



清华大学
Tsinghua University



粒子物理与核物理实验中的 数据分析

第一章：基本概念

杨振伟
清华大学



本章要点

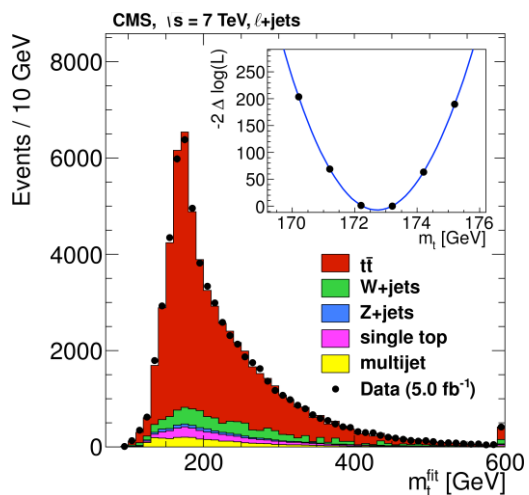
- 概率的定义与诠释
- 随机变量与概率密度
- 随机变量的函数
- 期待值、方差
- 不确定度的传递

物理实验的目的是什么？

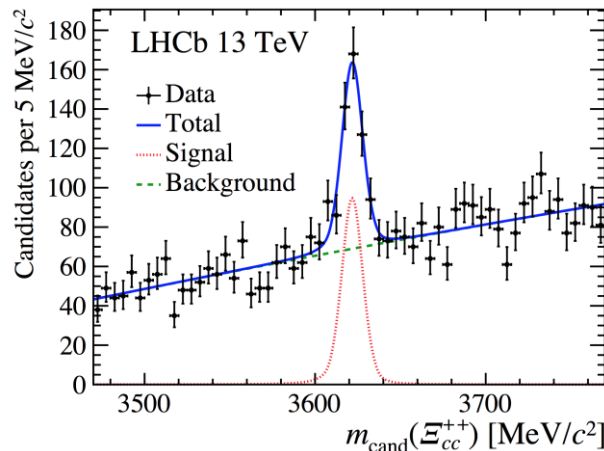
测量

+

发现



$$m_t = 173.49 \pm 1.07 \text{ GeV}/c^2$$

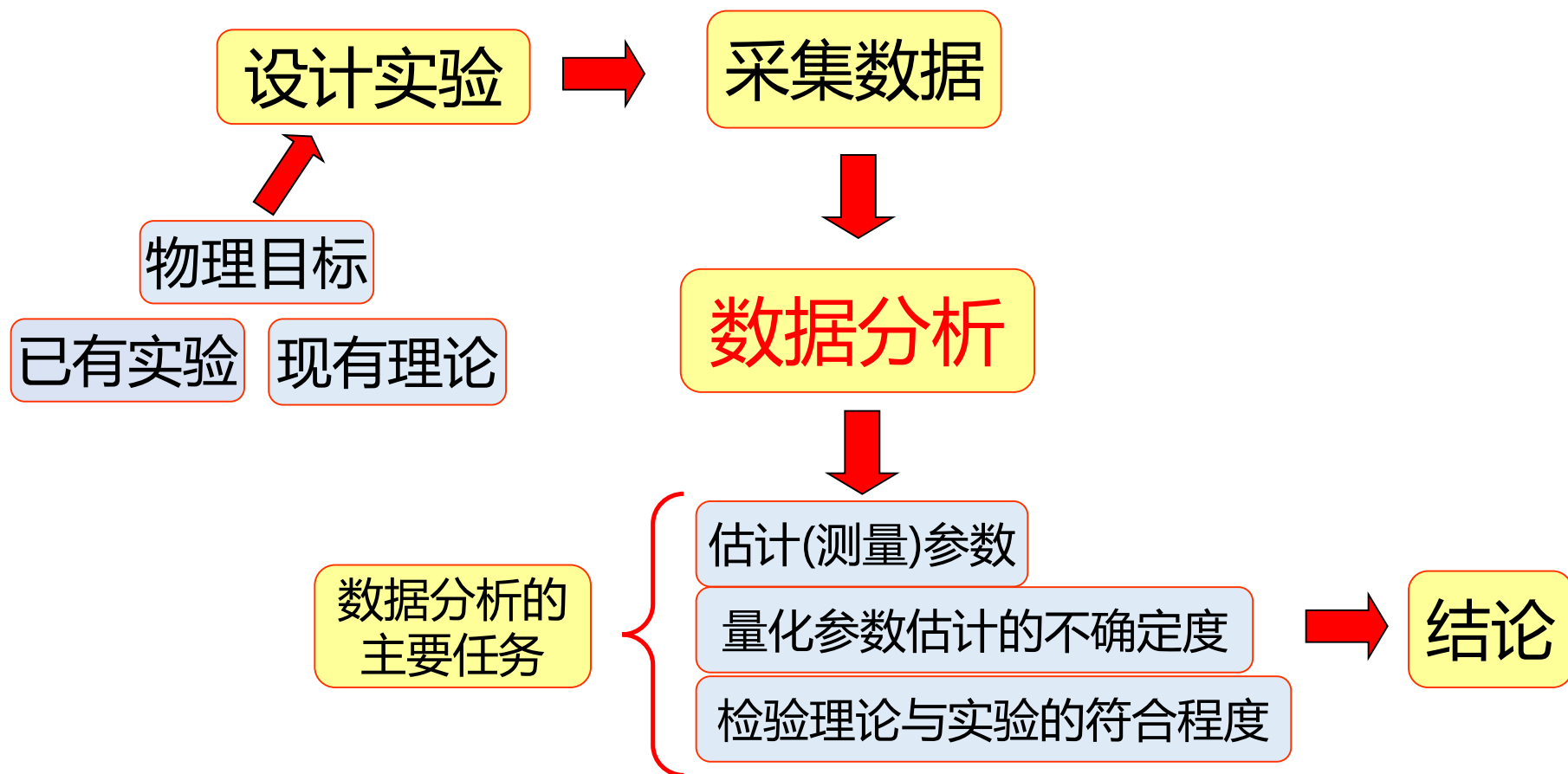


发现双粲重子 Ξ_{cc}^{++}

验证已有理论，或者揭示新规律

问题：如何实现？如何表述测量结果或发现？

如何实现？如何科学地给出结论？



概率与统计在所有过程都起着重要作用

随机性的来源

- 粒子物理与核物理中随机性的主要来源
 - 理论本身是非确定的
 - ✓ 量子力学 (...量力学)
 - 测量的随机不确定性
 - ✓ 即使没有量子效应也存在
 - 有些因素原则上可以确定但实际不确定
 - ✓ 受限于所需费用、时间等

这些不确定性可以用概率量化描述

概率与统计的差别



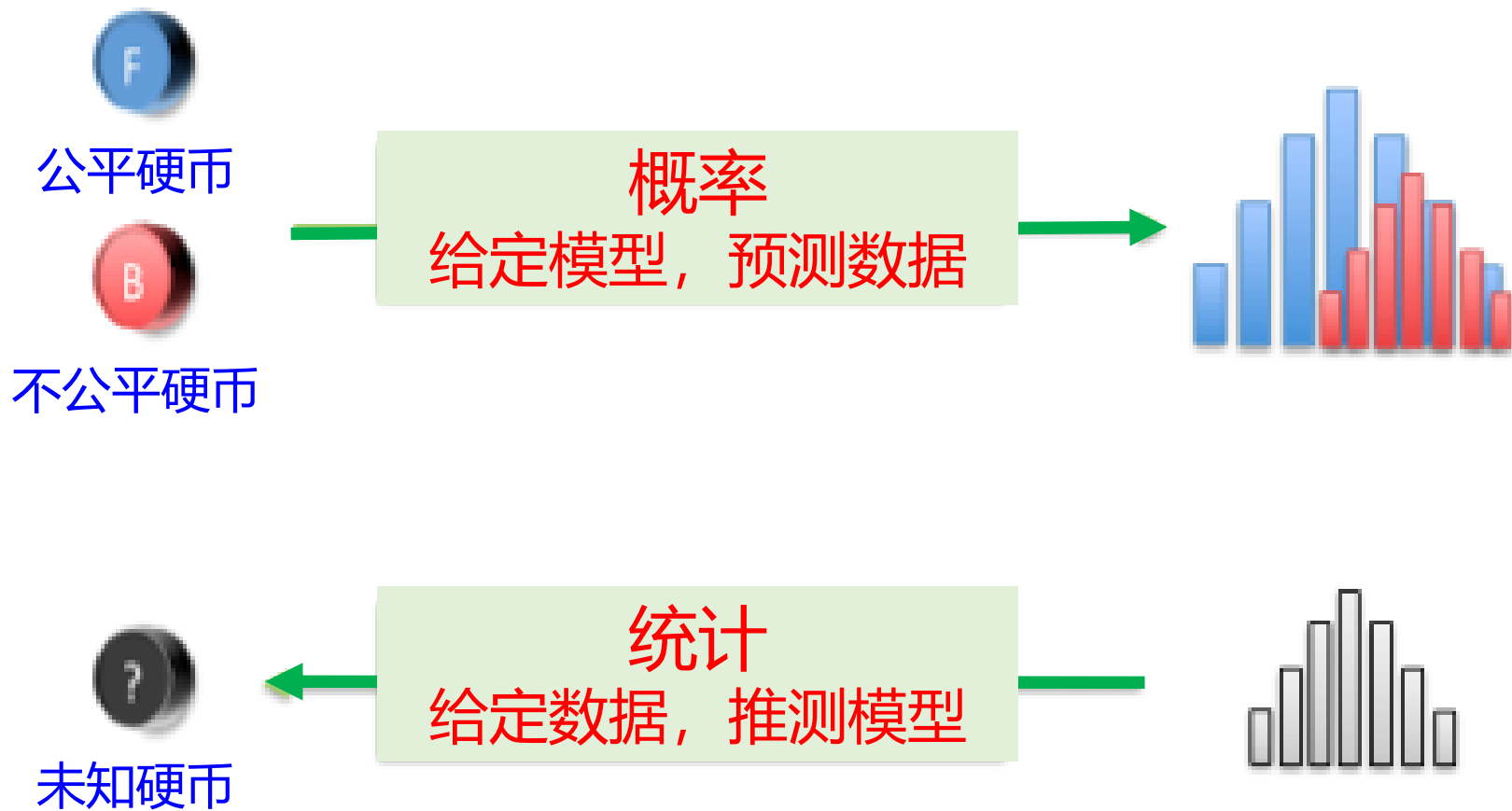
概率：给定模型，预测数据。

例如，已知桶里50白球，50黑球。
问：手中的球什么颜色？

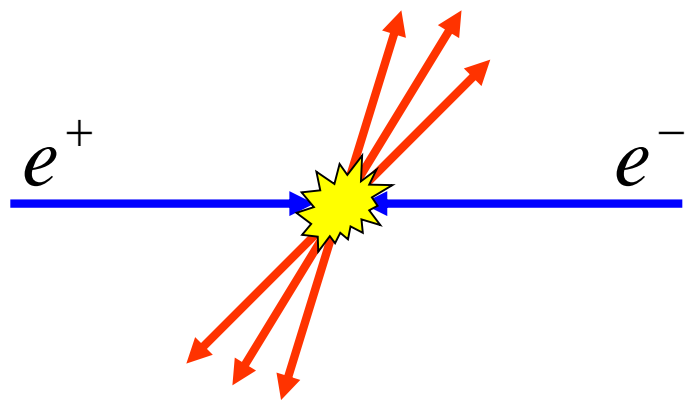
统计：给定数据，预测模型。

例如，已知手里4白球，6黑球。
问：桶中白球黑球的比例是多少？

概率与统计的差别



实验观测



观察某一类型的
 n 个事例

实验测量

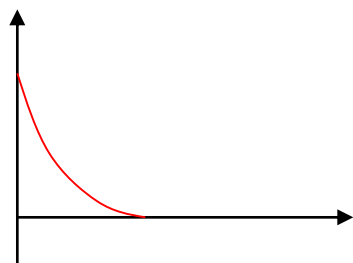
给出每个事例的特征量(能动量, 末态粒子数...).

理论预言

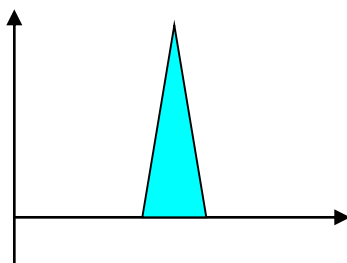
给出上述各特征量的分布, 而且可能还会包含自由参数。

实验数据 \rightarrow 统计分析 \rightarrow 物理结论

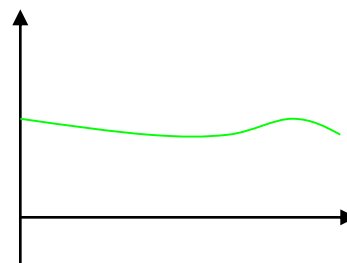
数据背后的物理图像是什么？



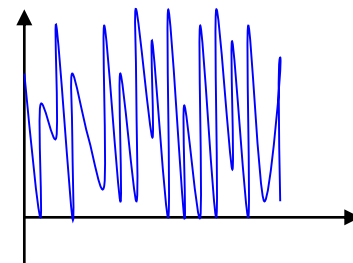
原初物理



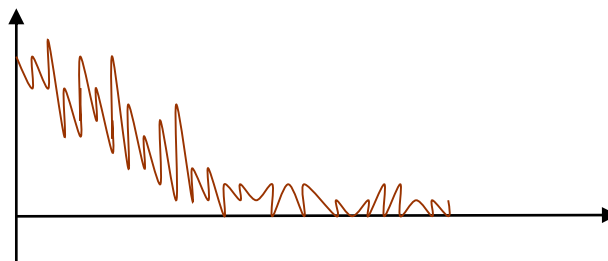
分辨率



探测效率



本底噪音



实验数据

数据分析专业术语：

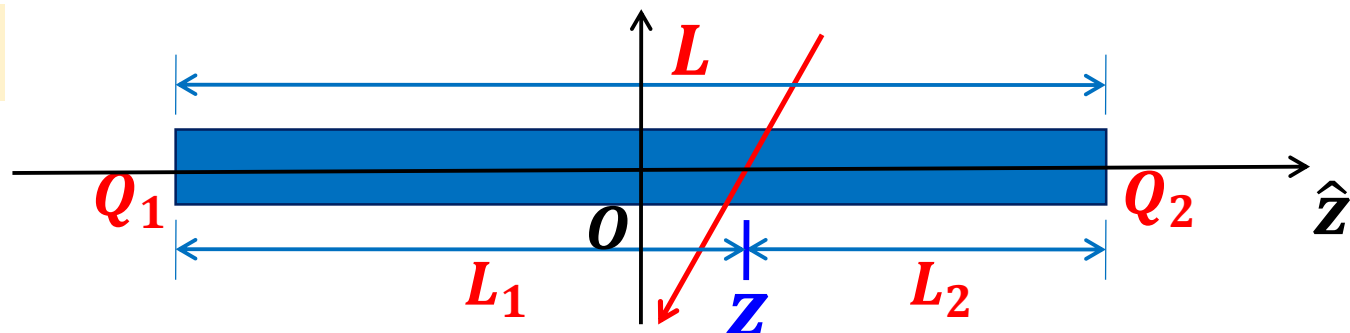
事例选择， 粒子鉴别， 选择条件， 信噪比优化， 无偏选择，
探测效率， 效率修正， 卷积分辨率， 解谱还原...

举例：测量闪烁体的衰减长度

光在闪烁体中传播时指数衰减 $Q = Q_0 e^{-\frac{L}{L_0}}$

L_0 ：闪烁体的衰减长度，表征闪烁体质量的一项重要指标。

实验测量



$$L_1 = 0.5L + z, \quad L_2 = 0.5L - z$$

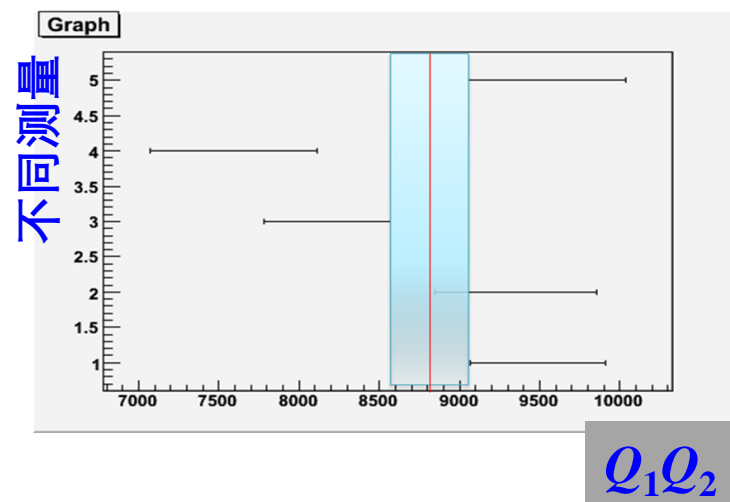
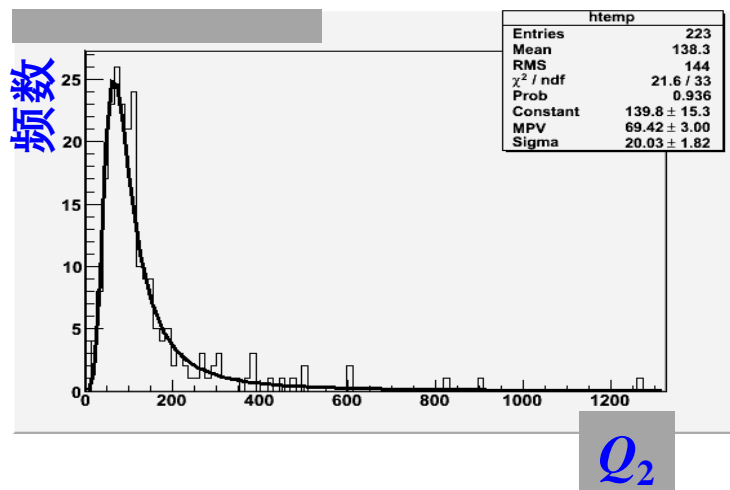
$Q_0 \propto E$: $0.5Q_0$ 向左, $0.5Q_0$ 向右

$$\begin{cases} Q_1 = 0.5Q_0 e^{-L_1/L_0} \\ Q_2 = 0.5Q_0 e^{-L_2/L_0} \end{cases} \rightarrow \begin{cases} Q_1 Q_2 = 0.25 Q_0^2 e^{-L/L_0} \\ Q_1/Q_2 = e^{-2z/L_0} \end{cases} \rightarrow \begin{cases} L_0 = L / \ln(Q_0^2 / 4Q_1 Q_2) \\ L_0 = -2z / \ln(Q_1 / Q_2) \end{cases}$$

举例：测量闪烁体衰减长度（续）

$$\begin{cases} Q_1 Q_2 = 0.25 Q_0^2 e^{-L/L_0} \\ Q_1/Q_2 = e^{-2z/L_0} \end{cases}$$

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。
理论上只要在给定某个位置 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。

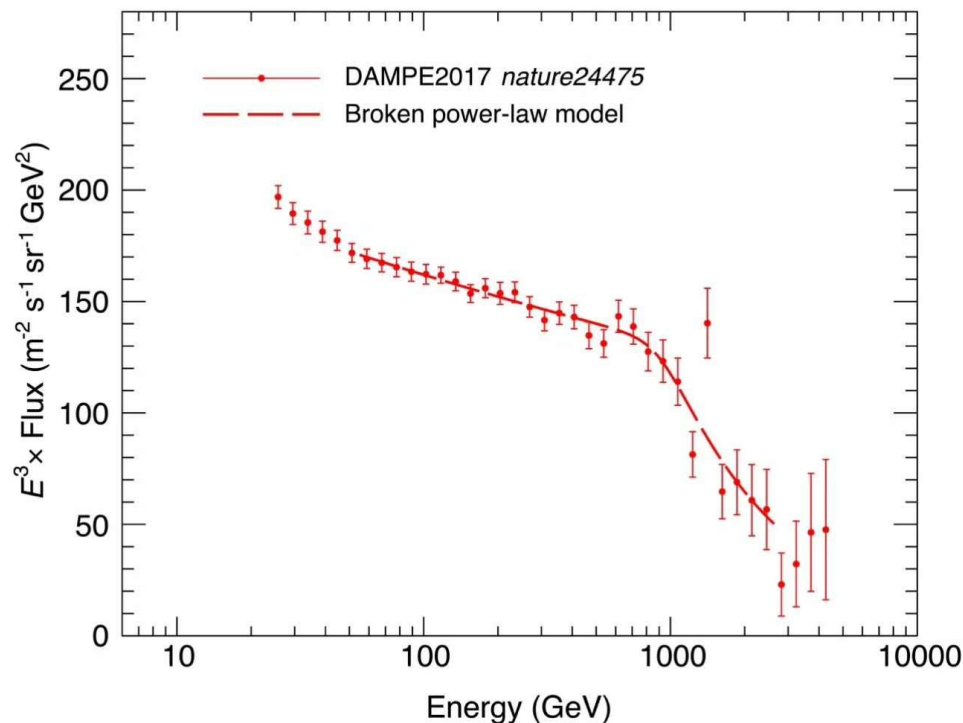


理论上是不变的 $Q_1 Q_2$ 值，
为什么每次测量都不相同？
能否认为 L_0 不是常数？



使用概率来量化结论！

举例：宇宙线测量与物理信号



问题1：如何确定能量测量的正确性？

问题2：如何确定1.4 TeV附近是物理信号还是统计涨落？

本章要点

➤ 概率的定义与诠释

- 随机事件
- 概率的定义
- 条件概率
- 概率的诠释
- 贝叶斯定理
- 全概率定理

➤ 随机变量与概率密度

➤ 随机变量的函数

➤ 期待值、方差

➤ 不确定度的传递

随机事件

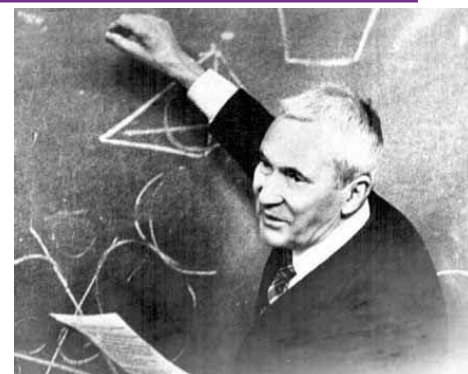
在一定的实验条件下，现象 A 可能发生，也可能不发生，并且只有发生或不发生这样两种可能性，这是偶然现象中一种比较简单的情形，我们把发生了现象 A 的事例称为随机事件 A ，简称事件 A 。随机事件也称随机事例。

概率的定义

柯尔莫哥洛夫公理

考虑一全集 S , 有子集 A, B, \dots

1. $P(A) \geq 0, \forall A \subset S$
2. $P(S) = 1$
3. $P(A \cup B) = P(A) + P(B)$: 如果 $A \cap B = \emptyset$



→ $P(A)$ 称为事件 A 的概率

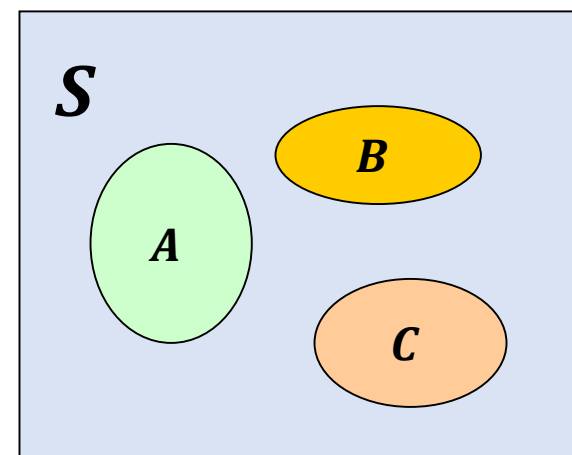
利用柯氏三公理可以得到

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

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

$$A \subset B \Rightarrow P(A) \leq P(B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



条件概率与独立性

给定事件 B 的条件下，事件 A 发生的条件概率定义为

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

要求 $P(B) \neq 0$

如果 $P(A \cap B) = P(A)P(B)$ ，则称 A 与 B 相互独立。

如果 A 与 B 相互独立，则

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



A 发生的概率与 B 是否发生没有关系

条件概率与独立性

问题：假设两个子集 A 和 B 满足 $A \cap B = \emptyset$,
那么 A 和 B 是否相互独立？

注意：两个子集互斥与独立定义不同。
互斥的两个事件必然相互不独立，其中一个事件发生，则另一个事件必然不发生。

概率的诠释

➤ 相对频率（频率论者）

假设 A, B, \dots 是可重复实验的一组结果，则概率是

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}$$

n_A : n 次重复实验中 A 出现的次数

n : 重复实验的次数

例如，量子力学、粒子散射、辐射衰变、宇宙线等

问题：相对频率的概率诠释满足柯氏公理吗？

概率的诠释

➤ 主观概率 (贝叶斯论者)

如果 A, B, \dots 是一些假设(真或假的一些陈述), 那么概率

$P(A)$ = 对 A 为真的信心程度(degree of belief)

✓ 两种解释皆与柯尔莫哥洛夫公理相符。

✓ 粒子物理与核物理实验中常用相对频率解释, 但是主观概率对不可重复现象可以提供更自然的处理:

系统不确定度, 某个粒子存在的概率, 置信区间的解释,

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。

- 频率的概率解释不适用于某些特殊情况，例如

如何理解天气预报“明天降水概率60%”？

如何理解顶夸克质量的测量结果 $m_t = 173.49 \pm 1.07 \text{ GeV}/c^2$ ？

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);
 - 在某个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!

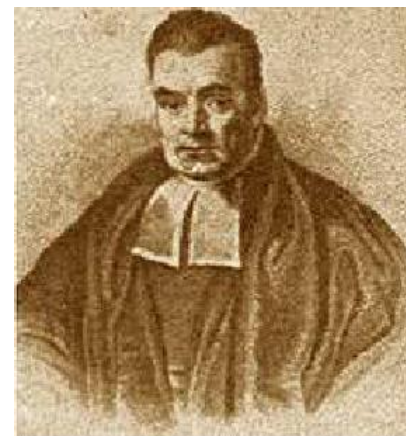
贝叶斯定理

假设 $P(A) \neq 0$ 且 $P(B) \neq 0$ ，根据条件概率的定义

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

利用 $P(A \cap B) = P(B \cap A)$ ，可得

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

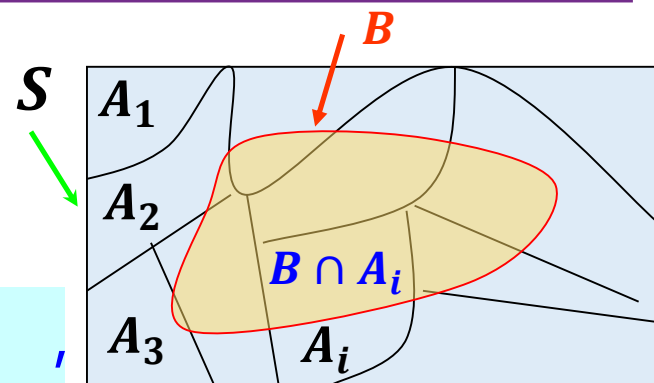


贝叶斯 (1702-1761)

这就是**贝叶斯定理**，Reverend Thomas Bayes 首先提出。

全概率定律

考虑样本空间 S 的一个子集 B 。
将样本空间划分为互斥的子集 A_i ，
使得 $\cup_i A_i = S$ 。



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

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

$$= \sum_i P(B|A_i)P(A_i)$$

利用条件概率公式

全概率定律

贝叶斯定理 \longrightarrow

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

例：如何利用贝叶斯定理

假设对任意一个人而言，感染上nCov的概率为

$$P(\text{nCov}) = 0.001$$

$$P(\text{no nCov}) = 0.999$$

验前概率，即任何检验之前

任何一次nCov检查的结果只有阴性(-)或阳性(+)两种

$$P(+|\text{nCov}) = 0.98$$

nCov感染患者阳性的概率

$$P(-|\text{nCov}) = 0.02$$

nCov感染患者阴性的概率

$$P(+|\text{no nCov}) = 0.03$$

nCov未感染患者阳性的概率

$$P(-|\text{no nCov}) = 0.97$$

nCov未感染患者阴性的概率

如果某人检查结果为阳性(+), 而他却觉得自己无明显感染渠道。那么他是否应担心自己真的感染了nCov?

例：如何利用贝叶斯定理(续)

我们想求的是验后概率 $P(\text{nCov}|+)$ ，即阳性结果条件下是nCov患者的概率。

利用贝叶斯定理，

nCoV患者阳性

$$P(\text{nCov}|+) = \frac{P(+|\text{nCov})P(\text{nCov})}{\underbrace{P(+|\text{nCov})P(\text{nCov}) + P(+|\text{no nCov})P(\text{no nCov})}_{\text{所有为阳性结果的人}}} = \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999} = 0.032$$

也就是说，你可能没什么问题！？

从个人角度看：对自己染上nCov结果的可信度为3.2%。
从医生角度看：象这样的人有3.2%感染上了nCov。



涉及到如何诠释结果（概率）的问题！

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

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

验后概率 似然性 先验概率

这是个“认识-实践-再认识-再实践”的迭代过程：



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

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

本章要点

➤ 概率的定义与诠释

➤ 随机变量与概率密度

- 随机变量
- 概率密度、直方图
- 联合概率密度
- 边缘概率密度
- 条件概率密度

➤ 随机变量的函数

➤ 期待值、方差

➤ 不确定度的传递

随机变量和概率密度函数

随机变量是样本空间元素的实值单值函数，可以是离散函数，也可以是连续函数

假设实验结果为连续值 x ， x 在 $[x, x + dx]$ 区间内的概率为

$$P(x \in [x, x + dx]) = f(x)dx$$



$f(x)$ 为概率密度函数(pdf)

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

归一化： x 一定是某个实数

对于离散结果 x_i ($i = 1, 2, \dots$):

$$P(x_i) = p_i$$

概率质量函数(pmf)

$$\sum_i P(x_i) = 1$$

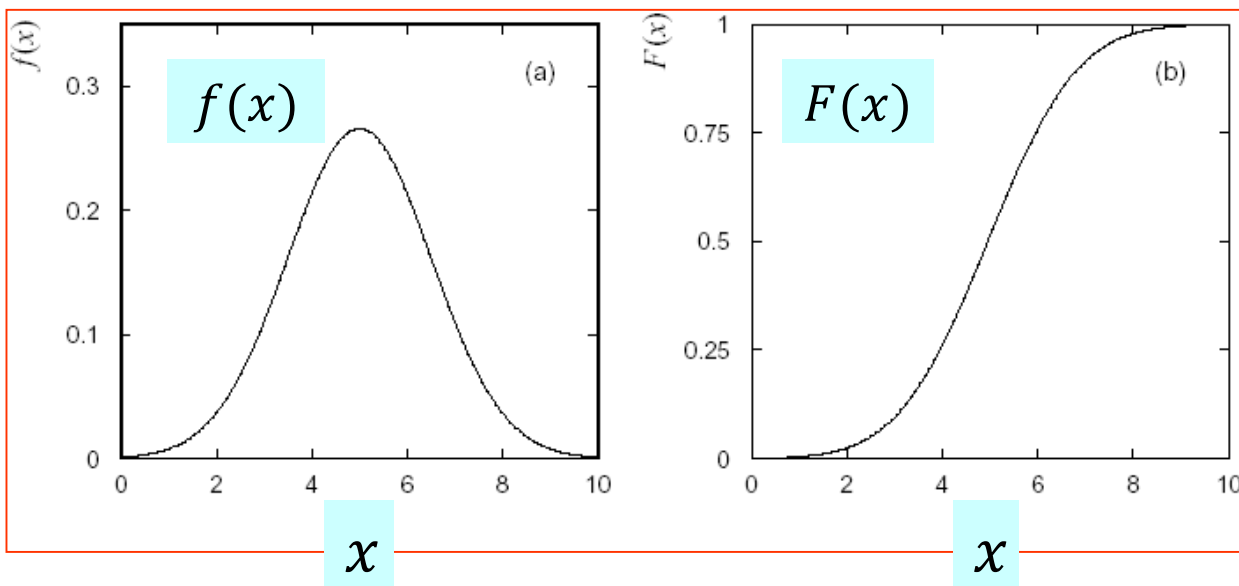
归一化

累积分布函数

实验结果小于等于 x 的概率

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

累积分布函数(cdf)



➤ 概率密度函数可以定义为

$$f(x) = \frac{\partial F(x)}{\partial x}$$

α 分位数、中位数与模

分位数 x_α 定义为随机变量 x 的值, 它使得

$$F(x_\alpha) = \alpha \quad (0 \leq \alpha \leq 1)$$

因此容易求出分位点

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

α 分位数 (α -quantile)
中位数 (median)
模 (mode)

随机变量 x 的中位数定义为

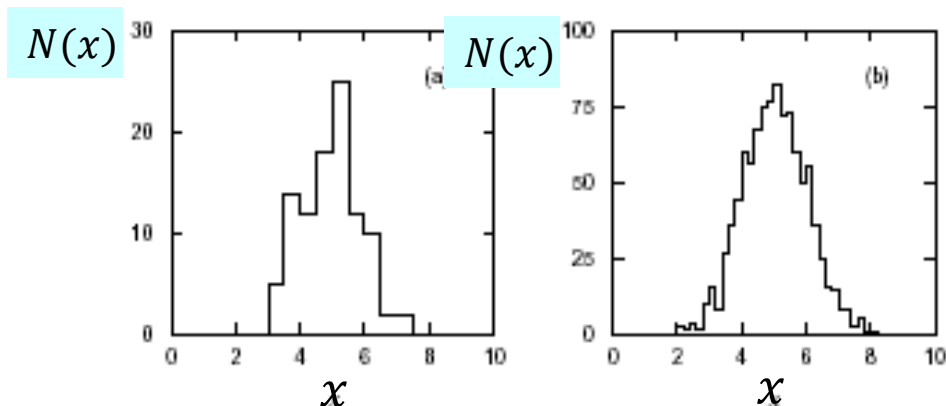
$$x_{1/2} = F^{-1}(1/2)$$

随机变量 x 被观测到大于或小于中位数的概率相等。

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 pdf 就是样本无穷大，区间宽度为零，且归一化到单位面积的直方图。

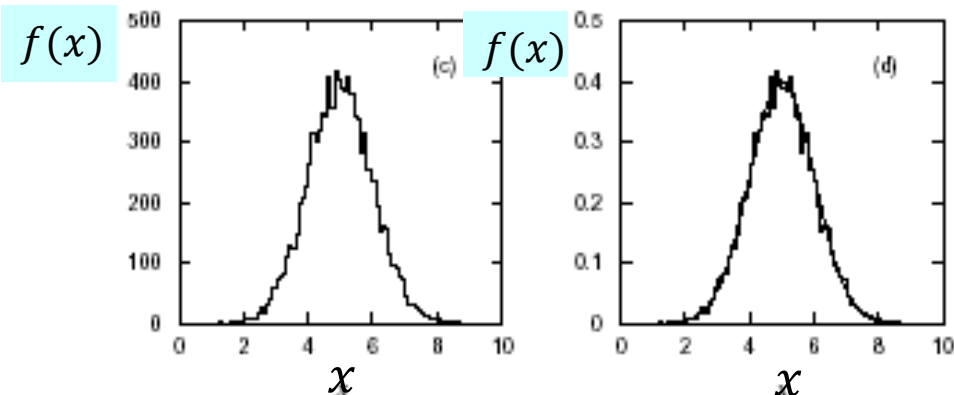


$$f(x) = \frac{N(x)}{n \cdot \Delta x}$$

$N(x)$ = 每个区间的频数

n = 填入直方图的总事例数

Δx = 区间的宽度



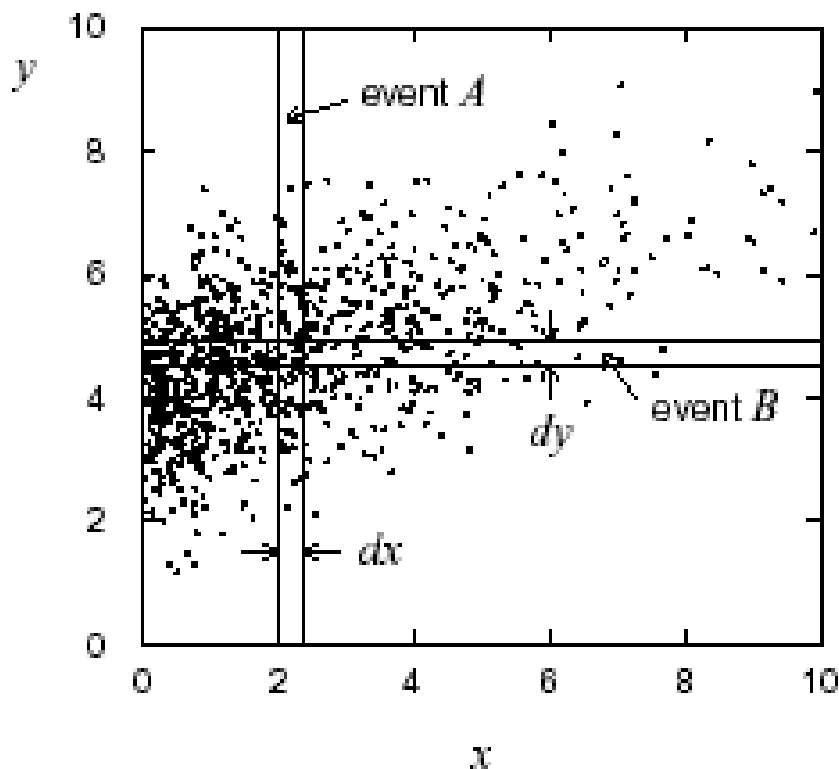
$$n \rightarrow \infty, \Delta x \rightarrow 0$$

$n \cdot \Delta x$ 有限

直方图在统计分析中非常重要，应准确理解它的含义。

多变量情形

观测量不止一个，例如 x 与 y



$$P(A \cap B) = \iint_{A \cap B} f(x, y) dx dy$$

$f(x, y)$ = 联合概率密度函数

$$\iint f(x, y) dx dy = 1$$

边缘概率密度

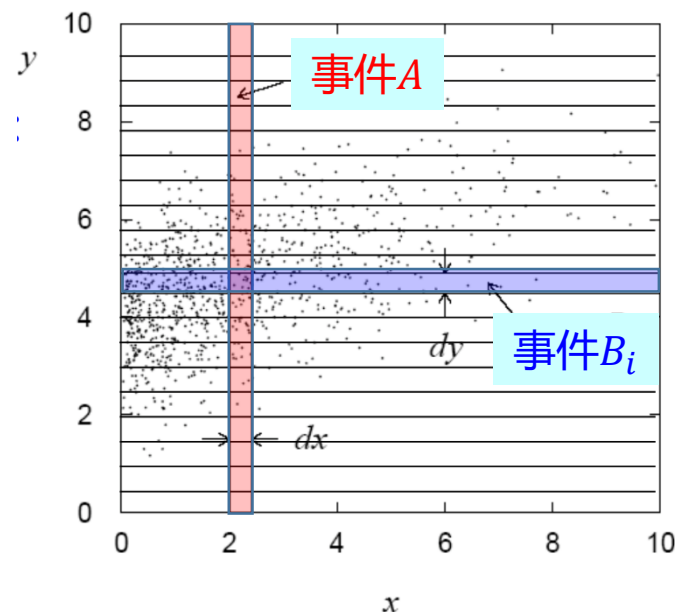
有时，我们只关心某个分量的pdf:

$$\begin{aligned} P(A) &= \sum_i P(A \cap B_i) \\ &= \sum_i \int_{B_i} f(x, y) dy dx \\ &\rightarrow \int f(x, y) dy dx \end{aligned}$$

$$f_x(x) = \int f(x, y) dy$$

类似地

$$f_y(y) = \int f(x, y) dx$$

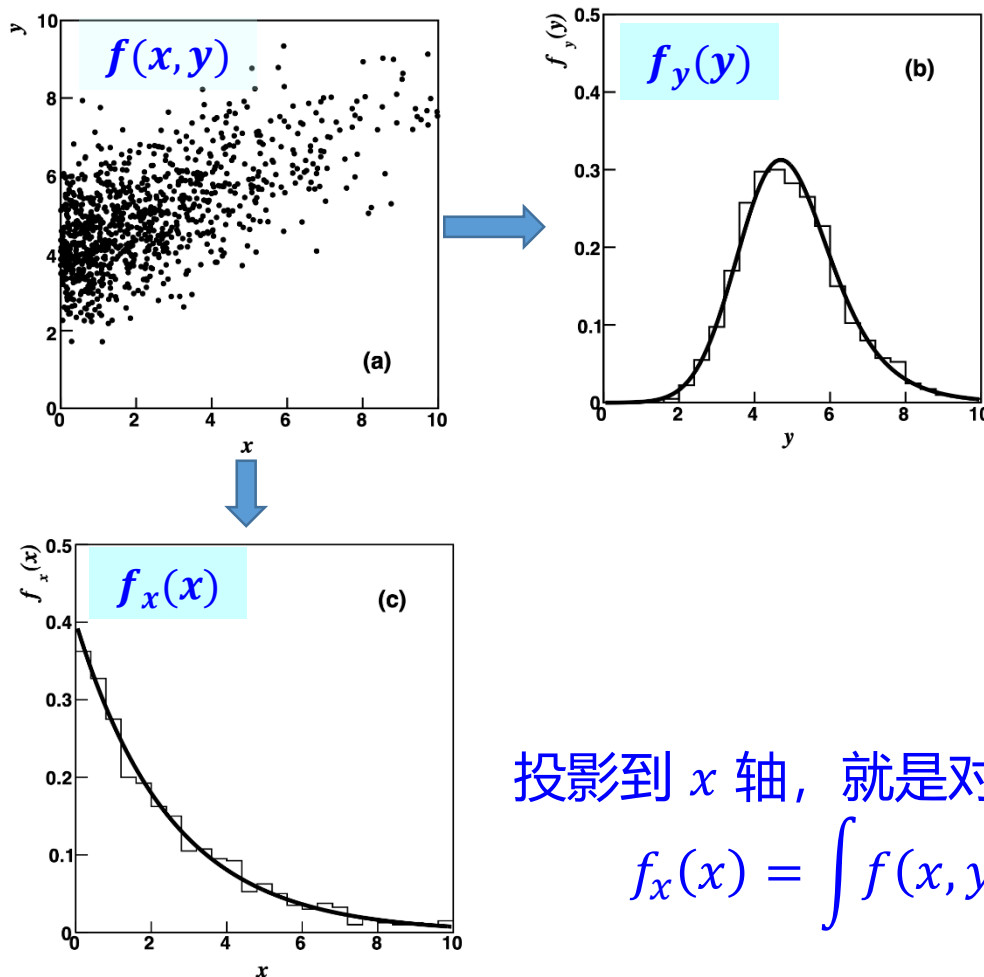


边缘概率密度

如果 $f(x, y) = f_x(x)f_y(y)$, 则称随机变量 x 和 y 相互独立

边缘概率密度

边缘概率密度：将联合概率密度投影某个坐标轴



投影到 y 轴，就是对 x 积分：

$$f_y(y) = \int f(x, y) dx$$

投影到 x 轴，就是对 y 积分：

$$f_x(x) = \int f(x, y) dy$$

条件概率密度函数

有时，我们关心联合pdf中某个变量为常数的情况。

→ 条件概率

根据条件概率的定义：

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

条件概率密度函数

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

贝叶斯定理可写为

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

如果 $f(x, y) = f_x(x)f_y(y)$ ，则 x, y 相互独立。

如果 $P(A \cap B) = P(A)P(B)$ ，则 A, B 相互独立。

条件概率密度函数(续)

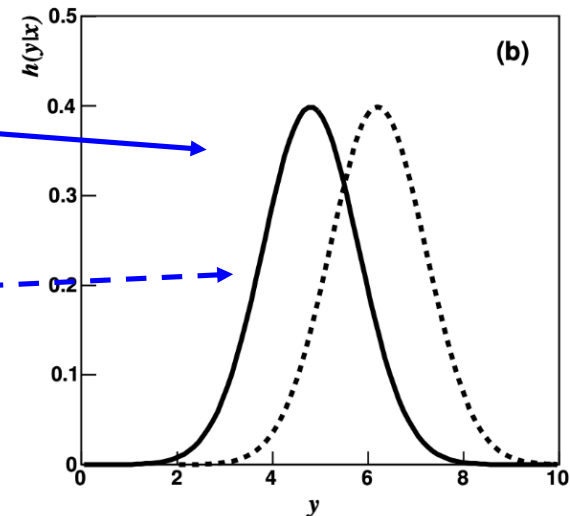
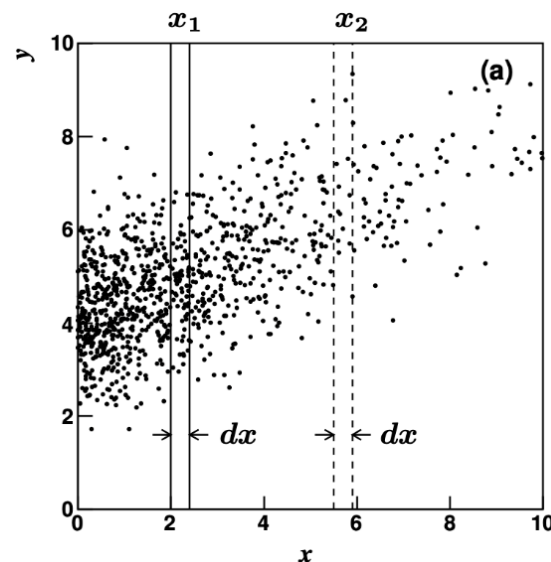
例如，已知联合概率密度 $f(x, y)$ 。
求条件概率密度 $h(y|x_1)$, $h(y|x_2)$ 。

做法：

1. 将 $f(x, y)$ 中的 x 固定为 x_1 或 x_2 ,
2. 除以相应的边缘概率密度 $f_x(x_1)$ 或 $f_x(x_2)$, 以保证归一化条件

$$h(y|x_1) = \frac{f(x_1, y)}{f_x(x_1)}$$

$$h(y|x_2) = \frac{f(x_2, y)}{f_x(x_2)}$$



举例：检查经验式概率密度函数

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），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

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

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

试判断哪一个可以用作概率密度函数？

本章要点

➤ 概率的定义与诠释

➤ 随机变量与概率密度

➤ 随机变量的函数

- 一维随机变量的函数
- 多维随机变量的函数
- 梅林卷积、傅立叶卷积
- 雅克比行列式

➤ 期待值、方差

➤ 不确定度的传递

数据分析中的问题

粒子物理与核物理实验中对动量的测量通常是分别测量

$$p_T = \sqrt{p_x^2 + p_y^2} \quad p_z \quad f(p_T, p_z)$$

在已知两分量测量值的概率密度函数情况下，总动量为

$$p = \sqrt{p_T^2 + p_z^2}$$

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。

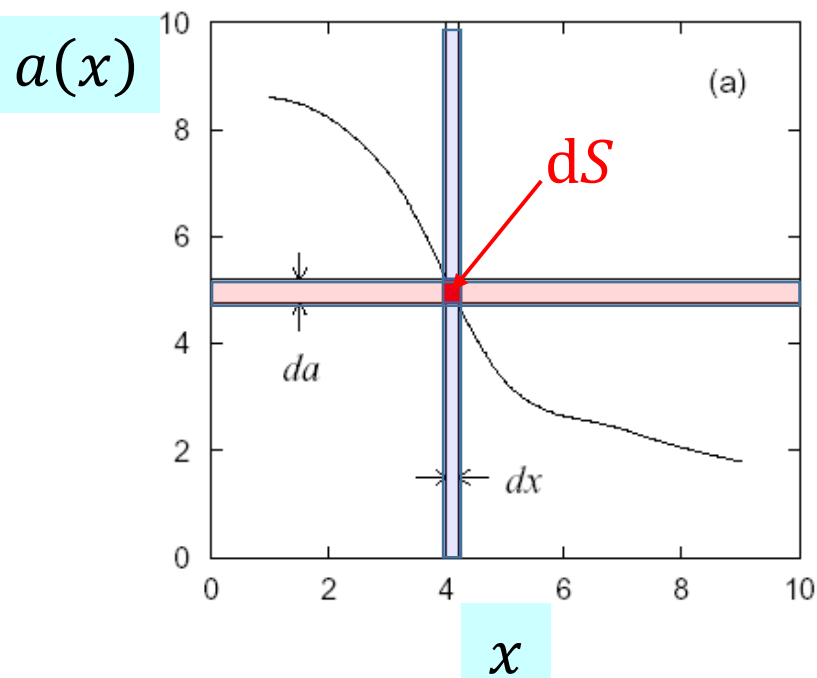
随机变量的函数自身也是一个随机变量。

例如，
 θ 和 $\cos \theta$

一维随机变量的函数

假设 x 服从概率密度 $f(x)$ ，对于函数 $a(x)$ ，其概率密度 $g(a)$ 为何？

假设函数 $a(x)$ 单调。



$a \in [a, a + da]$ 的概率 $g(a)da$ 等于 $x \in dS$ 的概率 $f(x)dx$

dS : 满足 $a \in [a, a + da]$ 的 x 的集合

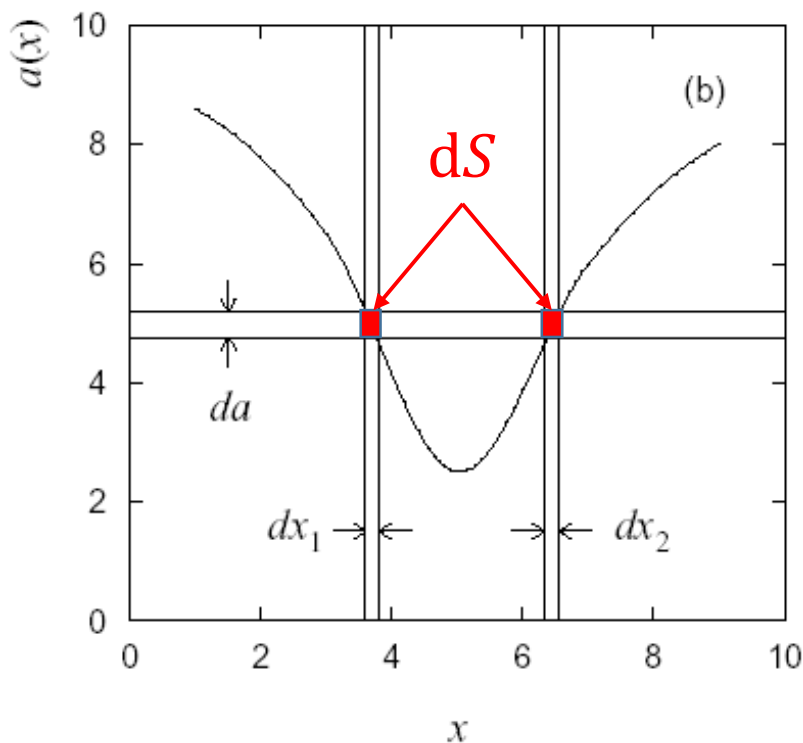
$$\begin{aligned} g(a)da &= \int_{dS} f(x)dx \\ &= \left| \int_{x(a)}^{x(a+da)} f(x')dx' \right| \\ &= \int_{x(a)}^{x(a) + \left| \frac{dx}{da} \right| da} f(x')dx' \end{aligned}$$

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

函数的逆不唯一情况

假如 $a(x)$ 的逆不唯一, dS 将包含多段 dx 区间, 需要全部考虑进来

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



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

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

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

多维随机变量的函数

考虑多维随机变量 $\vec{x} = (x_1, \dots, x_n)$ 与函数 $a(\vec{x})$ 。

已知 \vec{x} 的概率密度 $f(\vec{x}) = f(x_1, \dots, x_n)$ ，求 a 的概率密度 $g(a)$

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

dS : 在 $a(\vec{x}) = a'$ 和 $a(\vec{x}) = a' + da'$ 定义的两个曲面之间的 \vec{x} 空间范围

多维随机变量的函数: $z = xy$

如果两个随机变量 $x, y > 0$, 服从联合概率密度 $f(x, y)$, 考虑函数 $z = xy$, 其概率密度函数 $g(z)$ 是什么形式?

$$g(z)dz = \iint_{dS} f(x, y) dx dy = \int_0^\infty dx \int_{z/x}^{(z+dz)/x} f(x, y) dy$$

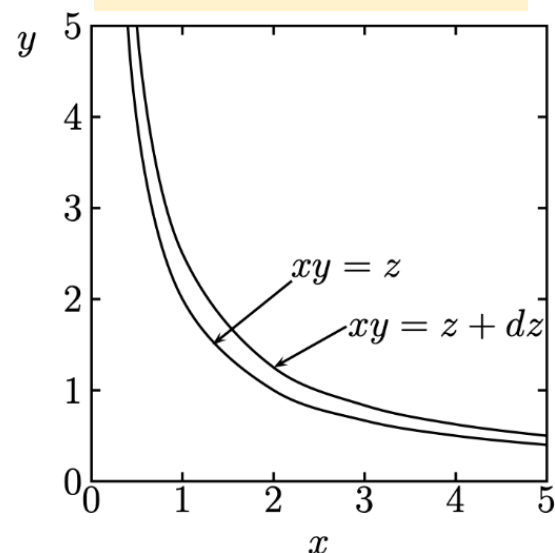
$$= \int_0^\infty dx \left[f\left(x, \frac{z}{x}\right) \frac{dz}{x} \right]$$

$$= \left[\int_0^\infty f\left(x, \frac{z}{x}\right) \frac{dx}{x} \right] dz$$

$$g(z) = \int_0^\infty f\left(x, \frac{z}{x}\right) \frac{dx}{x}$$

$$g(z) = \int_0^\infty f\left(\frac{z}{y}, y\right) \frac{dy}{y}$$

给定 x ,
当 $z \in (z, z + dz)$ 时,
 $y \in \left(\frac{z}{x}, \frac{z+dz}{x}\right)$ 。



多维随机变量的函数: $z = x + y$

如果两个随机变量 $x, y > 0$, 服从联合概率密度 $f(x, y)$, 考虑函数 $z = x + y$, 其概率密度函数 $g(z)$ 是什么形式?

$$\begin{aligned} g(z)dz &= \iint_{dS} f(x, y) dx dy = \int_0^\infty dx \int_{z-x}^{z+dz-x} f(x, y) dy \\ &= \int_0^\infty dx [f(x, z-x) dz] \\ &= \left[\int_0^\infty f(x, z-x) dx \right] dz \end{aligned}$$

给定 x ,
当 $z \in (z, z + dz)$ 时,
 $y \in (z - x, z + dz - x)$ 。



$$g(z) = \int_0^\infty f(x, z-x) dx$$

$$g(z) = \int_0^\infty f(z-y, y) dy$$

梅林卷积与傅立叶卷积

假设随机变量 x, y 相互独立, 分别服从 $g(x)$ 和 $h(y)$ 分布。

$z = xy$ 的概率密度函数 $f(z)$ 是什么形式?

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



梅林(Mellin)卷积

$$f = g \otimes h$$

$z = x + y$ 的概率密度函数 $f(z)$ 是什么形式?

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



傅立叶(Fourier)卷积

$$f = g \otimes h$$

多维随机变量的函数与雅可比行列式

考虑随机矢量 $\vec{x} = (x_1, \dots, x_n)$ ，联合概率密度为 $f(\vec{x})$ ，构造 n 个线性独立的函数 $\vec{a}(\vec{x}) = (a_1(\vec{x}), \dots, a_n(\vec{x}))$ ，并且其逆函数 $x_1(\vec{a}), \dots, x_n(\vec{a})$ 存在。

则 \vec{a} 的联合概率密度为

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

其中 J 是雅可比行列式

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial a_1} & \frac{\partial x_1}{\partial a_2} & \cdots & \frac{\partial x_1}{\partial a_n} \\ \frac{\partial x_2}{\partial a_1} & \frac{\partial x_2}{\partial a_2} & \cdots & \frac{\partial x_2}{\partial a_n} \\ \vdots & & & \vdots \\ & & \cdots & \frac{\partial x_n}{\partial a_n} \end{vmatrix}$$

对联合概率密度 $g(\vec{a})$ 积分掉其他不关心的变量，可以得到任意一个边缘概率密度 $g_i(a_i)$ 。
这是数据分析中误差传递的基础。

本章要点

- 概率的定义与诠释
- 随机变量与概率密度
- 随机变量的函数
- 期待值、方差
 - 均值、方差、标准差
 - 协方差、相关系数
- 不确定度的传递

期待值、方差、标准差

考虑概率密度为 $f(x)$ 的随机变量 x ，定义期待 (平均) 值为

$$E[x] = \int x f(x) dx$$

通常记为: $E[x] = \mu$

意为pdf的“重心”

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

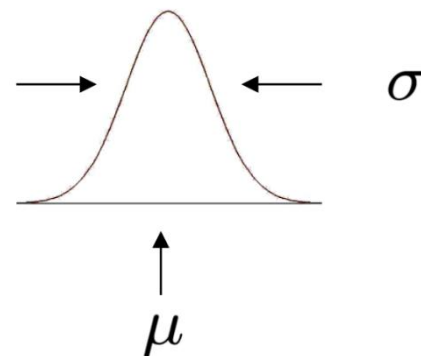
对概率密度为 $g(y)$ 的函数 $y(x)$, 有

$$E[y] = \int y g(y) dy = \int y(x) f(x) dx$$

方差定义为 $V[x] = E[(x - E[x])^2] = E[x^2] - \mu^2$

通常记为: $V[x] = \sigma^2$

标准差: $\sigma \equiv \sqrt{\sigma^2}$



协方差与相关系数

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

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

相关系数定义为

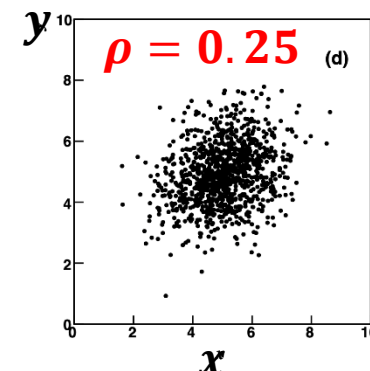
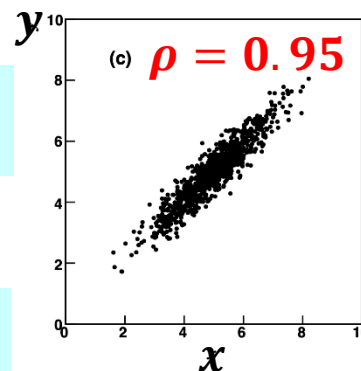
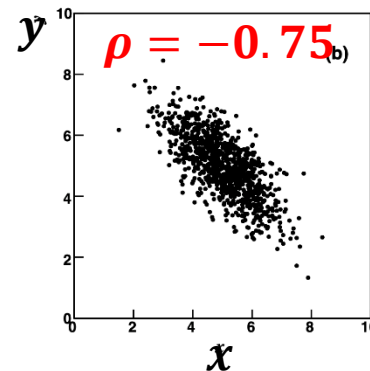
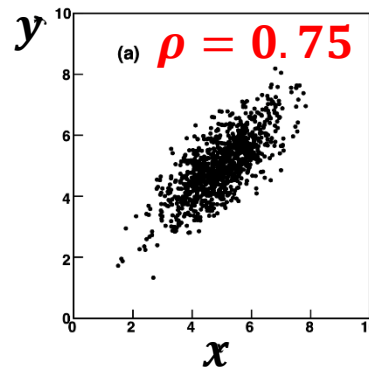
$$\rho_{xy} = \frac{\text{cov}[x, y]}{\sigma_x \sigma_y} \quad \text{无量纲}$$

如果 x, y 相互独立, 即

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

$$E[xy] = \iint xyf(x, y)dxdy = \mu_x\mu_y$$

则 $\text{cov}[x, y] = 0 \Rightarrow \rho_{xy} = 0$
即 x, y 不相关



判断题：

1. 随机变量 x 的数学期望 $E[x]$ 是 x 的函数。 【 】
2. 如果两个随机变量 x, y 不相关, 即 $\rho_{xy} = 0$, 则 x, y 相互独立。 【 】

例：样本均值

假设实验研究某核素衰变寿命，探测效率100%，共测量了 n 次，每次探测结果为 t_i 。求平均寿命(即寿命的期待值)。

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 $P(t_i)$ 是什么？

n 次测量出现 t_i 频率为一次，根据相对频率的概率诠释，

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

因此，平均寿命（或期待值）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 t_i 出现的频率为 m_i 次，结果会不同吗？

本章要点

- 概率的定义与诠释
- 随机变量与概率密度
- 随机变量的函数
- 期待值、方差
- 不确定度的传递
 - 不确定度的传递
 - 不确定度与相关性
 - 正交变换消除相关性

不确定度的传递(1)

假设我们对某个量测量了一组值 $\vec{x} = (x_1, \dots, x_n)$, 并得到其协方差 $V_{ij} = \text{cov}[x_i, x_j]$ (表征与 x_i 有关的测量不确定度)。

现考虑一函数 $y(\vec{x})$, 如何求其方差 $V[y]$?

硬核算法:

利用联合概率密度 $f(\vec{x})$ 求 y 的概率密度 $g(y)$, 然后计算

$$V[y] = E[y^2] - (E[y])^2$$

困难所在:

现实中通常不可行, 有时我们甚至完全清楚 $f(\vec{x})$ 的形式。

有没有简单可行的方法呢?

不确定度的传递(2)

假设我们已知 $\vec{\mu} = E[\vec{x}]$,

现实中通常只能根据测量得到 \vec{x} 的估计

对 $y(\vec{x})$ 在 $\vec{\mu}$ 处作一阶泰勒展开,

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

为了得到 $V[y]$, 我们需要计算 $E[y^2]$ 和 $E[y]$ 。

由于 $E[x_i - \mu_i] = 0$, $E[y(\vec{x})] \approx y(\vec{\mu})$

不确定度的传递(3)

$$\begin{aligned} E[y^2] &\approx y^2(\mu^2) + 2y(\vec{\mu}) \sum_{i=1}^n \left[\frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} E[x_i - \mu_i] \\ &\quad + E \left[\left(\sum_{i=1}^n \left[\frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i) \right) \left(\sum_{j=1}^n \left[\frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} (x_j - \mu_j) \right) \right] \\ &= y^2(\mu^2) + 0 + E \left[\sum_{i,j=1}^n \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)(x_j - \mu_j) \right] \\ &= y^2(\mu^2) + \sum_{i,j=1}^n \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij} \end{aligned}$$

因此, $y(\vec{x})$ 的方差为

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

不确定度传递公式
或
“误差传递公式”

不确定度的传递(4)

如果 x_i 不相关, 即 $V_{ij} = \sigma_i^2 \delta_{ij}$, 那么

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

这是特殊条件下的
“误差传递公式”

类似地, 对于 m 组函数 $\vec{y}(\vec{x}) = (y_1(\vec{x}), \dots, y_m(\vec{x}))$,

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

或者, 写成矩阵形式 $U = A V A^T$, 其中

$$A_{ij} = \left[\frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}}$$

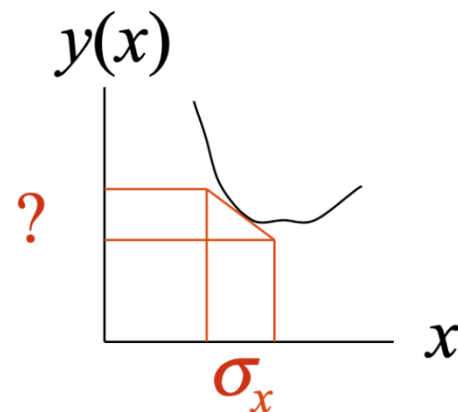
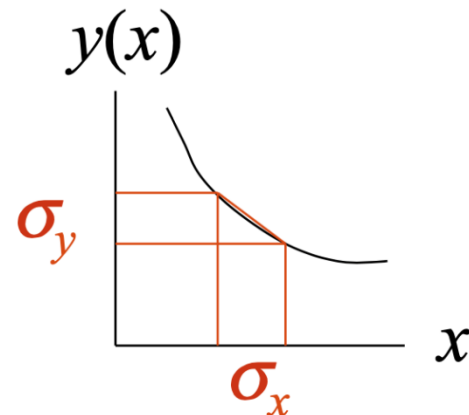
不确定度的传递(5)

不确定度传递公式告诉我们，如何用原始变量 \vec{x} 的协方差表示一组函数 $\vec{y}(\vec{x}) = (y_1(\vec{x}), \dots, y_m(\vec{x}))$ 的协方差。

局限性

- 只有当 $\vec{y}(\vec{x})$ 为线性时才严格成立；
- 如果函数在与 σ_i 差不多的范围内是非线性的，这个近似不再适用。

前面的推导没有要求 x_i 的概率密度的严格形式，例如，它可以不是高斯分布。



不确定度传递：特例

$$y = x_1 + x_2 \quad \longrightarrow \quad \sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2\text{cov}[x_1, x_2]$$

$$y = x_1 x_2 \quad \longrightarrow \quad \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + \frac{2\text{cov}[x_1, x_2]}{x_1 x_2}$$

如果 x_i 不相关：

- 和的不确定度的平方等于不确定度的平方和
- 积的相对不确定度的平方等于相对不确定度的平方和



注意：变量之间的相关性，可以引起巨大变化！

不确定度传递：特例

考虑 $y = x_1 - x_2$ ，其中： $\mu_1 = \mu_2 = 10$ ， $\sigma_1 = \sigma_2 = 1$ 。

1) $\rho = \frac{\text{cov}[x_1, x_2]}{\sigma_1 \sigma_2} = 0$

$$V[y] = 1^2 + 1^2 = 2$$



$$\sigma_y = 1.4$$

2) $\rho = 1$

$$V[y] = 1^2 + 1^2 - 2 = 0$$



$$\sigma_y = 0$$

即，对于 100% 相关的两个变量，其差的不确定度为零。

这种特征有时候是有益的：将共同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。

实践讨论题

- 利用普通的体重计称一个重量十公斤以内的物体，例如
 - 笔记本电脑
 - 装有大约10本书的书包
 - 行李箱
- 评价你的测量结果是否靠谱。如果不靠谱，有什么办法？



大亚湾反应堆中微子实验

- 大亚湾实验的主要目的是精确测量中微子振荡混合角 θ_{13} 。
- 中微子形态随运动距离的改变理论预言

$$\Delta I_r \sim \frac{\Delta S}{4\pi r^2} \cdot I \cdot P(\bar{\nu}_e \rightarrow \bar{\nu}_e) = \frac{\Delta S}{4\pi r^2} \cdot I \cdot f(\Delta m, \sin \theta_{13}) \cdot \sigma_{\text{xsec}} \cdot \varepsilon_{\text{eff}}$$

r : 探测器与反应堆的距离

ΔS : 探测器对应的面元

I : 反应堆发射中微子的通量

σ_{xsec} : 中微子与探测器的反应截面

ε_{eff} : 探测效率

$\Delta m, \sin \theta_{13}$: 待测量或参数

为了实现1%的探测精度，除了在振荡效应较大的远点探测器，大亚湾实验还在靠近反应堆处放置了两个近点探测器。这样做为什么有效？



坐标变换下的不确定度矩阵

实验上测量带电粒子动量通常是，测量粒子在探测器中各点的击中坐标 (x, y) ，然后拟合径迹。径迹往往用极坐标 (r, θ) 描述。一般来说， (x, y) 的测量不相关。 (r, θ) 是否相关？

两种坐标的变换关系：

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

$$V_{xy} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$

由于 $U_{r\theta} = AV_{xy}A^T$

$$A_{ij} = \left[\frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} \Rightarrow A = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix}$$

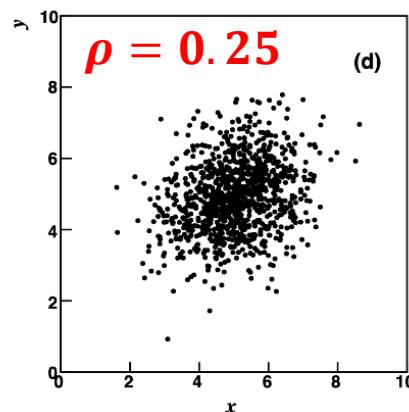
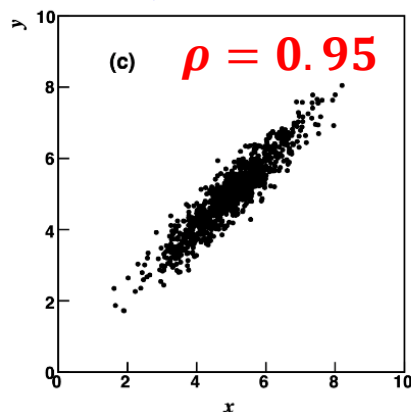
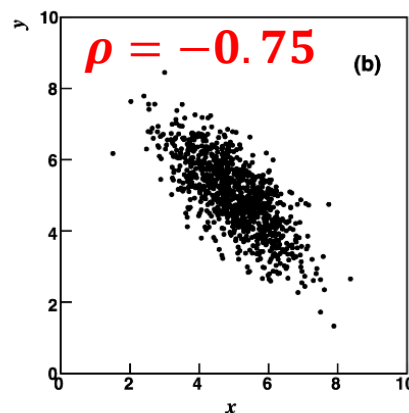
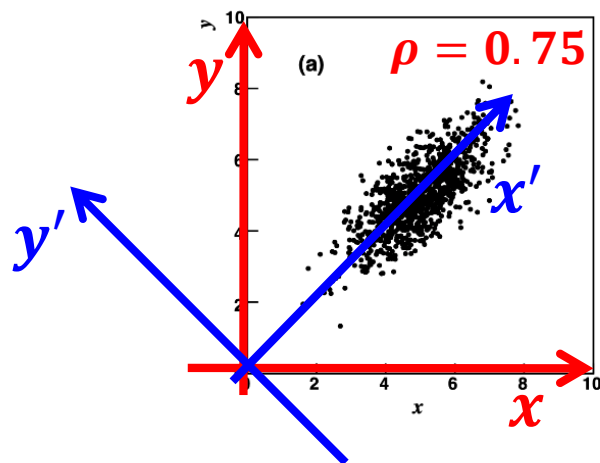
坐标变换后的协方差矩阵为

$$U_{r\theta} = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix} \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \begin{bmatrix} \frac{x}{r} & -\frac{y}{r^2} \\ \frac{y}{r} & \frac{x}{r^2} \end{bmatrix} = \frac{1}{r^2} \begin{bmatrix} x^2\sigma_x^2 + y^2\sigma_y^2 & \frac{xy(\sigma_y^2 - \sigma_x^2)}{r} \\ \frac{xy(\sigma_y^2 - \sigma_x^2)}{r} & x^2\sigma_x^2 + y^2\sigma_y^2 \end{bmatrix}$$

除非处处满足 $\sigma_x = \sigma_y$ ，否则 (r, θ) 有相关性。

不同坐标系下相关性的变化

通过转动坐标，随机变量的相关性会发生改变。



显然，通过将坐标系转动 45° ，上面的相关性在新坐标系下消失。

正交变换消除随机变量间的相关性

假设有 n 个随机变量 $\vec{x} = (x_1, \dots, x_n)$ 及其协方差矩阵 $V_{ij} = \text{cov}[x_i, x_j]$, 可以证明, 能够通过线性变换重新定义 n 个新的变量 $\vec{y} = (y_1, \dots, y_n)$, 使得对应的协方差矩阵 $U_{ij} = \text{cov}[y_i, y_j]$ 非对角元为零。

令这个变换为 $y_i = \sum_{j=1}^n A_{ij} x_j$ 或写成矩阵形式 $\vec{y} = A\vec{x}$

对应的协方差矩阵为

$$\begin{aligned} U_{ij} &= \text{cov}[y_i, y_j] \\ &= \\ \text{cov}\left[\sum_{k=1}^n A_{ik} x_k, \sum_{l=1}^n A_{jl} x_l\right] \\ &= \sum_{k,l=1}^n A_{ik} A_{jl} \text{cov}[x_k, x_l] \\ &= \sum_{k,l=1}^n A_{ik} A_{jl} V_{kl} \\ &= \sum_{k,l=1}^n A_{ik} V_{kl} A_{lj}^T \\ &= AVA^T \end{aligned}$$

非线性情况

$$\begin{aligned} U_{ij} &= \text{cov}[y_i, y_j] \\ &\approx \sum_{k,l=1}^n \left[\frac{\partial y_i}{\partial x_k} \frac{\partial y_j}{\partial x_l} \right]_{\vec{x}=\vec{\mu}} V_{kl} \end{aligned}$$

目标: 使矩阵 U 为对角矩阵

对角化变换后变量的协方差矩阵

选取适当的 A 使协方差矩阵 $U = AVA^T$ 对角化

先确定协方差矩阵 V 的本征列矢量 \vec{r}^i ($i = 1, \dots, n$)。解方程

$$V\vec{r}^i = \lambda_i\vec{r}^i \quad \text{或} \quad V_{kl}r_l^i = \lambda_i r_k^i$$

由于协方差矩阵总是对称的，可知本征矢量是正交的

$$\vec{r}^i \cdot \vec{r}^j = \sum_{k=1}^n r_k^i r_k^j = \delta_{ij}$$

变换矩阵 A 由本征矢量 \vec{r} 给出，即

$$\sum_{j=1}^n A_{ij}A_{jk}^T = \sum_{j=1}^n A_{ij}A_{jk}^T = \sum_{j=1}^n r_j^i r_j^k = \vec{r}^i \cdot \vec{r}^k = \delta_{ik}$$

正交变换后变量的协方差矩阵

因此，正则变换的协方差矩阵为

$$\begin{aligned} U_{ij} &= \sum_{k,l=1}^n A_{ik} V_{kl} A_{lj}^T \\ &= \sum_{k,l=1}^n r_k^i V_{kl} r_l^j \\ &= \sum_{k=1}^n r_k^i \lambda_j r_k^j \\ &= \lambda_j \vec{r}^i \cdot \vec{r}^j \\ &= \lambda_j \delta_{ij} \end{aligned}$$



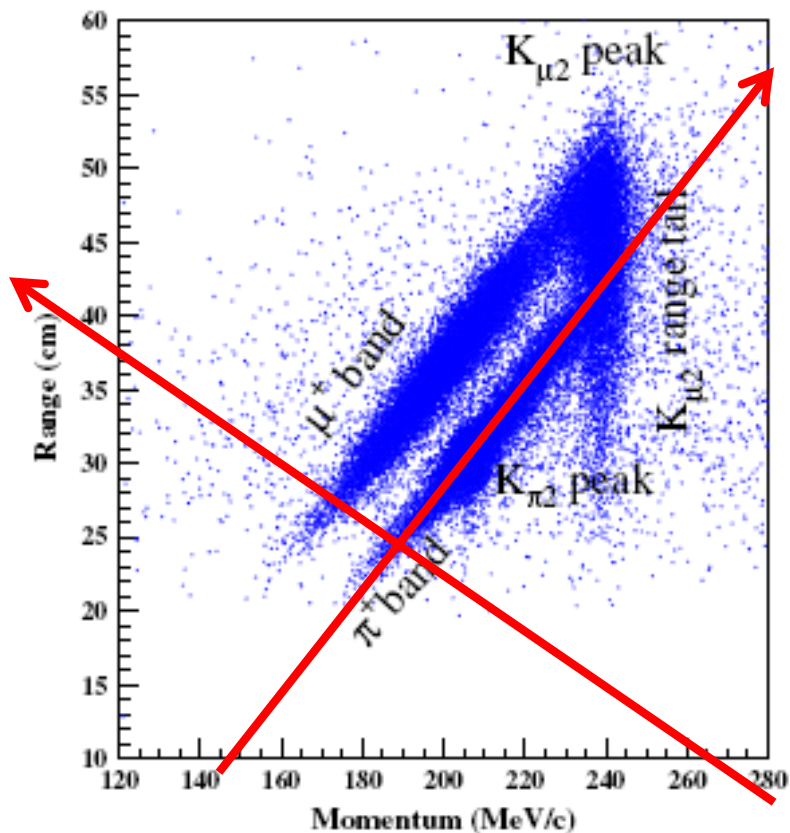
正交变换后，变量的方差由原协方差矩阵 V 的本征值给出。

对应于矢量的转动
不改变模的大小。

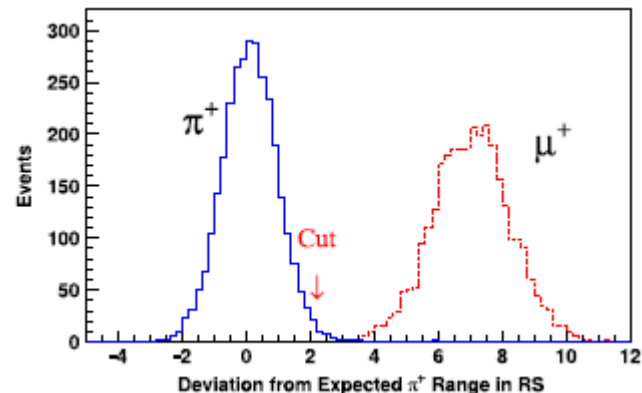
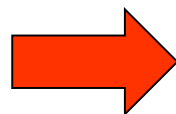
$$|\vec{y}|^2 = \vec{y}^T \vec{y} = \vec{x}^T A^T A \vec{x} = |\vec{x}|^2$$

尽管不相关的变量往往容易处理，但是对经过变换的变量的理解不一定容易。

带电粒子在闪烁体的射程



PHYSICAL REVIEW D 77, 052003 (2008)



在原来的定义下，可以得到粒子射程随动量大小的变化关系。通过转动变换，粒子的射程与动量发生了改变，没有物理含义，但是提供了一个很好的粒子类型甄别变量。

小结

➤ 概率的定义与诠释

- 随机事件
- 概率的定义与诠释（相对频率、主观）
- 条件概率、贝叶斯定理、全概率定理

➤ 随机变量与概率密度

- 随机变量、概率密度、直方图
- 联合概率密度、边缘概率密度、条件概率密度

➤ 随机变量的函数

- 一维（多维）随机变量的函数
- 梅林卷积、傅立叶卷积、雅克比行列式

➤ 期待值、方差

- 均值、方差、标准差
- 协方差、相关系数

➤ 不确定度的传递

- 不确定度的传递、不确定度与相关性
- 正交变换消除相关性