APMA 1650 – FINAL EXAM – SOLUTIONS

1. **STEP 1:** Let X be the number of occurrences of `PEYAM`

   X has values $0, 1, 2$ and so

$$E(X) = 0P(X = 0) + 1P(X = 1) + 2P(X = 2) = P(X = 1) + 2P(X = 2)$$

   STEP 2: $P(X = 1)$

   Total: 26^{10} possibilities

   Favorable: There are 6 places to put the string `PEYAM` and 26^5 ways of filling the remaining 5 spots, so $6 \times 26^5 - 1$ favorable possibilities. We subtract 1 because of the case `PEYAMPEYAM`

$$P(X = 1) = \frac{6 \times 26^5 - 1}{26^{10}}$$

   STEP 3: $P(X = 2)$

   Since the only possibility is the string `PEYAMPEYAM`

$$P(X = 2) = \frac{1}{26^{10}}$$

   STEP 4: Answer:

$$E(X) = P(X = 1) + 2P(X = 2) = \left(\frac{6 \times 26^5 - 1}{26^{10}}\right) + 2\left(\frac{1}{26^{10}}\right) = \frac{6 \times 26^5 + 1}{26^{10}}$$

2. **STEP 1:** Let X be the number of months in which the company sends out > 1 birthday email

Let X_i be the following indicator variable

$$X_i = \begin{cases} 1 & \text{if the company sends out} > 1 \text{ email in month } i \\ 0 & \text{otherwise} \end{cases}$$

Then $X = X_1 + \cdots + X_{12}$ and so

$$E(X) = E\left(X_1 + \cdots + X_{12}\right) = E(X_1) + \cdots + E(X_{12})$$

Moreover, for each i we have

$$E(X_i) = 0P(X_i = 0) + 1P(X_i = 1) = P(X_i = 1)$$

STEP 2: $P(X_i = 1)$

By the complement rule, we have

$$\begin{aligned} P(X_i = 1) =& 1 - P(\text{ the company sends out} \leq 1 \text{ email in month } i\ ) \\ =& 1 - \left(P(\text{ 0 emails }) + P(\text{ 1 email })\right) \end{aligned}$$

We can model this with Binom (n, p) where $n = 50$ employees and $p = \frac{1}{12}$ (birthday in month i) hence

$$\begin{aligned} P(X_i = 1) =& 1 - \left(\binom{50}{0}\left(\frac{1}{12}\right)^0\left(\frac{11}{12}\right)^{50} + \binom{50}{1}\left(\frac{1}{12}\right)^1\left(\frac{11}{12}\right)^{49}\right) \\ =& 1 - \left(\frac{11}{12}\right)^{50} - 50\left(\frac{1}{12}\right)\left(\frac{11}{12}\right)^{49} \end{aligned}$$

STEP 3: Answer:

Notice $E(X_i) = P(X_i = 1)$ doesn't depend on i and so

$$E(X) = 12E(X_i) = 12\left(1 - \left(\frac{11}{12}\right)^{50} - 50\left(\frac{1}{12}\right)\left(\frac{11}{12}\right)^{49}\right)$$

3. **STEP 1:** By definition of C we have

$$C = 10 + 2Y$$

Therefore:

$$E(C) = E(10 + 2Y) = E(10) + 2E(Y) = 10 + 2(3) = 16$$
$$\text{Var }(C) = \text{Var }(10 + 2Y) = \text{Var }(2Y) = 2^2 \text{ Var }(Y) = 4(2) = 8$$

STEP 2: By Chebyshev's Inequality, we have

$$P(|C - E(C)| \geq a) \leq \frac{\text{Var }(C)}{a^2} \Rightarrow P(|C - 16| \geq a) \leq \frac{8}{a^2}$$

We want the right-hand-side to be ≤ 0.08 and so it is enough to choose a such that

$$\frac{8}{a^2} \leq 0.08 \Rightarrow \frac{1}{a^2} \leq \frac{0.08}{8} = 0.01 \Rightarrow a^2 \geq \frac{1}{0.01} = 100 \Rightarrow a \geq 10$$

STEP 3: Answer

Therefore our interval is

$$|C - 16| < 10 \Rightarrow -10 < C - 16 < 10 \Rightarrow 6 < C < 26$$

So the desired interval is $(6, 26)$

4. (a) The triangle can be written as $0 \leq y \leq 1-x$ with $0 \leq x \leq 1$

The area of the triangle is $\frac{1}{2} \times 1 \times 1 = \frac{1}{2}$ and therefore

$$f(x,y) = \begin{cases} 2 & \text{if } 0 \leq y \leq 1-x \text{ and } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(b)

$$\text{Cov}\ (X,Y) = E(XY) - E(X)E(Y)$$

$$\begin{aligned} E(XY) &= \int_0^1 \int_0^{1-x} (xy)\ 2\, dy dx \\ &= \int_0^1 x \left(\int_0^{1-x} 2y dy \right) dx \\ &= \int_0^1 x \left[y^2 \right]_{y=0}^{y=1-x} dx \\ &= \int_0^1 x\,(1-x)^2\, dx \\ &= \int_0^1 x - 2x^2 + x^3 dx \\ &= \left[\frac{1}{2}x^2 - \frac{2}{3}x^3 + \frac{1}{4}x^4 \right]_0^1 \\ &= \frac{1}{2} - \frac{2}{3} + \frac{1}{4} \\ &= \frac{3}{4} - \frac{2}{3} \\ &= \frac{9-8}{12} \\ &= \frac{1}{12} \end{aligned}$$

$$f_X(x) = \int_0^{1-x} f(x,y) dy = \int_0^{1-x} 2dy = 2(1-x)$$

$$E(X) = \int_0^1 x\,(2(1-x))\, dx = \int_0^1 2x - 2x^2 dx = \left[x^2 - \frac{2}{3}x^3 \right]_0^1 = 1 - \frac{2}{3} = \frac{1}{3}$$

$$f_Y(y) = \int_0^{1-y} f(x,y)dy = \int_0^{1-y} 2dx = 2(1-y)$$

$$E(Y) = \int_0^1 y\,(2(1-y))\,dy = \frac{1}{3}$$

$$\text{Cov }(X,Y) = E(XY) - E(X)E(Y) = \left(\frac{1}{12}\right) - \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) = \frac{1}{12} - \frac{1}{9} = \frac{3-4}{36} = -\frac{1}{36}$$

(c) Because Cov $(X,Y) \neq 0$, X and Y are **not** independent

5. (a) Let $\overline{X} = \frac{X_1+\cdots+X_n}{n}$ and $\overline{Y} = \frac{Y_1+\cdots+Y_m}{m}$ then

$$\begin{aligned} E\left(\overline{X}\right) =& E\left(\frac{1}{n}(X_1+\cdots+X_n)\right) = \frac{1}{n}\left(E(X_1)+\cdots+E(X_n)\right) \\ =& \frac{1}{n}\left(\lambda+\cdots+\lambda\right) \qquad \text{Because } X_i \sim \text{ Poi } (\lambda) \\ =& \frac{1}{n}\left(n\lambda\right) = \lambda \end{aligned}$$

Similarly $E\left(\overline{Y}\right) = \lambda$

Hence $E(\hat{\lambda}) = E\left(a\overline{X}+b\overline{Y}\right) = aE(\overline{X}) + bE(\overline{Y}) = a\lambda + b\lambda = (a+b)\lambda$

In order for $\hat{\lambda}$ to be unbiased for λ we need $E(\hat{\lambda}) = \lambda$ and so $(a+b)\lambda = \lambda$ and therefore we need

$$a+b=1$$

(b) Because $\hat{\lambda}$ is unbiased, we have

$$\text{MSE }(\hat{\lambda}) = \left[\text{ Bias }(\hat{\lambda})\right]^2 + \text{ Var }(\hat{\lambda}) = \text{ Var }(\hat{\lambda})$$

$$\begin{aligned} \text{Var } \left(\overline{X}\right) =& \text{ Var } \left(\frac{1}{n}(X_1+\cdots+X_n)\right) = \frac{1}{n^2}\left(\text{ Var }(X_1)+\cdots+\text{ Var }(X_n)\right) \\ =& \frac{1}{n^2}\left(\lambda+\cdots+\lambda\right) \qquad \text{Because } X_i \sim \text{ Poi } (\lambda) \\ =& \frac{1}{n^2}\left(n\lambda\right) = \frac{\lambda}{n} \end{aligned}$$

Similarly Var $\left(\overline{Y}\right) = \frac{\lambda}{m}$

Hence Var $(\hat{\lambda}) =$ Var $(a\overline{X}+b\overline{Y}) = a^2$ Var $(\overline{X}) + b^2$ Var $(\overline{Y}) = a^2\left(\frac{\lambda}{n}\right) + b^2\left(\frac{\lambda}{m}\right)$

And therefore we get

$$\text{MSE }(\hat{\lambda}) = \left(\frac{a^2}{n}+\frac{b^2}{m}\right)\lambda$$

(c) From (b) with $m = n$ we have

$$\text{Var }(\hat{\lambda}) = \left(\frac{a^2}{n} + \frac{b^2}{n}\right)\lambda = \frac{1}{n}\left(a^2 + b^2\right)\lambda$$

Since $\hat{\lambda}$ is unbiased and $\lim_{n\to\infty}$ Var $(\hat{\lambda}) \to 0$ we get $\hat{\lambda}$ is consistent for λ

6. **STEP 1:** Since the sample size is small, we have to use the $t-$distribution, and so the 90% confidence interval is

$$[\hat{L}, \hat{U}] = \left[\overline{Y} - t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right), \overline{Y} + t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right]$$

The length of that confidence interval is

$$\left(\overline{Y} + t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right) - \left(\overline{Y} - t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right) = 2\left(t_{\alpha/2}\right)\left(\frac{S}{\sqrt{n}}\right)$$

STEP 2: Plug in the values:

$$1 - \alpha = 0.9 \Rightarrow \alpha = 0.1 \Rightarrow \alpha/2 = 0.05$$

We use the Student's $t-$distribution with $n - 1 = 15$ df and $p = 0.05$ to get $t_{\alpha/2} = 2$

Therefore we get

$$2\left(t_{\alpha/2}\right)\left(\frac{S}{\sqrt{n}}\right) < 4 \Rightarrow 4\left(\frac{S}{\sqrt{16}}\right) < 4 \Rightarrow S < 4$$

Therefore the largest sample standard deviation is $\boxed{S = 4}$

7. STEP 1: Likelihood Function

$$\begin{aligned}
L(Y_1, Y_2, \cdots, Y_n|p) =& p(Y_1)p(Y_2)\cdots p(Y_n) \\
=& \binom{m}{Y_1} p^{Y_1}(1-p)^{m-Y_1} \cdots \binom{m}{n} p^{Y_n}(1-p)^{m-Y_n} \\
=& \left[\binom{m}{Y_1}\cdots\binom{m}{Y_n}\right] p^{Y_1+\cdots+Y_n}(1-p)^{(m+\cdots+m)-(Y_1+\cdots+Y_n)} \\
=& Cp^{n\overline{Y}}(1-p)^{mn-n\overline{Y}} \qquad \text{where } C = \binom{m}{Y_1}\cdots\binom{m}{Y_n} \\
=& Cp^{n\overline{Y}}(1-p)^{n(m-\overline{Y})}
\end{aligned}$$

STEP 2: Logarithms

$$\ln\left(L(Y_1, \cdots, Y_n|p)\right) = \ln\left(Cp^{n\overline{Y}}(1-p)^{n(m-\overline{Y})}\right) = \ln(C) + n\overline{Y}\ln(p) + n(m-\overline{Y})\ln(1-p)$$

STEP 3: Derivative

$$\begin{aligned}
\frac{d}{dp}L(Y_1, \cdots, Y_n|p) =& \frac{d}{dp}\left(\ln(C) + n\overline{Y}\ln(p) + n(m-\overline{Y})\ln(1-p)\right) \\
=& \frac{n\overline{Y}}{p} - \frac{n(m-\overline{Y})}{1-p} = 0
\end{aligned}$$

$$\begin{aligned}
\frac{\cancel{n}\overline{Y}}{p} - \frac{\cancel{n}(m-\overline{Y})}{1-p} =& 0 \\
\frac{\overline{Y}}{p} =& \frac{m-\overline{Y}}{1-p} \\
(1-p)\overline{Y} =& (m-\overline{Y})p \\
\overline{Y} - \cancel{p\overline{Y}} =& mp - \cancel{\overline{Y}p} \\
p =& \frac{\overline{Y}}{m}
\end{aligned}$$

STEP 4: Answer: Therefore the MLE of p is

$$\hat{p} = \frac{\overline{Y}}{m}$$

8. (a) Parameter of interest: $\mu =$ mean age

(1) Alternative Hypothesis: $\mu < 30$
(2) Null Hypothesis: $\mu = 30$
(3) Test statistic: $\overline{Y} = 28$
(4) Rejection Region: $\left\{\overline{Y} \leq k\right\}$

(b) Since we know $\sigma = 5$ we get

$$\hat{\sigma} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{100}} = \frac{5}{10} = 0.5$$

Since the sample size is large we convert to Z

$$P(\overline{Y} \leq k) = 0.05 \Rightarrow P\left(\frac{\overline{Y} - 30}{\hat{\sigma}} \leq \frac{k - 30}{0.5}\right) = 0.05 \Rightarrow P\left(Z \leq \frac{k - 30}{0.5}\right) = 0.05$$

From the problem we get $P(Z \leq -2) = 0.05$ and so

$$\frac{k - 30}{0.5} = -2 \Rightarrow k = 30 + (0.5)(-2) = 30 - 1 = 29$$

Hence our rejection region is $\left\{\overline{Y} \leq 29\right\}$

Answer: Since the test statistic $\overline{Y} = 28$ is in the rejection region, since $28 \leq 29$ we reject the null hypothesis, and so our claim is supported with a level of $\alpha = 0.05$

(c) Here the observed value is $\overline{Y} = 28$

$$\begin{aligned} p\text{ -value } &= P(\overline{Y} \leq 28 \text{ given Null is true }) \\ &= P(\overline{Y} \leq 28 \text{ given } \mu = 30) \\ &= P\left(\frac{\overline{Y} - 28}{\hat{\sigma}} \leq \frac{28 - 30}{0.5}\right) \\ &= P(Z \leq -4) \end{aligned}$$

(d)

$$
\begin{aligned}
\beta =& P(\overline{Y} \text{ is outside RR when Alt is true })\\
=& P(\overline{Y} \geq k) \text{ given } \mu = 26\\
=& P(\overline{Y} \geq 29) \text{ given } \mu = 26\\
=& P\left(\frac{\overline{Y} - 26}{\hat{\sigma}} \geq \frac{29 - 26}{0.5}\right)\\
=& P(Z \geq 6)
\end{aligned}
$$