Chpt.5 Law of Large Number & Central Limit

第五章 大数定理及中心极限定理

上节回顾

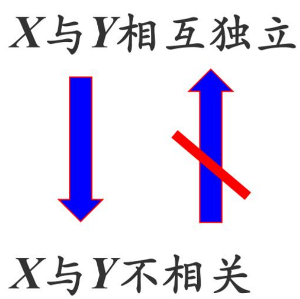
- 协方差：

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

- 相关系数：

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

- 不相关 vs 独立：



(X,Y)服从正态分布，则X,Y独立与X,Y不相关是等价的，充要条件都是 $\rho = 0$

- 矩
- 多维正态分布

设 $X \sim N(\mu, C)$，B 是一个 n 维的可逆矩阵，$Y = BX$

则 $Y \sim N(B\mu, BCB^T)$

思考题

设随机变量(X,Y)服从二元正态分布$N(0,1;1,4;-\frac{1}{2})$，求：

(1)$D(2X-Y)$; (2) $P(2X>Y)$; (3) (Z_1,Z_2)的分布，$Z_1=X+Y, Z_2=X-Y$.

解： 由题意知：$X\sim N(0,1)$, $Y\sim N(1,4)$, $\rho_{XY}=-\frac{1}{2}$.

(1) 由于 $\rho_{XY}=\dfrac{Cov(X,Y)}{\sqrt{D(X)D(Y)}}$,

从而 $Cov(X,Y)=\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}=-\frac{1}{2}\times 1\times 2=-1.$

故
$$\begin{aligned}D(2X-Y)&=D(2X)+D(-Y)+2Cov(2X,-Y)\\&=4D(X)+D(Y)-4Cov(X,Y)\\&=4\times 1+4-4\times(-1)=12.\end{aligned}$$

(2) $(X,Y)\sim N(0,1;1,4;\frac{-1}{2})$, 求： $P(2X>Y)$;

想到 $P(2X>Y)=\iint\limits_{2x>y} f(x,y)dxdy$, ——不可行！！

由于 $P(2X>Y)=P(2X-Y>0)$,

根据多元正态的性质2,由于(X,Y)服从二元正态分布, 故其分量的任意线性组合服从一元正态,即可得, $2X-Y\sim N(-1,12)$.

$E(2X-Y)=2E(X)-E(Y)$

故 $P(2X>Y)=P(2X-Y>0)$

$$=P(\frac{2X-Y-(-1)}{\sqrt{12}}>\frac{0-(-1)}{\sqrt{12}})=1-\Phi(\frac{1}{2\sqrt{3}}).$$

(3) $(X,Y)\sim N(0,1;1,4;\frac{-1}{2})$, 求：$(Z_1,Z_2)$的分布，$Z_1=X+Y, Z_2=X-Y$.

根据多元正态的性质3,即正态变量的线性变换不变性，可知

(Z_1,Z_2)也服从二元正态分布.

$$E(Z_1)=E(X)+E(Y)=1;\ E(Z_2)=E(X)-E(Y)=-1;$$

$$D(Z_1)=D(X+Y)=D(X)+D(Y)+2Cov(X,Y)=1+4+2\times(-1)=3;$$

$$D(Z_2)=D(X-Y)=D(X)+D(Y)-2Cov(X,Y)=1+4-2\times(-1)=7;$$

$$\rho_{Z_1Z_2}=\frac{Cov(Z_1,Z_2)}{\sqrt{D(Z_1)D(Z_2)}}=\frac{Cov(X,X)-Cov(X,Y)+Cov(Y,X)-Cov(Y,Y)}{\sqrt{3\times 7}}$$

$$=\frac{1-4}{\sqrt{3\times 7}}=-\sqrt{\frac{3}{7}};\qquad 则\ (Z_1,Z_2)\sim N(1,-1;3,7;-\sqrt{\frac{3}{7}}).$$

14

本章工作目标包括两个：

[1] 对概率论中的一些结论作出严格的证明；

[2] 为后面的统计作出准备。

概率论早期发展的目的：揭示随机现象的规律性.

概率与频率之间的关系 $\rightarrow$ 大数定律：研究无穷随机试验序列，刻画事件的概率与它发生的频率之间的关系。

大量的相互独立的随机因素的综合影响 $\rightarrow$ 中心极限定理：将观察的误差看作大量独立微小误差的累加，其分布渐近正态。

5.1 Law of Large Number

问题：频率→概率，如何定义这里的“趋向于”？

在相同的条件下，进行n次独立试验，其中事件A发生的次数记为n_A，定义频率 $f_n(A)=\dfrac{n_A}{n}$，我们说 $f_n(A)\to p$，p 就是事件A发生的概率。

用数列极限的语言来描述就是：

$$\forall\varepsilon>0,\exists N,\text{当}\,n>N\text{时},\left|\frac{n_A}{n}-p\right|\le\varepsilon$$

上式存在问题

$$\forall \varepsilon > 0, \exists N, \text{当} n > N \text{时}, \left| \frac{n_A}{n} - p \right| \le \varepsilon$$

其实这里的n_A是一个**随机变量**

因此，$\left| \frac{n_A}{n} - p \right| \le \varepsilon$ 表示的是一个事件，说 $n > N$ 时，总有 $\left| \frac{n_A}{n} - p \right| \le \varepsilon$

是不符合逻辑的，只能说以多大的概率上式成立。

$$P\left\{ \left| \frac{n_A}{n} - p \right| \le \varepsilon \right\} \to 1$$

$$\lim_{n\to\infty} P\left\{ \left| \frac{n_A}{n} - p \right| \le \varepsilon \right\} = 1$$

$\frac{n_A}{n}$依概率收敛于p

能否证明频率以概率收敛于概率？

从n重实验看起（n重实验意味着：　独立同分布　）

定义

$$X_i=\begin{cases}1 & \text{第}i\text{次实验中}A\text{出现}\\ 0 & A\text{未出现}\end{cases}$$

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

$$Y_n=\frac{n_A}{n}=\frac{1}{n}(X_1+X_2+\cdots+X_n)$$

则知道：

$$E(X_i)=p, D(X_i)=p(1-p)=\sigma^2$$

$$E(Y_n)=p,\quad D(Y_n)=\frac{1}{n}\sigma^2$$

$$P\left\{\left|\frac{n_A}{n}-p\right|\le\varepsilon\right\}=P\left\{\left|Y_n-p\right|\le\varepsilon\right\}$$

$$\ge 1-\frac{D(Y_n)}{\varepsilon^2}=1-\frac{1}{n}\left(\frac{\sigma^2}{\varepsilon^2}\right)$$

$$1\ge P\left\{\left|\frac{n_A}{n}-p\right|\le\varepsilon\right\}=P\left\{\left|Y_n-p\right|\le\varepsilon\right\}\ \ge 1-\frac{1}{n}\left(\frac{\sigma^2}{\varepsilon^2}\right)$$

$$\lim_{n\to\infty}P\left\{\left|\frac{n_A}{n}-p\right|\le\varepsilon\right\}=1$$

[Bernoulli大数定理] 设n_A是 n 次独立试验中事件A发生的次数，p 是事件A在一次试验中发生的概率，则对任意的$\varepsilon>0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|\le\varepsilon\right\}=1$$

定义： 一个随机变量的序列$Y_1,Y_2,\cdots,Y_n,\cdots$，如果对任意 $\varepsilon>0$，有 $\lim\limits_{n\to\infty}P\{|Y_n-p|\le\varepsilon\}=1$，则称序列Yn**依概率收敛**到$p$，记为

$$Y_n \xrightarrow{P} p(n\to\infty)$$

Remark（大数定律含义之一）： Bernoulli大数定理说明事件A发生的频率 n_A/n依概率收敛到事件的概率p。以严格的数学形式表达了我们的直观看法。在实际应用中，当试验次数足够大时，便可以用事件的频率来代替事件的概率p（即如何求一个抽象的概率p）

[车比雪夫大数定理] 设随机变量 $X_1,\cdots,X_n,\cdots$ 相互独立，且具有相同的数学期望与方差，记为 $E(X_i)=\mu \quad D(X_i)=\sigma^2 (i=1,2,\cdots,n,\cdots)$。则对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}(X_1+\cdots+X_n)-\mu\right|\le\varepsilon\right\}=1$$

也就是说，$\displaystyle\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} \mu$，当 $n\to+\infty$.

注：伯努利大数定理是车比雪夫大数定理的特殊形式

例题：

设随机变量$X_1, X_2, \cdots, X_n, \cdots$相互独立，且它们的分布律为

$$P\{X_i=\sqrt{i}\}=P\{X_i=-\sqrt{i}\}=\frac{1}{2i},\ P\{X_i=0\}=1-\frac{1}{i},\ i=1,\ 2,\ \cdots.$$

请讨论$\frac{1}{n}\sum_{i=1}^{n}X_i$的收敛性

$D(X)$存在 $\Rightarrow E(X)$存在 ✓　　$D(X)$存在 $\Leftarrow E(X)$存在 ✗

在切比雪夫大数定理里要求期望和方差都存在，
如果期望存在，方差不存在会如何？

类似的大数定理也是存在的，但要求这些随机变量是同分布的。

[辛钦大数定理]　设随机变量 $X_1,\cdots,X_n,\cdots$ 相互独立且同分布，数学期望存在，记为 μ。则对任意的 $\varepsilon>0$ 有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}(X_1+\cdots+X_n)-\mu\right|\le\varepsilon\right\}=1$$

也就是说，$\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} \mu$，当 $n\to+\infty$.

注意：这里我们没有要求它的方差存在。如果方差不存在，也可以使用辛钦大数定理。

例题：

设随机变量$X_1, X_2, \cdots, X_n, \cdots$, 相互独立同分布, $X_1 \sim U(-1,1)$. 则

(1) $\frac{1}{n}\sum_{i=1}^{n} X_i$, (2) $\frac{1}{n}\sum_{i=1}^{n} |X_i|$, (3) $\frac{1}{n}\sum_{k=1}^{n} X_i^2$ 分别依概率收敛吗？

如果依概率收敛，分别收敛于什么？（当 $n \to +\infty$时）

Remark（大数定律含义之二）：

满足一定条件的随机变量$X_1,\cdots,X_n,\cdots$，当n很大时，算术平均$\frac{1}{n}(X_1+\cdots+X_n)$接近于数学期望$E(X_i)=\mu$。

- **提供了求随机变量X的数学期望$E(X)$的近似值的方法:**

将随机变量X独立重复地观察n次, 记第k次观测值为X_k,

则$X_1,X_2,\cdots,X_n$相互独立, 且与X具有同样的分布.

那么, 当$E(X)$存在时, 由辛钦大数定律, 可知当n充分大时, 可将n次的平均$\frac{1}{n}\sum_{i=1}^{n}X_i$作为$E(X)$的近似.

思考：如何求南开大学学生的平均身高？

概括前面的几个定理，可以归结为两点：

1 频率稳定性：

事件A发生的频率依概率收敛到事件A的概率

事件A发生的概率为 p	$\Rightarrow \dfrac{n_A}{n} \xrightarrow{p} p \quad (n \to \infty)$
进行n次独立试验，A出现 n_A	

2 算术均值稳定性：

随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立	$\Rightarrow \dfrac{1}{n}(X_1 + X_2 + \cdots + X_n) \xrightarrow{p} \mu \quad (n \to \infty)$
具有相同的期望 μ 和方差 σ^2 或者同分布且期望 μ 存在	

依概率收敛的性质：

若 $X_n \xrightarrow{P} a,\ Y_n \xrightarrow{P} b$, 当 $n \to \infty$时, 函数$g(x,y)$在点(a,b)处连续, 那么

$g(X_n, Y_n) \xrightarrow{P} g(a,b)$, 当$n \to \infty$时.

如：当$n \to \infty$时, $X_n + Y_n \xrightarrow{P} a+b$,

$X_n \times Y_n \xrightarrow{P} a \times b$, $X_n / Y_n \xrightarrow{P} a/b\ (b \neq 0)$.

特别地, 若$X_n \xrightarrow{P} a$, $f(x)$在点a连续, 则

$f(X_n) \xrightarrow{P} f(a)$, 当$n \to \infty$时.

思考题

设随机变量$X_1, X_2, \cdots, X_n, \cdots$独立同分布，$X_1 \sim U(0,1)$，则$\sqrt[n]{X_1X_2\ldots X_n}$依概率收敛吗？如果依概率收敛，收敛于什么？