Lab 3: Bayesian Inference for Continuous Random Variables

2026

Each probability distribution is associated with a core name: `exp` is the core name of the exponential distribution, `beta` of the Beta distribution, `unif` of the uniform distribution, `gamma` of the Gamma distribution, and `norm` of the normal distribution.

The four basic associated functions are the pdf (probability density function), the cdf (cumulative distribution function), the quantile function, and the simulation procedure: adding the prefixes `d`, `p`, `q`, and `r`, respectively, to the core name. For example:

- `dexp(c)` is the probability density that X is 'near' c;
- `pexp(c)` is the probability that X is less than or equal to c.

Distribution	**Core**	**Parameters**
Binomial	`binom`	`size`, `prob`
Poisson	`pois`	`lambda`
Uniform	`unif`	`min`, `max`
Exponential	`exp`	`rate` = 1/mean
Beta	`beta`	`shape1`, `shape2`
Gamma	`gamma`	`shape`, `rate`
Normal	`norm`	`mean`, `sd`

If you are unsure how to use these commands, use `help()` to find more information.

Question 1: Exponential Distribution

Let X be the amount of time (in minutes) an eyesight tester spends with his or her client. The time is known to have an exponential distribution with the average amount of time equal to four minutes. Find the probability that an eyesight tester spends four to five minutes with a randomly selected client.

[0]Email: paulius.kazlauskas@evaf.vu.lt

Solution

Since the mean is 4 minutes, the rate parameter is $\lambda = 1/4 = 0.25$. The probability that the tester spends between four and five minutes with a client is:

$$\begin{aligned} P(4 < X < 5) &= P(X < 5) - P(X < 4) \\ &= \left(1 - e^{-\lambda \cdot 5}\right) - \left(1 - e^{-\lambda \cdot 4}\right) \\ &= e^{-\lambda \cdot 4} - e^{-\lambda \cdot 5} \end{aligned}$$

Because $\lambda = 0.25$:

$$P(4 < X < 5) = e^{-0.25\times 4} - e^{-0.25\times 5} = e^{-1} - e^{-1.25} \approx 0.0814$$

R Corner:

```r
# To find the CDF of the Exponential Distribution:
# P(X < q) = pexp(q, rate)

# Find P(X < 5)
P5 <- pexp(5, 0.25)

# Find P(X < 4)
P4 <- pexp(4, 0.25)

# P(4 < X < 5) = P(X < 5) - P(X < 4)
P45 <- P5 - P4
P45
```

Question 2: Memoryless Property of the Exponential

Let X be a random variable that has an exponential distribution with rate parameter $\lambda = 0.25$. Show by programming that:

$$P(X \leq 5 + 10 \mid X > 5) = P(X \leq 10)$$

Solution

By the definition of conditional probability:

$$P(X \leq 5 + 10 \mid X > 5) = \frac{P(5 < X \leq 5 + 10)}{P(X > 5)}$$

This equals $P(X \leq 10)$ because of the memoryless property of the exponential distribution: having already waited 5 minutes provides no information about how much longer you will wait.

Proof:

$$\begin{aligned}\frac{P(5 < X \leq 15)}{P(X > 5)} &= \frac{F(15) - F(5)}{1 - F(5)} \\ &= \frac{(1 - e^{-0.25\times 15}) - (1 - e^{-0.25\times 5})}{e^{-0.25\times 5}} \\ &= \frac{e^{-1.25} - e^{-3.75}}{e^{-1.25}} \\ &= 1 - e^{-2.5} = F(10) = P(X \leq 10)\end{aligned}$$

R Corner:

```
# Find P(X > 5)
A1 <- 1 - pexp(5, 0.25)

# Find P(5 <= X <= 5 + 10) = P(5 < X <= 15)
A2 <- pexp(15, 0.25) - pexp(5, 0.25)

# Find P(X <= 5 + 10 | X > 5) = A2 / A1
R1 <- A2 / A1

# Find P(X <= 10)
R2 <- pexp(10, 0.25)

# Compare R1 and R2 -- they should be equal
R1
R2
R1 == R2
```

Question 3: Standard Normal Distribution

Let Z have the standard normal distribution.

1. Find $P(0 \leq Z \leq 1.52)$
2. Find $P(Z \geq 2.11)$
3. Find $P(-1.45 \leq Z \leq 1.74)$
4. Find the 2.5 percentile and 97.5 percentile

Solution

1. Find $P(0 \leq Z \leq 1.52)$

Using the Table of Area under the standard normal density, $P(0 \leq Z \leq 1.52) = 0.4357$.

$$P(0 \leq Z \leq 1.52) = \Phi(1.52) - \Phi(0) = \Phi(1.52) - 0.5 = 0.9357 - 0.5 = 0.4357$$

```
# pnorm(q, mean = 0, sd = 1)
P1 <- pnorm(1.52, 0, 1) - pnorm(0, 0, 1)
P1
```

2. Find $P(Z \geq 2.11)$

$$P(Z \geq 2.11) = 1 - P(Z < 2.11) = 1 - \Phi(2.11) = 1 - 0.9826 = 0.0174$$

```r
P2 <- 1 - pnorm(2.11, 0, 1)
P2
```

3. Find $P(-1.45 \leq Z \leq 1.74)$

Using the symmetry of the standard normal:

$$\begin{aligned} P(-1.45 \leq Z \leq 1.74) &= P(-1.45 \leq Z \leq 0) + P(0 \leq Z \leq 1.74) \\ &= P(0 \leq Z \leq 1.45) + P(0 \leq Z \leq 1.74) \\ &= 0.4265 + 0.4591 = 0.8856 \end{aligned}$$

```r
P3 <- pnorm(1.74, 0, 1) - pnorm(-1.45, 0, 1)
P3
```

4. Find the 2.5 and 97.5 percentiles

```r
# 2.5 percentile
qnorm(0.025, 0, 1)

# 97.5 percentile
qnorm(0.975, 0, 1)
```

Question 4: Uniform Distribution

In winter, it takes between 0 and 60 minutes to start your car. What is the probability that it would take between 57 and 60 minutes?

Solution

Let X be the time (in minutes) to start the car. $X \sim \text{Uniform}(0, 60)$. The PDF is $f(x) = 1/(60 - 0) = 1/60$ for $x \in [0, 60]$.

$$P(57 \leq X \leq 60) = \frac{1}{60}\int_{57}^{60} dx = \frac{1}{60}(60 - 57) = \frac{3}{60} = \frac{1}{20} = 0.05$$

R Corner:

```r
# punif(q, min = 0, max = 1)
P1 <- punif(60, 0, 60) - punif(57, 0, 60)
P1
```

Question 5: Beta Distribution

Let Y be distributed according to the $\text{Beta}(10, 12)$ distribution.

1. Find $E(Y)$.
2. Find $\text{Var}(Y)$.
3. Generate a sample of 1000 observations from $\text{Beta}(10, 12)$ and calculate the mean and variance.
4. Find $P(Y > 5)$ using the normal approximation.
5. Find the exact value of $P(Y > 5)$ using the Beta distribution.

Solution

$Y \sim \text{Beta}(10, 12)$

1. Find $E(Y)$

$$E(Y) = \frac{\alpha}{\alpha + \beta} = \frac{10}{10 + 12} = \frac{10}{22} = \frac{5}{11} \approx 0.4545$$

2. Find $\text{Var}(Y)$

$$\text{Var}(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{10 \times 12}{22^2 \times 23} \approx 0.0108$$

```
a <- 10
b <- 12
Ey   <- a / (a + b)
Vary <- a * b / ((a + b)^2 * (a + b + 1))
Ey
Vary
```

3. Generate a sample of 1000 observations

```
Y <- rbeta(1000, 10, 12)
M <- mean(Y)
V <- var(Y)
M
V

# Plot histogram -- freq = FALSE converts y-axis to density
hist(Y, main = "Density", freq = FALSE)

# Overlay approximated normal density
x <- seq(0, 1, 0.01)
curve(dnorm(x, Ey, sqrt(Vary)), add = TRUE, col = "red")
```

4. Find $P(Y > 5)$ using the normal approximation

We approximate $Y \sim N(0.4545,\ 0.0108)$. Standardising:

$$P(Y > 5) = P\left(Z > \frac{5 - 0.4545}{\sqrt{0.0108}}\right) = P(Z > 43.73) \approx 0$$

Note: the Z-table only reaches about $z = 3.69$, beyond which $\Phi(z) \approx 1$, so $P(Z > 43.73) \approx 0$.

```r
P1 <- 1 - pnorm(5, Ey, sqrt(Vary))
P1
```

5. Find the exact value of $P(Y > 5)$ using the Beta distribution

Since the Beta distribution is bounded on $[0, 1]$, any value greater than 1 is impossible, so $P(Y > 5) = 0$ exactly.

```r
P2 <- 1 - pbeta(5, a, b)
P2
```

Question 6: Normal Distribution

Let Y be normally distributed with mean $\mu = 120$ and variance $\sigma^2 = 64$.

1. Find $P(Y \leq 130)$.
2. Find $P(Y \leq 135)$.
3. Find $P(114 \leq Y \leq 127)$.
4. Generate a sample of 1000 observations from $N(120, 64)$ and plot the histogram and approximated density.

Solution

Any normal distribution can be transformed into the standard normal by subtracting the mean and dividing by the standard deviation: $Z = (Y - \mu)/\sigma \sim N(0, 1)$.

Note that R's `rnorm` and `pnorm` take `sd` $= \sigma = \sqrt{\sigma^2}$, not the variance directly. Here $\sigma = \sqrt{64} = 8$.

1. Find $P(Y \leq 130)$

$$P(Y \leq 130) = P\left(Z \leq \frac{130 - 120}{8}\right) = P(Z \leq 1.25) = 0.5 + 0.3944 = 0.8944$$

```r
P1 <- pnorm(130, 120, sqrt(64))
P1
```

2. Find $P(Y \leq 135)$

$$P(Y \leq 135) = P\left(Z \leq \frac{135 - 120}{8}\right) = P(Z \leq 1.875) \approx 0.9696$$

```r
P2 <- pnorm(135, 120, sqrt(64))
P2
```

3. Find $P(114 \leq Y \leq 127)$

$$\begin{aligned} P(114 \leq Y \leq 127) &= P\left(\frac{114-120}{8} \leq Z \leq \frac{127-120}{8}\right) \\ &= P(-0.75 \leq Z \leq 0.875) \\ &= P(0 \leq Z \leq 0.75) + P(0 \leq Z \leq 0.875) \\ &\approx 0.2734 + 0.3106 = 0.5840 \end{aligned}$$

Note: the value for $P(Z \leq 0.875)$ is not available directly in the Z-table, so we use $P(Z \leq 0.88) \approx 0.3106$ as the closest tabulated value. In R, the exact value is used automatically.

```
P3 <- pnorm(127, 120, sqrt(64)) - pnorm(114, 120, sqrt(64))
P3
```

4. Generate a sample of 1000 observations and plot

```
# Generate the sample
y <- rnorm(1000, 120, sqrt(64))

# Plot histogram -- freq = FALSE converts y-axis to density
hist(y, main = "Density", freq = FALSE)

# Overlay the true normal density
x <- seq(90, 160, 0.5)
curve(dnorm(x, 120, sqrt(64)), add = TRUE, col = "red")
```

Question 7: Gamma Distribution

Let Y be distributed according to the Gamma$(12, 4)$ distribution.

1. Find $E(Y)$.
2. Find $\text{Var}(Y)$.
3. Find $P(Y \leq 4)$ using the normal approximation.
4. Find $P(Y \leq 4)$ using the Gamma distribution.
5. Find the 2.5 and 97.5 percentiles.

Solution

$Y \sim \text{Gamma}(r, v)$ with $r = 12$, $v = 4$.

1. Find $E(Y)$

$$E(Y) = \frac{r}{v} = \frac{12}{4} = 3$$

2. Find $\text{Var}(Y)$

$$\text{Var}(Y) = \frac{r}{v^2} = \frac{12}{16} = 0.75$$

```r
r <- 12
v <- 4

# Mean
Ey <- r / v
Ey

# Variance
Vary <- r / v^2
Vary
```

3. Find $P(Y \leq 4)$ using the normal approximation

We approximate $Y \sim N(3,\ 0.75)$. Standardising:

$$P(Y \leq 4) = P\left(Z \leq \frac{4-3}{\sqrt{0.75}}\right) = P(Z \leq 1.1547) \approx P(Z \leq 1.15) = 0.8749$$

```r
# P(Y <= 4) using the normal approximation
P1 <- pnorm(4, Ey, sqrt(Vary))
P1
```

4. Find $P(Y \leq 4)$ using the Gamma distribution

```r
# P(Y <= 4) using the gamma distribution (exact)
P2 <- pgamma(4, r, v)
P2

# Compare approximation and exact
P1
P2
```

5. Find the 2.5 and 97.5 percentiles

```r
# 2.5 percentile
qgamma(0.025, r, v)

# 97.5 percentile
qgamma(0.975, r, v)
```

8 Bonus Problems

Bonus Practice: Normal and Exponential

★ Bonus

A hospital records the time (in hours) between successive patient arrivals at an emergency unit. The inter-arrival times follow an exponential distribution with a mean of 0.5 hours.

1. Find the rate parameter λ. Write down the PDF and CDF.
2. Find $P(X > 1)$: the probability that the next patient arrives more than one hour from now.
3. Find $P(0.25 \leq X \leq 0.75)$.
4. **Memoryless property:** A patient just arrived 20 minutes ago and no new patient has come since. Show that the probability of waiting at least another 20 minutes is the same as the probability of waiting at least 20 minutes from scratch. Verify in R.
5. The hospital also tracks the *total* daily admissions, which are approximately normally distributed with mean $\mu = 48$ and variance $\sigma^2 = 36$.
   (a) Find $P(Y \leq 54)$.
   (b) Find $P(42 \leq Y \leq 57)$.
   (c) What is the 5th percentile of daily admissions? Interpret what this number means.

R: Compute all five parts in R. For part (e), use `pnorm` and `qnorm` directly — no manual standardisation needed.

Hints. For part (4), you need to show $P(X > 2/3 + 1/3 \mid X > 1/3) = P(X > 1/3)$ using the conditional probability formula and the CDF. For part (e.iii), think about what it means for a day to be "below the 5th percentile" in terms of hospital operations.

Bonus 1: Code Style — Naming and Spaces

★ Bonus

Chapter 4 of *R for Data Science* (2nd ed., Wickham et al., `https://r4ds.hadley.nz/workflow-style.html`) sets out the tidyverse style guide for writing clean, readable R code. Section 4.1 covers **naming** and Section 4.2 covers **spaces**.

The code below reproduces several calculations from today's lab, but it violates multiple style rules. Your tasks:

1. Identify every style violation. For each one, state which rule from the textbook it breaks (naming convention, spaces around operators, spaces after commas, spaces around `<-`, etc.).
2. Rewrite the entire block so it conforms to the tidyverse style guide.
3. Does the rewritten code produce different results? Why or why not?

```
# badly styled code -- find and fix all violations
R<-12
V<-4
```

```r
EY<-R/V
VarY<-R/V^2
P_approx<-pnorm(4,EY,sqrt(VarY))
P_exact <-pgamma(4,R,V)
PERCENTILE_low<-qgamma(0.025,R,V)
PERCENTILE_high<-qgamma(0.975,R,V)
cat("approx=" ,P_approx,"exact=",P_exact)
```

Bonus 2: Code Style — The Pipe Operator

★ Bonus

Section 4.3 of the same textbook chapter introduces the native pipe `|>` and its formatting rules: space before the pipe, pipe at the end of the line, indent continuation lines by two spaces.

The code below computes a small simulation study comparing the normal approximation to the exact Gamma CDF at several thresholds, written without pipes and with poor formatting. Your tasks:

1. Read Section 4.3 of the textbook. In 2–3 sentences, explain what the pipe does and what formatting rule each of the three "Avoid" examples in the book violates.
2. Rewrite the code below using `|>` wherever it improves readability, following all formatting rules from the textbook.
3. One of the thresholds produces a noticeably larger approximation error. Which one, and why does the normal approximation perform worse there?

```r
thresholds<-c(1,2,3,4,5)
r<-12; v<-4
Ey<-r/v; Vary<-r/v^2
exact<-sapply(thresholds,function(t) pgamma(t,r,v))
approx<-sapply(thresholds,function(t) pnorm(t,Ey,sqrt(Vary)))
errors<-round(abs(exact-approx),4)
results<-data.frame(threshold=thresholds,exact=round(exact,4)
  ,approx=round(approx,4),error=errors)
print(results)
```

Hint on part (3): the normal approximation is most accurate near the centre of the distribution and less accurate in the tails. Where is $\mu = 3$ relative to the five thresholds?

Bonus 3: Sectioning Comments

★ Bonus

Section 4.5 of the textbook explains how to use sectioning comments in R to organise longer scripts:

```
# Load data ----------------------------------------------

# Clean data ---------------------------------------------

# Compute summaries --------------------------------------
```

Take the full R code from today's lab (Questions 1–7) and organise it into a single well-structured R script using:

- Sectioning comments to separate each question.
- Correct naming conventions (`snake_case`, lowercase).
- Proper spacing around all operators and after all commas.
- The pipe `|>` wherever a chain of two or more operations on the same object makes the intent clearer.

Submit the script as a `.R` file. There is no single correct answer: the goal is a script a colleague could read and understand without asking you any questions.