STAT 150 HW #1 Chae Yeon Lee

problem 1

$X \sim$ Poisson (λ)

$Y \sim$ Poisson (μ)

a) $Z = X + Y$.

$$P(Z = S) = \sum_{K=0}^{S} P(X=K, Y=Z-K) = \sum_{K=0}^{S} e^{-\lambda} \frac{\lambda^K}{K!} \cdot e^{-\mu} \frac{\mu^{S-K}}{(S-K)!} = e^{-(\mu+\lambda)} \frac{1}{S!} \sum_{K=0}^{S} \frac{S!}{K!(S-K)!} \lambda^K \mu^{S-K}$$

$$= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^S}{S!}$$

Hence, $Z \sim$ Poisson $(\lambda+\mu)$.

$\Rightarrow$ by independence of X and Y.

b) $$P(X=K \mid Z=m) = \frac{P(X=K, Z=m)}{P(Z=m)} = \frac{P(X=K, Y=m-K)}{P(Z=m)} = \frac{P(X=K)P(Y=m-K)}{P(Z=m)}$$

$$= \frac{\frac{\cancel{e^{-\lambda}} \lambda^K}{K!} \frac{\cancel{e^{-\mu}} \mu^{m-K}}{(m-K)!}}{\frac{\cancel{e^{-(\lambda+\mu)}} (\lambda+\mu)^m}{m!}} = \frac{m!}{(m-K)!\,m!} \left(\frac{\lambda}{\lambda+\mu}\right)^K \left(\frac{\mu}{\lambda+\mu}\right)^{m-K}$$

this is binomial distribution with parameters. m, $p = \frac{\lambda}{\lambda+\mu}$

problem 2

a) If independent, $P(X=1 \cap Y=3) = P(X=1)P(Y=3)$. However, $P(X=1 \cap Y=3) = 1/6$

$P(X=1)P(Y=3) = \frac{1}{4} \cdot 1/3 = 1/12$

$P(X=1) = \frac{1}{12} + \frac{1}{6} = \frac{1}{4}$

$P(X=1 \cap Y=3) = 1/6 \neq P(X=1)P(Y=3)$

$P(Y=3) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

hence X and Y are not independent.

b) $E[X] = 1(1/12 + 1/6) + 2(1/6 + 1/3) + 3(1/12 + 1/6) = 1/4 + 2(1/2) + 3(1/4) = \boxed{2}$

$E[Y] = 2(1/12 + 1/6 + 1/12) + 3(1/6 + 1/6) + 4(1/3) = 2(1/3) + 3(1/3) + 4(1/3) = \boxed{3}$

$\mathrm{var}(X) = E(X^2) - E(X)^2 = 4.5 - 2^2 = \boxed{0.5}$

$E[X^2] = 1(1/12 + 1/6) + 4(1/6 + 1/3) + 9(1/12 + 1/6) = 1/4 + 4(1/2) + 9(1/4) = 14.5$

$\mathrm{var}(Y) = E(Y^2) - E(Y)^2 = \frac{29}{3} - 9 = \boxed{\frac{2}{3}}$

$E(Y^2) = 4(1/3) + 9(1/3) + 16(1/3) = \frac{4+9+16}{3} = \frac{29}{3}$

$\mathrm{cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = E((X-1)(Y-3))$

$= 1/2(1-1)(2-3) + 1/6(2-1)(2-3) + 1/12(3-1)(2-3) + 1/6(1-1)(3-3) + 1/6(3-1)(3-3) + 1/3(2-1)(4-3)$

$= -1/6 - 1/6 + 1/3 = \boxed{0}$

$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \boxed{0}$

c) $P(X=1) = 1/4$ $\quad P(X=1 \cap Y=2) = 1/4 \cdot 1/3 = 1/12 = P(X=1 \cap Y=3) = P(X=1 \cap Y=4)$

$P(X=2) = 1/2$ $\quad P(X=2 \cap Y=2) = P(X=2 \cap Y=3) = P(X=2 \cap Y=4) = 1/2 \cdot 1/3 = 1/6$

$P(X=3) = 1/4$ $\quad P(X=3 \cap Y=2) = P(X=3 \cap Y=3) = P(X=3 \cap Y=4) = 1/4 \cdot 1/3 = 1/12$

$P(Y=2) = 1/3$

$P(Y=3) = 1/3$

$P(Y=4) = 1/3$

y \ X	1	2	3
2	1/12	1/6	1/12
3	1/12	1/6	1/12
4	1/12	1/6	1/12

Problem 3

a) $$\sigma_{combined} = \sqrt{\frac{n_1(\sigma_1^2 + (\mu_1 - \mu_{combined})^2) + n_2(\sigma_2^2 + (\mu_2 - \mu_{combined})^2)}{n_1 + n_2}}$$

$$= \sqrt{\frac{n_1(100 + (\mu_1 - \mu_{com})^2) + n_2(100 + (\mu_2 - \mu_{com})^2)}{n_1 + n_2}}$$

$\mu_{combined} \in [\mu_1, \mu_2]$ since $\mu_{comb.}$ is also a weighted mean of μ_1 and μ_2.

Hence,

$$n_1(100 + (\mu_1 - \mu_{com})^2) + n_2(100 + (\mu_2 - \mu_{com})^2) > 100(n_1 + n_2)$$

$$\underbrace{\frac{n_1(100 + (\mu_1 - \mu_{com})^2) + n_2(100 + (\mu_2 - \mu_{com})^2)}{n_1 + n_2}} > 100$$

$\boxed{\therefore \sigma_{combined} > \sqrt{100} = 10.}$ $\quad$ (C) is the correct answer.

b) $$s_{total}^2 = \frac{30(10^2 + (75 - \bar{x}_{total})^2) + 20(100 + (60 - \bar{x}_{total})^2)}{30 + 20} = [30(100 + (75-69)^2) + 20(100 + (60-69)^2)]/50$$

$$= 154$$

$$\bar{x}_{total} = \frac{30 \cdot 75 + 20 \cdot 60}{50} = 69$$

$s_{total} = \sqrt{154} = \boxed{12.4096}$ $\quad$ $\boxed{12.41}$

problem 4

a) $S_n = w_1 + w_2 + \cdots + w_n$ $\quad i \le i \le n$

S_n is the sum of n independent copies of w_i with $i \in [1, n]$ where each w_i has geometric (p) distribution.

S_n is the time slot when nth success occurs in the sequence of iid coin tosses.

$S_n = K$ means nth success on the kth trial.

b) $P(S_n = K)$ for $K \ge 0$?

$$= \binom{K-1}{n-1} p^{n-1} \cdot (1-p)^{K-n} \cdot p$$

explanations by term

$\binom{K-1}{n-1}$: since kth toss must be a success, up to K-1 slots, there can be n-1 successes. The combinations are counted by $\binom{K-1}{n-1}$.

p^{n-1} : There must be n-1 successes in the first k-1 tosses.

$(1-p)^{K-n}$: in the total of K tosses, all tosses except K should be all failures.

p : the kth toss must be a success.

c) $E[S_n] = E[\sum_i x_i] = \frac{n}{p}$

each geometric distribution's expectation tells the expected # of tosses for 1 success. since we need n successes, and each w_i are independent of each other,

$E[S_n] = n \cdot E[w_i] = n \cdot 1/p$

$\text{var}(S_n) = \frac{n(1-p)}{p^2}$ because w_i are independent so we can multiply the $\text{var}(w_i) = \frac{1-p}{p^2}$ by n.

d) $P(S_n \ge 4n/p) \le \frac{E(S_n) \cdot p}{4n} = \frac{\cancel{n}}{p} \cdot \frac{\cancel{p}}{4\cancel{n}} = \boxed{\frac{1}{4}}$

e) $P(S_n \ge 4n/p)$ by Chebyshev

$$P(|S_n - E(S_n)| \ge \frac{4n}{p} - E(S_n)) \le \frac{\text{var}(S_n)}{(\frac{4n}{p} - E(S_n))^2} = \frac{\frac{n(1-p)}{p^2}}{(\frac{4n}{p} - \frac{n}{p})^2} = \frac{\frac{n(1-p)}{\cancel{p^2}}}{\frac{9n^2}{\cancel{p^2}}} = \boxed{\frac{1-p}{9n}}$$

f) since n is large, the distribution will follow a normal distribution.

$$P(S_n \ge 4n/p) = 1 - P(S_n < 4n/p) = 1 - \Phi\left(\frac{\frac{4n}{p} - \frac{n}{p}}{\sqrt{\frac{n(1-p)}{p^2}}}\right) = 1 - \Phi\left(\frac{\frac{3n}{p}}{\frac{\sqrt{n(1-p)}}{p}}\right) = \boxed{1 - \Phi\left(\frac{3n}{\sqrt{n(1-p)}}\right)}$$

9) $P(S_n \le (1+\epsilon)\frac{2n}{p})$

since weak law of large numbers applies to the sample mean, we need to convert S_n to $\overline{W_n}$.

[1]

For $P(S_n \le (1-\epsilon)\frac{2n}{p})$

$$\boxed{P\left(\frac{S_n}{n} \le \frac{(1\mp\epsilon)\cdot 2}{p}\right)} = P\left(\overline{W_n} - \frac{1}{p} \le \underbrace{\frac{(1\mp\epsilon)2}{p} - \frac{1}{p}}_{\epsilon'}\right)$$ For the law to hold, $\epsilon' \ge 0$.

~~HET~~

$$= P\left(\overline{W_n} - 1/p \le \frac{1}{p}(1-2\epsilon)\right)$$

① $\epsilon < 1/2 \Rightarrow$ by the law of large num, if $\epsilon < 1/2$, then $\epsilon' > 0$.

$$P(\overline{W_n} - 1/p \le \epsilon') = 1 - P(\overline{W_n} - 1/p > \epsilon') = 1 - 0 = \boxed{1}$$

$$\lim_{n\to\infty} P(|\overline{W_n} - 1/p| > \epsilon') = 0$$. therefore,

② $\epsilon = 1/2$ apply CLT.

$$P\left(\overline{W_n} - 1/p \le \frac{1}{p}(1-\cancel{2}\cdot 1/2)\right) = P(W_n - 1/p \le 0) = \Phi(0) = \boxed{0.5}$$
$$= P(Z \le 0)$$

③ $\epsilon > 1/2$.

$$P\left(\frac{\overline{W_n} - 1/p}{\sqrt{\frac{1-p}{np^2}}} \le \frac{(1-2\epsilon)\frac{1}{p}}{\sqrt{\frac{1-p}{np^2}}}\right) = \Phi\left(\frac{(1-2\epsilon)\cancel{1/p}}{\sqrt{\frac{1-p}{n\cancel{p^2}}}}\right) = \Phi\left(\frac{1-2\epsilon}{\sqrt{\frac{1-p}{n}}}\right)$$

since $n \to \infty$, $\Phi(\infty) = \boxed{1}$

[2] $P(S_n \le (1-\epsilon)\frac{2n}{p})$

$= P(\frac{S_n}{n} \le (1-\epsilon)\frac{2}{p})$ same rule applies as in [1] $P(\frac{S_n}{n} \le (1+\epsilon)\frac{2}{p})$ except the condition $\epsilon < 1/2$ and $\epsilon > 1/2$ are flipped.

same reasoning as the 3 cases above

① $\epsilon < 1/2$: $\Phi(\infty) = 1$

② $\epsilon = 1/2$: 0.5

③ $\epsilon > 1/2$: $\lim_{n\to\infty} P(|\overline{W_n} - 1/p| > \epsilon') = 0$

$1 - 0 = \boxed{1}$

② $P(S_n \le (1+\epsilon)\frac{2n}{p}) = P(\bar{\omega}_n - \frac{1}{p} \le \frac{(1+\epsilon)2}{p} - \frac{1}{p}) = P(\bar{\omega}_n - \frac{1}{p} \le \underbrace{\frac{1+2\epsilon}{p}}_{\epsilon'})$

since $\epsilon > 0$, $\frac{1+2\epsilon}{p} > 0$.

Hence, by the LLN,

$$\lim_{n\to\infty} P(|\bar{\omega}_n - 1/p| \le \epsilon') = 1 - \lim_{n\to\infty} P(|\bar{\omega}_n - 1/p| > \epsilon') = 1 - 0 = \boxed{1}$$

problem 5

a)

$F_x(x)$ is a continuous, non-decreasing (strictly increasing function). This guarantees that $F^{-1}(x)$ exists.

$Y = F(X)$

$F_Y(y) = P[Y \le y]$

$= P[F(X) \le y] = P[X \le F^{-1}(y)] = F[F^{-1}(y)] = y$

$F_Y(y) = y$ suggests that it is uniform (0,1).

b) ~~since~~ we can apply the inverse of the uniform cdf to F to get the distribution of the random variable X.

$Y = F(X) \Rightarrow P(Y \le xy)$

problem 6.

$MSE_c = \frac{1}{n}\sum_{i=1}^{n}(x_i-c)^2$ → convex function. so, where the der. is 0, it will be the function's minimum.

$\frac{d}{dc}[\frac{1}{n}\sum_{i=1}^{n}(x_i-c)^2] = \frac{1}{n}\sum_{i=1}^{n} 2(x_i-c)(-1) = \frac{2}{n}\sum_{i=1}^{n} c-x_i = \frac{2}{n}[nc - \sum_{i=1}^{n} x_i]$

at minimum, $\frac{d}{dc}[\frac{1}{n}\sum_{i=1}^{n}(x_i-c)^2) = 0$. so, $nc - \sum_{i=1}^{n} x_i$ should equal to 0.

$nc - \sum_{i=1}^{n} x_i = 0 \iff c = \frac{\sum_{i=1}^{n} x_i}{n} = \mu$

To get the $MSE_\mu = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$ This is precisely the definition of variance.

problem 7

a) False.

$\bar{x}_n = \frac{\sum_{i=1}^{n} x_i}{n}$ since we don't know exactly which value each x_i will take, we don't know if $\bar{x}_n$ will necessarily ~~tat~~ equal to $3-4\theta$.

~~Singular batch of~~.

A single batch of samples may not necessarily equal the mean of the random variable.

Here's the counter example. if $n=3$, and $x_1 = x_2 = x_3 = 1$, then $\bar{x}(n) = 1$.

but $E(X) = 3-4\theta$ where $0<\theta<1/3$.

Then, $5/3 < E(X) < 3$, and $E(X)$ cannot ~~ever~~ equal to $\bar{x}(n) = 1$.

b) True.

$$E\left[\frac{\sum_{i=1}^{n} x_i}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = E[X_i]$$ ✓

$$E(X_i) = E(X)$$

c) False; if n is large, then by the law of large number, $\bar{x}(n)$ would follow the normal distribution. With a continuous distribution, the probability that a random variable $\bar{x}(n)$ is exactly equal to a certain value is 0.

d) True. since the sum follows the normal distribution by the law of large numbers, on average, $\bar{x}(n)$ would equal to the mean of the normal distribution of X which is $E[X]$.

problem 8 $E[X] = \theta + 2(2\theta) + 3(1-3\theta) = 3-4\theta$

$\bar{x}_n = 3-4\theta$

$\bar{x}_n - 3 = -4\theta$

$\theta = \frac{3-\bar{x}_n}{4}$

→ $\bar{x}_n$ is ~~not~~ random variable, ~~and~~ ~~it is unbiased.~~ so the estimator $\hat{a}$ is random variable.

→ It is unbiased.

→ It is $\hat{a}$ consistent estimator because by LLN, $\bar{x}_n$ will tend to $\mu_{\bar{x}}$.

$$E(X^2) = \theta + 4(2\theta) + 9(1-2\theta) = 9 - 18\theta$$

standard error of estimator $\qquad var(X) = 9 - 18\theta - (3-4\theta)^2 = \underline{6\theta - 16\theta^2}$

$$\bar{x}_n = \frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow var(\bar{x}_n) = \frac{1}{n^2} \sum_{i=1}^{n} var(x_i) = \frac{1}{n}(6\theta - 16\theta^2)$$

$$var(\theta) = var\left(\frac{\bar{X}_n}{4}\right) = \frac{1}{16} var(\bar{x}_n) = \frac{1}{16n}(6\theta - 16\theta^2)$$

$$SE(\sigma) = \sqrt{\frac{6\theta - 16\theta^2}{16n}}$$