Homework_10

April 7, 2025

Probability for Data Science

UC Berkeley, Spring 2025

Michael Xiao and Ani Adhikari

CC BY-NC-SA 4.0

```python
[3]: 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 10 (Due Monday, April 7th at 5PM)

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. Poisson MGF

Let X have Poisson(μ) distribution, and let Y independent of X have Poisson (λ) distribution.

a) Derive the moment generating function of X. The answer is in the textbook. Your task here is to show why it's true.

b) Use the result of (a) to show that the distribution of $X + Y$ is Poisson.

1.3   2. Gamma Tail Bound

Before you do this exercise, carefully study a relevant example in the textbook. You will have to follow similar steps.

You will need the mgf of the gamma distribution. Also remember that you found the gamma mean and variance in Homework 9.

Let X have the gamma (r, λ) distribution.

a) Show that $P(X \geq 2E(X)) \leq \left(\frac{2}{e}\right)^r$.

b) Find Markov's and Chebyshev's bounds on $P(X \geq 2E(X))$.

c) [CODE] Fix $\lambda = 1$. Display overlaid plots of the following four graphs as functions of r, for r in the interval $(0.5, 15)$:

- The exact tail probability $P(X \geq 2E(X))$
- The bound in Part **a**: $\left(\frac{2}{e}\right)^r$
- Chebyshev's bound on $P(X \geq 2E(X))$
- Markov's bound on $P(X \geq 2E(X))$

The code uses `plt.plot` which you have used before. The expression `stats.gamma.cdf(x, r, scale=1)` evaluates to the cdf of the gamma $(r, 1)$ distribution at the point x.

```
[4]: # Answer to c
r = np.arange(0.05, 15, 0.1)


markov_bound = r/(2*r)


chebyshev_bound = 1/r


part_a_bound = (2/np.e)**r


# Use as many lines as you need for the exact values
exact = 1 - stats.gamma.cdf(2*r,r,scale=1)



plt.plot(r, exact, lw=2, label='Exact Chance')
plt.plot(r, part_a_bound, lw=2, label='Part (a) Bound')
plt.plot(r, chebyshev_bound, lw=2, label='Chebyshev Bound')
plt.plot(r, markov_bound, lw=2, label='Markov Bound')
plt.legend()
plt.xlabel('$r$')
plt.ylim(0, 1)
plt.xlim(0, 15)
plt.title('$P(X \geq 2E(X))$ for $X$ gamma $(r, 1)$');
```
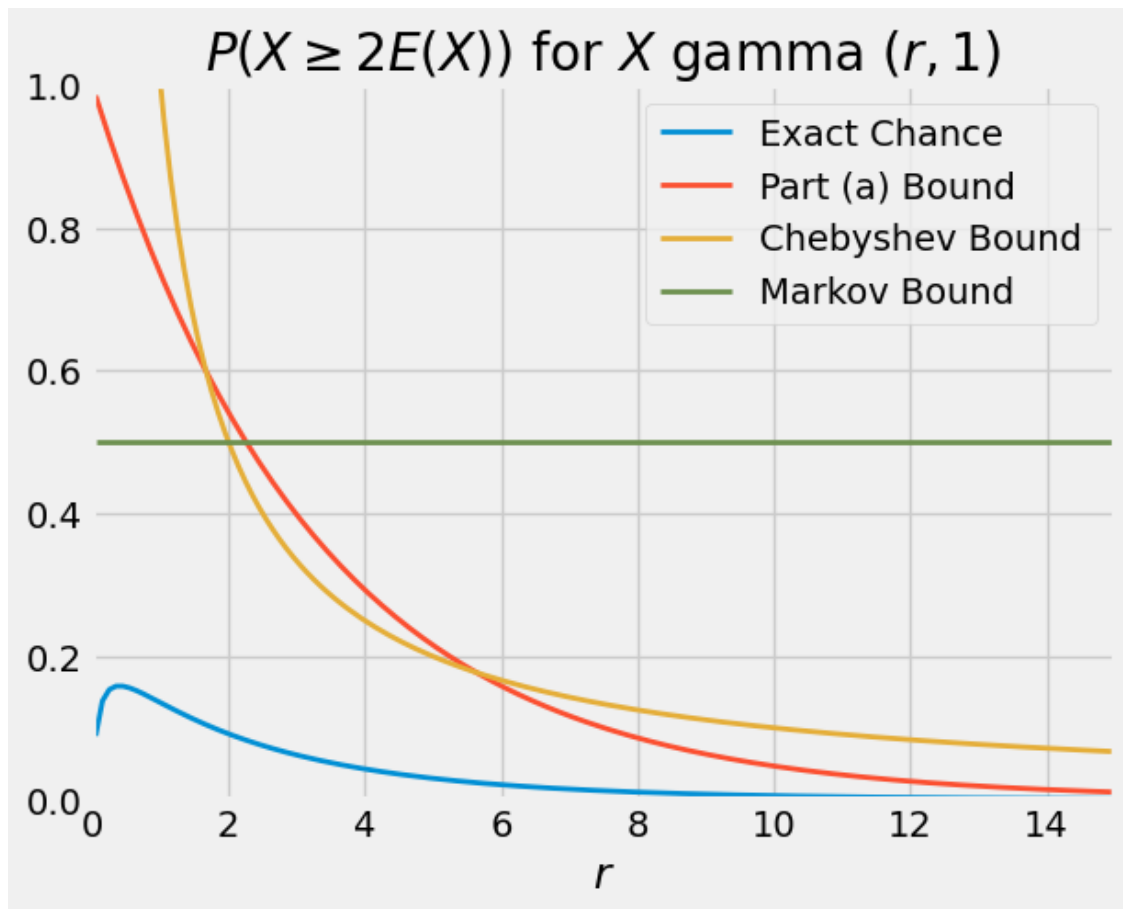
$P(X \ge 2E(X))$ for X gamma $(r, 1)$

Legend:
- Exact Chance
- Part (a) Bound
- Chebyshev Bound
- Markov Bound

r

1.4 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.

1.4.1 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.

1.4.2 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 10 Ed thread.

1.4.3 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 10 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 10 Ed thread.

1a. $X \sim \text{Poisson}(M)$

$P(X = k) = e^{-M} \cdot \dfrac{M^k}{k!}$

$M_x(t) = E[e^{tx}]$

$= \displaystyle\sum_{x=0}^{\infty} e^{tx} \cdot P(X = x)$

$= \displaystyle\sum_{x=0}^{\infty} e^{tx} \cdot e^{-M} \cdot \dfrac{M^x}{x!}$

$= e^{-M} \cdot \displaystyle\sum_{x=0}^{\infty} e^{tx} \cdot \dfrac{M^x}{x!}$

$= e^{-M} \cdot \displaystyle\sum_{x=0}^{\infty} \dfrac{(M e^t)^x}{x!}$

$= e^{-M} \underbrace{\displaystyle\sum_{x=0}^{\infty} \dfrac{k^x}{x!}}_{e^k}$

$k = M e^t$

$e^k = \displaystyle\sum_{x=0}^{\infty} \dfrac{k^x}{x!}$

$= 1 + \dfrac{k^1}{1!} + \dfrac{k^2}{2!} + \dots$

$= e^{-M} \cdot e^{M e^t}$

$= e^{M(e^t - 1)}$

1b. $X \sim \text{Poisson}(M)$
$Y \sim \text{Poisson}(\lambda)$ } independent

$X + Y \sim \text{Poisson}(M + \lambda)$?

$M_{x+y}(t) = E[e^{t(X+Y)}]$

$= E[e^{tX} \cdot e^{tY}]$

$= M_x(t) \cdot M_Y(t)$

$= e^{M(e^t - 1)} \cdot e^{\lambda(e^t - 1)}$

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

$\alpha = M + \lambda$

check : $Z \sim$ Poisson(α)

$M_Z(t) = e^{\alpha(e^t - 1)}$

$= e^{(M + \alpha)(e^t - 1)}$

$\therefore X + Y \sim$ Poisson$(M + \lambda)$

2a. Prove: $P(X \geq 2E[x]) \leq \left(\frac{2}{e}\right)^r$

$X \sim \text{Gamma}(r, \lambda)$

$$M_x(t) = \left(\frac{\lambda}{\lambda - t}\right)^r \qquad t < \lambda$$

$$E[x] = \frac{r}{\lambda}$$

$$2E[x] = \frac{2r}{\lambda}$$

$$P(X \geq c) \leq \min_{t>0} M_x(t) e^{-tc} \qquad \text{textbook } 19.4.4$$

$$P\left(X \geq \frac{2r}{\lambda}\right) \leq \min_{t>0} \frac{\left(\frac{\lambda}{\lambda - t}\right)^r}{e^{t \cdot 2r/\lambda}}$$

> numerator and denominator are both increasing functions

$$= y$$

$$y = \frac{\left(\frac{\lambda}{\lambda - t}\right)^r}{e^{t \cdot 2r/\lambda}}$$

$$\log(y) = \log\left(\frac{\left(\frac{\lambda}{\lambda - t}\right)^r}{e^{t \cdot 2r/\lambda}}\right)$$

$$= \ln\left(\frac{\lambda}{\lambda - t}\right)^r - \ln\left(e^{2rt/\lambda}\right)$$

$$= r \ln\left(\frac{\lambda}{\lambda - t}\right) - \frac{2rt}{\lambda}$$

$$\frac{d}{dt}\log(y) = r \cdot \frac{1}{\frac{\lambda}{\lambda - t}} \cdot \frac{\lambda}{(\lambda - t)^2} - \frac{2r}{\lambda}$$

$$= r \cdot \frac{\lambda - t}{\lambda} \cdot \frac{\lambda}{(\lambda - t)^2} - \frac{2r}{\lambda}$$

$$= \frac{r}{\lambda - t} - \frac{2r}{\lambda}$$

$$\therefore \quad 0 = r\left(\frac{1}{\lambda - t} - \frac{2}{\lambda}\right) \qquad\qquad y' = 0 \longrightarrow \text{minimum}$$

$$\frac{2}{\lambda} = \frac{1}{\lambda - t}$$

$$2\lambda - 2t = \lambda$$

$$2t = \lambda$$

$$t = \frac{\lambda}{2}$$

$$\therefore \quad y = \frac{\left(\dfrac{\lambda}{\lambda - \frac{\lambda}{2}}\right)^r}{e^{2r\lambda/2\lambda}}$$

$$= \frac{\left(\dfrac{2\lambda}{2\lambda - \lambda}\right)^r}{e^r}$$

$$= \frac{2^r}{e^r}$$

$$\therefore \quad P(x \geq 2E[x]) \leq \left(\frac{2}{e}\right)^r = P\left(x \geq \frac{2r}{\lambda}\right) \leq \left(\frac{2}{e}\right)^r$$

2b. $\quad P(X \geq c) \leq \dfrac{E[x]}{c} \qquad\qquad$ markov's

$$P(x \geq 2E[x]) \leq \frac{E[x]}{2E[x]}$$

$$P(x \geq \frac{2r}{\lambda}) \leq \frac{\frac{r}{\lambda}}{\frac{2r}{\lambda}}$$

$$\leq \frac{r}{\lambda} \cdot \frac{\lambda}{2r}$$

$$\leq \frac{1}{2}$$

$$\therefore P(X \geq 2E[X]) \leq \frac{1}{2}$$

$$P(|x - E[X]| \geq c) \leq \frac{Var(X)}{c^2} \qquad \text{chebyshev's}$$

$$X \sim Gamma(r, \lambda)$$
$$\therefore Var(X) = \frac{r}{\lambda^2}$$

$$P(|x - \frac{r}{\lambda}| \geq \frac{r}{\lambda}) \leq \frac{\frac{r}{\lambda^2}}{(\frac{r}{\lambda})^2}$$

$$\leq \frac{r}{\lambda^2} \cdot \frac{\lambda^2}{r^2}$$

$$\leq \frac{1}{r}$$

$$\therefore P(X \geq 2E[X]) \leq \frac{1}{r}$$