Econ 5023: Statistics for Decision Making

Univariate Statistics (IX): Monte Carlo Simulation and Parametric Distributions: Discrete, Binomial Distribution

Le Wang

Itinerary:

1. Brief intro to probability functions in R
2. Discrete Distribution: Binomial Distribution

Probability Distribution Functions in R

Table: Probability Distribution Functions in R

Distributions	**Root**
Binomial	`binom`
Poisson	`pois`
Beta	`beta`
Cauchy	`cauchy`
Chi-square	`chisq`
Exponential	`exp`
F	`f`
Gamma	`gamma`
Normal	`norm`
Student's t	`t`
Uniform	`unif`
Weibull	`weibull`

1. By prefixing a “d” to the function name in the table above, you can get **probability density** values (pdf).
2. By prefixing a “p”, you can get **cumulative probabilities** (cdf).
3. By prefixing a “q”, you can get **quantile** values.?
4. By prefixing an “r”, you can get **random numbers** from the distribution.

Discrete, Parametric Distribution (I): Binomial Distribution

This distribution describes the outcome of n independent trials in an experiment, which satisfies the following requirements (p.183 in the book)

1. Can take on only two values (outcomes): e.g., Success or Failure
2. The random variable, x, is the number of successes in a fixed number of trials (n).
3. The probability of success, p is fixed. (Does not vary over).
4. The trials are indepedent, meaning that the outcome of one trial does not affect the outcome of any other trial.

Probability Mass Function

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} \quad \text{where } x = 0, 1, 2, \ldots, n$$

Moments of this distribution:

$$\begin{aligned} \mu &= n \cdot p \\ \sigma^2 &= n \cdot p \cdot (1-p) \end{aligned}$$

Exmaple:
Suppose there are twelve multiple choice questions in an Economics class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four correct answers if a student attempts to answer every question at random.

Translations:

Exmaple:
Suppose there are twelve multiple choice questions in an Economics class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four correct answers if a student attempts to answer every question at random.

Translations:

1. Success is defined as whether you can guess it right.

Exmaple:

Suppose there are twelve multiple choice questions in an Economics class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four correct answers if a student attempts to answer every question at random.

Translations:

1. Success is defined as whether you can guess it right.
2. The probability of success, $p = \frac{1}{5} = 0.2$

Exmaple:
Suppose there are twelve multiple choice questions in an Economics class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four correct answers if a student attempts to answer every question at random.

Translations:

1. Success is defined as whether you can guess it right.
2. The probability of success, $p = \frac{1}{5} = 0.2$
3. The number of trials, $n = 12$

Exmaple:

Suppose there are twelve multiple choice questions in an Economics class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four correct answers if a student attempts to answer every question at random.

Translations:

1. Success is defined as whether you can guess it right.
2. The probability of success, $p = \frac{1}{5} = 0.2$
3. The number of trials, $n = 12$
4. The random variable, the number of success, is equal to 4.

Probability Mass Function

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} \quad \text{where } x = 0, 1, 2, \ldots, n$$

1. Success is defined as whether you can guess it right.
2. The probability of success, $p = \frac{1}{5} = 0.2$
3. The number of trials, $n = 12$
4. The random variable, the number of success, is equal to 4.

```
dbinom(4, size=12, prob=0.2)

## [1] 0.1328756
```

We can also calculate the probability of $x \leq 4$

```
dbinom(0, size=12, prob=0.2)+dbinom(1, size=12, prob=0.2)+d

## [1] 0.9274445
```

We can also calculate the probability of $x \leq 4$

```
dbinom(0, size=12, prob=0.2)+dbinom(1, size=12, prob=0.2)+dbinom(2, size=12, prob=0.2)+dbinom(3, size=12,

## [1] 0.9274445
```

Similarly,

```
pbinom(4, size = 12, prob = 0.2)

## [1] 0.9274445
```

Question: How will you manually simulate a Binomial random variable?

How do you even know the formula is correct? Or, the programmers actually coded it right?

Our Thought Process:

1. Generate a sample of 12 observations
2. Each observation represents whether it is a success
3. Calculate the sum of these 12 observations
4. Save the results
5. Repeat the process for many many times

```
sample <- sample(0:1,12, prob = c(0.8,0.2), replace = TRUE)
sum(sample)
```

```r
# Set seed so that we can reproduce these results
set.seed(123456)

# Initiate a vector to store results
x <- vector(0, 100)

# Create a sample of xs
for (i in c(1:10000)){
  sample <- sample(0:1,12, prob = c(0.8,0.2), replace = TRUE)
  x[i] <- sum(sample)
}

# Let's compare the characteristic of our simulated sample to the true value.
sum(x == 4)/length(x)
dbinom(4,size = 12, prob = 0.2)
```

```
# Set seed so that we can reproduce these results
set.seed(123456)

# Initiate a vector to store results
x <- vector(, 100)

# Create a sample of xs
for (i in c(1:10000)){
  sample <- sample(0:1,12, prob = c(0.8,0.2), replace = TRUE)
  x[i] <- sum(sample)
}

# Let's compare the characteristic of our simulated sample to the true value.
sum(x == 4)/length(x)
```

```
## [1] 0.1312
```

```
dbinom(4,size = 12, prob = 0.2)
```

```
## [1] 0.1328756
```