Date.　　/

不等式

- Markov 不等式 对 r.v. X 和 $\varepsilon>0$，有 $P(|X|\geq\varepsilon)\leq\frac{E|X|^\alpha}{\varepsilon^\alpha}$，$\alpha>0$

（证明）$P(|X|>\varepsilon)=EI_{(|X|^\alpha>\varepsilon^\alpha)}\leq E\left\{\left(\frac{|X|^\alpha}{\varepsilon^\alpha}\right)I_{(|X|^\alpha>\varepsilon^\alpha)}\right\}\leq E\left(\frac{|X|^\alpha}{\varepsilon^\alpha}\right)=\frac{E|X|^\alpha}{\varepsilon^\alpha}$，$\alpha>0$.

（推论）Chebyshev 不等式 $P(|X-EX|\geq\varepsilon)\leq\frac{var(X)}{\varepsilon^2}$.

- 内积不等式 设 EX^2 和 $EY^2<\infty$，则 $|E(XY)|\leq\sqrt{EX^2EY^2}$.（等号成立 $\Leftrightarrow$ $\exists a,b$ 不全为0，$aX+bY=0$ a.s.）

（引入线代内容）$\langle\vec{a},\vec{b}\rangle=\vec{a}\cdot\vec{b}$，$\langle X,Y\rangle\triangleq E(X,Y)$，$|X|^2=\langle X,X\rangle=EX^2$.

$E|XY|\leq\sqrt{EX^2EY^2}$，$|\vec{a}\cdot\vec{b}|\leq\sqrt{|\vec{a}|^2|\vec{b}|^2}$，$|\vec{a}||\vec{b}|\cos\theta\leq|\vec{a}||\vec{b}|$.

- Jensen 不等式（Jensen's inequality）

设 φ 是 $\mathbb{R}$ 上的凸函数（convex），X 和 $\varphi(X)$ 都有有限的期望，则有

$\varphi(EX)\leq E(\varphi(X))$. 对严格凸的 φ，等号成立当且仅当 $X=EX$. a.s.

证明：凸函数 → 对任意 a,b 都存在一常数 c，使得

$\varphi(b)-\varphi(a)\geq c(b-a)$. 令 $a=EX$，$b=X$，则有 $\varphi(X)-\varphi(EX)\geq c(X-EX)$.

两侧求期望得 $E(\varphi(X))-\varphi(EX)\geq cE(X-EX)=0$.

应用：EM 算法（Expectation-Maximization algorithm）；证明 MLE（maximum likelihood estimator）的一致性等.

条件期望、条件方差

期望（离散）若 $\sum_{i=1}^{\infty}|x_i|P(X=x_i|A)<\infty$，则 $E(X|A)=\sum_{i=1}^{\infty}x_iP(X=x_i|A)$.

（连续）若 $\int_{-\infty}^{\infty}|x|f_{X|A}(x|A)dx<\infty$，则 $E(X|A)=\int_{-\infty}^{\infty}xf_{X|A}(x|A)\,dx$.

条件期望的3种理解（具体解释见下页）

代数定义（平均值）　　几何（投影）　　选择最优化方案（预测）

- 例 灯泡已工作了 a 天，寿命 $X\sim\mathcal{E}(\lambda)$，求现在寿命的期望（$\frac{1}{\lambda}+a$）.

$E(X|X\geq a)=\int_a^{+\infty}xf(x|x\geq a)dx$. 其中 $f(x|X\geq a)=\frac{\lambda e^{-\lambda x}}{e^{-\lambda a}}=\lambda e^{\lambda(a-x)}$.

$E(X|X\geq a)=\int_a^{+\infty}\lambda e^{\lambda a}\cdot(xe^{-\lambda x})dx=\frac{1}{\lambda}+a$.

灯泡 $X\sim\mathcal{E}(\lambda)$，$Y\sim\mathcal{E}(\mu)$ 且相互独立，X 寿命短于 Y，求 $E(X|X<Y)$.

$E(X|X<Y)=\frac{E(XI_{X<Y})}{P(X<Y)}=\frac{\int_0^{+\infty}f(z)dz}{\int_0^{+\infty}f(x)dx\int_x^{+\infty}f(y)dy}$，$Z=XI_{X<Y}$，$f(z)=\lambda e^{-(\lambda+\mu)z}$.

$E(X|X<Y)=\frac{\frac{\lambda}{(\lambda+\mu)^2}}{\frac{\lambda}{\lambda+\mu}}=\frac{1}{\lambda+\mu}$.

Date. /

定理 设 $P(A)$ 是非负随机变量，则 $E(X|A)=\frac{E(XI_A)}{P(A)}$.

证明 $E(X|A)$ 是在 $P_A(\cdot)=P(\cdot|A)$ 下对 X 求期望，则 $E(X|A)=\int_0^{\infty}P(X>x|A)dx$

$E(X|A)=\frac{1}{P(A)}\int_0^{\infty}P(\{X>x\}\cap A)dx=\frac{1}{P(A)}\int_0^{\infty}P(XI_A>x)dx=\frac{E(XI_A)}{P(A)}$.

条件期望转化为无条件期望 + 事件概率组合操作.

定理 (全期望定理) (全概率公式的变形)

设 $A_1\dots A_n$ 构成 Ω 的一个分割. $E(X)=\sum_{i=1}^{n}P(A_i)E(X|A_i)$.

例 连续抛掷一枚均匀硬币，分别记 W_{HT}. W_{HH} 为等到第一个 HT. HH 的抛掷次数. 求 EW_{HT} 和 EW_{HH}. (4, 6)

背景: $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，给定 $Y=y$ 时 $X|Y=y\sim N(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),(1-\rho^2)\sigma_1^2)$.

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

在二元正态分布情况下恰能实现最优线性拟合.

$E(X|Y)$ 定义: (X,Y) 为随机向量，$E|X|<\infty$，若 $m(y)=E(X|Y=y)$，则称随机变量 $m(Y)$ 为给定 Y 对 X 的条件期望，记为 $E(X|Y)$.

(离散型) 若 $\sum|x_i|P(X=x_i|Y=y)<\infty$，则 $E(X|Y=y)=\sum_{i=1}^{\infty}x_iP(X=x_i|Y=y)$.

(连续型) 若 $\int_{-\infty}^{\infty}|x|f_{X|Y}(x|y)dx<\infty$，则 $E(X|Y=y)=\int_{-\infty}^{\infty}xf_{X|Y}(x|y)dx$.

考虑 $E(X+Y|X)=m(X)$，其中 $EY=\lambda$. $m(x)=E(X+Y|X=x)=E(Y|X=x)+x$

~~$m(x)=$~~ $E_A(X+Y)$ (令 $A=\{X=x\}$). X、Y 独立，则 $m(x)=x+\lambda$.

说明: $m(Y)$ 实际上是一个随 Y 变化的随机变量.

例. 设 X、Y 独立同分布，$X\sim P(\lambda)$，求 $E(X+Y|X)$ 和 $E(X|X+Y)$. $\Rightarrow E(X|X+Y=n)=\frac{n}{2}$.

1) $E_AX=\sum kP_A(k)$，$P_A(k)=P(X=k|X+Y=n)$.

2) $X|X+Y=n\sim B(n,\frac{1}{2})$，$X,Y\sim P(\lambda)$，$\frac{X+Y}{2}$ 也为 r.v.

3) $E(X+Y|X+Y=n)=n=E(X|X+Y=n)+E(Y|X+Y=n)=2E(X|X+Y=n)$.

那么 $E(X+Y|X)\sim P(\lambda)$. , $E(X|X+Y)\sim B(X+Y,\frac{1}{2})$.

统计学中的线性回归模型即基于条件期望思想. (希望用条件期望→期望)

(定理) 重期望法则: X,Y 为随机变量，$E(E(X|Y))=EX$.

(证明) $E(E(X|Y))=E(m(Y))=\sum_y m(y)P_Y(y)=\sum_y E(X|Y=y)P_Y(y)$ (接下页)

(以离散为例)

Date.　　/

$=\sum_y\sum_x xP_{X|Y}(x|y)P_Y(y)=\sum_x x\sum_y P_{X|Y}(x|y)P_Y(y)=\sum_x xP(x)=EX$.

常对有层次结构有简化作用，如Bayes网络.

条件期望的性质（X、Y为r.v，$g(x)$与$h(y)$为实函数，$E|X|<\infty$，$E|g(X)|<\infty$）

1° $|E(X|Y)|\le E(|X||Y)$.　（$E|X|\ge|EX|$）

2° $[E(X|Y)]^2\le E(X^2|Y)$.

3° $E[h(Y)g(X)|Y]=h(Y)E[g(X)|Y]$

4° $E[E(g(X)|Y)]=Eg(X)$.

5° 当X、Y独立时 $E[g(X)|Y]=Eg(X)$；

6° r.v. $X_1,\dots,X_n$的期望有限，$E(c+\sum_{i=1}^n c_iX_i|Y)=c+\sum_{i=1}^n c_iE(X_i|Y)$.

7° 若$X_1\le X_2$，则 $E(X_1|Y)\le E(X_2|Y)$.

例 X、Y为随机变量，$h(x)$是实函数，若$EX^2<\infty$，$Eh^2(Y)<\infty$，则

$E[(X-E(X|Y))h(Y)]=0$.

证明：由Cauchy-Schwarz不等式可得 $E(|Xh(Y)|)\le\sqrt{EX^2Eh^2(Y)}<\infty$.

利用条件期望性质，$E((X-E(X|Y))h(Y))=E(Xh(Y))-\cancel{E(X|Y)}\,E(E(X|Y)h(Y))$.

$=E(Xh(Y))-E(E(Xh(Y)|Y))=E(Xh(Y))-E(Xh(Y))=0$.

·最佳预测性质：设$EX^2<\infty$，$m(Y)=E(X|Y)$，则对任意实函数$g(Y)$有

$E(X-m(Y))^2\le E(X-g(Y))^2$. 等号成立当且仅当$g(Y)=m(Y)$ a.s.

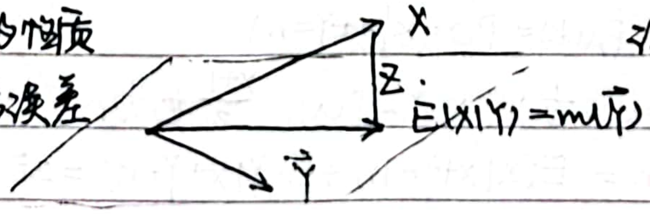
（证明）$E(g(Y))^2=\infty$时一定成立. 否则：首先$E^2(X|Y)\le E(X^2|Y)$，从而$E(m(Y))^2\le EX^2$. 有$E(h(Y))^2<\infty$.

令$h(Y)=m(Y)-g(Y)$，则$E(X-g(Y))^2=E(X-mY+mY-gY)^2=E(X-mY)^2+E(mY-gY)^2$

$+2E(X-mY)(mY-gY)=E(X-mY)^2+E(mY-gY)^2\ge E(X-mY)^2$.

·图形理解：E(X|Y)的性质　　设$Z=\vec{X}-E(X|Y)$.

可以证明E(X|Y)是均方误差意义下的最优预测.



~~E(X|Y)~~ E(X|Y)的意义：对正态分布而言 $E(X|Y)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(Y-\mu_2)$，最佳预测=最佳线性预测

说明：证明前必须指出~~若~~EX^2有限.

等号成立当且仅当$E[m(Y)-g(Y)]^2=0$，当且仅当$g(Y)=m(Y)$ a.s. 类比$var(X)\le E(X-c)^2$.

称$m(Y)=E(X|Y)$是X的最佳预测(optimal forecast)

若$E(X|Y)=C$，X、Y不一定独立：反例仍为$x^2+y^2\le 1$均匀分布.　　Date.　/

(定义) 若$E([X-E(X|Y)]^2|Y)$存在，称其为给定Y下X的条件方差，即

$$var(X|Y)=E\{[X-E(X|Y)]^2|Y\}.$$

(定理) 全方差法则：$var(X)=E(var(X|Y))+var(E(X|Y))$.

证明：$var(X)=E(X-EX)^2=E\{[X-E(X|Y)+E(X|Y)]-E(X-E(X|Y)+E(X|Y))\}^2$

$=E([X-E(X|Y)]^2+var(E(X|Y))=E(E\{[X-E(X|Y)]^2|Y\})+var(E(X|Y))=E(var(X|Y))+var(E(X|Y))$

例 (论文数问题)

tips：条件在随机变量上时，先用$X=x$定值计算，再将结果中定值换为随机变量X.

利用对称性；求和/积分换序等价.

二次型期望运算中用加减项组合出0均值随机变量.

协方差与相关

(定义) 设EX和EY存在，当$E|(X-EX)(Y-EY)|<\infty$时，称

$cov(X,Y)=E[(X-EX)(Y-EY)]$为X、Y的协方差(covariance).

$cov>0\Rightarrow$正相关，$cov=0\Rightarrow$不相关，$cov<0\Rightarrow$负相关.

性质　1° $cov(X,X)=var(X)$　　2° $cov(X,Y)=cov(Y,X)$

3° $cov(X,Y)=E(XY)-EXEY$ (常用计算公式)

4° 若c为某常数，则$cov(X,c)=0$.

5°(双线性) $cov(cX,Y)=c\,cov(X,Y)$，$cov(X,Y+Z)=cov(X,Y)+cov(X,Z)$

6° $var(X_1+X_2)=var(X_1)+var(X_2)+2cov(X_1,X_2)$.

($\Rightarrow var(X_1+\cdots X_n)=\sum_{i=1}^n var(X_i)+\sum_{i\ne j}cov(X_i,X_j)=\sum_{i=1}^n var(X_i)+2\sum_{i<j}cov(X_i,X_j)$).

例　多项分布：经典离散型随机向量例

同学们来自r个年级，两课堂随机点名(有放回)，记A_i表示点到第i个

由此$EXY=EXEY$不能推出X、Y独立，因为它只刻画了线性关联.

deli得力

Date. /

年级的同学，一堂课计划点名 n 次，用 X_i 表示 A_i 发生次数，则

$(X_1,\dots X_r)$ 的概率分布是 $P(X_1=k_1,\dots X_r=k_r)=\frac{n!}{k_1!\dots k_r!}p_1^{k_1}\dots p_r^{k_r}$.

(其中 $k_i\geq 0$, $p_i=P(A_i)$, $\sum_{i=1}^{r}k_i=n$)

$X_i\sim B(n,p_i)$, $var(X_i)=np_i(1-p_i)$；先考虑1,2项

$cov(X_1,X_2)\Rightarrow var(X_1+X_2)=var(X_1)+var(X_2)-2cov(X_1,X_2)$. $X_1+X_2\sim B(n,p_1+p_2)$.

或者用示性函数表示：$X_1=I_1+I_2+\dots I_n$, $X_2=J_1+\dots J_n$. $I_i\sim B(p_1)$ $J_i\sim B(p_2)$

$cov(\sum I_i,\sum J_j)=\sum_i\sum_j cov(I_i,J_j)=\sum_i\sum_j(E(I_iJ_j)-EI_iEJ_j)$

有放回抽取相互独立，且 I_iJ_j 在 $i=j$ 时一定为0.

$cov(\sum I_i,\sum J_j)=\sum_{i=j}EI_iJ_j+\sum_{i\neq j}EI_iJ_j-\sum_i\sum_j EI_iEJ_j=0+n(n-1)p_1p_2-n^2p_1p_2=-np_1p_2$.

说明：协方差的符号表示相关趋势，但大小不具有可比性（随单位变化）

(定义) $0<var(X)var(Y)<\infty$ 时 称 $corr(X,Y)=\frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$ 为

X、Y 的相关系数 (Pearson's correlation)，　对X、Y标准化

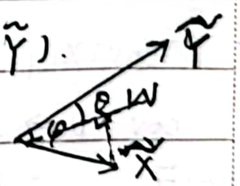
以二项正态分布为例 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)\Rightarrow corr(X,Y)=\rho$.

一般记作符号 ρ_{XY}. 相关系数具有可比性. 且XY独立 ⇔ XY不相关.

定理 $|\rho_{XY}|\leq 1$. $|\rho_{XY}|=1\Leftrightarrow$ 存在常数a、b、c 使得 $aX+bY=c$. a.s. ⇔ X、Y共线性

(深入理解) 定理：X Y相互独立，则X、Y不相关，即 $cov(X,Y)=0$.

考虑 $\tilde{X}=X-EX$, $\tilde{Y}=Y-EY$. 则 $E\tilde{X}=0$, $E\tilde{Y}=0$. $corr(X,Y)=corr(\tilde{X},\tilde{Y})$.



几何解释：$\langle\tilde{X},\tilde{Y}\rangle\triangleq E(\tilde{X}\tilde{Y})\leftrightarrow\langle\vec{a},\vec{b}\rangle\triangleq\vec{a}\cdot\vec{b}$.

$corr(X,Y)=corr(\tilde{X},\tilde{Y})=\frac{cov(\tilde{X},\tilde{Y})}{\sqrt{var(\tilde{X})}\sqrt{var(\tilde{Y})}}=\frac{E(\tilde{X}\tilde{Y})}{\sqrt{E\tilde{X}^2}\sqrt{E\tilde{Y}^2}}=\frac{\langle\tilde{X},\tilde{Y}\rangle}{|\tilde{X}||\tilde{Y}|}=\cos\varphi$.

一般 只有二元正态分布时 不相关⇔独立.

也就是说，X Y 可以不相互独立，但（比如垂直时）相关系数为0 ⇒ 没有线性关系.

经典反例：X、Y在单位圆 $x^2+y^2\leq 1$ 内均匀分布，则X、Y不相关也不独立.

归纳说明：不独立 → 线性关联(ρ)或非线性关联. 独立 → 不相关.

几何补充：令 $W=E(\tilde{X}|\tilde{Y})=m(\tilde{Y})$，它是X在Y所在平面上的投影.

$corr(X,Y)=\frac{\langle\tilde{X},\tilde{Y}\rangle}{|\tilde{X}||\tilde{Y}|}=\frac{\langle W,\tilde{Y}\rangle}{|\tilde{X}||\tilde{Y}|}=\frac{|W||\tilde{Y}|}{|\tilde{X}||\tilde{Y}|}\cos\theta$. $|\tilde{X}|$或$|\tilde{Y}|=0$时 corr不存在;

例：(X,Y)服从二元正态分布，当 $E(X|Y)=\mu$ (const). X、Y独立. 因为 $E(X)=E(E(X|Y))=\mu$.

且 $E(X-\mu)Y=E(E((X-\mu)Y|Y))=E[YE(X-\mu)|Y]=E[Y(E(X|Y)-\mu)]=EY\cdot 0=0$. $cov(X,Y)=0$.

二元正态分布下，独立⇔不相关.

概率母函数、矩母函数、特征母函数

(定义) 设X是取非负整值的随机向量，称 $g(s)=E(s^X)=\sum_{j=0}^{\infty}s^j P(X=j)$, $s\in[-1,1]$.

为X的概率母函数(probability-generating function). 约定 $0^0=1$.

背景: $X_1,\dots X_n$ 独立同分布，且 $\sim U(-1,1)$. 则 $S_n=X_1+\dots+X_n$ 的密度函数为

$f_{S_n}(x)=\frac{1}{2^n(n-1)!}\sum_{k=0}^{[\frac{n+x}{2}]}(-1)^k\binom{n}{k}(n+x-2k)^{n-1}$, $|x|\le n$. (之前用卷积推)

显然，概率母函数 $g(s)$ 在 $[-1,1]$ 上绝对收敛.

性质 1° $g(s)$ 为X的pgf, $g^{(k)}(s)$ 是其k阶导数. 则 $P(X=k)=\frac{g^{(k)}(0)}{k!}$, $k=0,1\dots$

(证明: 概率母函数与概率分布列相互唯一确定)

$\frac{g^{(k)}(0)}{k!}=\frac{g^{(k)}(s)}{k!}\Big|_{s=0}=\frac{1}{k!}\sum_{j=k}^{\infty}\frac{d^k}{ds^k}(s^j)P(X=j)\Big|_{s=0}=P(X=k)$.

2° $E(X)=g^{(1)}(1)$. 因为 $g^{(1)}(1)=\sum_{j=0}^{\infty}jP(X=j)=EX$.

3° 若 $EX<\infty$, 则 $\mathrm{var}X=g^{(2)}(1)+g^{(1)}(1)-[g^{(1)}(1)]^2$.

(证) $g^{(2)}(1)+g^{(1)}(1)-[g^{(1)}(1)]^2=\sum_{j=0}^{\infty}j(j-1)P(X=j)+\sum_{j=0}^{\infty}jP(X=j)-E^2X=E(X(X-1))+EX-E^2X=\mathrm{var}(X)$.

4° 若 $X_1,\dots X_n$ 相互独立, $g_i(s)=E(s^{X_i})$ 是 X_i 的概率母函数，则

$Y=X_1+\dots+X_n$ 的pgf为 $g_Y(s)=g_1(s)g_2(s)\dots g_n(s)$, $s\in[-1,1]$.

理解: pgf可以一次性处理多个随机变量，无需再用归纳法划分为2个一组处理.

- 二项分布 $B(n,p)$ 的pgf: $g(s)=\sum_{j=0}^{n}s^j\binom{n}{j}p^jq^{n-j}=(q+sp)^n$.

  则 设 $X_1,\dots X_m$ 相互独立且 $X_i\sim B(n_i,p)$, 则 $X_1+X_2+\dots X_m\sim B(n_1+n_2+\dots+n_m,p)$.

- 泊松分布 $P(\lambda)$ pgf $\Rightarrow$ $g(s)=\sum_{k=0}^{\infty}s^k\frac{\lambda^k}{k!}e^{-\lambda}=e^{\lambda(s-1)}$.

  设 $X_1,\dots X_m$ 相互独立且 $X_i\sim P(\lambda_i)$, 则 $X_1+X_2+\dots X_m\sim P(\lambda_1+\dots+\lambda_m)$.

- 几何分布 $G(p)$. $g(s)=\sum_{k=1}^{\infty}s^kpq^{k-1}=\frac{sp}{1-sq}$.

  设 $X_1,X_2,\dots X_m$ 相互独立且 $X_i\sim G(p)$. 则 $S_m=\sum X_i$, 其pgf: $g_m(s)=\left(\frac{sp}{1-sq}\right)^m$.

对其进行Taylor展开: $g_m(s)=(sp)^m\sum_{j=0}^{\infty}\frac{m(m+1)\dots(m+j-1)}{j!}(sq)^j=\sum_{k=m}^{\infty}\binom{k-1}{m-1}p^mq^{k-m}s^k$

于是得到Pascal分布 $P(S_m=k)=\binom{k-1}{m-1}p^mq^{k-m}$, $k=m,m+1\dots$

pgf本质上是同cdf的函数, $g(s)=Es^X=p_0s^0+p_1s^1+p_2s^2+\dots$ (由 p_i 构成pmf)

$g_X(s)\Rightarrow g_{X_1+\dots+X_n}$ $\begin{cases}\text{已知} & g(s,t)=E(s^Xt^Y).\\ \text{未知} & \text{置变量于指数，对iid可直接拆开.}\end{cases}$

(可以求导) PMF　Taylor展开，直接读函数

母函数可以直接处理iid $X_1+\dots+X_n$ 问题，取代卷积.

Date.

(随机变量个随机变量之和)

设 1) $\{X_j\}$ 是一列独立同分布的非负整值随机变量，且 X_1 的 pgf 为

$\varphi(s)=\sum_{k=0}^{\infty}P_k s^k$，$P_k=\mathbb{P}(X_1=k)$

2) N 为正整数随机变量，且其 pgf $G(s)=\sum_{n=1}^{\infty}s^n\mathbb{P}(N=n)$.

3) N 与 $\{X_j\}$ 相互独立.

则 $W\triangleq X_1+X_2+\dots X_N$ 的概率母函数 $H(s)=G(\varphi(s))$.

例 $X_1,\dots X_2$ 独立同分布，$X_i\sim G(p)$，$N\sim G(\tilde{p})$ 且与全部 X_i 独立，求 $S_N=X_1+\dots+X_N$.

几何分布的概率母函数 $g(s)=\sum_{k=1}^{\infty}s^k pq^{k-1}=\frac{sp}{1-sq}$. $g_N(s)=\frac{s\tilde{p}}{1-s\tilde{q}}$.

代入 $H(s)=G(\varphi(s))=\frac{sp\tilde{p}}{1-s(\tilde{q}+p\tilde{q})}=\frac{sp\tilde{p}}{1-s(1-p\tilde{p})}$，由可逆性可知 S_N 满足几何分布.

$S_N\sim G(p\tilde{p})$.

(有限个几何分布之和不是几何分布，但几何分布个几何分布和仍为几何分布)

(定义)(2维 pgf) 设 (X,Y) 是二维取非负整数值的随机向量，其分布列 $P_{ik}=\mathbb{P}(X=i,Y=k)$，

称 $g(s,t)=E(s^X t^Y)=\sum_{i=0}^{\infty}\sum_{k=0}^{\infty}P_{ik}s^i t^k$，$s,t\in[-1,1]$ 为 X、Y 的概率母函数.

性质 设 $g(s,t)$ 是 (X,Y) 的概率母函数，则

1° $|g(s,t)|\le g(1,1)=1$，$|s|\le1$，$|t|\le1$.

2° $g_{aX+bY+c}(s)=s^c g(s^a,s^b)$，a.b.c 均为常数

3° X 与 Y 独立 $\Leftrightarrow$ $g(s,t)=g_X(s)g_Y(t)$，$\forall s,t$；

4° $g(s,1)=g_X(s)$，$g(1,t)=g_Y(t)$. (有一个变量为 1 → 单变量分布)

5° $P_{ik}=\frac{1}{i!k!}\frac{\partial^{i+k}g(s,t)}{\partial s^i\partial t^k}\Big|_{s=t=0}$，$i,k=0,1,2\dots$

6° 若 $EX\cdot EY<\infty$，则 $EX=\frac{\partial g(s,t)}{\partial s}\Big|_{s=t=1}$，$EY=\frac{\partial g(s,t)}{\partial t}\Big|_{s=t=1}$.

7° 若 $EX^2,EY^2<\infty$，则 $EX^2=\frac{\partial^2 g}{\partial s^2}\Big|_{s=t=1}+\frac{\partial g}{\partial s}\Big|_{s=t=1}$，$EY^2=\frac{\partial^2 g}{\partial t^2}\Big|_{s=t=1}+\frac{\partial g}{\partial t}\Big|_{s=t=1}$，$EXY=\frac{\partial^2 g}{\partial s\partial t}\Big|_{s=t=1}$.

矩母函数

(定义) X 为随机向量，其矩母函数(moment generating function) $M_X(s)=E(e^{sX})$.

(理解) 利用 e^{sX} 的 Taylor 展开可以得到各阶矩 $EX^k=M^{(k)}(0)$.

离散型：$M_X(s)=\sum_j e^{sx_j}\mathbb{P}(X=x_j)$

连续型：$M_X(s)=\int_{-\infty}^{\infty}e^{sx}f(x)\,dx$

仅当 $E(e^{sX})<\infty$ 时 我们称 $M_X(s)$ 存在.

tips. Taylor展开 $M(s)$，e^{is}对应的系数就是 $P(X=i)$.

性质 设 $M(s)$ 为 X 的矩母函数，$M^{(k)}(s)$ 为其 k 阶导函数.

1° $Y=aX+b$ 的矩母函数为 $M_Y(s)=e^{sb}M(sa)$；

2° $EX^k=M^{(k)}(0)$，$k=1,2,\dots$

3° $M(0)=1$，当 X 仅取非负整数值时 $P(X=0)=\lim_{s\to-\infty}M(s)$.

4° 可逆性：若存在一正数 a，使得对 $\forall s\in[-a,a]$ 有 $M(s)<\infty$，则 $M(s)$ 唯一决定 X 的分布函数

5° 若 $X_1,\dots,X_n$ 相互独立，$M_{X_i}(s)=E(e^{sX_i})$ 是 X_i 的矩母函数，则

$Y=X_1+\dots+X_n$ 的矩母函数为 $M_Y(s)=M_{X_1}(s)\cdots M_{X_n}(s)$.

· PGF与MGF对比 （非负整数） （大部分r.v.）

$B(n,p)$ < pgf: $(q+ps)^n$；mgf: $(pe^s+q)^n$

$P(\lambda)$ < pgf: $e^{\lambda(s-1)}$；mgf: $e^{\lambda(e^s-1)}$

	PGF（非负整数）	MGF（大部分r.v.）	
定义	$ES^X=p_0s^0+p_1s^1+\dots$	$Ee^{sX}=\sum_{n=0}^{\infty}\frac{EX^n}{n!}s^n$	
可逆性	pgf ↔ pmf	（混合分布）mgf ↔ cdf	（注意 s 到 e^s 的替换）
$aX+b$	$\prod_{i=1}^{n} s^b g_X(s^a)$	$e^{sb}M_X(sa)$	$X_1+\dots+X_n$ (⊥)：$\prod_{i=1}^n M_i(s)$ 或 $\prod_{i=1}^n g_i(s)$.
$X_1+\dots+X_N$ (iid)	$g_N(s)$ 的 $s\to g_X(s)$	$M_N(s)$ 的 $e^s\to M_X(s)$	mgf有 $EX^k=M^{(k)}(0)$.
(X,Y)	$E(s^X,t^Y)$	$E(e^{sX},e^{tY})$.	

例 指数分布 $\mathcal{E}(\lambda)$　$f(x)=\lambda e^{-\lambda x}I_{(x\ge0)}$，求矩母函数 $M(s)$，EX，二阶矩 EX^2.

$s<\lambda$ 时有 $M(s)=\lambda\int_0^\infty e^{sx}e^{-\lambda x}dx=\lambda\frac{e^{(s-\lambda)x}}{s-\lambda}\Big|_0^\infty=\frac{\lambda}{\lambda-s}$.（$s\geq\lambda$ 时无意义）

$EX=M^{(1)}(0)=\frac{1}{\lambda}$，$EX^2=M^{(2)}(0)=\frac{2}{\lambda^2}$.

例（混合分布）银行的3位交易员为顾客服务时间均服从指数分布. 两位 $\lambda=6$，一位 $\lambda=4$.

随机选择一名交易员，求接受服务时间T的矩母函数.

$f_T(t)=(\frac{2}{3}6e^{-6x}+\frac{1}{3}4e^{-4x})$，$x>0$，$\frac{2}{3}$ 和 $\frac{1}{3}$ 表示概率混合而不是单次选择的时间分布.

$M_T(s)=\int_0^\infty e^{st}f(t)\,dt=\frac{2}{3}\frac{6}{6-s}+\frac{1}{3}\frac{4}{4-s}=\frac{2}{3}M(s)_{X_1}+\frac{1}{3}M(s)_{X_2}$.

反演：若已知 X 的矩母函数形如 $p_1M_{X_1}(s)+p_2M_{X_2}(s)$，则 X 是 X_1、X_2 的混合变量，

X_1、X_2 被选中概率分别为 p_1、p_2，$f_X(x)=p_1f_{X_1}(x)+p_2f_{X_2}(x)$.

例（正态分布）$Z=X+Y$，$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，相互独立. 求Z的pdf.

若 $V\sim N(0,1)$，$M_V(s)=e^{\frac{s^2}{2}}$；若 $W\sim N(\mu,\sigma^2)$，由线性 $M_W(s)=e^{\mu s+(\sigma^2s^2/2)}$

$\Rightarrow M_Z(s)=M_X(s)M_Y(s)=e^{(\mu_1+\mu_2)s+[(\sigma_1^2+\sigma_2^2)s^2/2]}$，$Z\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

可见独立的正态分布函数之和依然为正态分布.

· 若 $X\sim N(0,1)$，则 $EX^n=\begin{cases}0, & n\text{为奇数}\\ \frac{n!}{2^{\frac{n}{2}}(\frac{n}{2})!} & n\text{为偶数}\end{cases}$　抽象表述（相当于 $n!!$ 双阶乘）

Date. /

随机数个随机变量和的矩母函数

设$\{X_i\}$ iid，有共同矩母函数$M_X(s)$，N为取正整数值的随机变量，且独立于$\{X_i\}$。令$Y=\sum_{i=1}^{N}X_i$，$M_Y(s)=E(e^{sY})=E(E(e^{sY}|N=n))=E(M_X(s)^n)=\sum_{n=1}^{\infty}M_X^n(s)P(N=n)$。

$(M_N(s)=\sum_{n=1}^{\infty}(e^s)^n P(N=n)$，把$e^s$换成$M_X(s)$。相当于pgf $\xrightarrow{(x)}$ mgf $\xrightarrow{(e^x)}$ $M_X(s)$)。

例 $X_i\sim\mathcal{E}(\lambda)$，$N\sim G(p)$。$M_N(s)=\frac{pe^s}{1-(1-p)e^s}$，$M_X(s)=\frac{\lambda}{\lambda-s}$。

$M_Y(s)=\frac{pM_X(s)}{1-(1-p)M_X(s)}=\frac{p\lambda}{p\lambda-s}$，即$Y\sim\mathcal{E}(p\lambda)$。若$Y=X_1+\dots+X_n$，$M_Y(s)=[M_X(s)]^n$。

随机向量$\vec{X}=(X_1,\dots X_n)$对应的多元矩母函数$M_X(\vec{s})=E(e^{s_1X_1+\dots+s_nX_n})$。

特征函数

(定义) 对随机变量X，称$\varphi(t)=E(e^{itX})=E\cos(tX)+iE\sin(tX)$，$(x\in\mathbb{R})$

为X的特征函数(characteristic function, cf)，$i^2=-1$。

性质 设$\varphi(t)=E(e^{itX})$，1° $|\varphi(t)|\le\varphi(0)=1$，$\varphi(-t)=\overline{\varphi(t)}$。

2° $\varphi(t)$在$(-\infty,\infty)$上一致连续，即$|\varphi(t+h)-\varphi(t)|\le E|e^{ihX}-1|$。

3° 若$E(|X|^k)<\infty$，则$\varphi^{(k)}(t)=i^kE(X^ke^{itX})$，$\varphi^{(k)}(0)=i^kE(X^k)$。

4° 对任意常数a, b，有$\varphi_{aX+b}(t)=e^{itb}\varphi_X(at)$。

5° 若X_k有特征函数$\varphi_k(t)$，且$\{X_k\}$相互独立，则$Y=X_1+\dots X_n$的特征函数为$\varphi_Y(t)=\prod_{k=1}^{n}\varphi_k(t)$。

(可以把i看成未知常量，运算结束后代入复数)。相比矩母函数：特征函数一定存在。

- 二项分布$B(n,p)$ $\varphi(t)=\sum_{j=0}^{n}e^{itj}\binom{n}{j}p^jq^{n-j}=(q+pe^{it})^n$。
- 泊松分布$P(\lambda)$ $\varphi(t)=\sum_{k=0}^{\infty}e^{itk}\frac{\lambda^k}{k!}e^{-\lambda}=e^{\lambda(e^{it}-1)}$。

(对比三种母函数) (对比后，一般用矩母函数)

	pgf	mgf	cf
$B(n,p)$	$(q+ps)^n$	$(q+pe^s)^n$	$(q+pe^{it})^n$
$P(\lambda)$	$e^{\lambda(s-1)}$	$e^{\lambda(e^s-1)}$	$e^{\lambda(e^{it}-1)}$
$G(p)$	$\frac{sp}{1-sq}$	$\frac{e^sp}{1-e^sq}$	$\frac{e^{it}p}{1-e^{it}q}$

- 指数分布$\mathcal{E}(\lambda)$ $\varphi(t)=(1-\frac{it}{\lambda})^{-1}$。
- 均匀分布$U(a,b)$ $\varphi(t)=\frac{e^{itb}-e^{ita}}{it(b-a)}$，特别地$U(-c,c)$有$\varphi(t)=\frac{\sin ct}{ct}$
- 正态分布$N(\mu,\sigma^2)\to\varphi(t)=\exp(i\mu t-\frac{1}{2}\sigma^2t^2)$ $N(0,1)\to\varphi(t)=e^{-\frac{1}{2}\sigma^2t^2}=e^{-\frac{1}{2}t^2}$

Date. /

(正态分布特征函数运算) 先考虑 $N(0,1)$: $\phi(t)=E(e^{itX})=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{itx}e^{-\frac{x^2}{2}}dx$

$\phi(t)=e^{-\frac{t^2}{2}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{(x-it)^2}{2}}dx=e^{-\frac{t^2}{2}}$ (积分内为正态分布)

或者 $\phi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\cos(tx)e^{-\frac{x^2}{2}}dx \Rightarrow \phi'(t)=-t\phi(t)$, $\phi(0)=1 \Rightarrow \phi(t)=e^{-\frac{t^2}{2}}$.

对 $Y\sim N(\mu,\sigma^2)$: $X\sim N(0,1)$, $Y=\mu+\sigma X$. $E(e^{itY})=e^{it\mu}E(e^{it\sigma X})=\exp\{i\mu t-\frac{1}{2}\sigma^2t^2\}$.

设 $X_1\dots X_m$ 相互独立, $X_i\sim N(\mu_i,\sigma_i^2)$, 则 $Y=X_1+\dots X_m\sim N(\sum_{j=1}^m\mu_j,\sum_{j=1}^m\sigma_j^2)$.

- Cauchy 分布 $f(x)=\frac{1}{\pi(1+x^2)}$, $\varphi(t)=e^{-|t|}$.
- Laplace 分布 $f(x)=\frac{1}{2}e^{-|x|}$, $\varphi(t)=\frac{1}{1+t^2}$.

} 可直接使用,有对应关系.

特征函数一定存在且可逆的直观理解: 相当于将分布函数改变了坐标系.

随机变量的特征函数与分布函数相互唯一确定, 且均一定存在.

(定理) 逆转公式: 设 $\varphi(t)$ 是 X 的特征函数, $F(x)$ 是 X 的分布函数, 若 $F(x)$ 在 (a,b) 连续, 则 $\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)dt=F(b)-F(a)$.

(定理) 若 X 的特征函数满足 $\int_{\mathbb{R}}|\varphi(t)|dt<\infty$, 则 X 有连续密度函数 $f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\varphi(t)dt$.

(定理) 若 $E|X|^n<\infty$, 则 X 的特征函数 $\varphi(t)$ 满足: $\varphi(t)=\sum_{m=0}^{n}\frac{E[(itX)^m]}{m!}+o(t^n)$.

特别地, 若 $E(X^2)<\infty$, 则 $\varphi(t)=1+itEX-\frac{1}{2}t^2EX^2+o(t^2)$.

(矩不一定存在, 如 Cauchy 分布) 如果各阶矩都存在, 由 Taylor 展开, 唯一确定了分布.

区分: $EXY=EXEY \not\Rightarrow X\perp Y$, $E(X|Y)=C \not\Rightarrow X\perp Y$. 和只看一阶矩相同, 只考虑了一阶的独立性. 各阶矩都存在时才能唯一决定特征函数.

(定理) 混合分布: 设分布函数 $F_1(x),\dots F_m(x)$ 的特征函数分别为 $\varphi_1(x),\dots\varphi_m(x)$, $\lambda_i\ge0$, 且 $\sum_{k=1}^m\lambda_k=1$, 则 $\sum_{k=1}^m\lambda_kF_k$ 有特征函数 $\sum_{k=1}^m\lambda_k\varphi_k(t)$.

注: $Y=\sum_{k=1}^m\lambda_kF_k$ 表示 Y 以概率 λ_k 等于 X_k (其分布为 F_k), 对 $\{X_m\}$ 之间的独立性无要求. 需要指出的是 $Y\neq\sum_{k=1}^m\lambda_kX_k$.

- (定义) 设 $\vec{X}=(X_1,\dots X_n)$ 为随机向量, $\vec{X}$ 的特征函数定义为 $\varphi(\vec{t})=E(e^{i\vec{t}\vec{X}^T})$, $\vec{t}=(t_1,\dots t_n)$.

(定理) $\vec{X}=(X_1,\dots X_n)$ 为随机向量, $X_1,\dots X_n$ 相互独立的条件是 (充要条件)

$\varphi(\vec{t})=\varphi_1(t_1)\varphi_2(t_2)\dots\varphi_n(t_n)$. $\varphi_k(t_k)$ 是 X_k 的特征函数.

- 易错题: $\varphi_{X+Y}(t)=\varphi_X(t)\varphi_Y(t)$, $\forall t\in\mathbb{R}$, 能否说明 $X\perp Y$. (不能)

反例: Cauchy 分布 $f(x)=\frac{1}{\pi(1+x^2)}$, $Y=aX$. $\Phi_X(t)=e^{-|t|}$, $\phi_{X+Y}(t)=e^{-|t|}\cdot e^{-a|t|}=\Phi_X(t)\Phi_Y(t)$.

理解: $\varphi_{X+Y}(t)$ 只考虑了 $X=Y(=t)$ 情况, 不能覆盖整个二维空间.

deli得力

Date. /

$\frac{1}{n}\sum_{i=1}^{n}x_n$ = 样本均值

大数定律

(定理) (Weak law of large numbers, WLLN) 设 $\{X_n\}$ 是独立同分布的随机变量序列，且 $var(X_1)<\infty$，则对 $\forall\varepsilon>0$，$\lim_{n\to\infty}P(|\frac{1}{n}\sum_{i=1}^{n}X_i-EX_1|\geq\varepsilon)=0$.

也称 $\frac{1}{n}\sum_{k=1}^{n}X_k$ 依概率收敛到 EX_1，记作 $\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{P}EX_1$.

n很大时 有很大把握(但非100%) 断言 $\frac{1}{n}\sum_{i=1}^{n}X_i$ 很接近 EX_1. (连接偶然性与必然性)

(证明) $\{X_n\}$ 独立同分布，$var(X_1)<\infty$. 要证：$\forall\varepsilon>0$，$\lim_{n\to\infty}P(|\frac{1}{n}\sum_{i=1}^{n}X_i-EX_1|\geq\varepsilon)=0$.

利用切比雪夫不等式：对随机变量X和 $\varepsilon>0$，$P(|X-EX|\geq\varepsilon)\leq\frac{var(X)}{\varepsilon^2}$. 由同分布，

* $P(|\frac{1}{n}\sum_{i=1}^{n}X_i-EX_1|\geq\varepsilon)=P(|\frac{1}{n}\sum_{i=1}^{n}X_i-E(\frac{1}{n}\sum_{i=1}^{n}X_i)|\geq\varepsilon)\leq\frac{1}{\varepsilon^2}var(\frac{1}{n}\sum_{i=1}^{n}X_i)$,

$\frac{1}{\varepsilon^2}var(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n\varepsilon^2}var(X_1)\to0$, $n\to\infty$.

(定义) 若对 $\forall\varepsilon>0$，有 $\lim_{n\to\infty}P(|Y_n-Y|\geq\varepsilon)=0$，则称 Y_n 依概率收敛到 Y (convergence in probability)，记作 $Y_n\xrightarrow{P}Y$.

应用：(疫苗有效性问题) 设人群中疫苗有效的比例为p. 随机抽n个人，X_i 表示对第i个人有效. 样本中有效比例 $M_n=\frac{1}{n}\sum_{i=1}^{n}X_i$. 需抽多少人能保证误差大于3%的概率不超过5%，即 $P(|M_n-p|\geq0.03)\leq0.05$. 由上，$P(|M_n-p|\geq0.03)\leq\frac{var(X_i)}{n\varepsilon^2}\leq0.05$.

$var(X_i)=p(1-p)\leq\frac{1}{4}$. 故计算得取 $n=5556$ 时可行.

放宽条件：
- 独立同分布 (马尔可夫大数定律) $\{X_n\}$ iid → $\{X_n\}$ 同分布 → 无约束
- 方差有限 (辛钦大数定律) $var(X)<\infty\to E|X|<\infty\to nP(|X|\geq n)\to0$

定理扩展1 (Khinchin) $\{X_n\}$ 为iid且 $E|X_1|<\infty$，则 $\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{P}EX_1$.

扩展2 $\{X_n\}$ 是iid序列，则存在实数列 $\{a_n\}$ 使得 $\frac{1}{n}\sum_{k=1}^{n}X_k-a_n\xrightarrow{P}0$ 的充要条件是 $nP(|X_1|\geq n)\to0$. 此时可取 $a_n=E(X_1 I_{|X_1|<n})$.

· 强大数定律 (Strong law of large numbers, SLLN) 设 $\{X_n\}$ 是独立同分布的随机变量序列：1° 若 $E|X_1|<\infty$，则 $\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{a.s.}EX_1$ 几乎处处收敛.

2° 反之，若 $\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{a.s.}C$ 几乎处处收敛，则 $E|X_1|<\infty$ 且 $C=EX_1$.

以上所述为充要条件. 几乎处处收敛 → 以概率1收敛，$P(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k=EX_1)=1$.

(定义) 若 $P(\lim_{n\to\infty}Y_n=Y)=1$，则称 Y_n 几乎处处收敛到 Y (almost sure convergence)，记作 $Y_n\xrightarrow{a.s.}Y$ 或 $Y_n\to Y$ a.s.

$\{\lim_{n\to\infty}Y_n=Y\}=\{\omega:\lim_{n\to\infty}Y_n(\omega)=Y(\omega)\}\to\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\{\omega:|Y_k(\omega)-Y(\omega)|<\varepsilon\}=\lim_{n\to\infty}\inf A_n$ (下极限)

$A_n=\{|Y_n-Y|<\varepsilon\}$.

几乎处处收敛 可以推出依概率收敛，即 $P(\lim\limits_{n\to\infty} Y_n = Y) = 1 \Rightarrow \lim\limits_{n\to\infty} P(|Y_n - Y| > \varepsilon) = 0$.

对 $\forall \varepsilon > 0$，令 $A_n = \{\omega \mid |Y_n(\omega) - Y(\omega)| < \varepsilon\} = \{|Y_n - Y| < \varepsilon\}$.

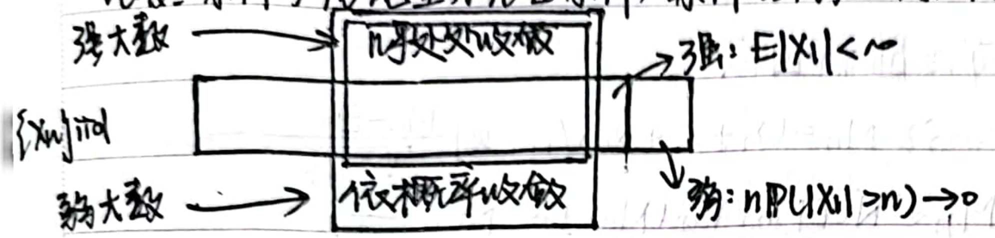
则 $\lim\limits_{n\to\infty} P(|Y_n - Y| \geq \varepsilon) = \lim\limits_{n\to\infty} P(A_n^c) = \lim\limits_{n\to\infty}(1 - P(A_n)) = 1 - \lim\limits_{n\to\infty} P(A_n)$.

故 $\lim\limits_{n\to\infty} P(|Y_n - Y| \geq \varepsilon) = 0 \Leftrightarrow$

(弱条件对应弱结论) $nP(|X| \geq n) \to 0 \Leftarrow E|X| < \infty$. (期望收敛是强的条件)

因为 $E|X| = \sum_k kP(|X|=k) \Rightarrow \sum_k kP(|X|=k) \to 0 \geq \sum_k nP(|X|=k) = nP(|X| \geq n)$.

比较：条件与结论互为充要条件，条件强弱不同，结论也强弱不同。



直观：采样次数越多，样本均值几乎一定渐近接近总体均值。

样本均值接近总体均值的可能性越来越大

(反例) $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ 服从弱大数定律，但不服从强大数定律，当 r.v. 序列 X，$X_1,\dots,X_n$ 独立同分布，$P(X=m) = P(X=-m) = \frac{C}{2m^2 \ln m}$，$m \geq 3$，$C$ 为常数。

理解：弱大数定律的充要条件为 $\lim\limits_{n\to\infty} nP(|X| \geq n) = 0$. 显然，

$$\lim_{n\to\infty} nP(|X| \geq n) = \lim_{n\to\infty} n\sum_{m=n}^{\infty} P(|X| = m).$$

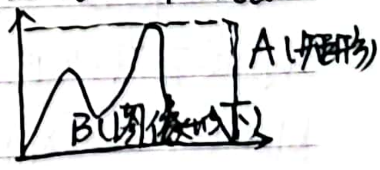
· 经验分布 设 $\{X_j\}$ 是概率空间 $(\Omega, \mathcal{F}, P)$ 上 iid 随机序列，用 x_j 表示 X_j 的观测值，即对某个确定的 $\omega \in \Omega$，有 $x_j = X_j(\omega)$. 则以概率1，$\{x_j\}$ 观测数列可以决定 X_j 的分布函数 $F(x)$.

(证明) 记 I_j 表示 $X_j \leq x$. $\{X_j\}$ 为 iid. $F(x) = P(X \leq x) = EI_j$. 考虑

$\frac{1}{n}\sum_{i=1}^n I_i \xrightarrow{a.s.} EI_i$，由大数定律，$\lim\limits_{n\to\infty} \frac{1}{n}\sum_{j=1}^n I_j = EI_j = EI_1 = P(X_1 \leq x) = F(x)$. a.s.

即以概率1 有 $\lim\limits_{n\to\infty} \frac{1}{n}\sum_{j=1}^n I_j = F(x)$. $\left(F_n(x) = \frac{1}{n}\sum_{i=1}^n I_i\right)$

· 例 Monte Carlo 蒙特卡洛方法：估计某复杂有界函数 $f(x)$ 在区间 $[a,b]$ 上的积分 $\int_a^b f(x)dx$.



从A中随机抽取n个点 $(X_1, Y_1) \dots (X_n, Y_n)$. I_j 表示 (X_j, Y_j) 在B中. 则以概率1 有 $\frac{1}{n}\sum_{j=1}^n I_j \Rightarrow EI_j = P(I_j = 1) = \frac{\int_a^b f(x)dx}{c(b-a)}$.

那么 $\int_a^b f(x)dx \approx c(b-a)\frac{1}{n}\sum_{j=1}^n I_j$，即可以用随机取样反映积分。

(随机模拟方法，用随机反映必然性)

Date.　　/

中心极限定理

(定理) Lindeberg - Lévy 中心极限定理

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

记$S_n=X_1+\dots+X_n$，则$\frac{S_n-E(S_n)}{\sqrt{Var(S_n)}}\xrightarrow{d}N(0,1)$.

注：$\forall x$，$F(x)$连续，$\lim\limits_{n\to\infty}F_n(x)=F_Y(x)$，则可记$F_n(x)\xrightarrow{d}F_Y(x)$，$Y_n\xrightarrow{d}Y$.（依分布收敛）（对每个x都收敛）

即$Y_n\approx N(0,1)$，$aY_n+b\approx N(b,a^2)$.

n充分大时提供近似分布$S_n\approx N(ES_n,Var(S_n))\approx N(n\mu,n\sigma^2)$.

直观解释：大量独立同分布的随机因素

· 定理变形：样本均值解释．记$M_n=(X_1+\dots+X_n)/n$，则$\frac{M_n-E(M_n)}{\sqrt{Var(M_n)}}\xrightarrow{d}N(0,1)$.

n充分大时提供近似分布$M_n\approx N(EM_n,Var(M_n))\approx N(\mu,\frac{\sigma^2}{n})$.

(证明) 进行标准化，令$Y_k=\frac{X_k-\mu}{\sigma}$．$\{Y_k\}$ iid，$EY_k=0$，$Var(Y_k)=1$，要证：$\frac{1}{\sqrt n}\sum_{k=1}^n Y_k\xrightarrow{d}N(0,1)$

依分布收敛等价于：$\lim\limits_{n\to\infty}\phi_n(t)=\phi_Y(t)$.

只需证：$\frac{1}{\sqrt n}\sum_{k=1}^n Y_k$的特征函数$\to e^{-\frac{t^2}{2}}$，$n\to\infty$.

由于特征函数$\begin{matrix}ax\to\phi(ax)\\ x\to\phi(x)\end{matrix}$，$\frac{1}{\sqrt n}\sum_{k=1}^n Y_k$特征函数为$[\phi(\frac{t}{\sqrt n})]^n$．$\frac{t}{\sqrt n}\to 0$，

$\phi(t)=Ee^{itY_1}$．$\lim\limits_{n\to\infty}\ln[\phi(\frac{t}{\sqrt n})]^n=\lim\limits_{x\to 0}\frac{\ln[\phi(tx)]}{x^2}\xrightarrow[\text{法则}]{\text{洛必达}}\frac{t^2}{2}\phi''(0)$.

$\phi''(0)=i^2EY^2=-EY^2$（二阶矩，为1）　则$[\phi(\frac{t}{\sqrt n})]^n\to e^{-\frac{t^2}{2}}$.

· 连续性定理 (Continuity theorem)（最常用、最重要定理之一）

X_n依分布收敛到$X\Leftrightarrow\lim\limits_{n\to\infty}\phi_n(t)=\phi(t)$，$\forall t\in\mathbb{R}$.

(定义) 若在$F(x)$的连续点x，有$\lim\limits_{n\to\infty}F_n(x)=F(x)$，则称$X_n$依分布收敛到X (convergence in distribution)，记作$X_n\xrightarrow{d}X$；或称F_n弱收敛到F (weak convergence)，记作$F_n\xrightarrow{w}F$.

· 例 疫苗有效性问题（更新）接上页计算的n=5556结果．

$M_n=(X_1+\dots+X_n)/n$，人群中疫苗有效的比例设为μ．现在希望：

求适当的n以保证$\mathbb{P}(|M_n-\mu|<0.03)\geq 0.95$．$\sigma=\sqrt{Var(X_1)}$，

$\mathbb{P}=\mathbb{P}(|\frac{M_n-\mu}{\sigma/\sqrt n}|<\frac{0.03}{\sigma/\sqrt n})\approx\mathbb{P}(|Z|<\frac{0.03\sqrt n}{\sigma})$，$\sigma\leq\frac12$，取$Z\sim N(0,1)$，

则$n=(\frac{1.96}{0.06})^2\approx 1068$即可．（$M_n\sim N(\mu,\frac{\sigma^2}{n})$）用大数定律是宽泛定范围，中心极限可以近似拟合正态分布 $(n\to\infty)$

使用中心极限定理必须说明条件：iid和方差有限.

理论近似：当前n项和计算复杂或无显示表达时

例：$\{X_n\}$ iid，$X_i \sim U(-1,1)$. $S_n = X_1 + X_2 + \cdots X_n$. 其密度函数

$$f_S(x) = \begin{cases} \frac{1}{2^n (n-1)!} \sum_{k=0}^{[\frac{n+x}{2}]} (-1)^k \binom{n}{k} (n+x-2k)^{n-1}, & |x| \le n \\ 0, & |x| > n \end{cases}$$

$S_n = nM_n \sim N(n\mu, n^2\frac{\sigma^2}{n})$（直接理解）

给出近似：$S_n \approx N(n\mu, n\sigma^2) = N(0, \frac{n}{3})$.（均匀分布 $\mu=0$，$\sigma^2=\frac{1}{3}$）

离散化修正：离散型随机变量上的应用

单点→区域，区域两端扩展. 以二项分布为例：

（定理）（de Moivre-Laplace，CLT）设 $S_n \sim B(n,p)$，n 充分大，$k, m \in N$，则

$$P(k \le S_n \le m) \approx \Phi\left(\frac{m-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right).$$ （Φ是标准正态cdf）

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

在计算单点概率时必须修正；其他情况下修正可以提高精度.

例（建造电影院）某地每日约有 $n=1600$ 人看电影，该地有一座旧电影院，预计建设完成后平均约有 $3/4$ 的观众将去新电影院. 建设规划座位数时要求座位尽可能多，但"空座多于或等于200"的概率不超过10%. 记座位数为m.

解：将每日看电影的观众按1~1600排序，$X_i=1$ 表示去新影院. $P(X_i=1)=\frac{3}{4}$，$P(X_i=0)=\frac{1}{4}$.

$\{X_i\}$ 为iid. 希望 $P(\sum_{i=1}^{1600} X_i \le m-200) \le 0.1$. 对 $\sum_{i=1}^{1600} X_i$ 用正态分布近似相当于

$N(\mu n, \sigma^2 n) \Rightarrow N(1200, 1200) \Rightarrow \Phi\left(\frac{m-200+0.5-1200}{10\sqrt{3}}\right) = 0.1$.（再用离散修正）

$\Phi(x)=0.1$，$x=-1.2816$，$m \approx 1377$.

- 从实例中可见：$n\to\infty$ 时正态近似越精确. 在实践中需要知道：样本量n多大时正态近似的结果可信. 这依赖于 X_i 是否接近正态分布，特别地，X_i 是否对称. 用正态近似计算 $P(S_n \le c)$ 时，若 c 在 S_n 均值附近，其精度会更高.

- 二项分布的泊松估计与正态估计

泊松近似：$np=\lambda \Rightarrow B(n,p) \to P(\lambda)$ $= X_{n1} + \cdots + X_{nn}$，$X_i \sim B(1,p)$，$\{X_{ni}\}$ iid.

条件：np 固定，$n\to\infty$，$p\to 0$.

正态近似：$B(n,p) = S_n \approx N(ES_n, Var(S_n)) = X_1 + \cdots + X_n$，$X_i \sim B(1,p)$ iid.

Date. /

条件：p固定，$n\to\infty$． $S_n=X_1+\dots+X_n\sim P(n)$， n很大可直接正态近似．

(对比) 若np固定且很大，比如$X_i\sim P(1)$， $S_n\approx N(ES_n, var(S_n))$．

例 $\{X_n\}$ iid，$X_n\sim P(1)$．$S_n=\sum_{i=1}^n X_i$．用CLT估计$\mathbb{P}(S_n=n)$．($EX=varX=\lambda$)

$ES_n=n$，$var(S_n)=\sqrt{n}$．$\mathbb{P}(S_n=n)\Rightarrow\mathbb{P}(n-0.5\le S_n\le n+0.5)\Rightarrow$

$=\mathbb{P}(\frac{n-0.5-n}{\sqrt{n}}\le\frac{S_n-n}{\sqrt{n}}\le\frac{n+0.5-n}{\sqrt{n}})=\Phi(\frac{0.5}{\sqrt{n}})-\Phi(\frac{-0.5}{\sqrt{n}})=2\Phi(\frac{0.5}{\sqrt{n}})-1$

由此 $\frac{n^n}{n!}e^{-n}=\mathbb{P}(S_n=n)\approx\int_{-\frac{1}{2\sqrt{n}}}^{\frac{1}{2\sqrt{n}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\approx\frac{1}{\sqrt{n}}\frac{1}{\sqrt{2\pi}}$．

得到stirling公式：$n!\approx\sqrt{2\pi n}(\frac{n}{e})^n$．

· 不是任意分布都能用中心极限定理，$X_1,\dots X_n$独立同分布且服从Cauchy分布．

记$M_n=\frac{1}{n}(X_1+\dots+X_n)$．$\Phi_X(t)=e^{-|t|}$，易证$\Phi_{M_n}(t)=e^{-|t|}$，$M_n$服从Cauchy分布．

说明：定理要求的"期望μ存在，方差$\sigma^2<\infty$"是必要条件．

收敛性总结 (依概率 / 几乎处处 / 依分布)

判断：已知两个r.v. X_1、X_2同分布，则$X_1=X_2$ (不一定正确)

不妨令X_1、X_2表示抛硬币．$X_1=\begin{cases}0 & H\\1 & T\end{cases}$，$X_2=\begin{cases}0 & T\\1 & H\end{cases}$．

$X_1(w=T)=1$而$X_2(w=T)=0$，则X_1、X_2依分布收敛，但显然不是几乎处处收敛．

也不是依概率收敛：显然不满足$\mathbb{P}(|X_n-X_2|\ge\varepsilon)\to0$．

收敛模式补充：L_p收敛 (不要求) 对$p>0$，若$\lim_{n\to\infty}E(|X_n-X|^p)=0$，

则称X_n在L_p下收敛到X，记作$X_n\xrightarrow{L_p}X$，或$X_n\to X$ in L_p．例如$\lim_{n\to\infty}E(|X_n-X|)=0\Leftrightarrow X_n\xrightarrow{L_1}X$．

各种收敛性之间关系： (又叫以概率1收敛)

$X_n\xrightarrow{a.s.}X$ (几乎处处收敛) $\Rightarrow$ $X_n\xrightarrow{P}X$ (依概率收敛) $\Rightarrow$ $X_n\xrightarrow{d}X$ (依分布收敛)

$X_n\xrightarrow{L_p}X$ (在L_p下收敛) $\Rightarrow$

依分布收敛是4种中最弱收敛．

依概率收敛得不到几乎处处收敛：$\mathbb{P}(\{w|\,|X_n(w)-X(w)|>\varepsilon\})\to0$．

$X(w)=0$，$X_1(w)=\begin{cases}1 & [0,\frac{1}{2}]\\0 & (\frac{1}{2},1]\end{cases}$ ……

也得不到L_1收敛．

收敛性质补充：考虑两个随机变量序列$\{X_i\}$和$\{Y_i\}$，X_n、Y_n分别依概率收敛，c为已知常数，则有：

1° cX_n、X_n+Y_n、$\max\{0, X_n\}$、$|X_n|$、X_nY_n都依概率收敛到各自极限。

2° 进一步，若$Y_n \neq 0$，则X_n/Y_n也依概率收敛到其对应极限。

几乎处处收敛和依概率收敛对细节有要求，但依分布收敛可互换细节，不影响整体

当收敛到常数时依概率与依分布收敛等价。

（定理）连续映射定理（Continuous mapping theorem）$\{X, X_n\}$是随机元序列，g连续，则

1° $X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$.　　2° $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$.

3° $X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$.　　（但是$X_n \xrightarrow{L^p} X \not\Rightarrow g(X_n) \xrightarrow{L^p} g(X)$）.

· Slutsky's theorem　设$X_n \xrightarrow{d} X$，$Y_n \xrightarrow{d} c$，则　（四则运算）

$X_n+Y_n \xrightarrow{d} X+c$，$X_nY_n \xrightarrow{d} cX$，$X_n/Y_n \xrightarrow{d} X/c\ (c\neq 0)$.

连续性定理 Continuity theorem：X的特征函数为$\phi(t)$，X_n的特征函数为$\phi_n(t)$，

* 则X_n依分布收敛到X的充要条件是$\lim_{n\to\infty}\phi_n(t) = \phi(t)$. $\forall t \in \mathbb{R}$.

例：T分布　$Z \sim N(0,1)$，$X \sim \chi^2_n$，X、Z独立．令$T = \frac{Z}{\sqrt{X/n}}$，则$T \sim t_n$.

对iid $Z_j \sim N(0,1)$，$X = Z_1^2 + \dots + Z_n^2$，$X \sim \chi^2_n$．证明$t_n \to N(0,1)$，$n \to \infty$：

（LLN）$\frac{1}{n}\sum_{i=1}^n Z_i^2 \xrightarrow{d/P} \cancel{\text{H}} EZ_i^2 = 1$.　（连续性定理）$\frac{1}{\sqrt{X/n}} \xrightarrow{d} \equiv 1$.

（Slutsky定理）$X_n/Y_n \xrightarrow{d} X/c \Rightarrow t_n \xrightarrow{d} Z = N(0,1)$.

例 二项分布用泊松分布近似用连续性定理证明：（不需要特征函数，可以取较简单的概率母函数/矩母函数近似）如下用概率母函数.

$B(n,p)$，$p = \frac{\lambda}{n}$．$\cancel{\lim_{n\to\infty}[(1+(s-1)\frac{\lambda}{n})}$ $\lim_{n\to\infty}(1-p_n+p_ns)^n = \lim_{n\to\infty}\left[(1+(s-1)\frac{\lambda}{n})^{\frac{n}{\lambda(s-1)}}\right]^{\lambda(s-1)}$

$= e^{\lambda(s-1)}$．取$np_n = \lambda$，$n\to\infty$时$B(n, p_n) \xrightarrow{d} P(\lambda)$.