PEKING UNIVERSITY

实验物理中的统计方法

第一章：基本概念

杨振伟

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

- 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

什么是概率?

生活中的概率统计问题：摇号2018

北京普通小汽车摇号

北京市小客车指标调控管理信息系统

个人普通指标配置

期号：201804

描述：2018年第04期个人普通小客车指标配置结果

指标配置日期：2018-08-26

实际指标配置时间：2018-08-26 10:15:42

有效编码总数：13603346

指标配置总数：6333

指标配置种子数：269109

$p \approx 0.05\%$

如果一直保持相同的中签概率，100次不中签的概率？

生活中的概率统计问题：摇号2021

北京普通小汽车摇号

北京市小客车指标调控管理信息系统

家庭和个人普通指标配置

期号：202101

描述：2021年第01期家庭和个人普通小客车指标配置结果

指标配置日期：2021-06-26

实际指标配置时间：2021-06-26 10:07:51

基数序号总数：39624353

指标配置总数：19100

指标配置种子数：413238

$$p \approx 0.048\%$$

北京市小客车指标调控管理信息系统

家庭和个人普通指标配置

期号：202102

描述：2021年第02期家庭和个人普通小客车指标配置结果

指标配置日期：2021-12-26

实际指标配置时间：2021-12-26 10:06:50

基数序号总数：43040012

指标配置总数：19100

指标配置种子数：757380

$$p \approx 0.044\%$$

如果一直保持相同的中签概率，100次不中签的概率？

生活中的概率统计问题：摇号2023

北京普通小汽车摇号

2023年第01期家庭和个人普通小客车指标配置结果

配置时间：2023-06-26 10:08:15
基数序号总数：49145235
指标配置总数：14300
指标配置种子数：317744

$$p \approx 0.029\%$$

2023年第02期家庭和个人普通小客车指标配置结果

配置时间：2023-12-26 10:07:37
基数序号总数：52541681
指标配置总数：14300
指标配置种子数：898617

$$p \approx 0.027\%$$

如果一直保持相同的中签概率，100次不中签的概率？

生活中的概率统计问题

“花伴侣”中这些可信度什么意思？

随机事件

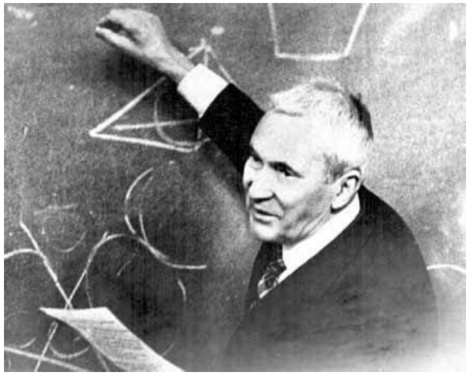
在一定的实验条件下，现象A可能发生，也可能不发生，并且只有发生或不发生这样两种可能性，这是偶然现象中一种比较简单的情形，我们把发生了现象A的事例称为随机事件A，简称事件A。
随机事件也称随机事例。

概率的定义

柯尔莫哥洛夫公理

考虑一全集S，有子集 $A, B, \ldots$

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



$\Rightarrow$ $P(A)$称为事件A的概率

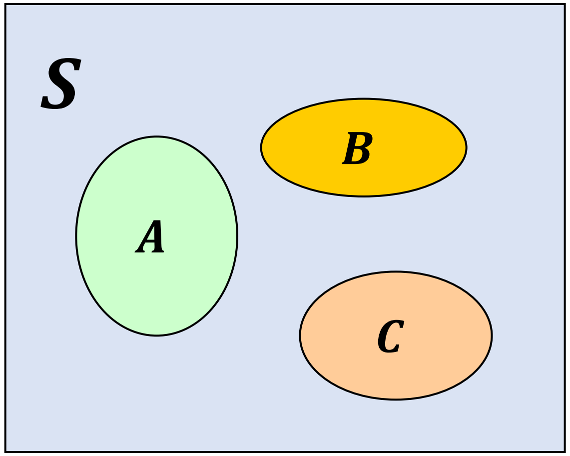
利用柯氏三公理可以得到

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

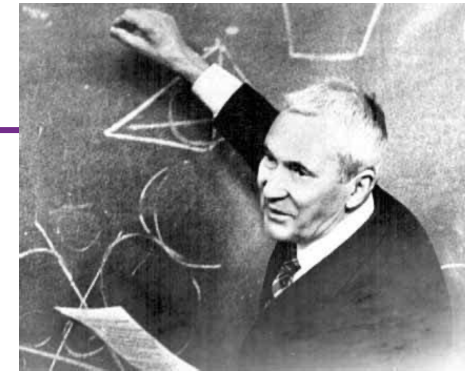
$P(A \cup \bar{A}) = 1$

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

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



更数学化的描述



1933年柯尔莫哥洛夫（Kolmogorov）基于集合论，给出了概率的公理化定义。

设S为样本空间，$\mathcal{F}$是由S的某些子集组成的一个事件域。如果对任意事件$A \in \mathcal{F}$，定义在$\mathcal{F}$上的一个实值函数$P(A)$满足：

(1) 非负性公理：若$A \in \mathcal{F}$，则$P(A) \geq 0$;

(2) 正则性（规范性）公理：$P(S) = 1$;

(3) 可列可加性公理：若$A_1, A_2, \cdots, A_n, \cdots$互斥，则$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$;

则称$P(A)$为事件A的概率。三元素$(S, \mathcal{F}, P)$称为概率空间。

概率的本质

概率的公理化定义描述了概率的本质：

概率是事件（集合）的函数。

若在事件域 $\mathcal{F}$ 上给出一个函数，只要这个函数满足柯氏公理，就可以称该函数为概率。

可以证明历史上出现过的频率定义、几何定义、古典定义、主观定义等，均满足柯氏公理。更准确地说，这些定义是概率的不同解释。

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

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

$\Rightarrow$ 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$A, B, \ldots$是可重复实验的一组结果，则概率是

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

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$A, B, \ldots$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 $m_t = 173.49 \pm 1.07\ \mathrm{GeV}/c^2$ ？

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

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

贝叶斯 (1702-1761)

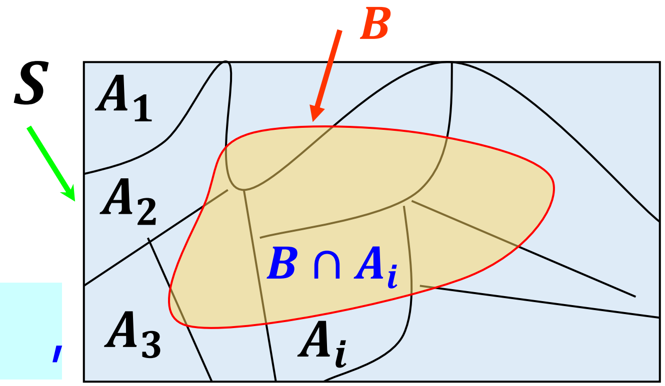
利用 $P(A \cap B) = P(B \cap A)$，可得

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

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 $\cup_i A_i = S$ 。



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

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

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

利用条件概率公式

全概率定律

贝叶斯定理

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

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

$$\underbrace{P(\text{理论}|\text{实验})}_{\text{验后概率}} = \underbrace{\frac{P(\text{实验}|\text{理论})}{P(\text{实验})}}_{\text{似然性}} \cdot \underbrace{P(\text{理论})}_{\text{先验概率}}$$

这是个“认识-实践-再认识-再实践”的迭代过程：

理论 + 实验 → 改进的理论 + 新实验 → 新改进的理论

- 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

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

$P(\mathrm{nCov}) = 0.001$
$P(\text{no nCov}) = 0.999$

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

$P(+|\mathrm{nCov}) = 0.98$　　nCov感染患者阳性的概率

$P(-|\mathrm{nCov}) = 0.02$　　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，

$$P(\text{nCov}|+)=\frac{\overbrace{P(+|\text{nCov})P(\text{nCov})}^{\boxed{\text{nCoV患者阳性}}}}{\underbrace{P(+|\text{nCov})P(\text{nCov})+P(+|\text{no nCov})P(\text{no nCov})}_{\boxed{\text{所有为阳性结果的人}}}}$$

$$=\frac{0.98\times 0.001}{0.98\times 0.001+0.03\times 0.999}=0.032$$

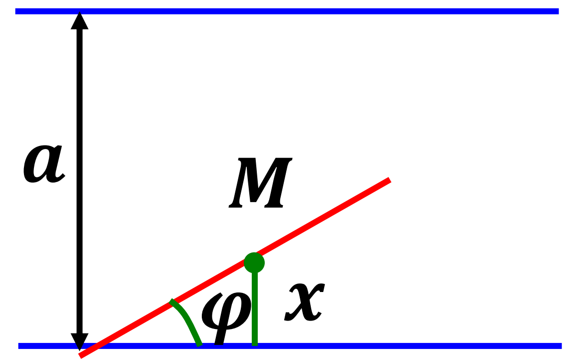
也就是说，你可能没什么问题！？

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AIDS。

涉及到如何诠释结果（概率）的问题！

例：蒲丰投针

- 1777年,法国科学家蒲丰(Buffon)提出了投针试验问题.平面上画有等距离为$\boldsymbol{a}$(>0)的一些平行直线,现向此平面任意投掷一根长为$\boldsymbol{b}$(<$\boldsymbol{a}$)的针,试求针与任一平行直线相交的概率。



解

以 x 表示针投到平面上时，针的中点 M 到最近一条平行线的距离，

φ表示该针与该平行直线夹角，

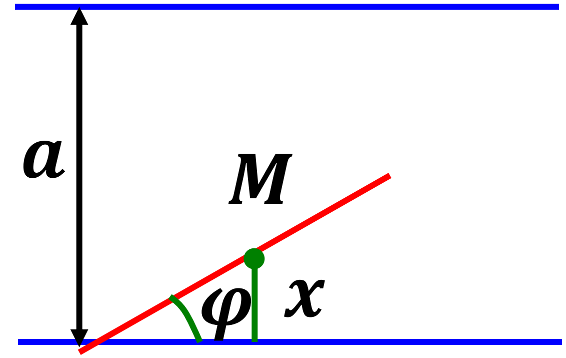
那么针落在平面上的位置可由(x, φ)完全确定。

例：蒲丰投针（续）

投针实验的所有可能结果与矩形区域

$$S=\left\{(x,\varphi)\middle|0\le x\le\frac{a}{2},0\le\varphi\le\pi\right\}$$

中的所有点一一对应

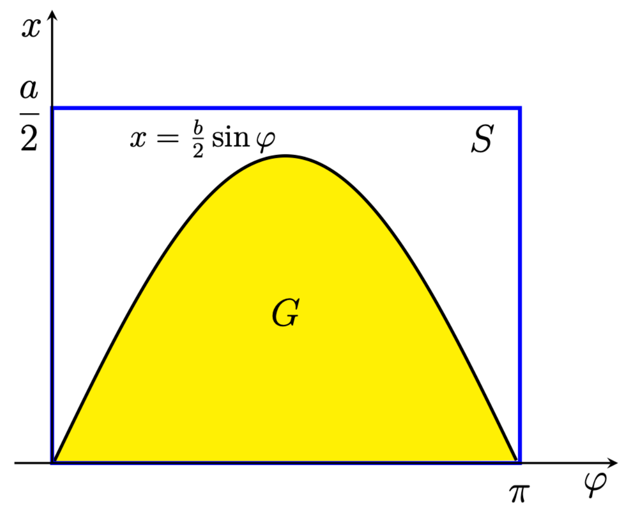


所关心的事件

$A=\{$ 针与任一平行直线相交 $\}$

发生的充分必要条件为满足

$$0\le x\le\frac{b}{2}\sin\varphi,0\le\varphi\le\pi$$



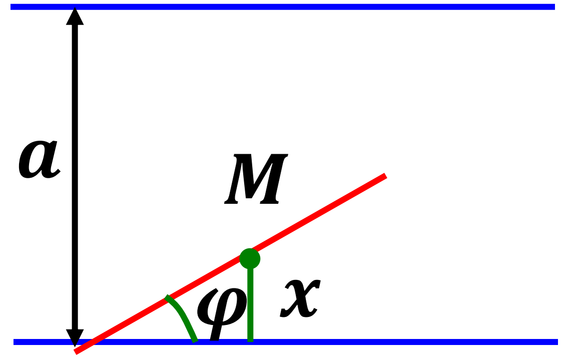
由投掷的任意性可知，这是一个几何概型问题。

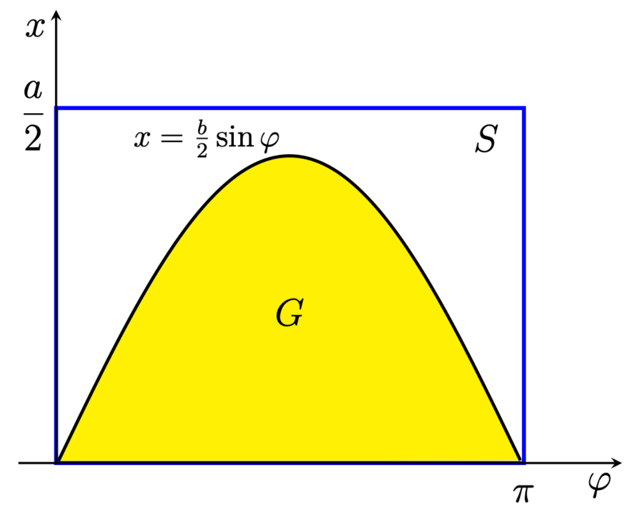
例：蒲丰投针（续）

$$P(A)=\frac{m(G)}{m(S)}=\frac{G\text{的面积}}{S\text{的面积}}$$

$$=\frac{\int_0^{\pi}\frac{b}{2}\sin\varphi d\varphi}{\frac{a}{2}\times\pi}$$

$$=\frac{b}{\frac{a}{2}\times\pi}=\frac{2b}{a\pi}$$





例：蒙提霍尔问题

- 蒙提霍尔问题，亦称蒙特霍问题或三门问题（英文：Monty Hall problem），是一个源自博弈论的数学游戏问题，以美国的电视游戏节目 *Let's Make a Deal* 而闻名。问题的名字来自该节目的主持人蒙提·霍尔（Monty Hall）。

例：蒙提霍尔问题

- 参赛者可看见三扇关闭的门，其中一扇门后面有一辆汽车，选中它即可赢得汽车；另外两扇门后面各藏有一只山羊。
  参赛者先选定一扇门，但在开启之前，知情的主持人会在其余两扇门中打开一个后面是山羊的门，并讯问参赛者是否更换其选择的门。
  问题是：参赛者更换选择是否会增加赢得汽车的机会？

例：蒙提霍尔问题

三门问题的贝叶斯解法：

不失一般性，可以始终将嘉宾选择的门标记为“1”，将主持人打开的门标记为“3”，将剩下的门标记为“2”。

第 i 个门后面是汽车的先验概率为

$$P(T_i) = 1/3$$

如果汽车在1号门，主持人打开3号门的条件概率为

$$P(O_3|T_1) = 1/2$$

如果汽车在2号门，主持人打开3号门的条件概率为

$$P(O_3|T_2) = 1$$

如果汽车在3号门，主持人打开3号门的条件概率为

$$P(O_3|T_3) = 0$$

例：蒙提霍尔问题

由贝叶斯定理，主持人打开3号门的条件下，在1号门和2号门发现汽车的条件概率分别为：

$$P(T_1|O_3) = \frac{P(O_3|T_1)P(T_1)}{\sum_{i=1}^{3} P(O_3|T_i)P(T_i)} = \frac{P(O_3|T_1)}{\sum_{i=1}^{3} P(O_3|T_i)}$$

$$= \frac{1/2}{1/2 + 1 + 0} = \frac{1}{3}$$

$$P(T_2|O_3) = \frac{P(O_3|T_2)P(T_2)}{\sum_{i=1}^{3} P(O_3|T_i)P(T_i)} = \frac{P(O_3|T_2)}{\sum_{i=1}^{3} P(O_3|T_i)}$$

$$= \frac{1}{1/2 + 1 + 0} = \frac{2}{3}$$

所以，嘉宾变更选择后获得汽车的概率更大。

案例分析：探测器触发效率

大型强子对撞机上的质子束流每 $25\ \mathrm{ns}$ 对撞一次，即事例率为$40\ \mathrm{MHz}$。假如每次对撞探测器的数据需占用约 $1\ \mathrm{MB}$ 磁盘空间，那么每秒钟需要记录$40\ \mathrm{TB}$。这显然不现实。

好在对撞事例中产生我们关心的信号的概率很小（假设约10^{-6}），因此可在硬件上编程设置一些筛选条件，只记录满足筛选条件的事例。这些筛选条件被称为“触发（trigger）”条件，应尽量满足：
1）信号被保留下来的概率尽量大（触发效率高）；
2）本底被保留下来的概率尽量小；
3）响应时间不能太长。

假设一级触发要求信号被保留下来的概率（效率）不低于99%，本底被保留下来的概率不高于10^{-4}。

求记录下来的事例中有多大概率是信号?

案例分析：探测器触发效率

解 用“sig”表示对撞中产生了信号；“bkg”表示对撞中没有信号产生，即只有本底；“w”表示对撞事例被记录了下来。

显然$P(\mathrm{sig}) = 10^{-6}$，$P(\mathrm{bkg}) = 1 - 10^{-6}$；
$P(\mathrm{w}|\mathrm{sig}) = 0.95$，$P(\mathrm{w}|\mathrm{bkg}) = 10^{-4}$。

记录下来的事例为信号的概率为

$$P(\mathrm{sig}|\mathrm{w}) = \frac{P(\mathrm{w}|\mathrm{sig})P(\mathrm{sig})}{P(\mathrm{w}|\mathrm{sig})P(\mathrm{sig}) + P(\mathrm{w}|\mathrm{bkg})P(\mathrm{bkg})}$$

$$= \frac{0.99\times10^{-6}}{0.99\times10^{-6} + 10^{-4}\times(1 - 10^{-6})}$$

$$= 9.8\times10^{-3}$$

即，记录下来的数据只有不到1%的事例含有信号。

“触发”的概念在很多领域都会用，例如各种类型的“预警”。

案例分析：古玩多赝品

假设古董市场中：

真货概率 $P(\text{真货})=10^{-5}$，

假货概率 $P(\text{假货})=1-P(\text{真货})=0.99999$。

如果古董被鉴别为真货，记为事件T，被鉴别为假货，记为事件F。

假设某资深玩家对古董鉴别能力比较高，当古董为真货时该玩家将其鉴别为真的条件概率为$P(T|\text{真货})=0.999$，当古董为假货时，将其鉴别为真的条件概率为$P(T|\text{假货})=0.01$。

那么，对任意一个古董，假如该玩家认为这是真古董，那么该古董实际上为假货的概率是多少?

案例分析：古玩多赝品

解：　需要计算的是，在鉴别为真货的条件下古董实际上为假货的条件概率$P(\text{假货}|T)$，用贝叶斯定理：

$$P(\text{假货}|T)=\frac{P(T|\text{假货})P(\text{假货})}{P(T|\text{假货})P(\text{假货})+P(T|\text{真货})P(\text{真货})}$$

$$=\frac{0.01\times0.99999}{0.01\times0.99999+0.999\times10^{-5}}$$

$$=0.999$$

真相：该资深玩家仔细鉴别后买下的古董中，

真货的概率仅0.1%。

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$[x, x+\mathrm{d}x]$区间内的概率为

$$P(x \in [x, x+\mathrm{d}x]) = f(x)\mathrm{d}x$$

$f(x)$ 为概率密度函数(pdf)

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

归一化：x 一定是某个实数

对于离散结果 x_i $(i = 1,2,\ldots)$：

$$P(x_i) = p_i$$

概率质量函数(pmf)

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

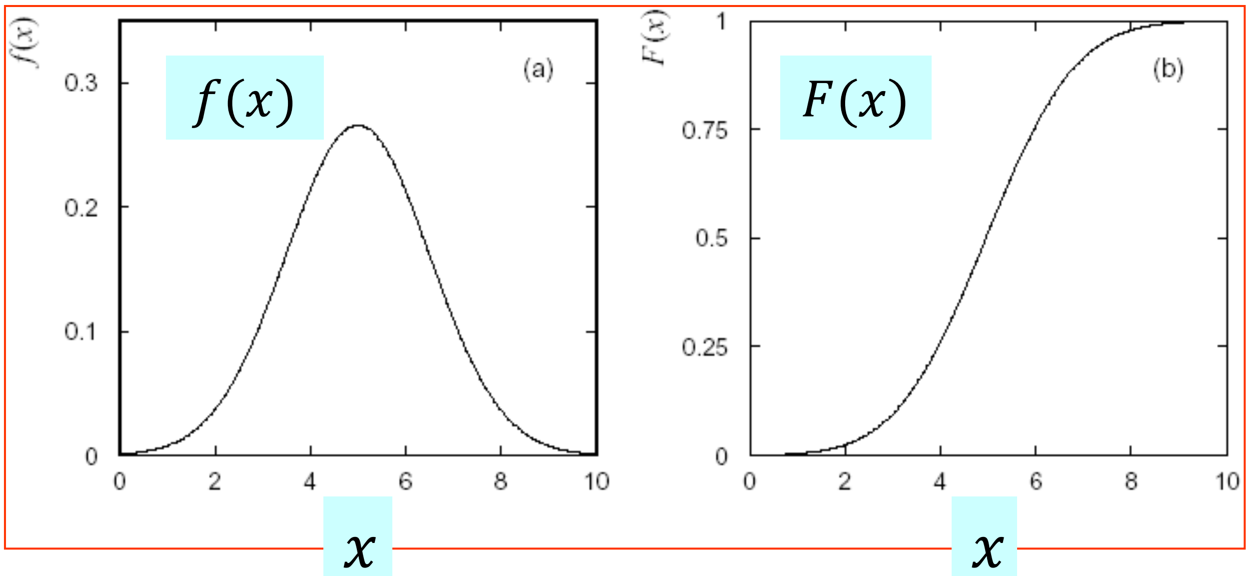
归一化

累积分布函数

- 实验结果小于等于 x 的概率

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

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、中位数与众数

分位数 x_α 定义为随机变量 x 的值，它使得

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

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（module）

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。

众数定义为使概率密度函数值达到极大的随机变量值。

α 分位数（quantile）

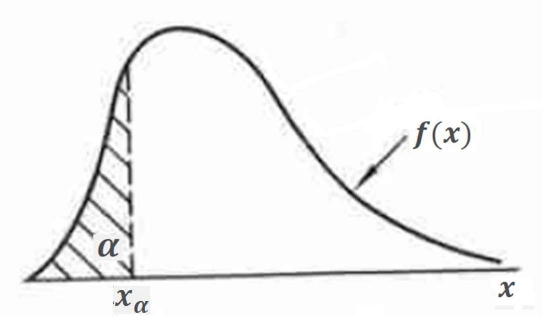
设X为随机变量，$0 < \alpha < 1$，若x_α满足

$$P(X \leq x_\alpha) = F(x_\alpha) = \alpha$$

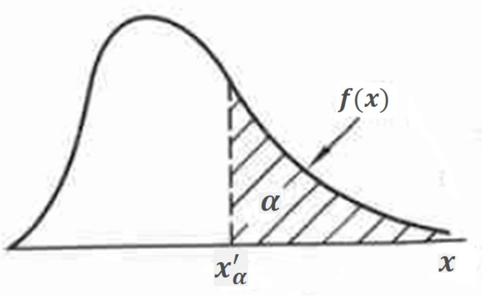
则称x_α为X的α-分位数，亦称下侧α-分位数。若x'_α满足

$$P(X \geq x'_\alpha) = \alpha$$

则x'_α被称为上侧α-分位数。



下侧α-分位数



上侧α-分位数

对连续变量$x'_\alpha = x_{1-\alpha}$，$x_\alpha = x'_{1-\alpha}$

中位数（median）与众数（mode）

随机变量 X 的中位数（或中值）定义为

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

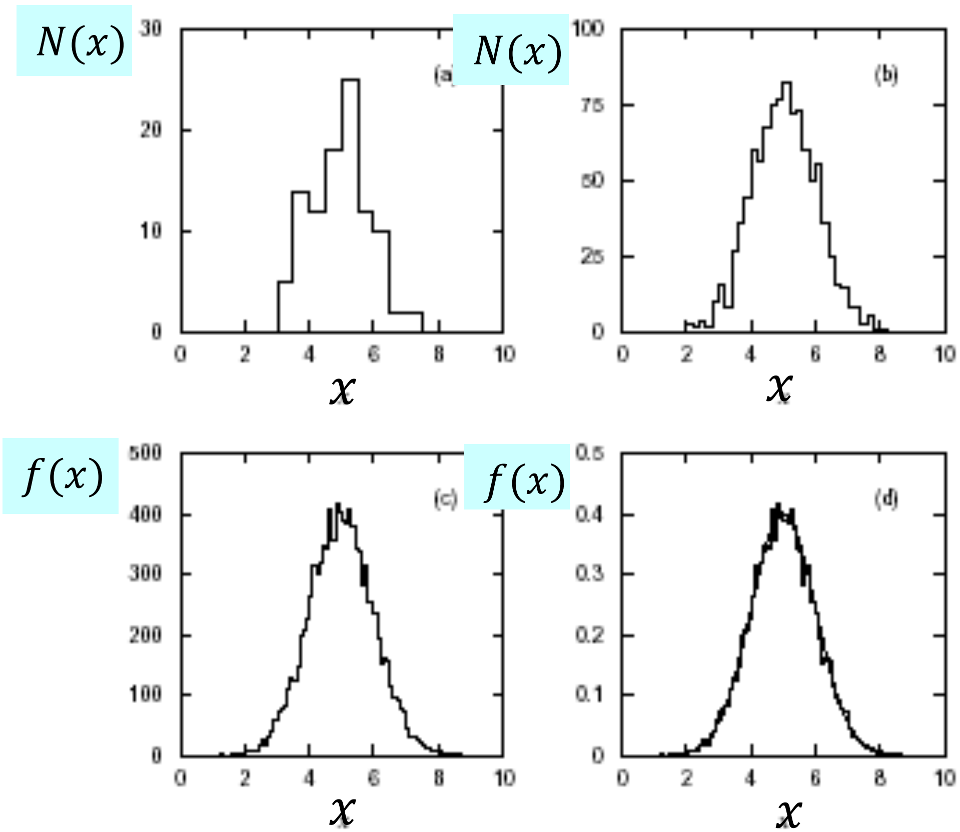
随机变量 X 被观测到大于或小于中位数的概率相等。

注意中位数与均值的区别。

众数定义为使概率密度函数值达到极大的随机变量值。
众数也被称为最可几值。

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。



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

$N(x)$ =每个区间的频数

n =填入直方图的总事例数

Δx =区间的宽度

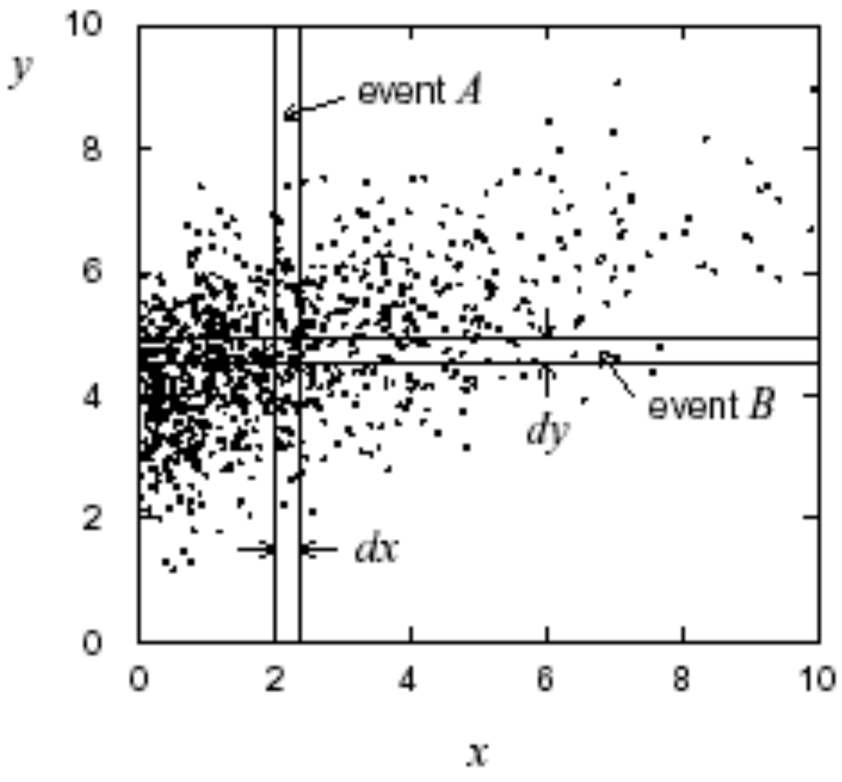
$$n \to \infty, \Delta x \to 0$$

$n\cdot\Delta x$有限

直方图在统计分析中非常重要，应准确理解它的含义。

多变量情形

观测量不止一个，例如 x 与 y



$$P(A \cap B) = f(x, y)\mathrm{d}x\mathrm{d}y$$

$f(x, y)$ =联合概率密度函数

$$\iint f(x, y)\mathrm{d}x\mathrm{d}y = 1$$

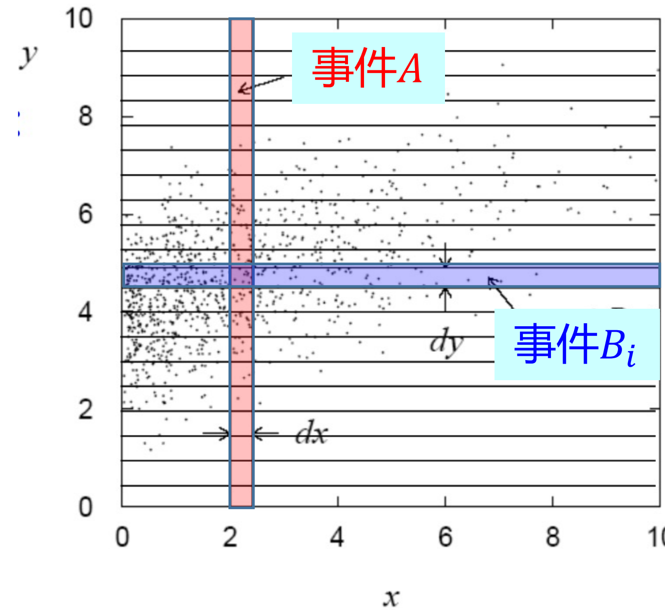
边缘概率密度

有时，我们只关心某个分量的pdf：

$$P(A) = \sum_i P(A \cap B_i)$$

$$= \sum_i f(x, y_i)\mathrm{d}y\,\mathrm{d}x$$

$$\to \int f(x, y)\mathrm{d}y\,\mathrm{d}x$$



$$f_x(x) = \int f(x, y)\mathrm{d}y$$

边缘概率密度

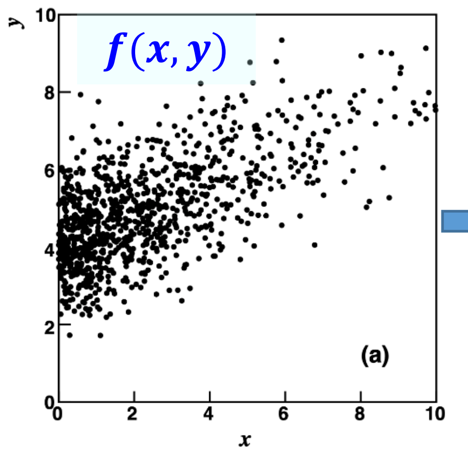
类似地

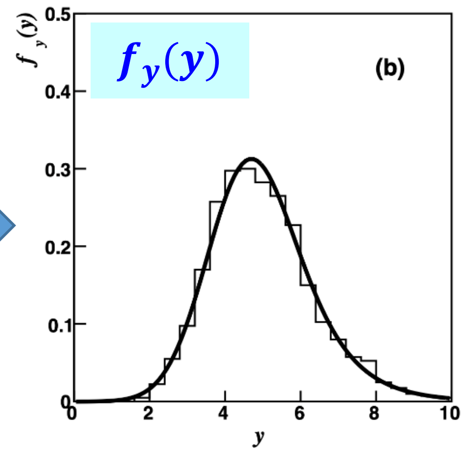
$$f_y(y) = \int f(x, y)\mathrm{d}x$$

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

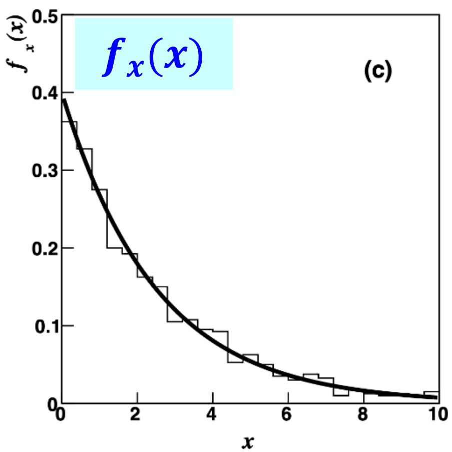
边缘概率密度：将联合概率密度投影某个坐标轴





投影到 y 轴，就是对x积分：

$$f_y(y) = \int f(x,y)\mathrm{d}x$$



投影到 x 轴，就是对y积分：

$$f_x(x) = \int f(x,y)\mathrm{d}y$$

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：
$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{f(x,y)\mathrm{d}x\mathrm{d}y}{f_x(x)\mathrm{d}x}$$

条件概率密度函数
$$h(y|x) = \frac{f(x,y)}{f_x(x)}, \qquad g(x|y) = \frac{f(x,y)}{f_y(y)}$$

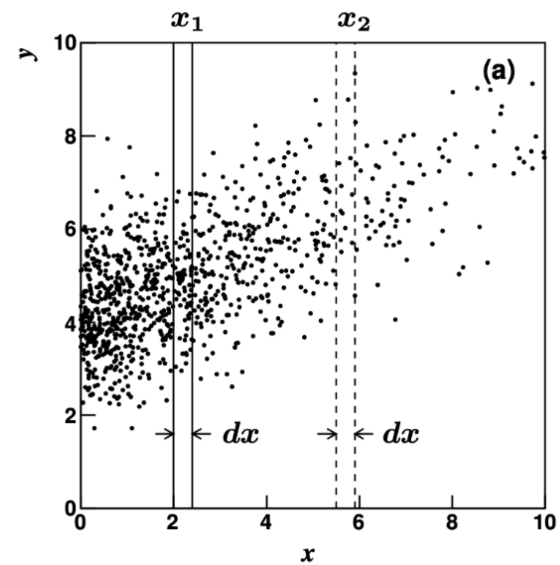
贝叶斯定理可写为
$$g(x|y) = \frac{h(y|x)f_x(x)}{f_y(y)}$$

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)

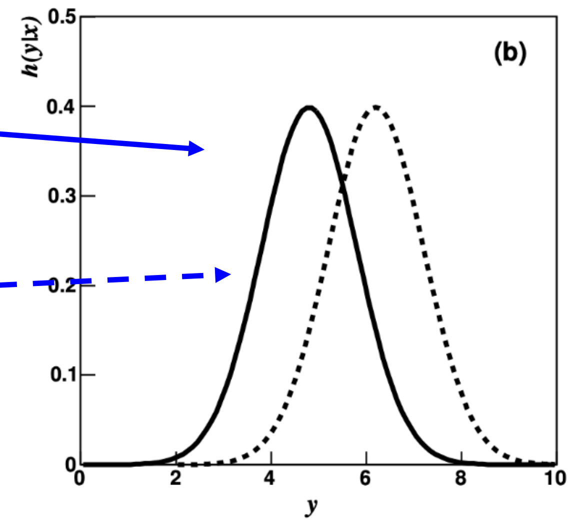
例如，已知联合概率密度$f(x,y)$。
求条件概率密度 $h(y|x_1)$，$h(y|x_2)$。



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

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

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



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

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \qquad p_{\mathrm{z}} \qquad f(p_{\mathrm{T}}, p_z)$$

在已知两分量测量值的概率密度函数情况下，总动量为

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

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。

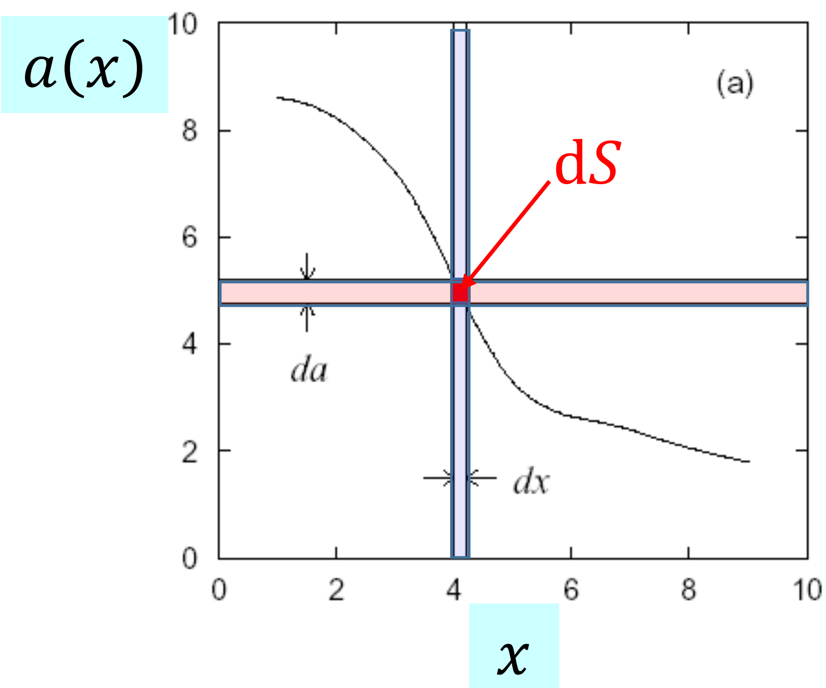
随机变量的函数自身也是一个随机变量。

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。



$a \in [a, a + \mathrm{d}a]$ 的概率 $g(a)\mathrm{d}a$
等于 $x \in \mathrm{d}S$ 的概率 $f(x)\mathrm{d}x$

$\mathrm{d}S$：满足$a \in [a, a + \mathrm{d}a]$的$x$的集合

$$g(a)\mathrm{d}a = \int_{\mathrm{d}S} f(x)\mathrm{d}x$$

$$= \left| \int_{x(a)}^{x(a+\mathrm{d}a)} f(x')\mathrm{d}x' \right|$$

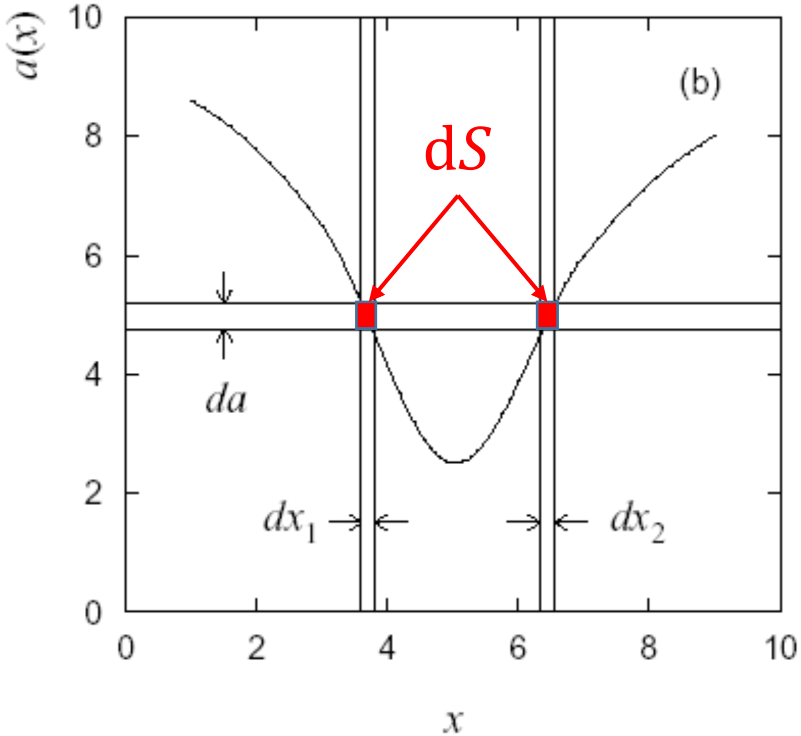
$$= \int_{x(a)}^{x(a)+\left|\frac{\mathrm{d}x}{\mathrm{d}a}\right|\mathrm{d}a} f(x')\mathrm{d}x'$$

$$\Rightarrow \quad \boldsymbol{g(a) = f\big(x(a)\big)\left|\frac{\mathrm{d}x}{\mathrm{d}a}\right|}$$

函数的逆不唯一情况

假如 $a(x)$ 的逆不唯一，$\mathrm{d}S$ 将包含多段 $\mathrm{d}x$ 区间，需要全部考虑进来

例如：$a = x^2$，$x = \pm\sqrt{a}$，$\mathrm{d}x = \pm\dfrac{\mathrm{d}a}{2\sqrt{a}}$



$$g(a)\mathrm{d}a = \int_{\mathrm{d}S} f(x)\mathrm{d}x$$

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

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

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$g(a)$

$$g(a')\mathrm{d}a' = \int \cdots \int_{\mathrm{d}S} f(x_1, \dots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

$\mathrm{d}S$：在$a(\vec{x}) = a'$和$a(\vec{x}) = a' + \mathrm{d}a'$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：$z = xy$

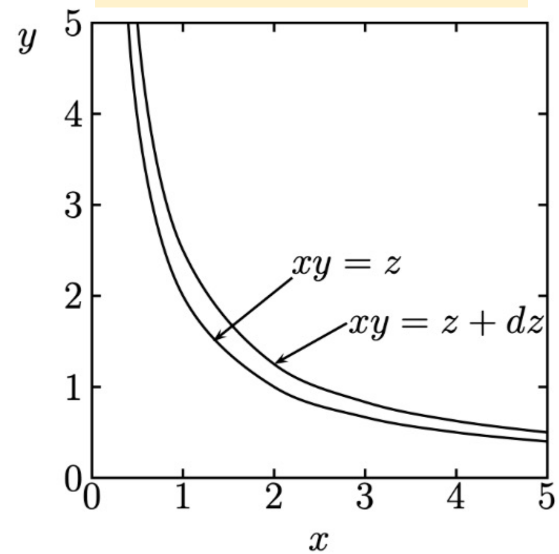
如果两个随机变量 $x, y > 0$，服从联合概率密度 $f(x,y)$，考虑函数 $z = xy$，其概率密度函数$g(z)$是什么形式?

$$g(z)\mathrm{d}z = \iint_{\mathrm{d}S} f(x,y)\mathrm{d}x\mathrm{d}y \;= \int_0^{\infty} \mathrm{d}x \int_{z/x}^{(z+\mathrm{d}z)/x} f(x,y)\mathrm{d}y$$

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

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

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



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

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

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?

$$g(z)\mathrm{d}z = \iint_{\mathrm{d}S} f(x,y)\mathrm{d}x\mathrm{d}y \quad = \int_0^{\infty} \mathrm{d}x \int_{z-x}^{z+\mathrm{d}z-x} f(x,y)\mathrm{d}y$$

$$= \int_0^{\infty} \mathrm{d}x\, [f(x, z-x)\mathrm{d}z]$$

$$= \left[\int_0^{\infty} f(x, z-x)\mathrm{d}x\right] \mathrm{d}z$$

给定x，
当$z \in (z, z+\mathrm{d}z)$时，
$y \in (z-x, z+\mathrm{d}z-x)$。

$$g(z) = \int_0^{\infty} f(x, z-x)\mathrm{d}x$$

$$g(z) = \int_0^{\infty} f(z-y, y)\mathrm{d}y$$

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?

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

→ 梅林(Mellin)卷积 $\boldsymbol{f = g \otimes h}$

$z = x + y$ 的概率密度函数$f(z)$是什么形式?

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

→ 傅立叶(Fourier)卷积 $\boldsymbol{f = g \otimes h}$

多维随机变量的函数与雅克比行列式

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 $\vec{a}(\vec{x}) = \left(a_1(\vec{x}), \dots, a_n(\vec{x})\right)$，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

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

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 $\boldsymbol{g(\vec{a})}$ 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 $\boldsymbol{g_i(a_i)}$ 。
这是数据分析中误差传递的基础。

例：坐标变换

在物理测量中，经常需要直角坐标系和极坐标系互换。假设x和y是相互独立的随机变量，均服从$N(0,1)$分布。

试证明变换为极坐标(ρ,φ)后，(ρ,φ)仍为相互独立的随机变量。其中

$$\rho=\sqrt{x^2+y^2},\qquad \rho>0$$

$$\varphi=\mathrm{arctg}\left(\frac{y}{x}\right),\qquad \varphi\in[0,2\pi]$$

坐标变换

证明： 利用变量变换法求解。

(x,y)的联合密度函数为 $f(x,y)=\frac{1}{2\pi}e^{-\frac{(x^2+y^2)}{2}}$

ρ和φ的反函数分别为， $x=\rho\cos\varphi$ 和 $y=\rho\sin\varphi$。

雅克比行列式

$$J=\begin{vmatrix}\frac{\partial x}{\partial \rho} & \frac{\partial y}{\partial \rho}\\ \frac{\partial x}{\partial \varphi} & \frac{\partial y}{\partial \varphi}\end{vmatrix}=\begin{vmatrix}\cos\varphi & \sin\varphi\\ -\rho\sin\varphi & \rho\cos\varphi\end{vmatrix}=-\rho$$

坐标变换

因此(ρ,φ)的联合密度函数为

$$f(\rho,\varphi)=f\big(x(\rho,\varphi),y(\rho,\varphi)\big)|J|$$

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{(\rho\sin\varphi)^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{(\rho cos\varphi)^2}{2}}\cdot\rho$$

$$=\frac{1}{2\pi}\rho e^{-\frac{\rho^2}{2}}\qquad\qquad \rho\geq 0,\qquad 0\leq\varphi\leq 2\pi$$

由于联合密度函数与φ无关，根据随机变量相互独立的定义，ρ与φ必然相互独立。

坐标变换

$$f(\rho,\varphi)=\frac{1}{2\pi}\rho e^{-\frac{\rho^2}{2}}\qquad \rho\geq 0,\qquad 0\leq\varphi\leq 2\pi$$

对$f(\rho,\varphi)$关于φ求积分，得到ρ的边缘密度函数：

$$f(\rho)=\int_0^{2\pi}\frac{1}{2\pi}\rho e^{-\frac{\rho^2}{2}}\mathrm{d}\varphi=\rho e^{-\frac{\rho^2}{2}},\qquad (\rho\geq 0)$$

（著名的瑞利分布）

对$f(\rho,\varphi)$关于ρ求积分，得到φ的边缘密度函数：

$$f(\varphi)=\int_0^{\infty}\frac{1}{2\pi}\rho e^{-\frac{\rho^2}{2}}\mathrm{d}\rho=\frac{1}{2\pi},\qquad (\varphi\in[0,2\pi])$$

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

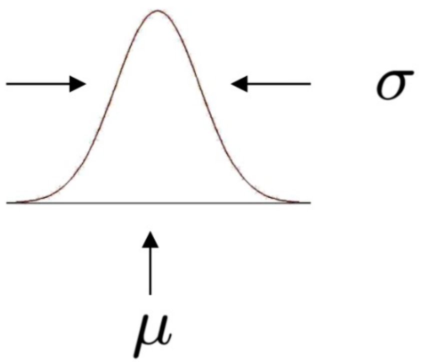
考虑概率密度为 $f(x)$ 的随机变量 x，定义期待 (平均)值为

$$\boldsymbol{E[x] = \int x f(x) \mathrm{d}x}$$

通常记为：$\boldsymbol{E[x] = \mu}$

意为pdf的“重心”

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



对概率密度为 $g(y)$ 的函数 $y(x)$，有

$$\boldsymbol{E[y] = \int y g(y) \mathrm{d}y = \int y(x) f(x)\, \mathrm{d}x}$$

方差定义为 $\boldsymbol{V[x] = E\left[(x - E[x])^2\right] = E\left[x^2\right] - \mu^2}$

通常记为：$\boldsymbol{V[x] = \sigma^2}$

标准差：$\boldsymbol{\sigma \equiv \sqrt{\sigma^2}}$

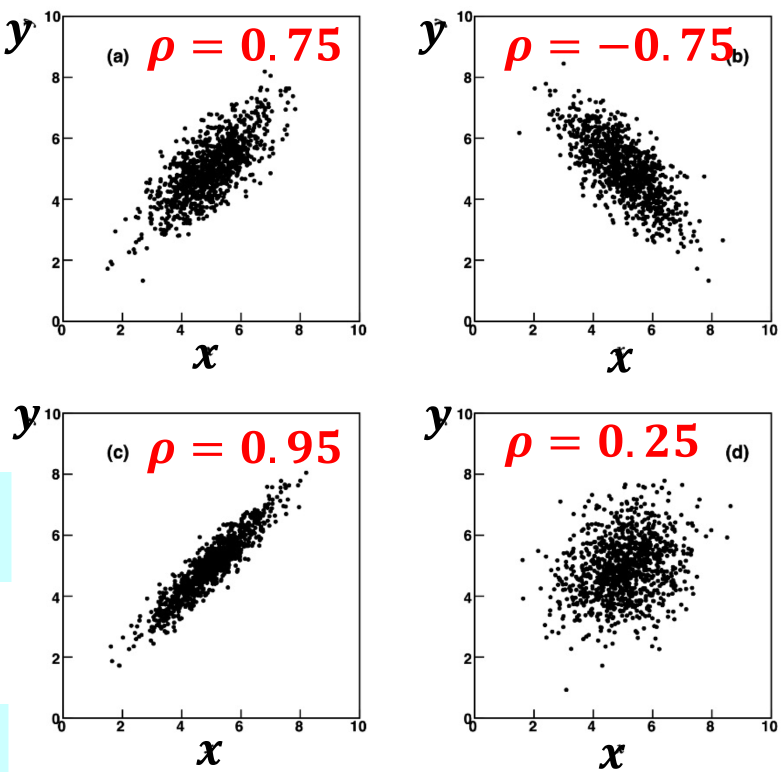
协方差与相关系数

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

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

相关系数定义为

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



如果 x, y 相互独立，即

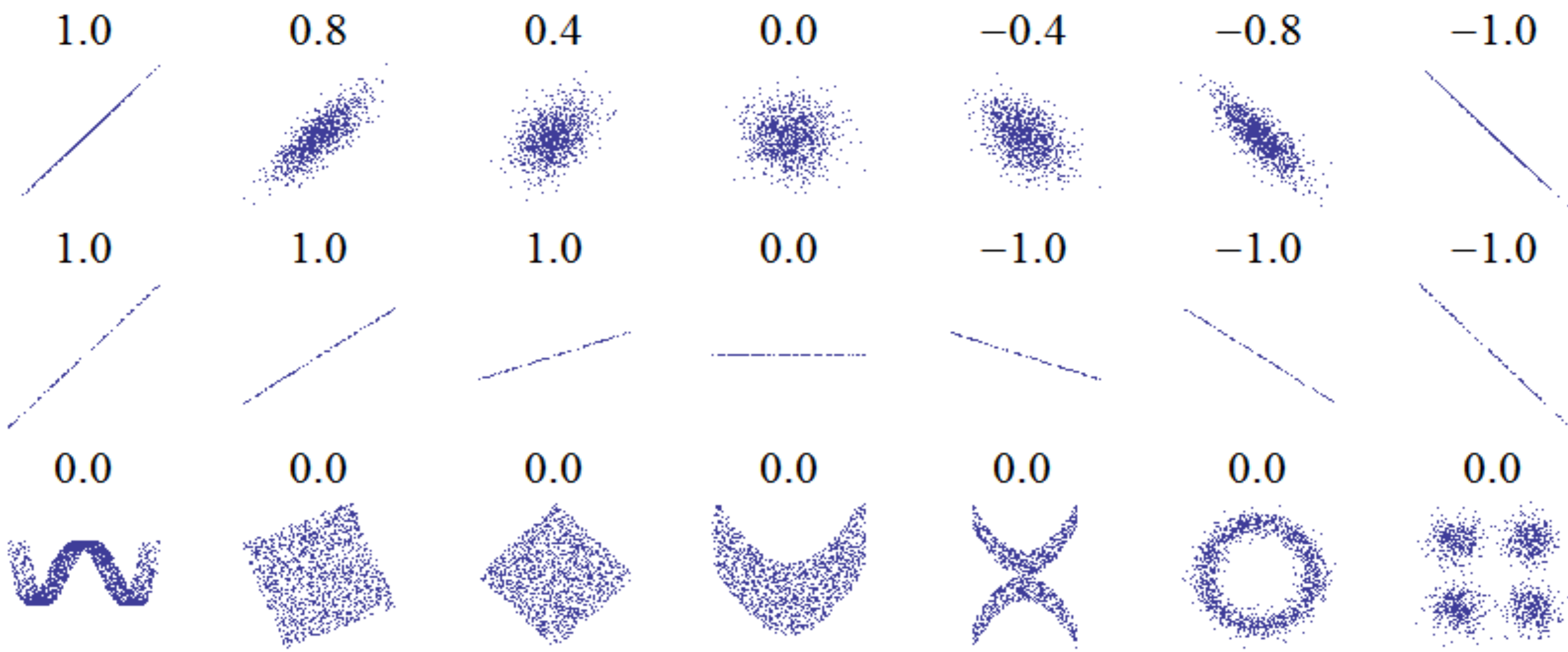
$$f(x,y) = f_x(x)f_y(y)$$

$$E[xy] = \iint xyf(x,y)\mathrm{d}x\mathrm{d}y = \mu_x\mu_y$$

则 $\mathrm{cov}[x,y] = 0$ ⇒ $\rho_{xy} = 0$

即x,y 不相关

相关性的例子



$\rho = 0$，但显然有一定的相关性

相关性的其他定义

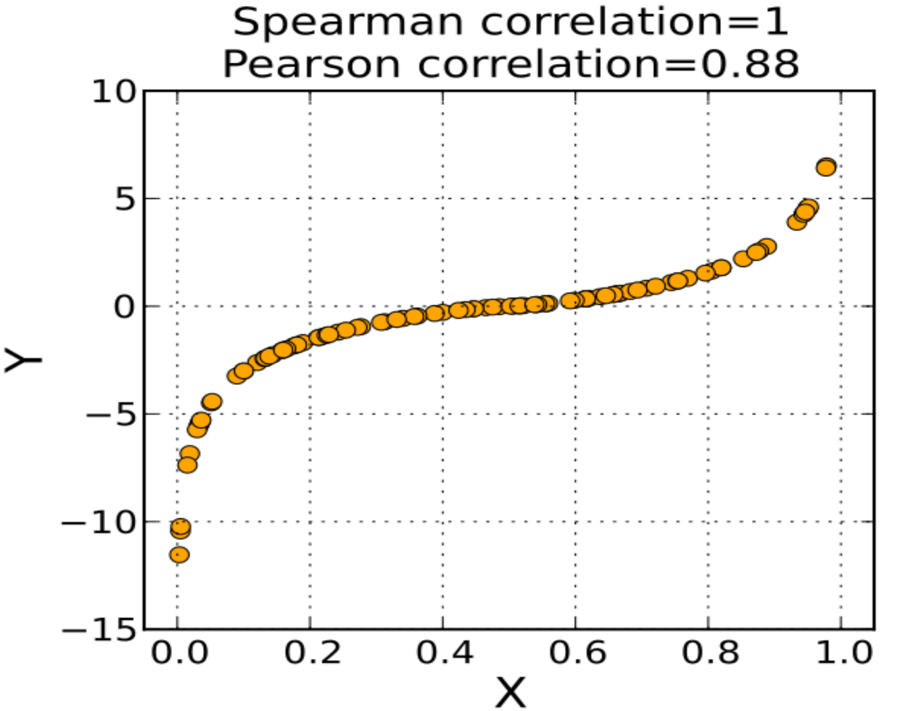
$$\rho_{xy} = \frac{\mathrm{cov}[x, y]}{\sigma_x \sigma_y}$$

皮尔逊相关系数

秩相关性（rank correlation）

- ➢ 比较两个样本的排序
  ➔ 可以更好地度量相关性

- ➢ 两种主要的秩相关系数：
  - Spearman’s rho
  - Kendall’s tau



Spearman’s rho $$r_s = \rho_{\mathrm{R}(X),\mathrm{R}(Y)} = \frac{\mathrm{cov}(\mathrm{R}(X),\mathrm{R}(Y))}{\sigma_{\mathrm{R}(X)}\sigma_{\mathrm{R}(Y)}}$$

原点矩和中心矩

设X为随机变量，k为正整数。若以下数学期望均存在，则称

$$\mu_k = E(X^k)$$

为X的k阶原点矩。称

$$\nu_k = E\left[\left(X - E(X)\right)^k\right]$$

为X的k阶中心矩。

1阶原点矩为数学期望，2阶中心矩为方差。

中心矩与原点矩的关系：

$$\nu_k = E\left[\left(X - E(X)\right)^k\right] = E[(X - \mu_1)^k] = \sum_{i=0}^{k} C_k^i \mu_i (-\mu_1)^{k-i}$$

其他特征数

- **变异系数**：设X为随机变量，且方差存在。则称 $C_V=\frac{\sqrt{\mathrm{Var}(X)}}{E(X)}$ 为X的变异系数。

C_V 是无量纲量，用于比较量纲不同的随机变量的波动大小。

- **偏度系数**：设随机变量X的前三阶矩存在，则比值$\beta_S=\frac{\nu_3}{\nu_2^{3/2}}=\frac{E\left[\left(X-E(X)\right)^3\right]}{[\mathrm{Var}(X)]^{3/2}}$被称为$X$的偏度系数，简称偏度。或记为$\gamma_1$ 。

β_S描述分布偏离对称性的程度。

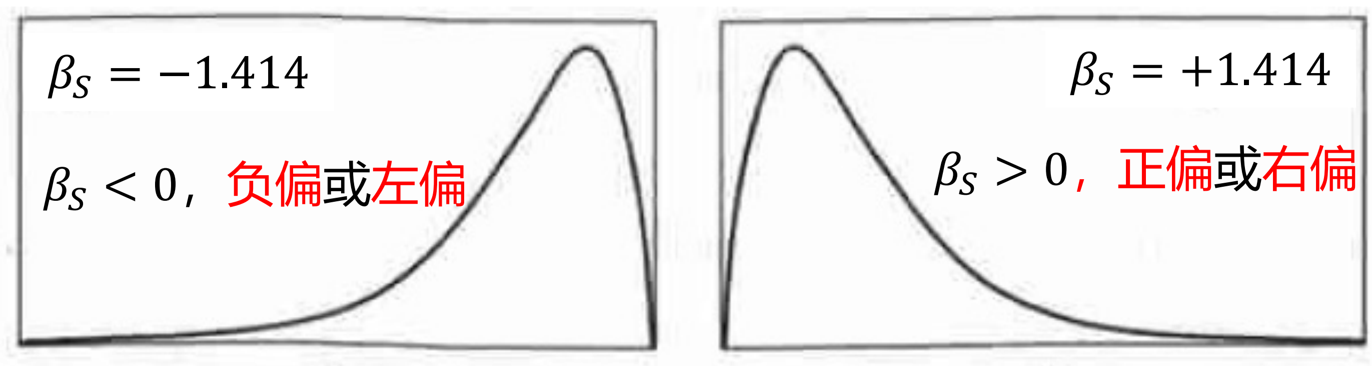
$\beta_S>0$称该分布为正偏或右偏；$\beta_S<0$称该分布为负偏或左偏。

- **峰度系数**：设随机变量X的前四阶矩存在，则 $\boldsymbol{\beta_k=\frac{\nu_4}{\nu_2^2}-3=\frac{E\left[\left(X-E(X)\right)^4\right]}{[\mathrm{Var}(X)]^2}-3}$ 被称为X的峰度系数，简称峰度。或记为γ_2 。

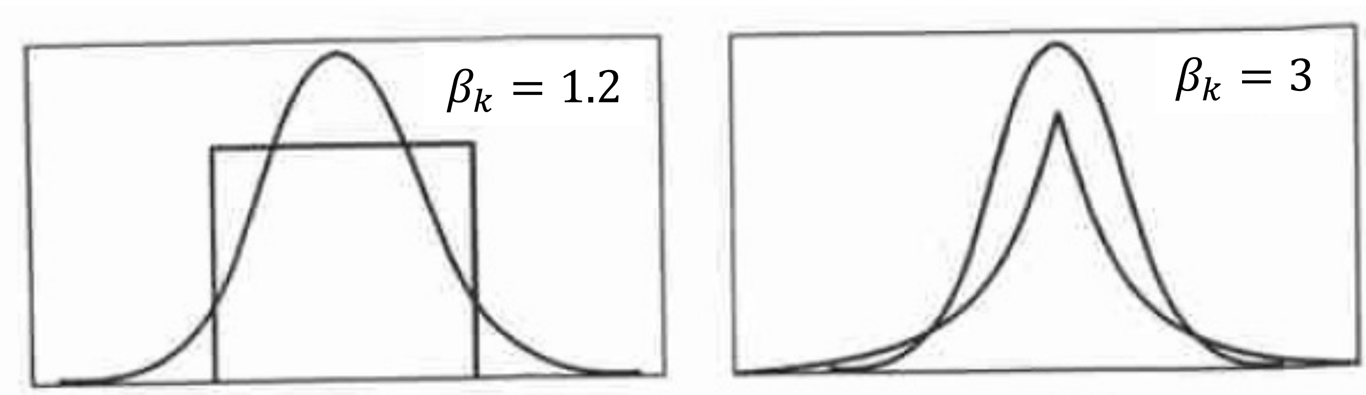
峰度描述分布的尖峭程度或尾部粗细程度，正态分布$\beta_k=0$。

偏度系数和峰度系数

- β_S描述分布偏离对称性的程度



- 峰度描述分布的尖峭程度或尾部粗细程度，正态分布 $\beta_k = 0$ 。



特征函数

设 X 是一个随机变量，则称 e^{itX} 的数学期望值，即

$$\varphi(t) = E\left[e^{itX}\right], \qquad -\infty < t < \infty$$

为随机变量 X 的特征函数。

对连续变量，
$$\varphi(t) = \int_{-\infty}^{\infty} e^{itX} f(x) \mathrm{d}x$$

这实际上是概率密度函数 $f(x)$ 的傅里叶变换。

对离散变量，
$$\varphi(t) = \sum_{k=0}^{\infty} e^{itx_k} p_k$$

由于$\left|e^{itX}\right| = 1$，随机变量 X 的特征函数总是存在。

特征函数的性质

1. $|\varphi(t)| \leq \varphi(0) = 1$

2. $\varphi(-t) = \overline{\varphi(t)}$

3. $\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at)$

4. 若随机变量X与Y独立，则

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$$

5. 若$E\left(X^l\right)$存在，则

$$\varphi^{(k)}(0) = i^k E[X^k], \quad 0 \leq k \leq l$$

因此，可以通过特征函数的导数求随机变量的矩：

$$E[X^k] = \frac{\varphi^{(k)}(0)}{i^k}$$

特征函数的性质（续）

1. 一致连续性　　　　2. 非负定性

3. 逆转公式：

设$F(x)$和$\varphi(t)$分别为随机变量X的分布函数和特征函数，则对$F(x)$的任意两个连续点$x_1 < x_2$，有

$$F(x_2) - F(x_1) = \lim_{\tau\to\infty} \frac{1}{2\pi} \int_{-\tau}^{\tau} \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi(t) \mathrm{d}t$$

4. 唯一性：

随机变量的分布函数由其特征函数唯一确定。

5. 若X为连续随机变量，其密度函数为$f(x)$，特征函数为$\varphi(t)$，若$\int_{-\infty}^{\infty} |\varphi(t)| \mathrm{d}t < \infty$，则

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) \mathrm{d}t$$

特征函数的用处

特征函数是处理概率论问题的有力工具，
其作用在于：

- 可将卷积运算化成乘法运算；
- 可将求各阶矩的积分运算化成微分运算；
- 可将求随机变量序列的极限分布化成一般的函数极限问题；
- ……….

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 $V_{ij} = \mathrm{cov}[x_i, x_j]$ （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 $V[y]$ ？

硬核计算法：

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，

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

为了得到 $V[y]$，我们需要计算 $E[y^2]$ 和 $E[y]$。

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

不确定度的传递(3)

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

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

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

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

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

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

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

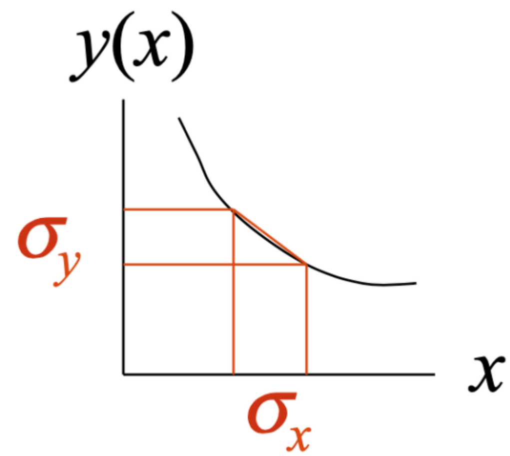
$$U_{kl} = \mathrm{cov}[y_k, y_l] \approx \sum_{i,j=1}^{n} \left[\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}$$

或者，写成矩阵形式 $U = AVA^T$，其中

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

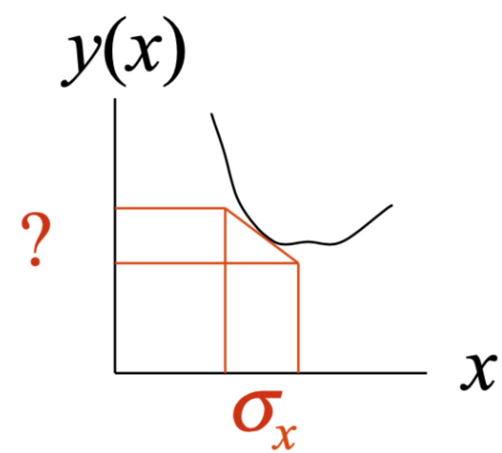
不确定度的传递(5)

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



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

不确定度传递：特例

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

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

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

$V[y] = 1^2 + 1^2 = 2$ $\longrightarrow$ $\sigma_y = 1.4$

2) $\rho = 1$

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

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？

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 $$U_{r\theta} = AV_{xy}A^T$$

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

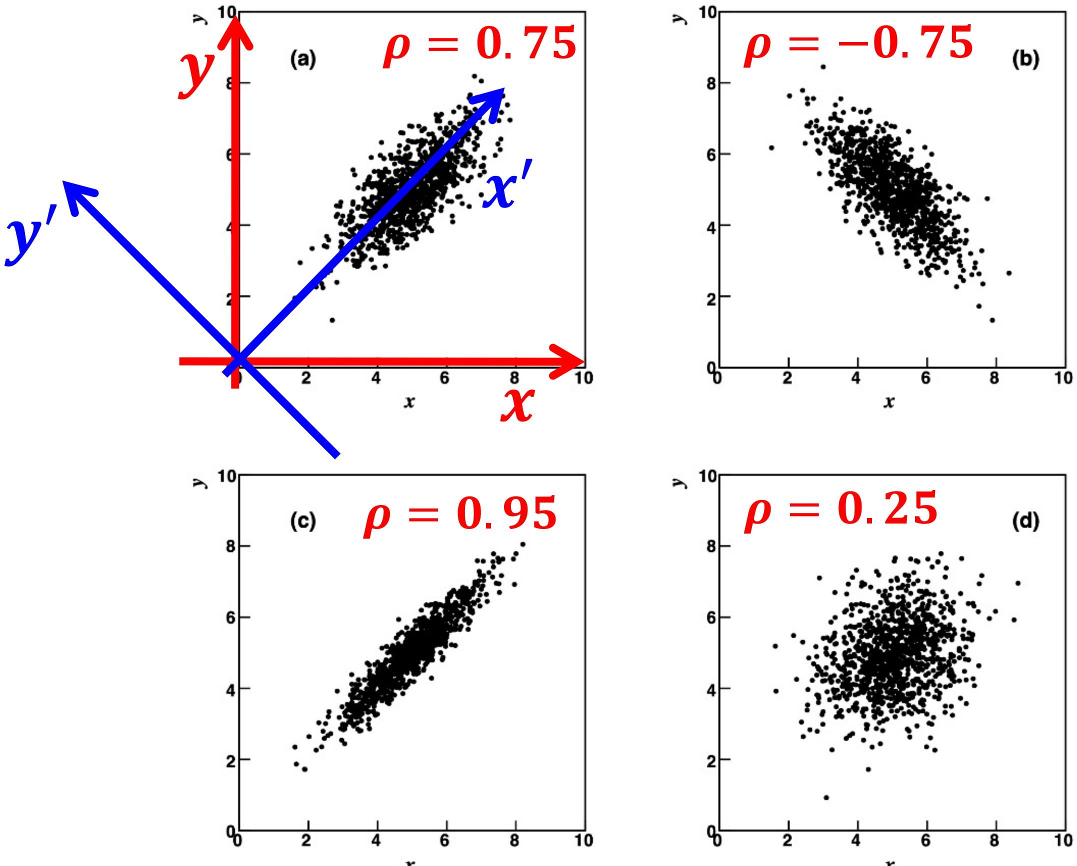
坐标变换后的协方差矩阵为

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

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 45^0，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 $\vec{x} = (x_1, \ldots, x_n)$ 及其协方差矩阵 $V_{ij} = \mathrm{cov}[x_i, x_j]$，可以证明，能够通过线性变换重新定义 n 个新的变量 $\vec{y} = (y_1, \ldots, y_n)$，使得对应的协方差矩阵 $U_{ij} = \mathrm{cov}[y_i, y_j]$非对角元为零。

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

对应的协方差矩阵为

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

非线性情况

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

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

$$A_{ij} = r_j^i, \qquad A_{ij}^T = r_i^j,$$

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

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

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

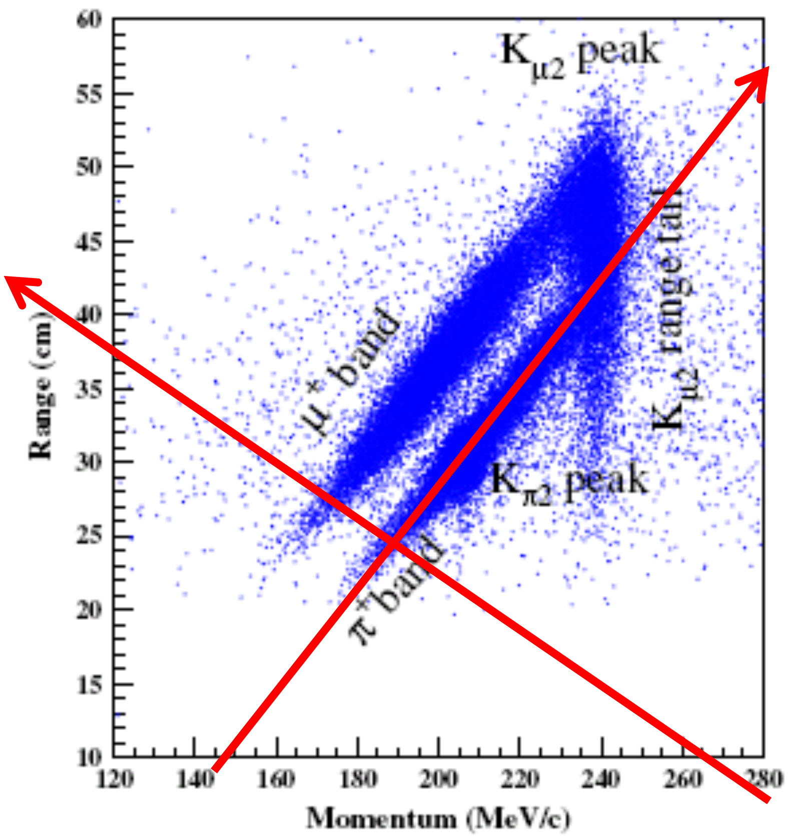
正交变换后，变量的方差由原协方差矩阵 V 的本征值给出。

对应于矢量的转动
不改变模的大小。

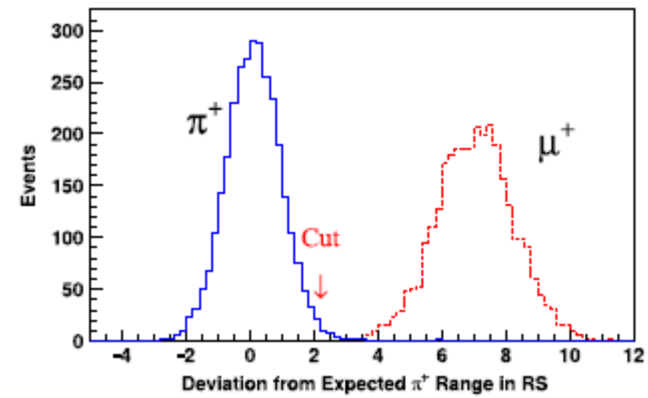
$$|\vec{y}|^2 = \vec{y}^T \vec{y} = \vec{x}^T A^T A \vec{x} = |\vec{x}|^2$$

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