计量经济学

第二讲：概率论复习

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室　粤海校区汇文楼2613
E-mail　huangjp@szu.edu.cn
Website　https://huangjp.com

主要内容

- 随机变量和其他基础概念
  - 随机性与概率
  - 随机变量、分布、期望值、方差
  - 两个随机变量间的关系：**条件概率**、独立性、相关性
- 重要的概率分布
  - 离散型概率分布
  - 连续型概率分布
- 随机抽样与大样本
  - **大数定律**
  - **中心极限定理**

随机变量

基本概念

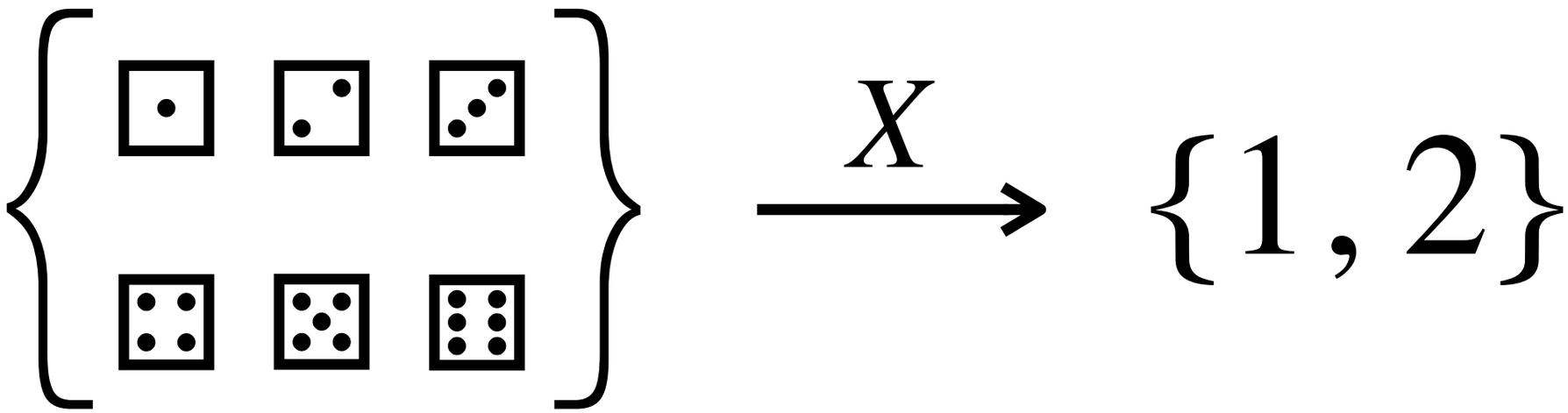
- 随机性（randomness）：偶然的、无法控制的、无法预测的
- 结果（outcome）：随机尝试或过程可能产生的互斥的后果
  - 扔硬币 → 正面或反面（互斥的）
  - 一周内晴天的总数 → 0, 1, 2, 3, 4, 5, 6, 7
- 概率（probability）：长期观测的结果发生的次数频率
- 样本空间（sample space）：所有可能结果的集合
  - { 正面，反面 }，{ 0, 1, 2, 3, 4, 5, 6, 7}
- 事件（event）：样本空间的子集，即一个或多个结果的集合
  - 一周内晴天数超过三天 → { 3, 4, 5, 6, 7 }

随机变量

Random variable

- **随机变量（random variable, r.v.）**是给一个随机尝试的所有结果各自对应一个实数值。换句话说，随机变量是在样本空间上定义的实数值函数。

- 离散型/连续型随机变量：随机变量取离散/连续值



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

概率与随机变量

- 概率是针对样本空间所含事件定义的

$$\Pr(\boxdot) = 1/6, \qquad \Pr(\{\boxdot, \boxdot, \boxdot\}) = 1/2$$

- 从随机变量的定义，我们可以得到随机变量每个值的概率，即

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

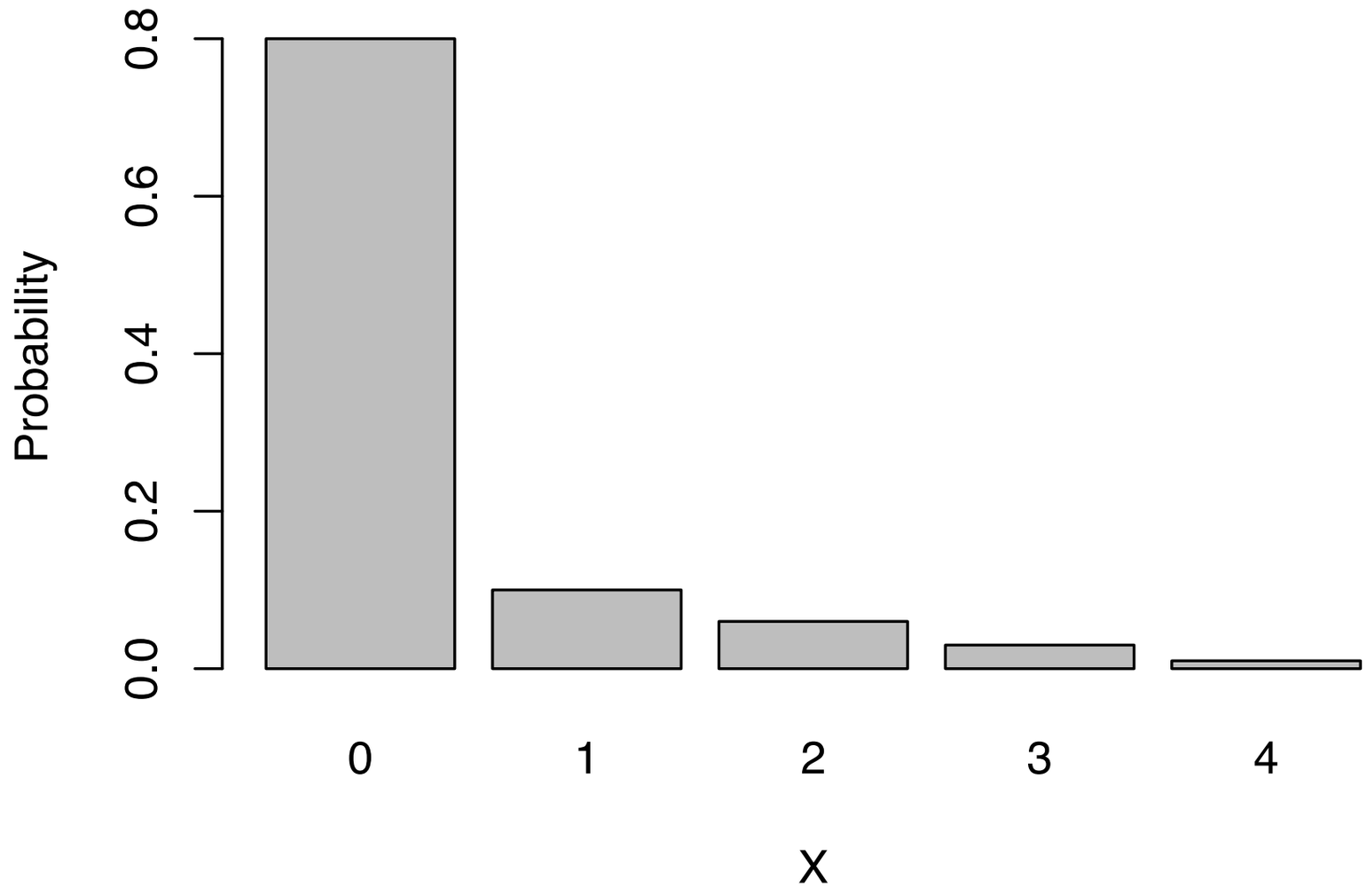
- 我们可以省略中间步骤，直接写成 $\Pr(X = 1) = 1/2$。

- 我们习惯性的用大写字母（如 X, Y）表示随机变量，用对应的小写字母（如 x, y）表示随机变量的取值。

离散型随机变量的概率分布

Probability distribution of a discrete r.v.

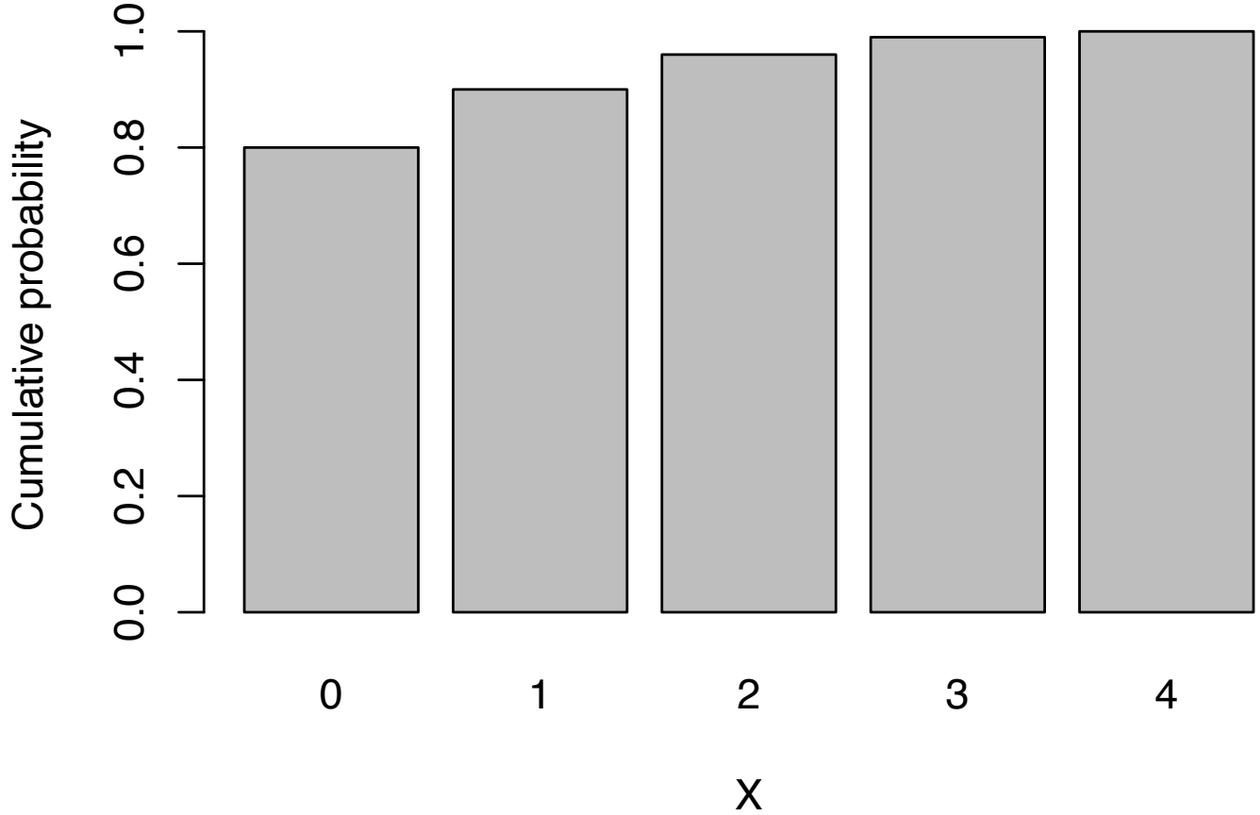
- 离散型随机变量的**概率分布**，是该随机变量所有可能取值及每个取值发生概率的列表（或函数）。



累积分布函数

Cumulative distribution function, c.d.f.

- **累积概率（cumulative probability）**指随机变量小于或等于某个特定值的概率。**累积概率分布**是所有可能的特定值及其累积概率的列表（或函数）。



伯努利分布

Bernoulli distribution

- 只能取两个值的随机变量为**伯努利随机变量**（Bernoulli random variable）。

  例如：抛硬币朝上的面、随机遇到路人的性别等

  在不失去一般性的情况下，我们可以令伯努利随机变量的取值为 0 和 1 .

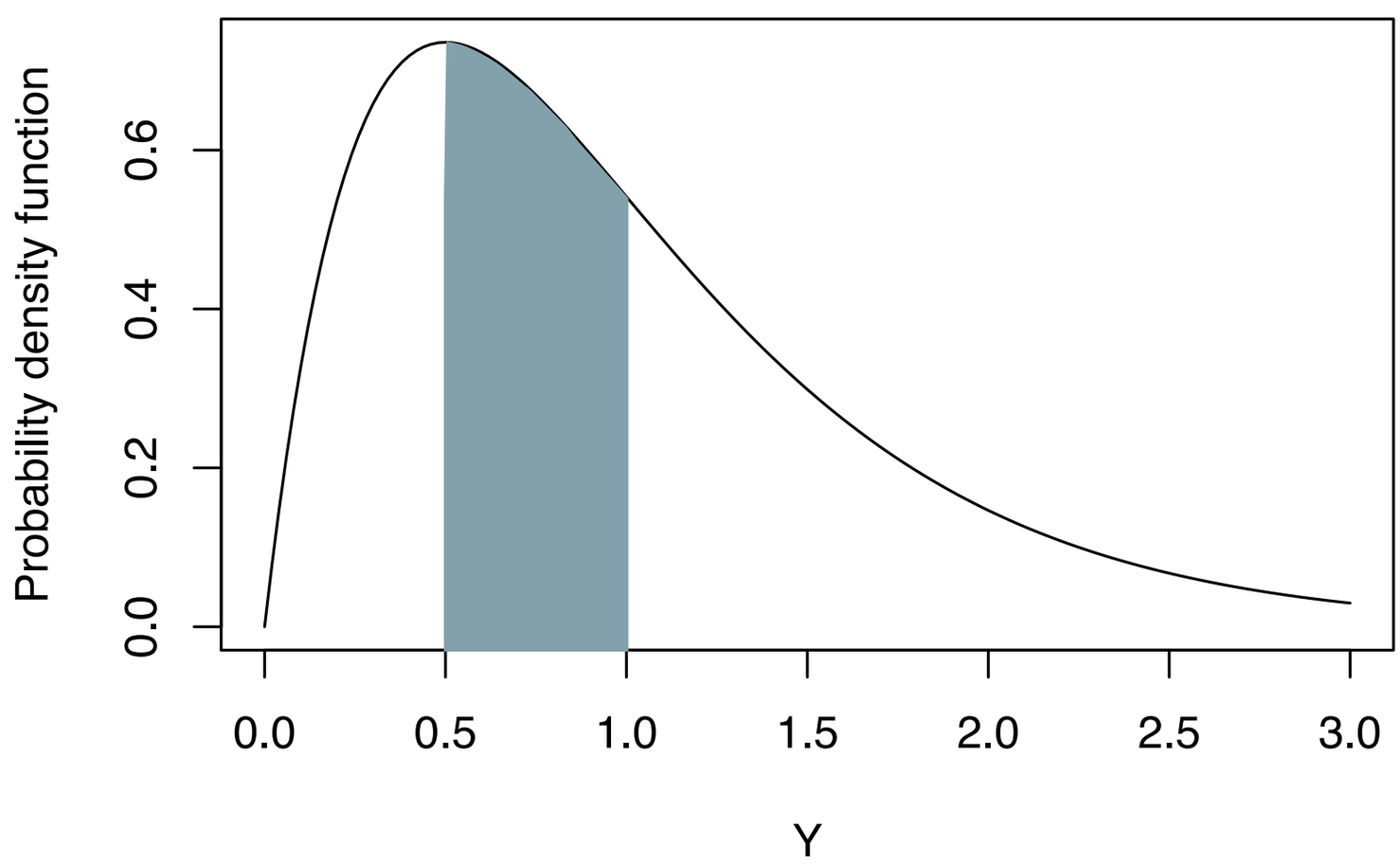
- 伯努利随机变量 G 所对应的分布为伯努利分布：

$$G = \begin{cases} 0, & \text{概率为 } p \\ 1, & \text{概率为 } 1-p \end{cases}$$

连续型随机变量的概率分布

Probability distribution of a continuous r.v.

- 连续型随机变量的取值是不可数的，因此无法列出所有取值和概率。此时我们用**概率密度函数（probability density distribution, p.d.f.）**表述其概率，即随机变量落入两点之间的概率等于这两点之间概率密度函数曲线下方的面积。

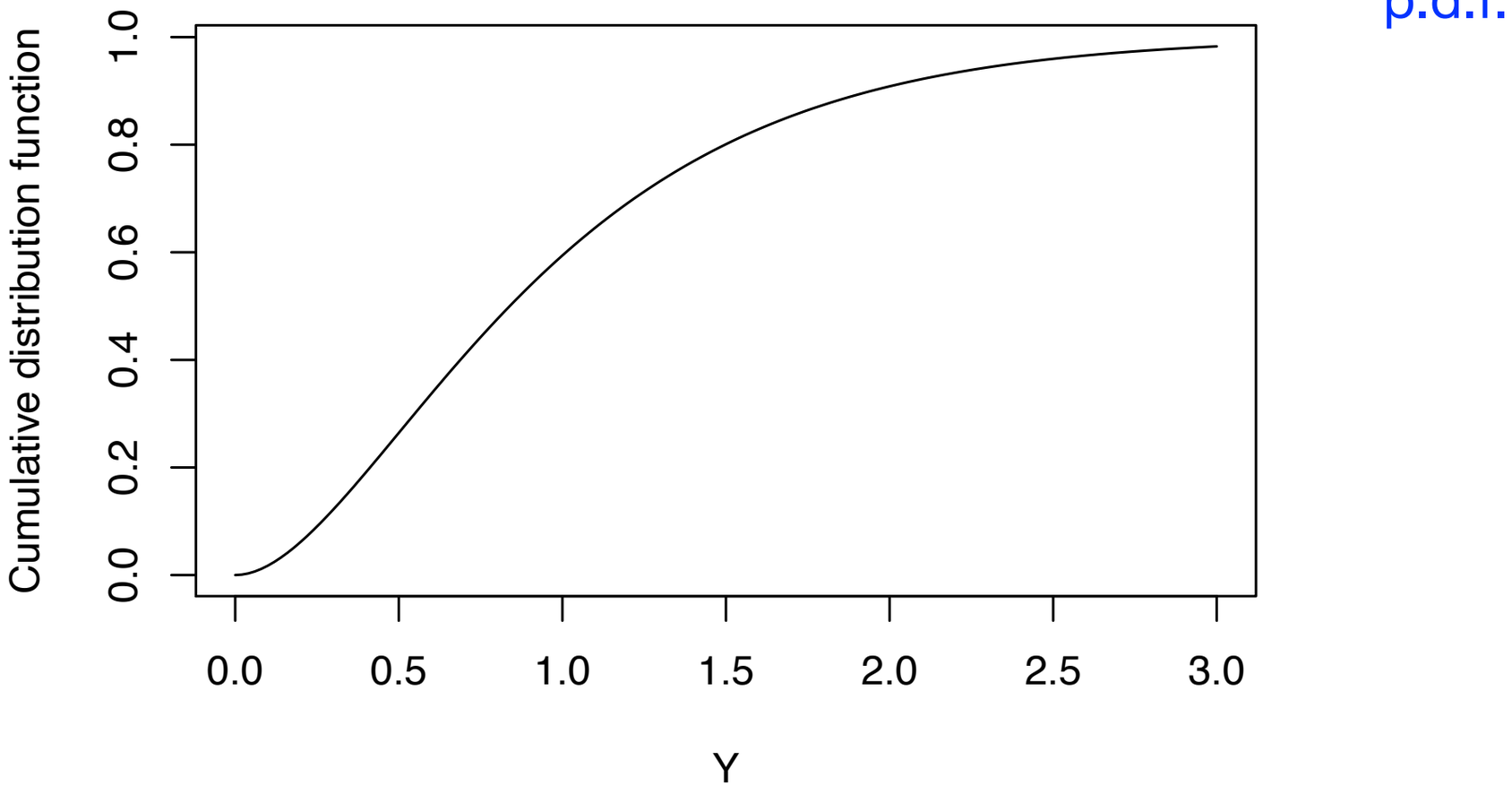


连续随机变量的累积分布函数

- 连续型随机变量**累积分布函数（cumulative distribution function, c.d.f.）**可以表述为

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

p.d.f.

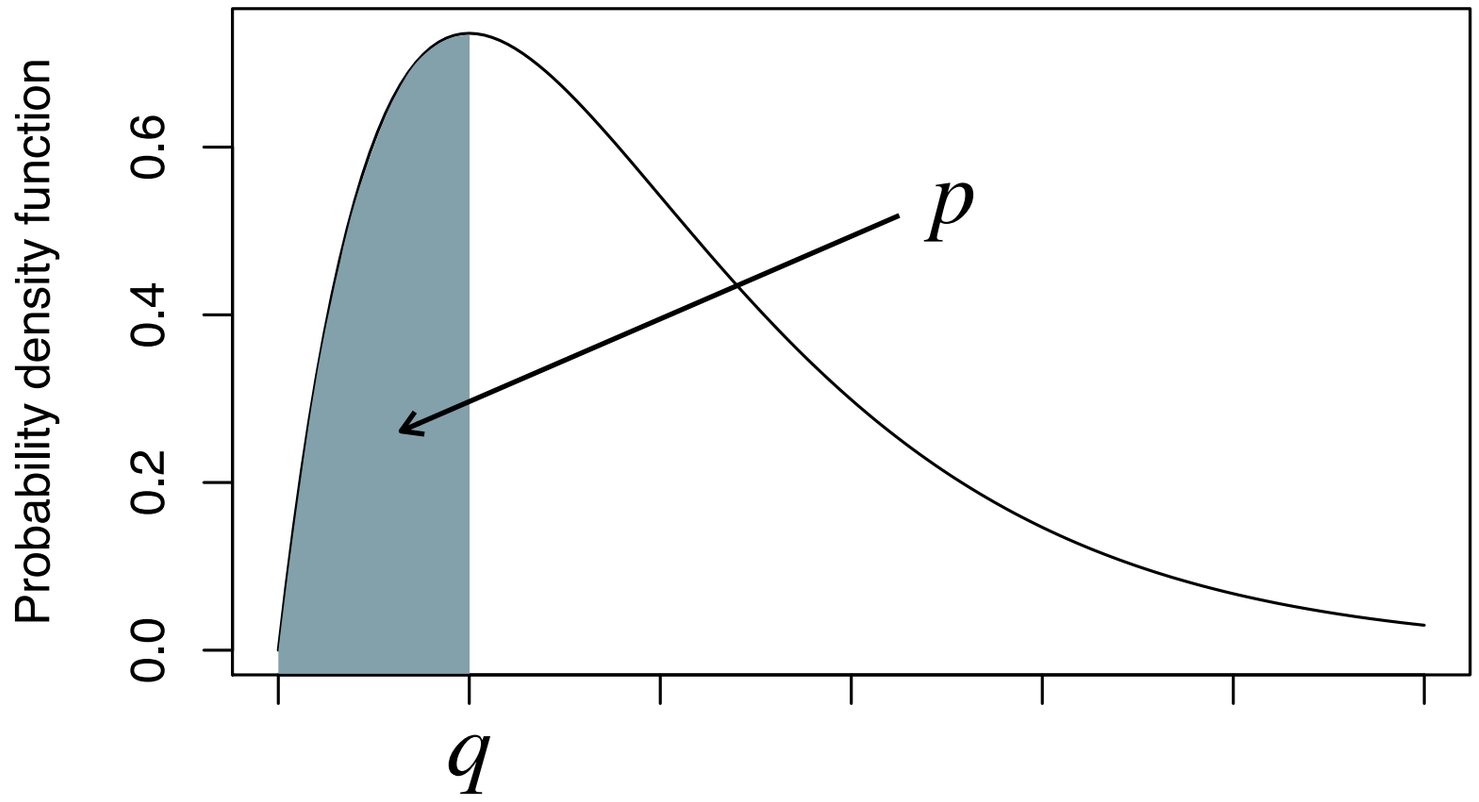


分位函数

Quantile function

- 累积分布函数的逆函数被称为分位函数，即给出累积概率值时，该函数返回随机变量取值范围的上限

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



期望值、方差、标准差

Expected value, variance, and standard deviation

- 期望值（或均值）μ_X

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

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

- 方差 σ_X^2

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

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

- 标准差

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

两个随机变量及其概率分布

- 我们经常需要同时考虑两个或更多的随机变量。例如，性别与收入，天气与农作物产量等。这就需要了解联合概率分布、边缘概率分布、条件概率分布的概念。

- **联合概率分布（joint probability distribution）**：两个随机变量同时取某些值时的概率分布，如 $(X, Y) = (x, y)$ 时。

- **边缘概率分布（marginal probability distribution）**：一个变量的概率分布。此术语是为了在多个变量时与联合分布区分开。

- **条件概率分布（conditional probability distribution）**：给定随机变量 X 取某特定值的条件下，另一随机变量 Y 的分布。

联合分布与边缘分布

		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	

X 与 Y 的联合概率：$\Pr(X = x, Y = y)$

X 的边缘概率：$\Pr(X = x) = \sum_{i=1}^{n} \Pr(X = x, Y = y_i)$

Y 的边缘概率：$\Pr(Y = y) = \sum_{i=1}^{n} \Pr(X = x_i, Y = y)$

条件概率与条件期望

Conditional probability and conditional expectation

- 当 $X = x$ 时 Y 的条件概率为

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

- 条件期望（conditional expectation）

$$E(Y \mid X = x) = \sum_{i=1}^{n} y_i \Pr(Y = y_i \mid X = x)$$

- 条件方差（conditional variance）

$$\mathrm{Var}(Y \mid X = x) = \sum_{i=1}^{n} \left[y_i - E(Y \mid X = x_i) \right]^2 \Pr(Y = y_i \mid X = x)$$

期望的迭代原则：$E(Y) = E[E(Y \mid X)]$

The law of iterated expectation

$$E(Y) = \sum_{i=1}^{n} y_i \Pr(Y = y_i) \qquad \text{期望值的定义}$$

$$= \sum_{i=1}^{n} y_i \sum_{j=1}^{m} \Pr(Y = y_i, X = x_j) \qquad \text{边缘概率的定义}$$

$$= \sum_{i=1}^{n} y_i \sum_{j=1}^{m} \Pr(Y = y_i \mid X = x_j) \Pr(X = x_j) \qquad \text{条件概率的定义}$$

$$= \sum_{j=1}^{m} \sum_{i=1}^{n} y_i \Pr(Y = y_i \mid X = x_j) \Pr(X = x_j) \qquad \text{调整计算顺序}$$

$$= \sum_{j=1}^{m} E(Y \mid X = x_j) \Pr(X = x_j) \qquad \text{条件期望的定义}$$

$$= E[E(Y \mid X)] \qquad \text{视 } E(Y \mid X) \text{ 为随机变量}$$

独立性

Independence

- 若知道两个随机变量 X 和 Y 中某一个变量的取值无法提供另一个变量的取值信息，则称 X 和 Y 独立分布（independently distributed）或**独立（independent）**。

- 数学定义：当

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

时，则 X 和 Y 独立。

- 若 X 和 Y 独立，则

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

协方差与相关系数

Covariance and correlation

- 协方差（covariance）是衡量两个变量同时变动成的的一个指标。

$$\begin{aligned}\sigma_{XY} &= \operatorname{cov}(X, Y) \\ &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} (x_i - \mu_X)(y_j - \mu_Y) \Pr(X = x_i, Y = y_j)\end{aligned}$$

- 相关系数（correlation）

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

独立性、条件期望与相关性

- 当 $\text{corr}(X, Y) = 0$ 时，我们说 X 和 Y 不相关（uncorrelated）。
- 独立性、条件期望和相关性之间存在下列关系：

$$X \text{ 和 } Y \text{ 独立} \quad \Rightarrow \quad E(Y \mid X) = \mu_Y \quad \Rightarrow \quad \text{corr}(X, Y) = 0$$

均值独立（mean independent）

- 第二部分的证明：

在不失去一般性的情况下，我们可以假设 $\mu_X = \mu_Y = 0$。则有

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] = E(XY) \\ &= E(YX) = E[E(YX \mid X)] \qquad \text{期望的迭代原则} \\ &= E[E(Y \mid X)\, X] = 0 \qquad \text{因为 } E(Y \mid X) = \mu_Y = 0 \end{aligned}$$

因此，$\text{corr}(X, Y) = 0$。

- 注意：X 和 Y 独立 $\quad \not\Leftarrow \quad E(Y \mid X) = \mu_Y \quad \not\Leftarrow \quad \text{corr}(X, Y) = 0$

关于随机变量之和

- 期望值

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y) = \mu_X + \mu_Y$$

- 方差

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + \mathrm{cov}(X,Y) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$$

当 X 和 Y 相互独立时，其协方差为零，则

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) = \sigma_X^2 + \sigma_Y^2$$

其他有用的公式

重要概念2.3

- 设 X, Y, Z 为随机变量，a, b, c 为常数，则有

$$\mathrm{E}(a + bX + cY) = a + b\mu_X + c\mu_Y,$$

$$\mathrm{var}(a + bY) = b^2\sigma_Y^2,$$

$$\mathrm{var}(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2,$$

$$\mathrm{E}(Y^2) = \sigma_Y^2 + \mu_Y^2,$$

$$\mathrm{cov}(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{VY},$$

$$\mathrm{E}(XY) = \sigma_{XY} + \mu_X\mu_Y,$$

$$|\mathrm{cov}(X, Y)| \leq 1 \text{ and } |\sigma_{XY}| \leq \sqrt{\sigma_X^2\sigma_Y^2}.$$

概率分布

二项分布

The binomial distribution

- 二项分布是拥有两个参数 n 和 p 的离散分布，其定义为

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n.$$

- 当 n 个随机变量 $X_1, X_2, \ldots, X_n$ 为独立同分布（independent and identically distributed, i.i.d），且都服从参数为 p 的伯努利分布时，他们的和 $X = X_1 + X_2 + \ldots + X_n$ 即服从参数为 n 和 p 的二项分布。

- 二项分布多用来描述反复发生的随机事件的发生次数。
例如：连续抛10次硬币时，4次正面朝上的概率。

正态分布

The normal distribution

- **正态分布**是一种连续分布。均值为 μ 方差为 σ^2 的正态分布 $N(\mu, \sigma^2)$ 的密度函数为

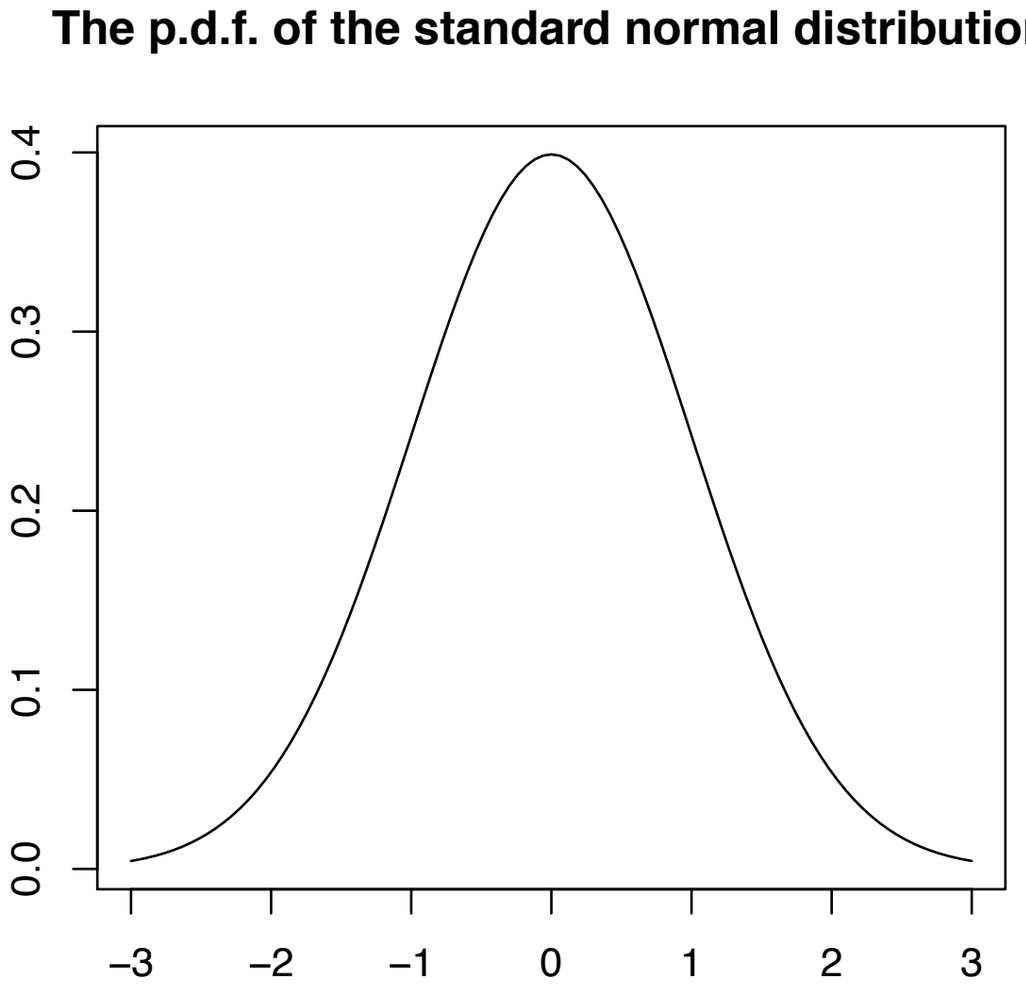
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big[-\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\Big]$$

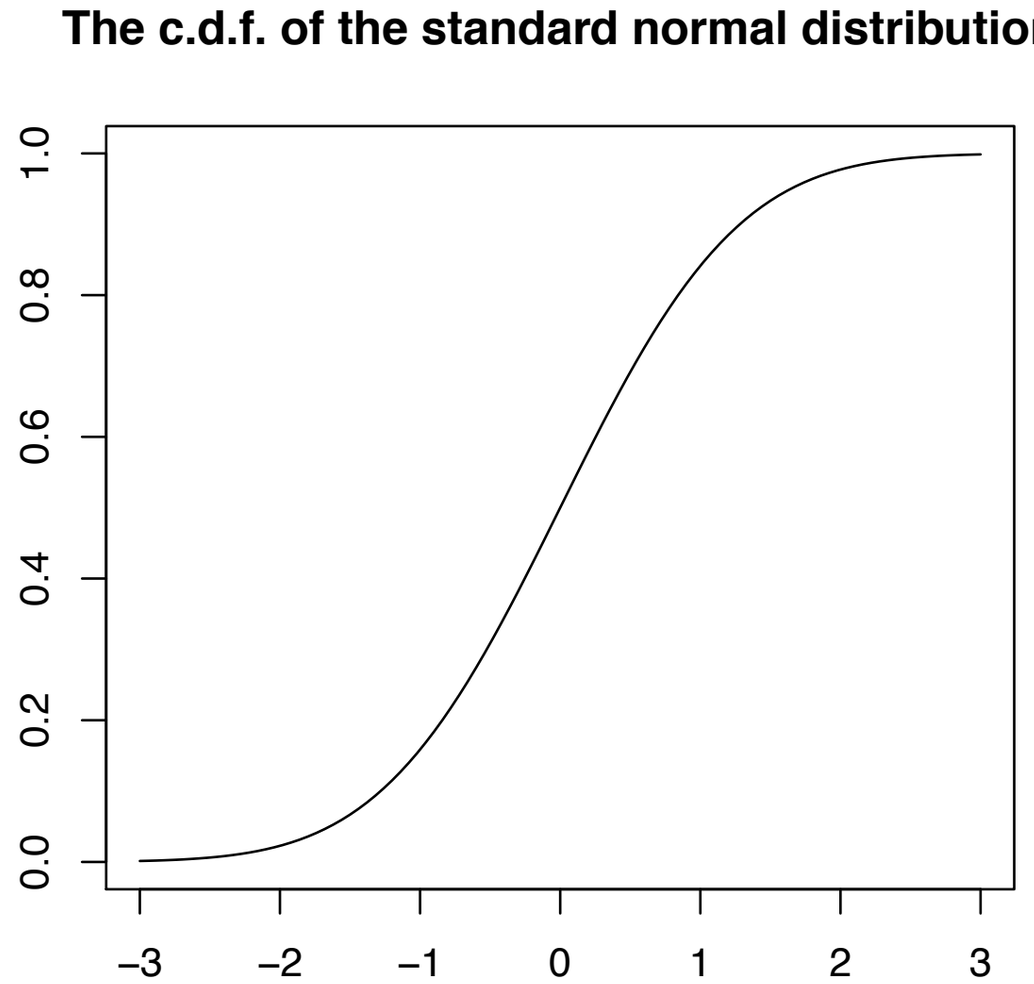
- $N(0,1)$ 被称为**标准正态分布（stantard normal distribution）**。服从标准正态分布的随机变量多用字母 Z 来表示，此时可将其累积分布函数写为

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

- 当二项分布的 n 足够大时，正态分布 $N(np, np(1-p))$ 可以作为该二项分布的近似。

正态分布的密度和累积分布函数





正态分布的概率

- 当 $X \sim N(\mu, \sigma^2)$ 时,

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

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

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

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

- 标准化（standardization）

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

卡方分布

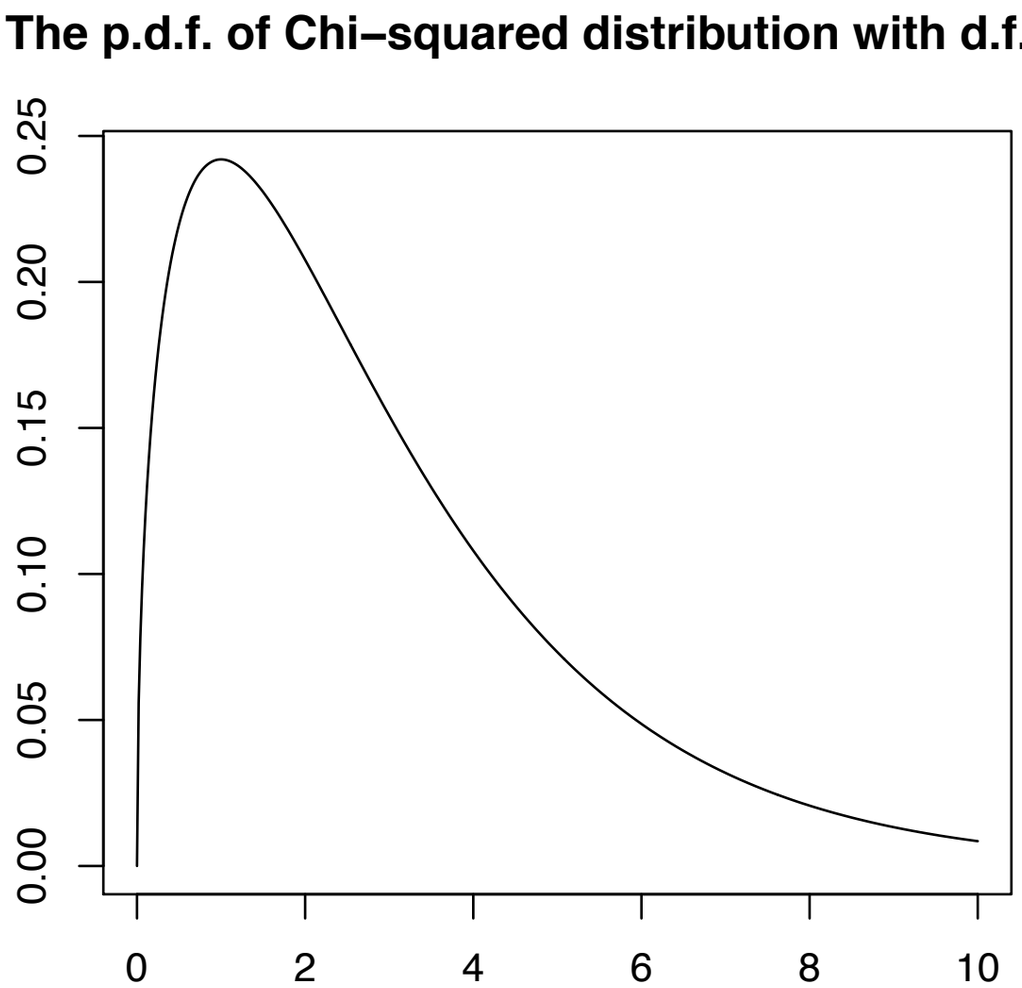
The chi-squared distribution

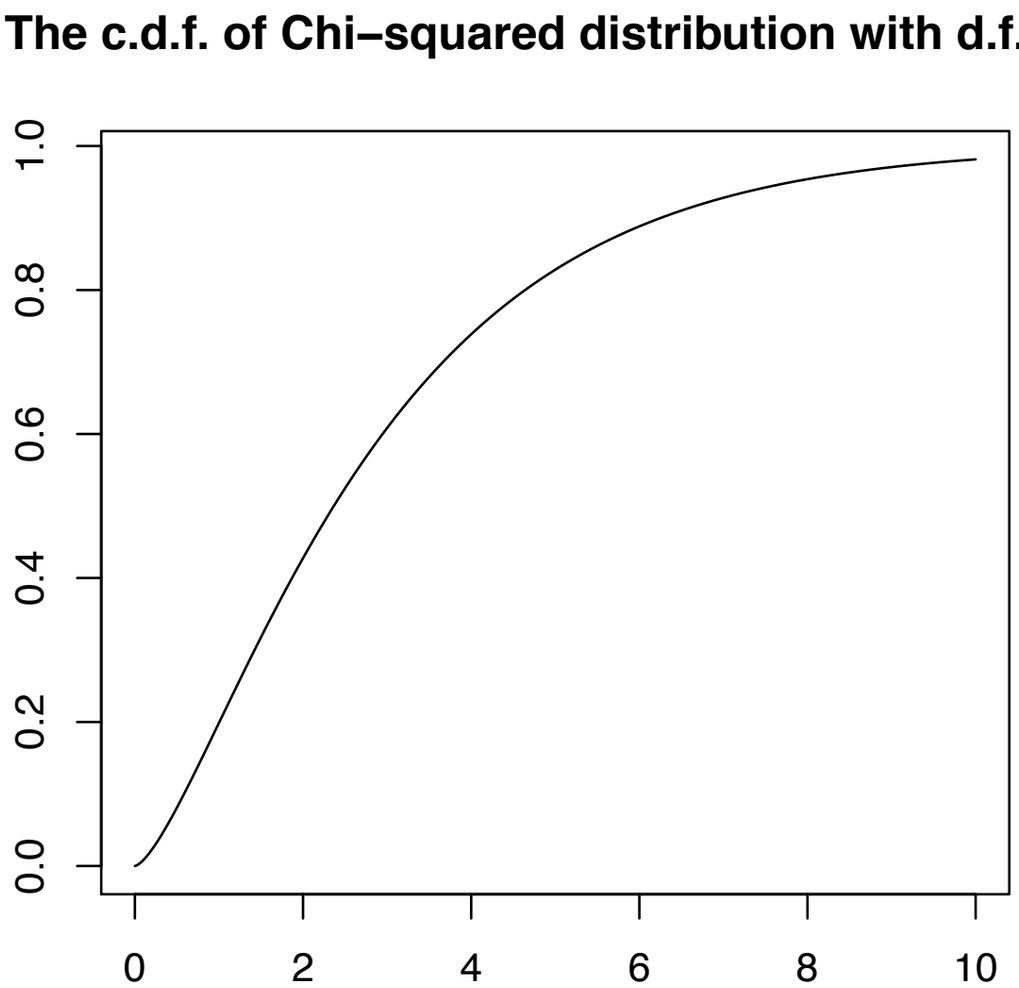
- m 个独立标准正态随机变量的平方和所服从的分布被称为**自由度（degree of freedom）为 m 的卡方分布**，记为 χ^2_m。这里 χ 为希腊字母，英语标记为 chi，发音为 kai（接近“忾”）。

- 例如，令 Z_1, Z_2, Z_3 为相互独立的标准正态随机变量，则 $Z_1^2 + Z_2^2 + Z_3^2$ 服从自由度为 3 的卡方分布。

- 在统计学和计量经济学中，某些类型的假设检验需要用到卡方分布。

卡方分布的密度和累积分布函数





Student's *t* 分布

- 令 Z 表示标准正态随机变量，W 表示服从自由度为 m 的卡方分布的随机变量，且 Z 和 W 独立。则随机变量

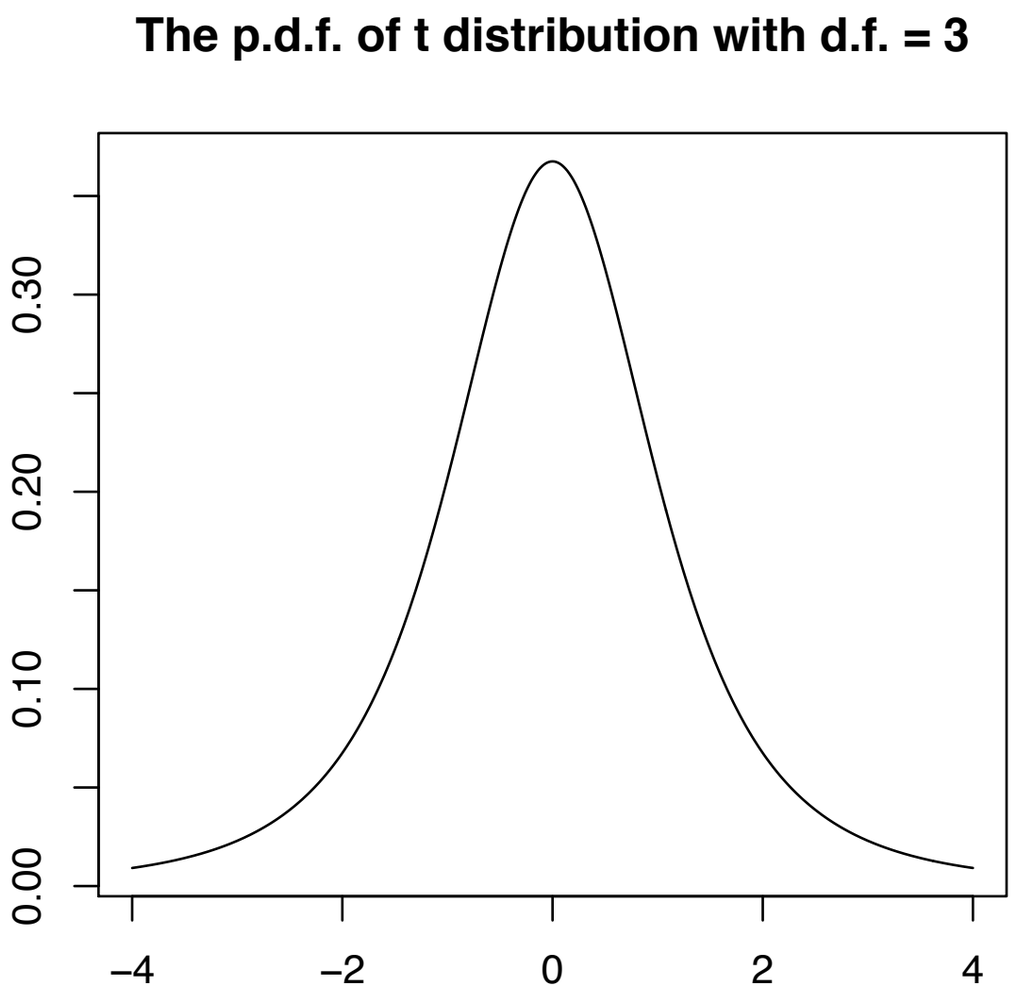
$$Z \Big/ \sqrt{W/m}$$

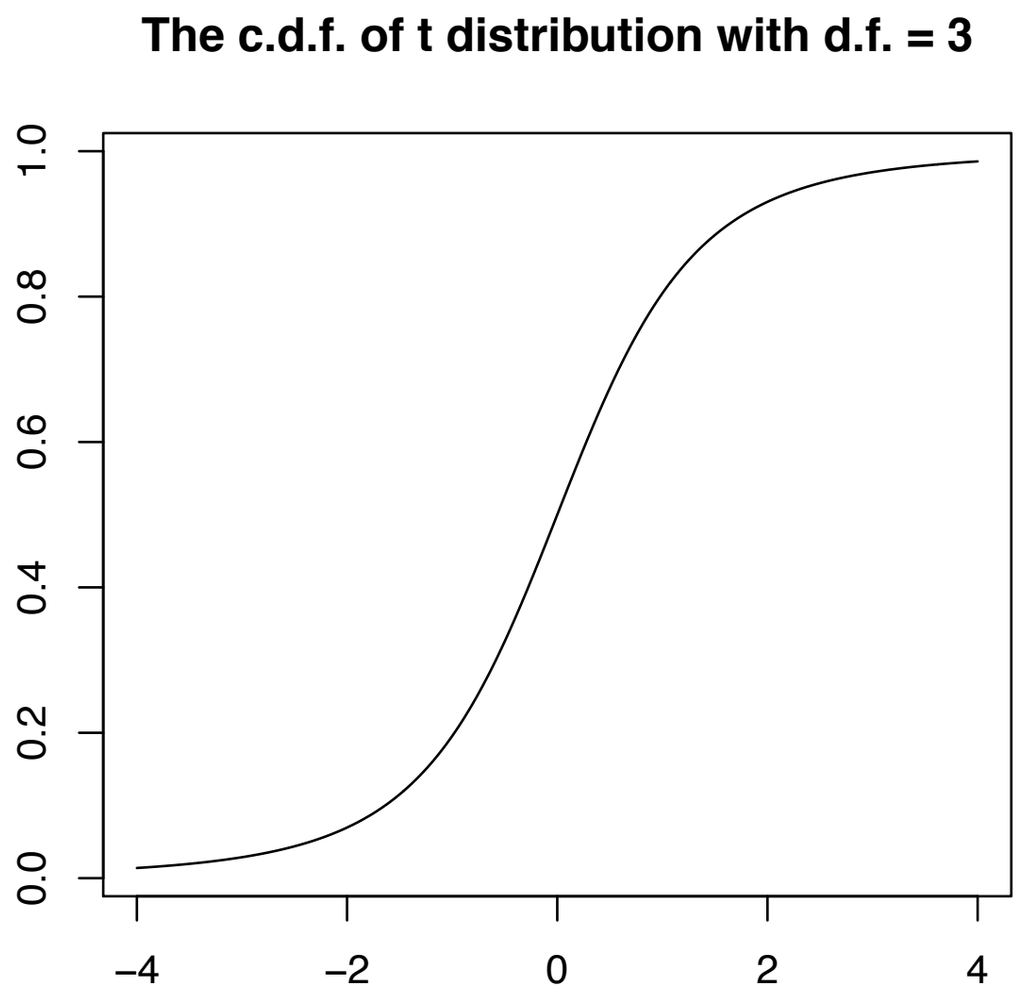
服从**自由度为 m 的 *t* 分布**，记为 t_m 。

- *t* 分布的密度函数形状与正态分布相似，但当自由度 m 较小时，*t* 分布尾部较厚，比正态分布更“平坦”。当 $m \geq 30$ 时可用标准正态分布近似 *t* 分布。t_∞ 等于标准正态分布。

- *t* 分布由英国人威廉·戈塞（William Sealy Gosset）于1908年发表。他当时就职的爱尔兰吉尼斯酿酒厂要求他以笔名发表该成果，他就取了“Student”为笔名。

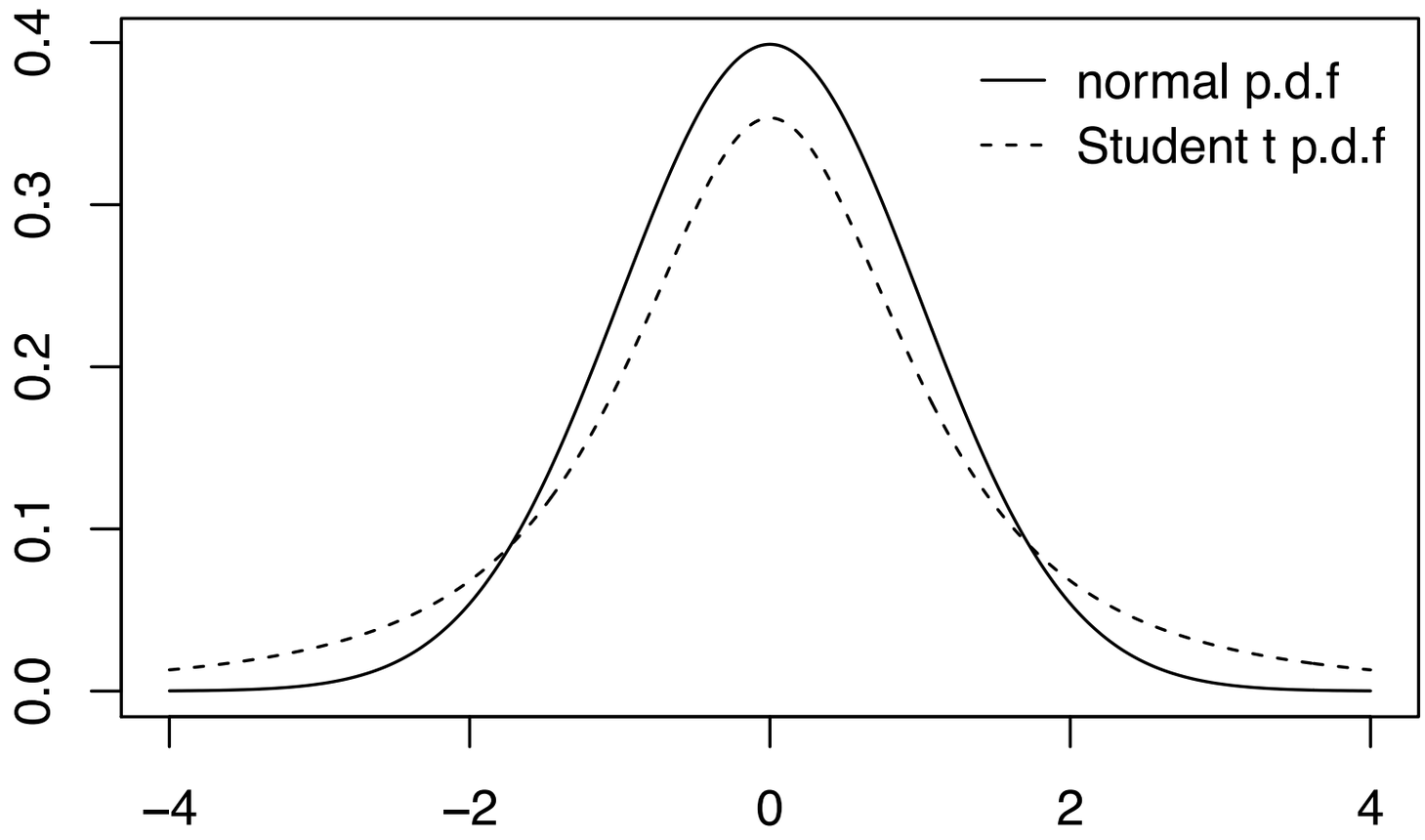
t 分布的密度和累积分布函数





t 分布与正态分布

- 当 t 分布的自由度较小时（$m < 20$），t 分布比标准正态分布拥有“更厚的尾部”。



F 分布

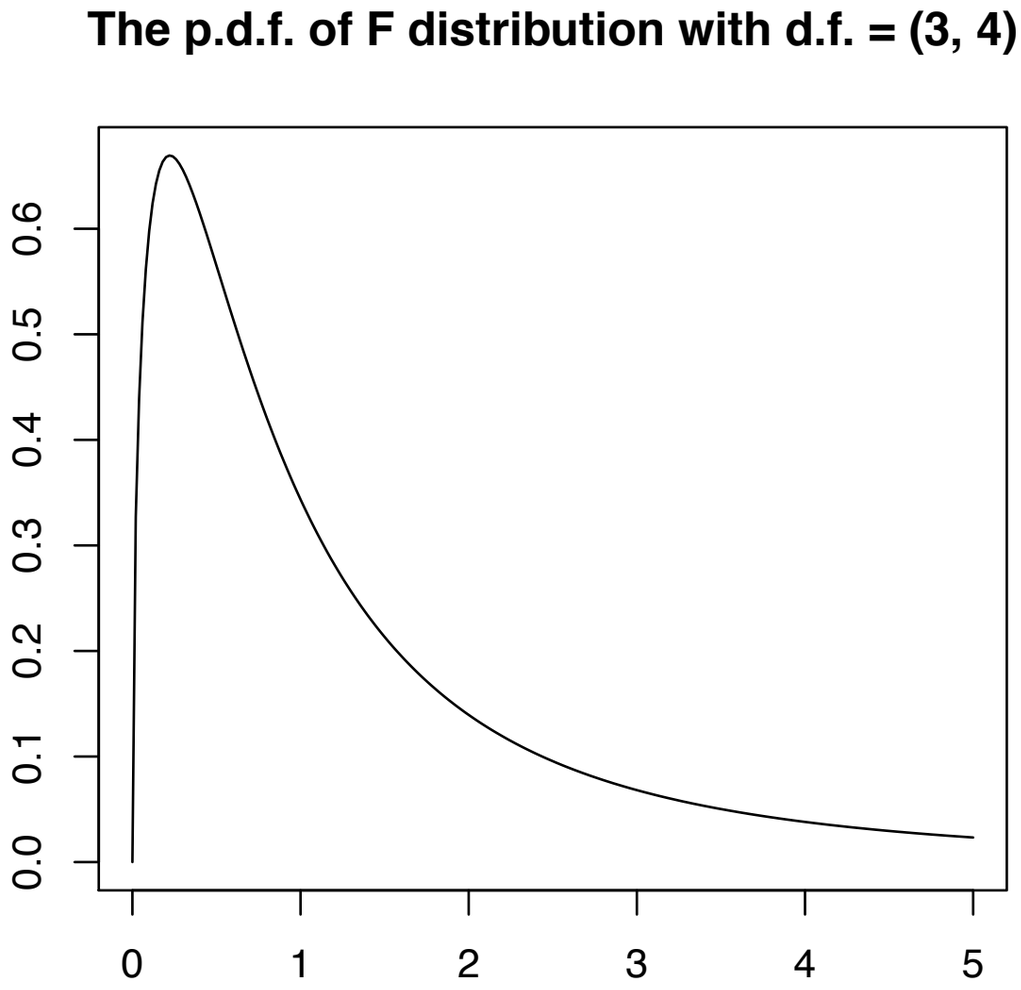
- 令 W 为自由度为 m 的卡方随机变量，V 为自由度为 n 的卡方随机变量，且 W 与 V 独立，则

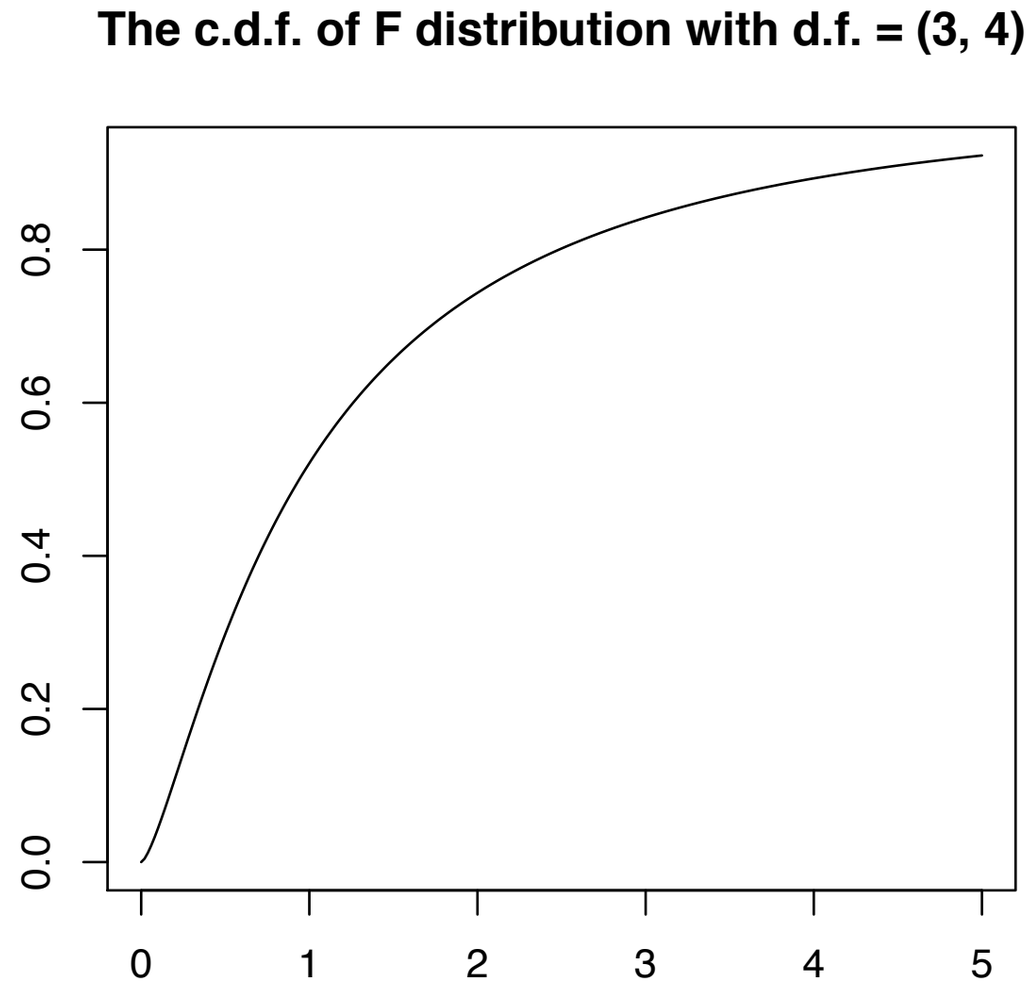
$$\frac{W/m}{V/n}$$

服从自由度为 m 和 n 的 *F* 分布，记为 $F_{m,n}$。

- 当 n 足够大时，$F_{m,n}$ 可用 $F_{m,\infty}$ 近似，而后者为 W/m 的分布。

F 分布的密度和累积分布函数





其他概率分布

- 离散分布
  - 离散均匀分布（discrete uniform distribution）
  - 泊松分布（Poisson distribution）
  - 负二项分布（negative binomial distribution）
    - 特例：几何分布（geometric distribution）
- 连续分布
  - 连续均匀分布
  - 对数正态分布（lognormal distribution）
  - Gamma 分布
    - 特例：指数分布（exponential distribution）
  - Beta 分布

随机抽样与大样本

随机抽样

Random sampling

- 简单随机抽样（simple random sampling）

  从总体（population）中随机选取 n 个样本，且使总体中的每个成员都有同等机会入选。

  - 例：随机从2021年4月中选择1天，然后在这一天早上8:00观察天气。重复此操作10次，即得到10个随机样本。此时总体是2021年4月每天的天气情况。

- 样本 $Y_1, Y_2, \ldots, Y_n$ 为独立同分布（i.i.d.）随机变量

  - 独立（independence）：$Y_1, Y_2, \ldots, Y_n$ 相互独立，即 Y_i 不受其他样本影响

  - 同分布（identically distributed）：$Y_1, Y_2, \ldots, Y_n$ 都服从同一分布，即总体的分布

样本均值与抽样分布

Sample average and sampling distribution

- 令 $Y_1, Y_2, \ldots, Y_n$ 为随机样本，因此（在获取观测值之前）是 i.i.d. 随机变量。

- $Y_1, Y_2, \ldots, Y_n$ 的均值 $\overline{Y}$ 也是随机变量，被称为**样本均值（sample mean）**：

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

- $\overline{Y}$ 的分布被称为**抽样分布（sampling distribution）**。

  抽样可以反复进行。每获得一次样本观测值，即可获得一个 $\overline{Y}$ 的观测值。通过多个 $\overline{Y}$ 的观测值可以推测 $\overline{Y}$ 的分布。

抽样分布

Sampling distribution

- 令 μ_Y 和 σ_Y^2 分别表示 Y_i 的期望值和方差。
- $\overline{Y}$ 的期望值和方差分别为

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

- 当总体服从正态分布时，$\overline{Y}$ 也服从正态分布。

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

- 当总体服不服从正态分布时，$\overline{Y}$ 的精确分布可能会非常复杂。

抽样分布的大样本近似

Large sample approximation of sampling distributions

- 当样本容量较大（即 n 足够大）时，我们可以得到抽样分布的近似：

  - **大数定律（the law of large numbers, LLN）**

    当样本容量较大时，$\overline{Y}$ 以非常高的概率逼近 μ_Y 。

- **中心极限定理（the central limit theorem, CLT）**

  当样本容量较大时，$\overline{Y}$ 的分布近似于正态分布 $N(\mu_Y, \sigma_Y^2 / n)$ 。

大数定律

The law of large numbers (LLN)

- 依概率收敛（convergence in probability）

  当 n 增大时，对任意常数 $c > 0$，如果随机变量 X 的取值落入区间 $(\alpha - c, \alpha + c)$ 的概率充分接近于 1，则说 X **依概率收敛**于 α，记为 $X \xrightarrow{p} \alpha$ 。更简洁的说法是 X 与 α **一致（consistent）**。

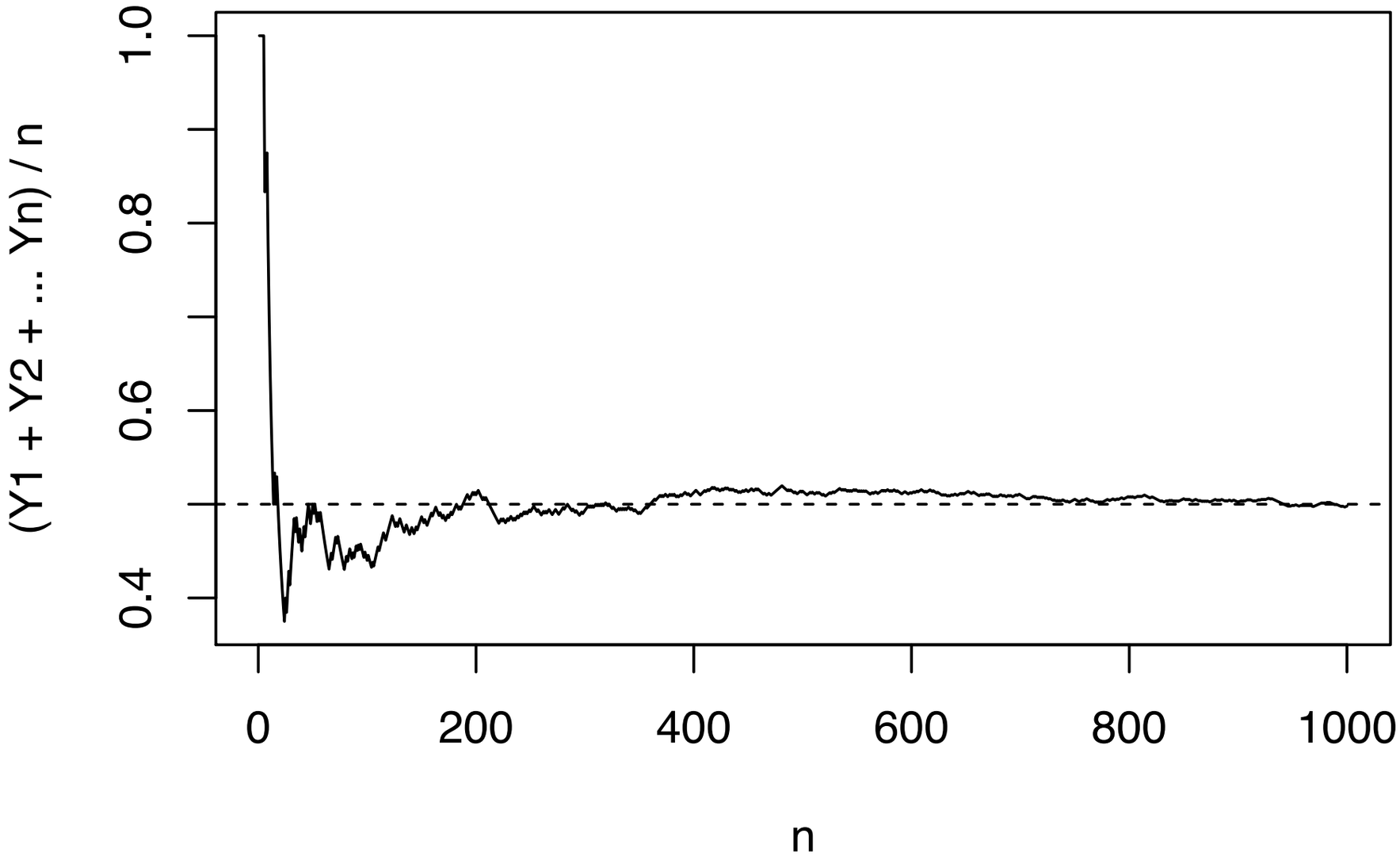
- 大数定律

  如果随机样本 $Y_1, Y_2, \ldots, Y_n$ 是独立同分布，且总体的期望值为 μ_Y，总体的方差为有限（$\sigma_Y^2 < \infty$）时，则 $\overline{Y} \xrightarrow{p} \mu_Y$ 。

注：存在不同版本的大数定律。这里介绍的为俗称的弱大数定律。

样本均值的变化趋势

- 以下为 n 次伯努利试验（例如连续抛硬币 n 次）的样本均值的变化



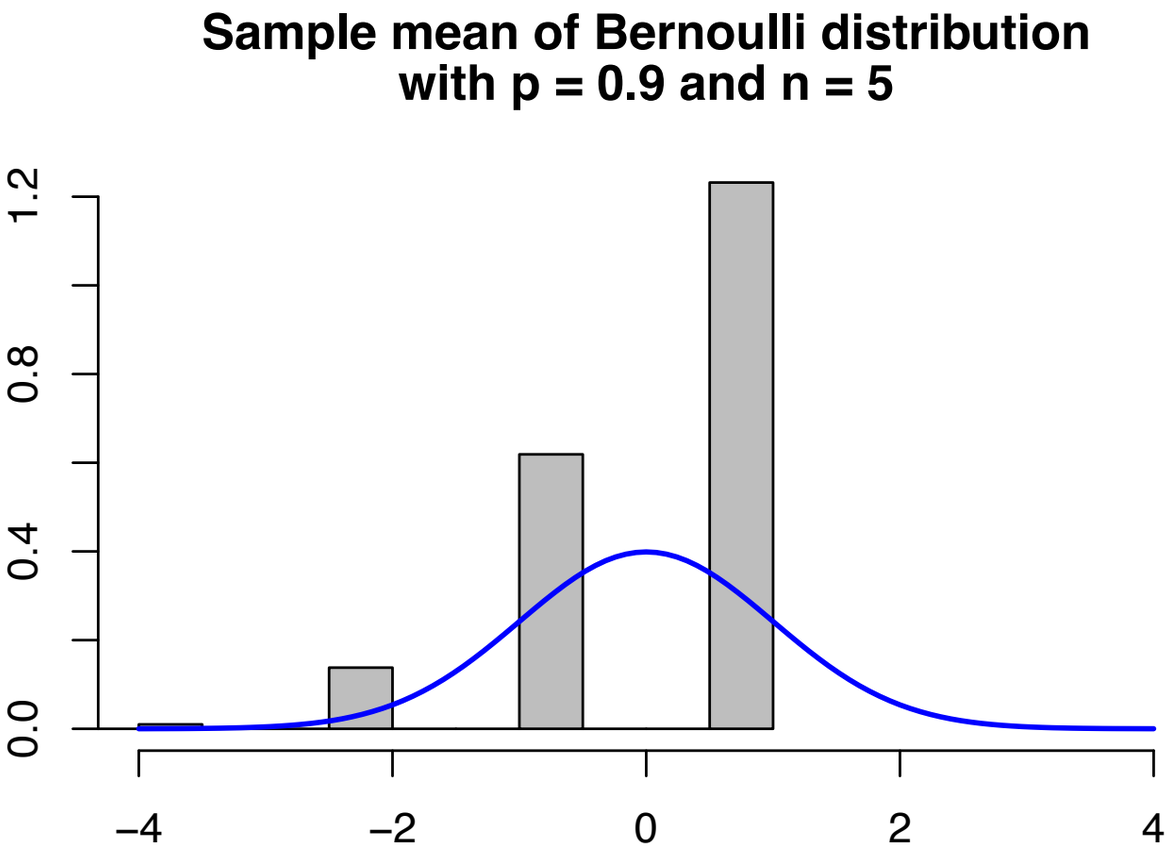
中心极限定理

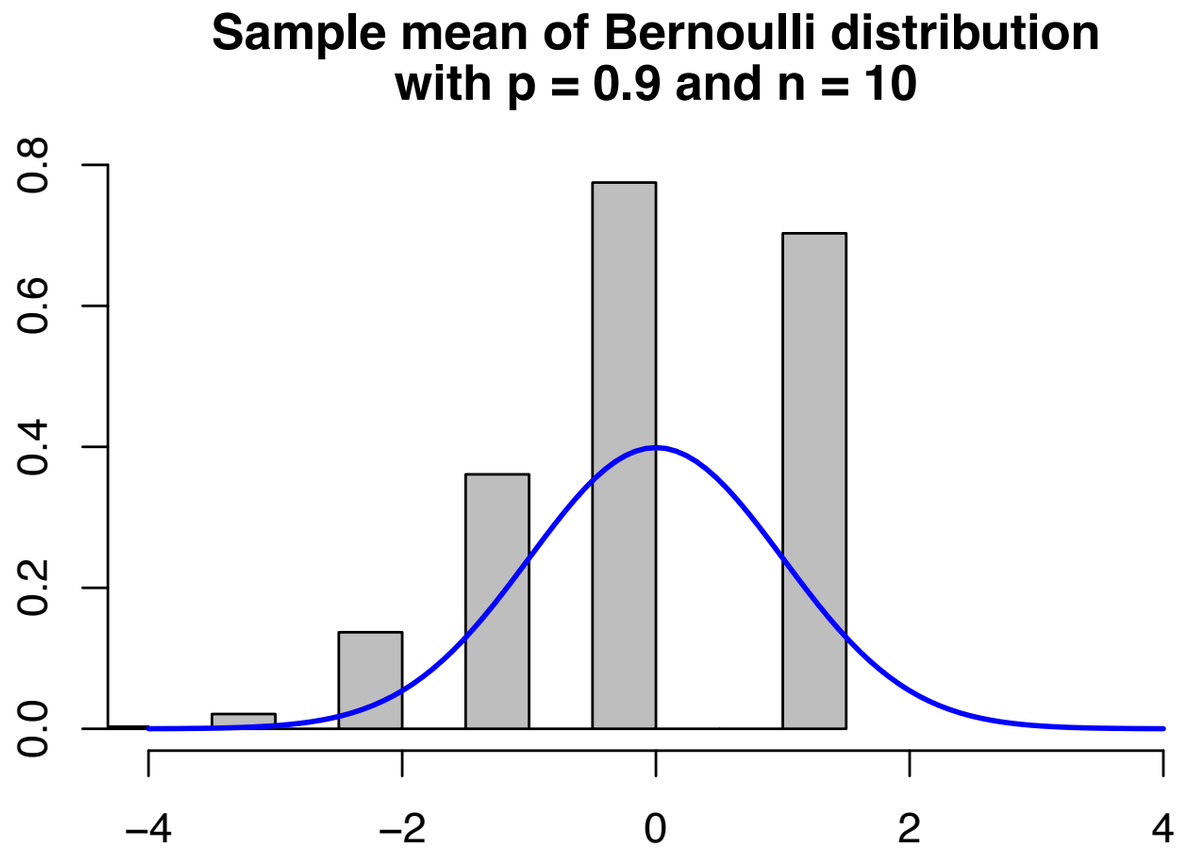
The central limit theorem (CLT)

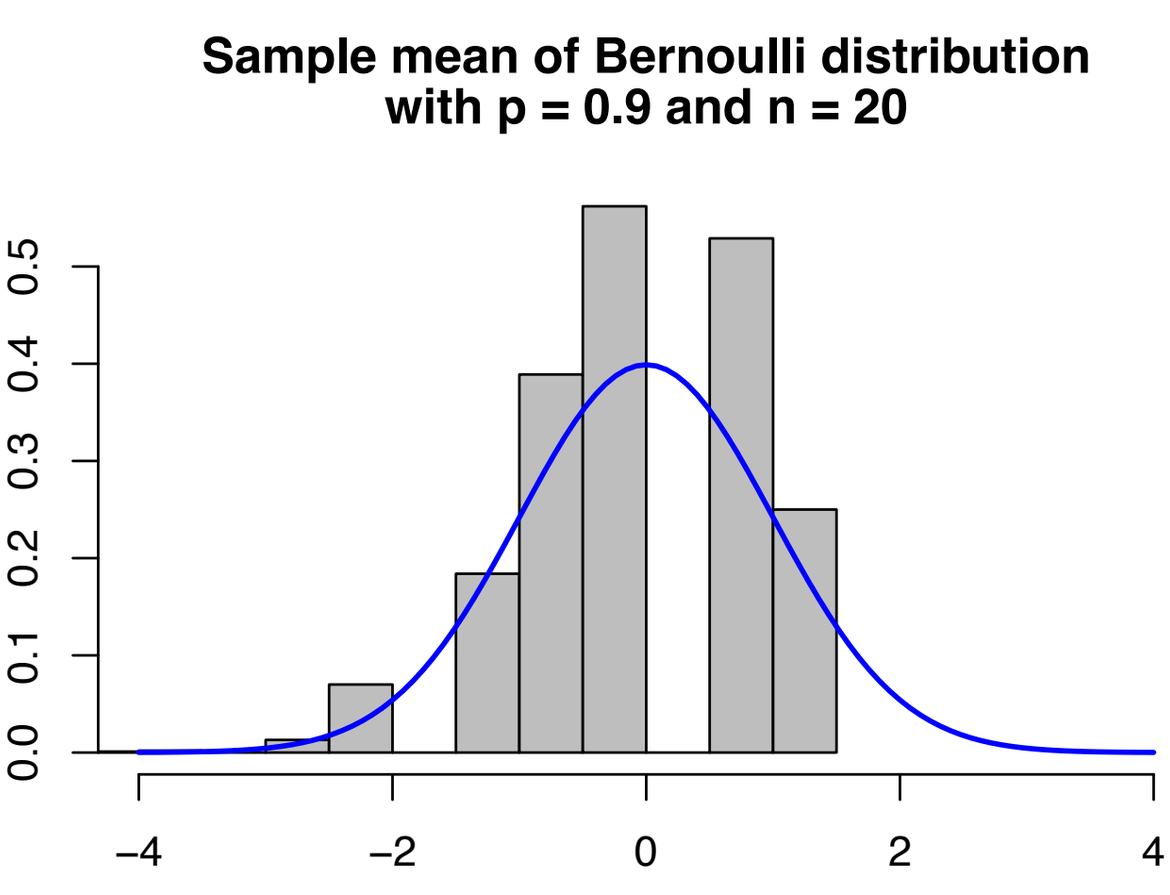
设随机样本 $Y_1, Y_2, \ldots, Y_n$ 是独立同分布，且总体的期望值为 μ_Y，总体的方差为 σ_Y^2 并满足 $0 < \sigma_Y^2 < \infty$ 。则当 $n \to \infty$ 时，$\overline{Y}$ 的分布近似于正态分布 $N(\mu_Y, \sigma_Y^2/n)$ 。或等价地，$\dfrac{\overline{Y} - \mu_Y}{\sigma/\sqrt{n}}$ 的分布近似于标准正态分布。

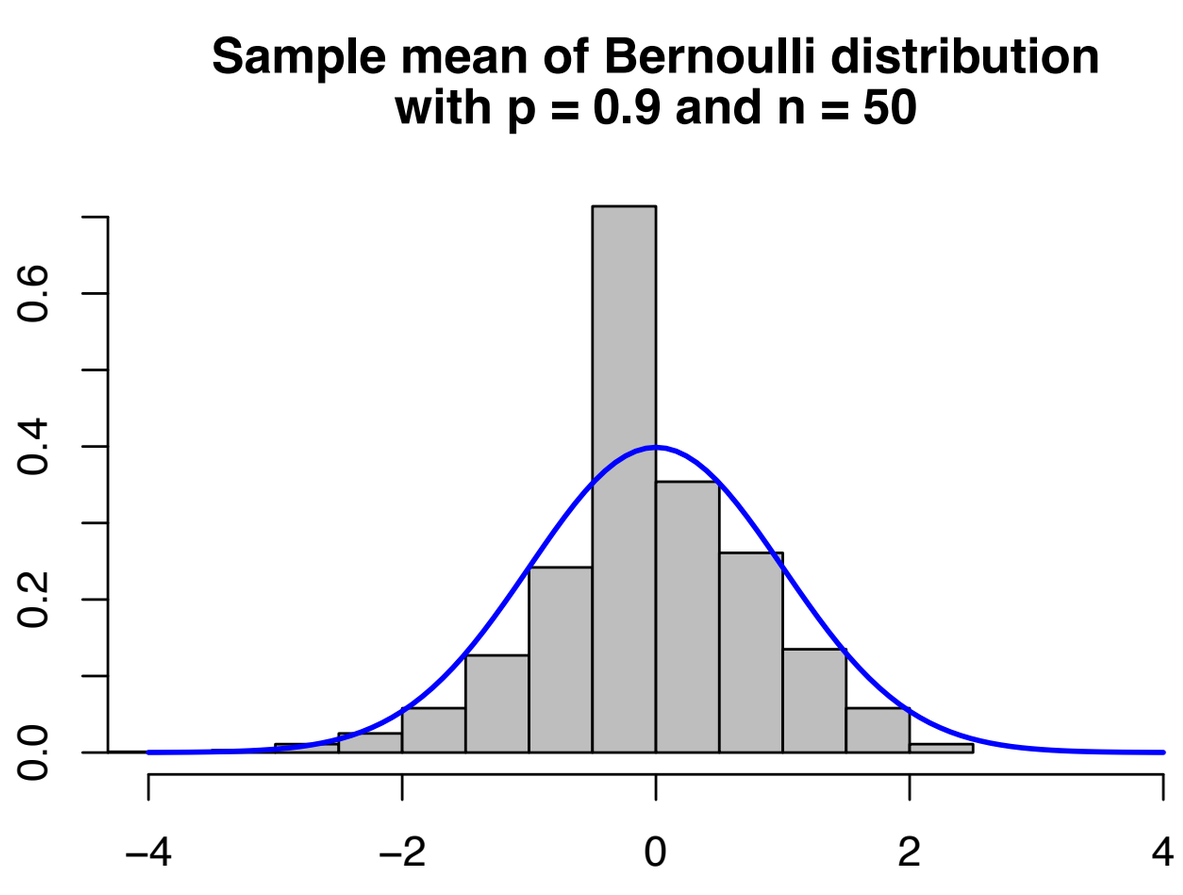
注：存在不同版本的中心极限定理。例如，在附加条件下，可以不要求 $Y_1, Y_2, \ldots, Y_n$ 为同分布（Lyapounov）。

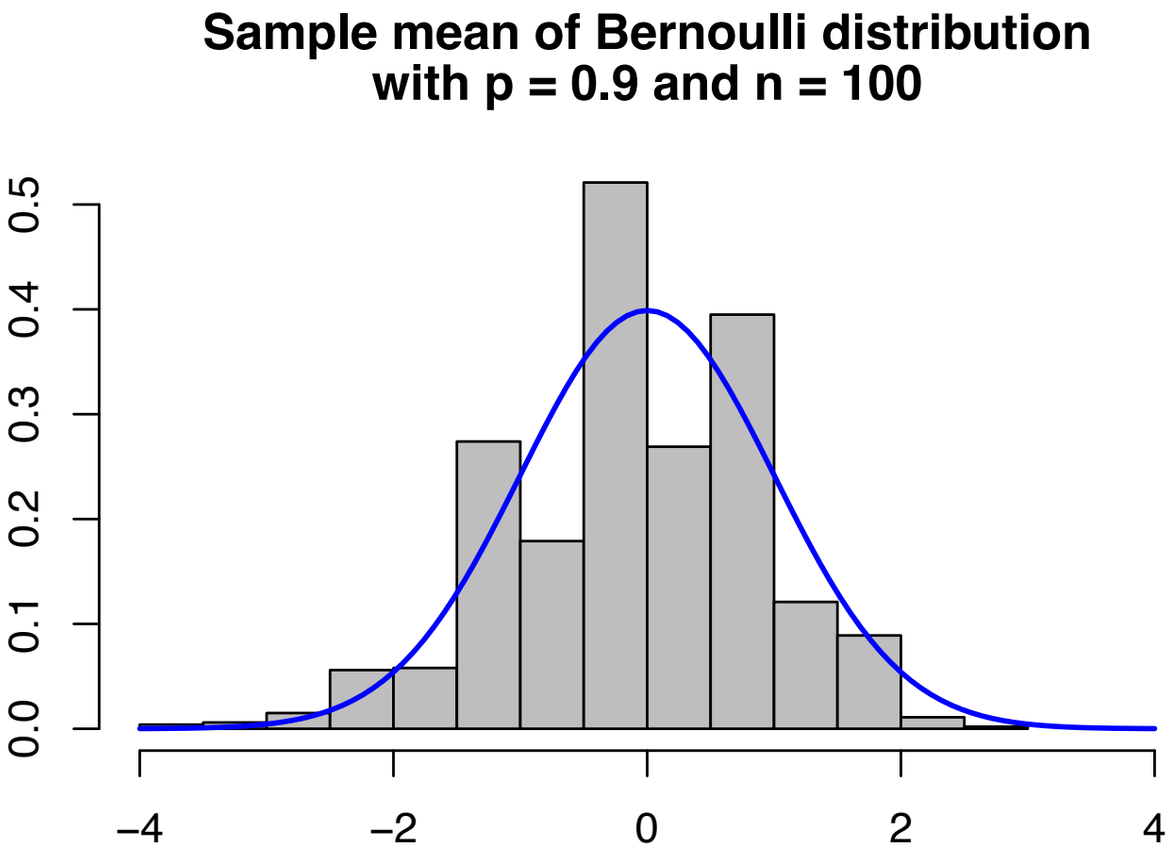
- 当 n 增大时，我们说 $\overline{Y}$ 服从**渐进正态分布（asymptotically normally distributed）**。一般情况下，$n > 30$ 即为足够大。

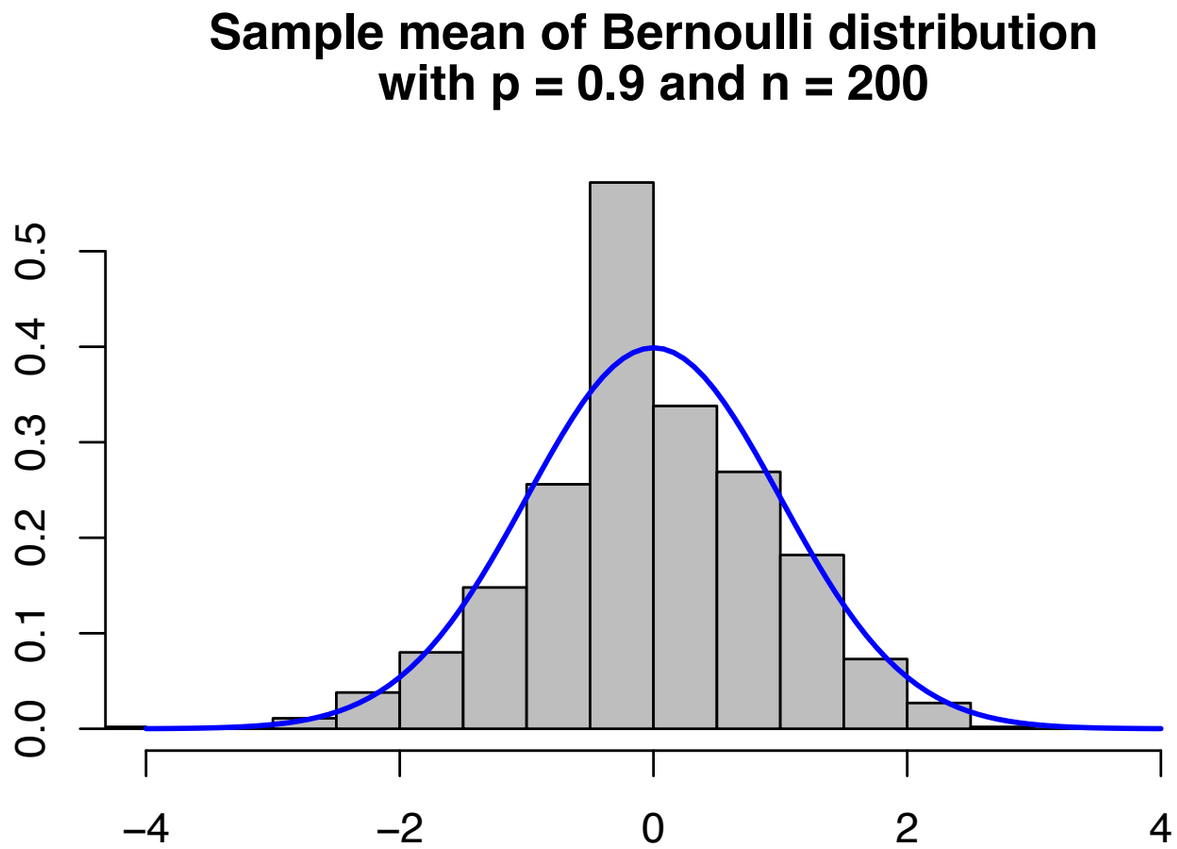

Sample mean of Bernoulli distribution
with p = 0.9 and n = 5
0.0
0.4
0.8
1.2
−4
−2
0
2
4

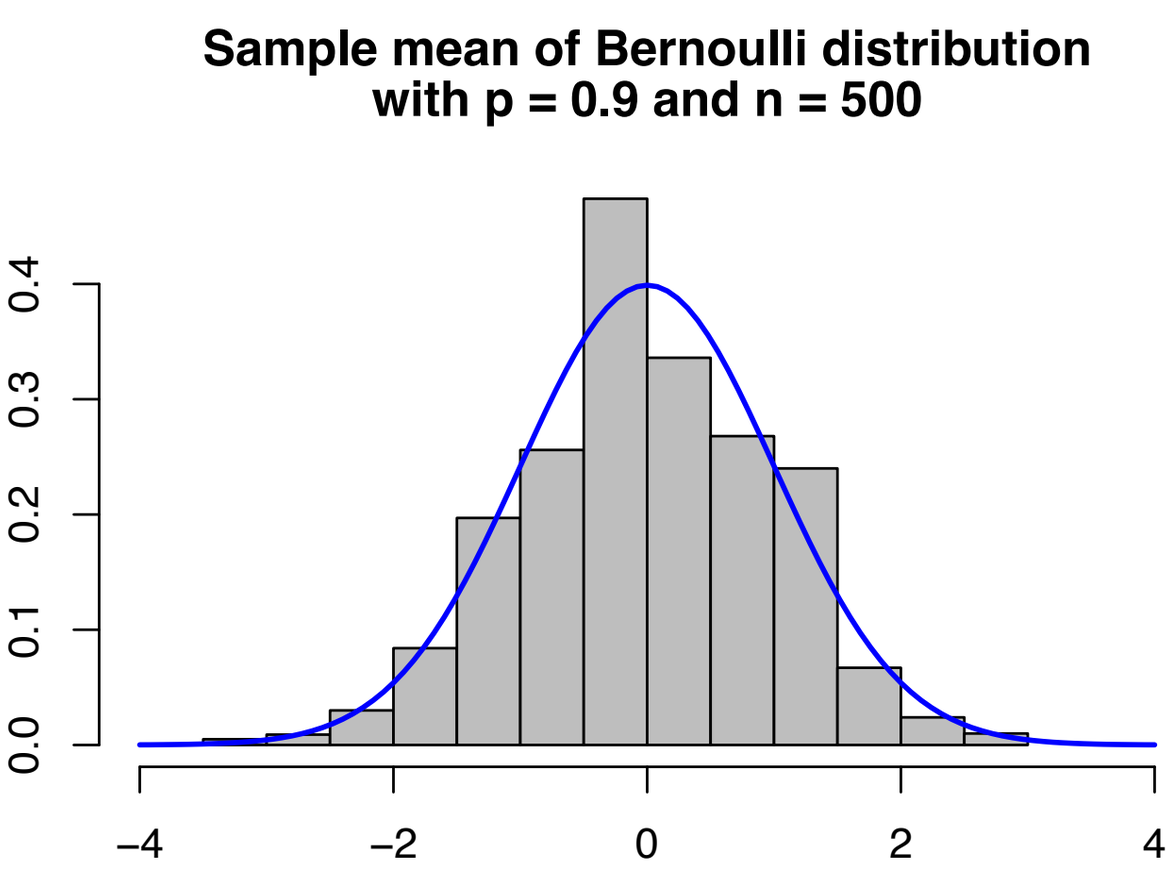

Sample mean of Bernoulli distribution
with p = 0.9 and n = 10
0.0
0.2
0.4
0.6
0.8
−4
−2
0
2
4

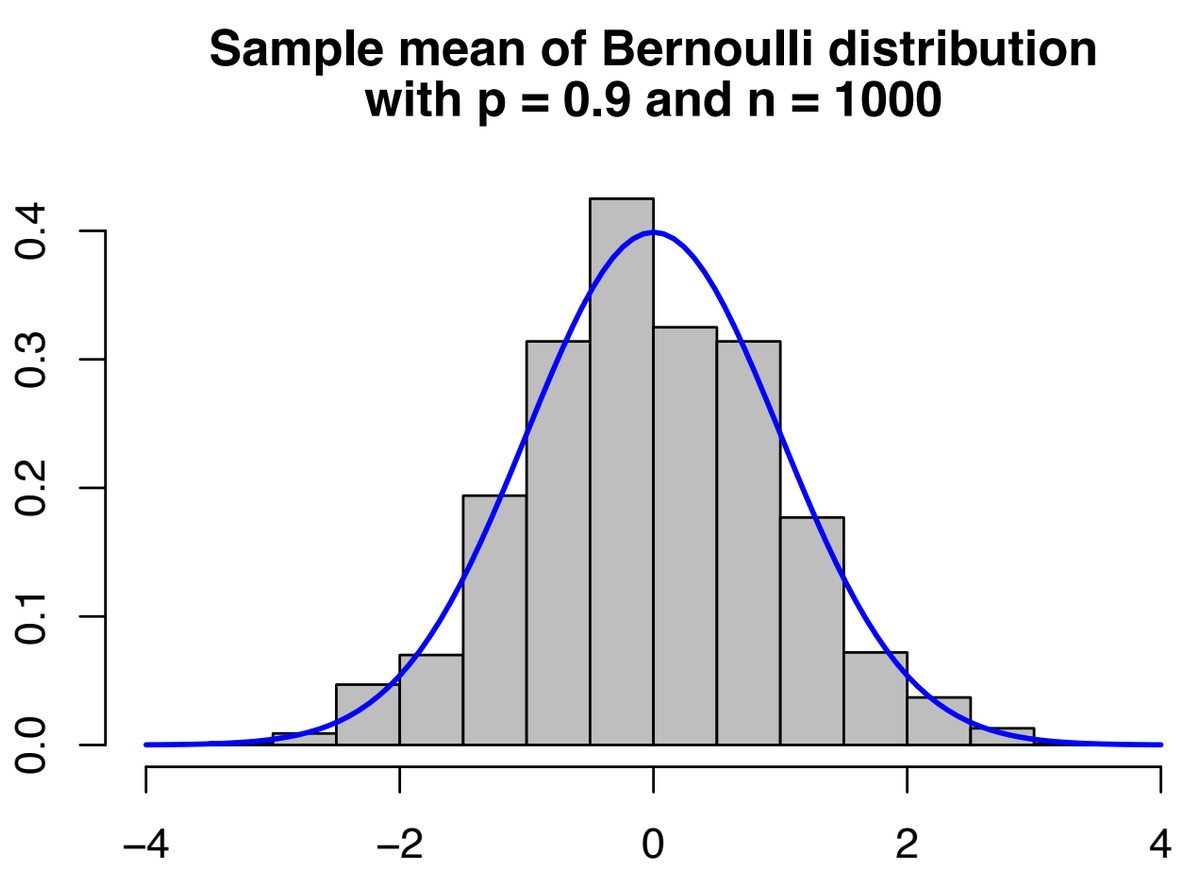

Sample mean of Bernoulli distribution
with p = 0.9 and n = 20
0.0
0.1
0.2
0.3
0.4
0.5
−4
−2
0
2
4


Sample mean of Bernoulli distribution
with p = 0.9 and n = 50
0.0
0.2
0.4
0.6
−4
−2
0
2
4


Sample mean of Bernoulli distribution
with p = 0.9 and n = 100
0.0
0.1
0.2
0.3
0.4
0.5
−4
−2
0
2
4


Sample mean of Bernoulli distribution
with p = 0.9 and n = 200
0.0
0.1
0.2
0.3
0.4
0.5
−4
−2
0
2
4


Sample mean of Bernoulli distribution
with p = 0.9 and n = 500
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4


Sample mean of Bernoulli distribution
with p = 0.9 and n = 1000
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

课后练习（不需要提交）

- 在下列各条件下，尝试利用 Excel 模拟中心极限定理。
  1. 总体服从 $p = 0.9$ 的伯努利分布。
  2. 总体服从 $[0, 5]$ 上的连续均匀分布。
  3. 总体服从 $\lambda = 1$ 的指数分布。
     指数分布是 c.d.f. 为 $F(x) = 1 - e^{-\lambda x}$ 的非负值连续分布。

- 分别观察各条件下达到理想近似效果所需的 n 的值。

扩展阅读

- 对于想巩固统计学所需要的概率知识的同学
  - Morris DeGroot & Mark Schervish, *Probability and Statistics*, Pearson.

- 对于想从理论层面全面学习概率论的同学
  - Sheldon Ross, *A First Course in Probability*, Pearson.