PSTAT 120B HW 4

Reading Outline

1) - A confidence interval is defined as interval estimators where it is a rule specifying the method for using the sample measurements to calculate two numbers that form the endpoints of the interval.

- They can be used if there is a high probability of enclosing our target parameter, giving us more confidence versus a point estimator which is almost certain to never equal to the target parameter.

2) - A confidence interval is often written to $(\hat{\theta}_L, \hat{\theta}_U)$,

Ex: By definition, a $1-\alpha$ confidence interval is:

$$P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1-\alpha$$

- Upper Confidence Bound is $\hat{\theta}_U$
- Lower Confidence Bound is $\hat{\theta}_L$
- Target Parameter is θ
- Confidence Coefficient is $1-\alpha$

3) - The pivotal quantity is a function of the sample measurements and the only unknown parameter θ

- For a random value Y, and for $c > 0$, then

$P(a \leq Y \leq b) = .7$ then $P(ac \leq Y \leq bc) = .7$

4) - $Z \sim \left(\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}\right)$ has a std normal dist

- Bounds should be as follows: $P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$

- Substitute Z value and solve to get $P(\hat{\theta} - Z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + Z_{\alpha/2}\sigma_{\hat{\theta}}) = 1-\alpha$, which results in a $100(1-\alpha)\%$ confidence interval

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

$\hat{\theta}_1 = Y_1$, $\hat{\theta}_2 = \frac{Y_1 + Y_2}{2}$

$\hat{\theta}_3 = \frac{Y_1 + 2Y_2}{3}$, $\hat{\theta}_4 = \min(Y_1, Y_2, Y_3)$

$\hat{\theta}_5 = \bar{Y}$

a) Unbiased implies $E(\hat{\theta}) - \theta = 0$

$\text{Exp}(Y_i) \Rightarrow E(Y_i) = \theta$

$E(\hat{\theta}_1) = \theta - \theta = 0$

$E(\hat{\theta}_2) = \frac{1}{2}\theta + \frac{1}{2}\theta - \theta = 0$

$E(\hat{\theta}_3) = \frac{1}{3}\theta + \frac{2}{3}\theta - \theta = 0$

$E(\hat{\theta}_4) = \frac{\theta}{n} - \theta \neq 0$

$E(\hat{\theta}_5) = \theta - \theta = 0$

$\boxed{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_5 \text{ are unbiased}}$

b) $\text{Exp}(Y) \Rightarrow Var(Y) = \theta^2$

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

$Var(\hat{\theta}_2) = (\frac{1}{2}\theta)^2 + (\frac{1}{2}\theta)^2 = \frac{1}{2}\theta^2$

$Var(\hat{\theta}_3) = \frac{1}{9}\theta^2 + \frac{4}{9}\theta^2 = \frac{5}{9}\theta^2$

$Var(\hat{\theta}_5) = (\frac{1}{3}\theta)^2 = \frac{1}{9}\theta^2$

$\boxed{\hat{\theta}_5 \text{ has the lowest variance}}$

c) $MSE(\hat{\theta}_1) = \theta^2$

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

$MSE(\hat{\theta}_3) = \frac{5}{9}\theta^2$

$MSE(\hat{\theta}_5) = \frac{1}{9}\theta^2$

(MSE is equal to variance since they are unbiased)

$\boxed{MSE(\hat{\theta}_5) \text{ has the lowest value of all estimators}}$

2) a) $E(n(Y/n)(1 - Y/n))$

$= nE((Y/n)(1 - Y/n))$

$= nE(Y/n) - nE(Y^2/n^2)$

$\Rightarrow np - \frac{1}{n}E(Y^2)$

$= np - \frac{1}{n}((1-p)np + n^2p^2)$

$= np - p(1-p) - np^2$

$np(1-p) \neq p(1-p)(n-1)$

b) In order for $(n-1)(1-p)p$ to become $np(1-p)$, we must multiply by $\frac{n}{n+1}$

Because of this, the estimate should be:

$$\frac{n^2}{n-1}\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right)$$

3.) $Y_1 \dots Y_n \overset{iid}{\sim} f(y)$

$\mathbb{E} Y_i \quad \bar{Y} = \hat{\mu}$

$P(|\bar{Y} - N| > k) \leq \frac{\theta^2}{nk}$

a) $P(\varepsilon > .01) \geq .99$

$$1 - \frac{\theta^2}{nk^2} = .99$$

$$\frac{\theta^2}{n(.01)^2} \leq .01$$

$$\boxed{\frac{\theta^2}{(.01)^3} \leq n}$$

b) i) When $\theta^2 = 1$,

$$\frac{1}{(.01)^3} \leq n$$

$n = 1{,}000{,}000$

ii) When $\theta^2 = 4$,

$$\frac{4}{(.01)^3} \leq n$$

$n = 4{,}000{,}000$

iii) When $\theta^2 = 9$,

$$\frac{9}{(.01)^3} \leq n$$

$n = 9{,}000{,}000$

c) When large population variances occur, the error of estimation should be higher as a result

4) $Z \sim N(1,0)$

$P(|\bar{Y}-\mu| > k) = 1 - P(-k \le \bar{Y}-\mu \le k)$

$= 1 - P\left(-\frac{k\sqrt{n}}{\sigma} \le Z \le \frac{k\sqrt{n}}{\sigma}\right)$

$= 2\Phi\left(-\frac{k\sqrt{n}}{\sigma}\right)$

a) $P\left(-\frac{k\sqrt{n}}{\sigma} \le Z \le \frac{k\sqrt{n}}{\sigma}\right)$

$\Rightarrow P\left(Z \le \frac{k\sqrt{n}}{\sigma}\right) - P\left(Z \le \frac{-k\sqrt{n}}{\sigma}\right)$

$\Rightarrow 1 - P\left(Z < -\frac{k\sqrt{n}}{\sigma}\right) - P\left(Z \le \frac{-k\sqrt{n}}{\sigma}\right)$

$= \boxed{1 - 2\Phi\left(\frac{-k\sqrt{n}}{\sigma}\right)}$

b) $P(|\bar{Y}-\mu| \le k) = 1 - 2\Phi\left(\frac{-k\sqrt{n}}{\sigma}\right)$

For $\varepsilon < .01 = k$, $1 - 2\Phi\left(\frac{-k\sqrt{n}}{\sigma}\right) = .99$

$\Rightarrow 2\Phi\left(\frac{-k\sqrt{n}}{\sigma}\right) = .01$

$\Rightarrow \Phi\left(\frac{-k\sqrt{n}}{\sigma}\right) = .005 \Rightarrow \Phi^{-1}(.005) = \frac{-k\sqrt{n}}{\sigma}$

$\Rightarrow \frac{-k\sqrt{n}}{\sigma} = -2.757 \quad \boxed{n = \left(\frac{2.757\,\sigma}{.01}\right)^2}$

c) $n_1 = \left(\frac{2.757}{.01}\right)^2 = \boxed{76010.49}$

$n_4 = \left(\frac{2.757(2)}{.01}\right)^2 = \boxed{304041.96}$

$n_9 = \left(\frac{2.757(3)}{.01}\right)^2 = \boxed{684094.41}$