GEOG 738 Discrete Choice Analysis - Exercise 3

Haoran Xu

2025-01-31

```r
library(dplyr) # A Grammar of Data Manipulation
library(ggplot2) # Create Elegant Data Visualisations Using the Grammar of Graphics
```

1. Define utility.

Utility is an abstract, latent measure of the pleasure, usefulness, enjoyment, or attractiveness associated with making a choice.

2. Describe in your own words the behavior described by utility maximization.

Utility maximization describes the process that individuals make choices in order to get the most satisfaction or benefit from the options available to them. For example, I am currently faced with a situation where I have to decide whether to move out of my current residence or still stay here. My to-be-decided choice will be based on the maximization of the utility I could get from one of the alternatives. To maximize the utility, a lot of factors, such as rent, moving cost, distance to campus, living conditions, all of which will be assessed by me personally.

3. What conditions are necessary for a function to be a valid probability distribution function? Consider the function shown in Figure below. This is called the triangle or tent function.

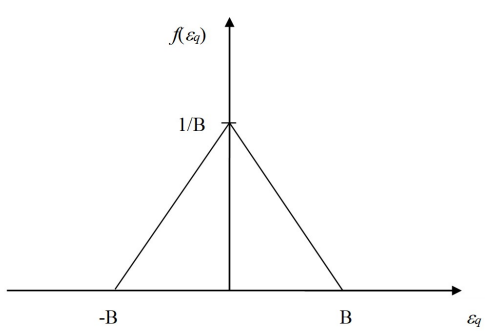


Figure 1: Triangle function

Condition 1: $f(x) \geq 0$ for all x Condition 2:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Condition 1 determines that probability would be greater or equal to 0. Condition 2 determines that the integral of probability is 1 (the area of the curve will also be 1).

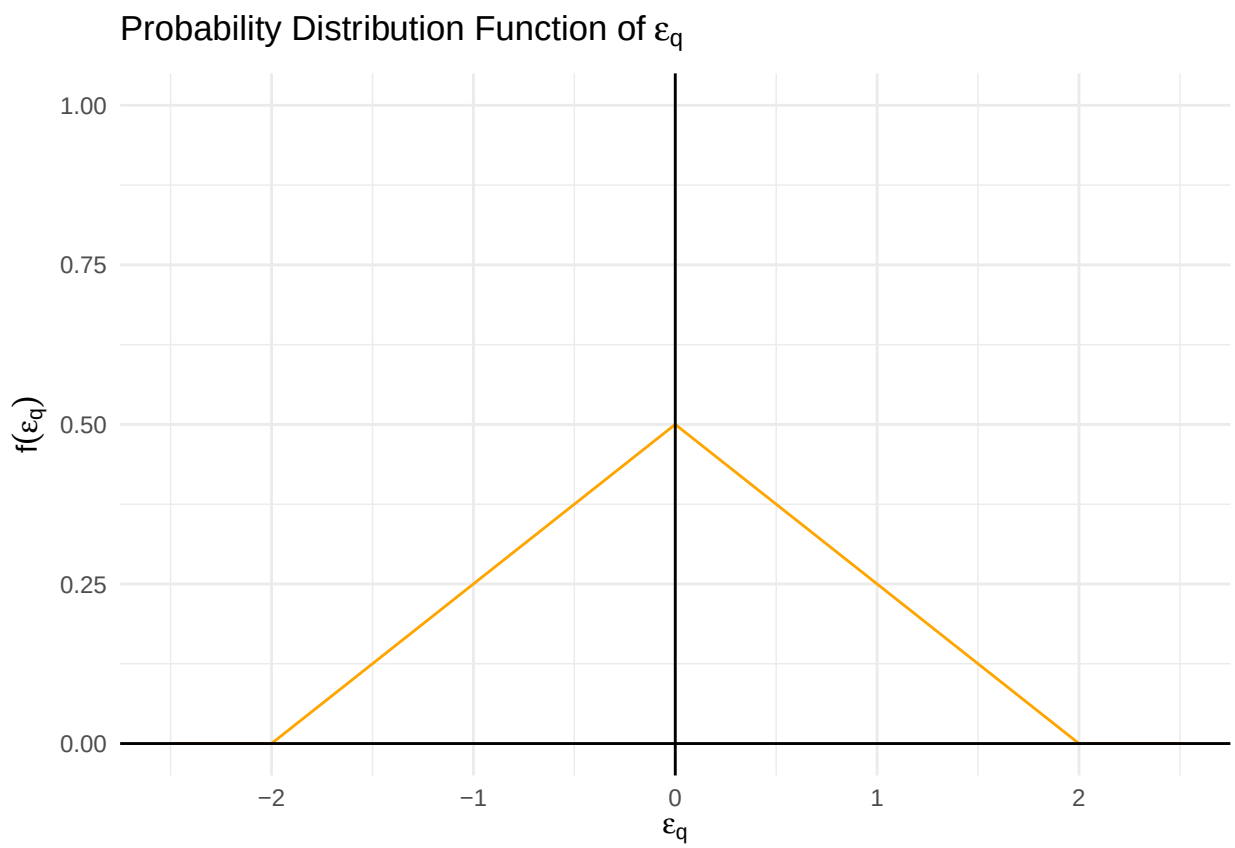
4. Show that the triangle function in the figure is a valid probability distribution function.

```r
triangle_function <- function(epsilon_q, B) {
  ifelse(abs(epsilon_q) <= B,
         (1 - abs(epsilon_q) / B) / B,
         0)
}

B <- 2

df1 <- data.frame(x =seq(from = -(B + 0.5),
                         to = B + 0.5,
                         by = 0.01)) %>%
  mutate(f = triangle_function(x, B))

# Plot
ggplot(data = df1,
       aes(x = x,
           y = f)) +
  geom_line(color = "orange") +
  ylim(c(0, 1)) +
  geom_hline(yintercept = 0) +
  geom_vline(xintercept = 0) +
  labs(title = expression("Probability Distribution Function of" ~ epsilon[q]),
       x = expression(epsilon[q]),
       y = expression(f(epsilon[q]))) +
  theme_minimal()
```




```
if (all(df1$f >= 0)) {
  print("All results (separated by 0.01) are equal to or greater than 0.")
} else {
  print("There are results that are less than 0.")
}
```

```
## [1] "All results (separated by 0.01) are equal to or greater than 0."
```

```
integrated_result <- integrate(triangle_function, lower = -B, upper = B, B = B)
print(paste("The integrated result is", integrated_result$value))
```

```
## [1] "The integrated result is 1"
```

5. Next, consider the following utility functions for two alternatives, namely i and j:

$$U_i = V_i + \epsilon_i U_j = V_j + \epsilon_j$$

Assume that the difference between the error terms below follows the triangle distribution:

$$\epsilon_q = \epsilon_i - \epsilon_j$$

Parting from the assumption above, derive a binary choice model for the probability of selecting alternative j.

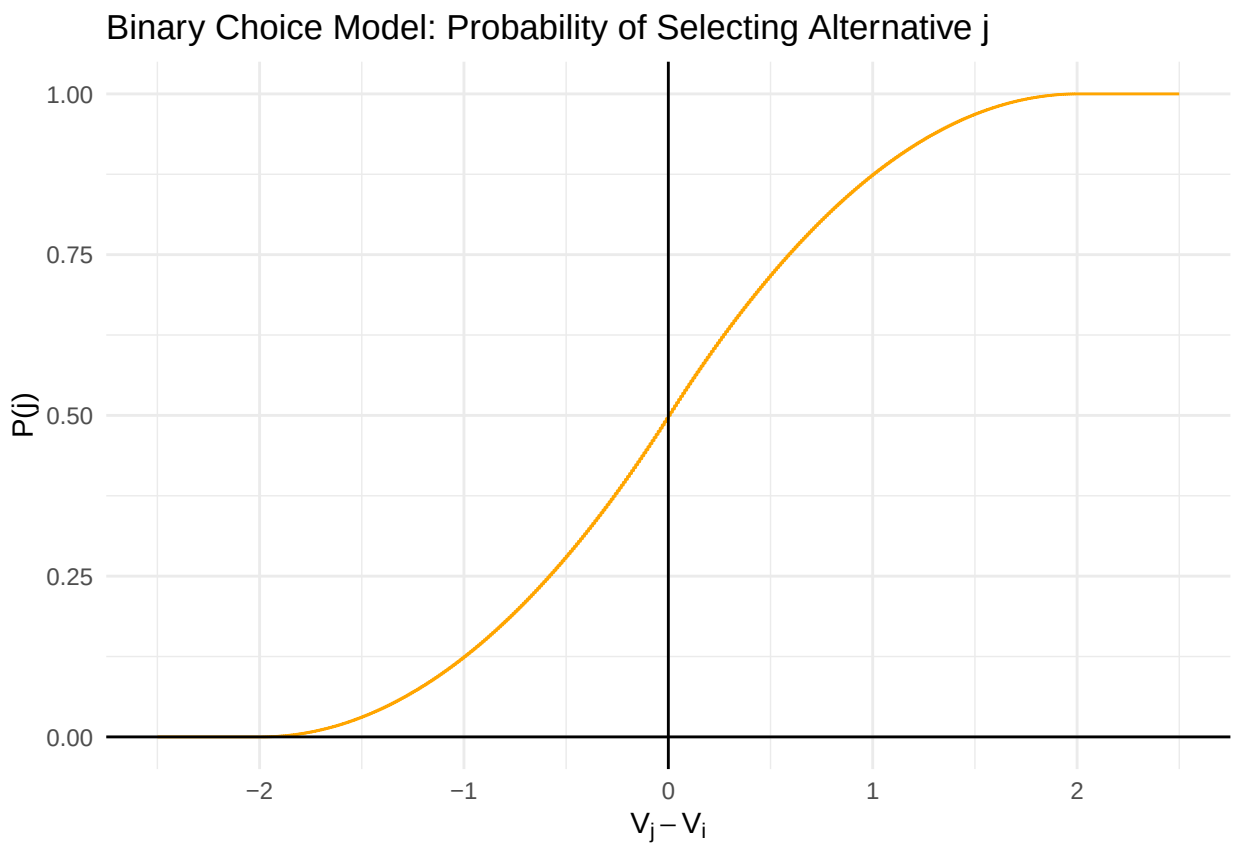
```
triangle_cdf <- function(epsilon_q, B) {
  ifelse(epsilon_q <= -B,
         0,
         ifelse(epsilon_q > -B & epsilon_q <= 0,
                (B + epsilon_q)^2 / (2 * B^2),
```

```r
                ifelse(epsilon_q > 0 & epsilon_q <= B,
                       1 - (B - epsilon_q)^2 / (2 * B^2),
                       1)))
}

# Define B
B <- 2

# Create a data frame
df2 <- data.frame(x = seq(from = -(B + 0.5),
                          to = B + 0.5,
                          by = 0.01)) %>%
  mutate(f = triangle_cdf(x, B))

# Plot
ggplot(data = df2,
       aes(x = x,
           y = f)) +
  geom_step(color = "orange") +
  ylim(c(0, 1)) +
  geom_hline(yintercept = 0) +
  geom_vline(xintercept = 0) +
  labs(title = "Binary Choice Model: Probability of Selecting Alternative j",
       x = expression(V[j] - V[i]),
       y = "P(j)") +
  theme_minimal()
```



The model illustrated above is the cumulative distribution function of P_j on $V_j - V_i$. The mathematical calculations can be derived from:

$$P_j = P(U_j > U_i) = P(V_j - V_i > \epsilon_i - \epsilon_j) = P(V_j - V_i > \epsilon_q) = F_{(V_j - V_i)}(\epsilon_q)$$