STAT 391 Homework 2

Chongyi Xu
University of Washington
STAT 391 Spring 2018
chongyix@uw.edu

---

1. Problem 1 - CDF's and densities

Let

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x^2, & 0 < x \leq 1 \\ 1, & 1 < x \end{cases}$$

and

$$G(x) = \begin{cases} 0, & x \leq 0 \\ 2x^2, & 0 < x \leq 0.5 \\ 1 - 2(1-x)^2, & 0.5 < x \leq 1 \\ 1, & 1 < x \end{cases}$$

be two cumulative distribution functions.

1. Plot F, G

```
import numpy as np
import matplotlib.pyplot as plt

x = np.arange(0,1,step=0.01)
F = [xx**2 for xx in x]
G = np.empty([len(x), 1])
for xx in x:
    if xx <= 0.5:
        G[x==xx] = 2 * xx**2
    elif xx <= 1:
```

```
        G[x==xx] = 1 - 2 * (1 - xx)**2
plt.figure()
plt.plot(x, F, 'r--', label='F(x)')
plt.plot(x, G, 'g^', label='G(x)')
plt.xlabel('x')
plt.title('F(x) and G(x)')
plt.legend()
plt.show()
```

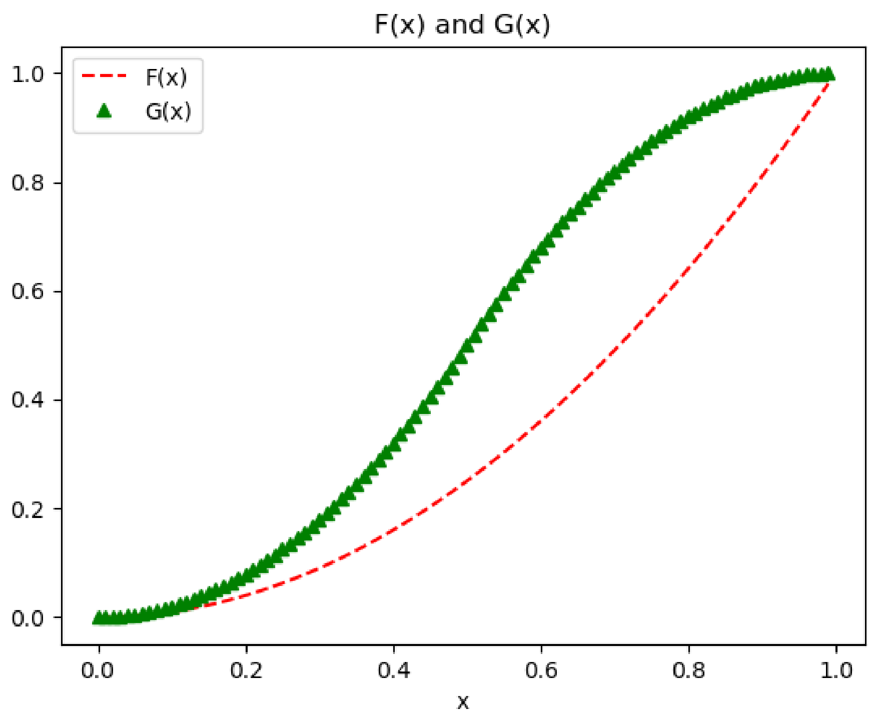


Figure 1: CDF: F, G

2. Compute their corresponding desensitize f, g. Plot them on a graph.

$$f_X(x) = \frac{d}{dx}F_X(x) \qquad = \begin{cases} 0, & x \leq 0 \\ 2x, & 0 < x \leq 1 \\ 0, & 1 < x \end{cases}$$

$$g_X(x) = \frac{d}{dx}G_X x \qquad = \begin{cases} 0, & x \leq 0 \\ 4x, & 0 < x \leq 0.5 \\ 4(1-x), & 0.5 < x \leq 1 \\ 0, & 1 < x \end{cases}$$

```
f = [2*xx for xx in x]
g = np.empty([len(x), 1])
for xx in x:
    if xx <= 0.5:
        g[x==xx] = 4 * xx
    elif xx <= 1:
        g[x==xx] = 4 * (1-xx)
plt.figure()
plt.plot(x, f, 'r--', label='f(x)')
plt.plot(x, g, 'g^', label='g(x)')
plt.xlabel('x')
plt.title('Density f(x) and g(x)')
plt.legend()
plt.show()
```

3. Denote by P_F and P_G the probability distribution defined by F, G. Find $a, a^{'}$ such that $P_F(0, a) = P_F(a, 1)$ and $P_G(0, a^{'}) = P_G(a^{'}, 1)$.

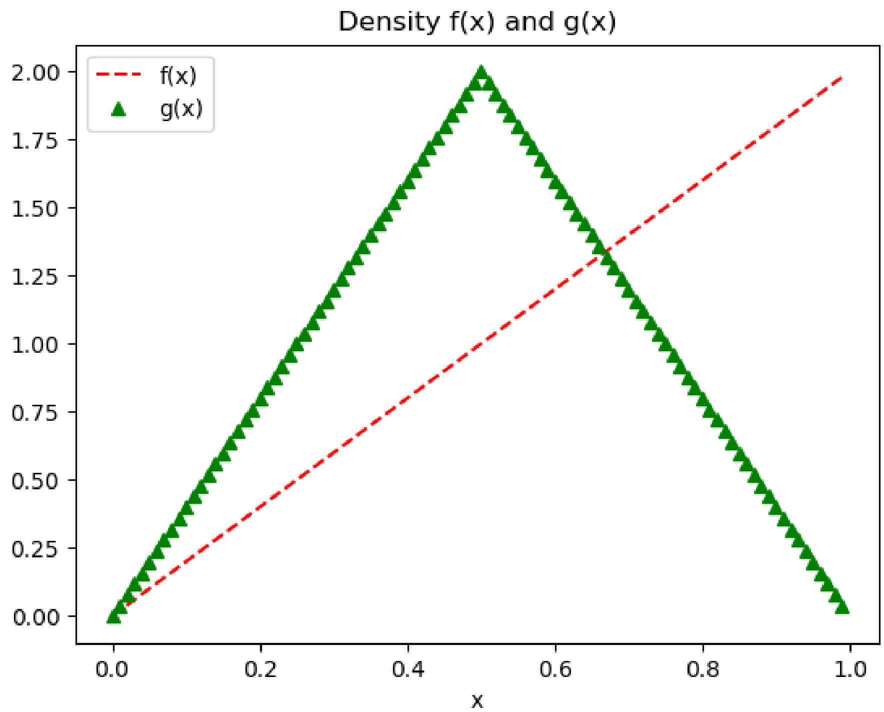


Figure 2: pdf: f, g

Since $P(\alpha, \beta) = P_X(x \leq \beta) - P_X(x \leq \alpha)$. Therefore,

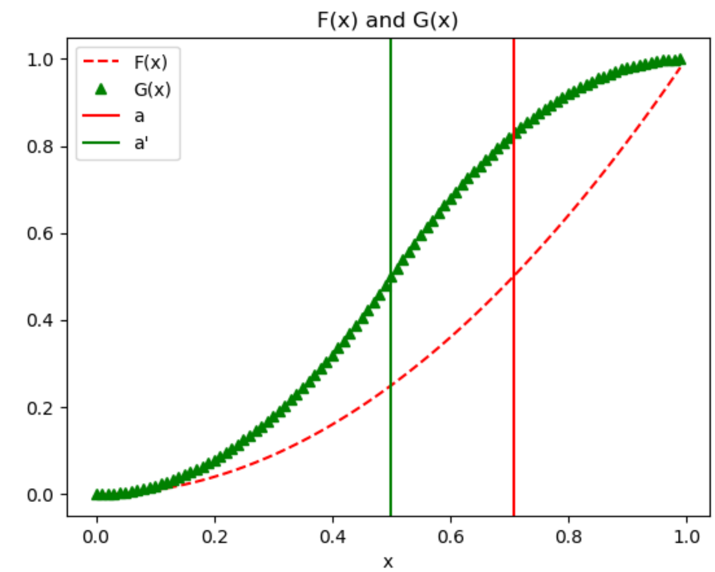
$$\begin{aligned}
P(0, a) &= P(a, 1) \\
F_X(a) - F_X(0) &= F_X(1) - F_X(a) \\
\Rightarrow 2F_X(a) &= F(0) + F(1) = 1 \\
F_X(a) &= \frac{1}{2} \\
a &= \frac{\sqrt{2}}{2} \\
G_X(a^{'}) - G_X(0) &= G_X(1) - G_X(a^{'}) \\
\Rightarrow 2G_X(a^{'}) &= 1 \\
G_X(a^{'}) &= \frac{1}{2} \\
a^{'} &= \frac{1}{2}
\end{aligned}$$

From the plot Figure 3b, we could see that the left part and right part has about same area under the curve on the sides of a and $a^{'}$, which is a way to verify my a and $a^{'}$ are correct.

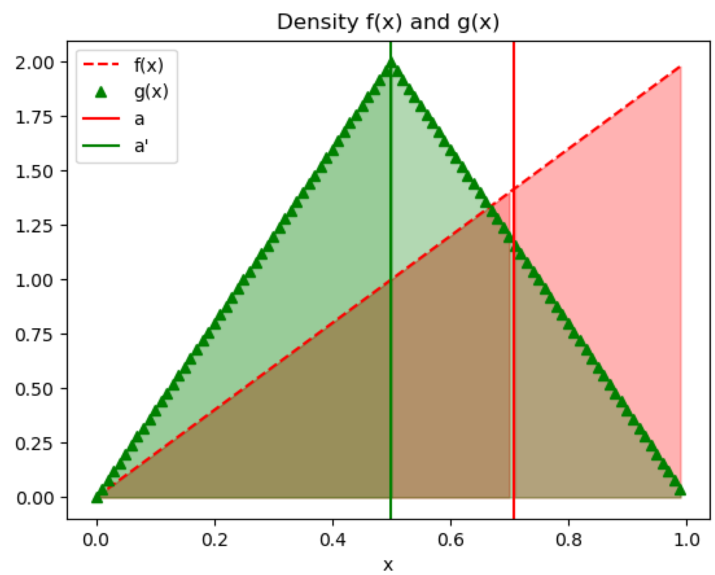
4. Find the probabilities of the following intervals $[0, 0.25]$, $[1, 1.75]$ under P_F, P_G.

$$\begin{aligned}
P_F(x \in [0, 0.25]) &= F_X(x = 0.25) - F_X(x = 0) \\
&= \frac{1}{16} \\
P_G(x \in [0, 0.25]) &= G_X(x = 0.25) - G_X(x = 0) \\
&= \frac{2}{16} = \frac{1}{8} \\
P_F(x \in [1, 1.75]) &= 1 - 1 = 0 = P_G(x \in [1, 1.75])
\end{aligned}$$

5. Find the shortest interval $[a_F, b_F]$ that has probability 0.1 under F. Find $[a_G, b_G]$ under G.



(a) CDF with a and a'



(b) PDF with a and a'

Figure 3: Plot with a and a'

```python
[af, ag, bf, bg] = [-float('inf'), -float('inf'),
    float('inf'), float('inf')]
x = np.arange(0,1,step=0.001)
F = np.array([xx**2 for xx in x])
G = np.empty([len(x), 1])
for xx in x:
    if xx <= 0.5:
        G[x==xx] = 2 * xx**2
    elif xx <= 1:
        G[x==xx] = 1 - 2 * (1 - xx)**2
G = G[:,0]
for i in range(len(x)):
    xa = x[i]
    for j in range(i+1, len(x)):
        xb = x[j]
        if (F[x==xb] - F[x==xa] == 0.1) and\
             (xb - xa < bf - af):
            [af, bf] = [xa, xb]
        if (G[x==xb] - G[x==xa] == 0.1) and\
            (xb - xa < bg - ag):
            [ag, bg] = [xa, xb]

print('[a_f, b_f]='+str([af,bf]))
print('[a_g, b_g]='+str([ag,bg]))
```

So within a step of 0.001, I found my shortest interval to be

$$[a_F, b_F] = [0.075, 0.325]$$
$$[a_G, b_G] = [0.025, 0.225]$$

```python
plt.figure()
plt.plot(x, f, 'r--', label='f(x)')
plt.plot(x, g, 'g^', label='g(x)')
plt.axvline(x=af, color='r', linestyle='solid')
plt.axvline(x=bf, color='r', linestyle='solid')
plt.axvline(x=ag, color='g', linestyle='solid')
plt.axvline(x=bg, color='g', linestyle='solid')
plt.fill_between(x, 0, f, where=np.logical_and(x>=af, x<=bf),
    color='red', alpha=0.3)
```

```
plt.fill_between(x, 0, g, where=np.logical_and(x>=ag, x<=bg),
    color='green', alpha=0.3)
plt.xlabel('x')
plt.title('Density f(x) and g(x)')
plt.legend()
plt.show()
```

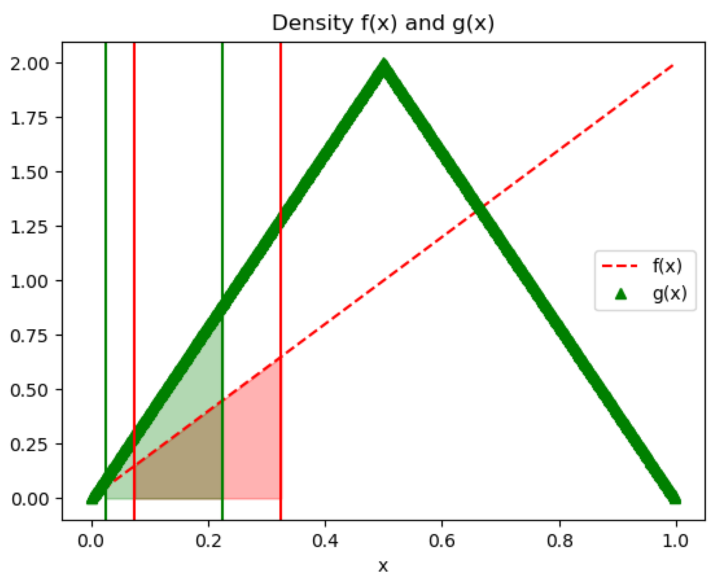


Figure 4: Intervals of a_F, b_F, a_G, b_G

Showing the intervals on the graphs of f, g on Figure 4

6. Calculate the means of f, g, denoted by $E_f[X], E_g[X]$.

```
print('E_f[x]=' + str(np.mean(f)))
print('E_g[x]=' + str(np.mean(g)))
```

```
E_f[x]=0.999
E_g[x]=1.0
```

2. Problem 2

1. Denote by p_n the probability of the interval $[n-1, n)$ under the exponential distribution. What is the expression of p_n as a function of γ and n? What is this expression if $\gamma = ln2$?

$$
\begin{aligned}
p_n &= Pr(x \in [n-1, n)) \\
&= Pr(x < n) - Pr(x \leq n-1) \\
&= 1 - e^{-\gamma n} - (1 - e^{-\gamma(n-1)}) \\
&= e^{-\gamma(n-1)} - e^{-\gamma n} \\
&= e^{-\gamma}(e^{n\gamma} - e^{n}) \qquad\qquad = e^{-\gamma n}(e^{\gamma} - 1)
\end{aligned}
$$

If $\gamma = ln2$, plugging in,

$$p_n = e^{-ln(2)}(e^{n \cdot ln(2) - e^n}) = \frac{1}{2}(2e^n - e^n)$$

2. What is the expression of $\frac{p_n}{p_{n+1}}$ as a function of γ and n? What is the expression if $\gamma = ln2$?

$$
\begin{aligned}
p_{n+1} &= e^{-\gamma}(e^{(n+1)\gamma} - e^{n+1}) \\
&= e^{-\gamma n}(1 - e^{-\gamma}) \\
\frac{p_n}{p_{n+1}} &= \frac{e^{\gamma} - 1}{1 - e^{-\gamma}}
\end{aligned}
$$

If $\gamma = ln2$, plugging in,

$$\frac{p_n}{p_{n+1}} = \frac{2-1}{1-1/2} = 2$$

3. Plot on the same graph the densities $f_\gamma(x)$ for $\gamma = ln2, ln3, ln4$.

```
import math
import numpy as np
import matplotlib.pyplot as plt

gamma = [math.log(2), math.log(3), math.log(4)]
```

```
x = np.arange(0, 10, step=0.01)
fx = np.empty([len(gamma), len(x)])
for k in range(len(gamma)):
    fx[k,] = [gamma[k] * math.exp(-gamma[k] * xx) for xx in x]

plt.figure()
plt.plot(x, fx[0,], 'r--', label='gamma=ln(2)')
plt.plot(x, fx[1,], 'g--', label='gamma=ln(3)')
plt.plot(x, fx[2,], 'b--', label='gamma=ln(4)')
plt.xlabel('x')
plt.title('Density f(x) for different gammas')
plt.legend()
plt.show()
```

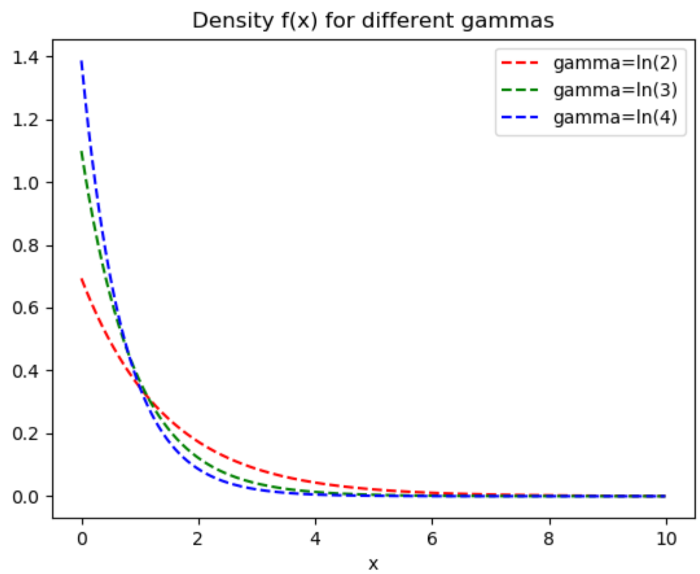


Figure 5: Densities of $f_\gamma(x)$ for γ =ln2,ln3,ln4

Figure 5 shows the density curves.

4. Let $g(x) = \frac{1}{Z}e^{-\gamma(x+3)}$, Evaluate the normalization constant Z as a function of γ. Evaluate the expression of the CDF G of this distribution.

Using the fact that the integral of a density function over the interval $(-\infty, +\infty)$ is 1.

$$\begin{aligned}
\int_{-3}^{\infty} \frac{1}{Z} e^{-\gamma(x+3)} dx &= 1 \\
\int_{-3}^{\infty} e^{-\gamma x - 3\gamma} dx &= Z \\
-\frac{1}{\gamma} e^{-\gamma x} \Big|_{-3}^{\infty} &= Z e^{3\gamma} \\
0 + \frac{1}{\gamma} e^{3\gamma} &= Z e^{3\gamma} \\
\Rightarrow Z &= \frac{1}{\gamma}
\end{aligned}$$

And therefore, we have our CDF $G_X(x)$ to be

$$\begin{aligned}
G_X(x) &= \int_{-3}^{x} \gamma e^{-\gamma(x+3)} dx \\
&= \gamma e^{-3\gamma} \int_{-3}^{x} e^{-\gamma x} dx \\
&= \gamma e^{-3\gamma} \frac{1}{-\gamma} e^{-\gamma x} \Big|_{-3}^{x} \\
&= -e^{-3\gamma} (e^{-\gamma x} - e^{3\gamma}) \\
&= 1 - e^{-\gamma(x+3)}
\end{aligned}$$

3. Problem 3

1. Make a sketch of densities f_X, f_Y.

```
import numpy as np
import matplotlib.pyplot as plt

x = np.arange(0,3,step=0.01)
fx= np.zeros([len(x), 1])
fy= np.zeros([len(x), 1])
for xx in x:
    if xx >= 0.75 and xx <= 1.5:
```

```
        fy[x==xx] = 1/(1.5-0.75)
    if xx >=1 and xx<= 3:
        fx[x==xx] = 1/(3-1)
fx = fx[:,0]
fy = fy[:,0]
plt.figure()
plt.plot(x, fx, 'r-', label='f_X')
plt.plot(x, fy, 'b-', label='f_Y')
plt.xlabel('x')
plt.title('Densities of X and Y')
plt.legend()
plt.show()
```

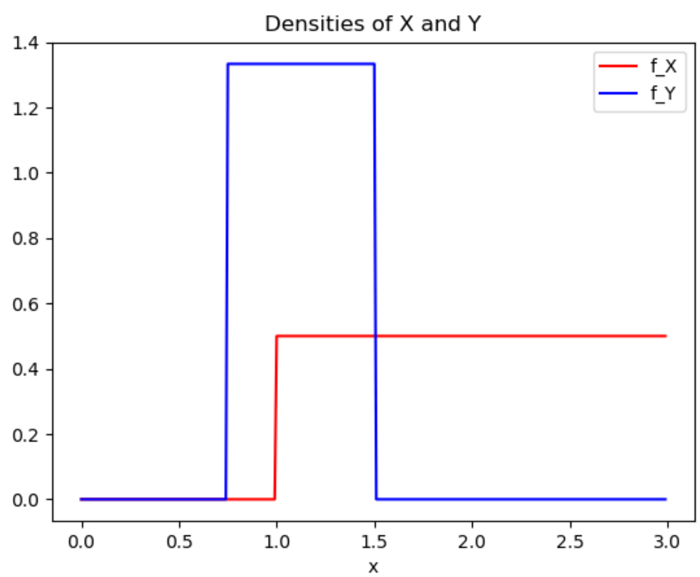


Figure 6: Densities of f_X and f_Y

Figure 6 shows the densities of rabbit's and fox's jump lengths.

2. What are the CDF's of x and y?

$$F_X(x) = \frac{x-1}{3-1} \qquad = \frac{x-1}{2}, x \in [1,3]$$
$$F_Y(y) = \frac{y-0.75}{1.5-0.75} \qquad = \frac{4y-3}{3}, y \in [0.75, 1.5]$$

```
Fx= np.zeros([len(x), 1])
Fy= np.zeros([len(x), 1])
for xx in x:
    if xx >= 0.75 and xx <= 1.5:
        Fy[x==xx] = (4*xx-3)/3
    if xx >=1 and xx<= 3:
        Fx[x==xx] = (xx-1)/2
Fx = Fx[:,0]
Fy = Fy[:,0]
plt.figure()
plt.plot(x, Fx, 'r-', label='F_X')
plt.plot(x, Fy, 'b-', label='F_Y')
plt.xlabel('x')
plt.title('CDF of X and Y')
plt.legend()
plt.show()
```

Figure 7 shows the CDF of rabbit's and fox's jump lengths.

3. What is the probability that a rabbit junps more than 2.5ft? What is the probability that a fox jumps less than 1ft?

$$\begin{aligned} Pr_X(x \geq 2.5) &= 1 - Pr_X(x \leq 2.5) \\ &= 1 - F_X(x = 2.5) \\ &= 1 - \frac{2.5-1}{2} \\ &= 0.25 Pr_Y(y \leq 1) \qquad = F_Y(y=1) \\ &= \frac{4*1-3}{3} \\ &\approx 0.33 \end{aligned}$$

4. A fox is $d = 1$ft away from an unsuspecting rabbit. What is the probability that the fox will catch the rabbit, if the fox jumps once

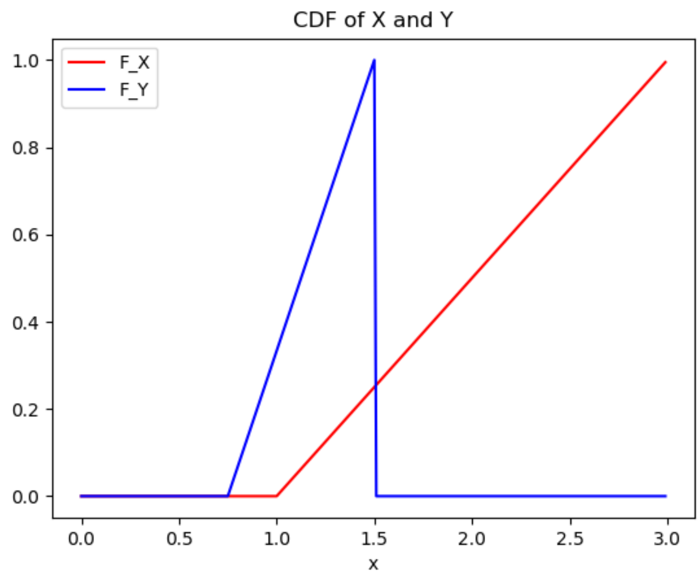


Figure 7: F_X and F_Y

directly towards the rabbit and the rabbit is too surprised to move? To catch a rabbit, this fox must land within 0.2 ft of the rabbit.

$$\begin{aligned}
Pr_Y(y \geq (d - 0.2)) &= 1 - Pr_Y(y \leq 0.8) \\
&= 1 - F_Y(y = 0.8) \\
&= 1 - \frac{4 * 0.8 - 3}{3} \\
&\approx 0.93
\end{aligned}$$

So the fox is has a probability of approximately 0.93 to catch the rabbit with a $d = 1$ ft away.

5. The same question, assume $d = 1.4ft$.

$$\begin{aligned}
Pr_Y(y \geq (d - 0.2)) &= 1 - Pr_Y(y \leq 1.2) \\
&= 1 - F_Y(y = 1.2) \\
&= 1 - \frac{4 * 1.2 - 3}{3} \\
&= 0.4
\end{aligned}$$

So the fox is has a probability of approximately 0.4 to catch the rabbit with a $d = 1.4$ ft away.

6. The fox is now only 0.5 ft from the rabbit, but the rabbit also takes a jump away from the fox. What is the probability that the fox will catch the rabbit, assuming that the fox jumped $y = 1$ ft?

   Since we are look for the probability that the fox will catch the rabbit, and we know that the fox jumped $y = 1$ ft. So, we need to find the probability that the rabbit is not able to flee from the fox. In the other word, the rabbit has jumped less than $y + 0.2 - 0.5 = y - 0.3$ ft.

$$\begin{aligned} Pr_X(x \leq (y - 0.3)) &= Pr_X(x \leq 0.7) \\ &= 0 \end{aligned}$$

   So the fox will not be able to catch the rabbit under this condition.

7. Make a plot of the probability that the fox catches the rabbit under the conditions in **f.**, as a function of y. Figure 8 shows the plot of

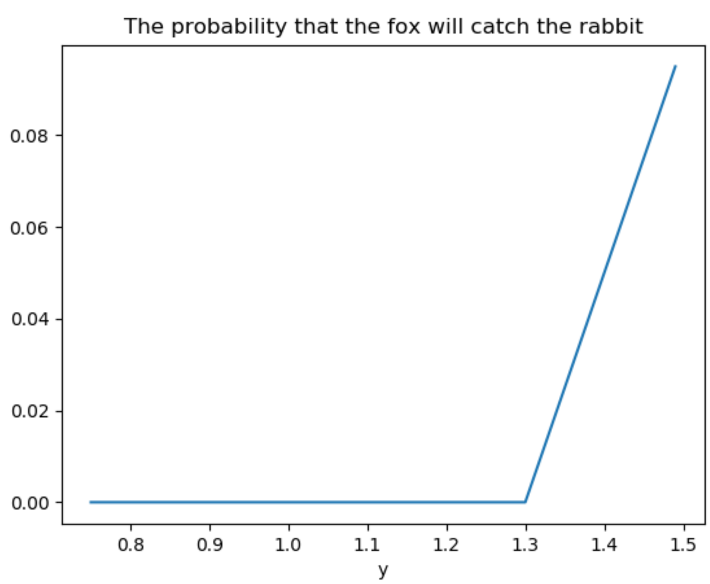


Figure 8: The probability that the fox catches the rabbit as a function of y

probability that the fox can catch the rabbit under the given condition with the chaning y.

4. Problem 4- Rayleigh Distribution

1. Write the formula of likelihood of a, $L(a)$.

$$L(a) = \prod_{i=1}^{n} f(r_i|a) = \prod_{i=1}^{n} \frac{r_i}{a^2} e^{-\frac{r_i^2}{2a^2}}$$

2. Take the logarithm of $L(a)$ to obtain the log-likelihood $l(a) = logL(a)$. Then compute the derivative

$$\begin{aligned} logL(a) &= log \prod_{i=1}^{n} \frac{r_i}{a^2} e^{-\frac{r_i^2}{2a^2}} \\ &= \sum_{i=1}^{n} log(\frac{r_i}{a^2} e^{-fracr_i^2 2a^2}) \\ &= \sum_{i} [log(r_i) - 2log(a) - \frac{r_i^2}{2a^2}] \\ &= \sum_{i} [log(r_i) - \frac{r_i^2}{2a^2}] - 2nlog(a) \end{aligned}$$

Then, the derivative $\partial l/\partial a$ will be

$$\frac{\partial l}{\partial a} = \sum_{i=1}^{n} \frac{r_i^2}{a^3} - \frac{2n}{a}$$

3. Now solve the equation $\frac{\partial l}{\partial a} = 0$ to obtain a formula for $a^{\mathrm{M\,L}}$ as a

function of the data.

$$\begin{aligned}
\frac{\partial l}{\partial a} = 0 &= \sum_{i=1}^{n} \frac{r_i^2}{a^3} - \frac{2n}{a} \\
&= \sum_i \frac{r_i^2}{a^2} - 2n \\
\frac{1}{a^2} &= \frac{2n}{\sum r_i} \\
a &= \sqrt{\frac{\sum_i r_i}{2n}}
\end{aligned}$$

4. Does this problem has sufficient statistics? What are they?

   In this problem, we don't have sufficient statistics directly, but we could use the data set $D = \{r_1, \cdots r_n\}$ to generate the sufficient statistics for us.

5. Problem 4 - Two random samples

1. Write the expression of $l(\gamma)$. Then maximize $l(\gamma)$ to obtain the expression of the Maximum Likelihood estimate $\gamma^{\mathrm{M\,L}}$

$$\begin{aligned}
l(\gamma) &= log(\prod_{i=1}^{n_1+n_2} \gamma e^{-\gamma t_i} \\
&= \sum_{i=1}^{n_1+n_2} (log(\gamma) - \gamma t_i) \\
&= (n_1 + n_2) log\gamma - \gamma \sum_{i}^{n_1+n_2} t_i
\end{aligned}$$

   Then, the derivative $\partial l / \partial \gamma$ will be

$$\begin{aligned}
\frac{\partial l}{\partial \gamma} &= (n_1 + n_2)\frac{1}{\gamma} - \sum_i t_i \\
\gamma^{\mathrm{M\,L}} &= \frac{n_1 + n_2}{\sum_{i=1}^{n_1+n_2} t_i}
\end{aligned}$$

2. Let $n_1 = 2, n_2 = 5$ James' times be $3.5, 0.8$ seconds, and Yali's times be $4.2, 0.5, 1.1, 2.0, 0.3$ seconds. Calculate the Maximum Likelihood of γ for these data.

$$
\begin{aligned}
\gamma^{\mathrm{M\,L}} &= \frac{n_1 + n_2}{\sum_{i=1}^{n_1+n_2} t_i} \\
&= \frac{2+5}{(3.5 + 0.8 + 4.2 + 0.5 + 1.1 + 2 + 0.3)} \\
&\approx 0.5645
\end{aligned}
$$

3. Suppose now that n1 = 2,000,000, n2 = 5,000,000 samples. James has left town to attend a conference after performing his part of the experiment, and has neglijently taken with him all his data $t_1, \cdots, t_{n_1}$ on his laptop. Yali needs to perform the estimation as in 1. but has no information about the experiment 2 James performed, except that he took some samples from the same f_γ as she. How many numbers must James send her, so that she can correctly estimate $\gamma^{\mathrm{M\,L}}$? What are these numbers and how should Yali use them?

   James just needs to send her one number which is the mean of the $t_1, \cdots, t_{n_1}$, $\bar{t}$.
   And then, she could calculate the $\gamma^{\mathrm{M\,L}}$

$$
\gamma_a^{\mathrm{M\,L}} = \frac{n_1 + n_2}{\sum_{i=1}^{n_2} t_i + n_1 * \bar{t}}
$$

6. Problem 5- Least Squares

Let $x_1, \cdots, x_n$ be real numbers and define by $g(z)$ the function

$$
g(z) = \sum_{i=1}^{n} (x_i - z)^2
$$

Show that the minimum of g is attained for

$$
z^* = \frac{1}{n} \sum_{i=1}^{n} x_i
$$

What is the value of $g(z^*)$?

$$\begin{aligned}
\frac{\partial g(z)}{\partial z} &= -2z \sum_{i=1}^{n} (x_i - z) \\
0 &= -2z^* \sum x_i + 2(z^*)^2 n \\
\sum x_i &= z^* n \\
z^* &= \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}
\end{aligned}$$

Then, we have $g(z^*)$ to be

$$g(z^*) = \sum_{i=1}^{n} (x_i - \bar{x})^2$$