Homework_08

March 19, 2025

Probability for Data Science

UC Berkeley, Spring 2025

Michael Xiao and Ani Adhikari

CC BY-NC-SA 4.0

1 Homework 8 (Due Monday, March 17th at 5PM)

```
[7]:  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')
```

1.0.1 Instructions

Your homeworks will generally have two components: a written portion and a portion that also involves code. Written work should be completed on paper, and coding questions should be done in the notebook. Start the work for the written portions of each section on a new page. You are welcome to LATEX your answers to the written portions, but staff will not be able to assist you with LATEX related issues.

It is your responsibility to ensure that both components of the homework are submitted completely and properly to Gradescope. **Make sure to assign each page of your pdf to the correct question. Refer to the bottom of the notebook for submission instructions.**

1.0.2 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.

1.0.3 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 selected for each question.

1.1.1 Plot_continuous

In this homework you will plot a number of densities using the `prob140` function `Plot_continuous` which takes as its arguments:

- a list consisting of two elements: the left endpoint and right endpoint of the plotting interval on the horizontal axis
- the name of a function that takes a numerical argument and returns the value of the density at that argument; this could be a function you have defined, or a `SciPy` function `stats.distribution_name.pdf`.

`Plot_continuous` takes other arguments as well, as you can see in the brief documentation, but you will not need those in the homework.

1.2 1. Widths of Confidence Intervals

In any part of this question that involves a sample size, you can assume the sample size is big enough for the Central Limit Theorem approximation to be good. You should answer on paper, but you can use the code cell provided below for arithmetic or to find standard normal percentages or percentiles. The appropriate libary has been imported at the top of this notebook.

a) A survey organization has used the methods of our class to construct an approximate 95% confidence interval for the mean annual income of households in a county. The interval runs from $\$66,000$ to $\$70,000$. If possible, find an approximate 98% confidence interval for the mean annual income of households in the county. If this is not possible, explain why not.

b) A survey organization is going to take a simple random sample of n voters from among all the voters in a state, to construct a 98% confidence interval for the proportion of voters who favor a proposition. Find an n such that the total width of the confidence interval (left end to right end) will be no more than 0.06. Remember that you can bound the variance of an indicator.

```
[8]: # Code cell for scratch work

    ...
```

[8]: Ellipsis

1.3 2. Density Operations

Let the density of X be given by

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find numerical values for each of the following. **You may not use Sympy or any other symbolic math programs or websites. Show all the calculus and algebra steps on paper**, and use the code cell below for the final numerical calculations.

a) $P(|X - 0.5| > 0.2)$

b) $P(|X - 0.5| > 0.2 \mid X > 0.4)$

c) $E(X)$

d) $SD(X)$

[9]: `# Code cell for calculating final numerical values`

1.4   3. Minimum and Average

In this exercise you will discover properties of the sample minimum and the sample average of i.i.d. samples taken from two well-known distributions. If necessary, you can leave answers in terms of the standard normal cdf Φ.

(a) Let $U_1, U_2, \ldots, U_n$ be i.i.d. uniform $(0, 1)$ random variables and let $X_n = \min\{U_1, U_2, \ldots, U_n\}$. Find the survival function of X_n and hence find the density of X_n.

(b) Let $T_1, T_2, \ldots, T_n$ be i.i.d. exponential (λ) random variables and let $Y_n = \min\{T_1, T_2, \ldots, T_n\}$. Find the survival function of Y_n and hence identify the distribution of Y_n as one of the famous ones; remember to provide the relevant parameters.

(c) Let $U_1, U_2, \ldots, U_n$ be i.i.d. uniform $(0, 1)$ random variables and let $V_n = \frac{1}{n} \sum_{i=1}^{n} U_i$. For large n, find an approximation to the survival function of V_n.

(d) Let $T_1, T_2, \ldots, T_n$ be i.i.d. exponential (λ) random variables and let $W_n = \frac{1}{n} \sum_{i=1}^{n} T_i$. For large n, find an approximation to the survival function of W_n.

1.5 4. Cauchy Density

This exercise is about a density that looks bell shaped but behaves quite differently from the normal. The diagram below will be helpful for Part **a**.

a) Suppose that a particle is fired from the origin in the (x, y)-plane in a straight line in a direction at a random angle θ to the x-axis. Let Y be the y-coordinate of the point where the particle hits the line $x = 1$. Show that if θ has the uniform distribution on $(-\pi/2, \pi/2)$, then the density of Y is

$$f_Y(y) \;=\; \frac{1}{\pi(1 + y^2)}, \quad -\infty < y < \infty$$

This is called the *Cauchy* density.

b) [**CODE**] Use `Plot_continuous` twice, to plot the Cauchy density over the interval $(-5, 5)$ and overlay the standard normal curve. Use `stats.cauchy.pdf` for the Cauchy density, make sure to use `'gold'` for the color of the normal curve. Refer to the description of `Plot_continuous` at the top of this homework.
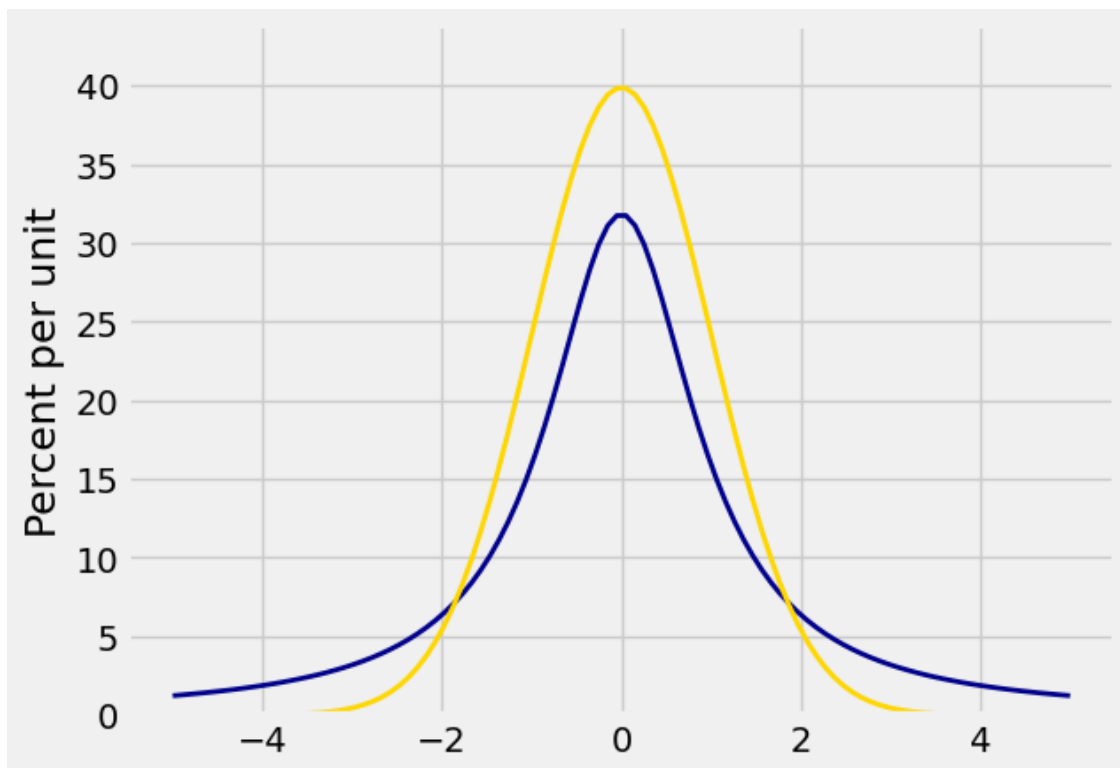
```
[10]: # Answer to b

      # Plotting interval on the horizontal axis: -5 to 5
      # The first plot should be the Cauchy.
      # The gold plot should be the standard normal curve.
      x = np.linspace(-5, 5, 1000)

      def cauchy(x):
          return stats.cauchy.pdf(x)

      def normal(x):
          return stats.norm.pdf(x)

      Plot_continuous([-5,5], cauchy)
      Plot_continuous([-5,5], normal, color='gold')
```

c) For Y with the Cauchy density, use calculus to show that $E(|Y|) = \infty$. Thus $E(Y)$ is undefined even though the density of Y is symmetric about 0.

The Cauchy curve is called the Witch of Agnesi. Skim the History and Applications sections of the Wikipedia article.

d) Explain what is being plotted by the code below, and discuss what you see in the graph in relation to the Weak Law of Large Numbers. Run the cell **several times** before you answer. You should also vary N.

[`stats.cauchy.rvs(size = N)` returns an array of N i.i.d. Cauchy random numbers.]

```
[11]: # Part d: run this several times
      N = 10000
      n = np.arange(1, N+1)
      x = stats.cauchy.rvs(size = N)
      y = np.cumsum(x)/n
      plt.plot(n, y, color='darkblue', lw=2)
      plt.plot([0, N], [0, 0], color='k', lw=2);
```
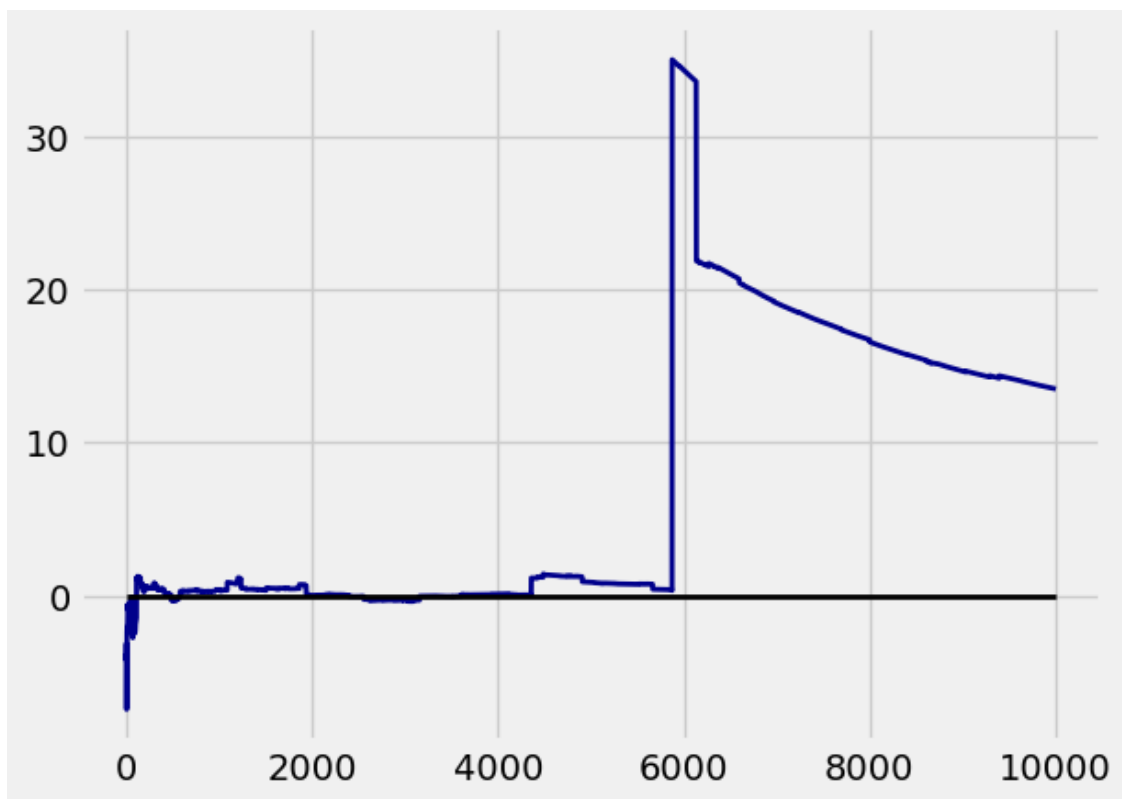
1.6   5. The Gamma Family

In this problem you will start with some calculus exercises and then develop one of the fundamental families of densities. An already-familiar distribution belongs to this family, as you will show in the next exercise (Problem 6). In the next exercise, you will also see that this family has been used to model individual infectiousness due to the novel coronavirus (COVID-19).

Assume that $r > 0$ and $\lambda > 0$ are constants.

a) The *Gamma function* of mathematics is defined by

$$\Gamma(r) = \int_0^\infty t^{r-1} e^{-t} dt$$

That letter is the upper case Greek letter Gamma. You can assume that the integral converges and that therefore $\Gamma(r)$ is a positive number.

Use integration by parts to show that

$$\Gamma(r+1) = r\Gamma(r), \quad r > 0$$

b) Use part (a) and induction to show that if r is a positive integer then $\Gamma(r) = (r-1)!$. Thus the Gamma function is a continuous extension of the factorial function.

c) Let X have density given by

$$f_X(t) = \begin{cases} \dfrac{1}{\Gamma(r)} t^{r-1} e^{-t}, & t > 0 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = \frac{1}{\lambda} X$. Show that the density of Y is

$$f_Y(t) = \begin{cases} \dfrac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t}, & t > 0 \\ 0 & \text{otherwise} \end{cases}$$

The density of Y is called the gamma (r, λ) density.

d) Use (c) and the textbook to confirm that if Z has the standard normal density then Z^2 has the gamma $(1/2, 1/2)$ density. Hence find $\Gamma(1/2)$, and then use (a) to find $\Gamma(3/2)$ and $\Gamma(5/2)$. Please don't leave any integrals in your answers.

Shapes of Gamma Densities: Part (c) shows that the gamma (r, λ) density is obtained as a change of scale (units of measurement) starting with a gamma $(r, 1)$ random variable and multiplying it by $1/\lambda$. That is why **stats** calls $1/\lambda$ the *scale parameter* of the gamma density. The parameter λ is called the *rate*.

The parameter r determines the shape of the curve and is called the *shape parameter*. To see why, it helps to look at some plots.

The method **stats.gamma.pdf** can be used with the following arguments to return the value of the density: - the value of t (possibly an array) at which to evaluate the density - r - λ specified as **scale=1/**

Run the cell below to see overlaid graphs of the several gamma densities with different values of r but the same rate λ. You will see why r is called the shape parameter.

[12]:
```python
# Part d
t = np.arange(0, 25, 0.01)
r_1 = stats.gamma.pdf(t, 1, scale=1/0.25)
r_1_5 = stats.gamma.pdf(t, 1.5, scale=1/0.25)
r_2 = stats.gamma.pdf(t, 2, scale=1/0.25)
r_2_5 = stats.gamma.pdf(t, 2.5, scale=1/0.25)
r_3 = stats.gamma.pdf(t, 3, scale=1/0.25)
plt.plot(t, r_1, lw=2, label='gamma (1, 0.25)')
plt.plot(t, r_1_5, lw=2, label='gamma (1.5, 0.25)')
plt.plot(t, r_2, lw=2, label='gamma (2, 0.25)')
plt.plot(t, r_2_5, lw=2, label='gamma (2.5, 0.25)')
plt.plot(t, r_3, lw=2, label='gamma (3, 0.25)')
plt.xlabel('$t$')
plt.legend()
plt.title('Gamma Densities: Same Rate and Different Shape Parameters');
```


Gamma Densities: Same Rate and Different Shape Parameters
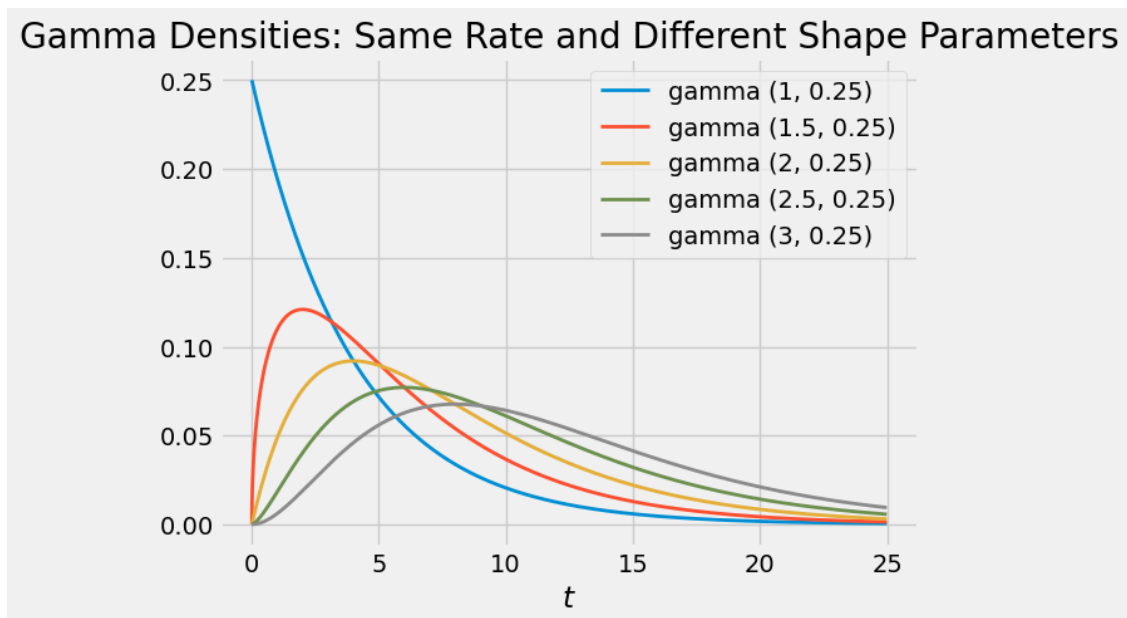
1.7 6. Mean and Variance of the Gamma

As in Question 3, let r and λ be positive constants.

a) Use Question 3(c) and properties of densities to evaluate

$$\int_0^\infty t^{r-1} e^{-\lambda t} dt$$

in terms of r, λ, and the Gamma function. For what follows, keep in mind that your formula is correct for all positive r.

b) Let T have gamma (r, λ) density. Use (a) and 5(a) to find a simple formula for $E(T)$ in terms of r and λ. Please don't leave any integrals or Gamma functions in your answer.

c) For T as in (b), find $E(T^2)$ in terms of the Gamma function.

d) Find $Var(T)$. **Make sure to simplify all integrals and gamma functions. As usual, show your work.**

e) Identify the gamma $(1, \lambda)$ density as one that has another famous name, and confirm that your answers to (b) and (c) are consistent with what you already know about that density.

f) [**CODE**] A report from Imperial College, UK, is believed to have spurred the US government to take steps to slow down the spread of Covid in March 2020. The assumptions of one of the main models is in the penultimate paragraph on Page 4. Note the sentence, "Individual infectiousness is assumed to be variable, described by a gamma distribution with mean 1 and shape parameter $\alpha = 0.25$."

Broadly speaking, infectiousness is quantified by the number of cases directly generated by one infected person. We denote this quantity as the random variable R. The model produced by researchers at Imperial College estimates the distribution of R under COVID-19, based on prior knowledge of similar viruses and respiratory diseases.

Plot the gamma distribution used in the model. The horizontal axis should start at 0 and extend to 2 standard deviations above the mean of 1. Make sure to label your x-axis.

We are not giving you any skeleton code for this. Use what you've learned from the code cells above, and please give your plot an appropriate title.

```
[14]: # Answer to f

mean = 1
sd = 2

t = np.linspace(0, mean + 2 * sd, 1000)
pdf = stats.gamma.pdf(t, 0.25, scale=1/0.25)

plt.plot(t, pdf)
plt.title('Gamma Distribution of COVID-19 Infectiousness');
plt.xlabel('Infectiousness (R)');
plt.ylabel('Probability Density');
```

Gamma Distribution of COVID-19 Infectiousness
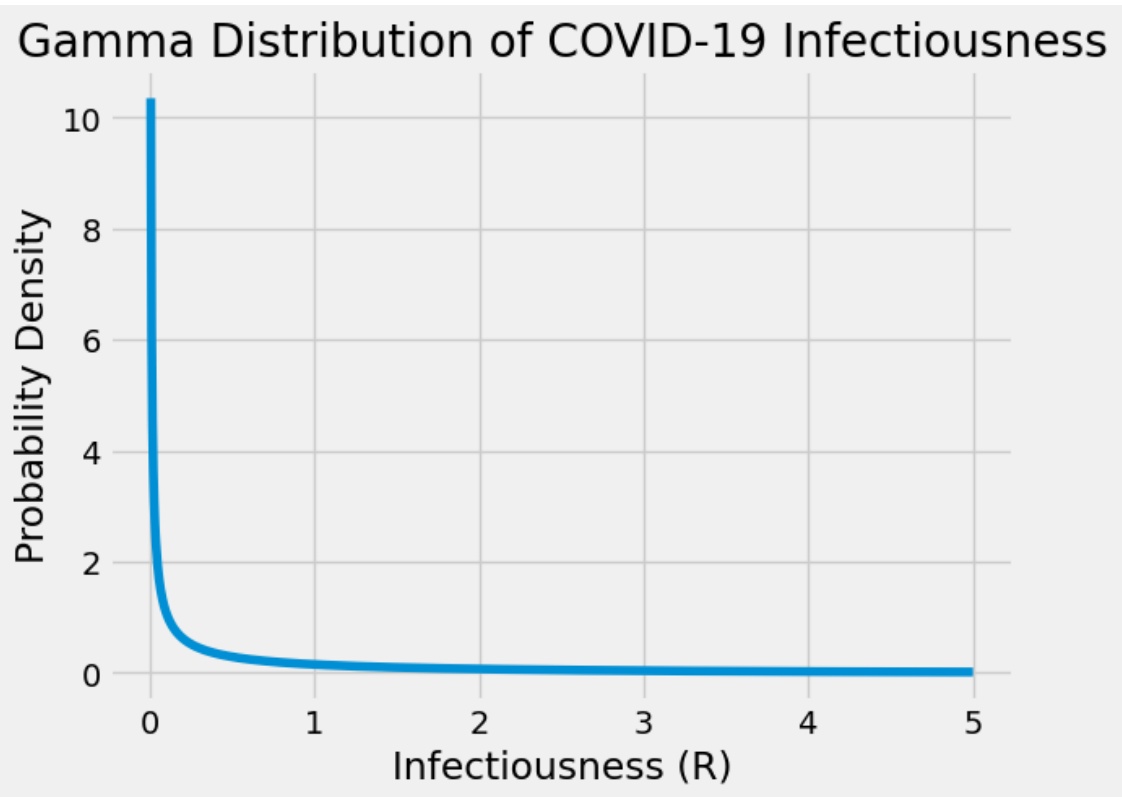
1.8 Submission Instructions

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

1.8.1 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.

1.8.2 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 8 Ed thread.

1.8.3 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 8 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 8 Ed thread.

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

[ ]:

1a. $\mu = \dfrac{66000 + 70000}{2}$

$= \dfrac{136000}{2}$

$= 68000$

98%
95%
?   68000   ?
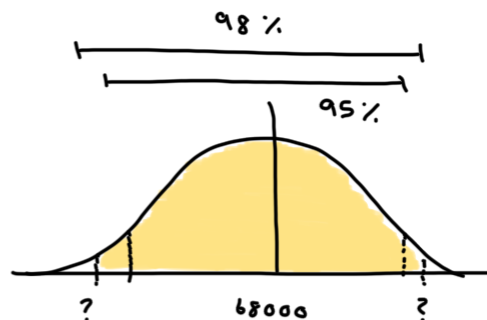
width of = $70000 - 66000$
95% CI $= 4000$

$\frac{1}{2}$ width $= \frac{1}{2}(4000)$
$= 2000$

95% CI z-score $= 1.96$

$(1.96)\dfrac{\sigma}{\sqrt{n}} = 2000$ margin of error

$\dfrac{\sigma}{\sqrt{n}} = \dfrac{2000}{1.96}$

≈ 1020.408

98% CI z-score $= 2.33$

98%
95%
65622.449   68000   70377.551
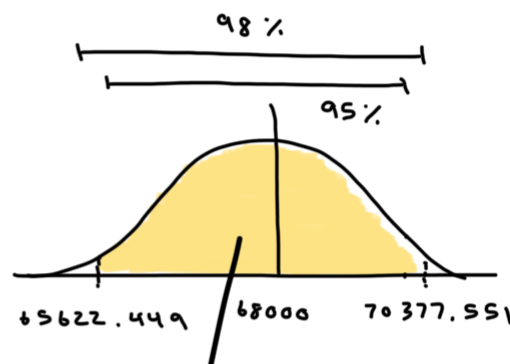
$M \pm z\dfrac{\sigma}{\sqrt{n}} = 68000 \pm (2.33) \times \left(\dfrac{2000}{1.96}\right)$

$= 68000 \pm 2377.551$

$= [65622.449, 70377.551]$

1b. $2 \cdot z\sqrt{\dfrac{\sigma}{\sqrt{n}}} \leq 0.06$

$2(2.33)\sqrt{\dfrac{p(1-p)}{n}} \leq 0.06$ 98% CI z-score $= 2.33$

$p = \frac{1}{2} \longrightarrow$ maximizes the variance,

which maximizes standard
deviation, which then
leads to the largest
possible confidence interval
↳ ensures n is large
enough

$$\therefore 4.66 \sqrt{\frac{0.5(1-0.5)}{n}} \leq 0.06$$

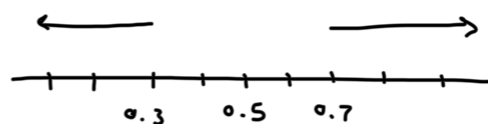$$(4.66)^2 \times \frac{0.25}{n} \leq (0.06)^2$$

$$\frac{5.4289}{n} \leq 0.0036$$

$$\frac{5.4289}{0.0036} \leq n$$

$$n \geq 1508.03$$

$\therefore$ n should be at least 1508 so
that the total width of the CI
will be no more than 0.06

2a. $P(|X - 0.5| > 0.2) = P(x < 0.5 - 0.2) + P(x > 0.5 + 0.2)$

$= P(X < 0.3) + P(X > 0.7)$

∴ arrows show anything more than 0.2 away from 0.5

$P(X < 0.3) = \int_0^{0.3} \frac{3}{2}(1 - x^2)\, dx$

$= \frac{3}{2}x - \frac{3}{2} \cdot \frac{1}{3} x^3 \Big|_0^{0.3}$

$= \frac{3}{2}(0.3) - \frac{1}{2}(0.3)^3 - \left(\frac{3}{2}(0) - \frac{1}{2}(0)^3 \right)$

$= 0.45 - 0.0135 - 0$

$= 0.4365$

$P(X > 0.7) = \int_{0.7}^{1} \frac{3}{2}(1 - x^2)\, dx$

$= \frac{3}{2}x - \frac{1}{2}x^3 \Big|_{0.7}^{1}$

$= \frac{3}{2}(1) - \frac{1}{2}(1)^3 - \left(\frac{3}{2}(0.7) - \frac{1}{2}(0.7)^3 \right)$

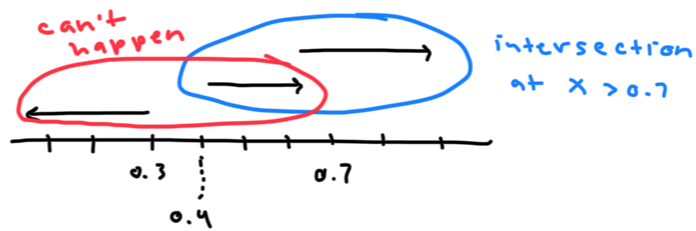$= \frac{3}{2} - \frac{1}{2} - \left(1.05 - 0.1715 \right)$

$= 0.1215$

∴ $P(X < 0.3) + P(X > 0.7) = 0.4365 + 0.1215$

$= 0.5580$

2b. $P(|X - 0.5| > 0.2 \mid X > 0.4)$

$= \dfrac{P(|X - 0.5| > 0.2 \text{ AND } X > 0.4)}{P(X > 0.4)}$

$P(|X - 0.5| > 0.2$ AND $X > 0.4) = P(X > 0.7)$

can't happen

intersection at $X > 0.7$

0.3

0.7

0.4

$P(X < 0.3)$ and $P(X > 0.4)$

can't be true at same time!

$$= \frac{P(X > 0.7)}{P(X > 0.4)}$$

$$P(X > 0.4) = \int_{0.4}^{1} \frac{3}{2}(1 - x^2)\, dx$$

$$= \frac{3}{2}x - \frac{1}{2}x^3 \Big|_{0.4}^{1}$$

$$= \frac{3}{2}(1) - \frac{1}{2}(1)^3 - \left(\frac{3}{2}(0.4) - \frac{1}{2}(0.4)^3\right)$$

$$= \frac{3}{2} - \frac{1}{2} - \left(0.6 - 0.032\right)$$

$$= 0.432$$

$$\therefore P(|X - 0.5| > 0.2 \mid X > 0.4) = \frac{0.1215}{0.432}$$

$$\simeq 0.281$$

2c.
$$E[X] = \int_{0}^{1} x \cdot \frac{3}{2}(1 - x^2)\, dx$$

$$= \frac{3}{2}\int_{0}^{1} x - x^3\, dx$$

$$= \frac{3}{2}\left[\frac{1}{2}x^2 - \frac{1}{4}x^4 \Big|_{0}^{1}\right]$$

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

$$= \frac{3}{4} - \frac{3}{8} - 0$$

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

2d. $SD(x) = \sqrt{Var(x)}$

$Var(x) = E[x^2] - (E[x])^2$

$$E[x^2] = \int_0^1 x^2 \cdot \frac{3}{2}(1-x^2)\,dx$$

$$= \frac{3}{2}\int_0^1 x^2 - x^4\,dx$$

$$= \frac{3}{2}\left[\frac{1}{3}x^3 - \frac{1}{5}x^5 \Big|_0^1\right]$$

$$= \frac{1}{2}(1)^3 - \frac{3}{10}(1)^5 - \left[\frac{1}{2}(0)^3 - \frac{3}{10}(0)^5\right]$$

$$= \frac{1}{2} - \frac{3}{10}$$

$$= \frac{1}{5}$$

$$Var(x) = \frac{1}{5} - \left(\frac{3}{8}\right)^2$$

$$= \frac{1}{5} - \frac{9}{64}$$

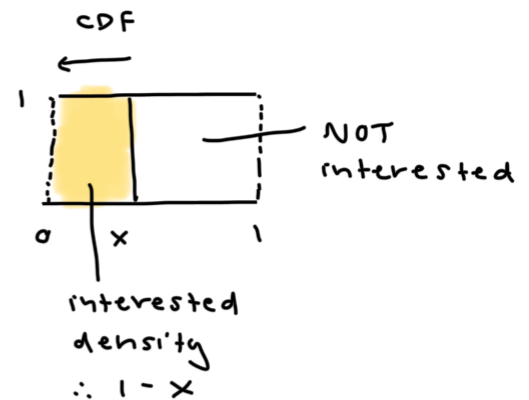$$\therefore SD(x) = \sqrt{\frac{1}{5} - \frac{9}{64}}$$

$$= \sqrt{0.0594}$$

$$= 0.2437$$

3a. $U_i \sim$ uniform $(0,1)$, $i = 1, \ldots, n$

$X_n = \min \{U_1, U_2, \ldots, U_n\}$

$$S_{X_n}(x) = P(X_n > x)$$
$$= P(\min \{U_1, U_2, \ldots, U_n\} > x)$$
$$= P(U_1 > x) P(U_2 > x) \ldots P(U_n > x)$$
$$= (P(U_1 > x))^n$$

$$P(U_1 > x) = 1 - P(U_1 \leq x)$$
$$= 1 - x$$

$$\therefore (1 - x)^n$$

CDF

NOT interested

interested
density
$\therefore 1 - x$

$$F_{X_n}(x) = 1 - S_{X_n}(x)$$
$$= 1 - (1-x)^n$$
$$F_{X_n}'(x) = 0 - n(1-x)^{n-1} \cdot -1$$
$$= n(1-x)^{n-1}, \quad 0 \leq x \leq 1$$

3b. $T_i \sim$ exponential (λ), $i = 1, 2, \ldots, n$

$Y_n = \min \{T_1, T_2, \ldots, T_n\}$

$$S_{Y_n}(y) = P(Y_n > y)$$
$$= P(\min \{T_1, T_2, \ldots, T_n\} > y)$$
$$= P(T_1 > y) P(T_2 > y) \ldots P(T_n > y)$$
$$= (P(T_1 > y))^n$$
$$= (e^{-\lambda y})^n \longrightarrow \text{discussion 8}$$
$$= e^{-n\lambda y} \qquad\qquad \text{EX 15.5}$$

$$\therefore Y_n \sim \text{Exponential} (n\lambda)$$

3c. $U_i \sim \text{Uniform}(0,1)$

$V_n = \frac{1}{n} \sum_{i=1}^{n} U_i$ sample mean

n is large $\rightarrow$ CLT $\rightarrow$ Normal distribution!

$E[U_i] = \frac{1}{2}$

$Var(U_i) = \frac{1}{12}$ $\Big\rangle$ textbook 15.3.1

$E[V_n] = E\left[\frac{1}{n} \sum_{i=1}^{n} U_i\right]$

$\qquad = \frac{1}{n} \sum_{i=1}^{n} E[U_i]$

$\qquad = \frac{1}{n} \cdot n \cdot \frac{1}{2}$

$\qquad = \frac{1}{2}$

$Var(V_n) = Var\left(\frac{1}{n} \sum_{i=1}^{n} U_i\right)$

$\qquad = \frac{1}{n^2} \sum_{i=1}^{n} Var(U_i)$ textbook 12.1.3

$\qquad\qquad\qquad\qquad\qquad\qquad Var(aX) = a^2 Var(X)$

$\qquad = \frac{1}{n^2} \cdot n \cdot \frac{1}{12}$

$\qquad = \frac{1}{12n}$

$V_n \sim \text{Normal}\left(\frac{1}{2}, \frac{1}{12n}\right)$

$S_{V_n}(v) = P(V_n > v)$

$\qquad = 1 - P(V_n \le v)$

$\qquad = 1 - \Phi\left(\frac{v - \frac{1}{2}}{\sqrt{1/(12n)}}\right)$ textbook 14.3.4

3d. $T_i \sim$ Exponential (λ)

$$W_n = \frac{1}{n} \sum_{i=1}^{n} T_i$$

n is large $\rightarrow$ CLT
$\downarrow$
Normal
distribution

$E[T_i] = \frac{1}{\lambda}$

$Var(T_i) = \frac{1}{\lambda^2}$ $\Big\}$ textbook 15.4.2

$$E[W_n] = E\left[\frac{1}{n} \sum_{i=1}^{n} T_i\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} E[T_i]$$

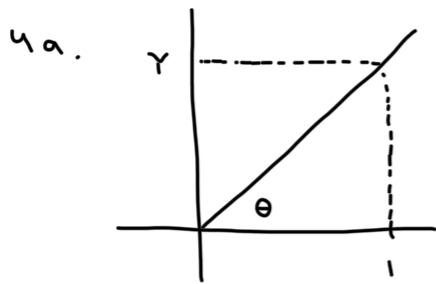$$= \frac{1}{n} \cdot n \cdot \frac{1}{\lambda}$$

$$= \frac{1}{\lambda}$$

$$Var(W_n) = Var\left[\frac{1}{n} \sum_{i=1}^{n} W_n\right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} Var(W_n)$$

$$= \frac{1}{n^2} \cdot n \cdot \frac{1}{\lambda^2}$$

$$= \frac{1}{n\lambda^2}$$

$$W_n \sim Normal\left(\frac{1}{\lambda}, \frac{1}{n\lambda^2}\right)$$

$$\delta_{w_n} = P(W_n > w)$$

$$= 1 - P(W_n \leq w)$$

$$= 1 - \Phi\left(\frac{w - \frac{1}{\lambda}}{\sqrt{1/(n\lambda^2)}}\right)$$

4a.

$$\tan\theta = \frac{y}{x}$$
$$y = x\tan\theta$$
$$y = (1)\tan\theta$$
$$\therefore Y = \tan\theta$$

∴ monotone function
↳ always increasing

$$\theta \sim \text{Uniform}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

$$f_\theta(\theta) = \frac{1}{\frac{\pi}{2} + \left(+\frac{\pi}{2}\right)} \qquad \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

$$= \frac{1}{\pi}$$

$$\int_{-\pi/2}^{\pi/2} \frac{1}{\pi}\, d\theta = 2\int_0^{\pi/2} \frac{1}{\pi}\, d\theta$$
$$= 2\left[\frac{1}{\pi}\theta \Big|_0^{\pi/2}\right]$$
$$= 2\left[\frac{1}{\pi}\left(\frac{\pi}{2}\right) - \frac{1}{\pi}(0)\right]$$
$$= 2\left[\frac{1}{2} - 0\right]$$
$$= 1 \quad \checkmark$$

$$\therefore f_\theta(\theta) \text{ is PDF}$$

Change of Variable
1. density of θ —— $y = \tan\theta$
2. derivative of
   transform —— $y = \tan\theta$
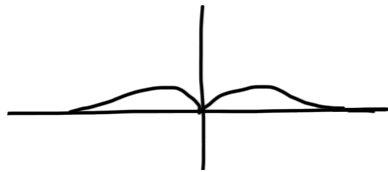 $$y' = \sec^2\theta$$
   $$= 1 + \tan^2\theta$$
   $$= 1 + y^2$$

$$f_{\tan\theta}(y) = \frac{f_\theta(\theta)}{g'(\theta)}$$

$$= \frac{\frac{1}{\pi}}{1+y^2}$$

$$= \frac{1}{\pi(1+y^2)} \qquad -\infty < y < \infty$$

4c. $E[|X|] = \displaystyle\int_{-\infty}^{\infty} |y| \frac{1}{\pi(1+y^2)} \, dy$

∴ symmetric

$$= \frac{2}{\pi} \int_{0}^{\infty} \frac{y}{1+y^2} \, dy$$

$$= \frac{2}{\pi} \int_{0}^{\infty} \frac{1}{u} \cdot \frac{1}{2} \, du \qquad\qquad u = 1+y^2$$

$$\qquad\qquad\qquad\qquad\qquad du = 2y$$

$$= \frac{1}{\pi} \left[\ln|u| \,\Big|_{0}^{\infty} \right] \qquad\qquad \frac{1}{2} du = y$$

$$= \frac{1}{\pi} \left[\ln|1 + y^2| \,\Big|_{0}^{\infty} \right]$$

$$= \frac{1}{\pi} \left[\underbrace{\left(\lim_{t \to \infty} \ln|1 + t^2| \right)}_{= \infty} - \underbrace{\ln|1 + (0)^2|}_{= 0} \right]$$

$$\therefore \quad \infty$$

4d. The code below is plotting an array of N iid Cauchy random numbers. In relation to the Weak Law of Large Numbers (WLLN), the sample mean of these values fluctuates widely, even as N increases, instead of converging to a fixed value. For example, distributions with a well-defined expected value (ie. normal distribution) should have a sample mean that stabilizes around the true mean as N increases. Yet, from part 4c, we know the Cauchy distribution has an undefined expected value because its expectation integral diverges. So, WLLN does NOT apply, as the

Sample mean doesn't settle to a single number. When we vary N, we still see that the sample mean continues fluctuating, which proves that the sample size does not stabilize like it does for distributions with finite means.

5a. $\Gamma(v) = \int_0^\infty t^{v-1} e^{-t} \, dt$

$\Gamma(v+1) = \int_0^\infty t^{(v+1)-1} e^{-t} \, dt$

$\qquad u = t^v \qquad\qquad dv = e^{-t} \, dt$

$\qquad\quad = \int_0^\infty t^v e^{-t} \, dt$

$\qquad du = v t^{v-1} dt \qquad v = -e^{-t}$

$\qquad\quad = -t^v e^{-t} - \int_0^\infty -e^{-t} v t^{v-1} \, dt$

$\qquad\quad = -t^v e^{-t} + v \underbrace{\int_0^\infty e^{-t} t^{v-1} \, dt}_{\Gamma(v)}$

$\qquad\quad = \left[-t^v e^{-t} \Big|_0^\infty \right] + v \Gamma(v)$

$\qquad\quad = \left[\lim_{t \to \infty} -t^v e^{-t} - \left(-(0)^v e^{-0} \right) \right] + v \Gamma(v)$

$\qquad\quad = 0 - 0 + v \Gamma(v)$

$\qquad\quad = v \Gamma(v)$

$\qquad \therefore \Gamma(v+1) = v \Gamma(v)$

5b. $\Gamma(v) = (v-1)!$

$\Gamma(1) = \int_0^\infty t^{1-1} e^{-t} \, dt$

$\qquad\quad = \int_0^\infty 1 \cdot e^{-t} \, dt$

$\qquad\quad = -e^{-t} \Big|_0^\infty$

$\qquad\quad = \left[\lim_{t \to \infty} -e^{-t} \right] - (-e^{-0})$

$\qquad\quad = 0 + 1$

$\qquad\quad = 1 \;\; = 0!$

$\Gamma(2) = \int_0^\infty t^{2-1} e^{-t} \, dt$

$\qquad\quad = \int_0^\infty t \, e^{-t} \, dt$

$\qquad\qquad\qquad\qquad\qquad u = t \qquad dv = e^{-t} \, dt$

$$= -te^{-t} - \int_0^\infty -e^{-t}\, dt \qquad\qquad du: dt \qquad v = -e^{-t}$$

$$= -te^{-t} - e^{-t}\Big|_0^\infty$$

$$= \left[\lim_{t\to\infty} \underbrace{-te^{-t}}_{0} \underbrace{- e^{-t}}_{0}\right] - \left(\underbrace{-0e^{-0}}_{0} \underbrace{- e^{-0}}_{-1}\right)$$

$$= 0 + 1$$

$$= 1 = 1!$$

$$\Gamma(3) = \int_0^\infty t^{3-1} e^{-t}\, dt$$

$$= \int_0^\infty t^2 e^{-t}\, dt \qquad\qquad u = t^2 \qquad dv = e^{-t}\, dt$$
$$\qquad\qquad\qquad\qquad\qquad\qquad du = 2t\, dt \qquad v = -e^{-t}$$

$$= -t^2 e^{-t} - \int_0^\infty -2te^{-t}\, dt$$

$$= -t^2 e^{-t} + 2 \underbrace{\int_0^\infty te^{-2t}\, dt}_{1 \text{ (answer above)}}$$

$$= -t^2 e^{-t}\Big|_0^\infty + 2(1)$$

$$= \left[\lim_{t\to\infty} \underbrace{-t^2 e^{-t}}_{0}\right] - \left(\underbrace{-(0)^2 e^{-0}}_{0}\right) + 2$$

$$= 2 = 2!$$

$$\therefore \Gamma(r+1) = r\,\Gamma(r)$$
$$= r(r-1)!$$
$$= r!$$

$$\therefore \Gamma(r) = (r-1)!$$

5c. $Y = \dfrac{1}{\lambda} X$

1. density of X $\qquad\qquad\qquad\qquad \dfrac{1}{\Gamma(r)} t^{r-1} e^{-t}$

2. inverse of

transformation _____ $y = \frac{1}{\lambda} x$

3. derivative of

$x = \lambda y$

transformation

$y = \frac{1}{\lambda} x$

$y' = \frac{1}{\lambda}$

$f_Y(t) = \dfrac{f_X(\lambda t)}{|g'(\lambda t)|}$

$f_{\frac{1}{\lambda}x}(t) = \dfrac{\dfrac{1}{\Gamma(r)} (\lambda t)^{r-1} e^{-\lambda t}}{\left|\dfrac{1}{\lambda}\right|}$ at $x = \lambda y$

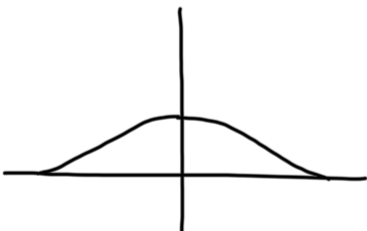$= \dfrac{1}{\Gamma(r)} (\lambda t)^{r-1} e^{-\lambda t} |\lambda|$

$= \dfrac{1}{\Gamma(r)} \lambda^{r-1+1} t^{r-1} e^{-\lambda t}$

$= \dfrac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} \qquad t > 0$

$\therefore f_Y(t) = \begin{cases} \dfrac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} & t > 0 \\ \\ 0 & \text{otherwise} \end{cases}$

5d. $Z \sim \text{Normal}(0, 1)$

$f_Z(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty \qquad$ textbook 16.1.3

∴ symmetry

$Y = Z^2$

∴ two-to-one function

1. density ——— $f_Z(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} Z^2}$

2. inverse ———————————— $y = Z^2$

3. derivative ————————— $Z = \pm\sqrt{y}$

$y = Z^2$

$y' = 2Z$

$f_Y(y) = \dfrac{f_Z(\sqrt{y})}{|g'(\sqrt{y})|} + \dfrac{f_Z(-\sqrt{y})}{|g'(-\sqrt{y})|}$

$= \dfrac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2}}{|2\sqrt{y}|} + \dfrac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-\sqrt{y})^2}}{|-2\sqrt{y}|}$ at $Z = \pm\sqrt{y}$

$= \dfrac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y}}{2\sqrt{y}} + \dfrac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y}}{2\sqrt{y}}$

$= 2\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \cdot \frac{1}{2\sqrt{y}} \right)$

$= \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}$ $y > 0$

Gamma (r, λ)

$f_Y(t) = \dfrac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t}$

∴ Gamma $(\frac{1}{2}, \frac{1}{2})$

$f_Y(t) = \dfrac{1}{\Gamma(\frac{1}{2})} \cdot \left(\frac{1}{2}\right)^{\frac{1}{2}} t^{\frac{1}{2}-1} e^{-\frac{1}{2}t}$

$$f_Y(y) = \frac{1}{\Gamma(\frac{1}{2})} \cdot \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y}$$

$$\frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} = \frac{1}{\Gamma(\frac{1}{2})} \cdot \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}$$

$$\frac{1}{\sqrt{2\pi}} = \frac{1}{\Gamma(\frac{1}{2})} \cdot \frac{1}{\sqrt{2}}$$

$$\frac{1}{\sqrt{\pi}} = \frac{1}{\Gamma(\frac{1}{2})}$$

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

$$\therefore \Gamma(r+1) = r\,\Gamma(r)$$

$$\Gamma(\tfrac{1}{2}+1) = \tfrac{1}{2}\Gamma(\tfrac{1}{2})$$

$$\Gamma(\tfrac{3}{2}) = \tfrac{1}{2}\sqrt{\pi}$$

$$\Gamma(\tfrac{3}{2}+1) = \tfrac{3}{2}\Gamma(\tfrac{3}{2})$$

$$\Gamma(\tfrac{5}{2}) = \tfrac{3}{2} \cdot \tfrac{1}{2}\sqrt{\pi}$$

$$= \tfrac{3}{4}\sqrt{\pi}$$

6a. $\int_0^\infty t^{v-1} e^{-\lambda t} dt$

$u = \lambda t \longrightarrow t = \frac{u}{\lambda}$

$du = \lambda\, dt$

$\frac{1}{\lambda} du = dt$

$= \int_0^\infty \left(\frac{u}{\lambda}\right)^{v-1} e^{-u} \cdot \frac{1}{\lambda}\, du$

$= \int_0^\infty u^{v-1} \cdot \left(\frac{1}{\lambda}\right)^{v-1+1} e^{-u}\, du$

$= \frac{1}{\lambda^v} \int_0^\infty u^{v-1} e^{-u}\, du$

$= \frac{1}{\lambda^v} \Gamma(v)$

6b. $T \sim gamma(v, \lambda)$

$E(T) = \int_0^\infty t \cdot F_Y(t)\, dt$

$= \int_0^\infty t \cdot \frac{\lambda^v}{\Gamma(v)} t^{v-1} e^{-\lambda t}\, dt$

$= \int_0^\infty \frac{\lambda^v}{\Gamma(v)} t^{v-1+1} e^{-\lambda t}\, dt$

$= \frac{\lambda^v}{\Gamma(v)} \underbrace{\int_0^\infty t^v e^{-\lambda t}\, dt}_{\frac{1}{\lambda^{v+1}} \Gamma(v+1)}$

idea: 6a.

$= \frac{\lambda^v \Gamma(v+1)}{\lambda^v \lambda^1 \Gamma(v)}$ idea: $\Gamma(v+1) = v\Gamma(v)$

$= \frac{v \Gamma(v)}{\lambda \Gamma(v)}$

$= \frac{v}{\lambda}$

6c. $E[T^2] = \int_0^\infty t^2 \cdot \frac{\lambda^v}{\Gamma(v)} t^{v-1} e^{-\lambda t} \, dt$

$= \frac{\lambda^v}{\Gamma(v)} \int_0^\infty t^{v-1+2} e^{-\lambda t} \, dt$

$= \frac{\lambda^v}{\Gamma(v)} \underbrace{\int_0^\infty t^{v+1} e^{-\lambda t} \, dt}_{\frac{1}{\lambda^{v+2}} \Gamma(v+2)}$

$= \frac{\lambda^v \, \Gamma(v+2)}{\Gamma(v) \, \lambda^v \, \lambda^2}$

$= \frac{(v+1) v \, \Gamma(v)}{\lambda^2 \, \Gamma(v)}$

$= \frac{v(v+1)}{\lambda^2}$

$\Gamma(v+2) = (v+1) \Gamma(v+1)$

$= (v+1) v \, \Gamma(v)$

6d. $\text{Var}(T) = E[T^2] - (E[T])^2$

$= \frac{v(v+1)}{\lambda^2} - \left(\frac{v}{\lambda}\right)^2$

$= \frac{v^2 + v - v^2}{\lambda^2}$

$= \frac{v}{\lambda^2}$

6e. Gamma$(1, \lambda)$

$f_T(t) = \frac{\lambda^1}{\Gamma(1)} t^{1-1} e^{-\lambda t}$

$= \frac{\lambda^1}{0!} e^{-\lambda t}$

$= \lambda e^{-\lambda t}$

$\therefore$ Exponential(λ)

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

$$= \frac{1}{\lambda}$$

$$E[T^2] = \frac{r(r+1)}{\lambda^2}$$

$$= \frac{1(1+1)}{\lambda^2}$$

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

$$Var(T) = E[T^2] - (E[T])^2$$

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

$$= \frac{1}{\lambda^2}$$

∴ matches exponential mean
+ variance (textbook 15.4.2)