

Assignment 3

Due date: Monday, November 2, 2026.

Total: 30 points

Coverage: [Chapter 4: Generating Random Variables](#).

Question	Format	Points
1	8 multiple-choice questions on random variable generation	8
2	8 multiple-choice questions on constructing and checking random variables	8
3	6 true/false questions	6
4	2 short-answer parts with four prompts each	8
Total		30

[Assignment PDF](#) | [Response template \(.qmd\)](#)

Submission

Use **R Markdown** or **Quarto**. Submit your **source file** (.Rmd or .qmd) and **rendered PDF** to the **Assignment 3** folder on **iCollege**. Name both files A3_LastName_FirstName with the appropriate extension, and include your name in the PDF.

Use the response template, or follow its answer table and question labels.

- **Questions 1-2:** Select one best answer, A, B, C, or D.
- **Question 3:** Write **True** or **False**.
- **Question 4:** Give a short explanation or calculation for each prompt.

Each item is worth **1 point**. Only Question 4 requires explanations or calculations.

R code and plots are optional. If you use R, include the code in your source file. Before generating random values, set a seed (e.g., `set.seed(8670)`) and keep that line in your code. Check that your source renders to PDF without errors.

Note: The questions use the main material in Chapter 4. Section 4.10 (**Supplementary**) and topics marked **optional** are not required. Normalization questions cover only the basic definitions and whether each form can be used for acceptance-rejection sampling.

Notation: $N(\mu, \sigma^2)$ uses variance as its second parameter. $\text{Gamma}(a, b)$ uses shape a and rate b .

Question 1: Generating Random Variables [8]

Select the **best answer**. Each item is worth **1 point**.

1a)

Consider the following R code:

```
x <- numeric(100)
for (i in seq_len(100)) {
  set.seed(8670)
  x[i] <- runif(1)
}
```

What is the main problem with using x as 100 independent uniform draws?

- A. The values must all be zero.
- B. The values are sampled without replacement.
- C. Resetting the seed repeats the same first draw in every iteration.
- D. A uniform generator cannot be used inside a loop.

1b)

Let $U \sim \text{Uniform}(0, 1)$. Which expression generates an exponential random variable with **mean 5**?

- A. $-\log(1 - U)/5$
- B. $-5 \log(1 - U)$
- C. $\exp(-5U)$
- D. $5U$

1c)

A normalized probability density must have:

- A. Maximum height 1.
- B. Mean 1.
- C. Variance 1.
- D. Total area 1.

1d)

Which statement best describes an unnormalized target?

- A. It has the same shape as the desired density, but its total area need not be 1.
- B. Its total area must always be 1.
- C. It must take negative values.
- D. Its maximum height must be 1.

1e)

Which statement about acceptance-rejection sampling is correct?

- A. The target must be normalized.
- B. The target must be unnormalized.
- C. Either form works, provided a valid envelope is used.
- D. An envelope is unnecessary if the target density is known.

1f)

A rejection sampler proposes $Y = 0.75$ and draws an independent $U = 0.60$. The acceptance probability for this proposal is 0.75. What should the sampler do?

- A. Accept and store 0.75.
- B. Accept and store 0.60.
- C. Reject because $U < 0.75$.
- D. Accept and store both values as draws from the target.

1g)

A rejection sampler uses normalized target and proposal densities with a valid bound $M = 2.5$. What is the expected number of proposals needed for 400 accepted observations?

- A. 160
- B. 400
- C. 600
- D. 1,000.

1h)

An envelope lies above the target at 100 grid points. What else must be checked before using it for rejection sampling? Assume the proposal covers the target support.

- A. The number of proposals is at least 100.
- B. The envelope is at least as high as the target throughout its support.
- C. The envelope has total area exactly 1.
- D. Every proposal is accepted.

Question 2: Constructing and Checking Random Variables [8]

Select the **best answer**. Each question is worth **1 point**.

2a)

Let $X_1 \sim \text{Poisson}(1.2)$ and $X_2 \sim \text{Poisson}(2.8)$ be independent. What is the distribution of $X_1 + X_2$?

- A. $\text{Poisson}(1.6)$
- B. $\text{Poisson}(3.36)$
- C. $\text{Poisson}(4)$
- D. $\text{Binomial}(4, 0.5)$

2b)

Suppose $G_1 \sim \text{Gamma}(2, 4)$ and $G_2 \sim \text{Gamma}(5, 4)$ are independent, with the second parameter denoting the rate. What is the distribution of $G_1/(G_1 + G_2)$?

- A. $\text{Beta}(2, 5)$
- B. $\text{Beta}(5, 2)$
- C. $\text{Gamma}(7, 4)$
- D. $\text{Unif}(0, 1)$

2c)

Let $Z_1, \dots, Z_6$ be independent standard normals. A programmer returns

$$W = \frac{Z_1}{\sqrt{(Z_1^2 + \dots + Z_6^2)/6}}$$

and calls the result a t_6 draw. Which statement identifies the error?

- A. The sum of six squared standard normals cannot be chi-square.
- B. A t generator requires a uniform numerator.
- C. Dividing the sum of squares by 6 is incorrect.
- D. The numerator is not independent of the chi-square denominator.

2d)

The target is a sample of independent pairs (X_1, X_2) from a mixture of two bivariate normal distributions, with probabilities 0.30 and 0.70. How should each pair be generated?

- A. Choose a separate component label for each coordinate, then draw from the corresponding marginal distributions.
- B. Choose one component with probabilities 0.30 and 0.70, then draw the whole pair from that component's bivariate distribution. Use a fresh label for each pair.
- C. Draw a pair from each component and return their coordinatewise weighted average.
- D. Choose one component once and use it for every pair in the dataset.

2e)

A discrete random variable takes values -2 , 1 , and 5 with probabilities 0.20 , 0.50 , and 0.30 . Using $F^{-1}(u) = \inf\{x : F(x) \geq u\}$, which value is returned when $u = 0.70$?

- A. 1
- B. -2
- C. 5
- D. No value is defined at a jump.

2f)

Each row of an $n \times 2$ matrix Z contains two independent standard normal draws. Let

$$R = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}.$$

If $X \leftarrow Z \%*\% R$, what is the covariance matrix of a row of X ?

- A. $\begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix}$
- B. $\begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$
- C. $\begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$
- D. $\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$

2g)

To generate a pair of independent standard normals, a programmer draws one $U \sim \text{Unif}(0, 1)$ and returns $X_1 = \Phi^{-1}(U)$ and $X_2 = \Phi^{-1}(U)$, where Φ is the standard normal CDF. Which conclusion is correct?

- A. The pair is correct because both marginal distributions are standard normal.
- B. Both marginals are correct, but the coordinates are equal and hence not independent.
- C. The coordinates are independent because the quantile function is deterministic.
- D. Neither coordinate has a standard normal distribution.

2h)

A generator is intended to produce observations from $\text{Uniform}(0, 1)$. Which result clearly indicates an error?

- A. The sample mean is 0.49 rather than exactly 0.50 .
- B. The sample variance is 0.084 rather than exactly $1/12$.
- C. The histogram bars are not all the same height.
- D. Some generated values are below 0 or above 1 .

Question 3: True or False [6]

Write **True** or **False** for each statement. Each statement is worth **1 point**. No explanation is required.

3a)

With the same code and computing environment, setting the same seed at the start reproduces the same simulation results.

3b)

If X has a discrete CDF F , then $F(X)$ is necessarily uniform on $(0, 1)$.

3c)

If independent gamma variables have the same rate, their sum is gamma with the sum of their shapes and that common rate.

3d)

In R, `rnorm(100, mean = 0, sd = 2)` targets a normal distribution with variance 2.

3e)

In rejection sampling, replacing a valid bound M by a larger valid bound preserves the target distribution of accepted values, although acceptance becomes less frequent.

3f)

A mixture of two bivariate normal distributions must itself have a bivariate normal distribution.

Question 4: Generating and Checking Samples [8]

Each item is worth **1 point**. Label your answers **4a.1-4a.4** and **4b.1-4b.4**. Show the requested calculations and keep explanations to one or two sentences.

4a) A normal mixture [4]

The target distribution is

$$0.25 N(-2, 1) + 0.75 N(2, 1),$$

where the second parameter is the variance. A student instead generates independent $A \sim N(-2, 1)$ and $B \sim N(2, 1)$ and returns

$$W = 0.25A + 0.75B.$$

1. Describe a correct procedure for generating one observation from the target mixture.
2. Find the mean of the target mixture and $E(W)$. Would checking only the mean detect the error?
3. Find the variance of the target mixture and $\text{Var}(W)$.
4. Both sets of simulated values are recentered and rescaled to have mean 0 and variance 1. Does this show that they have the same distribution? Name one additional distributional check.

4b) Checking an exponential generator [4]

The intended distribution is exponential with rate 2. A student uses:

```
set.seed(8670)
x <- rexp(10000, rate = 1 / 2)
```

The sample mean is about 2 and the sample variance is about 4. You do not need to run the code.

1. State the theoretical mean and variance of the intended distribution.
2. Identify the parameter error and write the corrected `rexp()` call.
3. Describe how you would use a histogram and the target density to check the generated values. What would you look for?
4. Would generating more observations with the same incorrect code fix the discrepancy? Explain briefly.