

Name: _____

Show all your work. No credit for lucky answers.

1. Given a random sample of size
- n
- from the density

$$f(y) = \begin{cases} e^{-(y-\theta)} & \text{if } \theta < y, \\ 0 & \text{else.} \end{cases}$$

where $\theta \in \mathbb{R}$ is unknown parameter. Find a constant a and b such that $\hat{\theta}_1 = a + \bar{Y}$ and $\hat{\theta}_2 = b + Y_{(1)}$ are both unbiased estimators of θ . Find the mean square error of $\hat{\theta}_1$ and $\hat{\theta}_2$? Which of the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ has smallest mean square error?

$$E[\bar{Y}] = E[Y] = \int_{\theta}^{\infty} y e^{-(y-\theta)} dy = \int_0^{\infty} (\theta + u) e^{-u} du = \theta + 1$$

$y - \theta = u$
 $dy = du$

$$\text{If } \theta = E[\hat{\theta}_1] = E[a + \bar{Y}] = a + \theta + 1, \quad \boxed{a = -1}$$

The cdf of Y is $F(y) = \int_{\theta}^y e^{-(y-\theta)} dy = 1 - e^{-(y-\theta)}$ if $y > \theta$

So, the cdf of $Y_{(1)}$ is

$$f_{Y_{(1)}}(y) = n(1 - F(y))^{n-1} f(y) = n(e^{-(y-\theta)})^{n-1} e^{-(y-\theta)} = n e^{-n(y-\theta)}, \text{ if } y > \theta.$$

$$E[Y_{(1)}] = \int_{\theta}^{\infty} y n e^{-n(y-\theta)} dy = \int_0^{\infty} (\theta + \frac{u}{n}) e^{-u} du = \theta + \frac{1}{n}$$

$n(y-\theta) = u, \quad y = \theta + \frac{u}{n}, \quad dy = \frac{du}{n}$

$$\text{If } \theta = E[\hat{\theta}_2] = E[b + Y_{(1)}] = b + \theta + \frac{1}{n}, \quad \boxed{b = -\frac{1}{n}}$$

$$E[Y^2] = \int_{\theta}^{\infty} y^2 e^{-(y-\theta)} dy = \int_0^{\infty} (\theta + u)^2 e^{-u} du = \theta^2 + 2\theta + 2$$

$y - \theta = u, \quad dy = du$

$$Var(Y) = E[Y^2] - (E[Y])^2 = \theta^2 + 2\theta + 2 - (\theta + 1)^2 = 1$$

$$MSE(\hat{\theta}_1) = Var(\hat{\theta}_1) = Var(a + \bar{y}) = Var(\bar{y}) = \frac{Var(y)}{n} = \frac{1}{n}$$

$$\boxed{MSE(\hat{\theta}_1) = \frac{1}{n}}$$

$$E[Y_{(1)}^2] = \int_{\theta}^{\infty} \frac{2}{n} e^{-n(y-\theta)} dy = \int_0^{\infty} \left(\theta + \frac{u}{n}\right)^2 e^{-u} du = \theta^2 + \frac{2\theta}{n} + \frac{2}{n^2}$$

$n(y-\theta) = u \quad dy = \frac{du}{n}$
 $y = \theta + \frac{u}{n}$

$$Var(Y_{(1)}) = E[Y_{(1)}^2] - (E[Y_{(1)}])^2 = \theta^2 + \frac{2\theta}{n} + \frac{2}{n^2} - \left(\theta + \frac{1}{n}\right)^2 = \frac{1}{n^2}$$

$$MSE(\hat{\theta}_2) = MSE\left(Y_{(1)} - \frac{1}{n}\right) = Var(Y_{(1)}) = \frac{1}{n^2}$$

$$\boxed{MSE(\hat{\theta}_2) = \frac{1}{n^2}}$$

$\hat{\theta}_2$ has a smallest mean square error, because $\frac{1}{n^2} \leq \frac{1}{n}$, for $n \geq 1$

(and $\frac{1}{n^2} < \frac{1}{n}$, for $n \geq 2$)

2. Suppose a sample of 16 wild adult female mosquitoes, *Culex Tarsalis*, was captured. Wing lengths were measured and the sample mean was 4.50 mm and the sample standard deviation was 0.28 mm. Find a 99% confidence interval for the mean wing length of adult female mosquitoes.

$$\bar{y} \pm t_{\frac{\alpha}{2}} s = 4.5 \pm (2.947) \frac{(0.28)}{\sqrt{16}} = 4.5 \pm \frac{0.82516}{4}$$
$$= [4.29371, 4.70629]$$

3. A manufacturer wishes to estimate the proportion of washing machines leaving the factory that is defective. How large a sample should she check in order to be 92 percent confident that the true proportion is estimated to within 0.012?

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$1.75 \sqrt{\frac{\frac{1}{2} \cdot \frac{1}{2}}{n}} = 0.012$$

$$n = \frac{(1.75)^2}{4 (0.012)^2} = 5316.84$$

$$\underline{n = 5317}$$

4. A new programming language is being evaluated for possible use at a large company. As part of the evaluation process, several standard programming tasks will be carried out using the new language, and the time required to complete each task recorded. From past projects using the old computer language, a different sample of 12 jobs were available. Assuming that these observations are i.i.d. normal with the same variance. These two samples (old and new language) are summarized as follows:

Language	sample size	sample mean	sample standard deviation
new	15	8.3	2.2
old	12	9.5	2.5

Calculate a 95% confidence interval for the difference between mean times of the new and old computer languages.

$$\bar{y}_1 - \bar{y}_2 \pm t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(14)(2.2)^2 + (11)(2.5)^2}{14 + 11} = 5.4609$$

$$9.5 - 8.3 \pm (2.060) \sqrt{5.4609} \sqrt{\frac{1}{15} + \frac{1}{12}}$$

$$= 1.2 \pm 1.8643 = (-0.6643, 3.0643)$$

5. Acme soft drinks has been plagued by poor quality in the bottles they have been supplied in the past 6 months. A little acceptance sampling will reduce this problem, they hope. So they propose to randomly sample 48 bottles from every shipment of 24,000 they are supplied in the future. If the average strength of these 48 bottles falls below 180 (internal pressure strength of 180 pounds per square inch), they will reject the shipment. Suppose a new shipment in fact is quite satisfactory: The 24,000 bottles have an average strength of 182, and a standard deviation of 4. What is the approximate chance that such a shipment would nevertheless be rejected by the sampling procedure?

prob of rejected shipment =

$$= P(\bar{Y} < 180) = P\left(N\left(0, 1\right) \leq \frac{180 - 182}{4/\sqrt{48}}\right)$$

$$= P(N(0, 1) \leq -3.46) = 0.000233$$

6. A Gallup poll conducted in November of 2002 showed that 15% of the 1017 surveyed in 2002 were very likely to do some or all of their Christmas shopping on line. A similar survey in 2001 showed that only 9% of 1030 surveyed were likely to do their holiday shopping on line. Find a 95% confidence interval for the increase in the proportion of all Americans doing holiday shopping on line (in other words estimate $p_{2002} - p_{2001}$).

$$\hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$0.15 - 0.09 \pm 1.96 \sqrt{\frac{0.15(0.85)}{1017} + \frac{(0.09)(0.91)}{1030}}$$

$$= 0.06 \pm 0.0280 = (0.032, 0.088)$$

7. In a large production run of millions of electronic chips, only 2% are defective. What is the chance that of 1000 chips pulled off the assembly line, 40 or more would be defective?

$$Y \sim \text{Bin}(1000, p = 0.02)$$

$$P(Y \geq 40) = P\left(N(0,1) \geq \frac{39.5 - 20}{\sqrt{1000(0.02)(0.98)}}\right)$$

$$= P(N(0,1) \geq 4.4046) = 0$$

8. I ski down a trail 8 times in one morning, and 1 time each of my runs from top to bottom. I get the following times (in seconds): 489, 469, 512, 493, 476, 481, 502, 496. Assuming that these observations are i.i.d. normal, find an 80% confidence interval for the variance of my run time from top to bottom.

$$\sum y_i = 3918, \quad \bar{y} = 489.75$$

$$\sum y_i^2 = 1920232$$

$$\sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{1}{n}(\sum y_i)^2 = 1920232 - \frac{(3918)^2}{8} = 1391.5$$

$$s^2 = \frac{\sum (y_i - \bar{y})^2}{7} = 198.7857$$

$$\chi_{1-\frac{\alpha}{2}}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi_{\frac{\alpha}{2}}^2$$

A 80% CI is

$$\left[\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}^2}, \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}^2} \right] = \left[\frac{1391.5}{12.0170}, \frac{1391.5}{2.83311} \right]$$

$$= [115.79, 491.1564]$$

9. Given a random sample of size n from the density

$$f(y) = \begin{cases} \frac{4\theta^4}{y^5} & \text{if } \theta < y, \\ 0 & \text{else.} \end{cases}$$

where $\theta > 0$ is unknown parameter. Show that $\hat{\theta} = Y_{(1)}$ is a consistent estimator of θ .

Method 1

$$P(Y_1 = \frac{4\theta^4}{y^5}, \text{ if } y > \theta, \quad 1 - F(y) = \int_y^{\infty} \frac{4\theta^4}{t^5} dt = \frac{\theta^4}{y^4}, \text{ if } y > \theta$$

Given $\varepsilon > 0$,

$$\begin{aligned} P(|\hat{\theta} - \theta| \geq \varepsilon) &= P(\hat{\theta} > \theta + \varepsilon) + P(\hat{\theta} < \theta - \varepsilon) = P(Y_{(1)} \geq \theta + \varepsilon) \\ &= P(Y_1 \geq \theta + \varepsilon, \dots, Y_n \geq \theta + \varepsilon) = (P(Y \geq \theta + \varepsilon))^n = \left(\frac{\theta}{\theta + \varepsilon}\right)^{4n} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

Method 2

$$f_{Y_{(1)}}(y) = n(1 - F(y))^{n-1} f(y) = n \left(\frac{\theta^4}{y^4}\right)^{n-1} \frac{4\theta^4}{y^5} = \frac{4n\theta^{4n}}{y^{4n+1}}, \text{ if } y > \theta$$

$$E[Y_{(1)}] = \int_{\theta}^{\infty} y \frac{4n\theta^{4n}}{y^{4n+1}} dy = \int_{\theta}^{\infty} 4n\theta^{4n} y^{-4n} dy = \frac{4n\theta^{4n} y^{-4n+1}}{-4n+1} \Big|_{\theta}^{\infty}$$

$$= \frac{4n\theta}{4n-1} \xrightarrow{n \rightarrow \infty} \theta$$

$$E[Y_{(1)}^2] = \int_{\theta}^{\infty} y^2 \frac{4n\theta^{4n}}{y^{4n+1}} dy = \int_{\theta}^{\infty} 4n\theta^{4n} y^{-4n+1} dy = \frac{4n\theta^{4n} y^{-4n+2}}{-4n+2} \Big|_{\theta}^{\infty}$$

$$= \frac{4n\theta^2}{4n-2}$$

$$\text{Var}(Y_{(1)}) = \frac{4n\theta^2}{4n-2} - \left(\frac{4n\theta}{4n-1}\right)^2 \xrightarrow{n \rightarrow \infty} \theta^2 - \theta^2 = 0$$

10. Given a random sample of size n from the density

$$f(y) = \begin{cases} \theta e^{-\theta y} & \text{if } 0 < y, \\ 0 & \text{else,} \end{cases}$$

where $\theta > 0$ is unknown parameter. Show that

$$\frac{n}{\sum_{j=1}^n Y_j}$$

is a consistent estimator of θ .

$$E[Y] = \int_0^{\infty} y \theta e^{-\theta y} dy = \int_0^{\infty} u e^{-u} \frac{du}{\theta} = \frac{1}{\theta}$$

$\theta y = u$

By the LLN, $\bar{Y} \xrightarrow{P} E[Y] = \frac{1}{\theta}$.

So, $\frac{1}{\bar{Y}} \xrightarrow{P} \theta$