

# Econometrics 1 *Applied Econometrics with R*

## Lecture 5: Review of Probability

---

黄嘉平

中国经济特区研究中心 讲师

办公室：文科楼1726

E-mail: [huangjp@szu.edu.cn](mailto:huangjp@szu.edu.cn)

Tel: (0755) 2695 0548

Office hour: Mon./Tue. 13:00-14:00

# Review of Probability

# Some basic definitions

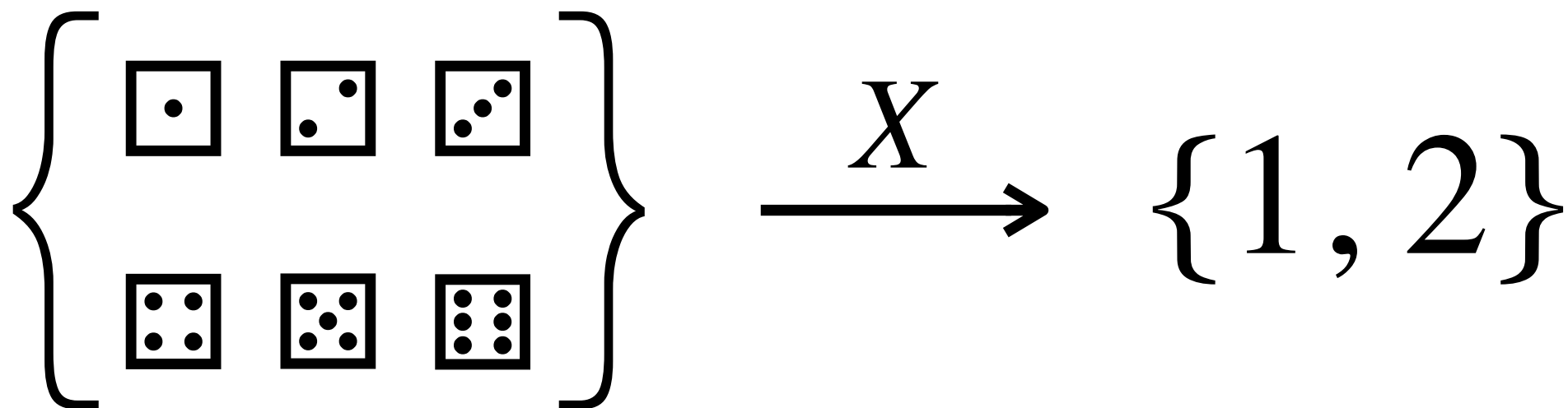
---

- *Randomness*: something you cannot control.
- *Outcome*: potential results of a random experiment or process.
- *Probability*: the frequency, or the proportion of time, that an outcome occurs in the long run.
- *Sample space*: the set of all possible outcomes.
- *Event*: a subset of the sample space.

# Random variable

---

- A random variable (r.v.) is a *mapping* from the sample space to a set of values.
- If the values of the r.v. is a discrete/continuous set, the r.v. is called a discrete/continuous random variable.



$$X(\square) = X(\square) = X(\square) = 1, \quad X(\square) = X(\square) = X(\square) = 2$$

# Probability and random variable

---

- Probabilities are defined for events (on sample space).

$$\Pr(\square) = 1/6, \quad \Pr(\{\square, \square, \square\}) = 1/2$$

- The probabilities of a random variable

$$\Pr(X = 1) = \Pr(\{\square, \square, \square\}) = 1/2$$

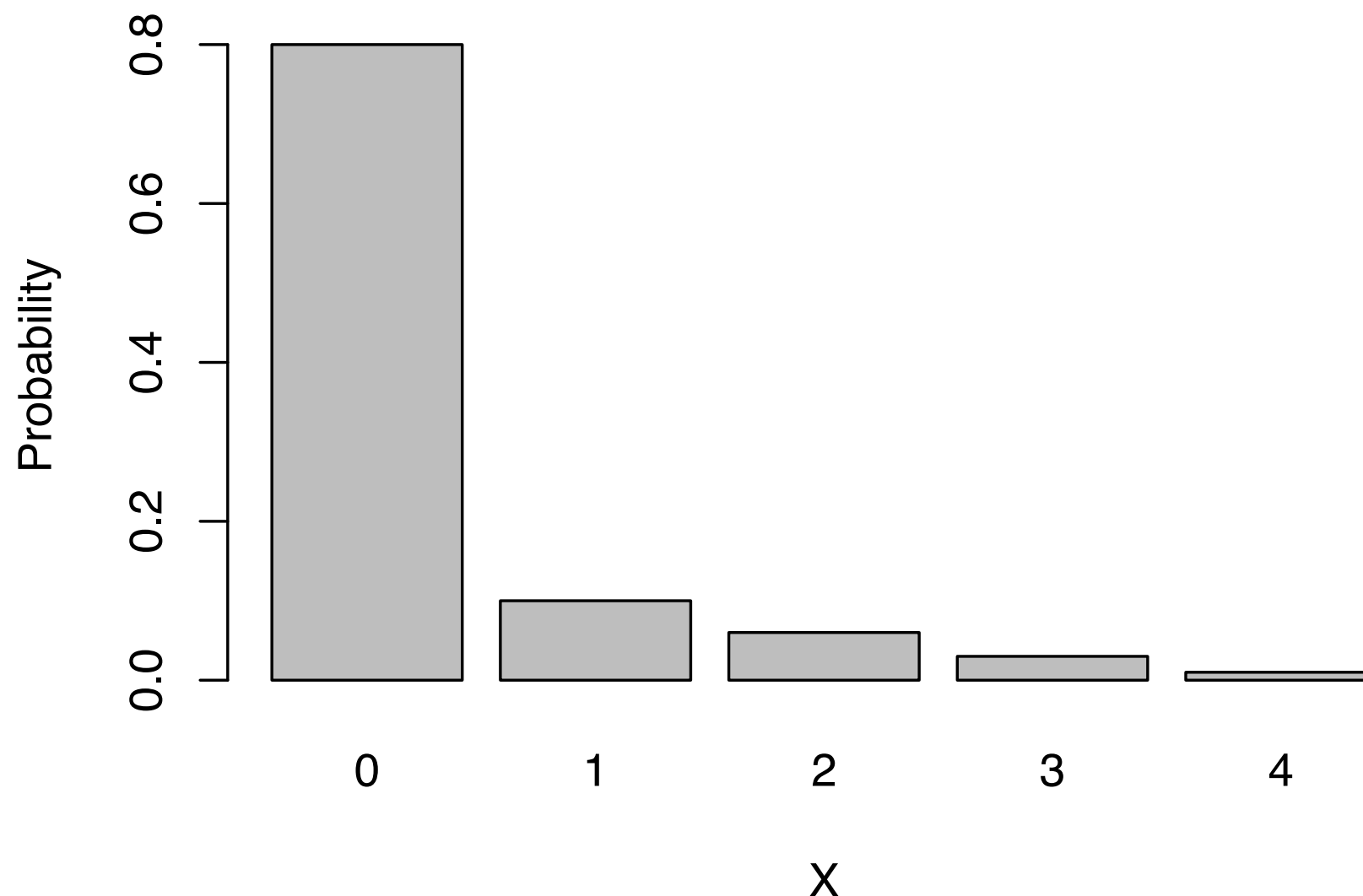
$$\Pr(X = 2) = \Pr(\{\square, \square, \square\}) = 1/2$$

- Usually we just write  $\Pr(X = 1) = 1/2$ .

# Probability distribution of a discrete r.v.

---

- The probability distribution of a discrete random variable is a list of all possible values of the variable and the probability that each value will occur.

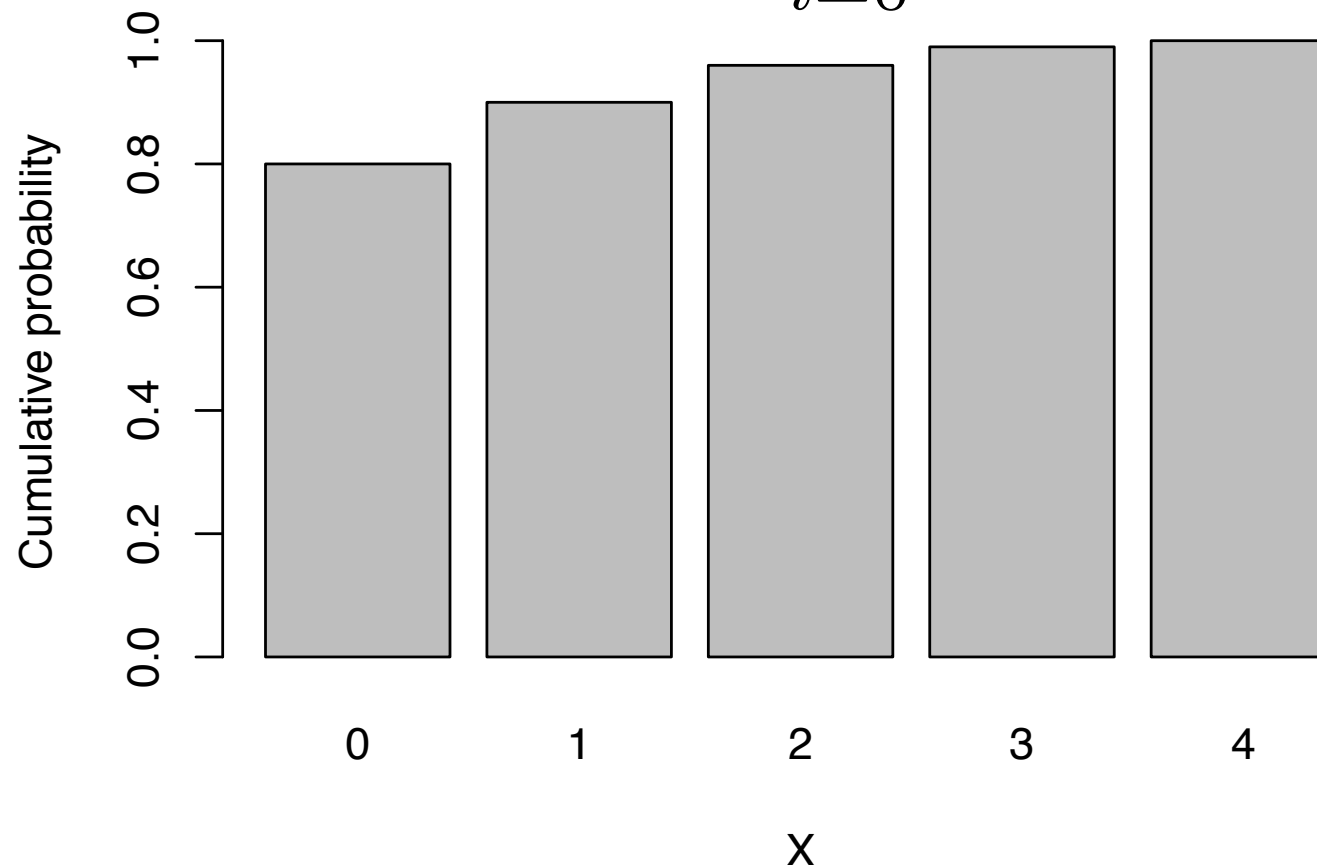


# Cumulative distribution function

---

- The cumulative distribution function (c.d.f) is a function describing the probability that the random variable is *less than or equal to* a particular value.

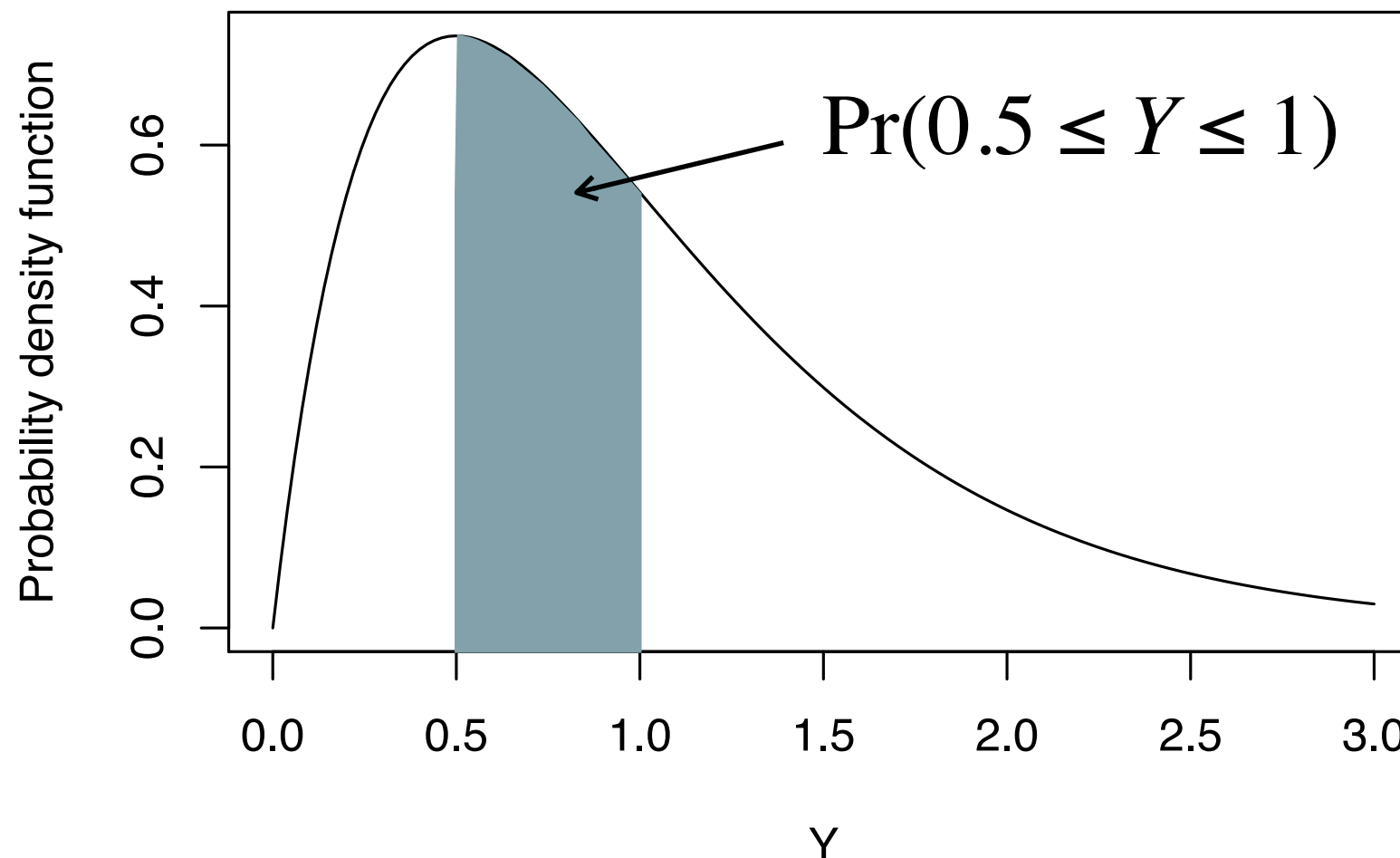
$$\Pr(X \leq n) = \sum_{i=0}^n \Pr(X = i)$$



# Probability distribution of a continuous r.v.

---

- The probability distribution of a continuous random variable is described by a *probability density function* (p.d.f.), where the area under it between any two points is the probability that the random variable falls between those two points.



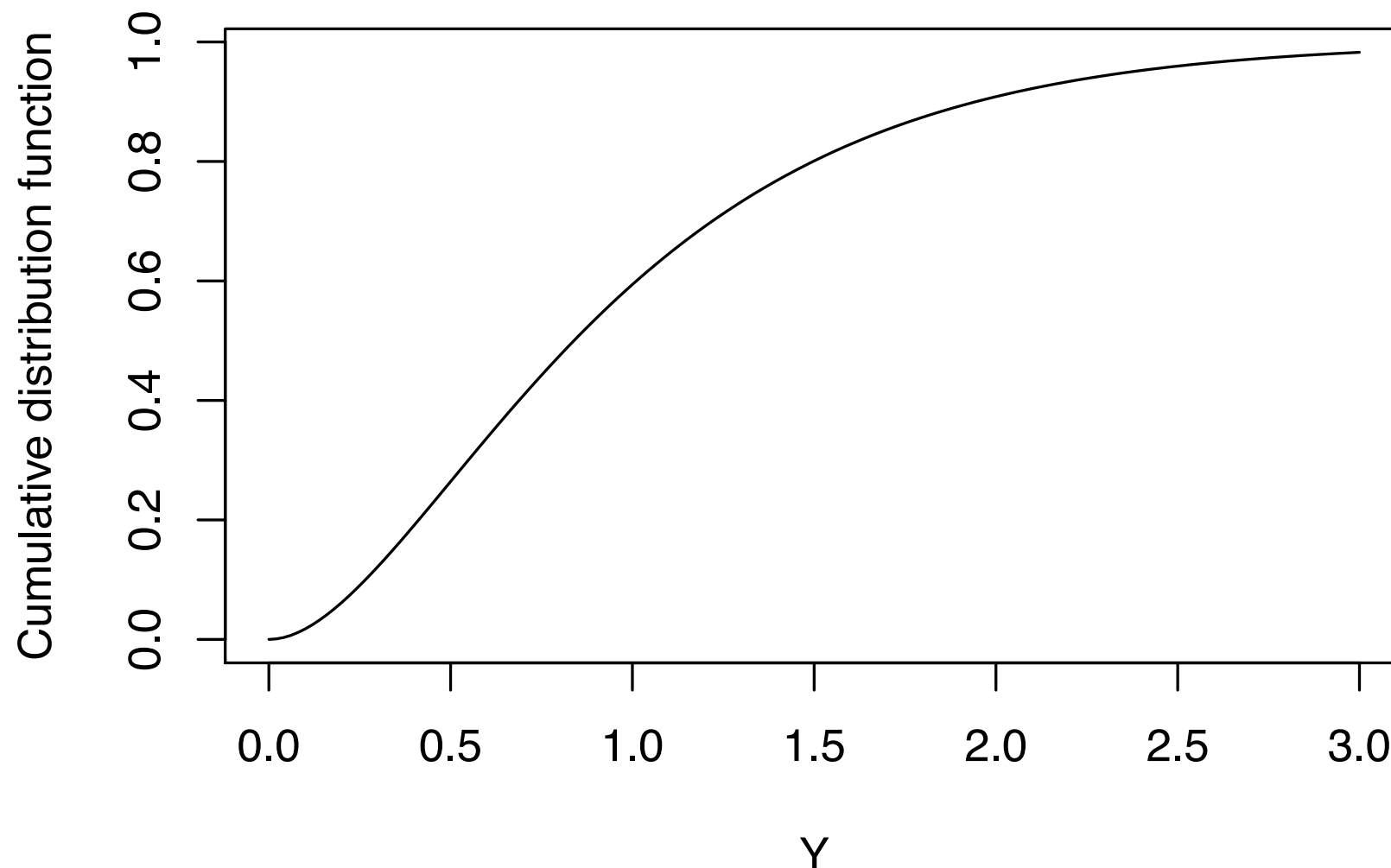


# The c.d.f. of a continuous random variable

---

- Given the p.d.f.  $f(y)$  of random variable  $Y$ , the c.d.f.  $F(y)$  of  $Y$  is defined by

$$F(y) := \Pr(Y \leq y) = \int_0^y f(u) du$$



- Expected value (or the mean)  $\mu_X$

$$\mathrm{E}(X) = \sum x_i \mathrm{Pr}(X = x_i)$$

$$\mathrm{E}(Y) = \int y f_Y(y) dy$$

- Variance  $\sigma_X^2$

$$\mathrm{var}(X) = \mathrm{E}[(X - \mu_X)^2] = \sum (x_i - \mu_X)^2 \mathrm{Pr}(X = x_i)$$

$$\mathrm{var}(Y) = \mathrm{E}[(Y - \mu_Y)^2] = \int (y - \mu_Y)^2 f_Y(y) dy$$

- Standard deviation

$$\sigma_X = \sqrt{\mathrm{var}(X)}$$

# Two discrete random variables

---

- Joint probability  $\Pr(X = x, Y = y)$

- Conditional probability

$$\Pr(Y = y \mid X = x) = \frac{\Pr(X = x, Y = y)}{\Pr(X = x)}$$

- $X$  and  $Y$  are independent if  $\Pr(Y = y \mid X = x) = \Pr(Y = y)$
- $X$  and  $Y$  are independent if and only if

$$\Pr(X = x, Y = y) = \Pr(X = x) \Pr(Y = y)$$

# Two discrete random variables

---

- Marginal distribution

$$\Pr(Y = y) = \sum_{i=1}^n \Pr(X = x_i, Y = y)$$

		X				Marginal probability of Y
		1	2	3	4	
Y	1	0.04	0.04	0.08	0.04	0.2
	2	0.01	0.03	0.2	0.06	0.3
	3	0.01	0.02	0.1	0.17	0.3
	4	0.04	0.01	0.12	0.03	0.2
Marginal probability of X		0.1	0.1	0.5	0.3	

# Two discrete random variables

---

- Covariance  $\sigma_{XY}$

$$\begin{aligned}\text{cov}(X, Y) &= \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \sum_i \sum_j (x_i - \mu_X)(y_j - \mu_Y) \Pr(X = x_i, Y = y_j)\end{aligned}$$

- Correlation

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

# Correlation and dependence

---

- $X$  and  $Y$  are said to be uncorrelated if  $\text{corr}(X, Y) = 0$ .
- $X$  and  $Y$  are independent  $\Rightarrow X$  and  $Y$  are uncorrelated

$$\begin{aligned}\text{cov}(X, Y) &= \sum_i \sum_j (x_i - \mu_X)(y_j - \mu_Y) \Pr(X = x_i, Y = y_j) \\ &= \sum_i \sum_j (x_i - \mu_X)(y_j - \mu_Y) \Pr(X = x_i) \Pr(Y = y_j) \\ &= \sum_i (x_i - \mu_X) \Pr(X = x_i) \sum_j (y_j - \mu_Y) \Pr(Y = y_j) \\ &= (\mathbb{E}(X) - \mu_X)(\mathbb{E}(Y) - \mu_Y) = 0\end{aligned}$$

# Correlation and dependence

---

- $X$  and  $Y$  are said to be uncorrelated if  $\text{corr}(X, Y) = 0$ .
- $X$  and  $Y$  are uncorrelated  $\nRightarrow$   $X$  and  $Y$  are independent
- Let  $X$  and  $Z$  be independent random variables such that

$X$			
Value	-1	0	1
Prob	0	1/2	1/2

$Z$			
Value	-1	0	1
Prob	1/2	0	1/2

and let  $Y = XZ$ . Verify that  $X$  and  $Y$  are uncorrelated and dependent.

# Important Distributions



# The normal distribution

---

- The p.d.f. of a normal distribution with mean  $\mu$  and variance  $\sigma^2$ , i.e.  $N(\mu, \sigma^2)$

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

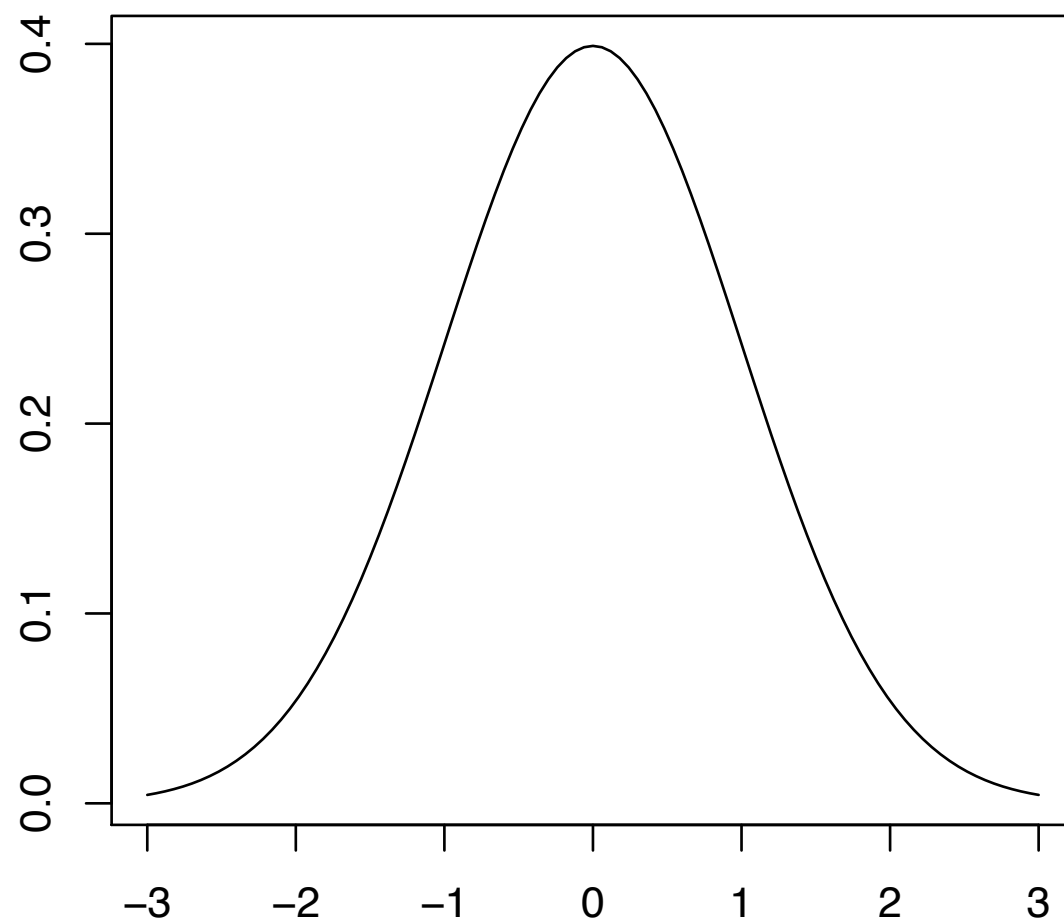
- The standard normal distribution is  $N(0, 1)$ . A standard normal random variable is usually denoted as  $Z$ , whose c.d.f is denoted by

$$\Pr(Z \leq z) = \Phi(z)$$

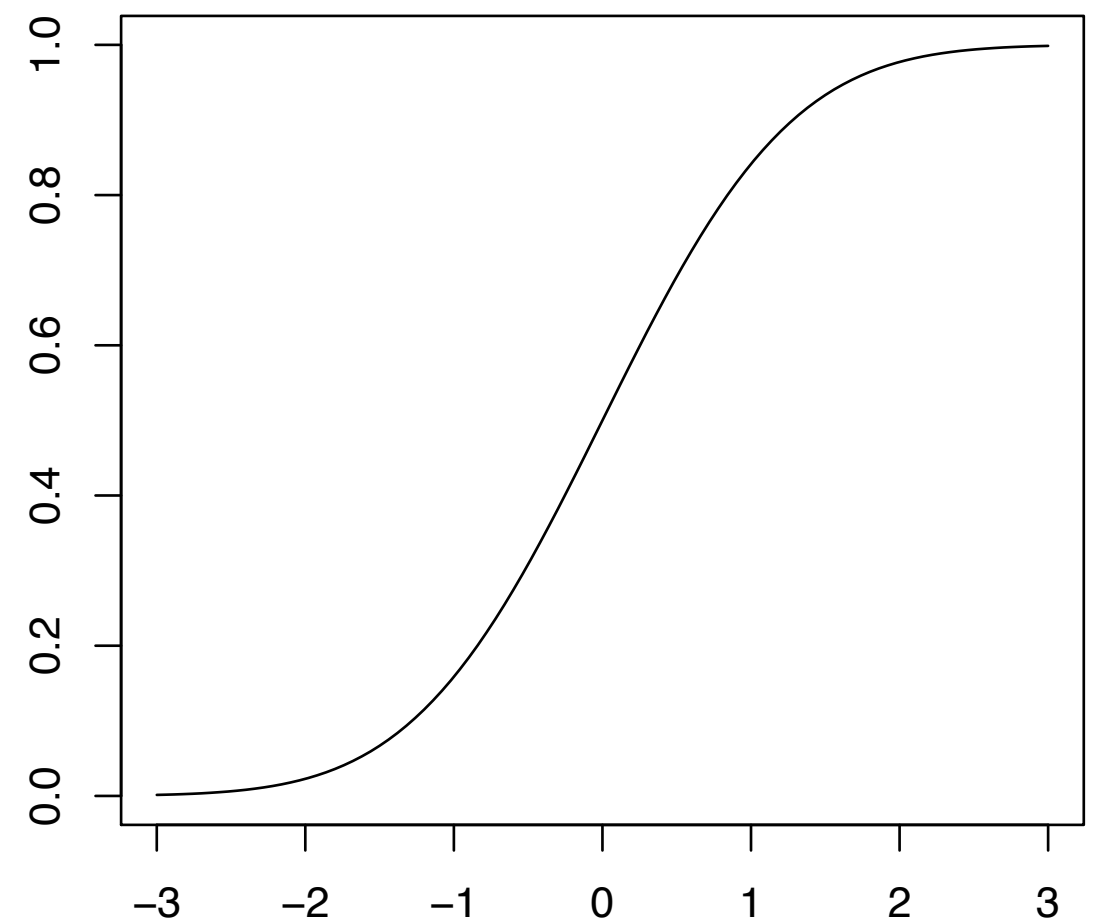
# The normal distribution

---

**The p.d.f. of the standard normal distribution**



**The c.d.f. of the standard normal distribution**



# The normal distribution and normal r.v.

---

- Some special probabilities for  $X \sim N(\mu, \sigma^2)$

$$\Pr(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.683$$

$$\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.954$$

$$\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997$$

$$\Pr(\mu - 1.96\sigma \leq X \leq \mu + 1.96\sigma) \approx 0.95$$

- Standardization of normal random variables

$$Z = (X - \mu) / \sigma$$

# The chi-squared distribution

---

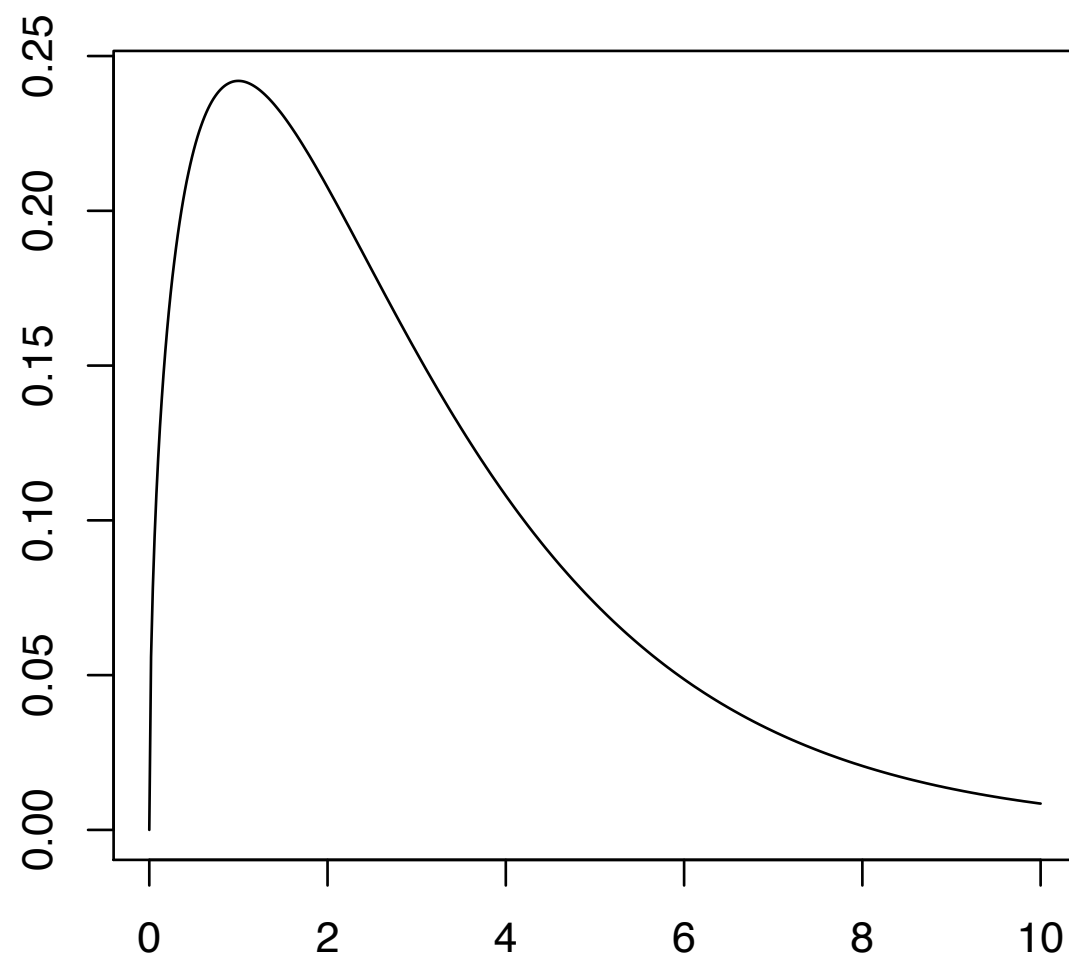
- The chi-squared distribution with  $m$  degrees of freedom, denoted by  $\chi_m^2$ , is the distribution of a sum of the squares of  $m$  independent standard normal random variables.
- Let  $Z_1, Z_2, Z_3$  be independent standard normal random variables. The c.d.f. of the chi-squared distribution with degree of freedom 3 is then

$$F_{\chi_3^2}(z) = \Pr(Z_1^2 + Z_2^2 + Z_3^2 \leq z)$$

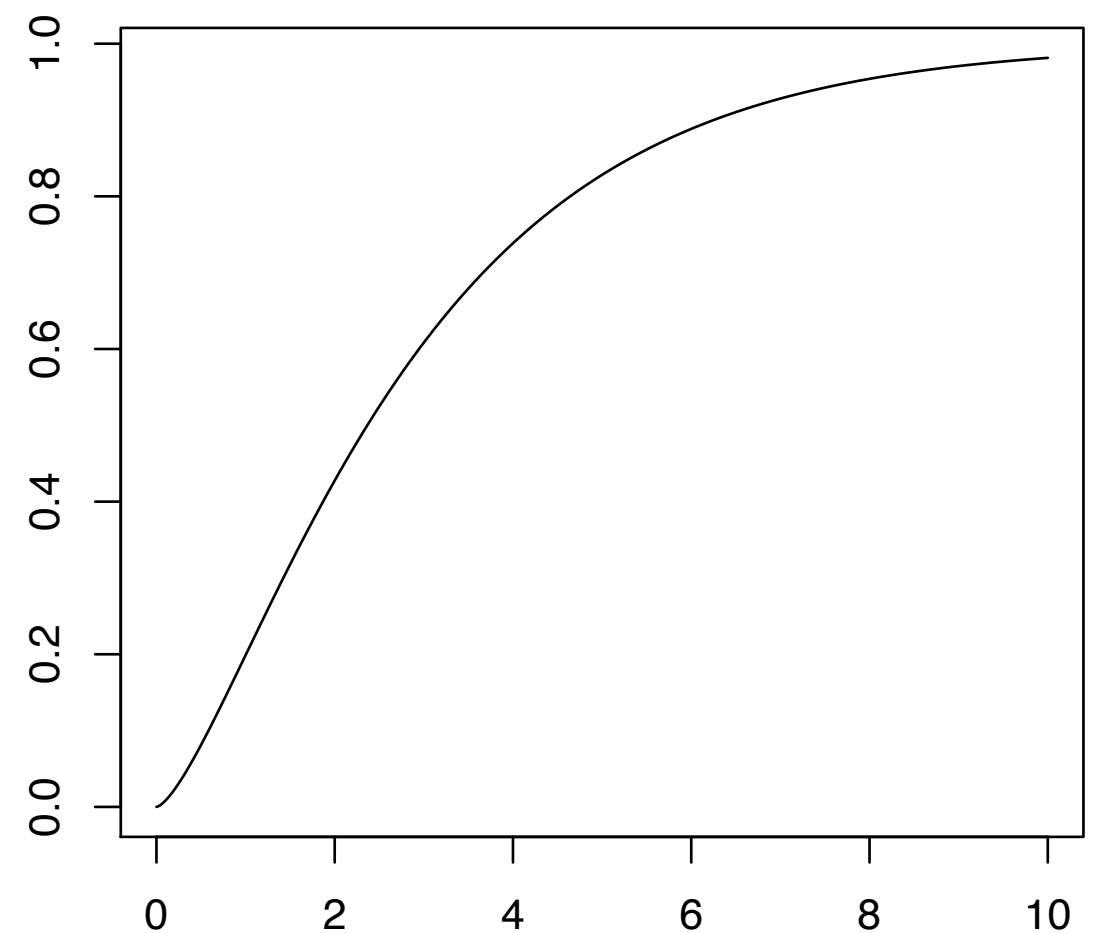
# The chi-squared distribution

---

**The p.d.f. of Chi-squared distribution with d.f. = 3**



**The c.d.f. of Chi-squared distribution with d.f. = 3**



# The Student $t$ distribution

---

- The Student  $t$  distribution with  $m$  degrees of freedom, denoted by  $t_m$ , is defined to be the distribution of the ratio of a standard normal r.v., divided by the squared root of an independently distributed chi-squared r.v. with  $m$  degrees of freedom divided by  $m$ .
- Let  $Z$  be a standard normal r.v. and  $W$  be a r.v. with a chi-squared distribution with d.f. =  $m$ , then the r.v.

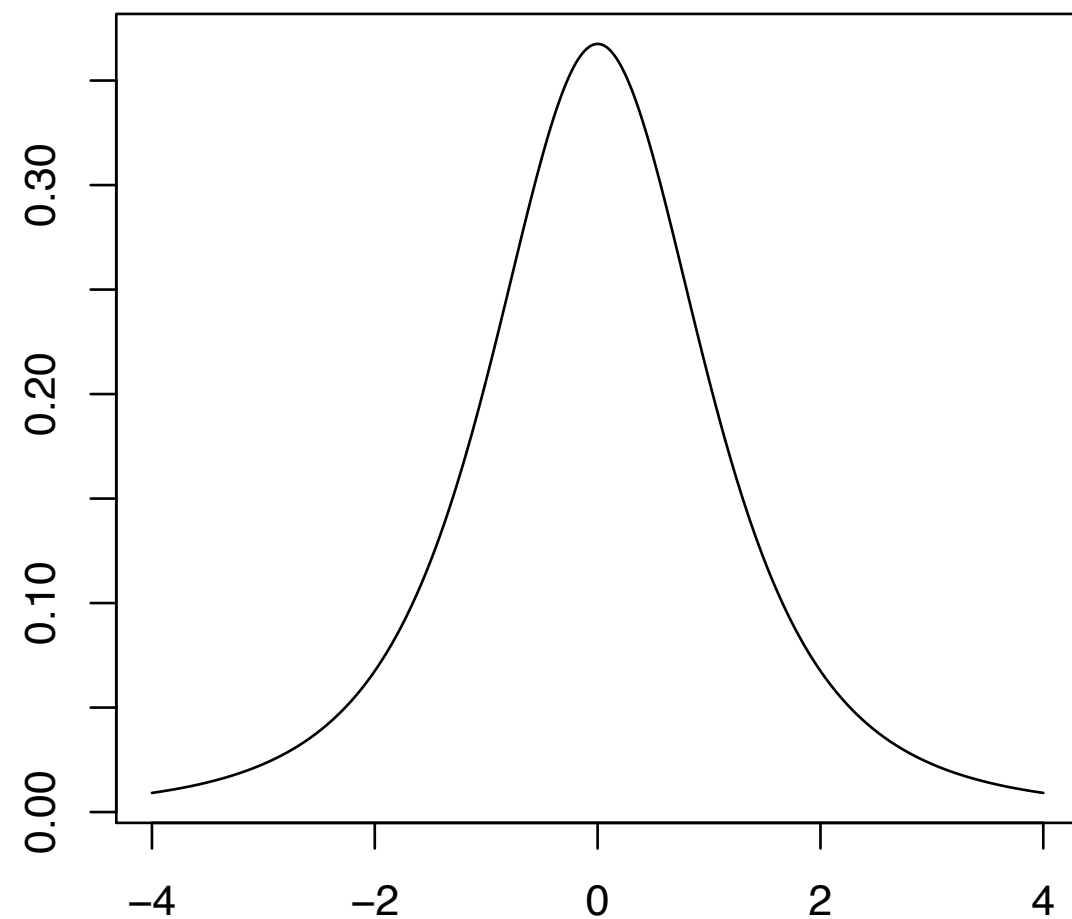
$$Z / \sqrt{W/m}$$

has a Student  $t$  distribution with d.f. =  $m$ .

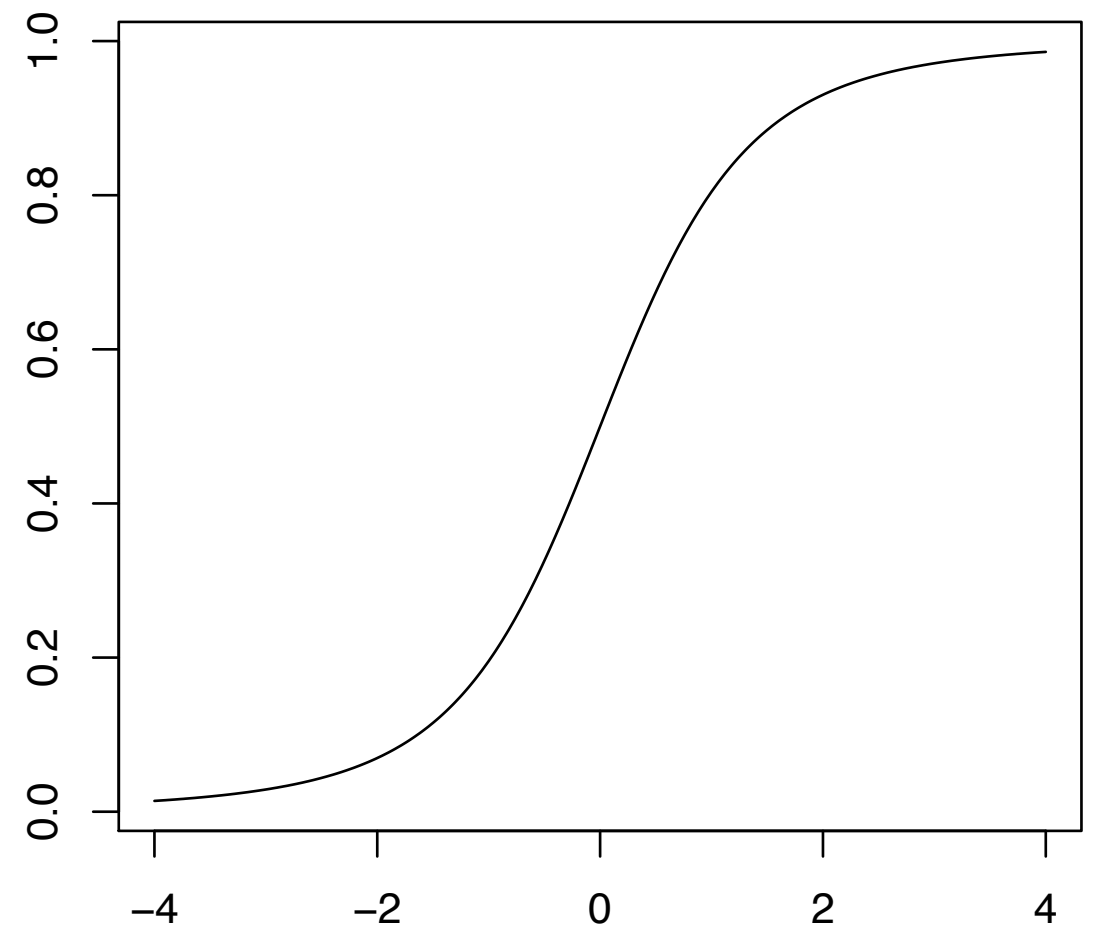
# The Student $t$ distribution

---

**The p.d.f. of  $t$  distribution with d.f. = 3**



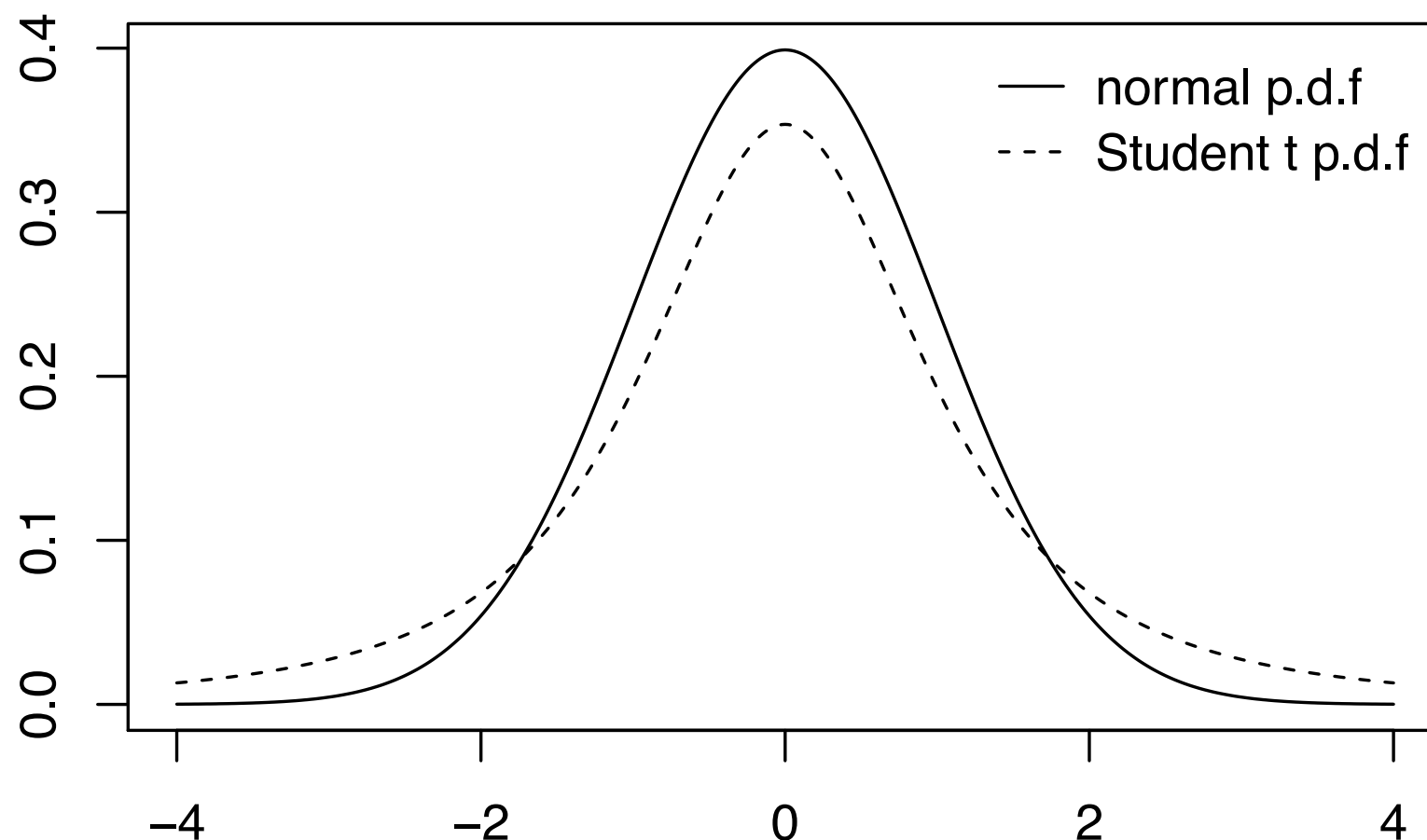
**The c.d.f. of  $t$  distribution with d.f. = 3**



# The $t$ distribution v.s. normal distribution

---

- When the degree of freedom of a  $t$  distribution is small ( $m < 20$ ), the  $t$  distribution has a “fatter tail” than a standard normal distribution.





# The $F$ distribution

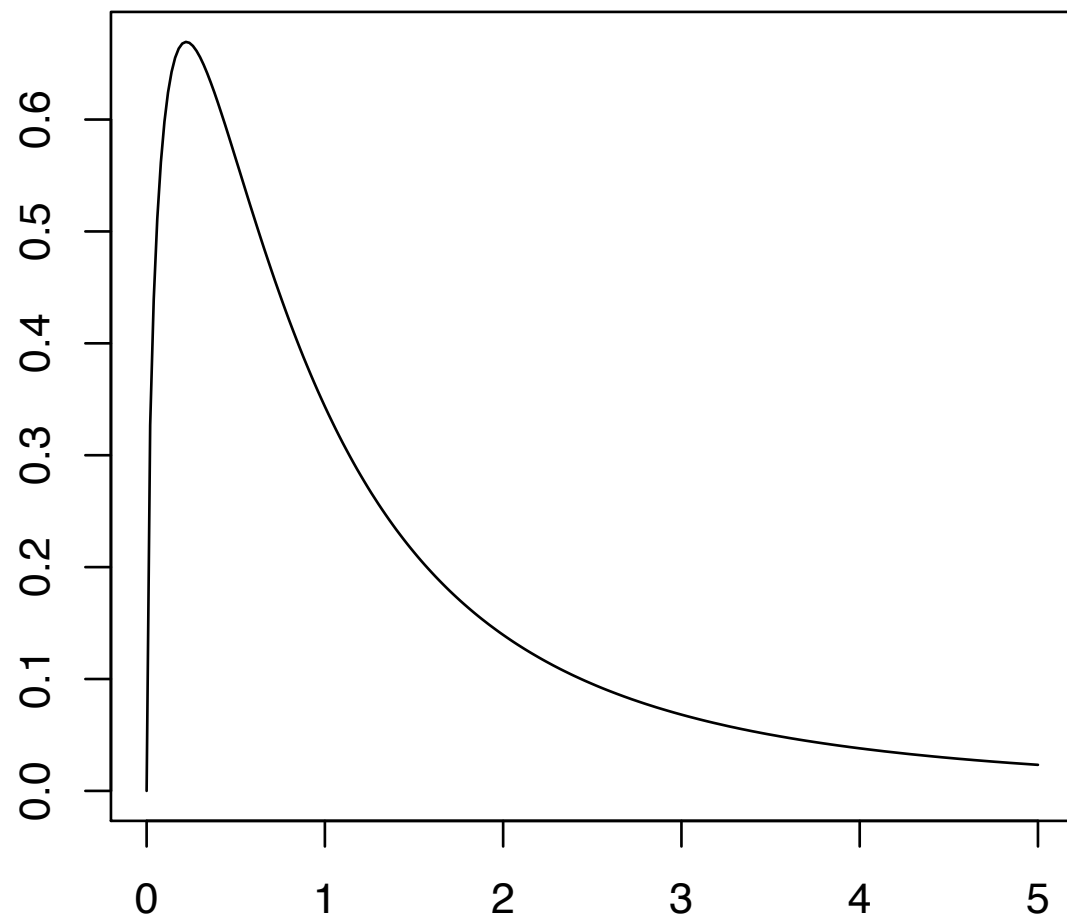
---

- The  $F$  distribution with  $m$  and  $n$  degrees of freedom, denoted by  $F_{m,n}$ , is defined to be the distribution of the ratio of a chi-squared r.v. with d.f. =  $m$  divided by  $m$ , to an independently distributed chi-squared r.v. with d.f. =  $n$  divided by  $n$ .
- Let  $W \sim \chi_m^2$  and  $V \sim \chi_n^2$ , then the random variable
$$\frac{W/m}{V/n}$$
has a distribution of  $F_{m,n}$ .

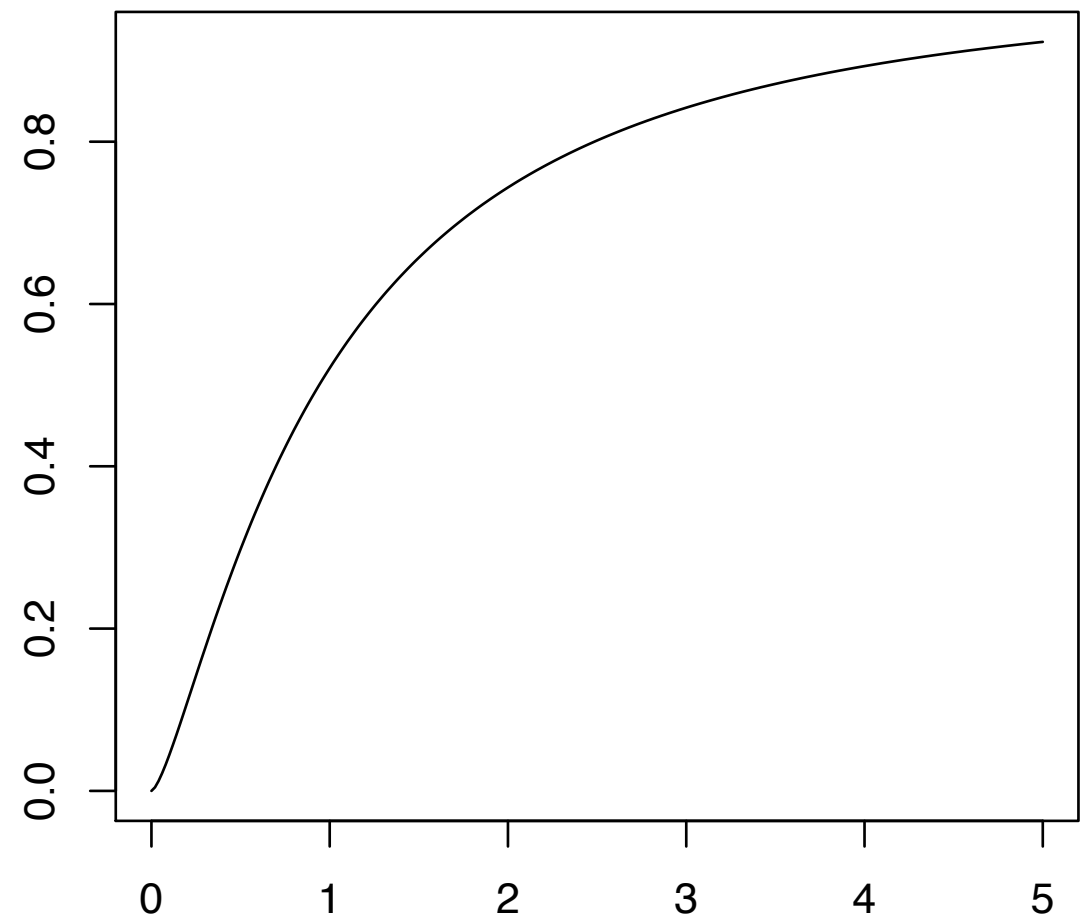
# The $F$ distribution

---

**The p.d.f. of  $F$  distribution with d.f. = (3, 4)**



**The c.d.f. of  $F$  distribution with d.f. = (3, 4)**



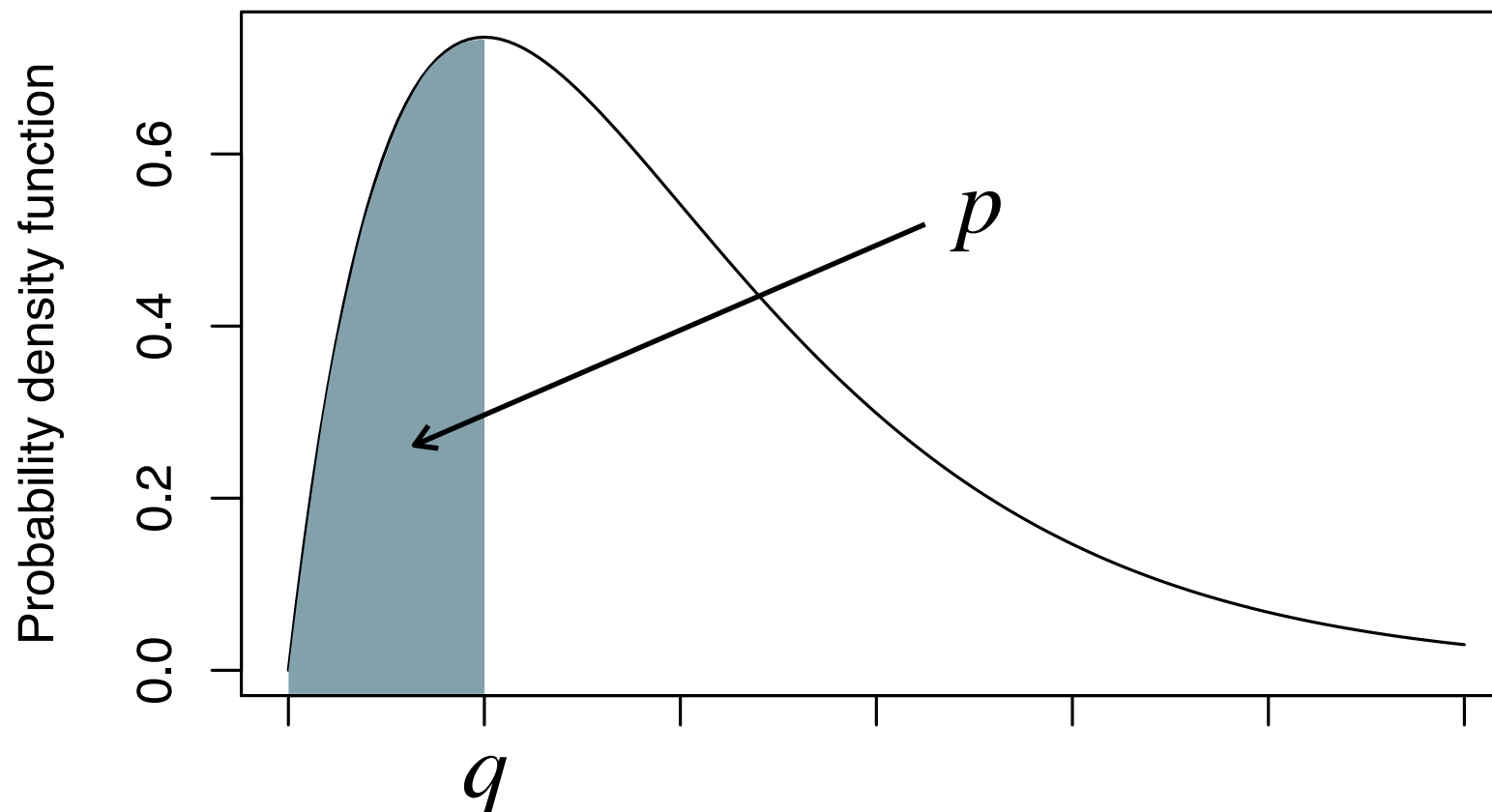
# Quantile function

---

- The quantile function is the inverse of a c.d.f.

$$p = \Pr(X \leq q) = F_X(q) \quad \Rightarrow \quad q = F_X^{-1}(p)$$

- $q$  is the point at which the cumulative probability of  $X$  is  $p$ .



# R commands for probability distributions

---

- General form: `d***`, `p***`, `q***`, `r***`
- `d***` — probability density function  
`p***` — cumulative distribution function  
`q***` — quantile function  
`r***` — generate a random number
- Specific distributions:
  - normal: `*** = norm`
  - chi-squared: `*** = chisq`
  - Student *t*: `*** = t`
  - F*: `*** = f`

# Large Random Samples

# Random sampling

---

- Simple random sampling:

Randomly choose  $n$  objects (the sample) from the population, where each member of the population is equally likely to be included in the sample.

- i.i.d. (independent and identically distributed) random variables:

independent — the outcome of  $X$  does not depend on the outcome of  $Y$

identical — the distribution of  $X$  and  $Y$  are the same

# Random sampling

---

- Let  $Y_1, Y_2, \dots, Y_n$  be a random sample, therefore they are i.i.d. random variables (before drawn).

- The sample average (or sample mean)

$$\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + \dots + Y_n) = \frac{1}{n} \sum_{i=1}^n Y_i$$

is also a random variable.

- The distribution of  $\bar{Y}$  is called the *sampling distribution* of  $\bar{Y}$ .

# Sampling distribution

---

- Let  $\mu_Y$  be the mean, and  $\sigma_Y^2$  be the variance of  $Y_i$ .
- Mean and variance of  $\bar{Y}$

$$E(\bar{Y}) = \mu_Y, \quad \text{var}(\bar{Y}) = \frac{\sigma_Y^2}{n}$$

- Generally the sampling distribution of  $\bar{Y}$  is complicated, but when the population distribution is normal, the sampling distribution is also normal.

$$Y_i \sim N(\mu_Y, \sigma_Y^2) \quad \Rightarrow \quad \bar{Y} \sim N(\mu_Y, \sigma_Y^2/n)$$



# Large sample approximations of sampling distributions

---

- When the size  $n$  of the sample is small, the exact distribution of  $\bar{Y}$  can be very complicated.
- When  $n$  is large (theoretically  $n \rightarrow \infty$ , in practice  $n > 30$ ), we can use the following tools to approximate sampling distribution:

The law of large numbers:  $\bar{Y} \xrightarrow{p} \mu_Y$

The central limit theorem:  $\bar{Y} \xrightarrow{d} N(\mu_Y, \sigma_Y^2/n)$

# The law of large numbers (LLN)

---

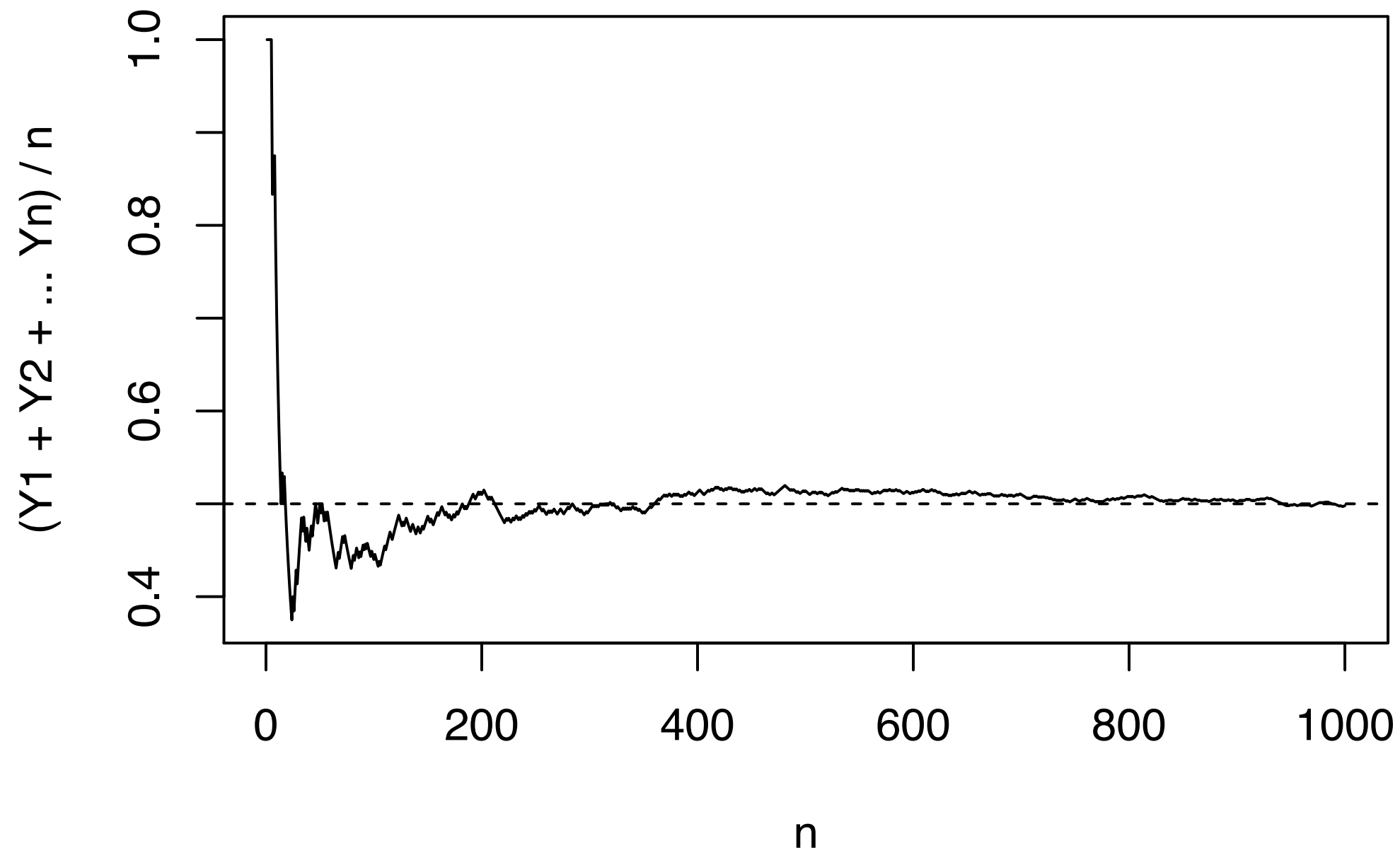
If the random sample are i.i.d., and the population variance is finite ( $\sigma_Y^2 < \infty$ ), then the sample mean  $\bar{Y}$  converges to the population mean  $\mu_Y$  in probability as the sample size increases ( $n \rightarrow \infty$ ).

- Converges in probability:  
The probability that  $\mu_Y - c < \bar{Y} < \mu_Y + c$  becomes arbitrarily close to 1 as  $n$  increases for any constant  $c > 0$ .
- There are other versions of LLN.

# Demonstrating the LLN

---

- The sample mean of  $n$  Bernoulli random variables (flipping a fair coin  $n$  times).



# The central limit theorem (CLT)

---

If the random sample are i.i.d., and the population variance is finite (  $\sigma_Y^2 < \infty$  ), then the distribution of  $\bar{Y}$  becomes arbitrarily well approximated by the normal distribution  $N(\mu_Y, \sigma_Y^2/n)$  as the sample size increases ( $n \rightarrow \infty$ ).

- The central limit theorem does not require the population distribution to be normal.
- Verify the central limit theorem with R using an arbitrary population distribution.

# Idea for programming

---

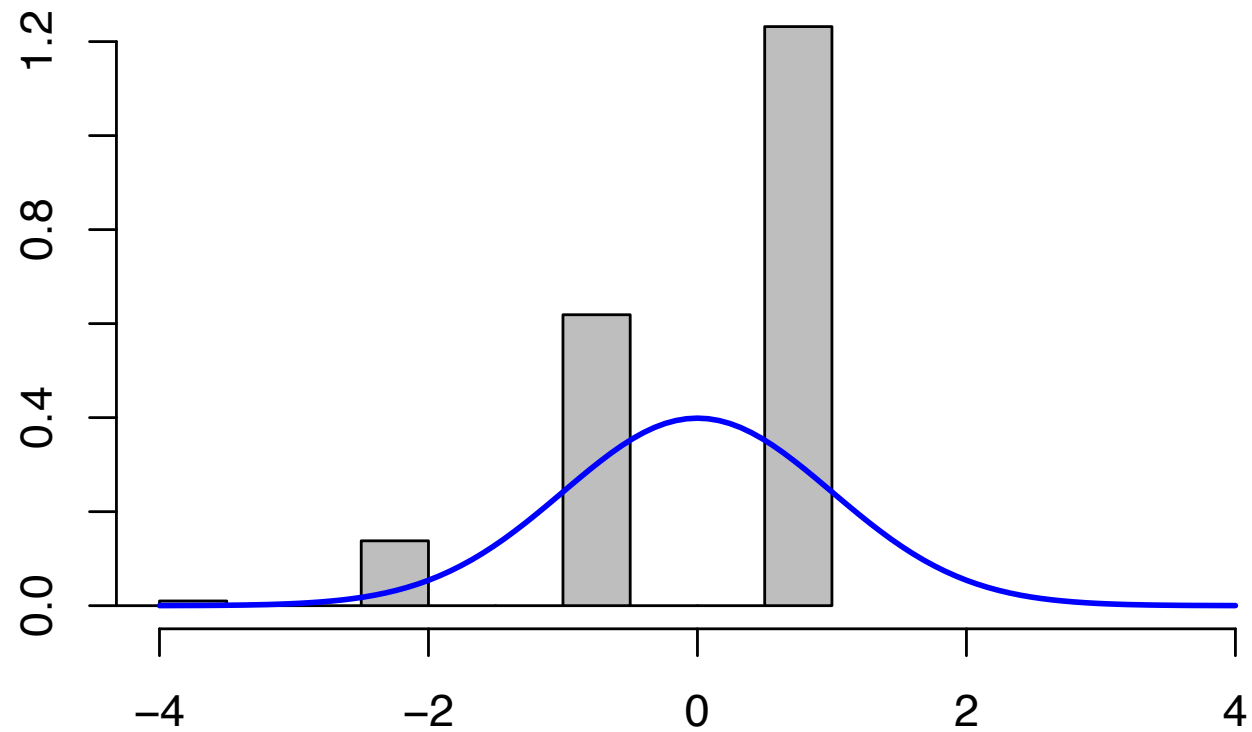
- The size  $n$  need to increase
- For each  $n$ :
  - We need a sufficient number of observations of sample mean
  - Draw the histogram of the sample means
  - Plot the density of the corresponding normal distribution

```
n <- c(2, 5, 10, 20, 50, 100, 200, 500, 1000) # sample size
k <- 2000    # number of samples for each size

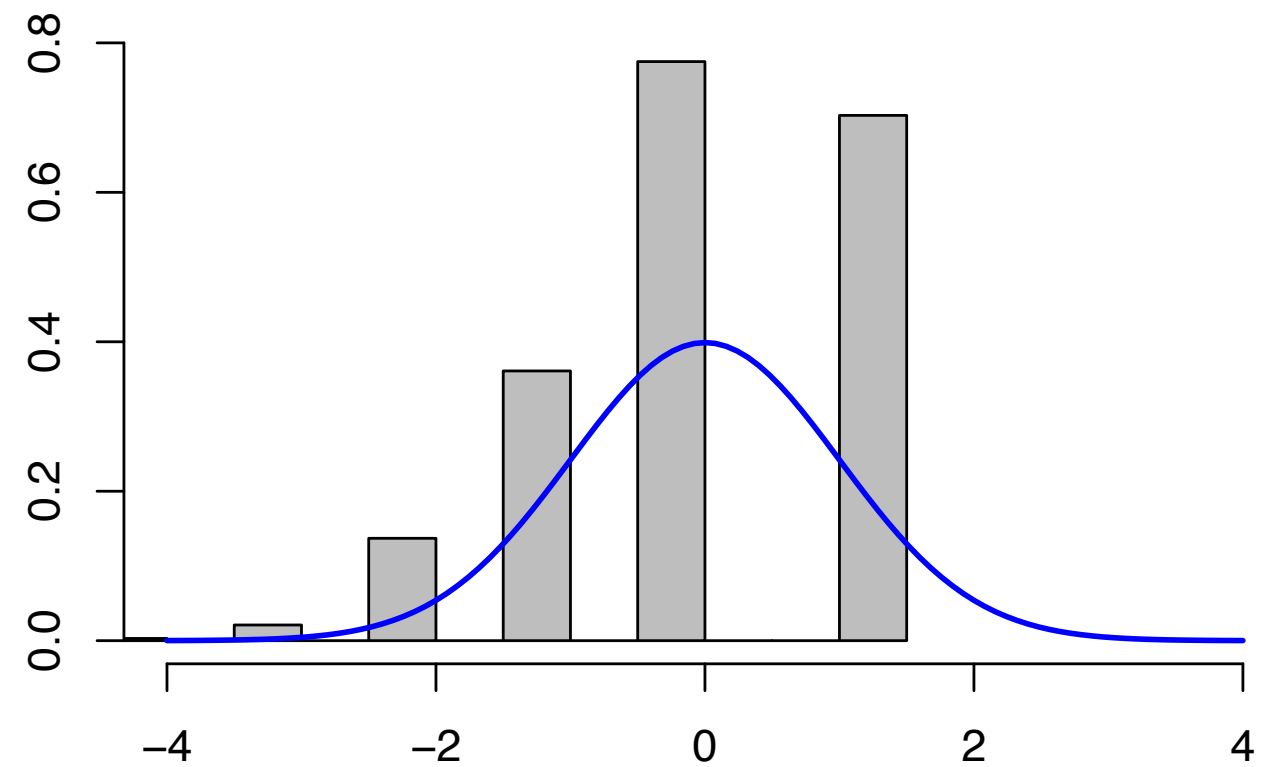
p <- 0.9    # probability of success for Bernoulli distribution
pmean <- p   # population mean
pvar <- p * (1-p) # population variance

for (i in 1:length(n)) {    # for each sample size
  smean <- rep(0, k)    # sample mean distribution (initialization)
  for (j in 1:k) {      # for each sample
    smean[j] <- mean(rbinom(n[i], 1, p))
    # calculate sample means
  }
  sd_smean <- (smean - pmean) / sqrt(pvar / n[i])
  # standardization
  hist(sd_smean, freq = FALSE, col = "grey",
        xlim = c(-4, 4))    # plot histogram
  curve(dnorm, add = TRUE, lwd = 2, col = "blue")
  # add standard normal density function
}
```

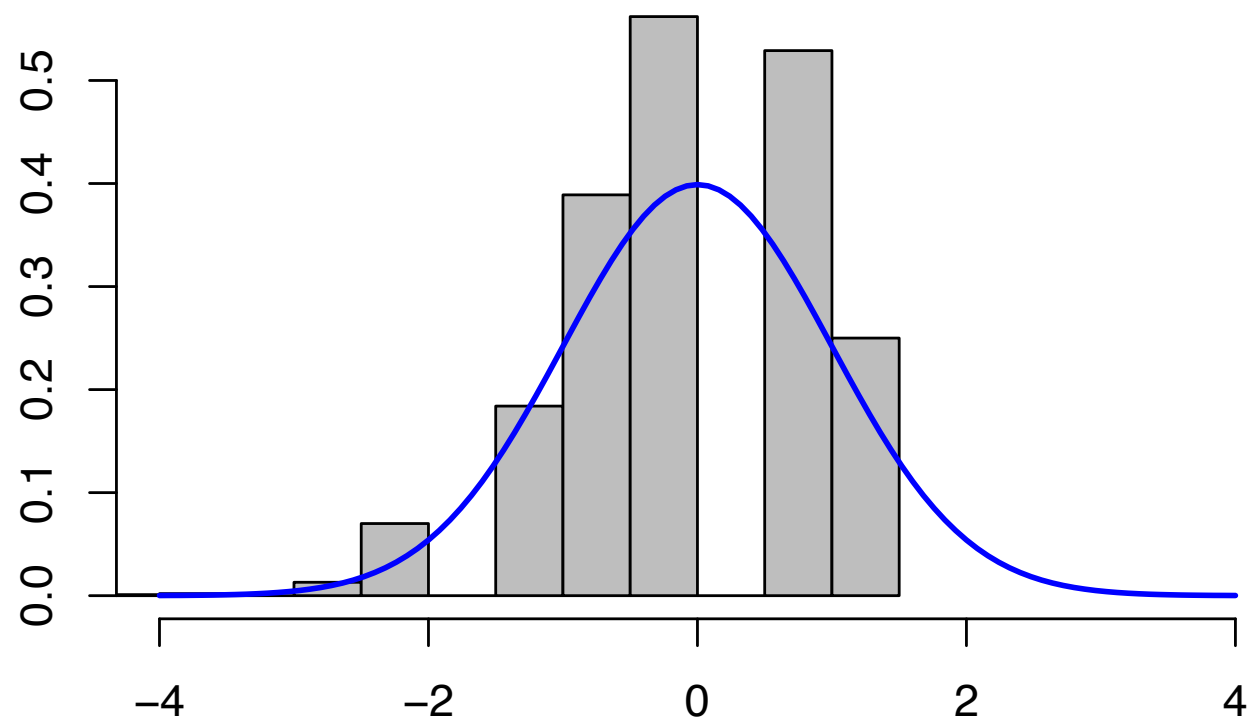
**Sample mean of Bernoulli distribution  
with  $p = 0.9$  and  $n = 5$**



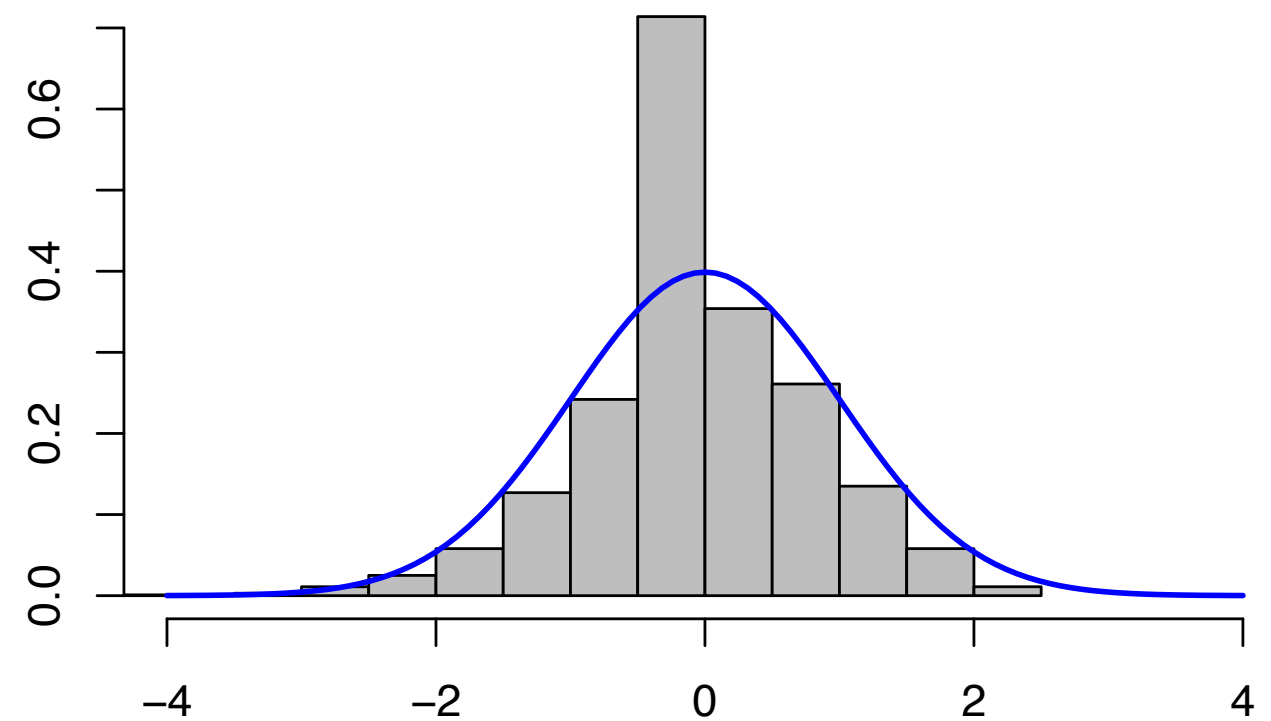
**Sample mean of Bernoulli distribution  
with  $p = 0.9$  and  $n = 10$**



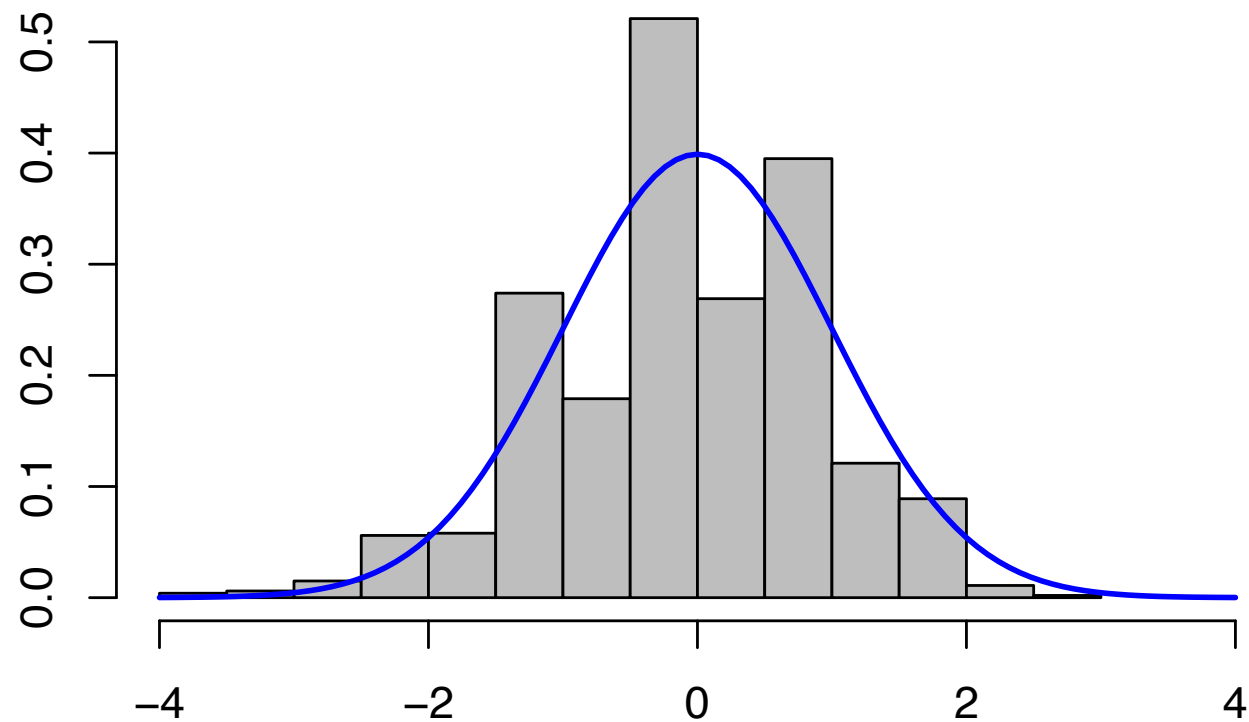
**Sample mean of Bernoulli distribution  
with  $p = 0.9$  and  $n = 20$**



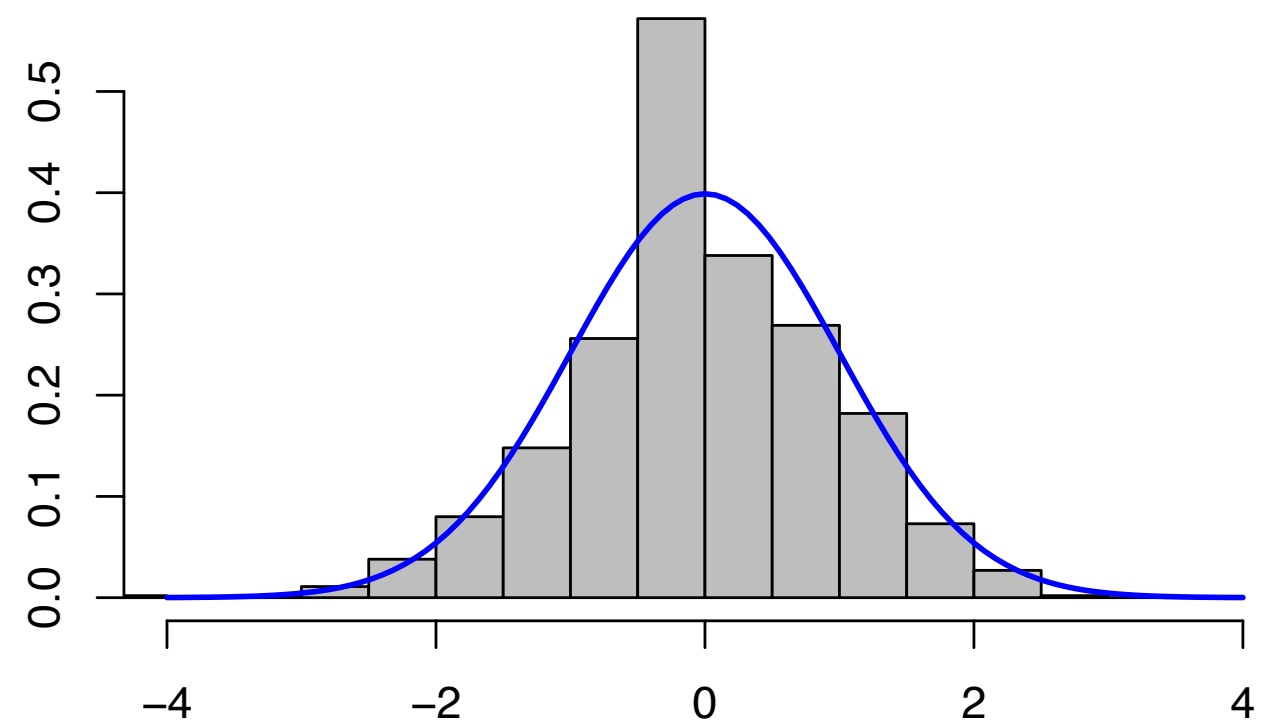
**Sample mean of Bernoulli distribution  
with  $p = 0.9$  and  $n = 50$**



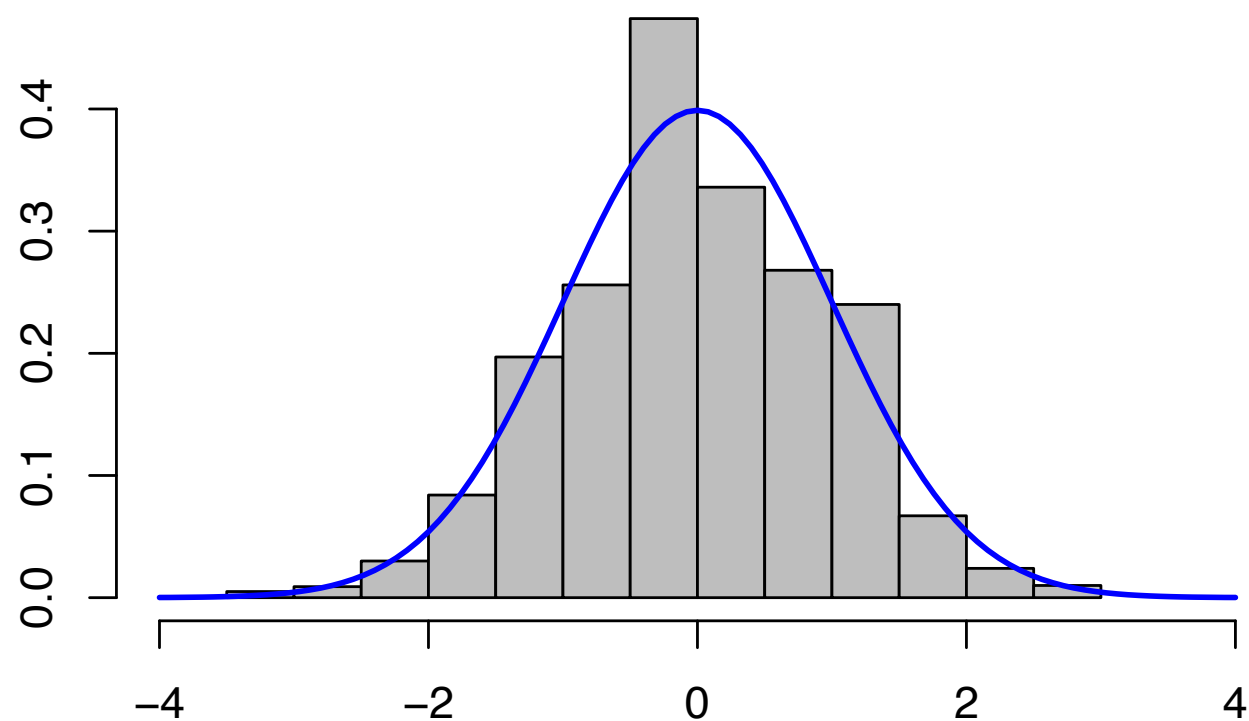
**Sample mean of Bernoulli distribution  
with  $p = 0.9$  and  $n = 100$**



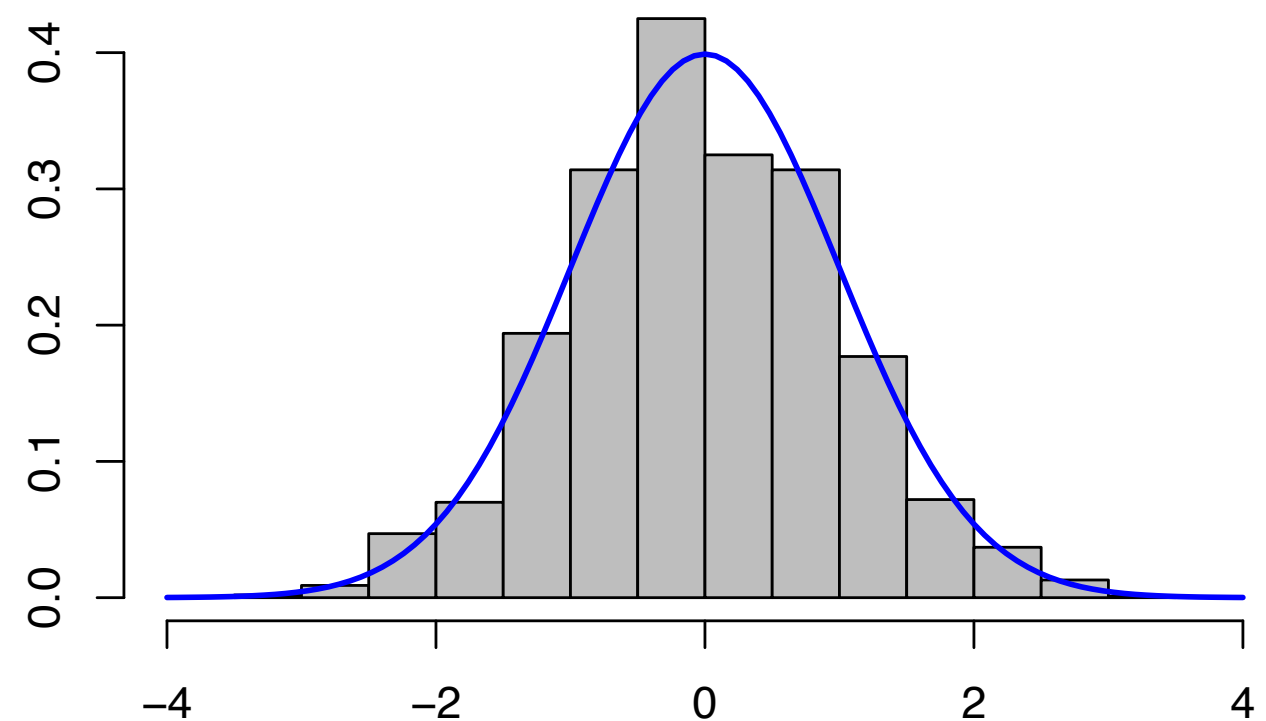
**Sample mean of Bernoulli distribution  
with  $p = 0.9$  and  $n = 200$**



**Sample mean of Bernoulli distribution  
with  $p = 0.9$  and  $n = 500$**



**Sample mean of Bernoulli distribution  
with  $p = 0.9$  and  $n = 1000$**





# Take home practice

---

- $X$  and  $Y$  are two random variables, find the assumptions on  $X$  and  $Y$  that make each of the following equation valid
  1.  $E(X + Y) = E(X) + E(Y)$
  2.  $E(XY) = E(X) \cdot E(Y)$
  3.  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
- Write a program to verify the Law of Large Numbers using i.i.d. chi-squared random variables.

# References

---

1. Stock, J. H. and Watson, M. M., *Introduction to Econometrics*, 3rd Edition, Pearson, 2012.
2. DeGroot, M. H. and Schervish, M. J., *Probability and Statistics*, 4th Edition, Pearson, 2012.