

ENGINEERS STAT LECTURE NOTES

CHAPTER 7

ESTIMATION

* An estimator is a function of random variables that aims to estimate an unknown parameter of the population. Namely, estimator is sample statistics.

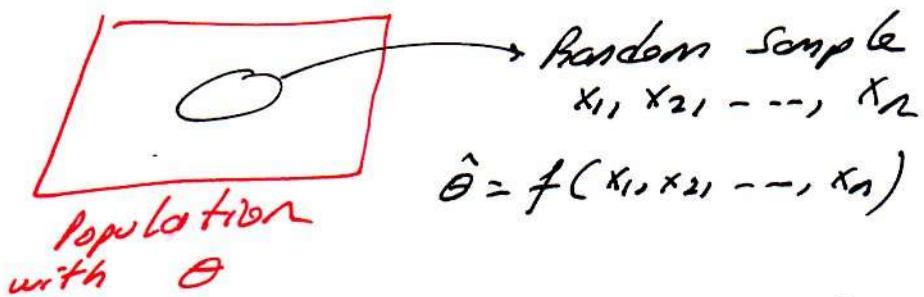
(Unknown Constant)
Population
Parameter

(Known Variable)
Sample
Statistic

$$\theta \xleftarrow{\text{Estimation}} \hat{\theta}$$

* $\hat{\theta}$ is unbiased estimator of θ if $E(\hat{\theta}) = \theta$.

This means, if we had a chance of taking all possible random samples from the population, their mean will be and calculate $\hat{\theta}$ for each sample, their mean will be θ .



* Unbiasedness is an important property but does NOT guarantee accuracy. Therefore, we use the estimator which has minimum variance among unbiased estimators.

- 7.1 Suppose X_1, X_2, X_3 denotes a random sample from the exponential distribution with density function

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x > 0 \\ 0 & \text{elsewhere} \end{cases} \quad \mu = \theta \quad \sigma^2 = \theta^2$$

Consider the following four estimators of θ :

$$\begin{aligned}\hat{\theta}_1 &= X_1 \\ \hat{\theta}_2 &= \frac{X_1 + X_2}{2} \\ \hat{\theta}_3 &= \frac{X_1 + 2X_2}{3} \\ \hat{\theta}_4 &= \bar{X}\end{aligned}$$

- a Which of the above estimators are unbiased for θ ?
- b Among the unbiased estimators of θ , which has the smallest variance?

7.1) $X \sim \text{Exponential}(\theta)$

$$E(X) = \theta \quad \text{Var}(X) = \theta^2$$

a) $E(\hat{\theta}_1) = E(X_1) = \theta \checkmark$

$$E(\hat{\theta}_2) = E\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2} \left[E(X_1) + E(X_2) \right] = \frac{1}{2} \cdot 2\theta = \theta \checkmark$$

$$E(\hat{\theta}_3) = E\left(\frac{X_1 + 2X_2}{3}\right) = \frac{1}{3} \left[E(X_1) + 2E(X_2) \right] = \frac{1}{3} \cdot 3\theta = \theta \checkmark$$

$$E(\hat{\theta}_4) = E(\bar{X}) = E\left[\frac{\sum X_i}{3}\right] = \frac{1}{3} \left[E(X_1) + E(X_2) + E(X_3) \right] = \frac{1}{3} \cdot 3\theta = \theta \checkmark$$

b) $\text{Var}(\hat{\theta}_1) = \theta^2$

$$\text{Var}(\hat{\theta}_2) = \text{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4} \left[\text{Var}(X_1) + \text{Var}(X_2) \right] = \frac{1}{4} \cdot 2\theta^2 \frac{\theta^2}{2}$$

$$\text{Var}(\hat{\theta}_3) = \frac{1}{9} \left[\text{Var}(X_1) + 4 \text{Var}(X_2) \right] = \frac{5\theta^2}{9}$$

$$\text{Var}(\hat{\theta}_4) = \frac{1}{9} \left[\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) \right] = \frac{\theta^2}{3} : \text{Minimum Variance}$$

- 7.2** The reading on a voltage meter connected to a test circuit is uniformly distributed over the interval $(\theta, \theta + 1)$, where θ is the true but unknown voltage of the circuit. Suppose X_1, \dots, X_n denotes a random sample of readings from this voltage meter.

- a Show that \bar{X} is a biased estimator of θ .
- b Find a function of \bar{X} that is an unbiased estimator of θ .

$$X \sim \text{Uniform}(\theta; \theta+1)$$

a) $E(X_i) = \frac{\theta+1+\theta}{2} = \theta + \frac{1}{2}$

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{\sum X_i}{n}\right) = \frac{1}{n} \cdot \sum E(X_i) = \frac{1}{n} \cdot \sum \left(\theta + \frac{1}{2}\right) = \frac{1}{n} \cdot \left(\theta + \frac{n}{2}\right) \\ &= \theta + \frac{1}{2} \end{aligned}$$

b) $E(\bar{X}) = \theta + \frac{1}{2}$

$$E(\bar{X}) - \frac{1}{2} = \theta$$

$E(\bar{X} - \frac{1}{2}) = \theta$ Then, $\hat{\theta} = \bar{X} - \frac{1}{2}$ is an unbiased estimator of θ .

- 7.4** The bias B of an estimator $\hat{\theta}$ is given by

$$B = |\theta - E(\hat{\theta})|$$

The mean squared error, or MSE, of an estimator $\hat{\theta}$ is given by

$$\text{MSE} = E(\hat{\theta} - \theta)^2$$

Show that

$$\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + B^2$$

[Note: $\text{MSE}(\hat{\theta}) = V(\hat{\theta})$ if $\hat{\theta}$ is an unbiased estimator of θ . Otherwise, $\text{MSE}(\hat{\theta}) > V(\hat{\theta})$.]

- 7.5** Refer to Exercise 7.2. Find $\text{MSE}(\bar{X})$ when \bar{X} is used to estimate θ .

$$\begin{aligned} 7.4) \quad E[(\hat{\theta} - \theta)^2] &= E(\theta^2 - 2\theta\hat{\theta} + \hat{\theta}^2) = E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2 \\ &= E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2 - E^2(\hat{\theta}) + E^2(\hat{\theta}) \\ &= \underbrace{E(\hat{\theta}^2)}_{=\text{Var}(\hat{\theta})} - \underbrace{E^2(\hat{\theta})}_{=[E(\hat{\theta}) - \theta]^2} + E^2(\hat{\theta}) = \text{Var}(\hat{\theta}) + B^2 \end{aligned}$$

(17)



7.5) $X \sim \text{Uniform}(\theta; \theta+1)$

$$E(X_i) = \frac{\theta+1+\theta}{2} = \theta + \frac{1}{2} \quad \text{Var}(X_i) = \frac{(\theta+1-\theta)^2}{12} = \frac{1}{12}$$

$$E(\bar{X}) = \theta + \frac{1}{2}; \quad B = E(\bar{X}) - \theta = \theta + \frac{1}{2} - \theta = \frac{1}{2}$$

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{\sum X_i}{n}\right) = \frac{1}{n^2} \cdot \sum \text{Var}(X_i) = \frac{1}{n^2} \cdot \frac{n}{12} = \frac{1}{12n}$$

$$MSE(\bar{X}) = \text{Var}(\bar{X}) + B^2 = \frac{1}{12n} + \left(\frac{1}{2}\right)^2 = \frac{1}{12n} + \frac{1}{4}$$

7.8 Suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are each unbiased estimators of θ , with $V(\hat{\theta}_1) = \sigma_1^2$ and $V(\hat{\theta}_2) = \sigma_2^2$. A new unbiased estimator for θ can be formed by

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

($0 \leq a \leq 1$). If $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent, how should a be chosen so as to minimize $V(\hat{\theta}_3)$?

$$7.8) \text{Var}(\hat{\theta}_3) = \text{Var}[a\hat{\theta}_1 + (1-a)\hat{\theta}_2] = a^2 \cdot \text{Var}(\hat{\theta}_1) + (1-a)^2 \cdot \text{Var}(\hat{\theta}_2)$$

$$\text{Var}(\hat{\theta}_3) = a^2 \cdot \sigma_1^2 + (1-a)^2 \cdot \sigma_2^2$$

$$\frac{d \text{Var}(\hat{\theta}_3)}{da} = 2a \sigma_1^2 - 2(1-a) \sigma_2^2 = 0$$

$$a \sigma_1^2 - (1-a) \sigma_2^2 = 0$$

$$a \sigma_1^2 - \sigma_2^2 + a \sigma_2^2 = 0$$

$$a(\sigma_1^2 + \sigma_2^2) = \sigma_2^2$$

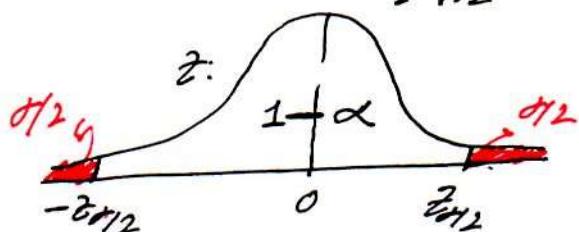
$$a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

One Sample Confidence Intervals.

Population Mean: μ

* Remember, for n sufficiently large (practically $n \geq 30$)
 Sample mean \bar{X} has a Normal distribution with
 mean μ and variance σ^2 . Then;

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ and}$$



$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

$$-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2} = 1 - \alpha$$

$$-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$-\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < -\mu < z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} - \bar{X}$$

$$P\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

So; $(1 - \alpha) \cdot 100\%$ Confidence Interval for μ is;

$$\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

This term is called Margin of Error.

* For small samples, assuming X_i has Normal distribution, we have; with degrees of freedom = $n-1$,

$$z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \quad \text{and}$$

$$\bar{X} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

is $(1-\alpha) \cdot 100\%$.

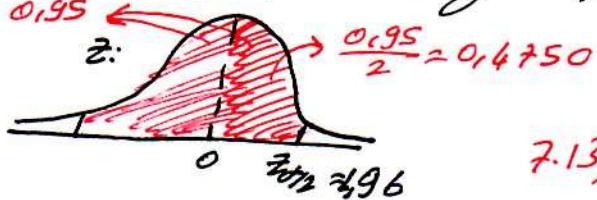
confidence interval for μ .

- 7.10 A random sample of 40 engineers was selected from among the large number employed by a corporation engaged in seeking new sources of petroleum. The hours worked in a particular week were determined for each engineer selected. These data had a mean of 46 hours and a standard deviation of 3 hours. For that particular week, estimate the mean hours worked for all engineers in the corporation, with a 95% confidence coefficient.

- 7.13 In the setting of Exercise 7.10, how many engineers should be sampled if it is desired to estimate the mean number of hours worked to within 0.5 hour with confidence coefficient 0.95?

7.10) $n=40 ; \bar{X}=46 ; s=3$

Since n is large, we may replace σ by s .



$$7.13) z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 0.95$$

95% C.I. for μ is;

$$\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$1.96 \cdot \frac{3}{\sqrt{40}} = 0.5$$

$$\sqrt{n} = \frac{1.96 \cdot 3}{0.5}$$

$$46 \pm 1.96 \cdot \frac{3}{\sqrt{40}}$$

$$(45.07 ; 46.93)$$

$$n = \left(\frac{1.96 \cdot 3}{0.5} \right)^2 = 138,3$$

139 ROUND UP!

- 7.21 The warpwise breaking strength measured on five specimens of a certain cloth gave a sample mean of 180 psi and a standard deviation of 5 psi. Estimate the true mean warpwise breaking strength for cloth of this type in a 95% confidence interval. What assumption is necessary for your answer to be valid?
- 7.22 Answer Exercise 7.21 if the same sample data resulted from a sample of
 a 10 specimens.
 b 100 specimens.

$$7.21) n = 5; \bar{X} = 180; s = 5$$

Assuming $X \sim \text{Normal}$, 95% C.I. for μ is,

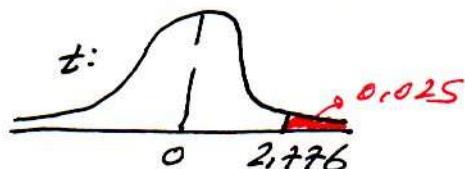
$$\bar{X} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$1 - \alpha = 0.95$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025; df = n - 1 = 4$$

$$(173.79; 186.21)$$



$$7.22) a) n = 10 \Rightarrow df = 9, t_{\alpha/2} = 2.262$$

$$180 \pm 2.262 \cdot \frac{5}{\sqrt{10}}$$

$$(176.42; 183.58)$$

$$b) n = 100 \Rightarrow \text{use } z \text{ table}, z_{\alpha/2} = 1.96$$

$$180 \pm 1.96 \cdot \frac{5}{\sqrt{100}}$$

$$(179.02; 180.98)$$

Note that, as sample size increases, the confidence interval become narrower. This is because, we have more information and so we have more accurate result.

Also note that t with ∞ degrees of freedom has $t_{\alpha/2} = 1.960$. z is limiting distribution of t .

Population Proportion: p

* Like sample mean, sample proportion has an approximately Normal distribution. This is due to "Normal Approximation to Binomial distribution".

We have, $\hat{p} \sim \text{Normal} \left(\mu_{\hat{p}} = p ; \sigma_{\hat{p}}^2 = \frac{p(1-p)}{n} \right)$

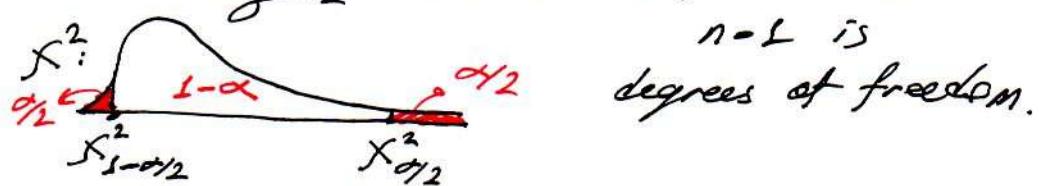
So, $(1-\alpha) \cdot 100\%$ Confidence Interval for p is;

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

Since we do NOT know p , we estimate variance of \hat{p} by $\frac{\hat{p}(1-\hat{p})}{n}$.

Population Variance: σ^2

* Remember, $V = \frac{(n-1) \cdot s^2}{\sigma^2} \sim \chi^2_{(n-1)}$ where



So; $P(X_{1-\alpha/2}^2 < V < X_{\alpha/2}^2) = 1 - \alpha$

$$X_{1-\alpha/2}^2 < \frac{(n-1) \cdot s^2}{\sigma^2} < X_{\alpha/2}^2$$

$$\frac{1}{X_{\alpha/2}^2} < \frac{\sigma^2}{(n-1) \cdot s^2} < \frac{1}{X_{1-\alpha/2}^2}$$

Then, $(1-\alpha) \cdot 100\%$ Confidence interval for σ^2 is;

$$\left(\frac{(n-1) \cdot s^2}{X_{\alpha/2}^2} ; \frac{(n-1) \cdot s^2}{X_{1-\alpha/2}^2} \right)$$

- 7.20 The Environmental Protection Agency has collected data on the LC50 (concentration killing 50% of the test animals in a specified time interval) measurements for certain chemicals likely to be found in freshwater rivers and lakes. For a certain species of fish, the LC50 measurements (in parts per million) for DDT in 12 experiments yielded the following:

$$16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4, 9$$

Assuming such LC50 measurements to be approximately normally distributed, estimate the true mean LC50 for DDT with confidence coefficient 0.90.

- 7.24 The variance of LC50 measurements is important, because it may reflect an ability (or inability) to reproduce similar results in identical experiments. Find a 90% confidence interval for σ^2 , the true variance of the LC50 measurements for DDT, using the data in Exercise 7.20.

7.20)

X_i	X_i^2
16	16^2
5	5^2
1	1
:	:
+	9
$\sum X_i = 108$	$\sum X_i^2 = 1426$

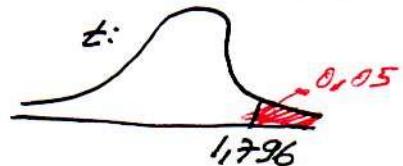
$$\bar{X} = \frac{\sum X_i}{n} = \frac{108}{12} = 9$$

$$S^2 = \frac{\sum X_i^2 - (\sum X_i)^2}{n-1} = \frac{1426 - \frac{108^2}{12}}{11} = 41,128$$

$$1-\alpha = 0,90$$

$$\alpha_{1/2} = 0,05$$

$$df = 12-1=11$$



90% C.I. for μ is;

$$\bar{X} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} : 9 \pm 1,796 \cdot \sqrt{\frac{41,128}{12}}$$

$$(5,675; 12,325)$$

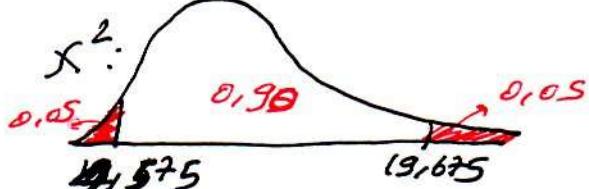
7.24)

$$1-\alpha = 0,90$$

$$\alpha_{1/2} = 0,05$$

$$1-\alpha_{1/2} = 0,95$$

$$df = n-1=11$$



90% C.I. for σ^2 is;

$$\left(\frac{(n-1)s^2}{X_{\alpha/2}^2}, \frac{(n-1).s^2}{X_{(1-\alpha/2)}^2} \right) : \left(\frac{11 \cdot 41,128}{19,675}, \frac{11 \cdot 41,128}{4,575} \right)$$

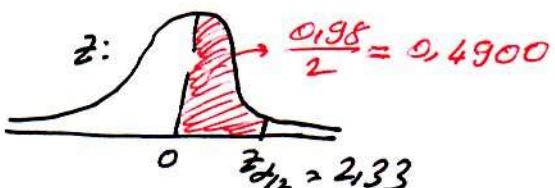
$$(22,994; 98,887)$$

(23)

- 7.17 Careful inspection of 70 precast concrete supports to be used in a construction project revealed 28 with hairline cracks. Estimate the true proportion of supports of this type with cracks in a 98% confidence interval.
- 7.18 Refer to Exercise 7.17. Suppose it is desired to estimate the true proportion of cracked supports to within 0.1, with confidence coefficient 0.98. How many supports should be sampled to achieve the desired accuracy?

$$7.17) \hat{p} = \frac{28}{70} = 0.4$$

$$n = 70$$



98% C.I. for p is;

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

$$0.4 \pm 2.33 \cdot \sqrt{\frac{0.4 \cdot (1-0.4)}{70}}$$

$$(0.264; 0.536)$$

$$7.18) 2.33 \cdot \sqrt{\frac{0.5(1-0.5)}{n}} = 0.1$$

$$\sqrt{n} = \frac{2.33 \cdot 0.5}{0.1} \Rightarrow n = \left(\frac{2.33 \cdot 0.5}{0.1} \right)^2 = 135.72$$

$$n = 136 \quad \text{ROUND UP!}$$

We use $\hat{p} = 0.5$ because this maximizes $\hat{p}(1-\hat{p})$.

If we had a prior information about p such that $p \in (0.1; 0.3)$, we would use $\hat{p} = 0.3$ for sample size determination.

- 7.31 The quantity

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (\text{or } z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}})$$

used in constructing confidence intervals for μ (or p) is sometimes called the *sampling error*. A *Time* (5 April 1993) article on religion in America reported that 54% of those between the ages of 18 and 26 found religion to be a "very important" part of their lives. The article goes on to state that the result comes from a poll of 1,013 people and has a sampling error of 3%. How is the 3% calculated and what is its interpretation? Can we conclude that a majority of people in this age group find religion to be very important?

$$\hat{p} = 0,54 ; n = 1013 ;$$

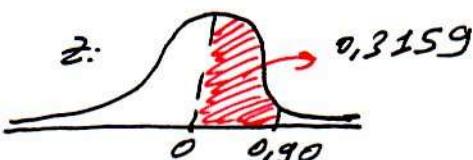
$(1-\alpha) \cdot 100\%$ C.I. for p is;

$$0,54 \pm 0,03$$

Then; $z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,03$

$$z_{\alpha/2} \cdot \sqrt{\frac{0,54 \cdot 0,46}{1013}} = 0,03 \Rightarrow z_{\alpha/2} = 0,90$$

$$1-\alpha = 2 \cdot 0,3159 = 0,6318$$



Although the C.I. $(0,51; 0,57)$ does NOT contain $0,5$ (more than 50% is majority), we cannot conclude majority of people because confidence level 63% is quite low.

Two Sample Confidence Intervals

* Remember, linear combinations of ^{NORMAL} random variables will have a normal distribution. Namely, if $X_1, X_2, \dots, X_k \sim \text{Normal}(\mu_i; \sigma_i^2)$ and if X_j are independent $j=1, 2, \dots, k$, then

$$W = a_1 X_1 + a_2 X_2 + \dots + a_k X_k$$

$$\mu_W = E(W) = a_1 E(X_1) + a_2 E(X_2) + \dots + a_k E(X_k)$$

$$= \sum a_j E(X_j) \quad \text{and}$$

$$\sigma_W^2 = \text{Var}(W) = a_1^2 \text{Var}(X_1) + \dots + a_k^2 \text{Var}(X_k)$$

$$= \sum a_j^2 \text{Var}(X_j)$$

Then, by central limit theorem, large sample confidence interval for μ_w will be

$$\hat{\mu}_w \pm \sqrt{\text{Var}(w)}$$

where $\hat{\mu}_w$ is an estimator of μ . (i.e. \bar{X}_w)

Population Means Difference;

(i) *Paired Samples.*

If units of the data are observed under two different conditions, this data is called "paired sample". The usual example for this is Before-After case.

Ex: A one-month diet is taken by 8 candidates and the following data is observed.

i	1	2	3	4	5	6	7	8
Before	93	105	86	97	73	81	78	95
After.	90	100	82	99	73	77	75	94

Find 95% C.I. for the diet's weight loss.

Ans:

$$\begin{array}{ccccccccc}
 & \text{Before} & \text{After} & D_i & D_i^2 & & & \\
 \hline
 93 & - & 90 & = 3 & 3^2 & & & \\
 105 & - & 100 & = 5 & 5^2 & & & \\
 \vdots & & \vdots & \vdots & \vdots & & & \\
 95 & & 94 & + 1 & 1^2 & & & \\
 & & & \hline
 & \sum D_i = 18 & \sum D_i^2 = 72 & & & & & \\
 & & & & & & & \\
 & & & & & \bar{D} = \frac{\sum D_i}{n} = \frac{18}{8} = 2,25 & & \\
 & & & & & S_D^2 = \frac{\sum D_i^2 - (\sum D_i)^2}{n-1} & & \\
 & & & & & & & \\
 & & & & & = \frac{72 - 18^2}{7} = 4,5 & & \\
 & & & & & S_D = \sqrt{4,5} = 2,12 & & \\
 \end{array}$$

$(1-\alpha) \cdot 100\%$ C.I. for $\mu_1 - \mu_2$ is;

$$\bar{D} \pm t_{df,2} \cdot \frac{s_D}{\sqrt{n}}$$

$$1-\alpha = 0,95$$

$$\alpha_{1/2} = 0,05$$

$$df = n - 1 = 7$$

$$2,25 \pm 2,365 \cdot \frac{2,12}{\sqrt{8}}$$

$$(0,477; 4,023)$$

Note that, after taking the differences, the C.I reduces to one sample confidence interval. The interpretation is, we are 95% confident that the diet's weight loss is between 0,477 kg and 4,023 kg in a month. We also assume that the distribution of weights are NORMAL. (Or directly the weight loss is NORMAL.)

(ii) Independent Samples
with large n OR known variances.

If we have two populations and if we take one sample from each population, these are independent samples. For large n , we have,

$$\bar{X}_1 \sim \text{Normal}(\mu_1, \sigma_1^2)$$

$$\bar{X}_2 \sim \text{Normal}(\mu_2, \sigma_2^2)$$

and $E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$

$$\text{Var}(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\text{So; } \bar{X}_1 - \bar{X}_2 \sim \text{Normal}(\mu_1 - \mu_2; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

and $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \text{Normal}(\mu=0; \sigma^2=1)$

Then, $(1-\alpha) \cdot 100\% \text{ C.I. for } \mu_1 - \mu_2 \text{ is;}$

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- 7.34** The abrasive resistance of rubber is increased by adding a silica filler and a coupling agent to chemically bond the filler to the rubber polymer chains. Fifty specimens of rubber made with a type I coupling agent gave a mean resistance measure of 92, the variance of the measurements being 20. Forty specimens of rubber made with a type II coupling agent gave a mean of 98 and a variance of 30 on resistance measurements. Estimate the true difference between mean resistances to abrasion in a 95% confidence interval.

- 7.35** Refer to Exercise 7.34. Suppose a similar experiment is to be run again with an equal number of specimens from each type of coupling agent. How many specimens should be used if we want to estimate the true difference between mean resistances to within 1 unit, with a confidence coefficient of 0.95?

7.34)

Type I

$$n_1 = 50$$

$$\bar{X}_1 = 92$$

$$\sigma_1^2 = 20$$

Type II

$$n_2 = 40$$

$$\bar{X}_2 = 98$$

$$\sigma_2^2 = 30$$

$$1-\alpha = 0.95$$



95% C.I. for $\mu_2 - \mu_1$ is;

$$(\bar{X}_2 - \bar{X}_1) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(98 - 92) \pm 1.96 \sqrt{\frac{20}{50} + \frac{30}{40}}$$

\leftarrow \rightarrow

$$(3.898; 8.102)$$

Note that, C.I. does NOT contain 0. So, we are 95% sure that Type II is better.

$$7.35) z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1 \quad \text{let, } n_1 = n_2 = n$$

$$1.96 \cdot \sqrt{\frac{20+30}{n}} = 1$$

$$\sqrt{n} = \frac{1.96 \cdot \sqrt{50}}{1} \Rightarrow n = 1.96^2 \cdot 50 = 192,08$$

$n = 193$ *Round Up!*

- 7.53 An electric circuit contains three resistors, each of a different type. Tests on 80 type I resistors showed a sample mean resistance of 9.1 ohms with a sample standard deviation of 0.2 ohm, tests on 88 type II resistors yielded a sample mean of 14.3 ohms and a sample standard deviation of 0.4 ohm, while tests on 112 type III resistors yielded a sample mean of 5.6 ohms and a sample standard deviation of 0.1 ohm. Find a 95% confidence interval for $\mu_I + \mu_{II} + \mu_{III}$, the expected resistance for the circuit. What assumptions are necessary for your answer to be valid?

7.53) Type I Type II Type III

$$n_1 = 80$$

$$n_2 = 88$$

$$n_3 = 112$$

$$\bar{x}_1 = 9.1$$

$$\bar{x}_2 = 14.3$$

$$\bar{x}_3 = 5.6$$

$$\sigma_1 = 0.2$$

$$\sigma_2 = 0.4$$

$$\sigma_3 = 0.1$$

$$z_{\alpha/2} = 1.96$$

$$E(\bar{x}_1 + \bar{x}_2 + \bar{x}_3) = \mu_I + \mu_{II} + \mu_{III}$$

$$\text{Var}(\bar{x}_1 + \bar{x}_2 + \bar{x}_3) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} + \frac{\sigma_3^2}{n_3}$$

Then; 95% C.I. for $\mu_I + \mu_{II} + \mu_{III}$ is;

$$(\bar{x}_1 + \bar{x}_2 + \bar{x}_3) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} + \frac{\sigma_3^2}{n_3}}$$

$$(9.1 + 14.3 + 5.6) \pm 1.96 \cdot \sqrt{\frac{0.2^2}{80} + \frac{0.4^2}{88} + \frac{0.1^2}{112}}$$

$$(28.904 ; 29.096)$$

(iii) Independent samples with small n AND unknown (but assumed equal) variances.

Assuming X_1 and X_2 have normal distribution, their pooled variances is;

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1) + (n_2-1)}$$

And $(1-\alpha) \cdot 100\%$. C.I. for $\mu_1 - \mu_2$ is;

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \cdot \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \text{ with } df = (n_1-1) + (n_2-1)$$

- 7.44** Unaltered, altered, and partly altered bitumens are found in carbonate-hosted lead-zinc deposits and may aid in the production of sulfide necessary to precipitate ore bodies in carbonate rocks. (See T. G. Powell, *Science*, 6 April 1984, p. 63.) The atomic hydrogen/carbon (H/C) ratios for 15 samples of altered bitumen had a mean of 1.02 and a standard deviation of 0.04. The ratios for 7 samples of partly altered bitumen had a mean of 1.16 and a standard deviation of 0.05. Estimate the difference between true mean H/C ratios for altered and partly altered bitumen, in a 98% confidence interval. Assume equal population standard deviations.

7.44)

Altered

$$n_1 = 15$$

$$\bar{X}_1 = 1,02$$

$$s_1 = 0,04$$

Partly Altered

$$n_2 = 7$$

$$\bar{X}_2 = 1,16$$

$$s_2 = 0,05$$

$$1-\alpha = 0,98$$

$$\alpha/2 = 0,01$$

$$df = 14 + 6 = 20$$



$$s_p^2 = \frac{14 \cdot 0,04^2 + 6 \cdot 0,05^2}{14+6} = 0,00187$$

98% C.I. for $\mu_2 - \mu_1$ is;

$$(1,16 - 1,02) \pm 2,528 \cdot \sqrt{0,00187 \left(\frac{1}{15} + \frac{1}{7} \right)}$$

$$(0,090; 2,130)$$



Population Variances Ratio.

$(1-\alpha) \cdot 100\%$ C.I. for σ_2^2 / σ_1^2 is;

$$\left(\frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{1}{F_{2,2}(v_1, v_2)}, \frac{\sigma_2^2}{\sigma_1^2} \cdot F_{d,2}(v_1, v_2) \right)$$

where $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ are degrees of freedom for numerator and denominator.

- 7.41 For a certain species of fish, the LC50 measurements (in parts per million) for DDT in 12 experiments were as follows, according to the EPA:

16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4, 9

Another common insecticide, Diazinon, gave LC50 measurements of 7.8, 1.6, and 1.3 in three independent experiments. Estimate the difference between the mean LC50 for DDT and the mean LC50 for Diazinon in a 90% confidence interval. What assumptions are necessary for your answer to be valid? Also estimate the true variance ratio in a 90% confidence interval.

7.41)

$n_1 = 12$ X_{1i} 16 5 21 : 9	X_{1i}^2 16^2 5^2 21^2 : 9^2	$\bar{X}_1 = \frac{108}{12} = 9$ $s_1^2 = \frac{1426 - \frac{108^2}{12}}{11}$ $s_1 = 6,424$ $\sum X_{1i} = 108$ $\sum X_{1i}^2 = 1426$	X_{2i} 7,8 1,6 1,3 + $\sum X_{2i} = 10,7$	X_{2i}^2 $7,8^2$ $1,6^2$ $1,3^2$ + $\sum X_{2i}^2 = 65,09$	$\bar{X}_2 = \frac{10,7}{3} = 3,57$ $s_2^2 = \frac{65,09 - \frac{10,7^2}{3}}{2} = 13,86$ $s_2 = 3,724$ $s_p^2 = \frac{9 \cdot 61,27 + 2 \cdot 13,86}{9+2} = 36,286$
---	---	--	--	---	--

DDT Diazinon

$n_1 = 12$ $n_2 = 3$

$\bar{X}_1 = 9$ $\bar{X}_2 = 3,57$

$s_1 = 6,424$ $s_2 = 3,724$

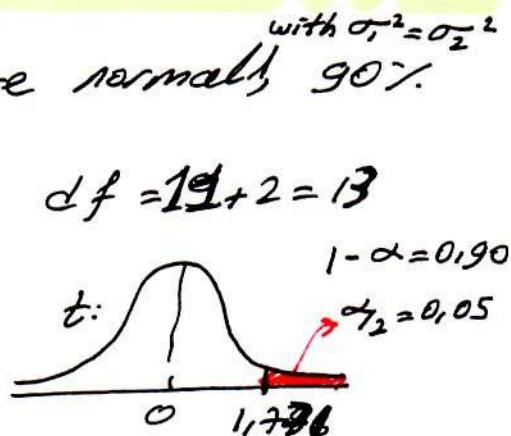
Assuming both populations are normal, 90% .

C.I. for $\mu_1 - \mu_2$ is;

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \cdot \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

$$(9 - 3,57) \pm 1,776 \sqrt{\frac{36,286}{12} + \frac{36,286}{3}}$$

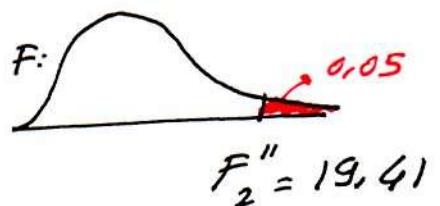
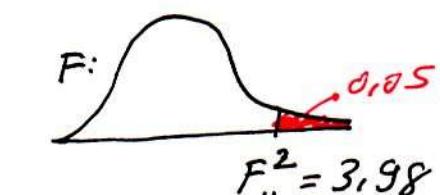
$$(-1,45; 14,58)$$



Since the C.I. contains 0, we can NOT claim that their means are different.

90% C.I. for σ_1^2 / σ_2^2 is;

$$\left(\frac{s_1^2}{s_2^2} \cdot \frac{1}{F''_{2; 0,05}} ; \frac{s_1^2}{s_2^2} \cdot F''_{11; 0,05} \right)$$

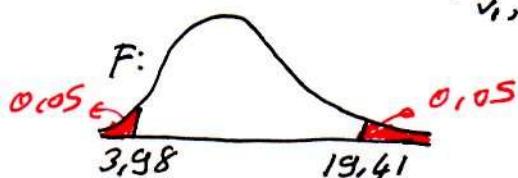


$$\left(\frac{48,27}{13,86} \cdot \frac{1}{19,41} ; \frac{41,27}{13,86} \cdot 3,98 \right)$$

$$(0,153; 11,85)$$

Since the C.I. contains 1, we can assume $\sigma_1^2 = \sigma_2^2$

Note that; $F_{v_2, \alpha}^{v_1} = \frac{1}{F_{v_1, 1-\alpha}^{v_2}}$. The idea is, like that of C.I. for σ^2 ,





Population Proportions Difference;

We have, $\hat{p}_1 \sim \text{Normal}(\rho_1; \frac{\rho_1(1-\rho_1)}{n_1})$ and

$$\hat{p}_2 \sim \text{Normal}(\rho_2; \frac{\rho_2(1-\rho_2)}{n_2})$$

So, $E(\hat{p}_1 - \hat{p}_2) = \rho_1 - \rho_2$ and

$$\text{Var}(\hat{p}_1 - \hat{p}_2) = \text{Var}(\hat{p}_1) + \text{Var}(\hat{p}_2) = \frac{\rho_1(1-\rho_1)}{n_1} + \frac{\rho_2(1-\rho_2)}{n_2}$$

Finally, the $(1-\alpha) \cdot 100\%$ C.I. for $\rho_1 - \rho_2$ is;

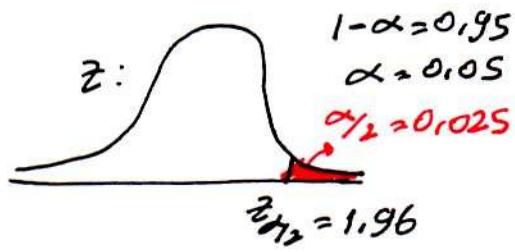
$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- 7.37** Bacteria in water samples are sometimes difficult to count, but their presence can easily be detected by culturing. In 50 independently selected water samples from a certain lake, 43 contained certain harmful bacteria. After adding a chemical to the lake water, another 50 water samples showed only 22 with the harmful bacteria. Estimate the true difference between the proportions of samples containing the harmful bacteria with a 95% confidence coefficient. Does the chemical appear to be effective in reducing the amount of bacteria?

- 7.38** In studying the proportion of water samples containing harmful bacteria, how many samples should be selected before and after a chemical is added if we want to estimate the true difference between proportions to within 0.1 with a 95% confidence coefficient? (Assume the sample sizes are to be equal.)

7.37)

<u>Before chemical</u>	<u>After chemical</u>
$X_1 = 43$	$X_2 = 22$
$n_1 = 50$	$n_2 = 50$
$\hat{p}_1 = \frac{43}{50} = 0,86$	$\hat{p}_2 = \frac{22}{50} = 0,44$



95% C.I. for $p_1 - p_2$ (reduction in amount of bacteria) is;

$$(0,86 - 0,46) \pm 1,96 \cdot \sqrt{\frac{0,86(1-0,86)}{50} + \frac{0,46(1-0,46)}{50}}$$

$$(0,252; 0,588)$$

Since the interval does NOT contain 0, we can be 95% sure that chemical reduced bacteria proportion.

7.38)

$$z_{0,025} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 0,1$$

we take $n_1 = n_2 = n$ and $\hat{p}_1 = \hat{p}_2 = 0,5$ (to max. variance)

$$1,96 \cdot \sqrt{\frac{0,5 \cdot 0,5}{n} + \frac{0,5 \cdot 0,5}{n}} = 0,1$$

$$\sqrt{n} = \frac{1,96 \cdot \sqrt{0,5}}{0,1} = 13,86 \Rightarrow n = 13,86^2 = 192,08$$

$\boxed{n = 193}$ *Round Up!*

7.52

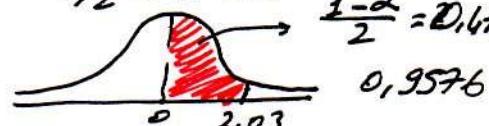
Time-Yankelovich surveys, regularly seen in the news magazine Time, report on telephone surveys of approximately 1,000 respondents. In December 1983, 60% of the respondents said that they worry about nuclear war. In a similar survey in June 1983, only 50% said that they worry about nuclear war. (See Time, 2 January 1984, p. 51.) The article reporting these figures says that when they are compared "the potential sampling error is plus or minus 4.5%." Explain how the 4.5% is obtained and what it means. Then estimate the true difference in these proportions in a 95% confidence interval.

$$n_1 = 1000 \quad \hat{p}_1 = 0,60 ; \quad n_2 = 1000 \quad \hat{p}_2 = 0,50$$

$$z_{0,025} \cdot \sqrt{\frac{0,60 \cdot 0,40}{1000} + \frac{0,50 \cdot 0,50}{1000}} = 0,045$$

$$0,0221 z_{0,025} = 0,045 \Rightarrow z_{0,025} = 2,03 \quad \frac{1-\alpha}{2} = 0,6788$$

confidence level is 96%.



95% C.I. is similar; $(0,60 - 0,50) \pm 0,045$

$$(0,055; 0,165)$$

34

Maximum Likelihood Estimation

let $f(x)$ is a pdf (or a pmf) with an unknown parameter θ . Given the data values $X_1 = x_1, \dots, X_n = x_n$, what value of θ maximizes this outcome? This is nothing, but maximum likelihood estimator $\hat{\theta}_{MLE}$:)

The procedure is as follows:

(i) Find likelihood function $L(\theta)$

$$L(\theta) = f(x_1, x_2, \dots, x_n; \theta) \stackrel{\text{by independence.}}{=} f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta)$$

(ii) Take \ln of $L(\theta)$: $\ln[L(\theta)]$. we make this because max. value of $L(\theta)$ and max. value of $\ln[L(\theta)]$ is at the same θ_0 value. But the latter is much more tractable.

(iii) Find $\hat{\theta}_{MLE}$ that maximizes $L(\theta)$. That is,

solve
for $\hat{\theta}$

$$\frac{\partial \ln[L(\theta)]}{\partial \theta} = 0$$

- 7.66 If X_1, \dots, X_n denotes a random sample from a Poisson distribution with mean λ , find the maximum likelihood estimator of λ .

- 7.67 Since $V(X_i) = \lambda$ in the Poisson case, it follows from the Central Limit Theorem that \bar{X} will be approximately normally distributed with mean λ and variance λ/n , for large n .

- a Use the above facts to construct a large-sample confidence interval for λ .
- b Suppose that 100 reinforced concrete trusses were examined for cracks. The average number of cracks per truss was observed to be four. Construct an approximate 95% confidence interval for the true mean number of cracks per truss for trusses of this type. What assumptions are necessary for your answer to be valid?



7.66) $X_i \sim \text{Poisson}(\lambda)$

$$f(x_i) = \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!}$$

$$\mathcal{L}(\lambda) = \frac{e^{-\lambda} \cdot \lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda} \cdot \lambda^{x_2}}{x_2!} \cdots \frac{e^{-\lambda} \cdot \lambda^{x_n}}{x_n!} = \frac{e^{-n\lambda} \cdot \lambda^{\sum x_i}}{\prod_{i=1}^n (x_i!)}$$

$$\ln \mathcal{L}(\lambda) = \ln \left[\frac{e^{-n\lambda} \cdot \lambda^{\sum x_i}}{\prod_{i=1}^n (x_i!)} \right] = \ln e^{-n\lambda} + \ln \lambda^{\sum x_i} - \ln \prod_{i=1}^n (x_i!)$$

$$\ln \mathcal{L}(\lambda) = -n\lambda + (\sum x_i) \ln \lambda - \sum \ln (x_i!)$$

$$\frac{\partial \ln \mathcal{L}(\lambda)}{\partial \lambda} = -n + \frac{\sum x_i}{\lambda} = 0$$

$$\sum x_i = n\lambda$$

$$\hat{\lambda}_{MLE} = \frac{\sum x_i}{n} = \bar{x}$$

7.67) $\bar{X} \sim \text{Normal}(\mu_{\bar{X}} = \lambda; \sigma_{\bar{X}}^2 = \frac{\lambda}{n})$

a) $P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1 - \alpha$ $z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$

$$-z_{\alpha/2} < \frac{\bar{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} < z_{\alpha/2}$$

$$\hat{\sigma}_{\bar{X}(\text{MLE})}^2 = \frac{\bar{x}}{n}$$

$$-z_{\alpha/2} < \frac{\bar{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} < z_{\alpha/2}$$

$$P\left(\bar{X} - z_{\alpha/2} \cdot \sqrt{\frac{\lambda}{n}} < \lambda < \bar{X} + z_{\alpha/2} \cdot \sqrt{\frac{\lambda}{n}}\right) = 1 - \alpha$$

b) $\bar{X} \pm z_{\alpha/2} \cdot \sqrt{\frac{\lambda}{n}} \Rightarrow 4 \pm 1,96 \sqrt{\frac{4}{100}} : 3,608; 4,392$

(36)

- 7.69 Suppose X_1, \dots, X_n denotes a random sample from the gamma distribution with a known α but unknown β . Find the maximum likelihood estimator of β .

7.69) $X_i \sim \text{Gamma}(\alpha; \beta)$

$$f(x_i) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot x_i^{\alpha-1} e^{-x_i/\beta}$$

$$L(\beta) = \left[\frac{1}{\Gamma(\alpha)\beta^\alpha} \right]^n \cdot \prod_{i=1}^n x_i^{\alpha-1} e^{-\frac{x_i}{\beta}}$$

$$\ln L(\beta) = -n \ln \beta - \alpha \cdot n \cdot \ln \beta + \sum \ln x_i^{\alpha-1} - \frac{\sum x_i}{\beta}$$

$$\frac{\partial \ln L(\beta)}{\partial \beta} = -\frac{n\alpha}{\beta} + \frac{\sum x_i}{\beta^2} = 0$$

$$\frac{\sum x_i}{\beta^2} = \frac{n\alpha}{\beta} \Rightarrow \hat{\beta}_{MLE} = \frac{\sum x_i}{\alpha \cdot n} = \frac{\bar{x}}{\alpha}$$

- 7.73 The number of improperly soldered connections per microchip in an electronics manufacturing operation follows a binomial distribution with $n = 20$ and p unknown. The cost of correcting these malfunctions, per microchip, is

$$C = 3X + X^2$$

Find the maximum likelihood estimate of $E(C)$ if \hat{p} is available as an estimate of p .

MLE has the property; if $\hat{\theta}$ is MLE of θ , then $f(\hat{\theta})$ is MLE of $f(\theta)$. we have,

$$X \sim \text{Binomial}(n=20, p)$$

$$E(X) = np = 20p \quad \text{and} \quad \text{Var}(X) = np(1-p) = 20p(1-p)$$

$$\text{Remember; } \text{Var}(X) = E(X^2) - E^2(X)$$

$$\text{Then; } E(X^2) = \text{Var}(X) + E^2(X) = 20p(1-p) + 400p^2$$

$$\text{Since } E(C) = E(3X + X^2) = 3E(X) + E(X^2)$$

$$\text{MLE of } E(C) \text{ is } \widehat{E(C)}_{MLE} = 3 \cdot 20\hat{p} + 20\hat{p}(1-\hat{p}) + 400\hat{p}^2 = 60\hat{p} + 20\hat{p}(1+19\hat{p})$$

Bayes Estimators.

* Remember from Probability, $P(A|B) = \frac{P(AB)}{P(B)}$

where $P(A|B)$ is probability of A given Event B had occurred. Also remember, if E_1, E_2, \dots, E_K are mutually exclusive and collectively exhaustive events;



$$P(B) = P(E_1 B) + P(E_2 B) + \dots + P(E_k B)$$

$$P(B) = P(E_1) P(B|E_1) + \dots + P(E_k) P(B|E_k)$$

$$P(B) = \sum_{i=1}^k P(E_i) \cdot P(B|E_i)$$

$$\text{Then; } P(E_j|B) = \frac{P(E_j B)}{P(B)} = \frac{P(E_j) P(B|E_j)}{\sum P(E_i) P(B|E_i)}$$

which is the Bayes' Rule.

Ex: In a factory, 30% of products are produced by Machine 1, 20% by Machine 2 and remaining by Machine 3. The defective proportions are 2%, 3% and 1% respectively. If an item is defective, what is the probability it is produced by Machine 2?

Ans:

$$\begin{aligned}
 &\rightarrow P(M_1) = 0,30 \quad P(D|M_1) = 0,02 \quad P(M_2|D) = \frac{0,20 \cdot 0,03}{0,30 \cdot 0,02 + 0,20 \cdot 0,03 + 0,50 \cdot 0,01} \\
 &\rightarrow P(M_2) = 0,20 \quad P(D|M_2) = 0,03 \quad P(M_2|D) = 0,353 \\
 &\rightarrow P(M_3) = 0,50 \quad P(D|M_3) = 0,01
 \end{aligned}$$



* So far, we assumed θ as a constant unknown parameter. Bayes estimator assumes that θ is a random variable with pdf (or pmf) $g(\theta)$. $g(\theta)$ is called prior pdf because it is assigned for θ prior to the collection of data x_1, \dots, x_n .

x_j has pdf (or pmf) $f(x_j|\theta)$. Then, the likelihood function is;

$$f(x_1, \dots, x_n | \theta) = f(x_1, \theta) \cdot \dots \cdot f(x_n, \theta).$$

We think that, having observed the data x_1, \dots, x_n has upgraded our prior pdf to posterior pdf, which is;

$$h(\theta | x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n | \theta) \cdot g(\theta)}{\int_{-\infty}^{\infty} f(x_1, \dots, x_n | \theta) \cdot g(\theta) d\theta}$$

Consider the similarity with Bayes Rule and posterior distribution. Only difference is Σ is replaced by \int when θ has a continuous distribution.

The expected value of posterior distribution is called Bayes Estimator. Namely;

$$\hat{\theta}_{\text{BAYES}} = \int_{-\infty}^{\infty} \theta \cdot h(\theta | x_1, \dots, x_n) d\theta.$$

The procedure to find Bayes Estimator is as follows:

(i) get $g(\theta)$ and $f(x_i|\theta)$ from the question.

Find $f(x_1, \dots, x_n|\theta) = f(x_1|\theta) \cdot \dots \cdot f(x_n|\theta)$.

(Sometimes it is directly given in the question).

(ii) Find ~~unadjusted~~^{posterior} distribution. Here, we sometimes use the fact that $\int_{-\infty}^{\infty} f(x) dx = 1$ if $f(x)$ is pdf of random variable X , to take the integration in the denominator.

$$h(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\theta) \cdot g(\theta)}{\int_{-\infty}^{\infty} f(x_1, \dots, x_n|\theta) \cdot g(\theta) d\theta}$$

$$(iii) \text{ Find } \hat{\theta}_{\text{BAYES}} = E(\theta|x_1, \dots, x_n) = \int_{-\infty}^{\infty} \theta \cdot h(\theta|x_1, \dots, x_n) d\theta$$

7.74 Refer to Example 7.24. Suppose that

$$g(p) = \begin{cases} 2 & 0 \leq p \leq \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

If two items are produced on a given day and one is defective, find the Bayes estimate of p .

7.75 Let Y denote the number of defects per yard for a certain type of fabric. For a given mean λ , Y has a Poisson distribution. But λ varies from yard to yard according to the density function

$$g(\lambda) = \begin{cases} e^{-\lambda} & \lambda > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find the Bayes estimator of λ .

7.74) $g(\theta) = 2 \quad 0 \leq \theta \leq \frac{1}{2} \quad \text{when } \theta = p$

~~BBB~~ $f(x_1, x_2|\theta) = \binom{2}{1} \theta^2 \cdot (1-\theta)^{2-1} = 2\theta(1-\theta)$

$$h \int_0^{1/2} f(x_1, x_2 | \theta) \cdot g(\theta) d\theta = \int_0^{1/2} \theta(1-\theta) \cdot 2d\theta = 4 \left(\frac{\theta^2}{2} - \frac{\theta^3}{3} \right) \Big|_0^{1/2}$$

$$= 4 \cdot \left(\frac{1}{8} - \frac{1}{24} \right) = \frac{8}{24} = \frac{1}{3}$$

(3)

$$h(\theta | x_1, x_2) = \frac{2\theta(1-\theta) \cdot 2}{1/3} = 12\theta(1-\theta) \quad 0 < \theta < 1/2$$

$$\hat{\theta}_{BAYES} = E(\theta | x_1, x_2) = \int_0^{1/2} \theta \cdot 12\theta(1-\theta) d\theta = 12 \int_0^{1/2} (\theta^2 - \theta^3) d\theta$$

$$= 12 \left(\frac{\theta^3}{3} - \frac{\theta^4}{4} \right) \Big|_0^{1/2} = 12 \cdot \left(\frac{1}{24} - \frac{1}{64} \right) = \underline{\underline{0,3225}}$$

7.75) $f(y|\lambda) = \frac{e^{-\lambda} \cdot \lambda^y}{y!}; g(\lambda) = e^{-\lambda} \quad \lambda > 0$

$$f(y|\lambda) \cdot g(\lambda) = \frac{e^{-2\lambda} \cdot \lambda^y}{y!}$$

$$I = \int_0^\infty f(y|\lambda) \cdot g(\lambda) d\lambda = \int_0^\infty \frac{e^{-2\lambda} \cdot \lambda^y}{y!} d\lambda$$

Remember, $X \sim \text{Gamma}(\alpha, \beta)$

$$f(x) = \frac{1}{\Gamma(\alpha) \beta^\alpha} \cdot x^{\alpha-1} \cdot e^{-x/\beta} \quad \text{and} \quad \Gamma(\alpha) = (\alpha-1)!$$

when α is an integer.

if we let $\beta = \frac{1}{2}; \alpha = y+1$, $x^{\alpha-1} \cdot e^{-x/\beta}$ will be a pdf.

$$f(x) = \frac{1}{\Gamma(y+1) \cdot (\frac{1}{2})^{y+1}} \cdot x^y \cdot e^{-2x}$$



$$I = \left(\frac{1}{2}\right)^{y+1} \cdot \int_0^{\infty} \frac{z}{y! \cdot \left(\frac{1}{2}\right)^{y+1}} \cdot \lambda^y \cdot e^{-2\lambda} d\lambda = \left(\frac{1}{2}\right)^{y+1}$$

$= 1$

Then; $h(\lambda|y) = \frac{\int_0^{\infty} f(y|\lambda) \cdot g(\lambda) d\lambda}{\int_0^{\infty} f(y|\lambda) \cdot g(\lambda) d\lambda} = \frac{e^{-2\lambda} \cdot \lambda^{y+1}}{y! \cdot \left(\frac{1}{2}\right)^{y+1}}$

and $\lambda|y \sim \text{Gamma}(\alpha = y+1; \beta = \frac{1}{2})$

$$E(\lambda|y) = \alpha \cdot \beta = (y+1) \cdot \frac{1}{2} = \frac{y+1}{2}$$