

1. On average, 7.5 users sign-up for an online social networking site each minute. What is the probability that:

- (a) More than 10 users will sign-up for the social networking site in the next minute?

X_1 is the number of users that sign-up in a minute

$$X_1 \sim \text{Poi}(\lambda_1) \text{ where } \lambda_1 = 7.5$$

$$P(X_1 > 10) = 1 - \sum_{i=0}^{10} \frac{\lambda_1^i}{i!} e^{-\lambda_1} = .138$$

- (b) More than 15 users will sign-up for the social networking site in the next 2 minutes?

X_2 is the number of users that sign-up in 2 minutes

$$X_2 \sim \text{Poi}(\lambda_2) \text{ where } \lambda_2 = 2 * 7.5 = 15$$

$$P(X_2 > 15) = 1 - \sum_{i=0}^{10} \frac{\lambda_2^i}{i!} e^{-\lambda_2} = .432$$

- (c) More than 20 users will sign-up for the social networking site in the next 3 minutes?

X_3 is the number of users that sign-up in 3 minutes

$$X_3 \sim \text{Poi}(\lambda_3) \text{ where } \lambda_3 = 3 * 7.5 = 22.5$$

$$P(X_3 > 20) = 1 - \sum_{i=0}^{10} \frac{\lambda_3^i}{i!} e^{-\lambda_3} = .653$$

2. Let X be a continuous random variable with probability density function:

$$f_X(x) = \begin{cases} c(2 - 2x^2) & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

- (a) What is the value of c in order for $f_X(x)$ to be a valid probability density function?

$$\int_{-1}^1 c(2 - 2x^2) dx = 1$$

$$c(2x - \frac{2}{3}x^3) \Big|_{-1}^1 = 1$$

$$c(\frac{4}{3} - \frac{-4}{3}) = 1$$

$$c = \frac{3}{8}$$

- (b) What is the cumulative distribution function (CDF) of X ?

$$\begin{aligned} F_X(a) &= P(X \leq a) = \int_{-\infty}^a f_X(x) dx = \int_{-1}^a c(2 - 2x^2) dx \\ &= c(2x - \frac{2}{3}x^3) \Big|_{-1}^a = c(-\frac{2}{3}a^3 + 2a + \frac{4}{3}) \\ &= -\frac{a^3}{4} + \frac{a}{4} + \frac{1}{6} \end{aligned}$$

(c) What is $E[X]$?

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-1}^1 xc(2 - 2x^2) dx \\ &= c \left(x^2 - \frac{x^4}{4} \right) \Big|_{-1}^1 = 0 \end{aligned}$$

3. Say we have a cable of length n . We select a point (chosen uniformly randomly) along the cable, at which we cut the cable into two pieces. What is the probability that the shorter of the two pieces of the cable is less than $\frac{1}{3}$ the size of the longer of the two pieces?

X is the length of the first piece of cable

$$\begin{aligned} X &\sim \text{Uni}(\alpha, \beta) \text{ where } \alpha = 0, \beta = n \\ f(x) &= \begin{cases} \frac{1}{\beta - \alpha} & \text{when } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases} \\ P(X < \frac{n}{3}) + P(X > \frac{2n}{3}) &= 2 \frac{\frac{n}{3}}{n} = \frac{2}{3} \end{aligned}$$

4. Let X be a normal (Gaussian) random variable with $\mu = 6$. If $P(X > 9) = 0.3$, what is the approximate value of $\text{Var}(X)$?

$$\begin{aligned} P(X > 9) &= 1 - P(X \leq 9) = .3 \\ 1 - \Phi\left(\frac{x - \mu}{\sigma}\right) &= .3 \\ \sigma &\approx 5.75 \\ \text{Var}(X) &= \sigma^2 \approx 33.06 \end{aligned}$$

5. The median of a continuous random variable having cumulative distribution function F is the value m such that $F(m) = 0.5$. That is, a random variable is just as likely to be larger than the median as it is to be smaller. Find the median of X in terms of the respective distribution parameters in each case.

(a) $X \sim \text{Uni}(a, b)$

$$\begin{aligned} P(X < m) &= P(X > m) = .5 \\ \frac{m - \alpha}{\beta - \alpha} &= .5 \\ m &= \frac{\alpha + \beta}{2} \end{aligned}$$

(b) $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} P(X < m) &= P(X > m) = .5 \\ m &= \mu \end{aligned}$$

(c) $X \sim \text{Exp}(\lambda)$

$$\begin{aligned} F(m) &= .5 \\ 1 - e^{-\lambda m} &= .5 \\ m &= -\frac{\ln(.5)}{\lambda} \end{aligned}$$

6. Let X_i be the number of weekly visitors to a web site in week i , where $X_i \sim N(2200, 52900)$ for all i . Assume that all X_i are independent of each other.

- (a) What is the probability that the total number of visitors to the web site in the next two weeks exceeds 5000?

Y is the number of visitors to the web site in a 2 week period

For X_i , $\sigma^2 = 52900 \Rightarrow \sigma = 230$

$$Y = 2X_i \sim N(2\mu, 2^2\sigma^2)$$

$$Y \sim N(4400, 211600)$$

$$P(Y > 5000) = 1 - \Phi\left(\frac{x - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{5000 - 4400}{460}\right) = 9.606 \times 10^{-2}$$

- (b) What is the probability that the total number of visitors exceeds 2000 in at least 2 of the next 3 weeks?

Z_i is the event that the total number of visitors exceeds 2000 in i of 3 weeks

$$P(Z_2 \cup Z_3) = P(Z_2) + P(Z_3)$$

$$P(Z_2) = \binom{3}{2} P(X_i > 2000)^2 P(X_i \leq 2000)$$

$$P(Z_3) = P(X_i > 2000)^3$$

$$P(X_i > 2000) = 1 - \Phi\left(\frac{x - \mu}{\sigma}\right) = .808$$

$$P(X_i \leq 2000) = \Phi\left(\frac{x - \mu}{\sigma}\right) = .192$$

$$P(Z_2) = .376$$

$$P(Z_3) = .527$$

$$P(Z_2 \cup Z_3) = .903$$

7. The joint probability density function of continuous random variables X and Y is given by:

$$f_{X,Y}(x,y) = \begin{cases} c \frac{y}{x} & \text{where } 0 < y < x < 1 \end{cases}$$

- (a) What is the value of c in order for $f_{X,Y}(x,y)$ to be a valid probability density?

$$\int_0^1 \int_0^x c \frac{y}{x} dy dx = 1$$

$$\int_0^1 \frac{c}{2} x dx = 1$$

$$\frac{c}{4} = 1 \Rightarrow c = 4$$

- (b) Are X and Y independent?

No, because the ranges of X and Y depend on each other.

- (c) What is the marginal density function of X ?

$$\begin{aligned} f_X(a) &= \int_0^a \frac{c}{a} y dy \\ &= \frac{c}{2a} y^2 \Big|_0^a \\ &= \frac{c}{2} a = 2a \end{aligned}$$

(d) What is the marginal density function of Y ?

$$\begin{aligned} f_Y(b) &= \int_b^1 cb \frac{1}{x} dx \\ &= cb \ln|x| \Big|_b^1 \\ &= -cb \ln(b) = -4b \ln(b) \end{aligned}$$

(e) What is $E[X]$?

$$\begin{aligned} E[X] &= \int_0^1 x f_X(x) dx \\ &= \int_0^1 2x^2 dx \\ &= \frac{2}{3} x^3 \Big|_0^1 = \frac{2}{3} \end{aligned}$$

(f) What is $E[Y]$?

$$\begin{aligned} E[Y] &= \int_0^1 y f_Y(y) dy \\ &= \int_0^1 -4y^2 \ln(y) dy \\ &= -\frac{4}{3} y^3 \ln(y) - \int_0^1 -\frac{4}{3} y^3 \frac{1}{y} dy && \text{by integration by parts} \\ &= -\frac{4}{3} y^3 \ln(y) + \int_0^1 \frac{4}{3} y^2 dy \\ &= \left[-\frac{4}{3} y^3 \ln(y) + \frac{4}{9} y^3 \right]_0^1 \\ &= \frac{4}{9} \end{aligned}$$

8. Let X , Y , and Z be independent random variables, where $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, and $Z \sim N(\mu_3, \sigma_3^2)$.

(a) Let $A = X + Y$. What is the distribution (along with parameter values) for A ?

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

(b) Let $B = 4X + 3$. What is the distribution (along with parameter values) for B ?

$$B \sim N(4\mu_1 + 3, 16\sigma_1^2)$$

(c) Let $C = aX - b^2Y + cZ$, where a , b , and c are real-valued constants. What is the distribution (along with parameter values) for C ?

$$\begin{aligned} \sum_{i=1}^N X_i &\sim N\left(\sum_{i=1}^N \mu_i, \sum_{i=1}^N \sigma_i^2\right) \\ C &\sim N(a\mu_1 - b^2\mu_2 + c\mu_3, a^2\sigma_1^2 + b^4\sigma_2^2 + c^2\sigma_3^2) \end{aligned}$$

9. Say we have a coin with unknown probability X of coming up heads when flipped. However, we believe (subjectively) that the prior probability (before seeing the results of any flips of the coin) of X is a beta distribution, where $E[X] = 0.5$ and $\text{Var}[X] = \frac{1}{36} \approx 0.0278$

- (a) What are the values of the parameters a and b (where $a, b > 1$) of the prior beta distribution for X ?

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

$$Var[X] = \frac{ab}{(a+b)^2(a+b+1)} = \frac{1}{36}$$

$$a = b = 4$$

- (b) Now say we flip the coin 12 times, obtaining 8 heads and 4 tails. What is the form (and parameters) of the posterior distribution (X | 12 flips resulting in 8 heads and 4 tails)?

$$a = a_{prior} + \text{number of successes (heads)} = 4 + 8 = 12$$

$$b = b_{prior} + \text{number of failures (tails)} = 4 + 5 = 9$$

$$X \sim \text{Beta}(12, 9)$$

- (c) What is $E[X | 12 \text{ flips resulting in 8 heads and 4 tails}]$?

$$E[X] = \frac{a}{a+b} = \frac{12}{21} = \frac{4}{7} \approx .571$$

- (d) What is $E[X | 12 \text{ flips resulting in 8 heads and 4 tails}]$?

$$Var[X] = \frac{ab}{(a+b)^2(a+b+1)} = \frac{6}{539} \approx .011$$

10. Say we have an array of n doubles, "arr[n]" (indexed from 0 to $n-1$), which contains independent and identically distributed non-negative real values (where each value in the array is unique). What is the expected number of times that "max update" (as noted by the comment in the code) is executed in the function (assuming the function is passed the array "arr" and its size is n)? Give an expression for the expectation.

```
double max(double arr[], int n)
{
    double max = -1; // note: all elements in arr[] are > -1
    for (int i = 0; i < n; i++)
    {
        if (arr[i] > max)
        {
            max = arr[i];
        }
    }
    return max;
}
```

max is updated whenever the i th number is the maximum of the subarray from 0 to i

X_i is the event that the i th number is the maximum of the subarray from 0 to i

$N_{updates}$ is the expected number of updates to the variable "max"

$$P(X_i) = \frac{1}{i+1}$$

$$N_{updates} = \sum_{i=0}^{n-1} 1 * \frac{1}{i+1}$$

11. [coding] Below are two sequences of 300 "coin flips" (H for heads, T for tails). One of these is a true sequence of 300 independent flips of a fair coin. The other was generated by a person typing out

Sequence 1:

Sequence 2:

A better way to gauge randomness is to analyze runs. The Wald-Wolfowitz test is a method for calculating the probability that a sequence has a given number of runs.

R is the number of runs in the sequence

n_1 is the number of occurrences of the first value, n_H number of heads in this case

n_2 is the number of occurrences of the other value, n_T number of tails in this case

$$P(R = r = 2k) = \frac{2 \binom{n_1-1}{k-1} \binom{n_2-1}{k-1}}{\binom{n_1+n_2}{n_1}} \quad \text{when } r \text{ is even}$$

$$P(R = r = 2k + 1) = \frac{2 \binom{n_1-1}{k} \binom{n_2-1}{k-1} + \binom{n_2-1}{k} \binom{n_1-1}{k-1}}{\binom{n_1+n_2}{n_1}} \quad \text{when } r \text{ is odd}$$

For Sequence 1,

$r = 216$

$$n_H = 148$$

$n_T = 257$

$$P(r = 216) = 5.248 \times 10^{-4}$$

For Sequence 2,

$$r = 148$$

$$n_H = 148$$

$n_T = 257$

$$P(r = 148) = 2.950 \times 10^{-3}$$

Sequence 2 is 5.62 times more likely to be random based on runs