ECO227Y5 Tutorial 15

William Hsu

Department of Economics
University of Toronto Mississauga

February 2nd, 2026

Question 7.21

Question: Refer to Exercise 7.13. Suppose that $n = 20$ observations are to be taken on $\ln(\mathrm{LC50})$ measurements and that $\sigma^2 = 1.4$. Let S^2 denote the sample variance of the 20 measurements.

Recall: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ where $n-1$ is the degrees of freedom

(a) Find a number b such that $P(S^2 \leq b) = 0.975$.

(b) Find a number a such that $P(a \leq S^2) = 0.975$.

(c) If a and b are as in parts (a) and (b), what is $P(a \leq S^2 \leq b)$?

$n = 20 \Rightarrow d.f = 19$, $\sigma^2 = 1.4$

a.) $P(S^2 \le b) = 0.975 \Rightarrow P\left(\frac{19 \cdot S^2}{1.4} \le \frac{19 \cdot b}{1.4}\right) = 0.975 \Rightarrow P\left(\chi^2_{19} \le \frac{19 \cdot b}{1.4}\right) = 0.975$

$\Rightarrow P\left(\chi^2_{19} > \frac{19 \cdot b}{1.4}\right) = \overset{1-0.975}{0.025} \Rightarrow \frac{19 \cdot b}{1.4} = 32.8523 \Rightarrow b = 2.42$

b.) $P(a \le S^2) = P(S^2 \ge a) = P\left(\chi^2_{19} \ge \frac{19 \cdot a}{1.4}\right) = 0.975$

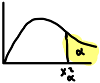
$\frac{19 \cdot a}{1.4} = 8.90655 \Rightarrow a = 0.656272$

c.) $P(a \le S^2 \le b) = P(S^2 \ge a) - P(S^2 \ge b)$

$= 0.975 - (1 - 0.975)$

$= 0.975 - 0.025 = 0.95$

How To Read χ^2 table: Right tailed table



χ^2_α is the point on your chi-square where α % of data lies to the right

df	$\chi^2_{0.995}$	...	$\chi^2_{0.950}$
1			
2			⋮
⋮			
8	. . .	. .	2.73264
⋮			

Ex. $P(\chi^2_8 \ge 2.733) \approx 0.950$

Question 7.26

Question: Refer to Exercise 7.11. Suppose that in the forest fertilization problem the population standard deviation of basal areas is not known and must be estimated from the sample. If a random sample of $n = 9$ basal areas is to be measured, find two statistics g_1 and g_2 such that

Recall: $\frac{\bar{y}-\mu}{s/\sqrt{n}} \sim t_{n-1}$, using s as σ not known

$$P\left[g_1 \le (\bar{Y} - \mu) \le g_2\right] = 0.90.$$

$n=9$, σ^2 not known, not dealing with estimating variance $\Rightarrow$ Student's T

d.f $= 9-1 = 8$. $\quad P\left[\frac{g_1}{s/3} \le t_8 \le \frac{g_2}{s/3}\right] = 0.90$

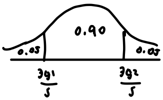
as $P(t_8 \ge 1.860) = 0.05$

From table: $P(-1.860 \le t_8 \le 1.860) = 0.90$

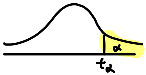
$-1.860 = \frac{3g_1}{s} \Rightarrow g_1 = -0.62\,s$

$1.860 = \frac{3g_2}{s} \Rightarrow g_2 = 0.62\,s$

note: T distributions are symmetric.



How To Read Student's T Table:



$P(t_7 \ge 1.895) = 0.05$

$t_{.100}$	$t_{.050}$	$\cdots$	$t_{.005}$	df
..	.		.	1
				2
				⋮
	1.895			7
				⋮
				∞

Question 7.29

Question: If Y is a random variable that has an F distribution with ν_1 numerator and ν_2 denominator degrees of freedom, show that $U = 1/Y$ has an F distribution with ν_2 numerator and ν_1 denominator degrees of freedom.

$Y \sim F_{v_1, v_2}$.

$U = \frac{1}{Y}$

Proof: Recall: If $W_1 \sim \chi^2_{v_1}$, $W_2 \sim \chi^2_{v_2}$. Then $X = \dfrac{\chi^2_{v_1}/v_1}{\chi^2_{v_2}/v_2} \sim F_{v_1, v_2}$

so $Y = \dfrac{\chi^2_{v_1}/v_1}{\chi^2_{v_2}/v_2} \Rightarrow \dfrac{1}{Y} = \dfrac{\chi^2_{v_2}/v_2}{\chi^2_{v_1}/v_1} = U \sim F_{v_2, v_1}$ ∎

Question 7.36

Question: Let S_1^2 denote the sample variance for a random sample of ten $\ln(\mathrm{LC50})$ values for copper and let S_2^2 denote the sample variance for a random sample of eight $\ln(\mathrm{LC50})$ values for lead, both samples using the same species of fish. The population variance for measurements on copper is assumed to be twice the corresponding population variance for measurements on lead. Assume S_1^2 to be independent of S_2^2.

(a) Find a number b such that

$$P\left(\frac{S_1^2}{S_2^2} \le b\right) = 0.95.$$

Recall: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

copper S_1^2, $n_1 = 10 \Rightarrow d.f._1 = 9$, σ_1^2 $\quad \sigma_1^2 = 2\sigma_2^2$

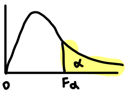
Lead S_2^2, $n_2 = 8 \Rightarrow d.f._2 = 7$, σ_2^2

$$P\left(\frac{S_1^2}{S_2^2} \le b\right) = 0.95 \Rightarrow P\left(\frac{\overbrace{\frac{(n_1-1)S_1^2}{\sigma_1^2}}^{\chi^2_{n_1-1}}/n_1-1}{\underbrace{\frac{(n_2-1)S_2^2}{\sigma_2^2}}_{\chi^2_{n_2-1}}/n_2-1} \le \frac{\overbrace{\frac{(n_1-1)}{\sigma_1^2}/n_1-1}^{=\frac{1}{\sigma_1^2}}}{\underbrace{\frac{(n_2-1)}{\sigma_2^2}/n_2-1}_{=\frac{1}{\sigma_2^2}}}\, b\right) = 0.95$$

$$\Rightarrow P\left(\frac{\chi^2_{n_1-1}/n_1-1}{\chi^2_{n_2-1}/n_2-1} \le \frac{\sigma_2^2}{\sigma_1^2} b\right) = 0.95 \Rightarrow P\left(F_{n_1-1,n_2-1} \le \frac{\sigma_2^2}{2\sigma_2^2} b\right) = 0.95$$

$$\Rightarrow P\left(F_{9,7} \le \frac{b}{2}\right) = 0.95 \Rightarrow P\left(F_{9,7} > \frac{b}{2}\right) = 0.05 \Rightarrow \frac{b}{2} = 3.68 \Rightarrow b = 7.36$$

How To Read F table:



numerator d.f

Denom df	α	1	2	...	6	7
1	0.050					
2	0.050					
⋮	⋮					
6	0.050		5.14			
7	.					

Read α beside the specific denominator d.f

$P(F_{2,6} \ge 5.14) = 0.050$

Question: Let S_1^2 denote the sample variance for a random sample of ten $\ln(\mathrm{LC50})$ values for copper and let S_2^2 denote the sample variance for a random sample of eight $\ln(\mathrm{LC50})$ values for lead, both samples using the same species of fish. The population variance for measurements on copper is assumed to be twice the corresponding population variance for measurements on lead. Assume S_1^2 to be independent of S_2^2 HINT: Use the result of Exercise 7.29 and notice that $P(U_1/U_2 \leq k) = P(U_2/U_1 \geq 1/k)$.

(b) Find a number a such that

$$P\left(a \leq \frac{S_1^2}{S_2^2}\right) = 0.95.$$

copper s_1^2, $n_1 = 10 \Rightarrow d.f._1 = 9$, σ_1^2 $\quad \sigma_1^2 = 2\sigma_2^2$

Lead s_2^2, $n_2 = 8 \Rightarrow d.f._2 = 7$, σ_2^2

- table's a's are small

$$P\left(a \leq \frac{s_2^2}{s_2^2}\right) \Rightarrow P\left(\frac{s_1^2}{s_2^2} \geq a\right) = 0.95 \Leftrightarrow P\left(\frac{s_2^2}{s_1^2} \leq \frac{1}{a}\right) = 0.95 \Rightarrow P\left(\frac{s_2^2}{s_1^2} > \frac{1}{a}\right) = 0.05$$

$$\frac{\sigma_1^2}{\sigma_2^2} = \frac{2\sigma_2^2}{\sigma_2^2} = 2$$

$$\Rightarrow P\left(\frac{\frac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2-1)}{\frac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)} > \frac{\frac{(n_2-1)}{\sigma_2^2}/(n_2-1)}{\frac{(n_1-1)}{\sigma_1^2}/(n_1-1)}\frac{1}{a}\right) = 0.05 \Rightarrow P\left(F_{7,9} > \frac{2}{a}\right) = 0.05$$

$$\Rightarrow \frac{2}{a} = 3.29 \Rightarrow a = 0.607$$

Question 7.36

Question: Let S_1^2 denote the sample variance for a random sample of ten $\ln(\mathrm{LC50})$ values for copper and let S_2^2 denote the sample variance for a random sample of eight $\ln(\mathrm{LC50})$ values for lead, both samples using the same species of fish. The population variance for measurements on copper is assumed to be twice the corresponding population variance for measurements on lead. Assume S_1^2 to be independent of S_2^2.

(c) If a and b are as in parts (a) and (b), find

$$P\left(a \leq \frac{S_1^2}{S_2^2} \leq b\right).$$

$P(a \le \frac{S_1^2}{S_2^2} \le b) = P(\frac{S_1^2}{S_2^2} \ge a) - P(\frac{S_1^2}{S_2^2} \ge b) = 0.95 - 0.05 = 0.90$

Question 7.43

Question: An anthropologist wishes to estimate the average height of men for a certain race of people. If the population standard deviation is assumed to be 2.5 inches and if she randomly samples 100 men, find the probability that the difference between the sample mean and the true population mean will not exceed 0.5 inch.

$n = 100$

$\sigma^2 = 2.5^2$ known

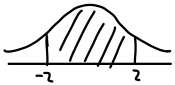
can use Z

Recall: $\dfrac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim Z$

$P(|\bar{y}-\mu| \le 0.5)$

$\Rightarrow P(-0.5 \le \bar{y}-\mu \le 0.5)$

$\Rightarrow P\left(\dfrac{-0.5}{\frac{2.5}{10}} \le Z \le \dfrac{0.5}{\frac{2.5}{10}}\right) = P(-2 \le Z \le 2) = 0.9544$



Question 7.45

Question: Workers employed in a large service industry have an average wage of \$7.00 per hour with a standard deviation of \$0.50. The industry has 64 workers of a certain ethnic group. These workers have an average wage of \$6.90 per hour. Is it reasonable to assume that the wage rate of the ethnic group is equivalent to that of a random sample of workers from those employed in the service industry? **[Hint:** Calculate the probability of obtaining a sample mean less than or equal to \$6.90 per hour.**]**

- use z

$\sigma^2 = 0.5^2$, $\mu = 7.00$

$n = 64$

$P(\bar{y} \le 6.9) = P\left(\frac{\bar{y} - 7.00}{\frac{0.5}{8}} \le \frac{6.9 - 7.00}{\frac{0.5}{8}} \right) \Rightarrow P(z \le -1.6) = P(z > 1.6) = 0.0548$

Question: An important aspect of a federal economic plan was that consumers would save a substantial portion of the money that they received from an income tax reduction. Suppose that early estimates of the portion of total tax saved, based on a random sampling of 35 economists, had mean 26% and standard deviation 12%.

(a) What is the approximate probability that a sample mean estimate, based on a random sample of $n = 35$ economists, will lie within 1% of the mean of the population of the estimates of all economists?

(b) Is it necessarily true that the mean of the population of estimates of all economists is equal to the percent tax saving that will actually be achieved?

a.) $n = 35$, $s^2 = 12^2$

$$P(|\bar{Y} - \mu| \le 1) \Rightarrow P\left(\frac{-1}{\frac{12}{\sqrt{35}}} \le \frac{\bar{y} - \mu}{\frac{12}{\sqrt{35}}} \le \frac{1}{\frac{12}{\sqrt{35}}} \right) \Rightarrow P(-0.49 \le z \le 0.49) = 1 - 2P(z > 0.49)$$

$$= 1 - 2 \cdot 0.3121$$

$$= 0.3758$$

since $n = 35 > 30$, we can use CLT, therefore use z instead of student's T

b.) no, mean estimates can very much differ from actualized tax saving.

Question 7.60

Question: The result in Exercise 7.58 holds even if the sample sizes differ. That is, if $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ constitute independent random samples from populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, then $\bar{X} - \bar{Y}$ will be approximately normally distributed, for large n_1 and n_2, with mean $\mu_1 - \mu_2$ and variance $(\sigma_1^2/n_1) + (\sigma_2^2/n_2)$.

The flow of water through soil depends on, among other things, the porosity (volume proportion of voids) of the soil. To compare two types of sandy soil, $n_1 = 50$ measurements are to be taken on the porosity of soil A and $n_2 = 100$ measurements are to be taken on soil B. Assume that $\sigma_1^2 = 0.01$ and $\sigma_2^2 = 0.02$. Find the probability that the difference between the sample means will be within 0.05 unit of the difference between the population means $\mu_1 - \mu_2$.

we know σ_1^2, σ_2^2 so use Z

$n_1 = 50$
$n_2 = 100$

$$P(|(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)| \le 0.05) = P(-0.05 \le (\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2) \le 0.05)$$

$$\sigma^2 = \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \Rightarrow P\left(\frac{-0.05}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \le Z \le \frac{0.05}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) = P\left(\frac{-0.05}{\frac{1}{50}} \le Z \le \frac{0.05}{\frac{1}{50}}\right) = P(-2.5 \le Z \le 2.5)$$

$$= 1 - 2P(Z > 2.5) = 1 - 0.0062$$
$$= 0.9876$$

sample size already accounted for in variance!

Question 8.6

Question: Suppose that $\underline{E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta}$, $V(\hat{\theta}_1) = \sigma_1^2$, and $V(\hat{\theta}_2) = \sigma_2^2$. Consider the estimator

$$\hat{\theta}_3 = a\hat{\theta}_1 + (1-a)\hat{\theta}_2.$$

(a) Show that $\hat{\theta}_3$ is an unbiased estimator for θ.

(b) If $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent, how should the constant a be chosen in order to minimize the variance of $\hat{\theta}_3$?

Definition: $\hat{\theta}$ is an unbiased estimator for θ if $E(\hat{\theta}) = \theta$

a.) $E(\hat{\theta}_3) = E(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) = aE(\hat{\theta}_1) + (1-a)E(\hat{\theta}_2) = a\theta + (1-a)\theta = \theta$

cov=0

$(x-y)^2 = x^2 - 2xy + y^2$

b.) $V(\hat{\theta}_3) = V(\underset{x}{a\hat{\theta}_1} + \underset{y}{(1-a)\hat{\theta}_2}) = a^2V(\hat{\theta}_1) + (1-a)^2V(\hat{\theta}_2) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2$

$\frac{dV(\hat{\theta}_3)}{da} = 0 \Rightarrow 2a\sigma_1^2 - 2(1-a)\sigma_2^2 = 0 \Rightarrow 2a\sigma_1^2 - 2\sigma_2^2 + 2a\sigma_2^2 = 0 \Rightarrow a = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$

UNIVERSITY OF TORONTO

Question 8.8

Question: Suppose that Y_1, Y_2, Y_3 denote a random sample from an exponential distribution with density function

$$f(y) = \begin{cases} \frac{1}{\theta} e^{-y/\theta}, & y > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

$F(y) = \int_0^y \frac{1}{\theta} e^{-y/\theta} dy \Rightarrow \frac{1}{\theta} \int_0^y e^{-y/\theta} dy \Rightarrow \frac{1}{\theta} \left[-\theta e^{-y/\theta} \right]_0^y \Rightarrow 1 - e^{-y/\theta}$

$F_{Y_{(1)}}(y) = P(Y_{(1)} \le y)$
$= 1 - P(Y_{(1)} \ge y)$
$= 1 - P(Y_i \ge y)^3$
$= 1 - (1 - P(Y_i \le y))^3$
$= 1 - (1 - F(y))^3$
$= 1 - [1 - (1 - e^{-y/\theta})]^3$
$= 1 - (e^{-3y/\theta})$
$\Rightarrow f_{Y_{(1)}}(y) = \frac{3}{\theta} e^{-3y/\theta}$

$\text{Exp}(\frac{\theta}{3})$

Consider the following five estimators of θ:

- order statistic

$$\hat{\theta}_1 = Y_1, \quad \hat{\theta}_2 = \frac{Y_1 + Y_2}{2}, \quad \hat{\theta}_3 = \frac{Y_1 + 2Y_2}{3}, \quad \hat{\theta}_4 = \min(Y_1, Y_2, Y_3), \quad \hat{\theta}_5 = \bar{Y}.$$

(a) Which of these estimators are unbiased?

$E(\hat{\theta}_1) = E(Y_1) = \theta$ unbiased

$E(\hat{\theta}_2) = E\left(\frac{Y_1 + Y_2}{2}\right) = \frac{1}{2}(E(Y_1) + E(Y_2)) = \frac{1}{2} \cdot 2\theta = \theta$ unbiased

$E(\hat{\theta}_3) = E\left(\frac{Y_1 + 2Y_2}{3}\right) = \frac{1}{3}(E(Y_1) + 2E(Y_2)) = \frac{1}{3}(\theta + 2\theta) = \theta$ unbiased

$E(\hat{\theta}_4) = E(Y_{(1)}) = \frac{\theta}{3} \neq \theta$ biased

$E(\hat{\theta}_5) = E(\bar{Y}) = E\left(\frac{Y_1 + Y_2 + Y_3}{3}\right) = \frac{1}{3}(3\theta) = \theta$

unbiased

Question 8.8

Question: Suppose that Y_1, Y_2, Y_3 denote a random sample from an exponential distribution with density function

$$f(y) = \begin{cases} \frac{1}{\theta} e^{-y/\theta}, & y > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Consider the following five estimators of θ:

$$\hat{\theta}_1 = Y_1, \quad \hat{\theta}_2 = \frac{Y_1 + Y_2}{2}, \quad \hat{\theta}_3 = \frac{Y_1 + 2Y_2}{3}, \quad \hat{\theta}_4 = \min(Y_1, Y_2, Y_3), \quad \hat{\theta}_5 = \bar{Y}.$$

(b) Among the unbiased estimators, which has the smallest variance?

$V(\hat{\theta}1) = \theta^2$ - no cov

$V(\hat{\theta}2) = \frac{1}{4}(V(Y_1) + V(Y_2)) = \frac{\theta^2}{2}$

$V(\hat{\theta}3) = \frac{1}{9}(V(Y_1) + 4V(Y_2)) = \frac{5\theta^2}{9}$

$V(\hat{\theta}5) = \frac{1}{9}(V(Y_1) + V(Y_2) + V(Y_3)) = \frac{\theta^2}{3}$

clearly $V(\hat{\theta}5)$ is smallest

Question 8.15

Question: Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a population whose density is given by

Recall: $MSE(\hat{\theta}) = V(\hat{\theta}) + (Bias(\hat{\theta}))^2$

$$f(y) = \begin{cases} 3\beta^3 y^{-4}, & \beta \leq y, \\ 0, & \text{elsewhere,} \end{cases}$$

$F(y) = \int_\beta^y 3\beta^3 y^{-4} dy$

$= 3\beta^3 \left[\frac{y^{-3}}{-3}\right]_\beta^y = -\beta^3 y^{-3} + 1$

where $\beta > 0$ is unknown. (This is one of the Pareto distributions introduced in Exercise 6.18.) Consider the estimator $\hat{\beta} = \min(Y_1, Y_2, \ldots, Y_n)$.

(a) Derive the bias of the estimator $\hat{\beta}$.

(b) Derive $\mathrm{MSE}(\hat{\beta})$.

a.) $\hat{\beta} = Y_{(1)}$. $F_{Y_{(1)}}(y) = P(Y_{(1)} \leq y) = 1 - P(Y_{(1)} \geq y) = 1 - P(Y_i \geq y)^n = 1 - (1 - P(Y_i \leq y))^n = 1 - (1 - F_{Y_i}(y))^n$

$\Rightarrow 1 - (1 - (-\beta^3 y^{-3} + 1))^n = 1 - (\beta^3 y^{-3})^n = 1 - \beta^{3n} y^{-3n} = 1 - \left(\frac{\beta}{y}\right)^{3n} \Rightarrow f_{Y_{(1)}}(y) = 3n \frac{\beta^{3n}}{y^{3n+1}}$

$E(\hat{\beta}) = \int_\beta^\infty y \cdot \frac{3n\beta^{3n}}{y^{3n+1}} dy = 3n\beta^{3n} \int_\beta^\infty \frac{1}{y^{3n}} dy = 3n\beta^{3n} \left[\frac{1}{-(3n-1) y^{3n-1}}\right]_\beta^\infty = 3n\beta^{3n} \left(\frac{1}{(3n-1)\beta^{3n-1}}\right) = \frac{3n}{3n-1}\beta$

Bias $= E(\hat{\beta}) - \beta = \beta\left(\frac{3n}{3n-1} - 1\right) = \frac{1}{3n-1}\beta$

$\Rightarrow MSE = \frac{3n}{3n-2}\beta^2 - \frac{9n^2}{(3n-1)^2}\beta^2 + \frac{1}{(3n-1)^2}\beta^2$

b.) $V(\hat{\beta}) = E(\hat{\beta}^2) - E(\hat{\beta})^2 = \int_\beta^\infty y^2 \, 3n \frac{\beta^{3n}}{y^{3n+1}} dy - \left(\frac{3n}{3n-1}\beta\right)^2 = \frac{3n}{3n-2}\beta^2 - \frac{9n^2}{(3n-1)^2}\beta^2$

Question 8.17

Question: If Y has a binomial distribution with parameters n and p, then $\hat{p}_1 = Y/n$ is an unbiased estimator of p. Another estimator of p is

$Y \sim Bin(n, p)$

$$\hat{p}_2 = \frac{Y+1}{n+2}.$$

(a) Derive the bias of $\hat{p}_2$.

(b) Derive $\mathrm{MSE}(\hat{p}_1)$ and $\mathrm{MSE}(\hat{p}_2)$.

(c) For what values of p is $\mathrm{MSE}(\hat{p}_1) < \mathrm{MSE}(\hat{p}_2)$?

a.) $E(\hat{p}_2) = E\left(\frac{Y+1}{n+2}\right) = \frac{1}{n+2}[E(Y)+1] = \frac{1}{n+2}(np+1) = \frac{np+1}{n+2}$

$\text{Bias} = E(\hat{p}_2) - p = \frac{np+1}{n+2} - p = \frac{np+1-np-2p}{n+2} = \frac{1-2p}{n+2}$

b.) $MSE(\hat{p}_1) = V(\hat{p}_1) + (Bias(\hat{p}_1))^2 = V(\hat{p}_1) = V\left(\frac{Y}{n}\right) = \frac{1}{n^2}V(Y) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$

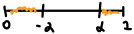
$MSE(\hat{p}_2) = V(\hat{p}_2) + (Bias(\hat{p}_2))^2 = V\left(\frac{Y+1}{n+2}\right) + \left(\frac{1-2p}{n+2}\right)^2 = \frac{np(1-p)}{(n+2)^2} + \left(\frac{1-2p}{n+2}\right)^2 = \frac{np(1-p)+(1-2p)^2}{(n+2)^2}$

c.) $\frac{p(1-p)}{n} < \frac{np(1-p)+(1-2p)^2}{(n+2)^2} \Rightarrow (n+2)^2 p(1-p) < n^2 p(1-p) + n(1-2p)^2$

$\Rightarrow (n^2+4n+4)(p-p^2) < n^2p - n^2p^2 + n(1-4p+4p^2) \Rightarrow n^2p + 4np + 4p - n^2p^2 - 4np^2 - 4p^2 < n^2p - n^2p^2 + n - 4np + 4np^2$

$\Rightarrow 8np - 8np^2 - 4p^2 + 4p - n < 0 \Rightarrow -p^2(8n+4) + p(8n+4) - n < 0 \Rightarrow p^2(8n+4) - p(8n+4) + n > 0$

$p = \frac{(8n+4) \pm \sqrt{(8n+4)^2 - 4(8n+4)n}}{2(8n+4)} = \frac{1}{2} \pm \sqrt{\frac{n+1}{8n+4}}$ call $\alpha, -\alpha$



so $p \in \left(0, \frac{1}{2} - \sqrt{\frac{n+1}{8n+4}}\right) \cup \left[\frac{1}{2} + \sqrt{\frac{n+1}{8n+4}}, 1\right]$

UNIVERSITY OF TORONTO