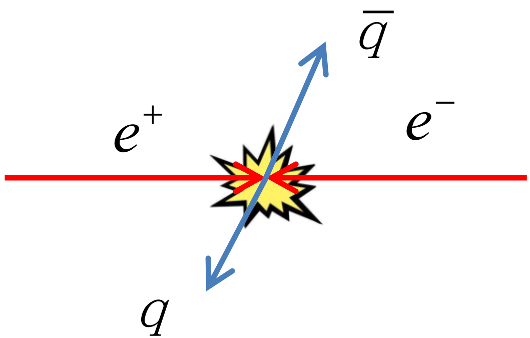
第四讲　基本概念

观察某一过程的*n*个事例



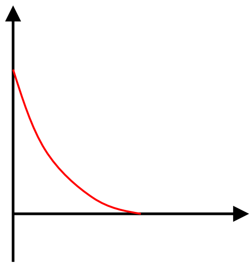
实验测量出每个事例的特征量(能动量，末态粒子数...)。

理论预言出上述各特征量的分布，而且可能还会包含某些 如相互作用耦合常数等自由参数。

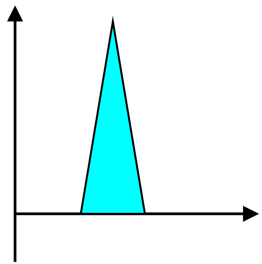
收集数据统计分析

估计参数值与相应的误差围，检验在何种程度上理与实验数据相符。

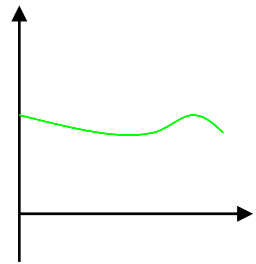
数据背后的物理图像是什么？



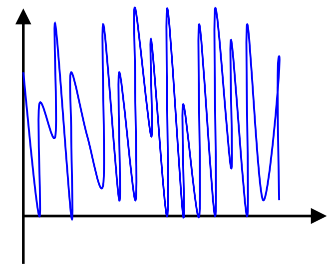
原初物理



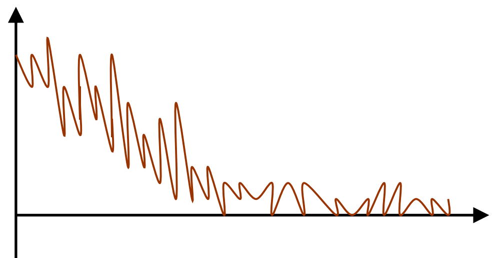
分辨率



探测效率



本底噪音



实验数据

数据分析专业术语：

事例选择，粒子鉴别，选择条件，信噪比优化，无偏选择，效率修正，卷积分辨率，解谱（像）还原...

如何科学地给出物理结论？

收集数据

数据分析

估计参数值与相应的误差范围，检验在何种程度上理论与实验数据相符。

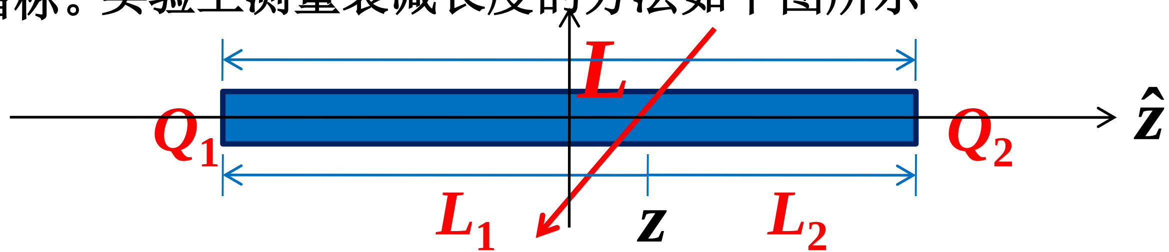
问题：如何评价这种检验？

举例：测量闪烁体衰减长度

光在闪烁体中传播时，具有下列衰减关系

$$Q = Q_0 \exp(-L / L_0)$$

其中，L_0 是闪烁体的衰减长度，它是表征闪烁体质量的一项重要指标。实验上测量衰减长度的方法如下图所示



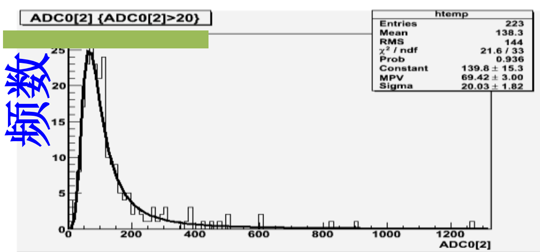
$$Q_0 \propto E, \qquad L_1 = 0.5L + z, \quad L_2 = 0.5L - z,$$
$$Q_1 = 0.5Q_0 \exp(-L_1 / L_0),$$
$$Q_2 = 0.5Q_0 \exp(-L_2 / L_0),$$

$$Q_1 Q_2 = 0.25Q_0^2 \exp(-L / L_0), \quad L_0 = -2z / \ln(Q_1 / Q_2)$$

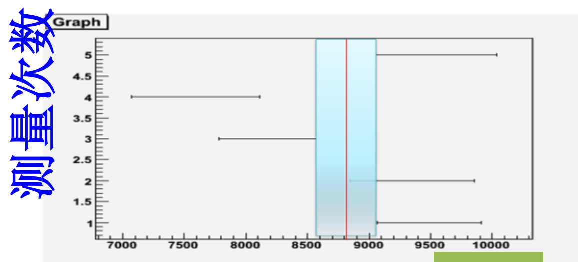
举例：测量闪烁体衰减长度

$$Q_1Q_2 = 0.25Q_0^2\exp(-L/L_0),\quad L_0 = -2z/\ln(Q_1/Q_2)$$

实验采用恒定光源，因此 Q_0 为常数，对待测闪烁体 L_0 也为常数。理论上只要在给定一个位置 z，测量闪烁体两端的电荷输出量即可。但在实际中，往往需要做多点测量。



Q_2



Q_1Q_2

理论上是不变的 Q_1Q_2值，
为什么每次测量都不相同？
能否认为 L_0不是常数？

使用概率来量化结论！

随机事例

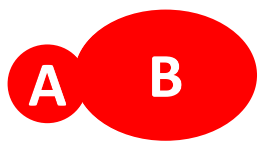
在一定的实验条件下，现象 ***A*** 可能发生，也可能不发生，并且只有发生或不发生这样两种可能性，这是偶然现象中一种比较简单的情形，我们把发生了现象 ***A*** 的事例称为随机事件 ***A***，简称事件 ***A***。也称随机事例

➢ 随机事例之间的关系

✓ A之逆事例

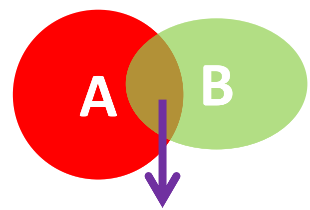
$$A \cap \bar{A} = 0$$

✓ A与B之并事例



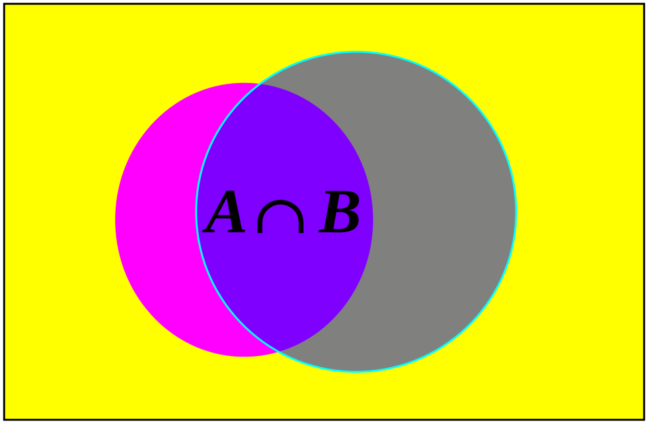
$$A \cup B$$

✓ A与B之积(交)事例



$$A \cap B$$

文恩图（Venn diagram）检验



$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$A \cup (A \cap B) = A$$

$$(A \cap B) \cup (A \cap \overline{B}) = A$$

$$A \cup B = (A \cap B) \cup (A \cap \overline{B}) \cup (\overline{A} \cap B)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

概率的定义

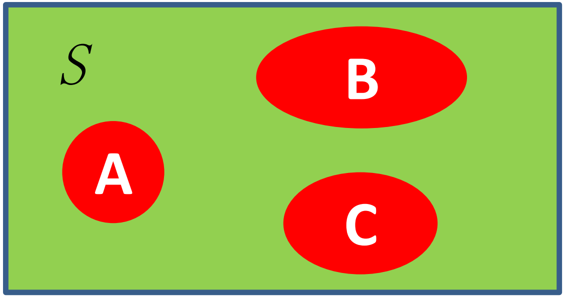
柯尔莫哥洛夫公理：考虑一全集 ***S*** 具有子集 ***A***，***B***，…

$$A \subset S, P(A) \geq 0$$
$$P(S) = 1$$
$$A \cap B = 0 \Rightarrow P(A \cup B) = P(A) + P(B)$$

***P*(*A*)称为事例*A*的概率**

从该公理与文恩图给出的结论可以导出下列概率公式

$$P(\overline{A}) = 1 - P(A)$$
$$P(A \cup \overline{A}) = 1$$
$$A \subset B \Rightarrow P(A) \leq P(B)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



条件概率

假设B出现的概率不为零，在给定B的情况下出现A的条件概率定义为

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

如果

$$P(A \cap B) = P(A)P(B)$$

则，表明A与B相互独立。

如果A与B相互独立，则有

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

结果与B无关

贝叶斯定理

根据条件概率的定义

$$P(A|B)=\frac{P(A\cap B)}{P(B)} \quad 与 \quad P(B|A)=\frac{P(B\cap A)}{P(A)}$$

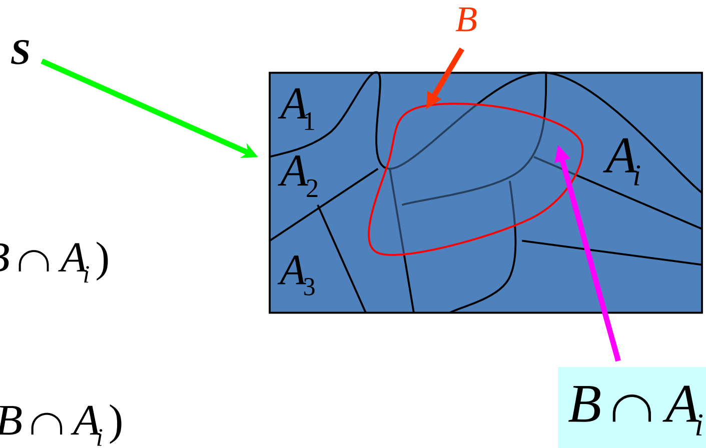
而 $P(A\cap B)=P(B\cap A)$ ，故

$$\boxed{P(A|B)=\frac{P(B|A)P(A)}{P(B)}}$$

贝叶斯定理由 **Reverend Thomas Bayes (1702-1761)** 首先提出。

全概率事例与贝叶斯定理

考虑在样本空间 $\boldsymbol{S}$ 中有一子集 $\boldsymbol{B}$。将样本空间分为互斥的子集 $\boldsymbol{A_i}$，使得



$$\cup_i A_i = \sum_i A_i = S$$

因此，

$$B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i)$$

表示成概率的形式为

$$P(B) = P(\cup_i (B \cap A_i)) = \sum_i P(B \cap A_i)$$

得到全概率事例公式

$$\boxed{P(B) = \sum_i P(B \mid A_i) P(A_i)}$$

贝叶斯定理

$$\boxed{P(A \mid B) = \frac{P(B \mid A) P(A)}{\sum_i P(B \mid A_i) P(A_i)}}$$

例子：如何利用贝叶斯定理

假设对任意一个人而言，感染上AIDS的概率为

$P(AIDS)=0.001$ 验前概率,即任何检验之前

$P(no\ AIDS)=0.999$

考虑任何一次AIDS检查的结果只有阴性(-)或阳性(+)两种

$P(+|AIDS)=0.98$ AIDS感染患者阳性的概率

$P(-|AIDS)=0.02$ AIDS感染患者阴性的概率

$P(+|no\ AIDS)=0.03$ AIDS未感染者阳性的概率

$P(-|no\ AIDS)=0.97$ AIDS未感染者阴性的概率

如果你的检查结果为阳性(+)，而你却觉得自己无明显感染渠道。那么你是否应担心自己真的感染上了AIDS？

例子：如何利用贝叶斯定理

利用贝叶斯定理，阳性结果条件下是AIDS患者的概率为

$$P(AIDS|+)=\frac{P(+|AIDS)P(AIDS)}{P(+|AIDS)P(AIDS)+P(+|no\ AIDS)P(no\ AIDS)}$$

$$=\frac{0.98\times0.001}{0.98\times0.001+0.03\times0.999}$$

$$=0.032\qquad(\text{验后概率})$$

AIDS患者阳性

所有为阳性结果的人

也就是说，你可能没什么问题！？

从你的观点上看：对自己染上AIDS结果的可信度为3.2%。

从医生角度上看：象你这样的人有3.2%感染上了AIDS。

涉及到如何诠释结果（概率）的问题！

概率含义的诠释

- **相对频率（频率论者）**

假设A，B，...是一组可重复实验的结果，则概率就是

$$P(A)=\lim_{n\to\infty}\frac{\text{结果为}A}{n\text{次实验}}$$

- **主观概率（贝叶斯论者）**

如果A，B，...是假设(是真或是假的各种陈述)，那么概率

P（A）：对A为真的信心程度

实际问题中，统计量总是有限的。**P(A)**完全 取决于A的划分与总统计量的大小。

造成概率的大小出现波动。因此，需要解决好A的定义及其误差

频率概率中的问题

- 实际问题中，统计量总是有限的。*P*(*A*)完全取决于*A* 的划分与总统计量的大小。

概率大小会出现波动。

需要解决好
•*A* 的定义
•适当的误差

- 该定义不适用于某些特殊情况

例如：我们可以说“明天有雨”。但是，如果我们根据概率频率定义说“明天可能有雨”，却是一个毫无科学意义的预报。

贝叶斯理论与主观概率

贝叶斯理论通常用于主观概率问题

$$P(理论|实验)=\frac{P(实验|理论)}{P(实验)}P(理论)$$

先验概率：P(理论)；　验后概率：P(理论 |实验)

似然性：P(实验 |理论)

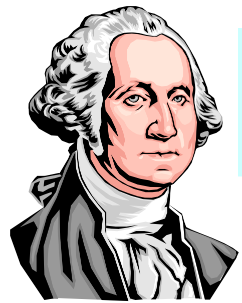
通过实验结果改进基于某一理论的信念(后验性的)

- 如果实验证明P(理论|实验)=0，则表明理论不能接受。
- 大的P(理论|实验)会增加对理论的信任度。
- 通过实验结果可以修改 P(理论)。
- 改进的P(理论)可应用于对重复实验结果的预测。
- P(理论|实验)对先验理论的依赖将最终消失。

主观概率中的问题

- 主观性：在对同一随机现象的描述中，我的P(理论)与你的P(理论)可能不同

理论家甲之理论A



理论家乙之理论B

- 使用主观概率的原因
  - 出于绝望 ✓
  - 出于无知 ✕
  - 出于懒惰 ？

主观概率的一些特点

主观概率有一些吸引人的地方，例如对于不可重复现象的处理中，显得比较自然

- 系统误差(重复实验时仍保持不变)；
- 在该事例出现的粒子是正电子；
- 自然界是超对称的；
- 明天将下雨(将来事件的不确定性)；
- 公元1500年元月一日北京下雨(过去事件的不确定性)。

结论中包含了主观上对事件为真的信念!

频率论者与主观概率

P(938.27195 < 质子质量 < 938.27211 MeV)是什么？

当以质量来判断一实际为质子的粒子类别时

- 频率论者：质子或非质子（不知道是哪个）
- 主观主义者(贝叶斯论者):68%是质子(对知识的陈述)

对主观概率而言，意味着

- 质子质量的不确定性与从100只球中有68只白球的球筐里能拿出白球的不确定性一样。

频率论者与主观概率

如果大多数贝叶斯论者说

- 巴西赢得2010年足球世界杯冠军的概率为68%
- 质子质量在938.27195-938.27211MeV内的概率为68%
- 大陆中国人2020年获诺贝尔奖的概率为68%

那么上述论断的68%就应该理解为结果为真的概率。

能否在频率定义中将质子质量在938.27195-938.27211MeV内理解成：在整个宇宙中，自然界给出了各种不同的质子质量，而它们中有68%在938.27195与938.27211MeV之间?

没问题...只不过这是对信心程度的一种表达。

艾滋病检验结果再认识

$$P(AIDS) = 0.001 \qquad \text{(验前概率)}$$

$$P(AIDS \mid +) = 0.032 \qquad \text{(验后概率)}$$

对于个人而言，**0.032** 是主观概率。如果没有其它额外的信息时，应把 **0.001** 当作相对频率解释。但是往往在病毒检验前，该相对频率被当作一种信念来处理个人是否患病。

如果还有其它额外的信息，应该给出不同的先验概率。这种贝叶斯统计的特点必定是主观的。例如，受检者有过吸毒历史。一旦验前概率改变，贝叶斯定理就会告诉患病的可能性。对阳性结果的诠释就会改变。

随机变量与概率密度函数

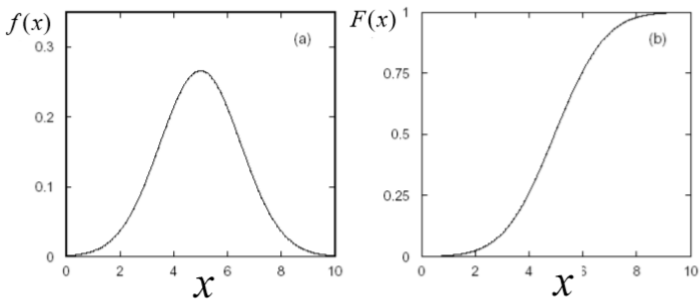
- **假设实验结果为x (记作样本空间中元素)，假设P为观测到x在[x,x+dx]范围内的概率**

$$P = f(x)dx$$

那么**概率密度函数p.d.f.** 定义为f(x)，它满足 $\int_{-\infty}^{+\infty} f(x)dx = 1$

定义累积分布函数：

$$F(x) = \int_{-\infty}^{x} f(x')dx'$$



对于离散型随机变量

$$f_i = P(x_i),\ \sum_{i=1}^{n} f_i = 1,\ F(x) = \sum_{x_i \le x} P(x_i)$$

α分位数、中值与模

分位点 x_α 定义为随机变量 x 的值，它使得

$$F(x_\alpha)=\alpha$$

这里 $0\le\alpha\le 1$。因此可以容易求出分位点

$$x_\alpha=F^{-1}(\alpha)$$

随机变量 x 的中值定义为

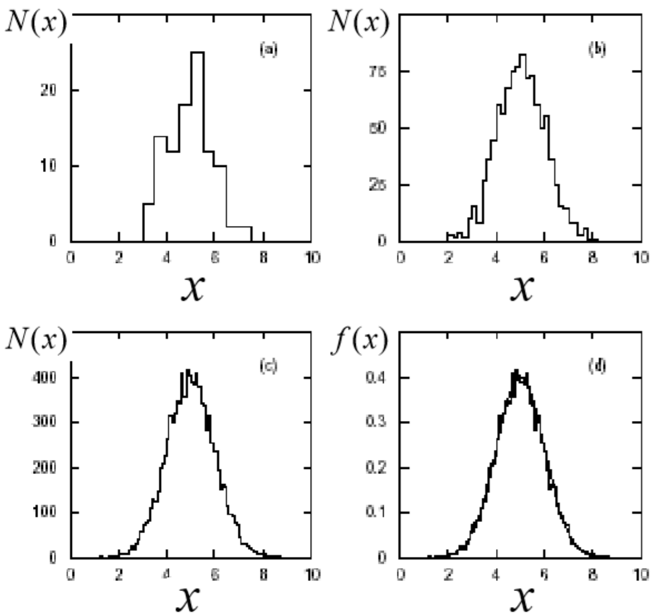
$$x_{1/2}=F^{-1}(1/2)$$

随机变量 x 被观测到大于或小于中值的概率是相等的。

模定义为使概率密度函数值达到极大的随机变量值。

直方图与概率密度函数

- 概率密度函数p.d.f. 就是拥有无穷大样本，区间宽度为零，而且归一化到单位面积的直方图。



$$f(x)=\frac{N(x)}{n\Delta x}$$

$N(x)=$ 每个区间的事例数(频数)

$n=$ 填入直方图的总事例数

$\Delta x=$ 区间的宽度

$$n\to\infty,\ \Delta x\to 0$$

$n\cdot\Delta x$有限

直方图在统计分析中非常重要，应准确理解它的含义。

条件概率密度函数

- 利用条件概率的定义，可得到

$$P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{f(x,y)dxdy}{f_x(x)dx}$$

定义条件概率的密度函数p.d.f. 为

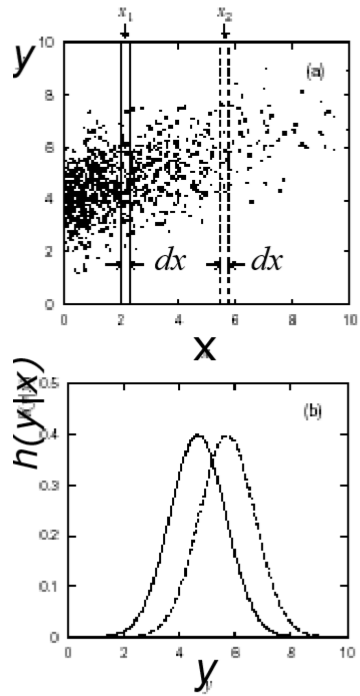
$$h(y|x)=\frac{f(x,y)}{f_x(x)},\quad g(x|y)=\frac{f(x,y)}{f_y(y)}$$

则贝叶斯定理可写为

$$g(x|y)=\frac{h(y|x)f_x(x)}{f_y(y)}$$

若x,y相互独立，则

$$f(x,y)=f_x(x)f_y(y)$$



提醒：概率都是条件概率

由柯尔莫哥洛夫公理，我们定义了概率 ***P*(*A*)**。

但在实际应用中，我们总是对 ***A*** 相对于许多样本空间的概率感兴趣，而不仅仅只是一个空间。因此，通常以记号

$$\boxed{P(A|S)}$$

来表示所进行的研究是在特定的样本空间 ***S*** 中，也就是 ***A*** 相对于 ***S*** 的条件概率。

因此，所有概率在实际应用中都是条件概率。

只有当 ***S*** 的选择是明白无误时，才能简单记为

$$\boxed{P(A|S)} \Rightarrow \boxed{P(A)}$$

证明举例：事例与逆事例

如果 $\boldsymbol{A}$ 是在 $\boldsymbol{S}$ 中的任意一个事例，则

$$P(\bar{A}) = 1 - P(A)$$

证明：由于 $\boldsymbol{A}$ 与 $\bar{\boldsymbol{A}}$ 根据定义是互斥的，并且从文恩图得到

$$A \cup \bar{A} = S$$

因此可以写出

$$\begin{aligned} P(A) + P(\bar{A}) &= P(A \cup \bar{A}) \\ &= P(S) \\ &= 1 \end{aligned} \Rightarrow P(\bar{A}) = 1 - P(A)$$

举例：检查给定概率的合理性

如果一个实验有三种可能并且互斥的结果 $\boldsymbol{A}$，$\boldsymbol{B}$ 和 $\boldsymbol{C}$，检查下列各种情况给出的概率值是否是合理的：

1) $P(A)=1/3, P(B)=1/3, P(C)=1/3$

2) $P(A)=0.64, P(B)=0.38, P(C)=-0.02$

3) $P(A)=0.35, P(B)=0.52, P(C)=0.26$

4) $P(A)=0.57, P(B)=0.24, P(C)=0.19$

举例：检查给定概率的合理性

如果一个实验有三种可能并且互斥的结果 ***A*，*B*** 和 ***C*，**
检查下列各种情况给出的概率值是否是合理的：

1) $P(A)=1/3, P(B)=1/3, P(C)=1/3$

2) $P(A)=0.64, P(B)=0.38, P(C)=-0.02$

3) $P(A)=0.35, P(B)=0.52, P(C)=0.26$

4) $P(A)=0.57, P(B)=0.24, P(C)=0.19$

结论：只有1）与4）是合理的。

评论：作为一个合格的实验研究人员，一定要具备判断结果是否合理的能力！

举例：检查经验概率密度函数

实验上经常经验性地从直方图中给出概率密度函数（例如通过拟合直方图分布等等），但是需要确定得到的函数是否满足概率密度函数的定义，例如

$$1)\ \ f(x)=\frac{x-2}{2}\quad 对于\ x=1,2,3,4$$

$$2)\ h(x)=\frac{x^2}{25}\qquad 对于\ x=0,1,2,3,4$$

试判断哪一个可以用作概率密度函数？

举例：检查经验概率密度函数

实验上经常经验性地从直方图中给出概率密度函数（例如通过拟合直方图分布等等），但是需要确定得到的函数是否满足概率密度函数的定义，例如

$$1)\ \ f(x)=\frac{x-2}{2} \quad 对于\ x=1,2,3,4$$

$$2)\ h(x)=\frac{x^2}{25} \quad 对于\ x=0,1,2,3,4$$

试判断哪一个可以用作概率密度函数？

答案：1）有负概率值；2）累积函数值大于1。因此，两者在给定的随机变量范围内都不能用作概率密度函数。

数据分析中的问题

粒子与核物理实验中对动量的测量通常是分别测量

$$p_{xy} \qquad p_z \qquad f(p_{xy}, p_z)$$

在已知两分量测量值的概率密度函数情况下，总动量为

$$p = \sqrt{p_{xy}{}^2 + p_z{}^2}$$

如何导出总动量的测量值的概率密度函数？

$$g(p)$$

是研究随机变量函数的p.d.f问题。

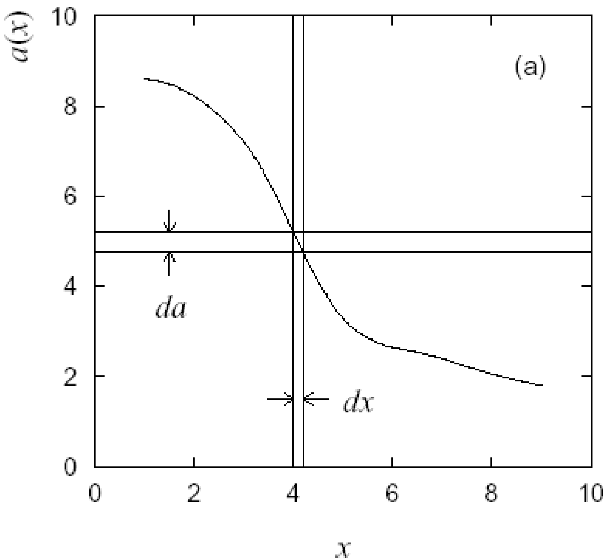
一维随机变量的函数

随机变量的函数自身也是一个随机变量。

例如：

θ与$\cos\theta$

假设 x 服从 p.d.f. $f(x)$，对于函数 $a(x)$，其p.d.f. $g(a)$为何？



$$g(a)da = \int_{dS} f(x)dx$$

$dS = a$ 在$[a, a+da]$内的 x 空间范围

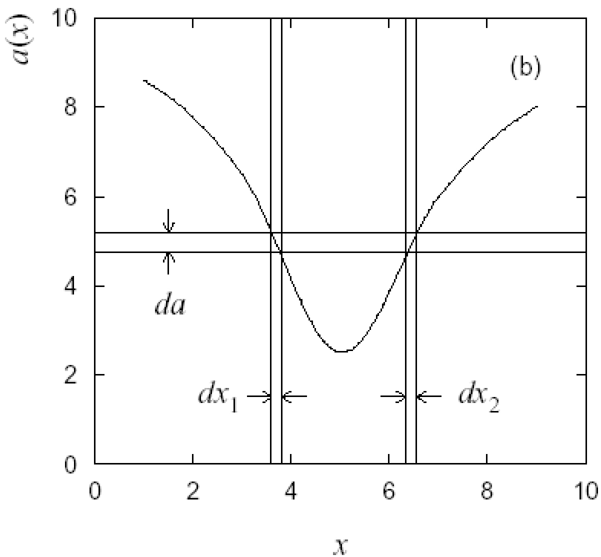
$$g(a)da = \left| \int_{x(a)}^{x(a+da)} f(x')dx' \right|$$

$$= \int_{x(a)}^{x(a)+\left|\frac{dx}{da}\right|da} f(x')dx'$$

$$\Rightarrow g(a) = f(x(a))\left|\frac{dx}{da}\right|$$

函数的逆不唯一情况

假如 $a(x)$ 的逆不唯一，则函数的 p.d.f. 应将 dS 中对应于 da 的所有 dx 的区间包括进来



例如: $a = x^2$, $x = \pm\sqrt{a}$, $dx = \pm\dfrac{da}{2\sqrt{a}}$

$$g(a)da = \int_{dS} f(x)dx$$

$$dS = \left[\sqrt{a}, \sqrt{a} + \frac{da}{2\sqrt{a}}\right] \cup \left[-\sqrt{a} - \frac{da}{2\sqrt{a}}, -\sqrt{a}\right]$$

$$g(a) = \frac{f(\sqrt{a})}{2\sqrt{a}} + \frac{f(-\sqrt{a})}{2\sqrt{a}}$$

多维随机变量的函数

考虑随机矢量 $\vec{x}=(x_1,\dots,x_n)$与函数 $a(\vec{x})$，对应的 **p.d.f.**

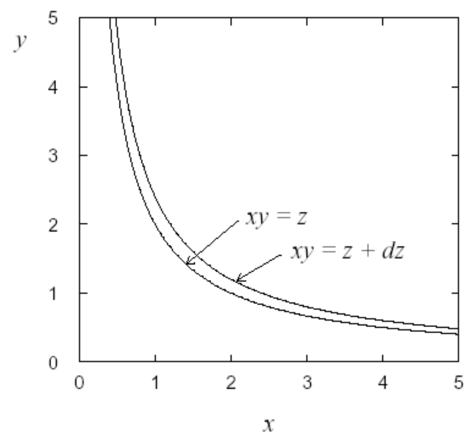
$$g(a')da' = \int\cdots\int_{dS} f(x_1,\dots,x_n)dx_1\dots dx_n$$

$dS=$在$a(\vec{x})=a'$与$a(\vec{x})=a'+da'$定义的曲面$\vec{x}$空间范围

如果两个独立变量 x 与 y，分别按 $g(x)$ 与 $h(y)$分布，那么函数 $z=xy$ 应具有何种形式？

$$f(x,y)=g(x)h(y)$$

$$\begin{aligned} f(z)dz &= \iint_{dS} f(x,y)dxdy \\ &= \iint_{dS} g(x)h(y)dxdy \\ &= \int_{-\infty}^{\infty} g(x)dx \int_{z/|x|}^{(z+dz)/|x|} h(y)dy \end{aligned}$$



多维随机变量的函数

$$f(z)=\int_{-\infty}^{\infty} g(x)h(\frac{z}{x})\frac{dx}{|x|}=\int_{-\infty}^{\infty} g(\frac{z}{y})h(y)\frac{dy}{|y|}$$

记作 g 与 h 的Mellin卷积 ⟹ $\boxed{f=g\otimes h}$

如果函数为 $z=x+y$，则应具有何种形式?

$$f(z)=\int_{-\infty}^{\infty} g(x)h(z-x)dx=\int_{-\infty}^{\infty} g(z-y)h(y)dy$$

记作 g 与 h 的傅立叶卷积 ⟹ $\boxed{f=g\otimes h}$

注意：通常将两者皆称为 g 与 h 的卷积，已相同记号表示。

多维随机变量的函数

考虑具有联合 *p.d.f.* 的随机矢量 $\vec{x}=(x_1,\ldots,x_n)$ ， 构造 n个线性独立的函数：$\vec{a}(\vec{x})=(a_1(\vec{x}),\ldots,a_n(\vec{x}))$，而且其逆函数 $x_1(\vec{a}),\ldots,x_n(\vec{a})$ 存在。那么 $\vec{a}$ 的联合 p.d.f. 为

$$\boxed{g(\vec{a})=|J|\,f(\vec{x})}$$

这里 J是雅可比行列式

$$J=\begin{vmatrix}\dfrac{\partial x_1}{\partial a_1} & \dfrac{\partial x_1}{\partial a_2} & \cdots & \dfrac{\partial x_1}{\partial a_n}\\[2ex] \dfrac{\partial x_2}{\partial a_1} & \dfrac{\partial x_2}{\partial a_2} & \cdots & \dfrac{\partial x_2}{\partial a_n}\\ \vdots & & & \vdots\\ & & \cdots & \dfrac{\partial x_n}{\partial a_n}\end{vmatrix}$$

$\Rightarrow$

任意一个函数 $g_i(a_i)$ 均可通过对函数 $g(\vec{a})$ 积分掉其它不用的变量而得到。是数据处理中误差传递的基础。

期待值及方差

考虑具有 **p.d.f.** $f(x)$ 的随机变量 x，定义期待(平均)值为

$$\boxed{E[x]=\int x\,f(x)dx}$$

通常记为：$E[x]=\mu$

注意: 它不是x的函数。

对离散型变量，有 $E[x]=\sum_{i=1}^{n}x_iP(x_i)$

对具有 **p.d.f.** $g(y)$ 的函数 $y(x)$ ，有

$$\boxed{E[y]=\int yg(y)dy=\int y(x)f(x)dx}$$

方差定义为

$$V[x]=E[(x-E[x])^2]=E[x^2]-\mu^2$$

通常记为：$V[x]=\sigma^2$

标准偏差：$\sigma\equiv\sqrt{\sigma^2}$

协方差与相关系数

定义协方差 $\mathrm{cov}[x,y]$ (也可用矩阵表示V_{xy})为

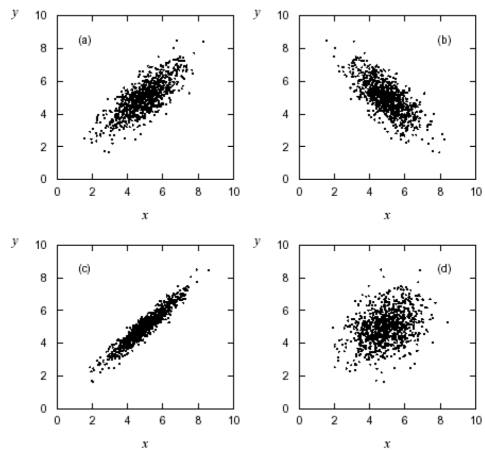
$$\boxed{\mathrm{cov}[x,y]=E[(x-\mu_x)(y-\mu_y)]=E[xy]-\mu_x\mu_y}$$

关联系数定义为

$$\boxed{\begin{array}{l}\rho_{xy}=\dfrac{\mathrm{cov}[x,y]}{\sigma_x\sigma_y},\\ -1\le\rho_{xy}\le 1\end{array}}$$

$\rho=0.75$ $\rho=-0.75$

$\rho=0.95$ $\rho=0.25$



如果 ***x***，***y*** 独立，即

$$f(x,y)=f_x(x)f_y(y)\ \Rightarrow\ \mathrm{cov}[x,y]=0$$

举例：样本平均值

假设实验上研究一核素衰变寿命，在探测效率为100%的情况下，每次探测到的寿命为 t_i，一共测量了 n 次，求平均寿命（也就是寿命的期待值）。

根据离散型期待值的定义 $E[t]=\sum_{i=1}^{n} t_i P(t_i)$ 问题的关键是 t_i 的概率密度函数是什么？

根据概率的相对频率定义，在 n 次测量中出现 t_i 频率为一次

$$P(t_i)=\frac{1}{n}$$

因此，期待值（或平均寿命）为 $E[t]=\sum_{i=1}^{n} t_i \frac{1}{n}=\frac{1}{n}\sum_{i=1}^{n} t_i$

思考：如果频率为 m_i 次，结果会不同吗？

不确定度传递

假设 $\vec{x} = (x_1, \ldots, x_n)$ 服从某一联合 **p.d.f.** $f(\vec{x})$ ，我们也许并不知道该函数形式，但假设有协方差 $V_{ij} = \mathrm{cov}[x_i, x_j]$ 和平均值 $\vec{\mu} = E[\vec{x}]$

现考虑一函数 $y(\vec{x})$，方差 $V[y] = E[y^2] - (E[y])^2$ 是什么？

将 $y(\vec{x})$ 在 $\vec{\mu}$ 附近按泰勒展开到第一级

$$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)$$

然后，计算 $E[y]$ 与 $E[y^2]$ …

不确定度传递

由于 $E[x_i - \mu_i] = 0$ 所以利用泰勒展开式可求

$$E[y(\vec{x})] \approx y(\vec{\mu})$$

$$E[y^2(\vec{x})] \approx y^2(\vec{\mu}) + 2y(\vec{\mu}) \cdot \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} E[x_i - \mu_i]$$

$$+ E\left[\left(\sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)\right)\left(\sum_{j=1}^{n} \left[\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} (x_j - \mu_j)\right)\right]$$

$$= y^2(\vec{\mu}) + \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}$$

不确定度传递

两项合起来给出 $y(\vec{x})$的方差

$$\boxed{\sigma_y^2 \equiv V[y] \approx \sum_{i,j=1}^{n}\left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}}$$

如果 x_i 之间是无关的，则 $V_{ij}=\sigma_i^2\delta_{ij}$，那么上式变为

$$\boxed{\sigma_y^2 \equiv V[y] \approx \sum_{i=1}^{n}\left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}}^2 \sigma_i^2}$$

类似地，对于m 组函数

$$\boxed{\vec{y}(\vec{x}) = (y_1(\vec{x}),\ldots,y_m(\vec{x}))}$$

不确定度传递

$$U_{kl} = \mathrm{cov}[y_k, y_l] \approx \sum_{i,j=1}^{n} \left[\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}$$

或者记为矩阵形式

$$U = AVA^T, \qquad A_{ij} = \left[\frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}}$$

注意：上式只对$\vec{y}(\vec{x})$为线性时是精确的，近似程度在函数非线性区变化比σ_i要大时遭到很大的破坏。另外，上式并不需要知道x_i的 **p.d.f.** 具体形式，例如，它可以不是高斯的。

不确定度传递的一些特殊情况

$$y = x_1 + x_2 \qquad \sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2\,\mathrm{cov}[x_1, x_2]$$

$$y = x_1 x_2 \qquad \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + 2\frac{\mathrm{cov}[x_1, x_2]}{x_1 x_2}$$

注意在相关的情况下，最终的误差会有很大的改变，例如当

$$y = x_1 - x_2,\ \mu_1 = \mu_2 = 10,\ \sigma_1 = \sigma_2 = 1$$

$$\begin{cases} \rho = 0: \quad E[y] = \mu_1 - \mu_2 = 0,\ V[y] = 1^2 + 1^2 = 2,\ \sigma_y = 1.4 \\ \rho = 1: \quad E[y] = \mu_1 - \mu_2 = 0,\ V[y] = 1^2 + 1^2 - 2 = 0,\ \sigma_y = 0 \end{cases}$$

这种特征有时候是有益的：将公共的或难以估计的不确定度，通过适当的数学处理将它们消掉，达到减小不确定度的目的。

坐标变换下的不确定度矩阵

实验上经常通过测量粒子在探测器中各点的击中坐标（x, y）来拟合在极坐标下的径迹（r, θ）。通常情况下，（x, y）的测量是不关联的。

$$r^2 = x^2 + y^2$$
$$\tan\theta = y / x$$

由于

$$U(r,\theta) = AV(x,y)A^T$$

因此，坐标变换后的不确定度矩阵为

$$\begin{pmatrix} \sigma_r^2 & \mathrm{cov}(r,\theta) \\ \mathrm{cov}(r,\theta) & \sigma_\theta^2 \end{pmatrix} = \begin{pmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{pmatrix} \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \begin{pmatrix} \frac{x}{r} & -\frac{y}{r^2} \\ \frac{y}{r} & \frac{x}{r^2} \end{pmatrix} = \frac{1}{r^2} \begin{pmatrix} x^2\sigma_x^2 + y^2\sigma_y^2 & \frac{xy}{r}(\sigma_y^2 - \sigma_x^2) \\ \frac{xy}{r}(\sigma_y^2 - \sigma_x^2) & \frac{1}{r^2}(y^2\sigma_x^2 + x^2\sigma_y^2) \end{pmatrix}$$

作业

1、证明： $P(A \cup B) = P(A) + P(B) - P(A \cap B).$

2、某粒子束流包含10^{-4}的电子，其余为光子。粒子通过某双层探测器，可能在2层都给出信号，也可能只有一层给出信号或者没有信号。电子(e)和光子（γ）在穿过该双层探测器给出0，1，或则2个信号的概率如下：

$$P(0|e) = 0.001 \qquad P(0|\gamma) = 0.99899$$

$$P(1|e) = 0.01 \qquad P(1|\gamma) = 0.001$$

$$P(2|e) = 0.989 \qquad P(2|\gamma) = 10^{-5}$$

问：1）如果只有1层给出信号，该粒子为光子的概率是多少？

2）如果2层都给出信号，该粒子为电子的概率是多少？

作业

3、考虑随机变量$\boldsymbol{x}$与常数α和β。证明：

$$E[\alpha x + \beta] = \alpha E[x] + \beta,$$

$$V[\alpha x + \beta] = \alpha^2 V[x].$$