

R for toxicology

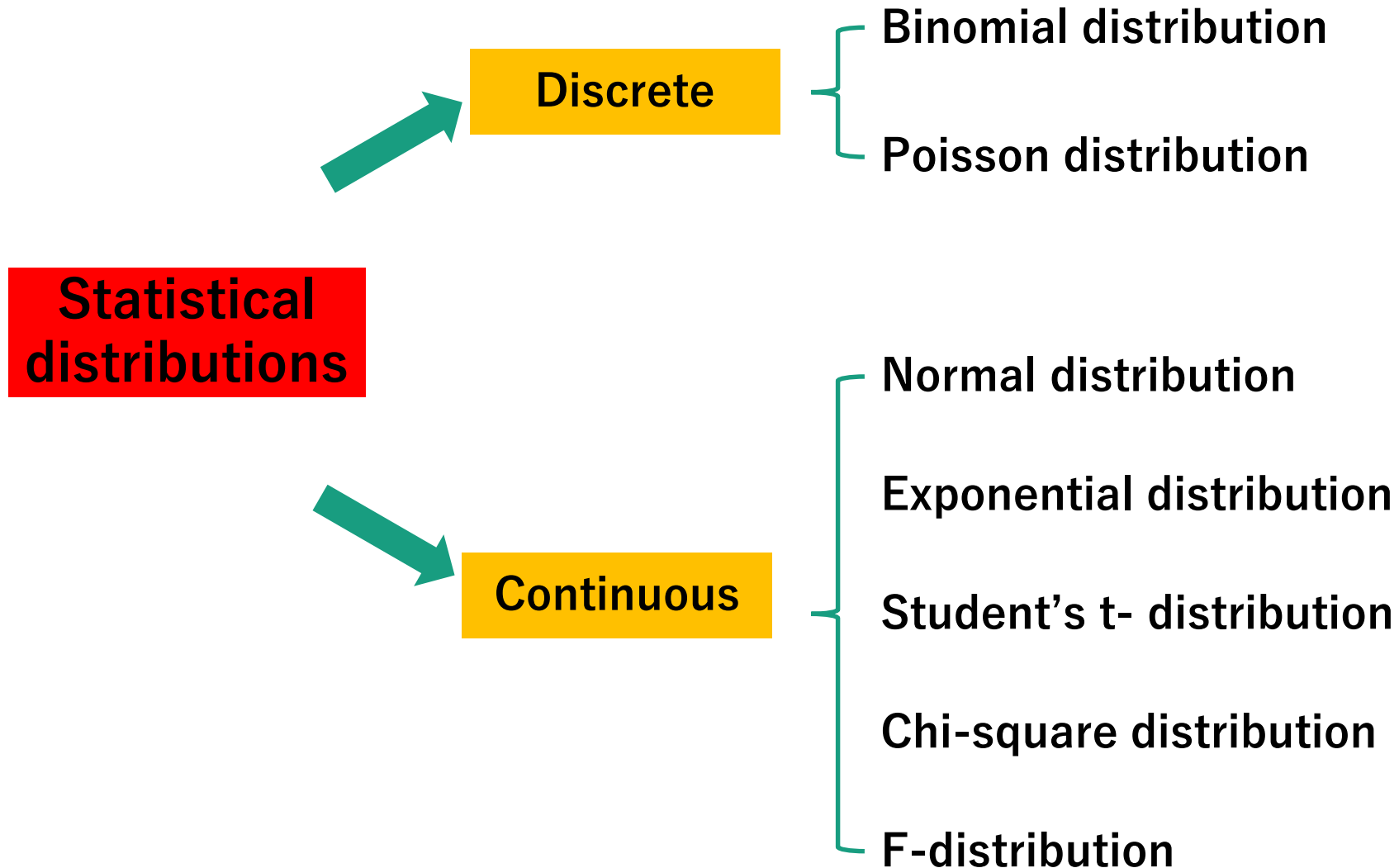
4. Basic statistics 1 distribution



Objective

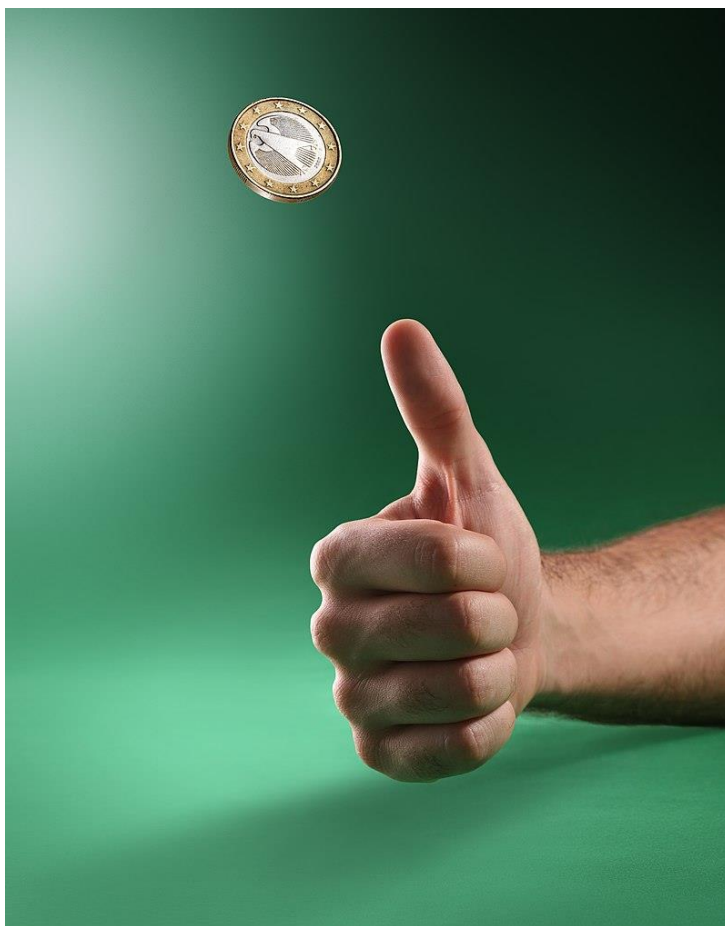
- **Understanding various distributions for data analysis**

Basic distributions



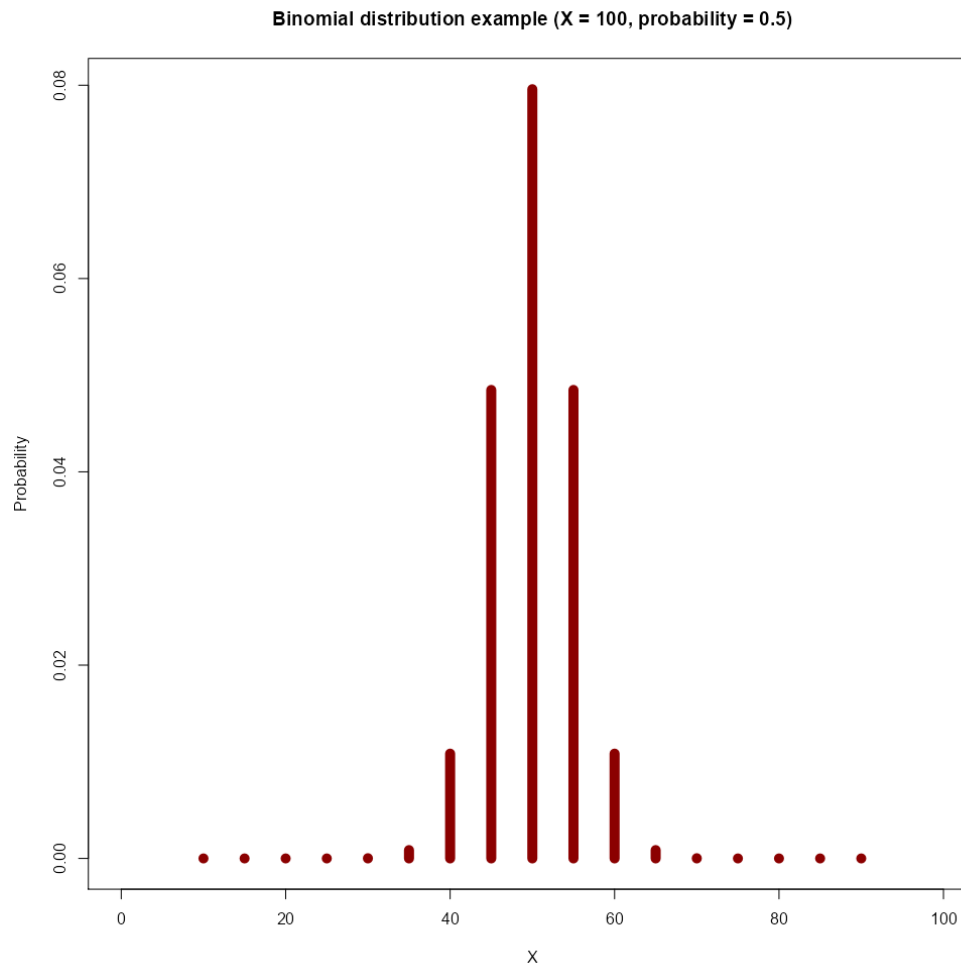
Binomial distribution

- Perform ' n ' times of Bernoulli trial with probability ' p '



Binomial distribution

■ Ex) Binomial distribution graph

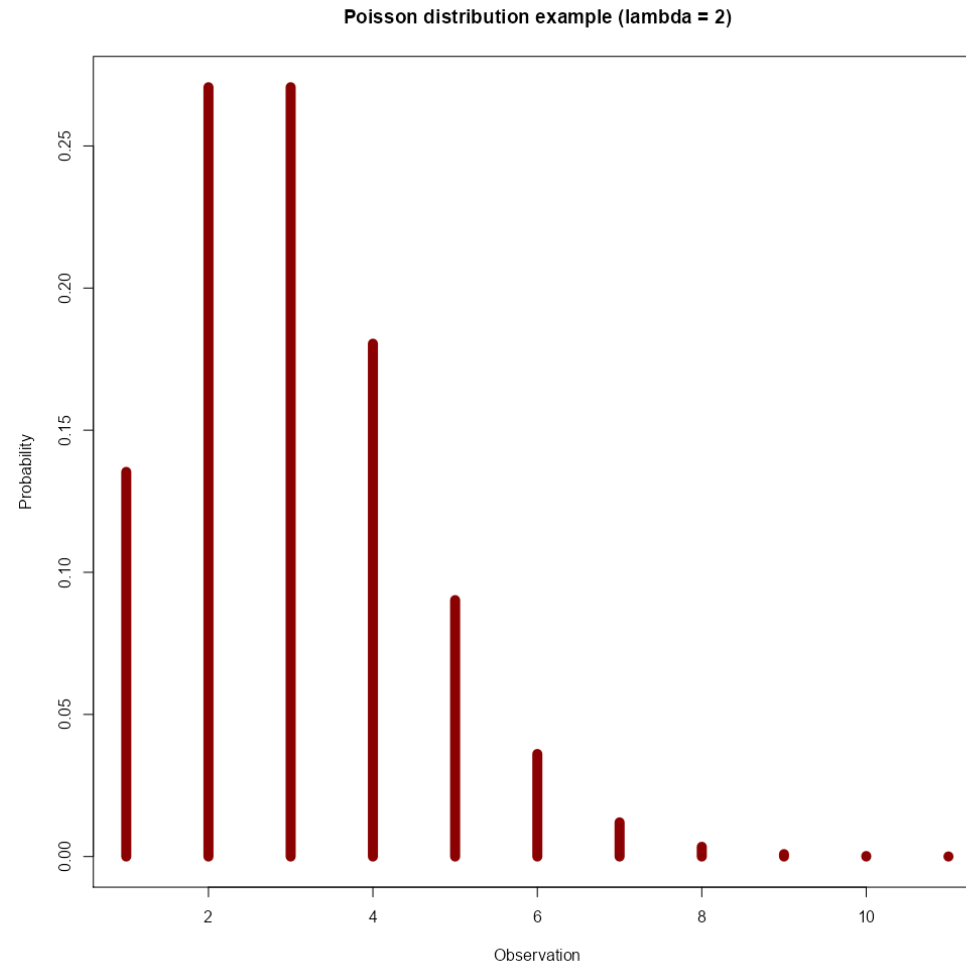


Poisson distribution

- An exponential function that was invented to calculate binomial distribution
- Suppose X is binomial distribution $B(n, p)$ and set $np = \lambda$. When n is big enough and p is close to 0, its shape is close to Poisson distribution.
- Poisson distribution is widely used when an event occurs randomly in a space for a certain period of time

Poisson distribution

■ Ex) Poisson distribution graph



Poisson distribution

Ex) The number of customers per hour at a store follows the Poisson distribution of $\lambda = 30$. What is the probability of the number of customers per hour is 27?

```
> dpois(x = 27, lambda = 30)*100  
[1] 6.553248
```

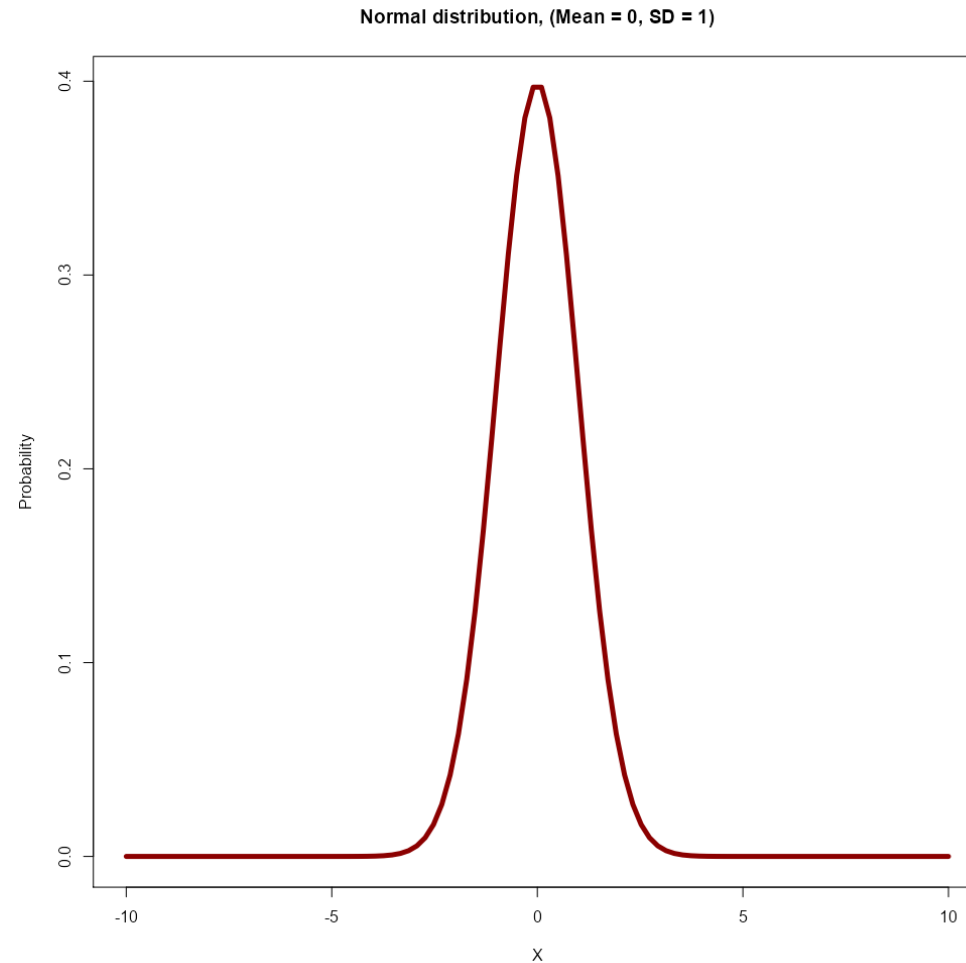
Around 6.55%

Normal distribution

- Distribution that plays essential roles in various statistical estimation
- When your sample size is big enough (normally, 30 or higher), all other distributions converge to a normal distribution (the central limit theorem)
- When mean is 0 and standard deviation is 1, we call it standard normal distribution

Normal distribution

■ Ex) Normal distribution graph

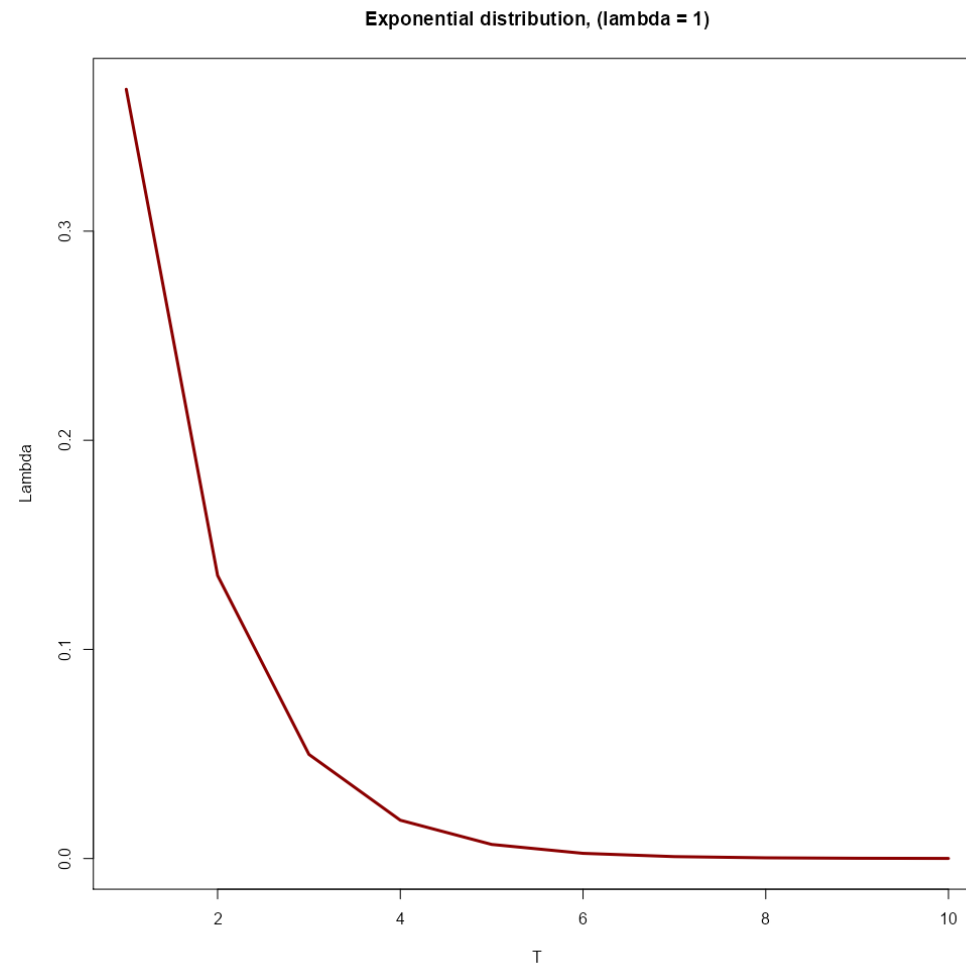


Exponential distribution

- The time it takes for an event to occur
- It looks similar to Poisson distribution but they are different!
- For example, the time it takes for a brand A washing machine to break down

Exponential distribution

Ex) Exponential distribution graph

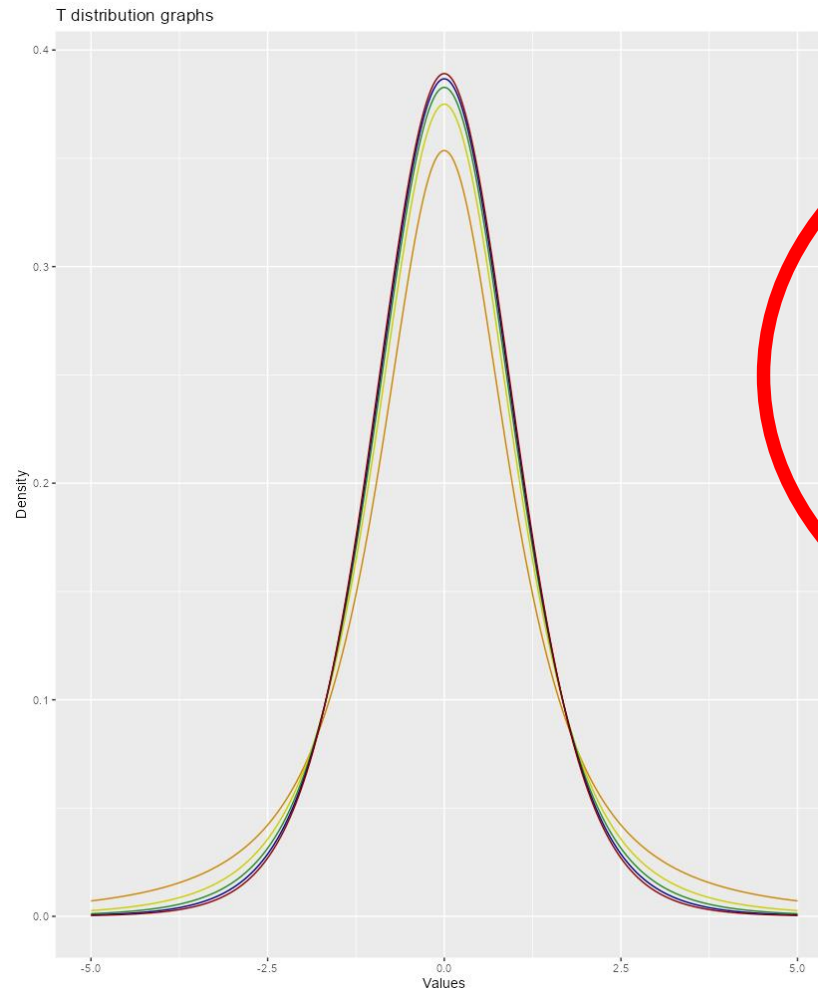


Student's t- distribution

- This distribution is used when you don't know distribution of statistical population is unknown or have small number of samples
- Commonly used when you estimate population mean
- When the number of samples are large enough ($N \geq 30$), it converges to a normal distribution

Student's t- distribution

Ex) Comparison of Student's t- distribution graphs



colour

- T distribution df = 10
- T distribution df = 2
- T distribution df = 4
- T distribution df = 6
- T distribution df = 8

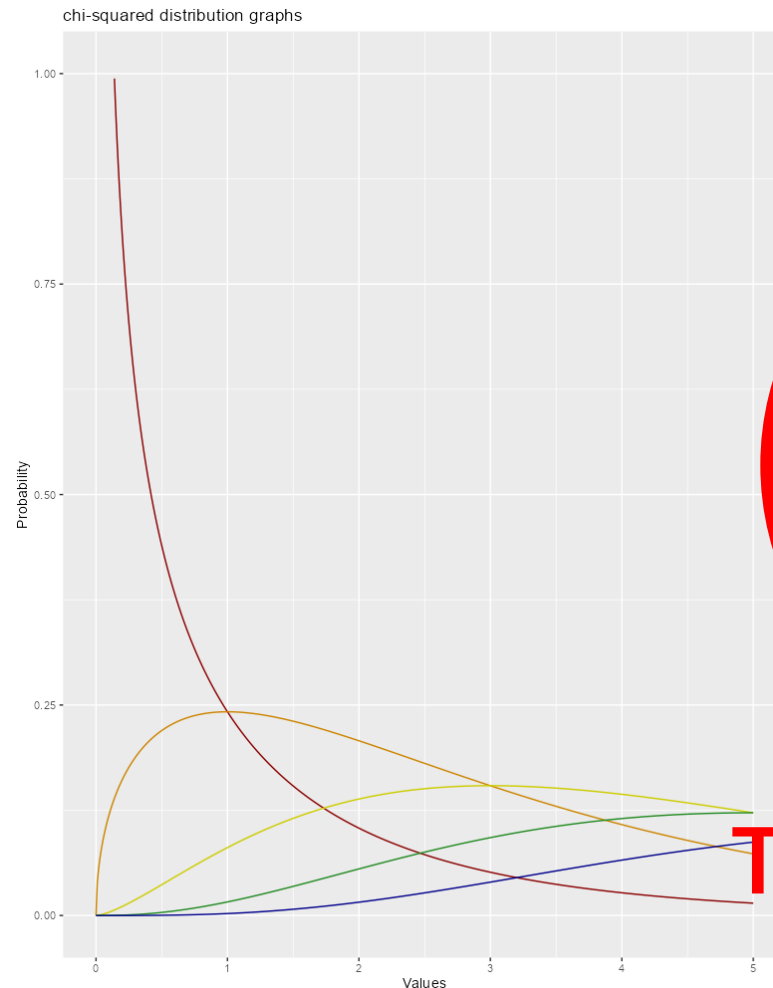
They use degree of freedom

Chi-square distribution

- This distribution is used when you estimate population variance
- It is normally used when you check suitability and independence of categorical data
Ex) The difference in types of desserts that men and women like

Chi-square distribution

Ex) Comparison of F-distribution graphs



colour

- chi-squared distribution df = 1
- chi-squared distribution df = 3
- chi-squared distribution df = 5
- chi-squared distribution df = 7
- chi-squared distribution df = 9

They use degree of freedom

F-distribution

- This distribution is used when you estimate similarity of variance
- This distribution is used when you perform ANOVA

F-distribution

Ex) Comparison of F-distribution graphs



colour

- F distribution df = 10 and 20
- F distribution df = 10 and 40
- F distribution df = 10 and 60
- F distribution df = 10 and 80

They use degree of freedom

References

- <https://rfriend.tistory.com/> (Korean)
- <https://rvisuall.tistory.com/133> (Korean)
- <https://www.geeksforgeeks.org/exponential-distribution-in-r-programming-dexp-pexp-qexp-and-rexp-functions/>