

# 第五讲 常用概率密度函数

---

# 二项分布

$N$  次独立测量，每次只有成功(概率始终为  $p$ )或失败(概率为 $1-p$ )两种可能，得到  $n$  次成功的概率为

$$f(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

平均值：

$$E[n] = \mu = \sum n f = Np$$

适用于仪器探测效率不确定度的计算

可以证明其满足归一化条件

$$\begin{aligned} \sum_n \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \\ = [(1-p) + p]^N = 1 \end{aligned}$$

方差：

$$\begin{aligned} V[n] &= \sigma^2 \\ &= E[(n - \mu)^2] \\ &= E[n^2] - E^2[n] \\ &= Np(1-p) \end{aligned}$$

# 二项分布的适用条件

1. 每次尝试仅有两种可能性;
2. 每次尝试的成功概率是一样的;
3. 不同次尝试的结果是独立的。

伯努利试验

考虑驾车人被停车检查有否不佩戴安全带的情况是否为一个

伯努利试验。两种结果：佩戴与不佩戴！

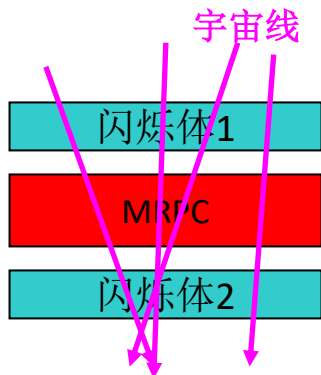
如果对所有车都一样，那么驾车人都有同样的概率不佩戴安全带！？（不同年龄人群都是一样的吗？）

检查不同驾车人都佩戴安全带，结果应该是独立的！？  
（对于同时同地的前后驾车人都是一样的吗？）

因此，根据数据采样情况，才能分清是否为伯努利试验，才能决定能否应用二项分布。

# 举例:在效率误差估计中的应用

## ● 多层阻性板室(MRPC)的探测效率



闪烁体1与2同时击中给出



穿过MRPC的粒子数 $N$

MRPC记录的击中数目 $N'$

$$p = \frac{N'}{N}$$
$$\Delta p = \frac{\Delta N'}{N} = \sqrt{\frac{p(1-p)}{N}}$$



MRPC探测效率  
测量值及其不确定度

# 二项分布指导决策

我们为大亚湾实验研制生产触发电子学板。按设计在一年内需要修理的电路板为**10%**。如果在实验所需的**20**块板中有**5**块在第一年使用时需要进行维修，那么这种故障率是否可以接受？

**解答：**首先找出在一年内**20**块板中有**5**块或更多出现问题需要进行维修的概率

$$\begin{aligned}\sum_{n=5}^{20} f(n; 20, 0.1) &= 1 - \sum_{n=0}^4 f(n; 20, 0.1) \\ &= 1 - \sum_{n=0}^4 \frac{20!}{n!(20-n)!} 0.1^n (1-0.1)^{20-n} \\ &= 1 - 0.9568 = 0.0432\end{aligned}$$

**5块以上发生故障的概率非常少，所以板的质量不能接受。**

# 从二项到多项分布

类似于二项分布, 但允许结果的可能性  $m$  大于两种, 概率为

$$\left\{ \begin{array}{l} \vec{p} = (p_1, p_2, \dots, p_m) \\ \sum_{i=1}^m p_i = 1 \end{array} \right. \quad \begin{array}{l} \text{尝试 } N \text{ 次, 结果为} \\ \text{可能性1: } n_1 \\ \text{可能性2: } n_2 \\ \dots \end{array} \quad \rightarrow \quad \vec{n} = (n_1, n_2, \dots, n_m)$$

得到  $(n_1, n_2, \dots, n_m)$  概率为

$$f(\vec{n}) = \frac{N!}{n_1! n_2! \dots n_m!} p_1^{n_1} p_2^{n_2} \dots p_m^{n_m}$$

$$\text{方差: } V[n_i] = Np_i(1 - p_i)$$

$$\text{协方差: } V_{ij} = -Np_i p_j \quad (i \neq j)$$

$$\text{平均值: } E[n_i] = Np_i$$

适用于直方图频数不确定性估计。

# 泊松分布

泊松分布是二项分布在  $N \rightarrow \infty$ ,  $p \rightarrow 0$  和  $Np = \text{常数 } \nu$  的极限形式。

$$f(n; \nu) = \frac{\nu^n}{n!} e^{-\nu}$$

平均值：

$$E[n] = \mu = \sum n f = \nu$$

方差：

$$\begin{aligned} V[n] &= \sigma^2 = E[(n - \mu)^2] \\ &= E[n^2] - E^2[n] \\ &= \nu \end{aligned}$$

著名的统计不确定度估计式

$$n \pm \sqrt{n}$$

# 泊松分布式是二项分布的近似

概率的第三公理：如果  $A_1, A_2, A_3, \dots$  是在空间  $S$  中互斥事例的有限或无限序列，则

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$$

$$\sum_{n=0}^{\infty} \frac{e^{-\nu} \nu^n}{n!} = e^{-\nu} \sum_{n=0}^{\infty} \frac{\nu^n}{n!} = e^{-\nu} e^{\nu} = 1 \quad (\text{函数为 } e^x \text{ 的麦克劳林公式})$$

$$\begin{aligned} f(n; N, \nu) &= \frac{N!}{n!(N-n)!} \left(\frac{\nu}{N}\right)^n \left(1 - \frac{n}{N}\right)^{N-n} \\ &= \frac{N(N-1)(N-2)\dots(N-n+1)}{n! N^n} \nu^n \left(1 - \frac{n}{N}\right)^{N-n} \\ &= \frac{(1 - \frac{1}{N})(1 - \frac{2}{N})\dots(1 - \frac{n-1}{N})}{n!} \nu^n \left(1 - \frac{\nu}{N}\right)^{N-n} \end{aligned}$$

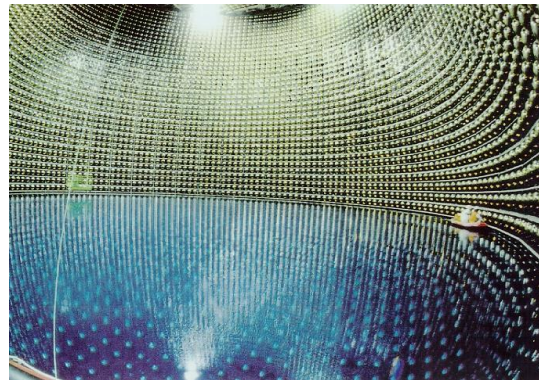
$$\left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{n-1}{N}\right) \xrightarrow{N \rightarrow \infty} 1$$

$$\begin{aligned} \left(1 - \frac{\nu}{N}\right)^{N-n} &= \left[ \left(1 - \frac{\nu}{N}\right)^{N/\nu} \right]^{\nu} \\ \times \left(1 - \frac{\nu}{N}\right)^{-n} &= \lim_{\substack{x = \frac{N}{\nu} \rightarrow \infty \\ \nu}} \left[ \left(1 + \frac{1}{x}\right)^{-x} \right]^{\nu} \\ &\rightarrow e^{-\nu} \end{aligned}$$



# 举例：光电倍增管暗电流影响

- 在有11146根PMT的探测器中，已知每根PMT暗电流产生的误响应为3.5kHz。求探测器在任意总长度为500 $\mu$ s时间段观察到每隔10ns PMT误击中数目分别为5和6的总次数



日本超级神冈中微子探测器

在10ns间隔观测到PMT误击中的平均数目为

$$10^{-8} \times 11146 \times 3500 = 0.4$$

500 $\mu$ s  
平均数

$$\begin{aligned} = 5 &\Rightarrow 5 \times 10^4 \times \frac{0.4^5}{5!} \times e^{-0.4} \cong 3(\text{次}) \\ = 6 &\Rightarrow 5 \times 10^4 \times \frac{0.4^6}{6!} \times e^{-0.4} \cong 0.2(\text{次}) \end{aligned}$$

2020-03-07 这一结论影响到我们在数据分析中应采取的对策。


# 二项分布与泊松分布

假设一学生站在路边想搭便车。过路的汽车平均频率为每分钟一辆，服从泊松分布。而每辆车让搭便车的概率为1%，计算该学生在过了60辆车以后还未能搭上车的可能性

$N=60, p=0.01, r=0$   特点：N大 p 小

根据二项分布：
$$\frac{60!}{0!(60-0)!} 0.01^0 (1-0.01)^{60-0} = 0.5472$$

根据泊松分布：
$$e^{-60 \times 0.01} \frac{(60 \times 0.01)^0}{0!} = 0.5488$$

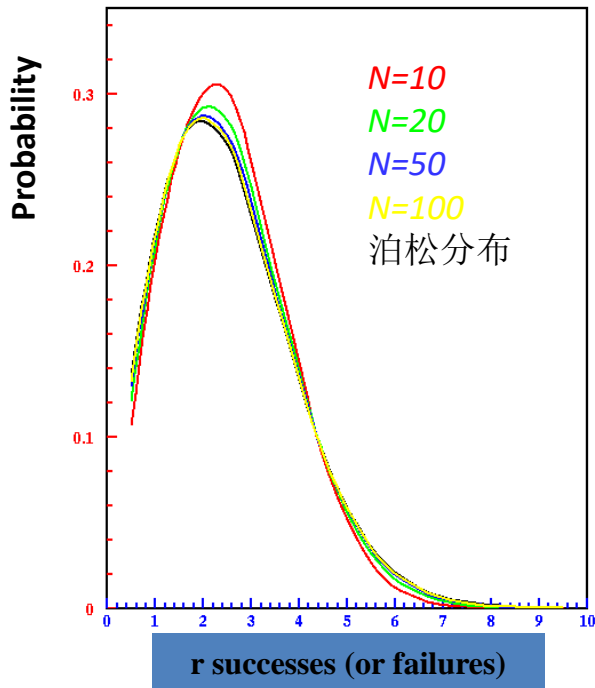


泊松分布是二项分布的近似。

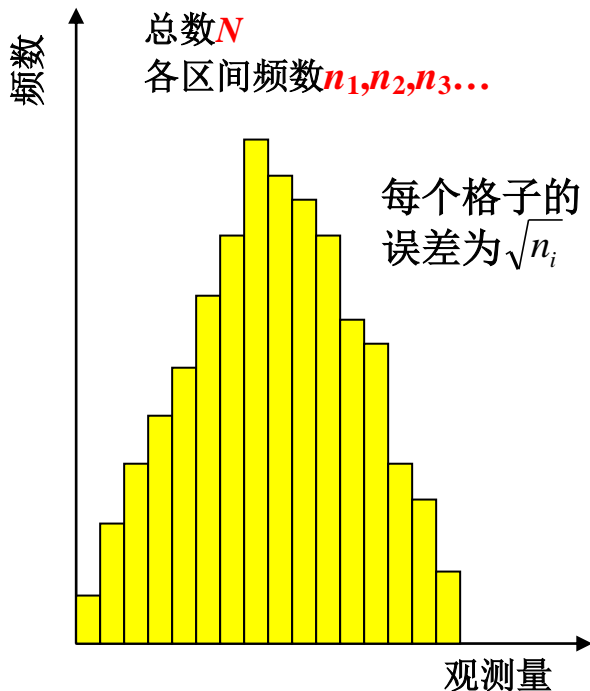
# 泊松分布是二项分布的近似

例如：对于以平均值为2的泊松分布而言，相当于二项分布中的 $Np=2$ 。

当 $N$ 值增大时，为了保持 $Np$ 不变， $p$ 值相应减小。可以从右图看出，当 $N$ 大于50时，两种分布的区别几乎可以忽略。



# 直方图中的误差处理



## 一个直方图可看成与

1. 一个事例总数满足泊松分布和在每个区间得到 $n_1, n_2, n_3, \dots$ 事例数为多项分布有关；
2. 或者是直方图中每个区间互相独立的泊松分布有关。

$$\Delta N = \sqrt{N}$$

或

$$\begin{aligned}(\Delta N)^2 &= (\Delta n_1)^2 + (\Delta n_2)^2 + (\Delta n_3)^2 + \dots \\ &= n_1 + n_2 + n_3 + \dots \\ &= N\end{aligned}$$

注意：当  $N < 5$  时误差估计会有很大的偏差。

# 高斯或正态分布

高斯函数具有连续性与对称性，概率密度为

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \text{记为 } N(\mu, \sigma)$$

$$\text{平均值 : } E[x] = \int xP(x)dx = \mu$$

$$\text{方差 : } V[x] = E[(x - \mu)^2] = E[x^2] - E^2[x] = \sigma^2$$

在所有统计问题扮演核心角色，应用于所有科学研究领域所涉及的分布。测量不确定度，特别是仪器误差通常用高斯函数来描述其概率分布。即使在应用中可能有不恰当的地方，仍然可提供与实际情况相近的很好近似。

# 中心极限定理

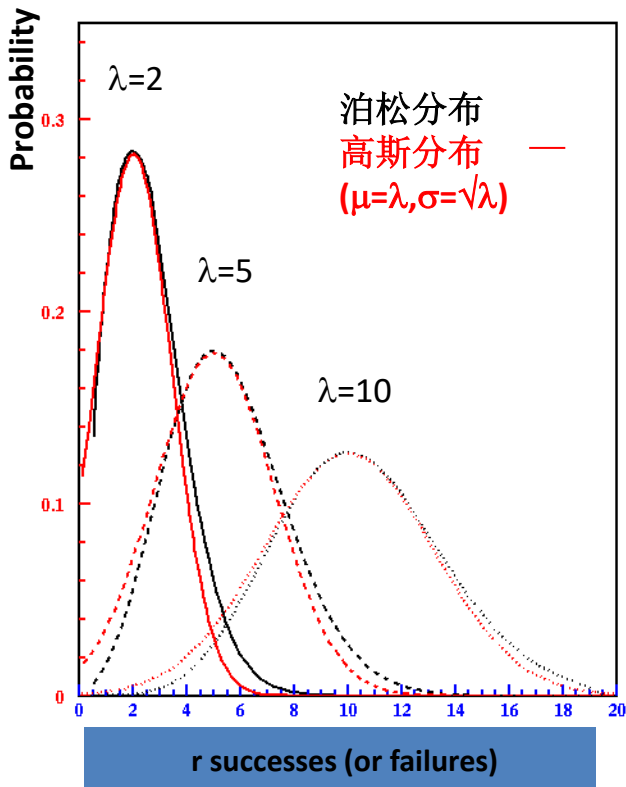
对于  $n$  个独立的随机变量  $x_i$ , 如果每个  $x_i$  都服从平均值为  $\mu_i$  和有限的方差  $\sigma_i^2$  分布, 那么变量

$$\frac{\left( \sum_{i=1}^n x_i - \sum_{i=1}^n \mu_i \right)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \xrightarrow{n \rightarrow \infty} \text{趋于 } N(0,1) \text{ 的正态分布}$$

因此, 如果

$$y = \sum_{i=1}^n x_i \Rightarrow E[y] = \sum_{i=1}^n \mu_i, V[y] = \sum_{i=1}^n \sigma_i^2$$

# 高斯分布与泊松分布



- ✓ 泊松分布只有非负整数定义。
- ✓ 高斯分布是连续且可延伸到正负无穷。
- ✓ 当泊松分布的平均值越大，与高斯分布的区别就越小。
- ✓ 实际应用时，当计数或事例数大于5时，可认为误差满足高斯分布。

# 多维高斯分布

对于随机变量  $\vec{x} = (x_1, \dots, x_n)$  其多维高斯函数概率密度为

$$f(\vec{x}; \vec{\mu}, V) = \frac{1}{(2\pi)^{n/2} |V|^{1/2}} \exp\left[-\frac{1}{2} (\vec{x} - \vec{\mu})^T V^{-1} (\vec{x} - \vec{\mu})\right]$$

相应的平均值与协方差为

$$E[x_i] = \mu_i, \quad \text{cov}[x_i, x_j] = V_{ij}$$

对于二维情形，其概率密度函数可表示为

$$f(x_1, x_2; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \quad \rho = \text{cov}[x_1, x_2]/(\sigma_1\sigma_2)$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)} \left[ \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1}\right) \left(\frac{x_2 - \mu_2}{\sigma_2}\right) \right] \right\}$$



## 二项、泊松与高斯分布的联系

$$f(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

$$f(n) = \frac{\nu^n}{n!} e^{-\nu}$$

二项分布

$N \rightarrow \infty, Np = \mu = \nu$

泊松分布

$N \rightarrow \infty$

$\mu \rightarrow \infty$

高斯分布

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

# 对数正态分布

如果连续变量  $y$  是具有均值为  $\mu$  方差为  $\sigma^2$  的高斯量，那么  $x = e^y$  服从对数正态分布。

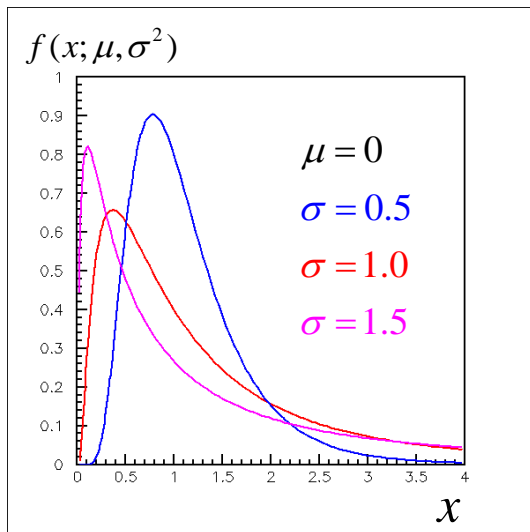
$$f(x; \mu, \sigma^2) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left(\frac{-(\log x - \mu)^2}{2\sigma^2}\right) & \text{对于 } x > 0, \sigma > 0 \\ 0 & \text{其它情况} \end{cases}$$

平均值：  $E[x] = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$

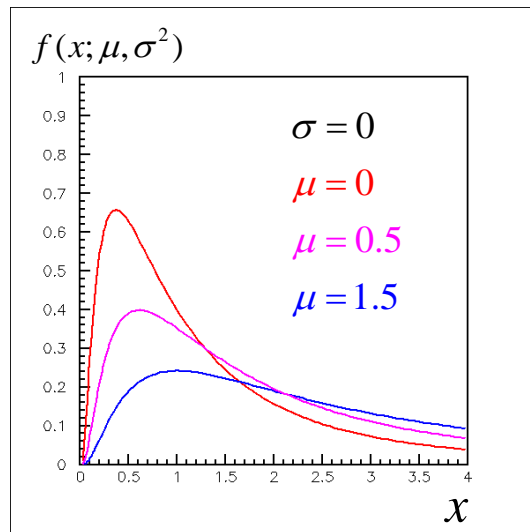
方差：  $V[x] = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$

# 不同均值和方差的对数正态分布

同均值不同方差



同方差不同均值



对数正态分布表示一个随机变量其对数服从正态（高斯）分布，提供了一个模型处理类似涉及许多小的倍增误差过程的误差。也适用于观测值是一个正比于过去观测的随机变量。

# 指定区间的对数正态分布计算

如果需要估计服从对数正态分布的随机变量在区间 ( $0 < a < b$ ) 的概率值, 需要计算

$$\int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left(\frac{-(\log x - \mu)^2}{2\sigma^2}\right) dx$$

对积分做代换  $y = \log(x)$ , 则可以得到所求的概率值

$$\begin{aligned} & \int_{\log a}^{\log b} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y - \mu)^2}{2\sigma^2}\right) dy \quad \text{令 } t = \frac{y - \mu}{\sigma} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\log b - \mu}{\sigma}} \exp\left(-\frac{t^2}{2}\right) dt - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\log a - \mu}{\sigma}} \exp\left(-\frac{t^2}{2}\right) dt \\ &= F\left(\frac{\log b - \mu}{\sigma}\right) - F\left(\frac{\log a - \mu}{\sigma}\right) \end{aligned}$$

可从正态分布  $N(0,1)$  表查中出积分值。

# 对数正态分布用于风险分析

核电站工程师必须采用模型来估计支撑蒸汽发电机的强度，以防止由于地震峰值加速度造成的破坏。专家的意见建议该强度的对数是具有  $\mu=4.0$  和  $\sigma^2=0.09$  正态分布。试估计但峰值加速度为 33 时，支撑系统依然可以承受的概率。

解答：

$$1 - F\left(\frac{\log(33) - 4.0}{0.30}\right) = 1 - F(-1.68) = 0.9535$$

或者说系统崩溃瓦解的概率为

$$F\left(\frac{\log(33) - 4.0}{0.30}\right) = F(-1.68) = 0.0465$$

注：大亚湾中微子实验隧道爆破最大允许当量也是采用类似方法估计。

# 均匀分布

在区间  $(a, b)$  上均匀分布的连续随机变量  $x$ ，其概率密度函数为

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta - \alpha} & \text{当 } \alpha \leq x \leq \beta \\ 0 & \text{其它} \end{cases}$$

$$\text{平均值: } E[x] = \mu = \frac{\alpha + \beta}{2}$$

均匀分布是用蒙特卡罗模拟随机现象的基础。

$$\text{方差: } V[x] = \sigma^2(x) = \frac{(\beta - \alpha)^2}{12}$$

# 指数分布

对于连续变量  $x$  ( $0 \leq x < \infty$ ) 指数分布,

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

平均值:  $E[x] = \frac{1}{\xi} \int_0^{\infty} x e^{-x/\xi} dx = \xi$

方差:  $V[x] = \frac{1}{\xi} \int_0^{\infty} (x - \xi)^2 e^{-x/\xi} dx = \xi^2$

常用于描述粒子寿命。

# $\chi^2$ - 分布

如果  $x_1, \dots, x_n$  是相互独立的高斯随机变量，按下列方式求和

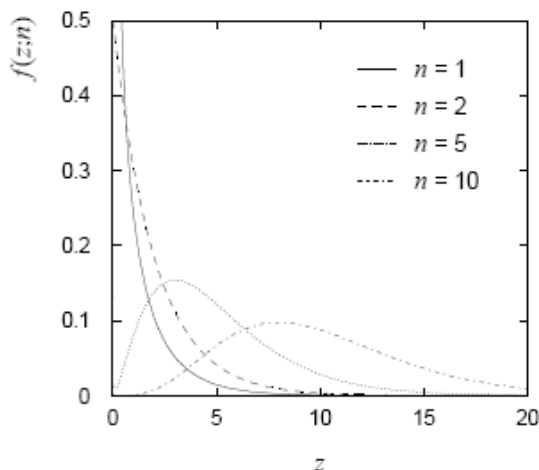
$$z = \sum_{i=1}^n (x_i - \mu_i)^2 / \sigma_i^2$$

服从自由度为  $n$  的  $\chi^2$  的 p.d.f 为

$$f(z; n) = \frac{z^{n/2-1}}{2^{n/2} \Gamma(n/2)} e^{-z/2}, (z \geq 0)$$

$\Gamma$  函数的定义为

$$\Gamma(r) \equiv \int_0^{\infty} x^{r-1} e^{-x} dx$$



平均值:  $E[z] = \mu = n$

202 方差:  $V[z] = \sigma^2 = 2n$

$\chi^2$ -分布通常用来检验假设与实际情况的符合程度。



# 柯西(布莱特-魏格纳)分布

对于连续随机变量  $x$  的柯西p.d.f.为

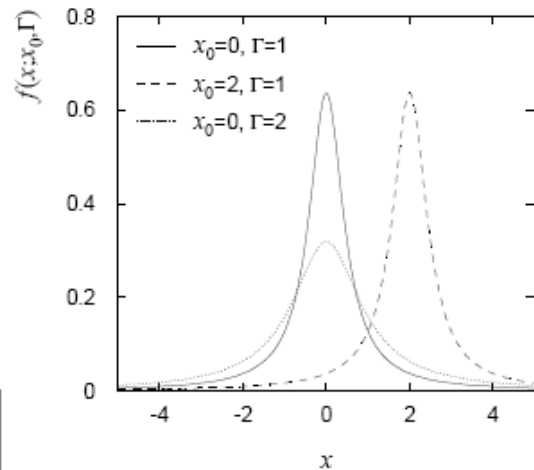
$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

是布莱特-魏格纳p.d.f.的一个特例

$$f(x; \Gamma, x_0) = \frac{1}{\pi} \frac{\Gamma/2}{\Gamma^2/4 + (x - x_0)^2}$$

其中,  $x_0$ ,  $\Gamma$ =(半高)宽度

在粒子物理中, 常用于描述“共振态”粒子的不变质量分布。



# 朗道分布

对于具有速度为 $\beta=v/c$ 的带电粒子穿过一层厚度为 $d$ 的物质，其能量损失 $\Delta$ 服从朗道p.d.f

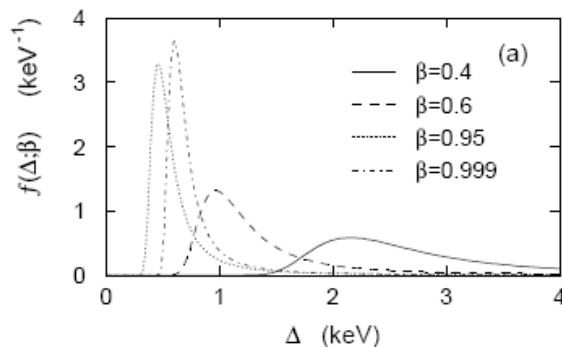
$$f(\Delta; \beta) = \frac{1}{\xi} \phi(\lambda),$$

$$\phi(\lambda) = \frac{1}{\pi} \int_0^\infty \exp(-u \log u - \lambda u) \sin(\pi u) du,$$

$$\lambda = \frac{1}{\xi} \left[ \Delta - \xi \left( \log \frac{\xi}{\varepsilon'} + 1 - \frac{1}{\sqrt{1-\beta^2}} \right) \right],$$

$$\xi = \frac{2\pi N_A e^4 z^2 \rho \Sigma Z}{m_e c^2 \Sigma A} \frac{d}{\beta^2},$$

$$\varepsilon' = \frac{I^2 (1-\beta^2) \exp(\beta^2)}{2m_e c^2 \beta^2}, \quad I = \text{平均激发能}$$



厚度 $d$ 增大时，趋于正态分布。

常用于描述粒子的电离能损  
或能量沉积。

# 如何通过实验对随机变量进行研究？

---

- ✓ 蒙特卡罗方法
- ✓ 随机数产生子
- ✓ 任意分布抽样之函数变换法与舍选法
- ✓ 蒙特卡罗方法中的精度问题
- ✓ 在粒子物理与核物理中的应用

# 蒙特卡罗方法简介

蒙特卡罗方法就是利用一系列随机数来计算各种概率大小和随机变量均值等等的数值分析技术。通常的步骤为：

- 1) 产生一系列在 $[0, 1]$ 之间均匀分布的随机数  $r_1, r_2, \dots, r_m$  。
- 2) 利用这些随机数按某些概率密度函数  $f(x)$  抽样生成我们感兴趣的另一随机序列  $x_1, x_2, \dots, x_n$ 。
- 3) 利用这些  $x$  值来估计  $f(x)$  的一些特性，例如：通过找到在区间  $a \leq x \leq b$  的  $x$  比例，给出积分值  $\int_a^b f(x)dx$ 。

第一层面上的应用： 蒙特卡罗计算 = 积分

第二层面上的应用： 蒙特卡罗变量 = “模拟的数据”

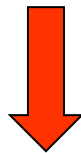
# 随机数的产生

用物理方法产生  
真正的随机数



- 不可重复
- 产生速度慢

用数学方法产生  
伪随机数



- 可以重复
- 产生的速度快

# 真随机数与伪随机数

美国兰德（RAND）公司在1950年代，利用真空管中产生的噪音制作了一个含十万个真正的随机数表，并运用于其开展的所有模拟研究中。

真正的随机数与伪随机数之间的区别在于：数据串是否具有可压缩性，即能否用更短的形式来表示。

真正的随机数是不可压缩的，非常不规则，以至于无法用更短的形式来表示它。

在粒子物理与核物理研究中，随机数的可重复性经常也是非常有用的，尤其是程序的调试(debugging)。

# 随机数产生子

目的是使在  $[0, 1]$  范围内产生的伪随机数满足：

均匀性；相互独立性；长周期性

## 乘同余法

$\xi_i \equiv \text{余数}(\lambda \xi_{i-1} / M) \Rightarrow \text{随机数 } r_i = \xi_i M^{-1}$   
 $\lambda$  与  $M$  为选定常数,  $\xi_0$  为随机数种子。

## 友情推荐

$M=2^K$	$\lambda=5^{2q+1}$	$\xi_0$	周期= $2^{K-2}$
$2^{32}$	$5^{13}$	1	$2^{30} \approx 10^9$
$2^{36}$	$5^{13}$	1	$2^{34} \approx 2 \cdot 10^{10}$
$2^{42}$	$5^{17}$	1	$2^{40} \approx 10^{12}$

# 用蒙特卡罗法计算积分

对于计算积分值  $\int_a^b f(x)dx$

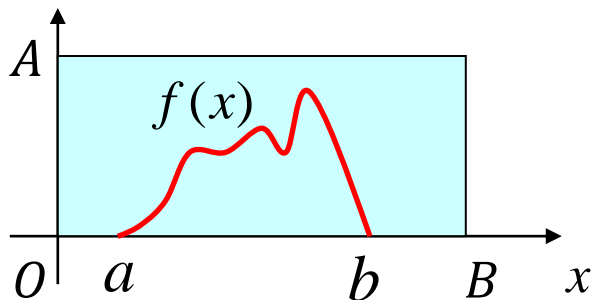
函数须解析可积

解析解:  $\int_a^b f(x)dx = F(x)|_{x=b} - F(x)|_{x=a}$

数值解:  $\int_a^b f(x)dx \approx \sum_{i=1}^n f(x_i) \frac{b-a}{n}$

自变量不能太多

蒙特卡罗方法:



在AB区间均匀投总数为  $N$  个点, 其中  $n$  个位于  $f(x)$  与  $x$  轴所围区域。

$$\int_a^b f(x)dx \approx (AO) \cdot (BO) \cdot \frac{n}{N}$$

对函数是否解析可积和  
是否太多自变量无要求



# 蒙特卡罗方法中的精度问题

采用蒙特卡罗方法(MC)计算积分与传统的梯形法相比有如下特点

## 一维积分:

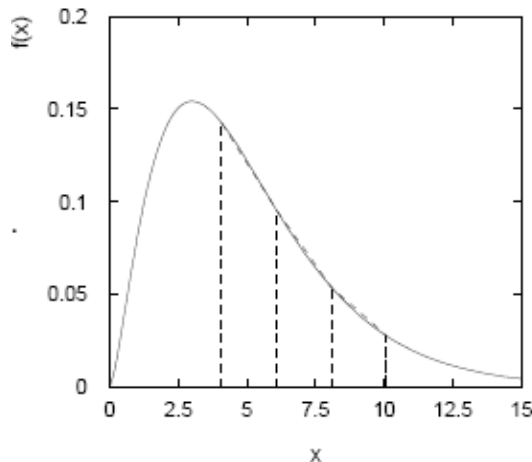
MC 精度:  $\propto 1/\sqrt{n}$ , ( $n$  为产生的随机数)

梯形法精度:  $\propto 1/n^2$ , ( $n$  为子区间的数目) ✓

## 多维积分:

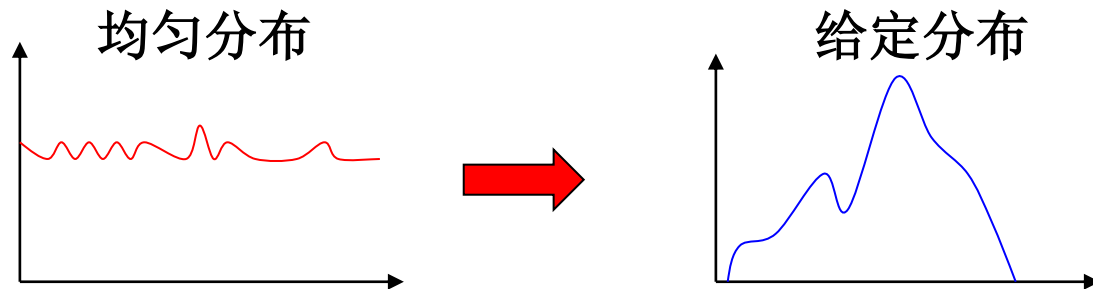
MC 精度:  $\propto 1/\sqrt{n}$  且与维数无关, ( $n$  为产生的随机数) ✓

梯形法精度:  $\propto 1/n^{2/\text{维数}}$ , ( $n$  为子区间的数目)



对于维数大于4的积分, 用蒙特卡罗方计算积分总是最好。

# 从均匀分布到任意分布的随机数



## 函数变换法

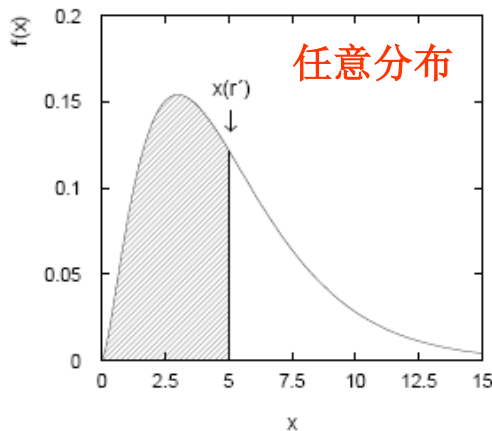
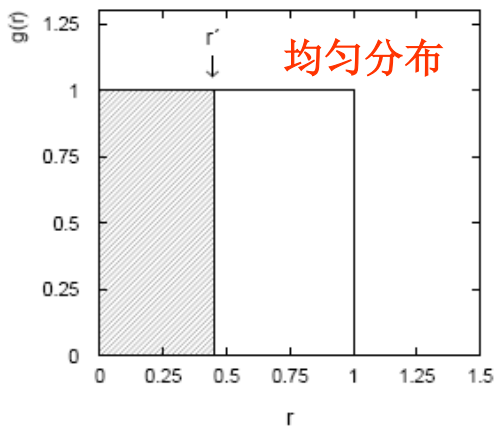
寻找某个函数，当函数的自变量取均匀分布值时，对应的函数值自动满足给定分布。

## 舍选法

从一个随机变量与对应概率密度函数最大值构成的二维均匀分布中，按概率密度函数与自变量关系曲线切割得到。

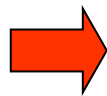
# 函数变换法

从在  $[0,1]$  均匀分布的随机数, 通过适当的变换  $x(r)$  找出服从  $f(x)$  分布的随机数  $x_1, x_2, \dots, x_n$



要求:  $P(r \leq r') = P(x \leq x(r'))$ ,

$$\text{即 } \int_{-\infty}^{r'} g(r) dr = r' = \int_{-\infty}^{x(r')} f(x') dx' = F(x(r'))$$



令  $F(x) = r$

解出  $x = F^{-1}(r)$

# 例子:指数分布抽样

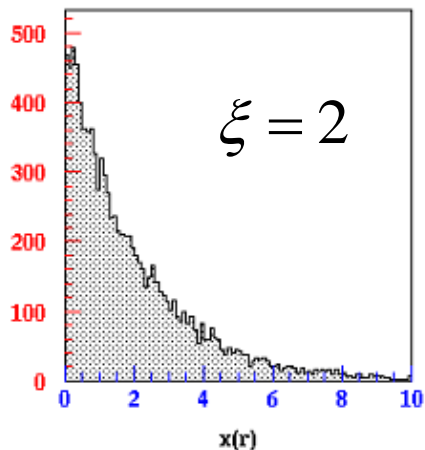
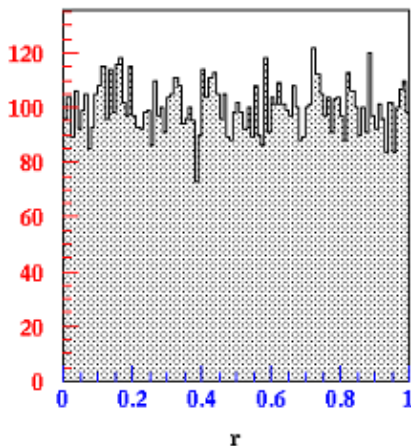
指数概率密度函数： $f(x, \xi) = \frac{1}{\xi} e^{-x/\xi} \quad (x \geq 0)$

令  $\int_0^{x(r)} \frac{1}{\xi} e^{-x'/\xi} dx' = r$  并解出  $x(r)$ , 得到

$$x(r) = -\xi \log(1-r)$$

随机变量  $r$  与  $x(r)$   
呈一一对应关系

抽样效率  
为100%。



# 舍选法

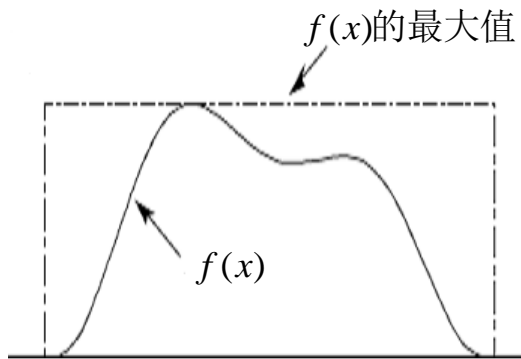
根据概率密度函数  $f(x)$  的自变量  $x$  取值范围,由第一个随机变量产生均匀分布的  $x$ ,即:  $x = x_{\min} + r_1(x_{\max} - x_{\min})$



产生第二个独立的随机变量,均匀分布在 0 与  $f_{\max}$  范围内,即  $u = r_2 f_{\max}$



如果  $u < f(x)$ ,则接受该  $x$  值,否则,拒绝该  $x$  值,从新进行抽样。



$x$

问题: 如何找到函数的最大值?

# 舍选法举例

```
subroutine acc_rej
real rvec(1)
call hbook1(10,'x(r)',100,0.,10.,0.)
call hbook1(20,'x(r)',100,0.,10.,0.)
call hbook2(30,'f(x) vs. x(r)',100,0.,10.,100,0.,1.1,0.)
fmax=-999.
do i=1,100
    call ranmar(rvec,1)
    r=0+rvec(1)*(10.-0.)
    f=0.5*exp(-r/2.)
    if(fmax.lt.f)fmax=f
end do
fmax=1.2*fmax
ntot=0
do i=1,10000
    call ranmar(rvec,1)
    r=0+rvec(1)*(10.-0.)
    z=0.5*exp(-r/2.)
```

找  $f(x)$  的最大值

PAW > call acc\_rej.f

PAW > zone 1 3; h/pl 10; h/pl 20; h/pl 30

```
if(z.gt.fmax)then
    fmax=z*1.2
    write(6,*)'z greater than fmax'
end if
call hfill(10,r,0.,1.0)
call ranmar(rvec,1)
u=rvec(1)*fmax
if(u.lt.z)then
    call hfill(20,r,0.,1.0)
    call hfill(30,r,u,1.0)
    ntot=ntot+1
end if
end do
write(6,*)'ntot=',ntot
return
end
```

# 舍选法举例

```
void acc_rej() {  
    TH1F *h11 = new TH1F("h11","",100,0,10);  
    TH1F *h12 = new TH1F("h12","",100,0,10);  
    TH2F *h2 = new TH2F("h2","",100,0,10,100,0,1);  
    double fmax=-999.;  
    for (int i=0;i<100;i++) {  
        double r = gRandom->Uniform(0,10);  
        double f = 0.5*exp(-r/2.);  
        if (fmax < f) fmax=f;  
    }  
    fmax *= 1.2;  
    cout << "fmax=" << fmax << endl;
```

```
root[0].x acc_rej.C  
root[1] h11->Draw();  
root[2] h12->Draw();  
root[3] h13->Draw();
```

## ROOT脚本

找  $f(x)$  的最大值

```
int ntot=0;  
for (int i=0;i<10000;i++) {  
    double r=gRandom->Uniform(0,10);  
    double z=0.5*exp(-r/2.);  
    if (z>fmax) fmax=1.2*z;  
    h11->Fill(r);  
    double u=gRandom->Uniform(0,fmax);  
    if (u<z) {  
        h12->Fill(r);  
        h2->Fill(r,u);  
        ntot += 1;  
    }  
}  
cout << "ntot=" << ntot << endl;  
}
```

# 常用概率密度分布函数的抽样

## 高斯(正态)分布

```
{ gROOT->Reset();  
  hx = new TH1F("hx","x dis.",  
                100,-10,10);  
  gRandom->SetSeed();  
  Double_t x;  
  const Double_t sigma=2.0;  
  const Double_t mean=1.0;  
  const Int_t kUPDATE = 1000;  
  for ( Int_t i=0; i<kUPDATE; i++) {  
    x=gRandom->Gaus(mean,sigma);  
    hx->Fill(x); }  
}
```

可以换为

```
x = gRandom->Rndm(i);  
x = gRandom->Uniform(xup);  
x = gRandom->Integer(lmax);  
x = gRandom->Landau(mean,sigma);  
x = gRandom->Binomial(ntot,prob);  
x = gRandom->Poisson(mean);  
x = gRandom->PoissonD(mean);  
x = gRandom->Exp(tau);  
x = gRandom->BreitWigner(me,sig);
```

产生平均值为mean  
标准偏差为sigma的  
高斯分布。

在ROOT环境下采用已有的分布，可以容易完成布置的练习。

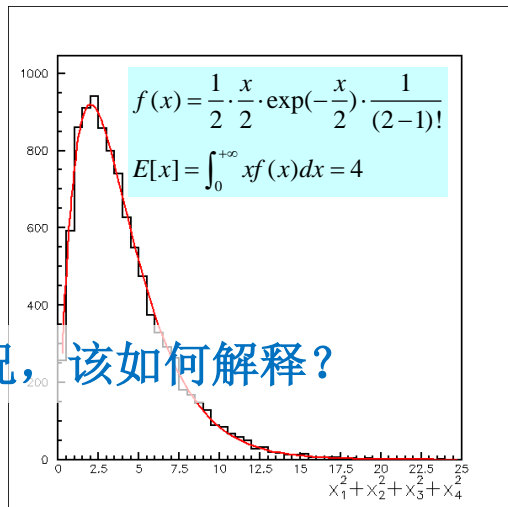
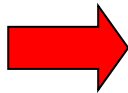
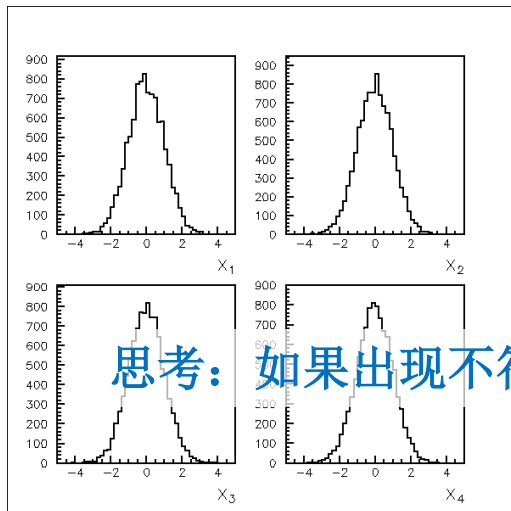


# 蒙特卡罗统计检验

例如常用来检验理论与实验符合好坏的 $\chi^2$  分布。

四个服从  $N(0,1)$  正态  
分布的且相互独立的  
随机变量平方和

一定符合自由度  
为 4 的  $\chi^2$  分布



# 粒子与核物理中模拟的应用

---

- 用于实验初期的设计阶段建模分析
- 用于了解实验可能遇到物理过程的基本特征
- 用于了解实验仪器自身所受到的各种影响因素与所影响的大小
- 用于数据分析阶段的系统分析
- ...

# 带电粒子在水中的输运过程模拟

给定带电粒子的四动量



单位厘米产生多少光子？



从均匀分布中产生满足一定波长分布的光子



沿朗伯-科夫光锥方向均匀给所有光子动量



每个光子开始在水中传播

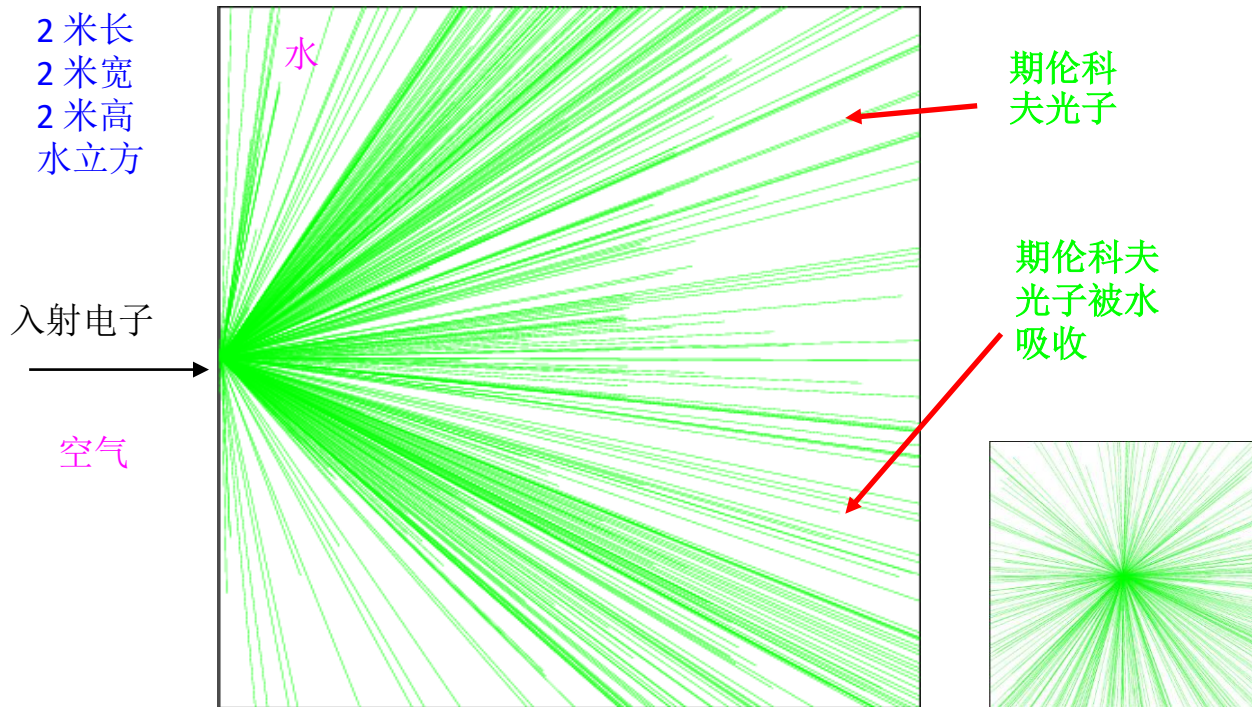


按光与水分子发生作用的概率抽样该光子是否被吸收或散射



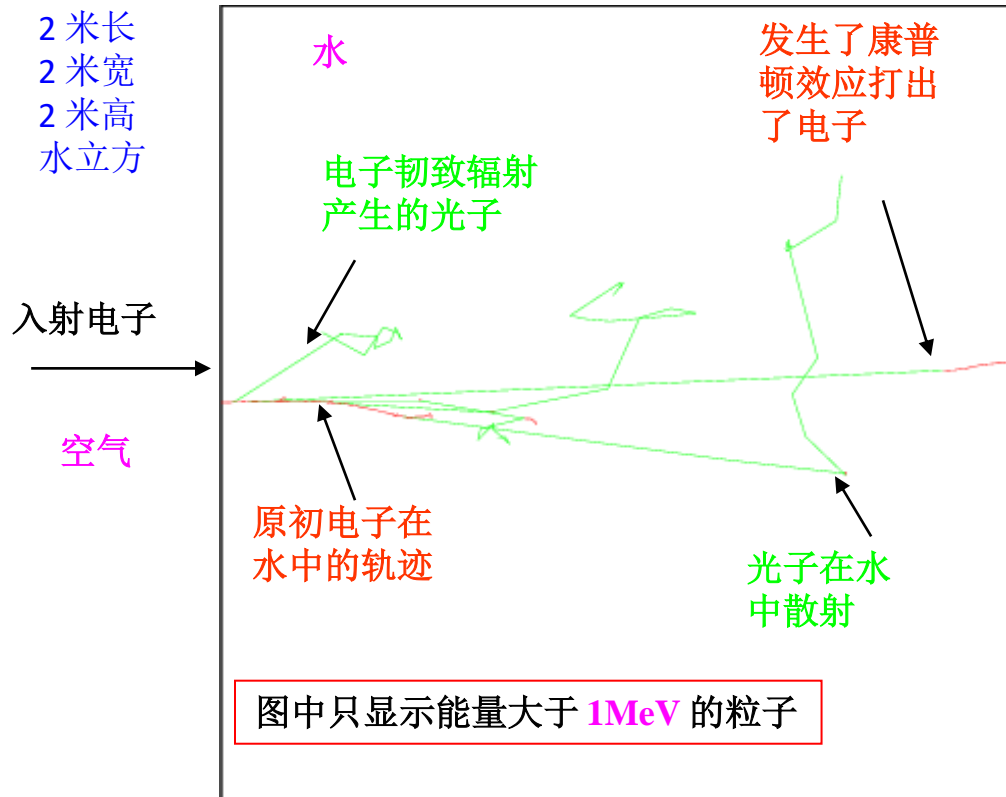
...

# 2 MeV 电子在水中的运输过程

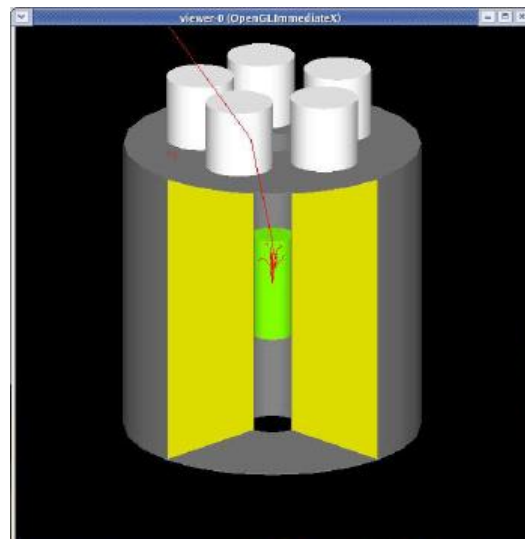


模拟结果显示了电子在水中发出期伦科夫光，损失能量直至被停止在水中的过程。

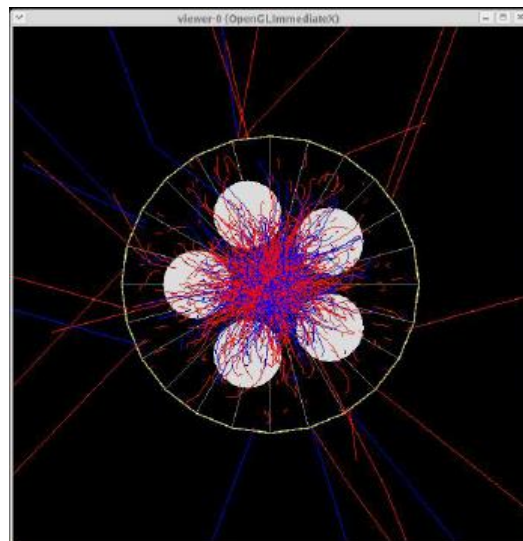
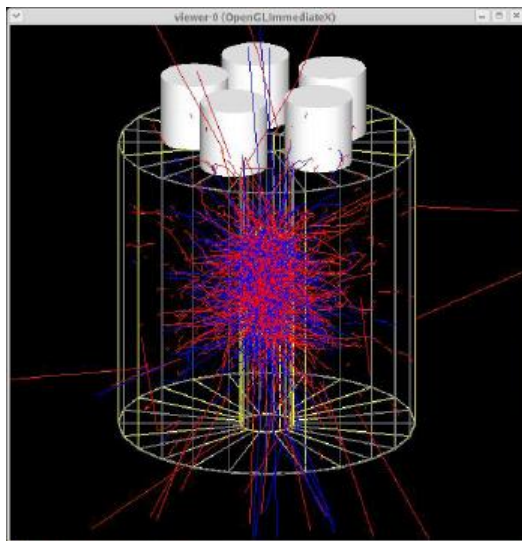
# 200 MeV 电子在水中的运输过程



# 探测器模拟（几何设置）



# 探测器模拟（物理过程）



这种模拟可以提供对探测器效率与预期性能的很好估计。

# 蒙特卡罗方法应用举例

如何确定在实验条件下，理论的概率密度函数

贝叶斯定理：
$$P(\text{理论} | \text{实验}) = \frac{P(\text{实验} | \text{理论})}{P(\text{实验})} P(\text{理论})$$

例如：一质量为  $m$  共振宽度为  $\Gamma$  的共振态在实验上观察到的概率密度函数是什么形式？

布莱特-魏格纳  
分布

⊗

探测器  
分辨率

⊗

探测  
效率

$$BW(M; m_0, \Gamma_0)$$

$$R(M' | M)$$

$$p(\varepsilon = 100\% | M)$$



# 应用举例

也就是说，对应于真实的  $M$ ，实际的  $M'$  应该是怎样一个分布

$$BW(M'; m_0, \Gamma_0) = \frac{\int_{-\infty}^{+\infty} BW(M; m_0, \Gamma_0) R(M' | M) p(\varepsilon = 100\% | M) dM}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} BW(M; m_0, \Gamma_0) R(M' | M) p(\varepsilon = 100\% | M) dM dM'}$$

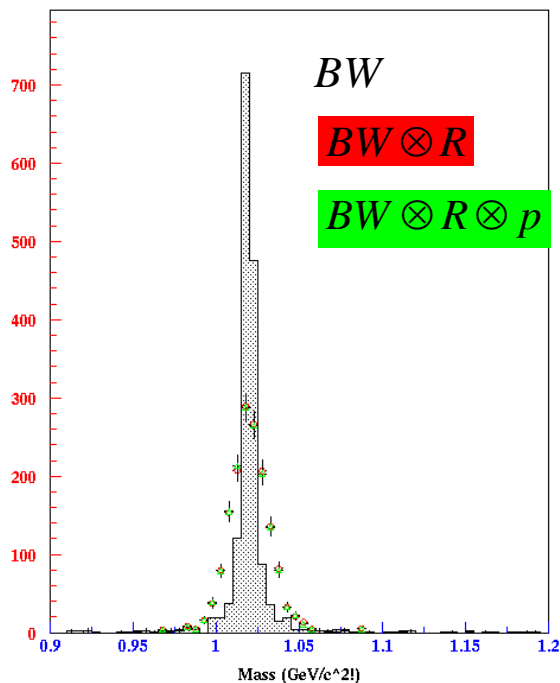
如果假设

$$m_0 = 1.19456 \text{ GeV}/c^2; \Gamma_0 = 0.00426 \text{ GeV}/c^2$$

$$R(M' | M) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(M' - M)^2}{2\sigma^2}\right), \sigma = 0.01 \text{ GeV}/c^2$$

$$p(\varepsilon = 100\% | M) = 0.9(1 - 0.1M^2)$$

# 应用举例



真实物理的图像在实验观测中会发生变化。如果探测器的影响可以用函数来表达，有时积分可积。但大多数数情况下，不能用函数表示时，蒙特卡罗方法可以给出最好的近似。

# 应用举例

应用蒙特卡罗方法的步骤：

步骤一：写出布莱特-魏格纳产生子

输出末态粒子的四动量

步骤二：输入末态粒子四动量，模拟粒子在探测器的响应

输出各子探测器响应

步骤三：输入各子探测器响应，重建探测粒子的四动量

输出探测粒子四动量

步骤四：输入探测粒子四动量，计算不变质量分布

⇒ 实验条件下预期的布莱特-魏格纳分布