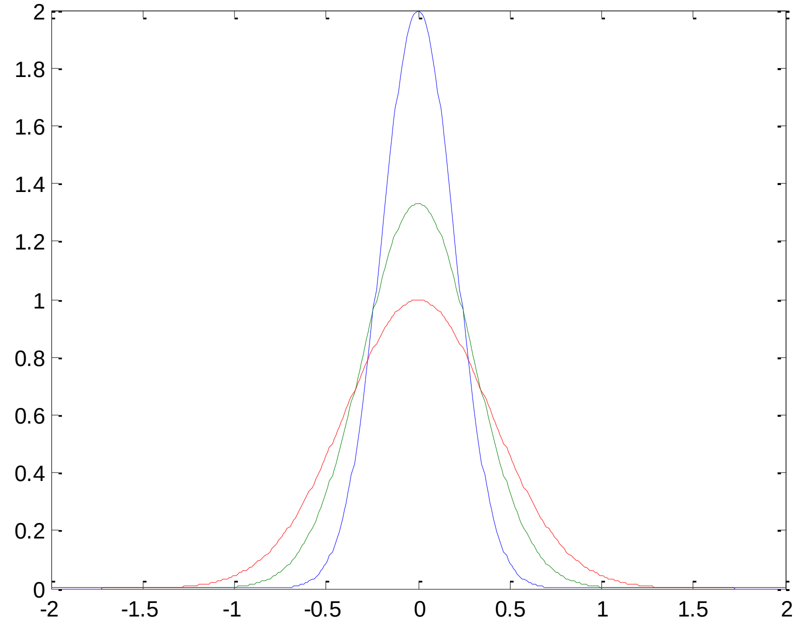
预备知识—概率论



概要

- 概要
- 基本概念
- 随机变量与概率分布
- 数字特征

概率论的起源：赌注分配问题

求助

帕斯卡

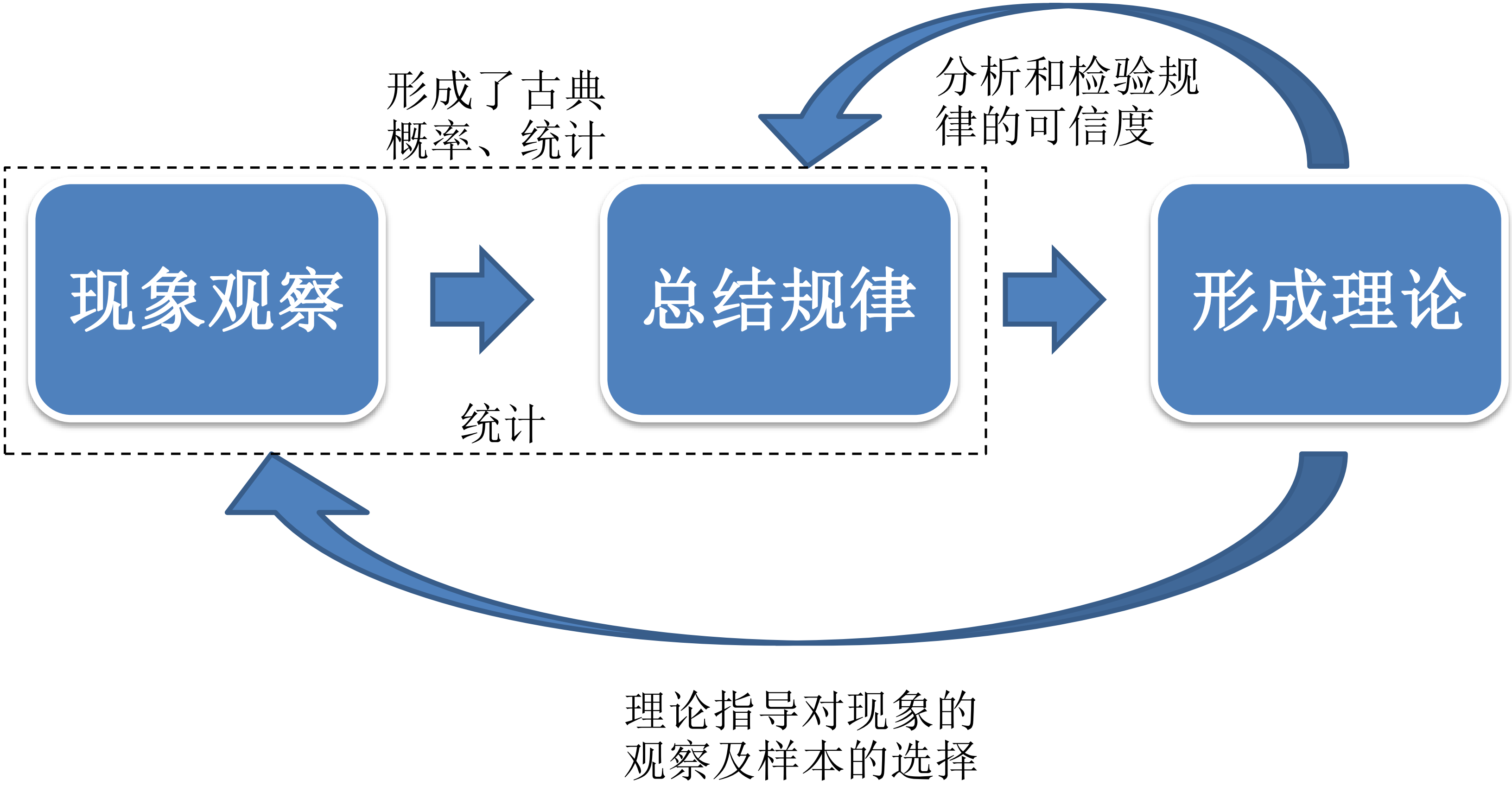
讨论

费马

17世纪中叶，法国贵族梅耳骑士与人掷骰子赌博，赌注分配有歧义

最后由瑞士数学家雅各布·伯努利彻底证明该问题，提出**大数定理**，出版《猜度术》

概率论起源与发展



要点：弄清统计与概率论之间的关系，避免混淆彼此概念

概率的几种解释

客观解释

- 又叫频率解释。概率描述了一个可以重复出现的事件的客观事实，用该事件发生的频率的极限来刻画

个人主义解释

- 也称为主观解释或贝叶斯解释，认为概率是某个人的偏好，它可以根据某人在带有不确定性结果的事件中表现出的行为来推算

必要性解释

- 又叫逻辑解释。此观点将概率视作命题与命题之间联系程度的度量，该联系程度是纯客观的，与人的作用无关，而是一种逻辑推理

随机试验

试验结果事先不能准确预言，三个特征

可以在相同条件下重复进行	每次试验结果不止一个，可预先知道试验所有可能结果	每次试验前不能确定那个结果会出现

样本空间

随机试验所有可能结果组成的集合，记为 Ω

事件

样本空间的子集A称为事件

集合运算

古典概率

- 随机试验中一切可能结果是**有限多个**；
- 每个结果出现的**可能性是相等的**；
- 则事件A发生的概率可表示为

$$P(A)=\frac{\text{事件A所包含的样本点个数}}{\text{样本空间中所含样本点个数}}$$

问题：如果随机试验所有可能的结果为无穷多个，该如何计算事件A的概率？

几何概率

- 计算**无穷个**基本事件的情形；
- 样本点具有**均匀分布**的性质；
- 设用L(Ω)作为区域Ω大小的量度，而区域Ω中任意可能出现的小区域A的量度用L(A)表示；
- 则事件A（或某一区域）发生的概率表示为

$$P(A)=\frac{L(A)}{L(\Omega)}$$

问题：如果事先不知道随机试验到底有多少可能的结果，又该如何计算事件A的概率？

统计概率

- 用于计算前两种随机概率概括不了的随机事件概率；
- 用事件的**频率**近似地去表达事件的**概率**；
- 若在同样的条件下，将随机试验独立的重复做n次，事件A出现了n_A次，则事件A的频率是

$$f_A = \frac{n_A}{n} \longrightarrow f_A \approx P(A)$$

- 当试验次数n增大时，其中大量的频率聚集在一个常数周围；
- 这个常数是客观存在的，反映了事件A出现可能性的大小，我们认为这个常数就是事件的**概率**。

(**大数定律对此结论的正确性给以了理论上的保证**)

概率论的发展

有限可能 → 无限可能 → 所有可能

古典概率　　几何概率　　统计概率

实现了从算术到数学的演进

概率空间

$X(\cdot)$

物理世界空间向抽象的数学空间的映射

要点：弄清概率的定义及其内涵，加深对随机变量含义的理解。了解如何确定一个随机变量的概率分布。

公理化定义概率

对于一个事件A∈样本空间Ω，赋予一个实数P，若满足：

1.0≤P(A) ≤1;

2.P(Ω)=1;

3.若A_1,A_2,……,A_k两两互斥，则

$$P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$$

我们称P(A)为事件A的一个概率。

概率空间

- 规定一个随机试验，所有样本点之集合构成样本空间Ω ，在样本空间中一个样本点或若干个样本点之**适当集合**F称为事件域，F中的每一个集合称为事件。若A ∈F，则P(A)就是事件A的概率，并称这三个实体的结合(Ω, F, P)为一个**概率空间**

条件概率

- 在事件B已发生这一条件下，事件A发生的概率。

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

全概率

- 若有N个互斥事件B_n(n=1,2,…,N)，它的并集等于整个样本空间，则

$$P(A)=\sum_{i=1}^{N}P(A\mid B_i)P(B_i)$$

贝叶斯公式

- 设事件A1，A2，…，An构成一个完备事件组，概率P(Ai)>0,i=1,2,……,n,对于任何一个事件B，若P(B)>0, 有

先验概率

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{\sum_{i=1}^{N} P(A_i)P(B \mid A_i)}$$

后验概率

事件A_1，A_2，……，A_n看作是导致事件B发生的“因素”，$P(A_i)$是在事件B已经出现这一信息得知前A_i出现的概率，通常称为**先验概率**。但是在试验中事件B的出现，有助于对导致事件B出现的各种“因素”发生的概率作进一步探讨，公式给出的$P(A_i \mid B)$是在经过试验获得事件B已经发生这个信息之后，事件A_i发生的概率，称为**后验概率**。后验概率依赖于试验中得到的新信息的具体情况（比如事件B发生还是不发生），并且给出在获得新信息之后，导致B发生的各种因素A_i发生情况的新知识，因此贝叶斯公式又称为**后验概率公式**或**逆概率公式**

已知某种疾病的发病率是0.001。现有一种试剂可以检验患者是否得病，准确率为0.99，即在患者确实得病的情况下，它有99%的可能呈阳性。它的误报率是5%，即在患者没有得病的情况下，它有5%的可能呈现阳性。现有一个病人的检验结果为阳性，请问他确实得病的可能性有多大？

假定A表示得病，那么P(A)为0.001。这就是“先验概率”，即没有做试验之前，预计的发病率或根据历史病例统计的发病率。再假定B表示测试结果为阳性。问题就是要计算P(A|B)。这就是"后验概率"，即做了试验以后，对发病率的估计。

$$P(A|B)=\frac{P(A)P(B|A)}{P(B|A)P(A)+P(B|\bar{A})P(\bar{A})}=\frac{0.001*0.99}{0.99*0.001+0.05*0.999}$$

$= 0.019$

问题：哪种试剂更好？

该如何下诊断结论？

另一种试剂的误报率为1%，P(A|B2)=0.09

续

	试剂1（B1）	试剂2（B2）
第二次检测	$P(\hat{A}\|B1) = \frac{P(\hat{A})P(B1\|\hat{A})}{P(B1\|\hat{A})P(\hat{A}) + P\left(B1\middle\|\bar{\hat{A}}\right)P(\bar{\hat{A}})} = \frac{0.019*0.99}{0.019*0.99+0.05*0.981} = 0.277$	$P(\hat{A}\|B2) = \frac{P(\hat{A})P(B2\|\hat{A})}{P(B2\|\hat{A})P(\hat{A}) + P\left(B2\middle\|\bar{\hat{A}}\right)P(\bar{\hat{A}})} = \frac{0.09*0.99}{0.09*0.99+0.01*0.91} = 0.907$
更新后的 $P(\hat{A})$	$P(\hat{A}) = 0.277$	$P(\hat{A}) = 0.907$

可以相信，如果第三次测试仍然采用试剂2且检测结果为阳性，则患病的概率必将趋近于100%

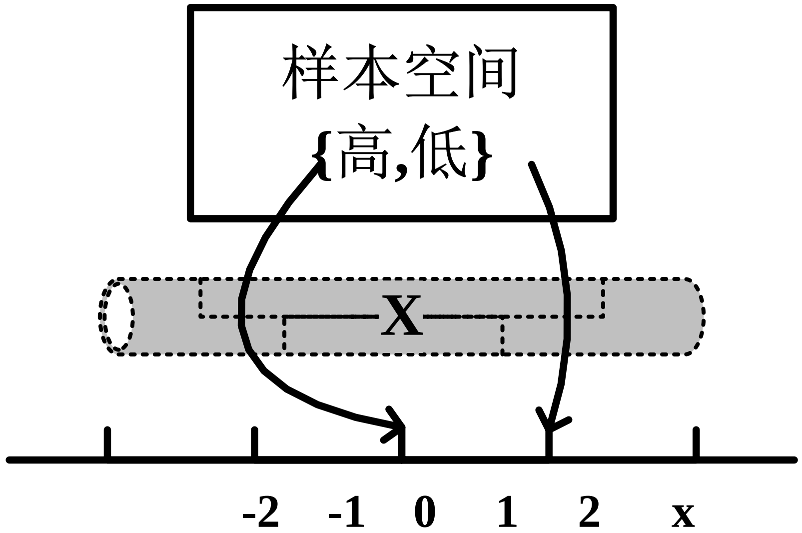
但是，假设第三次仍然采用试剂1，结果仍为阳性，则患病概率约为88%，低于试剂2第二次的概率。

独立事件

- 如果A、B为两个统计独立的事件，则

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

随机变量



定义:

设(Ω, F, P)是概率空间，X=X(e)是定义在Ω上的实函数，如果对任意实数x，{e:X(e) ≤x} ∈F，则称X(e)是F上的随机变量。

随机变量

事件 → 随机变量

离散型随机变量：

只取有限个数值或可列无穷多个值。

连续型随机变量：

从原样本空间到新样本空间的映射是某一个范围，是一段（或几段）实线（也可能是整个坐标轴），随机变量可以取值于某一区间中的任一数。

概率分布函数

分布函数（一个描述随机变量取值的概率分布情况的统一方法）

$$F(x) = P\{e : X(e) \le x\}, \ -\infty < x < \infty$$

性质：

1. F(x)是非降函数；

2. 0≤F(x) ≤1；

3. P{x_1<X ≤ x_2}=F(x_2)-F(x_1)；

4. F(x)是右连续，即F(x+0)=F(x)。

概率模型

通过对物理世界的抽象，建立了大量的模型，其中最常用的概率模型包括：

- 概率模型
  - 离散
    - 0-1分布
    - 二项分布
    - 泊松分布
  - 连续
    - 均匀分布
    - 指数分布
    - 正态分布

要点：理解各类主要概率分布模型的定义、概率分布函数、特性以及应用场合。了解如何从这些基本的概率模型演化到其他更广泛的场景（**随机变量的函数**）

概率分布

离散型随机变量的概率分布用分布列描述

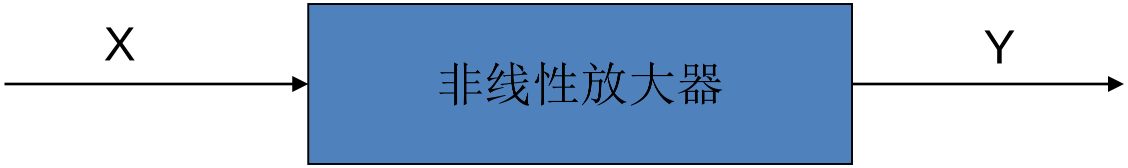
0—1分布	$P(X=1)=p, P(X=0)=q$
二项分布	$P(X=k)=C_n^k p^k q^{n-k}$
泊松分布	$P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$

连续型随机变量的概率分布用概率密度描述

均匀分布	$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b\\ 0, & \text{其它}\end{cases}$
正态分布	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}$
指数分布	$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$

随机变量函数的分布

在给定某任意的随机变量X，以及它的概率分布函数$F_X(x)$，希望进一步求出给定的随机变量的某些可测函数（如Y=g(X)）的概率分布函数。



Y的概率分布函数公式为

$$F_Y(y)=P(x:g(x)\le y, x\in\Omega_X)$$

如果上式右端概率的导数对于y处处存在，那么这个导数就给出了随机变量Y的概率密度

$$f_Y(y)=\frac{d}{dy}P(\{x:g(x)\le y, x\in\Omega_X\})$$

n维随机变量及其分布函数

设（Ω，F，P）是概率空间，**X**=**X**(e)＝$(X_1(e),\ldots,X_n(e))$是定义在Ω上的n维空间R^n中取值的向量函数。如果对于任意$X=(X_1,\ldots,X_n)\in R^n$，$\{e:X_1(e)\le x_1,\ldots,X_n(e)\le x_n\}\in F$，则称**X**=**X**(e)为**n维随机变量**。称

$$F(x)=F(x_1,\ldots,x_n)=P(e:X_1(e)\le x_1,\ldots,X_n(e)\le x_n)$$

为$X=(X_1,X_2,\ldots,X_n)$的**联合分布函数**

边际分布

若二维联合分布函数中有一个变元**趋于无穷**，则其极限函数便是一维分布函数，对于这种特殊性质，我们称其为**边际分布**。

对于任意两个随机变量X,Y，其联合分布函数为$F_{XY}(x,y)$，则

$$F_1(x) = F(x,\infty)$$
$$F_2(y) = F(\infty,y)$$

分别称$F_1(x)$和$F_2(y)$为$F_{XY}(x,y)$关于X和关于Y的**边际分布函数**。

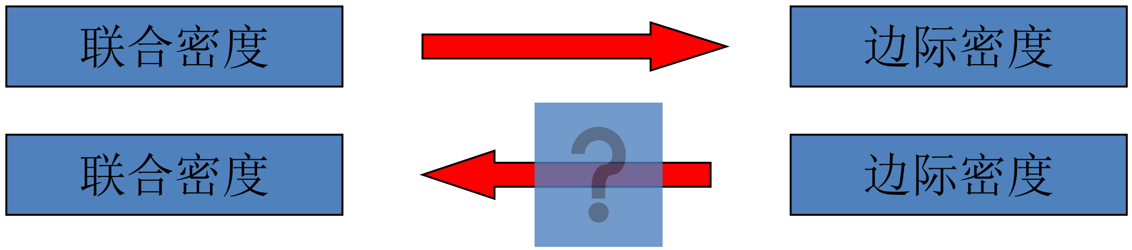
离散型随机变量（X，Y）边际分布函数计算如下

$$F_1(x) = F(x,\infty) = \lim_{y\to\infty} F_{XY}(x,y) = P(X \le x, Y < \infty) = P(X \le x)$$

连续型随机变量（X，Y）边际分布函数计算如下

$$F_2(y) = F\left(\infty, y\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} f(u,v)dudv$$

相互独立的随机变量



设X，Y是两个随机变量，若对任意实数x，y有

$$P(X \le x, Y \le y) = P((X \le x) \cap (Y \le y)) = P(X \le x)P(Y \le y)$$

则称X，Y为**相互独立的**随机变量。

若X，Y为相互独立随机变量，则有

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

条件分布

条件概率

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

条件分布函数

$$F_{X|B}(x|B)=P(X\le x|B)=\frac{P(X\le x\cap B)}{P(B)}$$

$$F_{X|Y}(x|Y=y)=\frac{\int_{-\infty}^{x}f_{XY}(u,y)du}{f_Y(y)}$$

两边对x微分

$$f_{X|Y}(x|y)=\frac{f_{XY}(x,y)}{f_Y(y)}$$

$$F_{X|Y}(x|Y=y)=\int_{-\infty}^{x}f_{X|Y}(u|y)du$$

随机变量的数字特征

- 统计平均与随机变量的数学期望
- 随机变量函数的期望值
- 方差
- 协方差
- 相关系数
- 独立与不相关

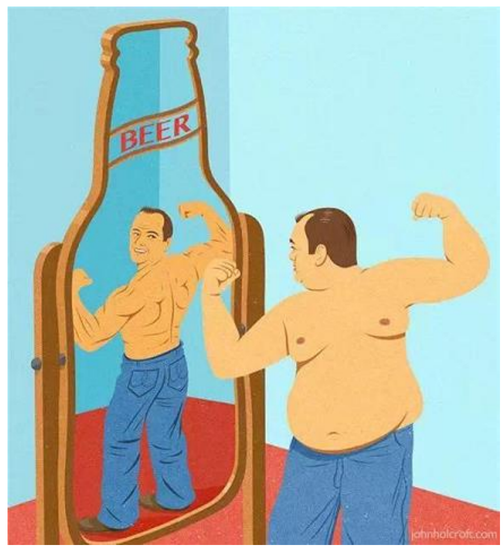
随机变量的数字特征

大数据

统计

数据工程
采样

理想中的
概率模型



现实中的
随机对象

要点：正确理解各个数字特征的定义，以及与概率模型之间的关系

为了“拉近”现实与理论的“距离”，定义了随机变量的数字特征，简化了概率问题求解

统计平均与数学期望

设离散随机变量X，它可能取4个值x_1，x_2，x_3，x_4，做试验 n 次，计算X的算术平均可得：

$$\overline{X} = \frac{1}{n}(x_1 n_1 + x_2 n_2 + x_3 n_3 + x_4 n_4) = \frac{1}{n}\sum_{k=1}^{4} x_k n_k = \sum_{k=1}^{4} x_k \left(\frac{n_k}{n}\right)$$

P(X=x_k)

$$\Rightarrow \quad \overline{X} = E(X) = \sum_{k=1}^{\infty} x_k P(X = x_k)$$

对于离散型随机变量可以用脉冲函数来表示其概率密度

$$f_X(x) = \sum_{k=1}^{\infty} P(X = x_k)\delta(x - x_k)$$

冲激函数

随机变量数学期望定义

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$$

随机变量函数的期望值

已知随机变量X的数学期望值，求随机变量函数Y=g(X)的数学期望，

$$E(Y)=E(g(X))=\int_{-\infty}^{\infty} y f_Y(y)dy=\int_{-\infty}^{\infty} g(x) f_X(x)dx$$

对于多维随机变量

$$\mathbf{X}=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \qquad \mathbf{Y}=g(\mathbf{X})$$

随机向量函数的数学期望

$$\boldsymbol{E}(\mathbf{Y})=\int_{-\infty}^{\infty} g(\mathbf{X}) f_X(x)dx$$

N维随机变量的数学期望

设$X_1,X_2, \ldots,X_n$为随机变量，求随机变量函数$Y=a_1X_1+a_2X_2+\ldots+a_nX_n$的数学期望。

$$
\begin{aligned}
E(Y) &= E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) \\
&= E(a_1X_1) + E(a_2X_2) + \cdots + E(a_nX_n) \\
&= a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)
\end{aligned}
$$

线性组合的随机变量的期望

已知随机变量X_1和X_2，求随机变量函数$Y=aX_1+bX_2$的数学期望

$$E(Y)=\int_{-\infty}^{\infty}\int(ax_1+bx_2)f_{X_1X_2}(x_1,x_2)dx_1dx_2$$

$$=a\int_{-\infty}^{\infty}\int x_1 f_{X_1X_2}(x_1,x_2)dx_1dx_2+b\int_{-\infty}^{\infty}\int x_2 f_{X_1X_2}(x_1,x_2)dx_1dx_2$$

$$=aE(X_1)+bE(X_2)$$

加权和的期望等于加权期望的和

求数学期望是线性运算

数学期望的线性运算不受独立条件限制

相互独立的随机变量的期望

已知随机变量X_1和X_2，求随机变量函数$Y=g_1(X_1)g_2(X_2)$的数学期望

$$E[Y]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x_1)g_2(x_2)f_{x_1x_2}(x_1,x_2)dx_1dx_2$$

假设两个随机变量X_1和X_2相互独立，则有

$$f_{x_1x_2}(x_1,x_2)=f_{x_1}(x_1)f_{x_1}(x_2)$$

因此，有

$$E[Y]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x_1)g_2(x_2)f_{x_1}(x_1)f_{x_2}(x_2)dx_1dx_2$$

$$=\int_{-\infty}^{\infty} g_1(x_1)f_{x_1}(x_1)dx_1\int_{-\infty}^{\infty} g_2(x_2)f_{x_2}(x_2)dx_2$$

$$=E[g_1(X_1)]E[g_1(X_2)]$$

K阶原点矩
k阶中心矩

随机变量X，若$E[X^k]<\infty$，称$E[X^k]$为X的**k阶原点矩**。

离散随机变量　　连续随机变量

$$E[X^k]=\sum_{i=1}^{n}x_i^k P(X=x_i)=\int_{-\infty}^{\infty}x^k f_X(x)dx$$

又若E[X]存在，且$E[|X-E[X]|^k]<\infty$，称

$$E[(X-E[X])^k]$$

为X的**k阶中心矩**。

离散随机变量　　连续随机变量

$$E[(X-E[X])^k]=\sum_{i=1}^{n}(x_i-E[X])^k P(X=x_i)=\int_{-\infty}^{\infty}(x-E[X])^k f_X(x)dx$$

K阶原点矩
k阶中心矩

一阶原点矩就是随机变量的数学期望，

$$EX \triangleq \int_{-\infty}^{\infty} x dF(x)$$

数学期望大致的描述了概率分布的中心。

二阶中心矩就是随机变量的方差，

$$DX \overset{def}{=} E(X - EX)^2$$

方差反映随机变量取值的离散程度。

$$\left.\begin{array}{l}\text{0—1分布}\\ \text{泊松分布}\\ \text{正态分布}\end{array}\right\}\text{数学期望和方差(见page3，表1-1)}$$

协方差

中心化的两个随机变量X-E[X]，Y-E[Y]的互相关矩称为随机变量X和Y的协方差，

$$Cov(X,Y) = E[(X - EX)(Y - EY)]$$
$$= E(XY) - E(X)E(Y)$$

协方差是描述随机现象中，随机变量X和Y概率相关的程度。

引入一个描述两个随机变量相关程度的系数

$$\rho_{XY} \overset{def}{=} \frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}}$$

ρ_{XY}称为归一化的协方差系数或相关系数。

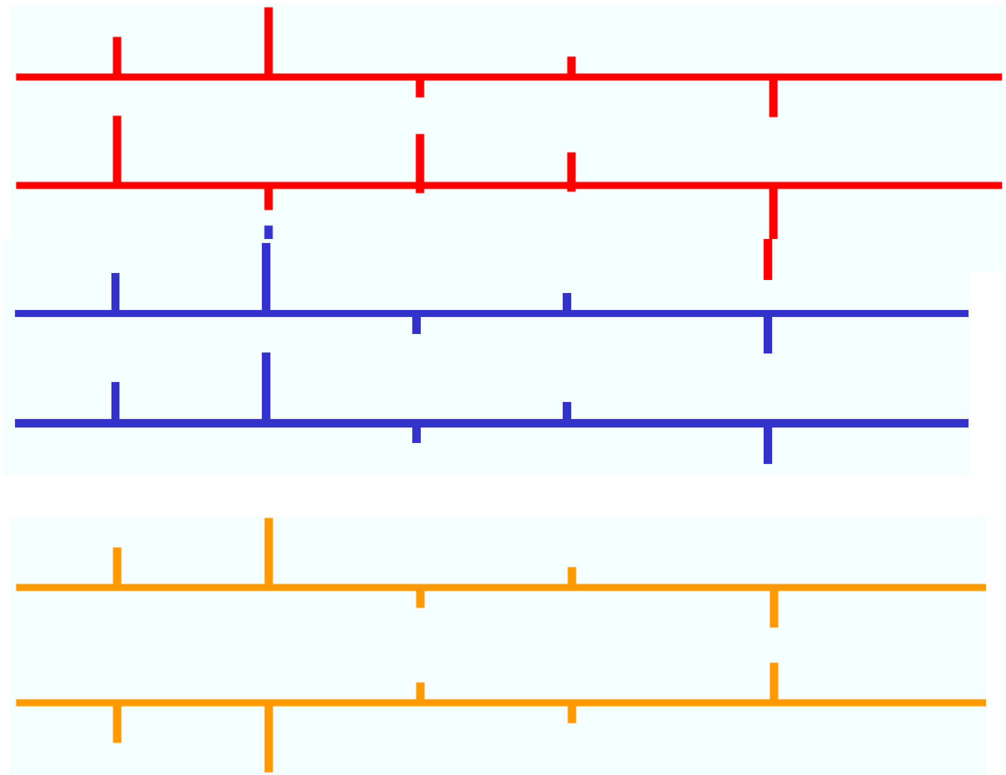
$$-1 \le \rho_{XY} \le 1$$

协方差

$$\rho \approx 0 \begin{cases} x \\ y \end{cases}$$

$$\rho \to 1 \begin{cases} x \\ y \end{cases}$$

$$\rho \to -1 \begin{cases} x \\ y \end{cases}$$



若ρ_{XY}＝0，则称随机变量X和Y不相关。

若两个随机变量X和Y的联合矩满足 $E[X^iY^j] = E[X^i]E[Y^j]$

则称随机变量X和Y统计独立

不相关与独立

统计独立 → 不相关

$$Cov(X,Y) = E[(X - EX)(Y - EY)]$$
$$= E[XY] - E[X]E[Y] = 0$$

统计独立 ← ? 不相关

特征函数

定义：设随机变量的概率密度为$f(x)$，称其傅氏逆变换

$$g(t)=E(e^{itx})=\int_{-\infty}^{\infty}e^{itx}f(x)dx$$

为X的特征函数。

而$f(x)=\dfrac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}g(t)dt$

对于离散随机变量，

$$g(t)=\sum_{k=1}^{\infty}e^{itx_k}p_k$$

性质：（介绍）

特征函数的性质

1) $g(0)=1, |g(t)|\le 1,\ g(-t)=\overline{g(t)}.$

2) $g(t)$在$(-\infty, +\infty)$上一致连续.

3)若随机变量X的n阶矩EX^n存在，则X的特征函数$g(t)$可微分n次，且当$k\le n$时，有$g^{(k)}(0)=i^k EX^k$.

4) $g(t)$是非负定函数，即对任意正整数n及任意实数$t_1,t_2,\cdots,t_n$和复数$z_1,z_2,\cdots,z_n$，有

$$\sum_{k,l=1}^{n} g(t_k-t_l)z_k\overline{z_l}\ge 0.$$

5)若$X_1,X_2,\cdots,X_n$是相互独立的随机变量，则$X=X_1+X_2+\cdots+X_n$的特征函数

$$g(t)=g_1(t)g_2(t)\cdots g_n(t)$$

其中$g_i(t)$是随机变量X_i的特征函数，$i=1,2,\cdots,n$.

6)随机变量的分布函数由其特征函数惟一确定.

例1.1

例1.1 设X服从B(n,p)，求X的特征函数g(t)及EX，EX²，DX。

根据特征函数的定义：

$$g(t) = E[e^{itx}]$$

$\because$ X服从二项分布$B(n, p)$

$\therefore$ 其分布列为：$p(X = k) = C_n^k p^k q^{n-k}, q = 1 - p, k = 0,1, \dots, n$

因此，$g(t) = \sum_{k=0}^{n} e^{itk} \bullet C_n^k p^k q^{n-k}$

$$= \sum_{k=0}^{n} C_n^k (pe^{it})^k \bullet q^{n-k}$$

$$= (pe^{it} + q)^n$$

$$EX = -ig'(0) = -i \frac{d}{dt}(pe^{it} + q)^n \Big|_{t=0}$$

$$= -i \cdot n(pe^{it} + q)^{n-1} \cdot pie^{it} \Big|_{t=0} = np$$

续例1.1

$$E[X^2] = (-i)^2 g''(0) = -g''(0) = -\frac{d^2}{dt^2}(pe^{it} + q)^n\Big|_{t=0}$$

$$= -\frac{d}{dt} inp(pe^{it} + q)^{n-1} \cdot e^{it}\Big|_{t=0}$$

$$= -inp(n-1)(pe^{it} + q)^{n-2} \cdot e^{it} \cdot pie^{it}\Big|_{t=0}$$

$$- inp \cdot (pe^{it} + q)^{n-1} \cdot ie^{it}\Big|_{t=0}$$

$$= n(n-1)p^2 + np = npq + n^2p^2$$

$$DX = E[X^2] - (EX)^2 = npq + n^2p^2 - n^2p^2 = npq$$

母函数

定义：设X是非负整型随机变量，分布列

$$p_k = P(X = k),\ k = 0,1,\ldots$$

称

$$P(s) \overset{def}{=} E(s^X) = \sum_{k=0}^{\infty} p_k s^k$$

为X的母函数。

性质：（介绍）

母函数的性质

1)非负整数值随机变量的分布列由其母函数惟一确定.

2)设$P(s)$是X的母函数，若EX存在，则

$$EX = P'(1),$$

若DX存在，则

$$DX = P''(1) + P'(1) - [P'(1)]^2.$$

3)独立随机变量之和的母函数等于母函数之积.

4)若$X_1, X_2, \cdots, X_n$是相互独立且同分布的非负整数值随机变量，N是与$X_1, X_2, \cdots$独立非负整数值随机变量，则$Y = \sum_{k=1}^{N} X_k$的母函数

$$H(s) = G(P(s)),$$

其中$G(s)$、$P(s)$分别是N、X_1的母函数.

小节

概率论中的基本概念——随机试验、样本空间、事件、概率、概率空间、条件概率、全概率。

随机变量及分布函数——随机变量、分布函数、随机变量函数的分布、n维随机变量、边际分布、条件分布。

随机变量的数字特征——统计平均、数学期望、方差、协方差、相关系数、相关性和统计独立。

作业

- 复习概率论与数理统计方面的知识。