R 語言教學

鍾旻錡, 陳柏瑜

Statistics with Recitation

NTU Econ

2020.12.23

Outline

Maximum Likelihood Estimation in R

Concept of MLE

Given random samples

$$\{X_i\}_{i=1}^{n} \overset{i.i.d.}{\sim} f_X(x;\theta)$$

where θ is a unknown parameter

As long as θ is given, we can know the probability of seeing a certain outcome x_i

Concept of MLE

To estimate θ, we simply consider the converse relationship between random samples and their parameters.

$f_X(x_i;\theta)$ tells the probability of seeing a certain random outcome x_i given θ

$L(\theta;x_i)$ tells the likelihood of a true parameter can be after we have seen a bunch of random samples x_i

D.G.P. and pseudo random numbers

We can "make up" data in order to understand the Maximum Likelihood Estimation.

First, we pretend knowing the true parameters.

Second, generate a bunch of random numbers from the above process. (D.G.P.)

Third, we pretend not knowing the above information.

Fourth, write down the likelihood function according to the distribution function, and do the numerical maximization.

One Parameter

D.G.P.

$$\{X_i\}_{i=1}^{n} \overset{i.i.d.}{\sim} Bernoulli(\mu)$$

where $n = 100$, $\mu = 0.87$

One Parameter

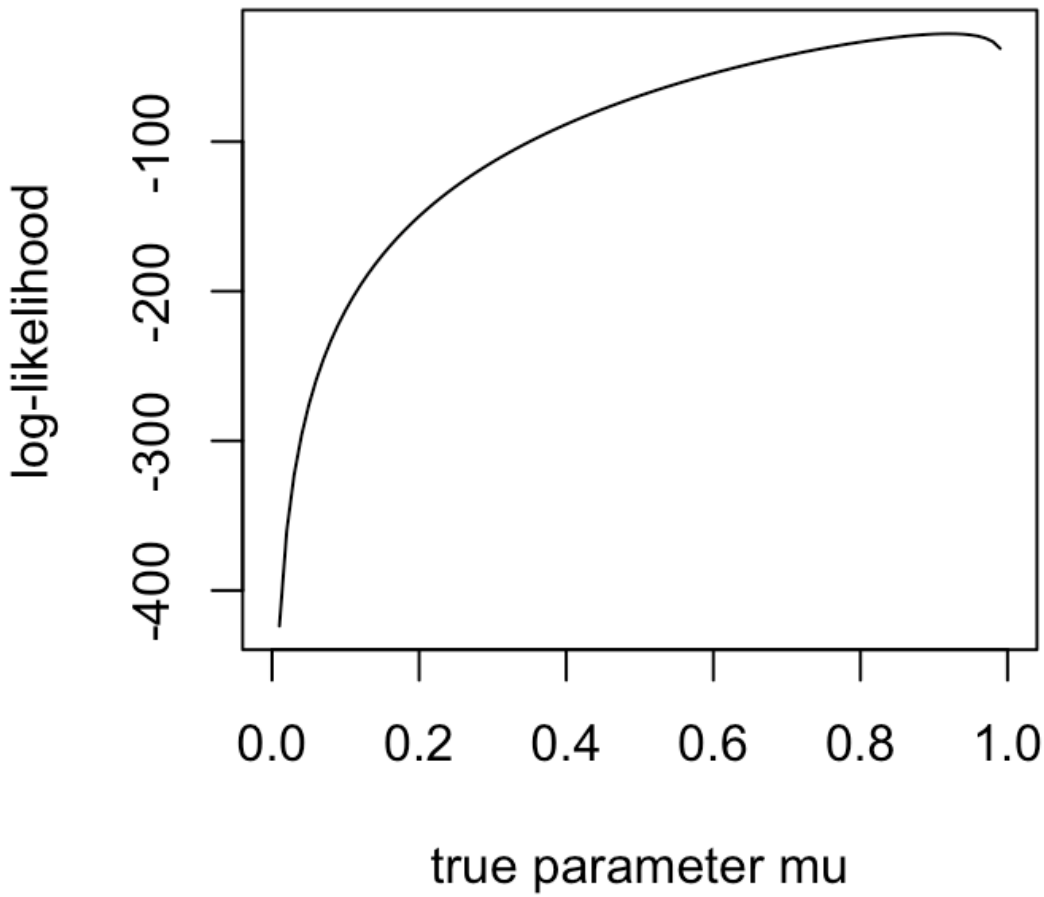
The likelihood function is:

$$L(\mu) = \prod_{i=1}^{n} \mu^{x_i}(1-\mu)^{1-x_i}$$

$$\ell(\mu) = \sum_{i=1}^{n} x_i log(\mu) + \sum_{i=1}^{n}(1-x_i) log(1-\mu)$$

Note that "taking log" is essential for the computer to do the maximization problem.

One Parameter: Graph



One Parameter: Numerical Maximization

In R, we can use ‘`nlm()`‘ function to do the "non-linear minimization."

Alternatively, we can use ‘`mle()`‘ in ‘`stat4`‘ package.

The previous one provides one parameter estimation, while the latter one provides estimation for multiple parameters.

Other numerical maximization functions such as ‘`optim()`‘ can also be used.

One Parameter: nlm() usage

```
nlm(objective function, staring points, print.level =
2, hessian = T)
```

- objective function: the function you want to minimize by deciding the variable θ
- staring points: the initial value for θ
- print.level: show all information for each iteration
- hessian: show the S.O.C.

One Parameter: Requirements to meet

Remember that numerical maximization may highly depends on:

1. starting points
2. smoothness of the objective function

And the iteration depends on:

1. gradient (F.O.C.)
2. Hessian (S.O.C.)

One Parameter: Practice 1

Given

$$Pr(X = x) = f_X(x) = p(1-p)^x, x = 0, 1, 2, 3, \ldots$$

Now, read the csv file `practice1.csv` provided as the random samples.

Find the maximum likelihood estimate for p with R

Two Parameters

D.G.P.

$$\{Y_i\}_{i=1}^{n} \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$$

where $n = 100, \mu = 0.9487, \sigma^2 = 9.487^2$

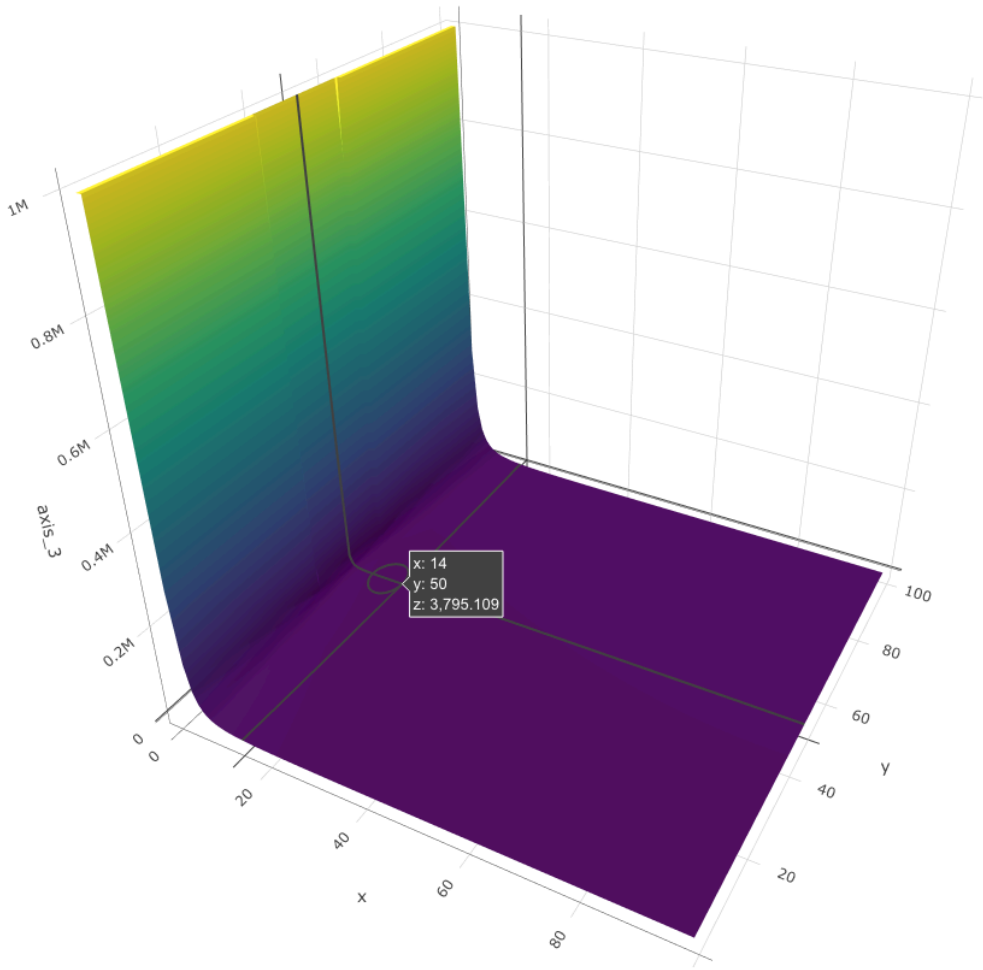
and ‘`set.seed(1234)`‘

Two Parameters

The likelihood function is:

```
mll = function(mu, sigma){
  logLikelihood <-0
  for(y in Y){
    logLikelihood <- logLikelihood + log(dnorm(y, mean = m
  }
  return(-logLikelihood)
}
```

Two Parameters: Graph



Two Parameter: mle() usage

```
mle(objective function, start)
```

- objective function: the function you want to minimize by deciding the variable θ
- start: a named list for θ

Two Parameter: Practice 2

Given

$$\{X_i\}_{i=1}^{n} \overset{i.i.d.}{\sim} F(v_1, v_2)$$

Suppose the D.G.P. is:

$n = 10000$, $v_1 = 2020$, $v_2 = 1223$, and $set.seed(20201223)$

Find the maximum likelihood estimate for v_1, v_2 with R

Interval Estimator and Interval Estimate

Interval Estimate

Given

$$\{X_i\}_{i=1}^{n} \overset{i.i.d.}{\sim} Bernoulli(\mu)$$

the D.G.P. is:

$n = 100, \mu = 0.87$

Construct the $(1-\alpha)100\%$ Interval Estimator / Estimate

Interval Estimate

If $\alpha = 0.05$, then the interval estimate is:

$$\bar{X}_n - z_{\frac{0.05}{2}} \frac{s_n}{\sqrt{n}}, \bar{X}_n + z_{\frac{0.05}{2}} \frac{s_n}{\sqrt{n}}$$

Write the above in R:

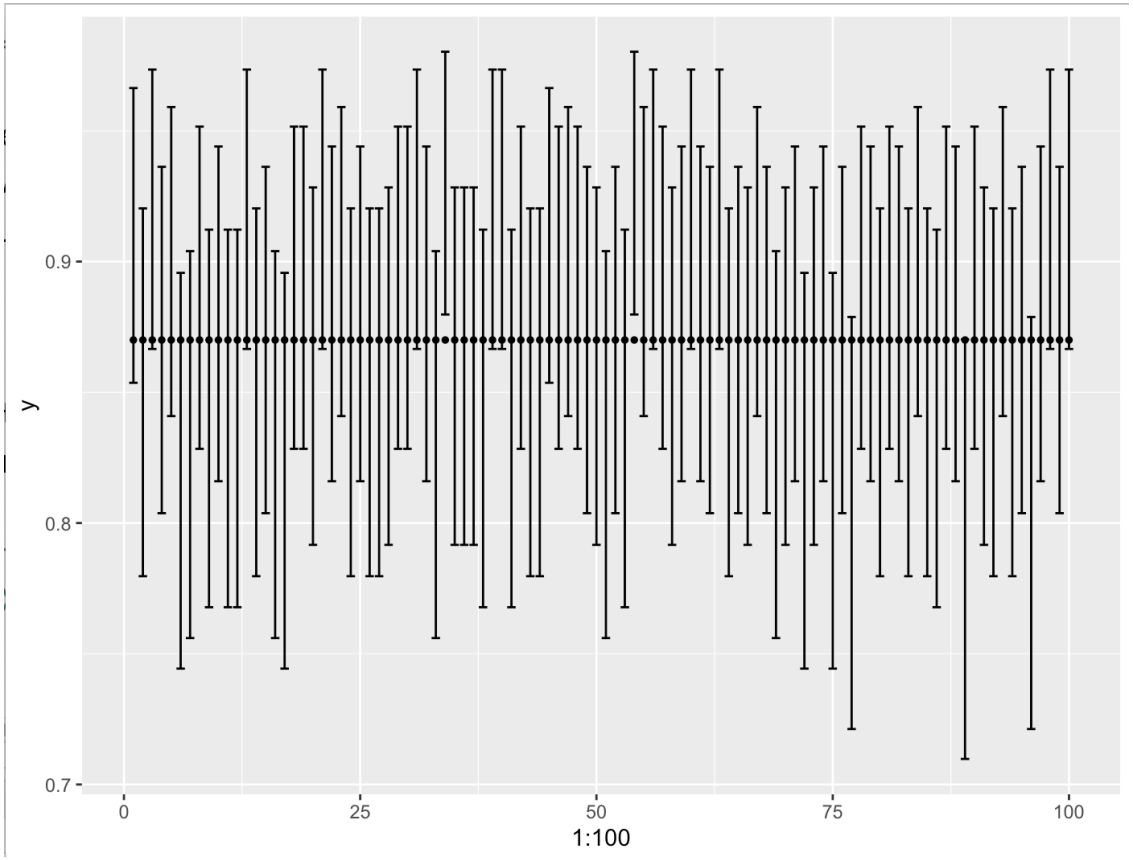
```
mean(X)-qnorm(0.975)*sd(X)/sqrt(n)
mean(X)+qnorm(0.975)*sd(X)/sqrt(n)
```

Interval Estimator

We may write a function that generates intervals by the same D.G.P., and then we can realize what "confidence" means.

```
confidence_interval = function(n=100, alpha=0.05){
  mu = 0.87
  X <- rbinom(n, size = 1, prob = mu)
  # qnorm(0.025); qnorm(0.975)
  l = mean(X)-qnorm(1-alpha/2)*sd(X)/sqrt(n)
  u = mean(X)+qnorm(1-alpha/2)*sd(X)/sqrt(n)
  return(c(l, u))
```

Interval Estimator



Interval Estimate: Practice 3

Consider

$$\{X_i\}_{i=1}^{n} \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$$

where both μ, σ^2 are unknown

Construct the 95% Interval Estimate for σ^2

Hypothesis Test

Hypothesis Test: Two Aspects

Given α,

1. Consider whether the parameter of interest is in the interval estimate I constructed
2. Suppose null hypothesis is true, then how unlikely it is that I can see the realized random samples?

Hypothesis Test: Aspect 1

Given

$$\{X_i\}_{i=1}^{n} \overset{i.i.d.}{\sim} Bernoulli(\mu)$$

the D.G.P. is:

$n = 100, \mu = 0.87$

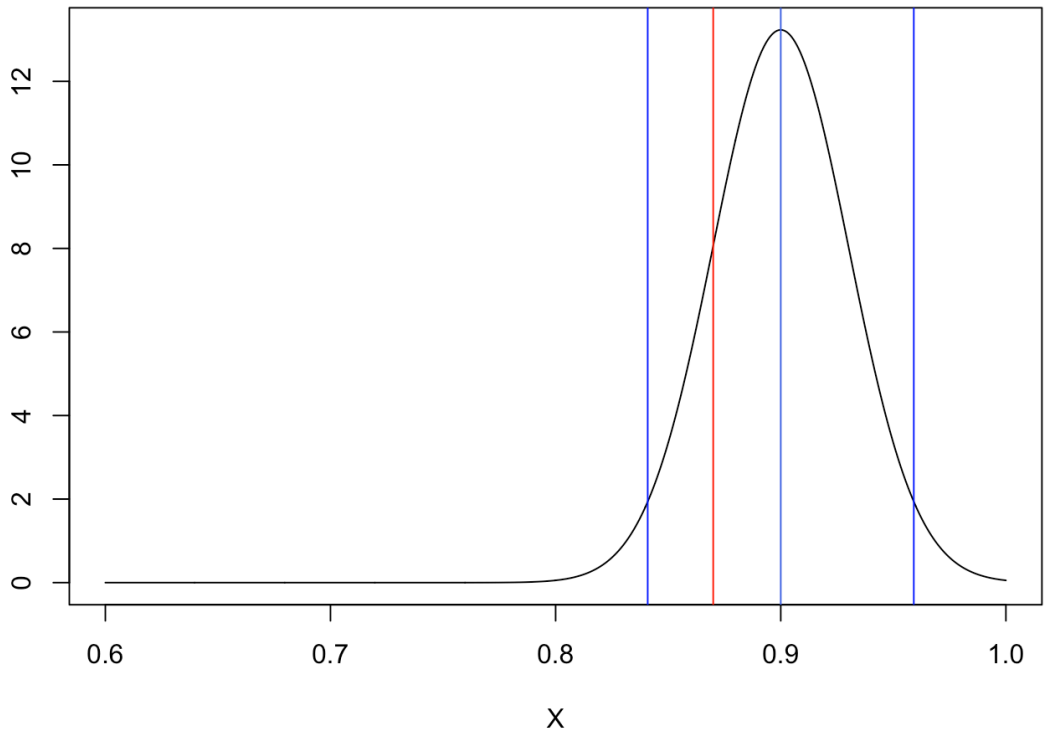
Construct the $(1-\alpha)100\%$ Interval Estimator / Estimate

Hypothesis Test: Aspect 1

If someone claims: the true parameter μ is exactly 0.87

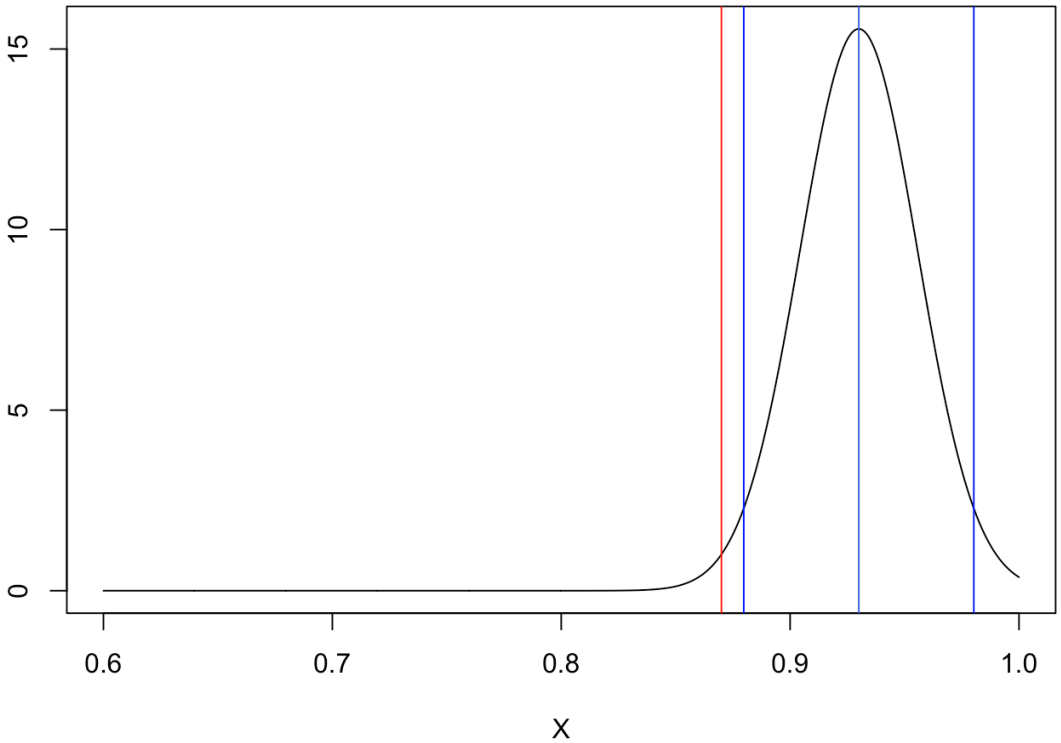
What can we say depends on the realized random samples?

Hypothesis Test: Aspect 1



Red: null hypothesis; Blue: random sample we saw

Hypothesis Test: Aspect 1



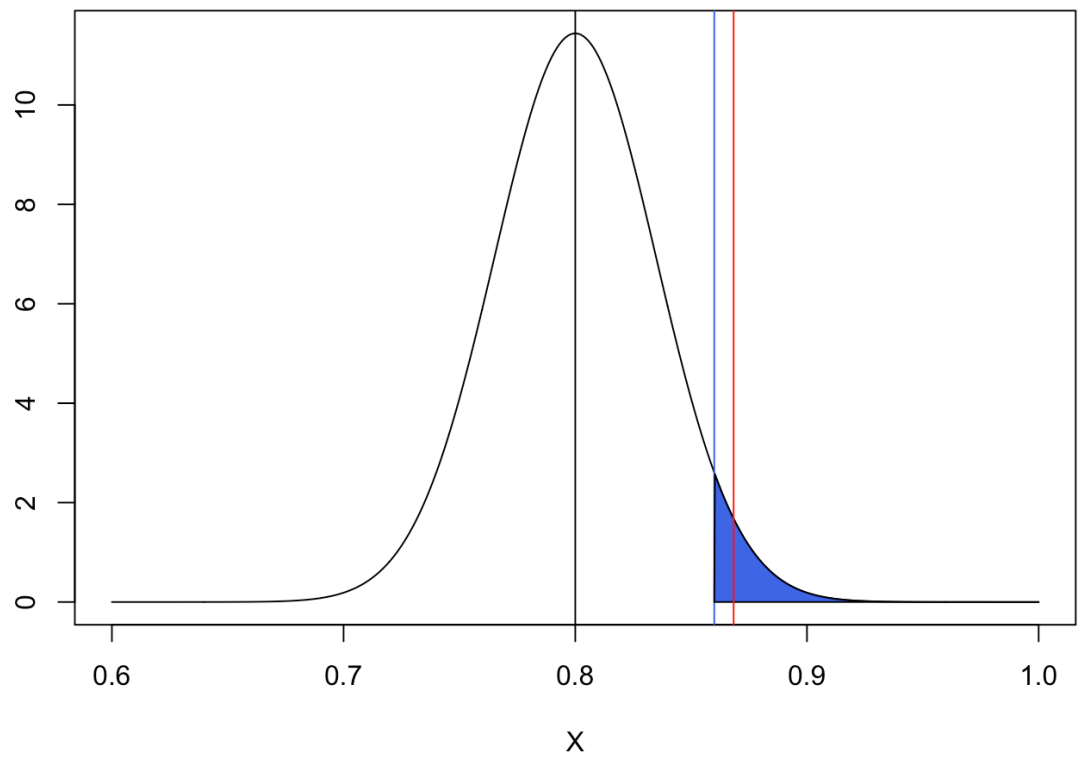
Red: null hypothesis; Blue: random sample we saw

Hypothesis Test: Aspect 2

If someone claims: the true parameter μ is exactly 0.87

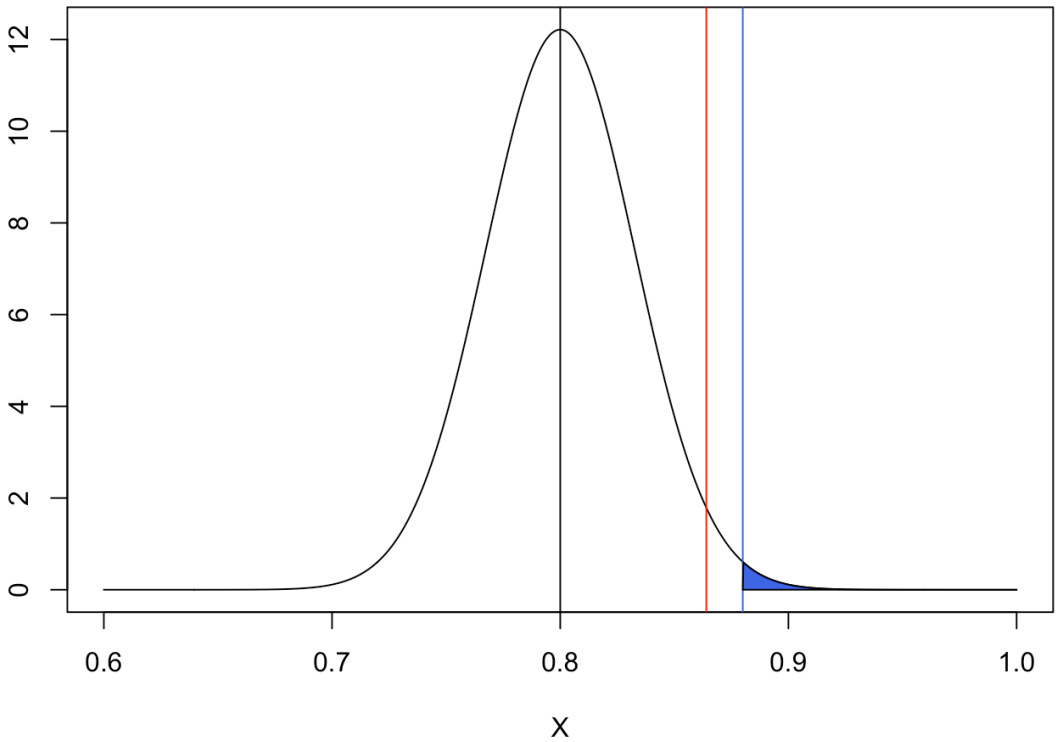
Now suppose what he said is true, then how could we verify his claim?

Hypothesis Test: Aspect 2



Red: significance level 0.05; Blue: random sample we saw; Black: null hypothesis

Hypothesis Test: Aspect 2



Red: significance level 0.05; Blue: random sample we saw; Black: null hypothesis

Hypothesis Test: Practice 4

Given

$$\{X_i\}_{i=1}^{n} \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$$

the D.G.P. is:

$n = 10000, \mu = \pi, \sigma^2 = \pi$

$$H_0 : \sigma^2 = 3$$

$$H_A : \sigma^2 \neq 3$$

Find the test statistics if $\alpha = 0.05$

Permutation Test

An Alternative Approach of Hypothesis Test

- Example in **Mathematical Statistics with Re-sampling and R** Ch3.1 (Chihara)
- Scientists invent a new drug that supposedly will inhibit a mouse's ability to run through a maze.
- The average time for the drug group is 25 s, and the average time for the control group is 20.33 s.
- The mean difference in times is 25 20.33 = 4.67 s.

Drug			Control		
30	25	20	18	21	22

An Alternative Approach of Hypothesis Test

We cannot tell for sure whether there is a real effect. What we do instead is to estimate how easily pure random chance would produce a difference this large. If that probability is small, then we conclude there is something other than pure random chance at work, and conclude that there is a real effect.

If what we saw is:

Drug			Control		
30	**25**	18	**20**	21	22

An Alternative Approach of Hypothesis Test

Original:

Drug			Control		
30	**25**	**20**	18	21	22

After permutating:

Drug			Control		
30	**25**	18	**20**	21	22

There are $\binom{6}{3}$ combinations for these samples.

An Alternative Approach of Hypothesis Test

Table 3.1 All possible distributions of {30, 25, 20, 18, 21, 22} into two sets.

Drug			Control			$\bar{X}_D$	$\bar{X}_C$	Difference in means
18	20	21	22	25	30	19.67	25.67	−6.00
18	20	22	21	25	30	20	25.33	−5.33
18	20	25	21	22	30	21	24.33	−3.33
18	20	30	21	22	25	22.67	22.67	0.00
18	21	22	20	25	30	20.33	25	−4.67
18	21	25	20	22	30	21.33	24	−2.67
18	21	30	20	22	25	23	22.33	0.67
18	22	25	20	21	30	21.67	23.67	−2.00
18	22	30	20	21	25	23.33	22	1.33
18	25	30	20	21	22	24.33	21	3.33
20	21	22	18	25	30	21	24.33	−3.33
20	21	25	18	22	30	22	23.33	−1.33
20	21	30	18	22	25	23.67	21.67	2.00
20	22	25	18	21	30	22.33	23	−0.67
20	22	30	18	21	25	24	21.33	2.67
20	25	30	18	21	22	**25**	**20.33**	**4.67** *
21	22	25	18	20	30	22.67	22.67	0.00
21	22	30	18	20	25	24.33	21	3.33
21	25	30	18	20	22	**25.33**	**20**	**5.33** *
22	25	30	18	20	21	**25.67**	**19.67**	**6.00** *

Rows where the difference in means exceeds the original value are highlighted.

An Alternative Approach of Hypothesis Test

- Among 20 possible combinations, there are only 3 of them has a more extreme difference in sample mean. (compared to 4.67)
- H_0 : There is no difference between "Drug" group and "Control" group.(i.e. There is no effect on drug.)
- H_A : There is a negative difference between "Drug" group and "Control" group.(i.e. There is an effect on drug that makes mice run through the maze faster.)
- The relative frequency of seeing a more extreme "difference in sample means" is the "p-value" under H_0

Is Permutation Feasible? Re-sampling!

- Extend to n mice, then there are $\binom{n}{\frac{n}{2}}$ combinations
- To calculate p-value, We need re-sampling

Permutation Test Example

- Chihara Ch3.3
- Investigating the consumption of hotwings and beer in a bar, while recording the gender.
- Goal: Do men consume more hot wings than women?

Table 1.7 Variables in data set `Beerwings`.

Variable	Description
Gender	Male or female
Beer	Ounces of beer consumed
Hot Wings	Number of hot wings eaten

Permutation Test Example

- 30 observations
- H_0 : Gender does not affect the consumption of hot wings
- H_A : Men do consume more hot wings.
- Under H_0, the following is one of the possible outcomes

Females					Males				
5	**6**	7	7	**8**	**4**	**5**	7	8	**9**
8	**11**	**12**	**13**	**14**	11	**12**	13	13	**13**
14	14	16	16	21	17	17	18	18	21

- Note that there are $\binom{30}{15} = 155,117,520$ combinations under H_0

Permutation Test Example

- We can create permutation resample since all possible outcomes are weighted with same probability.
- What we want is calculating p-value. We can draw some outcomes from the permutations and see whether they are extreme events or not.
- The number of extreme events over the number we draw (which is the relative frequency) is the approximation of p-value.

感謝大家聆聽