PROJECT 1

SANDRA

2025-10-24

Contents


```r
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.4     v tibble    3.3.0
## v purrr     1.1.0     v tidyr     1.3.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

0.1 BETA DISTRIBUTION

0.1.1 Discrete or Continuous?

Is it a discrete or continuous distribution? Continuous distribution

0.1.2 Assumptions of the Beta Distribution

The variable represents a proportion or probability between 0 and 1.

The shape parameters > 0 and > 0.

Observations (if used to estimate and) are independent.

There's a fixed underlying probability whose uncertainty we're modeling.

0.1.3 Real World Example

1. Election Probabilities – When people have different beliefs about the chance of a candidate or party winning, a Beta distribution can summarize these opinions and reflect both the most likely outcome and the uncertainty.
2. A/B Testing / Conversion Rates – In marketing or web design, if a fraction of users take a desired action (like clicking a button), the Beta distribution models the probability of success and captures the uncertainty due to the sample size.
3. Medical Trials – When testing a new treatment, the Beta distribution can represent the probability that a patient responds positively, especially useful when only a limited number of trial outcomes are observed.

0.1.4 Support

What are all possible values of the distribution? (also called the support of the distribution) 0 x 1 [0 & 1] **Explanation:** The Beta distribution is ideal for modeling quantities that are bounded between 0 and 1, such as probabilities, proportions, or percentages. For example, it can represent the probability of success in a **Bernoulli** trial or the proportion of a population with a certain characteristic.

0.1.5 Parameters

What are the parameters of the distribution? Provide alpha and beta

- the R names of the parameters :
- shape1 = α (alpha)
- shape2 = β (beta)
- verbal names for the parameters
- alpha is the first shape parameter -beta is the second shape parameter
- notation for the parameters Beta(,)

0.1.6 Parameter space

For each parameter, what is its range of possible values? (also called the parameter space) For the Beta distribution(,): > 0 >0 (parameter space: $(0, \infty)$)

> 0 >0 (parameter space: $(0, \infty)$)

0.1.7 Mean and variance

Express the following as a function of the parameter(s):

- The mean of the distribution mean = alpha/ alpha + beta
- The variance of the distribution variance = alpha * beta / ((alpha + beta)^2 * (alpha + beta + 1))

0.1.8 Plot of probability density function

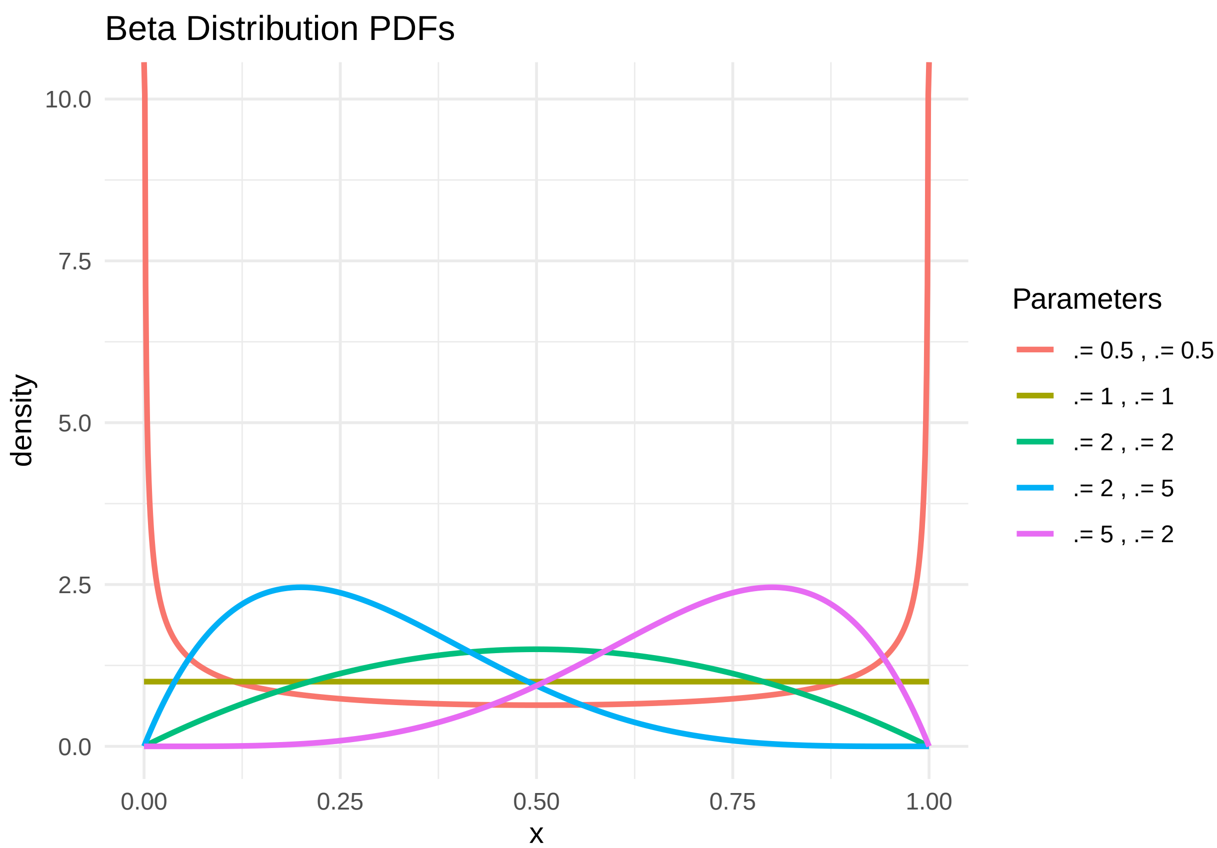
Plot the probability density function of the distributions at 4 different sets of parameter values.

```r
# Different parameter combinations
params <- list(
  c(1,1),    # Uniform distribution (flat)
  c(0.5,0.5), # U-shaped
  c(2,2),      # symmetric, bell-shaped
  c(2,5),       # skewed right
  c(5, 2)
)

x <- seq(0, 1, length.out = 1000)
df <- data.frame()  # initialize properly

for (p in params) {
  alpha <- p[1]; beta <- p[2]
  df <- rbind(df, data.frame(
    x = x,
    density = dbeta(x, alpha, beta),
    alpha = alpha,
    beta = beta
  ))
}
```

```r
ggplot(df, aes(x, density, color = paste(" =", alpha, ",  =", beta))) +
  geom_line(size=1) +
  labs(title = "Beta Distribution PDFs",
       color = "Parameters") +
  theme_minimal()
```



0.1.9 How does the distribution change with the parameters?

Describe how the distribution depends on changes in the parameters, including changes in the center, spread, and shape of the distribution.

Uniform case: = = 1 → Uniform (0,1) Every value between 0 and 1 is equally likely.

Skewness (who's bigger, or ?)

> → distribution skewed left (mass near 1).

< → distribution skewed right (mass near 0).

= → distribution is symmetric about 0.5.

Concavity (are and less than or greater than 1?)

, < 1 → U-shaped (high at the ends, low in the middle).

, > 1 → bell-shaped (peaked around the mean).

= = 1 → flat (uniform).

Concentration (variance control)

If you increase both and while keeping their ratio the same, the mean stays fixed but the variance shrinks.

Example:

(=2, =2) → symmetric, moderate variance.

(=20, =20) → still symmetric around 0.5, but much more concentrated around 0.5. Larger values of and (while keeping ratio constant) → distribution gets more concentrated around the mean (lower variance).

0.1.10 Distribution-specific questions and hints

You need only answer the questions about the distribution to which your group is assigned.

0.1.10.1 Beta

Hint: You should not worry about the parameter `ncp`. Leave it set at the default value of 0.

- For which values of the parameters is the density function concave up, concave down, and flat?

Concave Up: When < 1 and < 1, producing a U-shaped PDF with peaks at $= 0$ and $= 1$. Example: $= 0.5$, $= 0.5$.

Concave Down: When > 1 and > 1, resulting in a bell-shaped PDF. Example: $= 2$, $= 2$ (symmetric bell).

Flat: When $= 1$ and $= 1$, giving the Uniform(0,1) distribution with constant density $( ) = 1$ Skewed (mixed concavity): When one parameter is less than 1 and the other greater than 1. Example: $= 2$, $= 5$ (skewed right). .

- **Skewed (mixed concavity):** When one parameter is less than 1 and the other greater than 1. Example: $\alpha = 2, \beta = 5$ (skewed right).

0.1.11 Plot of Probability Density Function with Shape Explanations

```r
library(ggplot2)
library(dplyr)

# Define parameter combinations and shape labels
params <- data.frame(
  alpha = c(1, 0.5, 2, 2, 5),
  beta  = c(1, 0.5, 2, 5, 2),
  shape = c("Flat (Uniform)",
            "U-shaped (Concave Up)",
            "Bell-shaped (Concave Down)",
            "Skewed Right (Mixed Concavity)",
            "Skewed Left (Mixed Concavity)")
)

# Generate density data for each ( ,  )
x <- seq(0, 1, length.out = 1000)
df <- params %>%
  rowwise() %>%
  mutate(data = list(data.frame(
    x = x,
    density = dbeta(x, alpha, beta)
  ))) %>%
  unnest(cols = c(data))

# Create the faceted plot
ggplot(df, aes(x = x, y = density, color = shape)) +
  geom_line(size = 1.2) +
  facet_wrap(~shape, ncol = 2, scales = "free_y") +
```

```r
  labs(
    title = "Beta Distribution PDFs for Different Parameter Sets",
    x = "x",
    y = "Density"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    legend.position = "none",
    plot.title = element_text(face = "bold", hjust = 0.5)
  )
```