| Probability and Statistics for Engineers<br>Midterm 2 | | |
|---|---|---|
| Code : *CVE 303*<br>Acad.Year: *2018-19*<br>Semester : *Spring*<br>Date : *28.04.2019*<br>Time : *13:40*<br>Duration : *110 min* | Last Name:<br>Name: Solutions<br>Student ID: Signature:<br>7 QUESTIONS ON 5 PAGES<br>TOTAL 100 POINTS | #: |

| P1. (20) | P2. (20) | P3. (20) | P4. (20) | P5. (20) | | Total. (100) |
|---|---|---|---|---|---|---|
| | | | | | | |

**1.** *(10×2=20pts)* Short answer questions.

**(A)** If $\mathrm{E}[X]=3$ and $\mathrm{E}[Y]=2$ then what is $\mathrm{E}[X-Y]$?

$E[X-Y] = E[X] - E[Y]$

$= 3 - 2 = \boxed{1}$

**(B)** If $\mathrm{Var}[X]=3$ and $\mathrm{Var}[Y]=2$ then what is $\mathrm{Var}[X-Y]$? (Assume $X,Y$ are indep.)

$Var[X-Y] = Var[X] + (-1)^2 Var[Y]$

$= 3 + 2 = \boxed{5}$

**(C)** What does it mean if an estimator $\hat{\Theta}$ for the parameter $\theta$ is "unbiased"?

$\boxed{E[\hat{\Theta}] = \theta}$

**(D)** Why does the sample variance $s^2$ have "$(n-1)$" in the denominator?

If it had "n" in the denominator then it would be biased.

**(E)** What does it mean for $I$ to be the "$(1-\alpha)$ confidence interval", for the parameter $\theta$?

$\boxed{P(\theta \in I) = 1-\alpha}$

$\hookrightarrow P(\theta \notin I) = \alpha$

**(F)** What parameter is a $\chi^2$ confidence interval used for?

Variance $\boxed{\sigma^2}$

– or –

Standard Deviation $\boxed{\sigma}$

**(G)** When do you use a "$z$-test"?

- If $X$ is Normal with $\sigma$ known

– or –

- If #samples is large (maybe not Norm., $\sigma$ unknown)

**(H)** When do you use a "$t$-test"?

If $X$ is Normal with $\sigma$ unknown

(especially if #samples is not large)

**(I)** What is the difference between a "1-tailed" and "2-tailed" hypothesis test?

(1-tailed p-value) $= \frac{1}{2} *$ (2-tailed p-value)

1-tailed test

$H_0: \theta \geq \theta_0$ | $H_0: \theta \leq \theta_0$

$H_A: \theta < \theta_0$ | $H_A: \theta > \theta_0$

2-tailed test

$H_0: \theta = \theta_0$

$H_A: \theta \neq \theta_0$

**(J)** What probability is $\alpha$ measuring if we say a test has "significance level $\alpha$"?

$\alpha = P(\text{Type I Error})$

$= P(\text{Reject } H_0 \mid H_0 \text{ True})$

$\alpha$ is probability cutoff for p-value $P(\hat{\Theta} > \hat{\theta} \mid H_0 \text{ True})$

**2.** *(8×2=16pts)* The discrete joint random variable $(X, Y)$ has the joint probability mass function given to the right.

| $p(x,y)$ | $x=-1$ | $x=0$ | $x=1$ |
|---|---|---|---|
| $y=-1$ | $\frac{3}{20}$ | $\frac{3}{20}$ | $\frac{2}{20}$ |
| $y=0$ | $\frac{3}{20}$ | $\frac{1}{20}$ | $\frac{1}{20}$ |
| $y=1$ | $\frac{1}{20}$ | $\frac{4}{20}$ | $\frac{2}{20}$ |

Compute the following.

**(A)** The pmf for the statistic $X+Y$.

| $x+y$ | $-2$ | $-1$ | $0$ | $1$ | $2$ |
|---|---|---|---|---|---|
| $P(x+y)$ | $3/20$ | $3/20+3/20 = 6/20$ | $1/20+1/20+2/20 = 4/20$ | $4/20+1/20 = 5/20$ | $2/20$ |

**(A)** The pmf for the statistic $\min(X, Y)$.

| $\min(x,y)$ | $-1$ | $0$ | $1$ |
|---|---|---|---|
| $P(\min(x,y))$ | $1/20+3/20+3/20+3/20+2/20 = 12/20$ | $4/20+1/20+1/20 = 6/20$ | $2/20$ |

**3.** *(2×2=4pts)* The discrete random variable $X$ with the probability mass function given to the right is sampled two times to get $X_1$, $X_2$,

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $p(x)$ | $\frac{1}{6}$ | $\frac{2}{6}$ | $\frac{3}{6}$ |

Compute the following.

**(A)** The pmf for the statistic $T_0 = X_1 + X_2$

$P(T_0=0) = 1/6 \cdot 1/6 = 1/36$

$P(T_0=1) = 2 \cdot 1/6 \cdot 2/6 = 4/36$

$P(T_0=2) = 2 \cdot 1/6 \cdot 3/6 + 2/6 \cdot 2/6 = 10/36$

$P(T_0=3) = 2 \cdot 2/6 \cdot 3/6 = 12/36$

$P(T_0=4) = 3/6 \cdot 3/6 = 9/36$

| $T_0$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $p(t_0)$ | $1/36$ | $4/36$ | $10/36$ | $12/36$ | $9/36$ |

**(A)** The pmf for the statistic $\min(\{X_i\})$.

$P(\min=0) = 2 \cdot 1/6 \cdot 5/6 + 1/6 \cdot 1/6 = 11/36$

$P(\min=1) = 2 \cdot 2/6 \cdot 3/6 + 2/6 \cdot 2/6 = 16/36$

$P(\min=2) = 3/6 \cdot 3/6 = 9/36$

| $\min(\{X_i\})$ | 0 | 1 | 2 |
|---|---|---|---|
| $p(\min)$ | $11/36$ | $16/36$ | $9/36$ |

**4.** *(2×6=12pts)* Suppose $X$ with $\mathrm{E}[X] = 13$ and $\mathrm{Var}[X] = 36$ is sampled 100 times.

**(A)** What is the standard error of $\overline{X}$?

$$\sigma_{\bar{X}} = \sqrt{\sigma_X^2 / n} = \sqrt{36/100} = \boxed{6/10} = 3/5$$

**(B)** What is the (approx.) distribution of $\overline{X}$?

$$\boxed{\bar{X} \approx \text{Normal}(13, 6/10)}$$

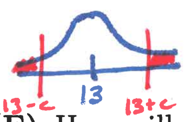
**(C)** How will the standard error of $\overline{X}$ change if $n$ is bigger?

If $n$ increases then

$\sigma_{\bar{X}} = \sqrt{\sigma^2/n}$ decreases.

**(D)** How will $\mathrm{P}(|\overline{X} - 13| > c)$ change if $n$ is bigger?

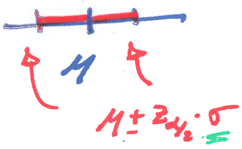
If $\sigma_{\bar{X}}$ decreases then




$P(|\bar{X}-13|>c)$ decreases.

**(E)** How will the $(1-\alpha)$ confidence interval for $\mu$ change if $n$ is bigger?

If $n$ increases then $(\sigma_{\bar{X}} \downarrow)$

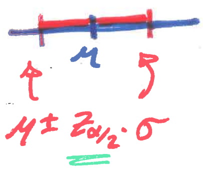



CI shrinks.

more narrow.

**(F)** How will the $(1-\alpha)$ confidence interval for $\mu$ change if $\alpha$ is bigger?

If $\alpha$ increases then $(z_{\alpha/2} \downarrow)$




CI shrinks.

more narrow.

**5.** *(2×4=8pts)* Give a short answer to the following.

**(A)** Write the formula for the sample mean using $X_1, X_2, ..., X_n$.

$$\boxed{\bar{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{\sum X_i}{n}}$$

**(B)** Write the formula for the sample variance using $X_1, X_2, ..., X_n$ and $\overline{X}$.

Similar to formula:

$$\mathrm{Var}[X] = E[(X-\bar{X})^2]$$

$$s^2 = \boxed{\frac{(X_1-\bar{X})^2 + \cdots + (X_n-\bar{X})^2}{n-1}}$$

**(C)** Write the formula for the sample mean using $T_0$.

Recall: $T_0 = X_1 + X_2 + \cdots + X_n$

$$\bar{X} = \boxed{\frac{1}{n} T_0}$$

**(D)** Write the formula for the sample variance using $T_0$ and $T_1$.

Recall: $T_1 = X_1^2 + X_2^2 + \cdots + X_n^2$

Similar to formula:

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2$$

$$s^2 = \boxed{\frac{1}{n-1}\left(T_1 - \frac{1}{n} T_0^2\right)}$$

**6.** *(3×3+11=20pts+6 Bonus)* A sample of size 100 is drawn from population, finding 36 occurrences.

**(A)** What is the point estimate for the population proportion $p$ from the sample data above?

$\hat{p} = X/n$ so $\hat{p} = x/n = \boxed{36/100}$

**(B)** What is the (approximate) standard error of the point estimator for $p$?

$\sigma_{\hat{p}} = \sqrt{\hat{p}\cdot\hat{q}/n}$ $\hat{p} = 36/100$, $\hat{q} = 1 - 36/100 = 64/100$ $\sigma_{\hat{p}} = \sqrt{\dfrac{36/100 \cdot 64/100}{100}} = \boxed{\dfrac{6.8}{1000}}$

This answer is fine, too.

**(C)** What is the (approximate) distribution of the point estimator for $p$?

$\boxed{\hat{P} \approx \text{Normal}\left(36/100, \dfrac{6.8}{1000}\right)}$

**(D)** Give a formula for the (approx.) 95% confidence interval for the population proportion $p$. (Your answer may use R commands such as `qnorm`, `pnorm`, `qt`, or `pt`, etc).

$p = \hat{p} \pm \sigma_{\hat{p}} \cdot z_{\alpha/2}$

$= \boxed{36/100 \pm \dfrac{6.8}{1000} \cdot \text{qnorm}(.025)}$

Note: $\dfrac{6.8}{1000} = \dfrac{64}{1000} = \dfrac{6}{125}$ if you like to simplify...

- **BONUS QUESTION:** It is possible to use the **binomial distribution** to compute a more precise 95% confidence interval for $p$ in the situation above.

The solution begins as follows: We may consider the function "dbinom$(x, 100, p)$" as a two variable function of $x$ and $p$. In fact, since $\int_p \sum_x \text{dbinom}(x, 100, p)\, dp = 1$, this gives a joint probability distribution for $(X, p)$.

**Q1:** Why is $\int_p \sum_x \text{dbinom}(x, 100, p)\, dp = 1$?

Since $\sum_x \text{dbinom}(x,100,p) = 1$, $\int_{p=0}^{p=1} \sum_x \text{dbinom}(x, 100, p)\, dp = \int_0^1 1\, dp = 1$

The confidence interval is found using $P(p \leqslant p_0 \mid X = 36)$

**Q2:** How can you compute $P(p \leqslant p_0 \mid X = 36)$?

$$P(p \le p_0 \mid X=36) = \frac{P(p \le p_0 \,\&\, X=36)}{P(X=36)} = \boxed{\frac{\int_0^{p_0} \text{dbinom}(36,100,p)\,dp}{\int_0^1 \text{dbinom}(36,100,p)\,dp}}$$

Note: Bottom integral $P(X=36) = \frac{1}{101}$

The left and right endpoints of the interval are $p_{\alpha/2}$ and $p_{1-\alpha/2}$.

**Q3:** What properties determine $p_{\alpha/2}$ and $p_{1-\alpha/2}$?

$p_{\alpha/2}$ is value so that $P(p \le p_{\alpha/2} \mid X=36) = \alpha/2$

$\hookrightarrow \int_0^{p_{\alpha/2}} \text{dbinom}(36, 100, p)\, dp = \frac{1}{101}\alpha/2$ ← A computer can easily find this by numerical approx. of integral!!!

Similarly for $p_{1-\alpha/2}$

**7.** *(5×4=20pts)* A normal population is sampled 15 times yielding a sample mean of $\overline{x} = 10$ and a sample standard deviation of $s = 3$. A similar population has mean $\mu = 15$.

Decide which test method should be used and give formulas for the **one-tailed** $p$-value.

**(A)** What are the null and alternative hypotheses (for the one-tailed test)?

Since $\bar{x} = 10 < \mu = 15$ we use hypotheses

$$\boxed{\begin{array}{l} H_0 : \mu \geq 15 \\ H_A : \mu < 15 \end{array}}$$

Note: In terms of test procedures this is equivalent to $\left\{\begin{array}{l} H_0 : \mu = 15 \\ H_A : \mu < 15 \end{array}\right.$

**(B)** Which test should be used and why?

We should use a (one-tailed) t-Test.

The distribution of $\bar{X}$ is Normal with unknown variance.

↳ Sample size n is too small to use z-Test (& Central Limit Thm)

**(C)** What is the test statistic?

Test statistic is sample mean $\bar{x} = \boxed{10}$

which normalizes to $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \boxed{\dfrac{10 - 15}{3/\sqrt{15}}} = \dfrac{-5}{3/\sqrt{15}}$

**(D)** Give a formula for the one-tailed $p$-value.
(Your answer should use R commands such as `qnorm`, `pnorm`, `qt`, or `pt`, etc).

$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{10 - 15}{3/\sqrt{15}} = \dfrac{-5}{3/\sqrt{15}}$ has $\underbrace{(15-1)}_{n-1} = 14$ df.

p-value $= \boxed{pt\left(\dfrac{-5}{3/\sqrt{15}}, 14\right)}$

pt(-5/(3/sqrt(15)), 14)