Sheil Sarda
ESE402 Assignment 1
Prof. Hassani
Sept. 21, 2021

Problem 1

Q1A] Using the rules of counting,

total # of combinations $= (\underbrace{6}_{\text{\# of colors/Peg}})^{4 \leftarrow \text{\# of pegs}} = 1296$

(since we are sampling 4 colors with replacement)

Since only one of the 1,296 unique combinations can be correct

$$P[\text{guessing correctly}] = \frac{1}{1{,}296}$$

Q1B] # of patterns possible when Placing 6 blocks in a line:

3 red, 3 green

6! ways of rearranging 6 distinct objects in a line

Out of these 6!, we must remove permutations that get double counted as a result of the 3 red blocks being interchangable

$$\Rightarrow \frac{6!}{3!} = \frac{720}{6} = 120$$

Similarly, remove permutations that are double-counted because of 3 greens being interchangable

$$= \frac{120}{3!} = \boxed{20} \text{ total patterns}$$

Now, if we add 3 more white blocks to the mix,

$$\text{\# of total patterns} = \frac{9!}{(3!)^3} = \frac{362{,}880}{216}$$

$$= \boxed{1680} \text{ total Patterns}$$

Problem 2A

Q 2A) To compute: $E[X]$

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

$$f(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}$$

$$\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha) \text{ for } \alpha > 0$$

Using the definition of Expected value:

$$E[X] = \int_0^{\infty} x f_x(x)\,dx$$

Substituting the PDF into $E[X]$:

$$= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x \cdot x^{\alpha-1} e^{-x/\beta}\,dx$$

$$= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} (\beta t)^{\alpha} e^{-t}\,d(\beta t) \qquad \text{where } t = \frac{x}{\beta}$$

$$= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} \beta^{\alpha} t^{\alpha} e^{-t} \beta\,dt$$

$$= \frac{\beta^{\alpha+1}}{\Gamma(\alpha)\beta^{\alpha}} \underbrace{\int_0^{\infty} t^{\alpha} e^{-t}\,dt}$$

$$= \frac{\beta}{\Gamma(\alpha)} \cdot \Gamma(\alpha+1) \leftarrow \text{by def}^{n} \text{ of } \Gamma \text{ function's PDF}$$

$$= \frac{\beta\,\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha\beta \cdot \frac{\cancel{\Gamma(\alpha)}}{\cancel{\Gamma(\alpha)}}$$

$$\Rightarrow \boxed{E[X] = \alpha\beta}$$

Problem 2B

Q2B] To Compute: $Var[X]$

We know: $Var[X] = E[X^2] - (E[X])^2$

First, we can compute $E[X^2]$: $\int_0^\infty x^2 f_x(x)\,dx$

$$= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^2 \cdot x^{\alpha-1} e^{-x/\beta}\,dx$$

$$= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-x/\beta}\,dx$$

$$= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty (\beta t)^{\alpha+1} e^{-t} \beta\,dt \qquad \text{where } t = x/\beta$$

$$= \frac{\beta^{\alpha+2}}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \underbrace{t^{\alpha+1} e^{-t}\,dt}$$

$$= \frac{\beta^{\alpha+2}}{\beta^\alpha \Gamma(\alpha)} \cdot \Gamma(\alpha+2)$$

$$= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \beta^2 = E[X^2]$$

Now, we can compute variance as: $E[X^2] - (E[X])^2$

$$= \frac{\Gamma(\alpha+2)\beta^2 - (\alpha\beta)^2 \Gamma(\alpha)}{\Gamma(\alpha)}$$

$$= \frac{\beta^2}{\Gamma(\alpha)}\left(\alpha \cdot (\alpha+1) \cdot \Gamma(\alpha) - \alpha^2 \Gamma(\alpha)\right)$$

$$= \beta^2 (\alpha)(\alpha + 1 - \alpha) = \boxed{\beta^2 \alpha = Var(X)}$$

Problem 2C

Q2C] PMF of $S \sim \chi^2_n$ where $n = \#$ of DoFs,

$$f_s(u) = \frac{1}{2} \cdot \frac{(u/2)^{n/2-1}}{\Gamma(n/2)} \cdot e^{-u/2}$$

Pattern matching the PMF of χ^2 w/ that of the Gamma(α,β)

$\alpha = \frac{n}{2}$ & $\beta = 2$

Thus, per the general case of $n=p$, $\boxed{\alpha = \frac{p}{2} \; ; \; \beta = 2}$

Problem 3

Determining MGF

Q3] Moment - generating function of

$X \sim \text{Bernoulli}(p)$

Use MGF to find (i) mean, (ii) variance

(iii) third moment $E[X^3]$

PMF $\Rightarrow f(x) = p^x (1-p)^{1-x}$

#MGF Given a random variable X, MGF: $M_X(t) = \mathbb{E}[e^{tX}]$

$\Rightarrow$ For Bernoulli, MGF: $\mathbb{E}[e^{t\underset{\text{Bernoulli}}{\underset{\uparrow}{X}}}]$

$= \sum_{n=0}^{1} p^x (1-p)^{1-x} \cdot e^{tx} \; dx$

$\Uparrow$

Σ instead of $\int$ since discrete distribution

$\Rightarrow M_X(t) = E(e^{tX}) = \sum_{n=0}^{1} Pr(X=n)\, e^{tn}$

$\Rightarrow M_X(t) = Pr(X=0)\, e^0 + Pr(X=1)\, e^t$

$\Rightarrow q + pe^t$

$\Rightarrow (1-p) + pe^t$

Mean

(i) Mean: 1^{st} moment

We know 1^{st} derivative of the MGF evaluated @ 0 gives us the 1^{st} moment

$\Rightarrow \frac{d}{dt}\left[(1-p) + pe^t\right] = 0 + \cancel{1}pe^t$

$= \cancel{1}pe^t$ @ $t=0$

$\Rightarrow pe^0 = p$

Variance

(ii) Variance: uses 2nd moment ; $E[X^2] - (E[X])^2$

$\Rightarrow \frac{d}{dt}[pe^t] = pe^t$ @ $t=0 \Rightarrow p$

2nd moment ↓ ↙ 1st moment squared

Thus variance $= p - p^2 = p(1-p)$

Third Moment

(iii) Third moment

$\rightarrow \frac{d}{dt}[pe^t] = \underline{\underline{pe^t}}$ @ $t=0$

$\Rightarrow \boxed{p}$

Problem 4

Q4] Recognize that the total weight of the 100 packages can be modeled by a ~~standard~~ normal distribution with the following parameters:

$$\sum X_i \sim N(3000, 100 \times 15^2)$$
$$= N(\underbrace{3000}_{\mu}, \underbrace{225,000}_{\sigma^2})$$

We wish to find the probability:

$$P\left(\frac{\sum X_i - \mu}{\sigma} > \underbrace{\frac{3500 - 3000}{\sqrt{15^2 \times 10^2}}}_{\text{Z-score}}\right)$$

The Z-score here evaluates to $3.\overline{33}$

Evaluating the probality @ this z-score:

$$P\left[\begin{matrix}\text{total weight of} \\ \text{100 packages} \\ \text{exceeds 3500}\end{matrix}\right] = 1 - 0.9995657701$$
$$= \boxed{0.00043423}$$

Problem 5A and Problem 5B

Q5A] To compute the 95% Confidence Interval,

we need to know: (1) Z-score = 1.96

(2) Standard Error $= \sqrt{\frac{p(1-p)}{n}} = 0.032496$

(3) Sample mean $= p = 0.12$

Thus, $\Pr\Big[\mu \in [\underbrace{\bar{x} - Z_{\frac{\alpha}{2}} \cdot \sigma}_{a}, \underbrace{\bar{x} + Z_{\frac{\alpha}{2}} \cdot \sigma}_{b}]\Big] = 1 - \underset{0.95}{\alpha}$

We can solve for the bounds of the range a, b:

$a = 0.12 - 1.96 \cdot (0.032496)$
$= 0.056308$
$b = 0.12 + 1.96 \cdot (0.032496)$
$= 0.183692$

$\Rightarrow \Pr\big[\mu \in [0.056308, 0.183692]\big] = 0.05$

Q5B] To compute 90% confidence Interval,

(1) Z-score = 1.645

(2) Standard Error $= \sqrt{\frac{p(1-p)}{n}} = 0.0271662$

(3) Sample mean $= p = 0.18$

We can solve for the bounds of the range a, b:
(Similar to Q5A)

$a = 0.18 - 1.645 \cdot (0.0271662) = 0.135312$
$b = 0.18 + 1.645 \cdot (0.0271662) = 0.224688$

$\Pr\big[\mu \in [0.135312, 0.224688]\big] = 0.1$

Problem 5C and Problem 5D

Q5C] To compute: $Var(\hat{d})$

Std. error $(\hat{d})$

i) $Var(\hat{d}) = Var(\hat{p}_1 - \hat{p}_2)$

$= Var(\hat{p}_1) + Var(\hat{p}_2)$ (applying linearity of variance since $\hat{p}_1$ & $\hat{p}_2$ are independent Bernoulli R Vs)

$$\Rightarrow Var(\hat{d}) = \frac{(0.12)(1-0.12)}{100} + \frac{(0.18)(1-0.18)}{200}$$

$$= 0.001056 + 0.000738$$

$$= \boxed{0.001794 = Var(\hat{d})}$$

ii) Standard error $(\hat{d}) = \sqrt{Var(\hat{d})} = \boxed{0.042356}$

Q5D] 99% Confidence Interval:

Z-score = 2.576

Standard Error = 0.042356

Sample mean = $E[\hat{p}_1 - \hat{p}_2] = 0.12 - 0.18 = -0.06$

Thus, similar to Q5A, we can compute the ranges of the CI [a, b]:

$a = -0.06 - 2.576(0.042356) = -0.169109$

$b = -0.06 + 2.576(0.042356) = 0.049109$

Interpretation: Since zero is the null value of the parameter, if the CI includes the null value, we cannot claim any statistically significant difference between $\hat{p}_1$ & $\hat{p}_2$ (2 groups)

95% Confidence Interval

Z-score = 1.96

range [a, b] ⇒ $a = -0.06 - 1.96(0.042356) = -0.143018$
$b = -0.06 + 1.96(0.042356) = 0.023018$

Interpretation: Since zero (null value) is still included in the [a, b] range, we cannot yet claim statistically significant difference between these 2 groups

90% Confidence Interval

Z-score = 1.645

range [a, b] ⇒ $a = -0.06 - 1.645(0.042356) = -0.129676$
$b = -0.06 + 1.645(0.042356) = 0.009676$

Interpretation: similar to the interpretation of the 99% & 95% CI, we still cannot claim any statistically significant difference since the null value (0) is included in the range [a, b].

Problem 6

Q6] To prove: $MSE(\hat{\theta}, \theta) = (bias(\hat{\theta}, \theta))^2 + Var(\hat{\theta})$

$$MSE(\hat{\theta}, \theta) = E[|\hat{\theta} - \theta|^2]$$
$$= E[(\hat{\theta}-\theta)^2] \quad (\because \text{ squaring ~~does~~ Absolute value})$$
$$= E[\hat{\theta}^2 + \theta^2 - 2\hat{\theta}\theta]$$
$$= E[\hat{\theta}^2] + E[\theta^2] - 2\theta E[\hat{\theta}] \quad (\text{Applying LOE})$$
$$= E[\hat{\theta}^2] + \theta^2 - 2\theta E[\hat{\theta}]$$

adding & subtracting this term

$$= \underline{(E[\hat{\theta}])^2} + \theta^2 - 2\theta E[\hat{\theta}] + E[\hat{\theta}^2] - (E[\hat{\theta}])^2$$

$$= \underbrace{(E[\hat{\theta}] - \theta)^2}_{\text{Bias}^2} + \underbrace{E[\hat{\theta}^2] - (E[\hat{\theta}])^2}_{\text{Variance}}$$

By definition, $\Rightarrow \boxed{MSE(\hat{\theta}, \theta) = (Bias(\hat{\theta}, \theta))^2 + Var(\hat{\theta})}$

Problem 7A

Q7A] Std. error of $\hat{p} = \sqrt{\dfrac{p(1-p)}{n}} = 0.043301$

Now, we can simplify the given equation:

$$P(|\hat{p} - p| \geq \delta) \approx 0.025 \quad (\text{converted to 1-tailed})$$
$$\rightarrow P(\hat{p} - p \geq \delta) \approx 0.0125$$
$$\rightarrow P\left(\frac{\hat{p} - p}{\sigma} \geq \frac{\delta}{\sigma}\right) \approx 0.0125 \quad (\text{converting } X \text{ to standard Normal})$$

Now, we can compute a Z-score that yields 0.0125 on the tail

$$\rightarrow Z = 2.2414 = \frac{\delta}{\sigma} \Rightarrow \delta = (2.414)(0.043301)$$
$$= 0.104529$$

Problem 7B

Q7B] Z-Score $= 1.96$
Sample mean $= 0.25$
Standard error $= 0.043301$

~~Q7B~~] Then, we can compute the bounds of the 95%
CI for p as follows:

$a = 0.25 - 1.96(0.043301) = 0.165130$

$b = 0.25 + 1.96(0.043301) = 0.334869$

Clearly, the above range contains the true value of p, which is given as 0.25