R Lab (Extra): Probability Distributions in R

Ikuma Ogura

August 23, 2019

Distributions Typically Used in Statistics

- In data analysis we often deal with classes of probability distributions whose specific shape can be determined with a few parameters.
- Example: by specifying the value of μ and σ, we can determine the exact shape of the normal distribution;

$$\mathrm{N}(\mathrm{x}|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\mathrm{x} - \mu)^2 \right)$$

- R contains useful functions to deal with probability distributions.
  - For each class of distribution, R has 4 types of functions.
    - ★ PMF/PDF
    - ★ CDF
    - ★ Quantile
    - ★ Random number generator

Functions Related with Probability Distributions in R (1)

- **Probability Mass Functions (PMF)/Probability Density Function (PDF)**
  - PMF of a discrete probability distribution $f(X = x)$ describes the probability that the random variable X takes the value x.
  - PDF of a continuous probability distribution $f(X = x)$ describes the *density* where X equals to x (i.e., relative likelihood of occurence of value x)
  - In R, PMF/PDF of a distribution is represented with a function starting from the letter `d`.
    - ★ e.g., PDF of a normal distributions can be computed with `dnorm()` function

Functions Related with Probability Distributions in R (2)

- **Cumulative Density Function (CDF)**
  - CDF of a probability distribution $F(X = x)$ describes the probability that X is equal to or smaller than x (i.e., $F(X = x) = \Pr(X \leq x)$)
  - Mathematically,

$$F'(X) = f(X)$$

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

  - In R, CDF of a distribution is represented with a function starting from the letter `p`
    - ⋆ e.g., CDF of a normal distributions can be computed with `pnorm()` function

Functions Related with Probability Distributions in R (3)

- **Inverse distribution function** or **Quantile function** is the inverse function of the CDF. Therefore, it takes the cumulative probability as the input and returns the value $X = x$ as the output.
  - In R, quantile function of a distribution is represented with a function starting from the letter `q`
    - ★ e.g., CDF of a normal distributions can be computed with `qnorm()` function

Functions Related with Probability Distributions in R (4)

- In R, we can simulate the value of X based on the PMF/PDF of a distribution
  - we call this practice as **random number generation**
- Functions to generate random numbers start from the letter `r`
  - e.g., we can generate random numbers from a normal distribution using the `rnorm()` command
- Strictly speaking, value of X simulated in R is not random; they are a deterministic sequence of numbers which look like random
  - To make sure that we can completely replicate the numbers, we determine the first number by specifying the seed through `set.seed()` command.
    - ★ e.g., First type `set.seed(123)` and then `rnorm(2)` into the R console. You'll get the two numbers -0.5604756 and -0.2301775 regardless of the computing environment.

Distributions Commonly Used in Statistics

Name	PMF/PDF	CDF	Quantile	Random Number
Uniform	`dunif`	`punif`	`qunif`	`runif`
Binomial	`dbinom`	`pbinom`	`qbinom`	`rbinom`
Poisson	`dpois`	`ppois`	`qpois`	`rpois`
Negative Binomial	`dnbinom`	`pnbinom`	`qnbinom`	`rnbinom`
Normal	`dnorm`	`pnorm`	`qnorm`	`rnorm`
Logistic	`dlogis`	`plogis`	`qlogis`	`rlogis`
t	`dt`	`pt`	`qt`	`rt`
F	`df`	`pf`	`qf`	`rf`
χ^2	`dchisq`	`pchisq`	`qchisq`	`rchisq`
Exponential	`dexp`	`pexp`	`qexp`	`rexp`
Weibull	`dweibull`	`pweibull`	`qweibull`	`rweibull`
Gamma	`dgamma`	`pgamma`	`qgamma`	`rgamma`

Examples

- The probability that a binomial random variable X equals to 3 when $n = 10$ and $p = 0.4$ is

```
dbinom(x = 3, size = 10, prob = 0.4)
```

```
## [1] 0.2149908
```

- The probability that a binomial random variable X is larger than 5 when $n = 13$ and $p = 0.3$ is

```
1 - pbinom(q = 5, size = 13, prob = 0.3)
```

```
## [1] 0.1653975
```

```
# or
pbinom(q = 5, size = 13, prob = 0.3, lower.tail = FALSE)
```

```
## [1] 0.1653975
```

Examples (cont.)

- The probability that a random variable X following the normal distribution with $\mu = 3$ and $\sigma^2 = 16$ falls between 5 and 7 is

```
pnorm(q = 7, mean = 3, sd = 4) - pnorm(q = 5, mean = 3, sd = 4)
```

```
## [1] 0.1498823
```

- If you want to find the value of $X \sim \mathrm{N}(-2, 4)$ where the cumulative probability just exceeds 0.4,

```
qnorm(p = 0.4, mean = -2, sd = 2)
```

```
## [1] -2.506694
```

Exercises!

- What is the probability of getting 4 heads out of 10 flips of a fair coin?
- What is the probability that a random variable $X \sim \mathrm{N}(4, 5)$ get values larger than 2 and smaller than 5?
- Generate 1,000 (quasi) random numbers from a Poisson distribution with $\lambda = 4$. Confirm that the mean and variance of the simulated values are close to 4.
  - What happens if we increase the number of random numbers to 100,000?