第三部分
特殊的概率密度函数

- 概率分布函数反映了随机变量的概率分布规律；
- 在概率论中处理概率分布时一般不涉及分布的物理来源，为在实验数据分析中正确地掌握和运用这些分布函数，需要：
  - 熟悉公式及运算规则；
  - 分布的物理意义；
- 实验数据处理中所用到的概率分布的来源：
  1. 实验所涉及到的物理问题本身的统计性质带来的，这类分布比较多样化，是和所处理的物理问题有直接的联系；
  2. 对实验测量结果作数据处理时所引进的。这一类分布比较标准化，且处理的方法也比较明确；
- 本章内容：
  - 数据处理过程中常用的概率分布函数，给出它们的定义、性质和实际应用

二项式分布

（Binomial distribution)

一、定义（亦称伯努利分布）：

考虑一个随机实验的两个互斥的结果：成功和失败，设成功的概率为***p***，则不成功的概率为***1-p=q***。在**n**次独立的实验中，有***r***次成功的概率为：

$$B(r;n,p)=\binom{n}{r}p^r(1-p)^{n-r},r=0,1,2,\cdots n$$

$$\binom{n}{r}\equiv\frac{n!}{r!(n-r)!}=\binom{n}{n-r}$$

二、性质：

1. 满足归一化条件

$$\sum_{r=0}^{n}B(r;n,p)=1$$

证：$\sum_{r=0}^{n}B(r;n,p)=\sum_{r=0}^{n}\binom{n}{r}p^r(1-p)^{n-r}=\sum_{r=0}^{n}\binom{n}{r}p^rq^{n-r}=(p+q)^n=1$

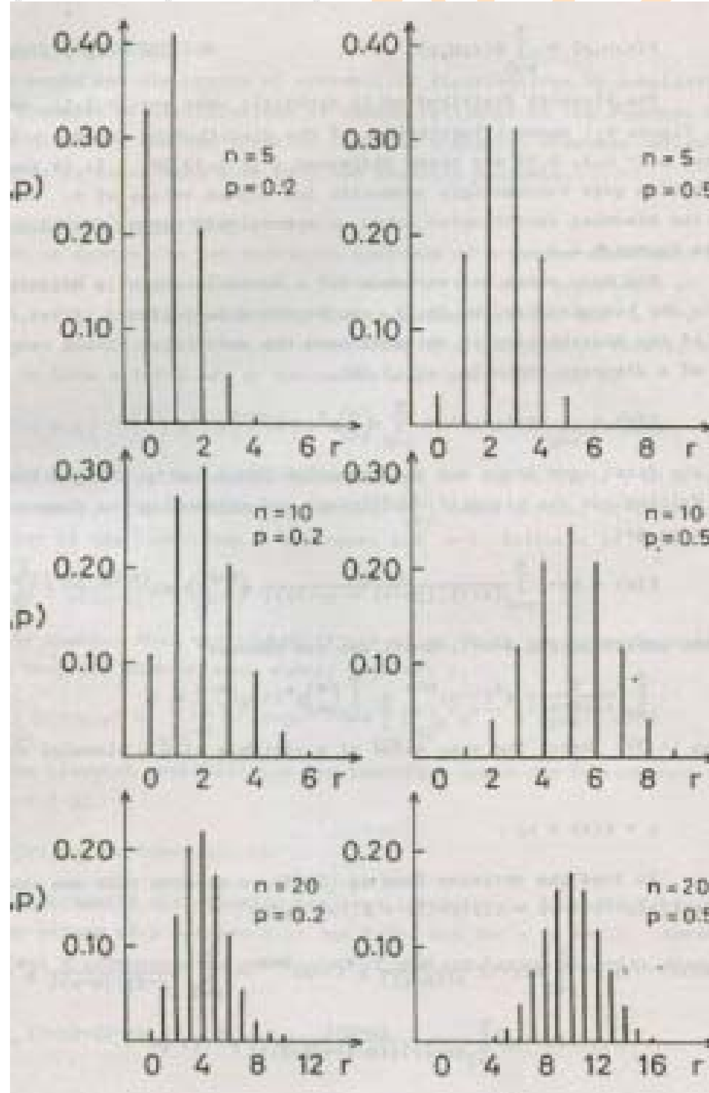
2. 在变换(*r,p*)←→(*n-r,1-p*)下保持不变：***B(r;n,p)=B(n-r;n,1-p)***

3. 当p=q=0.5时，是对称的;

   对于任意的p值，是非对称的;

   当n增大时，分布趋于对称;

   当n很大时，近似为正态分布



4. 服从二项式分布的随机变量*r* 的平均值和方差:

$$E(r) = np$$

$$V(r) \equiv E[r - E(r)]^2 = np(1-p) = npq$$

三、应用：

给出进行N次实验有*r*次成功的概率。

例1：直方图（Histogram）

考虑一直方图，设A表示一事例落入Bin *i*，A表示某事例落入直方图中其它的Bin，如果共有*n*个独立的事例，其中有r个事例落入Bin *i*，*n* -*r*个事例分布于其它的Bin ➔*r*服从二项式分布

Bin *i*中事例数*r*的期望值和方差：

$$\mu \equiv \mathrm{E}(r) = n\,p$$

$$\mathrm{V}(r) = n\,p\,(1 - p)$$

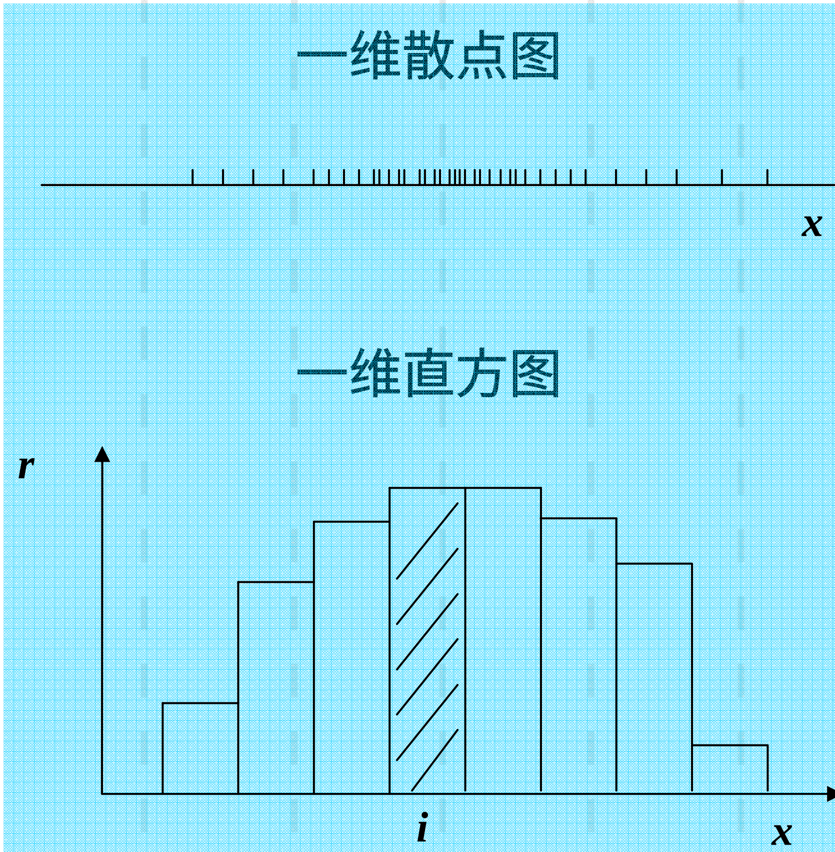
概率*p*是未知的，可由实验结果估计：

$$p = \hat{p} = \frac{r}{n}$$

*r*的标准偏差：

$$\sigma = \sqrt{V(r)} = \sqrt{r(1-\frac{r}{n})}$$

$$\sigma \to \sqrt{r}\,, n \to \infty$$



例2．设在某实验中，所期望的事例出现的概率为p。问，需要作多少次实验才能使至少有一个这样的事例出现的概率为α？

设在N次实验中共出现了X这样的事例。X服从二项式分布

$$B(X;N,p)=\binom{N}{X}p^{X}(1-p)^{n-X}$$

至少有一个这样的事例出现的概率：

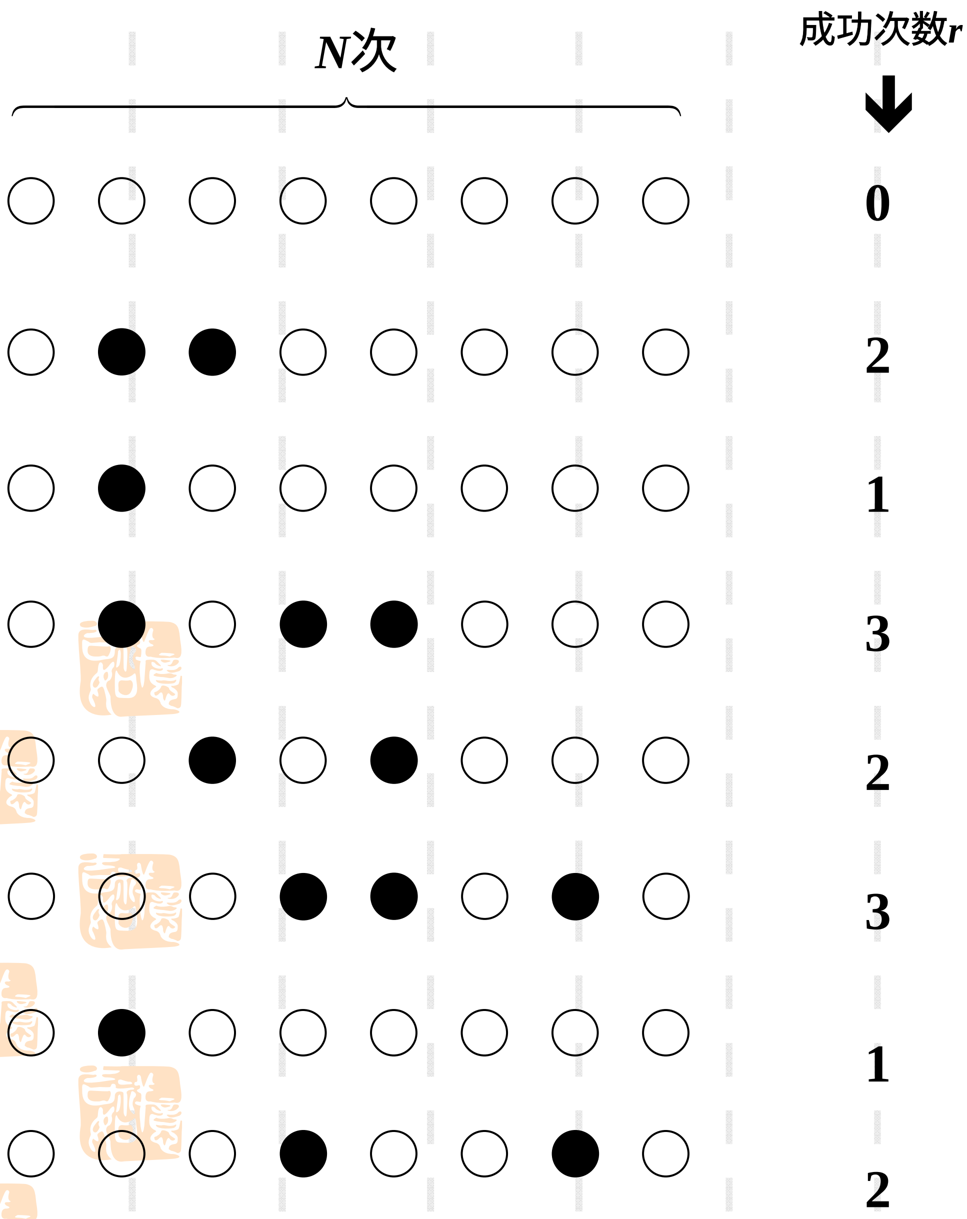
$$\boldsymbol{p}(\boldsymbol{X}\geq 1)=\sum_{\boldsymbol{X}=1}^{\boldsymbol{N}}\boldsymbol{B}(\boldsymbol{X};\boldsymbol{N},\boldsymbol{p})=1-\boldsymbol{B}(0;\boldsymbol{N},\boldsymbol{p})\geq\alpha$$

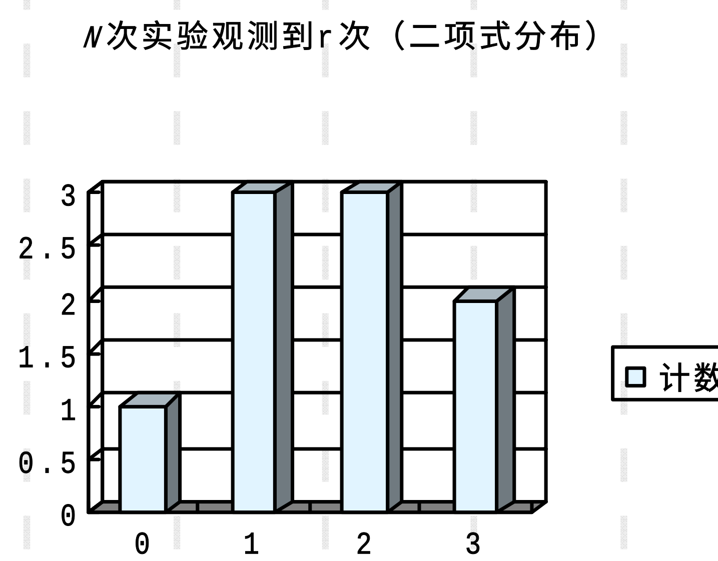
$$1-p(X=0)=1-B(0;N,p)\geq\alpha$$

$$B(0;N,p)=(1-p)^{N}$$

$$\therefore (1-p)^{N}\leq 1-\alpha$$

$$N\geq\left|\log(1-\alpha)/\log(1-p)\right|$$


N次
成功次数r
0
2
1
3
2
3
1
2


N次实验观测到r次（二项式分布）
3
2.5
2
1.5
1
0.5
0
0
1
2
3
计数

几何分布

作一系列独立的伯努利实验，前r-1次实验失败，第r次成功的概率：

$$g(r,p)=p(1-p)^{r-1}$$

不是从n次实验中抽取的。

负二项式分布

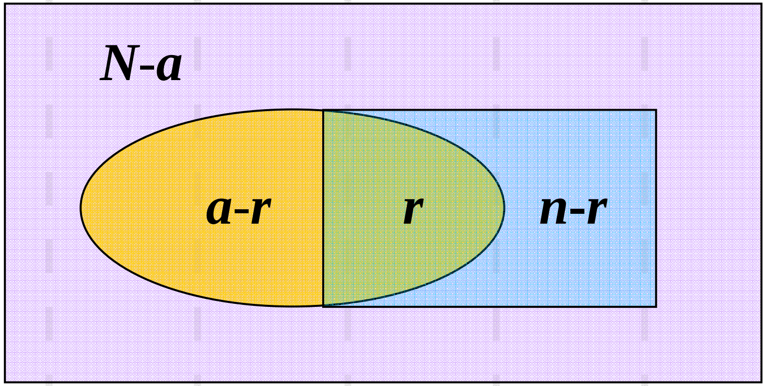
作一系列独立的伯努利实验，在第r次实验中事件是第k次成功，这类事件的概率为：

$$P_k(r;p)=\binom{r-1}{k-1}p^k(1-p)^{r-k}$$

超几何分布

N个元素，其中a个表示成功，N-a个表示失败，从N个元素中一次抽取n个元素，其中有r个成功，n-r个失败的概率为：

$$P(r;N,n,a)=\binom{N-a}{n-r}\binom{a}{r}\Big/\binom{N}{n}$$



超几何分布的期望值和方差为：

$$E(r)=\frac{na}{N}$$

$$V(r)=\frac{N-n}{N-1}\cdot\frac{na}{N}\cdot(1-\frac{a}{N})$$

当 $n \ll N$ 时，超几何分布近似为二项式分布

$$B(r;n,p)$$

其中 $p=\frac{a}{N}$。

第四章
特殊的概率密度函数

4.2 多项式分布
（Multinomial distribution)

多项式分布

（Multinomial distribution)

一、定义

设可能的实验结果可分成k组：A_1、A_2、…、A_k，每次实验结果落入某一组A_i的几率为p_i

$$\sum_{i=1}^{k} p_i = 1$$

如果共进行了n次独立的实验，实验结果落入各个组的次数为r_1、r_2、…、r_k的概率为 ($\sum_{i=1}^{k} r_i = n$)

$$M(\underline{r};n,\underline{p}) = \frac{n!}{r_1!r_2!..r_k!} p_1^{r_1} p_2^{r_2} \cdots p_k^{r_k}$$

二、性质

多项式分布是二项式分布的推广，除具有二项式分布的一些特性外，还具有以下的附加性质：

1）r_i的期望值：　　$E(r_i) = Np_i$

2）　r_i的方差：　　$v(r_i) = np_i\,(1 - p_i)$

3）　r_i和r_j的协方差：$cov(r_i, r_j) = -np_ip_j$

相关系数：

$$\rho(r_i, r_j) = \frac{\mathrm{cov}(r_i, r_j)}{\sigma_i \sigma_j} = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}$$

即：r_i和r_j总是负相关

一维直方图中，当bin宽度足够小时（$p_i \to 0$），r_i和r_j相关度很小。

4）当n很大时，多项式分布趋向于多维正态分布

三、应用：

<u>用于处理一次实验有多个可能的结果的情况</u>

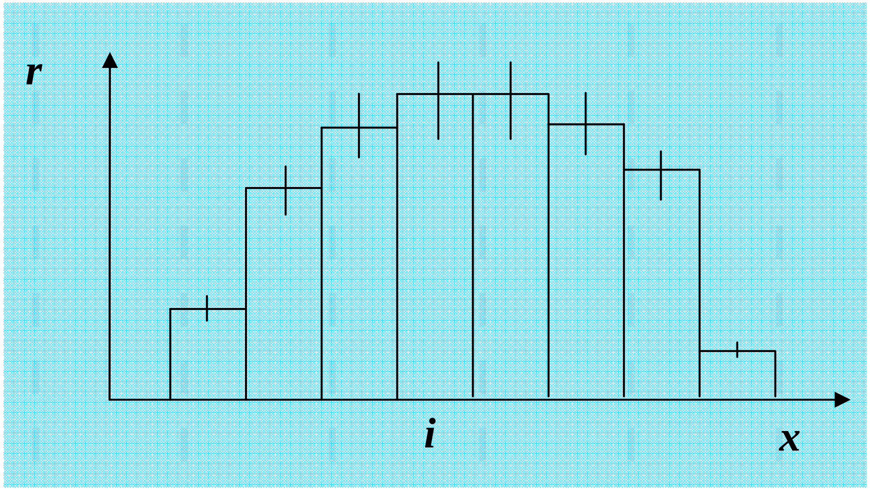
例：设有n个事例，分布于直方图的k个bin中，某事例落入bin i的概率为p_i，落入bin i的事例数为r_i，则k个bin中事例数分别为r_1、r_2、…、r_k的概率为多项式分布

r_i的期望值和方差：$E(r_i) = np_i \quad v(r_i) = np_i(1 - p_i)$

如果$p_i << 1$，即bin的数目k很大，则有$v(r_i) \approx np_i = r_i$

$$\sigma(r_i) \approx \sqrt{r_i}$$

带误差棒的一维直方图



第四章
特殊的概率密度函数

4.3 泊松分布

（Possion distribution)

泊松分布
(Possion distribution)

一、定义

泊松分布是二项式分布的极限形式：p→0，n→∞,但np=有限值μ.

根据Stirling公式，当n很大时

$$n!\cong\sqrt{2\pi}\,n^{n}e^{-n}$$

$$\frac{n!}{r!(n-r)!}p^{r}(1-p)^{n-r}\approx\frac{1}{r!}\frac{\sqrt{2\pi n}\,n^{n}e^{-n}}{\sqrt{2\pi(n-r)}(n-r)^{n-r}e^{-(n-r)}}\left(\frac{\mu}{n}\right)^{r}(1-\frac{\mu}{n})^{n-r}$$

$$\approx\frac{1}{r!}\frac{n^{n}}{(n-r)^{n}(n-r)^{-r}e^{r}}\mu^{r}(1-\frac{\mu}{n})^{n}$$

$$\approx\frac{1}{r!}\frac{1}{(1-\frac{r}{n})^{n}e^{r}}\mu^{r}(1-\frac{\mu}{n})^{n}$$

$$\approx\frac{1}{r!}\mu^{r}e^{-\mu}\qquad\Leftarrow e^{-x}\approx(1-\frac{x}{n})^{n}$$

$$p(r;\mu)=\frac{1}{r!}\mu^{r}e^{-\mu}\qquad r=0,1,2,\cdots$$

二、性质

- 期望值：E(γ)= μ
- 方差：V(γ)= μ

三、应用：

泊松分布给出在事例率为常数的情况下，在某一给定时间间隔内得到r个独立事例的概率。

例1. 气泡室中的气泡沿着带电粒子径迹的分布

设单位径迹长的上气泡的平均数目为常数g，假定

1. 在长度间隔[l, $l+\Delta l$]上最多只有一个气泡；
2. 在[l, $l+\Delta l$]这个间隔中找到一个气泡的概率正比于Δl；
3. 在两个不重迭的间隔中产生气泡的事件是互不相关的；

具有上述特点的随机过程就称为泊松过程。

由假设1和2，在 $[l, l+\Delta l]$ 中

有一个气泡的概率：$p_1(\Delta l)=g\Delta l$

没有气泡的概率：$p_0(\Delta l)=1- p_1(\Delta l)=1-g\Delta l$

根据假设3

$$p_0(l+\Delta l)= p_0(l) \cdot p_0 (\Delta l)$$ ←独立性

➔在$[l, l+\Delta l]$长度上没有气泡的概率＝在l长度上没有气泡的概率×在Δl长度上没有气泡的概率

$$\frac{p_0(l+\Delta l)-p_0(l)}{\Delta l}=-gp_0(l)$$

$$\Delta l \to 0 \quad \Rightarrow \frac{dp_0(l)}{dl}=-gp_0(l)$$

取边界条件$p_0(0)=1$,

$$p_0(l)=e^{-gl}$$

求在长度 l上观测到r个气泡的概率$p_r(l)$：

根据假定，在间隔$[l, l+\Delta l]$内最多只能有一个气泡

$$p_r(l+\Delta l)=\underbrace{p_r(l)\,p_0(\Delta l)}_{\text{r个气泡都在}l\text{内}}+\underbrace{p_{r-1}(l)\,p_1(\Delta l)}_{\text{r-1个气泡在}l\text{内，1个在}\Delta l}$$

$$\frac{dp_r(l)}{dl}=-gp_r(l)+gp_{r-1}(l)$$

$$\Rightarrow \quad p_r(l)=\frac{1}{r!}(gl)^r e^{-gl}$$

➔平均值$\mu=gl$的泊松分布

对r=0（在$[0,l]$中不产生气泡），概率是 $p_0(l)=e^{-gl}$

例2 放射源和本底辐射的叠加

从放射源中辐射出的粒子的数目服从泊松分布。

λ_x：单位时间内从放射源中辐射出的平均粒子数

γ_x：时间间隔t辐射出的粒子数目

$$p_r(r_x;\lambda_x t)=\frac{1}{r_x!}(\lambda_x t)^{r_x}e^{-\lambda_x t} \qquad \mu=\lambda_x t$$

如果将放射源放入一容器中，容器中的本底辐射服从$\mu=\lambda_b$的泊松分布

$$p_r(\gamma_b;\lambda_b t)=\frac{1}{\gamma_b!}(\lambda_b t)^{\gamma_b}e^{-\lambda_b t}$$

可测量量是来自放射源和本底的总粒子数，其分布为

$$p_r(\gamma;\lambda_x t,\lambda_b t)=p_r(\gamma-\gamma_b t,\lambda_t t)p_r(\gamma_b,\lambda_b t)$$

$$=\frac{1}{\gamma_b!}[(\lambda_x+\lambda_b)t]^{\gamma}e^{-(\lambda_x+\lambda_b)t}$$

服从泊松分布的变量的加法定理：几个独立的泊松分布变量的和还是泊松分布变量。

例3 计数器的计数分布

设计数器的计数效率为**p<1,**在时间间隔**t**内通过计数器的总粒子数**N**服从平均值为**v**的泊松分布。求在时间间隔内，计数器所记录到的粒子数的分布**p(r)**

要得到**r**个计数，必须至少有**r**个粒子通过探测器。对于一个取得的**N,**得到**r**个计数的概率服从二项式分布。

P(r)＝所有可以给出**r**个计数的概率之和

$$p(r)=\sum_{N=r}^{\infty}B(r,N,p)p(N,V)=\frac{1}{r!}\left(pv\right)^{r}e^{-pv}$$

➔**μ=pv**的泊松分布

例4 多项式分布和泊松分布间的关系

考虑有k个Bin的直方图，每个Bin中的事例数r_i服从多项式分布，设总事例数N服从平均值为ν的泊松分布，则联合概率密度

$$P(r_1,r_2\cdots,r_k,N)=M(\vec{r},N,\vec{p})\cdot P(N,\nu)$$

$$=\left(\frac{N!}{r_1!r_1!\cdots r_k!}p_1^{r_1}p_2^{r_2}\cdots p_k^{r_k}\right)\cdot\left(\frac{1}{N!}\nu^N e^{-\nu}\right)$$

$$=\left(\frac{1}{r_1!}(\nu p_1)^{r_1}e^{-\nu p_1}\right)\cdots\left(\frac{1}{r_k!}(\nu p_k)^{r_k}e^{-\nu p_k}\right)$$

$$=P(r_1,\nu p_1)\cdots P(r_k,\nu p_k)$$

$$\boldsymbol{N}=\sum_{i=1}^{k}\boldsymbol{r_i},\quad \sum_{i=1}^{k}\boldsymbol{p_i}=1$$

$$\nu^N=\nu^{\sum r_i}=\prod_{i=1}^{k}\nu^{r_i}$$

$$\boldsymbol{e}^{-\nu}=\boldsymbol{e}^{-\nu\sum_{i=1}^{k}p_i}=\prod_{i=1}^{k}\boldsymbol{e}^{-\nu p_i}$$

即：每个Bin中的事例是独立的泊松变量

$$E(r_i)=V(r_i)=\nu p_i=r_i$$

$$p_i=\frac{r_i}{E(N)}=\frac{r_i}{\nu}\qquad \therefore\sigma(r_i)=\sqrt{r_i}$$

第四章
特殊的概率密度函数

4.4 复合泊松分布

(Compound Possion distribution)

复合泊松分布

(Compound Possion distribution)

定义：

设是$r_1,r_2,\ldots$是一组N个独立的泊松变量，其平均值都为μ，n也是泊松变量，其平均值为ν，求

$$r=\sum_{i=1}^{n} r_i$$

的分布P(r)

根据边缘概率的定义，p(r)应为产生r个事例的所有的概率之和:

$$p(r)=\sum_{n=0}^{\infty}\left(\underbrace{p(r=\sum_{i=1}^{n} r_i;\mu)}_{}\cdot p(n,\nu)\right)$$

为n个独立的泊松变量的联合概率

根据泊松变量的加法定理

$$p(r=\sum_{i=1}^{n} r_i;\mu)=\frac{1}{r!}(n\mu)^r e^{-n\mu}$$

$$\therefore p(r)=\sum_{n=0}^{\infty}\left\{\left[\frac{1}{r!}(n\mu)^r e^{-n\mu}\right](\frac{1}{n!}\nu^n e^{-\nu})\right\}$$

性质：

$$E(r) = \mu\nu$$
$$V(r) = \mu(\mu+1)\nu$$

应用：

泊松型的随机过程触发另外一个泊松型的随机过程

例：云室中的液滴

带电粒子通过云室时，会受到一系列的散射，而每次散射过程都会引起液滴的产生。在一给定的径迹长度上，粒子受到的散射的次数服从泊松分布，每次散射所产生的液滴的数目也服从泊松分布。因此，在给定的径迹长度上所产生的液滴的数目**r**服从复合泊松分布。

μ：每次散射所产生的液滴的平均数目

ν：在给定的径迹长度上粒子所受到的散射的平均次数

均匀分布

(Uniform distribution)

概率密度函数：$x \in [a,b]$

$$f(x)=\begin{cases}0 & x<a\text{或}x>b\\ \dfrac{1}{b-a} & a\le x\le b\end{cases}$$

性质：

1、期望值 $E(x)=\int_a^b xf(x)dx=\frac{1}{2}(a+b)$

2、方差 $V(x)=\int_a^b [x-E(x)]^2 f(x)dx=\frac{1}{12}(b-a)^2$

3、累积分布 $F(x)=\int_a^x f(x')dx'=\dfrac{x-a}{b-a} \quad x\in[a,b]$

应用：

1、多丝室的位置分辨率：粒子在两丝间的击中位置分布是均匀分布：

丝距Δ＝b－a

位置分辨率：$\sigma=\sqrt{V(x)}=\dfrac{\Delta}{\sqrt{12}}$

2、均匀分布的随机数产生器

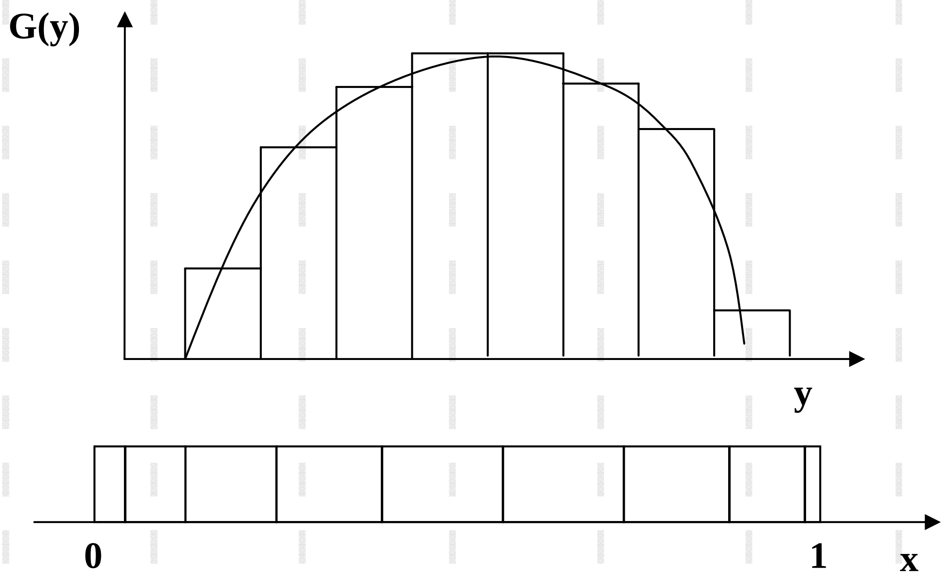
2、均匀分布的随机数产生器

任意连续分布的随机变量Y的概率密度函数为g(y)

令 $x=G(y)=\int_{-\infty}^{y} g(t)dt$

x的概率密度分布为 $f(x)=g(y)\left|\frac{dy}{dx}\right|=g(y)\left|\left(\frac{dG}{dy}\right)^{-1}\right|=1$

*x*是[0, 1]区间的均匀分布的随机变量，$y=G^{-1}(x)$是满足*g*(*y*)分布的随机变量



橡皮泥原有形状

橡皮泥压缩后的形状

第四章
特殊的概率密度函数

4.6 指数分布

(Exponential distribution)

指数分布
(Exponential distribution)

概率密度函数：

$$f(x;\beta)=\frac{1}{\beta}e^{-x/\beta}$$

性质：

期望值： $E(x) = \beta$

方差： $V(x) = \beta^2$

应用：

指数分布在粒子物理的应用非常广泛：衰变过程，衰减过程……

例：泡室中粒子径迹的距离分布

$\boxed{\text{在}[l, l+\Delta l]\text{中出现第一个气泡}}$

$\Updownarrow$

$\boxed{\text{在}[0,l]\text{中不出现气泡}} \;\cap\; \boxed{\text{在}[l, l+\Delta l]\text{中出现一个气泡概率}}$

根据泊松假设，两事件独立：

联合概率密度＝两事件概率密度之积

在[l, l+Δl] 内出现第一个气泡的概率为 $g\Delta l \bullet e^{-gl}$

在位置l 处单位长度内产生第一个气泡的概率（即概率密度）为

$$f(l) = ge^{-gl}$$

g为单位长度内平均气泡数目

例：一个放射源两次相继的核衰变之间时间间隔的分布

$$\boxed{\text{在}[t,\ t+\Delta t]\text{中发生第一次核衰变}}$$

$$\Updownarrow$$

$$\boxed{\text{在}[0,\ t]\text{中没有核衰变}} \quad \bigcap \quad \boxed{\text{在}[t,\ t+\Delta t]\text{中发生一次核衰变}}$$

根据泊松假设，两事件独立：

联合概率密度＝两事件概率密度之积

在[t, $t+\Delta t$] 内发生一次核衰变的概率为 $\lambda \Delta t \cdot e^{-\lambda t}$

在时刻t 单位时间内发生一次核衰变的概率密度为

$$f(t) = \lambda e^{-\lambda t}$$

λ为单位时间间隔内平均衰变次数

t 的平均值（称为核的平均寿命）为 $\tau \equiv E(t) = 1/\lambda$

两次衰变的时间间隔>t 的概率为 $1 - F(t) = e^{-\lambda t}$

正态分布（高斯分布）
（Normal or Gaussian distribution)

概率密度函数：

$$N(\mu,\sigma^2)\equiv f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}\quad(-\infty<x<\infty)$$

性质：

1、期望值: $\mathrm{E(x)}=\mu$

2、方差： $\mathrm{V(x)}=\sigma^2$

3、累积分布： $\mathrm{F(x)}=\Phi(\frac{x-\mu}{\sigma})$

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-\frac{1}{2}x^2}dx$$

➔误差函数

标准正态分布：（Standard Normal Distribution）N(0,1)

令 $$y=\frac{x-\mu}{\sigma}$$

得标准正态概率密度函数

$$\mathrm{N}(0,1)\equiv \boldsymbol{g}(\boldsymbol{y})=\frac{1}{\sqrt{2\pi}}\boldsymbol{e}^{-\frac{1}{2}\boldsymbol{y}^{2}}$$ ➔μ=0, σ=1的正态分布

累积标准正态分布函数：

$$\boldsymbol{G}(\boldsymbol{y})=\int_{-\infty}^{y}\boldsymbol{g}(\boldsymbol{y}')\boldsymbol{d}\boldsymbol{y}'=\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}\boldsymbol{e}^{-\frac{1}{2}\boldsymbol{y}'^{2}}\boldsymbol{d}\boldsymbol{y}'$$

$$G(-y)=1-G(y)$$

G(y)的应用：

1、设x是服从正态分布的随机变量，求x落于区间[a,b]内的概率

$$p(a \le x \le b) = p(x \le b) - p(x \le a)$$

$$= p(\frac{x-\mu}{\sigma} \le \frac{b-\mu}{\sigma}) - p(\frac{x-\mu}{\sigma} \le \frac{a-\mu}{\sigma})$$

$$= \int_{-\infty}^{b-\mu/\sigma} g(y')dy' - \int_{-\infty}^{a-\mu/\sigma} g(y')dy'$$

$$p(a \le x \le b) = G(\frac{b-\mu}{\sigma}) - G(\frac{a-\mu}{\sigma})$$

$$= G(\frac{b-\mu}{\sigma}) + G(\frac{\mu-a}{\sigma}) - 1 \qquad \Leftarrow \mathbf{G}(-\mathbf{y}) = 1 - \mathbf{G}(\mathbf{y})$$

1σ区间： $p(\mu-\sigma \le x \le \mu+\sigma) = 2G(1) - 1 = 0.6827$

2σ区间： $p(\mu-2\sigma \le x \le \mu+2\sigma) = 2G(2) - 1 = 0.9545$

3σ区间： $p(\mu-3\sigma \le x \le \mu+3\sigma) = 2G(3) - 1 = 0.9973$ 3σ 规则

2、已知概率值，求相对于平均值对称的区间 $[\mu - a, \mu + a]$

$$\gamma = G(\frac{a}{\sigma}) - G(\frac{-a}{\sigma}) = 2G(\frac{a}{\sigma}) - 1$$

$$G(\frac{a}{\sigma}) = \tfrac{1}{2}(1 + \gamma)$$

查表可得出 $\frac{a}{\sigma}$

γ = 0.9 　　　　a = 1.645
=0.95 　　　　= 1.960
=0.99 　　　　= 20576
=0.999 　　　　= 3.290

正态变量加法定理：

如果某一随机变量是一些正态变量的函数，该变量的分布形式是什么？

如果是线性函数➔ 加法定理

设$x_1,x_2,\ldots x_n$是相互独立的正态变量

$$x_i \to N(\mu_i,\sigma_i)$$

则 $$y=\sum_{i=1}^{n} a_i x_i$$

也是服从正态分布的变量，其平均值和方差分别为

$$E(y)=\sum_{i=1}^{n} a_i u_i$$

$$V(y)=\sum_{i=1}^{n} a^2{}_i \sigma^2{}_i$$

例：正态分布样本的样本平均值 $\bar{x}$和方差 s^2的特征。

设n个独立的随机变量都服从正态分布，其平均值和方差分别为μ和σ² 。对于由这n个量构成的正态样本

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

由正态变量的加法定理，样本平均值也是正态变量

$$E(\bar{x}) = \sum_{i=1}^{n} a_i u_i = \mu$$

$$V(\bar{x}) = \sum_{i=1}^{n} a^2{}_i \sigma^2{}_i = \frac{\sigma^2}{n}$$

$\leftarrow$ $a_i = \frac{1}{n}, \mu_i = \mu$

$\rightarrow$ $\bar{x}$ 的分布服从 $N(\mu, \frac{\sigma^2}{n})$

可以证明：

1、 $\dfrac{(n-1)s^2}{\sigma^2}$ 服从自由度为n－1的χ^2分布；

2、 $\bar{x}$ 和 s^2 是相互独立的随机变量

定理：

如果独立的随机变量服从相同的正态分布，则统计量 $\bar{x}$和 s^2是相互独立的；

反过来，如果随机样本的平均值和方差是相互独立的，则这一样本所代表的总体一定是正态分布。

中心极限定理（Central Limit Theorm）

设x_1，x_2，…… x_n是一组n个独立的随机变量，x_i的平均值和方差分别为μ_i和σ_i，则当n→∞时，变量

$$\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \mu_i\right) \Big/ \sqrt{\sum_{i=1}^{n} \sigma_i^2}$$

服从标准正态分布N(0,1)

例：高斯型随机变量产生器

设x 是在[0，1]之间均匀分布的随机数

$$E(x) = \mu = \tfrac{1}{2} \quad V(x) = \sigma^2 = \tfrac{1}{12}$$

对n个x的取值x_i（i=1,2,….n）定义随机变量

$$y = \left(\sum_{i=1}^{n} x_i - \frac{n}{2}\right) \Big/ \sqrt{\tfrac{1}{12} n}$$

在n→∞时，服从正态分布，在实际应用时，可取n=12

$$z = \sum^{12} x_i - 6$$

χ^2分布

(χ^2 distribution)

定义：

设$x_1,x_2,\ldots,x_n$，是一组n个相互独立的服从正态分布N(μ , σ²)的随机变量。这n个x_i构成容量为n的正态样本，所代表的正态总体的平均值和方差分别为 μ 和σ^2，定义

$$\chi^2=\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

变量χ^2的概率密度函数为

$$f(\chi^2,n)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\left(\chi^2\right)^{\frac{n}{2}-1}e^{-\frac{x^2}{2}}\quad 0\le\chi^2<\infty$$

➔自由度为n的χ^2分布

性质:

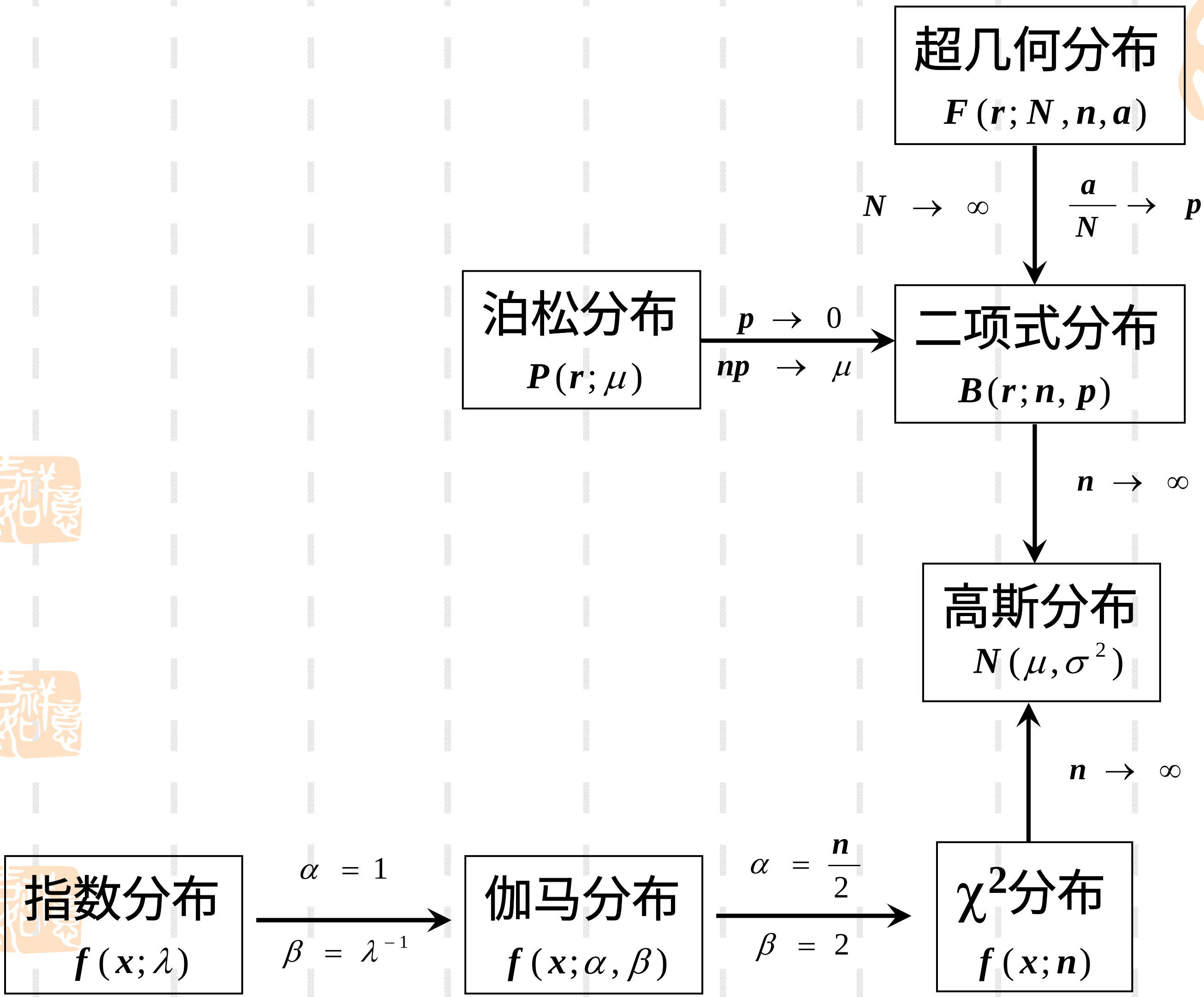
1、期望值：$E(\chi^2)=n$

2、方差：$V(\chi^2)=2n$

3、χ^2分布的概率值

$$F(\chi_\alpha^2,\nu)\equiv\int_0^{\chi_\alpha^2} f(\mu,\nu)d\mu=1-\alpha$$

几种分布的关系



超几何分布
$F(r;N,n,a)$
$N\to\infty$
$\frac{a}{N}\to p$
泊松分布
$P(r;\mu)$
$p\to 0$
$np\to\mu$
二项式分布
$B(r;n,p)$
$n\to\infty$
高斯分布
$N(\mu,\sigma^2)$
$n\to\infty$
指数分布
$f(x;\lambda)$
$\alpha=1$
$\beta=\lambda^{-1}$
伽马分布
$f(x;\alpha,\beta)$
$\alpha=\frac{n}{2}$
$\beta=2$
χ^2分布
$f(x;n)$