Homework_09

November 5, 2019

```
In [1]: # HIDDEN
        from datascience import *
        from prob140 import *
        import numpy as np
        import matplotlib.pyplot as plt
        plt.style.use('fivethirtyeight')
        %matplotlib inline
        from scipy import stats
        from sympy import *
        init_printing()
```

1 Homework 9

1.0.1 Instructions

Your homeworks 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. 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. Refer to the bottom of the notebook for submission instructions.

1.0.2 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.

1.0.3 1. Transformations of the Normal

Let Z have the standard normal density.

a) Find the density of $|Z|$. Don't try to use a formula. Just think about $P(|Z| \in dy)$ for positive y.

b) Find $E(|Z|)$. Calculating the integral is straightforward; you should just be able to recognize the indefinite integral.

c) [CODE] Complete the cell below to use `Plot_continuous` to plot the density of $|Z|$ and mark the expectation on the horizontal axis. In Data 8 you saw distributions of this shape when you simulated statistics such as "the absolute difference between the proportion of heads and $1/2$".

Start by defining a function `abs_norm_density` that takes a number y (positive or otherwise) as its argument and returns the density of $|Z|$ at y. Then plot the density over the interval $(-0.5, 4)$.

Assign $E(|Z|)$ to the name ev. Use `np.pi` for π.

1.0.4 Solution

1a) $f_{|Z|}(y) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$ for $y > 0$

1b) $\int_0^\infty y \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \Big|_0^\infty = 0 - \left(-\frac{2}{\sqrt{2\pi}}\right) = \sqrt{\frac{2}{\pi}}$

You could also recognize that $\int_0^\infty ye^{-\frac{1}{2}y^2} dz = 1$, since it is the total integral of the Rayleigh density.

In [2]: `#1c`

```
def abs_norm_density(y):
    if y > 0:
        return 2*stats.norm.pdf(y)
    else:
        return 0

Plot_continuous([-0.5, 4], abs_norm_density)

ev = (2/np.pi) **0.5
plt.scatter(ev, -0.01, marker = '^', color='red', s=50)

# Labels and axes; do not edit below this line
plt.xlabel('$y$')
plt.ylabel('Density at $y$')
plt.title('Density of $|Z|$')
plt.xlim(-0.6, 4)
plt.ylim(-0.02, 0.81);
```

Density of |Z|

d) A random variable whose log is normal is said to have a lognormal density. In the previous homework you examined the special case where the log is standard normal. Now let X have the normal (μ, σ^2) density and let $Y = e^X$. Use the change of variable formula for densities to find the density of Y.

1d) First note that the possible values of Y are in $(0, \infty)$. Here are the components of the change of variable formula.

- **The function:** Define $g(x) = e^x$. Note that g is smooth and increasing.
- **The derivative:** $g'(x) = e^x$.
- **The inverse function:** For $y > 0$ let $y = g(x) = e^x$. Now write x in terms of y, to get $x = \log(y)$.
- **The known density:** The density of X is $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

For $y > 0$ the change of variable formula says

$$f_Y(y) = f_X(x)\frac{1}{g'(x)} \quad \text{at } x = g^{-1}(y)$$

So for $y > 0$,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \frac{1}{e^x} \quad \text{at } x = \log(y)$$

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

3

e) [CODE] Complete the definitions in the cell below to get a plot of the lognormal density in the case $\mu = 0$ and $\sigma = 1$ and also (in gold) the lognormal density in the case $\mu = 0, \sigma = 0.5$. Use `np.log(w)` for $\log(w)$, and remember that `stats.norm.pdf(x, mu, sigma)` evaluates the normal density with mean μ and SD σ at the point x.

In [3]: `#1e`

```
# Density of e^Z where Z is standard normal
def lognorm_density_1(y):
    return (1/y) * stats.norm.pdf(np.log(y))

# Density of e^X where X is normal with mean 0 and SD 0.5
def lognorm_density_05(y):
    return (1/y) * stats.norm.pdf(np.log(y), 0, 0.5)

Plot_continuous([0.01, 7], lognorm_density_1)
Plot_continuous([0.01, 7], lognorm_density_05, color='gold')
plt.title('Lognormal Densities');
```
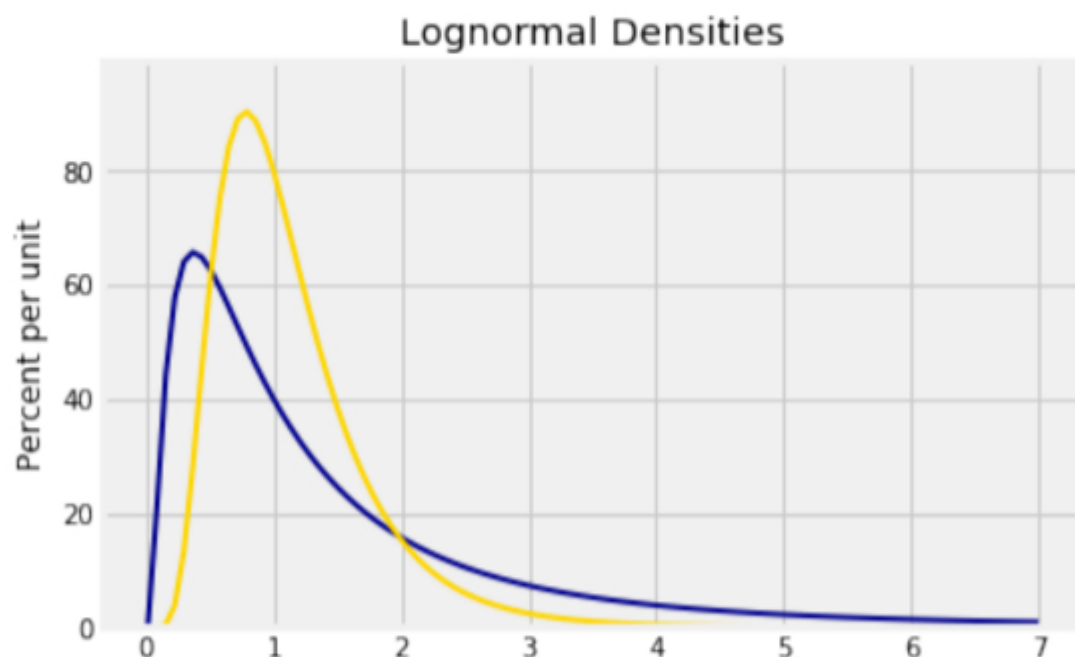

#newpage

1.0.5 2. Cauchy Density

Part (a) is from Pitman's text, page 310, where you'll find a useful diagram. But note that we are using Θ where he uses Φ, to avoid confusion with the standard normal cdf.

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.

1.0.6 Solution

2a) From the diagram, $Y = \tan(\Theta)$.

Let $y = g(\theta) = \tan(\theta)$

$-\infty < \tan(\theta) < \infty$ so the range of Y is $(-\infty, \infty)$

$\left| \frac{dg(\theta)}{d\theta} \right| = \left| \sec^2(\theta) \right| = \sec^2(\theta)$

$f_\Theta(\theta) = \frac{1}{\pi}$ for all $\theta \in (-\pi/2, \pi/2)$

By the change of variable formula,

$f_Y(y) = \frac{f_\Theta(\theta)}{\frac{dg(\theta)}{d\theta}} = \frac{1}{\pi \sec^2(\theta)}$ evaluated at $\arctan(y)$.
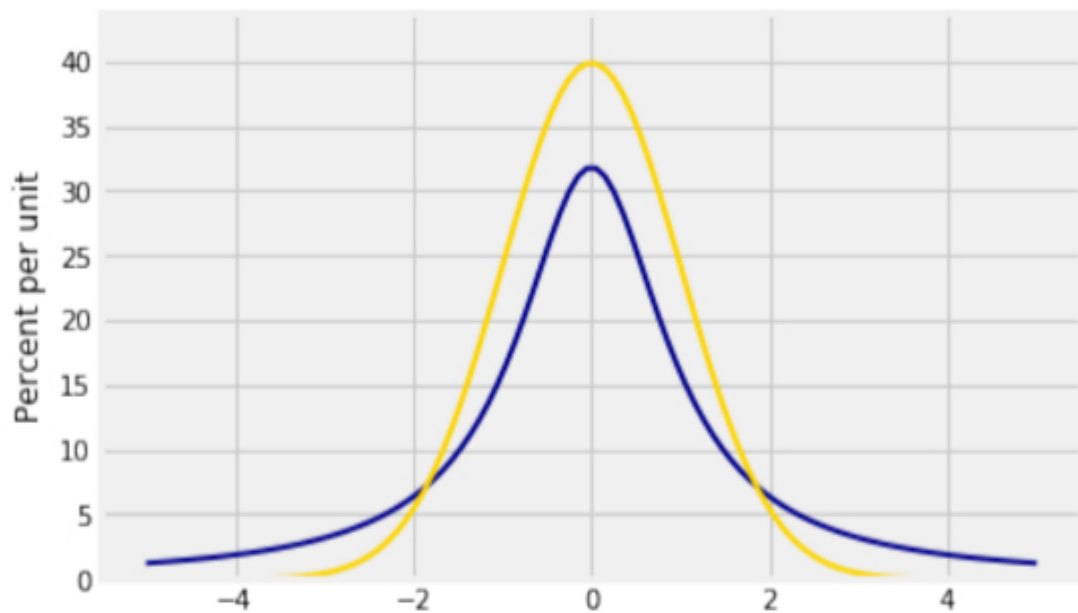
Using the diagram in the book, we see that $\sec(\theta) = \sqrt{\frac{1+y^2}{1}}$ so $\sec^2(\theta) = 1 + y^2$. So the angle whose tan is y is the angle whose $\sec^2$ is $1 + y^2$.

Hence $f_Y(y) = \frac{1}{\pi} \cdot \frac{1}{1+y^2} = \frac{1}{\pi(1+y^2)}$ for $y \in (-\infty, \infty)$

In [4]: #2b

```
# 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.

Plot_continuous([-5, 5], stats.cauchy.pdf)
Plot_continuous([-5, 5], stats.norm.pdf, color='gold')
```

5

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.

2c) $E(|Y|) = 2 \int_0^\infty \frac{y}{\pi(1+y^2)} dy = \int_0^\infty \frac{2y}{\pi(1+y^2)} dy$

Let $u = y^2$ so $du = 2y\,dy$. Substituting in u we get:

$\frac{1}{\pi} \int_0^\infty \frac{1}{1+u} du = \frac{1}{\pi} \log(1+u) \Big|_0^\infty = \infty$

2d) 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.]

```
In [5]: #2d: run this
        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);
```

6
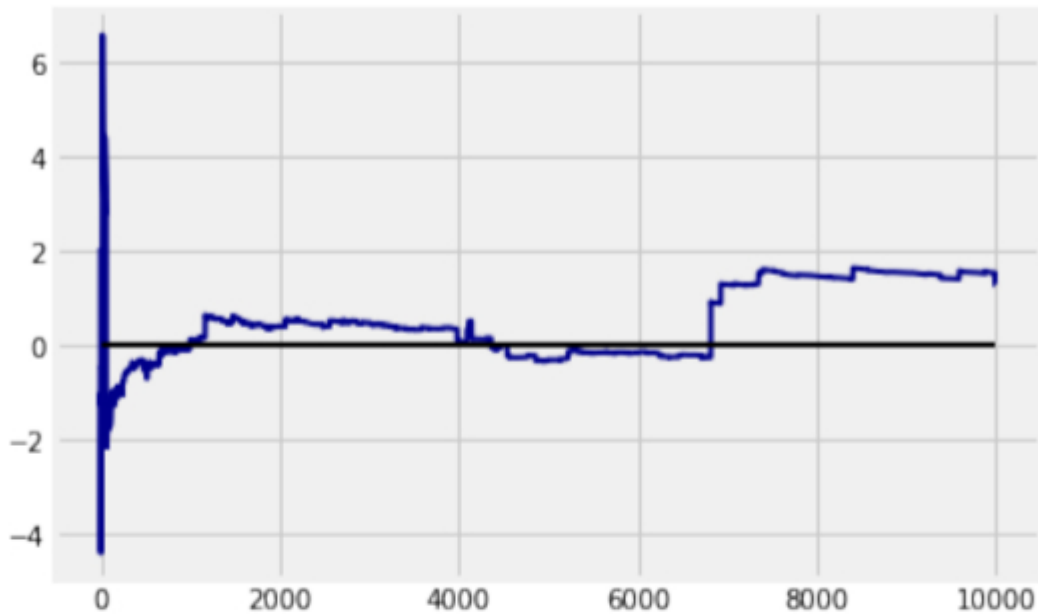
2d) There is a sequence of 10,000 i.i.d. Cauchy variables. The graph is plotting the successive sample means as the sample size increases. The graph doesn't settle down at all; the sample mean doesn't seem to get close to any number with high probability. The Weak Law doesn't apply because its conditions are violated: the expectation of the Cauchy is undefined.

#newpage

1.0.7   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. No SymPy.
a) Find $P(Y > 2X)$.
b) Find the marginal density of X.
c) Fill in the blanks (explain briefly): The joint density f above is the joint density of the _____ and _____ of ten independent uniform $(0,1)$ random variables.

1.0.8   Solution

3a) Draw the region. If you integrate y first and then x, then the limits of integration should be $2x$ to 1 for y and 0 to 0.5 for x.

$$\int_0^{0.5} \int_{2x}^1 90(y-x)^8 \, dy \, dx = \int_0^{0.5} \frac{90}{9}(y-x)^9 \Big|_{2x}^1 \, dx = \int_0^{0.5} 10[(1-x)^9 - x^9] \, dx = \frac{-10}{10}(1-x)^{10} - \frac{10}{10}x^{10} \Big|_0^{0.5} = \frac{511}{512}$$

3b) The possible values of X are $(0,1)$. For each fixed $x \in (0,1)$, draw the region to see that the limits of integration for y should be x to 1. For $x \in (0,1)$,

$$f_X(x) = \int_x^1 90(y-x)^8 dy = 10(y-x)^9 \big|_x^1 = 10(1-x)^9$$

3c) minimum; maximum

Let N be the minimum and M the maximum of 10 i.i.d. uniform $(0,1)$ random variables. For $0 < x < y < 1$,

$P(N \in dx, M \in dy)$

$\sim P(0$ variables before x, one variable in dx, 8 variables between x and y, one variable in dy, 0 variables after

This a multinomial probability:

$$\frac{10!}{0!1!8!1!0!} x^0 (1dx)^1 (y-x)^8 (1dy)^1 (1-y)^0 \sim 90(y-x)^8 dxdy$$

Since $P(N \in dx, M \in dy) \sim f_{N,M}(x,y)dxdy$ where $f_{N,M}$ is the joint density function of N and M, we have $f_{N,M}(x,y) = 90(y-x)^8$ for $0 < x < y < 1$.
#newpage

1.0.9   4. The Gamma Densities

In this problem you will start with some calculus exercises and then develop one of the funda-mental families of densities. 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

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

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

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

This is called the gamma (r, λ) density.

d) Use (c) and the formula in Section 16.3 of the Prob 140 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 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/λ

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

```
In [6]: 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();
```
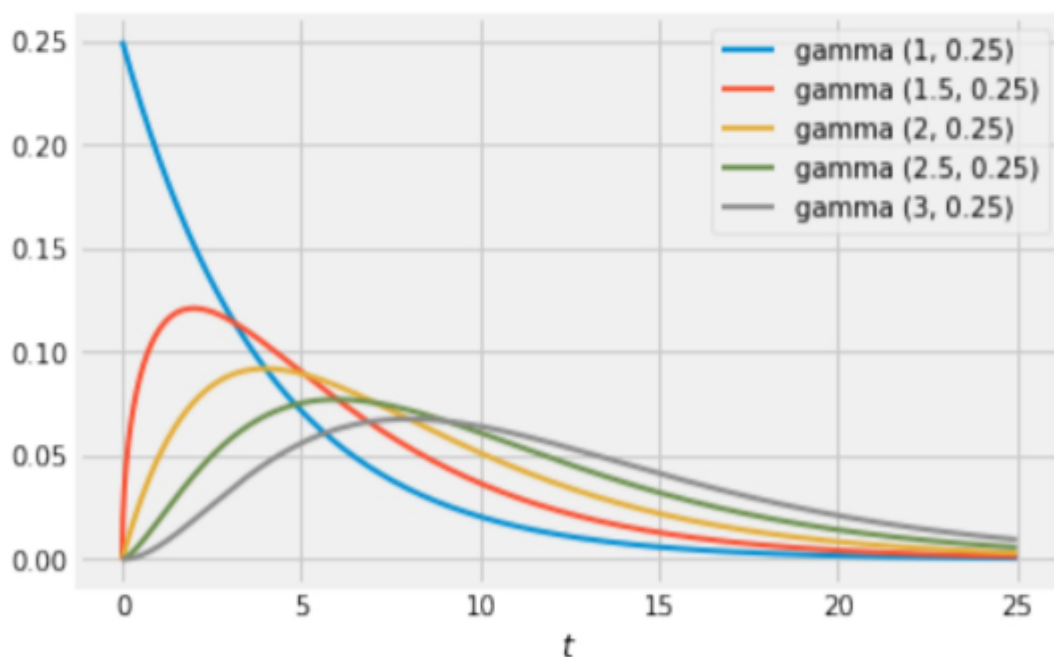
1.0.10   Solution

4a) $\Gamma(r+1) = \int_0^\infty t^r e^{-t} dt$

Use integration by parts: $u = t^r, du = rt^{r-1}, dv = e^{-t} dt, v = -e^{-t}$

$\Gamma(r+1) = \int_0^\infty t^r e^{-t} dt = -t^r e^{-t}\big|_0^\infty - \int_0^\infty -rt^{r-1}e^{-t}dt = 0 + r\int_0^\infty t^{r-1}e^{-t}dt = r\Gamma(r)$

4b) Base case: $\Gamma(1) = \int_0^\infty t^{1-1}e^{-t}dt = 1 = 0!$

Inductive Hypothesis: $\Gamma(n) = (n-1)!$

Inductive Step: Show $\Gamma(n+1) = n!$:

$\Gamma(n+1) = n\Gamma(n) = n(n-1)! = n!$

4c) The possible values of Y are $(0,\infty)$. Let $x > 0$ and let $y = g(x) = \frac{1}{\lambda}x$.

Then $x = g^{-1}(y) = \lambda y$ and $\frac{d}{dx}g(x) = \frac{1}{\lambda}$.

For $y > 0$, $f_Y(y) = \dfrac{f_X(x)}{1/\lambda}$ evaluated at $x = g^{-1}(y)$. So $f_Y(y) = \dfrac{\frac{1}{\Gamma(r)}(\lambda y)^{r-1}e^{-\lambda y}}{1/\lambda} = \frac{\lambda^r}{\Gamma(r)}y^{r-1}e^{-\lambda y}$.

4d) Let $W = Z^2$. Section 16.3 shows that for $w > 0$, $f_W(w) = \frac{1}{\sqrt{2\pi}}w^{-\frac{1}{2}}e^{-\frac{1}{2}w}$.

Thus for $w > 0$, $f_W(w) = cw^{r-1}e^{-\lambda w}$ for $r = 1/2$ and $\lambda = 1/2$. So W has the gamma $(1/2, 1/2)$ distribution.

To find the value of $\Gamma(1/2)$, examine the constant of integration. From the formula above, the constant is

$$\frac{1}{\sqrt{2\pi}} = \frac{\frac{1}{\sqrt{2}}}{\sqrt{\pi}} = \frac{2^{\frac{1}{2}}}{\sqrt{\pi}}.$$

By Part (c), the constant in the gamma $(1/2, 1/2)$ density is $\frac{2^{\frac{1}{2}}}{\Gamma(1/2)}$. So $\sqrt{\pi} = \Gamma(1/2)$.

By Part (a),

$\Gamma(3/2) = \Gamma(1/2+1) = (1/2)\Gamma(1/2) = \frac{1}{2}\sqrt{\pi}$

$\Gamma(5/2) = \Gamma(3/2+1) = (3/2)\Gamma(3/2) = \frac{3}{2}\cdot\frac{1}{2}\sqrt{\pi}$

#newpage

1.0.11   5. Mean and Variance of the Gamma

As in Question 4, let r and λ be positive constants.

  a) Use Question 4(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 4(a) to find a simple formula for $E(T)$ in terms of r and λ. Please don't leave any integrals in your answer.

  c) For T as in (b), find $E(T^2)$ and hence $Var(T)$. Again, there should be no integrals left in your answer.

  d) 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.

1.0.12   Solution

5a) $\int_0^\infty \frac{\lambda^r}{\Gamma(r)}t^{r-1}e^{-\lambda t}dt = 1$, so $\int_0^\infty t^{r-1}e^{-\lambda t}dt = \frac{\Gamma(r)}{\lambda^r}$

  5b) $E(T) = \int_0^\infty t\frac{\lambda^r}{\Gamma(r)}t^{r-1}e^{-\lambda t}dt = \frac{\lambda^r}{\Gamma(r)}\int_0^\infty t^r e^{-\lambda t}dt = \frac{\lambda^r}{\Gamma(r)}\frac{\Gamma(r+1)}{\lambda^{r+1}} = \frac{\lambda^r}{\Gamma(r)}\frac{r\Gamma(r)}{\lambda^{r+1}} = \frac{r}{\lambda}$

5c) $E(T^2) = \int_0^\infty t^2 \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} dt = \frac{\lambda^r}{\Gamma(r)} \int_0^\infty t^{r+1} e^{-\lambda t} dt = \frac{\lambda^r}{\Gamma(r)} \frac{\Gamma(r+2)}{\lambda^{r+2}} = \frac{\lambda^r}{\Gamma(r)} \frac{r(r+1)\Gamma(r)}{\lambda^{r+2}} = \frac{r^2+r}{\lambda^2}$

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

5d) The gamma $(1, \lambda)$ density at $t > 0$ is $\frac{\lambda^1}{\Gamma(1)} t^{1-1} e^{-\lambda t} = \lambda e^{-\lambda t}$ which is the exponential (λ) density.

Let X have the exponential (λ) density. Then

$E(X) = \frac{r}{\lambda} = \frac{1}{\lambda}$

$Var(X) = \frac{r}{\lambda^2} = \frac{1}{\lambda^2}$

#newpage

1.1 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.

1.1.1 Written Portion

- Scan all the pages into a PDF. You can use any scanner or a phone using an application. 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 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.

1.1.2 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 make a follow-up post on the general HW 9 Piazza thread.

1.1.3 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.**

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