数理统计

上海财经大学　统计与管理学院

Contents

第一章 概率论基础

一、随机事件与概率

1、概率的公理化定义

设Ω为一个样本空间，$\mathcal{F}$为Ω的某些子集组成的一个事件域。如果对任一事件 $A \in \mathcal{F}$，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、**正则性公理**：$P(\Omega) = 1$，

3、**可列可加性公理**：若$A_1, A_2, \ldots, A_n, \ldots$互不相容，则：

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

一、随机事件与概率

2、概率的性质

(1) $\boldsymbol{P(\varnothing)=0}$.

证明：$\Omega=\Omega\cup\varnothing$，根据可列可加性公理可得：

$$P(\Omega)=P(\Omega)+P(\varnothing)$$

可得：

$$P(\varnothing)=0$$

一、随机事件与概率

(2)(**有限可加性**)若有限个事件 $A_1, A_2, \dots, A_n$ 互不相容，则有：

$$\boldsymbol{P}\left(\bigcup_{i=1}^{n} \boldsymbol{A}_i\right) = \sum_{i=1}^{n} \boldsymbol{P}(\boldsymbol{A}_i)$$

证明：对于$A_1, A_2, \dots, A_n, \varnothing, \varnothing, \dots$应用可列可加性，

可得：$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1 \cup A_2 \cup \dots \cup A_n \cup \varnothing \cup \varnothing \cup \dots)$

$= P(A_1) + P(A_2) + \cdots + P(A_n) + P(\varnothing) + P(\varnothing) + \dots = P(A_1) + \dots + P(A_n)$

(3) 对任一事件A，有

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

证明：$\Omega = A \cup \bar{A}$，根据正则性和有限可加性，

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

一、随机事件与概率

(4) 若 $A \supset B$,则

$$\boldsymbol{P(A-B)=P(A)-P(B)}$$

证明：因为 $A \supset B$， $A = B \cup (A-B)$,两者互不相容,
由有限可加性可得:

$$P(A)=P(B)+P(A-B)$$

(5) 对任意两个事件A, B有:

$$\boldsymbol{P(A-B)=P(A)-P(AB)}$$

证明：$A-B=A-AB$,且$AB \subset A$，根据（4）可得

$$P(A-B)=P(A-AB)=P(A)-P（AB)$$

一、随机事件与概率

(6)(**加法公式**)对于任意两个事件A,B,有

$$\boldsymbol{P(A\cup B)=P(A)+P(B)-P(AB)}$$

证明: $A\cup B=A\cup(B-AB)$,

且A与$B-AB$互不相容,根据有限可加性和(5)得

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

一、随机事件与概率

3.古典概型常用公式

(1)条件概率

设A，B是两个事件，若$P(B) > 0$，则称

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

为“在事件B发生下事件A发生的条件概率”，简称条件概率。它满足概率的三条公理。

一、随机事件与概率

(2)乘法公式

i) 二元

若$P(B)>0$，则

$$P(AB)=P(B)P(A|B)$$

ii) n元

若$P(A_1A_2\cdots A_n)>0$，则

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}).$$

一、随机事件与概率

(3)全概率公式

设$B_1, B_2, \cdots B_n$ 互不相容，且$\bigcup_{i=1}^{n} B_i = \Omega$，如果$P(B_i) > 0$，$i = 1,2, \cdots, n$，则对任一事件$A$有

$$P(A) = \sum_{i=1}^{n} P(B_i) P(A|B_i).$$

全概率公式提供了计算复杂事件概率的一条有效途径。

一、随机事件与概率

(4)贝叶斯公式

设$B_1, B_2, \cdots B_n$互不相容，且$\cup_{i=1}^{n} B_i = \Omega$，如果$P(A) > 0$，$P(B_i) > 0$，$i = 1,2, \dots \dots, n$，则

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}, i = 1,2, \cdots n$$

在贝叶斯公式中，诸$P(B_i)$成为B_i的先验（试验以前）概率，而诸$P(B_i|A)$称为B_i的后验（试验以后）概率，它表示在“事件A发生”这个新信息后，对B_i 的概率作出的修正。

注：贝叶斯分析具有广泛的应用场景，比如垃圾邮件识别、疾病原因诊断等。

一、随机事件与概率

❖ **例1** （2012数1，3）设A，B，C是随机事件，A与C互不相容，$P(AB)=\frac{1}{2}$，$P(C)=\frac{1}{3}$，求$P(AB|\bar{C})$。

❖ **解:** 由A与C互不相容得$P(AC)=0$，从而$P(ABC)=0$.于是

$$P(AB|\bar{C})=\frac{P(AB\bar{C})}{P(\bar{C})}=\frac{P(AB)-P(ABC)}{1-P(C)}=\frac{\frac{1}{2}-0}{1-\frac{1}{3}}=\frac{3}{4}.$$

一、随机事件与概率

- **例2** 一盒晶体管中有8只合格品、2只不合格品。从中不放回地一只一只取出，试求第二次取出合格品的概率。
- **解：** 记事件A_i为“第i次取出合格品”，$i=1,2$。用全概率公式

$$P(A_2)=P(A_1)P(A_2|A_1)+P(\bar{A}_1)P(A_2|\bar{A}_1)$$

$$=\frac{8}{10}\times\frac{7}{9}+\frac{2}{10}\times\frac{8}{9}=\frac{4}{5}.$$

注：本例可以推广如下：袋中a个红球，b个黑球，不放回一只只取出，求第k次取出红球概率。小球模型也适用于彩票中奖问题分析。

一、随机事件与概率

- **例3** 甲、乙两人对同一目标进行射击，命中率分别为0.6,0.5。试在下列两种情况下，分别求“已知目标被命中，则甲射中”的概率：

(1)在甲、乙两人中随机地挑选一人，由他射击一次；

(2)甲、乙两人独立地各射击一次.

- **解:**（1）记事件A_1为“选中甲”，A_2为“选中乙”，B为“命中目标”。

$$P(A_1|B)=\frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{2}P(B|A_i)P(A_i)}=\frac{0.6\times0.5}{0.6\times0.5+0.5\times0.5}=\frac{6}{11}.$$

一、随机事件与概率

(2)记事件A为“甲射中”，B为“乙射中”，两人独立射击，故有$P(AB)=P(A)P(B)$，则所求概率为

$$P(A|A\cup B)=\frac{P\big(A\cap(A\cup B)\big)}{P(A\cup B)}=\frac{P(A)}{P(A)+P(B)-P(A)P(B)}=\frac{0.6}{0.6+0.5-0.6\times0.5}=\frac{3}{4}.$$

一、随机事件与概率

(5)独立性

i)两个事件的独立性　如果$P(AB)=P(A)P(B)$，则称事件A与B独立。否则称A与B不独立或相依。

ii)多个事件的独立性　设有n个事件$A_1, A_2, \cdots A_n$,如果对任意的$1 \le i < j < k < \cdots \ldots \le n$，以下等式成立

$$\begin{cases} P(A_iA_j) = P(A_i)P(A_j), \\ P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k), \\ \cdots \\ P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n), \end{cases}$$

则称此n个事件$A_1, A_2, \cdots A_n$相互独立。

一、随机事件与概率

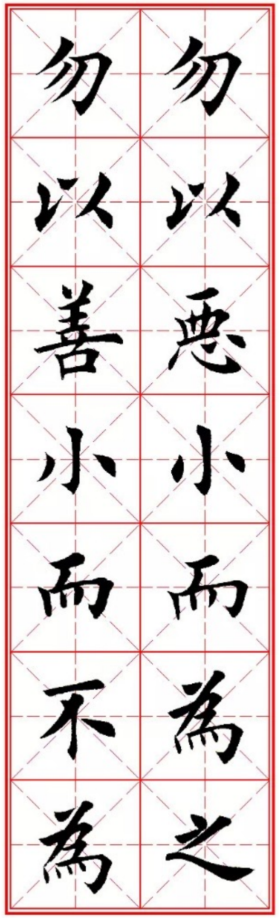
iii) n重独立重复试验　假如一个试验重复进行n次，并各次试验间相互独立，则称其为n重独立重复试验。假如一个试验只可能有两个结果：A与$\bar{A}$，则称其为**伯努利试验**。假如一个伯努利试验重复进行 n 次，并各次试验间相互独立，则称其为n重伯努利试验。

- **例4** 证明独立重复的小概率事件是必然发生的。

- **证明：** 设小概率事件A发生的概率为ε，即$P(A)=\varepsilon$，$\varepsilon>0$，则重复n次都不发生概率为$P(\bar{A})^n=(1-\varepsilon)^n$，则发生概率为$P=1-(1-\varepsilon)^n\xrightarrow{n\to\infty}1$，即必发生。

注： 吃路边摊和乱穿马路的人们要注意了！思考：举例可以发生，但发生的概率为0的事件.

思政案例：勿以恶小而为之，勿以善小而不为



蜀汉昭烈帝刘备遗诏出自《三国志·蜀书·先主传》

意思是不要因为是件较小的坏事就去做，不要因为是件较小的善事就毫不关心

思政案例：勿以恶小而为之，勿以善小而不为

实例：（小概率事件）

- 小概率事件是一个事件的发生概率很小，那么它在一次试验中是几乎不可能发生的，但在多次重复试验中是必然发生的。

- 比如现行某疾病诊断方法有p误诊概率，那么在以一个人口基数为n的样本中，至少有一个误诊的概率是多少？

思政案例：勿以恶小而为之，勿以善小而不为

n\p	0.01	0.001	10^(-6)	10^(-9)
100	0.6339677	0.09520785	9.999505e-05	9.999999e-08
1000	0.9999568	0.6323046	9.995007e-04	9.999995e-07
10^6	1	1	0.6321207	9.995001e-04
10^9	1	1	1	0.6321205
10^10	1	1	1	0.9999546

二、随机变量概念

1.随机变量(Random Variable, RV)

定义在样本空间Ω上的实值函数$X = X(\omega)$称为随机变量。

i) 仅取有限个或可列个值的随机变量称为离散随机变量;

ii) 取值充满某个区间(a, b)的随机变量称为连续随机变量，这里a可为$-\infty$，b可为$+\infty$.

二、随机变量概念

2.分布函数(Cumulative Distribution Function,CDF)

设X是一个随机变量，对任意实数x，称$F(x)=P(X\le x)$为X的分布函数，记为$X\ \sim F(x)$.分布函数具有如下三条基本性质：

i) 单调性　$F(x)$是单调非减函数，即对任意的$x_1<x_2$,有

$$F(x_1)\le F(x_2).$$

ii) 有界性　对任意的x，有$0\le F(x)\ \le 1$，且

$$F(-\infty)=\lim_{x\to-\infty}F(x)=0,F(+\infty)=\lim_{x\to+\infty}F(x)=1.$$

iii) 右连续性　$F(x)$是x的右连续函数，即对任意的x_0，有

$$\lim_{x\to x_0^+}F(x)=F(x_0)，即F(x_0+0)=F(x_0).$$

反之，可以证明：具有上述三条性质的函数$F(x)$一定是某一个随机变量的分布函数。

注：分布函数有两种定义$P(X\le x)$与$P(X<x)$，区别在于性质分别为右连续和左连续。本课件采用$P(X\le x)$定义分布函数。

二、随机变量概念

3.离散随机变量的概率分布列

若离散随机变量X的可能取值是$x_1, x_2, \cdots, x_n, \cdots$，则称$X$取$x_i$的概率

$$p_i = p(x_i) = P(X = x_i), i = 1,2, \cdots, n,$$

为X的概率分布列，简称分布列。分布列也可以用列表方式来表示：

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_n)$	$\cdots$

分布列$p(x_i)$具有如下两条基本性质：

i)非负性：$p(x_i) \geq 0$;　　ii)正则性: $\sum_{i=1}^{+\infty} p(x_i) = 1$.

二、随机变量概念

4.连续随机变量的概率密度函数

记连续随机变量X的分布函数为$F(x)$，若存在一个非负可积函数$f(x)$，使得对任意实数x，有$F(x)=\int_{-\infty}^{x}f(t)dt$，则称$f(x)$为$X$的概率密度函数，简称为密度函数。密度函数$f(x)$具有如下两条性质：

i) 非负性： $f(x)\geq 0$;

ii) 正则性： $\int_{-\infty}^{+\infty}f(x)dx=1.$

注：密度函数有常用的两种表达方式$f(x)$与$p(x)$，目前更为常用$f(x)$与对应大写相匹配，此外$F'(x)=f(x)$。

二、随机变量概念

5.常用公式

设随机变量X的分布函数为$F(x)$，则可用$F(x)$表示下列概率：

i) $P(X \le a) = F(a)$;

ii) $P(x < a) = F(a-0)$;

iii) $P(X > a) = 1 - P(X \le a) = 1 - F(a)$;

iv) $P(X = a) = P(X \le a) - P(X < a)$

$$= F(a) - F(a-0)$$

v) $P(X \ge a) = 1 - P(X < a) = 1 - F(a-0)$;

vi) $P(|X| < a) = P(-a < X < a)$

$$= P(X < a) - P(X \le -a)$$

$$= F(a-0) - F(-a).$$

二、随机变量概念

- **例5**（2011数1,3）设$F_1(x)$，$F_2(x)$为两个分布函数，其相应的概率密度$f_1(x)$，$f_2(x)$是连续函数，则必为概率密度的是（ ）

(A) $f_1(x)f_2(x)$　　(B)$2f_2(x)F_1(x)$

(C) $f_1(x)F_2(x)$　　(D)$f_1(x)F_2(x)+f_2(x)F_1(x)$

- **解：** 因为$F_1'(x)=f_1(x), F_2'(x)=f_2(x)$，于是

$$f_1(x)F_2(x)+f_2(x)F_1(x)=F_1'(x)F_2(x)+F_2'(x)F_1(x)$$
$$=[F_1(x)F_2(x)]',$$

从而

$$\int_{-\infty}^{+\infty}[f_1(x)F_2(x)+f_2(x)F_1(x)]\,dx=[F_1(x)F_2(x)]\Big|_{-\infty}^{+\infty}=1,$$

且 $f_1(x)F_2(x)+f_2(x)F_1(x)\geq 0$，故选(D)。

取 F_1,F_2服从 [0,1],[0,2]区间的均匀分布，反证前三项不成立。

二、随机变量概念

- **例6** 设随机变量X的密度函数为

$$\varphi(x) = Ae^{-|x|},\quad -\infty < x < +\infty,$$

试求：（1）系数A；（2）$P\{0 < X < 1\}$；

（3）X的分布函数.

- **解：**（1）由密度函数的性质

$$1 = \int_{-\infty}^{+\infty} Ae^{-|x|}dx = 2A\int_0^{+\infty} e^{-x}\,dx = -2Ae^{-x}\Big|_0^{+\infty} = 2A$$

解得 $A = \frac{1}{2}$.

（2）$P\{0 < X < 1\} = \int_0^1 \frac{1}{2}e^{-|x|}dx = -\frac{1}{2}e^{-x}\Big|_0^1 = \frac{1-e^{-1}}{2}$.

二、随机变量概念

（3）X的分布函数$F(x)=\int_{-\infty}^{x}\varphi(t)dt=\int_{-\infty}^{x}\frac{1}{2}e^{-|t|}dt$.

当$x<0$时，$F(x)=\int_{-\infty}^{x}\frac{1}{2}e^{t}dt=\frac{1}{2}e^{x}$;

当$x\geq 0$时，$F(x)=\int_{-\infty}^{x}\frac{1}{2}e^{-|t|}dt=\int_{-\infty}^{0}\frac{1}{2}e^{t}dt+\int_{0}^{x}\frac{1}{2}e^{-t}\,dt$

$$=1-\frac{1}{2}e^{-x};$$

故 $$F(x)=\begin{cases}\frac{1}{2}e^{x}, & x<0,\\ 1-\frac{1}{2}e^{-x}, & x\geq 0.\end{cases}$$

三、随机变量的数字特征

1.数学期望(Expectation)

设随机变量X的分布用分布列$p(x_i)$或用密度函数$f(x)$表示，若

$$\begin{cases} \sum_i |x_i| p(x_i) < +\infty, \text{当}X\text{为离散随机变量}, \\ \int_{-\infty}^{+\infty} |x| f(x) dx < +\infty, \text{当}X\text{为连续随机变量}, \end{cases}$$

则称$E(X) = \begin{cases} \sum_i x_i p(x_i), & \text{当}X\text{为离散随机变量}, \\ \int_{-\infty}^{+\infty} x f(x) dx, & \text{当}X\text{为连续随机变量} \end{cases}$

为X的数学期望，简称期望或均值，且称X的数学期望存在.否则称数学期望不存在.

三、随机变量的数字特征

数学期望的性质：

(1)X的某一函数$g(X)$的数学期望为

$$E[g(X)] = \begin{cases} \sum_i g(x_i)p(x_i), \text{在离散场合}, \\ \int_{-\infty}^{+\infty} g(x)f(x)dx, \text{在连续场合}; \end{cases}$$

(2)若c是常数，则$E(c) = c$;

(3)线性组合不变性(思考：n无穷大时成立吗?)

$$E\left(\sum_{i=1}^{n} k_i X_i\right) = \sum_{i=1}^{n} k_i E(X_i), \qquad \forall k_i \in R, i = 1,2,\cdots,n$$

(4)对任意常数a，有$E(aX) = aE(X)$;

(5)对任意的两个函数$g_1(x)$和$g_2(x)$，有

$$E[g_1(X) \pm g_2(X)] = E[g_1(X)] \pm E[g_2(X)].$$

三、随机变量的数字特征

- **例7** （2009数1）设随机变量X的分布函数为$F(x)=0.3\Phi(x)+0.7\Phi\left(\frac{x-1}{2}\right)$，其中$\Phi(x)$为标准正态分布函数，则$E(X)=$______.
- **解:** 随机变量X的概率密度为

$$f(x)=F'(x)=0.3\Phi'(x)+0.7\Phi'\left(\frac{x-1}{2}\right)=0.3\varphi(x)+0.35\varphi\left(\frac{x-1}{2}\right),$$

$$\mathrm{E}(X)=\int_{-\infty}^{+\infty}xf(x)\,dx=0.3\int_{-\infty}^{+\infty}x\varphi(x)\,dx+0.35\int_{-\infty}^{+\infty}x\varphi\left(\frac{x-1}{2}\right)dx.$$

注意到 $\int_{-\infty}^{+\infty}x\varphi(x)\,dx=0$（**思考：为什么**），以及

$$\int_{-\infty}^{+\infty}x\varphi\left(\frac{x-1}{2}\right)dx\xRightarrow{\frac{x-1}{2}=u}\int_{-\infty}^{+\infty}(1+2u)\varphi(u)\cdot 2du=2\int_{-\infty}^{+\infty}\varphi(u)du=2.$$

故$E(X)=0.3\times0+0.35\times2=0.7$.其中$\phi(x)$为标准正态密度函数。

注：本题还有其他解法。灵活使用正态分布性质。

三、随机变量的数字特征

❖ **例8** 某种灯管的使用寿命T是一个随机变量（单位：小时），其密度函数为

$$f(t)=\begin{cases}0.001e^{-0.001t}, & t>0,\\ 0, & t\le 0,\end{cases}$$

试求该种灯管的平均寿命。

❖ **解：** $E(T)=\int_0^{+\infty} t0.001e^{-0.001t}dt$

$$=-te^{-0.001t}\Big|_0^{+\infty}+\int_0^{+\infty}e^{-0.001t}dt$$

$$=-\frac{1}{0.001}e^{-0.001t}\Big|_0^{+\infty}=1000.$$

故其平均维修时间为1000小时.

注：该分布为失效率参数λ=0.001的指数分布。

三、分布的其他数字特征

- **例8-1** 设国际市场上对某种出口商品每年的需求量 X （单位：吨）是随机变量，它服从区间[2000,4000]上的均匀分布，每销售一吨商品，可为国家赚取外汇3万元；若销售不出，则每吨商品需贮存费1万元.问应组织多少货源，才能使国家收益期望最大.
- **解：** 设Y表示国家收益，假设组织货源t吨，显然应要求 $2000 \le t \le 4000$，则

$$Y = g(X,t) = \begin{cases} 3t, & X \ge t \\ 3X-(t-X), & X < t \end{cases}$$

又由题设知X的概率密度函数为

$$f_X(x) = \begin{cases} \dfrac{1}{2000}, 2000 \le x \le 4000 \\ 0, 其他 \end{cases}$$

于是Y的期望为

三、分布的其他数字特征

$$E(Y)=\int_{-\infty}^{+\infty} g(x,t)\, f_X(x)dx$$

$$=\frac{1}{2000}\left[\int_{2000}^{t}(4x-t)dx+\int_{t}^{4000}3tdx\right]$$

$$=\frac{1}{1000}(-t^2+7000t-4\times 10^6).$$

由$\frac{dE(Y)}{dt}=\frac{1}{1000}(-2t+7000)=0$，得$t=3500$.

由题意知应组织3500吨货源，才能使国家收益期望最大.

注： 需要注意，这里t是确定性的参变量，而X,Y是随机变量.

三、随机变量的数字特征

2.方差

随机变量X对其期望$E(X)$的偏差平方的数学期望（设其存在）

$$Var(X)=E[X-E(X)]^2$$

称为X的方差，方差的正平方根$\sigma(X)=\sigma_X=\sqrt{Var(X)}$称为$X$的标准差.

方差性质：(1)$Var(X)=E(X^2)-[E(X)]^2$;

(2)若c是常数，则$Var(c)=0$;

(3)若a，b常数，则$Var(aX+b)=a^2Var(X)$;

(4)若随机变量X的方差存在，则$Var(X)=0$的充要条件是X几乎处处为某个常数a，即$P(X=a)=1$.

注：$Var(X)$与$D(X)$用于表示方差，课件将统一采用$Var(X)$。

三、随机变量的数字特征

3.随机变量的标准化

对任意随机变量X，如果X的数学期望和方差存在，则称

$$Z=\frac{X-E(X)}{\sqrt{Var(X)}}$$

为X的标准化随机变量，此时有

$E(Z)=0, Var(Z)=1.$

注：若$X \sim N(\mu，\sigma^2)$，则$Z=\frac{X-\mu}{\sigma} \sim N(0，1)$.正态分布的标准化是数据分析常用的一种方法。**标准化优势是什么?**

三、随机变量的数字特征

❖ **例9**设X为随机变量，具有密度函数$f(x)=2x, 0\le x\le 1$，试求$E(X)$和$Var(X)$.

❖ **解：** 因为$E(X)=\int_{-\infty}^{+\infty} xf(x)\,dx=\int_0^1 x\cdot 2xdx=\frac{2}{3}$,

$$E(X^2)=\int_{-\infty}^{+\infty} x^2 f(x)\,dx=\int_0^1 x^2\cdot 2xdx=\frac{1}{2},$$

则 $Var(X)=E(X^2)-[E(X)]^2=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{1}{18}$.

❖ **注：** 此为$Beta(2,1)$分布，由$Beta$分布的数字特征可直接知：

$$E(X)=\frac{a}{a+b}=\frac{2}{2+1}=\frac{2}{3}$$

$$Var(X)=\frac{ab}{(a+b)^2(a+b+1)}=\frac{2\times 1}{(2+1)^2\times(2+1+1)}=\frac{1}{18}.$$

三、随机变量的数字特征

- **例10** (1)若$X\sim U(1,n)$，求$E(X)$和$Var(X)$；
  (2)若$X\sim disU(1,n)$，求$E(X)$和$Var(X)$.

- **解:** (1)若$X\sim U(1，n)$，则$f(x)=\frac{1}{n-1}$，$1\le x\le n$.

$$E(X)=\int_{-\infty}^{+\infty} xf(x)\,dx=\int_1^n x\cdot\frac{1}{n-1}dx=\frac{1}{n-1}\cdot\frac{n^2-1}{2}=\frac{n+1}{2}$$

$$E(X^2)=\int_{-\infty}^{+\infty} x^2f(x)\,dx=\int_1^n x^2\cdot\frac{1}{n-1}dx=\frac{1}{n-1}\cdot\frac{n^3-1}{3}=\frac{n^2+n+1}{3}$$

故 $Var(X)=E(X^2)-[E(X)]^2=\frac{n^2+n+1}{3}-\left(\frac{n+1}{2}\right)^2=\frac{(n-1)^2}{12}$

三、随机变量的数字特征

(2)若$X \sim disU(1, n)$，则$f(x) = \frac{1}{n}$, $x = 1,2,\cdots,n$.

$$E(X) = \sum_{x=1}^{n} x \cdot f(x) = \sum_{x=1}^{n} x \cdot \frac{1}{n} = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

$$E(X^2) = \sum_{x=1}^{n} x^2 \cdot f(x) = \sum_{x=1}^{n} x^2 \cdot \frac{1}{n} = \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6}$$

$$= \frac{(n+1)(2n+1)}{6}$$

故 $Var(X) = E(X^2) - [E(X)]^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}$

❖ **注：**注意区别均匀和离散均匀数字特征差异.

四、常用离散分布

1.二项分布(Binomial)

(1)若X的概率分布列为

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, k=0,1,\cdots,n.$$

则称X服从二项分布，记为$X\sim b(n,p)$，其中$0<p<1$.

(2)背景：n重伯努利试验中成功的次数X服从二项分布$b(n,p)$，其中p为一次伯努利试验中成功发生的概率.
如产品抽检抽取次品的概率等.

(3)$n=1$时的二项分布$b(1,p)$称为二点分布，或称$0-1$分布。因为当$X\sim b(1,p)$时，X可表示一次伯努利试验中成功的次数，它只能取值0与1.

四、常用离散分布

(4)二项分布$b(n,p)$的数学期望和方差分别是

$$E(X)=np,\quad Var(X)=np(1-p).$$

(5)若$X\sim b(n,p)$，则$Y=n-X\sim b(n,1-p)$，其中 $Y=n-X$ 是n重伯努利试验中失败的次数。

四、常用离散分布

2.泊松分布(Poisson)

(1)若X的概率分布列为

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},k=0,1,2,\cdots,$$

则称X服从泊松分布，记为$X \sim P(\lambda)$，（或记为$Pois(\lambda)$，其中参数$\lambda>0$。）

(2)背景：单位时间（或单位面积、单位产品等）上某稀有事件（这里稀有事件是指不经常发生的事件）发生的次数常服从泊松分布$P(\lambda)$，其中λ为该稀有事件发生的强度。如单位时间到银行办理业务的顾客数。

四、常用离散分布

(3)泊松分布$P(\lambda)$的数学期望和方差分别是

$$E(X)=\lambda,\qquad Var(X)=\lambda.$$

(4) **二项分布的泊松近似（泊松定理）**

在n重伯努利试验中，记事件A在一次试验中发生的概率为（与试验次数n有关），如果当$n\to+\infty$时，有$np_n\to\lambda$，则

$$\lim_{n\to+\infty}\binom{n}{k}p^k(1-p)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}.$$

四、常用离散分布*

3.超几何分布

4.几何分布

5.负二项分布

（略）

四、常用离散分布

名称与记号	分布列	期望	方差
$0-1$分布 $b(1,p)$	$p_k=p^k(1-p)^{1-k}, k=0,1$	p	$p(1-p)$
二项分布 $b(n,p)$	$p_k=\binom{n}{k}p^k(1-p)^{n-k}, k=0,1,\cdots,n.$	np	$np(1-p)$
泊松分布 $P(\lambda)$	$p_k=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,\cdots$	λ	λ
超几何分布* $h(n,N,M)$	$p_k=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \begin{matrix} k=0,1,\cdots,r, \\ r=\min\{M,n\}\end{matrix}$	$\frac{nM}{N}$	$\frac{nM(N-M)(N-n)}{N^2(N-1)}$
几何分布* $Ge(p)$	$p_k=(1-p)^{k-1}p, k=1,2,\cdots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
负二项分布 * $Nb(r,p)$	$p_k=\binom{k-1}{r-1}(1-p)^{k-r}p^r,$ $k=r,r+1,\cdots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

四、常用离散分布

- **例11** 设X服从参数为$\sqrt{5}$的$poisson$分布，则使得$P\{X=k\}$达到最大的k为多少?
- **解：** 由题知$X\sim P(\sqrt{5})$,

$$P\{X=k\}=\frac{\sqrt{5}^k}{k!}e^{-\sqrt{5}},k=0,1,2,\cdots,$$

$P\{X=k\}$达到最大，即$\begin{cases}P\{X=k\}\geq P\{X=k+1\}\\P\{X=k\}\geq P\{X=k-1\}\end{cases}$

代入并化简可得

$$\begin{cases}\dfrac{\sqrt{5}^k}{k!}\geq\dfrac{\sqrt{5}^{k+1}}{(k+1)!}\\\dfrac{\sqrt{5}^k}{k!}\geq\dfrac{\sqrt{5}^{k-1}}{(k-1)!}\end{cases}$$

解得$\sqrt{5}-1\leq k\leq\sqrt{5}$，又$k$为整数，所以$k=2$.

四、常用离散分布

❖ **例12** （2008数1,3）设随机变量X服从参数为1的泊松分布，则$P\{X=E(X^2)\}=$____.

❖ **解：** 由X服从参数为1的泊松分布知

$$E(X)=Var(X)=1,$$

因此

$$E(X^2)=Var(X)+[E(X)]^2=2.$$

所以$P\{X=E(X^2)\}=P\{X=2\}=\frac{1}{2}e^{-1}$,应填$\frac{1}{2}e^{-1}$.

五、常用连续分布

1.正态分布

(1)若X的密度函数（如下图）和分布函数分别为

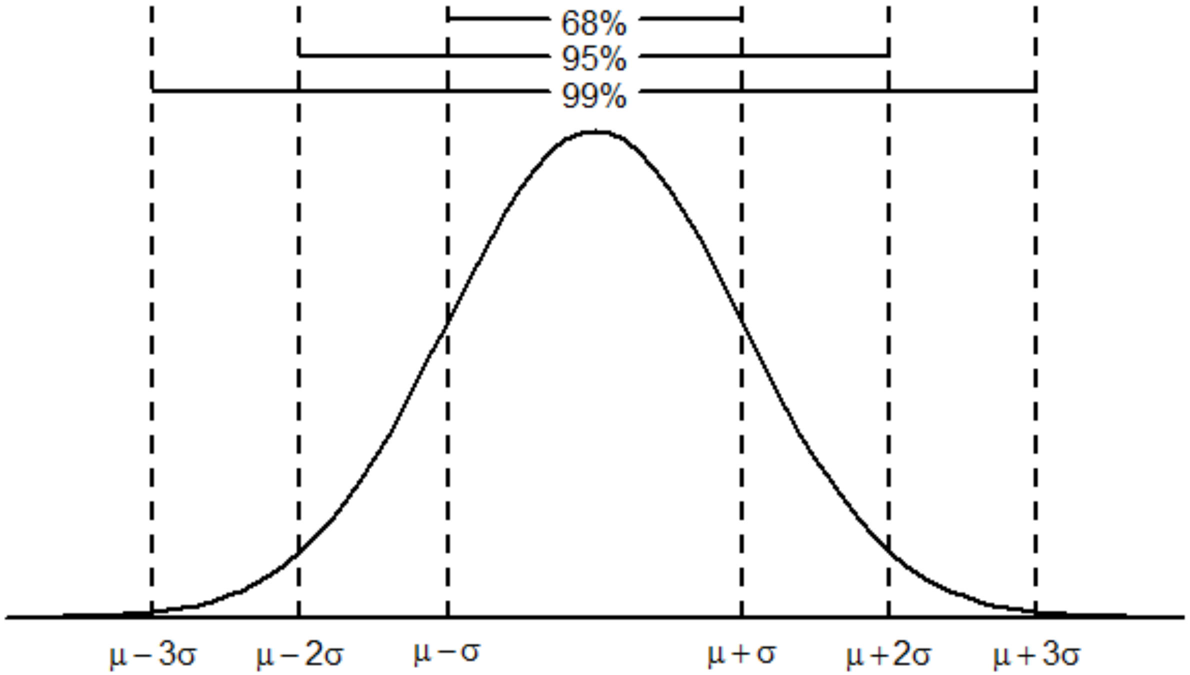
$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty,$$

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt,-\infty<x<\infty,$$

则称X服从正态分布，记作$X\sim N(\mu,\sigma^2)$，其中参数$-\infty<\mu<\infty,\sigma>0.$

五、常用连续分布

正态分布常用的概率



(2)背景：一个变量若是由大量微小的、独立的随机因素的叠加结果，则此变量一定是正态变量（服从正态分布的变量）。测量误差就是由量具偏差、测量环境的影响、测量技术的影响、测量人员的心理影响等等随机变量叠加而成的，所以测量误差常认为服从正态分布。最早由高斯研究误差理论发现。

五、常用连续分布

(3)**关于参数μ**：

- 正态分布的数学期望，即$E(X)=\mu$，称μ为正态分布的位置参数(Location parameter).
- μ是正态分布的对称中心，密度曲线在μ的两侧覆盖面积都为0.5，所以μ也是正态分布的中位数（见后面§2.7）.
- 若$X\sim N(\mu,\sigma^2)$，则X在离μ愈近取值的可能性愈大，离μ愈远取值的可能性愈小.

关于参数σ^2:

- 正态分布的方差，即$Var(X)=\sigma^2$.
- σ是正态分布的标准差，σ愈小，正态分布愈集中；σ愈大，正态分布愈分散. σ又称为正态分布的尺度(Scale)参数.
- 若$X\sim N(\mu,\sigma^2)$，则其密度函数$f(x)$在$\mu\pm\sigma$处有两个拐点.

五、常用连续分布

(4)称$\mu=0,\sigma=1$时的正态分布$N(0,1)$为标准正态分布.记Z为标准正态变量，$\varphi(z)$和$\Phi(z)$为标准正态分布的密度函数和分布函数. $\varphi(z)$和$\Phi(z)$满足:

- $\varphi(-z)=\varphi(z)$；
- $\Phi(-z)=1-\ \Phi(z)$.对$u>0$, $\Phi(u)$的值有表可查。
- $P(|Z|<z)=\Phi(z)-\Phi(-z)=2\Phi(z)-1.$

(5)标准化变换：若$X\sim N(\mu,\sigma^2)$，则

$Z=(X-\mu)/\sigma\sim N(0,1)$，其中$Z=(X-\mu)/\sigma$称为$X$的标准化变量。

五、常用连续分布

(6)若$X\sim N(\mu,\sigma^2)$，则对任意实数a与b，

$$P(X\le b)=\Phi\left(\frac{b-\mu}{\sigma}\right),$$

$$P(a<X)=1-\Phi\left(\frac{a-\mu}{\sigma}\right),$$

$$P(a<X\le b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right),$$

可见，涉及正态变量的概率计算，一般先做标准化，再查标准正态分布的分布函数表获得.

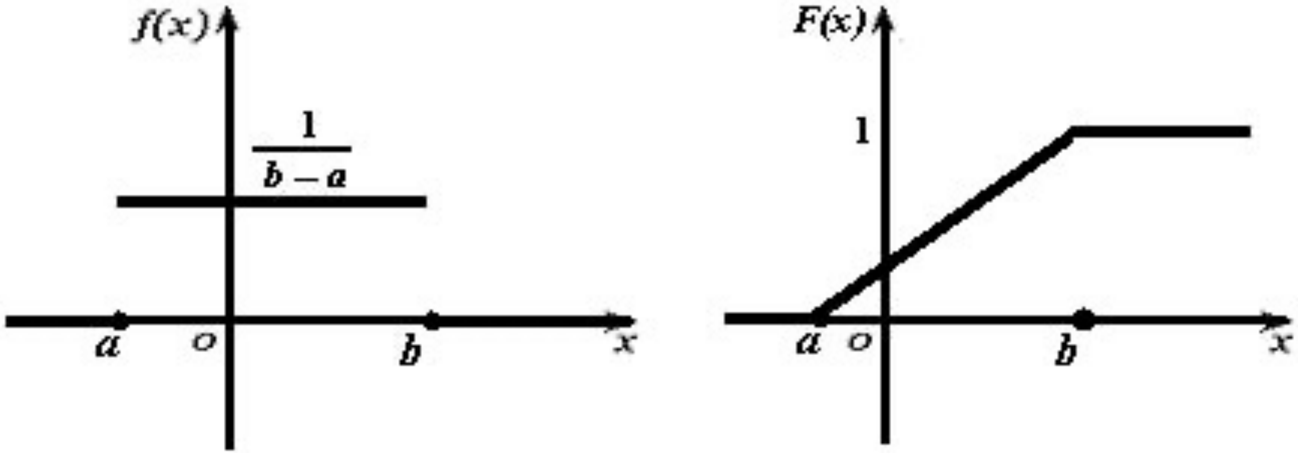
(7)正态分布的3σ原则：设$X\sim N(\mu,\sigma^2)$，则

$$P(|X-\mu|<k\sigma)=\Phi(k)-\Phi(-k)=\begin{cases}0.6826, k=1\\0.9545, k=2\\0.9973, k=3.\end{cases}$$

五、常用连续分布

2.均匀分布

(1)若X的密度函数和分布函数（如下图）分别为



$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b,\\ 0\,, & 其他,\end{cases}\qquad F(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\le x<b\\ 1, & x\ge b.\end{cases}$$

则称X服从区间(a,b)上的均匀分布，记作$X\sim U(a,b)$.

五、常用连续分布

(2)背景：向区间(a,b)随机投点，落点坐标X一定服从均匀分布$U(a,b)$.这里“随机投点”是指：点落在任意相等长度的小区间上的可能性是相等的.

(3)均匀分布$U(a,b)$的数学期望和方差分别是

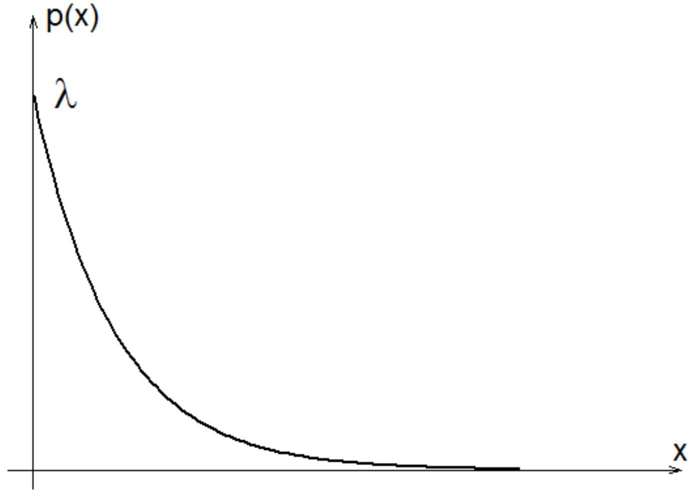
$$E(X)=\frac{a+b}{2},\qquad Var(X)=\frac{(b-a)^2}{12}$$

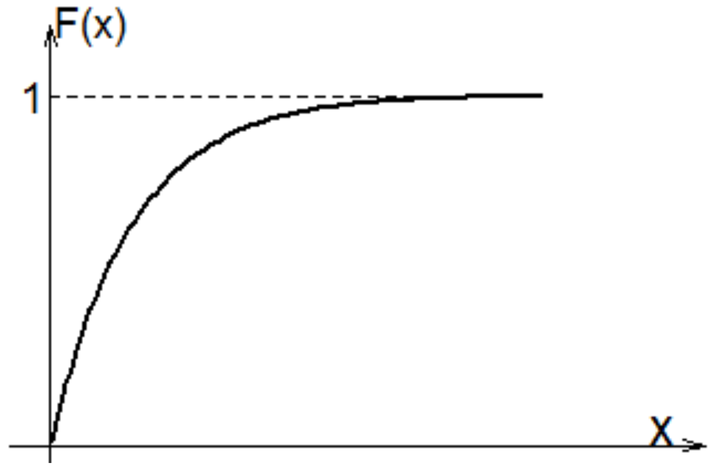
(4)称区间(0,1)上的均匀分布$U(0,1)$为标准均匀分布，它是导出其他分布随机数的桥梁.感兴趣的可查阅Ross著作*Simulation*。

五、常用连续分布

3.指数分布

(1)若X的密度函数和分布函数（如下图）分别为





$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0,\\ 0, & x<0,\end{cases}\qquad F(x)=\begin{cases}1-e^{-\lambda x}, & x\geq 0,\\ 0, & x<0,\end{cases}$$

则称X服从指数分布，记作$X\sim Exp(\lambda)$，其中参数$\lambda>0$.

五、常用连续分布

(2)背景：若一个元器件（或一台设备、或一个系统）遇到外来冲击时即告失效，则首次冲击来到的时间X（寿命）服从指数分布.很多产品的寿命可认为服从或近似服从指数分布.

(3)指数分布$Exp(\lambda)$的数学期望和方差分别为

$$E(X)=\frac{1}{\lambda}, Var(X)=\frac{1}{\lambda^2}$$

(4)指数分布的无记忆性：若$X\sim Exp(\lambda)$，则对任意$s>0$，$t>0$，有$P(X>s+t|X>s)=P(X>t)$.

五、常用连续分布

名称与记号	密度函数$f(x)$	期望	方差
正态分布$\boldsymbol{N(\mu,\ \sigma^2)}$	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, -\infty < x < +\infty$	μ	σ^2
均匀分布$U(\boldsymbol{a},\boldsymbol{b})$	$\frac{1}{b-a}, a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布$Exp(\lambda)$	$\lambda e^{-\lambda x}, x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
伽玛分布$Ga(\alpha,\lambda)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, x \geq 0$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$\boldsymbol{\chi^2(n)}$分布	$\frac{x^{n/2-1}e^{-x/2}}{\Gamma(n/2)2^{n/2}}, x \geq 0$	n	$2n$

五、常用连续分布*

名称与记号	密度函数$f(x)$	期望	方差
贝塔分布 $Be(\boldsymbol{a},\boldsymbol{b})$	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(\mathrm{b})}x^{a-1}(1-x)^{b-1}$, $0<x<1$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$
对数正态分布 $LN(\boldsymbol{\mu},\boldsymbol{\sigma}^2)$	$\frac{1}{\sqrt{2\pi}\sigma x}\exp\left\{-\frac{(lnx-\mu)^2}{2\sigma^2}\right\}$, $x>0$	$e^{\mu+\sigma^2/2}$	$e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$
柯西分布 $Cau(\boldsymbol{\mu},\boldsymbol{\lambda})$	$\frac{1}{\pi}\frac{\lambda}{\lambda^2+(x-\mu)^2}$, $-\infty<x<+\infty$	不存在	不存在
韦布尔分布 $Wei(\boldsymbol{m},\ \boldsymbol{\eta})$	$F(x)=1-\exp\left\{-\left(\frac{x}{\eta}\right)^m\right\}$, $x>0$	$\eta\Gamma(1+\frac{1}{m})$	$\eta^2\left[\Gamma\left(1+\frac{2}{m}\right)-\Gamma^2(1+\frac{1}{m})\right]$

五、常用连续分布

- **例13** 设K服从$(1,6)$上的均匀分布，求方程$x^2+Kx+1=0$有实根的概率.
- **解：** 方程$x^2+Kx+1=0$有实根的充要条件是

$$\{K^2-4\geq 0\}=\{K\leq -2\}\cup\{K\geq 2\}$$

而$K\sim U(1,6)$，因此所求概率为

$$P(K\leq -2)+P(K\geq 2)=0+\int_2^6\frac{1}{5}dx=\frac{4}{5}.$$

五、常用连续分布

- **例14** （2006数1,3）设随机变量X服正态分布$N(\mu_1,{\sigma_1}^2)$，Y服从正态分布$N(\mu_2,{\sigma_2}^2)$，且

$P\{|X-\mu_1|<1\}>P\{|Y-\mu_2|<1\}$，则必有（）

(A) $\sigma_1 < \sigma_2$ (B) $\sigma_1 > \sigma_2$

(C) $\mu_1 < \mu_2$ (D) $\mu_1 > \mu_2$

- **解：** 由题设可得$P\left\{\frac{|X-\mu_1|}{\sigma_1}<\frac{1}{\sigma_1}\right\}>P\left\{\frac{|Y-\mu_2|}{\sigma_2}<\frac{1}{\sigma_2}\right\}$，

则$2\Phi\left(\frac{1}{\sigma_1}\right)-1>2\Phi\left(\frac{1}{\sigma_2}\right)-1$，即$\Phi\left(\frac{1}{\sigma_1}\right)>\Phi\left(\frac{1}{\sigma_2}\right)$，

其中$\Phi(x)$是标准正态分布的分布函数.

又$\Phi(x)$是严格单调增加函数，所以$\frac{1}{\sigma_1}>\frac{1}{\sigma_2}$，即$\sigma_1 < \sigma_2$.选$A$.

五、常用连续分布

- **例15** 某校自主招生，共有10000人报考，假设考试成绩服从正态分布，且已知90分以上有359人，60分以下有1151人.现按考试成绩从高分到低分依次录用250人，试问被录取者中最低分为多少?
- **解：** 记X为考试成绩，则$X \sim N(\mu, \sigma^2)$.由频率估计概率知

$$0.0359 = P(X > 90) = 1 - \Phi(\frac{90-\mu}{\sigma}),$$

$$0.1151 = P(X < 60) = \Phi(\frac{60-\mu}{\sigma}),$$

上面两式可改写为 $0.9641 = \Phi\left(\frac{90-\mu}{\sigma}\right), 0.8849 = \Phi\left(\frac{\mu-60}{\sigma}\right).$

五、常用连续分布

再查表得$\frac{90-\mu}{\sigma}=1.8, \frac{\mu-60}{\sigma}=1.2$

由此解得$\mu=72, \sigma=10$.设被录用者中最低分为k，则由

$$0.025=P(X \geq k)=1-\Phi(\frac{k-72}{10})，或0.975=\Phi(\frac{k-72}{10}),$$

查表得$(k-72)/10 \geq 1.96$，从中解得$k \geq 91.6$，因此取被录用者中最低分为91.6分即可.

注： 当$p<0.5$时，满足等式$\Phi(x)=p$的x在标准正态分布函数表上不易查得，故改写此式为 $\Phi(-x)=1-p>0.5$，即可查得$-x$.

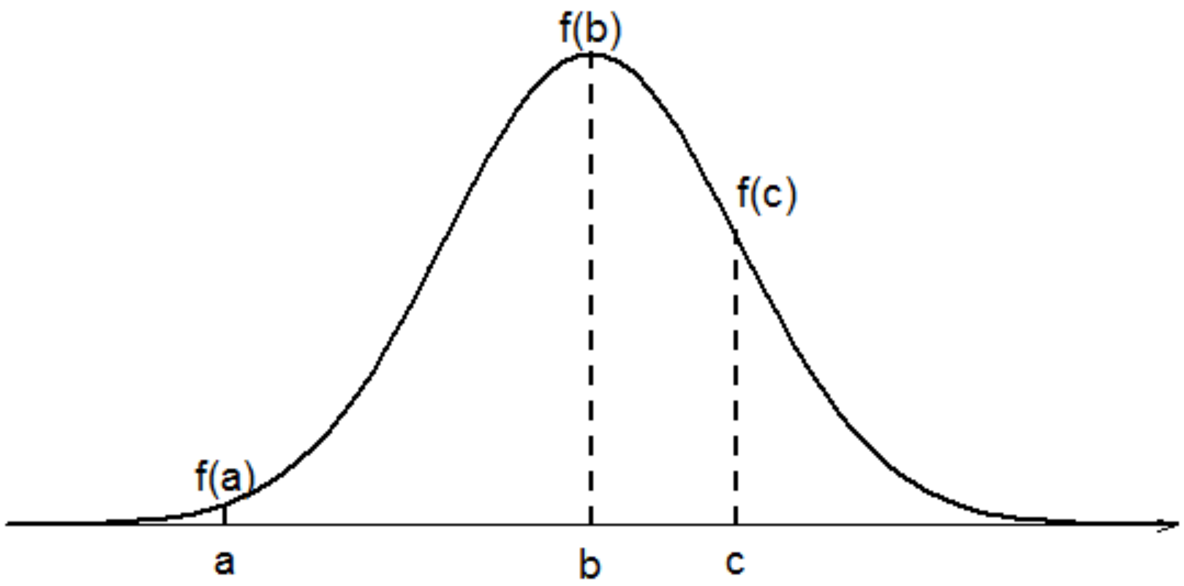
五、常用连续分布

- **例16**设随机变量$X \sim N(\mu,\sigma^2)$，$f(x)$为其密度函数，且a,b,c满足：$a<b<c, f(a)<f(c)<f(b)$,则

(A)$\frac{a+c}{2}<\mu<\frac{b+c}{2}$　　(B) $\frac{a+b}{2}<\mu<\frac{a+c}{2}$

(C)$a<\mu<b$　　(D)$b<\mu<c$

- **解：** 选A.结合画图用排除法，可得结论.



六、随机变量函数的分布

1. 设连续随机变量X的密度函数为$f_X(x), Y=g(X)$.

(1)若$y=g(x)$严格单调，其反函数$h(y)$有连续导数，则$Y=g(X)$的密度函数为

$$f_Y(y)=\begin{cases} f_X[h(y)]|h'(y)|, & a<y<b, \\ 0, & \text{其他}, \end{cases}$$

其中$a=\min\{g(-\infty), g(+\infty)\}$，$b=\max\{g(-\infty), g(+\infty)\}$.

(2)若$y=g(x)$在不相重叠的区间$I_1, I_2, \ldots$上逐段严格单调，其反函数$h_1(y), h_2(y), \ldots$有连续导数，则$Y=g(X)$的密度函数为

$$f_Y(y)=\sum_i f_X\big(h_i(y)\big)|h_i'(y)|$$

2.正态变量的线性变换仍为正态变量 若X服从正态分布$N(\mu,\sigma^2)$，则当$a\neq 0$时，有$Y=aX+b\sim N(a\mu+b, a^2\sigma^2)$.

六、随机变量函数的分布*

3.对数正态分布（略）

(1)若X的密度函数为

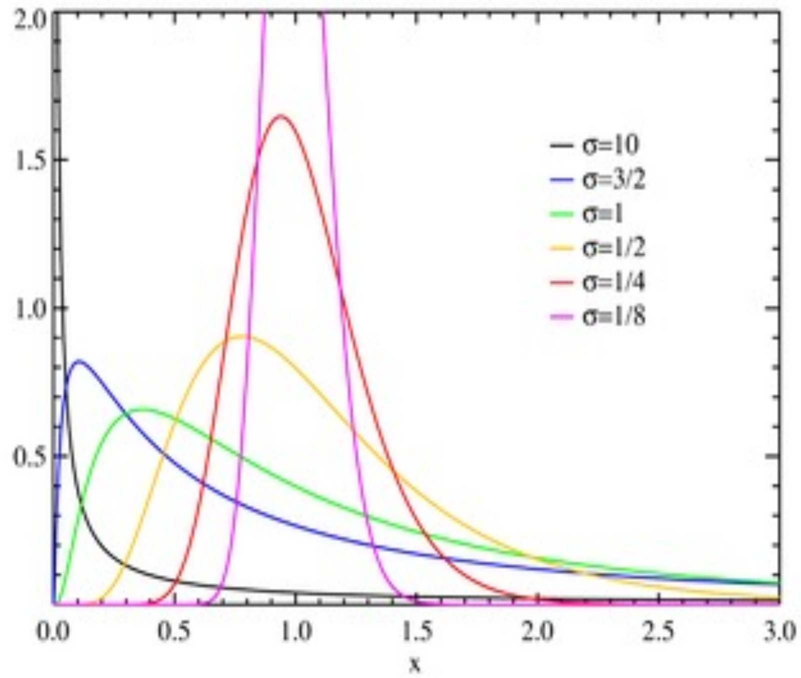
$$f_X(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}x\sigma}\exp\left\{-\dfrac{(\ln x-\mu)^2}{2\sigma^2}\right\}, & x>0\\ 0, & x\le 0.\end{cases}$$

则称X服从对数正态分布，记为$X\sim LN(\mu,\sigma^2)$，其中$-\infty<\mu<+\infty,\sigma>0$.对数正态分布的密度函数如图

(2)若$X\sim LN(\mu,\ \sigma^2)$，则$E(X)=\ e^{\mu+\sigma^2/2}$，

$$Var(X)=\ (e^{\sigma^2}-1)\ \cdot e^{2\mu+\sigma^2}.$$

(3)若$X\sim LN(\mu,\ \sigma^2)$，则$Y=logX\sim N(\mu,\sigma^2)$



六、随机变量函数的分布

- **例17** 设随机变量X服从参数为2的指数分布，试证$Y_1=e^{-2X}$和$Y_2=1-e^{-2X}$都服从区间(0,1)上的均匀分布。
- **证：** 因为X的密度函数为

$$f_X(x)=\begin{cases}2e^{-2x}, & x>0,\\ 0, & \text{其他},\end{cases}$$

又因为Y_1的可能取值范围是(0,1)，且$y_1=e^{-2x}$是严格单调减函数，其反函数为$x=h(y_1)=-0.5\ln y_1$，所以Y_1的密度函数为

六、随机变量函数的分布

$$f_{Y_1}(y)$$

$$= \begin{cases} f_X(-0.5lny_1)|-0.5/y_1|, & 0 < y_1 < 1 \\ 0, & \text{其他} \end{cases}$$

$$= \begin{cases} 1, & 0 < y_1 < 1, \\ 0, & \text{其他} \end{cases}$$

即$Y_1 \sim U(0,1)$.又由前面第4题，知$Y_2 = 1 - e^{-2X} = 1 - Y_1$也服从区间(0,1)上的均匀分布.结论得证.

注：若$Z \sim F(x)$，则$F(Z) \sim U(0,1)$.

七、多维随机变量及其分布

1.联合分布函数 对任意的n个实数$x_1, x_2, \dots, x_n$, n个事件 $X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n$同时发生的概率

$$F(x_1, x_2, \dots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n)$$

称为n维随机变量$(X_1, X_2, \dots, X_n)$的联合分布函数.

二维随机变量(X, Y)的联合分布函数$F(x, y) = P(X \le x, Y \le y)$具有如下四条基本性质:

(1)单调性 $F(x, y)$分别对x或y是单调不减的.

(2)有界性 对任意的x和y，有$0 \le F(x, y) \le 1$，且

$$F(-\infty, y) = F(x, -\infty) = 0，F(+\infty, +\infty) = 1.$$

(3)右连续性 对每个变量都是右连续的，即

$$F(x+0, y) = F(x, y)，F(x, y+0) = F(x, y).$$

七、多维随机变量及其分布

(4)非负性 对任意的$a < b$，$c < d$有

$$P(a < X \le b, c < Y \le d) = F(b,d) - F(a,d) - F(b,c) + F(a,c) \ge 0.$$

2.联合分布列

3.联合密度函数 如果存在二元非负函数$f(x,y)$，使得二维随机变量(X,Y)的分布函数$F(x,y)$可表示为

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dvdu,$$

则称(X,Y)为二维连续随机变量，称$f(x,y)$为(X,Y)的联合密度函数.

联合密度函数的基本性质:

(1)非负性:$f(x,y) \ge 0$; (2)正则性:$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1.$

七、多维随机变量及其分布*

4.条件分布

（1）离散型随机变量的条件分布

i)条件分布列　对一切使 $P(Y=y_j)=p_{\cdot j}=\sum_{i=1}^{+\infty}p_{ij}>0$ 的 y_j，称

$$p_{i|j}=P(X=x_i|Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\cdot j}},i=1,2\cdots$$

为给定$Y=y_j$条件下X的条件分布列.

同理，对一切使$P(X=x_i)=p_{i\cdot}=\sum_{j=1}^{+\infty}p_{ij}>0$的$x_i$，称

$$p_{j|i}=P(Y=y_j|X=x_i)=\frac{P(X=x_i,Y=y_j)}{P(X=x_i)}=\frac{p_{ij}}{p_{i\cdot}},j=1,2\cdots$$

为给定$X=x_i$条件下Y的条件分布列.

七、多维随机变量及其分布*

ii)条件分布函数 给定$Y=y_j$条件下X的条件分布函数为

$$F(x|y_j)=\sum_{x_i\le x}P(X=x_i|Y=y_j)=\sum_{x_i\le x}p_{i|j},$$

给定$X=x_i$条件下Y的条件分布函数为

$$F(y|x_i)=\sum_{y_j\le y}P(Y=y_j|X=x_i)=\sum_{y_j\le y}p_{j|i},$$

七、多维随机变量及其分布*

(2)连续随机变量的条件分布

i)条件密度函数

对于给定的x，若$f_X(x) > 0$，则称

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \qquad -\infty < y < +\infty$$

为Y关于$X = x$的**条件密度函数**.

类似地，若$f_Y(y) > 0$，则称

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \qquad -\infty < x < +\infty$$

为X关于$Y = y$的**条件密度函数**.

注：上式为条件概率的密度函数版本。

七、多维随机变量及其分布*

ii)条件分布函数

对一切使$f_Y(y)>0$的y，给定$Y=y$条件下X的条件分布函数为

$$F(x|y)=\int_{-\infty}^{x} f(u|y)du=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du.$$

类似对一切使$f_X(x)>0$的x，给定$X=x$条件下Y的条件分布函数分别为

$$F(y|x)=\int_{-\infty}^{y} f(v|x)dv=\int_{-\infty}^{y}\frac{f(x,v)}{f_X(x)}dv.$$

注：条件密度函数同样满足密度函数的所有性质。

七、多维随机变量及其分布*

5.多项分布

在n次独立重复试验中，如果每次试验有r个可能结果:$A_1, A_2, \ldots, A_r$,且每次试验中A_i发生的概率为 $p_i = P(A_i), i = 1,2,\ldots,r. p_1 + p_2 + \cdots + p_r = 1$. 记$X_i$为$n$次独立重复试验中$A_i$出现的次数，$i = 1,2,\cdots,r$.则$(X_1, X_2, \ldots, X_r)$服从**多项分布**，又称**$r$项分布**，记为$M(n, p_1, p_2, \cdots, p_r)$，其联合分布列为

$$P(X_1 = n_1, X_2 = n_2, \cdots, X_r = n_r) = \frac{n!}{n_1!\, n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

其中$n = n_1 + n_2 + \cdots + n_r$.

当$r = 2$时，即为二项分布.

七、多维随机变量及其分布*

6.二元均匀分布

设(X,Y)的联合密度函数为

$$f(x,\ y)=\begin{cases}\dfrac{1}{S(D)}, & (x,y)\in D\\ 0, & \text{其他}\end{cases}$$

其中D为一平面有界区域，$S(D)$为D的面积，则称(X,Y)服从二元均匀分布，常记为$(X,Y)\sim U(D)$.

注：本课件采用“多维”和“多元”表达同一概念。

七、多维随机变量及其分布

7.二元正态分布

如果二维随机变量(X,Y)的联合密度函数*为

$$f(x,y) = \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

$$-\infty < x,y < +\infty$$

则称(X,Y)服从二元正态分布，记为$(X,Y)\sim BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$.

其中$E(X)=\mu_1$， $E(Y)=\mu_2$;$Var(X)=\sigma_1^2$， $Var(Y)=\sigma_2^2$; $-1\le\rho\le 1$.

七、多维随机变量及其分布*

例18 （2010数1,3）设二维随机变量（X,Y）的概率密度为

$f(x,y)=Ae^{-2x^2+2xy-y^2}$， $-\infty<x<+\infty$， $-\infty<y<+\infty$

求常数A以及条件概率密度$f_{Y|X}(y|x)$.

解： 设由联合概率密度的性质有

$$1=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\,dxdy=A\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-2x^2+2xy-y^2}\,dxdy$$

$$=A\int_{-\infty}^{+\infty}e^{-x^2}dx\int_{-\infty}^{+\infty}e^{-(y-x)^2}\,d(y-x)$$

$$=A\times\sqrt{\pi}\times\sqrt{\pi}=A\pi,$$

得常数$A=\frac{1}{\pi}$，即$f(x,y)=\frac{1}{\pi}e^{-2x^2+2xy-y^2}$.

七、多维随机变量及其分布*

则$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,dy=\int_{-\infty}^{+\infty}\frac{1}{\pi}e^{-2x^2+2xy-y^2}\,dy$

$$=\frac{1}{\pi}e^{-x^2}\int_{-\infty}^{+\infty} e^{-(y-x)^2}\,d(y-x)=\frac{1}{\sqrt{\pi}}e^{-x^2},$$

因此

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{1}{\sqrt{\pi}}e^{-x^2+2xy-y^2},\quad -\infty<x<+\infty,$$

$$-\infty<y<+\infty.$$

注：本题充分利用积分$\int_{-\infty}^{+\infty} e^{-x^2}\,dx=\sqrt{\pi}$简化计算，思考该结论如何得到。

七、多维随机变量及其分布

- **例19** （2007数1,3）设二维随机变量（X,Y）的概率密度为

$$f(x,y)=\begin{cases}2-x-y, & 0<x<1,\ 0<y<1,\\ 0, & \text{其他},\end{cases}$$

（1）求$P\{X>2Y\}$;

（2）求$Z=X+Y$的概率密度$f_Z(z)$.

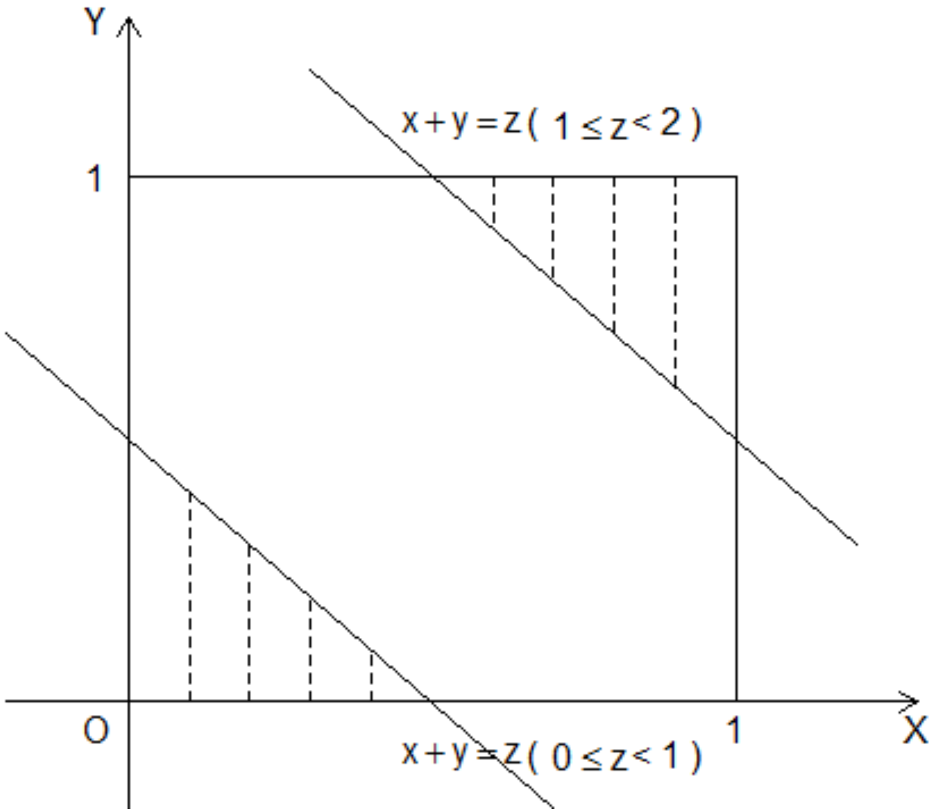
- **解：** (1)$P\{X>2Y\}=\iint_{x>2y} f(x,y)\,dxdy$

$$=\int_0^{\frac{1}{2}} dy\int_{2y}^{1}(2-x-y)\,dx=\frac{7}{24}.$$

七、多维随机变量及其分布

(2)先求Z的分布函数：

$$F_Z(z) = P\{X + Y \le z\} = \iint_{x+y\le z} f(x,y)\,dxdy.$$



七、多维随机变量及其分布

当$z<0$时，$F_Z(z)=0$；

当$0\le z<1$时，$F_Z(z)=\iint_{x+y\le z} f(x,y)\,dxdy$

$$=\int_0^z dy\int_0^{z-y}(2-x-y)\,dx=z^2-\frac{1}{3}z^3;$$

当$1\le z<2$时，$F_Z(z)=1-\iint_{x+y>z} f(x,y)\,dxdy$

$$=1-\int_{z-1}^1 dy\int_{z-y}^1(2-x-y)\,dx=1-\frac{1}{3}(2-z)^3\ ;$$

当$z\ge 2$时，$F_Z(z)=1.$

故$Z=X+Y$的概率密度

$$f_Z(z)=F'_Z(z)=\begin{cases}2z-z^2, & 0<z<1\\ (2-z)^2, & 1\le z<2\\ 0, & \text{其他.}\end{cases}$$

八、边际分布与随机变量的独立性

1.边际分布函数 若二维随机变量(X,Y)的联合分布函数为$F(X,Y)$，则称

$$F_X(x)=F(x,+\infty)=\lim_{y\to+\infty}F(x,y),-\infty<x<+\infty$$

为X的边际分布，称

$$F_Y(y)=F(+\infty,y)=\lim_{x\to+\infty}F(x,y),-\infty<y<+\infty$$

为Y的边际分布.

2.边际分布列 若二维离散随机变量(X,Y)的联合分布列为$\{p_{ij}\}$，则称$p_{i\cdot}=\sum_{j=1}^{+\infty}p_{ij}$，$i=1,2,\ldots$为$X$的边际分布列，称$p_{\cdot j}=\sum_{i=1}^{+\infty}p_{ij}$，$j=1,2,\ldots$为$Y$的边际分布列.

注：边际分布也称作边缘分布(Marginal distribution)。

八、边际分布与随机变量的独立性

3.边际密度函数 若二维连续随机变量(X,Y)的联合分布函数为$f(x,y)$，则称$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy,-\infty<x<+\infty$为$X$的边际密度函数，称

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx,-\infty<y<+\infty$$

为Y的边际密度函数.

4.注解

- 由高维联合分布可以获得低维的边际分布，反之不一定.
- 不同的联合分布可以有相同边际分布.
- 多项分布的边际分布仍为低维的多项分布或二项分布.
- 二维正态分布的边际分布为一维正态分布.

八、边际分布与随机变量的独立性

5.随机变量间的独立性

(1)设n维随机变量$(X_1, X_2, \dots, X_n)$的联合分布函数为$F(x_1, x_2, \cdots, x_n)$，且$F_{X_i}(x_i)$为X_i的边际分布函数.如果对任意n个实数$x_1, x_2, \cdots, x_n$，有

$$F(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

则称$X_1, X_2, \dots, X_n$相互独立.否则称$X_1, X_2, \dots, X_n$不相互独立，或相关.

(2)设n维离散随机变量$(X_1, X_2, \dots, X_n)$的联合分布列为$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$，且$P(X_i = x_i)$为$X_i$的边际分布列，$i = 1,2, \dots, n$.

八、边际分布与随机变量的独立性

如果对其任意n个取值$x_1, x_2, \cdots, x_n$，有

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

则称$X_1, X_2, \ldots, X_n$相互独立的.否则称$X_1, X_2, \ldots, X_n$不相互独立，或相关.

(3)设n维连续随机变量$(X_1, X_2, \ldots, X_n)$的联合密度函数为$f(x_1, x_2, \cdots, x_n)$，且$f_{X_i}(x_i)$为X_i的边际密度函数.如果对任意n个实数$x_1, x_2, \cdots, x_n$，有

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

则称$X_1, X_2, \ldots, X_n$相互独立的.否则称$X_1, X_2, \ldots, X_n$不相互独立，或相依.

八、边际分布与随机变量的独立性

❖ **例20** 设二维随机变量（X,Y）的概率密度为

$$f(x,y)=\begin{cases}\frac{k}{2}xe^{-x(1+y)}, & x>0,y>0\\ 0, & \text{其他}\end{cases}$$

（1）求系数k；

（2）求（X,Y）关于X和关于Y 的边缘密度；

（3）问X和Y是否相互独立.

❖ **解：**（1）由联合概率密度的性质有

$$1=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\,dxdy=\frac{k}{2}\int_0^{+\infty}\int_0^{+\infty}xe^{-x(1+y)}\,dxdy$$

$$=\frac{k}{2}\int_0^{+\infty}xe^{-x}[\int_0^{+\infty}e^{-xy}dy]dx=\frac{k}{2},$$

得常数$k=2$.

八、边际分布与随机变量的独立性

(2)由(1)得 $f(x,y)=\begin{cases} xe^{-x(1+y)}, & x>0, y>0 \\ 0, & 其他 \end{cases}$

则 $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,dy=\int_0^{+\infty} xe^{-x(1+y)}\,dy$

$$=xe^{-x}\int_0^{+\infty} e^{-xy}\,dy = xe^{-x}\left[-\frac{1}{x}e^{-xy}\Bigg|_{y=0}^{y=+\infty}\right]=e^{-x}.$$

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\,dx=\int_0^{+\infty} xe^{-x(1+y)}\,dx$$

$$=\frac{-1}{1+y}\int_0^{+\infty} x\,de^{-x(1+y)}$$

$$=\frac{-1}{1+y}\left[xe^{-x(1+y)}\Bigg|_{x=0}^{x=+\infty}-\int_0^{+\infty} e^{-x(1+y)}\,dx\right]$$

$$=\frac{-1}{1+y}\left[0-\left(\frac{-1}{1+y}\right)e^{-x(1+y)}\Bigg|_{x=0}^{x=+\infty}\right]=\frac{1}{(1+y)^2}.$$

八、边际分布与随机变量的独立性

故$f_X(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & x \le 0 \end{cases}.$

$$f_Y(y) = \begin{cases} \frac{1}{(1+y)^2}, & y > 0 \\ 0, & y \le 0 \end{cases}$$

(3) $f(x,y) \ne f_X(x) f_Y(y)$，所以X与Y不独立.

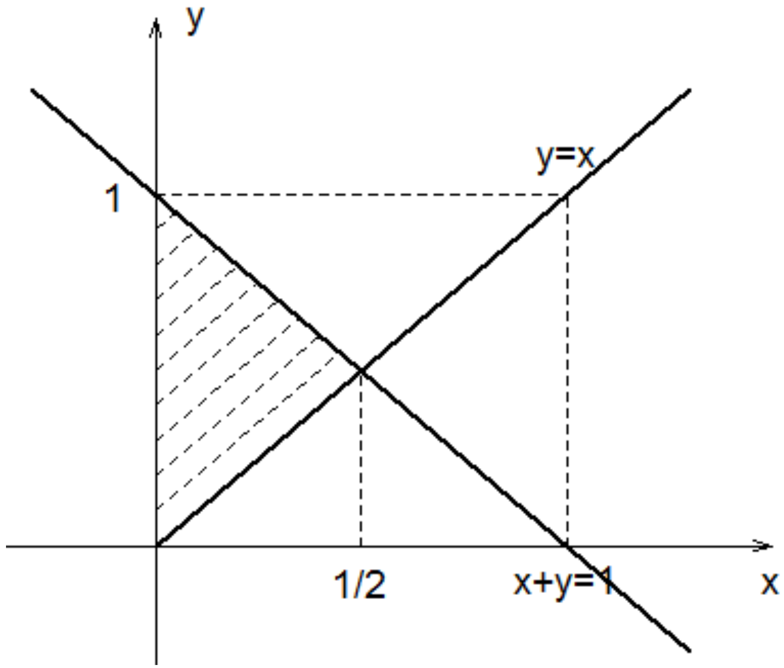
八、边际分布与随机变量的独立性

❖ **例21** 设(X,Y)的联合概率密度为

$$f(x,y)=\begin{cases}6x, & 0\le x\le y,\\ 0, & \text{其他}.\end{cases}$$

则$P\{X+Y\le 1\}=$______.

❖ **解：** $P\{X+Y\le 1\}=\int_0^{\frac{1}{2}}\int_x^{1-x}6x\,dydx=\frac{1}{4}$，故应填$\frac{1}{4}$.



八、边际分布与随机变量的独立性

- **例22** 设$X_1, X_2, \cdots, X_n$是独立同分布的正值随机变量，证明：

$$E\left(\frac{X_1+\cdots+X_k}{X_1+\cdots+X_n}\right)=\frac{k}{n}.$$

- **证：** 记$Y_j=X_j/\sum_{i=1}^n X_i$，则诸Y_j同分布，且由$|Y_j|\le 1$，知$E(Y_j)$存在且相等，$j=1,2,\cdots,n$.

又因为$1=E\left(\sum_{j=1}^n Y_j\right)=nE(Y_1)$,

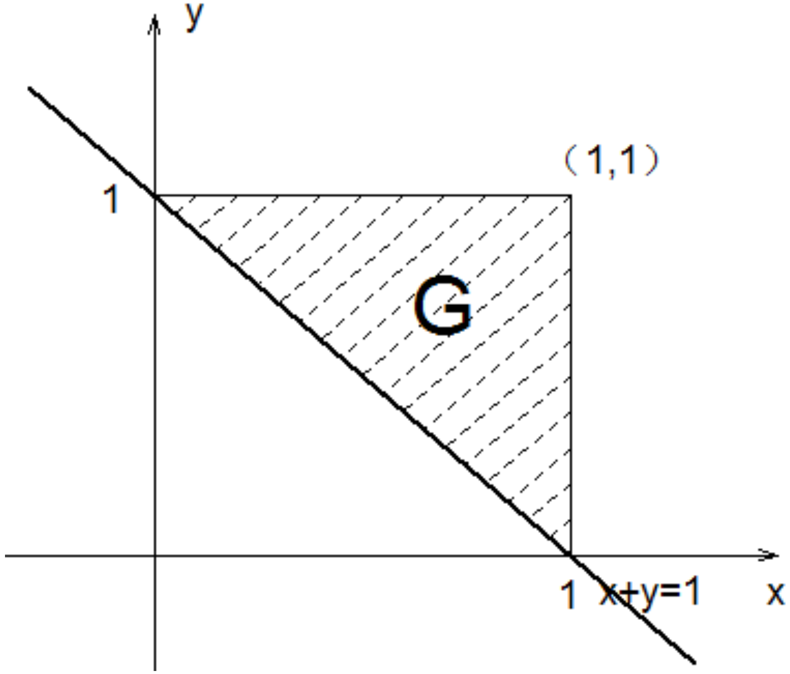
所以有$E\left(Y_j\right)=\frac{1}{n}, j=1,2,\ldots,n$.由此得

$$E\left(\frac{X_1+\cdots+X_k}{X_1+\cdots+X_n}\right)=E(Y_1+\cdots+Y_k)=kE(Y_1)=\frac{k}{n}.$$

八、边际分布与随机变量的独立性

❖ **例23** （2001数4）设随机变量X和Y的联合分布在以点（0,1），（1,0），（1,1）为顶点的三角形区域上服从均匀分布，试求$U = X + Y$的方差.

❖ **解：** 令三角形区域 $G = \{(x, y)|0 \le x \le 1, 1 - x \le y \le 1\}$，显然$G$的面积为$S_G = \frac{1}{2}$.



八、边际分布与随机变量的独立性

由均匀分布知（X, Y）的联合密度为

$$f(x,\ y)=\begin{cases}2, & (x,y)\ \in G\\ 0, & 其他\end{cases}$$

$$E(X+Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f(x,y)dxdy$$

$$=\iint_G 2(x+y)dxdy=2\int_0^1 dx\int_{1-x}^1 (x+y)dy=\frac{4}{3}.$$

$$E(X+Y)^2=\iint_G 2(x+y)^2dxdy$$

$$=2\int_0^1 dx\int_{1-x}^1 (x+y)^2dy=\frac{11}{6}.$$

因此

$$Var(U)=E(X+Y)^2-E^2(X+Y)=\frac{11}{6}-\left(\frac{4}{3}\right)^2=\frac{1}{18}.$$

九、多维随机变量函数的分布

1.最大值、最小值分布 （极值分布）

设$(X_1,X_2,\cdots,X_n)$是相互独立、同分布的n维连续随机变量，其共同的密度函数和分布函数分别为$f(x)$和$F(x)$，记

$$Y=\min\{X_1,X_2,\cdots,X_n\},Z=\max\{X_1,X_2,\cdots,X_n\}.$$

则

$$F_Y(y)=1-[1-F(y)]^n;$$
$$f_Y(y)=n[1-F(y)]^{n-1}f(y);$$
$$F_Z(z)=[F(z)]^n;$$
$$f_Z(z)=n[F(z)]^{n-1}f(z).$$

九、多维随机变量函数的分布*

2.变量变换法 （略） 若变换$\begin{cases} u = g_1(x,y) \\ v = g_2(x,y) \end{cases}$，存在唯一的反函数

$\begin{cases} x = x(u,v), \\ y = y(u,v), \end{cases}$且变换的雅可比行列式

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = \left(\frac{\partial(u,v)}{\partial(x,y)}\right)^{-1} = \left(\begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}\right)^{-1} \neq 0.$$

则二维连续随机变量(X,Y)的函数$\begin{cases} U = g_1(X,Y) \\ V = g_2(X,Y) \end{cases}$的联合密度函数

为$f_{U,V}(u,v) = f_{X,Y}\big(x(u,v), y(u,v)\big)|J|$.

九、多维随机变量函数的分布*

3.和的分布 卷积公式用于求随机变量$Z=X+Y$的分布.

(1)离散场合的卷积公式 $Z=X+Y$的分布列为

$$P(Z=z_k)=\sum_{i=1}^{+\infty}P(X=x_i,Y=z_k-x_i)$$

$$或=\sum_{j=1}^{+\infty}P\left(X=z_k-y_j,Y=y_j\right).$$

当X与Y独立时,

$$P(Z=z_k)=\sum_{i=1}^{+\infty}P(X=x_i)P(Y=z_k-x_i)$$

$$或=\sum_{j=1}^{+\infty}P\left(X=z_k-y_j\right)P(Y=y_j).$$

其中诸x_i为X的取值，诸y_j为Y的取值.

九、多维随机变量函数的分布*

(2)连续场合的卷积公式 $Z=X+Y$的密度函数为

$$f_Z(z)=\int_{-\infty}^{+\infty} f_{X,Y}(x,z-x)dx$$

$$或=\int_{-\infty}^{+\infty} f_{X,Y}(z-y,y)dy.$$

当X与Y独立时,

$$f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx$$

$$或=\int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy.$$

九、多维随机变量函数的分布

4.分布的可加性 若同一类分布的独立随机变量和的分布仍属于此类分布，则称此类分布具有可加性.以下一些常用分布具有可加性：

(1)二项分布：若$X\sim b(n,p)$，$Y\sim b(m,p)$，且X与Y独立，则$Z=X+Y\sim b(n+m,p)$.注意这里两个二项分布中的参数p要相同.

(2)泊松分布：若$X\sim P(\lambda_1)$，$Y\sim P(\lambda_2)$，且X与Y独立，则$Z=X+Y\sim P(\lambda_1+\lambda_2)$.

九、多维随机变量函数的分布

(3)正态分布：若$X\sim N(\mu_1,\sigma_1^2)$， $Y\sim N(\mu_2,\sigma_2^2)$，且$X$与$Y$独立，则$Z=X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

(4)伽玛分布：若$X\sim Ga(\alpha_1,\lambda)$，$Y\sim Ga(\alpha_2,\lambda)$，且$X$与$Y$独立，则$Z=X+Y\sim Ga(\alpha_1+\alpha_2,\lambda)$.注：这里两个伽玛分布中的尺度参数$\lambda$要相同.

(5)χ^2分布：若$X\sim\chi^2(n_1)$，$Y\sim\chi^2(n_2)$，且X与Y独立，则$Z=X+Y\sim\chi^2(n_1+n_2)$.

注：可加性有助于计算更为复杂的抽样分布，详见后续章节。

九、多维随机变量函数的分布

5.一些结论

(1)设$X_1, X_2, \ldots, X_n$独立同分布，都服从二点分布$b(1, p)$，则 $X_1 + X_2 + \cdots + X_n$服从二项分布$b(n, p)$.

(2)设$X_1, X_2, \ldots, X_n$独立同分布，都服从几何分布$Ge(p)$，则 $X_1 + X_2 + \cdots + X_n$服从负二项分布$Nb(n, p)$.

(3)设$X_1, X_2, \ldots, X_n$独立同分布，都服从指数分布$Exp(\lambda)$，则 $X_1 + X_2 + \cdots + X_n$服从伽玛分布$Ga(n, \lambda)$.

九、多维随机变量函数的分布

❖ **例24** 某种商品一周的需求量是一个随机变量，其密度函数为

$$f_1(t)=\begin{cases} te^{-t}, & t>0, \\ 0, & t\le 0. \end{cases}$$

设各周的需求量是相互独立的，试求

(1)两周需求量的密度函数$f_2(x)$；

(2)三周需求量的密度函数$f_3(x)$.

九、多维随机变量函数的分布

❖ **解：** 记X_i为第i周的需求量，$i=1,2,3$.根据题意知X_1, X_2, X_3相互独立，且密度函数都为$f_1(t)$. X_i服从伽玛分布$Ga\ (2,1)$，所以由伽玛分布的可加性知

(1)$X_1+X_2 \sim Ga(4,1)$，其密度函数为

$$f_2(x)=\frac{x^3}{6}e^{-x}, \qquad x>0.$$

(2)$X_1+X_2+X_3 \sim Ga(6,1)$，其密度函数为

$$f_3(x)=\frac{1}{\Gamma(6)}x^5e^{-x}=\frac{1}{120}x^5e^{-x}, \qquad x>0.$$

十、分布的其他数字特征

1. 二元随机变量函数的期望

设(X,Y)是二维随机变量，$z=g(x,y)$为实变量x,y的二元函数

(1)若(X,Y)是离散型随机变量，其分布律为：

$P\{X=x_i,Y=y_i\}=p_{ij}, i,j=1,2,\cdots$，且

$\sum_i\sum_j g(x_i,y_i)\,p_{ij}$绝对收敛，则

$$E(Z)=E[g(X,Y)]=\sum_i\sum_j g(x_i,y_i)\,p_{ij}.$$

(2)若(X,Y)是连续型随机变量，其密度函数为

$f(x,y)$，且$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)dxdy$绝对收敛,则

$$E(Z)=E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)dxdy.$$

十、分布的其他数字特征

2.k阶矩

(1)称$\mu_k = E(X^k)$为X的k阶**原点矩**.一阶原点矩就是数学期望;

(2)称$v_k = E\big(X - E(X)\big)^k$为$X$的$k$阶**中心矩**.二阶中心矩就是方差;

(3)前k阶中心矩可用原点矩表示，如

$$v_1 = 0;$$

$$v_2 = \mu_2 - \mu_1^2;$$

$$v_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3;$$

$$v_4 = \mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4.$$

3.变异系数 称比值$C_v(X) = \frac{\sqrt{Var(X)}}{E(X)}$为$X$的变异系数.变异系数是一个无量纲的量.

十、分布的其他数字特征

4.协方差 若$E[(X-E(X))(Y-E(Y))]$存在，则称

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

为X与Y的协方差.

- 当$Cov(X,Y)>0$时，称X与Y正相关，即X与Y有同时增加或同时减少的倾向.
- 当$Cov(X,Y)<0$时，称X与Y负相关，即有X增加而Y减少的倾向，或有X减少而Y增加的倾向.
- 当$Cov(X,Y)=0$时，称X与Y不相关.

协方差的性质:

(1) $Cov(X,Y)=Cov(Y,X)$;

(2) $Cov(X,Y)=E(XY)-E(X)E(Y)$;

十、分布的其他数字特征

(3)若X与Y相互独立，则$Cov(X,Y)=0$，反之不然;

(4) $Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)$；

(5)对任意的常数a，有$Cov(X,a)=0$；

(6)对任意的常数a，b，有

$$Cov(aX,bY)=abCov(X,Y)；$$

(7)对任意二维随机变量(X,Y)，有

$$Var(X\pm Y)=Var(X)+Var(Y)\pm 2Cov(X,Y).$$

对任意n个随机变量$X_1,X_2,\cdots,X_n$，有

$$Var\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}Var(X_i)+2\sum_{i=1}^{n}\sum_{j=1}^{i-1}Cov(X_i,X_j).$$

十、分布的其他数字特征

5.相关系数 设(X,Y)是一个二维随机变量，且$Var(X)>0, Var(Y)>0$.则称

$$Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

为X与Y的（线性）相关系数.

十、分布的其他数字特征

相关系数的性质:

(1)$-1 \le Corr(X,Y) \le 1$.

(2) $Corr(X,Y)$与$Cov(X,Y)$同号，或同为零.

(3) $Corr(X,Y) = Cov(X^*,Y^*)$，其中X^*,Y^*分别为X，Y的标准化随机变量.

(4) $Corr(X,Y) = \pm 1$的充要条件是X与Y间几乎处处有线性关系，即存在$a(a \ne 0)$与b，使得

$P(Y = aX + b) = 1$.其中当$Corr(X,Y) = 1$时，有$a > 0$；当$Corr(X,Y) = -1$时，有$a < 0$.

(5)在二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$场合，不相关与独立是等价的.

思政案例：相关性不一定等于因果性

在日常生活和数据分析中，我们可以得到大量相关性的结论，例如：

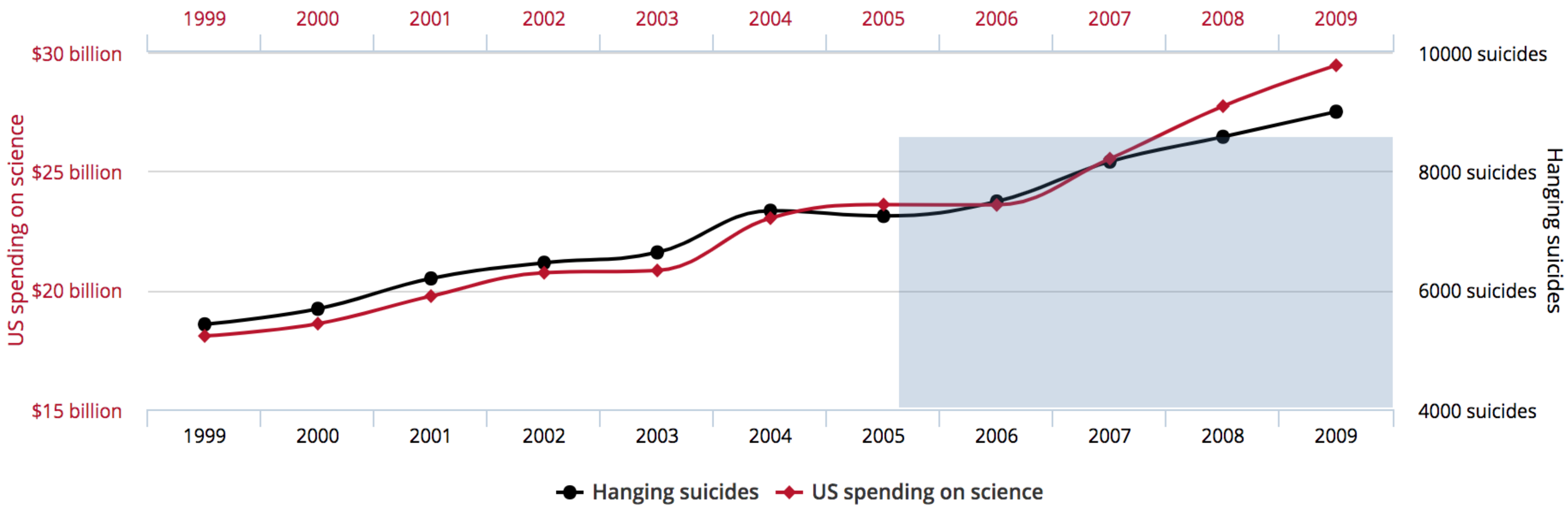
- 只要努力，就能成功
- 只要到了下班时间出公司大门，天就一定黑了
- 深圳交警表示，天秤、处女、天蝎座的人更喜欢违章
- 世界上不吃猪肉的人群中，人自爆的概率最大
- 据观察统计，消防车数量越多的火灾中，伤亡人数越多

此类结论的依据也似乎有很强的依据，即所谓的历史经验，甚至是大量的真实数据支持，我们通过各种模型分析得到以上种种结论。

但是这里面存在一个巨大的疑问就是，**“相关性一定等于因果性吗？”**让我们来看几个反常识的例子。

思政案例：相关性不一定等于因果性

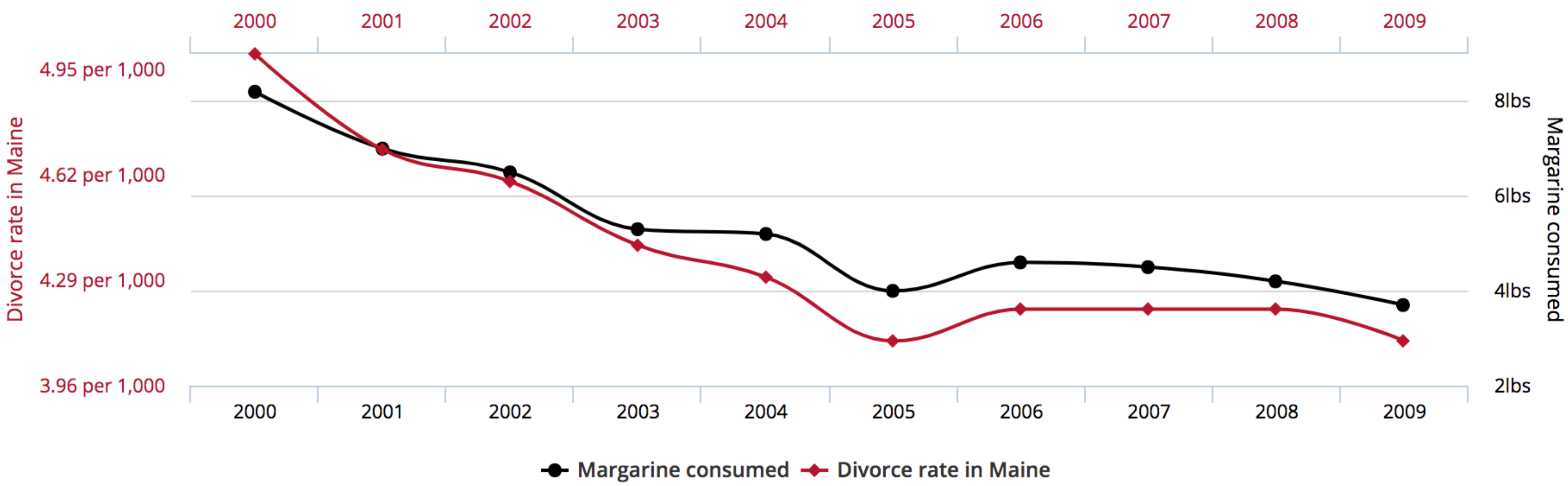
1）美国在科学、空间和技术上的开支 && 绞死、勒死和窒息的后果关联性



correlation=99.79%

思政案例：相关性不一定等于因果性

2）缅因州的离婚率与人均人造黄油消费量相关



correlation=99.26%

思政案例：相关性不一定等于因果性

3) 深圳交警表示，天秤、处女、天蝎座的人更喜欢违章

“我们发现上周天秤座、处女座、天蝎座携手违法数量居前三名。”吴凯峰告诉记者，通过从8月3日至8月9日行人、非机动车违法人的星座分类对比，发现天秤座、处女座、天蝎座位列违法量前三名（三个星座刚好也相连，从8月23日至11月22日），分别占违法总数的10.5%、9.63%和9.0%。金牛座、白羊座最“乖”，分别以673、661排在最后两位。
来源：深圳交警微门户

显然，深圳交警不会有意去做数据造假这种事情，但是星座决定论又显然违反了我们的常识。我们来分析一下其中的原因。

首先，报道里的数据分析存在错误：没有考虑星座在当地人口中的总体分布，而直接拿观测到的样本绝对数量进行比较。实际上正是因为这三个星座的人口，占当地总人口的比例较高（9-11月为北半球低纬度南部地区生育高峰），造成被抓到违章的人数较多。如果这个调查在北方，估计躺枪的会是双子、巨蟹和狮子座（6-8月）。

所以，星座与违章确实有关，源于星座人口比例，而不是星座“乖不乖”。

思政案例：相关性不一定等于因果性

上述讨论的这种"荒谬的相关性"还有其他大量的例子。显然，对上述讨论的情况，如果使用相关性分析，Peason相关分析等技术，很容易得出一个结论就是：

它们是彼此强烈相关的，看到其中任何一个事物发生，就有很大概率另一个事物也会伴随发生.

但显然，这个结论和我们的常识严重地相悖，为什么号称有坚实理论基础的统计算法，在这里似乎产生了一个非常荒谬的结论呢？难道统计学习是伪科学吗？

答案是否定的！真实的情况是统计学习没有问题，出问题的是使用它和理解它产生结论的方式

思政案例：相关性不一定等于因果性

伪相关性

假设两个事件A和B，最简单的因果性例子就是A导致了B，如下图所示：

因果性（causality）会导致相关性（correlation），反之则不成立。换言之，因果性是一个比相关性更强的概念。因果性是一种特殊的相关性。

统计学上把两个或者多个看起来高度相关，但却不存在因果性的现象叫做伪相关性（spurious correlation）。

思政案例：相关性不一定等于因果性

伪相关性的成因

- ✓ 存在其他变量（common response variables）同时对这两个事件造成（直接或者间接）影响，但我们无法观测到（意识到）它的存在。例如：

与冬天相比，夏天游泳的人变多且溺亡人数上升，同时吃冰淇淋的人数也增多，因此是否可以得出结论："冰淇淋"会导致"溺亡"？

这种听起来荒谬的结论就是因为错误地把相关性当成了因果性。我们没有观测到"夏天温度升高"这个因素同时会影响"游泳溺亡人数"和"吃冰淇淋人数"，而错误地认为两者间有直接的因果关系，相关性不是因果性的充要条件。

- ✓ 纯属巧合，尤其是处理小样本时

前面谈到的一张图，美国从1999-2009年在科技领域的支出，黑线是通过上吊、窒息等方式自杀的人数。我们发现其线性相关系数（[r=0.9978]）。r的取值范围为-1到1，越接近1越倾向于强正相关。显然，我们不会认为这两者之间存在因果性。且仔细观察后我们发现样本量仅仅是11个数据点，因此是巧合的可能性更大。

十、分布的其他数字特征

6.随机向量的数学期望与协方差阵 记n维随机向量为$\boldsymbol{X}=(X_1,X_2,\cdots,X_n)'$，以下假设所涉及的数学期望、方差、协方差均存在.

(1)随机向量$\boldsymbol{X}$的数学期望向量为

$$E(\boldsymbol{X})=(E(X_1),E(X_2),\cdots E(X_n))'.$$

(2)随机向量$\boldsymbol{X}$的协方差阵为

$$E\left[(\boldsymbol{X}-E(\boldsymbol{X}))(\boldsymbol{X}-E(\boldsymbol{X}))'\right]$$

$$=\begin{bmatrix} Var(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & Var(X_2) & \cdots & Cov(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n,X_1) & Cov(X_n,X_2) & \cdots & Var(X_n) \end{bmatrix}$$

也记以上的协方差阵为$Cov(\boldsymbol{X})$，或记成$\boldsymbol{\Sigma}$.

十、分布的其他数字特征

(3)随机向量$\boldsymbol{X}$的协方差阵$Cov(\boldsymbol{X}) = \left(Cov(X_i, X_j)\right)_{n\times n}$是一个对称的非负定矩阵.

(4)n元正态分布　设n维随机向量$\boldsymbol{X} = (X_1, X_2, \cdots, X_n)'$的协方阵为$\boldsymbol{\Sigma} = Cov(\boldsymbol{X})$，数学期望向量为$\boldsymbol{a} = (a_1, a_2, \cdots, a_n)'$.又记$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)'$，则由密度函数

$$p(x_1, x_2, \cdots, x_n) = p(\boldsymbol{x})$$
$$= \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{a})\}$$

定义的分布称为n元正态分布，记为$\boldsymbol{X}\sim N(\boldsymbol{a}, \boldsymbol{\Sigma})$.

十、分布的其他数字特征

- **例25-1** （2012数1）将长度为$1m$的木棒随机地截成两段，则两段长度的相关系数为______.
- **解：** 设两段长度分别为X,Y，则由已知条件有$X+Y=1$，则$Y=-X+1$，所以X,Y的相关系数为-1.

- **例25-2**设二维随机变量（X,Y）服从$N(\mu,\mu;\sigma^2,\sigma^2;0)$，则$E(XY^2)=$______.
- **解：** 由题可得，$Cov(X,Y)=0$.又，**在二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$场合，不相关与独立是等价的**.因此$X$与$Y$独立.故，

$$E(XY^2)=E(X)E(Y^2)=\mu(\mu^2+\sigma^2)$$

十、分布的其他数字特征

❖ **例26** 设二维随机变量(X,Y)服从单位圆内的均匀分布，其联合密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{\pi}, x^2+y^2<1,\\ 0, x^2+y^2\geq 1.\end{cases}$$

试证：X与Y不独立且X与Y不相关.

十、分布的其他数字特征

- **证：** 先求边际密度函数$p_X(x)$和$p_Y(y)$.

$$f_X(x)=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy=\frac{2}{\pi}\sqrt{1-x^2},-1<x<1,$$

$$f_Y(y)=\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}\frac{1}{\pi}dx=\frac{2}{\pi}\sqrt{1-y^2},-1<y<1.$$

所以由$f(x,y)\neq f_X(x)f_Y(y)$，知X与Y不独立.

又因为$f_X(x)$和$f_Y(y)$在对称区间上是偶函数，故$E(X)=E(Y)=0$，从而

$$Cov(X,Y)=E(XY)=\frac{1}{\pi}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}xydydx=0,$$

所以X与Y不相关.

注：不相关并不意味着独立。

十、分布的其他数字特征

- **例27**（2012数1，3）设二维离散型随机变量(X,Y)的概率分布为

X/Y	0	1	2
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	0	$\frac{1}{3}$	0
2	$\frac{1}{12}$	0	$\frac{1}{12}$

（1）求$P\{X=2Y\}$；（2）求$Cov(X-Y,Y)$．

- **解：**（1）$P\{X=2Y\}=P\{X=0,Y=0\}+$

$$P\{X=2,Y=1\}=\frac{1}{4}+0=\frac{1}{4}.$$

（2）易得X,Y的边际分布律为

十、分布的其他数字特征

X	0	1	2
P	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

Y	0	1	2
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

又XY的所有可能取值为0,1,4，且分布律为

XY	0	1	4
P	$\frac{7}{12}$	$\frac{1}{3}$	$\frac{1}{12}$

于是$E(X)=\frac{2}{3}$，$E(Y)=1$，$E(Y^2)=\frac{5}{3}$，$E(XY)=\frac{2}{3}$，故

$$Cov(X-Y,Y)=Cov(X,Y)-Cov(Y,Y)$$

$$=E(XY)-E(X)E(Y)-[E(Y^2)-E^2(Y)]$$

$$=\frac{2}{3}-\frac{2}{3}\times 1-\left[\frac{5}{3}-1^2\right]=-\frac{2}{3}.$$

十一、条件期望*

1.条件数学期望 条件分布的数学期望（若存在）称为条件期望，其定义如下：

$$E(X|Y=y)=\begin{cases}\sum_i x_i p(X=x_i|Y=y), & (X,Y)\text{为二维离散随机变量,}\\ \int_{-\infty}^{+\infty} xf(x|y)dx, & (X,Y)\text{为二维连续随机变量,}\end{cases}$$

$$E(Y|X=x)=\begin{cases}\sum_j y_j p(Y=y_j|X=x), & (X,Y)\text{为二维离散随机变量,}\\ \int_{-\infty}^{+\infty} yf(y|x)dy, & (X,Y)\text{为二维连续随机变量,}\end{cases}$$

十一、条件期望*

- 条件期望具有数学期望的一切性质.
- 条件期望$E(X|Y=y)$是y的函数，记为$g(y)$，它是另一个随机变量$g(Y)=E(X|Y)$的取值. $E(X|Y)$与$E(X|Y=y)$虽都称为条件期望，但含义不同.前者是特定的随机变量，后者是其取值.

2.全期望公式 设(X,Y)是二维随机变量，若$E(X)$存在,则$E(X)=E[E(X|Y)]$.注意：该公式中里面的期望是用条件分布$f(x|y)$计算的，外面的期望是用y的分布$f(y)$计算的.

十一、条件期望*

3.二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$的两个条件分布仍是正态分布，即

$$X|Y=y\sim N\left(g_1(y),\sigma_1^2(1-\rho^2)\right),$$

其中$g_1(y)=E(X|Y=y)=\mu_1+\frac{\rho\sigma_1}{\sigma_2}(y-\mu_2).$

$$Y|X=x\sim N\left(g_2(x),\sigma_2^2(1-\rho^2)\right),$$

其中$g_2(x)=E(Y|X=x)=\mu_2+\frac{\rho\sigma_2}{\sigma_1}(x-\mu_1).$

可见，在二维正态分布中，一个变量的条件期望是另一个变量取值的线性函数，常称为一元线性回归方程.

十一、条件期望*

❖ **例28** 设二维连续随机变量(X,Y)的联合密度函数为

$$f(x,y)=\begin{cases} x+y, & 0<x,y<1, \\ 0\,, & \text{其他}. \end{cases}$$

试求$E(X|Y=0.5)$.

❖ **解：** 先求条件密度函数$f(x|y)$.当$0<y<1$时，$f_Y(y)=\int_0^1(x+y)dx=0.5+y$.所以

$$f(x|y=0.5)=\begin{cases} x+0.5, & 0<x<1, \\ 0\,, & \text{其他}. \end{cases}$$

由此得

$$E(X|Y=0.5)=\int_0^1 x\cdot(x+0.5)dx=\frac{7}{12}.$$

十二、大数定律与中心极限定理

1.随机变量序列的收敛性

(1)依概率收敛　设$\{X_n\}$为一随机变量序列，X为一随机变量.如果对任意的$\varepsilon > 0$，有

$$\lim_{n\to+\infty} P\{|X_n - X| < \varepsilon\} = 1,$$

则称$\{X_n\}$依概率收敛于X，记作$X_n \xrightarrow{\mathrm{P}} X$.

(2)依概率收敛与服从大数定律间的关系　设$\{X_n\}$为一随机变量序列，记$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i, E(Y_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$.则$\{X_n\}$服从大数定律等价于$Y_n - E(Y_n) \xrightarrow{\mathrm{P}} 0$.

(3)依概率收敛的四则运算　如果$X_n \xrightarrow{\mathrm{P}} a, Y_n \xrightarrow{\mathrm{P}} $b,则有

十二、大数定律与中心极限定理

i) $X_n \pm Y_n \xrightarrow{\mathrm{P}} \mathrm{a} \pm b$;

ii) $X_n \times Y_n \xrightarrow{\mathrm{P}} \mathrm{a} \times b$;

iii) $X_n / Y_n \xrightarrow{\mathrm{P}} \mathrm{a}/b (b \neq 0)$.

(4)按分布收敛、弱收敛　设$\{F_n(x)\}$是随机变量序列$\{X_n\}$的分布函数列，$F(x)$为X的分布函数.若对$F(x)$的任一连续点x，都有$\lim\limits_{n\to\infty} F_n(x) = F(x)$，则称$\{F_n(x)\}$弱收敛于$F(x)$，记作$F_n(x) \xrightarrow{\mathrm{W}} \mathrm{F}(x)$.也称$\{X_n\}$按分布收敛于$X$，记作$X_n \xrightarrow{\mathrm{L}} X$.

(5)依概率收敛与按分布收敛间的关系

i) $X_n \xrightarrow{\mathrm{P}} X \Rightarrow X_n \xrightarrow{\mathrm{L}} X$.

ii) $X_n \xrightarrow{\mathrm{P}} c \Leftrightarrow X_n \xrightarrow{\mathrm{L}} c$. (其中c为常数).

十二、大数定律与中心极限定理

2.大数定律

(1)随机变量序列$\{X_n\}$服从大数定律 设$\{X_n\}$为随机变量序列，若对任意的$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right\} = 1.$$

则称$\{X_n\}$服从大数定律.

(2)伯努利大数定律 设μ_n为n重伯努利试验中事件A发生的次数，p为每次试验中A出现的概率，则对任意的$\varepsilon > 0$，有

$$\lim_{n\to\infty}\left\{\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right\} = 1,$$

十二、大数定律与中心极限定理

这是第一个大数定律，它表明：事件发生的频率是依概率收敛与该事件的概率，这就是“频率稳定于概率”的含义，也是“用频率去估计概率”的依据.

(3)切比雪夫大数定律 设$\{X_n\}$为一列两两不相关的随机变量序列，若每个X_i的方差存在，且有共同的上界，即$Var(X_i) \le c, \mathrm{i} = 1,2,\cdots$,则$\{X_n\}$服从大数定律.

切比雪夫不等式 设X为随机变量，且有有限方差，则对任意正数ε，有

$$P\{|X - E(X)| \ge \varepsilon\} \le \frac{Var(X)}{\varepsilon^2} \text{或} P\{|X - E(X)| < \varepsilon\} \ge 1 - \frac{Var(X)}{\varepsilon^2}.$$

十二、大数定律与中心极限定理

(4)马尔可夫大数定律　对随机变量序列$\{X_n\}$，若有

$$\frac{1}{n^2}Var(\sum_{i=1}^{n}X_i)\to 0(n\to\infty)$$

则$\{X_n\}$服从大数定律.上式被称为**马尔可夫条件.**

(5)辛钦大数定律　设$\{X_n\}$为一独立同分布的随机变量序列，若X_i的数学期望存在，则$\{X_n\}$服从大数定律.

十二、大数定律与中心极限定理

3.中心极限定理

(1)中心极限定理　研究独立随机变量和的极限分布在什么条件下为正态分布的命题.

(2)林德伯格-莱维中心极限定理　$\{X_n\}$设是独立同分布的随机变量序列，且$E(X_i)=\mu, Var(X_i)=\sigma^2>0$.记$Y_n^*=\frac{X_1+X_2+\cdots X_n-n\mu}{\sigma\sqrt{n}}$.

则对任意实数y，有

$$\lim_{n\to\infty} P(Y_n^*\le y)=\Phi(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{t^2}{2}}dt.$$

十二、大数定律与中心极限定理

(3)二项分布的正态近似

i)棣莫弗-拉普拉斯中心极限定理

设n重伯努利试验中，事件A在每次试验中出现的概率为$p(0<p<1)$,记μ_n为n此试验中事件A出现的次数，且记$Y_n^*=\frac{\mu_n-np}{\sqrt{np(1-p)}}$.

则对任意实数y，有

$$\lim_{n\to+\infty} P(Y_n^*\le y)=\Phi(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{t^2}{2}}dt.$$

十二、大数定律与中心极限定理*

ii)近似中的修正（略）

因为二项分布是离散分布，而正态分布是连续分布，所以用正态分布作为二项分布的近似计算中，作些修正可以提高精度.若$k_1 < k_2$均为整数，一般先作如下修正后再用正态近似.

$$\mathrm{P}(\mathrm{k}_1 \leq \mu_\mathrm{n} \leq k_2) = \mathrm{P}(\mathrm{k}_1 - 0.5 < \mu_\mathrm{n} < k_2 + 0.5)$$

$$= \Phi\left(\frac{k_2+0.5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1-0.5-np}{\sqrt{np(1-p)}}\right).$$

十二、大数定律与中心极限定理

- **例29** 某保险公司多年的统计资料表明，在索赔户中被盗索赔户占20%，以X表示在随意抽查的100个索赔户中因被盗向保险公司索赔的户数.

(1)写出X分布列;

(2)求被盗索赔户不少于14户且不多于30户的概率的近似值. *(P243 ex1)*

十二、大数定律与中心极限定理

❖ **解：** (1)X服从$n=100$，$p=0.2$的二项分布$b(100,0.2)$，即

$$P(X=k)=\binom{n}{k}0.2^k 0.8^{100-k}, k=0,1,2,\cdots n.$$

(2) 利用棣莫弗-拉普拉斯中心极限定理，有

$$\mathrm{P}(14 \le X \le 30)=\mathrm{P}(13.5<\mathrm{X}<30.5)$$

$$\approx \Phi\left(\frac{30.5-100\times0.2}{\sqrt{100\times0.2\times0.8}}\right)-\Phi\left(\frac{13.5-100\times0.2}{\sqrt{100\times0.2\times0.8}}\right)$$

$$=\Phi(2.625)-\Phi(-1.625)=\Phi(2.625)-1+\Phi(1.625)$$

$$=0.99565-1+0.948=0.9437.$$

这表明：被盗索赔户在14与30户之间的概率近似为0.9437.

十二、大数定律与中心极限定理

- **例30** 银行为支付某日即将到期的债券须准备一笔现金，已知这批债券共发放了500张，每张须付本息1000元，设持券人（1人1券）到期到银行领取本息的概率为0.4.问银行于该日应准备多少现金才能以99.9%的把握满足客户的兑换.

- **解：** 设X为该日到银行领取本息的总人数，则$X \sim B(500,0.4)$，所需支付现金为$1000X$，为使银行能以99.9%的把握满足客户的兑换，设银行该日应准备现金x元，则$P\{1000X \le x\} \ge 0.999$.由棣莫弗–拉普拉斯中心极限定理知：

十二、大数定律与中心极限定理

$$P\{1000X \le x\} = P\left\{X \le \frac{x}{1000}\right\}$$

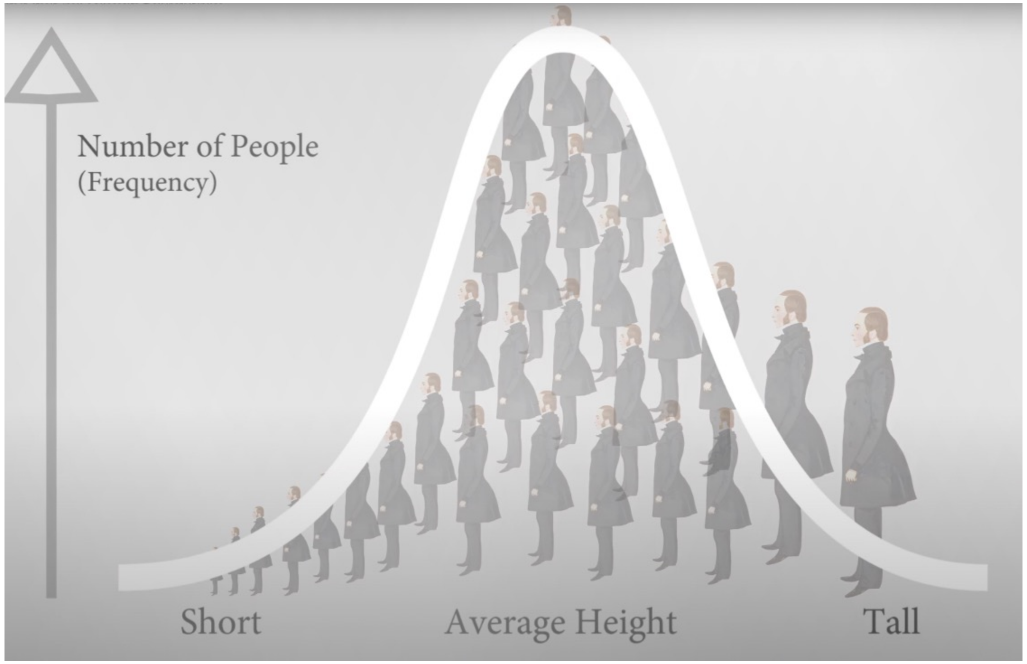
$$= P\left\{\frac{X - 500\times 0.4}{\sqrt{500\times 0.4\times 0.6}} \le \frac{\frac{x}{1000} - 500\times 0.4}{\sqrt{500\times 0.4\times 0.6}}\right\}$$

$$\approx \Phi\left(\frac{x - 20000}{2000\sqrt{30}}\right) \ge 0.999 = \Phi(3.1).$$

即$\frac{x-20000}{2000\sqrt{30}} \ge 3.1$，得$x \ge 233958.798$.

因此银行于该日应准备234000元现金才能以99.9%的把握满足客户的兑换.

思政案例：透过现象看本质



为什么一个人群的身高呈现正态分布?
为什么新生儿的体重呈现正态分布?
为什么人的智力呈现正态分布?
为什么考试成绩呈现正态分布?
…

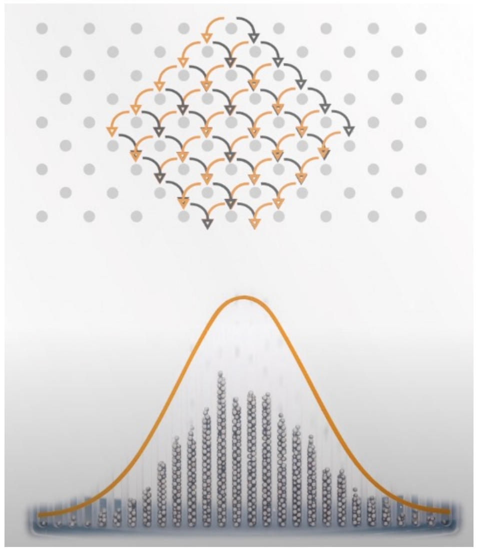
思政案例：透过现象看本质

高尔顿钉板

中心极限定理

正态分布体现了充分多的小独立随机事件叠加构成的结果，这解释了为什么正态分布在现实世界中如此常见。

- 身高是众多基因和环境作用的结果
- 考试成绩是每一次课堂和课后学习的结果
- 未来是过去无数次选择叠加的结果



二项分布的正态近似

$$C_{\mathrm{n}}^{k} p^{k}(1-p)^{n-k}$$

思政案例：透过现象看本质

中心极限定理: 在一些条件下, 独立随机变量之和的适当标准化形式收敛到正态分布。

- 1733年，法国数学家Abraham De Moivre在他”Doctrine of chances”一书中, 使用正态分布近似多次抛掷硬币正面朝上的次数。 这也是正态分布的第一次出现。
- 1812，法国数学家Pierre-Simon Laplace发表了重要的《Théorie Analytique des Probabilités》一书, 他拓展了De Moivre的想法，用正态分布来近似二项分布。
- 1900年, 俄国数学家Aleksandr Lyapunov在Chebyshev和Markov的工作基础上，利用Laplace所引入的特征函数方法，在Lyapunov条件下给出了第一个一般化的CLT(此时还不叫做CLT)的严格证明。
- 1920年， 匈牙利数学家 Georage Polya首次提出“zentraler Grenzwertsatz”(翻译为Certral Limit Theorem)一词。 “Central”一词在这里的使用源于它在概率论中的重要性。

思政案例：透过现象看本质

- 1922年，芬兰数学家Jarl W. Lindeberg在Lindeberg条件下证明了CLT。
- 1925年到1937年， 法国数学家Paul Pierre Lévy发展了特征函数理论，并使用其证明一般化的CLT。
- 1934年，英国数学家Alan Turing证明了与1922年Lindeberg CLT相似的结果，但在他提交这项工作之后才知道它已经被证明了。
- 1935年， 克罗地亚裔美国数学家William Feller发表论文讨论CLT成立的条件。他广泛地讨论了Lindeberg条件的充分性和必要性。

科学发现需要我们能够透过现象，思考并挖掘背后的道理。在这场跨越两百年的知识探索征程中，无数伟大的科学家做出了巨大的贡献，至今仍深刻影响着我们。

Thank You !