二项分布近似　n很大，p很小 $\longrightarrow$ 泊松分布近似、

n很大，p很大

n很大，p不大不小？

赌场游戏　A.B两人赌博，获胜概率 p, $1-p$，赌n局

假设 A赢局数 $x > np$，付给赌场 $x-np$ 元

否则　B付赌场 $np-x$ 元

问赌场净收的期望？

求A赢局数的概率　$b(i;n,p)$

Stirling公式　$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

假设$p=\frac{1}{2}$　$b(i;n,p)=\binom{n}{i}\left(\frac{1}{2}\right)^i$

$$b\left(\frac{n}{2};n,p\right)=\frac{n!}{\left(\frac{n}{2}\right)!\left(\frac{n}{2}\right)!}\left(\frac{1}{2}\right)^{\frac{n}{2}} \approx \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi\frac{n}{2}}^{\,2}\cdot\left(\frac{n}{2e}\right)^n}\cdot\left(\frac{1}{2}\right)^{n} = \frac{\sqrt{2\pi n}}{\pi n\left(\frac{1}{2}\right)^n}\left(\frac{1}{2}\right)^{n} = \sqrt{\frac{2}{\pi n}}$$

对于二项分布而言，最大在$(n+1)\cdot p$附近

$$p(|x-np|\le t)$$

$$b\left(\frac{n}{2}+d;n,p\right) = b(m+d;n,p) = \frac{2m!}{(m+d)!\,(m-d)!}\left(\frac{1}{2}\right)^{2m} = \frac{2m!}{m!\,m!}\cdot\frac{\prod_{i=1}^{d}(m-i+1)}{\prod_{i=1}^{d}(m+i)}\left(\frac{1}{2}\right)^{2m}$$

令$m=\frac{n}{2}$

$$= b(m;n,p)\frac{\prod_{i=1}^{d}(m-i+1)}{\prod_{i=1}^{d}(m+i)}$$

$$= b(m;n,p)\frac{\prod_{i=1}^{d}\left(1+\frac{-i+1}{m}\right)}{\prod_{i=1}^{d}\left(1+\frac{i}{m}\right)}$$

$1+x\approx e^x$

$$= b(m;n,p)\, e^{\sum_{i=-d+1}\frac{i}{m}+\sum_{i=-d}^{-1}\frac{i}{m}} \approx b(m;n,p)\,e^{-\frac{d^2}{m}}$$

$$b(m+d;n,p)=\sqrt{\frac{2}{n\pi}}\,e^{-\frac{d^2}{m}}$$

$$p(|x-np|\le d) = p\left(\frac{n}{2}-d\le x\le\frac{n}{2}+d\right) = \sum_{i=-d}^{d} b\left(\frac{n}{2}+i;n,p\right) = \sum_{i=-d}^{d}\sqrt{\frac{2}{n\pi}}\,e^{-\frac{i^2}{m}}$$

假设$p=\frac{1}{2}$

$$= \int_{-d}^{d}\sqrt{\frac{2}{n\pi}}\,e^{-\frac{i^2}{m}}\,di \qquad \left(m=\frac{n}{2}\right)$$

令$d=c\sqrt{n}$

$$= \int_{-c\sqrt{n}}^{c\sqrt{n}}\sqrt{\frac{2}{n\pi}}\,e^{-\frac{2i^2}{n}}\,di$$

令$x=\frac{2i}{\sqrt{n}}$　$(-c\sqrt{n}\le i\le c\sqrt{n},\ -2c\le x\le 2c)$

$$= \int_{-c\sqrt{n}}^{c\sqrt{n}}\sqrt{\frac{2}{n\pi}}\,\frac{\sqrt{n}}{2}\cdot e^{-\frac{2i^2}{n}}\,d\frac{2i}{\sqrt{n}}$$

$$= \int_{-2c}^{2c}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}\,dx$$

正态分布

定义 (Normal distribution)

设X是连续型随机变量，若X的概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\},-\infty<x<\infty$$

（其中$-\infty<\mu<\infty,\sigma>0$为常数）则称$X$服从参数为$\mu,\sigma^2$的正态分布（或高斯分布），记为 $X\sim N(\mu,\sigma^2)$，相应的概率密度为正态密度。

概率分布函数 $F(x)=\int_{-\infty}^{x} f(y)\,dy$

① 决定某一数量指标的随机因素很多，且相互独立，每个起的作用都不大 （误差分布，导致误差的因素很多且作用都不大）

② 许多分布可用正态分布来逼近
可用正态分布来导出

性质：① 对称性：关于 $x=\mu$ 对称

$F(\mu)=1-F(\mu)\Rightarrow F(\mu)=\frac{1}{2}$

$F(\mu-x)=$ ~~$1-F(\mu+x)=F(\mu+x)$~~ $1-F(\mu+x)$

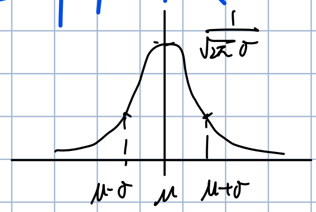
② 最大值 $\frac{1}{\sqrt{2\pi}\sigma}$

拐点 $\mu\pm\sigma$

渐近线 x轴

③ μ变化，左右移动

σ越大，越矮胖；σ越小，越瘦高



正态分布的3σ原则

$$P(-\sigma\le X-\mu\le\sigma)\approx 0.683$$
$$P(-2\sigma\le X-\mu\le 2\sigma)\approx 0.954$$
$$P(-3\sigma\le X-\mu\le 3\sigma)\approx 0.997$$

落在3σ之外的属于小概率事件，可以认为试验结果X在3σ之间

正态分布标准化

标准正态分布 $N(0,1)$

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$
$$\Phi(x)=\int_{-\infty}^{x}\varphi(y)\,dy=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy$$

一般正态分布 $X\sim N(\mu,\sigma^2)$ 分布函数 $F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-\mu)^2}{2\sigma^2}}dy$

令 $\frac{y-\mu}{\sigma}=t$

$$=\int_{-\infty}^{\frac{x-\mu}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$
$$=\Phi\left(\frac{x-\mu}{\sigma}\right)$$
任何一个正态分布都可以用标准正态分布来计算

正态分布概率计算公式 $P(a<X\le b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$

例2.3.10 (1) $P(400<X<700) = \Phi\left(\frac{700-600}{150}\right) - \Phi\left(\frac{400-600}{150}\right)$

$= \Phi\left(\frac{2}{3}\right) - \underline{\Phi\left(-\frac{4}{3}\right)}$

$1-\Phi\left(\frac{4}{3}\right)$

书上正态分布只有一半 (0.3)

但由正态分布对称性有 $\Phi(-x) = 1-\Phi(x)$

(2) $P(X>300) = 1-\Phi\left(\frac{300-600}{150}\right) = 1-\Phi(-2) = \Phi(2)$

(3) $P(X<a) = 0.1$

$= \Phi\left(\frac{a-600}{150}\right)$

表上没有0.1，查0.9

$1-\Phi\left(\frac{a-600}{150}\right) = 0.9$

$= \Phi\left(-\frac{a-600}{150}\right)$

定义 (2.3.6 上α分位数)

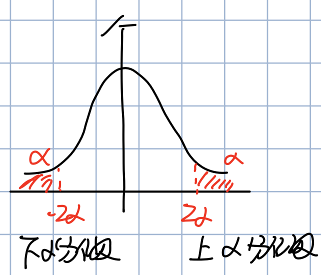
设$X \sim N(0,1)$，对给定的$\alpha(0<\alpha<1)$，若数z_α满足条件

$$P(X>z_\alpha) = \int_{z_\alpha}^{\infty} \phi(x)dx = \alpha$$

即

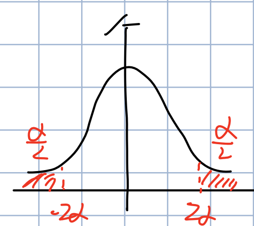
$$\Phi(z_\alpha) = 1-\alpha$$

则称z_α为标准正态分布的上α分位数。



定义 (双侧α分位数)

对于给定的$0<\alpha<1$，求出常数c_α使得 $P(|X|>c_\alpha) = \alpha$ 称c_α为双侧α分位数。



双侧α分位数

$C_\alpha = Z_{\frac{\alpha}{2}}$

Γ分布

定义 (Γ分布)

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x>0 \\ 0, x \le 0 \end{cases}$$

记$X \sim \Gamma(\alpha,\lambda)$。当$\alpha = n$为自然数时， 又称爱尔兰分布。

$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$

特殊取值: $\Gamma(1)=1$, $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$, $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$

特殊情况) 当$\alpha=1$时. $f(x) = \lambda e^{-\lambda x}$, 称为指数分布

当$\alpha=\frac{n}{2}$, $\lambda=\frac{1}{2}$时, 称为χ^2分布

例子

$P(-1<X<1) = \frac{5}{8}$

$P(-1<X\le x \mid -1<X<1) = \frac{x+1}{2}$

$= \frac{P(-1<X\le x)}{P(-1<X<1)} \Rightarrow P(-1<X\le x) = \frac{5}{8}\cdot\frac{x+1}{2}$

$x<-1$ $F(x)=0$

$-1\le x<1$ $F(x) = P(X=-1) + P(-1<X\le x) = \frac{1}{8} + \frac{5}{16}(x+1)$

$x\ge 1$ $F(x)=1$

§ 2.4 随机变量的函数及其分布

① X 是离散，$Y=g(X)$ 是离散

② X 连续，$Y=g(X)$ 都有可能

离散型随机变量函数的分布

若离散随机变量X的分布列为

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

则$Y=g(X)$也是一个离散型随机变量，其分布列可以表示为

Y	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_n)$	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

其中某些$g(x_i)$的取值相等时，将相等的值合并，并将对应的概率相加。

连续型随机变量函数的分布

1° $g(x)$ 严格单调　　$f(x)=f_X(x)$　　$-\infty<x<\infty$　　$a=\min\{g(-\infty),g(\infty)\}$　　$y\le a$　$F_Y(y)=0$

① 设 $g(x)\uparrow$，反函数为 $h(y)$ 严格单调可有反函数　　$b=\max\{g(-\infty),g(\infty)\}$　　$y\ge b$　$F_Y(y)=1$

$a<y<b$，$F(Y)=P(Y\le y)=P(g(x)\le y)=\int_{g(x)\le y} f_X(x)dx$

$=P(X\le h(y))=\int_{-\infty}^{h(y)} f_X(x)\,dx$　$\Rightarrow f_Y(y)=\begin{cases} f_X(h(y))\,h'(y) & y\in(a,b) \\ 0 & \text{else}\end{cases}$

对 y 求导 $\left(\int_{\Phi(y)}^{\varphi(y)} f(x)\,dx\right)'=f(\varphi(y))\cdot\varphi'(y)-f(\Phi(y))\cdot\Phi'(y)$

② $g(x)\downarrow$

$a<y<b$　$F_Y(y)=P(g(x)\le y)=P(X\ge h(y))=\int_{h(y)}^{\infty} f_X(x)\,dx\Rightarrow f_Y(y)=\begin{cases} -f_X(h(y))\cdot \underline{h'(y)}_{<0} & y\in(a,b) \\ 0 & \text{else}\end{cases}$

综合，有 $\begin{cases} f_X(h(y))\cdot|h'(y)|, & y\in(a,b) \\ 0 & \text{else}\end{cases}$

对于例 2.4.3　$Y=aX+b$

$f_Y(y)=f_X\left(\frac{y-b}{a}\right)\cdot\frac{1}{|a|}$

若 $X\sim N(\mu,\sigma^2)$，$Y=aX+b$

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}}\frac{1}{|a|}$$

$$=\frac{1}{\sqrt{2\pi}\,|a|\cdot\sigma}e^{-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}}\qquad Y\sim N(a\mu+b,\ a^2\sigma^2)$$

任何正态分布 的线性分布 也是正态分布

也可以用这个结论，进行正态分布的标准化

$X\sim N(\mu,\sigma^2)$　$Y=\frac{X-\mu}{\sigma}\sim N(0,1)$

$Y=aX+b$　$\begin{cases} a=\frac{1}{\sigma} \\ b=-\frac{\mu}{\sigma}\end{cases}$

例2.4.4 $X \sim N(\mu,\sigma^2)$，求 $Y=e^X$ $y\in(0,\infty)$

$y\le 0$，$f_Y(y)=0$

$y>0$ $f_Y(y)=f_X(\ln y)\frac{1}{|y|}=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}\frac{1}{|y|}$ 对数正态分布

$P(Y\le y)=P(e^X\le y)=P(X\le \ln y)=\Phi\left(\frac{\ln y-\mu}{\sigma}\right)$ 所有对数正态分布都可由标准正态分布计算

当g(x)为其它形式

例2.4.5 $Y=cX^2$

$$F_Y(y)=P(Y\le y)=P(cX^2\le y)=P\left(-\sqrt{\tfrac{y}{c}}\le X\le\sqrt{\tfrac{y}{c}}\right)=F_X\left(\sqrt{\tfrac{y}{c}}\right)-F_X\left(-\sqrt{\tfrac{y}{c}}\right)$$

$-\infty<x<\infty$

$0\le y<\infty$

由分布函数求概率密度函数：得 $f_Y(y)=f_X\left(\sqrt{\tfrac{y}{c}}\right)\left(\sqrt{\tfrac{y}{c}}\right)'-f_X\left(-\sqrt{\tfrac{y}{c}}\right)\cdot\left(-\sqrt{\tfrac{y}{c}}\right)'$

$$=\frac{1}{2\sqrt{cy}}\left(f_X\left(\sqrt{\tfrac{y}{c}}\right)+f_X\left(-\sqrt{\tfrac{y}{c}}\right)\right)\quad y\ge 0$$

2:当$g(X)$为其它形式时

分布函数微分法

- 化简$Y=g(X)$的分布函数$F_Y(y)=P(g(X)\le y)$，直到看出除有限多个点外导数$F'_Y(y)$存在且连续为止
- 求导数$F'_Y(y)$，令

$$f_Y(y)=\begin{cases}F'_Y(y), & \text{如果存在}\\ 0, & \text{否则}\end{cases}$$

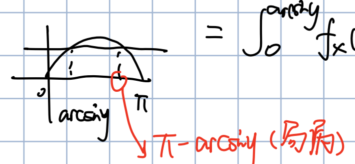
- 结论：$Y=g(X)$的概率密度即为$f_Y(y)$。

例2.4.6 $Y=\sin X$

$$F_Y(y)=P(Y\le y)=P(\sin X\le y)=P(0<X\le\arcsin y)+P(\pi-\arcsin y\le X<\pi)$$

$0<x<\pi$

$0<y<1$



$$=\int_0^{\arcsin y}f_X(x)\,dx+\int_{\pi-\arcsin y}^{\pi}f_X(x)\,dx$$

$$f_Y(y)=\begin{cases}\frac{2\arcsin y}{\pi^2}\frac{1}{\sqrt{1-y^2}}+\frac{2(\pi-\arcsin y)}{\pi^2}\frac{1}{\sqrt{1-y^2}}, & 0<y<1\\ 0 & \text{else}\end{cases}$$

例子

若随机变量X的分布函数$F_X(x)$为严格单调递增的连续函数，其反函数$F_X^{-1}(y)$ 存在，证明$Y = F_X(X)$ 服从$(0,1)$上的均匀分布$U(0,1)$。

Y为分布函数，取值范围在0-1之间

$y<0$，$F_Y(y)=0$

$y>1$，$F_Y(y)=1$

$0\le y<1$　$F_Y(y)=P(Y\le y)=P(F_X(X)\le y)=P(X\le F_X^{-1}(y))=F_X(F_X^{-1}(y))=y$

$Y=F_X(X)$ 服从 $U(0,1)$

任何连续随机变量都可通过其分布函数与均匀分布发生关系

随机数生成　$X\sim E(\lambda)$

$Y=F(X)=1-e^{-\lambda x}\sim U[0,1]$　反解出X，$1-e^{-\lambda x}=Y \Rightarrow X=-\frac{1}{\lambda}\ln(1-Y)$

从均匀分布随机数生成服从指数分布的随机数