CE 93—Engineering Data Analysis
Spring 2019

Hansen, Liu, Zhang

Answer all questions.

Partial credit generously given; show what you know!

Assume that the failure pressure is normally distributed

1. Data for the failure pressures of a certain type roof panels are obtained. A sample of 10 panels yields a mean failure pressure of 3 kPa and a sample standard deviation of 1 kPa. Based on these data:
   a) What is the MLE for the mean failure pressure of this type of roof panel (9 pts)?
   b) What is the 95% two-sided confidence interval for the mean failure pressure, assuming that the population variance ~~and~~ is known to be 1 kPa² (9 pts)?
   c) What is the 95% two-sided confidence interval for the mean failure pressure, assuming that the population variance is unknown (9 pts)?

a) let X denote the failure pressure. $X \sim N(\mu, \sigma^2)$

$$f_X(x_i) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2}\frac{(x_i-\mu_0)^2}{\sigma_0^2}\right)$$

$$L(\mu; \sigma^2, x_1, \dots x_n) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$LL(\mu; \sigma^2, x_1, \dots, x_n) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial LL(\mu; \sigma^2, x_1, \dots, x_n)}{\partial \mu} = +\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i-\mu) = 0$$

$$\Rightarrow \hat{\mu} = \boxed{\frac{1}{n}\sum_{i=1}^{n} x_i} \quad \text{or} \quad \boxed{\hat{\mu}_{MLE} = 3 \text{ kPa}}$$

b) $\bar{x} = 3$ kPa, $\sigma_{\bar{x}} = \sigma / \sqrt{n} = 1/\sqrt{10}$ kPa.

$\therefore$ 95% CI $= \bar{X} \pm Z_{\alpha/2}\sigma_{\bar{X}}$

$Z_{\alpha/2} = 1.96$ for $\frac{\alpha}{2} = 0.025$

$\therefore$ CI $= [3 - 1.96 \times \frac{1}{\sqrt{10}}, 3 + 1.96 \times \frac{1}{\sqrt{10}}] = \boxed{[2.3802, 3.6198]}$

c) $\bar{X} = 3$ kPa, $S = 1$ kPa.

$\therefore$ 95% CI $= \bar{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$

$t_{n-1,\alpha/2} = 2.262$ for $n-1 = 9$, $\frac{\alpha}{2} = 0.025$

$\therefore$ CI $= [3 - 2.262 \times \frac{1}{\sqrt{10}}, 3 + 2.262 \times \frac{1}{\sqrt{10}}]$

$= \boxed{[2.285, 3.7153]}$

2. A Bernoulli trial yields a success with probability p and a failure with probability $1-p$. Let the random variable X be the number of Bernoulli trials required until the first success is obtained. X follows a geometric distribution: $p_X(x)=(1-p)^{x-1}\cdot p$. Let $x_1, x_2, \ldots, x_n$ be a sample drawn from X.
   a) What is the likelihood function, $L(p)$ (10 pts)?
   b) What is the log likelihood function $LL(p)$ (10 pts)?
   c) Derive the MLE for p from b) (10 pts).

(a) $L(p; x_1, \ldots, x_n) = \prod_{i=1}^{n} (1-p)^{x_i-1} \cdot p$

$= (1-p)^{\left(\sum_{i=1}^{n} x_i\right)-n} \cdot p^n$

(b) $LL(p; x_1, \ldots, x_n) = \left[\left(\sum_{i=1}^{n} x_i\right) - n\right] \ln(1-p) + n \ln p$

(c) $\dfrac{\partial LL(p; x_1, \ldots, x_n)}{\partial p} = 0$

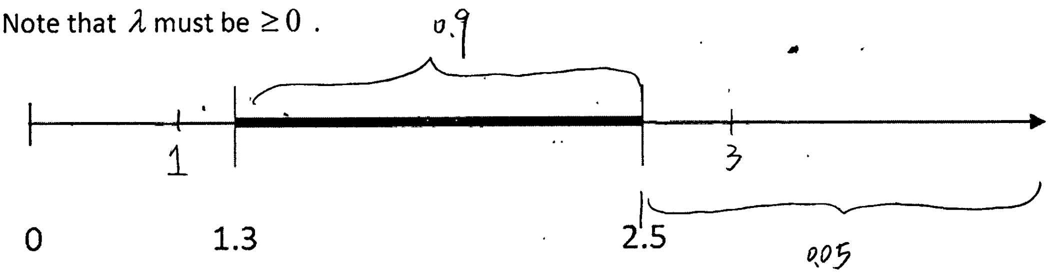
$\left[\left(\sum_{i=1}^{n} x_i\right) - n\right] \cdot \dfrac{-1}{1-p} + \dfrac{n}{p} = 0$

$\dfrac{n - \sum_{i=1}^{n} x_i}{p-p} = \dfrac{n}{p}$

$\cancel{np} - p\sum_{i=1}^{n} x_i = \cancel{np} - n$

$\boxed{p = \dfrac{n}{\sum_{i=1}^{n} x_i}}$

3. Data are obtained for the times, in minutes, between consecutive vehicles crossing a cordon line on a rural road. These times are assumed to follow an exponential distribution. The following represents that 90% (two-sided) confidence interval for the exponential parameter λ based on these data. The confidence interval is the thick portion of the number line between 1.3 and 2.5. Note that λ must be ≥ 0 .



Based on this confidence interval, answer the following questions (3 pts each):

a) What is the probability that λ falls outside the range [1.3, 2.5]?
b) What is the p-value at which we can reject $H0: \lambda = 2.5$ against $H1: \lambda \neq 2.5$.
c) Should we reject $H0: \lambda = 3$ against $H1: \lambda < 3$ at the .10 level? Why or why not? (Remember that in the text the null hypothesis would be $H0: \lambda \geq 3$.)
d) Should we reject $H0: \lambda = 1$ against $H1: \lambda < 1$ at the .10 level? Why or why not? (Remember that in the text the null hypothesis would be $H0: \lambda \geq 1$.)
e) What is the p-value for the hypothesis $H0: \lambda = 1.3$ against $H1: \lambda > 1.3$?(Remember that in the text the null hypothesis would be $H0: \lambda \leq 1.3$.)

a) 10%

b) 10%

c) We should reject the null hypothesis.
$\because$ p-value will be smaller than 5%, thus smaller than the significant level.
$\therefore$ Reject H_0.

d) We should not reject the null hypothesis
$\because$ p-value will be larger than 95%, thus larger than the significant level
$\therefore$ fail to reject. H_0.

e) p-value will be the area to the left of 1.3, $\therefore$ p-value = 5%.

4. Ammonium concentrations (in mg/L) at a large number of wells in the state of Iowa were measured. Of 349 alluvial wells, 182 has concentrations above 0.1. Of 143 quaternary wells, 112 has concentrations above 0.1.
   a) What is the 95% two-sided confidence interval for the proportion of quaternary wells with an ammonium concentration over 0.1? (7 pts)
   b) What is the 80% lower confidence interval for the proportion of alluvial wells with an ammonium concentration over 0.1? (7 pts)
   c) What is the 95% two-sided confidence interval for the difference in the proportions of quaternary wells and alluvial wells that have ammonium concentrations over 0.1? (7 pts)
   d) Test the null hypothesis that the proportions of quaternary wells and alluvial wells that have ammonium concentrations over 0.1 are equal, against the alternative hypothesis that they are not equal, as the 1% level of significance. (7 pts)

Traditional Method :

Given: $\hat{p}_A = \frac{182}{349} = 0.5215$ $\quad \hat{p}_B = \frac{112}{143} = 0.7832$

since there are at least 10 successes and 10 failures, I used 5.8 equation. (Traditional method)

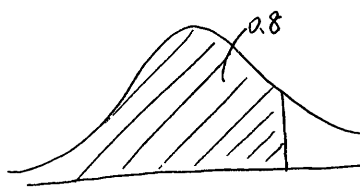
(a) $95\%\ CI = \hat{p}_B \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}$, $\quad z_{\frac{\alpha}{2}} = 1.96 \quad \frac{\alpha}{2} = 0.025$

$= \left[0.7832 - 1.96\sqrt{\frac{0.7832(1-0.7832)}{143}},\ 0.7832 + 1.96\sqrt{\frac{0.7832(1-0.7832)}{143}}\right]$

$= [0.7832 - 1.96 \times 0.03446,\ 0.7832 + 1.96 \times 0.03446]$

$\boxed{= [0.7157, 0.8507]}$

(b)



$80\%\ \text{Lower } CI = \left[0,\ \hat{p}_A + z_\alpha\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A}}\right]$ $\quad z_\alpha = 0.842$ for $\alpha = 0.2$

$= \left[0,\ 0.5215 + 0.842 \times \sqrt{\frac{0.5215(1-0.5215)}{349}}\right]$

$= [0,\ 0.5215 + 0.842 \times 0.02674]$

$\boxed{= [0,\ 0.5440]}$

(c) $95\%\ CI = \hat{p}_A - \hat{p}_B \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}$

① $\hat{p}_A - \hat{p}_B = \frac{182}{349} - \frac{112}{143} = -0.2617$

② $z_{\frac{\alpha}{2}} = 1.96$ for $\frac{\alpha}{2} = 0.025$

③ $\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}} = \sqrt{\frac{0.5215 \times 0.4785}{349} + \frac{0.7832 \times 0.2168}{143}} = \sqrt{0.001902} = 0.04362.$

$\therefore CI = [-0.2617 - 1.96 \times 0.04362,\ -0.2617 + 1.96 \times 0.04362] = \boxed{[-0.3472, -0.1762] \text{ or } [0.1762, 0.3472]}$

(d) $H_0: \hat{p}_A = \hat{p}_B$ $\quad \hat{p}_A = \frac{182}{349} = 0.5215$ $\quad \hat{p} = \frac{182+112}{349+143} = 0.5976$

$H_1: \hat{p}_A \neq \hat{p}_B$ $\quad \hat{p}_B = \frac{112}{143} = 0.7832$

$\alpha = 0.01$

$$z\text{-score} = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})(1/n_A + 1/n_B)}} = \frac{0.5215 - 0.7832}{\sqrt{0.5976 \times 0.4024 \times (\frac{1}{349} + \frac{1}{143})}}$$

$$= \frac{-0.2617}{\sqrt{0.002371}} = -5.3745$$

p-value = sum of the areas in the tails cut off by z and $-z$

$\therefore$ p-value $\approx 0 < 0.01$

$\therefore$ $\boxed{\text{Reject}}$ the null hypothesis.