Chương 0. Kiến thức chuẩn bị

0.1. Một số phân phối xác suất thông dụng

Phân phối Bernoulli

Giả sử kết quả khi thực hiện một phép thử chỉ xảy ra hai trường hợp: có hoặc không xảy ra tính chất nghiên cứu, với xác suất thành công là p.

Đặt $X = 1$ nếu tính chất nghiên cứu xảy ra và $X = 0$ nếu tính chất nghiên cứu không xảy ra. Khi đó, ta có bảng phân phối xác suất của X.

X	0	1
P	p	$1-p$

Các tham số đặc trưng của phân phối Bernoulli là

$$E(X) = p, Var(X) = p(1-p).$$

Ví dụ 1. Mô phỏng phân phối Bernoulli trong phần mềm R với 10 quan sát, xác suất thành công của tính chất nghiên cứu $p = 0.2$

Code R

```
# bernoulli distribution in r
rbinom(10, 1,.2)
```

```
[1] 1 0 0 0 0 1 0 0 0 0
```

Phân phối nhị thức (binomial distribution)

Thực hiện phép thử Bernoulli n lần. Gọi X là số lần xuất hiện tính chất nghiên cứu trong n lần. Khi đó, X là một biến ngẫu nhiên nhận các giá trị $0, 1, 2, \cdots, n$ với xác suất tương ứng

X	0	1	$\cdots$	n
P	$C_n^0 p^0 (1-p)^n$	$C_n^1 p^1 (1-p)^{n-1}$	$\cdots$	$C_n^n p^n (1-p)^0$

Các tham số đặc trưng của phân phối nhị thức

$$E(X) = np, Var(X) = np(1-p).$$

Ví dụ 2. Tính xác suất phân phối nhị thức X nhận giá trị bằng 5 khi thực hiện phép thử 10 lần và xác suất thành công bằng 0.5

Code R

```
# dbinom r - calculate binomial probability in r
dbinom(5, size=10, prob=0.5)
```

```
[1] 0.2460938
```

Ví dụ 3. Tính xác suất phân phối nhị thức X nhận giá trị nhỏ hơn hoặc bằng 5 khi thực hiện phép thử 10 lần với xác suất thành công bằng 0.5.

Code R

```
# pbinom in r - binomial probability r
pbinom(5,10,0.5)
```

```
[1] 0.6230469
```

Ví dụ 4. Tìm số lần thực hiện thành công khi xác suất thành công nhỏ hơn hoặc bằng số lần thực hiện thành công là 0.56, biết thực hiện phép thử 20 lần với xác suất thành công là 0.46

Code R

```
# r qbinom - inverse binomial distribution
qbinom(0.56,20,.46)
```

```
[1] 10
```

Phân phối chuẩn

Giả sử biến ngẫu nhiên X tuân theo phân phối chuẩn với trung bình μ và phương sai σ^2, trong đó $-\infty < \mu < \infty, \sigma^2 > 0$. Khi đó, hàm mật độ xác suất $f(x)$ có dạng

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Các tham số đặc trưng của phân phối chuẩn

$$E(X) = \mu, Var\,(X) = \sigma^2.$$

Trong trường hợp $\mu = 0, \sigma = 1$ thì phân phối chuẩn $N(0,1)$ được gọi là phân phối chuẩn tắc.

Ví dụ 5. Tính giá trị của hàm mật độ xác suất tuân theo phân phối chuẩn với trung bình bằng 2.5 và độ lệch chuẩn bằng 0.5 từ các giá trị -10 đến 10 với khoảng cách các điểm chia bảng 0.1:

Code R

```
# Create a sequence of numbers between -10 and 10 incrementing by 0.1.
x <- seq(-10, 10, by = .1)
# Choose the mean as 2.5 and standard deviation as 0.5.
y <- dnorm(x, mean = 2.5, sd = 0.5)
```

```
plot(x,y)
```

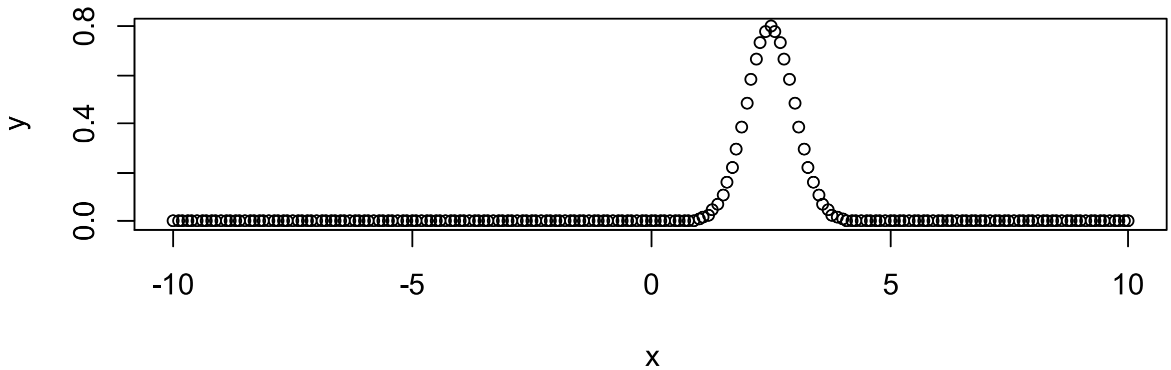
Các giá trị của hàm mật độ xác suất y:

```
  [1] 1.530786e-136 2.226901e-134 3.112546e-132 4.179831e-130 5.392993e-128 6.685429e-126
  [7] 7.962637e-124 9.111980e-122 1.001836e-119 1.058301e-117 1.074112e-115 1.047414e-113
 [13] 9.813304e-112 8.833655e-110 7.640008e-108 6.348563e-106 5.068568e-104 3.887974e-102
 [19] 2.865429e-100  2.029010e-98  1.380406e-96  9.023141e-95  5.666787e-93  3.419356e-91
 [25]  1.982348e-89  1.104190e-87  5.909296e-86  3.038477e-84  1.501082e-82  7.124939e-81
 [31]  3.249272e-79  1.423702e-77  5.993501e-76  2.424210e-74  9.420804e-73  3.517499e-71
 [37]  1.261851e-69  4.349213e-68  1.440262e-66  4.582477e-65  1.400836e-63  4.114365e-62
 [43]  1.161038e-60  3.147880e-59  8.200081e-58  2.052326e-56  4.935178e-55  1.140217e-53
 [49]  2.531048e-52  5.398107e-51  1.106142e-49  2.177752e-48  4.119402e-47  7.486661e-46
 [55]  1.307285e-44  2.193213e-43  3.535245e-42  5.475028e-41  8.146695e-40  1.164675e-38
 [61]  1.599766e-37  2.111233e-36  2.676974e-35  3.261221e-34  3.817198e-33  4.292767e-32
 [67]  4.638294e-31  4.815122e-30  4.802691e-29  4.602461e-28  4.237639e-27  3.748745e-26
 [73]  3.186222e-25  2.601923e-24  2.041461e-23  1.538920e-22  1.114600e-21  7.756224e-21
 [79]  5.185729e-20  3.331176e-19  2.055955e-18  1.219152e-17  6.945925e-17  3.802163e-16
 [85]  1.999676e-15  1.010454e-14  4.905711e-14  2.288313e-13  1.025551e-12  4.415980e-12
 [91]  1.826944e-11  7.261923e-11  2.773360e-10  1.017628e-09  3.587568e-09  1.215177e-08
 [97]  3.954639e-08  1.236524e-07  3.714724e-07  1.072207e-06  2.973439e-06  7.922598e-06
[103]  2.028170e-05  4.988494e-05  1.178861e-04  2.676605e-04  5.838939e-04  1.223804e-03
[109]  2.464438e-03  4.768176e-03  8.863697e-03  1.583090e-02  2.716594e-02  4.478906e-02
[115]  7.094919e-02  1.079819e-01  1.579003e-01  2.218417e-01  2.994549e-01  3.883721e-01
[121]  4.839414e-01  5.793831e-01  6.664492e-01  7.365403e-01  7.820854e-01  7.978846e-01
[127]  7.820854e-01  7.365403e-01  6.664492e-01  5.793831e-01  4.839414e-01  3.883721e-01
[133]  2.994549e-01  2.218417e-01  1.579003e-01  1.079819e-01  7.094919e-02  4.478906e-02
[139]  2.716594e-02  1.583090e-02  8.863697e-03  4.768176e-03  2.464438e-03  1.223804e-03
[145]  5.838939e-04  2.676605e-04  1.178861e-04  4.988494e-05  2.028170e-05  7.922598e-06
[151]  2.973439e-06  1.072207e-06  3.714724e-07  1.236524e-07  3.954639e-08  1.215177e-08
[157]  3.587568e-09  1.017628e-09  2.773360e-10  7.261923e-11  1.826944e-11  4.415980e-12
[163]  1.025551e-12  2.288313e-13  4.905711e-14  1.010454e-14  1.999676e-15  3.802163e-16
[169]  6.945925e-17  1.219152e-17  2.055955e-18  3.331176e-19  5.185729e-20  7.756224e-21
[175]  1.114600e-21  1.538920e-22  2.041461e-23  2.601923e-24  3.186222e-25  3.748745e-26
[181]  4.237639e-27  4.602461e-28  4.802691e-29  4.815122e-30  4.638294e-31  4.292767e-32
[187]  3.817198e-33  3.261221e-34  2.676974e-35  2.111233e-36  1.599766e-37  1.164675e-38
```

[193] 8.146695e-40 5.475028e-41 3.535245e-42 2.193213e-43 1.307285e-44 7.486661e-46

[199] 4.119402e-47 2.177752e-48 1.106142e-49

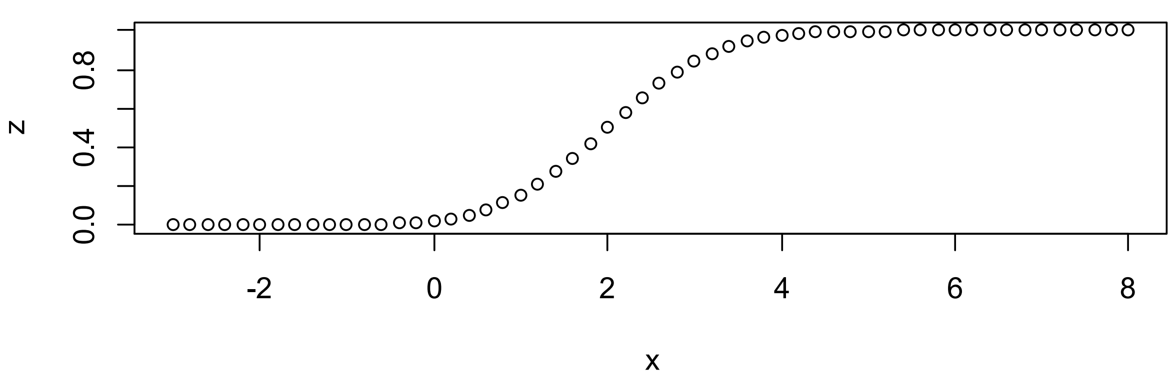
Đồ thị của hàm mật độ xác suất phân phối chuẩn $N(2.5, 0.5^2)$



Ví dụ 6. Xác định đồ thị của hàm phân phối xác suất tuân theo phân phối chuẩn $N(2,1^1)$

Code R

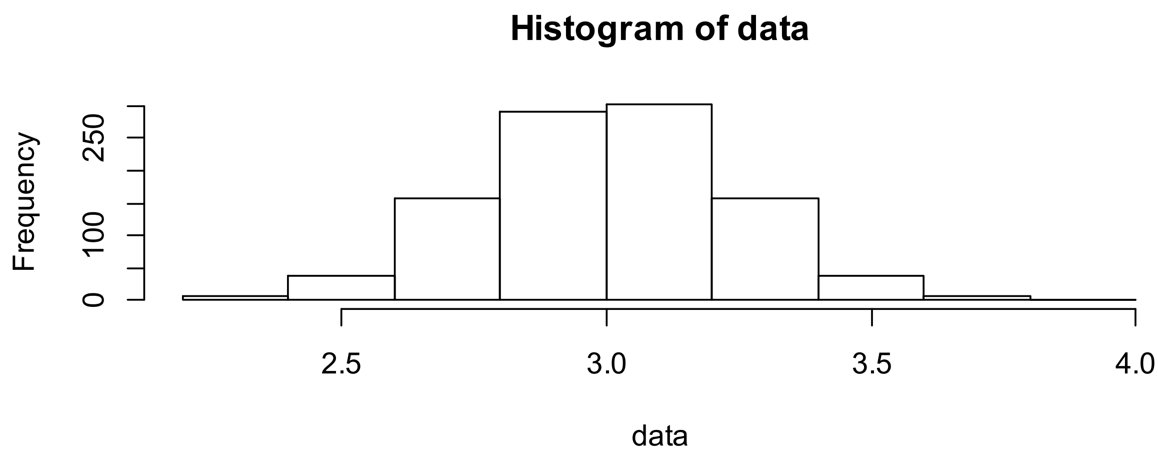
```
# Create a sequence of numbers between -3 and 8 incrementing by 0.2.
x <- seq(-3,8,by = .2)
# Choose the mean as 2 and standard deviation as 1.
z <- pnorm(x, mean = 2, sd = 1)
plot(x,z)
```



Ví dụ 7. Mô phỏng dữ liệu bao gồm 1000 quan sát từ phân phối chuẩn với trung bình bằng 3 và độ lệch chuẩn bằng 0.25 đồng thời vẽ đồ thị biểu diễn dữ liệu

Code R

```
data <- rnorm(1000, 3, .25)
hist(data)
```



Ví dụ 8. Tính giá trị xác suất của các giá trị nhỏ hơn 1.96 khi dữ liệu tuân theo phân phối chuẩn tắc $N(0,1)$

Code R

```
pnorm(1.96, 0, 1)
```

```
[1] 0.9750021
```

Ví dụ 9. Giá trị của phân phối chuẩn tắc mà xác suất của các giá trị nhỏ hơn giá trị cho trước bằng xác suất cố định

Code R

```
qnorm(0.975, 0, 1)
```

```
[1] 1.959964
```

0.2. Ước lượng hợp lý cực đại

Định nghĩa 1. Ước lượng hợp lý cực đại (MLE) $\hat{\theta}$ giá trị θ trong đó hàm hợp lý $L(\theta) = f(x|\theta)$ đạt giá trị supremum, tức là

$$sup_{\theta} f(x|\theta) = f(x|\hat{\theta})$$

Ước lượng hợp lý cực đại về trung bình của tổng thể tuân theo phân phối chuẩn và đã biết phương sai của tổng thể

Giả sử mẫu ngẫu nhiên $x_1, x_2, \cdots, x_n$ các quan sát độc lập và cùng tuân theo phân phối chuẩn $N(\mu, \sigma^2)$. Khi đó, ước lượng trung bình của tổng thể có dạng

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right] = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}\right]$$

Biến đổi tử số của số mũ ta có

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}\left[(x_i - \overline{X}) + (\overline{X} - \mu)\right]^2 = \sum_{i=1}^{n}(x_i - \overline{X})^2 + n(\overline{X} - \mu)^2$$

$$= (n-1)S^2 + n(\overline{X} - \mu)^2$$

Để hàm hợp lý đạt giá trị cực đại, ta cần biểu thức exp đạt giá trị cực đại, tức là $\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$ đạt giá trị cực tiểu, hay $\sum_{i=1}^{n}(x_i - \mu)^2$ đạt giá trị cực tiểu. Điều này đồng nghĩa với $n(\overline{X} - \mu)^2$ đạt giá trị cực tiểu, do $(n-1)S^2$ là hằng số.

Do đó, ước lượng hợp lý cực đại của trung bình tổng thể μ chính là trung bình mẫu $\overline{X}$.

0.3. Các bài toán cơ bản trong suy luận của thống kê cổ điển

Ước lượng điểm

Ước lượng khoảng

Định nghĩa 1. Ước lượng khoảng cho tham số θ với độ tin cậy $(1 - \alpha)$ là (θ_1, θ_2) nếu

$$P(\theta_1 < \theta < \theta_2) = 1 - \alpha.$$

Các trường hợp của khoảng tin cậy:

- Nếu khoảng tin cậy có dạng $(-\infty, \theta_2)$ thì gọi là khoảng tin cậy phía trái hay khoảng tin cậy tối đa.

- Nếu khoảng tin cậy có dạng $(\theta_1, +\infty)$ thì gọi là khoảng tin cậy phía phải hay khoảng tin cậy tối thiểu.
- Nếu khoảng tin cậy có dạng (θ_1, θ_2), với $\theta_1, \theta_2 \neq \pm\infty$ thì gọi là khoảng tin cậy hai phía hay khoảng tin cậy tối đa với độ dài khoảng tin cậy $l = \theta_2 - \theta_1$.

Một số bài toán ước lượng khoảng cho tham số của tổng thể có phân phối chuẩn

- Ước lượng trung bình μ của một tổng thể tuân theo phân phối chuẩn và đã biết phương sai σ^2 của tổng thể với độ tin cậy $(1-\alpha)$:

$$\mu \in \left(\overline{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

  Trong đó mẫu ngẫu nhiên từ tổng thể $\{x_1, x_2, \cdots, x_n\}$ bao gồm các quan sát độc lập và cùng tuân theo một phân phối xác suất chuẩn $N(\mu, \sigma^2)$.

- Ước lượng tỷ lệ p trên một tổng thể