第十六讲：数学期望

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

平均值与加权平均值

有甲乙两名射手，其射击技术可用下表表出

甲射手

击中环数	8	9	10
概　　率	0.3	0.1	0.6

乙射手

击中环数	8	9	10
概　　率	0.2	0.5	0.3

试问哪一个射手技术较好?

- 显然这个问题的答案不是一眼就可以看出来的;
- 这也表明分布列虽然完整的描述了随机变量，但却不够集中的反应其变化情况;
- 有必要找一些量来更集中，更概括的描述随机变量;
- 所要找的量多是某种平均值.

平均值与加权平均值

- 最常见的平均值求法:$\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$;
- 没有考虑每个数据相对重要性，比如一小学生考试成绩为：语文 95 分，数学 85 分，常识 60 分，若按上述计算其平均成绩为 $\bar{x} = 80$;
- 上述计算没有考虑到三个科目的相对重要性，不太能反映学生的真正成绩：如果在这个年级中，每周有语文 10 节课，数学 8 节课，常识 2 节课，则利用下述的平均计算其平均成绩似科更合理一些:

$$\bar{x}_w = 95 * \frac{10}{20} + 85 * \frac{8}{20} + 60 * \frac{2}{20} = 87.5$$

- 加权平均：给定权 w_i 满足 $\sum_{i=1}^{n} w_i = 1$, 则

$$\bar{x}_w = \sum_{i=1}^{n} w_i x_i$$

称为加权平均值.

有甲乙两名射手，其射击技术可用下表表出

甲射手

击中环数	8	9	10
概　　率	0.3	0.1	0.6

乙射手

击中环数	8	9	10
概　　率	0.2	0.5	0.3

考虑以概率为权重的加权平均:

$$\begin{aligned}\bar{x}_{\text{甲}} &= 8*0.3+9*0.1+10*0.6=9.3\\ \bar{x}_{\text{乙}} &= 8*0.2+9*0.5+10*0.3=9.1\end{aligned}$$

则平均起来，甲每枪射中 9.3 环，乙每枪射中 9.1 环，故甲的技术更好一些.

数学期望的定义

定义 1.1 设离散型随机变量的分布列为

$$\begin{pmatrix} x_1, & x_2, & \cdots, & x_k, & \cdots \\ p_1, & p_2, & \cdots, & p_k, & \cdots \end{pmatrix}$$

如果

$$\sum_{i=1}^{\infty} |x_i| p_i < \infty$$

则称

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) = \sum_{i=1}^{\infty} x_i p_i$$

为随机变量 X 的数学期望，或称该分布的数学期望，简称期望或均值。若级数 $\sum_{i=1}^{\infty} |x_i| p_i$ 不收敛，则称 X 的期望不存在.

从离散到连续

- 假设连续型随机变量 X 的分布密度为 $p(x)$;
- 考虑随机变量 X 的如下近似：记 $A_i := \{\omega : X(\omega) \in (x_i, x_{i+1}]\}$

$$\tilde{X} := \sum_i x_i I_{A_i}(\omega)$$

其中 $I_{A_i}(\omega) = \begin{cases} 1, & \omega \in A_i \\ 0, & \omega \notin A_i \end{cases}.$

显然 $\tilde{X}$ 为离散型随机变量，且

$$P(\tilde{X} = x_i) = \int_{x_i}^{x_{i+1}} p(x)dx \approx p(x_i)(x_{i+1} - x_i);$$

- 故 $\tilde{X}$ 的期望为

$$E(\tilde{X}) \approx \sum_i x_i p(x_i)(x_{i+1} - x_i) \to \int xp(x)dx$$

定义 1.2 设连续随机变量 X 的密度函数为 $p(x)$. 如果

$$\int_{-\infty}^{\infty} |x| p(x) dx < \infty,$$

则称

$$E(X) := \int_{-\infty}^{\infty} x p(x) dx$$

为随机变量 X 的数学期望，简称期望或均值。若 $\int_{-\infty}^{\infty} |x| p(x) dx$ 不收敛，则称 X 的期望不存在.

数学期望的一般情形

- 若随机变量 X 的分布函数为 $F(x)$;
- 类似于连续情形，考虑随机变量 X 的如下近似:

$$\tilde{X} := \sum_i x_i I_{A_i}(\omega)$$

其中 $A_i := \{\omega : X(\omega) \in (x_i, x_{i+1}]\}, I_{A_i}(\omega) = \begin{cases} 1, & \omega \in A_i \\ 0, & \omega \notin A_i \end{cases}$.
显然 $\tilde{X}$ 为离散型随机变量，且

$$P(\tilde{X} = x_i) = P(X \in (x_i, x_{i+1}]) = F(x_{i+1}) - F(x_i);$$

- 故 $\tilde{X}$ 的期望为

$$E(\tilde{X}) = \sum_i x_i \left[F(x_{i+1}) - F(x_i)\right] \to \int x dF(x)$$ 斯蒂尔切斯积分

数学期望的一般定义

定义 1.3 假设随机变量 X 的分布函数为 $F(x)$, 则定义

$$E(X) = \int_{-\infty}^{+\infty} x dF(x)$$

为随机变量 X 的数学期望。这里我们要求上述积分绝对收敛，否则称数学期望不存在.

$$I = \int_{-\infty}^{+\infty} g(x) dF(x)$$

- 当 $F(x)$ 为右连续的阶梯函数，在 $x_i (i = 1, 2, \cdots,)$ 具有跃度 p_i 时，上面的积分化为

$$I = \sum_i g(x_i) \Big(F(x_i) - F(x_{i-1}) \Big) = \sum_i g(x_i) p_i$$

- 当 $F(x)$ 存在导数 $F'(x) = p(x)$ 时，上述积分化为普通积分

$$I = \int_{-\infty}^{+\infty} g(x) p(x) dx$$

- 线性性质

$$\int_{-\infty}^{+\infty} (ag_1(x) + bg_2(x) dF(x) = a \int_{-\infty}^{+\infty} g_1(x) dF(x) + b \int_{-\infty}^{+\infty} g_2(x) dF(x)$$

- 若 $g(x) \geq 0$, $F(x)$ 单调不减，$b > a$, 则 $\int_a^b g(x) dF(x) \geq 0$

- 在引入上述相关的积分后，

$$
\begin{aligned}
F(x) &= P(X \leq x) = \int_{(-\infty,x]} dF(x) \\
P(X \in B) &= \int_B dF(x)
\end{aligned}
$$

- 数学期望的另一计算公式：关于概率测度的积分

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} xdF(x) \quad \text{注意到} F(x) = P(X(\omega) \leq x) := P(X^{-1}(x)) \\
&\xlongequal{x=X(\omega)} \int_\Omega X(\omega)dP(X^{-1}(X(\omega))) = \int_\Omega X(\omega)dP(\omega)
\end{aligned}
$$

第十七讲：数学期望 (续) 与方差

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

随机变量关于概率测度的积分

- 若 $X(\omega)=\sum_{i=1}^{n} a_i I_{A_i}(\omega), A_1,\cdots,A_n\in\mathcal{F}$, 即 X 为简单随机变量时，定义

$$\int_{\Omega} X(\omega)P(d\omega) := \sum_{i=1}^{n} a_i P(A_i)$$

- 若 $X(\omega)$ 为非负随机变量，则由之前随机变量的结构易知:

$$\text{存在简单随机变量序列} X_n(\omega)\uparrow X(\omega)$$

故可定义

$$\int_{\Omega} X(\omega)P(d\omega) := \lim_{n\to\infty}\int_{\Omega} X_n(\omega)P(d\omega)$$

- 若 $X(\omega)$ 为任一随机变量，则由于 $X(\omega)=X^+(\omega)-X^-(\omega)$, 故定义

$$\int_{\Omega} X(\omega)P(d\omega) := \int_{\Omega} X^+(\omega)P(d\omega)-\int_{\Omega} X^-(\omega)P(d\omega)$$

随机变量可积与积分存在的定义

定义 1.4 称随机变量 $X(\omega)$ 关于概率测度 P 的积分存在，如果

$$\int_\Omega X^+(\omega)P(d\omega) \text{ 与 } \int_\Omega X^-(\omega)P(d\omega) \text{不同时为} \infty$$

称随机变量 $X(\omega)$ 关于概率测度 P 可积，如果 $\int_\Omega X(\omega)P(d\omega) < \infty$，即

$$\int_\Omega X^+(\omega)P(d\omega) < \infty \text{ 且 } \int_\Omega X^-(\omega)P(d\omega) < \infty.$$

定义 1.5 若 $X(\omega)$ 是概率空间 $(\Omega, \mathcal{F}, P)$ 上的可积随机变量，则称

$$E(X) := \int_\Omega X(\omega)P(d\omega)$$

为随机变量 $X(\omega)$ 的数学期望.

随机变量期望的性质

1. 若 c 为常数，则 $E(c) = c$;
2. 若 $X \geq 0$, 则 $E(X) \geq 0$
3. 对任意的常数 a, 有 $E(aX) = aE(X)$;
4. 对任意两个函数 $g_1(x), g_2(x)$ 有
$$E(g_1(X) \pm g_2(X)) = E(g_1(X)) \pm E(g_2(X))$$
5. 对任意两个随机变量 X, Y, 有 $E(X+Y) = E(X) + E(Y)$;
6. 一般的，对任意两个随机变量 X, Y 及任给两个常数 a, b, 有
$$E(aX + bY) = aE(X) + bE(Y);$$
7. 更一般的有
$$E(\sum_{i=1}^{n} a_i X_i + b) = \sum_{i=1}^{n} a_i EX_i + b$$

定理 1.1 设 X 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，$\mathbf{F}$ 为随机变量 X 的分布，则对任意的 Borel 函数 $f(x)$ 有

$$\int_{\Omega} f(X(\omega))P(d\omega) = \int_{\mathbb{R}} f(x)\mathbf{F}(dx) \left(= \int_{\mathbb{R}} f(x)dF(x) \right).$$

特别的，若 $f(x) = x$，则有

$$E(X) = \int_{\Omega} X(\omega)P(d\omega) = \int_{\mathbb{R}} x\mathbf{F}(dx) = \int_{\mathbb{R}} xdF(x)$$

定理 1.2 设 X, Y 为概率空间 $(\Omega, \mathcal{F}, P)$ 上两个相互独立的随机变量，且 X, Y 均可积，则乘积 XY 也可积，且

$$E(XY) = E(X)E(Y)$$

定理 1.3 概率空间 $(\Omega, \mathcal{F}, P)$ 上两个随机变量 X, Y 相互独立的充要条件是，对于使得 $f(X)$ 与 $g(Y)$ 均可积的任何 Borel 函数 f, g 均有

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)].$$

定理 1.4 若随机变量 X 的分布函数为 $F(x)$, 则 X 的某一函数 $g(X)$ 的期望为

$$
\begin{aligned}
E[g(X)] &= \int_{-\infty}^{+\infty} g(x)dF(x) \\
&= \begin{cases} \sum_i g(x_i)[F(x_i) - F(x_{i-1})] = \sum_i g(x_i)P(X = x_i), \\ \\ \int_{-\infty}^{+\infty} g(x)p(x)dx. \end{cases}
\end{aligned}
$$

定理 1.5 若随机向量 $(X_1, \cdots, X_n)$ 的分布函数为 $F(x_1, \cdots, x_n)$，$g(x_1, \cdots, x_n)$ 为 n 元 Borel 函数，则

$$E[g(X_1, \cdots, X_n)] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, \cdots, x_n) dF(x_1, \cdots, x_n).$$

特别的，

$$E(X_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i dF(x_1, \cdots, x_n) = \int_{-\infty}^{+\infty} x_i dF_i(x_i)$$

其中 $F_i(x_i)$ 为 X_i 的分布函数. 更进一步，如果 $(X_1, \cdots, X_n)$ 为连续型随机向量即具有联合分布密度 $p(x_1, \cdots, x_n)$，则

$$E[g(X_1, \cdots, X_n)] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, \cdots, x_n) p(x_1, \cdots, x_n) dx_1 \cdots dx_n$$

$$E(X_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i p(x_1, \cdots, x_n) dx_1 \cdots dx_n = \int_{-\infty}^{+\infty} x_i p_{X_i}(x_i) dx_i$$

定义 1.6 假设随机向量 $(X_1, \cdots, X_n)$ 的每个分量 X_i 的数学期望都存在，则称

$$E(X) = (EX_1, EX_2, \cdots, EX_n)$$

为随机向量 X 的数学期望向量，简称为 X 的数学期望。这里

$$E(X_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i dF(x_1, \cdots, x_n) = \int_{-\infty}^{+\infty} x_i dF_i(x_i)$$

这里 $F_i(x_i)$ 为 X_i 的分布函数.

定理 1.6 (单调收敛定理) 若随机变量序列 $\{X_n\}$ 满足条件

$$0 \leq X_1(\omega) \leq X_2(\omega) \leq \cdots \leq X_n(\omega) \uparrow X(\omega), \omega \in \Omega,$$

则

$$\begin{aligned} \int_\Omega \lim_{n\to\infty} X_n(\omega)P(d\omega) &= \lim_{n\to\infty} \int_\Omega X_n(\omega)P(d\omega), \\ \text{i.e.} \quad E(\lim_{n\to\infty} X_n) &= \lim_{n\to\infty} E(X_n). \end{aligned}$$

- 设 $\{X_{nk}\}_{k\geqslant 1}$ 为定义 $E(X_n)$ 的简单随机变量列，且满足

$$0 \leqslant X_{11} \leqslant X_{12} \leqslant \cdots \leqslant X_{1k} \uparrow X_1$$

$$0 \leqslant X_{21} \leqslant X_{22} \leqslant \cdots \leqslant X_{2k} \uparrow X_2$$

$$\vdots$$

$$0 \leqslant X_{n1} \leqslant X_{n2} \leqslant \cdots \leqslant X_{nk} \uparrow X_n$$

$$\lim_{k\to\infty} E(X_{nk}) = E(X_n)$$

- 令 $\widetilde{X}_k = \max_{1\leqslant i\leqslant k} X_{ik}$，则 $\widetilde{X}_k$ 为非降简单随机变量序列，

$$X_{nk} \leqslant \widetilde{X}_k \leqslant X_k, \quad E(X_{nk}) \leq E\left(\widetilde{X}_k\right) \leq E(X_k)$$

单调收敛定理的证明 II

- 下证 $\lim_{k\to\infty}\widetilde{X}_k = X$, $\lim_{k\to\infty}E(\widetilde{X}_k) = \lim_{n\to\infty}E(X_n)$
  - 固定 n，令 $k\to+\infty$ 可得

$$X_n \le \lim_{k\to+\infty}\widetilde{X}_k \le X, \qquad E\left(X_n\right) \le \lim_{k\to+\infty}E\left(\widetilde{X}_k\right) \le \lim_{k\to+\infty}E(X_k)$$

  - 再令 $n\to+\infty$ 可得

$$\lim_{k\to\infty}\widetilde{X}_k = X, \qquad \lim_{k\to\infty}E(\widetilde{X}_k) = \lim_{n\to\infty}E(X_n)$$

- 结合 $E(\lim_{n\to\infty}X_n) = E(X) := \lim_{n\to\infty}E(\widetilde{X}_k)$ 可知定理得证.

期望的极限性质：Fatou 引理

定理 1.7 (Fatou 引理) 若 $\{X_n\}$ 是一随机变量序列，

- 如果存在可积随机变量 $\underline{Y}$ 使得 $X_n \geq \underline{Y}$, 则有

$$\begin{aligned} \int_\Omega \liminf_{n\to\infty} X_n(\omega)P(d\omega) &\leq \liminf_{n\to\infty} \int_\Omega X_n(\omega)P(d\omega), \\ \text{i.e.} \quad E(\liminf_{n\to\infty} X_n) &\leq \liminf_{n\to\infty} E(X_n). \end{aligned}$$

- 如果存在可积随机变量 $\overline{Y}$ 使得 $X_n \leq \overline{Y}$, 则有

$$\begin{aligned} \int_\Omega \limsup_{n\to\infty} X_n(\omega))P(d\omega) &\geq \limsup_{n\to\infty} \int_\Omega X_n(\omega))P(d\omega), \\ \text{i.e.} \quad E(\limsup_{n\to\infty} X_n) &\geq \limsup_{n\to\infty} E(X_n). \end{aligned}$$

Fatou 引理证明

- 如果 $\liminf_{n\to\infty} E(X_n) = +\infty$, 则不等式显然成立. 以下设 $\liminf_{n\to\infty} E(X_n)$ 有限.
- 令 $Y_n = \inf_{k\geq n}(X_k - \underline{Y}), n \geq 1$, 则 $\{Y_n\}$ 为非负随机变量的上升序列，且以 $\liminf_{n\to\infty}(X_n - \underline{Y})$ 为极限.
- 由单调收敛定理可得:

$$E\left[\liminf_{\to\infty}(X_n - \underline{Y})\right] = \lim_{n\to\infty} E\left[\inf_{k\geq n}(X_k - \underline{Y})\right] \leq \liminf_{n\to\infty} E(X_n - \underline{Y}).$$

- 消去有限的 $E(\underline{Y})$ 可证得结论
- 对 $\{-X_n\}$ 用下确界的结论可证得上确界的结论.

期望的极限性质：控制收敛定理

定理 1.8 (控制收敛定理) 设 $\{X_n\}$ 是一随机变量序列。如果存在可积随机变量 Y 使得 $|X_n| \le Y$, 且 $\lim_{n\to\infty} X_n(\omega) = X(\omega)$, 则随机变量 $X(\omega)$ 可积，且有

$$\begin{aligned} \int_\Omega \lim_{n\to\infty} X_n(\omega)P(d\omega) &= \lim_{n\to\infty}\int_\Omega X_n(\omega)P(d\omega), \\ \text{i.e.}\quad E(X) = E(\lim_{n\to\infty} X_n) &= \lim_{n\to\infty} E(X_n). \end{aligned}$$

- $|X_n| \le Y$ 蕴含 $|X| \le Y$, 故 X 可积.
- 取 $\underline{Y} = -Y$ 与 $\overline{Y} = Y$, 用 Fatou 引理便得到

$$E(X) \le \liminf_{n \to \infty} E\left(X_n\right) \le \limsup_{n \to \infty} E\left(X_n\right) \le E(X)$$

- 故 $\lim_n E\left(X_n\right)$ 存在且等于 $E(X)$, 定理至此得证.

定义 1.7 若随机变量 X^2 的期望 $E(X^2)$ 存在，则称

$$D(X) := E\left[(X - EX)^2\right]$$

为随机变量 X 的方差。有时我们也用 $Var(X)$ 来表示 X 的方差。称方差 $D(X)$ 的正平方根 $\sqrt{D(X)}$ 为随机变量 X 的标准差，记为 $\sigma(X)$ 或 σ_X.

定理 1.9 假设随机变量 X 的方差存在，则

$$\begin{aligned} D(X) &= E[(X - EX)^2] = \int_{-\infty}^{+\infty} (x - EX)^2 dF(x) \\ &= \begin{cases} \sum_i (x_i - EX)^2 [F(x_i) - F(x_{i-1})] = \sum_i (x_i - EX)^2 P(X = x_i) \\ \\ \int_{-\infty}^{+\infty} (x - EX)^2 p(x) dx \end{cases} \end{aligned}$$

1. $D(X) = E[(X - EX)^2] \geq 0$, $D(X) = 0$ 当且仅当 $P(X = EX) = 1$;
2. 常数的方差为 0 即 $D(c) = 0$, 其中 c 为常数;
3. $D(X) = E(X^2) - (EX)^2$: 事实上，根据方差的定义及期望的线性性质，我们有

$$
\begin{aligned}
D(X) &= E[(X - EX)^2] = E[X^2 - 2X \cdot EX + (EX)^2] \\
&= E(X^2) - 2EX \cdot EX + (EX)^2 \\
&= E(X^2) - (EX)^2
\end{aligned}
$$

方差的性质 (续)

4. 若 $E(X^2)=0$, 则 $E(X)=0, D(X)=0$. 事实上

$$0 \le D(X) = E(X^2) - (EX)^2 = -(EX)^2 \le 0$$

5. 若 a, b 为常数，则 $D(aX+b) = a^2 D(X)$, 事实上，

$$\begin{aligned} D(aX+b) &= E(aX+b-E(aX+b))^2 = E(aX-aEX)^2 \\ &= E[a^2(X-EX)^2] = a^2 E[(X-EX)^2] = a^2 D(X) \end{aligned}$$

6. $f(c) := E(X-c)^2$ 当且仅当 $c = E(X)$ 时取到最小值:由

$$f(c) = E(X-EX+EX-c)^2 = E(X-EX)^2 + [EX-c]^2$$

可得.

切比雪夫 (Chebyshev) 不等式

定理 1.10 设随机变量 X 的期望与方差均存在，则对任意的 $\epsilon > 0$ 有

$$P(|X - EX| \geq \epsilon) \leq \frac{D(X)}{\epsilon^2}$$

证明: 记 $a = EX$, 则

$$\begin{aligned} P(|X - a| \geq \epsilon) &= \int_{\{x:|x-a|\geq\epsilon\}} dF(x) \leq \int_{\{x:|x-a|\geq\epsilon\}} \frac{(x-a)^2}{\epsilon^2} dF(x) \\ &\leq \frac{1}{\epsilon^2} \int_{-\infty}^{+\infty} (x-a)^2 dF(x) = \frac{D(X)}{\epsilon^2} \end{aligned}$$

定理 1.11 若随机变量 X 的方差存在，则

$$D(X) = 0 \Leftrightarrow P(X = EX) = 1.$$

特别的，如果 $E(X^2) = 0$, 则 $E(X) = 0, D(X) = 0$, 从而 $P(X = 0) = 1$.

证明: 充分性显然，下面证必要性。设 $D(X) = 0$, 此时 EX 存在。注意到

$$\{|X - EX| > 0\} = \cup_{n=1}^{\infty} \{|X - EX| \geq \frac{1}{n}\}$$

故

$$\begin{aligned} P(|X - EX| > 0) &= P(\cup_{n=1}^{\infty}\{|X - EX| \geq \frac{1}{n}\}) \\ &\leq \sum_{n=1}^{\infty} P(|X - EX| \geq \frac{1}{n}) \leq \sum_{n=1}^{\infty} \frac{D(X)}{(1/n)^2} = 0 \end{aligned}$$

从而

$$P(X = EX) = 1 - P(|X - EX| > 0) = 1 - 0 = 1$$

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{n} kP(X=k) = \sum_{k=0}^{n} \frac{k\cdot n!}{k!(n-k)!}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n\cdot (n-1)!}{(k-1)!(n-1-(k-1))!}pp^{k-1}(1-p)^{n-1-(k-1)} \\
&= np\sum_{k=1}^{n} C_{n-1}^{k-1}p^{k-1}(1-p)^{n-1-(k-1)} \\
&= np
\end{aligned}
$$

特别的，对于两点分布 $(B(1,p))$ 随机变量 X, 有 $E(X)=p$. 事实上对于两点分布其期望为: $E(X)=0\cdot P(X=0)+1\cdot P(X=1)=p$.

计算二项分布期望的另一方法

前面在讲二项分布时，我们知道:

$$X = X_1 + X_2 + \cdots + X_n$$

其中 $X_k, k = 1, \cdots, n$ 为第 k 次伯努利试验成功的次数，显然服从两点分布。故

$$\begin{aligned} EX_i &= 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = p \\ EX_i^2 &= 0^2 \cdot P(X_i = 0) + 1^2 \cdot P(X_i = 1) = p \\ D(X_i) &= EX_i^2 - (EX_i)^2 = p - p^2 = p(1-p) \end{aligned}$$

再由期望的线性性知

$$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$$

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 P(X=k) = \sum_{k=1}^{n} k(k-1+1)C_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k(k-1)C_n^k p^k (1-p)^{n-k} + \sum_{k=1}^{n} kC_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} k(k-1)C_n^k p^k (1-p)^{n-k} + np \\
&= n(n-1)p^2 \sum_{k=2}^{n} C_{n-2}^{k-2} p^{k-2} (1-p)^{(n-2)-(k-2)} + np \\
&= n(n-1)p^2 + np
\end{aligned}$$

故

$$D(X) = E(X^2) - (EX)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

特别的，两点分布的方差为 $p(1-p)$.

泊松分布的期望与方差: $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=1}^{\infty} \lambda\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \lambda \\
E(X^2) &= \sum_{k=0}^{\infty} k^2P(X=k) = \sum_{k=1}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=1}^{\infty}[(k-1)+1]\frac{\lambda^k}{(k-1)!}e^{-\lambda} \\
&= \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda^2+\lambda \\
D(X) &= E(X^2)-(EX)^2 = \lambda^2+\lambda-\lambda^2 = \lambda
\end{aligned}
$$

几何分布的期望与方差:$P(X=k)=(1-p)^{k-1}p$

$$
\begin{aligned}
E(X) &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{\infty}\sum_{n=1}^{k}(1-p)^{k-1} = p\sum_{n=1}^{\infty}\sum_{k=n}^{\infty}(1-p)^{k-1} \\
&= p\sum_{n=1}^{\infty}(1-p)^{n-1}\frac{1}{1-(1-p)} = \sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{1}{p} \\
E(X^2) &= \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = p\left[\sum_{k=1}^{\infty} k(k+1)(1-p)^{k-1} - \sum_{k=1}^{\infty} k(1-p)^{k-1}\right] \\
&= 2p\sum_{k=1}^{\infty}\sum_{n=1}^{k} n(1-p)^{k-1} - p\sum_{k=1}^{\infty} k(1-p)^{k-1} \\
&= 2p\sum_{n=1}^{\infty} n\sum_{k=n}^{\infty}(1-p)^{k-1} - \frac{1}{p} = 2p\sum_{n=1}^{\infty} n(1-p)^{n-1}/p - \frac{1}{p} \\
&= \frac{2}{p^2} - \frac{1}{p} \\
D(X) &= E(X^2) - (EX)^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}
\end{aligned}
$$

$$
\begin{aligned}
E(X) &= \sum_{k=r}^{\infty} kC_{k-1}^{r-1}p^r(1-p)^{k-r} = \sum_{k=r}^{\infty} \frac{r}{p}\frac{k\cdot(k-1)!}{r\cdot(r-1)!(k-r)!}p^{r+1}(1-p)^{k-r} \\
&= \sum_{k=r}^{\infty} \frac{r}{p}C_{k+1-1}^{r+1-1}p^{r+1}(1-p)^{k-r} = \frac{r}{p}\sum_{l=r+1}^{\infty} C_{l-1}^{r+1-1}p^{r+1}(1-p)^{l-(r+1)} = \frac{r}{p} \\
E(X^2) &= \sum_{k=r}^{\infty} k^2C_{k-1}^{r-1}p^r(1-p)^{k-r} = \sum_{k=r}^{\infty} k(k+1-1)C_{k-1}^{r-1}p^r(1-p)^{k-r} \\
&= \sum_{k=r}^{\infty} k(k+1)C_{k-1}^{r-1}p^r(1-p)^{k-r} - \sum_{k=r}^{\infty} kC_{k-1}^{r-1}p^r(1-p)^{k-r} \\
&= \sum_{k=r}^{\infty} \frac{r(r+1)}{p^2}C_{k+2-1}^{r+2-1}p^{r+2}(1-p)^{(k+2)-(r+2)} - \frac{r}{p} \\
&= \frac{r(r+1)}{p^2} - \frac{r}{p} \\
D(X) &= E(X^2)-(EX)^2 = \frac{r(r+1)}{p^2} - \frac{r}{p} - \left(\frac{r}{p}\right)^2 = \frac{r(1-p)}{p^2}
\end{aligned}
$$

均匀分布的期望与方差:$X \sim U(a,b)$

- 若 $X \sim U(a,b)$, 即 $p(x) = \begin{cases} \dfrac{1}{b-a}, & x \in (a,b) \\ 0, & 其他 \end{cases}$, 故

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} xp(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{a+b}{2} \\
E(X^2) &= \int_{-\infty}^{+\infty} x^2p(x)dx = \int_a^b x^2\frac{1}{b-a}dx = \frac{a^2+ab+b^2}{3} \\
D(X) &= E(X^2) - (EX)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}
\end{aligned}
$$

标准正态分布的期望与方差:$X \sim N(0,1)$

- 若 $X \sim N(0,1)$ 即 $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 则

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} xp(x)dx = \int_{-\infty}^{+\infty} x\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0 \\
D(X) &= E[(X-EX)^2] = E(X^2) = \int_{-\infty}^{+\infty} x^2 p(x)dx \\
&= \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} xd(-e^{-\frac{x^2}{2}}) \\
&= \frac{1}{\sqrt{2\pi}} \left(-xe^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \right) \\
&= 1
\end{aligned}
$$

- 若 $X \sim N(\mu, \sigma^2)$, 则易知 $\dfrac{X-\mu}{\sigma} \sim N(0,1)$, 故

$$
\begin{aligned}
E(X) &= E(\sigma \cdot \frac{X-\mu}{\sigma} + \mu) = \sigma E(\frac{X-\mu}{\sigma}) + \mu = \mu \\
D(X) &= D(\sigma \cdot \frac{X-\mu}{\sigma} + \mu) = \sigma^2 D(\frac{X-\mu}{\sigma}) = \sigma^2
\end{aligned}
$$

Gamma 分布的期望

- 若 $X\sim\Gamma(\alpha,\lambda)$ 即 $p(x)=\begin{cases}\dfrac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$，故

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} xp(x)dx=\int_0^{+\infty} x\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}dx\\
&= \int_0^{+\infty}\frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)}\frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}x^{\alpha+1-1}e^{-\lambda x}dx=\frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)}=\frac{\alpha}{\lambda};\\
E(X^2) &= \int_{-\infty}^{+\infty} x^2p(x)dx=\int_0^{+\infty} x^2\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}dx\\
&= \int_0^{+\infty}\frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)}\frac{\lambda^{\alpha+2}}{\Gamma(\alpha+2)}x^{\alpha+2-1}e^{-\lambda x}dx=\frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)}=\frac{\alpha(\alpha+1)}{\lambda^2}\\
D(X) &= E(X^2)-(EX)^2=\frac{\alpha(\alpha+1)}{\lambda^2}-\left(\frac{\alpha}{\lambda}\right)^2=\frac{\alpha}{\lambda^2}
\end{aligned}
$$

指数分布与 χ^2 分布的期望与方差

- 若 $X \sim \Gamma(1,\lambda)$ 分布即 X 服从指数分布，此时 $\alpha = 1$

$$
\begin{aligned}
E(X) &= \frac{\alpha}{\lambda} = \frac{1}{\lambda} \\
E(X^2) &= \frac{\alpha(\alpha+1)}{\lambda^2} = \frac{2}{\lambda^2} \\
D(X) &= \frac{\alpha}{\lambda^2} = \frac{1}{\lambda^2}
\end{aligned}
$$

- 若 $X \sim \Gamma(\frac{n}{2},\frac{1}{2})$ 即 X 服从 $\chi^2(n)$ 分布，此时 $\alpha = n/2, \lambda = 1/2$

$$
\begin{aligned}
E(X) &= \frac{\alpha}{\lambda} = n \\
E(X^2) &= \frac{\alpha(\alpha+1)}{\lambda^2} = n(n+2) \\
D(X) &= \frac{\alpha}{\lambda^2} = 2n
\end{aligned}
$$

第十八讲：协方差与相关系数

Email: xzhangseu@seu.edu.cn

东南大学　　数学学院

定义 1.8 对于随机向量 $X=(X_1,\cdots,X_n)$, 类似于随机向量期望的定义, 我们定义随机向量 X 的方差为 $D(X):=(D(X_1),D(X_2),\cdots,D(X_n))$.

定义 1.9 设随机向量 $(X_1,\cdots,X_n)$ 的每个分量 X_i 的方差均存在，称

$$cov(X_i,X_j):=E[(X_i-EX_i)(X_j-EX_j)],i,j=1,2,\cdots,n$$

为 X_i 与 X_j 的协方差. 而将协方差构成的 $n\times n$ 方阵

$$B=(b_{ij}),\quad b_{ij}=cov(X_i,X_j)$$

称为随机向量的协方差阵。一般来说，

- 若 $cov(X_i,X_j)>0$ 时，称 X_i 与 X_j 正相关;
- 若 $cov(X_i,X_j)<0$ 时，称 X_i 与 X_j 负相关;
- 若 $cov(X_i,X_j)=0$ 时，称 X_i 与 X_j 不相关或零相关.

- $cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j)$: 事实上

$$
\begin{aligned}
cov(X_i, X_j) &= E[(X_i - EX_i)(X_j - EX_j)] \\
&= E\Big[X_iX_j - X_iE(X_j) - X_jE(X_i) + E(X_i)E(X_j)\Big] \\
&= E(X_iX_j) - E(X_i)E(X_j)
\end{aligned}
$$

- 若 a 为常数，则 $cov(X_i, a) = 0$;
- 对任意常数 a, b, 有 $cov(aX_i, bX_j) = ab \cdot cov(X_i, X_j)$;
- $cov(X_i + X_j, X_k) = cov(X_i, X_k) + cov(X_j, X_k)$.

- $cov(X_i,X_i)=E(X_i-EX_i)^2=D(X_i)$;
- $D(X_1+X_2+\cdots+X_n)=\sum_{i=1}^n D(X_i)+2\sum_{1\le i<j\le n}cov(X_i,X_j)$: 事实上

$$
\begin{aligned}
D(\sum_{i=1}^n X_i) &= E(\sum_{i=1}^n X_i - E\sum_{i=1}^n X_i)^2 = E(\sum_{i=1}^n (X_i-EX_i))^2 \\
&= E\Big[\sum_{i=1}^n (X_i-EX_i)^2 + 2\sum_{1\le i<j\le n}(X_i-EX_i)(X_j-EX_j)\Big] \\
&= \sum_{i=1}^n D(X_i) + 2\sum_{1\le i<j\le n} cov(X_i,X_j)
\end{aligned}
$$

- 特别的,

$$D(X_i+X_j)=D(X_i)+D(X_j)+2cov(X_i,X_j)$$

- 对称性: $cov(X_i, X_j) = cov(X_j, X_i)$, 由定义可直接推得，从而协方差阵 B 是对称矩阵;
- 协方差矩阵 B 是非负定矩阵，事实上，对任意向量 $y = (y_1, y_2, \cdots, y_n)$,

$$
\begin{aligned}
yBy' &= \sum_{i,j} y_i y_j b_{ij} = \sum_{i,j} y_i y_j E[(X_i - EX_i)(X_j - EX_j)] \\
&= E\sum_{i,j} y_i(X_i - EX_i) \cdot y_j(X_j - EX_j) \\
&= E[\sum_i y_i(X_i - EX_i)]^2 \geq 0
\end{aligned}
$$

例 1.1 设二维随机变量 (X,Y) 的联合密度为

$$p(x,y)=\begin{cases}3x, & 0<y<x<1,\\ 0, & \text{其他}\end{cases}$$

试求 $cov(X,Y)$.

解: 由 $cov(X,Y)=E(XY)-EXEY$ 知，我们只需计算 $E(XY),EX,EY$ 的值.

$$\begin{aligned}
EX &= \int_{-\infty}^{+\infty} xp_X(x)dx=\int_{-\infty}^{+\infty} x\int_{-\infty}^{+\infty} p(x,y)dydx\\
&= \int_0^1 x\int_0^x 3xdydx=\int_0^1 3x^3dx=3/4\\
EY &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yp(x,y)dydx=\int_0^1\int_0^x 3xydydx=3/8\\
E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyp(x,y)dydx=\int_0^1\int_0^x xy\cdot 3xdydx=3/10\\
cov(X,Y) &= \frac{3}{10}-\frac{3}{4}\cdot\frac{3}{8}=\frac{3}{160}>0
\end{aligned}$$

例 1.2 设二维随机变量 (X, Y) 的联合密度为

$$p(x,y)=\begin{cases}\dfrac{x+y}{3}, & 0<x<1, 0<y<2,\\ 0, & \text{其他}\end{cases}$$

试求 $D(2X-3Y+8)$.

解: 注意到

$$\begin{aligned}D(2X-3Y+8) &= D(2X-3Y)=D(2X)+D(-3Y)+2cov(2X,-3Y)\\ &= 2^2D(X)+(-3)^2D(Y)+2\cdot 2\cdot(-3)cov(X,Y)\\ &= 4D(X)+9D(Y)-12cov(X,Y)\end{aligned}$$

所以我们需要计算:$E(X), E(X^2), E(Y), E(Y^2), E(XY)$.

$$\begin{aligned}E(X) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xp(x,y)dydx=\int_0^1\int_0^2 x\frac{x+y}{3}dydx=5/9\\ E(X^2) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2p(x,y)dydx=\int_0^1\int_0^2 x^2\frac{x+y}{3}dydx=7/18\end{aligned}$$

类似的可计算

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yp(x,y)dydx = \int_0^1\int_0^2 y\frac{x+y}{3}dydx = 11/9 \\
E(Y^2) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2p(x,y)dydx = \int_0^1\int_0^2 y^2\frac{x+y}{3}dydx = 16/9 \\
E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyp(x,y)dydx = \int_0^1\int_0^2 xy\frac{x+y}{3}dydx = 2/3
\end{aligned}
$$

从而

$$
\begin{aligned}
D(X) &= 7/18 - (5/9)^2 = 13/162, \\
D(Y) &= 16/9 - (11/9)^2 = 23/81, \\
cov(X,Y) &= 2/3 - 5/9 \cdot 11/9 = -1/81 \\
D(2X-3Y+8) &= 4 \cdot 13/162 + 9 \cdot 23/81 - 12 \cdot (-1/81) = 245/81
\end{aligned}
$$

若 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 即其联合分布密度为

$$
\begin{aligned}
p(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times \\
&\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}
\end{aligned}
$$

则由上一章的知识知:$X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$, 故

$$
\begin{aligned}
&E(X)=\mu_1, \quad E(Y)=\mu_2 \\
&b_{11}=cov(X,X)=D(X)=\sigma_1^2, \\
&b_{22}=cov(Y,Y)=D(Y)=\sigma_2^2
\end{aligned}
$$

$$b_{12} = cov(X,Y) = E[(X-EX)(Y-EY)] = \iint (x-\mu_1)(y-\mu_2)p(x,y)dxdy$$

$$\xlongequal[v=(y-\mu_2)/\sigma_2]{u=(x-\mu_1)/\sigma_1} \iint \frac{\sigma_1\sigma_2 uv}{2\pi\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)\}dudv$$

$$= \iint \frac{\sigma_1\sigma_2 uv}{2\pi\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}\left[(u-\rho v)^2+(1-\rho^2)v^2\right]\}dudv$$

$$\xlongequal[t=v]{s=(u-\rho v)/\sqrt{1-\rho^2}} \frac{\sigma_1\sigma_2}{2\pi} \iint (\sqrt{1-\rho^2}st+\rho t^2)\exp\{-\frac{s^2+t^2}{2}\}dsdt$$

$$= \frac{\sigma_1\sigma_2}{2\pi}\left(\sqrt{1-\rho^2}\int_{-\infty}^{+\infty} se^{-\frac{s^2}{2}}ds\int_{-\infty}^{+\infty} te^{-\frac{t^2}{2}}dt + \rho\int_{-\infty}^{+\infty} e^{-\frac{s^2}{2}}ds\int_{-\infty}^{+\infty} t^2e^{-\frac{t^2}{2}}dt\right)$$

$$= \frac{\sigma_1\sigma_2}{2\pi}(\sqrt{1-\rho^2}\cdot 0+\rho\sqrt{2\pi}\cdot\sqrt{2\pi}) = \rho\sigma_1\sigma_2$$

故二维正态分布的协方差矩阵为

$$B = \begin{pmatrix} \sigma_1^2, & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2, & \sigma_2^2 \end{pmatrix}$$

定义 1.10 设 X, Y 为方差存在的两个随机变量，则称

$$r := \frac{cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为 X 与 Y 的相关系数.

相关系数的性质: $r := \dfrac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$

- $r=0 \Leftrightarrow cov(X,Y)=0$ 即 X,Y 不相关;
- 若记 $X^* := \dfrac{X-EX}{\sqrt{D(X)}}$, $\quad Y^* := \dfrac{Y-EY}{\sqrt{D(Y)}}$, 则 $EX^*=EY^*=0$, 从而

$$
\begin{aligned}
r &:= \frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{E[(X-EX)(Y-EY)]}{\sqrt{D(X)}\sqrt{D(Y)}} \\
&= E(X^*Y^*) = E[(X^*-EX^*)(Y^*-EY^*)] \\
&= cov(X^*,Y^*)
\end{aligned}
$$

- $|r|\le 1$, 且
  - $r=1$ 当且仅当 $P(X^*=Y^*)=1$;
  - $r=-1$ 当且仅当 $P(X^*=-Y^*)=1$.

柯西 - 施瓦兹 (Cauchy-Schwarz) 不等式

定理 1.12 对任意的随机变量 X 与 Y 都有

$$[E(XY)]^2 \leq E(X^2)E(Y^2)$$

等式成立当且仅当存在常数 t_0 使得

$$P(Y = t_0 X) = 1$$

证明: 对任意的实数 t, 定义

$$u(t) := E(tX - Y)^2 = t^2E(X^2) - 2tE(XY) + E(Y^2).$$

显然，对一切的 $t \in R$, $u(t) \ge 0$, 故方程 $u(t) = 0$ 或者没有实根或者有重根，从而

$$\Delta = [2E(XY)]^2 - 4E(X^2)E(Y^2) \le 0$$

即

$$[E(XY)]^2 \le E(X^2)E(Y^2)$$

上述不等式等号成立，当且仅当 $\Delta = 0$ 即 $u(t) = 0$ 存在一个重根 t_0, 这时

$$u(t_0) = E(t_0X - Y)^2 = 0$$

从而由可知

$$P(t_0X - Y = 0) = 1 \Leftrightarrow P(Y = t_0X) = 1.$$

相关系数绝对值小于等于 1 的证明

- 首先由 $X^* = \dfrac{X - EX}{\sqrt{D(X)}}, Y^* = \dfrac{Y - EY}{\sqrt{D(Y)}}$ 知:

$$
\begin{aligned}
EX^* &= EY^* = 0 \\
D(X^*) &= D\left(\frac{X - EX}{\sqrt{D(X)}}\right) = \frac{1}{D(X)}D(X - EX) = \frac{1}{D(X)}D(X) = 1 \\
D(Y^*) &= D\left(\frac{Y - EY}{\sqrt{D(Y)}}\right) = \frac{1}{D(Y)}D(Y - EY) = \frac{1}{D(Y)}D(Y) = 1
\end{aligned}
$$

- 其次，由 $r = cov(X^*, Y^*)$ 知

$$
\begin{aligned}
|r| &= |cov(X^*, Y^*)| = |E(X^*Y^*) - E(X^*)E(Y^*)| \le \sqrt{E[(X^*)^2]E[(Y^*)^2]} \\
&= \sqrt{D(X^*)D(Y^*)} = 1;
\end{aligned}
$$

- $|r| = 1$ 当且仅当存在 t_0 使得 $P(Y^* = t_0X^*) = 1$, 而此时有

$$r = E(X^*Y^*) = t_0E[(X^*)^2] = t_0$$

随机变量不相关时期望与方差的性质

定理 1.13 对于随机变量 X, Y, 下面四个事实是等价的:

1. X, Y 不相关;
2. $r = 0$ 即相关系数为 0;
3. $E(XY) = E(X)E(Y)$;
4. $D(X + Y) = D(X) + D(Y)$.

证明: (1) 与 (2) 等价是显然的，根据定义即可得. 由于

$$cov(X, Y) = E(XY) - E(X)E(Y)$$

故 $cov(X, Y) = 0$ 当且仅当 $E(XY) = E(X)E(Y)$, 即 (1) 和 (3) 等价. 另外，又由于

$$D(X + Y) = D(X) + D(Y) + 2cov(X, Y)$$

故 $cov(X, Y) = 0$ 当且仅当 $D(X + Y) = D(X) + D(Y)$, 即 (1) 与 (4) 等价.

独立与不相关的联系：独立必定不相关

定理 1.14 若 X, Y 独立，则 X 与 Y 不相关.

证明: 我们仅对连续型随机变量给出证明。因为 X, Y 独立，故其联合分布密度 $p(x,y) = p_X(x)p_Y(y)$, 从而

$$\begin{aligned} E(XY) &= \iint xyp(x,y)dxdy = \iint xyp_X(x)p_Y(y)dxdy \\ &= \int xp_X(x)dx \int yp_Y(y)dy = E(X)E(Y) \end{aligned}$$

从而 X, Y 不相关.

定理 1.15 若 X, Y 独立，则

$$E(XY) = E(X)E(Y), \quad D(X+Y) = D(X) + D(Y)$$

更一般的，我们有若 $X_1, \cdots, X_n$ 为相互独立的随机变量，则.

$$\begin{aligned} E(X_1 \cdots X_n) &= E(X_1)E(X_2)\cdots E(X_n); \\ D(\sum_{i=1}^{n} X_i) &= \sum_{i=1}^{n} D(X_i) \end{aligned}$$

独立与不相关的联系：不相关未必独立

例 1.3 设随机变量 $X \sim N(0, \sigma^2)$, 并令 $Y = X^2$, 显然 X, Y 不独立，但是此时 X, Y 不相关。事实上，此时

$$cov(X, Y) = E(X \cdot X^2) - E(X)E(X^2) = 0.$$

例 1.4 设 θ 为 $[0,2\pi]$ 上的均匀分布，a 为一固定常数，令

$$X=\cos\theta,\quad Y=\cos(\theta+a).$$

则我们有

$$\begin{aligned}
E(X) &= \int_0^{2\pi}\cos t\cdot\frac{1}{2\pi}dt=0,\quad E(Y)=\int_0^{2\pi}\cos(t+a)\cdot\frac{1}{2\pi}dt=0\\
E(X^2) &= \int_0^{2\pi}\cos^2 t\cdot\frac{1}{2\pi}dt=\frac{1}{2},\quad E(Y^2)=\int_0^{2\pi}\cos^2(t+a)\cdot\frac{1}{2\pi}dt=\frac{1}{2}\\
E(XY) &= \int_0^{2\pi}\cos t\cos(t+a)\cdot\frac{1}{2\pi}dt=\frac{1}{2}\cos a
\end{aligned}$$

因此 $r=\cos a$. 故

- $a=0$ 时，$r=1, X=Y$;
- $a=\pi$ 时，$r=-1, X=-Y$;
- $a=\dfrac{\pi}{2}$ 或 $\dfrac{3\pi}{2}$ 时，$r=0$, X, Y 不相关，但此时 $X^2+Y^2=1$, 因此不独立.

定理 1.16 对于二维正态分布 (X, Y),X 与 Y 不相关与 X, Y 独立等价.

证明: 因为独立必然不相关，故我们仅需证在不相关条件下，X, Y 独立即可. 对于二维正态分布，我们有

$$cov(X, Y) = \rho\sigma_1\sigma_2$$

从而 X, Y 不相关，当且仅当 $\rho = 0$, 此时

$$\begin{aligned}
p(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\
&\quad \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\} \cdot \frac{1}{\sqrt{2\pi}\sigma_2}\exp\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\} \\
&= p_X(x)P_Y(y)
\end{aligned}$$

故 X, Y 独立，定理得证.

第十九讲：条件数学期望与母函数

Email: xzhangseu@seu.edu.cn

东南大学　　数学学院

- 在第二章中对于任何有正概率的事件 B, 我们引入了条件概率 $P(\cdot|B)$ 及条件分布 $F(x|B) := P(X \le x|B)$;
- 类似于概率与分布函数，对于上述的条件概率及条件分布，我们也可以定义相应的条件数学期望:

定义 1.11 如果下述积分绝对收敛，则称

$$E(X|B) := \int_{-\infty}^{+\infty} x dF(x|B)$$

为已知事件 B 发生后 X 的条件数学期望.

条件数学期望的两类特殊情形：离散型与连续型

定理 1.17 若 X, Y 均为离散型随机变量，并且条件数学期望定义中的 B 选为 $B := \{Y = y_j\}$ 的形式，则

$$E(X|Y = y_j) = \int_{-\infty}^{+\infty} xdF(x|Y = y_j) = \sum_i x_i P(X = x_i|Y = y_j)$$

定理 1.18 若 X, Y 均为连续型随机变量，并且条件数学期望定义中的 B 选为 $B := \{Y = y\}$ 的形式，则

$$E(X|Y = y) = \int_{-\infty}^{+\infty} xdF(x|Y = y) = \int_{-\infty}^{+\infty} xp(x|y)dx$$

定理 1.19 设 $g(x)$ 为 Borel 函数，则 $g(X)$ 关于 $Y=y$ 的条件期望为

$$E(g(X)|Y=y)=\int_{-\infty}^{+\infty}g(x)dF(x|y)$$

定义 1.12 假设 X 的方差存在，则称

$$\begin{aligned} D(X|Y=y) &:= E[(X-E(X|Y=y))^2|Y=y] \\ &= \int_{-\infty}^{+\infty}(x-E(X|Y=y))^2dF(x|y) \end{aligned}$$

为给定 $Y=y$ 后 X 的条件方差.

例 1.5 若 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, 则

$$E(X|Y=y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2).$$

证明: 在第二章中，我们知道给定 $Y=y$ 下，X 的条件分布服从

$$N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2))$$

故

$$E(X|Y=y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)$$

一般来说，给定 $Y=y$ 后 X 的条件数学期望是 Y 的可能值 y 的函数，记之为

$$\varphi(y) := E(X|Y=y)$$

如果再将 y 用 Y 代回，就得到一个随机变量 $\varphi(Y)$, 相应的，我们记

$$E(X|Y) := \varphi(Y)$$

并称随机变量 $E(X|Y)$ 为 X 关于 Y 的条件数学期望.

条件数学期望的性质：重期望

定理 1.20 设 (X,Y) 为二维随机向量，则对于 Borel 函数 $g(x)$ 有

$$E[E(g(X)|Y)] = E(g(X))$$

证明： 我们仅对二维连续型随机变量给出证明。令 $\varphi(y) := E(g(X)|Y=y)$，则 $E(g(X)|Y) = \varphi(Y)$，从而

$$\begin{aligned}
\varphi(y) &= E(g(X)|Y=y) = \int_{-\infty}^{+\infty} g(x)dF(x|y) \\
&= \int_{-\infty}^{+\infty} g(x)p(x|Y=y)dx = \int_{-\infty}^{+\infty} g(x)\frac{p(x,y)}{p_Y(y)}dx \\
E[E(g(X)|Y)] &= E(\varphi(Y)) = \int_{-\infty}^{+\infty} \varphi(y)p_Y(y)dy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)\frac{p(x,y)}{p_Y(y)}p_Y(y)dxdy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)p(x,y)dxdy = E(g(X))
\end{aligned}$$

重期望的不同表达形式

- 上述定理的重期望公式可写成如下形式

$$
\begin{aligned}
E(g(X)) &= \int_{-\infty}^{+\infty} E[g(X)|Y=y]dF_Y(y) \\
&= \begin{cases} \sum_j E(g(X)|Y=y_j)P(Y=y_j) \\ \int_{-\infty}^{+\infty} E(g(X)|Y=y)p_Y(y)dy \end{cases}
\end{aligned}
$$

- 特别的，若 $g(X)=X$, 则有

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} E[X|Y=y]dF_Y(y) \\
&= \begin{cases} \sum_j E(X|Y=y_j)P(Y=y_j) \\ \int_{-\infty}^{+\infty} E(X|Y=y)p_Y(y)dy \end{cases}
\end{aligned}
$$

巴格达窃贼问题

例 1.6 一个窃贼被关在有 3 个门的地牢中。其中 1 号门通向自由，出 1 号门后走 3 个小时便回到地面；2 号门通向一个地道，在此地道走 5 个小时后返回地牢；3 号门通向一个更长的地道，沿这个地道走 7 个小时后回到地牢。如果窃贼每次选择 3 个门的可能性总相等，求他为获自由而奔走的平均时间.

解: 设窃贼需要走 X 小时到达地面，则 X 的所有可能取值为

$$3, 5+3, 7+3, 5+5+3, 5+7+3, 7+7+3, \cdots,$$

则显然要写出 X 的分布列是困难的，所以无法直接求 $E(X)$. 但是如果我们引入 Y 表示窃贼每次对 3 个门的选择，则

$$P(Y=1)=P(Y=2)=P(Y=3)=1/3$$

$$E(X|Y=1)=3, E(X|Y=2)=5+E(X), E(X|Y=3)=7+E(X)$$

从而由重期望公式可得

$$E(X)=\sum_{i=1}^{3}E(X|Y=i)P(Y=i)=5+\frac{2}{3}E(X)\Rightarrow E(X)=15$$

随机个随机变量和的期望

定理 1.21 设 $X_1, X_2, \cdots$，为一列独立同分布的随机变量，随机变量 N 只取正整数值，且 N 与 $\{X_n\}$ 独立，则

$$E\left(\sum_{i=1}^{N} X_i\right) = E(X_1)E(N)$$

证明: 由重期望公式可知

$$\begin{aligned}
E\left(\sum_{i=1}^{N} X_i\right) &= E\left[E\left(\sum_{i=1}^{N} X_i|N\right)\right] = \sum_{n=1}^{+\infty} E\left(\sum_{i=1}^{N} X_i|N=n\right)P(N=n) \\
&= \sum_{n=1}^{+\infty} E\left(\sum_{i=1}^{n} X_i|N=n\right)P(N=n) = \sum_{n=1}^{+\infty} E\left(\sum_{i=1}^{n} X_i\right)P(N=n) \\
&= \sum_{n=1}^{+\infty} nE(X_1)P(N=n) = E(X_1)E(N)
\end{aligned}$$

定义 1.13 如果 $E|X|^k < +\infty$, 则称 $m_k := E(X^k)$ 为随机变量 X(及其分布) 的 k 阶原点矩。而称 $c_k := E(X-EX)^k$ 为随机变量 (及其分布) 的 k 阶中心矩.

定理 1.22 当 $E|X|^k < +\infty$ 时有，

$$c_k = \sum_{i=0}^{k} C_k^i (-m_1)^{k-i} m_i, \quad m_k = \sum_{i=0}^{k} C_k^i c_{k-i} m_1^i.$$

证明: 注意到

$$\begin{aligned}(X-EX)^k &= \sum_{i=0}^{k} C_k^i X^i (-EX)^{k-i}, \\ X^k = (EX+X-EX)^k &= \sum_{i=0}^{k} C_k^i (EX)^i (X-EX)^{k-i}\end{aligned}$$

对上面两式取期望即可得证.

母函数的定义

定义 1.14 对任何实数列 $\{p_n\}$, 如果幂级数

$$G(s)=\sum_{n=0}^{\infty}p_n s^n \tag{1}$$

的收敛半径 $s_0>0$, 则称 $G(s)$ 为数列 $\{p_n\}$ 的母函数。特别当 $\{p_n\}$ 为某非负整值随机变量 X 的概率分布时，(1) 式至少在区间 $[-1,1]$ 上绝对收敛且一致收敛，此时有

$$G(s)=E(s^X),$$

称此 $G(s)$ 为随机变量 X 或其概率分布 $\{p_n\}$ 的母函数.

- 已知 $G(s)$, 如何确定 $\{p_n\}$?
- 注意到

$$G(0) = p_0, \quad G^{(n)}(s) = n!p_n + \sum_{k=n+1}^{\infty} k(k-1)\cdots(k-n+1)p_k s^{k-n}$$

- 故

$$p_n = \frac{1}{n!}G^{(n)}(0)$$

- 非负整值概率分布 $\{p_n\}$ 与其母函数 $G(s)$ 是一一对应的，母函数可以作为描述这种分布的一种工具.

母函数与随机变量矩的关系

定理 1.23 设非负随机变量 X 的母函数为 $G(s)$, 如果 $E(X)$ 与 $E(X^2)$ 有限，则

$$G'(1) = E(X), \quad G''(1) = E(X^2) - E(X)$$

常见离散型分布的母函数

- Poisson 分布的母函数

$$G(s) = E(s^X) = \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda(s-1)}$$

- 几何分布的母函数

$$G(s) = E(s^X) = \sum_{k=1}^{\infty} s^k q^{k-1} p = \frac{ps}{1-qs}$$

独立和的母函数

定理 1.24 如果 X, Y 为相互独立的随机变量，它们分别有概率分布 $\{a_n\}, \{b_n\}$ 及对应的母函数 $A(s), B(s)$, 则它们的和 $X+Y$ 的母函数为

$$C(s) = A(s)B(s).$$

更进一步，如果 $X_1, \cdots, X_n$ 相互独立，其对应的母函数分别为 $A_1(s)$, $\cdots, A_n(s)$, 则 $X_1 + \cdots + X_n$ 的母函数为

$$C(s) = \Pi_{i=1}^{n} A_i(s)$$

证明: 注意到由 X, Y 相互独立可知 s^X, s^Y 也相互独立，从而

$$C(s) = E(s^{X+Y}) = E(s^X s^Y) = E(s^X)E(s^Y) = A(s)B(s)$$

例 1.7 设 X 服从二项分布 $B(n, p)$, 则 X 的母函数为 $G(s) = (q + ps)^n$.

定理 1.25 设 $\{X_k\}$ 为相互独立同分布的非负整值随机变量序列，其共同的母函数为 $G(s)$. 如果 N 为另一非负整值随机变量，其母函数为 $F(s)$. 则当 N 与每一个 X_k 均独立时，$X=\sum_{k=1}^{N} X_k$ 的母函数为

$$H(s)=F(G(s)).$$

证明: 由重期望公式可得

$$\begin{aligned}
H(s) &= E(s^X)=E(E(s^X|N))=\sum_{n=0}^{\infty}P(N=n)E(s^{\sum_{k=1}^{N}X_k}|N=n)\\
&= \sum_{n=0}^{\infty}P(N=n)E(s^{\sum_{k=1}^{n}X_k}|N=n)=\sum_{n=0}^{\infty}P(N=n)E(s^{\sum_{k=1}^{n}X_k})\\
&= \sum_{n=0}^{\infty}P(N=n)[G(s)]^n=E(G(s)^N)\\
&= F(G(s))
\end{aligned}$$

- 对 $H(s)$ 求导可得

$$H'(s) = F'(G(s))G'(s)$$

- 令 $s=1$ 并注意到 $G(1)=1$ 可得

$$E(X) = E(N)E(X_1)$$

- 再对 $H'(s)$ 求导可得

$$H''(s) = F''(G(s))[G'(s)]^2 + F'(G(s))G''(s)$$

- 令 $s=1$ 可得

$$\begin{aligned} E(X^2) - E(X) &= [E(N^2) - E(N)][E(X_1)]^2 + E(N)[E(X_1^2) - E(X_1)] \\ &= E(N)D(X_1) + E(N^2)[E(X_1)]^2 - E(N)E(X_1) \end{aligned}$$

- 从而

$$D(X) = E(X^2) - [E(X)]^2 = E(N)D(X_1) + D(N)[E(X_1)]^2$$

第二十讲：特征函数

Email: xzhangseu@seu.edu.cn

东南大学　　数学学院

复随机变量及其期望

定义 1.15 若 $X(\omega), Y(\omega)$ 为定义在 Ω 上的实值随机变量，则称 $Z(\omega) = X(\omega) + iY(\omega)$ 为复随机变量；称 $\overline{Z} = X(\omega) - iY(\omega)$ 为 $Z(\omega)$ 的复共轭随机变量；称 $|Z| := \sqrt{X^2 + Y^2}$ 为复随机变量 Z 的模.

定义 1.16 若随机变量 X, Y 的期望 $E(X), E(Y)$ 都存在，则复随机变量 Z 的数学期望定义为 $E(Z) := E(X) + iE(Y)$.

- $Z_1 = X_1 + iY_1$ 与 $Z_2 = X_2 + iY_2$ 独立当且仅当 (X_1, Y_1) 与 (X_2, Y_2) 独立；
- $E(e^{iX}) = E(\cos X) + iE(\sin X)$;
- $|e^{iX}| = \sqrt{\cos^2 X + \sin^2 X} = 1$;
- 若 X, Y 独立，则 e^{iX} 与 e^{iY} 也独立.

特征函数的定义

定义 1.17 设 X 是一个随机变量，称

$$\begin{aligned}\varphi(t) &:= E(e^{itX}) = E[\cos(tX)] + iE[\sin(tX)] \\ &= \int_{-\infty}^{+\infty} \cos(tx)dF(x) + i\int_{-\infty}^{+\infty} \sin(tx)dF(x) \\ &= \int_{-\infty}^{+\infty} e^{itx}dF(x), \quad -\infty < t < +\infty\end{aligned}$$

为 X 的特征函数。特别的，

- 如果 X 为离散型随机变量，其分布列为 $p_k = P(X = x_k)$，则

$$\varphi(t) = \sum_{k=1}^{+\infty} e^{itx_k}p_k, \quad -\infty < t < +\infty;$$

- 如果 X 为连续型随机变量，其分布密度为 $p(x)$，则

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx}p(x)dx, \quad -\infty < t < +\infty;$$

1. 由 $\varphi(t)=E[\cos(tX)]+iE[\sin(tX)]$ 知，特征函数总是存在的;
2. $\varphi(0)=1, |\varphi(t)|\le 1$: $\varphi(0)=E(1)=1$, 而

$$\begin{aligned}|\varphi(t)| &= |E[\cos(tX)]+iE[\sin(tX)]| = \left[(E[\cos(tX)])^2+(E[\sin(tX)])^2\right]^{1/2} \\ &\le \left[E[\cos^2(tX)]+E[\sin^2(tX)]\right]^{1/2}=1;\end{aligned}$$

3. $\varphi(-t)=\overline{\varphi(t)}$: $\varphi(-t)=E[\cos(tX)]-iE[\sin(tX)]=\overline{\varphi(t)}$;
4. 若 $Y=aX+b$, 其中 a,b 为常数，则 $\varphi_Y(t)=e^{ibt}\varphi_X(at)$, 事实上，

$$\begin{aligned}&\varphi_Y(t)=E[e^{it(aX+b)}]=E[\cos(t(aX+b))]+iE[\sin(t(aX+b))]\\ &=E[\cos(taX)\cos(tb)-\sin(taX)\sin(tb)]+iE[\sin(taX)\cos(tb)+\cos(taX)\sin(tb)]\\ &=\cos(tb)\big(E[\cos(taX)]+iE[\sin(taX)]\big)+i\sin(tb)\big(E[\cos(taX)]+iE[\sin(taX)]\big)\\ &=e^{itb}Ee^{itaX}=e^{itb}\varphi_X(at)\end{aligned}$$

特征函数的性质 II

5. 若 X, Y 独立，则 $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$, 事实上,

$$\begin{aligned}
&\varphi_{X+Y}(t) = E(e^{it(X+Y)}) = E(e^{itX}e^{itY}) \\
&= E[(\cos(tX) + i\sin(tX))(\cos(tY) + i\sin(tY))] \\
&= E[(\cos(tX) + i\sin(tX))\cos(tY)] + iE[(\cos(tX) + i\sin(tX))\sin(tY)] \\
&= E[\cos(tY)]E[e^{itX}] + iE[\sin(tY)]E[e^{itX}] \\
&= E(e^{itX})E(e^{itY})
\end{aligned}$$

6. 更一般的，若 $X_1, X_2, \cdots, X_n$ 相互独立，则对于 $Y = X_1 + \cdots + X_n$ 的特征函数有

$$\varphi_Y(t) = \Pi_{k=1}^{n}\varphi_{X_k}(t)$$

7. 如果 $E(X^k)$ 存在，则

$$\varphi^{(k)}(t) = i^k E[X^k e^{itX}], \quad \varphi^{(k)}(0) = i^k E[X^k]$$

事实上，

$$\frac{d^k}{dt^k}\varphi(t) = \frac{d^k}{dt^k}E[e^{itX}] = E[\frac{d^k}{dt^k}e^{itX}] = E[(iX)^k e^{itX}] = i^k E[X^k e^{itX}]$$

8. $\varphi(t)$ 在 $(-\infty,+\infty)$ 上一致连续：

$$
\begin{aligned}
&|\varphi(t+h)-\varphi(t)| = |E(e^{i(t+h)X})-E(e^{itX})| = |E[e^{itX}(e^{ihX}-1)]| \\
&\le E|e^{itX}(e^{ihX}-1)| = E|e^{ihX}-1| = \int_{-\infty}^{+\infty} |e^{ihx}-1|dF_X(x) \\
&= \int_{|x|>A} |e^{ihx}-1|dF_X(x) + \int_{-A}^{A} |e^{ihx}-1|dF_X(x) \\
&\le 2\int_{|x|>A} dF_X(x) + \int_{-A}^{A} |hx|dF_X(x) \\
&\le 2P(|X|>A) + hAP(|X|\le A) \le 2P(|X|>A) + hA
\end{aligned}
$$

选取 A 使得 $P(|X|>A)\le \epsilon/4$, $h\le\delta=\dfrac{\epsilon}{2A}$, 则只要 $|h|\le\delta$ 必有

$$
|\varphi(t+h)-\varphi(t)| \le 2P(|X|>A) + hA \le \epsilon
$$

9. $\varphi(t)$ 是非负定函数，即对任意的正整数 n 及 n 个实数 $t_1,\cdots,t_n$ 及 n 个复数 $z_1,\cdots,z_n$，均有

$$\sum_{k=1}^{n}\sum_{j=1}^{n}\varphi(t_k-t_j)z_k\bar{z}_j\geq 0$$

证明: 由特征函数的定义可得

$$\begin{aligned}\sum_{k=1}^{n}\sum_{j=1}^{n}\varphi(t_k-t_j)z_k\bar{z}_j &= \sum_{k=1}^{n}\sum_{j=1}^{n}z_k\bar{z}_jE[e^{i(t_k-t_j)X}]=\sum_{k=1}^{n}\sum_{j=1}^{n}E[z_ke^{it_kX}\bar{z}_j\overline{e^{it_jX}}]\\ &= E[\sum_{k=1}^{n}z_ke^{it_kX}\sum_{j=1}^{n}\overline{z_je^{it_jX}}]=E[|\sum_{k=1}^{n}z_ke^{it_kX}|^2]\geq 0\end{aligned}$$

常用分布的特征函数 I

- 单点分布 $P(X=a)=1$: $\varphi(t)=e^{ita}$;
- 两点分布 $P(X=x)=p^x(1-p)^{1-x}, x=0,1$: $\varphi(t)=pe^{it}+(1-p)$;
- 泊松分布:$\varphi(t)=\sum_{k=0}^{+\infty}e^{ikt}\frac{\lambda^k}{k!}e^{-\lambda}=e^{\lambda e^{it}}e^{-\lambda}=e^{\lambda(e^{it}-1)}$;
- 均匀分布 $U(a,b)$:$\varphi(t)=\int_a^b\frac{e^{itx}}{b-a}dx=\frac{e^{ibt}-e^{iat}}{it(b-a)}$;
- 标准正态分布:

$$\begin{aligned}
\varphi(t) &= \int_{-\infty}^{+\infty} e^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \\
&= \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{+\infty}\cos(tx)e^{-\frac{x^2}{2}}dx + i\int_{-\infty}^{+\infty}\sin(tx)e^{-\frac{x^2}{2}}dx\right) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\cos(tx)e^{-\frac{x^2}{2}}dx
\end{aligned}$$

对 $\varphi(t)$ 求导可得

$$\begin{aligned}\varphi'(t) &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x\sin(tx)e^{-\frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\sin(tx)de^{-\frac{x^2}{2}}\\ &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} t\cos(tx)e^{-\frac{x^2}{2}}dx = -t\varphi(t).\end{aligned}$$

从而

$$\frac{\varphi'(t)}{\varphi(t)} = -t, \ \ i.e., \ \ [\ln\varphi(t)]' = -t.$$

对上式两边积分可得

$$\ln\varphi(t) = \ln\varphi(t) - \ln\varphi(0) = -\frac{t^2}{2} + C.$$

因此, $\varphi(t) = e^{-\frac{t^2}{2}+C}$. 再利用 $\varphi(0) = 1$ 可得

$$\varphi(t) = e^{-t^2/2}.$$

常用分布的特征函数 III

- 指数分布:

$$\varphi(t)=\int_0^\infty e^{itx}\lambda e^{-\lambda x}dx=\left(1-\frac{it}{\lambda}\right)^{-1}$$

- 二项分布: $Y\sim B(n,p)$. 注意到 $Y=X_1+\cdots+X_n$, 其中 X_i 为独立同分布的随机变量，且 $X_i\sim B(1,p)$, 故

$$\varphi_Y(t)=\Pi_{k=1}^n(pe^{it}+q)=(pe^{it}+q)^n,\quad q=1-p$$

- 一般正态分布: $Y\sim N(\mu,\sigma^2)$, 则由 $X:=\dfrac{Y-\mu}{\sigma}\sim N(0,1)$ 及 $Y=\sigma X+\mu$ 可得

$$\varphi_Y(t)=e^{i\mu t}\varphi_X(\sigma t)=e^{i\mu t-\frac{\sigma^2t^2}{2}}$$

常用分布的特征函数 IV

- 伽玛分布:$Y \sim \Gamma(n, \lambda)$. 注意到 $Y = X_1 + \cdots + X_n$, 其中 X_i 独立同指数分布即 $X_i \sim \exp(\lambda)$, 故

$$\varphi_Y(t) = \Pi_{k=1}^{n}(1 - \frac{it}{\lambda})^{-1} = (1 - \frac{it}{\lambda})^{-n}$$

更进一步，当 $Y \sim \Gamma(\alpha, \lambda)$, 我们有

$$\varphi_Y(t) = (1 - \frac{it}{\lambda})^{-\alpha}$$

- $\chi^2(n)$ 分布：因为 $\chi^2(n) = \Gamma(n/2, 1/2)$, 故其特征函数为

$$\varphi(t) = (1 - 2it)^{-n/2}$$

例 1.8 利用特征函数的方法求 $\Gamma(\alpha,\lambda)$ 分布的数学期望与方差.

解: Γ 分布的特征函数及其导数分别为

$$\begin{aligned}
\varphi(t) &= (1-\frac{it}{\lambda})^{-\alpha} \\
\varphi'(t) &= \frac{i\alpha}{\lambda}(1-\frac{it}{\lambda})^{-\alpha-1}, \quad \varphi'(0)=\frac{i\alpha}{\lambda} \\
\varphi''(t) &= \frac{i^2\alpha(\alpha+1)}{\lambda^2}(1-\frac{it}{\lambda})^{-\alpha-2}, \quad \varphi''(0)=-\frac{\alpha(\alpha+1)}{\lambda^2}
\end{aligned}$$

故由 $\varphi^{(k)}(0)=i^kE(X^k)$ 知,

$$\begin{aligned}
E(X) &= \frac{\varphi'(0)}{i}=\frac{\alpha}{\lambda}, \quad E(X^2)=\frac{\varphi''(0)}{i^2}=\frac{\alpha(\alpha+1)}{\lambda^2} \\
D(X) &= E(X^2)-(EX)^2=\frac{\alpha(\alpha+1)}{\lambda^2}-\frac{\alpha^2}{\lambda^2}=\frac{\alpha}{\lambda^2}
\end{aligned}$$

定理 1.26 (逆转公式) 设 $F(x)$ 与 $\varphi(t)$ 分别为随机变量 X 的分布函数和特征函数，则对 $F(x)$ 的任意两个连续点 $x_1 < x_2$ 有

$$F(x_2) - F(x_1) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi(t)dt$$

定理 1.27 (唯一性定理) 随机变量的分布函数由其特征函数唯一决定.

证明: 在逆转公式中令 $x_2 = x, x_1 = y \to -\infty$ 可得

$$F(x) = \lim_{y\to-\infty} \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ity} - e^{-itx}}{it} \varphi(t)dt$$

而由分布函数本身是右连续可知，分布函数任一点的取值可由连续点上的值唯一决定.

引理 1.1 (狄利克雷积分) 令

$$I(a,T)=\int_0^T \frac{\sin at}{t}dt, \quad \mathrm{sgn}\{a\}:=\begin{cases}-1, & a<0,\\ 0, & a=0,\\ 1, & a>0,\end{cases}$$

则

$$\lim_{T\to+\infty} I(a,T)=\lim_{T\to+\infty}\int_0^T \frac{\sin at}{t}dt=\frac{\pi}{2}\mathrm{sgn}\{a\}$$

证明: 注意到如果作积分换元 $y=at$ 可得

$$\lim_{T\to+\infty} I(a,T)=\mathrm{sgn}\{a\}\lim_{T\to+\infty} I(1,T)$$

只需证明

$$\lim_{T\to+\infty} I(1,T) = \lim_{T\to+\infty} \int_0^T \frac{\sin t}{t} dt = \frac{\pi}{2}.$$

注意到 $\frac{1}{t} = \int_0^{+\infty} e^{-ut} du$, 故

$$\begin{aligned} I(1,T) &= \int_0^T \int_0^{+\infty} e^{-ut} \sin t du dt = \int_0^{+\infty} \left[\int_0^T e^{-ut} \sin t dt \right] du \\ &= \int_0^{+\infty} \left[\frac{1}{1+u^2} - \frac{1}{1+u^2} e^{-uT}(u \sin T + \cos T) \right] du \\ &= \frac{\pi}{2} - \int_0^{+\infty} \frac{u \sin T + \cos T}{1+u^2} e^{-uT} du \to \frac{\pi}{2} \end{aligned}$$

证明: 考虑积分

$$
\begin{aligned}
J(T) &= \frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}\varphi(t)dt \\
&= \frac{1}{2\pi}\int_{-T}^{T}\left[\int_{-\infty}^{+\infty}\frac{e^{-itx_1}-e^{-itx_2}}{it}e^{itx}dF(x)\right]dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[\int_{-T}^{T}\frac{e^{it(x-x_1)}-e^{it(x-x_2)}}{it}dt\right]dF(x) \\
&= \frac{1}{\pi}\int_{-\infty}^{+\infty}\left[\int_{0}^{T}\frac{\sin t(x-x_1)}{t}dt-\int_{0}^{T}\frac{\sin t(x-x_2)}{t}dt\right]dF(x) \\
&= \frac{1}{\pi}\int_{-\infty}^{+\infty}\left[I(x-x_1,T)-I(x-x_2,T)\right]dF(x) \\
&\to \frac{1}{\pi}\frac{\pi}{2}\int_{-\infty}^{+\infty}\left[\operatorname{sgn}\{x-x_1\}-\operatorname{sgn}\{x-x_2\}\right]dF(x) \qquad T\to+\infty \\
&= \int_{x_1}^{x_2}dF(x)=F(x_2)-F(x_1)
\end{aligned}
$$

定理 1.28 (逆转公式一般情形) 设 $F(x)$ 与 $\varphi(t)$ 分别为随机变量 X 的分布函数和特征函数，则对任意的 $x_1, x_2 \in R$ 有

$$\frac{F(x_2-0)+F(x_2)}{2}-\frac{F(x_1-0)+F(x_1)}{2}=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}\varphi(t)dt$$

连续型分布情形下的逆转公式

定理 1.29 若 X 为连续型随机变量，其密度函数为 $p(x)$，其特征函数为

$$\varphi(t)=\int_{-\infty}^{+\infty}e^{itx}p(x)dx,$$

如果 $\int_{-\infty}^{+\infty}|\varphi(t)|dt<+\infty$，则

$$p(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-itx}\varphi(t)dt$$

证明: 由分布密度的定义可知

$$\begin{aligned}p(x) &= \lim_{h\to 0}\frac{F(x+h)-F(x)}{h}=\lim_{h\to 0}\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{e^{-itx}-e^{-it(x+h)}}{it\cdot h}\varphi(t)dt\\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\lim_{h\to 0}\frac{e^{-ith}-1}{-it\cdot h}e^{-itx}\varphi(t)dt=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-itx}\varphi(t)dt\end{aligned}$$

第二十一讲: 随机变量的收敛性

张鑫

Email: x.zhang.seu@foxmail.com

东南大学　数学学院

- 几乎必然收敛: 如果

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1,$$

则称随机变量列 $\{X_n\}$ 几乎必然收敛到 X, 记作 $X_n \xrightarrow{a.e.} X$, 或 $X_n \to X, a.s.$

- 依概率收敛: 如果对任意的 $\epsilon > 0$ 有

$$\lim_{n\to\infty} P(|X_n(\omega) - X(\omega)| > \epsilon) = 0,$$

则称 X_n 依概率收敛到 X, 记作 $X_n \xrightarrow{P} X$.

- r 阶平均收敛: 设 $r > 0$ 为常数，如果随机变量 X 与 $X_n(n \geq 1)$ 的 r 阶矩皆有限，并且有

$$\lim_{n\to\infty} E|X_n - X|^r = 0,$$

则称 $\{X_n\}$ 为 r 阶平均收敛到 X, 简称 r 阶收敛，记作 $X_n \xrightarrow{L^r} X$.

几乎必然收敛的刻划与性质

- 若记 $E := \{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$, 则

$$E = \cap_{m=1}^{\infty} \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{m}\} \in \mathcal{F};$$

- $X_n \overset{a.s.}{\to} X$ 等价于 $P(E) = 1$ 或 $P(\bar{E}) = 0$;
- $X_n \overset{a.s.}{\to} X$ 的充分必要条件是对任意的 $\epsilon > 0$ 有

$$\lim_{k\to\infty} P(\cup_{n=k}^{\infty} |X_n - X| \geq \epsilon) = 0$$

- 若 $X_n \overset{a.s.}{\to} X$, 则 $X_n \overset{P}{\to} X$.

依概率收敛不能推出几乎必然收敛

例 2.1 取 $\Omega=[0,1), \mathcal{F}$ 为区间 $[0,1)$ 上的 Borel 集类，P 为 Lebesgue 测度。令

$$X_{11}(\omega)=1;\quad X_{21}(\omega)=\begin{cases}1, & \omega\in[0,\frac{1}{2}),\\ 0, & \omega\in[\frac{1}{2},1);\end{cases}\qquad X_{22}(\omega)=\begin{cases}1, & \omega\in[\frac{1}{2},1),\\ 0, & \omega\in[0,\frac{1}{2});\end{cases}$$

一般地，对每个自然数 k, 将 $[0,1)$ 等分为 k 个区间，并取

$$X_{ki}(\omega)=\begin{cases}1, & \omega\in[\frac{i-1}{k},\frac{i}{k}),\\ 0, & \text{其他},\end{cases}\qquad i=1,\cdots,k$$

定义

$$Y_1=X_{11}, Y_2=X_{21}, Y_2=X_{22}, Y_3=X_{31},\cdots$$

对于随机变量序列 Y_n, 对任意的 $\epsilon>0$ 有

$$P(|X_{ki}(\omega)|\geq\epsilon)\leq\frac{1}{k}\to 0,$$

但是收敛的点集 $E=\emptyset$, 即 Y_n 不几乎必然收敛.

引理 2.1 (Markov 不等式) 设随机变量 X 的 r 阶矩有限，则对任意 $\epsilon > 0$ 有

$$P(|X| \geq \epsilon) \leq \frac{1}{\epsilon^r} E|X|^r$$

定理 2.1 若 $X_n \xrightarrow{L^r} X$, 则 $X_n \xrightarrow{P} X$. 反之不真.

例 2.2 仍取区间 $(0,1)$ 上的几何概率空间 $(\Omega, \mathcal{F}, P)$. 固定 $r > 0$, 令

$$X_n(\omega) = \begin{cases} n^{1/r}, & \omega \in (0, 1/n), \\ 0, & \omega \in [1/n, 1), \end{cases} \qquad X(\omega) = 0.$$

易知 $X_n \xrightarrow{a.s.} X$, 从而 $X_n \xrightarrow{P} X$, 但 $X_n \overset{L^r}{\nrightarrow} X$.

定义 2.1 设随机变量 $X, X_1, X_2, \cdots$, 的分布函数分别为 $F(x), F_1(x), F_2(x), \cdots$,, 如果对 $F(x)$ 的任一连续点 x 均有

$$\lim_{n\to} F_n(x) = F(x), \tag{2}$$

则称 $\{F_n(x)\}$ 弱收敛于 $F(x)$, 记作 $F_n(x) \xrightarrow{W} F(x)$. 此时也称 $\{X_n\}$ 按分布收敛于 X, 记作 $X_n \xrightarrow{W} X$.

注 2.1 后面我们用 C_F 表示函数 $F(x)$ 的连续点集，即

$$C_F := \{x : x为F(x)的连续点\}$$

例 2.3 考虑随机变量序列 $\{X_n\}$ 服从如下退化分布

$$P(X_n = \frac{1}{n}) = 1, n = 1, 2, \cdots,$$

它们的分布函数分别为 $F_n(x) = \begin{cases} 0, & x < \frac{1}{n} \\ 1, & x \geq \frac{1}{n} \end{cases}$

- 则在点点收敛要求下，$\{F_n(x)\}$ 的极限函数为

$$g(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

  显然上述极限不满足右连续性，即 $g(x)$ 不是一个分布函数.
- 因为 $F_n(x)$ 在 $x = 1/n$ 处有跳跃，故当 $n \to \infty$ 时，跳跃点的位置趋于 0,很自然的认为 $\{F_n(x)\}$ 应该收敛于点 $x = 0$ 的退化分布，即 $F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$. 但是对任意的 n, 有 $F_n(0) = 0$, 而 $F(0) = 1$, 所以 $\lim_{n\to\infty} F_n(0) = 0 \neq 1 = F(0)$, 而这个不收敛的点恰为 $F(x)$ 的不连续点.

定理 2.2 若 $X_n \xrightarrow{P} X$, 则 $X_n \xrightarrow{W} X$.

证明: 设随机变量 $X, X_1, X_2, \cdots$, 的分布函数分别为 $F(x), F_1(x), F_2(x), \cdots$, 只需证在 $F(x)$ 的连续点上均有 $F_n(x) \to F(x)$ 即可. 而要证明这一点我们只需要说明，对任意的 x 均有

$$F(x-0) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x+0)$$

注意到，对任意的 $x' < x$,

$$\begin{aligned}\{X \le x'\} &= \{X_n \le x, X \le x'\} \cup \{X_n > x, X \le x'\} \\ &\subset \{X_n \le x\} \cup \{|X_n - X| \ge x - x'\}\end{aligned}$$

从而有 $F(x') \le F_n(x) + P(|X_n - X| \ge x - x')$. 再由 $X_n \xrightarrow{P} X$ 知

$$F(x') \le \liminf_{n\to\infty} F_n(x) \Rightarrow F(x-0) \le \liminf_{n\to\infty} F_n(x)$$

类似可证，当 $x'' > x$ 时，$\limsup_{n\to\infty} F_n(x) \le F(x+0)$.

依分布收敛无法推出依概率收敛

例 2.4 设随机变量 X 的分布列为

$$P(X = -1) = P(X = 1) = 1/2$$

令 $X_n = -X$, 则 X_n 与 X 同分布，即 X_n 与 X 有相同的分布，故 $X_n \xrightarrow{W} X$, 但对任意的 $\epsilon \in (0,2)$,

$$P(|X_n - X| \geq \epsilon) = P(2|X| \geq \epsilon) = 1 \nrightarrow 0$$

定理 2.3 若 c 为常数，则 $X_n \xrightarrow{P} c \Leftrightarrow X_n \xrightarrow{W} c$

只需证明充分性即可。记 X_n 的分布函数为 $F_n(x), n = 1, 2, \cdots$ ，常数 c 的分布函数为 $F(x) = \begin{cases} 0, & x < c, \\ 1, & x \geq c. \end{cases}$ 故对任意的 $\epsilon > 0$,

$$
\begin{aligned}
P(|X_n - c| \geq \epsilon) &= P(X_n \geq c + \epsilon) + P(X_n \leq c - \epsilon) \\
&\leq P(X_n > c + \epsilon/2) + P(X_n \leq c - \epsilon) \\
&= 1 - F_n(c + \epsilon/2) + F_n(c - \epsilon) \\
&\to 1 - F(c + \epsilon/2) + F(c - \epsilon) = 0
\end{aligned}
$$

从而 $X_n \xrightarrow{P} c$.

弱收敛的充分条件

定理 2.4 设 $F(x), F_n(x)(n \geq 1)$ 均为分布函数，如果对于 $F(x)$ 的连续点集 C_F 的某个稠密子集 D 上的一切 x 均有

$$\lim_{n\to\infty} F_n(x) = F(x), \forall x \in D \tag{3}$$

则 $F_n \xrightarrow{W} F$.

证明: 由 $D \subset C_F$ 立得必要性. 下设(3)式成立. 对任何 $x \in C_F$, 取

$$y < x < z \text{且} y, z \in D,$$

则有

$$F_n(y) \leq F_n(x) \leq F_n(z). \tag{4}$$

上式令 $n \to \infty$ 并利用(3)式可得

$$F(y) = \lim_n F_n(y) \leq \underline{\lim}_n F_n(x) \leq \overline{\lim}_n F_n(x) \leq \lim_n F_n(z) = F(z).$$

再令 $y \uparrow x$ 及 $z \downarrow x$ 便得证(2)式. 即 $F_n \xrightarrow{W} F$ 。证毕。

定理 2.5 (Helly 第一定理) 任意分布函数列 $F_n(x)$ 均存在子列 $F_{n(k)}(x)$ 以及一个有界、非降、右连续的函数 $F(x)$ 使得对任意的 $x \in C_F$, 均有

$$\lim_{k\to\infty} F_{n(k)}(x) = F(x).$$

注 2.2 极限函数 $F(x)$ 不一定为分布函数. 比如, 若 $a+b+c=1$, $G(x)$ 为分布函数, 令 $F_n(x) = a1_{(x\geq n)} + b1_{(x\geq -n)} + cG(x)$, 则易知

$$F_n(x) \to F(x) = b + cG(x),$$

但

$$\lim_{x\downarrow-\infty} F(x) = b \neq 0, \quad \lim_{x\uparrow\infty} F(x) = b + c = 1 - a \neq 1$$

- 令 $q_1, q_2, \ldots$ 为有理数排列, $m_0(j) \equiv j$. 由于对每个 k,

$$F_m(q_k) \in [0,1], \quad \forall m \geq 1.$$

因此, 存在 $m_{k-1}(j)$ 的子列 $m_k(j)$ 使得

$$\lim_{j\to\infty} F_{m_k(j)}(q_k) = G(q_k).$$

- 令 $F_{n(k)}(x) := F_{m_k(k)}(x)$, 则

$$\lim_{k\to\infty} F_{n(k)}(q) = G(q), \quad \forall q \in \mathbb{Q}.$$

- 令 $F(x) = \inf\{G(q) : q \in \mathbb{Q}, q > x\}$, 则易知 $F(x)$ 右连续. 事实上,

$$\begin{aligned}\lim_{x_n \downarrow x} F(x_n) &= \inf\{G(q) : q \in \mathbb{Q}, q > x_n \text{ 对某个} n\} \\ &= \inf\{G(q) : q \in \mathbb{Q}, q > x\} = F(x)\end{aligned}$$

- 右连续性证明方法 2: 对任何 $\varepsilon > 0$, 总存在 $q_k > x$ 使 $G(q_k) < F(x) + \varepsilon$, 且对于一切 $y \in (x, q_k)$ 有

$$F(x) \le F(y) < G(q_k) < F(x) + \varepsilon.$$

先令 $y \downarrow x$ ， 再令 $\varepsilon \downarrow 0$ 得 $F(x) \le F(x+0) \le F(x)$. 右连续得证.

- 令 $x \in C_F$ 并取 $r_1, r_2, r_3 \in \mathbb{Q}$ 使得 $r_1 < r_2 < x < r_3$, 则

$$F(x) - \epsilon < F(r_1) \le F(r_2) \le F(x) \le F(r_3) < F(x) + \epsilon$$

- 由于

$$F_{n(k)}(r_2) \to G(r_2) \ge F(r_1), \quad F_{n(k)}(r_3) \to G(r_3) \le F(r_3),$$

因此, 当 k 充分大时,

$$F(x) - \epsilon < F_{n(k)}(r_2) \le F_{n(k)}(x) \le F_{n(k)}(r_3) < F(x) + \epsilon$$

- 综上可知

$$\lim_{k\to\infty} F_{n(k)}(x) = F(x), \quad x \in C_F.$$

胎紧 (tight)

定义 2.2 设随机变量 X_n 的分布函数为 $F_n(x)$, 我们称分布函数列 $\{F_n(x)\}_{n\geq 1}$ 为胎紧的 (tight), 如果对任意的 $\epsilon > 0$, 存 M_ϵ 使得

$$\sup_{n\geq 1} P(|X_n| > M_\epsilon) \leq \epsilon.$$

定理 2.6 分布函数列 $\{F_n(x)\}_{n\geq 1}$ 是胎紧的 (tight) 的当且仅当对任意的 $\epsilon > 0$, 存 M_ϵ 使得 $\sup_{n\geq 1}\left(1 - F_n(M_\epsilon) + F_n(-M_\epsilon)\right) \leq \epsilon$.

证明: 充分性, 若 $\sup_{n\geq 1}\left(1 - F_n(M_\epsilon) + F_n(-M_\epsilon)\right) \leq \epsilon$, 则由

$$\begin{aligned}\sup_{n\geq 1} P(|X_n| > M_\epsilon) &= \sup_{n\geq 1}\left(1 - F_n(M_\epsilon) + P(X_n < M_\epsilon)\right)\\ &\leq \sup_{n\geq 1}\left(1 - F_n(M_\epsilon) + F_n(-M_\epsilon)\right) \leq \epsilon\end{aligned}$$

知分布函数列 $\{F_n(x)\}_{n\geq 1}$ 是胎紧的.反之, 取 $K_\epsilon = M_\epsilon + \delta, \delta > 0$, 则

$$\begin{aligned}\sup_{n\geq 1}\left(1 - F_n(K_\epsilon) + F_n(-K_\epsilon)\right) &\leq \sup_{n\geq 1}\left(1 - F_n(M_\epsilon) + P(X_n < -M_\epsilon)\right)\\ &= \sup_{n\geq 1} P(|X_n| > M_\epsilon) \leq \epsilon\end{aligned}$$

定理 2.7 分布函数列 $\{F_n(x)\}_{n\geq 1}$ 的子序列收敛到某个分布函数 $F(x)$ 当且仅当 $\{F_n(x)\}_{n\geq 1}$ 是胎紧的.

证明:

- 假设 $\{F_n(x)\}_{n\geq 1}$ 是胎紧的, 且 $F_{n(k)}(x) \to F(x)$. 下证 $F(x)$ 为分布函数,即证: $F(+\infty) - F(-\infty) = 1$.
- 令 $r, s \in C_F$ 且 $r < -M_\epsilon, s > M_\epsilon$, 由 $F_n(r) \to F(r)$, $F_n(s) \to F(s)$ 得

$$\begin{aligned} 1 - F(s) + F(r) &= \lim_{k\to\infty} \left(1 - F_{n(k)}(s) + F_{n(k)}(r)\right) \\ &\leq \lim_{k\to\infty} \sup_{k\geq 1} \left(1 - F_{n(k)}(M_\epsilon) + F_{n(k)}(-M_\epsilon)\right) \\ &\leq \sup_{n\geq 1} \left(1 - F_n\left(M_\epsilon\right) + F_n\left(-M_\epsilon\right)\right) \leq \epsilon \end{aligned}$$

- 令 $r \to -\infty, s \to \infty$ 后再令 $\epsilon \to 0$ 可得 $F(+\infty) - F(-\infty) = 1$

定理证明续

- 若极限函数是分布函数，下面用反证法证明 $\{F_n(x)\}_{n\geq 1}$ 是胎紧的.
- 若 $\{F_n(x)\}_{n\geq 1}$ 不是胎紧的，则存在 $\epsilon > 0$ 以及子列 $n(k)$ 使得对任意的 k 均有
$$1 - F_{n(k)}(k) + F_{n(k)}(-k) \geq \epsilon,$$
- 由 Helly 第一定理可知，$\{F_{n(k)}(x)\}_{k\geq 1}$ 存在子列 $\{F_{n(k_j)}\}_{j\geq 1} \to F(x)$
- 令 $r, s \in C_F$，$r < 0 < s$，则

$$\begin{aligned} 1 - F(s) + F(r) &= \lim_{j\to\infty} \left(1 - F_{n(k_j)}(s) + F_{n(k_j)}(r)\right) \\ &\geq \liminf_{j\to\infty} \left(1 - F_{n(k_j)}\left(k_j\right) + F_{n(k_j)}\left(-k_j\right)\right) \geq \epsilon \end{aligned}$$

- 令 $s \to \infty, r \to -\infty$ 可知 $F(x)$ 不是分布函数，与假设矛盾.

定理 2.8 (Helly 第二定理) 设分布函数列 $\{F_n(x)\}$ 弱收敛于分布函数 $F(x)$ 当且仅当对任何有界连续函数 $g(x)$ 有

$$\int_{\mathbb{R}} g(x)dF_n(x) \to \int_{\mathbb{R}} g(x)dF(x).$$

或等价的

$$E[g(X_n)] \to E[g(X)].$$

定理 2.9 若分布函数列 $\{F_n(x)\}_{n\geq 1}$ 弱收敛于分布函数 $F(x)$, 则存在分布函数为 $F_n(x)$ 的随机变量序列 $\{Y_n\}_{n\geq 1}$ 及分布函数为 $F(x)$ 的随机变量 Y, 使得 $Y_n \to Y, a.s.$.

证明:

- 令 $\Omega = (0,1), \mathcal{F} = \mathcal{B}((0,1))$, P 为 $(0,1)$ 上的 Lebesgue 测度.
- 令

$$
\begin{aligned}
& Y_n(y) = F_n^{-1}(y) := \sup\{x : F_n(x) < y\} \\
& Y(y) = F^{-1}(y) := \sup\{x : F(x) < y\} \\
& a_y = \sup\{x : F(x) < y\}, \quad b_y = \inf\{x : F(x) > y\} \\
& \Omega_0 = \{y : (a_y, b_y) = \emptyset\}
\end{aligned}
$$

- 由于 (a_y, b_y) 互不相交且每个非空区间均包含不同的有理数, 因此可知 $\Omega - \Omega_0$ 中的元素个数可数, 从而 $P(\Omega - \Omega_0) = 0$.

引理证明 (续)

- 下证, 对任意的 $y \in \Omega_0$, 我们有 $F_n^{-1}(y) \to F^{-1}(y)$.
- 首先, 我们有 $\liminf_{n\to\infty} F_n^{-1}(y) \geq F^{-1}(y)$,
  - 令 $x \in C_F$ 使得 $x < F^{-1}(y)$, 则由 $y \in \Omega_0$ 知 $F(x) < y$
  - 由 $x \in C_F$ 以及 $F_n(x) \to F(x)$ 可知, 当 n 充分大时, $F_n(x) < y$, 也即 $F_n^{-1}(y) \geq x$
  - 令 $n \to \infty$ 取下极限后再令 $x \to F^{-1}(y)$ 可得本结果.
- 其次, 我们可证 $\limsup_{n\to\infty} F_n^{-1}(y) \leq F^{-1}(y)$
  - 令 $x \in C_F$ 使得 $x > F^{-1}(y)$, 则由 $y \in \Omega_0$ 知 $F(x) > y$
  - 由 $x \in C_F$ 以及 $F_n(x) \to F(x)$ 可知, 当 n 充分大时, $F_n(x) > y$, 也即 $F_n^{-1}(y) \leq x$
  - 令 $n \to \infty$ 取上极限后再令 $x \to F^{-1}(y)$ 可得本结果.

- 必要性
  - 由引理可知，存在 $\{Y_n\}_{n\geq 1}$ 及 Y 使得 $Y_n \to Y, a.s.$
  - 由于 $g(x)$ 是连续函数，从而 $g(Y_n) \to g(Y), a.s.$
  - 从而由控制收敛定理可知

$$E[g(X_n)] = E[g(Y_n)] \to E[g(Y)] = E[g(X)],$$

也即

$$\int_{\mathbb{R}} g(x)dF_n(x) \to \int_{\mathbb{R}} g(x)dF(x).$$

- 充分性
  - 令

$$g_{x,\epsilon}(y)=\begin{cases}1 & y\le x\\ 0 & y\ge x+\epsilon\\ \text{linear} & x\le y\le x+\epsilon\end{cases}$$

  - 由 $g_{x,\epsilon}$ 的定义及连续性可知

$$\limsup_{n\to\infty}P\left(X_n\le x\right)\le\limsup_{n\to\infty}Eg_{x,\epsilon}\left(X_n\right)=Eg_{x,\epsilon}\left(X\right)\le P\left(X\le x+\epsilon\right)$$

$$\liminf_{n\to\infty}P\left(X_n\le x\right)\ge\liminf_{n\to\infty}Eg_{x-\epsilon,\epsilon}\left(X_n\right)=Eg_{x-\epsilon,\epsilon}\left(X\right)\ge P\left(X\le x-\epsilon\right)$$

  - 令 $\epsilon\to 0$ 可得

$$\lim\sup_{n\to\infty}P\left(X_n\le x\right)\le P\left(X\le x\right),$$

$$\lim\inf_{n\to\infty}P\left(X_n\le x\right)\ge P\left(X<x\right)\overset{x\in C_F}{=}P\left(X\le x\right)$$

  - 由上可得, 对任意的 $x\in C_F$, 我们有

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

定理 2.10 (连续性定理) 分布函数列 $\{F_n(x)\}$ 弱收敛到分布函数 $F(x)$ 的充分必要条件是: 相应的特征函数列 $\{\varphi_n(t)\}$ 逐点收敛到相应的特征函数 $\varphi(t)$.

证明:

- 必要性显然. 事实上, 由 $\cos(tx), \sin(tx)$ 为有界连续函数以及 Helly 第二定理可知

$$
\begin{aligned}
\varphi_n(t) &:= \int_{\mathbb{R}} e^{itx} dF_n(x) \\
&= \int_{\mathbb{R}} \cos(tx) dF_n(x) + i \int_{\mathbb{R}} \sin(tx) dF_n(x) \\
&\to \int_{\mathbb{R}} \cos(tx) dF(x) + i \int_{\mathbb{R}} \sin(tx) dF(x) \\
&= \int_{\mathbb{R}} e^{itx} dF(x) =: \varphi(t).
\end{aligned}
$$

连续性定理证明 (续)

下证充分性. 先证 $\{F_n(x)\}_{n\geq 1}$ 是胎紧的.

- 注意到

$$u^{-1}\int_{-u}^{u}(1-\varphi_n(t))\,dt = u^{-1}\int_{-u}^{u}\int_{\mathbb{R}}\left(1-e^{itx}\right)dF_n(x)dt$$

$$= u^{-1}\int_{\mathbb{R}}\int_{-u}^{u}\left(1-e^{itx}\right)dtdF_n(x) = 2\int_{\mathbb{R}}(1-\frac{\sin ux}{ux})dF_n(x)$$

$$\geq 2\int_{|x|\geq\frac{2}{u}}(1-\frac{1}{|ux|})dF_n(x) \geq \int_{|x|\geq\frac{2}{u}}dF_n(x) = P(|X_n|\geq\frac{2}{u})$$

- 另一方面,

$$\lim_{n\to\infty}u^{-1}\int_{-u}^{u}(1-\varphi_n(t))\,dt = u^{-1}\int_{-u}^{u}(1-\varphi(t))\,dt \xrightarrow{u\to 0} 0$$

- 从而

$$P(|X_n|\geq\frac{2}{u}) \leq u^{-1}\int_{-u}^{u}(1-\varphi_n(t))\,dt \leq 2\epsilon$$

连续性定理证明 (续)

其次，证明对任意的有界连续函数 $g(x)$ 均有 $E[g(X_n)] \to E[g(X)]$.

- 首先注意到，对于 $\{F_n(x)\}_{n\geq 1}$ 的任一子列 $\{F_{n(k)}(x)\}_{k\geq 1}$，若 $F_{n(k)}(x) \xrightarrow{W} F(x)$，则分布函数 $F(x)$ 的特征函数必为 $\varphi(t)$.
- 由 $\{F_n(x)\}_{n\geq 1}$ 的胎紧性可知，$\{F_n(x)\}_{n\geq 1}$ 的任一子列 $\{F_{n(k)}(x)\}_{k\geq 1}$ 均存在子列 $\{F_{n(k_j)}(x)\}_{j\geq 1}$ 弱收敛到某个分布函数 $F(x)$，且 $F(x)$ 的特征函数为 $\varphi(t)$.
- 由于 $\{F_{n(k_j)}(x)\}_{j\geq 1}$ 弱收敛到某个分布函数 $F(x)$，故由 Helly 第二定理可知，对任意的有界连续函数 $g(x)$ 均有

$$\int g(x)dF_{n(k_j)}(x) \to \int g(x)dF(x)$$

- 故综上可知，$\int g(x)dF_n(x)$ 的任一子列 $\int g(x)dF_{n(k)}(x)$ 均存在子列 $\int g(x)dF_{n(k_j)}(x) \to \int g(x)dF(x)$. 从而，$\int g(x)dF_n(x) \to \int g(x)dF(x)$，即

$$E[g(X_n)] \to E[g(X)].$$

第二十二讲：大数定律

张鑫

Email: x.zhang.seu@foxmail.com

东南大学　　数学学院

- 记 S_n 为 n 重伯努利试验中事件 A 出现次数，称 $\frac{S_n}{n}$ 为事件 A 出现的频率；
- 记 p 为一次试验中 A 发生的概率；
- 概率是频率的稳定值，何谓稳定值？似乎与频率的极限有关；
- 考虑 $|\frac{S_n}{n} - p|$, 如果能类似于函数列极限情形即对任意的 $\epsilon > 0$, 存在 N 使得当 $n \geq N$ 时

$$|\frac{S_n}{n} - p| < \epsilon, \forall \omega \in \Omega.$$

- 但是上述情形无法做到，事实上，对任意的 $\epsilon < 1 - p$,

$$P(|\frac{S_n}{n} - p| < \epsilon) \leq P(\frac{S_n}{n} - p < \epsilon) \leq P(\frac{S_n}{n} - p < 1 - p)$$
$$= P(\frac{S_n}{n} < 1) = 1 - P(\frac{S_n}{n} = 1) = 1 - P(X_1 = 1, \cdots, X_n = 1)$$
$$= 1 - p^n < 1$$

定理 2.11 设 S_n 为 n 重伯努利试验中事件 A 出现次数，p 为一次试验中 A 发生的概率，则对任意的 $\epsilon > 0$，有

$$\lim_{n\to\infty} P(|\frac{S_n}{n} - p| < \epsilon) = 1, \quad \text{或等价的} \quad \lim_{n\to\infty} P(|\frac{S_n}{n} - p| \geq \epsilon) = 0$$

证明：因为 $S_n \sim B(n,p)$，故 $S_n \sim B(n,p), E(\frac{S_n}{n}) = p, D(\frac{S_n}{n}) = \frac{p(1-p)}{n}$.
从而由切比雪夫不等式可得

$$1 \geq P(|\frac{S_n}{n} - p| < \epsilon) \geq 1 - \frac{D(\frac{S_n}{n})}{\epsilon^2} = 1 - \frac{p(1-p)}{n\epsilon^2} \to 1 \quad n \to \infty$$

故

$$\lim_{n\to\infty} P(|\frac{S_n}{n} - p| < \epsilon) = 1$$

伯努利大数定律的含义

- 随着 n 的增大，事件 A 发生的频率 $\frac{S_n}{n}$ 与其概率 p 的偏差大于预先给定的精度 ϵ 的可能性越来越小，要多小有多小，这便是频率稳定于概率的含义;
- 事实上，由切比雪夫不等式可得

$$P(|\frac{S_n}{n} - p| \geq \epsilon) \leq \frac{D(\frac{S_n}{n})}{\epsilon^2} = \frac{p(1-p)}{\epsilon^2}\frac{1}{n}$$

- 伯努利大数定律提供了用频率来确定概率的理论依据;
- 比如要估计某种产品的不合格品率 p, 则可从该种产品中随机抽取 n 件，当 n 很大时，这 n 件产品中的不合格品的比例可作为不合格品率 p 的估计值.

伯努利大数定律的另一形式

- 注意到，在伯努利大数定律中，如果记 $X_i = \begin{cases} 1, & \text{第}i\text{试验}A\text{发生} \\ 0, & \text{第}i\text{试验}A\text{不发生} \end{cases}$，则 $S_n = \sum_{i=1}^{n} X_i$，并且

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i, p = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

- 伯努利大数定律的结论为：对任意的 $\epsilon > 0$，有

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}[X_i - E(X_i)]| < \epsilon) = 1$$

也即

$$\frac{1}{n}\sum_{i=1}^{n}[X_i - E(X_i)] \xrightarrow{P} 0$$

(强) 大数定律的一般形式

定义 2.3 设有一随机变量序列 $\{X_n\}$, 如果对任意的 $\epsilon > 0$, 有

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}[X_i - E(X_i)]| < \epsilon) = 1 \text{ 即 } \frac{1}{n}\sum_{i=1}^{n}[X_i - E(X_i)] \xrightarrow{P} 0$$

则称该随机变量序列 $\{X_n\}$ 满足大数定律。如果进一步有

$$\frac{1}{n}\sum_{i=1}^{n}[X_i - E(X_i)] \xrightarrow{a.s.} 0$$

则称 $\{X_n\}$ 满足强大数定律.

切比雪夫大数定律

定理 2.12 设 $\{X_n\}$ 为一列两两不相关的随机变量序列，若每个 X_i 的方差存在且有共同的上界 M，即 $D(X_i) \le M, \forall i \in \mathbb{N}$，则 $\{X_n\}$ 服从大数定律，即对任意的 $\epsilon > 0$，有

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \epsilon) = 1$$

特别的，若 $\{X_n\}$ 独立同分布，且 $D(X_n) < \infty$，则 $\{X_n\}$ 服从大数定律.

证明： 注意到

$$\begin{aligned} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| \ge \epsilon) &\le \frac{D(\frac{1}{n}\sum_{i=1}^{n} X_i)}{\epsilon^2} \\ &= \frac{1}{\epsilon^2}\frac{1}{n^2}\sum_{i=1}^{n} D(X_i) \\ &\le \frac{1}{\epsilon^2}\frac{M}{n} \to 0 \end{aligned}$$

定理 2.13 设随机变量序列 $\{X_n\}$ 满足如下马尔科夫条件

$$\frac{1}{n^2}D(\sum_{i=1}^{n}X_i) \to 0$$

则 $\{X_n\}$ 服从大数定律，即对任意的 $\epsilon > 0$, 有

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}X_i - \frac{1}{n}\sum_{i=1}^{n}E(X_i)| < \epsilon) = 1$$

定理 2.14 设 $\{X_n\}$ 为一列独立同分布随机变量序列，若每个 X_i 的期望存在，则 $\{X_n\}$ 服从大数定律，即对任意的 $\epsilon > 0$, 有

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \epsilon) = 1.$$

特别的，如果 $E|X_i|^k, k = 1, 2, \cdots,$ 存在，则 $\{X_n^k\}$ 服从大数定律.

证明：设 $\{X_n\}$ 独立同分布，且 $E(X_i) = a, i = 1, \cdots$,, 则只需要证明

$$Y_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} a, \quad n \to \infty$$

由依概率收敛的性质，仅需证 $\varphi_{Y_n}(t) \to e^{iat}$ 即可。事实上，由

$$\begin{aligned}\varphi_{Y_n}(t) &= [\varphi(\frac{t}{n})]^n = [1 + \varphi'(0)\frac{t}{n} + o(\frac{t}{n})]^n = [1 + ia\frac{t}{n} + o(\frac{t}{n})]^n \\ &= \exp\{\frac{\ln(1 + ia\frac{t}{n} + o(\frac{t}{n}))}{1/n}\} \to e^{iat}\end{aligned}$$

定理 2.15 对于事件列 $\{A_j\}$，有

- 若 $\sum_{n=1}^{\infty} P(A_n) < \infty$，则

$$P(\limsup_{n\to\infty} A_n) = \lim_{k\to\infty} P(\cup_{n=k}^{\infty} A_n) = 0$$

- 若 $\{A_n\}$ 相互独立，则 $\sum_{n=1}^{\infty} P(A_n) = \infty$ 当且仅当

$$P(\limsup_{n\to\infty} A_n) = \lim_{k\to\infty} P(\cup_{n=k}^{\infty} A_n) = 1$$

证明： 1. 注意到 $\lim_{k\to\infty} P(\cup_{n=k}^{\infty} A_n) \leq \lim_{k\to\infty} \sum_{n=k}^{\infty} P(A_n) = 0.$

2. 由于

$$\begin{aligned}
P(\cup_{n=k}^{\infty} A_n) &= \lim_{m\to\infty} P(\cup_{n=k}^{m} A_n) = \lim_{m\to\infty} (1 - P(\cap_{n=k}^{m} \overline{A}_n)) \\
P(\cap_{n=k}^{m} \overline{A}_n) &= \Pi_{n=k}^{m} P(\overline{A}_n) = \Pi_{n=k}^{m} (1 - P(A_n)) \\
&\leq \Pi_{n=k}^{m} \exp(-P(A_n)) = \exp(-\sum_{n=k}^{m} P(A_n)) \xrightarrow{m\to\infty} 0
\end{aligned}$$

定理 2.16 (Borel 强大数定律) 设 S_n 为 n 重伯努利试验中事件 A 出现次数，p 为一次试验中 A 发生的概率，则

$$\frac{S_n}{n} \xrightarrow{a.s.} p$$

分析: 记 $A_n = \{|S_n/n - p| \geq \varepsilon\}$, 则定理中的几乎必然收敛的充分必要条件是

$$P\left(\limsup_n A_n\right) = 0.$$

再 Borel-Cantelli 引理可知, 只需证明

$$\sum_{n=1}^{\infty} P\left(A_n\right) < +\infty.$$

若像证明弱大数定律那样用切贝谢夫不等式，我们只能得到

$$P\left(A_n\right) \leq \frac{1}{\varepsilon^2} D\left(\frac{S_n}{n}\right) = \frac{pq}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Borel 强大数定律证明

- 用 $r=4$ 时的马尔科夫不等式可得 $P(A_n) \leq \frac{1}{\varepsilon^4} E\left|\frac{S_n}{n}-p\right|^4$.
- 将 S_n 写为 n 个独立伯努利分布随机变量之和 $S_n=\sum_{k=1}^n X_k$, 从而

$$E\left[\left|\frac{S_n}{n}-p\right|^4\right]=\frac{1}{n^4}\sum_{1\leq i,j,k,l\leq n} E\left[(X_i-p)\left(X_j-p\right)(X_k-p)(X_l-p)\right].$$

- 注意 $X_1,\cdots,X_n$ 相互独立，故上式和号中仅有

$$E\left[(X_i-p)^4\right]=pq\left(p^3+q^3\right)\neq 0,$$

$$E\left[(X_i-p)^2\left(X_j-p\right)^2\right]=p^2q^2\neq 0,(i\neq j).$$

- 注意到前者共 n 项，后者共 $C_n^2C_4^2=3n(n-1)$ 项，故我们有

$$E\left[\left|\frac{S_n}{n}-p\right|^4\right.=\frac{pq}{n^4}\left[n\left(p^3+q^3\right)+3pq\left(n^2-n\right)\right]<\frac{1}{4n^2}.$$

- 代入马等科夫不等式右方可得 $P(A_n)\leq\frac{1}{4n^2\varepsilon^4}$, 定理得证.

注 2.3

- 上述定理证明，本质上只用到独立同分布序列的四阶中心矩 $E\left[(X_k-p)^4\right]$ 以及 $E\left[(X_k-p)^2\left(X_j-p\right)^2\right]$ 有限, 并没有用到伯努利情形的其它特性. 所以，在四阶中心矩有限的前提下，其证明可以照搬到一般独立同分布情形.
- 在如下定理中, 我们在明显减弱的条件下，得到一般独立同分布场合的强大数定律.
  - 证明中采用恰当选取子序列的方法, 仅用切比雪夫不等式就得到了满足 Borel-Cantelli 引理要求的估计式.
  - 而且运用著名的“截尾法”, 连切比雪夫不等式中“方差有限”的要求都可以绕过.

定理 2.17 假定 $\{X_n\}$ 是相互独立同分布的随机变量序列，如果它们有有限的期望 $E(X_n)=a$, 则强大数定律成立，即 $n\to\infty$ 时有

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

分析

- 记部分和 $S_n = \sum_{k=1}^n X_k$. 如非负情形成立, 则有

$$\frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^n X_k^+ - \frac{1}{n}\sum_{k=1}^n X_k^- \xrightarrow{a.s.} E\left(X_1^+\right) - E\left(X_1^-\right) = a.$$

故不妨假定 $\{X_k\}$ 是非负随机变量列.

- 注意到

$$P\left\{\frac{1}{n}\left|S_n - E\left(S_n\right)\right| \geq \varepsilon\right\} \leq \frac{E[\sum_{i=1}^n (X_i - E(X_i))]^k}{\varepsilon^k n^k}, k = 2, 4$$

- 高阶矩不存在, 需要截断
  - 方法 1: $X_n^*(\omega) = \begin{cases} X_n(\omega), & X_n(\omega) \leq M \\ 0, & X_n(\omega) > M. \end{cases} = X_n(\omega) 1_{[0,M]}\left(X_n(\omega)\right).$
  - 方法 2: $X_n^*(\omega) = \begin{cases} X_n(\omega), & X_n(\omega) \leq n, \\ 0, & X_n(\omega) > n. \end{cases} = X_n(\omega) 1_{[0,n]}\left(X_n(\omega)\right).$
  - $S_n^* = \sum_{i=1}^n X_i^*$

分析 (续)

- 对截断方法 1:
  - 若 $P\{X_1 > M\} > 0$, 则

$$\sum_{n=1}^{\infty} P(X_n \neq X_{\mathrm{n}}^*) = \sum_{n=1}^{\infty} P(X_n > M) = \sum_{n=1}^{\infty} P(X_1 > M) = +\infty,$$

  - 无法推出 $X_n - X_n^* \to 0, a.s.$. 从而无法得到 $\dfrac{S_n - S_n^*}{n} \to 0, a.s.$
- 对截断方法 2:
  - $\sum P(X_n \neq X_{\mathrm{n}}^*) = \sum P(X_1 > n) \leq \int_0^{+\infty} P(X_1 > t)dt = E\left(X_1\right) < +\infty,$
  - 故 $X_n - X_n^* \to 0, a.s.$. 更进一步有, $\dfrac{S_n - S_n^*}{n} \to 0, a.s.$
  - 注意到

$$\frac{S_n}{n} = \frac{S_n - S_n^*}{n} + \frac{S_n^* - E(S_n^*)}{n} + \frac{E(S_n^*)}{n}$$

  - $E(X_n^*) = E[X_n(\omega)1_{[0,n]}\left(X_n(\omega)\right)] = E[X_1(\omega)1_{[0,n]}\left(X_1(\omega)\right)] \xrightarrow{n\to\infty} E(X_1)$
  - 从而 $\dfrac{E(S_n^*)}{n} = \dfrac{\sum_{i=1}^{n} E(X_i^*)}{n} \to E(X_1) = a$

- 只需证

$$\frac{S_n^* - E(S_n^*)}{n} \xrightarrow{a.s.} 0$$

- 注意到

$$P\left(\frac{1}{n}\left|S_n^* - E\left(S_n^*\right)\right| \geq \varepsilon\right) \leq \frac{D(S_n^*)}{\varepsilon^2 n^2} = \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n} D(X_i^*)$$

$$\leq \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n}\int_0^\infty x^2 1_{[0,i]}(x)dF(x) \leq \frac{1}{\varepsilon^2 n}\int_0^\infty x^2 1_{[0,n]}(x)dF(x)$$

- 从而

$$\sum_{n=1}^{\infty} P\left(\frac{1}{n}\left|S_n^* - E\left(S_n^*\right)\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\int_0^\infty x^2 \sum_{n=1}^{\infty}\frac{1}{n} 1_{[0,n]}(x)dF(x)$$

- 若 $\sum_{n=1}^{\infty} \frac{1}{n} 1_{[0,n]}(x) < \frac{M}{x}$, 则由 X_i 期望存在可知上式小于 ∞, 问题得证.

分析 (续)

- 遗憾的是，我们无法做到 $\sum_{n=1}^{\infty}\frac{1}{n}1_{[0,n]}(x)<\frac{M}{x}$.
- 但我们如果将 n 换成特定的子列 $u_n=[\theta^n]\geq\frac{\theta^n}{2}$，并令 N 是使 $u_n\geq x$ 的最小自然数，我们则有可能做到

$$\sum_{n=1}^{\infty}\frac{1}{u_n}1_{[0,u_n]}(x)=\sum_{n=N}^{\infty}\frac{1}{u_n}\leq 2\sum_{n=N}^{\infty}\theta^{-n}=M\theta^{-N}\leq\frac{M}{x}$$

- 综上，我们可以证明，对任意的 $\theta>1$ 及相应的 $u_n=[\theta^n]$，我们均有

$$\frac{S_{u_n}^*-E(S_{u_n}^*)}{u_n}\xrightarrow{a.s.}0$$

- 从而可证

$$\frac{S_{u_n}}{u_n}=\frac{S_{u_n}-S_{u_n}^*}{u_n}+\frac{S_{u_n}^*-E(S_{u_n}^*)}{u_n}+\frac{E(S_{u_n}^*)}{u_n}\xrightarrow{a.s.}a$$

- 若 $u_n \le k \le u_{n+1}$, 则由 X_k 的非负性知

$$\frac{u_n}{u_{n+1}}\frac{S_{u_n}}{u_n} \le \frac{S_k}{k} \le \frac{u_{n+1}}{u_n}\frac{S_{u_{n+1}}}{u_{n+1}}$$

- 注意 $u_{n+1}/u_n \to \theta$, 故令 n 及相应的 k 趋于无穷可得, 以概率 1 有

$$\frac{1}{\theta}\cdot a \le \liminf_{k\to\infty}\frac{S_k}{k} \le \limsup_{k\to\infty}\frac{S_k}{k} \le \theta\cdot a.$$

- 上式对任意 $\theta > 1$ 均成立, 特别的取 $\theta = (\ell+1)/\ell$, 则以概率 1 有

$$\frac{\ell}{\ell+1}\cdot a \le \liminf_{k\to\infty}\frac{S_k}{k} \le \limsup_{k\to\infty}\frac{S_k}{k} \le \frac{\ell+1}{\ell}\cdot a.$$

- 对每个 $\ell \ge 1$, 记上式不成立的例外集为 E_ℓ, 且取 $E = \bigcup_{\ell=1}^{\infty} E_\ell$.
- 显然 $P(E) = 0$, 且对任意的 $\omega \notin E$, 上式对一切 $\ell \ge 1$ 成立.
- 令 $\ell \to \infty$ 便得证对任意的 $\omega \in \bar{E}$ 恒有

$$\lim_k \frac{S_k}{k} = \lim_n \frac{1}{n}\sum_{k=1}^{n} X_k = a$$

证明

- 不妨假定 $\{X_k\}$ 是非负随机变量列.
- 截断: $X_n^*(\omega) = \begin{cases} X_n(\omega), & X_n(\omega) \le n, \\ 0, & X_n(\omega) > n. \end{cases} = X_n(\omega)1_{[0,n]}(X_n(\omega))$.
- 令 $S_n^* = \sum_{i=1}^n X_i^*$ 及 $u_n = [\theta^n]$
- 证明: $\dfrac{1}{u_n} S_{u_n} \xrightarrow{a.s.} a$.
  - 只需证明

$$\frac{S_n - S_n^*}{n} \xrightarrow{a.s.} 0, \quad \frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n} \xrightarrow{a.s.} 0, \quad \frac{E(S_n^*)}{n} \to a$$

- 由 θ 的任意性证明

$$\frac{S_n}{n} \xrightarrow{a.s.} a.$$

问题 2.1 设 $0 \le f(x) \le 1$, 计算 $f(x)$ 在区间 $[0,1]$ 上的积分近似值

$$J = \int_0^1 f(x)dx$$

解: 注意到

$$J = \int_0^1 f(x)dx = \int_0^1 \int_0^{f(x)} dydx = P(Y \le f(X)) := p$$

其中 (X, Y) 为方形区域 $[0,1] \times [0,1]$ 上的均匀分布。此时我们有

- X, Y 分别服从 $[0,1]$ 上的均匀分布且 X, Y 独立;
- 故计算积分的值实际上就是计算事件 $A = \{Y \le f(X)\}$ 的概率 p;
- (X, Y) 为方形区域 $[0,1] \times [0,1]$ 均匀分布，因此可以看成在方形区域内的随机投点试验;
- 由伯努利大数定律知，我们可以用重复随机投点试验中事件 A 出现的频率：即随机点 (X, Y) 落入区域 $\{(x,y) : y \le f(x)\}$ 中的频率作为 p 的估计值.

1. 先用计算机产生 $(0,1)$ 均匀分布的 $2n$ 个随机数，组成 n 对随机数 $(x_i, y_i), i = 1, \cdots, n$, 这里的 n 可以很大;
2. 对 n 对数据 $(x_i, y_i), i = 1, \cdots, n$, 记录满足 $y_i \le f(x_i)$ 的次数，即事件 A 发生的频数 S_n;
3. 根据频数 S_n, 计算频率 $\dfrac{S_n}{n}$, 即为积分 J 的近似值.

问题 2.2 设 $c \le g(x) \le d, x \in [a,b]$, 计算 $g(x)$ 在区间 $[a,b]$ 上的积分近似值

$$J_{a,b} = \int_a^b g(x)dx$$

解:

$$\begin{aligned} J_{a,b} &= \int_a^b g(x)dx \xlongequal{x=a+(b-a)y} \int_0^1 g(a+(b-a)y)(b-a)dy \\ &= (b-a)(d-c)\int_0^1 \left(\frac{g(a+(b-a)y)-c}{d-c} + \frac{c}{d-c}\right)dy \\ &= S_0 \int_0^1 f(y)dy + c(b-a) \end{aligned}$$

其中

$$S_0 = (b-a)(d-c), \quad f(y) = \frac{g(a+(b-a)y)-c}{d-c}$$

定理 2.18 (Kolmogorov 极大不等式) 设随机变量 $X_1, \cdots, X_n$ 相互独立，它们都有零期望与有限方差。令 $S_n = \sum_{k=1}^n X_k$，则对任意的 $\epsilon > 0$ 有

$$P(\max_{1\le k\le n} |S_k| \ge \epsilon) \le \frac{1}{\epsilon^2} D(S_n) \tag{5}$$

- 对 $k=1,\cdots,n$, 令

$$A_k=\left\{\omega:|S_k|\geq\varepsilon;\text{ 但 }\left|S_j\right|<\varepsilon, j=1,\cdots,k-1\right\}.$$

- 则 A_k 两两不相容, 且

$$\left\{\max_{I\leq k\leq n}|S_k|\geq\varepsilon\right\}=\bigcup_{k=1}^{n}A_k$$

- 故由 $P\left(A_k\right)=E\left(1_{A_k}\right)$ 及 A_k 的定义可推出

$$P\left\{\max_{1\leq k\leq n}|S_k|\geq\varepsilon\right\}=\sum_{k=1}^{n}P\left(A_k\right)\leq\frac{1}{\varepsilon^2}\sum_{k=1}^{n}E\left(S_k^2 1_{A_k}\right). \qquad (6)$$

- 注意到 $S_k^2 1_{A_k}$ 是 $\xi_1,\cdots,X_k$ 的函数，它与 $S_n-S_k=\sum_{i=k+1}^{n}\xi_i$ 相互独立. 故

$$E\left[\left(S_n-S_k\right)S_k 1_{A_k}\right]=E\left[S_n-S_k\right]E\left[S_k 1_{A_k}\right]=0.$$

- 于是 (6) 式可进一步化为

$$
\begin{aligned}
& P\left\{\max_{1\le k\le n}\left|S_k\right|\ge\varepsilon\right\} \\
\le & \frac{1}{\varepsilon^2}\sum_{k=1}^{n}\left\{E\left(S_k^2 1_{A_k}\right)+2E\left[\left(S_n-S_k\right)S_k 1_{A_k}\right]+E\left[\left(S_n-S_k\right)^2 1_{A_k}\right]\right\} \\
= & \frac{1}{\varepsilon^2}E\left[S_n^2\sum_{k=1}^{n}1_{A_k}\right]\le\frac{1}{\varepsilon^2}E\left(S_n^2\right)=\frac{1}{\varepsilon^2}D\left(S_n\right).
\end{aligned}
$$

引理 2.2 考虑实数列 $\{c_n\}$ 及其部分和数列 $s_n = \sum_{k=1}^{n} c_k$.

1. 若 $c_n \to c$, 则 $s_n/n \to c$;
2. 设 $b_m = \sup\{|s_{m+k} - s_m| : k \geq 1\}, b = \inf\{b_m : m \geq 1\}$, 则 $\sum_{n=1}^{\infty} c_n$ 收敛的充分必要条件是 $b = 0$;
3. (柯罗奈克引理) 若 $\sum_{n=1}^{\infty} c_n/n$ 收敛, 则 $s_n/n \to 0$.

1. - 对任 $\varepsilon > 0$, 选 n_1 使 $n > n_1$ 时 $|c_n - c| < \varepsilon/2$;
   - 再选 $n_2 > n_1$ 使 $n > n_2$ 则 $\sum_{k=1}^{n_1} |c_k - c|/n < \varepsilon/2$.
   - 当 $n > n_2$ 时有

$$\left|\frac{s_n - c}{n}\right| \leq \sum_{k=1}^{n} \frac{|c_n - c|}{n} \tag{7}$$

$$= \sum_{k=1}^{n_1} \frac{|c_k - c|}{n} + \sum_{k=n_1+1}^{n} \frac{|c_k - c|}{n} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{8}$$

2. - 若 $b = 0$, 则对任意 $\varepsilon > 0$, 存在 n_1 使 $b_{n_1} < \varepsilon$. 从而对一切 $k \geq 1$ 有

$$\left|s_{n_1+k} - s_{n_1}\right| < \varepsilon.$$

   - 这表明 $\sum_{n=1}^{\infty} c_n$ 收敛。逆命题证明类似.

引理证明

3.

- 令
$$t_0=0,\quad t_n=\sum_{k=1}^{n}\frac{c_k}{k},n\geq 1$$
- 由 $c_k=k\left(t_k-t_{k-1}\right)$ 可得
$$s_{n+1}=\sum_{k=1}^{n+1}kt_k-\sum_{k=1}^{n+1}kt_{k-1}=-\sum_{k=1}^{n}t_k+(n+1)t_{n+1},\ \ n\geq 1$$
- 故
$$\frac{s_{n+1}}{n+1}=-\frac{n}{n+1}\cdot\frac{1}{n}\sum_{k=1}^{n}t_k+t_{n+1},\ \ n\geq 1 \tag{9}$$
- 现设 $\sum_{n=1}^{\infty}c_n/n$ 收敛，即其部分利序列 $\{t_n\}$ 收敛到有穷的极限 t.
- 由已证得的 1 知序列 $\left\{\frac{1}{n}\sum_{k=1}^{n}t_k\right\}$ 收敛到 t.
- 由此在(9) 式中令 $n\to\infty$ 便得证 $s_n/n\to 0$.

定理 2.19 (Kolmogorov 强大数定律) 设 $\{X_n\}$ 是独立随机变量序列，满足

$$\sum_{k=1}^{\infty} \frac{D(X_k)}{k^2} < +\infty, \tag{10}$$

则强大数定律成立，即当 $n \to \infty$ 时有

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

- 令 $X_k' = X_k - E(X_k)$, 则 $E\left(X_k'\right) = 0, D\left(X_k'\right) = D(X_k)$.
- 条件 (10) 式对独立序列 $\left\{X_k'\right\}$ 仍成立.
- 问题化为证明 $n \to \infty$ 时有

$$\frac{1}{n}\sum_{k=1}^{n} X_k' \xrightarrow{a.s} 0.$$

- 取 $S_n = \sum_{k=1}^{n} X_k'/k$, 并令

$$b_m = \sup\left\{\left|S_{m+k} - S_m\right|, k = 1, 2, \cdots\right\},$$
$$b = \inf\left\{b_m, m = 1, 2, \cdots\right\}.$$

- 对任意 $\varepsilon > 0$ 及自然数 n, m, 由 Kolmogorov 极大不等式(5) 可得

$$P\left\{\max_{1\le k\le n}\left|S_{m+k} - S_m\right| \ge \varepsilon\right\} \le \frac{1}{\varepsilon^2}\sum_{k=m+1}^{m+n}\frac{D\left(X_k\prime\right)}{k^2}. \tag{11}$$

- 注意到 $n \to \infty$ 时,

$$\max_{1 \le k \le n} |S_{m+k} - S_m| \uparrow b_m \ge b, (m \ge 1)$$

- 故对一切 $m \ge 1$ 有

$$\{b > \varepsilon\} \subset \{b_m > \varepsilon\} \subset \bigcup_{n=1}^{\infty} \left\{ \max_{1 \le k \le n} |S_{m+k} - S_m| \ge \varepsilon \right\}.$$

- 于是用概率的下连续性及 (11) 式可得

$$P\{b > \varepsilon\} \le P\{b_m > \varepsilon\} \le \frac{1}{\varepsilon^2} \sum_{k=m+1}^{\infty} \frac{D(X_k\prime)}{k^2}.$$

- 由条件 (10)式以及在上式中令 $m \to \infty$ 可得: $P\{b > \varepsilon\} = 0.$

- 取 $\varepsilon = 1/\ell$, 并记使 $b > 1/\ell$ 成立的零概集为 E_ℓ.
- 令 $E = \bigcup_{\ell=1}^{\infty} E_\ell$, 则 $P(E) = 0$, 且对任意的 $\omega \notin E$ 均有

$$b(\omega) \le 1/\ell, \forall l \ge 1.$$

- 故对任意的 $\omega \notin E$ 恒有 $b(\omega) = 0$. 进而, 在 $\bar{E}$ 上 $\sum_{k=1}^{\infty} X_k'/k$ 收敛.
- 故在 $\overline{E}$ 上

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k' = 0.$$

第二十三讲：中心极限定理

张鑫

Email: x.zhang.seu@foxmail.com

东南大学　　数学学院

中心极限定理：问题的由来

- 实际生活中，我们会经常会进行误差分析，而误差的产生是由大量微小的相互独立的随机因素叠加而成。举个例子来说：机床加工机械轴使其直径符合规定要求，加工后的机械轴的直径会受以下各种随机因素的影响:
  - 机床振动与转速的影响;
  - 刀具装配与磨损的影响;
  - 钢材成份与产地的影响;
  - 量具的误差，测量技术的影响;
  - 车间的温度，湿度，照明，工作电压的影响等等.
- 若记每个因素使每个加工轴的直径产生的误差为 $X_i, i=1,\cdots,n$, 则最终对加工轴直径产生的误差为

$$Y_n = \sum_{i=1} X_i$$

- 显然在实际中，我们必然要关心当 $n\to\infty$ 时 Y_n 的分布是什么？
- 为什么我们不根据 X_i 的分布用卷积公式来计算 Y_n 的分布？

一个例子

例 2.5 设 $\{X_n\}$ 为独立同分布的随机变量序列，其共同的分布为区间 $(0,1)$ 上的均匀分布。记 $Y_n = \sum_{i=1}^{n} X_i$, 试计算 Y_n 的分布.

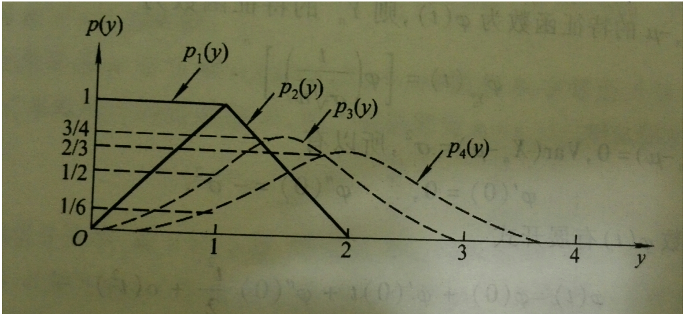

图 4.4.1 均匀分布的卷积

问题：寻求 Y_n 的极限分布

- 从上面均匀分布的例子可知:

$$E(Y_n) = \sum_{i=1}^{n} E(X_i) = \frac{n}{2} \to \infty$$

$$D(Y_n) = \sum_{i=1}^{n} D(X_i) = \frac{n}{12} \to \infty$$

- 考虑 Y_n 的极限分布时，需对 Y_n 进行标准化:

$$Y_n^* := \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}}$$

- 在什么情况下

$$\lim_{n\to\infty} P(Y_n^* \le y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2} du$$

定理 2.20 (林德伯格 - 莱维 (Lindeberg-Lé vy) 中心极限定理) 设 $\{X_n\}$ 是独立同分布的随机变量序列，且 $E(X_i) = \mu, D(X_i) = \sigma^2 > 0$ 存在，若记

$$Y_n^* = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}},$$

则对任意的 $y \in R$ 有

$$\lim_{n\to\infty} P(Y_n^* \leq y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2} du$$

证明: 要证 Y_n^* 依分布收敛于标准正态分布，只需要说明其特征函数收敛于标准正态分布的特征函数即可. 注意到 $Y_n^* = \sum_{i=1}^n \frac{X_i - \mu}{\sigma\sqrt{n}}$. 故若记 $Z_i := X_i - \mu$ 的特征函数为 $\varphi(t)$, 则

$$
\begin{aligned}
\varphi(t) &= \varphi(0) + \varphi'(0)t + \varphi''(0)\frac{t^2}{2} + o(t^2) \\
&= 1 + iE(Z_i)t + i^2E(Z_i^2)\frac{t^2}{2} + o(t^2) \\
&= 1 - \frac{\sigma^2 t^2}{2} + o(t^2) \\
\varphi_{Y_n^*} &= \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n\sigma^2}\right)\right]^n \to e^{-t^2/2}
\end{aligned}
$$

标准正态随机数的产生

- 在随机模拟中经常会需要产生正态随机数，下面我们介绍一种利用中心极限定理通过 $(0,1)$ 上的均匀分布的随机数来产生正态分布 $N(\mu,\sigma^2)$ 随机数的方法;
- 中心极限定理理论分析：设 $X_1,\cdots,X_n,\cdots$ 为一列独立同 $(0,1)$ 上的均匀分布的随机变量序列，则 $E(X_i)=1/2, D(X_i)=1/12$, 故

$$\frac{\sum_{i=1}^{n} X_i - n/2}{\sqrt{n/12}} \to \sim N(0,1)$$

  特别的，若 $n=12$, 则 $X_1+X_2+\cdots+X_{12}-6$ 近似服从标准正态分布.
- 产生正态随机数的方法:
  - 从计算机中产生 12 个 $(0,1)$ 上均匀分布随机数，记为 $x_1,\cdots,x_{12}$;
  - 计算 $y=x_1+\cdots+x_{12}-6$, 由上面分析知可将 y 看成 $N(0,1)$ 的一个随机数;
  - 计算 $z=\mu+\sigma y$, 可将 z 看成 $N(\mu,\sigma^2)$ 分布的随机数;
  - 重复上面步骤 n 次，即可得到 n 个 $N(\mu,\sigma^2)$ 分布随机数.

定理 2.21 (棣莫弗 - 拉普拉斯中心极限定理) 设 n 重伯努利试验中，事件 A 在每次试验中出现的概率为 $p \in (0,1)$，记 S_n 为 n 重伯努利试验中事件 A 出现的次数，则对任意的 $y \in R$ 均有

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le y\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2} du$$

在上述定理中，若记 $\Phi(y) = \beta, Y_n^* := \dfrac{S_n - np}{\sqrt{np(1-p)}}$，则

$$P(Y_n^* \le y) \approx \Phi(y) = \beta$$

独立不同分布下的中心极限定理

- 设 $\{X_n\}$ 是一个相互独立的随机变量序列，其期望与方差有限分别为

$$E(X_i) = \mu_i, \quad D(X_i) = \sigma_i^2, \quad i = 1, 2, \cdots ,$$

- 考虑随机变量的和 $Y_n = X_1 + \cdots + X_n$, 则

$$E(Y_n) = \mu_1 + \cdots + \mu_n, \quad \sigma(Y_n) := \sqrt{D(Y_n)} = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$$

- 考虑 Y_n 的标准化:$Y_n^* := \dfrac{Y_n - E(Y_n)}{\sigma(Y_n)} = \sum_{i=1}^n \dfrac{X_i - \mu_i}{\sigma(Y_n)}$
- 要求 Y_n^* 中的各项 $\dfrac{X_i - \mu_i}{\sigma(Y_n)}$ 均匀的小，即对任意的 $\tau > 0$, 要求事件

$$A_{ni} = \left\{ \left| \frac{X_i - \mu_i}{\sigma(Y_n)} \right| > \tau \right\} = \{|X_i - \mu_i| > \tau\sigma(Y_n)\}$$

发生的可能性小或直接要求其概率趋于 0.
- 为达目的，我们要求 $\lim_{n\to\infty} P(\max_{1\le i\le n} |X_i - \mu_i| > \tau\sigma(Y_n)) = 0$

注意到

$$
\begin{aligned}
& P(\max_{1\le i\le n} |X_i-\mu_i| > \tau\sigma(Y_n)) = P(\cup_{i=1}^{n} |X_i-\mu_i| > \tau\sigma(Y_n)) \\
\le & \sum_{i=1}^{n} P(|X_i-\mu_i| > \tau\sigma(Y_n)) = \sum_{i=1}^{n} \int_{|x-\mu_i|>\tau\sigma(Y_n)} dF_i(x) \\
\le & \frac{1}{\tau^2}\cdot\frac{1}{\sigma(Y_n)^2}\sum_{i=1}^{n}\int_{|x-\mu_i|>\tau\sigma(Y_n)} (x-\mu_i)^2 dF_i(x)
\end{aligned}
$$

故只要下面的 林德伯格条件 满足

$$
\lim_{n\to\infty}\frac{1}{\sigma(Y_n)^2}\sum_{i=1}^{n}\int_{|x-\mu_i|>\tau\sigma(Y_n)} (x-\mu_i)^2 dF_i(x) = 0
$$

就可保证 Y_n^* 中各加项充分的小

定理 2.22 设独立随机变量序列 $\{X_n\}$ 满足林德伯格条件

$$\lim_{n\to\infty}\frac{1}{\sigma(Y_n)^2}\sum_{i=1}^{n}\int_{|x-\mu_i|>\tau\sigma(Y_n)}(x-\mu_i)^2dF_i(x)=0 \qquad (12)$$

则对任意的 x, 有

$$\lim_{n\to\infty}P\left(\frac{1}{\sigma(Y_n)}\sum_{i=1}^{n}(X_i-\mu_i)\le x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}du$$

注 2.4 假设 $\{X_n\}$ 独立同分布且方差有限，则容易验证上述的林德伯格条件满足，从而我们先前在独立同分布情形下的中心极限定理是林德伯格中收极限定理的特例.

引理 2.3 对一切 $x \in R$ 及非负整数 n 有

$$\left| e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} \right| \leq \left[\frac{|x|^{n+1}}{(n+1)!} \right] \wedge \left[\frac{2|x|^n}{n!} \right]$$

引理证明

- 用分部积分可得

$$\int_0^x (x-s)^n e^{is} ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1}\int_0^x (x-s)^{n+1} e^{is} ds \tag{13}$$

- 递归可得知，对一切 $n \geq 0$ 有

$$e^{ix} = \sum_{k=0}^{n} \frac{(ix)^k}{k!} + \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds \tag{14}$$

- 由此推出估计式 $\left|e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}\right| \leq \frac{|x|^{n+1}}{(n+1)!}$
- 将(13)式中 n 换为 $n-1$，解出右方积分并代入(14)便推出

$$e^{ix} = \sum_{k=0}^{n} \frac{(ix)^k}{k!} + \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}\left(e^{is}-1\right) ds$$

- 故可得另一估计式 $\left|e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}\right| \leq \frac{2|x|^n}{n!}$
- 联合上述两个估计式可得引理结论.

注 2.5 若引理中的 $n=0,1,2$, 则有

$$\left|e^{ix}-1\right| \leq |x| \wedge 2$$

$$\left|e^{ix}-1-ix\right| \leq \frac{x^2}{2} \wedge (2|x|)$$

$$\left|e^{ix}-1-ix+\frac{x^2}{2}\right| \leq \left(\frac{1}{6}|x|^3\right) \wedge x^2$$

林德伯格中心极限定理证明

- 只需证 $Y_n^* := \frac{\sum_{k=1}^n (X_k - E[X_k])}{\sqrt{D(\sum_{k=1}^n X_k)}}$ 的特征函数列收敛到标准正态分布的特征函数，即

$$\lim_{n\to\infty} \phi_n(t) = \lim_{n\to\infty} E[e^{itY_n^*}] = \exp\{-\frac{t^2}{2}\}. \tag{15}$$

- 对一切 $n \geq 1$ 及 $k = 1, \cdots, n$，令 $X_{nk} = (X_k - \mu_k) / \sigma(Y_n)$，则

$$E(X_{nk}) = 0, \quad D(X_{nk}) = \sigma_k^2 / \sigma(Y_n)^2.$$

- 以 $F_{nk}(x)$ 及 $\phi_{nk}(t)$ 分别表 X_{nk} 的分布函数及特征函数.
- 已知的林德伯格条件(12)式化为

$$\lim_n \sum_{k=1}^n \int_{(|x|\geq\tau)} x^2 dF_{nk}(x) = 0, \ \forall \tau > 0 \tag{16}$$

- 而待证的(15)式变为 $n \to \infty$ 时有

$$\phi_n(t) = \prod_{k=1}^n \phi_{nk}(t) \longrightarrow \exp\{-\frac{t^2}{2}\}, \forall t \in R$$

林德伯格中心极限定理证明

- 注意到

$$\phi_{nk}(t) \approx 1 + \phi'_{nk}(0)t + \frac{1}{2}\phi''_{nk}(0)t^2 = 1 + iE[X_{nk}]t + \frac{1}{2}i^2E[X_{nk}^2]t^2$$

$$= 1 - \frac{t^2\sigma_k^2}{2\sigma(Y_n)^2} \approx \exp\{-\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}\}$$

$$\exp\{-\frac{t^2}{2}\} = \exp\{-\frac{t^2}{2}\sum_{k=1}^{n}\frac{\sigma_k^2}{\sigma(Y_n)^2}\} = \prod_{k=1}^{n}\exp\{-\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}\}$$

- 因此, 欲证 $\phi_n(t) = \prod_{k=1}^{n}\phi_{nk}(t) \longrightarrow e^{-\frac{t^2}{2}}, \forall t \in R$, 只需证明

$$|\prod_{k=1}^{n}\phi_{nk}(t) - \prod_{k=1}^{n}(1 - \frac{t^2\sigma_k^2}{2\sigma(Y_n)^2})| \longrightarrow 0$$

$$|\prod_{k=1}^{n}(1 - \frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}) - \prod_{k=1}^{n}\exp\{-\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}\}| \longrightarrow 0$$

证明: $\left|\prod_{k=1}^{n}\phi_{nk}(t)-\prod_{k=1}^{n}\left(1-\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}\right)\right|\to 0$

- 由引理2.3可知，对一切 $t\in R$ 有

$$\left|e^{itx}-\left(1+itx-\frac{1}{2}t^2x^2\right)\right|\le|tx|^2\wedge|tx|^3$$

- 注意到 X_{nk} 二阶矩有限，两边对 F_{nk} 取积分得

$$\left|\phi_{nk}(t)-\left(1-\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}\right)\right|\le\int_{-\infty}^{\infty}|tx|^2\wedge|tx|^3dF_{nk}(x)$$

- 而对任何 $\tau>0$ ，上式右方不超过

$$\begin{aligned}&\int_{(|x|<\tau)}|tx|^3dF_{nk}(x)+\int_{(|x|\ge\tau)}(tx)^2dF_{nk}(x)\\&\le\tau|t|^3\frac{\sigma_k^2}{\sigma(Y_n)^2}+t^2\int_{(|x|\ge\tau)}x^2dF_{nk}(x)\end{aligned}$$

- 由于 $\sum_{k=1}^{n}\sigma_k^2/\sigma(Y_n)^2=D\left(Y_n^*\right)=1$, 求和便推出

$$\sum_{k=1}^{n}\left|\phi_{nk}(t)-\left(1-\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}\right)\right|\le\tau|t|^3+t^2\sum_{k=1}^{n}\int_{(|x|\ge\tau)}x^2dF_{nk}(x)$$

- 上式先令 $n \to \infty$ 再令 $\tau \downarrow 0$, 由(16)式得

$$\sum_{k=1}^n |\phi_{nk}(t) - (1 - \frac{t^2\sigma_k^2}{2\sigma(Y_n)^2})| \to 0, t \in R$$

- 故

$$|\prod_{k=1}^n \phi_{nk}(t) - \prod_{k=1}^n (1 - \frac{t^2\sigma_k^2}{2\sigma(Y_n)^2})| \le \sum_{k=1}^n |\phi_{nk}(t) - (1 - \frac{t^2\sigma_k^2}{2\sigma(Y_n)^2})| \to 0$$

得证.

- 注意到当 $|z|\le 1/2$ 时有

$$|e^z-1-z|\le \frac{1}{2}\sum_{k=2}^{\infty}|z|^k=\frac{1}{2}|z|^2/(1-|z|)\le z^2$$

- 取 $z=\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}$，可得

$$\begin{aligned}
&\left|\prod_{k=1}^n\left(1-\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}\right)-\prod_{k=1}^n \exp\{-\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}\}\right|\\
&\le \sum_{k=1}^n\left|\exp\{-\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}\}-\left(1-\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}\right)\right|\\
&\le \frac{t^4}{4}\sum_{k=1}^n\left(\frac{\sigma_k^2}{\sigma(Y_n)^2}\right)^2\le \frac{t^4}{4}\left(\max_{1\le k\le n}\frac{\sigma_k^2}{\sigma(Y_n)^2}\right)\sum_{k=1}^n\frac{\sigma_k^2}{\sigma(Y_n)^2}\\
&\le \frac{t^4}{4}\left(\max_{1\le k\le n}\frac{\sigma_k^2}{\sigma(Y_n)^2}\right)
\end{aligned}$$

- 若 $\max_{1\le k\le n}\frac{\sigma_k^2}{\sigma(Y_n)^2}\to 0$，则定理得证.

证明：$\max_{1\le k\le n}\frac{\sigma_k^2}{\sigma(Y_n)^2}\to 0$

- 对任何 $\tau>0$ 有

$$\frac{\sigma_k^2}{\sigma(Y_n)^2}=D\left(X_{nk}\right)=\int_{-\infty}^{+\infty}x^2dF_{nk}(x)=\int_{|x|<\tau}x^2dF_{nk}(x)+\int_{|x|\ge\tau}x^2dF_{nk}(x)$$
$$\le\tau^2+\int_{(|x|\ge\tau)}x^2dF_{nk}(x)\le\tau^2+\sum_{k=1}^{n}\int_{(|x|\ge\tau)}x^2dF_{nk}(x)$$

- 从而

$$\max_{1\le k\le n}\frac{\sigma_k^2}{\sigma(Y_n)^2}\le\tau^2+\sum_{k=1}^{n}\int_{(|x|\ge\tau)}x^2dF_{nk}(x)$$

- 再次令 $n\to\infty$ 及 $\tau\downarrow 0$，由(16)式得

$$\lim_{n}\left\{\max_{1\le k\le n}\frac{\sigma_k^2}{\sigma(Y_n)^2}\right\}=0$$

- 进而，若 $z=\frac{t^2\sigma_k^2}{2\sigma(Y_n)^2}$，当 n 充分大时 $|z|\le 1/2$.

定理 2.23 设 $\{X_n\}$ 为独立随机变量序列，如果存在常数列 $\{L_n\}$ 使得

$$\max_{1\le k\le n} |X_k| \le L_n, \text{ 且 } \lim_n \frac{L_n}{\sigma(Y_n)} = 0$$

则 Lindeberg 条件成立，从而 $\{X_n\}$ 满足中心极限定理.

- 由假设知，对于 $\tau > 0$ ，存在 N ，使当 $n \ge N$ 时有 $2L_n < \tau\sigma(Y_n)$.
- 因此当 $n \ge N$ 时对一切 $k = 1, \cdots, n$ 有 $\Omega = \{|X_k - \mu_k| < \tau\sigma(Y_n)\}$.
- 故 $n \ge N$ 时有

$$\begin{aligned}
&\frac{1}{\sigma(Y_n)^2}\sum_{k=1}^n \int_{|x-\mu_k|\ge\tau\sigma(Y_n)} (x-\mu_k)^2\, dF_k(x) \\
&= \frac{1}{\sigma(Y_n)^2}\sum_{k=1}^n \int_{-\infty}^{\infty} (x-\mu_k)^2\, 1_{\{|x-\mu_k|\ge\tau\sigma(Y_n)\}}(x) dF_k(x) \\
&= \frac{1}{\sigma(Y_n)^2}\sum_{k=1}^n E[(X_k-\mu_k)^2 1_{\{|x-\mu_k|\ge\tau\sigma(Y_n)\}}(X_k)] = 0.
\end{aligned}$$

- 这表明林德伯格条件(12) 式成立，故 $\{X_n\}$ 满足中心极限定理.

定理 2.24 设 $\{X_n\}$ 为独立随机变量序列，若存在 $\delta > 0$ 满足

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

则 $\lim_{n\to\infty} P(\frac{1}{\sigma(Y_n)} \sum_{k=1}^{n} (X_k - \mu_k) \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du.$

证明： 只要验证林德伯格条件成立. 实际上对任何 $\tau > 0$ 有

$$\lim_{n} \frac{1}{\sigma(Y_n)^2} \sum_{k=1}^{n} \int_{(|x-\mu_k| \geq \tau\sigma(Y_n))} (x-\mu_k)^2 \, dF_k(x)$$

$$\leq \lim_{n} \frac{1}{\sigma(Y_n)^2 \, (\tau\sigma(Y_n))^{\delta}} \sum_{k=1}^{n} \int_{(|x-\mu_k| \geq \tau\sigma(Y_n))} |x-\mu_k|^{2+\delta} \, dF_k(x)$$

$$\leq \lim_{n} \frac{1}{\tau^{\delta}} \frac{1}{\sigma(Y_n)^{2+\delta}} \sum_{k=1}^{n} \int_{\Re} |x-\mu_k|^{2+\delta} \, dF_k(x) = 0.$$

定理证毕.