概率论与数理统计（5）

清华大学

2020 年春季学期

二维随机变量函数的数学期望

- 若二维随机变量 (X, Y) 的联合分布列为 $P(X = x_i, Y = y_j)$ 或者联合密度函数为 $p(x, y)$, 则 $Z = g(X, Y)$ 的数学期望为

$$E(Z) = E(g(X, Y)) = \begin{cases} \sum_i \sum_j g(x_i, y_j) P(X = x_i, Y = y_j), \\ \iint g(x, y) p(x, y) \, dx dy. \end{cases}$$

- 若二维随机变量 (X, Y) 的联合分布列为 $P(X = x_i, Y = y_j)$ 或者联合密度函数为 $p(x, y)$，则 $Z = g(X, Y)$ 的数学期望为

$$E(Z) = E(g(X, Y)) = \begin{cases} \sum_i \sum_j g(x_i, y_j) P(X = x_i, Y = y_j), \\ \iint g(x, y) p(x, y)\, dx dy. \end{cases}$$

- 或者，我们已经找到 $Z = g(X, Y)$ 的分布列或密度函数，则

$$E(Z) = \begin{cases} \sum_k z_k P(Z = z_k) \\ \int_{-\infty}^{\infty} z p(z)\, dz. \end{cases}$$

- 若二维随机变量 (X, Y) 的联合分布列为 $P(X = x_i, Y = y_j)$ 或者联合密度函数为 $p(x, y)$, 则 $Z = g(X, Y)$ 的数学期望为

$$E(Z) = E(g(X, Y)) = \begin{cases} \sum_i \sum_j g(x_i, y_j) P(X = x_i, Y = y_j), \\ \iint g(x, y) p(x, y)\, dxdy. \end{cases}$$

- 或者，我们已经找到 $Z = g(X, Y)$ 的分布列或密度函数，则

$$E(Z) = \begin{cases} \sum_k z_k P(Z = z_k) \\ \int_{-\infty}^{\infty} z p(z)\, dz. \end{cases}$$

- 若 $g(X, Y) = X$, 得到的是 X 边际分布的期望:
$E(X) = \int_{-\infty}^{\infty} x p_X(x)\, dx.$

- 若二维随机变量 (X, Y) 的联合分布列为 $P(X=x_i, Y=y_j)$ 或者联合密度函数为 $p(x, y)$, 则 $Z=g(X, Y)$ 的数学期望为

$$E(Z)=E(g(X, Y))=\begin{cases}\sum_i\sum_j g(x_i, y_j)P(X=x_i, Y=y_j),\\ \iint g(x, y)p(x, y)\,dxdy.\end{cases}$$

- 或者，我们已经找到 $Z=g(X, Y)$ 的分布列或密度函数，则

$$E(Z)=\begin{cases}\sum_k z_kP(Z=z_k)\\ \int_{-\infty}^{\infty} zp(z)\,dz.\end{cases}$$

- 若 $g(X, Y)=X$, 得到的是 X 边际分布的期望：
$E(X)=\int_{-\infty}^{\infty} xp_X(x)\,dx.$
- 若 $g(X, Y)=(X-E(X))^2$, 得到的是 X 边际分布的方差。

例子

- 在长度为 a 的线段上任取两点的平均长度:

例子

- 在长度为 a 的线段上任取两点的平均长度：X, Y 服从 $(0, a)$ 上的均匀分布，且相互独立，

例子

- 在长度为 a 的线段上任取两点的平均长度：X, Y 服从 $(0, a)$ 上的均匀分布，且相互独立，

$$
\begin{aligned}
E(|X-Y|) &= \int_0^a \int_0^a |x-y| \frac{1}{a^2} dxdy \\
&= \frac{1}{a^2}\Big[\int_0^a \int_0^x (x-y) dydx + \int_0^a \int_x^a (y-x) dydx\Big] \\
&= \frac{1}{a^2}\int_0^a (x^2 - ax + \frac{a^2}{2}) dx = \frac{a}{3}.
\end{aligned}
$$

例子

- 在长度为 a 的线段上任取两点的平均长度：X, Y 服从 $(0,a)$ 上的均匀分布，且相互独立，

$$\begin{aligned}
E(|X-Y|) &= \int_0^a \int_0^a |x-y| \frac{1}{a^2}\, dxdy \\
&= \frac{1}{a^2}\Big[\int_0^a \int_0^x (x-y)\, dydx + \int_0^a \int_x^a (y-x)\, dydx\Big] \\
&= \frac{1}{a^2}\int_0^a (x^2 - ax + \frac{a^2}{2})\, dx = \frac{a}{3}.
\end{aligned}$$

- $X_i \sim Exp(\lambda)$, $i=1,2$, 且相互独立， $Y=\max\{X_1, X_2\}$.

例子

- 在长度为 a 的线段上任取两点的平均长度：X, Y 服从 $(0,a)$ 上的均匀分布，且相互独立，

$$
\begin{aligned}
E(|X-Y|) &= \int_0^a \int_0^a |x-y| \frac{1}{a^2} dxdy \\
&= \frac{1}{a^2}\Big[\int_0^a \int_0^x (x-y)\,dydx + \int_0^a \int_x^a (y-x)\,dydx\Big] \\
&= \frac{1}{a^2}\int_0^a (x^2 - ax + \frac{a^2}{2})\,dx = \frac{a}{3}.
\end{aligned}
$$

- $X_i \sim Exp(\lambda)$, $i=1,2$, 且相互独立，$Y=\max\{X_1, X_2\}$.
- $p_Y(y) = 2(1-e^{-\lambda y})\lambda e^{-\lambda y}$, $y>0$.

例子

- 在长度为 a 的线段上任取两点的平均长度：X, Y 服从 $(0,a)$ 上的均匀分布，且相互独立，

$$E(|X-Y|)=\int_0^a\int_0^a |x-y|\frac{1}{a^2}dxdy$$
$$=\frac{1}{a^2}\Big[\int_0^a\int_0^x (x-y)dydx+\int_0^a\int_x^a (y-x)dydx\Big]$$
$$=\frac{1}{a^2}\int_0^a (x^2-ax+\frac{a^2}{2})dx=\frac{a}{3}.$$

- $X_i\sim Exp(\lambda)$, $i=1,2$, 且相互独立，$Y=\max\{X_1,X_2\}$.
- $p_Y(y)=2(1-e^{-\lambda y})\lambda e^{-\lambda y}$, $y>0$.

$$E[\max\{X_1,X_2\}]=\int_0^\infty y2\lambda(1-e^{-\lambda y})e^{-\lambda y}dy$$
$$=\frac{2}{\lambda}\int_0^\infty \lambda ye^{-\lambda y}d(\lambda y)-\frac{1}{2\lambda}\int_0^\infty 2\lambda ye^{-2\lambda y}d(2\lambda y)=(\frac{2}{\lambda}-\frac{1}{2\lambda})\Gamma(2)$$

数学期望和方差的性质

- $E(X+Y)=E(X)+E(Y)$:

- $E(X+Y)=E(X)+E(Y)$:

$$\begin{aligned} E(X+Y) &= \iint_{\mathbb{R}^2} (x+y)p(x,y)\,dxdy \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} p(x,y)\,dydx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} p(x,y)\,dxdy \\ &= \int_{-\infty}^{\infty} xp_X(x)\,dx + \int_{-\infty}^{\infty} yp_Y(y)\,dy = E(X)+E(Y). \end{aligned}$$

- $E(X+Y)=E(X)+E(Y)$:

$$
\begin{aligned}
E(X+Y) &= \iint_{\mathbb{R}^2}(x+y)p(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} p(x,y)\,dydx+\int_{-\infty}^{\infty} y\int_{-\infty}^{\infty} p(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty} xp_X(x)\,dx+\int_{-\infty}^{\infty} yp_Y(y)\,dy=E(X)+E(Y).
\end{aligned}
$$

- 若 X 与 Y 相互独立，则 $E(XY)=E(X)E(Y)$.

- $E(X+Y)=E(X)+E(Y)$:

$$
\begin{aligned}
E(X+Y) &= \iint_{\mathbb{R}^2} (x+y)p(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} p(x,y)\,dydx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} p(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty} xp_X(x)\,dx + \int_{-\infty}^{\infty} yp_Y(y)\,dy = E(X)+E(Y).
\end{aligned}
$$

- 若 X 与 Y 相互独立，则 $E(XY)=E(X)E(Y)$.

$$
\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyp_X(x)p_Y(y)\,dxdy \\
&= \int_{-\infty}^{\infty} xp_X(x)\,dx \int_{-\infty}^{\infty} yp_Y(y)\,dy = E(X)E(Y).
\end{aligned}
$$

数学期望和方差的性质

- 若 X 与 Y 相互独立，则 $Var(X+Y) = Var(X) + Var(Y)$.

数学期望和方差的性质

- 若 X 与 Y 相互独立，则 $Var(X+Y) = Var(X) + Var(Y)$.

$$
\begin{aligned}
& Var(X+Y) \\
&= E[(X+Y-E(X+Y))^2] = E[(X-E(X)+Y-E(Y))^2] \\
&= E[(X-E(X))^2] + E[(Y-E(Y))^2] + 2E[(X-E(X))(Y-E(Y)) \\
&= Var(X) + Var(Y).
\end{aligned}
$$

- 若 X 与 Y 相互独立，则 $Var(X+Y) = Var(X) + Var(Y)$.

$$
\begin{aligned}
& Var(X+Y) \\
& = E[(X+Y-E(X+Y))^2] = E[(X-E(X)+Y-E(Y))^2] \\
& = E[(X-E(X))^2] + E[(Y-E(Y))^2] + 2E[(X-E(X))(Y-E(Y)) \\
& = Var(X) + Var(Y).
\end{aligned}
$$

- 前面的三个等式对任意有限个随机变量都成立：
  - $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$,
  - 若 $X_i, i = 1, \ldots, n$ 相互独立，则

$$E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n),$$

$$Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n).$$

例子: 又是抛硬币

- 连续抛硬币一百万次，连续 6 个正面接着连续 6 个反面出现的次数的数学期望值是?

例子: 又是抛硬币

- 连续抛硬币一百万次，连续 6 个正面接着连续 6 个反面出现的次数的数学期望值是?
- 考虑随机变量 I_j, $I_j = 1$ 如果从第 j 次开始（包括 j）连续抛出了 6 个正面紧接着连续 6 个反面，否则为 0.

例子: 又是抛硬币

- 连续抛硬币一百万次，连续 6 个正面接着连续 6 个反面出现的次数的数学期望值是?
- 考虑随机变量 I_j, $I_j = 1$ 如果从第 j 次开始（包括 j）连续抛出了 6 个正面紧接着连续 6 个反面，否则为 0.
- 那么出现的次数为 $\sum_{j=1}^{1000000-11} I_j$,

例子: 又是抛硬币

- 连续抛硬币一百万次，连续 6 个正面接着连续 6 个反面出现的次数的数学期望值是?
- 考虑随机变量 I_j, $I_j = 1$ 如果从第 j 次开始（包括 j）连续抛出了 6 个正面紧接着连续 6 个反面，否则为 0.
- 那么出现的次数为 $\sum_{j=1}^{1000000-11} I_j$,
- $E(I_j) = \frac{1}{2^{12}}$.?
- $E(\sum_{j=1}^{1000000-11} I_j) = \sum_{j=1}^{1000000-11} E(I_j) = \frac{1000000-11}{2^{12}} = 244.14.$

回到双色球问题

- 假设有 N 个人各买了一注，考虑随机变量 X_i, $X_i = 1$ 若第 i 个人中了大奖，否则 $X_i = 0$.

回到双色球问题

- 假设有 N 个人各买了一注，考虑随机变量 X_i, $X_i = 1$ 若第 i 个人中了大奖，否则 $X_i = 0$.
- 中大奖的人数为 $Y_N = \sum_{i=1}^{N} X_i$, $E(Y_N) = Np$, $Var(Y_N) = Np(1-p)$.

回到双色球问题

- 假设有 N 个人各买了一注，考虑随机变量 X_i, $X_i = 1$ 若第 i 个人中了大奖，否则 $X_i = 0$.
- 中大奖的人数为 $Y_N = \sum_{i=1}^{N} X_i$, $E(Y_N) = Np$, $Var(Y_N) = Np(1-p)$.
- 一个中大奖的人都没有的概率是

$$P(Y_N = 0) \leqslant P(|Y - E(Y)| \geqslant Np) \leqslant \frac{Var(Y)}{N^2p^2} = \frac{N(1-p)p}{N^2p^2}.$$

- $p = \frac{1}{17721088}$, $N = 300000000$, $P(Y_N = 0) \leqslant 0.059$.

回到双色球问题

- 假设有 N 个人各买了一注，考虑随机变量 X_i, $X_i=1$ 若第 i 个人中了大奖，否则 $X_i=0$.
- 中大奖的人数为 $Y_N=\sum_{i=1}^N X_i$, $E(Y_N)=Np$, $Var(Y_N)=Np(1-p)$.
- 一个中大奖的人都没有的概率是

$$P(Y_N=0)\leqslant P(|Y-E(Y)|\geqslant Np)\leqslant \frac{Var(Y)}{N^2p^2}=\frac{N(1-p)p}{N^2p^2}.$$

- $p=\frac{1}{17721088}$, $N=300000000$, $P(Y_N=0)\leqslant 0.059$.更精确一点的计算?

回到双色球问题

- 假设有 N 个人各买了一注，考虑随机变量 X_i, $X_i=1$ 若第 i 个人中了大奖，否则 $X_i=0$.
- 中大奖的人数为 $Y_N=\sum_{i=1}^N X_i$, $E(Y_N)=Np$, $Var(Y_N)=Np(1-p)$.
- 一个中大奖的人都没有的概率是

$$P(Y_N=0)\leqslant P(|Y-E(Y)|\geqslant Np)\leqslant \frac{Var(Y)}{N^2p^2}=\frac{N(1-p)p}{N^2p^2}.$$

- $p=\frac{1}{17721088}$, $N=300000000$, $P(Y_N=0)\leqslant 0.059$.更精确一点的计算?
- $P(Y_N=0)=(1-p)^{30000000}=4.44463*10^{-8}$.

大数定律：抢先版

- 考虑 n 个独立同分布随机变量，$X_1, \ldots, X_n$，（如 n 次独立重复试验），其方差 σ^2 和期望 μ 均存在（有限）。

大数定律：抢先版

- 考虑 n 个独立同分布随机变量, $X_1, \ldots, X_n$, （如 n 次独立重复试验），其方差 σ^2 和期望 μ 均存在（有限）.
- $E(\frac{1}{n}(X_1 + \cdots + X_n)) = \mu$.
- $Var(\frac{1}{n}(X_1 + \cdots + X_n)) = \frac{\sigma^2}{n}$.

大数定律：抢先版

- 考虑 n 个独立同分布随机变量, $X_1, \ldots, X_n$, （如 n 次独立重复试验）, 其方差 σ^2 和期望 μ 均存在（有限）.
- $E(\frac{1}{n}(X_1 + \cdots + X_n)) = \mu$.
- $Var(\frac{1}{n}(X_1 + \cdots + X_n)) = \frac{\sigma^2}{n}$.
- 对于 $\forall \epsilon > 0$,

$$P(|\frac{1}{n}\sum_{i=1}^{n} X_n - \mu| \geqslant \epsilon) \leqslant \frac{\sigma^2}{n\epsilon^2} \to 0, \quad n \to \infty.$$

大数定律：抢先版

- 考虑 n 个独立同分布随机变量, $X_1, \ldots, X_n$, （如 n 次独立重复试验），其方差 σ^2 和期望 μ 均存在（有限）.
- $E(\frac{1}{n}(X_1 + \cdots + X_n)) = \mu$.
- $Var(\frac{1}{n}(X_1 + \cdots + X_n)) = \frac{\sigma^2}{n}$.
- 对于 $\forall \epsilon > 0$,

$$P(|\frac{1}{n}\sum_{i=1}^{n} X_n - \mu| \geqslant \epsilon) \leqslant \frac{\sigma^2}{n\epsilon^2} \to 0, \quad n \to \infty.$$

- $\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu, \quad n \to \infty.$

协方差

- (X, Y) 为二维随机变量，则称

$$Cov(X, Y) = E\Big[(X - E(X))(Y - E(Y))\Big],$$

为 X 与 Y 的协方差。

协方差

- (X, Y) 为二维随机变量，则称

$$Cov(X, Y) = E\Big[(X - E(X))(Y - E(Y))\Big],$$

为 X 与 Y 的协方差。

- $Cov(X, Y) = Cov(Y, X)$.

协方差

- (X, Y) 为二维随机变量，则称

$$Cov(X, Y) = E\Big[(X - E(X))(Y - E(Y))\Big],$$

  为 X 与 Y 的协方差。
- $Cov(X, Y) = Cov(Y, X)$.
- $Cov(X, Y) > 0$, 称 X 与 Y 正相关。

协方差

- (X, Y) 为二维随机变量，则称

$$Cov(X, Y) = E\Big[(X - E(X))(Y - E(Y))\Big],$$

为 X 与 Y 的协方差。

- $Cov(X, Y) = Cov(Y, X)$.
- $Cov(X, Y) > 0$, 称 X 与 Y 正相关。
- $Cov(X, Y) < 0$, 称 X 与 Y 负相关。

协方差

- (X, Y) 为二维随机变量，则称

$$Cov(X, Y) = E\Big[(X - E(X))(Y - E(Y))\Big],$$

  为 X 与 Y 的协方差。
- $Cov(X, Y) = Cov(Y, X)$.
- $Cov(X, Y) > 0$，称 X 与 Y 正相关。
- $Cov(X, Y) < 0$，称 X 与 Y 负相关。
- $Cov(X, Y) = 0$，称 X 与 Y 不相关。

协方差的性质

- $Cov(X, Y) = E(XY) - E(X)E(Y)$.

协方差的性质

- $Cov(X,Y) = E(XY) - E(X)E(Y)$.

$$\begin{aligned} Cov(X,Y) &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

协方差的性质

- $Cov(X, Y) = E(XY) - E(X)E(Y)$.

$$\begin{aligned} Cov(X, Y) &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

若 $Cov(X, Y) = 0$, $E(XY) = E(X)E(Y)$.

协方差的性质

- $Cov(X,Y)=E(XY)-E(X)E(Y)$.

$$\begin{aligned} Cov(X,Y) &= E[XY-XE(Y)-YE(X)+E(X)E(Y)] \\ &= E(XY)-E(X)E(Y). \end{aligned}$$

若 $Cov(X,Y)=0$, $E(XY)=E(X)E(Y)$.
- 若 X 与 Y 相互独立，则 $Cov(X,Y)=0$, 反之不然。

协方差的性质

- $Cov(X,Y) = E(XY) - E(X)E(Y)$.

$$\begin{aligned} Cov(X,Y) &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

  若 $Cov(X,Y) = 0$, $E(XY) = E(X)E(Y)$.
- 若 X 与 Y 相互独立，则 $Cov(X,Y) = 0$, 反之不然。
  - $X \sim N(0,1)$, $Y = X^2$, $Cov(X,Y) = E(X^3) - E(X)E(Y) = 0$.

协方差的性质

- $Cov(X,Y)=E(XY)-E(X)E(Y)$.

$$\begin{aligned} Cov(X,Y) &= E[XY-XE(Y)-YE(X)+E(X)E(Y)] \\ &= E(XY)-E(X)E(Y). \end{aligned}$$

  若 $Cov(X,Y)=0$, $E(XY)=E(X)E(Y)$.
- 若 X 与 Y 相互独立，则 $Cov(X,Y)=0$, 反之不然。
  - $X\sim N(0,1)$, $Y=X^2$, $Cov(X,Y)=E(X^3)-E(X)E(Y)=0$.
  - $X\sim U(0,2\pi)$, $Y=\cos X$, $Z=\sin X$, $Cov(Y,Z)=0$.

协方差的性质

- $Cov(X,Y)=E(XY)-E(X)E(Y)$.

$$\begin{aligned} Cov(X,Y) &= E[XY-XE(Y)-YE(X)+E(X)E(Y)] \\ &= E(XY)-E(X)E(Y). \end{aligned}$$

  若 $Cov(X,Y)=0$, $E(XY)=E(X)E(Y)$.
- 若 X 与 Y 相互独立，则 $Cov(X,Y)=0$, 反之不然。
  - $X\sim N(0,1)$, $Y=X^2$, $Cov(X,Y)=E(X^3)-E(X)E(Y)=0$.
  - $X\sim U(0,2\pi)$, $Y=\cos X$, $Z=\sin X$, $Cov(Y,Z)=0$.
- $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$

协方差的性质

- $Cov(X,Y)=E(XY)-E(X)E(Y)$.

$$\begin{aligned} Cov(X,Y) &= E[XY-XE(Y)-YE(X)+E(X)E(Y)] \\ &= E(XY)-E(X)E(Y). \end{aligned}$$

  若 $Cov(X,Y)=0$, $E(XY)=E(X)E(Y)$.
- 若 X 与 Y 相互独立，则 $Cov(X,Y)=0$, 反之不然。
  - $X\sim N(0,1)$, $Y=X^2$, $Cov(X,Y)=E(X^3)-E(X)E(Y)=0$.
  - $X\sim U(0,2\pi)$, $Y=\cos X$, $Z=\sin X$, $Cov(Y,Z)=0$.
- $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$

$$\begin{aligned} & Var(X+Y) \\ &= E(X+Y-E(X+Y))^2 \\ &= E(X-E(X))^2+E(Y-E(Y))^2+2E(X-E(X))(Y-E(Y)). \end{aligned}$$

协方差的性质

- $Cov(X, Y) = E(XY) - E(X)E(Y)$.

$$\begin{aligned} Cov(X, Y) &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

若 $Cov(X, Y) = 0$, $E(XY) = E(X)E(Y)$.

- 若 X 与 Y 相互独立，则 $Cov(X, Y) = 0$, 反之不然。
  - $X \sim N(0,1)$, $Y = X^2$, $Cov(X, Y) = E(X^3) - E(X)E(Y) = 0$.
  - $X \sim U(0, 2\pi)$, $Y = \cos X$, $Z = \sin X$, $Cov(Y, Z) = 0$.
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

$$\begin{aligned} &Var(X + Y) \\ &= E(X + Y - E(X + Y))^2 \\ &= E(X - E(X))^2 + E(Y - E(Y))^2 + 2E(X - E(X))(Y - E(Y)). \end{aligned}$$

若 $Cov(X, Y) = 0$, $Var(X + Y) = Var(X) + Var(Y)$.

协方差的性质

- $\forall a \in \mathbb{R},\ Cov(X, a) = 0.$

协方差的性质

- $\forall a \in \mathbb{R},\ Cov(X, a) = 0.$
- $\forall a, b \in \mathbb{R},\ Cov(aX, bY) = abCov(X, Y).$

协方差的性质

- $\forall a \in \mathbb{R},\ Cov(X, a) = 0.$
- $\forall a, b \in \mathbb{R},\ Cov(aX, bY) = abCov(X, Y).$
- 分配律：$Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$。

协方差的性质

- $\forall a \in \mathbb{R},\ Cov(X, a) = 0.$
- $\forall a, b \in \mathbb{R},\ Cov(aX, bY) = abCov(X, Y).$
- 分配律：$Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$。

$$
\begin{aligned}
Cov(X + Y, Z) &= E((X + Y)Z) - E(X + Y)E(Z) \\
&= E(XZ) + E(YZ) - E(X)E(Z) - E(Y)E(Z) \\
&= Cov(X, Z) + Cov(Y, Z).
\end{aligned}
$$

- 施瓦茨（Schwarz）不等式：$[Cov(X, Y)]^2 \leqslant \sigma_X^2\sigma_Y^2.$

协方差的性质

- $\forall a \in \mathbb{R},\ Cov(X, a) = 0.$
- $\forall a, b \in \mathbb{R},\ Cov(aX, bY) = abCov(X, Y).$
- 分配律：$Cov(X+Y, Z) = Cov(X, Z) + Cov(Y, Z)$。

$$\begin{aligned} Cov(X+Y, Z) &= E((X+Y)Z) - E(X+Y)E(Z) \\ &= E(XZ) + E(YZ) - E(X)E(Z) - E(Y)E(Z) \\ &= Cov(X, Z) + Cov(Y, Z). \end{aligned}$$

- 施瓦茨（Schwarz）不等式：$[Cov(X, Y)]^2 \leqslant \sigma_X^2\sigma_Y^2$.考虑

$$g(t) = E[t(X-E(X))+(Y-E(Y))]^2 = t^2\sigma_X^2+2tCov(X, Y)+\sigma_Y^2 \geqslant 0,$$

协方差的性质

- $\forall a \in \mathbb{R},\ Cov(X, a) = 0.$
- $\forall a, b \in \mathbb{R},\ Cov(aX, bY) = abCov(X, Y).$
- 分配律：$Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$。

$$\begin{aligned} Cov(X+Y, Z) &= E((X+Y)Z) - E(X+Y)E(Z) \\ &= E(XZ) + E(YZ) - E(X)E(Z) - E(Y)E(Z) \\ &= Cov(X, Z) + Cov(Y, Z). \end{aligned}$$

- 施瓦茨（Schwarz）不等式：$[Cov(X, Y)]^2 \leqslant \sigma_X^2 \sigma_Y^2$.考虑

$$g(t) = E[t(X - E(X)) + (Y - E(Y))]^2 = t^2\sigma_X^2 + 2tCov(X, Y) + \sigma_Y^2 \geqslant 0,$$

则

$$(2Cov(X, Y))^2 - 4\sigma_X^2\sigma_Y^2 \leqslant 0.$$

计算协方差

- 设二维随机变量 (X, Y) 在矩形
$$G=\{(x,y):0\leqslant x\leqslant 2,\ 0\leqslant y\leqslant 1\},$$
上服从均匀分布，记
$$U=\begin{cases}1, & X>Y,\\ 0, & X\leqslant Y,\end{cases}\qquad V=\begin{cases}1, & X>2Y,\\ 0, & X\leqslant 2Y.\end{cases}$$
求 U 和 V 的协方差。

计算协方差

- 设二维随机变量 (X, Y) 在矩形
$$G = \{(x, y) : 0 \leqslant x \leqslant 2,\ 0 \leqslant y \leqslant 1\},$$
上服从均匀分布，记
$$U = \begin{cases} 1, & X > Y, \\ 0, & X \leqslant Y, \end{cases} \qquad V = \begin{cases} 1, & X > 2Y, \\ 0, & X \leqslant 2Y. \end{cases}$$
求 U 和 V 的协方差。
- $p(x, y) = \begin{cases} \frac{1}{2}, & (x, y) \in G \\ 0, & \text{其他}. \end{cases}$

计算协方差

- 设二维随机变量 (X, Y) 在矩形
$$G = \{(x, y) : 0 \leqslant x \leqslant 2,\ 0 \leqslant y \leqslant 1\},$$
上服从均匀分布，记
$$U = \begin{cases} 1, & X > Y, \\ 0, & X \leqslant Y, \end{cases} \qquad V = \begin{cases} 1, & X > 2Y, \\ 0, & X \leqslant 2Y. \end{cases}$$
求 U 和 V 的协方差。
- $p(x, y) = \begin{cases} \frac{1}{2}, & (x, y) \in G \\ 0, & \text{其他}. \end{cases}$
- $P(U = 1) = P(X > 1) + \int_0^1 \int_0^x 0.5dydx = 0.75,$
$P(U = 0) = 0.25,\ P(V = 1) = \int_0^2 \int_0^{x/2} 0.5dydx = 0.5,$
$P(V = 0) = 0.5.$

计算协方差

- 设二维随机变量 (X, Y) 在矩形
$$G = \{(x, y) : 0 \leqslant x \leqslant 2,\ 0 \leqslant y \leqslant 1\},$$
上服从均匀分布，记
$$U = \begin{cases} 1, & X > Y, \\ 0, & X \leqslant Y, \end{cases} \quad V = \begin{cases} 1, & X > 2Y, \\ 0, & X \leqslant 2Y. \end{cases}$$
求 U 和 V 的协方差。
- $p(x, y) = \begin{cases} \frac{1}{2}, & (x, y) \in G \\ 0, & \text{其他}. \end{cases}$
- $P(U = 1) = P(X > 1) + \int_0^1 \int_0^x 0.5 dydx = 0.75,$
$P(U = 0) = 0.25,\ P(V = 1) = \int_0^2 \int_0^{x/2} 0.5 dydx = 0.5,$
$P(V = 0) = 0.5.$
- $P(UV = 1) = \int_0^2 \int_0^{x/2} 0.5 dydx = 0.5.$

计算协方差

- 设二维随机变量 (X, Y) 在矩形
$$G = \{(x, y) : 0 \leqslant x \leqslant 2,\ 0 \leqslant y \leqslant 1\},$$
上服从均匀分布，记
$$U = \begin{cases} 1, & X > Y, \\ 0, & X \leqslant Y, \end{cases} \qquad V = \begin{cases} 1, & X > 2Y, \\ 0, & X \leqslant 2Y. \end{cases}$$
求 U 和 V 的协方差。
- $p(x, y) = \begin{cases} \frac{1}{2}, & (x, y) \in G \\ 0, & \text{其他}. \end{cases}$
- $P(U = 1) = P(X > 1) + \int_0^1 \int_0^x 0.5dydx = 0.75,$
$P(U = 0) = 0.25,\ P(V = 1) = \int_0^2 \int_0^{x/2} 0.5dydx = 0.5,$
$P(V = 0) = 0.5.$
- $P(UV = 1) = \int_0^2 \int_0^{x/2} 0.5dydx = 0.5.$
- $Cov(U, V) = E(UV) - E(U)E(V) = 0.5 - 0.75 * 0.5 = 0.125.$

相关系数

- 设 (X,Y) 是一个二维随机变量，且 $Var(X)=\sigma_X^2>0$, $Var(Y)=\sigma_Y^2>0$, 则称

$$Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}=\frac{Cov(X,Y)}{\sigma_X,\sigma_Y},$$

为 X 与 Y 的相关系数。

相关系数

- 设 (X, Y) 是一个二维随机变量，且 $Var(X) = \sigma_X^2 > 0$, $Var(Y) = \sigma_Y^2 > 0$, 则称

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X, Y)}{\sigma_X, \sigma_Y},$$

为 X 与 Y 的相关系数。

- 相关系数是相应标准化随机变量的协方差：

相关系数

- 设 (X, Y) 是一个二维随机变量，且 $Var(X) = \sigma_X^2 > 0$, $Var(Y) = \sigma_Y^2 > 0$, 则称

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)\, Var(Y)}} = \frac{Cov(X, Y)}{\sigma_X, \sigma_Y},$$

为 X 与 Y 的相关系数。

- 相关系数是相应标准化随机变量的协方差：

$$X' = \frac{X - \mu_X}{\sigma_X}, \quad Y' = \frac{Y - \mu_Y}{\sigma_Y},$$

相关系数

- 设 (X, Y) 是一个二维随机变量，且 $Var(X) = \sigma_X^2 > 0$，$Var(Y) = \sigma_Y^2 > 0$，则称

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{Cov(X, Y)}{\sigma_X, \sigma_Y},$$

  为 X 与 Y 的相关系数。

- 相关系数是相应标准化随机变量的协方差：

$$X' = \frac{X - \mu_X}{\sigma_X}, \qquad Y' = \frac{Y - \mu_Y}{\sigma_Y},$$

$$Cov(X', Y') = Cov\left(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}\right) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

相关系数的性质

- 有界性：$|Corr(X, Y)| \leqslant 1$.

相关系数的性质

- 有界性：$|Corr(X,Y)| \leqslant 1$.
- $Corr(X,Y) = \pm 1$ 当且仅当 X 与 Y 几乎处处线性相关，即存在 $a \neq 0, b$, 使得
$$P(X = aY + b) = 1.$$

相关系数的性质

- 有界性：$|Corr(X,Y)| \leqslant 1$.
- $Corr(X,Y) = \pm 1$ 当且仅当 X 与 Y 几乎处处线性相关，即存在 $a \neq 0, b$, 使得
$$P(X = aY + b) = 1.$$
  - 充分性：若 $Y = aX + b$, 则 $Corr(X,Y) = \frac{a\sigma_X^2}{|a|\sigma_X^2} = \begin{cases} 1, & a > 0 \\ -1, & a < 0. \end{cases}$.

相关系数的性质

- 有界性：$|Corr(X,Y)| \leqslant 1$.
- $Corr(X,Y) = \pm 1$ 当且仅当 X 与 Y 几乎处处线性相关，即存在 $a \neq 0, b$, 使得
$$P(X = aY + b) = 1.$$
  - 充分性：若 $Y = aX + b$, 则 $Corr(X,Y) = \frac{a\sigma_X^2}{|a|\sigma_X^2} = \begin{cases} 1, & a > 0 \\ -1, & a < 0. \end{cases}$.
  - 必要性：$Var(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}) = 2[1 \pm Corr(X,Y)]$.

相关系数的性质

- 有界性：$|Corr(X,Y)| \leqslant 1$.
- $Corr(X,Y) = \pm 1$ 当且仅当 X 与 Y 几乎处处线性相关，即存在 $a \neq 0, b$, 使得
$$P(X = aY + b) = 1.$$

  - 充分性：若 $Y = aX + b$, 则 $Corr(X,Y) = \frac{a\sigma_X^2}{|a|\sigma_X^2} = \begin{cases} 1, & a > 0 \\ -1, & a < 0. \end{cases}$.
  - 必要性：$Var(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}) = 2[1 \pm Corr(X,Y)]$.
  若 $Corr(X,Y) = 1$, 则
$$Var(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}) = 0 \Rightarrow P(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = c) = 1.$$

相关系数的性质

- 有界性：$|Corr(X,Y)| \leqslant 1$.
- $Corr(X,Y) = \pm 1$ 当且仅当 X 与 Y 几乎处处线性相关，即存在 $a \neq 0, b$, 使得

$$P(X = aY + b) = 1.$$

  - 充分性：若 $Y = aX + b$, 则 $Corr(X,Y) = \frac{a\sigma_X^2}{|a|\sigma_X^2} = \begin{cases} 1, & a > 0 \\ -1, & a < 0. \end{cases}$.
  - 必要性：$Var(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}) = 2[1 \pm Corr(X,Y)]$.
    若 $Corr(X,Y) = 1$, 则

$$Var(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}) = 0 \Rightarrow P(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = c) = 1.$$

    若 $Corr(X,Y) = -1$, 则

$$Var(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}) = 0 \Rightarrow P(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} = c) = 1.$$

协方差或者相关系数的一个应用: 投资组合

- 有金融产品 X 和 Y，考虑投资组合 $X + Y$.
- 回报期望为 $E(X) + E(Y)$.
- 组合的方差为 $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$.
- 在回报期望一样的时候，希望风险越小越好。
- 应该选择 $Cov(X, Y) < 0$ 的产品。

二维正态分布的相关系数

- $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

二维正态分布的相关系数

- $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.
-

$$
\begin{aligned}
Cov(X, Y) &= E[(X - E(X))(Y - E(Y))] \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_1)(y-\mu_2) \\
&\times \quad \exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} \\
&\quad + \frac{(y-\mu_2)^2}{\sigma_2^2}]\} dxdy
\end{aligned}
$$

- $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.
-

$$
\begin{aligned}
Cov(X, Y) &= E[(X - E(X))(Y - E(Y))] \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_1)(y-\mu_2) \\
&\times \quad \exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} \\
&\quad + \frac{(y-\mu_2)^2}{\sigma_2^2}]\}dxdy
\end{aligned}
$$

将中括号内化为

$$
(\frac{x-\mu}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2})^2 + (1-\rho^2)(\frac{y-\mu_2}{\sigma_2})^2.
$$

二维正态分布的相关系数

- 作变量代换

$$\begin{cases} u = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}\right) \\ v = \frac{y-\mu_2}{\sigma_2}, \end{cases} \quad \text{即} \quad \begin{cases} x-\mu_1 = \sigma_1(u\sqrt{1-\rho^2}+\rho v), \\ y-\mu_2 = \sigma_2 v, \end{cases}$$

二维正态分布的相关系数

- 作变量代换

$$\begin{cases} u = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}\right) \\ v = \frac{y-\mu_2}{\sigma_2}, \end{cases} \quad \text{即} \quad \begin{cases} x-\mu_1 = \sigma_1(u\sqrt{1-\rho^2} + \rho v), \\ y-\mu_2 = \sigma_2 v, \end{cases}$$

- $|J| = \sigma_1\sigma_2\sqrt{1-\rho^2}$.

- 作变量代换

$$\begin{cases} u = \frac{1}{\sqrt{1-\rho^2}}(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}) \\ v = \frac{y-\mu_2}{\sigma_2}, \end{cases} \quad 即 \quad \begin{cases} x-\mu_1 = \sigma_1(u\sqrt{1-\rho^2}+\rho v), \\ y-\mu_2 = \sigma_2 v, \end{cases}$$

- $|J| = \sigma_1\sigma_2\sqrt{1-\rho^2}$.
-

$$\begin{aligned} & Cov(X, Y) \\ & = \frac{\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(uv\sqrt{1-\rho^2}+\rho v^2)\exp\{-\frac{1}{2}(u^2+v^2)\}dudv \\ & = \rho\sigma_1\sigma_2. \end{aligned}$$

二维正态分布的相关系数

- 作变量代换

$$\begin{cases} u = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}\right) \\ v = \frac{y-\mu_2}{\sigma_2}, \end{cases} \quad 即 \quad \begin{cases} x-\mu_1 = \sigma_1(u\sqrt{1-\rho^2}+\rho v), \\ y-\mu_2 = \sigma_2 v, \end{cases}$$

- $|J| = \sigma_1\sigma_2\sqrt{1-\rho^2}$.
-

$$\begin{aligned} & Cov(X,Y) \\ &= \frac{\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(uv\sqrt{1-\rho^2}+\rho v^2)\exp\{-\frac{1}{2}(u^2+v^2)\}dudv \\ &= \rho\sigma_1\sigma_2. \end{aligned}$$

- $Corr(X,Y) = \rho$.

- 作变量代换

$$\begin{cases} u = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}\right) \\ v = \frac{y-\mu_2}{\sigma_2}, \end{cases} \quad \text{即} \quad \begin{cases} x-\mu_1 = \sigma_1(u\sqrt{1-\rho^2}+\rho v), \\ y-\mu_2 = \sigma_2 v, \end{cases}$$

- $|J| = \sigma_1\sigma_2\sqrt{1-\rho^2}$.
-

$$\begin{aligned} & Cov(X, Y) \\ & = \frac{\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(uv\sqrt{1-\rho^2}+\rho v^2)\exp\{-\frac{1}{2}(u^2+v^2)\}dudv \\ & = \rho\sigma_1\sigma_2. \end{aligned}$$

- $Corr(X, Y) = \rho$.
- 二维正态分布 (X, Y)：X, Y 不相关与相互独立等价。(非常特殊)

- 若 n 维随机变量 $X=(X_1,\ldots,X_n)$ 的每个分量的数学期望都存在，则 n 维随机变量的数学期望为：

$$E(X)=(E(X_1),\ldots,E(X_n)),$$

- 其协方差矩阵为：

$$\begin{aligned}&E[(X-E(X))(X-E(X))']\\ &=\begin{pmatrix} Var(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n)\\ Cov(X_2,X_1) & Var(X_2) & \cdots & Cov(X_2,X_n)\\ \vdots & \vdots & & \vdots\\ Cov(X_n,X_1) & Cov(X_n,X_2) & \cdots & Var(X_n)\end{pmatrix}.\end{aligned}$$

随机向量的数学期望和协方差矩阵

- 若 n 维随机变量 $X=(X_1,\ldots,X_n)$ 的每个分量的数学期望都存在，则 n 维随机变量的数学期望为：

$$E(X)=(E(X_1),\ldots,E(X_n)),$$

- 其协方差矩阵为：

$$E[(X-E(X))(X-E(X))']$$
$$=\begin{pmatrix} Var(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & Var(X_2) & \cdots & Cov(X_2,X_n) \\ \vdots & \vdots & & \vdots \\ Cov(X_n,X_1) & Cov(X_n,X_2) & \cdots & Var(X_n) \end{pmatrix}.$$

- 协方差矩阵为非负定对称矩阵。

n 维正态分布

- n 维正态分布的密度函数有以下形式：

$$p(x_1,\ldots,x_n)=p(x)=\frac{1}{(2\pi)^{n/2}|B|^{1/2}}\exp\{-\frac{1}{2}(x-a)'B^{-1}(x-a)\},$$

其中 B 是其协方差矩阵，a 为其数学期望（向量）。

- n 维正态分布的密度函数有以下形式:

$$p(x_1,\ldots,x_n)=p(x)=\frac{1}{(2\pi)^{n/2}|B|^{1/2}}\exp\{-\frac{1}{2}(x-a)'B^{-1}(x-a)\},$$

其中 B 是其协方差矩阵，a 为其数学期望（向量）。

- 二维时

$$a=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\quad B=\begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{pmatrix}.$$

离散随机变量的条件分布

- 二维离散随机变量 (X, Y) 的联合分布列为

$$p_{ij} = P(X = x_i, Y = y_j), \quad i, j = 1, 2, \ldots.$$

- 对一切使得 $P(Y = y_j) = p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij} > 0$ 的 y_j，称

$$p_{i|j} = P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \ldots,$$

为给定 $Y = y_j$ 条件下 X 的条件分布列。

离散随机变量的条件分布

- 二维离散随机变量 (X, Y) 的联合分布列为

$$p_{ij} = P(X = x_i, Y = y_j), \quad i, j = 1, 2, \ldots.$$

- 对一切使得 $P(Y = y_j) = p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij} > 0$ 的 y_j，称

$$p_{i|j} = P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \ldots,$$

为给定 $Y = y_j$ 条件下 X 的条件分布列。为什么是分布列?

- 二维离散随机变量 (X, Y) 的联合分布列为

$$p_{ij} = P(X = x_i, Y = y_j), \quad i, j = 1, 2, \ldots.$$

- 对一切使得 $P(Y = y_j) = p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij} > 0$ 的 y_j，称

$$p_{i|j} = P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \ldots,$$

为给定 $Y = y_j$ 条件下 X 的条件分布列。为什么是分布列?

- 给定 $Y = y_j$ 条件下 X 的条件分布函数为

$$F(x|y_j) = \sum_{x_i \leqslant x} P(X = x_i | Y = y_j) = \sum_{x_i \leqslant x} p_{i|j}.$$

例子

- 随机变量 X 与 Y 相互独立，且 $X \sim P(\lambda_1)$, $Y \sim P(\lambda_2)$. 求 $X + Y = n$ 条件下 X 的条件分布。

例子

- 随机变量 X 与 Y 相互独立，且 $X \sim P(\lambda_1)$，$Y \sim P(\lambda_2)$. 求 $X+Y=n$ 条件下 X 的条件分布。
- $X+Y \sim P(\lambda_1+\lambda_2)$.
- $$\begin{aligned} P(X=k|X+Y=n) &= \frac{P(X=k, X+Y=n)}{P(X+Y=n)} \\ &= \frac{P(X=k)P(X=n-k)}{P(X+Y=n)} \\ &= \frac{\frac{\lambda_1^k}{k!}e^{-\lambda_1}\frac{\lambda_2^{n-k}}{(n-k)!}e^{-\lambda_2}}{\frac{(\lambda_1+\lambda_2)^n}{n!}e^{-\lambda_1-\lambda_2}} \\ &= \frac{n!}{k!(n-k)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}. \end{aligned}$$

例子

- 随机变量 X 与 Y 相互独立，且 $X \sim P(\lambda_1)$，$Y \sim P(\lambda_2)$. 求 $X+Y=n$ 条件下 X 的条件分布。
- $X+Y \sim P(\lambda_1+\lambda_2)$.
- $$\begin{aligned} P(X=k|X+Y=n) &= \frac{P(X=k, X+Y=n)}{P(X+Y=n)} \\ &= \frac{P(X=k)P(X=n-k)}{P(X+Y=n)} \\ &= \frac{\frac{\lambda_1^k}{k!}e^{-\lambda_1}\frac{\lambda_2^{n-k}}{(n-k)!}e^{-\lambda_2}}{\frac{(\lambda_1+\lambda_2)^n}{n!}e^{-\lambda_1-\lambda_2}} \\ &= \frac{n!}{k!(n-k)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}. \end{aligned}$$
- 这是二项分布。

连续随机变量的条件分布

- 二维随机变量 (X, Y) 的联合密度函数为 $p(x, y)$, 边际密度函数分别为 $p_X(x)$, $p_Y(y)$. 对于使得 $p_Y(y) > 0$ 的 y, 在给定 $Y = y$ 的条件下 X 的条件分布函数和条件密度函数分别为

$$F(x|y) = \int_{-\infty}^{x} \frac{p(u, y)}{p_Y(y)} du, \quad p(x|y) = \frac{p(x, y)}{p_Y(y)}.$$

- 二维随机变量 (X, Y) 的联合密度函数为 $p(x, y)$, 边际密度函数分别为 $p_X(x)$, $p_Y(y)$. 对于使得 $p_Y(y) > 0$ 的 y, 在给定 $Y = y$ 的条件下 X 的条件分布函数和条件密度函数分别为

$$F(x|y) = \int_{-\infty}^{x} \frac{p(u, y)}{p_Y(y)} du, \quad p(x|y) = \frac{p(x, y)}{p_Y(y)}.$$

$$\begin{aligned} P(X \leqslant x| Y = y) &= \lim_{h \to 0} P(X \leqslant x|y \leqslant Y \leqslant y + h) \\ &= \lim_{h \to 0} \frac{P(X \leqslant x, y \leqslant Y \leqslant y + h)}{P(y \leqslant Y \leqslant y + h)} \\ &= \lim_{h \to 0} \frac{\int_{-\infty}^{x} \int_{y}^{y+h} p(x', y') dy' dx'}{\int_{y}^{y+h} p_Y(y') dy'} \\ &= \int_{-\infty}^{x} \frac{p(x', y)}{p_Y(y)} dx'. \end{aligned}$$

例子

- $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

例子

- $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.
- X 在 $Y = y$ 条件下的条件密度函数为

$$
\begin{aligned}
& p(x|y) \\
& = \frac{p(x, y)}{p_Y(y)} \\
& = \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2})\}}{\frac{1}{\sqrt{2\pi}\sigma_2} \exp\{-\frac{(y-\mu_2)}{2\sigma_2^2}\}} \\
& = \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\{-\frac{1}{2\sigma_1^2(1-\rho^2)}[x - (\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2))]^2\} \\
& \sim N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2), \sigma_1^2(1 - \rho^2)).
\end{aligned}
$$

连续场合的全概率公式和贝叶斯公式

- $p(x, y) = p_X(x)p(y|x) = p_Y(y)p(x|y)$.

连续场合的全概率公式和贝叶斯公式

- $p(x, y) = p_X(x)p(y|x) = p_Y(y)p(x|y)$.
- 全概率公式:

$$p_Y(y) = \int_{-\infty}^{\infty} p_X(x)p(y|x)\,dx,$$

$$p_X(x) = \int_{-\infty}^{\infty} p_Y(y)p(x|y)\,dy.$$

连续场合的全概率公式和贝叶斯公式

- $p(x, y) = p_X(x)p(y|x) = p_Y(y)p(x|y)$.
- 全概率公式:

$$p_Y(y) = \int_{-\infty}^{\infty} p_X(x)p(y|x)\,dx,$$

$$p_X(x) = \int_{-\infty}^{\infty} p_Y(y)p(x|y)\,dy.$$

- 贝叶斯公式:

$$p(x|y) = \frac{p_X(x)p(y|x)}{\int_{-\infty}^{\infty} p_X(x)p(y|x)\,dx}.$$

$$p(y|x) = \frac{p_Y(y)p(x|y)}{\int_{-\infty}^{\infty} p_Y(y)p(x|y)\,dy}.$$

条件期望

- 条件分布的数学期望称为条件期望，即

$$E(X|\,Y=y)=\begin{cases}\sum_i x_i P(X=x_i|\,Y=y), & \text{离散情形}\\ \int_{-\infty}^{\infty} xp(x|y)\,dx, & \text{连续情形}.\end{cases}$$

$$E(\,Y|X=x)=\begin{cases}\sum_j y_j P(\,Y=y_j|X=x), & \text{离散情形}\\ \int_{-\infty}^{\infty} yp(y|x)\,dy, & \text{连续情形}.\end{cases}$$

条件期望

- 条件分布的数学期望称为条件期望，即

$$E(X|Y=y)=\begin{cases}\sum_i x_i P(X=x_i|Y=y), & \text{离散情形}\\ \int_{-\infty}^{\infty} xp(x|y)\,dx, & \text{连续情形.}\end{cases}$$

$$E(Y|X=x)=\begin{cases}\sum_j y_j P(Y=y_j|X=x), & \text{离散情形}\\ \int_{-\infty}^{\infty} yp(y|x)\,dy, & \text{连续情形.}\end{cases}$$

- 条件期望 $E(X|Y=y)$ 是 y 的函数，若令 $g(y)=E(X|Y=y)$，则 $g(Y)=E(X|Y)$ 是一个随机变量，我们称为 X 关于 Y 的条件期望。

条件期望的性质

- 重期望公式 $E(X) = E(E(X|Y))$.

- 重期望公式 $E(X) = E(E(X|Y))$.

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xp(x,y)\,dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xp(x|y)p_Y(y)\,dydx \\
&= \int_{-\infty}^{\infty}\Big[\int_{-\infty}^{\infty} xp(x|y)\,dx\Big]p_Y(y)\,dy \\
&= \int_{-\infty}^{\infty} E(X|Y=y)p_Y(y)\,dy = E(E(X|Y)).
\end{aligned}$$

条件期望的性质

- 重期望公式 $E(X) = E(E(X|Y))$.

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xp(x, y)\,dy dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xp(x|y)p_Y(y)\,dy dx \\
&= \int_{-\infty}^{\infty}\Big[\int_{-\infty}^{\infty} xp(x|y)\,dx\Big]p_Y(y)\,dy \\
&= \int_{-\infty}^{\infty} E(X|Y=y)p_Y(y)\,dy = E(E(X|Y)).
\end{aligned}
$$

条件期望的性质

- 如果 X 与 Y 相互独立，则 $E(X|Y) = E(X)$.

$$E(X|Y=y) = \int xp(x|y)\,dx = \int xp(x)\,dx = E(X).$$

条件期望的性质

- 如果 X 与 Y 相互独立，则 $E(X|Y) = E(X)$.

$$E(X|Y=y) = \int xp(x|y)\,dx = \int xp(x)\,dx = E(X).$$

- $E(f(Y)Z|Y) = f(Y)E(Z|Y)$.

条件期望的性质

- 如果 X 与 Y 相互独立，则 $E(X|Y) = E(X)$.

$$E(X|Y=y) = \int xp(x|y)\,dx = \int xp(x)\,dx = E(X).$$

- $E(f(Y)Z|Y) = f(Y)E(Z|Y)$.

$$\begin{aligned} E(f(Y)Z|Y=y) &= \int_{-\infty}^{\infty} f(y)zp(f(y)z|y)\,dz \\ &= f(y)\int_{-\infty}^{\infty} zp(z|y)\,dz = f(y)E(Z|Y=y). \end{aligned}$$

随机个随机变量和的数学期望

- 设 $X_1, \dots,$ 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 $\{X_1, \dots\}$ 独立，则

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

随机个随机变量和的数学期望

- 设 $X_1,\dots,$ 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 $\{X_1,\dots\}$ 独立，则

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

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

概率论与数理统计（6）

清华大学

2020 年春季学期

依概率收敛

- 设 $\{X_n\}$ 为一随机变量序列，X 为一随机变量，如果对任意的 $\epsilon > 0$, 有
$$P(|X_n - X| \geqslant \epsilon) \to 0, \quad n \to \infty,$$
则称序列 $\{X_n\}$ 依概率收敛于 X, 记作 $X_n \xrightarrow{P} X$.

依概率收敛

- 设 $\{X_n\}$ 为一随机变量序列，X 为一随机变量，如果对任意的 $\epsilon > 0$, 有
$$P(|X_n - X| \geqslant \epsilon) \to 0, \quad n \to \infty,$$
则称序列 $\{X_n\}$ 依概率收敛于 X, 记作 $X_n \xrightarrow{P} X$.
- 依概率收敛的含义是：绝对偏差 $|X_n - X|$ 小于任何给定量的可能性随着 n 的增大越来越接近 1，其等价于
$$P(|X_n - X| < \epsilon) \to 1, \quad n \to \infty.$$

依概率收敛的运算规则

- $\{X_n\}$, $\{Y_n\}$ 为两个随机变量序列，a, b 为两个常数，若

$$X_n \xrightarrow{P} a, \quad Y_n \xrightarrow{P} b,$$

则有

  - $X_n + Y_n \xrightarrow{P} a + b$;

依概率收敛的运算规则

- $\{X_n\}, \{Y_n\}$ 为两个随机变量序列，a, b 为两个常数，若

$$X_n \xrightarrow{P} a, \quad Y_n \xrightarrow{P} b,$$

则有

  - $X_n + Y_n \xrightarrow{P} a + b;$
  - $X_n \times Y_n \xrightarrow{P} a \times b;$

依概率收敛的运算规则

- $\{X_n\}$, $\{Y_n\}$ 为两个随机变量序列，a, b 为两个常数，若

$$X_n \xrightarrow{P} a, \quad Y_n \xrightarrow{P} b,$$

则有

  - $X_n + Y_n \xrightarrow{P} a + b$;
  - $X_n \times Y_n \xrightarrow{P} a \times b$;
  - 若 $b \neq 0$, 则 $X_n / Y_n \xrightarrow{P} a/b$.

依概率收敛的运算规则

- $X_n + Y_n \xrightarrow{P} a + b$;

依概率收敛的运算规则

- $X_n + Y_n \xrightarrow{P} a + b$;
-

$$\{|X_n + Y_n - (a+b)| \geqslant \epsilon\} \subset \left\{(|X_n - a| \geqslant \frac{\epsilon}{2}) \cup (|Y_n - b| \geqslant \frac{\epsilon}{2})\right\},$$

依概率收敛的运算规则

- $X_n + Y_n \xrightarrow{P} a + b$;
- $$\{|X_n + Y_n - (a+b)| \geqslant \epsilon\} \subset \left\{(|X_n - a| \geqslant \frac{\epsilon}{2}) \cup (|Y_n - b| \geqslant \frac{\epsilon}{2})\right\},$$
- $$0 \leqslant P(|X_n + Y_n - (a+b)| \geqslant \epsilon) \leqslant P((|X_n - a| \geqslant \frac{\epsilon}{2}) + P(|Y_n - b| \geqslant \frac{\epsilon}{2}) \to 0,$$

依概率收敛的运算规则

- $X_n \times Y_n \xrightarrow{P} a \times b.$

依概率收敛的运算规则

- $X_n \times Y_n \xrightarrow{P} a \times b.$
- $X_n \xrightarrow{P} 0, \Rightarrow \ X_n^2 \xrightarrow{P} 0.$ 因为 $P(|X_n^2| \geqslant \epsilon) = P(|X_n| \geqslant \sqrt{\epsilon}).$

依概率收敛的运算规则

- $X_n \times Y_n \xrightarrow{P} a \times b.$
- $X_n \xrightarrow{P} 0, \Rightarrow \quad X_n^2 \xrightarrow{P} 0.$ 因为 $P(|X_n^2| \geqslant \epsilon) = P(|X_n| \geqslant \sqrt{\epsilon}).$
- $X_n \xrightarrow{P} a, \Rightarrow \forall c \in \mathbb{R}, \quad cX_n \xrightarrow{P} ca$ 因为对于 $c \neq 0$ （$c = 0$ 的情况显然）

$$P(|cX_n - ca| \geqslant \epsilon) = P(|X_n - a| \geqslant \epsilon/c).$$

依概率收敛的运算规则

- $X_n \times Y_n \xrightarrow{P} a \times b.$
- $X_n \xrightarrow{P} 0, \Rightarrow \quad X_n^2 \xrightarrow{P} 0.$ 因为 $P(|X_n^2| \geqslant \epsilon) = P(|X_n| \geqslant \sqrt{\epsilon}).$
- $X_n \xrightarrow{P} a, \Rightarrow \forall c \in \mathbb{R}, \quad cX_n \xrightarrow{P} ca$ 因为对于 $c \neq 0$ （$c = 0$ 的情况显然）

$$P(|cX_n - ca| \geqslant \epsilon) = P(|X_n - a| \geqslant \epsilon/c).$$

- $X_n \xrightarrow{P} a, \Rightarrow X_n^2 \xrightarrow{P} a^2$ 因为 $X_n^2 - a^2 = (X_n - a)^2 - 2a(X_n - a).$

依概率收敛的运算规则

- $X_n \times Y_n \xrightarrow{P} a \times b$.
- $X_n \xrightarrow{P} 0, \Rightarrow \quad X_n^2 \xrightarrow{P} 0$. 因为 $P(|X_n^2| \geqslant \epsilon) = P(|X_n| \geqslant \sqrt{\epsilon})$.
- $X_n \xrightarrow{P} a, \Rightarrow \forall c \in \mathbb{R}, \quad cX_n \xrightarrow{P} ca$ 因为对于 $c \neq 0$ （$c = 0$ 的情况显然）
$$P(|cX_n - ca| \geqslant \epsilon) = P(|X_n - a| \geqslant \epsilon/c).$$

- $X_n \xrightarrow{P} a, \Rightarrow X_n^2 \xrightarrow{P} a^2$ 因为 $X_n^2 - a^2 = (X_n - a)^2 - 2a(X_n - a)$.
- $X_n \times Y_n = \frac{1}{2}[(X_n + Y_n)^2 - X_n^2 - Y_n^2] \xrightarrow{P} \frac{1}{2}[(a + b)^2 - a^2 - b^2]$.

依概率收敛的运算规则

- $b \neq 0,\ X_n/Y_n \xrightarrow{P} a/b.$

依概率收敛的运算规则

- $b \neq 0,\ X_n/Y_n \xrightarrow{P} a/b.$
- 首先 $1/Y_n \xrightarrow{P} 1/b$:

依概率收敛的运算规则

- $b \neq 0,\ X_n/Y_n \xrightarrow{P} a/b.$
- 首先 $1/Y_n \xrightarrow{P} 1/b$:

$$
\begin{aligned}
P(|1/Y_n - 1/b| \geqslant \epsilon) &= P(|\frac{Y_n - b}{Y_n b}| \geqslant \epsilon) \\
&= P\Big(|\frac{Y_n - b}{b^2 + b(Y_n - b)}| \geqslant \epsilon, |Y_n - b| < \epsilon\Big) \\
&\quad + P\Big(|\frac{Y_n - b}{b^2 + b(Y_n - b)}| \geqslant \epsilon, |Y_n - b| \geqslant \epsilon\Big) \\
&\leqslant P(|\frac{Y_n - b}{b^2 - \epsilon|b|}| \geqslant \epsilon) + P(|Y_n - b| \geqslant \epsilon).
\end{aligned}
$$

依概率收敛的运算规则

- $b \neq 0$, $X_n/Y_n \xrightarrow{P} a/b$.
- 首先 $1/Y_n \xrightarrow{P} 1/b$:

$$\begin{aligned}
P(|1/Y_n - 1/b| \geqslant \epsilon) &= P(|\frac{Y_n - b}{Y_n b}| \geqslant \epsilon) \\
&= P\Big(|\frac{Y_n - b}{b^2 + b(Y_n - b)}| \geqslant \epsilon, |Y_n - b| < \epsilon\Big) \\
&\quad + P\Big(|\frac{Y_n - b}{b^2 + b(Y_n - b)}| \geqslant \epsilon, |Y_n - b| \geqslant \epsilon\Big) \\
&\leqslant P(|\frac{Y_n - b}{b^2 - \epsilon|b|}| \geqslant \epsilon) + P(|Y_n - b| \geqslant \epsilon).
\end{aligned}$$

- $X_n/Y_n = X_n \times (1/Y_n)$.

按分布收敛

- 设随机变量序列 $\{X_n\}$ 的分布函数为 $\{F_n(x)\}$, X 的分布函数为 $F(x)$. 若对于 $F(x)$ 的任一连续点 x, 都有

$$F_n(x) \to F(x), \quad n \to \infty,$$

则称 $\{F_n(x)\}$ 弱收敛于 $F(x)$, 记作 $F_n(x) \xrightarrow{W} F$, 同时称 $\{X_n\}$ 按分布收敛于 X, 记作

$$X_n \xrightarrow{L} X.$$

两种收敛的关系

- $X_n \xrightarrow{P} X, \Rightarrow X_n \xrightarrow{L} X.$

两种收敛的关系

- $X_n \xrightarrow{P} X, \Rightarrow X_n \xrightarrow{L} X.$

$$X_n \xrightarrow{P} X \;\Rightarrow\; F(x-0) \leqslant \liminf_{n\to\infty} F_n(x) \leqslant \limsup_{n\to\infty} F_n(x) \leqslant F(x+0).$$

两种收敛的关系

- $X_n \xrightarrow{P} X, \Rightarrow X_n \xrightarrow{L} X.$

$$X_n \xrightarrow{P} X \;\Rightarrow\; F(x-0) \leqslant \liminf_{n\to\infty} F_n(x) \leqslant \limsup_{n\to\infty} F_n(x) \leqslant F(x+0).$$

$$\begin{aligned} \text{当}\, x' < x,\; \{X \leqslant x'\} &= \{X_n \leqslant x, X \leqslant x'\} \cup \{X_n > x, X \leqslant x'\} \\ &\subset \{X_n \leqslant x\} \cup \{|X_n - X| \geqslant x - x'\}, \end{aligned}$$

- $X_n \xrightarrow{P} X, \Rightarrow X_n \xrightarrow{L} X.$

$$X_n \xrightarrow{P} X \;\Rightarrow\; F(x-0) \leqslant \liminf_{n\to\infty} F_n(x) \leqslant \limsup_{n\to\infty} F_n(x) \leqslant F(x+0).$$

$$\begin{aligned} \text{当}\, x' < x,\ \{X \leqslant x'\} &= \{X_n \leqslant x, X \leqslant x'\} \cup \{X_n > x, X \leqslant x'\} \\ &\subset \{X_n \leqslant x\} \cup \{|X_n - X| \geqslant x - x'\}, \end{aligned}$$

$$F(x') \leqslant F_n(x) + P(|X_n - X| > x - x') \Rightarrow F(x') \leqslant \liminf_{n\to\infty} F_n(x).$$

两种收敛的关系

- $X_n \xrightarrow{P} X, \Rightarrow X_n \xrightarrow{L} X.$

$$X_n \xrightarrow{P} X \;\Rightarrow\; F(x-0) \leqslant \liminf_{n\to\infty} F_n(x) \leqslant \limsup_{n\to\infty} F_n(x) \leqslant F(x+0).$$

$$\begin{aligned}\text{当}\, x' < x,\ \{X \leqslant x'\} &= \{X_n \leqslant x, X \leqslant x'\} \cup \{X_n > x, X \leqslant x'\} \\ &\subset \{X_n \leqslant x\} \cup \{|X_n - X| \geqslant x - x'\},\end{aligned}$$

$$F(x') \leqslant F_n(x) + P(|X_n - X| > x - x') \Rightarrow F(x') \leqslant \liminf_{n\to\infty} F_n(x).$$

$$F(x-0) \leqslant \liminf_{n\to\infty} F_n(x).$$

两种收敛的关系

- $X_n \xrightarrow{P} X, \Rightarrow X_n \xrightarrow{L} X.$

$$X_n \xrightarrow{P} X \;\Rightarrow\; F(x-0) \leqslant \liminf_{n\to\infty} F_n(x) \leqslant \limsup_{n\to\infty} F_n(x) \leqslant F(x+0).$$

$$\text{当} x' < x,\ \{X \leqslant x'\} = \{X_n \leqslant x, X \leqslant x'\} \cup \{X_n > x, X \leqslant x'\}$$
$$\subset \{X_n \leqslant x\} \cup \{|X_n - X| \geqslant x - x'\},$$

$$F(x') \leqslant F_n(x) + P(|X_n - X| > x - x') \Rightarrow F(x') \leqslant \liminf_{n\to\infty} F_n(x).$$

$$F(x-0) \leqslant \liminf_{n\to\infty} F_n(x).$$

类似地，$x < x'$, $\{X_n \leqslant x\} = \{X < x', X_n \leqslant x\} \cup \{X > x', X_n \leqslant x\}$

$$\limsup_{n\to\infty} F_n(x) \leqslant F(x+0).$$

两种收敛的关系

- $X_n \xrightarrow{P} X, \Rightarrow X_n \xrightarrow{L} X.$

$$X_n \xrightarrow{P} X \;\Rightarrow\; F(x-0) \leqslant \liminf_{n\to\infty} F_n(x) \leqslant \limsup_{n\to\infty} F_n(x) \leqslant F(x+0).$$

$$\text{当} x' < x,\ \{X \leqslant x'\} = \{X_n \leqslant x, X \leqslant x'\} \cup \{X_n > x, X \leqslant x'\}$$
$$\subset \{X_n \leqslant x\} \cup \{|X_n - X| \geqslant x - x'\},$$

$$F(x') \leqslant F_n(x) + P(|X_n - X| > x - x') \Rightarrow F(x') \leqslant \liminf_{n\to\infty} F_n(x).$$

$$F(x-0) \leqslant \liminf_{n\to\infty} F_n(x).$$

类似地，$x < x'$, $\{X_n \leqslant x\} = \{X < x', X_n \leqslant x\} \cup \{X > x', X_n \leqslant x\}$

$$\limsup_{n\to\infty} F_n(x) \leqslant F(x+0).$$

- 一般情况下按分布收敛不能推出依概率收敛：考虑
$X:\ P(X=-1)=0.5,\ P(X=1)=0.5,\ X_n = -X.$

两种收敛的关系

- 若 c 为常数，$X_n \xrightarrow{L} c \Rightarrow X_n \xrightarrow{P} c.$

两种收敛的关系

- 若 c 为常数，$X_n \xrightarrow{L} c \Rightarrow X_n \xrightarrow{P} c$.
- $X = c$ 的分布函数为

$$F(x) = \begin{cases} 0, & x < c, \\ 1, & x \geqslant c. \end{cases}.$$

- 若 c 为常数，$X_n \xrightarrow{L} c \Rightarrow X_n \xrightarrow{P} c$.
- $X = c$ 的分布函数为

$$F(x) = \begin{cases} 0, & x < c, \\ 1, & x \geqslant c. \end{cases}.$$

对于任意 $\epsilon > 0$, 有

$$\begin{aligned} P(|X_n - c| \geqslant \epsilon) &= P(X_n \geqslant c + \epsilon) + P(X_n \leqslant c - \epsilon) \\ &\leqslant P(X_n > c + \epsilon/2) + P(X_n \leqslant c - \epsilon) \\ &= 1 - F_n(c + \epsilon/2) + F_n(c - \epsilon) \\ &= 0, \quad n \to \infty. \end{aligned}$$

两种收敛的关系

- 若 c 为常数，$X_n \xrightarrow{L} c \Rightarrow X_n \xrightarrow{P} c$.
- $X = c$ 的分布函数为

$$F(x) = \begin{cases} 0, & x < c, \\ 1, & x \geqslant c. \end{cases}.$$

对于任意 $\epsilon > 0$, 有

$$\begin{aligned} P(|X_n - c| \geqslant \epsilon) &= P(X_n \geqslant c + \epsilon) + P(X_n \leqslant c - \epsilon) \\ &\leqslant P(X_n > c + \epsilon/2) + P(X_n \leqslant c - \epsilon) \\ &= 1 - F_n(c + \epsilon/2) + F_n(c - \epsilon) \\ &= 0, \quad n \to \infty. \end{aligned}$$

- X_n 依分布收敛到 X, Y_n 依分布收敛到 Y, $X_n + Y_n$ 是否依分布收敛到 $X + Y$?

依概率 1 收敛或者几乎必然收敛

- 如果 $P(\lim_{n\to\infty} X_n = X) = 1$, 则称 X_n 依概率 1 收敛于 X, 记为

$$X_n \to X, \ \ a.s.$$

依概率 1 收敛或者几乎必然收敛

- 如果 $P(\lim_{n\to\infty} X_n = X) = 1$, 则称 X_n 依概率 1 收敛于 X, 记为

$$X_n \to X, \ a.s.$$

-

$$X_n \to X, \ a.s. \quad \text{当且仅当} \quad \lim_{n\to\infty} P(\cup_{k\geqslant n}|X_n - X| \geqslant \epsilon) = 0.$$

依概率 1 收敛或者几乎必然收敛

- 如果 $P(\lim_{n\to\infty} X_n = X) = 1$, 则称 X_n 依概率 1 收敛于 X, 记为
$$X_n \to X,\ a.s.$$

- $$X_n \to X,\ a.s. \quad \text{当且仅当} \quad \lim_{n\to\infty} P(\cup_{k\geqslant n}|X_n - X| \geqslant \epsilon) = 0.$$

- $\forall \epsilon > 0,\ A_n^\epsilon = \{|X_n - X| \geqslant \epsilon\},\ A^\epsilon = \cap_{n=1}^\infty \cup_{k\geqslant n} A_k^\epsilon$, 则
$$\{X_n \to X\}^c = \cup_{m=1}^\infty A^{1/m}.$$

- $$0 = P(\{X_n \to X\}^c) \leftrightarrow P(\cup_{m=1}^\infty A^{1/m}) = 0 \leftrightarrow P(A^{1/m}) = 0, \quad \forall m \geqslant 1$$
$$\leftrightarrow \lim_{n\to\infty} P(\cup_{k\geqslant n}|X_k - X| \geqslant \epsilon) = 0.$$

依概率 1 收敛于依概率收敛

- $X_n \to X,\ a.s. \Rightarrow X_n \xrightarrow{P} X.$

依概率 1 收敛于依概率收敛

- $X_n \to X,\ a.s. \Rightarrow X_n \xrightarrow{P} X.$
- 反之不然。

依概率 1 收敛于依概率收敛

- $X_n \to X,\ a.s. \Rightarrow X_n \xrightarrow{P} X.$
- 反之不然。考虑 $S = (0, 1)$ 的均匀分布，令随机变量

$$X_n(x) = \begin{cases} 1, & x \in (\frac{k}{2^i}, \frac{k+1}{2^i}), \\ 0, & else, \end{cases}$$

其中 i, k 满足 $n = 2^i + k$. 令 $X = 0$.

依概率 1 收敛于依概率收敛

- $X_n \to X,\ a.s. \Rightarrow X_n \xrightarrow{P} X.$
- 反之不然。考虑 $S = (0,1)$ 的均匀分布，令随机变量

$$X_n(x) = \begin{cases} 1, & x \in (\frac{k}{2^i}, \frac{k+1}{2^i}), \\ 0, & else, \end{cases}$$

其中 $i,\ k$ 满足 $n = 2^i + k$. 令 $X = 0$.显然任意 x 都有无限个 n 使得 $X_n(x) = 1$, 所以 X_n 不依概率 1 收敛于 0. 但依概率收敛于 0:

依概率 1 收敛于依概率收敛

- $X_n \to X,\ a.s. \Rightarrow X_n \xrightarrow{P} X.$
- 反之不然。考虑 $S=(0,1)$ 的均匀分布，令随机变量

$$X_n(x)=\begin{cases}1, & x\in(\frac{k}{2^i},\frac{k+1}{2^i}),\\ 0, & else,\end{cases}$$

其中 i, k 满足 $n=2^i+k$. 令 $X=0$.显然任意 x 都有无限个 n 使得 $X_n(x)=1$, 所以 X_n 不依概率 1 收敛于 0. 但依概率收敛于 0:

$$P(|X_n-0|\geqslant\epsilon)\leqslant\frac{1}{2^{i(n)}}\to 0,\quad n\to\infty.$$

- $X_n \to X,\ a.s. \Rightarrow X_n \xrightarrow{P} X.$
- 反之不然。考虑 $S=(0,1)$ 的均匀分布，令随机变量

$$X_n(x)=\begin{cases}1, & x\in(\frac{k}{2^i},\frac{k+1}{2^i}),\\ 0, & else,\end{cases}$$

其中 i, k 满足 $n=2^i+k$. 令 $X=0$.显然任意 x 都有无限个 n 使得 $X_n(x)=1$, 所以 X_n 不依概率 1 收敛于 0. 但依概率收敛于 0:

$$P(|X_n-0|\geqslant\epsilon)\leqslant\frac{1}{2^{i(n)}}\to 0,\quad n\to\infty.$$

几乎必然收敛不作考试要求

特征函数的定义

- 若 X 为一随机变量，则称

$$\varphi(t) = E(e^{itX}), \quad -\infty < t < \infty,$$

为 X 的特征函数。

特征函数的定义

- 若 X 为一随机变量，则称

$$\varphi(t) = E(e^{itX}), \quad -\infty < t < \infty,$$

为 X 的特征函数。

- $e^{itx} = \cos tx + i \sin tx$, $|e^{itX}| = 1$，所以 $E(|e^{itX}|) = 1$, $E(e^{itX})$ 总是存在的。

特征函数的定义

- 若 X 为一随机变量，则称

$$\varphi(t) = E(e^{itX}), \quad -\infty < t < \infty,$$

为 X 的特征函数。

- $e^{itx} = \cos tx + i\sin tx$, $|e^{itX}| = 1$，所以 $E(|e^{itX}|) = 1$, $E(e^{itX})$ 总是存在的。
- 离散情形：X 的分布列为 $p_k = P(X = x_k)$, $k = 1, 2, \ldots$，则 X 的特征函数为

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

特征函数的定义

- 若 X 为一随机变量，则称

$$\varphi(t) = E(e^{itX}), \quad -\infty < t < \infty,$$

为 X 的特征函数。

- $e^{itx} = \cos tx + i \sin tx$, $|e^{itX}| = 1$，所以 $E(|e^{itX}|) = 1$, $E(e^{itX})$ 总是存在的。
- 离散情形：X 的分布列为 $p_k = P(X = x_k)$, $k = 1, 2, \ldots$，则 X 的特征函数为

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

连续情形：X 的密度函数为 $p(x)$, 则 X 的特征函数为

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

常见分布的特征函数一

- 单点分布：$P(X=a)=1$, $\varphi(t)=E(e^{itX})=e^{ita}$.

常见分布的特征函数一

- 单点分布：$P(X=a)=1$, $\varphi(t)=E(e^{itX})=e^{ita}$.
- $0-1$ 分布：$P(X=1)=p$, $P(X=0)=1-p$, $\varphi(t)=pe^{it}+(1-p)$.

常见分布的特征函数一

- 单点分布：$P(X=a)=1$, $\varphi(t)=E(e^{itX})=e^{ita}$.
- $0-1$ 分布：$P(X=1)=p$, $P(X=0)=1-p$, $\varphi(t)=pe^{it}+(1-p)$.
- 泊松分布 $P(\lambda)$: $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,\ldots,$

$$\varphi(t)=E(e^{itX})=\sum_{k=0}^{\infty}e^{ikt}\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!}e^{-\lambda}=e^{\lambda(e^{it}-1)}.$$

常见分布的特征函数一

- 单点分布：$P(X=a)=1$, $\varphi(t)=E(e^{itX})=e^{ita}$.
- $0-1$ 分布：$P(X=1)=p$, $P(X=0)=1-p$, $\varphi(t)=pe^{it}+(1-p)$.
- 泊松分布 $P(\lambda)$: $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,\ldots,$

$$\varphi(t)=E(e^{itX})=\sum_{k=0}^{\infty}e^{ikt}\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!}e^{-\lambda}=e^{\lambda(e^{it}-1)}.$$

- 均匀分布 $U(a,b)$: 其密度函数为 $p(x)=\begin{cases}\frac{1}{b-a}, & x\in(a,b),\\ 0, & else\end{cases}$, 则

$$\varphi(t)=E(e^{itX})=\int_a^b\frac{e^{itx}}{b-a}dx=\frac{e^{ibt}-e^{iat}}{it(b-a)}.$$

常见分布的特征函数二

- 标准正态分布 $N(0,1)$: 密度函数为 $p(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, $-\infty<x<\infty$.

- 标准正态分布 $N(0,1)$: 密度函数为 $p(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, $-\infty<x<\infty$.

$$\begin{aligned}\varphi(t)&=\int_{-\infty}^{\infty}e^{itx}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx=\int_{-\infty}^{\infty}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\sum_{n=0}^{\infty}\frac{(itx)^n}{n!}dx\\&=\sum_{n=0}^{\infty}\frac{(it)^n}{n!}\int_{-\infty}^{\infty}\frac{x^n e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx.\\&=\sum_{m=0}^{\infty}\frac{(it)^{2m}}{(2m)!}\frac{(2m)!}{2^m m!}=\sum_{m=0}^{\infty}(-\frac{t^2}{2})^m\frac{1}{m!}=e^{-\frac{t^2}{2}}.\end{aligned}$$

常见分布的特征函数三

- 指数分布 $Exp(\lambda)$，其密度函数为

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leqslant 0, \end{cases}$$

常见分布的特征函数三

- 指数分布 $Exp(\lambda)$，其密度函数为

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leqslant 0, \end{cases}$$

- 其特征函数为

$$\begin{aligned} \varphi(t) &= E(e^{itX}) = \int_0^{\infty} e^{itx} \lambda e^{-\lambda x} dx \\ &= \lambda[\int_0^{\infty} \cos(tx) e^{-\lambda x} dx + i \int_0^{\infty} \sin(tx) e^{-\lambda x} dx] \\ &= \lambda[\frac{\lambda}{\lambda^2 + t^2} + i \frac{t}{\lambda^2 + t^2}] = (1 - \frac{it}{\lambda})^{-1}. \end{aligned}$$

- $\int_0^{\infty} cos(tx) e^{-\lambda x} dx = -\frac{1}{\lambda} \cos(tx) e^{-\lambda x}|_0^{\infty} - \frac{1}{\lambda} \int_0^{\infty} t \sin tx e^{-\lambda x} dx.$

特征函数的性质一

- $|\varphi(t)| \leqslant \varphi(0) = 1.$

特征函数的性质一

- $|\varphi(t)| \leqslant \varphi(0) = 1.$
- $\varphi(-t) = \overline{\varphi(t)}.$

特征函数的性质一

- $|\varphi(t)| \leqslant \varphi(0) = 1$.
- $\varphi(-t) = \overline{\varphi(t)}$.
- 若 $Y = aX + b$, 其中 a, b 为常数，则 $\varphi_Y(t) = e^{ibt}\varphi_X(at)$.

- $|\varphi(t)| \leqslant \varphi(0) = 1$.
- $\varphi(-t) = \overline{\varphi(t)}$.
- 若 $Y = aX + b$, 其中 a, b 为常数，则 $\varphi_Y(t) = e^{ibt}\varphi_X(at)$.

$$E(e^{i(aX+b)t}) = \int_{-\infty}^{\infty} e^{iaxt+ibt} p_X(x)\,dx = e^{ibt}\int_{-\infty}^{\infty} e^{ix(at)} p_X(x)\,dx.$$

- 独立随机变量和的特征函数为每个随机变量的特征函数的积：若 X 与 Y 相互独立，则

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

特征函数的性质一

- $|\varphi(t)| \leqslant \varphi(0) = 1$.
- $\varphi(-t) = \overline{\varphi(t)}$.
- 若 $Y = aX + b$, 其中 a, b 为常数，则 $\varphi_Y(t) = e^{ibt}\varphi_X(at)$.

$$E(e^{i(aX+b)t}) = \int_{-\infty}^{\infty} e^{iaxt+ibt} p_X(x)\,dx = e^{ibt}\int_{-\infty}^{\infty} e^{ix(at)} p_X(x)\,dx.$$

- 独立随机变量和的特征函数为每个随机变量的特征函数的积：若 X 与 Y 相互独立，则

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

$$\begin{aligned} E(e^{i(X+Y)t}) &= \int_{-\infty}^{\infty} e^{i(x+y)t} p_X(x) p_Y(y)\,dxdy \\ &= \int_{-\infty}^{\infty} e^{ixt} p_X(x)\,dx \int_{-\infty}^{\infty} e^{iyt} p_Y(y)\,dy. \end{aligned}$$

常见分布的特征函数四

- 二项分布 $b(n,p)$: 若 $Y \sim b(n,p)$, 则 $Y = X_1 + \cdots + X_n$，其中 X_i 是独立同分布的 $0-1$ 分布随机变量，所以我们有

$$\varphi_Y(t) = \prod_{k=1}^{n} \varphi_{X_i}(t) = (pe^{it} + 1 - p)^n.$$

常见分布的特征函数四

- 二项分布 $b(n,p)$: 若 $Y \sim b(n,p)$, 则 $Y = X_1 + \cdots + X_n$，其中 X_i 是独立同分布的 $0-1$ 分布随机变量，所以我们有

$$\varphi_Y(t) = \prod_{k=1}^{n} \varphi_{X_i}(t) = (pe^{it} + 1 - p)^n.$$

- 正态分布 $N(\mu, \sigma^2)$： $X = \frac{Y-\mu}{\sigma} \sim N(0,1)$, 则 $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$, 所以

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

常见分布的特征函数四

- 二项分布 $b(n,p)$: 若 $Y \sim b(n,p)$, 则 $Y = X_1 + \cdots + X_n$，其中 X_i 是独立同分布的 $0-1$ 分布随机变量，所以我们有

$$\varphi_Y(t) = \prod_{k=1}^{n} \varphi_{X_i}(t) = (pe^{it} + 1 - p)^n.$$

- 正态分布 $N(\mu,\sigma^2)$： $X = \frac{Y-\mu}{\sigma} \sim N(0,1)$, 则 $Y = \sigma X + \mu \sim N(\mu,\sigma^2)$, 所以

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

- 伽玛分布 $Ga(n,\lambda)$: $Y \sim Ga(n,\lambda)$, 则 $Y = X_1 + \cdots + X_n$，其中 X_i 为独立同分布的服从指数分布 $Exp(\lambda)$ 的随机变量，则

$$\varphi_Y(t) = \prod_{k=1}^{n} \varphi_{X_i}(t) = (1 - \frac{it}{\lambda})^{-n}.$$

特征函数的性质二

- 一致连续性：随机变量 X 的特征函数 $\varphi(t)$ 在 $(-\infty, \infty)$ 上一致连续，即

特征函数的性质二

- 一致连续性：随机变量 X 的特征函数 $\varphi(t)$ 在 $(-\infty, \infty)$ 上一致连续，即任意 $\epsilon > 0$，存在 $\delta > 0$，使得若 $|x-y| \leqslant \delta$,，则
$$|\varphi(x) - \varphi(y)| \leqslant \epsilon.$$

特征函数的性质二

- 一致连续性：随机变量 X 的特征函数 $\varphi(t)$ 在 $(-\infty,\infty)$ 上一致连续，即任意 $\epsilon>0$，存在 $\delta>0$，使得若 $|x-y|\leqslant\delta,$，则
$$|\varphi(x)-\varphi(y)|\leqslant\epsilon.$$

- 对于任意 $a>0$ 和 $t,h\in\mathbb{R}$，有
$$|\varphi(t+h)-\varphi(t)|=\left|\int_{-\infty}^{+\infty}(e^{ihx}-1)e^{itx}p(x)\,dx\right|$$
$$=\int_{-\infty}^{\infty}|e^{ihx}-1|p(x)\,dx\leqslant\int_{-a}^{a}|e^{ihx}-1|p(x)\,dx+2\int_{|x|>a}p(x)\,dx.$$

- 一致连续性：随机变量 X 的特征函数 $\varphi(t)$ 在 $(-\infty,\infty)$ 上一致连续，即任意 $\epsilon>0$, 存在 $\delta>0$, 使得若 $|x-y|\leqslant\delta$,, 则

$$|\varphi(x)-\varphi(y)|\leqslant\epsilon.$$

- 对于任意 $a>0$ 和 $t,h\in\mathbb{R}$, 有

$$|\varphi(t+h)-\varphi(t)|=|\int_{-\infty}^{+\infty}(e^{ihx}-1)e^{itx}p(x)\,dx|$$

$$=\int_{-\infty}^{\infty}|e^{ihx}-1|p(x)\,dx\leqslant\int_{-a}^{a}|e^{ihx}-1|p(x)\,dx+2\int_{|x|>a}p(x)\,dx.$$

对于任意 $\epsilon>0$, 先取 a 充分大使得 $2P(|X|>a)\leqslant\frac{\epsilon}{2}$, 再取 $h=\frac{\epsilon}{2a}$,

$$\int_{-a}^{a}|e^{ihx}-1|p(x)\,dx=\int_{-a}^{a}|e^{ihx/2}(e^{ihx/2}-e^{-ihx/2})|p(x)\,dx$$

$$\leqslant|2\sin hx/2|2a<2ah.$$

特征函数的性质三

- 非负定性：随机变量 X 的特征函数 $\varphi(t)$ 是非负定的，即对任意正整数 n 及 n 个实数 $t_1, \ldots, t_n$，和 n 个复数 $z_1, \ldots, z_n$，有

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

特征函数的性质三

- 非负定性：随机变量 X 的特征函数 $\varphi(t)$ 是非负定的，即对任意正整数 n 及 n 个实数 $t_1, \ldots, t_n$，和 n 个复数 $z_1, \ldots, z_n$，有

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

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{j=1}^{n}\varphi(t_k - t_j) z_k \bar{z}_j &= \sum_{k=1}^{n}\sum_{j=1}^{n} z_k \bar{z}_j \int_{-\infty}^{\infty} e^{i(t_k - t_j)x} p(x)\, dx \\
&= \int_{-\infty}^{\infty} \sum_{k=1}^{n}\sum_{j=1}^{n} z_k \bar{z}_j e^{i(t_k - t_j)x} p(x)\, dx \\
&= \int_{-\infty}^{\infty} (\sum_{k=1}^{n} z_k e^{it_k x})(\sum_{j=1}^{n} e^{-it_j x} \bar{z}_j) p(x)\, dx \\
&= \int_{-\infty}^{\infty} |\sum_{k=1}^{n} z_k e^{it_k x}|^2 p(x)\, dx \geqslant 0.
\end{aligned}$$

特征函数的性质四

- 若 $E(X^m)$ 存在，则特征函数 $\varphi(t)$ m 次可导，且对于 $1 \leqslant k \leqslant m$, 有

$$\varphi^{(k)}(0) = i^k E(X^k).$$

特征函数的性质四

- 若 $E(X^m)$ 存在，则特征函数 $\varphi(t)$ m 次可导，且对于 $1 \leqslant k \leqslant m$，有

$$\varphi^{(k)}(0) = i^k E(X^k).$$

$$\varphi^{(k)}(t) = \int_{-\infty}^{\infty} \frac{d^m}{dt}(e^{ixt})p(x)\,dx = \int_{-\infty}^{\infty} i^k x^k e^{itx} p(x)\,dx.$$

特征函数的性质四

- 若 $E(X^m)$ 存在，则特征函数 $\varphi(t)$ m 次可导，且对于 $1 \leqslant k \leqslant m$，有

$$\varphi^{(k)}(0) = i^k E(X^k).$$

$$\varphi^{(k)}(t) = \int_{-\infty}^{\infty} \frac{d^m}{dt}(e^{ixt})p(x)\,dx = \int_{-\infty}^{\infty} i^k x^k e^{itx} p(x)\,dx.$$

特别地，我们有，

$$E(X) = \frac{\varphi'(0)}{i}, \quad Var(X) = -\varphi''(0) + (\varphi'(0))^2.$$

特征函数的性质四

- 若 $E(X^m)$ 存在，则特征函数 $\varphi(t)$ m 次可导，且对于 $1 \leqslant k \leqslant m$，有

$$\varphi^{(k)}(0) = i^k E(X^k).$$

$$\varphi^{(k)}(t) = \int_{-\infty}^{\infty} \frac{d^m}{dt}(e^{ixt})p(x)\,dx = \int_{-\infty}^{\infty} i^k x^k e^{itx} p(x)\,dx.$$

特别地，我们有，

$$E(X) = \frac{\varphi'(0)}{i}, \quad Var(X) = -\varphi''(0) + (\varphi'(0))^2.$$

这两条公式是证明大数定律和中心极限定理的关键！

特征函数唯一决定分布函数

- 逆转公式：设 $F(x)$ 和 $\varphi(t)$ 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

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

特征函数唯一决定分布函数

- 逆转公式：设 $F(x)$ 和 $\varphi(t)$ 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

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

- 唯一性定理：随机函数的分布函数由其特征函数唯一决定。

特征函数唯一决定分布函数

- 逆转公式：设 $F(x)$ 和 $\varphi(t)$ 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

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

- 唯一性定理：随机函数的分布函数由其特征函数唯一决定。对于 $F(x)$ 的每一个连续点 x，让 y_n 沿着 $F(x)$ 的连续点趋向 $-\infty$，则由逆转公式有

$$\begin{aligned} F(x) &= \lim_{y_n\to-\infty} F(x) - F(y_n) \\ &= \lim_{y_n\to-\infty} \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ity_n} - e^{-itx}}{it} \varphi(t)\, dt. \end{aligned}$$

特征函数唯一决定密度函数

- 设 X 为连续随机变量，其密度函数为 $p(x)$, 特征函数为 $\varphi(t)$. 若

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

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

特征函数唯一决定密度函数

- 设 X 为连续随机变量，其密度函数为 $p(x)$, 特征函数为 $\varphi(t)$. 若

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

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

-

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

特征函数唯一决定密度函数

- 设 X 为连续随机变量，其密度函数为 $p(x)$, 特征函数为 $\varphi(t)$. 若

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

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

-

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

- 这个公式一般被称为傅里叶逆变换公式。

一些例子

- 相互独立正态分布的可加性：$X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2)$，$X$ 与 Y 相互独立，则 $X + Y$ 的特征函数为

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{it\mu_1 - \frac{\sigma_1^2 t^2}{2}} e^{it\mu_2 - \frac{\sigma_2^2 t^2}{2}} = e^{it(\mu_1+\mu_2) - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}},$$

所以 $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

- 相互独立正态分布的可加性：$X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2)$，$X$ 与 Y 相互独立，则 $X+Y$ 的特征函数为

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{it\mu_1 - \frac{\sigma_1^2 t^2}{2}} e^{it\mu_2 - \frac{\sigma_2^2 t^2}{2}} = e^{it(\mu_1+\mu_2) - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}},$$

所以 $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

- 伽玛分布的可加性：$X \sim Ga(\alpha_1, \lambda)$，$Y \sim Ga(\alpha_2, \lambda)$，$X$ 与 Y 相互独立，则 $X+Y$ 的特征函数为

$$\varphi_{X+Y}(t) = (1-\frac{it}{\lambda})^{-\alpha_1}(1-\frac{it}{\lambda})^{-\alpha_2} = (1-\frac{it}{\lambda})^{-\alpha_1-\alpha_2}.$$

一些例子

- 相互独立正态分布的可加性：$X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2)$，$X$ 与 Y 相互独立，则 $X+Y$ 的特征函数为

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{it\mu_1 - \frac{\sigma_1^2 t^2}{2}} e^{it\mu_2 - \frac{\sigma_2^2 t^2}{2}} = e^{it(\mu_1+\mu_2) - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}},$$

  所以 $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.
- 伽玛分布的可加性：$X \sim Ga(\alpha_1, \lambda)$，$Y \sim Ga(\alpha_2, \lambda)$，$X$ 与 Y 相互独立，则 $X+Y$ 的特征函数为

$$\varphi_{X+Y}(t) = (1 - \frac{it}{\lambda})^{-\alpha_1}(1 - \frac{it}{\lambda})^{-\alpha_2} = (1 - \frac{it}{\lambda})^{-\alpha_1-\alpha_2}.$$

  所以 $X+Y \sim Ga(\alpha_1+\alpha_2, \lambda)$.

一些例子

- 某概率分布的特征函数为 $\varphi(t) = e^{-|t|}$, 求对应的分布?

一些例子

- 某概率分布的特征函数为 $\varphi(t) = e^{-|t|}$, 求对应的分布?
- $$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}e^{-|t|}dt = \frac{1}{2\pi}\int_{0}^{\infty} e^{-(1+ix)t}dt + \frac{1}{2\pi}\int_{-\infty}^{0} e^{(1-ix)t}dt$$
  $$= \frac{1}{2\pi}(\frac{1}{1+ix} + \frac{1}{1-ix}) = \frac{1}{\pi(1+x^2)}.$$

一些例子

- 某概率分布的特征函数为 $\varphi(t) = e^{-|t|}$, 求对应的分布?
- $$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}e^{-|t|}dt = \frac{1}{2\pi}\int_{0}^{\infty} e^{-(1+ix)t}dt + \frac{1}{2\pi}\int_{-\infty}^{0} e^{(1-ix)t}dt$$
$$= \frac{1}{2\pi}(\frac{1}{1+ix} + \frac{1}{1-ix}) = \frac{1}{\pi(1+x^2)}.$$
- 这是柯西分布。

特征函数与弱收敛

- 分布函数列 $\{F_n(x)\}$ 弱收敛于分布函数 $F(x)$ 的充分必要条件是 $\{F_n\}$ 的特征函数序列 $\{\varphi_n(t)\}$ 弱收敛于 $F(x)$ 的特征函数 $\varphi(t)$.

特征函数与弱收敛

- 分布函数列 $\{F_n(x)\}$ 弱收敛于分布函数 $F(x)$ 的充分必要条件是 $\{F_n\}$ 的特征函数序列 $\{\varphi_n(t)\}$ 弱收敛于 $F(x)$ 的特征函数 $\varphi(t)$.
- X_λ 服从参数为 λ 的泊松分布，则

$$\lim_{\lambda\to\infty} P\left(\frac{X_\lambda-\lambda}{\sqrt{\lambda}}\leqslant x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt.$$

- 分布函数列 $\{F_n(x)\}$ 弱收敛于分布函数 $F(x)$ 的充分必要条件是 $\{F_n\}$ 的特征函数序列 $\{\varphi_n(t)\}$ 弱收敛于 $F(x)$ 的特征函数 $\varphi(t)$.
- X_λ 服从参数为 λ 的泊松分布，则

$$\lim_{\lambda\to\infty} P(\frac{X_\lambda-\lambda}{\sqrt{\lambda}} \leqslant x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt.$$

- X_λ 的特征函数为 $\varphi_\lambda(t) = e^{\lambda(e^{it}-1)}$，则 $\frac{X_\lambda-\lambda}{\sqrt{\lambda}}$ 的特征函数为

$$g_\lambda(t) = \varphi_\lambda(\frac{t}{\sqrt{\lambda}})e^{-i\sqrt{\lambda}t} = e^{\lambda(e^{i\frac{t}{\sqrt{\lambda}}}-1)-i\sqrt{\lambda}t}.$$

- 分布函数列 $\{F_n(x)\}$ 弱收敛于分布函数 $F(x)$ 的充分必要条件是 $\{F_n\}$ 的特征函数序列 $\{\varphi_n(t)\}$ 弱收敛于 $F(x)$ 的特征函数 $\varphi(t)$.
- X_λ 服从参数为 λ 的泊松分布，则

$$\lim_{\lambda\to\infty} P(\frac{X_\lambda-\lambda}{\sqrt{\lambda}} \leqslant x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt.$$

- X_λ 的特征函数为 $\varphi_\lambda(t)=e^{\lambda(e^{it}-1)}$，则 $\frac{X_\lambda-\lambda}{\sqrt{\lambda}}$ 的特征函数为

$$g_\lambda(t)=\varphi_\lambda(\frac{t}{\sqrt{\lambda}})e^{-i\sqrt{\lambda}t}=e^{\lambda(e^{i\frac{t}{\sqrt{\lambda}}}-1)-i\sqrt{\lambda}t}.$$

指数部分 $\lambda(e^{i\frac{t}{\sqrt{\lambda}}}-1)-i\sqrt{\lambda}t=$
$\lambda(1+\frac{it}{\sqrt{\lambda}}-\frac{t^2}{2\lambda}+O(\frac{1}{\lambda^{3/2}})-1)-i\sqrt{\lambda}t\to-\frac{t^2}{2},\quad \lambda\to\infty.$

概率论与数理统计（7）

清华大学

2020 年春季学期

几点回顾

- 切比雪夫不等式：对于任意 $c > 0$, $P(|X - E(X)| \geqslant c) \leqslant \frac{1}{c^2} Var(X)$.
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$.
- 特征函数的在 $t = 0$ 处的泰勒展开：
  $\varphi(t) = 1 + iE(X) + \frac{1}{2} i^2 E(X^2) + \dots$.
- 依概率收敛与依分布收敛，当极限是常数时，二者等价。
- 如果 X_n 的特征函数 $\varphi_n(t)$ 收敛到 X 的特征函数时，X_n 依分布收敛到 X.
- 若 X 与 Y 相互独立，则 $X + Y$ 的特征函数为
  $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.
- $Y = aX + b$, 则 Y 的特征函数为 $\varphi_Y(t) = e^{ibt}\varphi_X(at)$.

泊松定理（重制版）

- 对于任意的 $n \in \mathbb{N}$, 考虑 n 个独立随机变量 $X_{n1}, \ldots, X_{nn}$,

$$P(X_{nk} = 1) = p_{nk}, \quad P(X_{nk} = 0) = 1 - p_{nk}.$$

若当 $n \to \infty$ 时，$\max_{1 \leqslant k \leqslant n} p_{nk} \to 0$, 且 $\sum_{k=1}^{n} p_{nk} \to \lambda$，则若令 $S_n = X_{n1} + \cdots + X_{nn}$, 有对于任意 $m = 0, 1, \ldots,$

$$P(S_n = m) \to \frac{\lambda^m}{m!} e^{-\lambda}, \quad n \to \infty.$$

泊松定理（重制版）

- 对于任意的 $n \in \mathbb{N}$, 考虑 n 个独立随机变量 $X_{n1}, \ldots, X_{nn}$,

$$P(X_{nk} = 1) = p_{nk}, \quad P(X_{nk} = 0) = 1 - p_{nk}.$$

若当 $n \to \infty$ 时，$\max_{1 \leqslant k \leqslant n} p_{nk} \to 0$, 且 $\sum_{k=1}^{n} p_{nk} \to \lambda$，则若令 $S_n = X_{n1} + \cdots + X_{nn}$, 有对于任意 $m = 0, 1, \ldots,$

$$P(S_n = m) \to \frac{\lambda^m}{m!} e^{-\lambda}, \quad n \to \infty.$$

- $\varphi_{S_n}(t) = E(e^{itS_n}) = \prod_{k=1}^{n} (p_{nk} e^{it} + 1 - p_{nk}) = \prod_{k=1}^{n} (1 + p_{nk}(e^{it} - 1)),$

泊松定理（重制版）

- 对于任意的 $n \in \mathbb{N}$, 考虑 n 个独立随机变量 $X_{n1}, \ldots, X_{nn}$,

$$P(X_{nk} = 1) = p_{nk}, \quad P(X_{nk} = 0) = 1 - p_{nk}.$$

若当 $n \to \infty$ 时，$\max_{1 \leqslant k \leqslant n} p_{nk} \to 0$, 且 $\sum_{k=1}^{n} p_{nk} \to \lambda$，则若令 $S_n = X_{n1} + \cdots + X_{nn}$, 有对于任意 $m = 0, 1, \ldots$,

$$P(S_n = m) \to \frac{\lambda^m}{m!} e^{-\lambda}, \quad n \to \infty.$$

- $\varphi_{S_n}(t) = E(e^{itS_n}) = \prod_{k=1}^{n} (p_{nk} e^{it} + 1 - p_{nk}) = \prod_{k=1}^{n} (1 + p_{nk}(e^{it} - 1))$, 而由条件假设 $Y(n, t) = \prod_{k=1}^{n} (1 + p_{nk}(e^{it} - 1)) \to e^{\lambda(e^{it} - 1)}$.

泊松定理（重制版）

- 对于任意的 $n \in \mathbb{N}$, 考虑 n 个独立随机变量 $X_{n1}, \dots, X_{nn}$,

$$P(X_{nk} = 1) = p_{nk}, \quad P(X_{nk} = 0) = 1 - p_{nk}.$$

若当 $n \to \infty$ 时，$\max_{1 \leqslant k \leqslant n} p_{nk} \to 0$, 且 $\sum_{k=1}^{n} p_{nk} \to \lambda$，则若令 $S_n = X_{n1} + \cdots + X_{nn}$, 有对于任意 $m = 0, 1, \dots,$

$$P(S_n = m) \to \frac{\lambda^m}{m!} e^{-\lambda}, \quad n \to \infty.$$

- $\varphi_{S_n}(t) = E(e^{itS_n}) = \prod_{k=1}^{n} (p_{nk} e^{it} + 1 - p_{nk}) = \prod_{k=1}^{n} (1 + p_{nk}(e^{it} - 1))$, 而由条件假设 $Y(n, t) = \prod_{k=1}^{n} (1 + p_{nk}(e^{it} - 1)) \to e^{\lambda(e^{it}-1)}$.

$$\ln Y(n, t) = \sum_{k=1}^{n} \ln(1 + p_{nk}(e^{it} - 1)) = \sum_{k=1}^{n} p_{nk}(e^{it} - 1) + O(\sum_{k=1}^{n} p_{nk}^2).$$

- (伯努利大数定律) 设 S_n 为 n 重伯努利试验中事件 A 发生的次数，p 为每次试验中 A 发生的概率，则对任意的 $\epsilon > 0$, 有

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

大数定律

- (伯努利大数定律) 设 S_n 为 n 重伯努利试验中事件 A 发生的次数, p 为每次试验中 A 发生的概率, 则对任意的 $\epsilon > 0$, 有

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

也就是说 $\frac{S_n}{n} \xrightarrow{P} p$.

大数定律

- (伯努利大数定律) 设 S_n 为 n 重伯努利试验中事件 A 发生的次数，p 为每次试验中 A 发生的概率，则对任意的 $\epsilon > 0$, 有

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

也就是说 $\frac{S_n}{n} \xrightarrow{P} p$.

- $S_n \sim b(n,p)$, $E(\frac{S_n}{n}) = p$, $Var(\frac{S_n}{n}) = \frac{p(1-p)}{n}$.

大数定律

- (伯努利大数定律) 设 S_n 为 n 重伯努利试验中事件 A 发生的次数，p 为每次试验中 A 发生的概率，则对任意的 $\epsilon > 0$, 有

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

  也就是说 $\frac{S_n}{n} \xrightarrow{P} p$.

- $S_n \sim b(n, p)$, $E(\frac{S_n}{n}) = p$, $Var(\frac{S_n}{n}) = \frac{p(1-p)}{n}$.由切比雪夫不等式有

$$P(|\frac{S_n}{n} - p| \geqslant \epsilon) \leqslant \frac{p(1-p)}{n\epsilon^2}.$$

  也就是 $1 - \frac{p(1-p)}{n\epsilon^2} \leqslant P(|\frac{S_n}{n} - p| < \epsilon) \leqslant 1$.

- (伯努利大数定律) 设 S_n 为 n 重伯努利试验中事件 A 发生的次数, p 为每次试验中 A 发生的概率, 则对任意的 $\epsilon > 0$, 有

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

也就是说 $\frac{S_n}{n} \xrightarrow{P} p$.

- $S_n \sim b(n, p)$, $E(\frac{S_n}{n}) = p$, $Var(\frac{S_n}{n}) = \frac{p(1-p)}{n}$.由切比雪夫不等式有

$$P(|\frac{S_n}{n} - p| \geqslant \epsilon) \leqslant \frac{p(1-p)}{n\epsilon^2}.$$

也就是 $1 - \frac{p(1-p)}{n\epsilon^2} \leqslant P(|\frac{S_n}{n} - p| < \epsilon) \leqslant 1.$

- 可以通过大量重复试验得到真实概率的近似值。

蒙特卡罗方法求积分一

- 设二维随机变量服从正方形 $\{0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$ 上的均匀分布，则 X 与 Y 相互独立，且均服从 [0,1] 上的均匀分布，

蒙特卡罗方法求积分一

- 设二维随机变量服从正方形 $\{0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$ 上的均匀分布，则 X 与 Y 相互独立，且均服从 [0,1] 上的均匀分布，考虑 $[0,1]$ 上的函数 $f(x)$，且 $0 \leqslant f(x) \leqslant 1$，令 $A = \{Y \leqslant f(X)\}$，

蒙特卡罗方法求积分一

- 设二维随机变量服从正方形 $\{0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$ 上的均匀分布，则 X 与 Y 相互独立，且均服从 [0,1] 上的均匀分布，考虑 $[0,1]$ 上的函数 $f(x)$，且 $0 \leqslant f(x) \leqslant 1$，令 $A = \{Y \leqslant f(X)\}$，则

$$P(A) = \int_0^1 \int_0^{f(x)} dydx = \int_0^1 f(x)\,dx.$$

- 设计伯努利试验：每次任取 $(x, y) \in [0,1] \times [0,1]$，记录 A 发生的次数。频率近似积分 $\int_0^1 f(x)\,dx$.

蒙特卡罗方法求积分一

- 设二维随机变量服从正方形 $\{0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$ 上的均匀分布，则 X 与 Y 相互独立，且均服从 [0,1] 上的均匀分布，考虑 $[0,1]$ 上的函数 $f(x)$, 且 $0 \leqslant f(x) \leqslant 1$, 令 $A = \{Y \leqslant f(X)\}$, 则

$$P(A) = \int_0^1 \int_0^{f(x)} dydx = \int_0^1 f(x)\,dx.$$

- 设计伯努利试验：每次任取 $(x, y) \in [0,1] \times [0,1]$, 记录 A 发生的次数。频率近似积分 $\int_0^1 f(x)\,dx$.
- 对于一般区间 $[a, b]$ 上的函数 $g(x)$, 若 $c \leqslant g(x) \leqslant d$, 则令

$$f(y) = \frac{1}{d-c}[g(a+(b-a)y) - c]$$

$$\int_a^b g(x)\,dx = \int_0^1 g(a+(b-a)y)(b-a)\,dy = \frac{d-c}{b-a}\int_0^1 f(y)\,dy + (b-a)c.$$

大数定律的一般形式

- 一随机变量序列 $\{X_n\}$ 若满足性质：对于任意的 $\epsilon > 0$, 均有

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

则称该随机变量序列 $\{X_n\}$ 服从大数定律。

大数定律的一般形式

- 一随机变量序列 $\{X_n\}$ 若满足性质：对于任意的 $\epsilon > 0$, 均有

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

  则称该随机变量序列 $\{X_n\}$ 服从大数定律。
- (切比雪夫大数定律) 若 $\{X_n\}$ 为两两互不相关的随机变量序列，若每个 X_n 的方差存在，且有共同上界，即

$$Var(X_n) \leqslant c < \infty, n = 1, 2, \ldots$$

  则该序列服从大数定律。

大数定律的一般形式

- 一随机变量序列 $\{X_n\}$ 若满足性质：对于任意的 $\epsilon > 0$，均有

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

  则称该随机变量序列 $\{X_n\}$ 服从大数定律。

- （切比雪夫大数定律）若 $\{X_n\}$ 为两两互不相关的随机变量序列，若每个 X_n 的方差存在，且有共同上界，即

$$Var(X_n) \leqslant c < \infty, n = 1, 2, \ldots$$

  则该序列服从大数定律。

- 伯努利大数定律是切比雪夫大数定律的特例。

切比雪夫大数定律的证明

- $$Var(\frac{1}{n}\sum_{i=1}^{n}X_i) = \frac{1}{n^2}\sum_{i=1}^{n}Var(X_i) \leqslant \frac{c}{n}.$$
- 对于任意的 $\epsilon > 0$, 有切比雪夫不等式有,

$$P(|\frac{1}{n}\sum_{i=1}^{n}X_i - \frac{1}{n}\sum_{i=1}^{n}E(X_i)| < \epsilon) \geqslant 1 - \frac{c}{n\epsilon^2}.$$

例子

- $\{X_n\}$ 是独立同分布随机变量，$E(X_n^4) < \infty$, 则对于任意的 $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))^2 - \frac{1}{n}\sum_{i=1}^{n} Var(X_i)| < \epsilon) = 1.$$

例子

- $\{X_n\}$ 是独立同分布随机变量，$E(X_n^4) < \infty$, 则对于任意的 $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))^2 - \frac{1}{n}\sum_{i=1}^{n} Var(X_i)| < \epsilon) = 1.$$

- 令 $Y_n = (X_n - E(X_n))^2$,

例子

- $\{X_n\}$ 是独立同分布随机变量，$E(X_n^4) < \infty$, 则对于任意的 $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))^2 - \frac{1}{n}\sum_{i=1}^{n} Var(X_i)| < \epsilon) = 1.$$

- 令 $Y_n = (X_n - E(X_n))^2$, 则
$Var(Y_n) = E\Big(X_n - E(X_n)\Big)^4 - (Var(X_n))^2 < \infty.$

例子

- $\{X_n\}$ 是独立同分布随机变量，$E(X_n^4) < \infty$, 则对于任意的 $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))^2 - \frac{1}{n}\sum_{i=1}^{n} Var(X_i)| < \epsilon) = 1.$$

- 令 $Y_n = (X_n - E(X_n))^2$, 则 $Var(Y_n) = E\Big(X_n - E(X_n)\Big)^4 - (Var(X_n))^2 < \infty$.满足切比雪夫大数定律的条件。

例子

- $\{X_n\}$ 是独立同分布随机变量，$E(X_n^4) < \infty$，则对于任意的 $\epsilon > 0$，

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))^2 - \frac{1}{n}\sum_{i=1}^{n} Var(X_i)| < \epsilon) = 1.$$

- 令 $Y_n = (X_n - E(X_n))^2$，则 $Var(Y_n) = E\Big(X_n - E(X_n)\Big)^4 - (Var(X_n))^2 < \infty$.满足切比雪夫大数定律的条件。
- $\frac{1}{n}\sum_{i=1}^{n} Var(X_i) = Var(X)$，方差统计量的理论基础。

马尔可夫大数定律

- 对于随机变量序列 $\{X_n\}$，若

$$\frac{1}{n^2}Var(\sum_{k=1}^{n}X_k) \to 0, \quad n \to \infty,$$

则该随机变量序列满足大数定律。

马尔可夫大数定律

- 对于随机变量序列 $\{X_n\}$，若

$$\frac{1}{n^2}Var(\sum_{k=1}^{n}X_k)\to 0,\quad n\to\infty,$$

则该随机变量序列满足大数定律。

- 设 $\{X_n\}$ 为一同分布，方差存在的随机变量序列，且 X_n 仅与相邻的 X_{n-1} 和 X_{n+1} 相关，而与其他的不相关，则该序列服从大数定律。

马尔可夫大数定律

- 对于随机变量序列 $\{X_n\}$，若

$$\frac{1}{n^2} Var(\sum_{k=1}^{n} X_k) \to 0, \quad n \to \infty,$$

则该随机变量序列满足大数定律。

- 设 $\{X_n\}$ 为一同分布，方差存在的随机变量序列，且 X_n 仅与相邻的 X_{n-1} 和 X_{n+1} 相关，而与其他的不相关，则该序列服从大数定律。
-

$$\frac{1}{n^2} Var(\sum_{k=1}^{n} X_k) = \frac{1}{n^2}[\sum_{k=1}^{n} Var(X_k)+2\sum_{i=1}^{n-1} Cov(X_i, X_{i+1})] \to 0, \quad n \to \infty$$

马尔可夫大数定律

- 对于随机变量序列 $\{X_n\}$，若

$$\frac{1}{n^2}Var(\sum_{k=1}^{n}X_k)\to 0,\quad n\to\infty,$$

则该随机变量序列满足大数定律。

- 设 $\{X_n\}$ 为一同分布，方差存在的随机变量序列，且 X_n 仅与相邻的 X_{n-1} 和 X_{n+1} 相关，而与其他的不相关，则该序列服从大数定律。
-

$$\frac{1}{n^2}Var(\sum_{k=1}^{n}X_k)=\frac{1}{n^2}[\sum_{k=1}^{n}Var(X_k)+2\sum_{i=1}^{n-1}Cov(X_i,X_{i+1})]\to 0,\quad n\to\infty$$

- 跟相邻的任意有限 m 个相关，也成立。

- 对于随机变量序列 $\{X_n\}$，若

$$\frac{1}{n^2}Var(\sum_{k=1}^{n}X_k) \to 0, \quad n \to \infty,$$

则该随机变量序列满足大数定律。

- 设 $\{X_n\}$ 为一同分布，方差存在的随机变量序列，且 X_n 仅与相邻的 X_{n-1} 和 X_{n+1} 相关，而与其他的不相关，则该序列服从大数定律。
-

$$\frac{1}{n^2}Var(\sum_{k=1}^{n}X_k) = \frac{1}{n^2}[\sum_{k=1}^{n}Var(X_k)+2\sum_{i=1}^{n-1}Cov(X_i,X_{i+1})] \to 0, \quad n \to \infty$$

- 跟相邻的任意有限 m 个相关，也成立。可以更一般（作业）。

辛钦大数定律

- 若 $\{X_n\}$ 为独立同分布的随机变量序列，且 X_n 的数学期望存在，则该序列服从大数定律。

辛钦大数定律

- 若 $\{X_n\}$ 为独立同分布的随机变量序列，且 X_n 的数学期望存在，则该序列服从大数定律。
- $$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} a = E(X_1), \quad n \to \infty.$$

辛钦大数定律

- 若 $\{X_n\}$ 为独立同分布的随机变量序列，且 X_n 的数学期望存在，则该序列服从大数定律。
- $$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} a = E(X_1), \quad n \to \infty.$$
- X_n 的特征函数为 $\varphi(t) = \varphi(0) + \varphi'(0)t + o(t) = 1 + iat + o(t)$.

辛钦大数定律

- 若 $\{X_n\}$ 为独立同分布的随机变量序列，且 X_n 的数学期望存在，则该序列服从大数定律。
- $$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} a = E(X_1), \quad n \to \infty.$$
- X_n 的特征函数为 $\varphi(t) = \varphi(0) + \varphi'(0)t + o(t) = 1 + iat + o(t).$
- $Y_n = \frac{1}{n}\sum_{k=1} X_k$ 的特征函数为

辛钦大数定律

- 若 $\{X_n\}$ 为独立同分布的随机变量序列，且 X_n 的数学期望存在，则该序列服从大数定律。
- $$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} a = E(X_1), \quad n \to \infty.$$
- X_n 的特征函数为 $\varphi(t) = \varphi(0) + \varphi'(0)t + o(t) = 1 + iat + o(t)$.
- $Y_n = \frac{1}{n}\sum_{k=1} X_k$ 的特征函数为
$$\varphi_{Y_n}(t) = [\varphi(\frac{t}{n})]^n = (1 + ia\frac{t}{n} + o(\frac{t}{n}))^n \to e^{iat}.$$
- e^{iat} 为常值（随机变量）$X = a$ 的特征函数，所以
$$Y_n \xrightarrow{L} a \Rightarrow Y_n \xrightarrow{P} a.$$

辛钦大数定律

- 若 $\{X_n\}$ 为独立同分布的随机变量序列，且 X_n 的数学期望存在，则该序列服从大数定律。
- $$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} a = E(X_1), \quad n \to \infty.$$
- X_n 的特征函数为 $\varphi(t) = \varphi(0) + \varphi'(0)t + o(t) = 1 + iat + o(t)$.
- $Y_n = \frac{1}{n}\sum_{k=1} X_k$ 的特征函数为
$$\varphi_{Y_n}(t) = [\varphi(\frac{t}{n})]^n = (1 + ia\frac{t}{n} + o(\frac{t}{n}))^n \to e^{iat}.$$
- e^{iat} 为常值（随机变量）$X = a$ 的特征函数，所以
$$Y_n \xrightarrow{L} a \Rightarrow Y_n \xrightarrow{P} a.$$
- 注意：这里并不要求方差存在！

一个例子：数学期望必须存在！

- 设随机变量 $X_1, \ldots, X_n$ 相互独立且服从同一柯西分布，其密度函数均为
$$p(x) = \frac{1}{\pi(1+x^2)},$$
考虑
$$Y = \frac{1}{n}(X_1 + \cdots + X_n)$$
的密度函数。

一个例子：数学期望必须存在！

- 设随机变量 $X_1, \ldots, X_n$ 相互独立且服从同一柯西分布，其密度函数均为
$$p(x) = \frac{1}{\pi(1+x^2)},$$
考虑
$$Y = \frac{1}{n}(X_1 + \cdots + X_n)$$
的密度函数。
- X_i 的特征函数为 $\varphi(t) = e^{-|t|}$,

一个例子：数学期望必须存在！

- 设随机变量 $X_1, \ldots, X_n$ 相互独立且服从同一柯西分布，其密度函数均为
$$p(x) = \frac{1}{\pi(1+x^2)},$$
考虑
$$Y = \frac{1}{n}(X_1 + \cdots + X_n)$$
的密度函数。
- X_i 的特征函数为 $\varphi(t) = e^{-|t|}$,
- $X = X_1 + \cdots + X_n$ 的特征函数为 $\varphi_X(t) = (e^{-|t|})^n = e^{-n|t|}$,

一个例子：数学期望必须存在！

- 设随机变量 $X_1, \ldots, X_n$ 相互独立且服从同一柯西分布，其密度函数均为
$$p(x) = \frac{1}{\pi(1+x^2)},$$
考虑
$$Y = \frac{1}{n}(X_1 + \cdots + X_n)$$
的密度函数。
- X_i 的特征函数为 $\varphi(t) = e^{-|t|}$，
- $X = X_1 + \cdots + X_n$ 的特征函数为 $\varphi_X(t) = (e^{-|t|})^n = e^{-n|t|}$，
- Y 的特征函数为 $\varphi_Y(t) = \varphi_X(\frac{t}{n}) = e^{-|t|}$，

一个例子：数学期望必须存在！

- 设随机变量 $X_1, \ldots, X_n$ 相互独立且服从同一柯西分布，其密度函数均为
$$p(x) = \frac{1}{\pi(1+x^2)},$$
考虑
$$Y = \frac{1}{n}(X_1 + \cdots + X_n)$$
的密度函数。
- X_i 的特征函数为 $\varphi(t) = e^{-|t|}$，
- $X = X_1 + \cdots + X_n$ 的特征函数为 $\varphi_X(t) = (e^{-|t|})^n = e^{-n|t|}$，
- Y 的特征函数为 $\varphi_Y(t) = \varphi_X(\frac{t}{n}) = e^{-|t|}$，
- 所以 Y 的密度函数为 $p(y) = \frac{1}{\pi(1+y^2)}$，$-\infty < y < \infty$.

蒙特卡罗方法求积分二

- 设 X 为 $[a, b]$ 上的均匀分布，则 $Y = f(X)$ 的数学期望是

$$E(Y) = \int_a^b f(x)\,dx.$$

蒙特卡罗方法求积分二

- 设 X 为 $[a,b]$ 上的均匀分布，则 $Y=f(X)$ 的数学期望是

$$E(Y)=\int_a^b f(x)\,dx.$$

- $X_i \sim U(a,b)$ 相互独立，则 $f(X_i)$ 相互独立。

蒙特卡罗方法求积分二

- 设 X 为 $[a,b]$ 上的均匀分布，则 $Y=f(X)$ 的数学期望是

$$E(Y)=\int_a^b f(x)\,dx.$$

- $X_i \sim U(a,b)$ 相互独立，则 $f(X_i)$ 相互独立。
-

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \int_a^b f(x)\,dx.$$

蒙特卡罗方法求积分二

- 设 X 为 $[a,b]$ 上的均匀分布，则 $Y=f(X)$ 的数学期望是

$$E(Y)=\int_a^b f(x)\,dx.$$

- $X_i \sim U(a,b)$ 相互独立，则 $f(X_i)$ 相互独立。
- $$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \int_a^b f(x)\,dx.$$
- 可以推广到多重积分。这种计算方法在二重以上的积分才比较有意义。

- X_n 是独立同分布的随机变量序列，$\frac{1}{n}\sum_{i=1}^{n} X_n^k \to ?, \quad k \geqslant 1$

样本矩：抢先版

- X_n 是独立同分布的随机变量序列，$\frac{1}{n}\sum_{i=1}^{n} X_n^k \to ?, \quad k \geqslant 1$
- 只要 $E(X_n^k)$ 存在，有辛欣大数定律，它收敛到 $E(X^k)$.

样本矩：抢先版

- X_n 是独立同分布的随机变量序列，$\frac{1}{n}\sum_{i=1}^n X_n^k \to ?, \quad k \geqslant 1$
- 只要 $E(X_n^k)$ 存在，有辛欣大数定律，它收敛到 $E(X^k)$.
- $E(X^3)$ 存在，$f(x_1, x_2, x_3)$ 是连续函数，则

$$f(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{i=1}^n X_i^2, \frac{1}{n}\sum_{i=1}^n X_i^3) \to f(E(X), E(X^2), E(X^3)).$$

样本矩：抢先版

- X_n 是独立同分布的随机变量序列，$\frac{1}{n}\sum_{i=1}^n X_n^k \to ?, \quad k \geqslant 1$
- 只要 $E(X_n^k)$ 存在，有辛欣大数定律，它收敛到 $E(X^k)$.
- $E(X^3)$ 存在，$f(x_1, x_2, x_3)$ 是连续函数，则

$$f(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{i=1}^n X_i^2, \frac{1}{n}\sum_{i=1}^n X_i^3) \to f(E(X), E(X^2), E(X^3)).$$

- 这是后面学习的矩法构造统计估计量的基础。

强大数定律

- (博雷尔强大数定律) 设 X_n 为独立同分布随机变量序列，且
$$P(X_n=1)=p,\quad P(X_n=0)=1-p,$$
则
$$\frac{1}{n}\sum_{k=1}^{n}X_k\to p,\ a.s.$$
- (柯尔莫哥洛夫强大数定律) 设 X_n 为独立同分布随机变量序列，且 $E(|X_n|)<\infty$，则
$$\frac{1}{n}\sum_{k=1}^{n}X_k\to E(X_1),\ a.s.$$

强大数定律

- (博雷尔强大数定律)设 X_n 为独立同分布随机变量序列，且

$$P(X_n = 1) = p, \quad P(X_n = 0) = 1 - p,$$

  则

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to p, \ a.s.$$

- (柯尔莫哥洛夫强大数定律)设 X_n 为独立同分布随机变量序列，且 $E(|X_n|) < \infty$，则

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to E(X_1), \ a.s.$$

  本页内容不做考试要求.

中心极限定理

- 中心极限定理考察随机变量和 $Y_n = X_1 + \cdots + X_n$ 的分布。

中心极限定理

- 中心极限定理考察随机变量和 $Y_n = X_1 + \cdots + X_n$ 的分布。
- 标准化：$Y_n^* = \frac{Y_n - E(Y_n)}{\sqrt{Var(Y_n)}}$.

- 中心极限定理考察随机变量和 $Y_n = X_1 + \cdots + X_n$ 的分布。
- 标准化：$Y_n^* = \frac{Y_n - E(Y_n)}{\sqrt{Var(Y_n)}}$.
- 林德伯格-莱维（Lindeberg-Levy）中心极限定理：设 $\{X_n\}$ 是独立同分布的随机变量序列，且 $E(X_n) = \mu$, $Var(X_n) = \sigma^2$ 均存在，若记

$$Y_n^* = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}},$$

则对任意实数 y, 有，

$$\lim_{n\to\infty} P(Y_n^* \leqslant y) = \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{t^2}{2}}\, dt.$$

Lindeberg-Levy 中心极限定理的证明

- 因为 $E(X_n-\mu)=0$, $Var(X_n-\mu)=\sigma^2$, 所以随机变量 $X_n-\mu$ 的特征函数可表示为

$$\varphi(t)=\varphi(0)+\varphi'(0)t+\frac{1}{2!}\varphi''(0)t^2+o(t^2)=1-\frac{\sigma^2t^2}{2}+o(t^2).$$

- 因为 $E(X_n-\mu)=0$, $Var(X_n-\mu)=\sigma^2$, 所以随机变量 $X_n-\mu$ 的特征函数可表示为

$$\varphi(t)=\varphi(0)+\varphi'(0)t+\frac{1}{2!}\varphi''(0)t^2+o(t^2)=1-\frac{\sigma^2t^2}{2}+o(t^2).$$

- Y_n^* 的特征函数为

$$\varphi_{Y_n^*}(t)=(\varphi(\frac{t}{\sigma\sqrt{n}}))^n=[1-\frac{t^2}{2n}+o(\frac{t^2}{n})]^n\to e^{-\frac{t^2}{2}}.$$

- $e^{-\frac{t^2}{2}}$ 为标准正态分布的特征函数，所以 Y_n^* 按分布收敛到一个标准正态分布随机变量。

例子

- X_n 是独立同分布的随机变量序列，$a = E(X_n^2)$, $b = E(X_n^4)$,

$$\frac{\sum_{i=1}^{n} X_n^2}{\sqrt{n}} \sim ?, \quad , n >> 1,$$

例子

- X_n 是独立同分布的随机变量序列，$a = E(X_n^2)$，$b = E(X_n^4)$，

$$\frac{\sum_{i=1}^{n} X_n^2}{\sqrt{n}} \sim ?, \quad , n >> 1,$$

- $Var(X_n^2) = b - a^2$,

例子

- X_n 是独立同分布的随机变量序列，$a = E(X_n^2)$，$b = E(X_n^4)$，

$$\frac{\sum_{i=1}^{n} X_n^2}{\sqrt{n}} \sim ?, \quad , n >> 1,$$

- $Var(X_n^2) = b - a^2$，所以

$$\frac{\sum_{i=1}^{n}(X_n^2 - a)}{\sqrt{b - a^2}\sqrt{n}} \to N(0,1).$$

例子

- X_n 是独立同分布的随机变量序列，$a = E(X_n^2)$, $b = E(X_n^4)$,

$$\frac{\sum_{i=1}^n X_n^2}{\sqrt{n}} \sim ?, \quad , n >> 1,$$

- $Var(X_n^2) = b - a^2$, 所以

$$\frac{\sum_{i=1}^n (X_n^2 - a)}{\sqrt{b - a^2}\sqrt{n}} \to N(0,1).$$

$$\frac{\sum_{i=1}^n X_n^2}{\sqrt{n}} \sim N(a\sqrt{n}, (b - a^2)).$$

独立不同分布下的中心极限定理

- 林德伯格中心极限定理：设 $\{X_n\}$ 为一个相互独立的随机变量序列，且它们具有有限的数学期望和方差：

$$E(X_n) = \mu_n, \quad Var(X_n) = \sigma_n^2, \quad n = 1, 2, \ldots.$$

令 $B_n = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$, 若对于任意的 $\tau > 0$，有

$$\lim_{n\to\infty} \frac{1}{\tau^2 B_n^2} \sum_{k=1}^{n} \int_{|x-\mu_i|>\tau B_n} (x-\mu_i)^2 p_i(x)\, dx = 0,$$

则对于任意 $x \in \mathbb{R}$, 有

$$\lim_{n\to\infty} P(\frac{1}{B_n} \sum_{i=1}^{n} (X_i - \mu_i) \leqslant x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

例子：林德伯格条件的验证

- 设 X_n 为独立同分布的随机变量序列，且数学期望和方差均存在，则

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-\mu_i|>\tau B_n}(x-\mu_k)^2 p_k(x)\,dx=\frac{n}{n\sigma^2}\int_{|x-\mu|>\tau\sqrt{n}\sigma}(x-\mu)^2 p(x)\,dx.$$

例子：林德伯格条件的验证

- 设 X_n 为独立同分布的随机变量序列，且数学期望和方差均存在，则

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-\mu_i|>\tau B_n}(x-\mu_k)^2 p_k(x)\,dx = \frac{n}{n\sigma^2}\int_{|x-\mu|>\tau\sqrt{n}\sigma}(x-\mu)^2 p(x)\,dx.$$

- 因为方差有限，即

$$Var(X_k) = \int_{-\infty}^{\infty}(x-\mu)^2 p(x)\,dx < \infty,$$

- 所以有

$$\lim_{n\to\infty}\int_{|x-\mu|>\tau\sigma\sqrt{n}}|x-\mu|^2 p(x)\,dx = 0.$$

独立不同分布下的中心极限定理

- (李雅普诺夫中心极限定理)设 $\{X_n\}$ 为独立随机变量序列，若存在 $\delta > 0$，满足

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

则对于任意 $x \in \mathbb{R}$，有

$$\lim_{n\to\infty} P(\frac{1}{B_n} \sum_{i=1}^{n} (X_i - \mu_i) \leqslant x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

- (李雅普诺夫中心极限定理) 设 $\{X_n\}$ 为独立随机变量序列，若存在 $\delta > 0$，满足

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

则对于任意 $x \in \mathbb{R}$，有

$$\lim_{n\to\infty} P\left(\frac{1}{B_n} \sum_{i=1}^{n} (X_i - \mu_i) \leqslant x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

  - 满足 Lindeberg 条件：

$$\begin{aligned} E(|X_k - \mu_k|^{2+\delta}) &= \int_{-\infty}^{\infty} |x-\mu_k|^{2+\delta} p_k(x)\, dx \\ &\geqslant \int_{\{x:|x-\mu_k|\geqslant \tau B_n} |x-\mu_k|^{2+\delta} p_k(x)\, dx \\ &\geqslant \tau^\delta B_n^\delta \int_{\{x:|x-\mu_k|\geqslant \tau B_n\}} |x-\mu_k|^{2} p_k(x)\, dx. \end{aligned}$$

中心极限定理的应用

- 误差分析：在计算机中，所有实数 x 都是用由一定位数的小数 x' 来近似表示。这个近似过程产生误差 $\epsilon = x - x'$.

中心极限定理的应用

- 误差分析：在计算机中，所有实数 x 都是用由一定位数的小数 x' 来近似表示。这个近似过程产生误差 $\epsilon = x - x'$.一般假设

$$\epsilon \sim U(-0.5 \times 10^{-k}, 0.5 \times 10^{-k}).$$

中心极限定理的应用

- 误差分析：在计算机中，所有实数 x 都是用由一定位数的小数 x' 来近似表示。这个近似过程产生误差 $\epsilon = x - x'$.一般假设

$$\epsilon \sim U(-0.5 \times 10^{-k}, 0.5 \times 10^{-k}).$$

- n 个实数 x_i 的和 S, 及其近似 S'，则

$$S - S' = \sum_{i=1}^{n} x_i - x_i' = \sum_{i=1}^{n} \epsilon_i.$$

中心极限定理的应用

- 误差分析：在计算机中，所有实数 x 都是用由一定位数的小数 x' 来近似表示。这个近似过程产生误差 $\epsilon = x - x'$.一般假设

$$\epsilon \sim U(-0.5 \times 10^{-k}, 0.5 \times 10^{-k}).$$

- n 个实数 x_i 的和 S, 及其近似 S'，则

$$S - S' = \sum_{i=1}^{n} x_i - x_i' = \sum_{i=1}^{n} \epsilon_i.$$

- 当 n 足够大时，近似地有 $\frac{\sum_{i=1}^{n} \epsilon_i}{\sqrt{n\frac{10^{-2k}}{12}}} \sim N(0,1)$. 则

$$P(|\sum_{i=1}^{n} \epsilon_i| \leqslant z) \approx 2\Phi(\frac{z\sqrt{12}}{\sqrt{n10^{-2k}}}) - 1.$$

中心极限定理的应用

- 二项分布的正态近似——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 $(0 < p < 1)$，记 S_n 为 n 次试验中事件 A 出现的次数，则

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant y\right) = \Phi(y) = \frac{1}{2\pi}\int_{-\infty}^{y} e^{-t^2/2}\,dt.$$

中心极限定理的应用

- 二项分布的正态近似——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 $(0 < p < 1)$，记 S_n 为 n 次试验中事件 A 出现的次数，则

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant y\right) = \Phi(y) = \frac{1}{2\pi}\int_{-\infty}^{y} e^{-t^2/2}\,dt.$$

- p 比较小时，用泊松分布近似比较合理。

中心极限定理的应用

- 二项分布的正态近似——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 $(0 < p < 1)$，记 S_n 为 n 次试验中事件 A 出现的次数，则

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant y\right) = \Phi(y) = \frac{1}{2\pi}\int_{-\infty}^{y} e^{-t^2/2}dt.$$

- p 比较小时，用泊松分布近似比较合理。$np > 5$ 且 $n(1-p) > 5$ 时用正态分布近似比较合理。

中心极限定理的应用

- 二项分布的正态近似——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 $(0 < p < 1)$，记 S_n 为 n 次试验中事件 A 出现的次数，则

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant y\right) = \Phi(y) = \frac{1}{2\pi}\int_{-\infty}^{y} e^{-t^2/2}\,dt.$$

- p 比较小时，用泊松分布近似比较合理。$np > 5$ 且 $n(1-p) > 5$ 时用正态分布近似比较合理。
- 修正近似：若 $a < b$ 均为正整数，则

$$P(a \leqslant S_n \leqslant b) = P(a - 0.5 \leqslant S_n \leqslant b + 0.5).$$

中心极限定理的应用

- 二项分布的正态近似——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 （$0 < p < 1$)，记 S_n 为 n 次试验中事件 A 出现的次数，则

$$\lim_{n\to\infty} P(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant y) = \Phi(y) = \frac{1}{2\pi}\int_{-\infty}^{y} e^{-t^2/2}dt.$$

- p 比较小时，用泊松分布近似比较合理。$np > 5$ 且 $n(1-p) > 5$ 时用正态分布近似比较合理。
- 修正近似：若 $a < b$ 均为正整数，则

$$P(a \leqslant S_n \leqslant b) = P(a - 0.5 \leqslant S_n \leqslant b + 0.5).$$

当 n 相对比较小时，一般更好的近似为

$$P(a \leqslant S_n \leqslant b) \approx \Phi(\frac{b + 0.5 - np}{\sqrt{np(1-p)}}) - \Phi(\frac{a - 0.5 - np}{\sqrt{np(1-p)}}).$$

二项分布的正态近似

- $P(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant y) \approx \Phi(y) = \beta.$

二项分布的正态近似

- $P(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant y) \approx \Phi(y) = \beta.$
- 给定 n, y, 求 β: 一个复杂系统由 100 个相互独立的部件组成，每个部件正常工作的概率为 0.9. 系统要正常工作，至少需要 85 个部件正常工作，那系统正常工作的概率大约是?

二项分布的正态近似

- $P(\frac{S_n-np}{\sqrt{np(1-p)}} \leqslant y) \approx \Phi(y) = \beta.$
- 给定 n, y, 求 β: 一个复杂系统由 100 个相互独立的部件组成，每个部件正常工作的概率为 0.9. 系统要正常工作，至少需要 85 个部件正常工作，那系统正常工作的概率大约是?

$$P(Y_n \geqslant 85) \approx 1 - \Phi(\frac{85-0.5-90}{3}) = \Phi(5.5/3) \approx= 0.9664.$$

二项分布的正态近似

- $P(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant y) \approx \Phi(y) = \beta.$
- 给定 n, y, 求 β: 一个复杂系统由 100 个相互独立的部件组成，每个部件正常工作的概率为 0.9. 系统要正常工作，至少需要 85 个部件正常工作，那系统正常工作的概率大约是?

$$P(Y_n \geqslant 85) \approx 1 - \Phi(\frac{85 - 0.5 - 90}{3}) = \Phi(5.5/3) \approx= 0.9664.$$

- 给定 n,β, 求 y: 某工厂有 200 台机器，在一个小时内，每台机器有 70% 的时间在工作，工作时消耗 15KW 电能，问要多少电能才能以 95% 的可能性保证工厂正常工作呢?

- $P(\frac{S_n-np}{\sqrt{np(1-p)}} \leqslant y) \approx \Phi(y) = \beta$.
- 给定 n, y, 求 β: 一个复杂系统由 100 个相互独立的部件组成，每个部件正常工作的概率为 0.9. 系统要正常工作，至少需要 85 个部件正常工作，那系统正常工作的概率大约是?

$$P(Y_n \geqslant 85) \approx 1 - \Phi(\frac{85-0.5-90}{3}) = \Phi(5.5/3) \approx= 0.9664.$$

- 给定 n,β, 求 y: 某工厂有 200 台机器，在一个小时内，每台机器有 70% 的时间在工作，工作时消耗 15KW 电能，问要多少电能才能以 95% 的可能性保证工厂正常工作呢? Y 为工作的机器数，$Y \sim b(200, 0.7)$, $E(Y) = 140$, $Var(Y) = 42$.

二项分布的正态近似

- $P(\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant y) \approx \Phi(y) = \beta.$
- 给定 n, y, 求 β: 一个复杂系统由 100 个相互独立的部件组成，每个部件正常工作的概率为 0.9. 系统要正常工作，至少需要 85 个部件正常工作，那系统正常工作的概率大约是?

$$P(Y_n \geqslant 85) \approx 1 - \Phi(\frac{85 - 0.5 - 90}{3}) = \Phi(5.5/3) \approx= 0.9664.$$

- 给定 n, β, 求 y: 某工厂有 200 台机器，在一个小时内，每台机器有 70% 的时间在工作，工作时消耗 15KW 电能，问要多少电能才能以 95% 的可能性保证工厂正常工作呢? Y 为工作的机器数，$Y \sim b(200, 0.7)$, $E(Y) = 140$, $Var(Y) = 42$.则要求

$$P(15Y_n \leqslant y) \approx \Phi(\frac{y/15 + 0.5 - 140}{\sqrt{42}}) \geqslant 0.95.$$

二项分布的正态近似

- $P(\frac{S_n-np}{\sqrt{np(1-p)}} \leqslant y) \approx \Phi(y) = \beta.$
- 给定 n, y, 求 β: 一个复杂系统由 100 个相互独立的部件组成，每个部件正常工作的概率为 0.9. 系统要正常工作，至少需要 85 个部件正常工作，那系统正常工作的概率大约是?

$$P(Y_n \geqslant 85) \approx 1 - \Phi(\frac{85 - 0.5 - 90}{3}) = \Phi(5.5/3) \approx = 0.9664.$$

- 给定 n,β, 求 y: 某工厂有 200 台机器，在一个小时内，每台机器有 70% 的时间在工作，工作时消耗 15KW 电能，问要多少电能才能以 95% 的可能性保证工厂正常工作呢? Y 为工作的机器数，$Y \sim b(200, 0.7)$, $E(Y) = 140$, $Var(Y) = 42$.则要求

$$P(15Y_n \leqslant y) \approx \Phi(\frac{y/15 + 0.5 - 140}{\sqrt{42}}) \geqslant 0.95.$$

$$\frac{y/15 + 0.5 - 140}{\sqrt{42}} \geqslant 1.645, \quad \Rightarrow y \geqslant 2252.$$

- 给定 y, β, 求 n:先要调查中央电视台新闻联播在清华大学的收视率 p, 调查人员准备将所有受访对象中观看该节目的人员频率 $\hat{p}$ 来估计 p, 现在要以 90% 的把握认为 $\hat{p}$ 和真实的 p 的差异不超过 5%。需要访问多少人?

二项分布的正态近似

- 给定 y, β, 求 n:先要调查中央电视台新闻联播在清华大学的收视率 p, 调查人员准备将所有受访对象中观看该节目的人员频率 $\hat{p}$ 来估计 p, 现在要以 90% 的把握认为 $\hat{p}$ 和真实的 p 的差异不超过 5%。需要访问多少人?
- Y 为访问对象中收看新闻联播的人数, 则 $Y \sim b(n, p)$.

二项分布的正态近似

- 给定 y, β, 求 n:先要调查中央电视台新闻联播在清华大学的收视率 p, 调查人员准备将所有受访对象中观看该节目的人员频率 $\hat{p}$ 来估计 p, 现在要以 90% 的把握认为 $\hat{p}$ 和真实的 p 的差异不超过 5%。需要访问多少人?
- Y 为访问对象中收看新闻联播的人数, 则 $Y \sim b(n,p)$.则要求为

$$P(|\frac{Y}{n} - p| \leqslant 0.05) \geqslant 0.9,$$

- 给定 y, β, 求 n:先要调查中央电视台新闻联播在清华大学的收视率 p, 调查人员准备将所有受访对象中观看该节目的人员频率 $\hat{p}$ 来估计 p, 现在要以 90% 的把握认为 $\hat{p}$ 和真实的 p 的差异不超过 5%。需要访问多少人?
- Y 为访问对象中收看新闻联播的人数, 则 $Y \sim b(n,p)$.则要求为

$$P(|\frac{Y}{n} - p| \leqslant 0.05) \geqslant 0.9,$$

$$P(|\frac{Y}{n} - p| \leqslant 0.05) = P(|\frac{(Y - np)}{\sqrt{np(1-p)}}| \leqslant \frac{0.05n}{\sqrt{np(1-p)}})$$
$$\approx 2\Phi(0.05\sqrt{\frac{n}{p(1-p)}}) - 1 \geqslant 0.95 \Rightarrow 0.05\sqrt{\frac{n}{p(1-p)}} \geqslant 1.645.$$

二项分布的正态近似

- 给定 y, β, 求 n:先要调查中央电视台新闻联播在清华大学的收视率 p, 调查人员准备将所有受访对象中观看该节目的人员频率 $\hat{p}$ 来估计 p, 现在要以 90% 的把握认为 $\hat{p}$ 和真实的 p 的差异不超过 5%。需要访问多少人?
- Y 为访问对象中收看新闻联播的人数, 则 $Y \sim b(n,p)$.则要求为

$$P(|\frac{Y}{n} - p| \leqslant 0.05) \geqslant 0.9,$$

$$P(|\frac{Y}{n} - p| \leqslant 0.05) = P(|\frac{(Y - np)}{\sqrt{np(1-p)}}| \leqslant \frac{0.05n}{\sqrt{np(1-p)}})$$

$$\approx 2\Phi(0.05\sqrt{\frac{n}{p(1-p)}}) - 1 \geqslant 0.95 \Rightarrow 0.05\sqrt{\frac{n}{p(1-p)}} \geqslant 1.645.$$

$$n \geqslant p(1-p)\frac{1.645^2}{0.05^2}, \Rightarrow n \geqslant 0.25 \times 1082.41 \approx 271.$$

- (Berry-Esseen 定理) 考虑独立同分布随机变量序列 $\{X_n\}$, 假设 $E(X_n)=0$, $Var(X_n)=\sigma^2<\infty$, 其 $E(|X_1|^3)<\infty$. 令

$$S_n=\frac{X_1+\cdots+X_n}{\sigma\sqrt{n}},$$

其分布函数为 $F_n(x)$, 则有

$$\sup_x|F_n(x)-\Phi(x)|\leqslant C\frac{E|X_1|^3}{\sigma^3\sqrt{n}},$$

而 $\frac{1}{\sqrt{2\pi}}\leqslant C\leqslant 0.8$.

- Berry-Esseen 定理的结论一般情况下不能被加强：考虑独立同分布随机变量序列 $\{X_n\}$,

$$P(X_n = 1) = 0.5, \quad P(X_n = -1) = 0.5.$$

中心极限定的逼近速度

- Berry-Esseen 定理的结论一般情况下不能被加强：考虑独立同分布随机变量序列 $\{X_n\}$,

$$P(X_n = 1) = 0.5, \quad P(X_n = -1) = 0.5.$$

则对于任意 n,

$$2P(\sum_{k=1}^{2n} X_k < 0) + P(\sum_{k=1}^{2n} X_k = 0) = 1.$$

中心极限定的逼近速度

- Berry-Esseen 定理的结论一般情况下不能被加强：考虑独立同分布随机变量序列 $\{X_n\}$,

$$P(X_n = 1) = 0.5, \quad P(X_n = -1) = 0.5.$$

则对于任意 n,

$$2P(\sum_{k=1}^{2n} X_k < 0) + P(\sum_{k=1}^{2n} X_k = 0) = 1.$$

所以

$$|P(\sum_{k=1}^{2n} X_k < 0) - 0.5| = \frac{1}{2}P(\sum_{k=1}^{2n} X_k = 0) = \frac{1}{2}C_{2n}^{n}2^{-2n} \sim \frac{1}{\sqrt{2\pi 2n}}.$$

中心极限定的逼近速度

- Berry-Esseen 定理的结论一般情况下不能被加强：考虑独立同分布随机变量序列 $\{X_n\}$,

$$P(X_n = 1) = 0.5, \quad P(X_n = -1) = 0.5.$$

则对于任意 n,

$$2P(\sum_{k=1}^{2n} X_k < 0) + P(\sum_{k=1}^{2n} X_k = 0) = 1.$$

所以

$$|P(\sum_{k=1}^{2n} X_k < 0) - 0.5| = \frac{1}{2}P(\sum_{k=1}^{2n} X_k = 0) = \frac{1}{2}C_{2n}^{n}2^{-2n} \sim \frac{1}{\sqrt{2\pi 2n}}.$$

Stirling 公式：$n! \sim \sqrt{2\pi n}e^{-n}n^n$.

中心极限定的逼近速度

- Berry-Esseen 定理的结论一般情况下不能被加强：考虑独立同分布随机变量序列 $\{X_n\}$,

$$P(X_n = 1) = 0.5, \quad P(X_n = -1) = 0.5.$$

则对于任意 n,

$$2P(\sum_{k=1}^{2n} X_k < 0) + P(\sum_{k=1}^{2n} X_k = 0) = 1.$$

所以

$$|P(\sum_{k=1}^{2n} X_k < 0) - 0.5| = \frac{1}{2} P(\sum_{k=1}^{2n} X_k = 0) = \frac{1}{2} C_{2n}^{n} 2^{-2n} \sim \frac{1}{\sqrt{2\pi 2n}}.$$

Stirling 公式：$n! \sim \sqrt{2\pi n} e^{-n} n^n$. $C_{2n}^{n} = \frac{(2n)!}{(n!)^2} \sim 2^{2n} \frac{1}{\sqrt{\pi n}}$.

概率论与数理统计（8）

清华大学

2020 年春季学期

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。
- 几个时髦的名词：

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。
- 几个时髦的名词：
  - 数据挖掘 (Data Mining);

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。
- 几个时髦的名词：
  - 数据挖掘 (Data Mining)；
  - 机器学习（Machine learning)，深度学习（deep learning);

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。
- 几个时髦的名词:
  - 数据挖掘 (Data Mining);
  - 机器学习（Machine learning), 深度学习（deep learning);
  - 大数据科学（Big Data);

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。
- 几个时髦的名词:
  - 数据挖掘 (Data Mining);
  - 机器学习（Machine learning), 深度学习（deep learning);
  - 大数据科学（Big Data);
- 这门课只涉及基本概念和方法。只涉及数据分析，不涉及数据收集。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。
- 个体是单个研究对象，如说某某的身高，成绩之类的。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。
- 个体是单个研究对象，如说某某的身高，成绩之类的。
- 我们想了解总体的性质，如数学期望，方差，某些特殊性质所占的比例等等。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。
- 个体是单个研究对象，如说某某的身高，成绩之类的。
- 我们想了解总体的性质，如数学期望，方差，某些特殊性质所占的比例等等。
- 可以把所有的个体都研究一遍。没有任何遗漏，错误。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。
- 个体是单个研究对象，如说某某的身高，成绩之类的。
- 我们想了解总体的性质，如数学期望，方差，某些特殊性质所占的比例等等。
- 可以把所有的个体都研究一遍。没有任何遗漏，错误。当总体的数量太大时，不大实际。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。
- 个体是单个研究对象，如说某某的身高，成绩之类的。
- 我们想了解总体的性质，如数学期望，方差，某些特殊性质所占的比例等等。
- 可以把所有的个体都研究一遍。没有任何遗漏，错误。当总体的数量太大时，不大实际。
- 随机抽样方法。

简单随机抽样

- 随机地从总体中抽出 n 个个体，记为 $x_1, \ldots, x_n$。它们就是总体的一个样本，n 称为样本容量，样本中的个体称为样品。

简单随机抽样

- 随机地从总体中抽出 n 个个体，记为 $x_1, \ldots, x_n$。它们就是总体的一个样本，n 称为样本容量，样本中的个体称为样品。
- 当然样本容量越大越接近总体。

简单随机抽样

- 随机地从总体中抽出 n 个个体，记为 $x_1, \ldots, x_n$。它们就是总体的一个样本，n 称为样本容量，样本中的个体称为样品。
- 当然样本容量越大越接近总体。
- 样本要具有随机性：每个样品的分布应该与总体的相同。

简单随机抽样

- 随机地从总体中抽出 n 个个体，记为 $x_1, \ldots, x_n$。它们就是总体的一个样本，n 称为样本容量，样本中的个体称为样品。
- 当然样本容量越大越接近总体。
- 样本要具有随机性：每个样品的分布应该与总体的相同。
- 样本要有独立性：即 $x_1, \ldots, x_n$ 相互独立。

简单随机抽样

- 随机地从总体中抽出 n 个个体，记为 $x_1, \ldots, x_n$。它们就是总体的一个样本，n 称为样本容量，样本中的个体称为样品。
- 当然样本容量越大越接近总体。
- 样本要具有随机性：每个样品的分布应该与总体的相同。
- 样本要有独立性：即 $x_1, \ldots, x_n$ 相互独立。若总体的分布函数为 $F(x)$，则样本容量为 n 的样本联合分布函数为

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F(x_i).$$

- 经验分布函数：设 $x_1,\ldots,x_n$ 是取自总体分布函数为 $F(x)$ 的样本，如果将样本观测值由小到大排列，为 $x_{(1)},\ldots,x_{(n)}$，则称其为有序样本，用有序样本定义的函数

$$F_n(x)=\begin{cases}0, & x<x_{(1)}, \\ \frac{k}{n}, & x_k\leqslant x<x_{(k+1)}, \quad k=1,\ldots,n-1, \\ 1, & x\geqslant x_{(n)},\end{cases}$$

为经验分布函数。

- 经验分布函数：设 $x_1, \ldots, x_n$ 是取自总体分布函数为 $F(x)$ 的样本，如果将样本观测值由小到大排列，为 $x_{(1)}, \ldots, x_{(n)}$，则称其为有序样本，用有序样本定义的函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_k \leqslant x < x_{(k+1)}, \quad k = 1, \ldots, n-1, \\ 1, & x \geqslant x_{(n)}, \end{cases}$$

为经验分布函数。

- 若定义 $I_i(x) = \begin{cases} 1, & x_i \leqslant x, \\ 0, & x_i > x, \end{cases}$,

- 经验分布函数：设 $x_1, \ldots, x_n$ 是取自总体分布函数为 $F(x)$ 的样本，如果将样本观测值由小到大排列，为 $x_{(1)}, \ldots, x_{(n)}$，则称其为有序样本，用有序样本定义的函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_k \leqslant x < x_{(k+1)}, \quad k = 1, \ldots, n-1, \\ 1, & x \geqslant x_{(n)}, \end{cases}$$

为经验分布函数。

- 若定义 $I_i(x) = \begin{cases} 1, & x_i \leqslant x, \\ 0, & x_i > x, \end{cases}$，则 $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_i(x)$

- 经验分布函数：设 $x_1, \ldots, x_n$ 是取自总体分布函数为 $F(x)$ 的样本，如果将样本观测值由小到大排列，为 $x_{(1)}, \ldots, x_{(n)}$，则称其为有序样本，用有序样本定义的函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_k \leqslant x < x_{(k+1)}, \quad k = 1, \ldots, n-1, \\ 1, & x \geqslant x_{(n)}, \end{cases}$$

为经验分布函数。

- 若定义 $I_i(x) = \begin{cases} 1, & x_i \leqslant x, \\ 0, & x_i > x, \end{cases}$，则 $F_n(x) = \frac{1}{n}\sum_{i=1}^n I_i(x) \to F(x)$.

- 经验分布函数：设 $x_1, \ldots, x_n$ 是取自总体分布函数为 $F(x)$ 的样本，如果将样本观测值由小到大排列，为 $x_{(1)}, \ldots, x_{(n)}$，则称其为有序样本，用有序样本定义的函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_k \leqslant x < x_{(k+1)}, \quad k = 1, \ldots, n-1, \\ 1, & x \geqslant x_{(n)}, \end{cases}$$

为经验分布函数。

- 若定义 $I_i(x) = \begin{cases} 1, & x_i \leqslant x, \\ 0, & x_i > x, \end{cases}$，则 $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_i(x) \to F(x)$.
- (格里纹科定理)：$P(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0) = 1$.

样本数据的整理和显示

- 频数频率表
- 直方图
- 茎叶图

样本数据的整理和显示

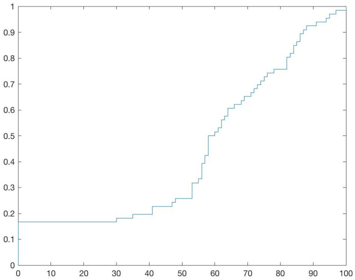
- 频数频率表
- 直方图
- 茎叶图

```
>> stemleafplot(V)
  0 | 0 0 0 0 0 0 0 0 0 0 0
  1 |
  2 |
  3 | 0 5
  4 | 1 1 7 8
  5 | 3 3 3 3 5 6 6 6 6 7 7 8 8 8 8 8
  6 | 0 1 2 2 3 4 4 6 8
  7 | 1 2 3 4 5 6 8
  8 | 2 2 2 3 4 4 5 6 6 7 8
  9 | 1 4 5 7
 10 | 0
```

样本数据的整理和显示

- 频数频率表
- 直方图
- 茎叶图

```
>> stemleafplot(V)
  0 | 0 0 0 0 0 0 0 0 0 0 0
  1 |
  2 |
  3 | 0 5
  4 | 1 1 7 8
  5 | 3 3 3 3 5 6 6 6 6 7 7 8 8
  6 | 0 1 2 2 3 4 4 6 8
  7 | 1 2 3 4 5 6 8
  8 | 2 2 2 3 4 4 5 6 6 7 8
  9 | 1 4 5 7
 10 | 0
```



统计量及其分布

- 设 $x_1, \ldots, x_n$ 为某总体的样本，若样本函数 $T = T(x_1, \ldots, x_n)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布为抽样分布。

统计量及其分布

- 设 $x_1, \ldots, x_n$ 为某总体的样本，若样本函数 $T = T(x_1, \ldots, x_n)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布为抽样分布。
- $\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_i$.

统计量及其分布

- 设 $x_1, \ldots, x_n$ 为某总体的样本，若样本函数 $T = T(x_1, \ldots, x_n)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布为抽样分布。
- $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_i$.
- $\frac{1}{n}\sum_{k=1}^{n}(x_i - \mu)^2$，其中 μ 是总体分布的数学期望 (已知)。

- 设 $x_1, \ldots, x_n$ 为某总体的样本，若样本函数 $T = T(x_1, \ldots, x_n)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布为抽样分布。
- $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_i$.
- $\frac{1}{n}\sum_{k=1}^{n}(x_i - \mu)^2$，其中 μ 是总体分布的数学期望 (已知)。
- $\frac{1}{n}\sum_{k=1}^{n}(x_i - \bar{x})^2$.

统计量及其分布

- 设 $x_1, \ldots, x_n$ 为某总体的样本，若样本函数 $T = T(x_1, \ldots, x_n)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布为抽样分布。
- $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_i$.
- $\frac{1}{n}\sum_{k=1}^{n}(x_i - \mu)^2$，其中 μ 是总体分布的数学期望 (已知)。
- $\frac{1}{n}\sum_{k=1}^{n}(x_i - \bar{x})^2$.
- $\frac{1}{n-1}\sum_{k=1}^{n}(x - \bar{x})^2$

样本均值

- 设 $x_1, \ldots, x_n$ 为某总体的样本，则其算数平均称为样本均值，一般用 $\bar{x}$ 表示，即 $\bar{x} = \frac{x_1 + \cdots + x_n}{n}$.

样本均值

- 设 $x_1, \ldots, x_n$ 为某总体的样本，则其算数平均称为样本均值，一般用 $\bar{x}$ 表示，即 $\bar{x} = \frac{x_1+\cdots+x_n}{n}$.
- 若把样本中的数据与样本均值之差称为偏差，则样本所有偏差之和为 0，即 $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.

样本均值

- 设 $x_1,\ldots,x_n$ 为某总体的样本，则其算数平均称为样本均值，一般用 $\bar{x}$ 表示，即 $\bar{x}=\frac{x_1+\cdots+x_n}{n}$.
- 若把样本中的数据与样本均值之差称为偏差，则样本所有偏差之和为 0，即 $\sum_{i=1}^{n}(x_i-\bar{x})=0$.

$$\sum_{i=1}^{n}(x_i-\bar{x})=\sum_{i=1}^{n}x_i-n\bar{x}=0.$$

样本均值

- 设 $x_1, \ldots, x_n$ 为某总体的样本，则其算数平均称为样本均值，一般用 $\bar{x}$ 表示，即 $\bar{x} = \frac{x_1+\cdots+x_n}{n}$.
- 若把样本中的数据与样本均值之差称为偏差，则样本所有偏差之和为 0，即 $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = 0.$$

- 数据观测值与均值的偏差平方和最小，即若考虑 $f(c) = \sum_{i=1}^{n}(x_i - c)^2$，则 $c = \bar{x}$ 取到最小值。

样本均值

- 设 $x_1, \ldots, x_n$ 为某总体的样本，则其算数平均称为样本均值，一般用 $\bar{x}$ 表示，即 $\bar{x} = \frac{x_1+\cdots+x_n}{n}$.
- 若把样本中的数据与样本均值之差称为偏差，则样本所有偏差之和为 0，即 $\sum_{i=1}^n (x_i - \bar{x}) = 0$.

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = 0.$$

- 数据观测值与均值的偏差平方和最小，即若考虑 $f(c) = \sum_{i=1}^n (x_i - c)^2$，则 $c = \bar{x}$ 取到最小值。

$$\begin{aligned}\sum_{i=1}^n (x_i - c)^2 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - c)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - c)^2 + 2\sum_{i=1}^n (x_i - \bar{x})(\bar{x} - c)\end{aligned}$$

样本均值的抽样分布

- 若总体分布为 $N(\mu,\sigma^2)$, 则 $\bar{x}\sim N(\mu,\sigma^2/n)$.

样本均值的抽样分布

- 若总体分布为 $N(\mu, \sigma^2)$, 则 $\bar{x} \sim N(\mu, \sigma^2/n)$.

$$x_1 + \cdots + x_n \sim N(n\mu, n\sigma^2), \quad \bar{x} \sim N(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}).$$

样本均值的抽样分布

- 若总体分布为 $N(\mu, \sigma^2)$, 则 $\bar{x} \sim N(\mu, \sigma^2/n)$.

$$x_1 + \cdots + x_n \sim N(n\mu, n\sigma^2), \quad \bar{x} \sim N(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}).$$

- 对于一般的分布，若 $E(x) = \mu$, $Var(x) = \sigma^2$, 则 $E(\bar{x}) = \mu$, $Var(\bar{x}) = \frac{\sigma^2}{n}$.

样本均值的抽样分布

- 若总体分布为 $N(\mu, \sigma^2)$, 则 $\bar{x} \sim N(\mu, \sigma^2/n)$.

$$x_1 + \cdots + x_n \sim N(n\mu, n\sigma^2), \quad \bar{x} \sim N(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}).$$

- 对于一般的分布，若 $E(x) = \mu$, $Var(x) = \sigma^2$, 则 $E(\bar{x}) = \mu$, $Var(\bar{x}) = \frac{\sigma^2}{n}$. 当 n 充分大时，近似地 $\bar{x} \sim N(\mu, \sigma^2/n)$..

样本均值的抽样分布

- 若总体分布为 $N(\mu,\sigma^2)$, 则 $\bar{x}\sim N(\mu,\sigma^2/n)$.

$$x_1+\cdots+x_n\sim N(n\mu,n\sigma^2),\quad \bar{x}\sim N(\frac{n\mu}{n},\frac{n\sigma^2}{n^2}).$$

- 对于一般的分布，若 $E(x)=\mu$, $Var(x)=\sigma^2$, 则 $E(\bar{x})=\mu$, $Var(\bar{x})=\frac{\sigma^2}{n}$. 当 n 充分大时，近似地 $\bar{x}\sim N(\mu,\sigma^2/n)$..

$$\frac{\sum_{i=1}^n(x_i-\mu)}{\sqrt{n}\sigma}\tilde{\sim}N(0,1),\quad \bar{x}=\frac{\sigma}{\sqrt{n}}\frac{\sum_{i=1}^n(x_i-\mu)}{\sqrt{n}\sigma}+\mu\tilde{\sim}N(\mu,\frac{\sigma^2}{n}).$$

样本均值的抽样分布

- 若总体分布为 $N(\mu,\sigma^2)$, 则 $\bar{x}\sim N(\mu,\sigma^2/n)$.

$$x_1+\cdots+x_n\sim N(n\mu,n\sigma^2),\quad \bar{x}\sim N(\frac{n\mu}{n},\frac{n\sigma^2}{n^2}).$$

- 对于一般的分布，若 $E(x)=\mu$, $Var(x)=\sigma^2$, 则 $E(\bar{x})=\mu$, $Var(\bar{x})=\frac{\sigma^2}{n}$. 当 n 充分大时，近似地 $\bar{x}\sim N(\mu,\sigma^2/n)$..

$$\frac{\sum_{i=1}^n(x_i-\mu)}{\sqrt{n}\sigma}\tilde{\sim}N(0,1),\quad \bar{x}=\frac{\sigma}{\sqrt{n}}\frac{\sum_{i=1}^n(x_i-\mu)}{\sqrt{n}\sigma}+\mu\tilde{\sim}N(\mu,\frac{\sigma^2}{n}).$$

- 我们学过的分布中，还有哪些是可以精确知道 $\bar{x}$ 的分布的?

样本均值的抽样分布

- 若总体分布为 $N(\mu,\sigma^2)$, 则 $\bar{x}\sim N(\mu,\sigma^2/n)$.

$$x_1+\cdots+x_n\sim N(n\mu,n\sigma^2),\quad \bar{x}\sim N(\frac{n\mu}{n},\frac{n\sigma^2}{n^2}).$$

- 对于一般的分布，若 $E(x)=\mu$, $Var(x)=\sigma^2$, 则 $E(\bar{x})=\mu$, $Var(\bar{x})=\frac{\sigma^2}{n}$. 当 n 充分大时，近似地 $\bar{x}\sim N(\mu,\sigma^2/n)$..

$$\frac{\sum_{i=1}^n(x_i-\mu)}{\sqrt{n}\sigma}\tilde{\sim}N(0,1),\quad \bar{x}=\frac{\sigma}{\sqrt{n}}\frac{\sum_{i=1}^n(x_i-\mu)}{\sqrt{n}\sigma}+\mu\tilde{\sim}N(\mu,\frac{\sigma^2}{n}).$$

- 我们学过的分布中，还有哪些是可以精确知道 $\bar{x}$ 的分布的？指数分布，伽玛分布，$\chi^2(n)$ 分布。(具有独立可加性)。

样本方差与样本标准差

- 设 $x_1, \ldots, x_n$ 为取自某总体的样本，则它关于均值平均偏差和为

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

称为样本方差，其算数根 $s_n = \sqrt{s_n^2}$ 称为样本标准差。

样本方差与样本标准差

- 设 $x_1, \dots, x_n$ 为取自某总体的样本，则它关于均值平均偏差和为

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

称为样本方差，其算数根 $s_n = \sqrt{s_n^2}$ 称为样本标准差。当 n 不大时，一般会用

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

s^2 为无偏样本方差。对应的 $s = \sqrt{s^2}$ 代替 s_n.

样本方差

- 设总体 X 有 $E(X)=\mu$, $Var(X)=\sigma^2<\infty$, $x_1,\ldots,x_n$ 为从该总体取出的样本, 则

$$E(\bar{x})=\mu,\quad Var(\bar{x})=\sigma^2/n,\quad E(s^2)=\sigma^2.$$

样本方差

- 设总体 X 有 $E(X)=\mu$, $Var(X)=\sigma^2<\infty$, $x_1,\ldots,x_n$ 为从该总体取出的样本, 则

$$E(\bar{x})=\mu,\quad Var(\bar{x})=\sigma^2/n,\quad E(s^2)=\sigma^2.$$

-

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}(x_i^2-2x_i\bar{x}+\bar{x}^2)=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$$
$$=n(\sigma^2+\mu^2)+n(\mu^2-\frac{\sigma^2}{n})=(n-1)\sigma^2.$$

样本方差

- 设总体 X 有 $E(X)=\mu$, $Var(X)=\sigma^2<\infty$, $x_1,\dots,x_n$ 为从该总体取出的样本，则

$$E(\bar{x})=\mu, \quad Var(\bar{x})=\sigma^2/n, \quad E(s^2)=\sigma^2.$$

-

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}(x_i^2-2x_i\bar{x}+\bar{x}^2)=\sum_{i=1}^{n}x_i^2-n\bar{\bar{x}}^2$$
$$=n(\sigma^2+\mu^2)+n(\mu^2-\frac{\sigma^2}{n})=(n-1)\sigma^2.$$

$$E(s_n^2)=\frac{n-1}{n}\sigma^2, \quad E(s^2)=\sigma^2.$$

- 设 $x_1, \ldots, x_n$ 是取自总体 X 的样本，$x_{(i)}$ 称为该样本的第 i 个次序统计量，它的取值是将样本观测值有小到大排列后的第 i 个观测值，其中 $x_{(1)} = \min\{x_1, \ldots, x_n\}$ 为该样本的最小次序统计量，$x_{(n)} = \max\{x_1, \ldots, x_n\}$ 为该样本的最大次序统计量。$(x_{(1)}, \ldots, x_{(n)})$ 为该样本的次序统计量。

次序统计量及其抽样分布

- 设 $x_1, \ldots, x_n$ 是取自总体 X 的样本，$x_{(i)}$ 称为该样本的第 i 个次序统计量，它的取值是将样本观测值有小到大排列后的第 i 个观测值，其中 $x_{(1)} = \min\{x_1, \ldots, x_n\}$ 为该样本的最小次序统计量，$x_{(n)} = \max\{x_1, \ldots, x_n\}$ 为该样本的最大次序统计量。$(x_{(1)}, \ldots, x_{(n)})$ 为该样本的次序统计量。
- 一般情况下，简单样本的样品是独立同分布的，而次序统计量既不独立也不同分布。

次序统计量及其抽样分布

- 设 $x_1, \ldots, x_n$ 是取自总体 X 的样本，$x_{(i)}$ 称为该样本的第 i 个次序统计量，它的取值是将样本观测值有小到大排列后的第 i 个观测值，其中 $x_{(1)} = \min\{x_1, \ldots, x_n\}$ 为该样本的最小次序统计量，$x_{(n)} = \max\{x_1, \ldots, x_n\}$ 为该样本的最大次序统计量。$(x_{(1)}, \ldots, x_{(n)})$ 为该样本的次序统计量。
- 一般情况下，简单样本的样品是独立同分布的，而次序统计量既不独立也不同分布。
- 如，若总体为均匀分布 $[a, b]$，想知道 a，b，可以考虑 $Y = \min\{x_1, \ldots, x_n\}$ 和 $Y' = \max\{x_1, \ldots, x_n\}$。如，极差 $R = x_{(n)} - x_{(1)}$.

单个次序统计量的抽样分布

- 设总体 X 的分布函数为 $F(x)$, 密度函数为 $p(x)$, 则 $x_{(k)}$ 的密度函数为

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k}p(x).$$

单个次序统计量的抽样分布

- 设总体 X 的分布函数为 $F(x)$, 密度函数为 $p(x)$, 则 $x_{(k)}$ 的密度函数为
$$p_k(x) = \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k}p(x).$$

- 第 k 个值落在区间 $(x, x+\Delta]$ 内，即
$$\begin{cases} k-1\text{个小于}x, \\ 1\text{个落入}(x, x+\Delta], \\ n-k\text{个大于}x+\Delta. \end{cases}$$

一共有 $\binom{n}{k} \times k = \frac{n!}{(n-k)!(k-1)!}$ 种可能性，所有
$$F_k(x+\Delta) - F_k(x) \approx \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k}(p(x)\Delta).$$

单个次序统计量的抽样分布

- 最小次序统计量 $x_{(1)}$ 的密度函数为

$$p_1(x) = n(1 - F(x))^{n-1} p(x).$$

单个次序统计量的抽样分布

- 最小次序统计量 $x_{(1)}$ 的密度函数为

$$p_1(x) = n(1 - F(x))^{n-1} p(x).$$

- 最大次序统计量 $x_{(n)}$ 的密度函数为

$$p_n(x) = n(F(x))^{n-1} p(x).$$

单个次序统计量的抽样分布

- 最小次序统计量 $x_{(1)}$ 的密度函数为

$$p_1(x) = n(1 - F(x))^{n-1} p(x).$$

- 最大次序统计量 $x_{(n)}$ 的密度函数为

$$p_n(x) = n(F(x))^{n-1} p(x).$$

- 若总体分布为均匀分布 $U(0,1)$, 则第 k 个次序统计量的密度函数为

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, \quad 0 < x < 1.$$

单个次序统计量的抽样分布

- 最小次序统计量 $x_{(1)}$ 的密度函数为

$$p_1(x) = n(1 - F(x))^{n-1} p(x).$$

- 最大次序统计量 $x_{(n)}$ 的密度函数为

$$p_n(x) = n(F(x))^{n-1} p(x).$$

- 若总体分布为均匀分布 $U(0,1)$, 则第 k 个次序统计量的密度函数为

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, \quad 0 < x < 1.$$

这是贝塔分布 $Be(k, n-k+1)$。

单个次序统计量的抽样分布

- 最小次序统计量 $x_{(1)}$ 的密度函数为

$$p_1(x) = n(1 - F(x))^{n-1} p(x).$$

- 最大次序统计量 $x_{(n)}$ 的密度函数为

$$p_n(x) = n(F(x))^{n-1} p(x).$$

- 若总体分布为均匀分布 $U(0,1)$, 则第 k 个次序统计量的密度函数为

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, \quad 0 < x < 1.$$

这是贝塔分布 $Be(k, n-k+1)$。一般贝塔分布的密度函数为

$$p(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad E(Be) = \frac{a}{a+b}.$$

多个次序统计量及其抽样分布

- 次序统计量 $(x_{(i)}, x_{(j)})$, $i < j$, 的联合分布密度函数为

$$p_{i,j}(y,z) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(y)]^{i-1}[F(z)-F(y)]^{j-i-1} \times [1-F(z)]^{n-j}p(y)p(z), \quad y \leqslant z.$$

多个次序统计量及其抽样分布

- 次序统计量 $(x_{(i)}, x_{(j)})$, $i < j$, 的联合分布密度函数为

$$p_{i,j}(y,z) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(y)]^{i-1}[F(z)-F(y)]^{j-i-1} \\ \times [1-F(z)]^{n-j}p(y)p(z), \quad y \leqslant z.$$

- 对于 $y < z$，考虑事件 $\{x_{(i)} \in (y, y+\Delta y],\ x_{(j)} \in (z, z+\Delta z]\}$, 则有

$$\begin{cases} i-1\text{小于}y, \\ 1\text{个落到}(y, y+\Delta y], \\ j-i-1\text{个落到}(y+\Delta y, z], \\ 1\text{个落到}(z, z+\Delta z], \\ n-j\text{大于}z+\Delta z. \end{cases}$$

有 $\binom{n}{i} \times i \times \binom{n-i}{j-i} \times (j-i)$ 中可能的组合。

样本极差的抽样分布

- 设总体的分布为均匀分布 $U(0,1)$，则 $(x_{(1)}, x_{(n)})$ 的联合分布密度为

$$p(y,z) = n(n-1)(z-y)^{n-2}.$$

样本极差的抽样分布

- 设总体的分布为均匀分布 $U(0,1)$，则 $(x_{(1)}, x_{(n)})$ 的联合分布密度为

$$p(y,z) = n(n-1)(z-y)^{n-2}.$$

- 样本极差为 $R = x_{(n)} - x_{(1)}$, 则给定 $x_{(1)} = y$,

$$p(y,r) = n(n-1)r^{n-2}, \quad y > 0, r > 0, y + r < 1.$$

则

$$p_R(r) = \int_0^{1-r} n(n-1)r^{n-2}dy = n(n-1)r^{n-2}(1-r).$$

样本极差的抽样分布

- 设总体的分布为均匀分布 $U(0,1)$，则 $(x_{(1)}, x_{(n)})$ 的联合分布密度为

$$p(y,z) = n(n-1)(z-y)^{n-2}.$$

- 样本极差为 $R = x_{(n)} - x_{(1)}$，则给定 $x_{(1)} = y$，

$$p(y,r) = n(n-1)r^{n-2}, \quad y > 0, r > 0, y + r < 1.$$

则

$$p_R(r) = \int_0^{1-r} n(n-1)r^{n-2}\,dy = n(n-1)r^{n-2}(1-r).$$

- 一般分布方法类似，不一定能用初等函数表示。

样本分位数与样本中位数

- 分布函数 $F(x)$ 的 p-分位数为 x_p 满足 $F(x_p)=p$.

样本分位数与样本中位数

- 分布函数 $F(x)$ 的 p-分位数为 x_p 满足 $F(x_p) = p$.
- 样本中位数 $m_{0.5}$ 定义为

$$m_{0.5} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{n 为奇数}, \\ \frac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right), & \text{n 为偶数}. \end{cases}.$$

样本分位数与样本中位数

- 分布函数 $F(x)$ 的 p-分位数为 x_p 满足 $F(x_p) = p$.
- 样本中位数 $m_{0.5}$ 定义为

$$m_{0.5} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{n 为奇数,} \\ \frac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right), & \text{n 为偶数.} \end{cases}.$$

- 样本 p 分位数 m_p 定义为

$$m_p = \begin{cases} x_{([np+1])}, & \text{np 不为整数,} \\ \frac{1}{2}\left(x_{(np)} + x_{(np+1)}\right), & \text{np 为整数.} \end{cases}$$

样本分位数的抽样分布

- 设总体密度函数为 $p(x)$, x_q 为其 q 分位数，$p(x)$ 在 x_q 处连续且 $p(x_q) > 0$, 则当 $n \to \infty$ 时样本 q 分位数 m_q 的渐进分布为

$$m_q \dot\sim N(x_q, \frac{q(1-q)}{np^2(x_q)}).$$

特别地，对于样本中位数，

$$m_{0.5} \dot\sim N(x_{0.5}, \frac{1}{4np^2(x_q)}).$$

样本分位数的抽样分布

- （大概原因）假设密度函数为正，即 $\forall x \in \mathbb{R}$, $p(x) > 0$. 考虑 $Y_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leqslant x) = \frac{1}{n}\sum_{i=1}^{n} I_i(x)$.

样本分位数的抽样分布

- （大概原因）假设密度函数为正，即 $\forall x \in \mathbb{R}$, $p(x) > 0$. 考虑 $Y_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leqslant x) = \frac{1}{n}\sum_{i=1}^n I_i(x)$. $E(I_i(x)) = F_X(x)$, $Var(I_i(x)) = (1 - F_X(x))F_X(x)$.

样本分位数的抽样分布

- (大概原因）假设密度函数为正，即 $\forall x \in \mathbb{R}$, $p(x) > 0$. 考虑 $Y_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leqslant x) = \frac{1}{n}\sum_{i=1}^{n} I_i(x)$.$E(I_i(x)) = F_X(x)$, $Var(I_i(x)) = (1 - F_X(x))F_X(x)$. 所以有

$$\sqrt{n}(Y_n(x) - F(x)) \to A \sim N(0, F_X(x)(1 - F_X(x))).$$

考虑 $g(t) = F_X^{-1}(t)$, $0 < t < 1$. 则 $g'(t) = \frac{1}{p(F_X^{-1}(t))}$. 则，

$$\sqrt{n}(F_X^{-1}(Y_n(x)) - F_X^{-1}(F_X(x))) \to B \sim N(0, \frac{F_X(x)(1 - F_X(x))}{(p(F_X^{-1}(F_X(x))))^2}),$$

令 $x = X_{(nq)}$，则 $F_X^{-1}(F_X(X_{(nq)})) = X_{(nq)}$ 而且 $|x_q - F_X^{-1}(Y_n(X_{(nq)}))| = O(\frac{1}{n}) \to 0$, 于是 $\sqrt{n}(X_{(nq)} - x_q) \to C \sim N(0, \frac{q(1-q)}{p^2(x_q)})$.

样本分位数的抽样分布

- (大概原因) 假设密度函数为正, 即 $\forall x \in \mathbb{R}$, $p(x) > 0$. 考虑 $Y_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leqslant x) = \frac{1}{n}\sum_{i=1}^n I_i(x)$.$E(I_i(x)) = F_X(x)$, $Var(I_i(x)) = (1 - F_X(x))F_X(x)$. 所以有

$$\sqrt{n}(Y_n(x) - F(x)) \to A \sim N(0, F_X(x)(1 - F_X(x))).$$

考虑 $g(t) = F_X^{-1}(t)$, $0 < t < 1$. 则 $g'(t) = \frac{1}{p(F_X^{-1}(t))}$. 则,

$$\sqrt{n}(F_X^{-1}(Y_n(x)) - F_X^{-1}(F_X(x))) \to B \sim N(0, \frac{F_X(x)(1 - F_X(x))}{(p(F_X^{-1}(F_X(x))))^2}),$$

令 $x = X_{(nq)}$, 则 $F_X^{-1}(F_X(X_{(nq)})) = X_{(nq)}$ 而且 $|x_q - F_X^{-1}(Y_n(X_{(nq)}))| = O(\frac{1}{n}) \to 0$, 于是 $\sqrt{n}(X_{(nq)} - x_q) \to C \sim N(0, \frac{q(1-q)}{p^2(x_q)})$.本页内容不做考试要求

五线概括与箱线图

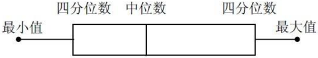
- 考虑 $x_{(1)}$，$m_{0.25}$，$m_{0.5}$，$m_{0.75}$，$x_{(n)}$. $m_{0.25}$ 与 $m_{0.75}$ 有称四分位数。

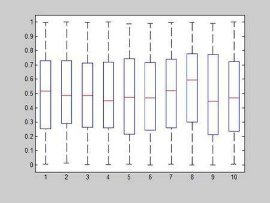
五线概括与箱线图

- 考虑 $x_{(1)}$，$m_{0.25}$，$m_{0.5}$，$m_{0.75}$，$x_{(n)}$. $m_{0.25}$ 与 $m_{0.75}$ 有称四分位数。
- 箱线图：

五线概括与箱线图

- 考虑 $x_{(1)}$，$m_{0.25}$，$m_{0.5}$，$m_{0.75}$，$x_{(n)}$. $m_{0.25}$ 与 $m_{0.75}$ 有称四分位数。
- 箱线图：





三大抽样分布

- 卡方 (χ^2) 分布：设 $X_1, \ldots, X_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $\chi^2 = X_1^2 + \cdots + X_n^2$ 的分布称为自由度为 n 的 χ^2 分布，记为 $\chi^2(n)$.

三大抽样分布

- 卡方 (χ^2) 分布：设 $X_1, \ldots, X_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $\chi^2 = X_1^2 + \cdots + X_n^2$ 的分布称为自由度为 n 的 χ^2 分布，记为 $\chi^2(n)$.
- $X_i^2 \sim Ga(1/2, 1/2)$, $\chi^2 \sim Ga(n/2, 1/2)$.

三大抽样分布

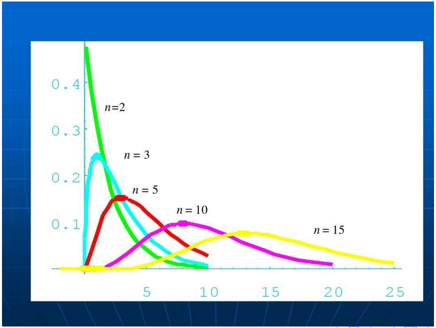
- 卡方 (χ^2) 分布：设 $X_1, \ldots, X_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $\chi^2 = X_1^2 + \cdots + X_n^2$ 的分布称为自由度为 n 的 χ^2 分布，记为 $\chi^2(n)$.
- $X_i^2 \sim Ga(1/2, 1/2)$, $\chi^2 \sim Ga(n/2, 1/2)$.

$$p(y) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} y^{n/2-1} e^{-y/2}.$$

三大抽样分布

- 卡方 (χ^2) 分布：设 $X_1, \ldots, X_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $\chi^2 = X_1^2 + \cdots + X_n^2$ 的分布称为自由度为 n 的 χ^2 分布，记为 $\chi^2(n)$.
- $X_i^2 \sim Ga(1/2, 1/2)$, $\chi^2 \sim Ga(n/2, 1/2)$.

$$p(y) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} y^{n/2-1} e^{-y/2}.$$

- 

卡方分布

- $x_1, \ldots, x_n$ 时来自正态分布 $N(\mu, \sigma^2)$ 的一个样本, 其中 μ 时已知量。考虑统计量
$$T = \sum_{i=1}^{n} (x_i - \mu)^2.$$
- $$\frac{T}{\sigma^2} = \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma})^2 \sim \chi^2(n).$$

卡方分布

- $x_1, \ldots, x_n$ 时来自正态分布 $N(\mu, \sigma^2)$ 的一个样本，其中 μ 时已知量。考虑统计量
$$T = \sum_{i=1}^{n} (x_i - \mu)^2.$$
- $$\frac{T}{\sigma^2} = \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma})^2 \sim \chi^2(n).$$
- $\chi^2(n)$ 分布的特征函数为 $(1-2it)^{-n/2}$，则 $(1-i2\sigma^2 t)^{-n/2}$ 为 T 的特征函数。

卡方分布

- $x_1, \ldots, x_n$ 时来自正态分布 $N(\mu, \sigma^2)$ 的一个样本，其中 μ 时已知量。考虑统计量

$$T = \sum_{i=1}^{n} (x_i - \mu)^2.$$

- $$\frac{T}{\sigma^2} = \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma})^2 \sim \chi^2(n).$$

- $\chi^2(n)$ 分布的特征函数为 $(1 - 2it)^{-n/2}$，则 $(1 - i2\sigma^2 t)^{-n/2}$ 为 T 的特征函数。
- $Ga(\alpha, \lambda)$： $p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$, $(1 - \frac{it}{\lambda})^{-\alpha}$.

卡方分布

- $x_1, \ldots, x_n$ 时来自正态分布 $N(\mu, \sigma^2)$ 的一个样本，其中 μ 时已知量。考虑统计量

$$T = \sum_{i=1}^{n} (x_i - \mu)^2.$$

- $$\frac{T}{\sigma^2} = \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma})^2 \sim \chi^2(n).$$

- $\chi^2(n)$ 分布的特征函数为 $(1 - 2it)^{-n/2}$，则 $(1 - i2\sigma^2 t)^{-n/2}$ 为 T 的特征函数。
- $Ga(\alpha, \lambda)$: $p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$, $(1 - \frac{it}{\lambda})^{-\alpha}$.
- $T \sim Ga(n/2, \frac{1}{2\sigma^2})$. $p_T(t) = \frac{1}{(2\sigma^2)^{n/2}\Gamma(n/2)} e^{-\frac{t}{2\sigma^2}} t^{n/2-1}$.

卡方分布

- 设 $x_1, \ldots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_i - \bar{x})^2.$$

卡方分布

- 设 $x_1, \ldots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_i - \bar{x})^2.$$

- $\bar{x}$ 与 s^2 相互独立。

卡方分布

- 设 $x_1, \dots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_i - \bar{x})^2.$$

- $\bar{x}$ 与 s^2 相互独立。
- $\bar{x} \sim N(\mu, \sigma^2/n)$.

卡方分布

- 设 $x_1,\ldots,x_n$ 是来自正态分布 $N(\mu,\sigma^2)$ 的样本，其样本均值和样本方差分别为

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\quad s^2=\frac{1}{n-1}\sum_{k=1}^{n}(x_i-\bar{x})^2.$$

- $\bar{x}$ 与 s^2 相互独立。
- $\bar{x}\sim N(\mu,\sigma^2/n)$.
- $\frac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$.

卡方分布

- 设 $x_1, \ldots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为

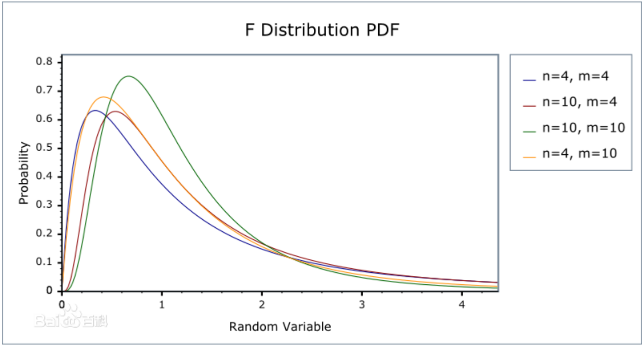
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_i - \bar{x})^2.$$

- $\bar{x}$ 与 s^2 相互独立。
- $\bar{x} \sim N(\mu, \sigma^2/n)$.
- $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$.
- 记号：$\chi^2_{1-\alpha}(n)$ 为自由度为 n 的 χ^2 分布的 $1-\alpha$ 分位数：$P(\chi^2 \leqslant \chi^2_{1-\alpha}(n)) = 1-\alpha$.

F 分布

- 若 $X_1 \sim \chi^2(m)$, $X_2 \sim \chi^2(n)$, 且 X_1 与 X_2 相互独立，则 $F = \frac{X_1/m}{X_2/n}$ 的分布为自由度为 m 和 n 的 F 分布，记为 $F \sim F(m, n)$.

F 分布

- 若 $X_1 \sim \chi^2(m)$, $X_2 \sim \chi^2(n)$, 且 X_1 与 X_2 相互独立，则 $F = \frac{X_1/m}{X_2/n}$ 的分布为自由度为 m 和 n 的 F 分布，记为 $F \sim F(m,n)$.
- 其密度函数为 $\frac{\Gamma(\frac{m+n}{2})(\frac{m}{n})^{m/2-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} y^{m/2-1}(1+\frac{m}{n}y)^{-\frac{m+n}{2}}$
- 

F 分布

- 记 $F_{1-\alpha}(m,n)$ 为自由度为 m, n 的 F 分布的 $1-\alpha$ 分位数。则

$$F_\alpha(n,m)=\frac{1}{F_{1-\alpha}(m,n)}.$$

F 分布

- 记 $F_{1-\alpha}(m,n)$ 为自由度为 m, n 的 F 分布的 $1-\alpha$ 分位数。则

$$F_\alpha(n,m) = \frac{1}{F_{1-\alpha}(m,n)}.$$

$$F \sim F(m,n), \quad \frac{1}{F} \sim F(n,m),$$

F 分布

- 记 $F_{1-\alpha}(m,n)$ 为自由度为 m, n 的 F 分布的 $1-\alpha$ 分位数。则

$$F_\alpha(n,m)=\frac{1}{F_{1-\alpha}(m,n)}.$$

$$F\sim F(m,n),\quad \frac{1}{F}\sim F(n,m),$$

$$\alpha=P(\frac{1}{F}\leqslant F_\alpha(n,m))=P(F\geqslant\frac{1}{F_\alpha(n,m)})$$

F 分布

- 记 $F_{1-\alpha}(m,n)$ 为自由度为 m, n 的 F 分布的 $1-\alpha$ 分位数。则

$$F_\alpha(n,m)=\frac{1}{F_{1-\alpha}(m,n)}.$$

$$F\sim F(m,n),\quad \frac{1}{F}\sim F(n,m),$$

$$\alpha=P(\frac{1}{F}\leqslant F_\alpha(n,m))=P(F\geqslant\frac{1}{F_\alpha(n,m)})$$

$$P(F\leqslant\frac{1}{F_\alpha(n,m)})=1-\alpha.$$

F 分布

- 设 $x_1, \ldots, x_m$ 为来自 $N(\mu_1, \sigma_1^2)$ 的样本, $y_1, \ldots, y_n$ 为来自 $N(\mu_2, \sigma_2)$ 的样本，而且这两个样本相互独立，s_x^2 和 s_y^2 为它们的样本方差，则

$$F = \frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2} \sim F(m-1, n-1).$$

F 分布

- 设 $x_1, \ldots, x_m$ 为来自 $N(\mu_1, \sigma_1^2)$ 的样本, $y_1, \ldots, y_n$ 为来自 $N(\mu_2, \sigma_2)$ 的样本，而且这两个样本相互独立，s_x^2 和 s_y^2 为它们的样本方差，则

$$F = \frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2} \sim F(m-1, n-1).$$

- $s_x^2/\sigma_1^2 \sim \chi^2(m-1)$, $s_y^2/\sigma_2^2 \sim \chi^2(n-1)$.

F 分布

- 设 $x_1, \ldots, x_m$ 为来自 $N(\mu_1, \sigma_1^2)$ 的样本, $y_1, \ldots, y_n$ 为来自 $N(\mu_2, \sigma_2)$ 的样本，而且这两个样本相互独立，s_x^2 和 s_y^2 为它们的样本方差，则

$$F = \frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2} \sim F(m-1, n-1).$$

- $s_x^2/\sigma_1^2 \sim \chi^2(m-1)$, $s_y^2/\sigma_2^2 \sim \chi^2(n-1)$.
- $F \sim F(m-1, n-1)$.

F 分布-例子

- 设 $x_1, \ldots, x_{15}$ 是总体 $N(0, \sigma^2)$ 的一个样本，求

$$y = \frac{x_1^2 + \cdots + x_{10}^2}{2(x_{11}^2 + \cdots + x_{15}^2)}$$

的分布。

F 分布-例子

- 设 $x_1, \ldots, x_{15}$ 是总体 $N(0, \sigma^2)$ 的一个样本，求

$$y = \frac{x_1^2 + \cdots + x_{10}^2}{2(x_{11}^2 + \cdots + x_{15}^2)}$$

  的分布。
- $(x_1^2 + \cdots + x_{10}^2)/\sigma^2 \sim \chi^2(10).$

F 分布-例子

- 设 $x_1, \ldots, x_{15}$ 是总体 $N(0, \sigma^2)$ 的一个样本，求

$$y = \frac{x_1^2 + \cdots + x_{10}^2}{2(x_{11}^2 + \cdots + x_{15}^2)}$$

  的分布。
- $(x_1^2 + \cdots + x_{10}^2)/\sigma^2 \sim \chi^2(10)$.
- $(x_{11}^1 + \cdots + x_{15}^2)/\sigma^2 \sim \chi^2(5)$.

F 分布-例子

- 设 $x_1, \ldots, x_{15}$ 是总体 $N(0, \sigma^2)$ 的一个样本，求

$$y = \frac{x_1^2 + \cdots + x_{10}^2}{2(x_{11}^2 + \cdots + x_{15}^2)}$$

  的分布。
- $(x_1^2 + \cdots + x_{10}^2)/\sigma^2 \sim \chi^2(10)$.
- $(x_{11}^1 + \cdots + x_{15}^2)/\sigma^2 \sim \chi^2(5)$.
- $y = \frac{(x_1^2 + \cdots + x_{10}^2)/\sigma^2/10}{(x_{11}^1 + \cdots + x_{15}^2)/\sigma^2/5} \sim F(10, 5)$.

t 分布

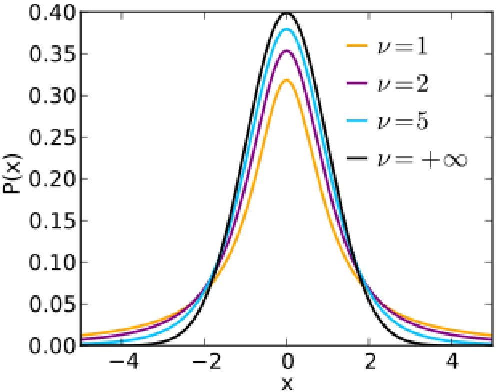
- 设随机变量 X 与 Y 相互独立且 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 则 $t = \frac{X_1}{\sqrt{Y/n}}$ 的分布为自由度为 n 的 t 分布。记 $t \sim t(n)$.

t 分布

- 设随机变量 X 与 Y 相互独立且 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 则 $t = \frac{X_1}{\sqrt{Y/n}}$ 的分布为自由度为 n 的 t 分布。记 $t \sim t(n)$.
- 其密度函数为 $\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{y^2}{n})^{-\frac{n+1}{2}}$, $y \in \mathbb{R}$.

t 分布

- 设随机变量 X 与 Y 相互独立且 $X\sim N(0,1)$, $Y\sim\chi^2(n)$, 则 $t=\frac{X_1}{\sqrt{Y/n}}$ 的分布为自由度为 n 的 t 分布。记 $t\sim t(n)$.
- 其密度函数为 $\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{y^2}{n})^{-\frac{n+1}{2}}$, $y\in\mathbb{R}$.



-

t 分布

- 自由度为 1 的 t 分布是柯西分布。

t 分布

- 自由度为 1 的 t 分布是柯西分布。
- $n > 1$，期望存在，均为 0，$n > 2$ 时方差存在，为 $n/(n-2)$.

t 分布

- 自由度为 1 的 t 分布是柯西分布。
- $n > 1$，期望存在，均为 0，$n > 2$ 时方差存在，为 $n/(n-2)$.
- $n \geqslant 30$ 时，可用标准正态分布逼近。

t 分布

- 自由度为 1 的 t 分布是柯西分布。
- $n > 1$，期望存在，均为 0，$n > 2$ 时方差存在，为 $n/(n-2)$.
- $n \geqslant 30$ 时，可用标准正态分布逼近。
- $t_{1-\alpha} = -t_\alpha$.
- $x_1, \ldots, x_n$ 为正态分布 $N(\mu, \sigma^2)$ 的一个样本，则 $t = \frac{\sqrt{n}(\bar{x}-\mu)}{s} \sim t(n-1)$.

t 分布

- 自由度为 1 的 t 分布是柯西分布。
- $n>1$，期望存在，均为 0，$n>2$ 时方差存在，为 $n/(n-2)$.
- $n \geqslant 30$ 时，可用标准正态分布逼近。
- $t_{1-\alpha}=-t_\alpha$.
- $x_1,\ldots,x_n$ 为正态分布 $N(\mu,\sigma^2)$ 的一个样本，则 $t=\frac{\sqrt{n}(\bar{x}-\mu)}{s}\sim t(n-1)$.

$$\frac{\sqrt{n}(\bar{x}-\mu)}{s}=\frac{(\bar{x}-\mu)/(\sigma/\sqrt{n})}{\sqrt{(n-1)s^2/\sigma^2/(n-1)}}.$$

t 分布

- 自由度为 1 的 t 分布是柯西分布。
- $n > 1$，期望存在，均为 0，$n > 2$ 时方差存在，为 $n/(n-2)$.
- $n \geqslant 30$ 时，可用标准正态分布逼近。
- $t_{1-\alpha} = -t_\alpha$.
- $x_1, \ldots, x_n$ 为正态分布 $N(\mu, \sigma^2)$ 的一个样本，则 $t = \frac{\sqrt{n}(\bar{x}-\mu)}{s} \sim t(n-1)$.

$$\frac{\sqrt{n}(\bar{x}-\mu)}{s} = \frac{(\bar{x}-\mu)/(\sigma/\sqrt{n})}{\sqrt{(n-1)s^2/\sigma^2/(n-1)}}.$$

- $y_1, \ldots, y_m$ 为 $N(\mu', \sigma^2)$ 的样本，则若记 $s_w^2 = \frac{(n-1)s_x^2+(m-1)s_y^2}{m+n-2}$，则

$$\frac{(\bar{x}-\bar{y})-(\mu-\mu')}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2).$$

概率论与数理统计（9）

清华大学

2020 年春季学期

充分统计量

- 假设总体分布具有未知参数 θ, $T(x_1,\ldots,x_n)$ 是一个样本的统计量，它被称为充分的如果没有其他其他同一样本的统计量能比它得到更多关于未知参数的信息。
- 设 $F(\theta,x_1,\ldots,x_n)$ 是统计量 T 的抽样分布，我们考虑 $T=t$ 情形下的 $X=(x_1,\ldots,x_n)$ 的条件分布 $F(\theta,X|T=t)$, 则有以下两种可能
    - $F(\theta,X|T=t)$ 与未知参数 θ 有关;
    - $F(\theta,X|T=t)$ 与未知参数 θ 无关。
- 充分统计量的定义：设 $x_1,\ldots,x_n$ 是来自某总体的样本，总体的分布为 $F(x;\theta)$, 统计量 $T=T(x_1,\ldots,x_n)$ 称为 θ 的充分统计量如果在给定 T 的取值后，$x_1,\ldots,x_n$ 的条件分布于 θ 无关。

- 设总体的分布为 $b(1,\theta)$, $X_1,\ldots,X_n$ 为样本，令 $T=X_1+\cdots+X_n$，则其为 θ 的充分统计量。

$$
\begin{aligned}
&P(X_1=x_1,\ldots,X_n=x_n|T=t)\\
&=\frac{P(X_1=x_1,\ldots,X_{n-1}=x_n,X_n=t-\sum_{i=1}^{n-1}x_i)}{P(\sum_{i=1}^{n}X_i=t)}\\
&=\frac{\prod_{i=1}^{n-1}\theta^{x_i}(1-\theta)^{1-x_i}\theta^{t-\sum_{i=1}^{n-1}x_i}(1-\theta)^{1-t+\sum_{i=1}^{n-1}x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}}\\
&=\frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-5}}=\frac{1}{\binom{n}{t}}
\end{aligned}
$$

充分统计量

- 若总体的分布为 $N(\mu,1)$, 则 $T=\bar{x}$ 为充分统计量。
-

$$p(x_1,\ldots,x_{n-1},t;\mu)\sim\exp\{\frac{-1}{2}[\sum_{i=1}^{n-1}(x_i-\mu)^2+(nt-\sum_{i=1}^{n-1}x_i-\mu)^2]\}$$

$$\sim\exp\{\frac{-1}{2}[n(t-\mu)^2+\sum_{i=1}^{n-1}x_i^2+(\sum_{i=1}^{n-1}x_i-nt)^2-nt^2]\}$$

- $p_{\bar{x}}(t)\sim\exp\{-\frac{n}{2}(t-\mu)^2\}$.

充分统计量

- 总体为均匀分布 $U(0,\theta)$，则 $x_{(n)}$ 是关于 θ 的充分统计量。

$$p(x_1;\theta)\ldots p(x_n;\theta)=\begin{cases}1/\theta^n, & 0<\min\{x_i\}\leqslant\max\{x_i\}<\theta,\\ 0, & else.\end{cases}$$

$g(t,\theta)=\frac{1}{\theta^n}I_{t<\theta}$，$h(X)=1$.

- 总体为 $N(\mu,\sigma^2)$，$\theta=(\mu,\sigma^2)$，则若 $t=\sum_{i=1}^n x_i$ 和 $T=\sum_{i=1}^n x_i^2$，则 (t,T) 为充分统计量。

$$p(X;\theta)\sim e^{-\frac{\sum_{i=1}^n(x_i-\mu)^2}{2\sigma^2}}\sim e^{-\frac{n\mu^2}{2\sigma^2}}e^{-\frac{\sum_{i=1}^n x_i^2-2\mu\sum_{i=1}^n x_i}{2\sigma^2}}$$

$$g(t,T,\theta)\sim e^{-\frac{n\mu^2}{2\sigma^2}}e^{-\frac{T-2\mu t}{2\sigma^2}},\quad h(X)=1.$$

- 寻找充分统计量：总体为 $p(x;\theta)=mx^{m-1}\theta^{-m}e^{-(\frac{x}{\theta})^m}$, $x>0$, $\theta>0$, $m>0$. m 是已知量，$\theta>0$ 为未知量。

$$p(x_1,\ldots,x_n;\theta)=m^n(x_1\ldots x_n)^{m-1}\theta^{-mn}e^{-\sum_{i=1}^n x_i^m/\theta^m}.$$

若令 $T=\sum_{i=1}^n x_i^m$, 取 $g(t,\theta)=\theta^{-nm}e^{-t/\theta^m}$, $h(x_1,\ldots,x_n)=m^n(\prod_{i=1}^n x_i)^{m-1}$, 则有因子分解定理得知，$T=\sum_{i=1}^n x_i^m$ 是 θ 的充分统计量。

无偏差估计

- 对于任一总体，样本均值都是总体的均值（数学期望）的无偏差估计。

$$E(\bar{x}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = E(X).$$

- 样本方差 $s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 不是总体方差的无偏差估计:

$$E(s_n^2) = \frac{n-1}{n}\sigma^2.$$

不过，$E(s_n^2) \to \sigma^2$，$n \to \infty$。这情形，我们称之为渐进无偏差估计。

- 对样本方差稍微修整以下 $s^2 = \frac{n}{n-1}s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. s^2 为总体方差的无偏差估计。

- 统计量的无偏差性不具有函数不变性。即，如果 $\hat{\theta}$ 是参数 θ 的无偏差估计，我们不能直接认为 $g(\hat{\theta})$ 是 $g(\theta)$ 的无偏差估计。如 s^2 是方差的无偏差估计，$s=\sqrt{s^2}$ 一般不是标准差的无偏差估计。
- 假设总体是 $N(\mu,\sigma^2)$，则 $Y=\frac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$，则

$$E(s)=\frac{\sigma}{\sqrt{n-1}}E(Y^{1/2})=\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\sigma.$$

在这种情形下，s 是标准差的渐进无偏差估计。也可以修正为无偏差估计

$$s/\Big(\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\Big).$$

缩小估计中偏差的方法：刀切法

- 设 $T(x)$ 是基于样本 $x=(x_1,\ldots,x_n)$ 的关于参数 $g(\theta)$ 的统计量，且 $E_\theta T(x)=g(\theta)+O(\frac{1}{n})$. 用 x_{-i} 表示在 x 中删除 x_i 后得到的样本向量，即 $(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$。定义刀切统计量为：

$$T_J(x)=nT(x)-\frac{n-1}{n}\sum_{i=1}^{n}T(x_{-i}).$$

- $E_\theta T_J(x)=g(\theta)+O(\frac{1}{n^2})$. 而且其方差也不会增大。
- 大概的原因：假设 $E_\theta T(x)=g(\theta)+\frac{a_1}{n}+\frac{a_2}{n^2}+\cdots$. 则

$$\begin{aligned}E_\theta T_J &= (ng(\theta)+\frac{na_1}{n}+\frac{na_2}{n^2})-\frac{n-1}{n}(ng(\theta)+\frac{na_1}{n-1}+\frac{na_2}{(n-1)^2})\\ &= g(\theta)+\frac{a_2}{n}-\frac{a_2}{n-1}+O(\frac{1}{n^3}).\end{aligned}$$

- 有时候能得到无偏差估计。如果总体为 $b(1,\theta)$，而 $g(\theta)=\theta^2$. 考虑 $T(x)=\bar{x}^2$.

$$E(T(x))=\theta^2+\frac{\theta(1-\theta)}{n}.$$

应用刀切法，

$$T(x_{-i})=(\frac{n\bar{x}-x_i}{n-1})^2=\frac{n^2\bar{x}^2+x_i^2-2nx_i\bar{x}}{(n-1)^2}.$$

于是

$$T_J(x)=n\bar{x}^2-\frac{n-1}{n}\sum_{i=1}^{n}\frac{n^2\bar{x}^2+x_i^2-2nx_i\bar{x}}{n-1}=\frac{n\bar{x}^2}{n-1}-\frac{\sum_{i=1}^{n}x_i^2}{n(n-1)}.$$

$$E(T_J)=\frac{n}{n-1}(\theta^2+\frac{\theta(1-\theta)}{n})-\frac{n(\theta^2+\theta(1-\theta))}{n(n-1)}=\theta^2.$$

不可估参数

- 当总体中的某未知参数存在无偏差估计，则称该参数为可估的，否则称其为不可估的。
- 总体为 $b(1,p)$, 则参数 $\theta = \frac{1}{p}$ 是不可估的。
- $T = x_1 + \cdots + x_n$ 是充分统计量。若 $\hat{\theta}$ 是 θ 的无偏差估计，则对于任意 $p > 0$,

$$\frac{1}{p} = E_\theta(\hat{\theta}) = E(E(\hat{\theta}|T=t)) = \sum_{i=1}^{n} E(\hat{\theta}|T=i)\binom{n}{i}p^i(1-p)^{n-i}.$$

显然 $E(\hat{\theta}|T=t)$ 不可能都等于 0，则上式为 p 的 $n+1$ 次多项式，最多 $n+1$ 个 p 是等号成立。也就是不可能对所有可能的 p 都成立。矛盾。

无偏差估计的有效性

- 假设 θ_1 和 θ_2 均是参数 θ 的无偏差估计，若有

$$Var_\theta(\theta_1) \leqslant Var_\theta(\theta_2)$$

且至少有一个 θ 使得不等号严格成立，则称 θ_1 比 θ_2 有效。

- 如总体的均值为 μ，方差为 σ^2。$x_1, \ldots, x_n$ 是一个样本，则 $\mu_3 = \frac{1}{3}(x_1 + x_2 + x_3)$ 和样本均值 $\bar{x}$ 都是 μ 的无偏差估计，

$$Var(\mu_3) = \frac{\sigma^2}{3}, \quad Var(\bar{x}) = \frac{\sigma^2}{n}.$$

当样本容量大于 3 时，$\bar{x}$ 比 μ_3 有效。

无偏差估计的有效性

- 若 $x_1, \ldots, x_n$ 为总体 $U(0,\theta)$ 的样本，则 $\theta_1 = \frac{n+1}{n} x_{(n)}$ 是 θ 的无偏差估计。

$$E(\theta_1) = \frac{n+1}{n} E(x_{(n)}) = \frac{n+1}{n} \frac{n}{n+1} \theta = \theta.$$

$$Var(\theta_1) = (\frac{n+1}{n})^2 Var(x_{(n)}) = \frac{\theta^2}{n(n+2)}.$$

- $\theta_2 = 2\bar{x}$ 也是 θ 的无偏差估计。

$$E(\theta_2) = 2E(\bar{x}) = 2\frac{\theta}{2} = \theta$$

$$Var(\theta_2) = \frac{4}{n} \frac{\theta^2}{12} = \frac{\theta}{3n}.$$

- 样本容量大于 1 时，θ_1 比 θ_2 有效。

矩估计

- 用样本矩来估计总体的矩。
- 用样本的矩的函数来估计总体矩的函数。
- 随机变量的矩：$E(X^k)$(k-阶原点矩), $E((X-E(X))^k)$(k 阶中心矩).
- 样本的矩：k-阶样本原点矩：

$$\frac{1}{n}\sum_{i=1}^{n} x_i^k.$$

k-阶样本中心矩：

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k.$$

概率函数已知时未知参数的矩估计

- 假设总体具有已知的概率函数 $p(x;\theta_1,\dots,\theta_k)$. 且总体的 k-阶原点矩 μ_k 存在。又假设 $\theta_j=\theta_j(\mu_1,\dots,\mu_k)$, 则 θ_j 的矩估计为

$$\hat{\theta}_j=\theta_j(a_1,\dots,a_k),\qquad a_i=\frac{1}{n}\sum_{k=1}^{n}x_k^i.$$

如果要估计 $\eta=g(\theta_1,\dots,\theta_k)$, 用矩估计方法得到的统计量为

$$\hat{\eta}=g(\hat{\theta}_1,\dots,\hat{\theta}_k).$$

- $\mu=E(X)$, $\hat{\mu}=\bar{x}$.
- $\sigma^2=Var(X)=E(X^2)-(E(X))^2$, $\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}x_i^2-(\bar{x})^2$.

矩估计

- 总体为二点分布 $b(1,p)$, $\theta = \frac{1}{p}$, $\hat{\theta} = \frac{1}{\bar{x}}$.

矩估计

- 总体为二点分布 $b(1,p)$, $\theta = \frac{1}{p}$, $\hat{\theta} = \frac{1}{\bar{x}}$.
- 总体为指数分布 $Exp(\lambda)$: $p(x:\lambda) = \lambda e^{-\lambda x}$, $x > 0$. $E(X) = \frac{1}{\lambda}$. 所以 $\lambda = \frac{1}{E(X)}$.

$$\hat{\lambda} = \frac{1}{\bar{x}}.$$

$Var(X) = \frac{1}{\lambda}$, 则 $\lambda = \frac{1}{\sqrt{Var(X)}}$,

$$\lambda = \frac{1}{s}.$$

- 矩估计可以是多个形式。尽量采取使用低阶矩的矩估计。

矩估计

- 总体为均匀分布 $U(a,b)$, a 和 b 都是未知参数。

$$EX = \frac{a+b}{2}, \quad Var(X) = \frac{(b-a)^2}{12}.$$

$$a = EX - \sqrt{3\,Var(X)}, \quad b = E(X) + \sqrt{3\,Var(X)}.$$

  a, b 的矩估计为

$$\hat{a} = \bar{x} - \sqrt{3}s, \hat{b} = \bar{x} + \sqrt{3}s.$$

- $\hat{a}_1 = x_{(1)}$, $\hat{b}_1 = x_{(n)}$.
- 都是渐进无偏估计。

统计估计量的相合性

- 若 $\theta_n = \hat{\theta}_n(x_1, \ldots, x_n)$ 是 θ 的一个估计量，n 是样本容量，若对任意的 $\epsilon > 0$, 都有

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \geqslant \epsilon) = 0$$

  则称 θ_n 为参数 θ 的相合估计。
- 总体为正态分布 $N(\mu, \sigma^2)$，$\bar{x}$ 为 μ 的相合估计，s_n^2 与 s^2 均是 σ_2^2 的相合估计.
- 总体为 $U(a, b)$ 分布，则 $x_{(1)}$ 为 a 的相合估计，$x_{(n)}$ 是 b 的相合估计。

统计量的相合性

- 设 $\theta_n = \hat{\theta}_n(x_1, \ldots, x_n)$ 是 θ 的一个估计量，如果 $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta, \quad \lim_{n\to\infty} Var(\hat{\theta}_n) = 0$, 则 $\hat{\theta}_n$ 是 θ 的相合估计。

$$P(|\hat{\theta}_n - E(\hat{\theta}_n)| \geqslant \frac{\epsilon}{2}) \leqslant \frac{4}{\epsilon^2} Var(\hat{\theta}_n).$$

当 n 充分大时 $|E(\hat{\theta}_n) - \theta| < \frac{\epsilon}{2}$.

$$\{|\hat{\theta}_n - E\hat{\theta}_n| < \epsilon/2\} \subset \{|\hat{\theta}_n - \theta| < \epsilon\}.$$

- 若 $\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nk}$ 是 $\theta_1, \ldots, \theta_k$ 的相合估计，$\eta = g(\theta_1, \ldots, \theta_k)$ 则 $\hat{\eta}_n = g(\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nk})$ 是 η 的相合估计。也就是说，相合估计具有连续函数不变性。
- 矩估计一般都具有相合性：由大数定律，样本原点矩是原点矩的相合估计（在满足某阶矩的有界性的前提下）。？为何?