# 《初等概率论》第 11 讲

邓 婉 璐

清华大学
统计学研究中心

November 2, 2018

# 独立性的判定

设随机向量 $\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$ 有分布函数 $F(\mathbf{x}_1, \mathbf{x}_2)$、密度函数 $f(\mathbf{x}_1, \mathbf{x}_2)$、特征函数 $\phi(\mathbf{t}_1, \mathbf{t}_2)$；又设 $\mathbf{X}_i$ 的分布函数、密度函数、特征函数分别为 $F_i(\mathbf{x}_i)$, $f_i(\mathbf{x}_i)$ 和 $\phi_i(\mathbf{t}_i)$.

**定理 1.1**

随机向量 $\mathbf{X}_1$ 和 $\mathbf{X}_2$ 相互独立的充要条件是下列条件之一成立：

1. $F(\mathbf{x}_1, \mathbf{x}_2) = F_1(\mathbf{x}_1)F_2(\mathbf{x}_2)$;
2. $f(\mathbf{x}_1, \mathbf{x}_2) = f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)$;
3. $\phi(\mathbf{t}_1, \mathbf{t}_2) = \phi_1(\mathbf{t}_1)\phi_2(\mathbf{t}_2)$.

# 多元正态分布

**定义 2.1 (multivariate normal distribution)**

设 $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_n)^{\mathrm{T}}$ 是 $n$ 维常数列向量，$\mathbf{B}$ 是 $n \times m$ 常数矩阵，$\varepsilon_1, \varepsilon_2, ..., \varepsilon_m$ 是相互独立且服从标准正态分布的随机变量. 如果

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\varepsilon},$$

其中 $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_m)^{\mathrm{T}}$，就称 $\mathbf{X} = (X_1, X_2, ..., X_n)^{\mathrm{T}}$ 服从$n$-维正态分布，记作 $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{B}\mathbf{B}^{\mathrm{T}})$. 特别地，当矩阵 $\mathbf{B}\mathbf{B}^{\mathrm{T}}$ 退化时，还称相应的正态分布是退化的正态分布.

♣ 如果 $\mathbf{X}$ 服从多元正态分布，则 $\mathbf{X}$ 的任何分量 $(X_{j_1}, ..., X_{j_k})^{\mathrm{T}}$ 也服从（多元）正态分布.

♣ $\mathrm{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})$ 是单位矩阵 $\mathbf{I}$, 所以

$$E(\mathbf{X}) = \boldsymbol{\mu} + \mathbf{B}\,E(\boldsymbol{\varepsilon}) = \boldsymbol{\mu};$$

$$\Sigma \stackrel{\text{def}}{:=} E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathrm{T}}] = \mathbf{B}\mathbf{I}\mathbf{B}^{\mathrm{T}} = \mathbf{B}\mathbf{B}^{\mathrm{T}}.$$

# 多元正态分布

♣ $\varepsilon$ 的特征函数

$$\phi_{\boldsymbol{\varepsilon}}(\mathbf{t}) = E\exp(\mathrm{i}\mathbf{t}^{\mathrm{T}}\boldsymbol{\varepsilon}) = \prod_{j=1}^{n} \exp\left(-t_j^2/2\right) = \exp(-\mathbf{t}^{\mathrm{T}}\mathbf{t}/2),$$

其中 $\mathbf{t} = (t_1, t_2, ..., t_n)^{\mathrm{T}} \in R^n$.

♣ $\mathbf{X}$ 的特征函数

$$\begin{aligned}
\phi_{\mathbf{X}}(\mathbf{t}) &= E\exp(\mathrm{i}\mathbf{t}^{\mathrm{T}}\mathbf{X}) = E\exp\left[\mathrm{i}(\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu} + \mathbf{t}^{\mathrm{T}}\mathbf{B}\boldsymbol{\varepsilon})\right] \\
&= \exp(\mathrm{i}\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu})E\exp\left[\mathrm{i}(\mathbf{t}^{\mathrm{T}}\mathbf{B})\boldsymbol{\varepsilon}\right] \\
&= \exp\left[\mathrm{i}\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu} - \frac{1}{2}(\mathbf{t}^{\mathrm{T}}\mathbf{B})(\mathbf{t}^{\mathrm{T}}\mathbf{B})^{\mathrm{T}}\right] \\
&= \exp\left[\mathrm{i}\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^{\mathrm{T}}\Sigma\mathbf{t}\right].
\end{aligned}$$

♣ 从中可以看出，$\mathbf{X}$ 的期望和协方差矩阵唯一决定了 $\mathbf{X}$ 的特征函数. 由于随机向量的特征函数与分布函数是相互唯一决定的，所以 $\mathbf{X}$ 的分布由 $\boldsymbol{\mu}$ 和 $\Sigma$ 唯一决定. 用 $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ 表示 $\mathbf{X}$ 服从均值是 $\boldsymbol{\mu}$，协方差矩阵是 $\Sigma$ 的正态分布.

# 多元正态分布

**定理 2.1**

$\mathbf{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ 的充要条件 是对任何 $\mathbf{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}} \in R^n$,

$$Y := \mathbf{a}^{\mathrm{T}}\mathbf{X} \sim \mathcal{N}(\mathbf{a}^{\mathrm{T}}\boldsymbol{\mu},\, \mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a}).$$

证明. $(\Longrightarrow)$. $Y$ 的特征函数

$$\begin{aligned}\phi_Y(t) &= E\exp(\mathrm{i}tY) = E\exp\left[\mathrm{i}(t\mathbf{a}^{\mathrm{T}})\mathbf{X}\right] \\ &= \exp\left[\mathrm{i}t(\mathbf{a}^{\mathrm{T}}\boldsymbol{\mu}) - \tfrac{1}{2}t^2\mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a}\right]. \qquad (1)\end{aligned}$$

所以 $Y \sim \mathcal{N}(\mathbf{a}^{\mathrm{T}}\boldsymbol{\mu},\, \mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a})$.

$(\Longleftarrow)$ 在 (1) 中取 $t = 1$ 得

$$E\exp(\mathrm{i}\mathbf{a}^{\mathrm{T}}\mathbf{X}) = \exp\left[\mathrm{i}\mathbf{a}^{\mathrm{T}}\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a}\right].$$

故 $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

# 多元正态分布

**定理 2.2**

如果 $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$，则对任意常数矩阵 $\mathbf{A}$ 和常数向量 $\mathbf{b}$，只要 $\mathbf{b}+\mathbf{AX}$ 有意义，则 $\mathbf{Y} = \mathbf{b}+\mathbf{AX} \sim \mathcal{N}(\mathbf{b}+\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}^{\mathrm{T}})$.

**证明.** $\mathbf{Y} = \mathbf{b} + \mathbf{AX} = (\mathbf{b} + \mathbf{A}\boldsymbol{\mu}) + (\mathbf{AB})\varepsilon$，即 $\mathbf{Y}$ 服从多元正态分布. 计算其均值和协方差即可得.

**定理 2.3**

设 $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$，如果

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix},$$

且 $\mathbf{X}_1, \boldsymbol{\mu}_1$ 和方阵 $\Sigma_{11}$ 的行数相同，则 $\mathbf{X}_1$ 和 $\mathbf{X}_2$ 独立，而且

$$\mathbf{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \Sigma_{11}), \quad \mathbf{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_{22}).$$

# 多元正态分布

证明. $\mathbf{X}$ 的特征函数

$$
\begin{aligned}
\phi(\mathbf{t}) = \phi(\mathbf{t}_1, \mathbf{t}_2) &= \exp\left[\mathrm{i}\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}^{\mathrm{T}}\Sigma\mathbf{t}\right] \\
&= \exp\left[\mathrm{i}\mathbf{t}_1^{\mathrm{T}}\boldsymbol{\mu}_1 - \tfrac{1}{2}\mathbf{t}_1^{\mathrm{T}}\Sigma_{11}\mathbf{t}_1 + \mathrm{i}\mathbf{t}_2^{\mathrm{T}}\boldsymbol{\mu}_2 - \tfrac{1}{2}\mathbf{t}_2^{\mathrm{T}}\Sigma_{22}\mathbf{t}_2\right] \\
&= \exp\left[\mathrm{i}\mathbf{t}_1^{\mathrm{T}}\boldsymbol{\mu}_1 - \tfrac{1}{2}\mathbf{t}_1^{\mathrm{T}}\Sigma_{11}\mathbf{t}_1\right] \times \exp\left[\mathrm{i}\mathbf{t}_2^{\mathrm{T}}\boldsymbol{\mu}_2 - \tfrac{1}{2}\mathbf{t}_2^{\mathrm{T}}\Sigma_{22}\mathbf{t}_2\right] \\
&= \phi_1(\mathbf{t}_1)\phi_2(\mathbf{t}_2).
\end{aligned}
$$

故结论成立.

# 多元正态分布

**定理 2.4**

如果 $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$，则 $(X_1, X_2, ..., X_n)$ 相互独立的充要条件是 $\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2)$.

**定理 2.5**

当 $\Sigma$ 正定时，$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ 有联合密度函数

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi}\,)^n \sqrt{\det(\Sigma)}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right].$$

证明. 因为 $\Sigma$ 是正定的，所以存在可逆方阵 $\mathbf{B}$ 使得 $\Sigma = \mathbf{B}\mathbf{B}^{\mathrm{T}}$，且 $\mathbf{X} = \boldsymbol{\mu} + \mathbf{B}\varepsilon$，其中 $\varepsilon \sim \mathcal{N}(0, \mathbf{I})$. 易得 $\varepsilon$ 的密度函数为

$$f_\varepsilon(\mathbf{y}) = \frac{1}{(\sqrt{2\pi}\,)^n} \exp\left(-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y}\right).$$

注意到 $\{\mathbf{X} = \mathbf{x}\} = \{\varepsilon = \mathbf{B}^{-1}(\mathbf{x}-\boldsymbol{\mu})\}$，且映射:

$$\mathbf{y} = \mathbf{B}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

# 多元正态分布

是可逆的，其雅克比行列式为

$$\left|\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right| = |\mathbf{B}^{-1}| = \frac{1}{\sqrt{\det(\Sigma)}}.$$

因此, $\mathbf{X}$ 的密度函数为：

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{(\sqrt{2\pi}\,)^n} \exp\left\{-\tfrac{1}{2}[\mathbf{B}^{-1}(\mathbf{x}-\boldsymbol{\mu})]^{\mathrm{T}}[\mathbf{B}^{-1}(\mathbf{x}-\boldsymbol{\mu})]\right\} \left|\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right| \\ &= \frac{1}{(\sqrt{2\pi}\,)^n\sqrt{\det(\Sigma)}} \exp\left[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]. \end{aligned}$$

**定理 2.6**

设 $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$，$\det(\Sigma) > 0$ 和分块矩阵

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

其中 $\mathbf{X}_1, \boldsymbol{\mu}_1$ 和方阵 $\Sigma_{11}$ 的行数相同，则在条件 $\mathbf{X}_1 = \mathbf{x}_1$ 下，$\mathbf{X}_2$ 服从多元正态分布

$$\mathcal{N}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1),\ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right).$$

**证明.** 令

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} I_{11} & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{22} \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 - \boldsymbol{\mu}_1 \\ \mathbf{X}_2 - \boldsymbol{\mu}_2 \end{pmatrix}$$
$$\sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}\right)$$

故 $\mathbf{Y}_1$ 与 $\mathbf{Y}_2$ 独立. 注意到 $\mathbf{Y}_2 = \mathbf{X}_2 - \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{Y}_1$,

# 多元正态分布

想法就是构造新的与条件独立的随机变量，使得条件分布即为分布！！

所以

$$\mathbf{X}_2 = \mathbf{Y}_2 + \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1).$$

从而，

$$\begin{aligned}
&\mathbb{P}(\mathbf{X}_2 \le \mathbf{x}_2 | \mathbf{X}_1 = \mathbf{x}_1) \\
&= \mathbf{P}(\mathbf{Y}_2 + \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \le \mathbf{x}_2 | \mathbf{X}_1 = \mathbf{x}_1) \\
&= \mathbf{P}(\mathbf{Y}_2 + \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \le \mathbf{x}_2 | \mathbf{Y}_1 = \mathbf{x}_1 - \boldsymbol{\mu}_1) \\
&= \mathbf{P}(\mathbf{Y}_2 + \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \le \mathbf{x}_2).
\end{aligned}$$

又 $\mathbf{Y}_2 \sim \mathcal{N}(0, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$，故

$$\begin{aligned}
\mathbf{X}_2|_{\mathbf{X}_1=\mathbf{x}_1} &\overset{d}{=} \mathbf{Y}_2 + \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \\
&\sim \mathcal{N}\Big(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1),\ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Big).
\end{aligned}$$

## ♣ A. 收敛模式 (mode of convergence)

设 $(\Omega, \mathscr{F}, \mathbb{P})$ 是概率空间，$X_n$ 和 $X$ 是随机变量，其分布函数分别为 $F_n(x)$ 和 $F(x)$，即

$$F_n(x) = \mathbb{P}(X_n \le x), \quad F(x) = \mathbb{P}(X \le x).$$

**定义 3.1 (数列的收敛)**

设 $\{a_i\}$ 是实数列，$a$ 为一实数，如果对 $\forall \varepsilon > 0$，存在正整数 $N$，使得当 $n \ge N$ 时，有

$$|a_n - a| \le \varepsilon,$$

则称数列 $a_n$ 收敛于 $a$，记作 $\lim_{n\to\infty} a_n = a$.

# 收敛性

**定义 3.2 (convergence in distribution)**

如果在 $F(x)$ 的连续点 $x$, 有

$$\lim_{n\to\infty} F_n(x) = F(x),$$

则称 $X_n$依分布收敛到 $X$(convergence in distribution), 记作 $X_n \xrightarrow{d} X$, 或者称 $F_n$ 弱收敛到 $F$ (weak convergence), 记作 $F_n \xrightarrow{w} F$.

**定义 3.3 (convergence in probability)**

如果对 $\forall \varepsilon > 0$, 有

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0,$$

则称 $X_n$依概率收敛到 $X$(convergence in probability), 记作 $X_n \xrightarrow{p} X$, 或 $X_n \to X$ in prob.

# 收敛性

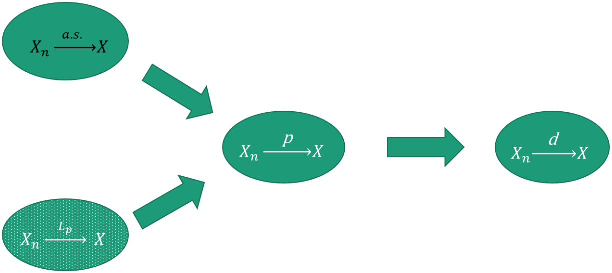
**定义 3.4 (almost sure convergence)**

如果
$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1,$$
则称 $X_n$几乎处处收敛到 $X$(almost sure convergence)，记作 $X_n \xrightarrow{\text{a.s.}} X$, 或 $X_n \to X$ a.s.

**定义 3.5 ($L_p$ convergence)**

对 $p>0$, 如果
$$\lim_{n\to\infty} E(|X_n - X|^p) = 0,$$
则称 $X_n$在 $L_p$ 下收敛到 $X$, 记作 $X_n \xrightarrow{L_p} X$, 或 $X_n \to X$ in $L_p$.

## ♣ B. 各种收敛之间的关系


$X_n \xrightarrow{a.s.} X$
$X_n \xrightarrow{L_p} X$
$X_n \xrightarrow{p} X$
$X_n \xrightarrow{d} X$

# 收敛性

**定理 3.1 (a.s. implies in prob.)**

如果 $X_n \xrightarrow{a.s.} X$，则 $X_n \xrightarrow{p} X$.

**证明.** 注意到，对 $\forall \varepsilon > 0$,

$$\left\{\lim_{n\to\infty} X_n = X\right\} = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\{|X_k - X| < \varepsilon\}.$$

因此，利用概率的连续型可得

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) &\leq \lim_{n\to\infty} \mathbb{P}(\cup_{k=n}^{\infty}|X_k - X| \geq \varepsilon) \\
&= \mathbb{P}\Big(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{|X_k - X| \geq \varepsilon\}\Big) \\
&= 1 - \mathbb{P}\Big(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\{|X_k - X| < \varepsilon\}\Big) \\
&= 1 - 1 = 0.
\end{aligned}$$

# 收敛性

**定理 3.2 ($L_p$ implies in prob.)**

如果 $X_n \xrightarrow{L_p} X$, 则 $X_n \xrightarrow{p} X$.

**证明.** 对 $\forall \varepsilon > 0$, 利用 Markov 不等式可得.

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) \leq \lim_{n\to\infty} \frac{1}{\varepsilon^p} E(|X_n - X|^p) = 0.$$

**定理 3.3 (in prob. implies in dist.)**

如果 $X_n \xrightarrow{p} X$, 则 $X_n \xrightarrow{d} X$.

**证明.** 令 $F_n(x) = \mathbb{P}(X_n \leq x)$ 和 $F(x) = \mathbb{P}(X \leq x)$. 首先,

$$\begin{aligned} F_n(x) &= \mathbb{P}(X_n \leq x, |X_n - X| \leq \varepsilon) + \mathbb{P}(X_n \leq x, |X_n - X| > \varepsilon) \\ &\leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon) \\ &= F(x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon). \end{aligned}$$

另一方面，

$$
\begin{aligned}
&F_n(x)\\
&= 1-\mathbb{P}(X_n > x)\\
&= 1-\mathbb{P}(X_n > x, |X_n - X| \le \varepsilon) - \mathbb{P}(X_n > x, |X_n - X| > \varepsilon)\\
&\ge 1-\mathbb{P}(X > x-(X_n - X), |X_n - X| \le \varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon)\\
&\ge 1-\mathbb{P}(X > x-\varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon)\\
&= F(x-\varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon).
\end{aligned}
$$

因此，

$$
\begin{aligned}
F(x-\epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \le F_n(x)&\\
&\le F(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon).
\end{aligned}
$$

令 $n \to \infty$,

$$
F(x-\epsilon) \le \liminf_n F_n(x) \le \limsup_n F_n(x) \le F(x+\epsilon).
$$

# 收敛性

如果 $F(x)$ 在 $x$ 处连续，则当 $\epsilon \downarrow 0$，有 $F(x-\epsilon) \uparrow F(x)$ 和 $F(x+\epsilon) \downarrow F(x)$. 所以

$$\lim_{n\to\infty} F_n(x) = F(x),$$

即，$X_n \xrightarrow{d} X$.

# 收敛性

## ♣ C. 反例

**例 3.1**

$$X_n \xrightarrow{d} X \nRightarrow X_n \xrightarrow{p} X.$$

解. $\{X_n\}$ 独立同分布 $\mathcal{N}(0,1)$, 则 $X_n \xrightarrow{d} X_1$. 但是

$$\mathbb{P}(|X_n - X_1| > 1) = \mathbb{P}(|\mathcal{N}(0,1)| > \frac{1}{\sqrt{2}}) > 0.$$

**例 3.2**

$X_n \xrightarrow{d} C \Longleftrightarrow X_n \xrightarrow{p} C$, 其中 $C$ 为常数.

# 收敛性

**例 3.3**

$$X_n \xrightarrow{p} X \nRightarrow X_n \xrightarrow{L_p} X.$$

**解.** 取概率空间 $(\Omega, \mathscr{F}, \mathbb{P}) = ((0,1), \mathscr{B}((0,1)), \lambda)$. 定义

$$X_n(\omega) = \begin{cases} n, & \text{if } \omega \in (0, n^{-p}], \\ 0, & \text{if } \omega \in (n^{-p}, 1), \end{cases} \tag{2}$$

且 $X \equiv 0$. 则, 对 $\forall \varepsilon > 0$, 当 $n \to \infty$,

$$\begin{aligned} \mathbb{P}(|X_n - 0| > \varepsilon) &= \mathbb{P}(X_n > \varepsilon) = \mathbb{P}(X_n = n) \\ &= \mathbb{P}(\{\omega : \omega \in (0, n^{-p}]\}) \\ &= \lambda((0, n^{-p}]) = n^{-p} \to 0. \end{aligned}$$

但是 $E(|X_n - 0|^p) = 1$.

**例 3.4**

$$X_n \xrightarrow{L_p} X \nRightarrow X_n \xrightarrow{L_q} X \ (p < q).$$

# 收敛性

**例 3.5**

$$X_n \xrightarrow{\text{a.s.}} X \nRightarrow X_n \xrightarrow{L_p} X.$$

解. $X_n$ 的定义见 (2). 对 $\forall \omega \in \Omega$, 都有 $X_n(\omega) \to 0 \equiv X$. 但 $E(|X_n - 0|^p) = 1$.

**例 3.6**

$$X_n \xrightarrow{L_p} X \nRightarrow X_n \xrightarrow{\text{a.s.}} X.$$

解. 假设 $\{X_n\}$ 独立, 且 $X_n$ 定义如下:

$$\mathbb{P}(X_n = 0) = 1 - n^{-1}, \quad \mathbb{P}(X_n = 1) = n^{-1}.$$

则 $X_n \xrightarrow{L_1} 0$. 但是 $X_n \overset{\text{a.s.}}{\nrightarrow} 0$. 令 $A_n = \{X_n = 1\}$, 则

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

由 Borel-Cantelli 引理, “$\{A_n\}$ 有无穷多个发生” 的概率为 1.

# 收敛性

**定理 3.4 (Continuous mapping theorem)**

设 $\{X, X_n\}$ 是随机元序列，$g$ 连续，则

1. $X_n \xrightarrow{\text{a.s.}} X \Longrightarrow g(X_n) \xrightarrow{\text{a.s.}} g(X)$.
2. $X_n \xrightarrow{p} X \Longrightarrow g(X_n) \xrightarrow{p} g(X)$.
3. $X_n \xrightarrow{d} X \Longrightarrow g(X_n) \xrightarrow{d} g(X)$.

♣ 但是 $X_n \xrightarrow{L_p} X \not\Rightarrow g(X_n) \xrightarrow{L_p} g(X)$.

# 收敛性

**定理 3.5 (Slutsky's theorem)**

假设 $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{p} c$，则

1. $X_n + Y_n \xrightarrow{d} X + c$;
2. $X_n Y_n \xrightarrow{d} cX$;
3. $X_n / Y_n \xrightarrow{d} X/c$, $(c \neq 0)$.

证明. 【方法 1】逐条证明之.

【方法 2】(Coupling method) 先证明 $(X_n, Y_n) \xrightarrow{d} (X, c)$，然后应用 "Continuous mapping theorem" $g(x, y) = x + y$; $xy$; $x/y$, 即可得到结论.

# 大数定律

♣ 样本均值估计总体均值靠不靠谱？

假定 $X_1, X_2, \ldots,$ 是随机变量序列，令 $S_n = X_1 + \ldots + X_n$.

**Questions**

1. $\frac{S_n}{b_n} - a_n \xrightarrow{p} 0$ 的（充要）条件是什么？
2. $\frac{S_n}{b_n} - a_n \xrightarrow{\text{a.s.}} 0$ 的（充要）条件是什么？

其中 $\{a_n\}, \{b_n\}$ 是非随机序列，且 $0 < b_n \uparrow \infty$.

♣ 极限理论 (LLN, CLT, etc.) 的用处很多：

1. 为期望（或概率）和试验序列之间的联系提供了合理的解释；
2. 为序列和提供近似分布；
3. 在统计推断中发挥主要作用.

# 大数定律

**定理 4.1 (Weak law of large numbers，WLLN)**

设 $\{X_n\}$ 是独立同分布的随机变量序列，则存在实数列 $\{a_n\}$ 使得

$$\frac{1}{n}\sum_{k=1}^{n} X_k - a_n \xrightarrow{p} 0$$

的充分必要条件是

$$n\mathbb{P}(|X_1| \geq n) \to 0.$$

此时，可取 $a_n = E(X_1 I_{\{|X_1|<n\}})$.

**推论 4.1**

设 $\{X_n\}$ 是 i.i.d.，且 $E|X_1| < \infty$，则

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{p} EX_1. \tag{3}$$

# 大数定律

令 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $a = EX_1$. 则 (3) 等价于：对 $\forall\varepsilon$,

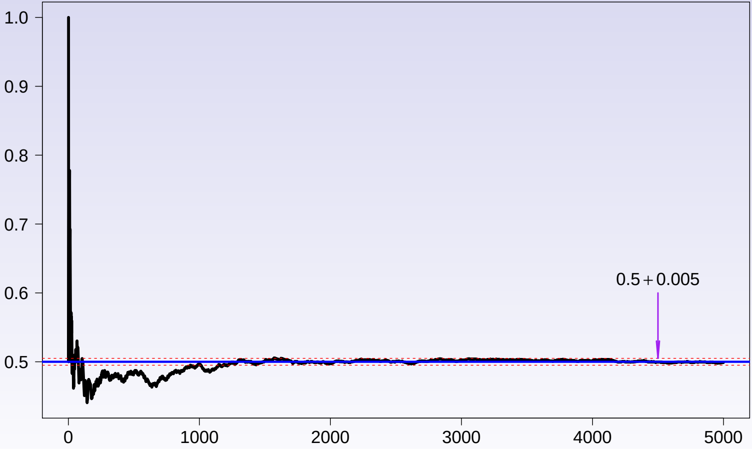
$$\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - a| \geq \varepsilon) = 0.$$

即：不论给定怎么小的 $\varepsilon > 0$, $\bar{X}_n$ 与 $a$ 偏离有否可能达到 $\varepsilon$ 或者更大呢？这是有可能！但是当 $n$ 很大时，出现这种较大偏差的可能性很小，以致于当 $n$ 很大时，我们有很大的 (然而不是百分之百的) 把握断言 $\bar{X}_n$ 很接近 $a$.

比如独立同分布的 $X_i \sim B(1, 0.5)$ (掷硬币), $i = 1, 2, \ldots$. 则下图给出了 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ 的一次实现.

# 大数定律





**例 4.1**

假设人群中学习过统计学的比例为 $f$. 我们随机抽取 $n$ 个人，用 $X_i$ 表示第 $i$ 个人是否学习过统计学：即如果学过，则 $X_i = 1$; 否则 $X_i = 0$. 则这 $n$ 个人的样本中学习过统计学的比例为 $M_n = (X_1 + \ldots + X_n)/n$. 我们如何通过 $M_n$ 得到对 $f$ 的认识？

根据 WLLN，我们可以用 $M_n$ 来近似估计 $f$. 不妨设我们的目标是得到误差率小于 1% 的 95% 置信区间，即:

$$P(|M_n - f| \geq 0.01) \leq 0.05.$$

由切比雪夫不等式 (Chebyshev's inequality):

$$P(|M_n - f| \geq 0.01) \leq \frac{\sigma^2_{M_n}}{0.01^2} = \frac{\sigma^2_x}{n(0.01)^2} \leq \frac{1}{4n(0.01)^2}.$$

所以我们可以找一个保守的方法，即抽取 $n = 50,000$ 的人，就可保证 $P(|M_n - f| \geq 0.01) \leq 0.05$.

# 大数定律



**定理 4.2 (Strong law of large numbers，SLLN)**

设 $\{X_n\}$ 是独立同分布的随机变量序列，则

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} C$$

的充分必要条件是 $EX_1$ 存在，且 $EX_1 = C$.

♣ 直观区别：

1. 弱大数定律想证明：采样的次数越多，平均值接近真实期望值的可能性越来越大；
2. 强大数定律想证明：采样的次数越多，平均值一定越来越接近真实期望值；

# 大数定律

**定理 4.3 (Marcinkiewicz SLLN(选修))**

设 $\{X_n\}$ 是独立同分布的随机变量序列，$0 < r < 2$，则

$$\frac{1}{n^r}\sum_{k=1}^{n}(X_k - a) \xrightarrow{\text{a.s.}} 0$$

当且仅当是 $E(|X_1|^r) < \infty$ 存在，其中 $a = (EX_1)I_{\{1\le r<2\}} + 0I_{\{0<r<1\}}$.

**定理 4.4 (Kolmogorov(选修))**

设 $\{X_n\}$ 是独立的随机变量序列，有相同的期望 $\mu = EX_k$,

如果 $\displaystyle\sum_{k=1}^{\infty}\frac{\mathrm{var}(X_k)}{k^2} < \infty,$ 则 $\displaystyle\frac{1}{n}\sum_{k=1}^{n}X_k \xrightarrow{\text{a.s.}} \mu.$

# 中心极限定理



假定 $X_1, X_2, \ldots,$ 是一列随机变量序列，令 $S_n = X_1 + \ldots + X_n$.

**Questions**

1. $\frac{S_n}{b_n} - a_n \xrightarrow{d} Y$ 的（充要）条件是什么？

其中 $\{a_n\}, \{b_n\}$ 是非随机序列，且 $0 < b_n \uparrow \infty$; $Y$ 是非退化的随机变量.

**定理 5.1 (Central limit theorem, CLT)**

设 $\{X_n\}$ 是独立同分布的随机变量序列，其期望为 $\mu$, 方差为 $\sigma^2 < \infty$，则

$$\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu) \xrightarrow{d} \mathcal{N}(0,1).$$

♣ 虽然在一般情况下很难求出 $X_1 + \ldots + X_n$ 的分布的确切形式，但 $n$ 较大时，可以通过 $\Phi(\cdot)$ 给出其近似值.

# 中心极限定理

证明. 令 $Y_k = (X_k - \mu)/\sigma$. 则 $\{Y_k\}$ 是独立同分布，其期望为 0，方差为 1. 用 $\phi(t) = E\mathrm{e}^{\mathrm{i}tY_1}$ 表示 $Y_1$ 的特征函数. 注意到

$$\phi'(0) = \mathrm{i}EY_1 = 0, \quad \phi''(0) = \mathrm{i}^2 EY_1^2 = -1.$$

利用 Taylor 展开，可得

$$\phi(t) = \phi(0) + \phi'(0)t + \tfrac{1}{2}\phi''(0)t^2 + o(t^2).$$

所以 $\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^n (X_k - \mu)$ 的特征函数为

$$\begin{aligned}\phi_n(t) &= E\exp(\mathrm{i}t(Y_1 + \ldots + Y_n)/\sqrt{n}) \\ &= [\phi(t/\sqrt{n})]^n \\ &= \left(1 - \frac{t^2}{2n} + o(t^2/n)\right)^n \\ &\to \mathrm{e}^{-t^2/2}, \quad (n \to \infty)\end{aligned}$$

由连续性定理 (Continuity theorem) 可知，结论成立.

# 中心极限定理

**例 5.1**

假设人群中学习过统计学的比例为 $f$. 我们随机抽取 $n$ 个人，用 $X_i$ 表示第 $i$ 个人是否学习过统计学：即如果学过，则 $X_i = 1$; 否则 $X_i = 0$. 则这 $n$ 个人的样本中学习过统计学的比例为 $M_n = (X_1 + ... + X_n)/n$. 我们如何通过 $M_n$ 得到对 $f$ 的认识？

作为对比，我们的目标仍然是得到误差率小于 1% 的 95% 置信区间，即：$P(|M_n - f| \geq 0.01) \leq 0.05$.
则利用中心极限定理，我们可以得到比切比雪夫不等式更精确的推断：

$$P(|M_n - f| \geq 0.01) = P(|\frac{X_1 + ... + X_n - nf}{\sqrt{n}\sigma}| \geq \frac{0.01\sqrt{n}}{\sigma_x})$$

$$\approx P(|Z| \geq \frac{0.01\sqrt{n}}{\sigma_x}) \leq P(|Z| \geq 0.02\sqrt{n}),$$

其中 $Z \sim \mathcal{N}(0,1)$. 取 $n = (\frac{1.96}{0.02})^2 = 9604$ 即可.

# 中心极限定理

♣ 应用到二项分布上

设 $X_i \sim B(1,p)$, $i = 1, 2, \ldots$, $i.i.d.$. 令 $S_n = X_1 + \ldots + X_n$.

例如，$n = 36, p = 0.5$.

- $P(S_n \leq 21) = P(S_n < 22)$. 用哪个做正态估计，21 还是 22？
- 单点估计为 0? $P(S_n = 19) \approx 0$

# 中心极限定理

## 定理 5.2 (de Moivre-Laplace, CLT, 1716)

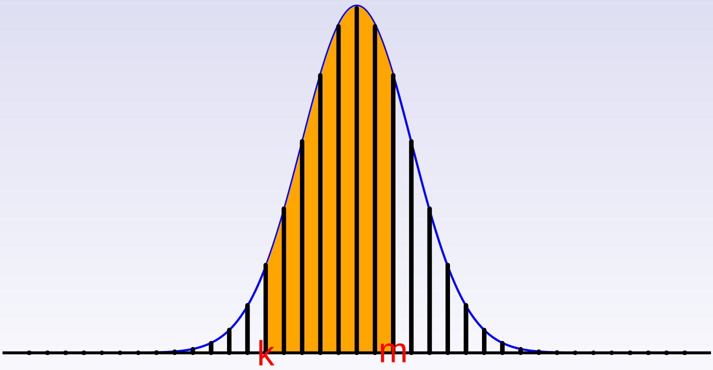
设 $S_n \sim B(n, p)$, $n$ 充分大, $k$ 和 $m$ 是非负整数, 则

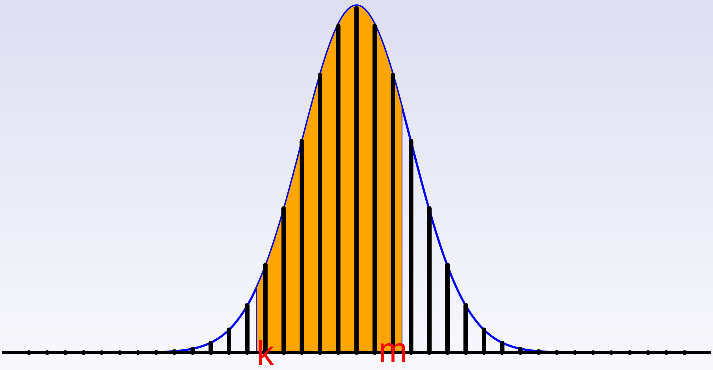
$$\mathbb{P}(k \le S_n \le m) \approx \Phi\left(\frac{m-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right).$$

其修正形式【近似更为精确】:

$$\mathbb{P}(k \le S_n \le m) \approx \Phi\left(\frac{m+0.5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-0.5-np}{\sqrt{np(1-p)}}\right).$$

更重要的是: 修正形式可以计算单点概率 $\mathbb{P}(S_n = k)$.

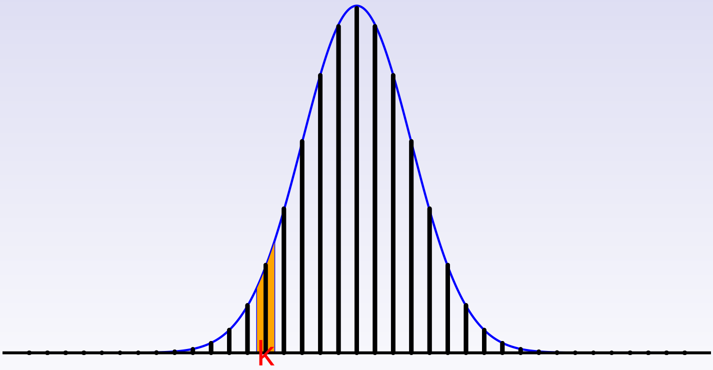

k
m



k
m

# 中心极限定理





# 中心极限定理

**例 5.2**

设 $S_n \sim B(36, 0.5)$，求概率 $\mathbb{P}(S_n \le 21)$.

**解.** 其精确的概率为：

$$\mathbb{P}(S_n \le 21) = \sum_{k=0}^{21} \binom{36}{k} 0.5^{36} = 0.8785.$$

利用中心极限定理，若端点不经修正，则近似概率为：

$$\mathbb{P}(S_n \le 21) \approx \Phi\left(\frac{21 - np}{\sqrt{np(1-p)}}\right) = \Phi((21 - 18)/3) = 0.8413.$$

修正之后的近似为：

$$\mathbb{P}(S_n \le 21) \approx \Phi\left(\frac{21.5 - np}{\sqrt{np(1-p)}}\right) = \Phi(1.17) = 0.8789995.$$

另：$\mathbb{P}(S_n = 19) = 0.1251$, $\mathbb{P}(S_n = 19) \approx 0.124$.

**例 5.3**

设某地区内原有一家小型电影院，因不敷需要，拟筹建一所较大型的. 设据分析，该地区每日平均看电影者约有 $n = 1600$ 人，且预计新电影院建成开业后，平均约有 $3/4$ 的观众将去这新电影院。现该影院在计划其座位数时，要求座位数尽可能多，但“空座达到 200 或更多”的概率又不超过 10%. 问设多少座位为好?

**解**. 设把每日看电影的人排号为 1,2,...,1600，且令 $X_i = 1$ 如果第 $i$ 个观众去新电影院，否则 $= 0$，$i = 1, \ldots, 1600$. 则按照假定有 $\mathbb{P}(X_i = 1) = 3/4$, $\mathbb{P}(X_i = 0) = 1/4$. 又假定各观众去不去电影院是独立选择，则 $X_1, X_2, \ldots, X_{1600}$ 是独立随机变量. 现设座位数为 $m$, 则按要求

$$\mathbb{P}(X_1 + \cdots + X_{1600} \le m - 200) \le 0.1$$

在这个条件下，取 $m$ 最大. 这显然就是在上式取等号时，因为 $np = 1600 \times 3/4 = 1200$, $\sqrt{np(1-p)} = 10\sqrt{3}$, 所以，

# 中心极限定理

按照修正近似公式，$m$ 应满足条件

$$\Phi\big((m-200+0.5-1200)/(10\sqrt{3})\big)=0.1$$

当 $\Phi(x)=0.1$ 时，$x=-1.2816$. 由

$$\frac{m-200+0.5-1200}{10\sqrt{3}}=-1.2816$$

得 $m=1377.31\approx 1377.$

♣ 从上述定理与实例中可以看出，当 $n\to\infty$, 正态近似就会越精确，但是在实践中，样本量 $n$ 是固定的，有限的. 所以须知道 $n$ 多大时正态近似的结果是可信的. 很遗憾，没有简单普遍的准则来判断. 这要依赖于 $X_i$ 的分布是否与正态分布接近，特别地，还依赖于 $X_i$ 的分布是否对称. 比如 $X_i\sim\mathcal{U}(-1,1)$ 或 $\mathcal{E}(\lambda)$. 进一步，使用正态近似计算 $\mathbb{P}(S_n\le c)$ 的时候，其近似程度与 $c$ 的值有关，一般说来，如果 $c$ 在 $S_n$ 均值的附近，其精度会更高一些.

## ♣ 二项分布的泊松估计与正态估计

回忆之前讲过的二项分布可用泊松分布近似（矩母函数推导）.

- 令 $n \to \infty$，应用 CLT?
- 说明泊松分布 = 正态分布?

事实上，对二项分布 $B(n, p)$：

- 若 $p$ 固定，$n \to \infty$：可用正态分布近似.（若 $p$ 接近 $\frac{1}{2}$，则较小的 $n$ 就很近似了；若 $p$ 接近 0 或 1，则需较大的 $n$ 才有比较好的近似.
- 若 $np$ 固定，大小适当，$n \to \infty, p \to 0$：可用泊松分布近似.

实际应用中（举例）：

- 若 $p = 1/10, n = 500$：可用正态分布近似.
- 若 $p = 1/100, n = 100$：可用泊松分布近似.

**定理 5.3 (Lyapunov, CLT (选修))**

设 $\{X_n\}$ 是独立的随机变量序列，如果存在 $\delta > 0$，使得

$$\lim_{n\to\infty} \frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E(|X_k - EX_k|^{2+\delta}) = 0,$$

其中 $B_n^2 = \sum_{k=1}^{n} \mathrm{var}(X_k)$，则

$$\frac{1}{B_n} \sum_{k=1}^{n} (X_k - EX_k) \xrightarrow{d} \mathcal{N}(0,1).$$

# 中心极限定理

**定理 5.4 (Lindeberg-Feller, CLT (选修))**

设 $\{X_n\}$ 是独立的 r.v.s.，则：方差序列 $\{\sigma_k^2 := \text{var}(X_k)\}$ 满足

$$B_n^2 = \sum_{k=1}^{n} \text{var}(X_k) \to \infty, \quad \frac{\sigma_n^2}{B_n^2} \to 0 \tag{4}$$

并且中心极限定理

$$\frac{1}{B_n} \sum_{k=1}^{n} (X_k - EX_k) \xrightarrow{d} \mathcal{N}(0,1)$$

成立的充分必要条件是 Lindeberg 条件成立，即：对 $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} E\left\{(X_k - EX_k)^2 I_{\{|X_k - EX_k| > \varepsilon B_n\}}\right\} = 0.$$

♣ 条件 (4)$\Longleftrightarrow \lim\limits_{n\to\infty} \max\limits_{1\le k\le n} \frac{\sigma_k^2}{B_n^2} = 0$ (Feller 条件)

# 小结

知识点

- 极限定理（均值、分布）：大数定律、中心极限定理
- 不同收敛性的定义（不要求掌握相关定理的证明、不要求掌握 $L_p$ 收敛）
- 多种独立性的判定法则

技巧

- 用连续分布函数近似表达离散分布函数时，需适当修正
- 不等式控制偏差范围.
- 在不失一般性情况下 (WLOG)，利用标准化形式证明，可简化证明
- 利用随机变量特征函数（或矩母函数、概率母函数）的极限确认随机变量极限的分布

# 作业

打 * 的题目是选做，不算成绩，因而不必写入作业：

- 教材第 5 章 5, 6*; 9, 10, 11, 14*.
- 设 $(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$，且二维矩阵 $\Sigma$ 正定，求 $X+Y$ 与 $X-Y$ 独立的充分必要条件.
- 若 $X_n \xrightarrow{p} Y_1$, $X_n \xrightarrow{p} Y_2$，证明 $Y_1 = Y_2$, $a.s.$.
- 若 $X_n \xrightarrow{p} X$, $Y_n \xrightarrow{p} Y$，证明 $X_n + Y_n \xrightarrow{p} X + Y$. *
- 设全世界有 $n$ 个家庭，每个家庭有 $k$ 个小孩的概率都是 $p_k$. 设 $p_k$ 满足 $\sum_{k=0}^{c} p_k = 1$. 如果各个家庭的小孩数是相互独立的，计算一个小孩来自有 $k$ 个小孩的家庭的概率.
- 利用中心极限定理证明 $\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{2}$.

*Thank you!*