# 概率统计B

## Probability and Statistics

张思容
zhangsirong@buaa.edu.cn

数学与系统科学学院，北京航空航天大学
School of Mathematics and System Sciences, BUAA

October 17, 2014

# Review：回顾

RECALL:

- 概率模型=样本空间+概率律
- 古典模型：一个因素影响下的概率模型。
- 复杂系统(或重复试验): 很多个(独立)因素影响下的概率模型?

TODAY

- 随机变量与分布函数;
- 二项分布。

Galton box: (bean machine)

# 随机变量的定义

什么是随机变量？ random variable

- 直接解释: 对每一个样本结果给出一个(实)数值。
- 数学解释；定义在一个样本空间上的函数。
  对比：函数的自变量与因变量。

**EXAMPLE (投硬币)**

投硬币三次，样本空间有*8*个元素。记每个结果中正面*(H)*的个数为$X$. $X$是个随机变量。

$X$取值$0, 1, 2, 3 \in \mathcal{R}$. 同理可定义随机变量$Y$为每个结果中反面的个数。

**Definition (定义:随机变量)**

随机变量$X$是定义在样本空间$\Omega$上的实值函数:即每一个样本点$\omega$,$X(\omega)$是个实数。 通常用$X, Y, Z$或$\xi, \eta$等表示。

例子：人体正常体温$T$ =多少摄氏度。华氏度$= 18/10c + 32$?

# 概率律的转移

## Remark (push-back)

随机变量引导概率律转移到实数上。
设$x \in \mathcal{R}$, $X^{-1}(x) = \{\omega_1, \omega_2, \dots\}$.
则$P(X = x) = \sum_{\omega_i} P(\omega_i), \omega_i \in \Omega$.

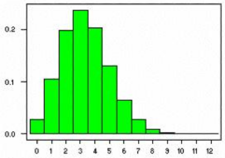
- $X$取值为离散的,称为离散随机变量, 概率律变成一个概率序列, 称为分布列;
- $X$取值包含连续区间, 概率律呢?
  注意:$P(X = x_0) = 0!$, 概率律变成一个(密度)函数。

## EXAMPLE (人体温度)

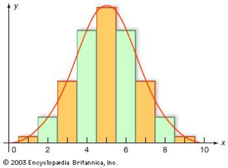
正常体温在$\mathcal{R}$上的有意义表示$36.8 \le X \le 37.2$(事件), 概率计算:
$P(36.8 \le X \le 37.2) = P(\{\omega : 36.8 \le X(\omega) \le 37.2\})$。
区间足够小? 得到一个密度函数。

# 分布函数的定义

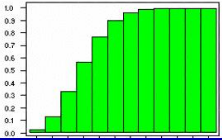
离散分布




连续分布




累积分布




**Definition (概率分布函数)**

给定随机变量$X$，定义函数

$$F(x) = P(X \leq x), -\infty \leq x \leq +\infty$$

称$F$为$X$的概率分布函数，简称分布函数或累积分布函数$CDF$。

分布函数的充分必要条件：

- $0 \leq F(x) \leq 1$.
- $F(x)$单调非减
- $F(x)$右连续；***

通常随机变量对应一些常见分布函数，直接称随机变量为某个分布。

# 随机变量的注解***

- 随机变量不仅仅是一个函数$X:\Omega \to \mathcal{R}$;
  更重要的是得到分布函数。可以记为一个变换$(\Omega, P) \to (\mathcal{R}, F)$。
  事实上分布函数更重要。
- 随机变量的函数还是随机变量(复合函数)。$\Omega \to \mathcal{R} \to \mathcal{R}$
- 区分分布函数可以用函数的数字特征：期望，方差等等。
- 随机变量分类：离散随机变量，非离散随机变量(我们一般仅仅考虑其中的连续随机变量);

# 分布函数的判断

分布函数的性质

- $F(-\infty) = 0, F(+\infty) = 1$
- $P(a \leq X \leq b) = F(b) - F(a)$
- $P(X = x) = F(x) - F(x^{-})$ ***

EXAMPLE

设$F(x) = a + be^{-x}, x \geq 0$, $F(x)$是否可以是某个随机变量的分布函数?特别求概率$P(X \leq \ln 2)$.

$a = 1, b = -1, P = 1/2$
说明： 有很多分布函数，但常见只有几种。

# 随机变量与决策

**EXAMPLE (轮盘赌 Roulette)**

美式轮盘赌有38个数值，最简单的玩法押一个数值一元，可赢得35元。问一次押注的期望赢钱多少？

解答： $EX = 35\frac{1}{38} + (-1)\frac{37}{38} = -1/19 \approx -0.053$

# 定义

**Definition (离散随机变量)**

随机变量取值为有限或可数个时，称为离散随机变量。
记其值为$x_1.x_2,\dots,x_k,\dots$。
特别记$P(X=x_k)=p_k, k=1,2,\dots,$ $P$是$X$对应的概率分布，称$\{p_k\}$为概率分布列。

- 常记为

| $X$ | $x_1$ | $x_2$ | $x_3$ | $\dots$ |
|---|---|---|---|---|
| $P$ | $p_1$ | $p_2$ | $p_3$ | $\dots$ |

- $p_k \geq 0, \sum p_k = 1.$
  注：$\Omega = \bigcup_i \{\omega | X(\omega) = x_i\}$
- 分布函数$F(x)$是阶跃函数。

# 二项分布

**EXAMPLE (两点分布，Bernouli贝努利分布)**

$X$取值为0或1时的概率分布是
$P(x=1)=p, P(X=0)=1-p,$ 记为$B(1,p)$.
模型：最简单随机变量=一次试验是否成功。

**Definition (二项分布或二项随机变量)**

随机变量的取值为$0,1,2,\ldots,N$,且满足
$P(X=k)=C_N^k p^k q^{N-k}, k=1,2,\ldots,N, p+q=1.$
称其满足二项分布，记为$B(N,p)$.

- 模型：$N$次独立试验中的成功次数。
- 称为二项分布因为其值为满足二项式定理的各项。
  $(p+q)^N=\sum C_N^k p^k q^{N-k}$,可递推计算,$N$充分大时,需近似计算;
- 应用： 投票权大小(委员会组成); 比赛规则(三局两胜,五局三胜?)...

# 应用例子

## EXAMPLE (基因遗传)

设某个生物特征(眼睛颜色，或左撇子)由一对基因决定。$D$为显性基因，$r$为隐性基因。后代从父母中各得到一个基因。只要有一个显性基因，就必然出现该生物特征。问一对混合型父母(基因为$Dr$)的四个后代中有三个有该生物特征的概率是多少?

解答： 服从二项$B(4,3/4)$，$P(X=3)=C_4^3(\frac{3}{4})^3\frac{1}{4}=\frac{27}{64}$.

## EXAMPLE (安全)

设飞机上每个引擎的失效概率相同为$p$且互相独立.如果一个飞机上的一半引擎正常，则飞机可以正常运行。问$p$为多少时，用四个引擎的飞机比用两个引擎安全?

解答： 四个引擎正常个数服从$B(4,p)$，飞机正常飞行即$X\geq 2$, $P(X\geq 2)=1-(1-p)^4-4p(1-p)^3$;类似两个引擎飞机正常飞行$Y\geq 1$,$P(Y\geq 1)=1-(1-p)^2$。计算可得$p\geq 2/3$时，四个引擎的飞机更安全

# $N$次独立试验中的分布

- 模型：$N$次独立试验 或$N$次有放回抽签。
- 二项分布: $X=k$,试验中的成功次数, $B(N,p)$
  $X=X_1+X_2+\cdots+X_N$,其中$X_i$是第$i$次试验是否成功的两点分布。
- 几何分布： $Y=k$,首次成功时的试验次数。
  [几何分布]: 随机变量取值为$1,2,\ldots,\infty$,且满足
  $P(Y=k)=p(1-p)^{k-1}, k=1,2,\ldots,\infty$.
- ***负二项分布： $Y=k$,第$r$次成功时的试验次数。$NB(r,p)$
  $Y=Y_1+Y_2+\cdots+Y_r$, $Y_i$一次成功时的试验次数。
  [负二项分布] 随机变量取值为$1,2,\ldots,\infty$,且满足
  $P(Y=k)=C_{k-1}^{r-1}p^r(1-p)^{k-r}, k=1,2,\ldots,\infty$.

***Banach火柴盒问题：有两个火柴盒(各有$N$个)，随机选取一个,直到发现一个盒子空了为止。问另外一个盒子有$k$个火柴的概率。
答案： 固定一个盒子： $Y\sim NB(N+1,1/2)$,试验次数$Y=2N+1-k$.
$P(Y=2N+1-k)=C_{2N-k}^{N}(1/2)^{2N-k+1}$.
两个盒子的概率为上面两倍。

# 无放回抽签的分布***

**Definition (超几何分布)**

随机变量取值为$0, 1, 2, \ldots, n$,且满足
$P(X=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k=1,2,\ldots,n;$
称其满足超几何分布，记为$H(n,M,N)$.

- 模型：从$(N,M)$大样本中抽取小样本$n$,$M$是$N$中特别一类(如次品),则$n$中次品个数$k$服从超几何分布。特别$n=1, k=1$有抽签原理!.
- ***较大$N,M$时,可用二项分布逼近。$B(n,p), p=M/N$,即有放回与无放回差别不大。
- 实用例子：估计某地区的某动物总数$N$。
  捕捉一批动物$M$，做标记放回， 过一段时间，再捕捉一批动物$n$,其中有标记的为$k$, 利用超几何分布，估计$N$。
  概率为$P(X=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$是关于$N$的一个序列，
  我们求最大值对应的$N$,计算有$N=nM/k$. 称为最大似然估计。

# 作业

北航教材:
P44 习题二. 1,4,6,8,12,13

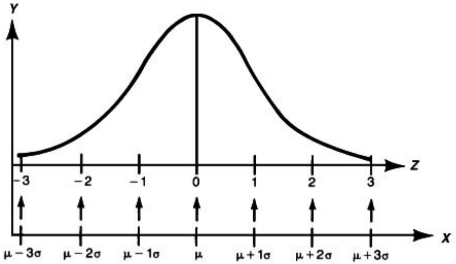
# Review：回顾

RECALL:

- 随机变量与分布函数
- 离散分布：两点分布，二项分布，几何分布，

TODAY

- 泊松分布；
- 连续随机变量：
- 常见连续分布：均匀分布，指数分布，正态分布；

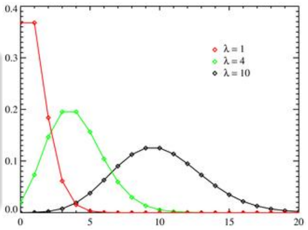



The Translation of $X$ to $Z$ by the Transformation $Z = (X-\mu)/\sigma$

*Figure 3*

正态分布，高斯分布，Bell曲线(钟型曲线)，

# 定义




**Definition (泊松分布)**

随机变量取值为$0, 1, 2, \ldots, k, \infty$,满足

$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \ldots; \lambda$为正常数.

称其满足泊松分布，记为$\mathcal{P}(\lambda)$.

- 模型：大量试验中的小概率事件发生次数。
  每年发生战争(或地震)的次数,某一小时进入某邮局的顾客数； 寿命超过100岁的人数；
- 逼近二项分布：当$n$大,$p$小，$np$合适大小时，可用$\lambda = np$的泊松分布近似计算。
  简单证明：
- 例子：设二项分布$B(100, 0.01)$，计算
  $P(X = 5) = C_{100}^5 \cdot 0.01^5 \cdot 0.99^{95} = 0.00290$; 用Poisson分布$\lambda = 1 = 100 * 0.01$逼近， 有 $P(Y = 5) = 1/5 \cdot e^{-1} = 0.00306$, 误差不大。

# 应用例子

- (事故发生) 设某高速公路段每天发生事故的次数服从$\lambda = 3$的泊松分布，求今天不发生事故的概率。
  解答： $P(X = 0) = e^{-3} \approx 0.05.$
- (印刷错误)设某本书任一页有印刷错误的次数服从$\lambda = 1$的泊松分布， 求某一页出现至少一个错误的概率
  解答： $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1} \approx 0.633$

# 泊松范例***

泊松范例：$n$个小概率事件发生的概率$p_i$,所有事件互相独立或弱相依，则事件发生次数服从 $\lambda = \sum p_i$的泊松分布。
例如： 配对问题 $N$个人得到自己作业的次数近似服从$\lambda = N * 1/N = 1$的泊松分布.

**EXAMPLE (生日问题***)**

任意两个人生日相同的事件是小概率，弱相依的； 设发生事件的次数服从$\lambda = C_n^2 \frac{1}{365}$的泊松分布，求$n$个人里至少两个人同生日的概率。

解答： 一次试验为任选两个人为同一天生日；$p = 1/365$, 试验次数(任选两个人)$C_n^2$. 出现两个人生日相同次数$X$服从泊松分布。
有$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda}$
如果要求$P(X \geq 1) > 0.5$,即$e^{-n(n-1)/730} \geq 0.5$; 计算有$n \geq 23$.
多少人中有三个人生日相同？ $\lambda = C_n^3(\frac{1}{365})^2$.

# 连续随机变量的定义

从离散到连续

- 样本空间：有限(或可数) $\rightarrow$ 不可数(一般是区间)
- 概率律：概率分布列$p_i$ $\rightarrow$ 概率密度函数 $f(x)$
  $p_i = P(X = x_i)$; $P(x \leq X \leq x + \delta) \sim f(x)\delta$

例子：打靶的环数:$X = 10, 9, 8, ..., [0, 10]$.

**Definition (连续随机变量)**

随机变量$X$取值为实数中区间时， 如果存在非负函数$f(x)$,使得任意$-\infty \leq a < b \leq \infty$, 有 $P(a < X \leq b) = \int_a^b f(x)dx$, 称$X$为*(绝对)*连续随机变量。 特别称$f(x)$是$X$的概率密度函数*probability density*。

性质：

- 密度函数的充分必要条件：$0 \leq f(x), \int_{-\infty}^{\infty} f(x)dx = 1.$
- 分布函数与密度函数 $F'(x) = f(x)$(除有限点).

# 密度函数的判断

密度函数的性质

- $f(x) \geq 0, \int_{-\infty}^{\infty} f(x)dx = 1$
- $F(x) = \int_{-\infty}^{x} f(t)dt$

EXAMPLE

设$f(x) = a\sin x, \pi \geq x \geq 0$, $f(x)$是否可以是某个随机变量的密度函数?特别求其分布函数.

$a = 1/2$
$F(x) = 1/2(1 - \cos x), 0 \leq x < \pi.$

# 均匀分布

### Definition (uniform 分布)

如果连续随机变量$X$的密度函数满足， $f(x)=\frac{1}{b-a}, x\in(a,b)$, $f(x)=0$, 其他情形。 记为$X\sim\mathcal{U}(a,,b)$.
特别称$I_{(a,b)}=1, x\in(a,b)$为示性函数, $f(x)=\frac{1}{b-a}I_{(a,b)}$

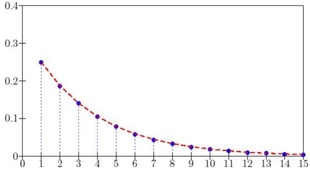
模型: 等可能事件的推广. 特别有分段均匀函数(离散简化: 两点分布)

### EXAMPLE (随机到达)

设某路公共车从5点开始每十五分钟发车；假设一个人在8点到8:30之间随机到达车站， 问 (a) 他等候时间小于5分钟的概率 (b)他等候时间大于10分钟的概率。

解答 设到达时间$X$服从均匀分布$\mathcal{U}(0,30)$；(a)
$P=P(10<X<15)+P(25<X<30)=\int_{10}^{15}\frac{1}{30}dx+\int_{25}^{30}\frac{1}{30}dx=1/3$
(b)类似 $P=P(0<X<5)+P(15<X<20)=1/3$

# 指数分布




**Definition (指数分布)**

如果连续随机变量$X$的密度函数满足，
$f(x) = \lambda e^{-\lambda x}, x \geq 0$, $f(x) = 0$, 其他情形。 记为$X \sim \varepsilon(\lambda)$.

- $F(\alpha) = 1 - e^{-\lambda\alpha}, \alpha \geq 0$;
  几何分布与指数分布: $F_{geom}(N) = 1 - (1-p)^N \sim F(N\delta)$
- 模型：小概率事件之间的间隔时间(等待时间);
  下一次类似汶川地震发生的时间；人或机器的寿命；排队时间；
- (定理) 指数分布的无后效性：$P(X > s + t | X > s) = P(X > t)$
- 排队等待概率:设有两个窗口，两个人正在办理，每个人的办理时间服从指数分布$\varepsilon(\lambda)$, 设你是下一个，则你最后办完的概率是 1/2.
  注：排队中不必随便换队！

# 指数分布与失效函数

**Definition (失效函数:失效率，死亡率)**

定义$\lambda(t)=\frac{f(t)}{1-F(t)}$为分布$F$的失效函数。

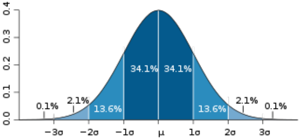
注记： 解释：寿命到达$t$时刻后在$dt$内失效的概率.
指数分布的失效函数为常数$\lambda$,称为死亡率。一般情况是个函数。

**EXAMPLE (吸烟与寿命)**

一般可设人的寿命满足死亡率为$\lambda$的指数分布。研究说明吸烟者的死亡率是非吸烟者的两倍。 设一个不吸烟的50岁人的正常死亡率为1/30,则同龄的吸烟者死亡率为1/15. 计算两个人活到60岁的概率。

解答： $P(X>10)=1-F_X(10)=e^{-1/3}\approx 0.71$
$P(Y>10)=1-F_Y(10)=e^{-2/3}\approx 0.51.$

# 正态分布


**Definition (正态分布,高斯分布，Bell曲线 Normal)**

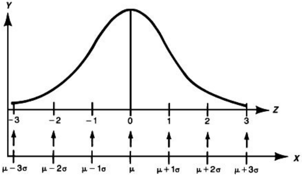
如果连续随机变量$X$的密度函数满足，

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in R$,称$X$服从参数为$(\mu, \sigma)$的正态分布。记为$X \sim N(\mu, \sigma^2)$.

特别称$N(0,1)$为标准正态分布。记密度函数为$\Phi(x)$.

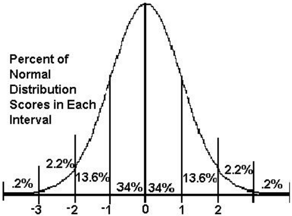
- 二项分布的逼近: $n$充分大当可用正态分布近似计算。
  Demoivre-Laplace 定理：$B(n,p) \to N(np, np(1-p))$
- 模型：大量物理或生物数据的特征指标
  高斯分布:研究测量误差的分布；IQ分数，成绩；
- 特别$\mu$表示平均值，$\sigma$表示差异大小；
  $f(x)$关于$x = \mu$对称，$f(\mu) = 1/(\sqrt{2\pi\sigma^2})$最大;

# 标准正态分布及分位数




The Translation of $X$ to $Z$ by the Transformation $Z = (X-\mu)/\sigma$

*Figure 3*




- 正态分布累积分布函数没有解析表达式！
- 正态分布的线性变换：
  $X \sim N(0,1), Y = \sigma X + \mu \sim N(\mu, \sigma^2)$
- 标准正态分布表： $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
  特别有$z_\alpha$称为$\alpha$分位数:满足
  $P(Z \le z_\alpha) = \alpha, 0 \le \alpha \le 1$.
- 正态分布的取值分布:$\mu \pm \sigma, 68.27\%$,
  $\mu \pm 2\sigma, 95.45\%$, $\mu \pm 3\sigma, 99.73\%$,
  $P(|X-\mu| \le 6\sigma) = 99.99999980\%$; 六西格玛管理!

应用(考试成绩)： 希望一个好的考试成绩是正态分布,设成绩分类为$A, B, C, D, F$, 可以给成绩如下: $A = \{X > \mu + \sigma\}, 15.67\%$,
$B = \{\mu < X < \mu + \sigma\}, 34.13\%$, $C = \{\mu - \sigma < X < \mu\}, 34.13\%$
$D = \{\mu - 2\sigma < X < \mu - \sigma\}, 13.59\%$, $F = \{X < \mu - 2\sigma\}, 2.28\%$.

# 其他连续分布***

- $\Gamma(\alpha,\beta)$分布: $\alpha$次成功的等待时间。
  $\Gamma(n,\beta)=\epsilon_1(\beta)+\epsilon_2(\beta)+\ldots.$
  离散模型：负二项分布。
- Weibull分布：最弱环节的寿命(可靠性) $F(x)=1-e^{-(x/\eta)^\beta}$.
- Beta分布：$f(x)=\frac{1}{\beta(a,b)}x^{a-1}(1-x)^{b-1}, 0<x<1$
  $\beta(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma a+b}$.
- 其他复合随机变量的分布：对数正态分布, ...

# 作业

北航教材:
P44 习题二. 14,18,20,23,27,33,35,37

# 标准正态分布

**Remark (标准正态分布)**

$N(0,1)$为标准正态分布,密度函数满足， $f(x) = \frac{1}{\sqrt{2\pi}} exp^{-\frac{x^2}{2}}, x \in R$,
特别记分布函数为$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$.

- 结论： $\int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi}$.
- 密度函数性质：关于0点对称，0点最大，两边单调。
- 分布函数性质：$\Phi(0) = 1/2, \Phi(-x) = 1 - \Phi(x)$.
- 分位数点：$\Phi(-1.96) = 0.025, \Phi(-1.645) = 0.05$
  $P(X > z_{1-\alpha}) = \alpha,\ P(|X| > z_{1-\alpha/2}) = \alpha$

$3\sigma$,$6\sigma$管理的概率： $X \sim N(\mu, \sigma^2)$.
则$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 0.9974$.

# 正态分布与信号噪音***

## Remark

一般假设噪音服从正态分布$N(0, \sigma^2)$(与其他随机变量独立)。

## EXAMPLE (信号传输)

假设传输二进制信号0, 1, 发送信号为$s$,接收信号为$R$, 由于有随机噪音$N \sim N(0, 1)$影响, 可设$R = s + N$。如果信号是1,发送$s = 2$，信号是0，发送$s = -2$。指定接收信号的判断法则：$R \geq 0.5, s^* = 1$, $R < 0.5, s^* = 0$. 求该法则的错误概率。

## 解答:.

两类错误: 信号为0,判断为1, 信号为1,判断为0. 即$P(s^* = 1|s = -2) = P(R \geq 0.5) = P(-2 + N > 0.5) = \Phi(-2.5) \approx 0.062$,
类似有$P(s^* = 0|s = 2) = P(R < -1.5) = \Phi(-1.5) \approx 0.0668$. □

# 如何构造新的随机变量?

一个随机变量：函数的复合 $g(X)$

- 设$X$是随机变量,$g$是$R \to R$的函数，$Y = g(X) : \Omega \to R \to R$，$Y$是随机变量;
- 离散分布: $p_0, p_1, p_2 \dots$
  例子: $Y = X^2$, $Y = 4 \to X = \pm 2$,
  $P(Y = 4) = P(g^{-1}(4)) = P(X = 2) + P(X = -2)$;
  参见第四章例1.

多个随机变量: 随机变量的运算 $X + Y, XY, X/Y$, 需要随机向量的分布，见第三章。

# 复合随机变量的分布

已知$X$的分布，问$Y=g(X)$的分布；

- $g$是$R\to R$的连续函数；　$Y:\Omega\to R\to R$，　$Y$是随机变量；
- 离散分布：求原像。例子 $Y=X^2, Y=4\to X=\pm 2$, $P(Y=4)=P(g^{-1}(4))=P(X=2)+P(X=-2)$;
- 连续分布：分布函数求原像 $F(y)=F(g(x))$; $g^{-1}(y)$可能有多个值！ $P(g(X)\le y)\sim P(\{X\le g^{-1}(y)\})$

**EXAMPLE (正态分布的线性变换)**

设$X\sim N(0,1)$,则$Y=\sigma X+\mu$服从$N(\mu,\sigma^2)$。

**EXAMPLE ($\chi^2$分布)**

设$X\sim N(0,1)$,则$Y=X^2$服从$\chi^2$分布。 $f_Y(y)=\frac{1}{\sqrt{2\pi}}y^{-1/2}e^{-y/2}$.

# 复合随机变量的密度函数

**Theorem (复合随机变量的密度函数)**

设$X$的密度函数$f(x), Y = g(X)$;$Y$的取值范围为$D$, 对任意$y \in D$, 设其原像可以写成有限点(即$\sharp(g^{-1}(y))$有限).每一个可以写成一个逆函数$h_i(y) = x_i$; 则 $Y$的密度函数为 $f_Y(y) = \sum_{i=1}^{n} f(h_i(y))|h_i'(y)|$。

注记：参见教材4.2节定理1.推广到分段单调函数即可。

特例证明:

假设$g(x)$是严格单调增的(必然有唯一原像),设逆函数$h(y)$.
$F_y(y) = P(Y \leq y) = P(g(X) \leq y) = P(x \leq h(y)) = F_x(h(y))$,
两边求导数有$f_y(y) = f_x(h(y))h'(y)$. □

例子: $X \sim N(0,1)$, $Y = X^2$.
每一个$y$有两个原像，可以定义两个逆函数$h_1 = \sqrt{y}, h_2 = -\sqrt{y}$; 计算有$f_Y(y) = \Phi(h_1(y))|h_1'(y)| + \Phi(h_2(y))|h_2'(y)| = \frac{e^{-y/2}}{\sqrt{2\pi y}}$.

# 期望与方差

**EXAMPLE (寿命预期)**

已知一个中国人(50岁)继续生存的时间服从指数分布$\epsilon(1/25)$. 问一个50岁人的预期寿命是多少? 大多数人的寿命范围? $\mu \pm s, 75 \pm 25.$

- 随机变量的期望: $EX = \sum x_i p_i$
  即平均值,盈利: 常记为$\mu$;
  例子: Roulette美式轮盘赌有38个数值, 最简单的玩法押一个数值一元, 可赢得35元。问一次押注的期望赢钱多少?
- 随机变量的方差: $Var(X) = DX = \sum (x_i - EX)^2 p_i$
  即变异, 风险: 通常定义标准方差$s = \sqrt{DX}$,记数据为$\mu \pm s$.
  例子续: 如果同上的豪华轮盘赌的最低押金是十元(可赢350元), 你希望参加吗?
- 还有其他特征吗? 偏态 skewness, 峰态: kurtosis;

# 分布的特征: 期望 expectation

**Definition (离散随机变量期望)**

设离散随机变量$X$的概率分布列为$p_i$,且$\sum_{i=0}^{+\infty}|x_i|p_i < +\infty$,
称$EX = \sum_{i=0}^{+\infty} x_i p_i$为$X$的期望(值).

- 随机变量的平均特征:又称均值 (概率加权平均);
- 期望可能无穷,有限取值随机变量有有限期望;
- 掷骰子的期望值是$EX = (1+2+3+4+5+6)/6 = 3.5$

常见离散分布的期望;

- 两点分布(贝努利分布) $EX = p$
- 二项分布$B(N,p)$, $EX = np$
- 泊松分布$\mathcal{P}(\lambda)$,$EX = \lambda$
- 几何分布$EX = 1/p$;
- 超几何分布$H(n,M,N)$, $EX = nM/N$.

# 随机变量的特征

正态分布的特征$N(\mu,\sigma^2)$: $\mu$表示平均值，$\sigma$表示差异大小。

**Definition (连续随机变量期望)**

设连续随机变量$X$的密度函数为$f(x)$,且$\int_{i=-\infty}^{+\infty}|x|f(x)dx<+\infty$,
称$EX=\int_{-\infty}^{+\infty}xf(x)dx$为$X$的期望(值).

注：离散随机变量：$EX=\sum_{i=0}^{+\infty}x_i p_i$;
离散到连续：$EX=\sum xP(x\leq X\leq x+dx)\approx\sum x(f(x)dx)$

**Definition (方差)**

设随机变量$X$的期望$EX=\mu$有限, 则称$Var(X)=E(X-\mu)^2$为$X$的方差.又记为$DX$.

可以验证: 正态分布有$EX=\mu$, $Var(X)=\sigma^2$. 注：还有其他特征,参见第五章。

# 数学期望和矩

**Definition (连续随机变量期望)**

设连续随机变量$X$的密度函数为$f(x)$,且$\int_{i=-\infty}^{+\infty}|x|f(x)dx < +\infty$,
称$EX = \int_{-\infty}^{+\infty} xf(x)dx$为$X$的期望(值).

- 从离散到连续：级数求和到积分。
- 积分可能不存在或无穷；
- 推广：$EX^n = \int_{-\infty}^{+\infty} x^n f(x)dx$,称为$X$的$n$阶矩(moment)。
- 特别$E(X-EX)^n$称为$X$的$n$阶中心矩，二阶中心矩即方差$DX = E(X-EX)^2 = EX^2 - (EX)^2$.
- 基本性质：(积分性质决定)单调性,线性复合。

***$E$又称为期望算子(线性):给定一个随机变量(分布函数)得到一个数。参见泛函分析。

# 一般期望的计算公式

**Proposition (复合函数)**

设$Y=g(X)$,如果$X$是离散随机变量，$EY=\sum_{i=0}^{\infty} g(x_i)p_i$; 如果$X$是连续随机变量，$EY=\int_{-\infty}^{+\infty} g(x)f(x)dx$.

- 离散情形证明: $\sum_j g(x_j)p(x_j)=\sum_i \sum_{g(x_j)=y_i} y_i p(x_j)=EY$.
- 连续情形证明：***

**Lemma (非负随机变量的期望)**

设$Y\geq 0$,则$EY=\int_0^{\infty} P(Y>y)dy$.

**Proof.**

$Eg(X)=\int_0^{\infty} P(g(x)>y)dy=\int_0^{\infty}\int_{g(x)>y} f(x)dxdy$

交换积分$Eg(x)=\int_{x:g(x)>0}\int_0^{g(x)} dyf(x)dx=\int_{x:g(x)>0} g(x)f(x)dx$. □

# 常见随机变量的期望与方差

常见离散分布;

- 两点分布，Bernouli贝努利分布 $EX = p, VarX = pq$
- 二项分布$B(N, p)$， $EX = np, VarX = npq$
- 泊松分布$\mathcal{P}(\lambda)$,$EX = \lambda, VarX = \lambda$
- 几何分布$EX = 1/p, VarX = q/p^2$;

常见连续分布:

- 均匀分布 $\mathcal{U}(a, , b)$,$EX = \frac{a+b}{2}, VarX = \frac{(b-a)^2}{12}$;
- 指数分布 $\varepsilon(\lambda)$, $EX = 1/\lambda, VarX = 1/\lambda^2$;

# 实例：期望

### EXAMPLE (淘宝店盈利估计)

设淘宝店销售一种商品。假设卖一件盈利$a$元，没卖掉损失$b$元。设商品的需求量服从某个分布$F$, 问库存(订购)多少商品可以得到最大平均利润？ 特别如果$F$是指数分布，应该库存多少商品？

解答: 设$Y$是商品的需求， $P(Y \leq k) = F(k)$.
设库存量为$k$,则盈利随机变量$Z$满足；
$Z = aY - b(k - Y), Y < k, Z = ak, Y \geq k$
设$f(y)$为密度函数；平均利润
即$m(k) = EZ = \int_0^k (ay - b(k - y))f(y)dy + \int_k^\infty akf(y)dy$
计算有$m(k) = (a + b)\int_0^k yf(y)dy + ak - (a + b)k\int_0^k f(y)dy$;
即求$m(k)$的极值点。计算有$m'(k) = a - (a + b)\int_0^k f(y)dy, m''(k) < 0$,
极大值点$m'(k) = 0$,有$F(k) = \frac{a}{a+b}$; 特别当$F$为指数分布，$1 - e^{-\lambda k} = \frac{a}{a+b}$; $k = (\lambda)^{-1}\ln\frac{a+b}{b}$.

# 作业

北航教材:
P89 习题四. 1， 7,13
P118 习题五。 2， 3， 6， 9