}(\theta)$ is maximized when θ is minimized.

$\theta \le x_i \le 2\theta$ for all $i = 1, \ldots, n$.

order $X_1, \ldots, X_n$ in order:

$\theta < X_1 \le X_2 \le \cdots < 2\theta$

$\theta < X_{(1)}$ and $X_{(n)} < 2\theta \Rightarrow \frac{X_{(n)}}{2} < \theta < X_{(1)}$

In order to minimize θ, we use $\frac{X_{(n)}}{2}$.

therefore, $\hat{\theta}_{MLE} = \frac{\max_i x_i}{2}$

(d) Using the sufficient statistic $T(X)$ from part (a), find an unbiased estimator whose MSE is at least as good as $\hat{\theta}$.

$$\tilde{\theta} = E[\hat{\theta} | T] = \frac{2}{3} E(X_1 | \min_i X_i, \max_i X_i)$$

(Let $A = \min X_i$ and $B = \max X_i$

$$= \frac{2}{3} E(X_1 | A = \min_i X_i, B = \max_i X_i)$$

Law of Total probability is used:

$$E(X_1 | A = \min_i X_i, B = \max_i X_i)$$

$$= E(X_1 | A = \min_i X_i, B = \max_i X_i, X_1 = \overset{a}{A}) \cdot \overset{1/n}{P(X_1 = \min_i X_i)} +$$
$$E(X_1 | A = \min_i X_i, B = \max_i X_i, X_1 = \overset{b}{B}) \cdot \overset{1/n}{P(X_1 = \max_i X_i)} +$$
$$\overset{\frac{a+b}{2}}{E(X_1 | A = \min_i X_i, B = \max_i X_i, X_1 \neq A, X_1 \neq B)} \cdot \overset{1-\frac{2}{n}}{P(X_1 \neq \min_i X_i \cap X_1 \neq \max_i X_i)}$$

$$= \frac{a}{n} + \frac{b}{n} + \frac{a+b}{2}\left(1 - \frac{2}{n}\right) = \frac{a+b}{2}$$

$$\tilde{\theta} = \frac{\min_i X_i + \max_i X_i}{\cancel{2}} \cdot \frac{\cancel{2}}{3} = \boxed{\frac{\min_i X_i + \max_i X_i}{3}}$$

(e) Can you use $T(X)$ to improve $\hat{\theta}_{MLE}$ as well? why or why not?

$$E[\hat{\theta}_{MLE} | T] = E\left(\frac{\max_i X_i}{2} \,\Big|\, (\min_i X_i, \max_i X_i)\right) = \frac{\max_i X_i}{2}$$

Taking a conditional expectation doesn't affect the value of the estimator

(f) How do you think the mean-squared error of the resulting estimator from part (d) compares to the MSE of $\hat{\theta}_{MLE}$? Feel free to simulate the experiment in R to help you interpret/convince yourself of your intuition.

The MSE of $\hat{\theta}_{MLE}$ would be higher than the estimator from part (d) because while $\hat{\theta}_{MLE}$ only uses the maximum of all observation, d's answer takes minimum and maximum into calculation. Maximum is not sufficient because minimum needs to be also taken into account for sufficient statistics, as verified in (a). Hence, estimator from (d) is better than $\hat{\theta}_{MLE}$.

Problem 2

Let $X_1, \ldots, X_n$ be an i.i.d sample from a distribution with the density function

$$f(x;\theta) = \frac{\theta}{(1+x)^{(\theta+1)}} \qquad 0 < \theta < \infty \text{ and } 0 \le x < \infty$$

Find a sufficient statistic for θ.

$$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} \frac{\theta}{(1+x_i)^{\theta+1}} = \frac{\theta^n}{[\underbrace{\prod_{i=1}^{n}(1+x_i)}_{U}]^{\theta+1}} = \frac{\theta^n}{U^{\theta+1}}$$

$$\boxed{U = \prod_{i=1}^{n}(1+x_i)} \leftarrow$$ U is the sufficient statistics for θ.

$$g(U,\theta) = \frac{\theta^n}{U^{\theta+1}}$$

Problem 3

Let $X_1, X_2, X_3 \sim Bernoulli(p)$. Let T, U and S be statistics. Table below provides values of the statistics for each possible outcome of X_1, X_2, X_3. Given the information in the table below is T sufficient? what about U? What about S? Explain why or why not.

$$\frac{(1-p)^3}{p(t=0)}$$

$$\frac{1}{3} = \frac{p(1-p)^2}{p(1-p)^2 \cdot 3}$$

(x_1,x_2,x_3)	t	u	s	$f((x_1,x_2,x_3)\mid t)$	$f(x_1,x_2,x_3\mid U)$	$f(x_1,x_2,x_3\mid S)$
(0, 0, 0)	0	0	0	1	1	
(0, 0, 1)	1/3	1	1	1/3	1/3	
(0, 1, 0)	1/3	1	1	1/3	1/3	
(1, 0, 0)	1/3	1	65	1/3	1/3	
(0, 1, 1)	2/3	65	65	1/3	1/2	p/(p+1)
(1, 0, 1)	2/3	65	65	1/3	1/2	
(1, 1, 0)	2/3	201	201	1/3	1	
(1, 1, 1)	1	92	92	1	1	

For S:

$$\frac{P(X_1=0, X_2=1, X_3=1 \mid S=65)}{P(S=65)} = \frac{p^2(1-p)}{p(1-p)^2 + (1-p)p^2 + p^2(1-p)} = \frac{p^2(1-p)}{p(1-p)[(1-p)+p+p]} = \frac{\cancel{p}\,p^{\cancel{2}}\cancel{(1-p)}}{\cancel{p}(p+1)\cancel{(1-p)}} = \frac{p}{p+1}$$

→ since this is in a function of p, S is not a sufficient statistics.

For t: $P(X_1=0, X_2=0, X_3=0 \mid t=0) = \frac{(1-p)^3}{(1-p)^3} = 1$

$$P((X_1,X_2,X_3)=(0,0,1) \mid t=1/3) = P((X_1,X_2,X_3)=(0,1,0) \mid t=1/3) = P((X_1,X_2,X_3)=(1,0,0) \mid t=1/3)$$

$$= \frac{(1-p)^2 p}{(1-p)^2 p (3)} = \frac{1}{3}$$

$$P((X_1,X_2,X_3)=(0,1,1) \mid t=2/3) = P((X_1,X_2,X_3)=(1,0,1) \mid t=2/3) = P((X_1,X_2,X_3)=(1,1,0) \mid t=2/3)$$

$$= \frac{p^2(1-p)}{p^2(1-p)(3)} = \frac{1}{3}$$

$$P((X_1,X_2,X_3)=(1,1,1) \mid t=1) = \frac{p^3}{p^3} = 1$$

For u:

$$P((X_1,X_2,X_3)=(0,0,0) \mid U=0) = 1$$

$$P((X_1,X_2,X_3)=(0,0,1) \mid U=1) = P((X_1,X_2,X_3)=(0,1,0) \mid U=1) = P((X_1,X_2,X_3)=(1,0,0) \mid U=1)$$

$$= \frac{p(1-p)^2}{p(1-p)^2 \cdot 3} = 1/3$$

$P((X_1,X_2,X_3)=(0,1,1)\mid U=65) = P((X_1,X_2,X_3)=(1,0,1)\mid U=65)$

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

$$P((X_1,X_2,X_3)=(1,1,0)\mid U=201) = \frac{p^2(1-p)}{p^2(1-p)} = 1$$

$$P((X_1,X_2,X_3)=(1,1,1)\mid U=92) = \frac{p^3}{p^3} = 1$$

u and t are sufficient statistics because none of the computed values are in terms of p. S is not a sufficient statistics.

Problem 4

Suppose $X_1, \ldots, X_n$ when $n > 2$ are i.i.d $Bernoulli(p)$ where $0 < p < 1$ is unknown.

(a) Find a sufficient statistic $T(X_1, \ldots, X_n)$ for p.

(b) Show that $\mathbb{1}\{X_1 = 1, X_2 = 0\}$ is an unbiased estimator of $p(1-p)$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function.

(c) Use the Rao-Blackwell theorem to improve the above estimator.

(a) $$f(x|\theta) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i} = \left(\frac{p}{1-p}\right)^{\sum x_i}\cdot(1-p)^n$$

$f(x|\theta)$ depends only on $x_1, \ldots, x_n$ through the sufficient statistics $\sum_{i=1}^n x_i$ and

$f(x|\theta)$ is the form $g(\sum x_i, p)\,h(x)$ where $h(x) = 1$ and $g(t,\theta) = \left(\frac{p}{1-p}\right)^t(1-p)^n$.

(b) $$\mathbb{1}\{X_1=1, X_2=0\} = \begin{cases} 1 & \text{if } X_1 = 1 \text{ and } X_2 = 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$P((X_1=1, X_2=0)=1) = P(X_1=1)\,P(X_2=0) = p^1(1-p)^{1-1}\cdot p^0\cdot(1-p)^{1-0} = p(1-p)$$

by independence

$$E[\mathbb{1}\{X_1=1, X_2=0\}] = 1\cdot P((X_1=1,X_2=0)=1) + 0\cdot P((X_1=1,X_2=0)=0) = p(1-p)$$

since $E[\mathbb{1}\{X_1=1, X_2=0\}] = (1-p)p$, it is an unbiased estimator for $p(1-p)$.

(c) $\tilde{\theta} = E(\hat{\theta}\mid T)$ where $MSE(\tilde{\theta}) \le MSE(\hat{\theta})$

$$\tilde{\theta} = E[\mathbb{1}_{x_i=0} \mid \sum x_i = t] = 1\cdot P((X_1=0, X_2=1)\mid \sum x_i = t) = \frac{P((X_1=0, X_2=1)\cap \sum_{i=1}^n x_i = t)}{P(\sum_{i=1}^n x_i = t)}$$

$$= \frac{P(X_1=0)\,P(X_2=1)\,P(\sum_{i=3}^n x_i = t-1)}{P(\sum_{i=1}^n x_i = t)}$$

$\sum_{i=3}^n x_i \sim$ binomial $(n-2, p)$

$\sum_{i=1}^n x_i \sim$ binomial (n, p)

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

$$= \binom{n-2}{t-1}\Big/\binom{n}{t}$$

$$= \frac{(n-2)!}{(t-1)!\,(n-t-1)!} \times \frac{(n-t)!\,(t)!}{n!} = \frac{t(n-t)}{n(n-1)} = \frac{\sum_{i=1}^n x_i\left(n - \sum_{i=1}^n x_i\right)}{n(n-1)}$$

since $t = \sum_{i=1}^n x_i$

$$= \frac{n\bar{x}(n - n\bar{x})}{n(n-1)} = \frac{n^2\bar{x}(1-\bar{x})}{n(n-1)} = \boxed{\frac{n\bar{x}(1-\bar{x})}{n-1}}$$ improved estimator.

Problem 5

Consider the coin from the lecture with probability of heads either $p = 0.5$ or $p = 0.7$. Suppose that instead of tossing the coin 10 times, the coin was tossed until a head came up and the total number of tosses X was recorded.

(a) Which outcomes favor $H_0 : p = 0.5$ over $H_1 : p = 0.7$? Do this only based on the probabilities of each outcome.

(b) Which outcomes favor H_1 over H_0?

(c) For your decision rule, what is the probability of falsely rejecting H_0? This is called the Type I error. What is the significance level of your test?

(d) For your decision rule, what is the probability of not rejecting H_0 when it is actually false? This is called the Type II error.

(a)

X	1	2	3	4
p=0.5	0.5	0.25	0.125	...
p=0.7	0.7	0.21	0.063	...

$\boxed{X \geq 2}$

p=0.5's probability are better than p=0.7's probability for $x \geq 2$.

(b) $\boxed{X=1}$ is the only case when coin is more likely to be fair than biased towards head.

(c) Type I:

X	1	2	3	4
p=0.5	0.5	0.25	0.125	...
p=0.7	0.7	0.21	0.063	...

α = significance level = 0.5

p(type I error) = $p(H_0 \mid H_1)$ = 0.5

d) p(Type II error) = 1 − 0.7 = 0.3

When H_0 is false is when X=1. $p(X=1 \mid p(\text{head})=0.7)$ = 0.7

Hence, p(failing to reject the null) = 1 − 0.7 = 0.3 ✓

(a) Suppose the university will save Evans if you can conclude, at a significance level of $\alpha = 0.05$, that the true proportion of undergrads that do not want Evans demolished is higher than 0.8. So, you randomly ask $n = 100$ undergrads whether they want to keep dear Evans and conduct the following hypothesis test:

$$H_0 : p = 0.8, \quad H_1 : p > 0.8$$

Calculate the rejection region, or the range of $\hat{p}$ for which we would reject the null hypothesis and conclude that more than 80% of the undergraduate population wants to save Evans. $\hat{p}$ is the proportion of students who want to keep Evans in your sample. Use the normal approximation when necessary.

$$\frac{\hat{p} - 0.8}{\sqrt{(0.8)(0.2)/100}} \geq 1.65$$

$$\hat{p} \geq (1.65)(\sqrt{0.0016}) + 0.8 = \boxed{0.866}$$

(b) Now Suppose the university believes exactly 80% of students want to keep Evans. So using the same random sample of $n = 100$ undergrads you conduct the following hypothesis test:

$$H_0 : p = 0.8, \quad H_1 : p \neq 0.8$$

Calculate the new rejection region with the same significance level $\alpha = 0.05$? Use the normal approximation.

$$P(|z| \geq 1.96)$$

$$|z| = \left| \frac{\hat{p} - 0.8}{\sqrt{0.0016}} \right| \geq 1.96$$

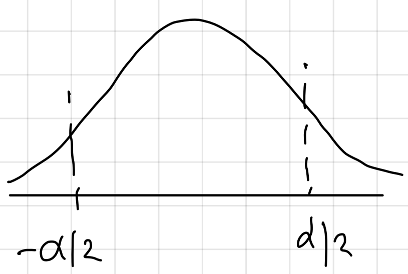
$$|\hat{p} - 0.8| \geq 1.96(\sqrt{0.0016})$$

$$\hat{p} \geq 0.8 - 0.0784 = \boxed{0.7216}$$

$$\hat{p} \leq 0.0784 + 0.8 = \boxed{0.8784}$$

$$0.7216 \leq \hat{p} \leq 0.8784$$

(c) Construct the $(1 - \alpha)$ confidence interval using the same 100 students and assume normal approximation when necessary. How is this confidence interval (approximately) related to your hypothesis test in part (b)?



$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \rightarrow \frac{\sigma}{\sqrt{n}} z = \bar{x} - \mu_0$$

$$\sigma/\sqrt{n} = 0.08$$

$$\left(\hat{p} - 0.08\, z(\alpha/2),\ \hat{p} + 0.08\, z(\alpha/2)\right)$$

The interval with confidence level $1-\alpha$ is the acceptance region of a z-test with significance level α.