1. (a) **Solution:**
   The number of users should be a poisson distribution which is
   $X \sim Poi(\lambda)$ where $\lambda = 5.5$
   The probability that more than 7 users will sign-up for the social networking site in the next miniute is
   $$P(E) = 1 - P(E^c) = 1 - P\{X \leq 7\} = 1 - \sum_{i=0}^{7} \frac{e^{-5.5}5.5^i}{i!}$$
   Answer:
   We can also use normal distribution to approximate the probability
   $$P\{X > 7.5\} = 1 - \Phi(\frac{7.5-5.5}{\sqrt{5.5}}) = 1 - 0.8032 = 0.1968$$

   (b) **Answer:**
   The probability that more than 13 users will sign-up for the social networking site in the next 2 minute is
   $$P\{X > 13.5\} = 1 - P\{X \leq 13.5\} = 1 - \Phi(\frac{13.5-5.5\times2}{\sqrt{5.5\times2}}) = 1 - 0.7734 = 0.2266$$

   (c) **Answer:**
   The probability that more than 15 users will sign-up for the social networking site in the next 3 minute is
   $$P\{X > 15.5\} = 1 - P\{X \leq 15.5\} = 1 - \Phi(\frac{15.5-5.5\times3}{\sqrt{5.5\times3}}) = \Phi(0.2462) = 0.5987$$

2. (a) **Solution:**
   $$\int_0^1 \int_0^x c\frac{y}{x}dydx = \int_0^1 \frac{1}{2}c\frac{x^2}{x}dx = \frac{1}{4}cx^2\big|_0^1 = \tfrac{1}{4}c = 1$$
   Answer:
   $c = 4$

   (b) **Solution:**
   $$f_X(x) = \int_0^x 4\frac{y}{x}dy = 2c\frac{y^2}{x}\big|_0^x = 2x$$
   $$f_Y(y) = \int_y^1 4\frac{y}{x}dx = cy\ln(x)\big|_y^1 = -4y\ln(y)$$
   $$f_{XY}(x,y) \neq f_X(x)f_Y(y)$$
   Answer:
   X and Y are not independent

   (c) **Answer:**
   $f_X(x) = 2x$

   (d) **Answer:**
   $f_Y(y) = -4y\ln(y)$

   (e) **Answer:**
   $$E[X] = \int_0^1 2x^2dx = \tfrac{2}{3}x^3\big|_0^1 = \tfrac{2}{3}$$

   (f) **Answer:**
   $$E[Y] = \int_0^1 -4y^2\ln(y)dy = -4(\ln(y)\int_0^1 y^2dy - \int_0^1 (\frac{1}{y}\frac{1}{3}y^3)dy)$$
   $$E[Y] = -4(\ln(y)\frac{1}{3}y^3 - \frac{1}{9}y^3)\big|_0^1 = \tfrac{4}{9}$$

3. (a) **Answer:**

$$P\{X_1 + X_2 \geq 5000\} = P\{\frac{X_1+X_2-2200\times2}{\sqrt{52900\times2}} \geq \frac{5000-2200\times2}{\sqrt{52900\times2}}\} = 1 - \Phi(1.84) = 0.033$$

To be more accurate, we should use > 4999.5 instead of ≥ 5000. The result in this particular case has no significant difference.

(b) **Solution:**

$$P\{X \geq 2000\} = 1 - \Phi(\frac{2000-2200}{\sqrt{52900}}) = \Phi(0.87) = 0.808$$

Answer:

The probability that the weekly number of visitors exceeds 2000 in at least 2 of the next 3 weeks is

$$\binom{3}{2}p^2(p-1) + \binom{3}{1}p^3 = 0.904$$

4. (a) **Answer:**

$$A \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

(b) **Answer:**

$$B \sim N(5\mu_1 + 2, 5\sigma_1^2)$$

(c) **Solution:**

$$C_a = aX$$
$$C_b = -bY$$
$$C_c = c^2Z$$
$$C_a \sim N(a\mu_1, a\sigma_1^2)$$
$$C_b \sim N(-b\mu_1, b\sigma_2^2)$$
$$C_c \sim N(c^2\mu_1, c^2\sigma_3^2)$$

Answer:

$$C \sim N(a\mu_1 - b\mu_2 + c^2\mu_3, a\sigma_1^2 + b\sigma_2^2 + c^2\sigma_3^2)$$

5. (a) **Answer:**

$$f_{XY}(x,y) = \begin{cases} \frac{1}{6}\frac{1}{x} & x \in \{1,2,3,4,5,6\}, y \leq x \\ 0 & \text{Otherwise} \end{cases}$$

(b) **Answer:**

$$P(X = j | Y = i) = \begin{cases} \frac{1}{6}\frac{1}{j} & i \leq j \\ 0 & i > j \end{cases}$$

(c) **Solution:**

$$f_X(x) = \begin{cases} \frac{1}{6} & x \in \{1,2,3,4,5,6\} \\ 0 & x \notin \{1,2,3,4,5,6\} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{6}\sum_{i=y}^{6}\frac{1}{y} & y \in \{1,2,3,4,5,6\} \\ 0 & y \notin \{1,2,3,4,5,6\} \end{cases}$$

$$f_{XY}(x,y) \neq f_X(x)f_Y(y)$$

Answer:

X and Y are NOT independent

6. **Solution:**

Since the expectation of the distance are symmetrical when the package is dropped off in the 4 different quadrants. Let D' = the distance the robot travels to get to the package in first quadrants.

$$f'_D(d) = \frac{1}{4}(\frac{1}{5})^2$$

$$E[D'] = \frac{1}{100} \int_0^5 \int_0^5 (x+y)dxdy = \frac{1}{100} \int_0^5 \frac{25}{2} + 5ydy = \frac{125}{100}$$

Answer:

$E[D] = 4E[D'] = 4\frac{125}{100} = 5$

7. **Solution:**

The probability that "max update" is executed in each loop $p_i = P(E_i) = \frac{1}{i+1}$, because the probability that all of the previous independent numbers is

$\frac{i!}{(i+1)!} = \frac{1}{i+1}$

Let

$I_i = \begin{cases} 1 & E_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$

Answer:

The expected number of times that "max update" is executed is

$$E[X] = \sum_{i=0}^{n-1} E[I_i] = \sum_{i=0}^{n-1} p_i = \sum_{i=1}^{n} \frac{1}{i}$$

8. (a) **Solution:**

$E[X] = \frac{a}{a+b} = 0.5$

$Var[X] = \frac{ab}{(a+b)^2(a+b+1)}$

Answer:

$a = b = 4$

(b) **Solution:**

$X|13$ flips resulting in 8 heads and 5 tails $\sim Beta(4+8, 4+5)$

Answer:

$f(x) = \begin{cases} \frac{1}{B(12,9)} x^{11}(1-x)^8 & \text{if} 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$

(c) **Solution:**

$a' = 4 + 8 = 12$

$b' = 4 + 4 = 8$

Answer:

$E[X|12 \text{ flips resulting in 8 heads and 4 tails}] = \frac{a'}{a'+b'} = 0.6$

(d) **Answer:**

$Var(X|12 \text{ flips resulting in 8 heads and 4 tails}) = \frac{a'b'}{(a'+b')^2(a'+b'+1)} = 0.0114$

9. **Solution:**

By post processing the 2 data file we can find

For randomFlips1.txt

$P_1\{Flip = Head\} = 0.49$

$P_1\{Flip = Tail\} = 0.51$

$P_1\{Flip = Tail—previous flip = head\} = 0.694$

$P_1\{Flip = Tail—previous flip = Tail\} = 0.329$

$P_1\{Flip = Head—previous flip = head\} = 0.306$

$P_1\{Flip = Head—previous flip = Tail\} = 0.671$

From the previous results, we can find that the flips are not independent to the previous flip result.

For example

Let

$P_1(H') = P_1\{\text{previous flip} = \text{head}\}$
$P_1(HH') = P_1(H|H')P(H') = 0.306 \times 0.49 = 0.15$
$P_1(H)p_1(H') = 0.51 \times 0.51 = 0.26 \neq P_1(HH')$
For randomFlips2.txt
$P\{\text{Flip} = \text{Head}\} = 0.503$
$P\{\text{Flip} = \text{Tail}\} = 0.497$
$P\{\text{Flip} = \text{Tail—previous flip} = \text{head}\} = 0.51$
$P\{\text{Flip} = \text{Tail—previous flip} = \text{Tail}\} = 0.5$
$P\{\text{Flip} = \text{Head—previous flip} = \text{head}\} = 0.49$
$P\{\text{Flip} = \text{Head—previous flip} = \text{Tail}\} = 0.5$
$P_2(HH') = P_2(H|H')P(H') = 0.25$
$P_2(H)p_2(H') = P_2(HH')$
For the 2nd file, we can find that the flips are independent to the previous flip result.
Answer:
Sequence 2 is a better random generator, because of the independence between flip results in row.

10. (a) **Answer:**

$E[X] = 7.41$

$E[Y] = 8.03$

(b) **Answer:**

$E[X^2] = 58.84$

$E[Y^2] = 68.17$

(c) **Solution:**
$\mu_A = 7.41$, $\sigma_A^2 = E[X^2] - E[X]^2 = 3.98$
$\mu_B = 8.03$, $\sigma_B^2 = E[Y^2] - E[Y]^2 = 3.68$
Answer:

$X \sim N(7.41, 3.98)$

$Y \sim N(8.03, 3.68)$

(d) **Solution:**
Given writer is A, for each keystroke in email, the probability that the duration equals to the timing information is $P(X_i = x_i|A) = P(E_i|A) = \frac{1}{\sqrt{2\pi}\sigma_A}e^{-(x_i-\mu_A)^2/2\sigma_A^2}\epsilon$
$\ln(P(E_i|A)) = \ln(\frac{\epsilon}{\sqrt{2\pi}\sigma_A}) - (x_i - \mu_A)^2/2\sigma_A^2$
So, the probability that the email is written given writer is A is
$\ln(P(E|A)) = \ln(\prod\limits_{i=1}^{n} P(E_i|A) = \sum\limits_{i=1}^{n} \ln(P(E_i|A))$
$\ln(P(E|A)) = n\ln(\frac{\epsilon}{\sqrt{2\pi}\sigma_A}) - \sum\limits_{i=1}^{n}(x_i - \mu_A)^2/2\sigma_A^2 = -248.57$
The probability that the email is written given writer is B is
$\ln(P(E|B)) = n\ln(\frac{\epsilon}{\sqrt{2\pi}\sigma_B}) - \sum\limits_{i=1}^{n}(x_i - \mu_A)^2/2\sigma_B^2 = -243.80$

The probability that the writer is A given the email has been written is
$P(A|E) = \frac{P(AE)}{P(E)} = \frac{P(E|A)P(A)}{P(E)}$
The probability that the writer is B given the email has been written is
$P(B|E) = \frac{P(E|B)P(B)}{P(E)}$
$\frac{P(A|E)}{P(B|E)} = \frac{0.5P(E|A)}{0.5P(E|B)} = \frac{P(E|A)}{P(E|B)}$
$\frac{P(A|E)}{P(B|E)} = e^{\ln(P(E|A))-\ln(P(E|B))}$

Answer:

$$\frac{P(A|E)}{P(B|E)} = e^{-4.77} = 0.0085$$

11. (a) **Answer:**

$$P\{X > 7.5\} = 1 - P\{X \le 7.5\} = 1 - \sum_{i=0}^{7} P\{X = i\}$$

$$\boxed{P\{X > 7.5\} = 0.06}$$

(b) **Solution:**

$$\int_{7.5}^{30} \frac{K}{x} dx = 1$$

$$K \ln(x)\big|_{0.75}^{30} = 0.06$$

$$\int_{7.5}^{30} \frac{C}{x^3} dx = 1$$

$$-C \frac{1}{2x^2}\big|_{0.75}^{30} = 0.06$$

Answer:

$$\boxed{K = 0.043}$$

$$\boxed{C = 7.2}$$

(c) **Answer:**

$$P_1\{X > 10\} = 0.043 \ln(x)\big|_{10}^{30} = 0.047$$

$$P_2\{X > 10\} = -7.2 \frac{1}{2x^2}\big|_{10}^{30} = 0.032$$

(d) **Solution:**

$$E_1[X] = \sum_{i=0}^{7} P\{X = i\}i + \int_{7.5}^{30} \frac{K}{x} x \, dx$$

$$E_1[X] = 3.02 + 0.043x\big|_{7.5}^{30} = 3.99$$

$$E_2[X] = \sum_{i=0}^{7} P\{X = i\}i + \int_{7.5}^{30} \frac{C}{x^3} x \, dx$$

$$E_2[X] = 3.02 - 7.2\frac{1}{x}\big|_{7.5}^{30} = 3.74$$

Answer:

$$\boxed{E_1[S] = 3.99}$$

$$\boxed{E_2[S] = 3.74}$$

(e) **Solution:**

$$E_1[X^2] = \sum_{i=0}^{7} P\{X = i\}i^2 + \int_{7.5}^{30} \frac{K}{x} x^2 \, dx$$

$$E_1[X^2] = 12.06 + \frac{0.043}{2}x^2\big|_{7.5}^{30} = 30.2$$

$$E_2[X^2] = \sum_{i=0}^{7} P\{X = i\}i^2 + \int_{7.5}^{30} \frac{C}{x^3} x^2 \, dx$$

$$E_2[X^2] = 12.06 + 7.2 \ln(x)\big|_{7.5}^{30} = 22.04$$

Answer:

$$\boxed{E_1[S^2] = 30.2}$$

$$\boxed{E_2[S^2] = 22.04}$$