Homework_09

March 31, 2025

Probability for Data Science

UC Berkeley, Spring 2025

Michael Xiao and Ani Adhikari

CC BY-NC-SA 4.0

```
[1]: import warnings
     warnings.filterwarnings('ignore')

     from prob140 import *
     from datascience import *
     import numpy as np
     from scipy import stats

     import matplotlib.pyplot as plt
     %matplotlib inline
     import matplotlib
     matplotlib.style.use('fivethirtyeight')
```

1 Homework 9 (Due Monday, March 31st at 5 PM)

1.0.1 How to Do Your Homework

The point of homework is for you to try your hand at using what you've learned in class. The steps to follow:

- Go to lecture and sections, and also go over the relevant text sections before starting on the homework. This will remind you what was covered in class, and the text will typically contain examples not covered in lecture. The weekly Study Guide will list what you should read.
- Work on some of the practice problems before starting on the homework.
- Attempt the homework problems by yourself with the text, section work, and practice materials all at hand. Sometimes the week's lab will help as well. The two steps above will help this step go faster and be more fruitful.
- At this point, seek help if you need it. Don't ask how to do the problem — ask how to get started, or for a nudge to get you past where you are stuck. Always say what you have already tried. That helps us help you more effectively.

- For a good measure of your understanding, keep track of the fraction of the homework you can do by yourself or with minimal help. It's a better measure than your homework score, and only you can measure it.

1.0.2 Rules for Homework

- Every answer should contain a calculation or reasoning. For example, a calculation such as $(1/3)(0.8) + (2/3)(0.7)$ or `sum([(1/3)*0.8, (2/3)*0.7])` is fine without further explanation or simplification. If we want you to simplify, we'll ask you to. But just $\binom{5}{2}$ by itself is not fine; write "we want any 2 out of the 5 frogs and they can appear in any order" or whatever reasoning you used. Reasoning can be brief and abbreviated, e.g. "product rule" or "not mutually exclusive."
- You may consult others (see "How to Do Your Homework" above) but you must write up your own answers using your own words, notation, and sequence of steps.
- We'll be using Gradescope. You must submit the homework according to the instructions at the end of homework set.

1.1 We will not grade assignments which do not have pages correctly selected for each question.

1.2 1. Working with a Joint Density

Let the joint density of X and Y be

$$f(x, y) = \begin{cases} c(x + y) & \text{for } 0 < x < 2, 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

In the parts below, **do the calculus and show all your work**. Don't use SymPy.

(a) Find c.

(b) Find $P(Y > X + 1)$.

(c) Find the density of X.

(d) Are X and Y identically distributed? Explain.

(e) Are X and Y independent? Explain.

(f) Find $E(XY)$.

1.3 2. Functions of Exponential Random Variables

Let X and Y be independent exponential random variables with rates λ and μ respectively.

a) Let $W = \min(X, Y)$. Find the distribution of W. Start by remembering which of the following is easiest when working with minima: the density, the cdf, or the survival function.

b) Let c be a positive constant. Results about exponential random variables in Sections 16.1 and 17.2 of the textbook to find $P(X > cY)$ without integration.

c) Use part (b) to find the cdf of $\frac{X}{Y}$.

1.4   3. Functions of Uniform Random Variables

Let X and Y have joint density

$$f(x, y) = \begin{cases} 90(y - x)^8, & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

In what follows, please do the calculus yourself and **show all your work**. No `SymPy`.

a) Find $P(Y > 2X)$.

b) Find the marginal density of X and remember to provide the possible values. Recognize the density as a member of a famous family and state its name and parameters.

c) Fill in the blanks (**prove your answer**): The joint density f above is the joint density of the _____ and _____ of ten independent uniform $(0, 1)$ random variables.

[Follow the steps in this calculation in the textbook.]

1.5   4. Peter Meets Paul

Peter and Paul agree to meet at a restaurant at noon. Peter arrives at time normally distributed with mean 12:00 noon and SD 3 minutes. Paul arrives at a time normally distributed with mean 12:02 P.M. and SD 4 minutes.

Find the chances below assuming that the two arrival times are independent.

- First, write a formula for the chance in terms of the standard normal cdf Φ.
- Then use a code cell to find the numerical value. You do not have to turn in any coding work for this question.

a) P(Peter arrives before Paul)

b) P(both men arrive within 3 minutes of noon)

c) P(the two men arrive within 3 minutes of each other)

```
[2]:  # Calculation for a
      phi_1 = stats.norm.cdf(2/5)
      phi_1
```

[2]: 0.65542174161032418

```
[3]:  # Calculation for b
      phi_1 = stats.norm.cdf(1)
      phi_2 = stats.norm.cdf(-1)

      phi_3 = stats.norm.cdf(1/4)
      phi_4 = stats.norm.cdf(-5/4)

      p = (phi_1 - phi_2) * (phi_3 - phi_4)
      p
```

[3]: 0.33660452709071237

```
[4]:  # Calculation for c
      phi_1 = stats.norm.cdf(1)
      phi_2 = stats.norm.cdf(-1/5)

      p = phi_1 - phi_2
      p
```

[4]: 0.42060445550764597

1.6 5. Rayleigh Facts and an Application

Let R have the Rayleigh density given by $f(r) = re^{-\frac{1}{2}r^2}$ for $r > 0$.

Refer to Section 18.1 and Section 18.4 for ways in which this density arises.

The point of Parts **a** and **b** is for you *not to reinvent wheels*, and also for you to notice that math calculation can be reduced by using probability facts. The probabilistic approach to math is powerful, so please follow the approach in the instructions. We will not give credit for other approaches.

a) Write $E(R)$ as an integral but don't find its value yet. Let Z be standard normal. You know the numerical value of the variance of Z. Write the variance of Z as an integral, compare with your integral for $E(R)$, and use the comparison to find the value of $E(R)$.

b) Use properties of R^2 to find $E(R^2)$ without any further integration, and hence find $Var(R)$.

c) Suppose two shots are fired at a target. Assume each shot hits with independent normally distributed coordinates, with the same means and equal unit variances. Let D be the distance between the point where the two shots strike. Find $E(D)$ and $Var(D)$.

[Your calculations will go faster if you remember that a normal $(0, \sigma^2)$ variable can be written as σZ where Z is standard normal, and if you use Parts **a-b**.]

1.7 Submission Instructions

Many assignments throughout the course will have a written portion and a code portion. Please follow the directions below to properly submit both portions.

1.7.1 Written Portion

- Scan all the pages into a PDF. You can use any scanner or a phone using applications such as CamScanner. Please **DO NOT** simply take pictures using your phone.
- Please start a new page for each question. If you have already written multiple questions on the same page, you can crop the image in CamScanner or fold your page over (the old-fashioned way). This helps expedite grading.
- It is your responsibility to check that all the work on all the scanned pages is legible.
- If you used LaTeX to do the written portions, you do not need to do any scanning; you can just download the whole notebook as a PDF via LaTeX.

1.7.2 Code Portion

- Save your notebook using `File > Save and Checkpoint`.
- Generate a PDF file using `File > Download As > PDF via LaTeX`. This might take a few seconds and will automatically download a PDF version of this notebook.
    - If you have issues, please post a follow-up on the general Homework 9 Ed thread.

1.7.3 Submitting

- Combine the PDFs from the written and code portions into one PDF. Here is a useful tool for doing so.
- Submit the assignment to Homework 9 on Gradescope.

- **Make sure to assign each page of your pdf to the correct question.**
- **It is your responsibility to verify that all of your work shows up in your final PDF submission.**

If you are having difficulties scanning, uploading, or submitting your work, please read the Ed Thread on this topic and post a follow-up on the general Homework 9 Ed thread.

1.8 We will not grade assignments which do not have pages selected for each question.

[ ]:

1a. $f(x, y) = \begin{cases} c(x+y) & 0 < x < 2, \ 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$

$$1 = \int_0^2 \int_0^2 c(x+y) \, dy \, dx$$

$$= c \int_0^2 \left[xy + \tfrac{1}{2}y^2 \Big|_0^2 \right] dx$$

$$= c \int_0^2 \left[2x + \tfrac{1}{2}(2)^2 - 0 \right] dx$$

$$= c \int_0^2 2x + 2 \, dx$$

$$= c \left[x^2 + 2x \Big|_0^2 \right]$$

$$= c \left[(2)^2 + 2(2) - 0 \right]$$

$$= c \left[4 + 4 \right]$$

$$1 = 8c$$

$$c = \tfrac{1}{8}$$

$$f(x, y) = \begin{cases} \tfrac{1}{8}(x+y) & 0 < x < 2, \ 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

1b. $P(Y > X + 1) = \int_0^1 \int_{x+1}^2 \tfrac{1}{8}(x+y) \, dy \, dx$

$y = x + 1$

$(0, 1) \, (1, 2)$

$$= \tfrac{1}{8} \int_0^1 \left[xy + \tfrac{1}{2}y^2 \Big|_{x+1}^2 \right] dx$$

$$= \tfrac{1}{8} \int_0^1 2x + \tfrac{1}{2}(2)^2 - \left(x(x+1) + \tfrac{1}{2}(x+1)^2 \right) dx$$

$\therefore y > x + 1$

$$= \tfrac{1}{8} \int_0^1 2x + 2 - x^2 - x - \tfrac{1}{2}x^2 - x - \tfrac{1}{2} \, dx$$

when

$0 < x < 1$

$$= \tfrac{1}{8} \int_0^1 -\tfrac{3}{2}x^2 + \tfrac{3}{2} \, dx$$

$$= \tfrac{1}{8} \left[-\tfrac{1}{2}x^3 + \tfrac{3}{2}x \Big|_0^1 \right]$$

$$= \frac{1}{8} \left[-\frac{1}{2}(1)^3 + \frac{3}{2}(1) - 0 \right]$$

$$= \frac{1}{8} \left[-\frac{1}{2} + \frac{3}{2} \right]$$

$$= \frac{1}{8}$$

1c. $\quad f_x(x) = \int_0^2 \frac{1}{8}(x+y) \, dy$

$$= \frac{1}{8} \left[xy + \frac{1}{2}y^2 \Big|_0^2 \right]$$

$$= \frac{1}{8} \left[2x + \frac{1}{2}(2)^2 - 0 \right]$$

$$= \frac{2x+2}{8}$$

$$= \frac{x+1}{4} \qquad 0 < x < 2$$

1d. $\quad f_Y(y) = \int_0^2 \frac{1}{8}(x+y) \, dx$

$$= \frac{1}{8} \left[\frac{1}{2}x^2 + yx \Big|_0^2 \right]$$

$$= \frac{1}{8} \left[\frac{1}{2}(2)^2 + 2y - 0 \right]$$

$$= \frac{1}{8} \left[2 + 2y \right]$$

$$= \frac{y+1}{4} \qquad 0 < y < 2$$

$$\therefore \quad f_x(x) = f_Y(y)$$
$$\hookrightarrow X \text{ and } Y \text{ are identically}$$
$$\text{distributed}$$

1e. $f(x,y) = f_X(x) \, f_Y(y)$?

$$\frac{x+y}{8} = \frac{x+1}{4} \times \frac{y+1}{4}$$

$$\frac{x+y}{8} = \frac{(x+1)(y+1)}{16}$$

$$\frac{x+y}{8} \neq \frac{xy + x + y + 1}{16}$$

∴ X and Y are NOT independent

1f. $E[XY] = \int_0^2 \int_0^2 xy \cdot \frac{1}{8}(x+y) \, dy \, dx$

$\quad = \frac{1}{8} \int_0^2 \int_0^2 x^2 y + x y^2 \, dy \, dx$

$\quad = \frac{1}{8} \int_0^2 \left[\frac{1}{2} x^2 y^2 + \frac{1}{3} x y^3 \Big|_0^2 \right] dx$

$\quad = \frac{1}{8} \int_0^2 \left[\frac{1}{2} x^2 (2)^2 + \frac{1}{3} x (2)^3 - 0 \right] dx$

$\quad = \frac{1}{8} \int_0^2 2x^2 + \frac{8}{3} x \, dx$

$\quad = \frac{1}{8} \left[\frac{2}{3} x^3 + \frac{4}{3} x^2 \Big|_0^2 \right]$

$\quad = \frac{1}{8} \left[\frac{2}{3} (2)^3 + \frac{4}{3} (2)^2 - 0 \right]$

$\quad = \frac{1}{8} \left[\frac{16}{3} + \frac{16}{3} \right]$

$\quad = \frac{1}{8} \left[\frac{32}{3} \right]$

$\quad = \frac{4}{3}$

2a. $X \sim \text{Exponential}(\lambda)$

$Y \sim \text{Exponential}(M)$

$W = \min(X, Y)$

idea: survival function! $P(W > w)$
$(1 - \text{cdf})$ $1 - P(W \leq w)$

$$P(W > w) = P(\min(X, Y) > w)$$
$$= P(X > w \text{ and } Y > w)$$
$$= P(X > w) P(Y > w)$$

$P(X \leq w) = 1 - e^{-\lambda w}$ $P(Y \leq w) = 1 - e^{-Mw}$
$P(X > w) = e^{-\lambda w}$ $P(Y > w) = e^{-Mw}$

$$= e^{-\lambda w} \times e^{-Mw}$$
$$= e^{-\lambda w - Mw}$$
$$S_w(w) = e^{-(\lambda + M) w}$$

$$F_w(w) = 1 - e^{-(\lambda + M) w}$$

$\therefore W \sim \text{Exponential}(\lambda + M)$

2b. $P(X > cY) = ?$

$$\therefore P(cY > y) = P(Y > \frac{y}{c})$$
$$= e^{-M(y/c)}$$
$$= e^{-(M/c) y}$$

$$\therefore \ cY \ \sim \ \text{Exponential} \left(\frac{M}{c} \right)$$

$$P(X > Y) = \frac{M}{\lambda + M} \qquad \text{textbook } 17.2.2$$

$$= \frac{\frac{M}{c}}{\lambda + \frac{M}{c}}$$

$$= \frac{M}{c\lambda + M}$$

2c. $\quad P\left(\frac{X}{Y} \le z \right) = P(X \le zY)$

$$= 1 - P(X > zY) \qquad \text{idea: same as part b}$$

$$F_Z(z) = 1 - \frac{M}{z\lambda + M} \qquad z \ge 0$$

3a.

$$f(x,y) = \begin{cases} 90(y-x)^8 & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

$P(Y > 2X) \longrightarrow$ bounds

$$0 < 2x < 1$$
$$0 < x < \tfrac{1}{2}$$

$$2x < y < 1$$

$$= \int_0^{1/2} \int_{2x}^1 90(y-x)^8 \, dy \, dx$$

$y = 1$
$u = 1-x$

$y = 2x$
$u = x$

$u = y - x$
$du = dy$

$$= 90 \int_0^{1/2} \int_x^{1-x} u^8 \, du \, dx$$

$$= 90 \int_0^{1/2} \left[\tfrac{1}{9} u^9 \Big|_x^{1-x} \right] dx$$

$$= 10 \int_0^{1/2} (1-x)^9 - x^9 \, dx$$

$$= 10 \left[\underbrace{\int_0^{1/2} (1-x)^9 \, dx}_{①} - \underbrace{\int_0^{1/2} x^9 \, dx}_{②} \right]$$

① $\displaystyle \int_0^{1/2} (1-x)^9 \, dx$

$$= \int_1^{1/2} u^9 \cdot - du$$

$u = 1-x$
$-du = dx$

$x = \tfrac{1}{2}$
$u = \tfrac{1}{2}$

$x = 0$
$u = 1$

$$= \int_{1/2}^1 u^9 \, du$$

$$= \tfrac{1}{10} u^{10} \Big|_{1/2}^1$$

$$= \tfrac{1}{10} \left((1)^{10} - \left(\tfrac{1}{2}\right)^{10} \right)$$

$$= \tfrac{1}{10} \left(1 - \tfrac{1}{1024} \right)$$

$$= \frac{1023}{10240}$$

② $\int_0^{1/2} x^9 \, dx$

$= \frac{1}{10} x^{10} \Big|_0^{1/2}$

$= \frac{1}{10} \left(\left(\frac{1}{2}\right)^{10} - 0 \right)$

$= \frac{1}{10} \left(\frac{1}{1024} \right)$

$= \frac{1}{10240}$

$\therefore \quad = 10 \left(\frac{1023}{10240} - \frac{1}{10240} \right)$

$= \frac{10230}{10240} - \frac{10}{10240}$

$= \frac{10220}{10240}$

$= \frac{511}{512}$

3b. $f_X(x) = \int_0^1 f(x,y) \, dy$

when $0 < x < 1$
$\quad \hookrightarrow x < y < 1$

$= \int_x^1 90 (y-x)^8 \, dy$

$= 90 \int_0^{1-x} u^8 \, du$

$y = 1$
$v = 1-x$

$u = y - x$
$du = dy$

$= 90 \left[\frac{1}{9} u^9 \Big|_0^{1-x} \right]$

$= 10 \left[(1-x)^9 - 0 \right]$

$= 10 (1-x)^9 \qquad 0 < x < 1$

$y = x$
$v = 0$

$10 (1-x)^9 \longrightarrow 10 x^{1-1} (1-x)^{10-1}$

$\therefore \quad \text{Beta}(1, 10)$

3c. $U_1, U_2, \dots, U_{10} \sim$ iid Uniform $(0,1)$

$$X = \min(U_1, \dots, U_{10})$$
$$Y = \max(U_1, \dots, U_{10})$$

1 of $U_1, \dots, U_{10}$ at x = dx

1 of $U_1, \dots, U_{10}$ at y = dy

8 of $U_1, \dots, U_{10}$ in (x,y) = $(y-x)$

0 of $U_1, \dots, U_{10}$ outside $[x,y]$ = 0

$$\therefore \frac{10!}{1!\,1!\,8!\,0!} (dx)^1 (dy)^1 (y-x)^8 (0)^0$$

$$= \frac{10 \times 9 \times 8!}{8!} \, dx \, dy \, (y-x)^8$$

$$f_{XY}(x,y) \, dx \, dy = 90 (y-x)^8 \, dx \, dy$$

$$f_{XY}(x,y) = 90 (y-x)^8, \quad 0 < x < y < 1$$

$\therefore$ The joint density f above is the joint density of the minimum and maximum of ten independent uniform $(0,1)$ random variables

4a. P(Peter arrives before Paul)

Peter: $X \sim \text{Normal}(0, 3^2)$

$\sim \text{Normal}(0, 9)$

Paul: $Y \sim \text{Normal}(2, 4^2)$

$\sim \text{Normal}(2, 16)$

∴ X and Y are independent

$P(X < Y) = P(X - Y < 0)$

$X - Y \sim \text{Normal}(0 - 2, 9 + 16)$

$\sim \text{Normal}(-2, 25)$

$$= \Phi\left(\frac{0 - (-2)}{\sqrt{25}}\right)$$

$$= \Phi\left(\frac{2}{5}\right)$$

≈ 0.6554

Normal $(E[X], Var(X))$

$SD(X) = \sqrt{Var(X)}$

4b. $-3 \leq X \leq 3$

$-3 \leq Y \leq 3$

⟩ Paul and Peter both arrive within 3 min of noon

↳ disregard Paul's relative to his mean?

∴ NOT $(-1 \leq Y \leq 5)$

$P\left(\begin{array}{c}\text{both men arrive within} \\ 3 \text{ minutes of noon}\end{array}\right) = P(-3 \leq X \leq 3) \, P(-3 \leq Y \leq 3)$

$P(-3 \leq X \leq 3) = P(X \leq 3) - P(X \leq -3)$

$$= \Phi\left(\frac{3-0}{3}\right) - \Phi\left(\frac{-3-0}{3}\right)$$

$$= \Phi(1) - \Phi(-1)$$

$$P(-3 \le Y \le 3) = P(Y \le 3) - P(Y \le -3)$$

$$= \Phi\left(\frac{3-2}{4}\right) - \Phi\left(\frac{-3-2}{4}\right)$$

$$= \Phi\left(\frac{1}{4}\right) - \Phi\left(-\frac{5}{4}\right)$$

$$= \left(\Phi(1) - \Phi(-1)\right) \times \left(\Phi\left(\frac{1}{4}\right) - \Phi\left(-\frac{5}{4}\right)\right)$$

$$\approx 0.3366$$

4c. $P\left(\begin{array}{l}\text{the two men arrive within} \\ \text{3 min of each other}\end{array}\right)$

$$P(|x-Y| \le 3) = P(-3 \le x - Y \le 3)$$

$$= P(x-Y \le 3) - P(x-Y \le -3)$$

$$= \Phi\left(\frac{3 + (-2)}{5.}\right) - \Phi\left(\frac{-3 + (+2)}{5}\right)$$

$$= \Phi(1) - \Phi\left(-\frac{1}{5}\right)$$

$$\approx 0.4206$$

5a. $R \sim$ Rayleigh

$$f(r) = re^{-\frac{1}{2}r^2} \qquad r > 0$$

$$E[R] = \int_0^\infty r \cdot re^{-\frac{1}{2}r^2} \, dr$$

$$= \int_0^\infty r^2 e^{-\frac{1}{2}r^2} \, dr$$

$Z \sim$ Normal $(0, 1)$

$$Var(Z) = E[Z^2] - (E[Z])^2$$
$$= 1 - 0 \qquad\qquad \text{textbook} \quad 18.1.3$$
$$Var(Z) = 1$$

$$Var(Z) = E[Z^2] = \int_{-\infty}^\infty z^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \, dz$$

$$1 = 2 \int_0^\infty z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \, dz$$

$$1 = \frac{2}{\sqrt{2\pi}} \int_0^\infty z^2 e^{-\frac{1}{2}z^2} \, dz$$

$$\frac{\sqrt{2\pi}}{2} = \int_0^\infty z^2 e^{-\frac{1}{2}z^2} \, dz$$

$$\therefore \quad z = r$$

$$\frac{\sqrt{2\pi}}{2} = \int_0^\infty r^2 e^{-\frac{1}{2}r^2} \, dr$$

$$\therefore \quad E[R] = \frac{\sqrt{2\pi}}{2}$$

5b. $E[R^2] = \int_0^\infty r^2 \cdot r e^{-\frac{1}{2}r^2} dr$

$\qquad = \underbrace{\int_0^\infty 2u e^{-u} du}$

$\qquad\qquad\qquad \therefore \text{ gamma!}$

$$\begin{array}{ll} r = 0 & r = \infty \\ U = 0 & U = \infty \end{array}$$

$U = r^2 \rightarrow 2U = r^2$

$dU = r\, dr \qquad \sqrt{2U} = r$

$\Gamma(2) = 2\int_0^\infty u^{2-1} e^{-U} du \qquad \text{textbook } 18.3$

$\Gamma(2) = (2-1)!$

$\qquad = 1!$

$\qquad = 1$

$\therefore E[R^2] = 2(1)$

$\qquad\qquad = 2$

$\therefore \operatorname{Var}(R) = E[R^2] - (E[R])^2$

$\qquad\qquad = 2 - \left(\frac{\sqrt{2\pi}}{2}\right)^2$

$\qquad\qquad = 2 - \frac{\pi}{2}$

5c. $\quad \text{shot } 1 = (X_1, Y_1)$

$\qquad \text{shot } 2 = (X_2, Y_2)$

$\therefore X_1, X_2, Y_1, Y_2 \sim \text{Normal } (0, 1)$

$D = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$

$X = X_2 - X_1 \qquad\qquad Y = Y_2 - Y_1$

$X \sim \text{Normal } (0-0, 1+1) \qquad Y \sim \text{Normal}(0-0, 1+1)$

$\quad \sim \text{Normal } (0, 2) \qquad\qquad \sim \text{Normal } (0, 2)$

$\therefore D = \sqrt{X^2 + Y^2}$

Since D has Rayleigh's distribution

$$\therefore D = \sigma R$$

$$E[D] = \sqrt{2} \ E[R]$$

$$= \sqrt{2} \times \frac{\sqrt{2\pi}}{2}$$

$$= \frac{2\sqrt{\pi}}{2}$$

$$= \sqrt{\pi}$$

$$Var(D) = \sigma^2 \ Var(R)$$

$$= (\sqrt{2})^2 \ (2 - \frac{\pi}{2})$$

$$= 2 \ (2 - \frac{\pi}{2})$$

$$= 4 - \pi$$