MAST20006 Probability for Statistics /MAST90057 Elements of Probability

Assignment 5, Semester 1 2020

Due time: 4pm, Friday June 5.

Name:

Student ID:

New completion process

- Note this assignment is being handled using a similar process to that now planned for the final exam so you can become familiar with it.

  To complete this assignment, you need to write your solutions into the blank answer spaces following each question in this assignment PDF.

- If you have a printer (or can access one), then you should print out the assignment template, handwrite your solutions into the answer spaces and then scan your assignment to a PDF file using a scanning app on your mobile phone for upload.

- If you do not have a printer and you know how to annotate a PDF using an iPad/Android tablet/Graphics tablet, then annotate your answers directly onto the assignment PDF and save a copy for submission.

- If unfortunately you cannot manage to complete using either of the above methods, you may handwrite your answers as normal on blank paper and then scan for submission.

  In order to submit this way you will need to **mimic the template** by numbering your blank pages following the template as closely as possible, and noting which answers to write on each of those pages. Remember to keep the two cover pages in your submission with just your name and student ID stated on the first page.

  Also you can note on pages where you run out of room that further working is appended and include additional working pages **but only at the end** after the template pages.

- Whether you complete on paper or by annotating, if you find you are unable to answer the whole question in the answer space provided then you can append additional handwritten solutions to the **end** of your templated assignment. If you do this you **MUST** make a note in the correct answer space or page for the question, warning the marker that you have appended additional remarks at the end.

- When finished, submit your assignment PDF to GradeScope below by first selecting your PDF file and then clicking on 'Upload PDF'.

- A **poor presentation penalty** of 10% will apply unless your submitted assignment meets all of the following requirements:

  - it is a single PDF with all pages in correct template order and the correct way up, and with any blank pages with additional working added only at the end of the template pages;
  - has all pages clearly readable;
  - has all pages cropped to the A4 borders of the original page and is imaged from directly above to avoid excessive 'keystoning'

  These requirements are easy to meet if you use a scanning app on your phone and take some care with your submission - please review it before submitting to double check you have satisfied all of the above requirements.

- The **strict** submission deadline is **4pm Melbourne time on Friday 5 June**. You have longer than of the normal one week to complete this assignment. Consequently late assignments will **NOT** be accepted. We recommend you submit at least a day before the due date to avoid any technical delays. If there are extenuating, eg medical, circumstances, contact the Tutorial Coordinator.

- There are **5 assignment questions** and **6 Maple project questions**, of which some randomly chosen questions will be marked. The assignment part is worth 4% of the total mark and the Maple project is worth 10% of the total mark. Note you are expected to submit answers to **all** questions, otherwise **a mark penalty will apply**.

- Please attach your Maple worksheet at the end of your assignment. **Your first line of code should be your student ID number**. For instructions on how to save a Maple worksheet as a PDF file, visit `https://www.maplesoft.com/support/help/Maple/view.aspx?path=worksheet%2Fmanaging%2FexportPDF`

  For instruction on how to merge your assignment PDF and the PDF of your Maple worksheet into a single PDF document, use for example this online tool: `https://smallpdf.com/merge-pdf`

- Working and reasoning **must** be given to obtain full credit. Give clear and concise explanations. Clarity, neatness, and style count.

* * *

1. Let $X_1, X_2, \ldots, X_n$ be a random sample (i.e. $X_1, X_2, \ldots, X_n$ are independent) of size n from a geometric distribution with success probability $p = 0.8$.

   (a) Find the mgf $M_{Y_4}(t)$ of $Y_4 = X_1 + X_2 + X_3 + X_4$ using the geometric mgf. Then name the distribution of Y_4.

Solution:

$$M_{Y_4}(t) = (M_{X_1}(t))^4 = \left(\frac{0.8e^t}{1 - 0.2e^t}\right)^4, \quad t < \ln(5).$$

So $Y_4 \stackrel{d}{=} \text{NegBin}(r = 4, p = 0.8)$.

   (b) Find the mgf $M_{Y_n}(t)$ of $Y_n = X_1 + X_2 + \cdots + X_n$. Then name the distribution of Y_n.

Solution:

$$M_{Y_n}(t) = (M_{X_1}(t))^n = \left(\frac{0.8e^t}{1 - 0.2e^t}\right)^n, \quad t < \ln(5).$$

So $Y_n \stackrel{d}{=} \text{NegBin}(r = n, p = 0.8)$.

(c) Find the mgf $M_{\bar{Y}_n}(t)$ of the sample mean $\bar{Y}_n = \frac{Y_n}{n}$.

Solution:

$$\begin{aligned} M_{\bar{Y}_n}(t) &= E(e^{t(Y_n/n)}) \\ &= M_{Y_n}(t/n) \\ &= \left(\frac{0.8e^{t/n}}{1-0.2e^{t/n}}\right)^n, \quad t < n\ln(5). \end{aligned}$$

(d) Find the limit $\lim_{n\to\infty} M_{\bar{Y}_n}(t)$ using the result of (c). What distribution does the limiting mgf correspond to?

Solution:

First, using Taylor expansion $e^{ax} = 1 + ax + o(ax)$ and the result $\lim_{n\to\infty}[1 + an^{-1} + o(n^{-1})]^{bn} = e^{ab}$,

$$\begin{aligned}
\lim_{n\to\infty}\left(\frac{0.8e^{\frac{t}{n}}}{1-0.2e^{\frac{t}{n}}}\right)^{n} &= \lim_{n\to\infty}\left(\frac{5}{4}e^{-\frac{t}{n}}-\frac{1}{4}\right)^{-n} \\
&= \lim_{n\to\infty}\left(1-\frac{5t}{4n}+o(n^{-1})\right)^{-n} \\
&= e^{\frac{5}{4}t}.
\end{aligned}$$

The above result can also be obtained by using l'Hospital's rule or Maple.

Thus, $\lim_{n\to\infty} M_{\bar{Y}_n}(t) = \lim_{n\to\infty}\left(\frac{0.8e^{t/n}}{1-0.2e^{t/n}}\right)^{n} = e^{(5/4)t}$.

The function $e^{(5/4)t}$ is the mgf of a degenerate distribution where the random variable has only one possible value, $5/4$. This suggests that the sample mean $\bar{Y}_n$ converges in probability to $5/4$ which is the mean of geometric distribution with success probability $p = 0.8$. This results from the Law of Large Numbers.

(e) Let

$$Z_n = \sqrt{n}\left(\frac{\bar{Y}_n - \frac{5}{4}}{\frac{\sqrt{5}}{4}}\right) = \frac{4}{5}\sqrt{5n}\,\bar{Y}_n - \sqrt{5n}.$$

Find $M_{Z_n}(t)$, the mgf of Z_n. Then use Maple or a theoretical argument to find the limiting mgf $\lim_{n\to\infty} M_{Z_n}(t)$. What is the limiting distribution of Z_n?

Solution:

$$\begin{aligned} M_{Z_n}(t) &= E\left(\exp\left\{\frac{4}{5}\sqrt{5n}\bar{Y}_n\,t - \sqrt{5n}\,t\right\}\right) \\ &= e^{-\sqrt{5n}\,t} M_{\bar{Y}_n}(\frac{4}{5}\sqrt{5n}t) \\ &= e^{-\sqrt{5n}\,t}\left(\frac{0.8e^{4\sqrt{5n}t/(5n)}}{1-0.2e^{4\sqrt{5n}t/(5n)}}\right)^n, \quad t < \frac{\sqrt{5n}}{4}\ln(5). \end{aligned}$$

Using Maple, we obtain

$$\begin{aligned} \lim_{n\to\infty} M_{Z_n}(t) &= \lim_{n\to\infty} e^{-\sqrt{5n}\,t}\left(\frac{0.8e^{4\sqrt{5n}t/(5n)}}{1-0.2e^{4\sqrt{5n}t/(5n)}}\right)^n \\ &= e^{\frac{1}{2}t^2}, \end{aligned}$$

which is the mgf of $N(0,1)$. Therefore, the limiting distribution of Z_n is standard normal $N(0,1)$. This results from the central limit theorem.

2. Let X_1, X_2 and X_3 be a random sample of size $n = 3$ from the exponential distribution with pdf $f(x) = \frac{1}{2}e^{-x/2}$, $0 < x < \infty$. Find

   (a) $P(1 < X_1 < 2,\ 2 < X_2 < 3,\ 3 < X_3 < 4)$.

Solution:

$$P(1 < X_1 < 2,\ 2 < X_2 < 3,\ 3 < X_3 < 4)$$

$$= P(1 < X_1 < 2)P(2 < X_2 < 3)P(3 < X_3 < 4)$$

$$= \int_1^2 \frac{1}{2}e^{-x_1/2}dx_1 \cdot \int_2^3 \frac{1}{2}e^{-x_2/2}dx_2 \cdot \int_3^4 \frac{1}{2}e^{-x_3/2}dx_3$$

$$= (e^{-0.5} - e^{-1})(e^{-1} - e^{-1.5})(e^{-1.5} - e^{-2})$$

$$= 0.00303.$$

(b) $E[X_1 X_2^2 (X_3 - 2)^2]$.

Solution:

- $E(X_1) = E(X_2) = E(X_3) = \int_0^\infty x \cdot \frac{1}{2} e^{-x/2} dx = 2.$

- $E(X_1^2) = E(X_2^2) = E(X_3^2) = \int_0^\infty x^2 \cdot \frac{1}{2} e^{-x/2} dx = 8.$

- $E[(X_1 - E(X_1))^2] = E[(X_2 - E(X_2))^2] = E[(X_3 - E(X_3))^2] = 8 - 2^2 = 4.$

- So

$$\begin{aligned} E[X_1 X_2^2 (X_3 - 2)^2] &= E[X_1] E[X_2^2] E[(X_3 - 2)^2] \\ &= 2 \cdot 8 \cdot 4 \\ &= 64 \end{aligned}$$

(using independence of X_1, X_2, X_3).

3. Let X_1, X_2, X_3 denote a random sample of size $n = 3$ from a distribution with the Poisson pmf

$$f(x) = \frac{2^x}{x!}e^{-2}, \quad x = 0, 1, 2, 3, \ldots.$$

(a) Compute $P(X_1 + X_2 + X_3 = 1)$.

Solution:

The only possible values that (X_1, X_2, X_3) can take so that $X_1 + X_2 + X_3 = 1$ are (1,0,0), (0,1,0) and (0,0,1).

Hence

$$\begin{aligned} & P(X_1 + X_2 + X_3 = 1) \\ = \; & P(X_1 = 1, X_2 = 0, X_3 = 0) + P(X_1 = 0, X_2 = 1, X_3 = 0) \\ & + P(X_1 = 0, X_2 = 0, X_3 = 1) \\ = \; & 3P(X_1 = 1)P(X_1 = 0)^2 = 3 \cdot \frac{2^1}{1!}e^{-2}\left(\frac{2^0}{0!}e^{-2}\right)^2 \\ = \; & 6e^{-6} \\ = \; & 0.01487. \end{aligned}$$

(b) Find the moment-generating function of $Z = X_1 + X_2 + X_3$ using the Poisson mgf of X_1. Then name the distribution of Z.

Solution:

$M_{X_1}(t) = M_{X_2}(t) = M_{X_3}(t) = e^{2(e^t-1)}$.

So the mgf for Z is

$$M_Z(t) = M_{X_1}(t)M_{X_2}(t)M_{X_3}(t) = \left(e^{2(e^t-1)}\right)^3 = e^{6(e^t-1)}.$$

Therefore, $Z \stackrel{d}{=} \text{Poisson}(\lambda = 6)$.

(c) Find the probability $P(X_1 + X_2 + X_3 = 4)$ using the result of (b).

Solution:

$$\begin{aligned} P(X_1 + X_2 + X_3 = 4) &= P(Z = 4) \\ &= \frac{6^4}{4!}e^{-6} \\ &= 54e^{-6} \\ &= 0.1339. \end{aligned}$$

(d) If $Y = \max\{X_1, X_2, X_3\}$, find the probability $P(Y \le 2)$.

Solution:

$$\begin{aligned}
P(Y \le 2) &= P(\max\{X_1, X_2, X_3\} \le 2) \\
&= P(X_1 \le 2,\ X_2 \le 2,\ X_3 \le 2) \\
&= P(X_1 \le 2)^3 \\
&= \left[\frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} + \frac{2^2}{2!}e^{-2}\right]^3 \\
&= 125e^{-6} \\
&= 0.3098
\end{aligned}$$

4. If $E(X) = 16$ and $E(X^2) = 292$, use Chebyshev's inequality to determine

   (a) A lower bound for $P(8 < X < 24)$.

Solution:

$\sigma^2 = \text{Var}(X) = 292 - 16^2 = 36.$

So we have

$$\begin{aligned} P(8 < X < 24) &= P(|X - 16| < 8) \\ &= P(|X - 16| < \frac{8}{6} \cdot 6) \\ &\geq 1 - \frac{1}{(8/6)^2} \\ &= \frac{7}{16} \end{aligned}$$

(b) An upper bound for $P(|X-16| \geq 18)$.

Solution:

$$\begin{aligned} P(|X-16| \geq 18) &= P(|X-16| \geq 3 \cdot 6) \\ &\leq \frac{1}{3^2} = \frac{1}{9}. \end{aligned}$$

5. Suppose that the distribution of the weight of a prepackaged "1-pound bag" of carrots is $N(1.18, 0.07^2)$ and the distribution of the weight of a prepackaged "3-pound bag" of carrots is $N(3.22, 0.09^2)$. Now independently select at random three 1-pound bags of carrots with weights being X_1, X_2 and X_3 respectively. Also randomly select one 3-pound bag of carrots with weight being W. Let $Y = X_1 + X_2 + X_3$.

   (a) Find the mgf of Y.

Solution:

$$\begin{aligned} M_Y(t) &= M_{X_1}(t)^3 \\ &= [\exp(1.18t + \frac{1}{2}0.07^2t^2)]^3 \\ &= \exp(3.54t + 0.00735t^2). \end{aligned}$$

(b) Find the distribution of Y, the total weight of the three 1-pound bags of carrots selected.

Solution:

$$Y \stackrel{d}{=} N(3 \cdot 1.18, 3 \cdot 0.07^2) = N(3.54, 0.0147).$$

(c) Find the probability $P(Y < W)$, i.e., the probability that the sum of weights of three 1-pound bags randomly selected is smaller than the weight of one 3-pound bag randomly selected.

Solution:

$$Y - W \stackrel{d}{=} N(3.54 - 3.22, 0.0147 + 0.09^2) = N(0.32, 0.0228).$$

So

$$\begin{aligned} P(Y < W) &= P(Y - W < 0) \\ &= P\left(\frac{Y - W - 0.32}{\sqrt{0.0228}} < \frac{0 - 0.32}{\sqrt{0.0228}}\right) \\ &= P(Z < -2.1193) \\ &= 0.017 \end{aligned}$$

MAPLE QUESTIONS

1. Please always `simplify` your final results in doing the following questions.
2. Use `unassign('obj')` if you want to remove an assigned `obj`.
3. Use `restart` to clean up your worksheet if it seems "messed up" and you get lost.
4. After `restart`, don't forget typing `with(Statistics):`

* * *

1. (*Banach Matchbox Problem*) A cigarette smoker carries two matchboxes, one in his right pocket and one in his left pocket. Whenever he wants to smoke, he selects a pocket at random and takes a match from the box in that pocket. Suppose each box contains 100 matches initially. Let X be the number of matches left in a box when the smoker for the first time discovers that the other box is empty. It can be shown that the pmf of X is

$$P(X = x) = \binom{200-x}{100} 2^{x-200}, \quad x = 0, 1, 2, \ldots, 100.$$

Complete the following tasks and keep 4 significant digits after the decimal point in your answers.

(a) Find the probability $P(X \geq 10)$.

Solution:

$P(X \geq 10) = 0.4694$

(b) Find the mean $E(X)$.

Solution:

$E(X) = 10.3260$

(c) Find $E(X^2)$.

Solution:

$E(X^2) = 169.0219$

(d) Find $E[(X+1)^{-2}]$.

Solution:

$E[(X+1)^{-2}] = 0.08862$

2. Let a continuous random variable X have the following pdf

$$f(x) = \frac{2}{9}(x+1)(2-x), \quad -1 < x < 2.$$

(a) Find the cdf $F(x)$ of X.

Solution:

$$F(x) = \begin{cases} 0, & x \le -1 \\ \frac{4}{9}x + \frac{7}{27} - \frac{2}{27}x^3 + \frac{1}{9}x^2 = \frac{1}{27}(7-2x)(x+1)^2, & -1 < x \le 2 \\ 1, & x > 2 \end{cases}$$

(b) Find the probability $P(-3.5 < X < 1.0)$.

Solution:

$P(-3.5 < X \leq 1.0) = \frac{20}{27} = 0.7407.$

(c) Find the mean $E(X)$.

Solution:

$E(X) = 1/2.$

(d) Find the mgf $M(t) = E[\exp(tX)]$.

Solution:

$M(t) = \frac{2}{9t^3}(3e^{-t}t + 2e^{-t} + 3e^{2t}t - 2e^{2t}), \quad -\infty < t < \infty.$

(e) Find the third moment $E(X^3)$.

Solution:

$E(X^3) = 4/5.$

(f) Let $Y = X^2$. Find the range of Y and the pdf $g(y)$ of Y.

Solution: $0 \leq Y < 4$

$$g(y) = \begin{cases} \frac{4}{9\sqrt{y}} - \frac{2}{9}\sqrt{y}, & 0 \leq y < 1 \\ \frac{2}{9\sqrt{y}} - \frac{1}{9}\sqrt{y} + \frac{1}{9} = \frac{(\sqrt{y}+1)(2-\sqrt{y})}{9\sqrt{y}}, & 1 \leq y < 4 \\ 0, & \text{elsewhere.} \end{cases}$$

3. Consider continuous random variables X and Y which have the following joint pdf

$$f(x, y) = x + y, \quad 0 < x < 1,\ 0 < y < 1.$$

(a) Find the marginal pdf $f_1(x)$ of X.

Solution:

$f_1(x) = \int_0^1 (x + y)dy = x + \frac{1}{2}, \quad 0 < x < 1.$

(b) Find the mean $E(X)$.

Solution:

$E(X) = \int_0^1 x(x + \frac{1}{2})dx = \frac{7}{12}.$

(c) Find the variance $\mathrm{Var}(X)$.

Solution:

$\mathrm{Var}(X) = \int_0^1 (x - \frac{7}{12})^2 (x + \frac{1}{2}) dx = \frac{11}{144}$.

(d) Find the covariance $\mathrm{Cov}(X, Y)$. (Note that $E(X) = E(Y)$.)

Solution:

$\mathrm{Cov}(X, Y) = \int_0^1 \int_0^1 (x - \frac{7}{12})(y - \frac{7}{12})(x + y) dx dy = -\frac{1}{144}$

(e) Find the conditional pdf $h(y|x)$ of Y given $X = x$.

Solution:

$h(y|x) = \frac{x+y}{x+\frac{1}{2}}, \quad 0 < y < 1.$

(f) Find the condition probability $P(Y \le \frac{1}{2} | X = \frac{1}{3})$.

Solution:

$P(Y \le \frac{1}{2} | X = \frac{1}{3}) = \int_0^{1/2} h(y|x = \frac{1}{3}) dy = \frac{7}{20}$.

(g) Find the probability $P(\frac{1}{2} < X + Y < \frac{3}{2})$.

Solution:

$$\begin{aligned} P(\frac{1}{2} < X+Y < \frac{3}{2}) &= 1 - \int_0^{1/2}\int_0^{-x+1/2}(x+y)dydx \\ &= -\int_{1/2}^{1}\int_{-x+3/2}^{1}(x+y)dydx \\ &= 1 - \frac{1}{24} - \frac{5}{24} \\ &= \frac{3}{4}. \end{aligned}$$

4. Let X have the normal distribution $N(\mu, \sigma^2)$ and let $Y = e^X$.

(a) Find the range of Y and the pdf g(y) of Y .

Solution:

$Y > 0$,

$$g(y) = \begin{cases} 0 & y \leq 0 \\ \frac{1}{\sqrt{2}\,\pi\sigma\,y} e^{-\frac{(\ln(y)-\mu)^2}{2\,\sigma^2}} & y > 0. \end{cases}$$

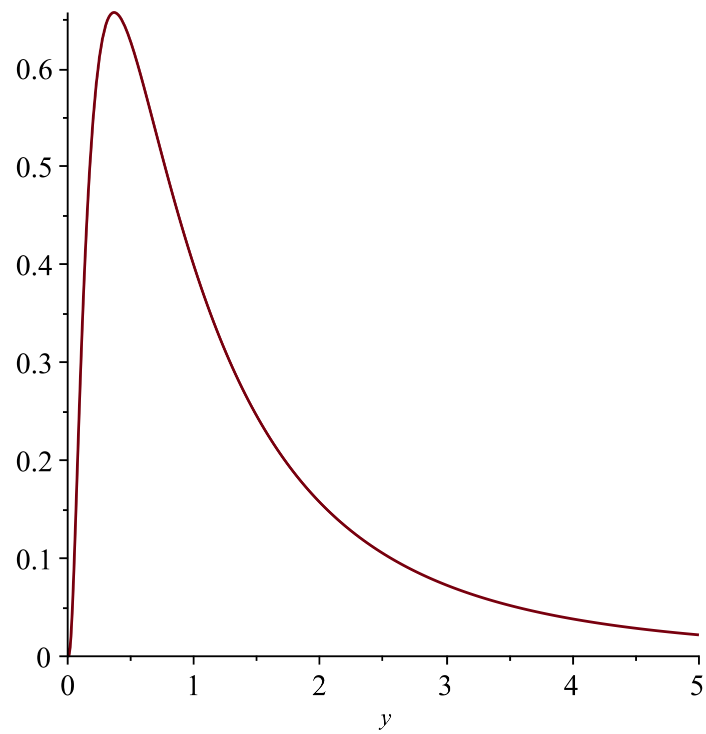
(b) Find the third moment of Y, $E[Y^3]$.

Solution:

$$E[Y^3] = e^{9/2\,\sigma^2+3\,\mu}$$

(c) *In the next four subquestions, we assume that $\mu = 0$ and $\sigma = 1$.*
Sketch the graph of the pdf of Y for $0 < y \leq 5$ (use Maple to generate the graph and copy it the best you can in the answer box).

Solution:



(d) What is the mean of Y? Keep 3 significant digits after the decimal point in your result.

Solution:

$E[Y] = e^{1/2} = 1.649$

(e) What is the median of Y? What can you conclude about $P(Y \leq 1)$?

Solution:

$m = 1$, which implies $P(Y \leq 1) = 0.5$.

(f) What is $P(Y > 2)$? Keep 3 significant digits after the decimal point in your result.

Solution:

$P(Y > 2) = 1 - P(Y \leq 2) = 0.244.$

5. Let X be a continuous random variable with the density function

$$f(x) = \begin{cases} \frac{2}{x^2}, & 1 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find the kth moment of X for $k \geq 1$.

Solution:

For $k = 1$, $E[X^k] = E[X] = 2\ln(2)$, and for $k \geq 2$, $E[X^k] = \frac{2^k - 2}{k-1}$.

(b) Find the cdf of X and explain how you could use it to simulate a realisation/observation from X.

Solution:

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

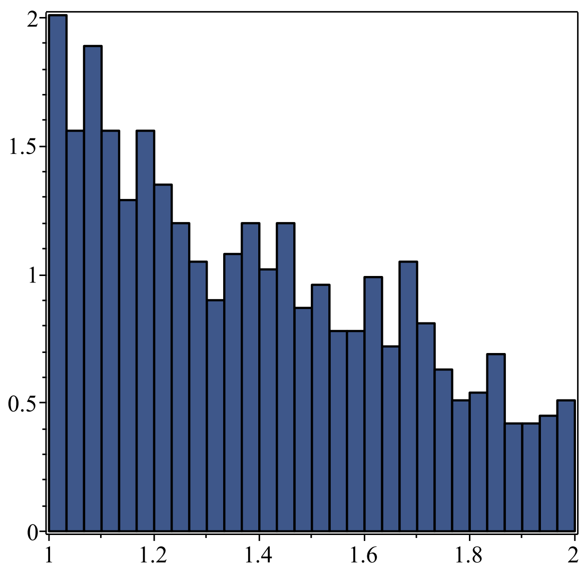
The inverse of $F_X(x)$ on $1 < x < 2$, is $F^{-1}(y) = 2/(2-y)$ for $0 < y < 1$. So an observation x from X can be obtained by generating an observation y from a uniform distribution on $(0, 1)$ and letting $x := F^{-1}(y)$.

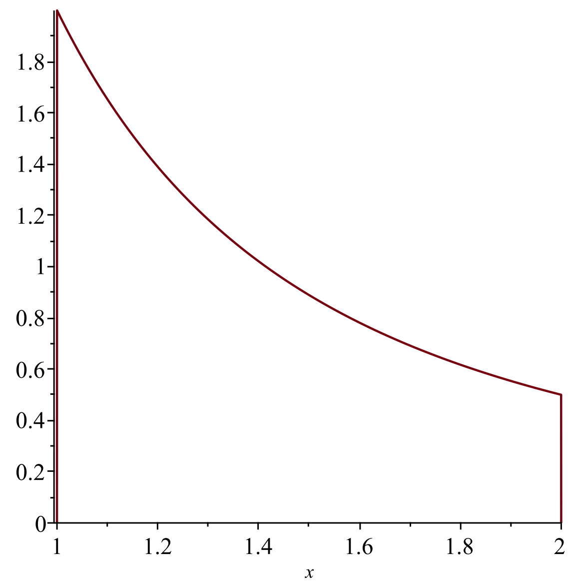
(c) Generate a sample of 1000 observations from the distribution of X and plot the corresponding histogram. Compare the shape of the histogram with that of the density of X.

For this question, sketch the graph of the density and write your comment on the comparison of this graph with the histogram in the box below; refer to your attached code for the histogram.

Solution:

The left graph shows the histogram and the right graph shows the density. The histogram is a fair approximation of the density.





(d) Let $Z = \ln(X)$. Find the mgf $M_Z(t)$ of Z. For which values of t is it well defined? Compare its expression with your answer to (b) and comment.

Solution:

$M_Z(t) = \frac{2^t-2}{t-1} = E[e^{t\ln(X)}] = E[X^t]$. It is well defined for all t.

(e) Find $E[Z]$.

Solution:

$$E[Z] = -\ln(2) + 1$$

6. Let X be a Gamma distribution with mean 8 and variance 32.

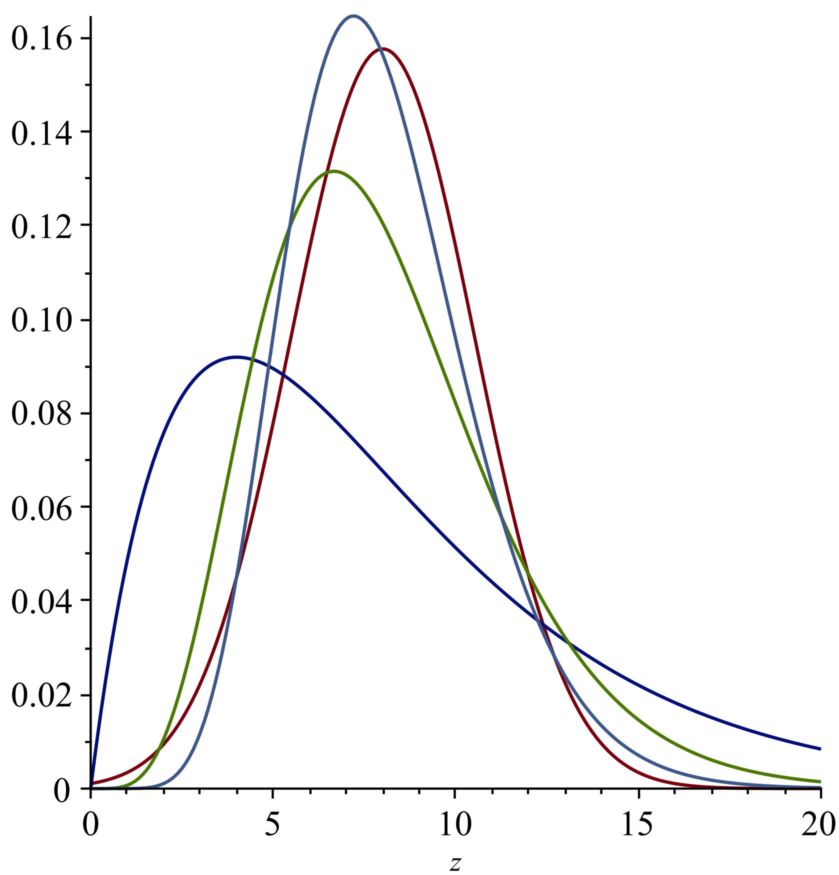
   Let $X_1, X_2, \ldots, X_n$ be n independent random variables with the same distribution as X. Let $Y_n = \sum_{i=1}^{n} X_i/n$ be the sample mean.

   Sketch the graphs of the pdf of X and the pdf of Y_n for $n = 3, 5$. What do you observe? Can you compare Y_n to a known distribution when n is large? Elaborate on your answer.

Solution:

A Gamma distribution with mean 8 and variance 32 has parameters $\theta = 2$ and $\alpha = 4$.

The dark blue curve is the pdf of X, the green curve is the density of Y_3, the blue curve is the density of Y_5 and the red curve is the density of a normal distribution with mean 8 and variance $32/n$.



For large n, Y_n becomes close to the normal distribution with mean 8 and variance $32/n$.