概　　率　　论

Probability

肖磊，2024年12月3日

已学知识点

- 第一章　事件与概率
  - 随机现象与统计规律性

    ① 概率的频率解释依然是当今最通行的解释.

    ② 描述频率趋近于概率的大数定律总是概率论的第一大数定律.

    ③ 实际当中用频率作为概率的估计是十分自然的.
  - 样本空间与事件

符号	集合论含义	概率论含义
Ω	空间或全集	样本空间或必然事件
Φ	空集	不可能事件
ω	元素	样本点
A	子集	随机事件
$\omega \in A$	ω 是 A 的元素	事件 A 包含样本点 ω
$A \subset B$	A 是 B 的子集	A 发生则 B 发生
$AB = \Phi$	A, B 不相交	A, B 不可能同时发生
$A \cup B$	并集	A, B 至少有一个发生
$A \cap B$	交集	A, B 同时发生
$A - B$	差集	A 发生而 B 不发生
$\bar{A}$	余集	A 不发生

已学知识点

- 第一章　事件与概率
  - 古典概型 (等可能概率模型)：(1) 样本空间样本点有限；(2) 每个样本点等可能出现.
    - 计数方法：排列组合.
    - 三个基本性质：非负性、规范性、有限可加性.
  - 几何概率：以等可能性定义概率，处理无限场合，概率是几何体的测度之比.
    - 三个基本性质：非负性、规范性、可列可加性.
  - 概率空间：$\left(\Omega, \mathscr{F}, P\right)$
    - 难点和要点：事件域 $\mathscr{F}$ 的选择，太小不能满足需要，太大难以定义概率.
    - 选择包含我们关注的所有事件的 σ 域，保证事件对交、并、逆、差作可列次运算的封闭性.
    - 在这种 σ 域上，能定义满足非负、规范和可列可加性的概率测度.

已学知识点

- 第二章 条件概率与统计独立性

  - 条件概率：$P(A\,|\,B) = \dfrac{P(AB)}{P(B)}$.

    - 乘法公式：$P\left(A_1A_2\cdots A_n\right) = P\left(A_1\right)\cdot P\left(A_2\middle|\,A_1\right)\cdot P\left(A_3\middle|\,A_1A_2\right)\cdot\cdots\cdot P\left(A_n\middle|\,A_1A_2\cdots A_{n-1}\right)$

    - 全概率公式：$P(B) = \sum_{i=1}^{\infty} P\left(A_i\right)\cdot P\left(B\,\middle|A_i\right)$

    - Bayes 公式：$P\left(A_k\,\middle|B\right) = \dfrac{P\left(A_k\right)\cdot P\left(B\,\middle|A_k\right)}{\sum_{i=1}^{\infty} P\left(A_i\right)\cdot P\left(B\,\middle|A_i\right)}$

$$\left.\begin{array}{l} A_i\cap A_j = \Phi\quad (i\neq j) \\ \bigcup_{i=1}^{\infty} A_i = \Omega \end{array}\right\}$$

已学知识点

- 第二章 条件概率与统计独立性
  - 事件独立性：两个事件 $P(AB) = P(A) \cdot P(B)$. 三个事件 $\begin{cases} P(AB) = P(A) \cdot P(B) \\ P(AC) = P(A) \cdot P(C) \\ P(BC) = P(B) \cdot P(C) \\ P(ABC) = P(A) \cdot P(B) \cdot P(C) \end{cases}$
  - 试验独立性：一个试验的结果对其它各试验的可能结果的概率都无影响.
  - Bernoulli 试验 E：概率空间 $\left(\Omega, \mathscr{F}, P\right)$，其中

$$A \subset \Omega, \quad \mathscr{F} = \left\{\Phi, A, \bar{A}, \Omega\right\}, \quad P(A) = p, P\left(\bar{A}\right) = q, \quad (p > 0, q > 0, p + q = 1)$$

- Bernoulli 分布
- 二项分布
- 几何 分布
- Pascal 分布
- 多项分布
- Poisson 分布

已学知识点

- 第三章 随机变量与分布函数
  - 随机变量 (r.v.) 与分布函数 (c.d.f.)：
    - 随机变量 ξ 是概率空间 $(\Omega, \mathscr{F}, P)$ 中定义域为 Ω、取值于 $\mathbb{R}$ 的一个单值 Borel 函数.
    - 分布函数 $F(x) = P\{\xi(\omega) < x\}$ ，$-\infty < x < \infty$ 是单调不降、取值 $[0, 1]$ 的右连续函数. 它完整描述了随机变量，是研究的主要对象.
    - 随机变量依取值 $\begin{cases} \text{离散型：} & \text{分布律 (mass function)、分布列} \\ \text{连续型：} & \text{概率密度 (p.d.f.)} \end{cases}$
    - 主要分布： $\begin{cases} \text{离散型：} & \text{Bernoulli, binomial, Poisson, hyper-geometric, geometric} \\ \text{连续型：} & \text{uniform, exponential, normal, } \Gamma \end{cases}$
    - 概率计算： $\begin{cases} \text{离散型：} & P\{x \in D\} = \sum\limits_{x_i \in D} p_i, \quad P\{(x, y) \in D\} = \sum\limits_{(x_i, y_j) \in D} p_{ij} \\ \text{连续型：} & P\{x \in D\} = \int\limits_D p(x)\mathrm{d}x, \quad P\{(x, y) \in D\} = \iint\limits_D p(x, y)\mathrm{d}x\mathrm{d}y \end{cases}$

已学知识点

- 第三章 随机变量与分布函数

随机向量，随机变量的独立性：

- 随机向量即多元随机变量
  - 联合分布： 联合分布函数、联合分布律、联合密度
  - 边际分布： 边际分布函数、边际分布律、边际密度
  - 条件分布： 条件分布函数、条件分布律、条件密度
  - 独立性： 与事件独立性几乎完全相同

- 主要分布：
  - 离散型： 多项分布、多元超几何分布
  - 连续型： 二元均匀分布、二元正态分布

$$
\begin{aligned}
(\xi, \eta) &\sim N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right) \\
&\sim N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{pmatrix}\right)
\end{aligned}
\quad\Longrightarrow\quad
\begin{cases}
\xi \sim N\left(\mu_1, \sigma_1^2\right), \qquad \eta \sim N\left(\mu_2, \sigma_2^2\right) \\
\left(\eta \,\middle|\, \xi = x\right) \sim N\left(\mu_2 + \rho\dfrac{\sigma_2}{\sigma_1}\left(x-\mu_1\right), \sigma_2^2\left(1-\rho^2\right)\right) \\
\left(\xi \,\middle|\, \eta = y\right) \sim N\left(\mu_1 + \rho\dfrac{\sigma_1}{\sigma_2}\left(y-\mu_2\right), \sigma_1^2\left(1-\rho^2\right)\right) \\
\xi \text{ 与 } \eta \text{ 相互独立} \Longleftrightarrow \rho = 0
\end{cases}
$$

已学知识点

第三章 随机变量与分布函数

随机变量的函数及其分布：

- 随机变量的函数什么情况下还是随机变量
- 离散型易：

$$\begin{cases} \eta = g(\xi): & \text{对应法} \\ \zeta = g(\xi, \eta): & \text{表上作业法，独立情形和的卷积公式} \end{cases}$$

- 连续型难：

$$\eta = g(\xi): \text{直接法 } F_\eta(y) = P\{\eta < y\} = \int_{g(x)<y} p_\xi(x)\mathrm{d}x$$

$$\zeta = g(\xi, \eta): \text{直接法 } F_\zeta(z) = P\{\zeta < z\} = \iint_{g(x, y)<z} p(x, y)\mathrm{d}x\mathrm{d}y$$

和 (卷积公式)、差、商、积、 max ，min 的公式

$$\begin{cases} \zeta_1 = g_1(\xi, \eta) \\ \zeta_2 = g_2(\xi, \eta) \end{cases}: \text{变换法} \begin{cases} u = g_1(x, y) \\ v = g_2(x, y) \end{cases} \Longrightarrow \begin{cases} x = h_1^{-1}(u, v) \\ y = h_2^{-1}(u, v) \end{cases} \Longrightarrow J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

$$q_{\zeta_1, \zeta_2}(u, v) = p\left(h_1^{-1}(u, v), h_2^{-1}(u, v)\right) \cdot |J|$$

已学知识点

- 第四章 数字特征与特征函数

  - 数学期望

    - 定义：$E(\xi) = \begin{cases} \sum\limits_i x_i \cdot p_i, & \text{离散型} \\ \int\limits_{-\infty}^{\infty} x \cdot p_\xi(x)\mathrm{d}x, & \text{连续型} \end{cases}$，绝对收敛.

    - 随机变量函数的数学期望：$E\left[g(\xi)\right] = \begin{cases} \sum\limits_i g\left(x_i\right) \cdot p_i, & \text{离散型} \\ \int\limits_{-\infty}^{\infty} g(x) \cdot p_\xi(x)\mathrm{d}x, & \text{连续型} \end{cases}$

$$E\left[g\left(\xi, \eta\right)\right] = \begin{cases} \sum\limits_i \sum\limits_j g\left(x_i, y_j\right) \cdot p_{ij}, & \text{离散型} \\ \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} g\left(x, y\right) \cdot p_{(\xi, \eta)}(x, y)\mathrm{d}x\mathrm{d}y, & \text{连续型} \end{cases}$$

数学期望的性质：

① $E(c) = c$

② $a \le \xi \le b \implies a \le E(\xi) \le b$

③ $\xi \le \eta \implies E(\xi) \le E(\eta)$

④ $E\left(a\xi + b\eta\right) = a \cdot E\left(\xi\right) + b \cdot E\left(\eta\right)$

⑤ ξ, η 独立 $\implies E(\xi\eta) = E(\xi) \cdot E(\eta)$

已学知识点

- 第四章 数字特征与特征函数

  - 方差：$D(\xi)=E\left\{\left[\xi-E(\xi)\right]^2\right\}$ 标准差：$\sqrt{D(\xi)}$ 计算公式：$D(\xi)=E(\xi^2)-\left[E(\xi)\right]^2$

  - Chebyshev 不等式：

$$P\left\{\left|\xi-E(\xi)\right|\geq\varepsilon\right\}\leq\frac{D(\xi)}{\varepsilon^2}\iff P\left\{\left|\xi-E(\xi)\right|<\varepsilon\right\}\geq 1-\frac{D(\xi)}{\varepsilon^2}$$

  - 协方差：

$$\operatorname{cov}(\xi,\eta)\triangleq E\left\{\left[\xi-E(\xi)\right]\left[\eta-E(\eta)\right]\right\}=E(\xi\eta)-E(\xi)\cdot E(\eta)$$

  - 和的方差：$D(X\pm Y)=D(X)+D(Y)\pm 2\operatorname{cov}(X,Y)$

方差的性质：

① 设 C 是常数，则 $D(C)=0$.

② 设 ξ 是随机变量，c 为常数，则 $D(\xi+c)=D(\xi)$.

③ 设 ξ 是随机变量，c 为常数，则 $D(c\xi)=c^2D(\xi)$.

④ 若 $c\neq E(\xi)$，则 $D(\xi)<E\left[(\xi-c)^2\right]$.

⑤ $D(\xi)=0 \iff P\{\xi=C\}=1$.

⑥ 若 ξ 与 η 独立，则 $D(\xi\pm\eta)=D(\xi)+D(\eta)$.

协方差的性质：

① $\operatorname{cov}(X,X)=D(X)$.

② $\operatorname{cov}(X,Y)=\operatorname{cov}(Y,X)$

③ $\operatorname{cov}(aX,bY)=ab\cdot\operatorname{cov}(X,Y)$，$a$, b 为常数.

④ $\operatorname{cov}(X_1+X_2,Y)=\operatorname{cov}(X_1,Y)+\operatorname{cov}(X_2,Y)$.

⑤ 若 X, Y 相互独立，则 $\operatorname{cov}(X,Y)=0$.

⑥ 若 C 为常数，则 $\operatorname{cov}(X,C)=0$.

已学知识点

- 第四章 数字特征与特征函数

  - 常见分布的数学期望与方差：

分布	符号表示	数学期望	方差
二项分布	$\xi \sim b(n, p)$	$E(\xi) = np$	$D(\xi) = np(1-p)$
Bernoulli 分布	$\xi \sim b(1, p)$	$E(\xi) = p$	$D(\xi) = p(1-p)$
Poisson 分布	$\xi \sim P(\lambda)$	$E(\xi) = \lambda$	$D(\xi) = \lambda$
均匀分布	$\xi \sim U[a, b]$	$E(\xi) = \frac{a+b}{2}$	$D(\xi) = \frac{(b-a)^2}{12}$
Gamma 分布	$X \sim \Gamma(r, \lambda)$	$E(\xi) = \frac{r}{\lambda}$	$D(\xi) = \frac{r}{\lambda^2}$
指数分布	$\xi \sim \mathrm{Exp}(\lambda)$	$E(\xi) = \frac{1}{\lambda}$	$D(\xi) = \frac{1}{\lambda^2}$
χ^2 分布	$X \sim \chi_n^2$	$E(\xi) = n$	$D(\xi) = 2n$
正态分布	$\xi \sim N(\mu, \sigma^2)$	$E(\xi) = \mu$	$D(\xi) = \sigma^2$

已学知识点

- 第四章 数字特征与特征函数
  - 相关系数：$\rho_{\xi\eta} \triangleq \dfrac{\operatorname{cov}(\xi,\eta)}{\sqrt{D(\xi)}\sqrt{D(\eta)}}$
  - Cauchy-Schwarz 不等式：$\left|E(\xi\eta)\right|^2 \le E(\xi^2)\cdot E(\eta^2)$
  - 矩：
$$\begin{cases} k\text{ 阶原点矩：}\ m_k = E(\xi^k), \quad k=1,2,3,\ldots \\ k\text{ 阶中心矩：}\ c_k = E\left\{[\xi - E(\xi)]^k\right\}, \quad k=2,3,\ldots \\ k+l\text{ 阶混合原点矩：}\ E(\xi^k\eta^l) \\ k+l\text{ 阶混合中心矩：}\ E\left\{[\xi-E(\xi)]^k[\eta-E(\eta)]^l\right\} \end{cases}$$
  - 定理：中心矩和原点矩可以相互表示.
$$\begin{cases} c_k = \sum_{i=0}^{k}\binom{k}{i} m_i(-m_1)^{k-i} \\ m_k = \sum_{i=0}^{k}\binom{k}{i} c_{k-i}\, m_1^i \end{cases}$$

相关系数的性质：

① $\left|\rho_{\xi\eta}\right| \le 1.$

② 等价命题：$\operatorname{cov}(\xi,\eta)=0$；$\xi$ 与 η 不相关；
$E(\xi\cdot\eta)=E(\xi)\cdot E(\eta)$；
$D(\xi+\eta)=D(\xi)+D(\eta)$.

③ 若随机变量 ξ 与 η 独立，则 ξ 与 η 不相关.

④ 二元正态分布的独立性和不相关性是等价的.

⑤ 二值随机变量的不相关性与独立性等价.

已学知识点

- 第四章 数字特征与特征函数

$\xi \backslash \eta$	y_1	y_2	$\cdots$	y_j	$\cdots$	$p_1(\bullet)$
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$\cdots$	$p(x_1, y_j)$	$\cdots$	$p_1(x_1)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$\cdots$	$p(x_2, y_j)$	$\cdots$	$p_1(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
x_i	$p(x_i, y_1)$	$p(x_i, y_2)$	$\cdots$	$p(x_i, y_j)$	$\cdots$	$p_1(x_i)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
$p_2(\bullet)$	$p_2(y_1)$	$p_2(y_2)$	$\cdots$	$p_2(y_j)$	$\cdots$	1

- 条件期望：

$$\begin{cases} E\left(\eta \,\middle|\, \xi = x_i\right) = \sum\limits_j y_j \cdot P\left\{\eta = y_j \,\middle|\, \xi = x_i\right\} = \sum\limits_j y_j \cdot \dfrac{p\left(x_i, y_j\right)}{p_1\left(x_i\right)} \\ E\left(\xi \,\middle|\, \eta = y_j\right) = \sum\limits_i x_i \cdot P\left\{\xi = x_i \,\middle|\, \eta = y_j\right\} = \sum\limits_i x_i \cdot \dfrac{p\left(x_i, y_j\right)}{p_2\left(y_j\right)} \end{cases}$$

- 连续型随机变量的条件期望：$E\left(\eta \,\middle|\, \xi = x\right) = \int_{-\infty}^{+\infty} y \cdot p\left(y \,|\, x\right) \mathrm{d}y$

  - η 关于 ξ 的条件期望 $E(\eta|\xi)$ 是一个随机变量，它在 $E\left(\eta \,\middle|\, \xi = x\right)$ 处的密度函数为 $p_\xi(x)$.

- 重期望公式：$E(\xi) = E\left[E\left(\xi \,\middle|\, \eta\right)\right]$.

已学知识点

第四章 数字特征与特征函数

- **复随机变量**：若 ξ 与 η 是概率空间 $(\Omega, \mathscr{F}, P)$ 上的实值随机变量，称 $\zeta = \xi + i\eta$ 为复随机变量.
  - 可平行定义或建立一系列结果：如 $E(\zeta) = E(\xi) + iE(\eta)$、独立性等.
  - Euler 公式：$e^{ix} = \cos x + i \sin x$.
- **特征函数**：$f_\xi(t) = E\left(e^{it\xi}\right) = \int_{-\infty}^{+\infty} e^{itx} dF_\xi(x) = \begin{cases} \sum_{j=1}^{\infty} p_j \cdot e^{itx_j}, & \text{离散型} \\ \int_{-\infty}^{+\infty} e^{itx} \cdot p_\xi(x)\, dx, & \text{连续型} \end{cases}$

特征函数的性质：

① 特征函数 $f(t)$ 有如下性质：$|f(t)| \le f(0) = 1$，$f(-t) = \overline{f(t)}$.

② 特征函数 $f(t)$ 在 $(-\infty, +\infty)$ 一致连续.

③ 对于任意的正整数 n 以及任意实数 $t_1, t_2, \ldots, t_n$、任意复数 $\lambda_1, \lambda_2, \ldots, \lambda_n$，均有 $\sum_{k=1}^{n} \sum_{j=1}^{n} f\left(t_k - t_j\right) \lambda_k \overline{\lambda_j} \ge 0$

④ 两个相互独立的随机变量之和的特征函数等于它们各自特征函数的乘积.

⑤ 设随机变量 ξ 的 n 阶矩存在，则它的特征函数可微分 n 次，且当 $k \le n$ 时：$f^{(k)}(0) = i^k E\left(\xi^k\right)$.

⑥ 设 $\eta = a\xi + b$，这里 a, b 为常数，则 $f_\eta(t) = e^{ibt} \cdot f_\xi(at)$.

常见分布的特征函数：

① 退化分布 $I_c(x)$：$f(t) = e^{itc}$

② 二项分布 $b(n, p)$：$f(t) = \left(pe^{it} + 1 - p\right)^n$

③ Poisson 分布：$f(t) = e^{\lambda\left(e^{it} - 1\right)}$

④ Γ 分布：$f(t) = \left(1 - \frac{it}{\lambda}\right)^{-r}$

⑤ 正态分布：$f(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$

已学知识点

- 第四章 数字特征与特征函数

  - 特征函数和分布函数相互唯一确定.

    - 分布函数可以确定特征函数：$f_\xi(t) \triangleq E\left(\mathrm{e}^{it\xi}\right) = \int_{-\infty}^{+\infty} \mathrm{e}^{itx}\, \mathrm{d}F_\xi(x) = \begin{cases} \sum_{j=1}^{\infty} p_j \cdot \mathrm{e}^{itx_j}, & \text{离散型} \\ \int_{-\infty}^{+\infty} \mathrm{e}^{itx} \cdot p_\xi(x)\,\mathrm{d}x, & \text{连续型} \end{cases}$
    - (唯一性定理) 分布函数由特征函数唯一确定.
    - 若特征函数 $f(t)$ 绝对可积，则相应分布函数 $F(x)$ 的导数存在并连续，并且 $F'(x) = p(x) = \dfrac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-itx} f(t)\,\mathrm{d}t$.

  - 分布函数的再生性：两个具有同一类型分布的独立随机变量之和的分布仍是这种类型的分布，且对应的参数等于两个随机变量相应参数之和.

    - 二项分布、Poisson 分布、正态分布、Γ 分布具有再生性.
    - 分解问题：对于正态分布、Poisson 分布，分解问题成立.

已学知识点

- 第五章 极限定理

  - **大数定律**：若随机变量序列 $\{\xi_k\}$ 满足 $\lim\limits_{n\to\infty} P\left\{\left|\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k - a\right| < \varepsilon\right\} = 1$，则称它服从大数定律.

  - **中心极限定理**：若独立的随机变量序列 $\{\xi_k\}$ 满足

$$\lim_{n\to\infty} P\left\{\frac{\sum\limits_{i=1}^{n}\xi_i - \sum\limits_{i=1}^{n}E(\xi_i)}{\sqrt{\sum\limits_{i=1}^{n}D(\xi_i)}} < x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t,$$

  则称它服从中心极限定理.

  - **问题**：寻找大数定律、中心极限定理成立的条件.

已学知识点

第五章 极限定理

$\mu_n \sim b(n, p)$

- Bernoulli 大数定律：设 μ_n 是 n 重 Bernoulli 试验中事件 A 发生的次数，$p = P\{A\}$，则 $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right\} = 1 \quad 或 \quad \lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| \geq \varepsilon\right\} = 0$$

- Chebyshev 大数定律：设 $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ 是两两不相关的随机变量序列，每一随机变量都有有限的方差，且它们有公共的上界 $D(\xi_1) \leq C, D(\xi_2) \leq C, \ldots, D(\xi_n) \leq C, \ldots$，则 $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}E(\xi_k)\right| < \varepsilon\right\} = 1$$

- Markov 大数定律：对于随机变量序列 $\xi_1, \xi_2, \ldots, \xi_n, \ldots$，若 $\frac{1}{n^2}\cdot D\left(\sum_{k=1}^{n}\xi_k\right) \longrightarrow 0$，则 $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}E(\xi_k)\right| < \varepsilon\right\} = 1$$

已学知识点

第五章 极限定理

- **独立同分布下的大数定律**：设 $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ 是独立同分布的随机变量序列，且 $E(\xi_i) = \mu$，$D(\xi_i) = \sigma^2, i = 1, 2, \ldots$ 则 $\forall \varepsilon > 0$，

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \mu\right| < \varepsilon\right\} = 1.$$

- **Poisson 大数定律**：在一个独立试验序列中，如果事件 A 在第 k 次试验中出现的概率等于 p_k，以 μ_n 表示在前 n 次试验中事件 A 出现的次数，则对任意 $\varepsilon > 0$，都有

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n} - \frac{p_1 + p_2 + \cdots + p_n}{n}\right| < \varepsilon\right\} = 1.$$

大数定律的重要意义：

- 大数定律建立了大量重复独立试验中事件出现频率具有稳定性，使得概率有了客观意义.
- 大数定律在偶然性与必然性之间架起了桥梁，是概率论学科最有特色的命题.
  - 要求 n 很大，即它是一类极限定理.
  - 讨论的是平均值，这是概率论的特色.
  - 建立概率接近于 0 或 1 的规律，这是概率论研究中特别要强调的.
  - 规律的产生是大量独立或弱相关因素积累的结果 —— 统计独立性的概念.
- 参数估计：大数定律是重要理论基础之一.

已学知识点

- 第五章 极限定理

$\mu_n \sim b(n, p)$

▶ De Moivre - Laplace 定理：若 μ_n 是 n 次 Bernoulli 试验中事件 A 出现的次数，$0 < p < 1$，则对任意有限区间 $[a, b]$：

① 当 $a \le x_k \equiv \dfrac{k-np}{\sqrt{np(1-p)}} \le b$ 及 $n \to \infty$ 时，一致地有 $P\{\mu_n = k\} \div \left(\dfrac{1}{\sqrt{np(1-p)}} \cdot \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_k^2}\right) \longrightarrow 1.$

局部极限定理

② 当 $n \to \infty$ 时，一致地有 $P\left\{a \le \dfrac{\mu_n - np}{\sqrt{np(1-p)}} < b\right\} \longrightarrow \displaystyle\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \qquad (-\infty < x < +\infty).$

积分极限定理

$$P\{\mu_n = k\} \approx \frac{1}{\sqrt{np(1-p)}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_k^2}, \quad x_k = \frac{k-np}{\sqrt{np(1-p)}}; \quad n \to \infty$$

$$\frac{\mu_n - np}{\sqrt{np(1-p)}} \sim N(0, 1), \quad n \to \infty \quad \Longrightarrow \quad \mu_n \sim N(np, np(1-p)), \quad n \to \infty$$

▶ 应用：用频率估计概率时的计算问题

已学知识点

- 第五章 极限定理

  - 分布函数弱收敛：$F_n(x) \xrightarrow{w} F(x) \iff$ 在 $F(x)$ 的每一连续点上都有 $\lim\limits_{n\to\infty} F_n(x) = F(x)$.
  - Levy - Cramer 连续性定理：分布函数列 $F_n(x) \xrightarrow{w} F(x) \iff$ 对应的特征函数列 $\{f_n(t)\}$ 在任意有限区间内一致收敛到某个函数 $f(t)$.
  - 随机变量的收敛性：

① 依分布收敛 $\xi_n \xrightarrow{L} \xi$ ： $F_{\xi_n}(x) \xrightarrow{W} F_\xi(x)$

② 依概率收敛 $\xi_n \xrightarrow{p} \xi$ ： $\lim\limits_{n\to\infty} P\left\{\left|\xi_n - \xi\right| \geq \varepsilon\right\} = 0$

③ r 阶矩收敛 $\xi_n \xrightarrow{r} \xi$ ： $\lim\limits_{n\to\infty} E\left(\left|\xi_n - \xi\right|^r\right) = 0$， $r > 0$

④ 几乎处处收敛 $\xi_n \xrightarrow{a.s.} \xi$ ： $P\left\{\lim\limits_{n\to\infty} \xi_n(\omega) = \xi(\omega)\right\} = 1$

5.2 收敛性

一、分布函数弱收敛

二、连续性定理

三、随机变量的收敛

四、Khintchin 大数定律与 Kolmogorov 大数定律

四、Khintchin 大数定律与 Kolmogorov 大数定律

定理：(Khintchin 大数定律) 设 $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ 是相互独立的随机变量序列，且服从相同的分布，均具有有限的数学期望 $E\left(\xi_n\right) = a$，则对任意的 $\varepsilon > 0$，都有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - a\right| < \varepsilon\right\} = 1, \text{即 } \frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{p} a\,, \quad n \to \infty$$

- 前面我们通过 Chebyshev 不等式建立起了多种大数定律，但都假定了方差的存在性. 在独立同分布场合，并不需要有这个要求，这就是有名的辛钦 (Khintchin) 大数定律.

四、Khintchin 大数定律与 Kolmogorov 大数定律

定理：(Kolmogorov 大数定律) 设 $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ 是独立同分布 (i.i.d.) 的随机变量序列，则

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{\text{a.s.}} a\ , \quad n \to \infty$$

当且仅当数学期望 $E\left(\xi_k\right)$ 存在，且 $E\left(\xi_k\right) = a.$

四、Khintchin 大数定律与 Kolmogorov 大数定律

- 大数定律的意义：
  - 为寻找随机变量的期望值提供了一条实际可行的途径.

定理：(Khintchin 大数定律) 设 $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ 是相互独立的随机变量序列，且服从相同的分布，均具有有限的数学期望 $E(\xi_n) = a$，则对任意的 $\varepsilon > 0$，都有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - a\right| < \varepsilon\right\} = 1, \text{ 即 } \frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{p} a\,, \quad n\to\infty$$

  - 为寻找随机事件的概率提供了一条实际可行的途径.

定理：(Bernoulli 大数定律) 设 μ_n 是 n 重 Bernoulli 试验中事件 A 发生的次数，$p = P\{A\}$，则 $\forall\varepsilon > 0$，

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right\} = 1 \quad \text{或} \quad \lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| \geq \varepsilon\right\} = 0$$

五、大数定律的应用

- 用 Monte Carlo 方法计算定积分 $J=\int_a^b g(x)\mathrm{d}x.$

计算机可生成 $U[a,\,b]$ 的随机数 $\{\xi_k\}$

任取相互独立、服从 $[a,\,b]$ 上均匀分布的随机变量 $\{\xi_k\}$ ： $\xi_k \overset{\text{i.i.d.}}{\sim} U[a,\,b]$， $k=1,\,2,\,\ldots$

$\Longrightarrow$ $\{g(\xi_k)\}$ 也是 i.i.d. 的随机变量， $k=1,\,2,\,\ldots$

$$E\left[g(\xi_k)\right]=\int_a^b g(x)\cdot\frac{1}{b-a}\mathrm{d}x=\frac{1}{b-a}\int_a^b g(x)\cdot\mathrm{d}x=\frac{J}{b-a}\Longrightarrow J=(b-a)\cdot E\left[g(\xi_k)\right]$$

$\Longrightarrow$ 只要能求得 $E\left[g(\xi_k)\right]$，即可得到 J 的值

大数定律保证了这种算法失效的概率为零！

定理：(Khintchin 大数定律) 设 $\xi_1,\,\xi_2,\,\ldots,\,\xi_n,\,\ldots$ 是相互独立的随机变量序列，且服从相同的分布，均具有有限的数学期望 $E(\xi_n)=a$，则对任意的 $\varepsilon>0$，都有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-a\right|<\varepsilon\right\}=1,\ \text{即}\ \frac{1}{n}\sum_{k=1}^{n}\xi_k\xrightarrow{p}a\,,\quad n\to\infty$$

$$\Longrightarrow\quad \frac{g(\xi_1)+g(\xi_2)+\cdots+g(\xi_n)}{n}\xrightarrow{p}E\left[g(\xi_k)\right]$$

五、大数定律的应用

- 用 Monte Carlo 方法计算定积分 $J = \int_a^b g(x)\mathrm{d}$

```
f = function(n = 10, a = -1, b = 1) {
  x = runif(100000, min = a, max = b) # 从 [a, b] 抽取均匀分布的
  g = function(x) 1/(sqrt(2 * pi)) * exp(-x^2/2) # 定义 g 为被积函数
  y = g(x) # 计算 g(x) 的值
  data.frame('近似值' = (b - a) * mean(y), '真值' = pnorm(b, 0, 1) -
}
f()
f(n = 100)
f(n = 1000)
f(n = 10000)
f(n = 10000, a = -2, b = 2)
f(n = 10000, a = -3, b = 3)
```

```
> f()
     近似值      真值
1 0.6823968 0.6826895
> f(n = 100)
     近似值      真值
1 0.6821788 0.6826895
> f(n = 1000)
     近似值      真值
1 0.6822648 0.6826895
> f(n = 10000)
     近似值      真值
1 0.6832235 0.6826895
> f(n = 10000, a = -2, b = 2)
     近似值      真值
1 0.9543783 0.9544997
> f(n = 10000, a = -3, b = 3)
    近似值      真值
1 1.000106 0.9973002
```

五、大数定律的应用

- 用 Monte Carlo 方法计算定积分 $J = \int_a^b g(x)\mathrm{d}x.$

  - 这种通过概率论的想法构造模型从而实现数值计算的方法，随着电子计算机的发展，已形成一种新的计算方法 —— 概率计算方法，亦称蒙特卡洛 (Monte Carlo) 方法. 它在原子物理、公用事业理论中发挥了不少作用，这个方法的理论根据之一就是大数定律！
  - 至于计算积分，Monte Carlo 方法的实用场合是计算重积分

$$J = \overbrace{\int \cdots\cdots \int}^{m}_{K} g(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

其中 $\boldsymbol{x}$ 是 m 维空间中的点.

5.3 中心极限定理

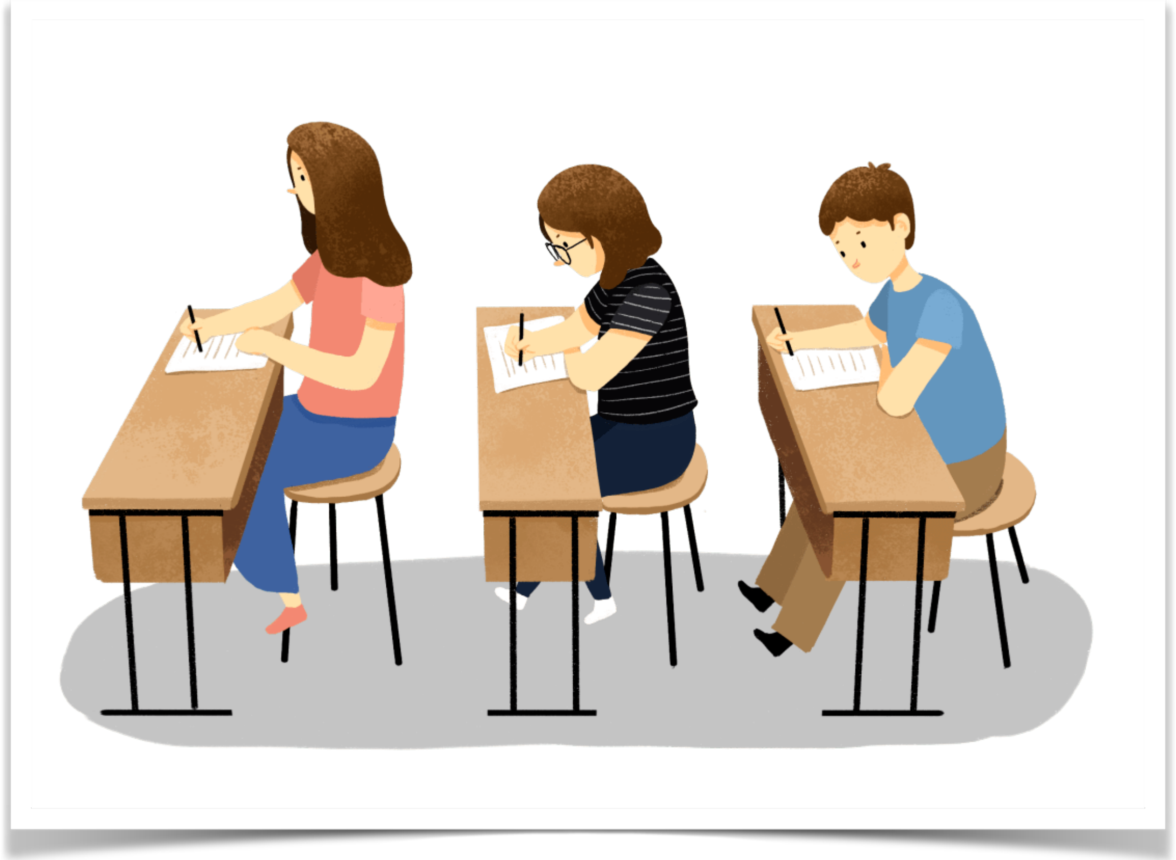
一、中心极限定理的客观背景

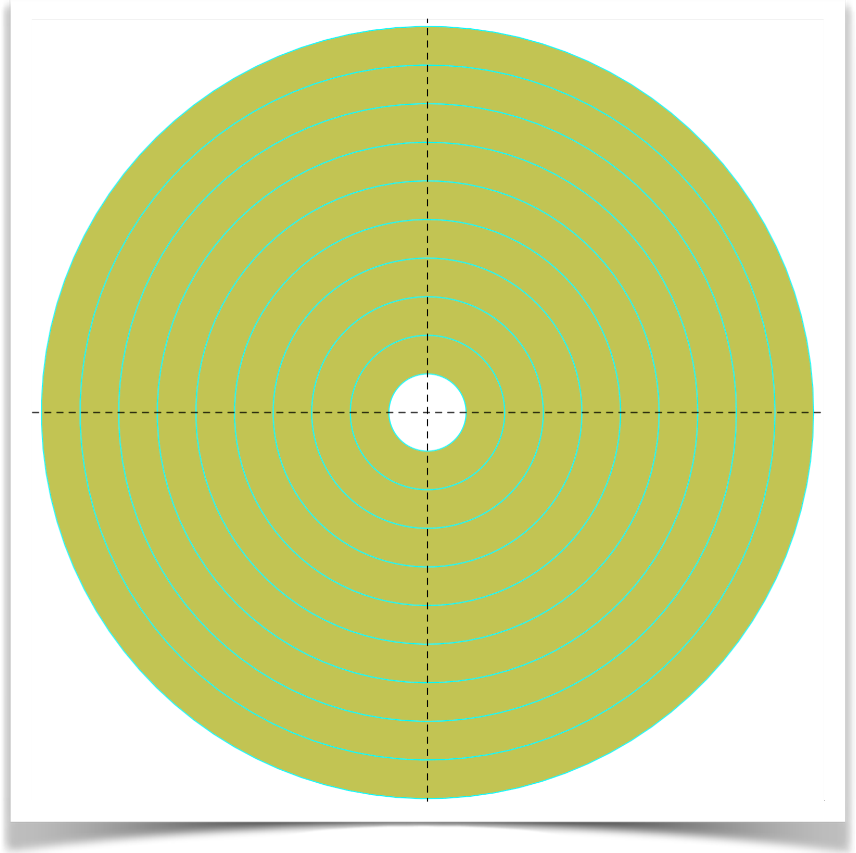
二、Lindburg-Levy 中心极限定理

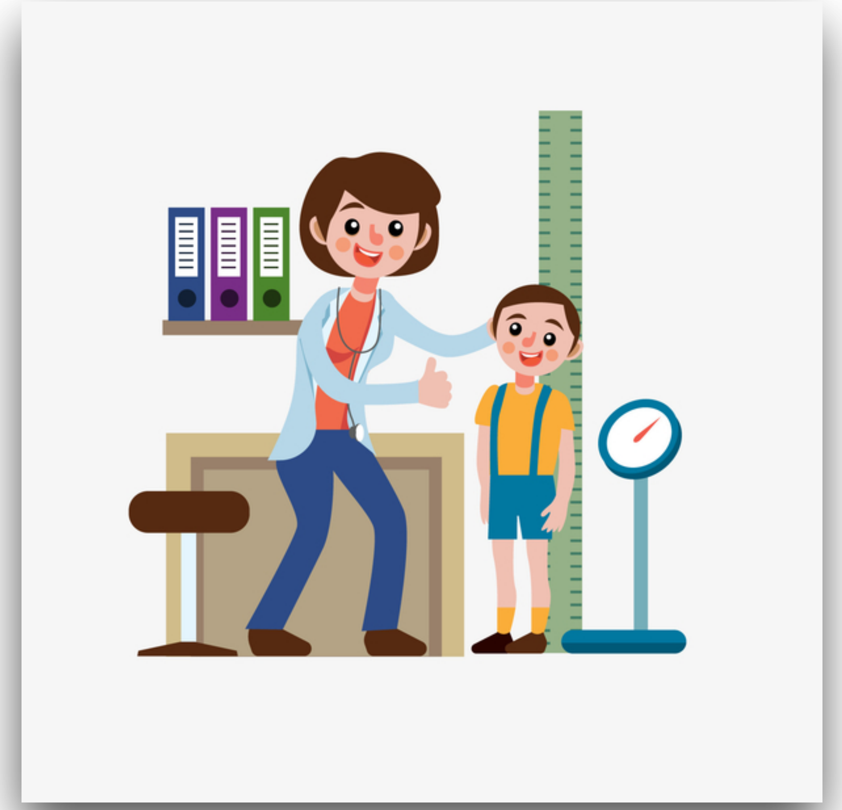
三、中心极限定理的应用

一、中心极限定理的客观背景

- 自从 Gauss 指出测量误差服从正态分布之后，人们发现正态分布在自然现象中极为常见.
  - 身高、体重.
  - 射击的落点与目标的偏差.
  - 考试成绩.







- 在实际问题中，许多随机变量是由相互独立的随机因素的综合 (和) 影响所形成，而每一个别因素在总影响中所起的作用不很大，则这种量通常会服从或近似服从正态分布.
- 在数理统计中，经常假定总体服从正态分布，其合理性需要有依据.

二、Lindburg-Levy 中心极限定理

定理： (Lindburg-Levy 中心极限定理) 设 $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ 是独立同分布 (i.i.d.) 的随机变量序列,

且 $E\left(\xi_k\right) = \mu$, $D\left(\xi_k\right) = \sigma^2 < \infty$, $k = 1, 2, \ldots,$ 则

$$\lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} \xi_k - n\mu}{\sqrt{n}\ \sigma} \le x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{\frac{t^2}{2}} \mathrm{d}t$$

即

$$\frac{\sum_{k=1}^{n} \xi_k - n\mu}{\sqrt{n}\ \sigma} \xrightarrow{L} \xi \sim N(0\ ,\ 1)\ ,\quad n \to \infty$$

证明略.

二、Lindburg-Levy 中心极限定理

- 设 $\{\xi_n\}_{n=1}^{\infty}$ 是 i.i.d. 的以 p 为参数的 Bernoulli 随机变量序列. $\Longrightarrow$ $S_n = \sum\limits_{k=1}^{n} \xi_k \sim b(n, p)$

定理：(Lindburg-Levy 中心极限定理) 设 $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ 是 i.i.d. 的随机变量序列，且 $E\left(\xi_k\right) = \mu,\ D\left(\xi_k\right) = \sigma^2 < \infty$ ，　$k = 1, 2, \ldots,$　则

$$\lim_{n\to\infty} P\left\{ \frac{\sum\limits_{k=1}^{n} \xi_k - n\mu}{\sqrt{n}\ \sigma} \le x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{t^2}{2}} dt,$$

即

$$\frac{\sum\limits_{k=1}^{n} \xi_k - n\mu}{\sqrt{n}\ \sigma} \xrightarrow{L} \xi \sim N(0\ ,\ 1)\ ,\quad n \to \infty.$$

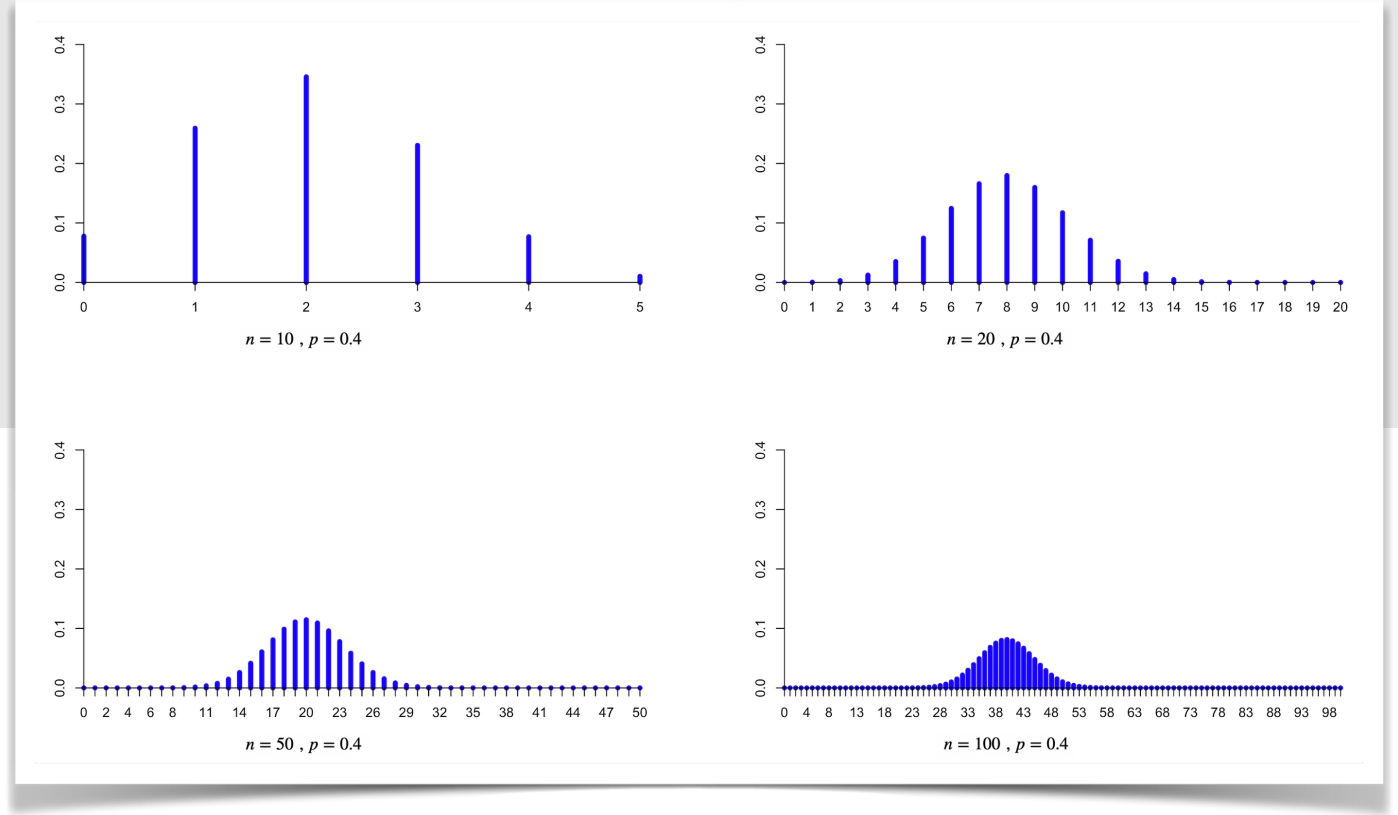
$$\Longrightarrow \frac{\sum\limits_{k=1}^{n} \xi_k - np}{\sqrt{np(1-p)}} \xrightarrow{L} N(0\ ,\ 1)\ ,\quad n \to \infty$$

$$\Longrightarrow \sum_{k=1}^{n} \xi_k \xrightarrow{L} N\left(np\ ,\ np(1-p)\right)\ ,\quad n \to \infty$$

二、Lindburg-Levy 中心极限定理

- 设 $\{\xi_n\}_{n=1}^{\infty}$ 是 i.i.d. 的以 p 为参数的 Bernoulli 随机变量序列. $\Longrightarrow$ $S_n = \sum_{k=1}^{n} \xi_k \sim b(n, p)$

```
f = function(n = 5, p = 0.4) {
  x = dbinom(0:n, n, p)
  plot(0:n, x, type = 'h', axes = FALSE, xlim = c(0, n), ylim = c(0, 0.4), xlab = '', ylab = '',
     col = 'blue', lwd = 5)
  axis(1, at = 0:n, pos = 0)
  axis(2, pos = 0)
  }
par(mfrow = c(2, 2))
f()
f(n = 20)
f(n = 50)
f(n = 100)
```



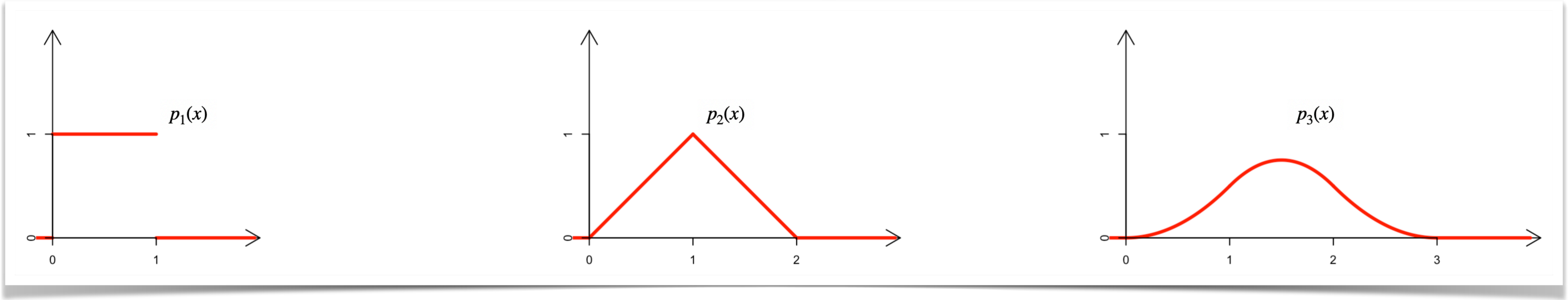
$$\Longrightarrow \quad \sum_{k=1}^{n} \xi_k \xrightarrow{L} N\left(np \,,\, np(1-p)\right) \,, \quad n \to \infty$$

二、Lindburg-Levy 中心极限定理

- 设 $\{\xi_n\}_{n=1}^{\infty}$ 是 i.i.d. 的 $U(0\ ,\ 1)$ 随机变量序列. $S_n \triangleq \sum_{k=1}^{n} \xi_k \xrightarrow{\text{p.d.f.}} p_n(x)$

$$p_1(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{其它} \end{cases},$$

$$p_2(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 \le x < 2 \\ 0, & \text{其它} \end{cases},$$



$$p_3(x) = \begin{cases} \frac{1}{2}x^2, & 0 < x < 1 \\ \frac{1}{2}\left(x^2 - 3(x-1)^2\right), & 1 \le x < 2 \\ \frac{1}{2}\left(x^2 - 3(x-1)^2 + 3(x-2)^2\right), & 2 \le x < 3 \\ 0, & \text{其它} \end{cases}$$

三、中心极限定理的应用

- 某车间有 200 台车床，它们独立地工作着，开工率各为 0.6 ，开工时耗电各为 1 千瓦，问供电所至少要提供给这个车间多少电力才能以 99.9% 的概率保证这个车间不会因供电不足而影响生产?

某一台车床工作与否是 Bernoulli 试验

ξ_k	1	0
p_j	0.6	0.4

$k = 1, 2, \ldots, 200$

同时在工作的车床数 $\eta = \xi_1 + \xi_2 + \cdots + \xi_{200} \sim b(200, 0.6)$

问题：求供电数 r 使得 $P\{\eta \le r\} = P\left\{\sum_{k=1}^{200} \xi_k \le r\right\} \ge 0.999 \iff P\left\{\frac{\eta - 120}{\sqrt{48}} \le \frac{r-120}{\sqrt{48}}\right\} \ge 0.999$

Lindburg-Levy 中心极限定理 $\Longrightarrow$ $\eta = \sum_{k=1}^{200} \xi_k \xrightarrow{L} N(200 \times 0.6, 200 \times 0.6 \times 0.4) = N(120, 48)$

$$\Phi\left(\frac{r-120}{\sqrt{48}}\right) \ge 0.999$$

qnorm(0.999, 0, 1, lower.tail = TRUE)

```
> qnorm(0.999, 0, 1, lower.tail = TRUE)
[1] 3.090232
```

$$\frac{r-120}{\sqrt{48}} \ge 3.09 \Longrightarrow r \ge 141.4$$

三、中心极限定理的应用

- 某车间有 200 台车床，它们独立地工作着，开工率各为 0.6，开工时耗电各为 1 千瓦，问供电所至少要提供给这个车间多少电力才能以 99.9% 的概率保证这个车间不会因供电不足而影响生产?

$$\Leftrightarrow \boxed{P\left\{\frac{\eta-120}{\sqrt{48}} \le \frac{r-120}{\sqrt{48}}\right\}} \ge 0.999$$

$$\Phi\left(\frac{r-120}{\sqrt{48}}\right) \ge 0.999$$

$$\frac{r-120}{\sqrt{48}} \ge 3.09 \implies r \ge 141.4$$

该结果表明 $P\{\eta \le 142\} \ge 0.999$，所以若供电 142 千瓦，由于供电不足而影响生产的可能性小于 0.001，这相当于 8 小时工作中有半分钟受影响，在一般企业中是可以接受的.

三、中心极限定理的应用

- 掷一枚均匀的骰子，为了至少有 95% 的把握使 2 点向上的频率与概率之差落在 0.01 的范围之内，问至少需要投掷多少次?

掷一枚均匀骰子的结果是 Bernoulli 试验

ξ_k	1	0
p_j	$\frac{1}{6}$	$\frac{5}{6}$

$$\Longrightarrow \quad E(\xi_k)=\frac{1}{6}, \quad D(\xi_k)=\frac{5}{36}$$

出现 2 点向上的总数 $\eta=\xi_1+\xi_2+\cdots+\xi_n \sim b\left(n, \frac{1}{6}\right)$

$$\Longrightarrow \quad \frac{\frac{\eta}{n}-\frac{1}{6}}{\sqrt{\frac{5}{36n}}} \xrightarrow{L} N(0, 1)$$

问题： 求投掷次数 n 使得 $P\left\{\left|\frac{\eta}{n}-\frac{1}{6}\right| \le 0.01\right\} \ge 0.95$

Lindburg-Levy 中心极限定理 $\Longrightarrow \quad \eta=\sum_{k=1}^{n}\xi_k \xrightarrow{L} N\left(\frac{1}{6}n, \frac{5}{36}n\right) \Longrightarrow \quad \frac{\eta}{n}-\frac{1}{6} \xrightarrow{L} N\left(0, \frac{5}{36n}\right)$

$$\Longrightarrow \quad \left(\frac{\eta}{n}-\frac{1}{6}\right)\sqrt{\frac{36n}{5}} \xrightarrow{L} N(0, 1)$$

三、中心极限定理的应用

- 掷一枚均匀的骰子，为了至少有 95% 的把握使 2 点向上的频率与概率之差落在 0.01 的范围之内，问至少需要投掷多少次?

问题： 求投掷次数 n 使得 $P\left\{\left|\frac{\eta}{n}-\frac{1}{6}\right|\le 0.01\right\}\ge 0.95$

$$\Longleftrightarrow \boxed{P\left\{\left|\left(\frac{\eta}{n}-\frac{1}{6}\right)\sqrt{\frac{36n}{5}}\right|\le 0.01\sqrt{\frac{36n}{5}}\right\}}\ge 0.95$$

$$=\Phi\left\{0.01\sqrt{\frac{36n}{5}}\right\}-\Phi\left\{-0.01\sqrt{\frac{36n}{5}}\right\}$$

$$=2\Phi\left\{0.01\sqrt{\frac{36n}{5}}\right\}-1\ge 0.95 \Longrightarrow \Phi\left\{0.01\sqrt{\frac{36n}{5}}\right\}\ge 0.975$$

qnorm(0.975, 0, 1, lower.tail = TRUE)

```
> qnorm(0.975, 0, 1, lower.tail = TRUE)
[1] 1.959964
```

$$\Longrightarrow 0.01\sqrt{\frac{36n}{5}}\ge 1.959964 \Longrightarrow n\ge 5335.36$$ 至少需要投掷 5336 次

$$\Longrightarrow \left(\frac{\eta}{n}-\frac{1}{6}\right)\sqrt{\frac{36n}{5}}\xrightarrow{L} N(0\ ,\ 1)$$

三、中心极限定理的应用

- 某自动售货机出售三种蛋糕，价格分别为 1 (元)、1.2 (元)、1.5 (元)，根据以往经验知道这三种蛋糕售出的概率分别为 0.3、0.2、0.5. 某天售出了 300 只蛋糕，问

① 这一天收入至少 390 元的概率.

② 这一天售出价格为 1.2 (元) 的蛋糕多于 60 只的概率.

ξ_k	1	1.2	1.5
p_j	0.3	0.2	0.5

pnorm(0.78326, 0, 1, lower.tail = FALSE)

```
> pnorm(0.78326, 0, 1, lower.tail = FALSE)
[1] 0.2167372
```

$$\Longrightarrow \begin{cases} E(\xi_k) = 1\times 0.3 + 1.2\times 1.2 + 1.5\times 0.5 = 1.29 \\ D(\xi_k) = E(\xi_k^2) - \left[E(\xi_k)\right]^2 = 1^2\times 0.3 + 1.2^2\times 1.2 + 1.5^2\times 0.5 - 1.29^2 = 0.0489 \end{cases}$$

这一天收入至少 390 元的概率 $\Longrightarrow$ $P\{\xi_1+\xi_2+\cdots+\xi_{300}\geq 390\} = P\left\{\dfrac{\sum\limits_{k=1}^{300}\xi_k - 300\times 1.29}{\sqrt{300\times 0.0489}} \geq \dfrac{390-300\times 1.29}{\sqrt{300\times 0.0489}}\right\}$

Lindburg-Levy 中心极限定理 $\Longrightarrow$ $\sum\limits_{k=1}^{300}\xi_k \xrightarrow{L} N(300\times 1.29\ ,\ 300\times 0.0489)$

$$\approx 1-\Phi(0.78326)$$
$$= 0.2167$$

三、中心极限定理的应用

- 某自动售货机出售三种蛋糕，价格分别为 1 (元)、1.2 (元)、1.5 (元)，根据以往经验知道这三种蛋糕售出的概率分别为 0.3、0.2、0.5. 某天售出了 300 只蛋糕，问

① 这一天收入至少 390 元的概率.

② 这一天售出价格为 1.2 (元) 的蛋糕多于 60 只的概率.

ξ_k	1	1.2	1.5
p_j	0.3	0.2	0.5

n_1, n_2, n_3 售出三种蛋糕的只数 $\Longrightarrow$ $n_1+n_2+n_3=300$ $\Longrightarrow$ (n_1, n_2, n_3) 服从多项分布 $\Longrightarrow$ $n_2 \sim b(300, 0.2)$

售出价格为 1.2 (元) 的蛋糕多于 60 只的概率 $\Longrightarrow$ $P\{n_2>60\}=P\left\{\dfrac{n_2-300\times 0.2}{\sqrt{300\times 0.2\times 0.8}}>\dfrac{60-300\times 0.2}{\sqrt{300\times 0.2\times 0.8}}\right\}$

De Moivre - Laplace 定理 $\Longrightarrow$ $\dfrac{n_2-300\times 0.2}{\sqrt{300\times 0.2\times 0.8}} \xrightarrow{L} N(0, 1)$

$$\approx 1-\Phi(0)=\frac{1}{2}$$