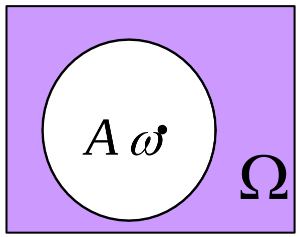
一、随机事件及其概率

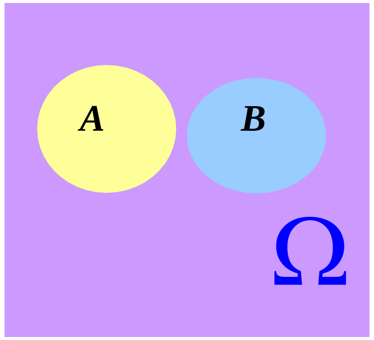
随机试验

- 在许多问题中，一次试验只有两个结果：A或者$\bar{A}$，这类试验称为伯努利试验(Bernoulli trials).
- 随机试验E中的每一个基本的结果(outcome)称为基本事件(simple event, elementary event)或样本点(sample)ω.
- 全体样本点构成的集合(set)称为试验E的样本空间(sample space)或结果空间（outcome space）Ω.
- 具有某种性质的样本点构成的样本空间的一个子集称为随机事件(random event),简称事件(event). 事件通常用A, B,C等表示.



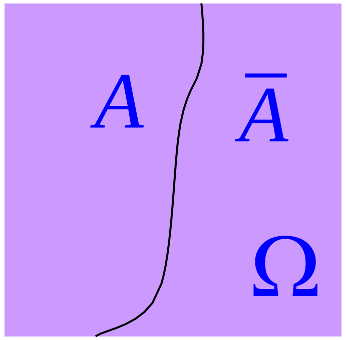
事件的互斥与对立关系

- 互不相容(互斥, disjoint, mutually exclusive)



若事件A，B不能同时发生，即 $A\cap B=\Phi$,
则称事件A与B互不相容.

- 对立事件（补事件，逆事件，complement）



对于事件A，由所有不包含在A中的样本点所组成的事件称为A的对立事件，记为$\bar{A}$.

事件A不发生 ⟺ 对立事件$\bar{A}$发生.

$$\boxed{A+\bar{A}=\Omega}$$

$$\boxed{A\bar{A}=\Phi}$$

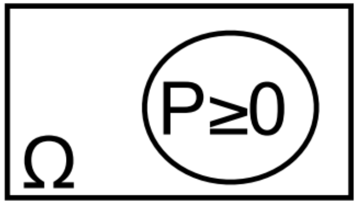
若事件A与B对立，则事件A与B互不相容。 反之，不一定成立。

mutually exhaustive and exclusive

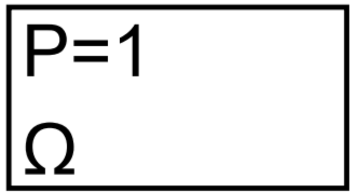
概率的公理化定义

设随机试验E的样本空间为Ω，若对E的每一事件A，都有一个实数$P(A)$与之对应，并且满足下列三条公理（Axioms），则称$P(A)$为事件A的概率.

(1)非负性（Non-negativity）对每一个事件A，有 $P(A)\geq 0$



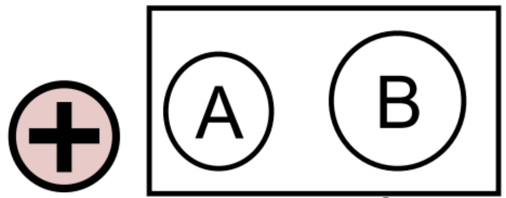
(2)规范性（Unitarity） $P(\Omega)=1$



(3)可列可加性（Countable Addition rule）

对任意可数个两两互不相容事件$A_1, A_2, \dots,$ 有

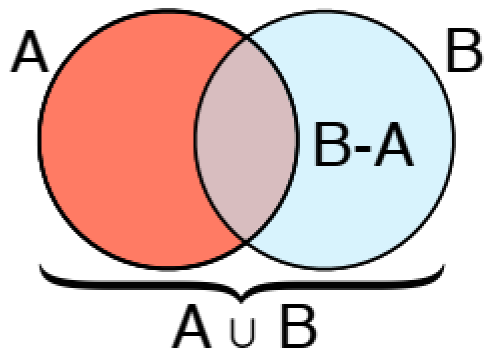
$$P\left(\sum_{i=1}^{+\infty} A_i\right)=\sum_{i=1}^{+\infty} P(A_i).$$



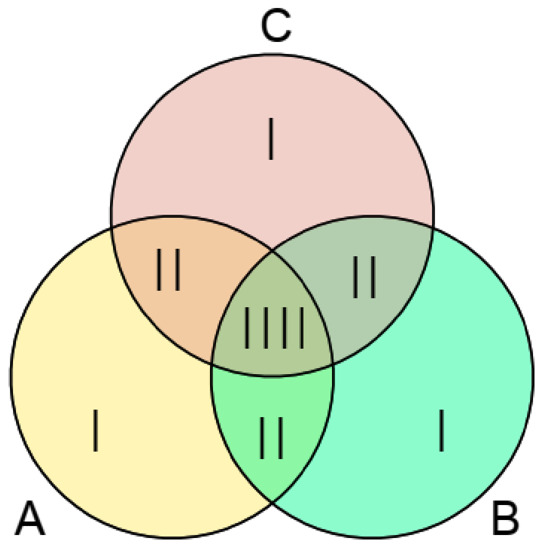
加法公式

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

注： $P(AB) = P(A \cap B) = P(A, B)$



$$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & + P(A \cap B \cap C) \end{aligned}$$



一般的加法公式(Principle of Inclusion-Exclusion)

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{1 \le i \le n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j) + \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

条件概率

设A，B是两个事件，且$P(A)>0$,

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

称为在事件A发生的条件下，事件B发生的条件概率(conditional probabilities).

乘法公式(product rule)

设A, B与$A_1, A_2, \ldots, A_n$都是样本空间Ω中的事件. 则

(1) $P(AB)=P(B)P(A|B)$ （这里要求$P(B)>0$）.

$\quad\quad = P(A)P(B|A)$ （这里要求$P(A)>0$）.

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

（这里要求$P(A_1A_2\ldots A_{n-1})>0$） （Chain rule ）

注：P(S|A,B,C) 相当于 P(S|ABC)，P(S|A∩B∩C)

事件的独立性

设A和B是样本空间Ω中的两个事件，如果有

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

则称事件A与事件B相互独立，简称独立(independent，记为A∐B).

否则称A与B不独立.

事件A与事件B相互独立表示其中一个事件的发生与否不影响另一个事件发生的概率.

必然事件、不可能事件与任何事件相互独立.

独立性的判断方法：

方法1：根据试验实际情况判断。例如，有放回的抽取方式中两次抽取结果是相互独立的。

方法2：验证$P(AB)=P(A)P(B)$

n个事件的独立性

设有n个事件$A_1, A_2, \ldots, A_n$，如果对任意的

$1 \le i < j < k < \ldots \le n$，以下$2^n - n - 1$个等式均成立.

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

则称n个事件$A_1, A_2, \ldots, A_n$相互独立(mutually independent).

- n个事件相互独立，则一定两两独立。
- n个事件两两独立不能保证n个事件相互独立。
- n个相互独立的事件中的任意部分事件(两个及以上)仍是相互独立的。
- 相互独立的n个事件中的任意部分事件换成对立事件后，所得到的n个事件仍相互独立.

加法公式(Principle of Inclusion-Exclusion)

$P(\bigcup_{i=1}^{n} A_i) = \sum_{1\le i\le n} P(A_i) - \sum_{1\le i<j\le n} P\left(A_i \cap A_j\right) + \cdots + (-1)^{n-1} P(\bigcap_{i=1}^{n} A_i)$

n=2时， $P(A \cup B) = P(A) + P(B) - P(AB)$

乘法公式(Chain rule)

$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$

(这里要求$P(A_1A_2\cdots A_{n-1}) > 0$)

n=2时， $P(AB) = P(B)P(A|B)$ (要求$P(B) > 0$)

n个事件$A_1, A_2, \cdots, A_n$互不相容时，$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{1\le i\le n} P(A_i)$， $P\left(\bigcap_{i=1}^{n} A_i\right) = 0$

n=2时， $P(A \cup B) = P(A) + P(B)$， $P(AB) = 0$

n个事件$A_1, A_2, \cdots, A_n$**相互独立**时， $P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n}(1 - P(A_i))$，

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

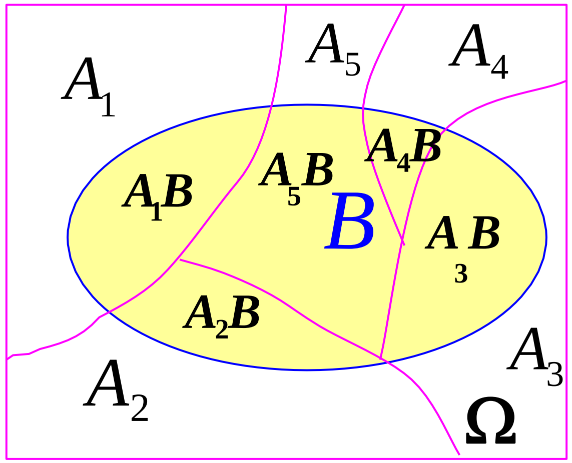
n=2时， $P(A \cup B) = P(A) + P(B) - P(A)P(B)$， $P(AB) = P(A)P(B)$

全概率公式(Law of Total Probability， LOTP)

设$A_1, A_2, \ldots, A_n$(n可以是∞)是样本空间Ω中的n个事件，若它们两两互不相容，而且 $\sum_{k=1}^{n} A_k = \Omega$，则称它们是样本空间$\Omega$的一个分割(Partition)，亦或完备事件组.

全概率公式 设事件$A_1, A_2, \ldots, A_n$ (n可以是∞)是Ω的一个分割，且$P(A_k) > 0(i = 1, 2, \ldots, n)$，则对$\Omega$中的任一事件$B$，有

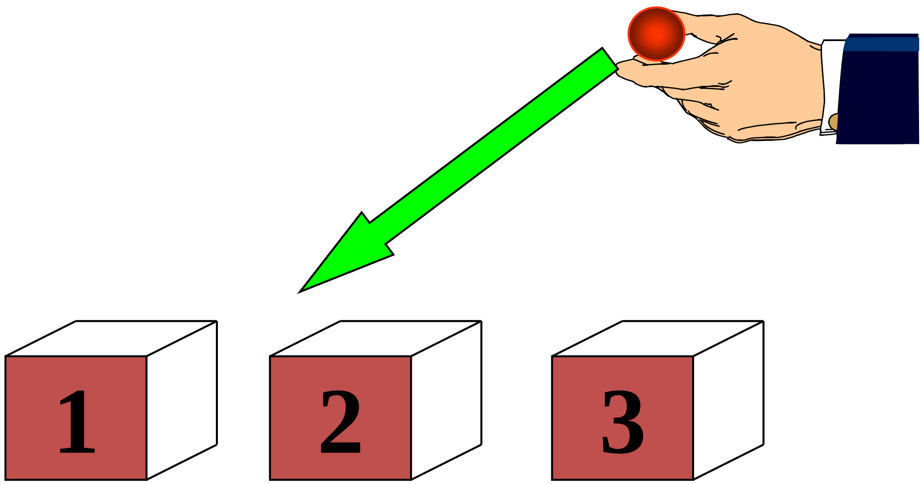
$$P(B) = \sum_{k=1}^{n} P(A_k)P(B \mid A_k)$$



某一事件B的发生有各种可能的原因A_i(i=1,2,…,n)，如果B是由原因A_i所引起，则B发生的概率是 $P(BA_i)=P(A_i)P(B|A_i)$. 每一原因都可能导致B发生，故B发生的概率是各原因引起B发生概率的总和，即全概率公式（由因到果）.

贝叶斯公式(Bayes' rule)

某人从任一箱中任意摸出一球,发现是红球,求该球是取自1号箱的概率,或者问该球取自哪号箱的可能性最大?



可以使用贝叶斯定理来计算每个箱子的后验概率。

贝叶斯定理的公式为:

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

其中:

- A 是事件"球来自1号箱";
- B 是事件"摸到的是红球";
- $P(A|B)$ 是给定球是红球的情况下，它来自1号箱的概率。

计算步骤:

1. **先验概率**：我们假设从每个箱子中摸到球的概率是相等的。所以，$P(1) = P(2) = P(3) = \frac{1}{3}$。
2. **似然函数**：即在每个箱子中摸到红球的概率。
   - $P(B|1) = \frac{r_1}{r_1+b_1}$
   - $P(B|2) = \frac{r_2}{r_2+b_2}$
   - $P(B|3) = \frac{r_3}{r_3+b_3}$
3. **边际概率**：摸到红球的总概率，可以通过加法法则计算:

   $$P(B) = P(B|1)P(1) + P(B|2)P(2) + P(B|3)P(3)$$

   代入数值:

   $$P(B) = \frac{1}{3}\left(\frac{r_1}{r_1+b_1} + \frac{r_2}{r_2+b_2} + \frac{r_3}{r_3+b_3}\right)$$

4. **后验概率**：根据贝叶斯定理，计算球来自1号箱的概率:

   $$P(1|B) = \frac{P(B|1)P(1)}{P(B)} = \frac{\frac{r_1}{r_1+b_1} \cdot \frac{1}{3}}{\frac{1}{3}\left(\frac{r_1}{r_1+b_1} + \frac{r_2}{r_2+b_2} + \frac{r_3}{r_3+b_3}\right)}$$

已知结果求原因(由果到因，推理)

贝叶斯公式 设事件$A_1, A_2, \ldots, A_n$ (n可以是∞)是Ω

分割. 则对Ω中的任一事件B（$P(B) > 0$），有

$$P(A_k|B) = \frac{P(A_k)P(B \mid A_k)}{\sum_{j=1}^{n} P(A_j)P(B \mid A_j)}, \quad k = 1,2,\cdots,n$$

二、随机变量及其分布

随机变量

$$X : \Omega \mapsto \mathbb{R}$$

定义域：Ω， 值域：R

设试验 E 的样本空间为Ω，如果对于每一个样本点 $\omega \in \Omega$，都有唯一实数 $X(\omega)$ 与之对应，则称单值实函数 $X = X(\omega)$ 为样本空间Ω上的随机变量(random varable,简写为r.v.).

这样就可以利用实数 $X(\omega)$ 描述样本点 ω，利用实数的子集来描述事件. 随机变量 $X(\omega)$ 简记为 X.

基本事件 $\{\omega：X(\omega)=x\}$ 简记为 $\{X=x\}$ 或 $X=x$

事件 $\{\omega：X(\omega)<=x\}$ 简记为 $\{X<=x\}$ 或 $X<=x$

随机变量的分类

离散型　可能取值的个数为有限个(finite)或无限可列个(countably infinite).

连续型　可能的取值充满某个(有限,无限)区间.

分布函数

Cumulative distribution function (CDF)

设X是定义在样本空间Ω上的随机变量，称函数$F(x)=P\{X\le x\}$, $-\infty<x<\infty$为随机变量X的概率分布函数(probability distribution function)，简称分布函数. 简记为$X\sim F(x)$.

分布函数的性质

(1) $F(x)$单调不减，即$\forall\ a<b$， $F(a)\le F(b)$

(2) $0\le F(x)\le 1$ $\quad\lim\limits_{x\to+\infty}F(x)=1,\ \lim\limits_{x\to-\infty}F(x)=0$

(3) $F(x)$右连续，即$\lim\limits_{x\to x_0^+}F(x)=F(x_0)$

性质(1)--(3)是鉴别一个函数是否是某r.v的分布函数的充分必要条件.

离散随机变量

p_k满足

$$p_k \geq 0, k=1,2,\ldots$$

$$\sum_k p_k = 1$$

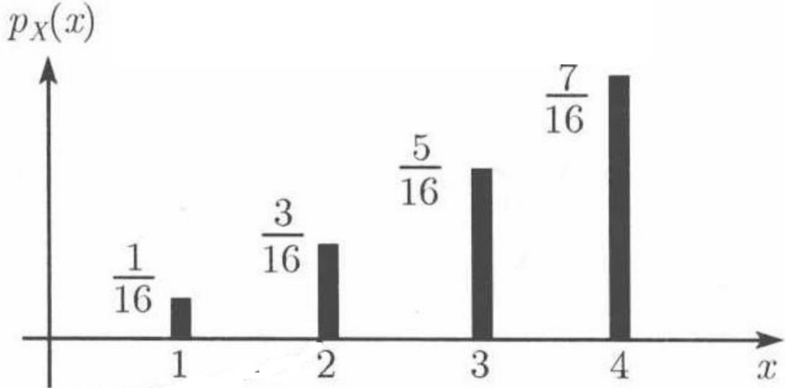
设X是样本空间Ω上的随机变量．若集合$\{X(\omega):\omega\in\Omega\}=\{x_1,x_2,\ldots,x_n,\ldots\}$是有限集或无限可列集，则称$X$是离散型随机变量.

$p_k=P(X=x_k)$，$k=1,2,\ldots$ 称为离散型随机变量X的概率分布或分布律（Probability mass functions, pmf)

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

为离散型随机变量X的概率分布表（或分布列）.

离散型随机变量X的概率分布图



两点分布

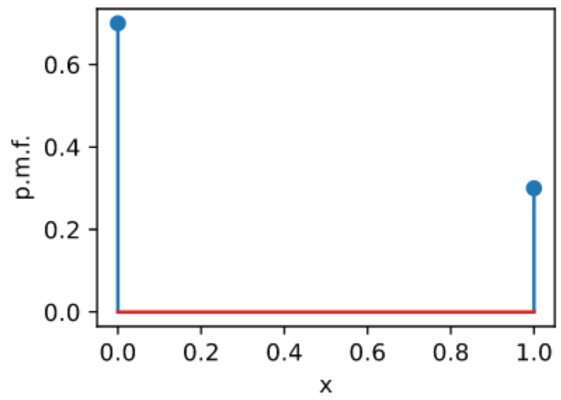
只有两种可能结果的随机实验（A发生，或$\bar{A}$发生），如新生婴儿是男还是女、明天是否下雨等，随机变量X表示一个事件发生A的次数，只可能取0与1两个值,它的分布列为

X	0	1
P	$1-p$	p

则称X服从两点分布/(0 – 1) 分布/ Bernoulli分布，记为$X \sim \mathrm{Bern}(p)$.

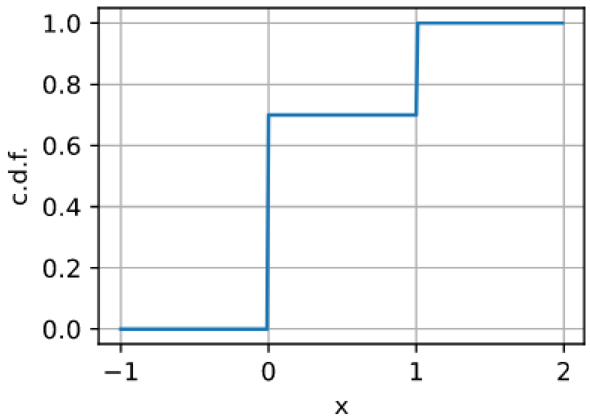
符号“～”表示“服从于”

X称为Bernoulli random variable.



$$P\{X=k\}=p^k(1-p)^{(1-k)},\quad \text{其中} k=0,1$$

$$F(x)=\begin{cases}0 & x<0,\\ 1-p & 0\le x<1,\\ 1 & x>=1.\end{cases}$$



几何分布(Geometric distribution)

在伯努利试验中，直到事件A第一次发生时所做的试验次数X服从几何分布.(**First Success distribution**，第一次成功的分布)

若随机变量X的分布列为：$0<p<=1$

$$P\{X=k\}=(1-p)^{k-1}p,\quad 其中k=1,2,\ldots.$$

则称X服从几何分布，记为$X\sim G(p)$.

$$\sum_{k=1}^{\infty}(1-p)^{k-1}\cdot p \ = 1$$

For $r\in(-1,1)$, $\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$.

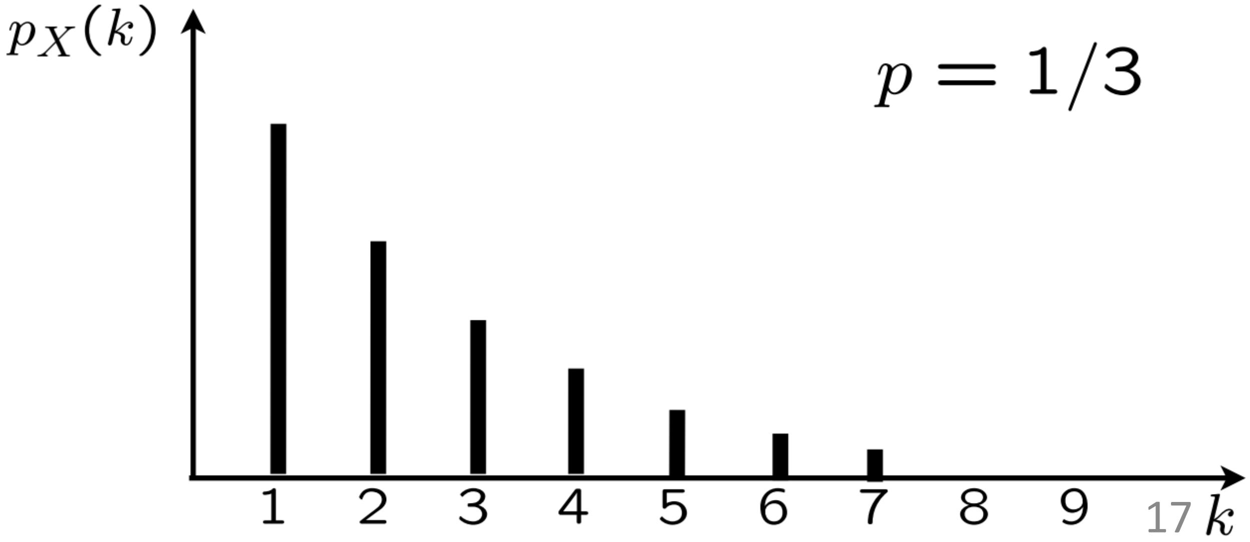
$$P(X=1)=p$$

$$P(X>1)=1-p$$

$$P(X\geq k)=(1-p)^{k-1}$$

$$P(X>k)=(1-p)^{k}$$

$$P(X\leq k)=1-(1-p)^{k}$$



超几何分布

设有共有N件产品，其中M件次品，今从中任取n件（不放回抽取），则这n件产品中的次品数X是一个离散型随机变量，其概率分布为

$$P(x=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \qquad k=0,1,\ldots,\min(M,n)$$

称为超几何分布**(Hypergeometric distribution).**

二项分布(Binomial)

在n重贝努利试验中，事件A出现的次数记为X服从二项分布。

随机变量X的概率分布为

$$P(X=k)=C_n^k p^k(1-p)^{n-k}, \qquad 0 \le k \le n$$

其中$0<p<1$. 称X服从参数为 n, p 的二项分布, 记作$X \sim B(n, p)$.

当$n=1$时，$X \sim B(1, p)$，即是两点分布.

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p) \Rightarrow Y \equiv \sum_{i=1}^{n} X_i \sim \text{B}(n, p)$$

(独立同分布，iid，independently identically distribution)

- 抛1次硬币，正面朝上的次数（0或1），服从伯努利分布。
- 抛n次硬币，正面朝上的次数，服从二项分布。

扩展：多类别(categorical)分布、多项(multinomial)分布

抛1次骰子，第k面朝上的次数(k=1,2,...,6)（0或1），这是多类别分布

抛n次骰子，第1面朝上出现m_1次,...,第6面朝上出现了m_6次的概率，这是多项分布。

泊松分布(Poisson)

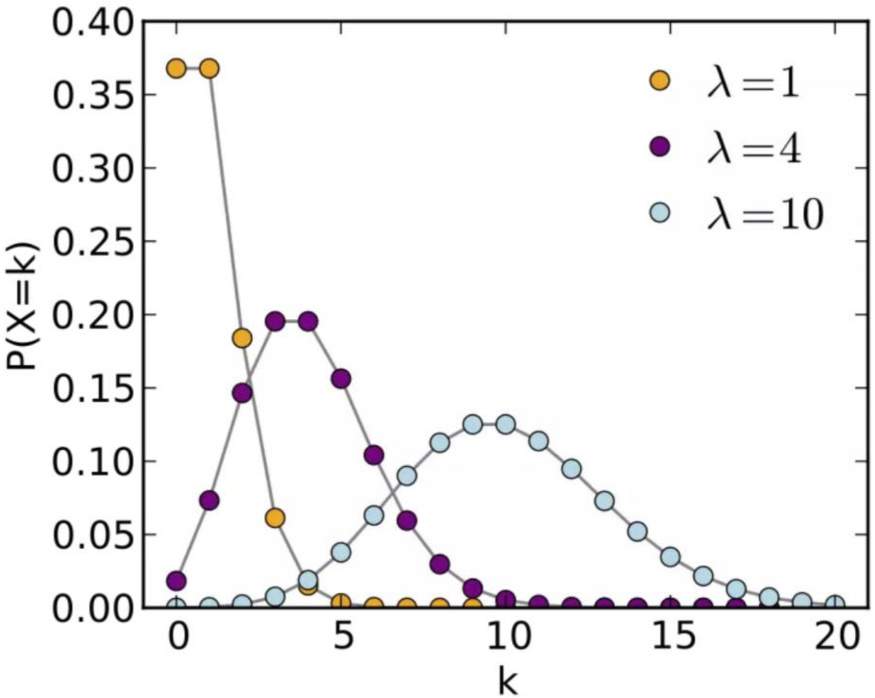
若随机变量X的分布列为

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\ldots \quad \lambda>0$$

则称X服从参数为λ的泊松分布. 记为$X\sim P(\lambda)$.

泊松分布(稀有分布律)主要用于估计某稀有事件(*rare events*)在特定时间内或空间中发生的次数.

λ是过去某段时间或某个空间内随机事件发生的平均次数。
时间单位改变了， λ值也应随之改变。



Possion定理

设随机变量 $X_n(n=1,2,\cdots)$ 服从参数为n, p_n 的二项分布，

若 $\lim_{n\to\infty} np_n = \lambda$, 则有 $\lim\limits_{n\to\infty} P(X_n = k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$.

二项分布的极限分布是 Poisson 分布

若$X \sim B(n, p)$, 则当n 较大，p 较小, 则

$$C_n^k p^k (1-p)^{n-k} \approx e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \cdots$$

$n > 10$, $p < 0.1$时近似效果较好

二项分布 $\xrightarrow{\ \boldsymbol{n}\text{很大}, \boldsymbol{p}\text{很小}\ }$ 泊松分布

几种离散型分布之间的关系

从a件正品和b件次品中抽取产品，每次抽一件（每次试验都只有两种结果，伯努利实验）：

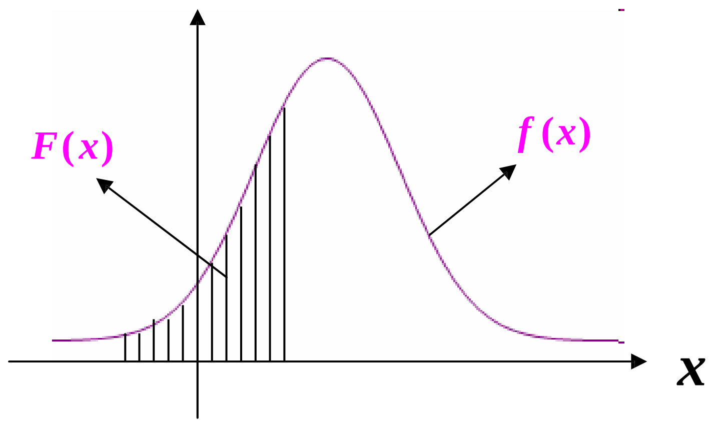
- 抽一件产品：出现的次品数X =0或1，服从两点分布。
- 抽n件产品（放回）：出现的次品数X = 0，1, ..., min(b,n)，服从二项分布。
- 当n 较大，次品率p较小时，出现的次品数X = 0,1, ...,可用泊松分布$(\lambda = np)$近似二项分布
- 连续抽取（放回），直到第一次出现次品时所做的试验次数X= 1,2,…，服从几何分布。
- 抽n件产品（不放回）：出现次品数X = 0，1, ..., min(b, n) 服从超几何分布。

连续型随机变量及密度函数

设$F(x)$是随机变量X的概率分布函数(*cumulative distribution function,* cdf)，若存在可积函数$f(x)\geq 0$，对任意实数x，有

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\mathrm{d}t \qquad -\infty < x < +\infty$$

称X为连续型随机变量，$f(x)$为X的概率密度函数 (Probability Density Function，pdf).



若$f(x)$在点x处连续，则$f(x) = F'(x) = \frac{d}{dx}\int_{-\infty}^{x} f(t)dt$

概率密度函数应满足：

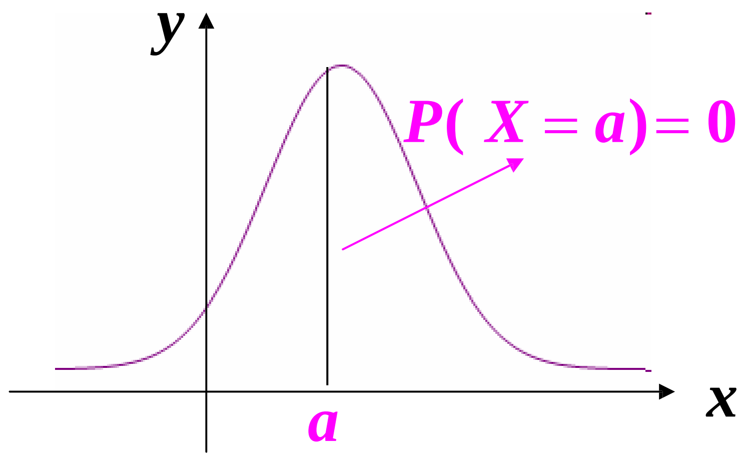
- $f(x) \ge 0$，但某些x的函数值不一定小于或等于1
- $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = F(+\infty) = 1$

连续型随机变量取任一指定值的概率为0

即：$P(X = a) = 0$，a为任一指定值

$$P(X = a) = \lim_{\Delta x \to 0} P(a \le X \le a + \Delta x)$$

$$= \lim_{\Delta x \to 0} \int_a^{a+\Delta x} f(x)dx = 0$$



概率为0 (或1) 的事件未必不发生(或一定发生)

$P(\varnothing)=0$，但由$P(A)=0$，不能推出$A = \varnothing$

$P(\Omega)=1$，但由$P(B)=1$，不能推出$B = \Omega$

例如

B是$X \in (-\infty, a) \cup (a, \infty)$

A是$\{X = a\}$

均匀分布(continuous uniform distribution)

如果连续型随机变量 X 具有密度函数

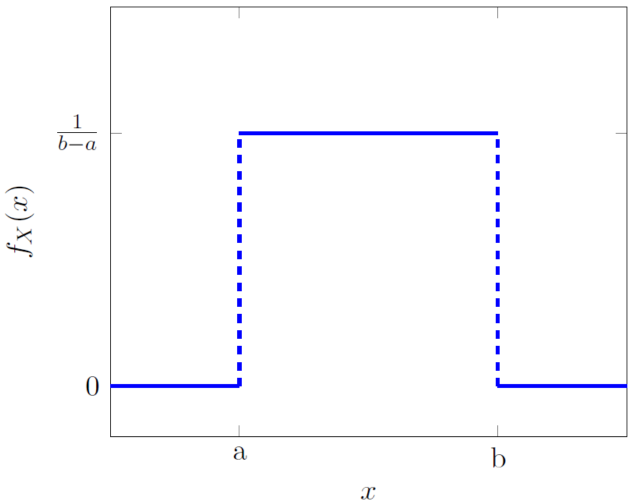
$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b\\ 0 & \text{其他}\end{cases}$$

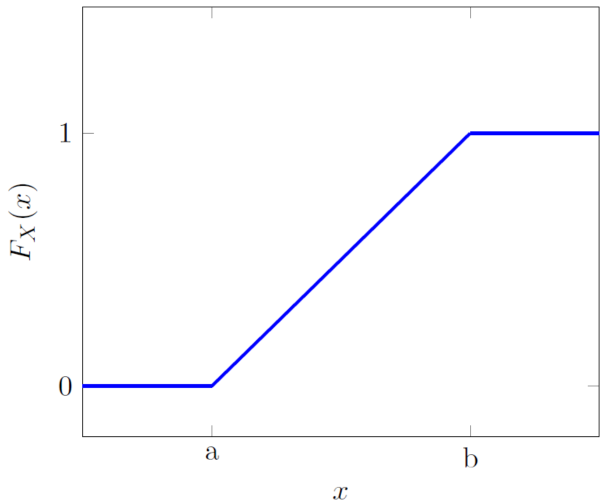
在指定区间内的所有值都具有相同的概率。

其中 a,b 是有限数，则称 X 是 $[a,b]$ 上的均匀分布, 记作 $X\sim U\,[a,b]$

分布函数为

$$F(x)=\begin{cases}0 & x<a,\\ \dfrac{x-a}{b-a} & x\in[a,b)\\ 1 & x>=b\end{cases}$$

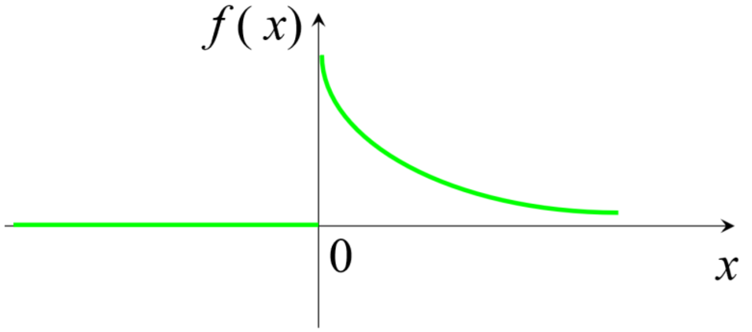




指数分布(Exponential Distribution)

如果连续型随机变量X具有密度函数

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

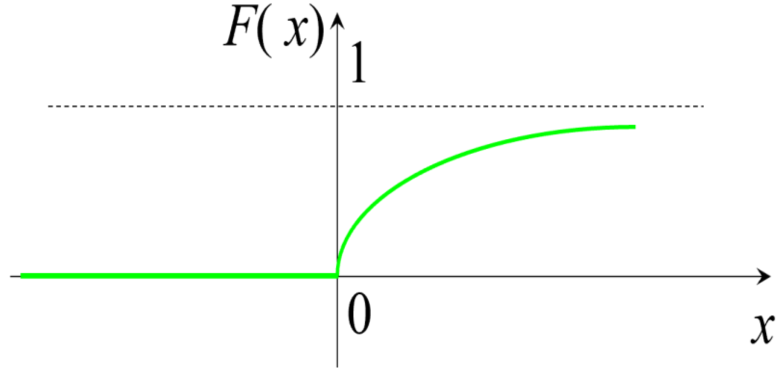


$\lambda>0$为常数，则称X服从参数为λ的指数分布，

简记为$X\sim E(\lambda)$.

其分布函数为

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



正态分布(Normal/Gaussian)

如果连续型随机变量X的密度函数

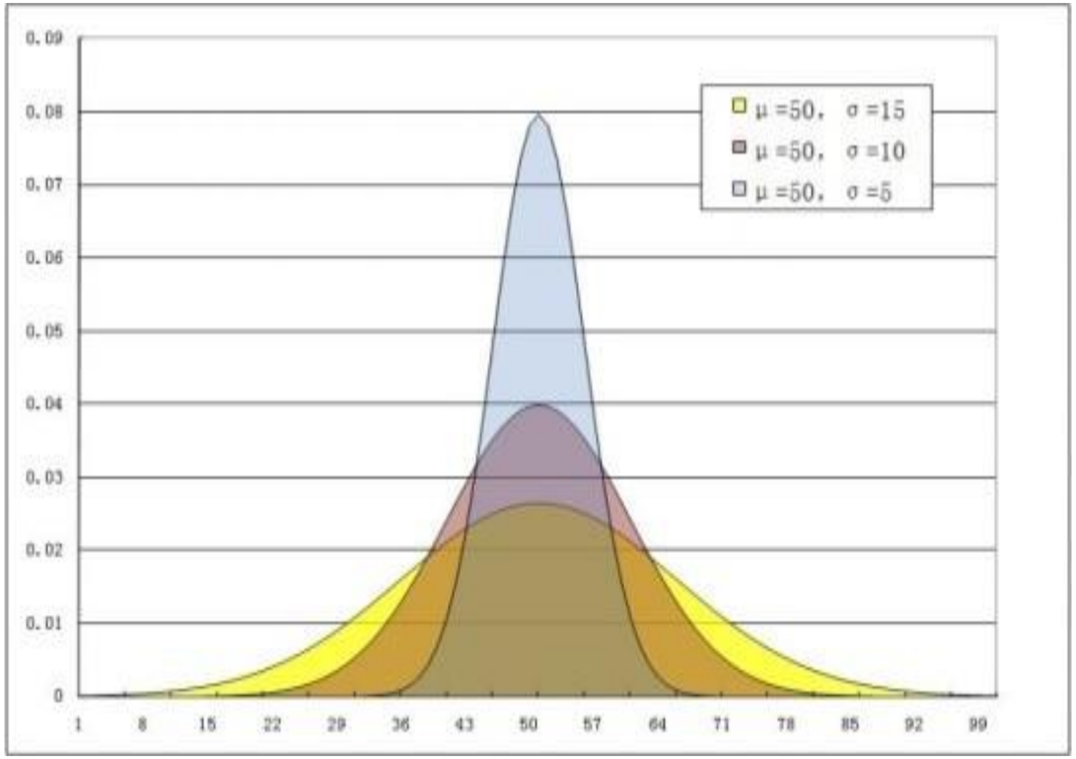
$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\qquad -\infty<x<+\infty$$

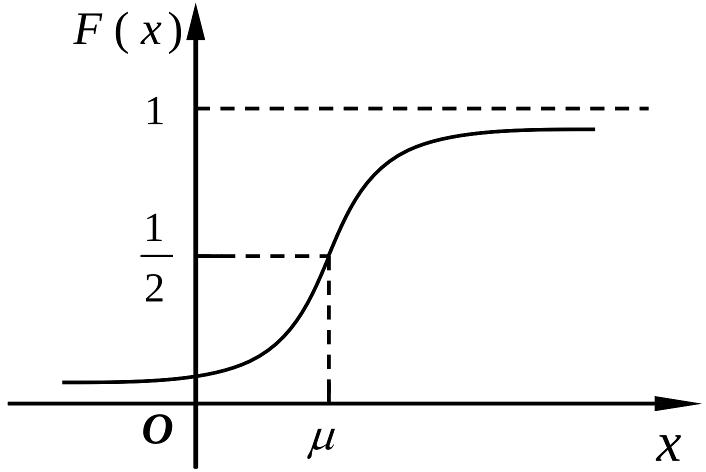
其中μ,σ为常数，且$\sigma>0$，则称随机变量X服从参数μ,σ的正态分布，记为$X\sim N(\mu,\sigma^2)$.

正态分布的分布函数为

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

积分没有解析表达式。





$f(x)$的特点：

- 在$x=\mu$处达到极大值；

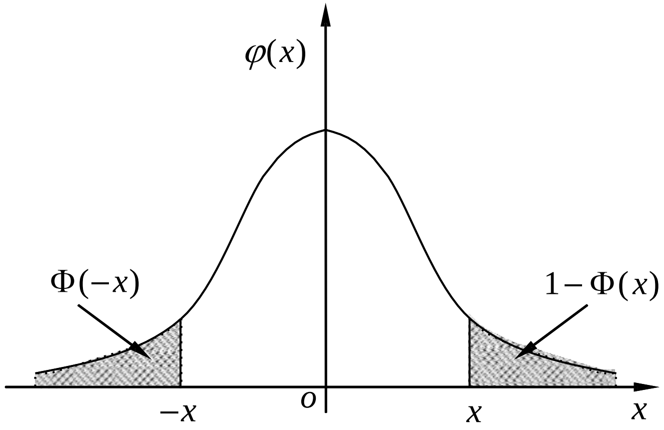
$$f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$$

- σ越小，曲线在$x=\mu$附近越陡峭.
- 正态分布的密度曲线是一条关于钟形曲线，两头小，中间大，左右对称（关于$x=\mu$对称），以x轴为渐近线，呈单峰状. $f(\mu+x)=f(\mu-x)$
- 在$x=\mu\pm\sigma$时，曲线 $f(x)$ 在对应的点处有拐点
- μ决定了图形的中心位置，σ 决定了图形中峰的陡峭程度.

标准正态分布(Standard Normal distribution)

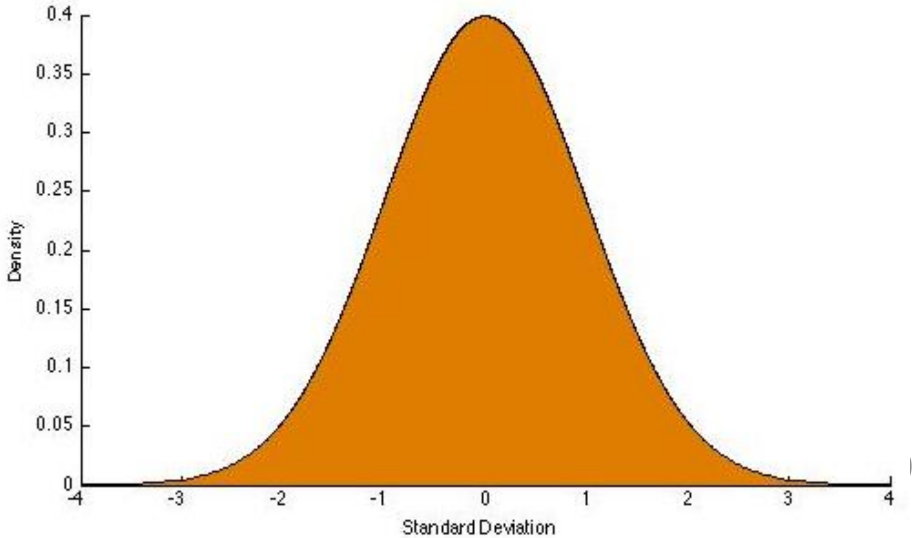
当$\mu=0$, $\sigma=1$时，称为标准正态分布，记为 $X \sim N(0,1)$，相应的分布密度函数及分布函数分别记为$\varphi(x)$及$\Phi(x)$.

$$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \qquad -\infty < x < +\infty$$



$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\mathrm{d}t \qquad -\infty < x < +\infty$$

$$=\begin{cases} \text{查表}\Phi(x) & x>0 \\ 1-\Phi(-x) & x<0 \end{cases}$$



一般的正态分布可以通过线性变换转化为标准正态分布.

设$X \sim N(\mu, \sigma^2)$，则$Y = \frac{X-\mu}{\sigma} \sim \mathbf{N(0,1)}$

将标准正态分布的分布函数制成表（z-table），就可以解决一般正态分布的概率计算问题. z-score $= \frac{X-\mu}{\sigma}$

对一般的正态分布：$X \sim N(\mu, \sigma^2)$ 其分布函数 $F(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t$

作变量代换 $s = \frac{t-\mu}{\sigma}$ $\Longrightarrow$ $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

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

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

$$\boldsymbol{P(|X-\mu| \le k\sigma) = F(\mu+k\sigma) - F(\mu-k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1}$$

设随机变量 $X \sim N(\mu, \sigma^2)(\sigma > 0)$, 记 $p = P\{X \leqslant \mu + \sigma^2\}$,

则 [填空1] (填ABCD之一)

(A) p 随着 μ 的增加而增加.

(B) p 随着 σ 的增加而增加.

(C) p 随着 μ 的增加而减少.

(D) p 随着 σ 的增加而减少.

作答

离散型随机变量函数的分布

设X是离散型随机变量，其分布列为

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

则$Y=g(X)$的可能取值为$g(x_1)$，… ， $g(x_n)$，… 也是离散型，其分布列为

Y	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_n)$	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

注意：当某些$g(x_i)$相等时，应把它们适当合并.

$$P(Y=y_i)=\sum_{k:\,g(x_k)=y_i} p_k, \quad i=1,2,\cdots$$

连续型随机变量函数的分布

设X是连续型随机变量，其分布函数与密度函数分别为$F_X(x)$、$f_X(x)$，求$Y=g(X)$的分布函数$F_Y(y)$、密度函数$f_Y(y)$.

注意：$F_Y(y)$、$f_Y(y)$只能含有y，不能含有x；

根据x的取值范围确定y的取值范围确定

从分布函数出发推导

$$F_Y(y)=P(Y\le y)=P(g(X)\le y)$$

$$f_Y(y)=F_Y'(y)$$

对于连续型随机变量，求Y=$g(X)$ 的分布的关键是把事件{ $g(X)\le y$ }转化为X的不等式，利用 X 的分布来求 $F_Y(y)=P(g(X)\le y)$

注意:积分区域$D=\{x:g(x)\le y\}$的确定，可能随着y的取值不同而改变。

定理 设X是连续型随机变量，取值范围为区间(a, b)（有限或无限），密度函数为$f_X(x)$，$Y = g(X)$.

(1) 若$y = g(x)$在(a, b)上**严格单调**，**反函数**$x = g^{-1}(y) = h(y)$连续可导. 则$Y=g(X)$是连续型随机变量，其密度函数为

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

其中 $\alpha = \min(g(a), g(b)),\quad \beta = \max(g(a), g(b))$.

(2) 若$y = g(x)$在(a, b)的不相重叠的区间I_1，I_2，…，**逐段严格单调**，其反函数$h_1(y)$，$h_2(y)$，…，均连续可导，则$Y = g(X)$是连续型随机变量，其密度函数为

$$f_Y(y) = f_X[h_1(y)]|h_1'(y)| + f_X[h_2(y)]|h_2'(y)| + \dots$$

例： 设 $X \sim N(\mu, \sigma^2)$，$Y = aX + b$，则

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{1}{a}(y - b)\right)$$

$$= \frac{1}{\sqrt{2\pi}\sigma|a|} e^{-\frac{(y-a\mu-b)^2}{2a^2\sigma^2}} \qquad -\infty < y < \infty$$

$$Y \sim N(a\mu + b,\ a^2\sigma^2)$$

$$\boxed{X \sim N(\mu, \sigma^2)\text{，则 } Y = aX + b \sim N(a\mu + b,\ a^2\sigma^2)}$$

特别地，$a = \frac{1}{\sigma}$，$b = -\frac{\mu}{\sigma}$

$$\boxed{\text{若 } X \sim N(\mu, \sigma^2)\text{，则 } Y = \frac{X-\mu}{\sigma} \sim N(0,1)}$$

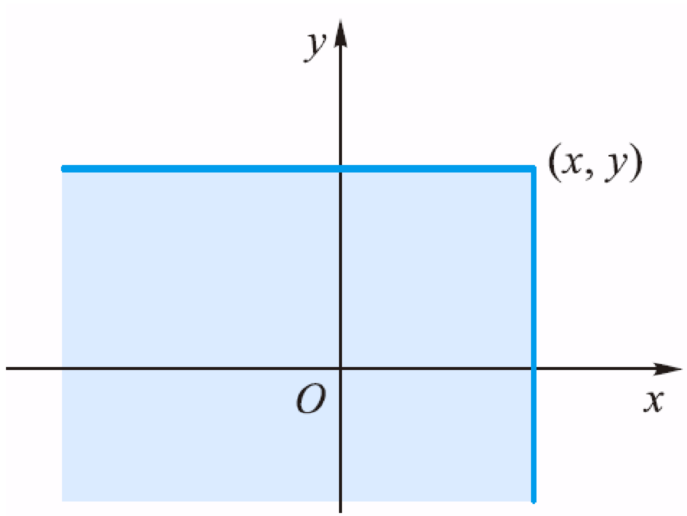
二维随机变量的联合分布函数(joint cdf)

设 (X,Y) 为二维随机变量, 对任意实数 x,y, 二元函数

$$F(x,y)=P(X\leqslant x, Y\leqslant y)$$

称为二维随机变量 (X,Y) 的**联合分布函数**, 简称为 (X,Y) 的分布函数.

用平面上的点 (x, y) 表示二维随机变量 (X, Y) 的一组可能的取值，则 $F(x, y)$ 表示随机点落在以 (x,y) 为顶点而位于该点左下方的无穷矩形区域 D 内的概率.



联合分布函数的性质

① $0 \le F(x, y) \le 1$

$F(+\infty, +\infty) = 1$

$F(x, -\infty) = 0$

$F(-\infty, y) = 0$

$F(-\infty, -\infty) = 0$

②对每个变量单调不减

固定 x , 对任意的 $y_1 < y_2$, $F(x, y_1) \le F(x, y_2)$

固定 y , 对任意的 $x_1 < x_2$, $F(x_1, y) \le F(x_2, y)$

③对每个变量右连续

$F(x_0, y_0) = F(x_0 + 0, y_0)$ $F(x_0, y_0) = F(x_0, y_0 + 0)$

二维离散型随机变量

若二维随机变量(X,Y)所有可能的取值为有限个或无限可列个，则称(X,Y)为二维离散型随机变量.

描述二维离散型随机变量的概率特性常用**联合概率分布**。

设(X,Y)的所有可能的取值为(x_i, y_j)，$i,j=1,2,\cdots$

称$P\left(X=x_i, Y=y_j\right)=p_{ij}$，$i,j=1,2,\cdots$ 为二维随机变量(X,Y)的联合概率分布，简称 概率分布 或 分布律

$$0 \le p_{ij} \le 1, \quad i,j=1,2,\cdots$$

$$\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty} p_{ij}=1$$

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$
$\vdots$	$\cdots$		$\cdots$		$\cdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\cdots$		$\cdots$		$\cdots$

二维连续型随机变量

设二维随机变量(X,Y)的分布函数为 $F(x,y)$, 若存在非负可积函数 $f(x,y)$，使得对于任意实数 x , y 有

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

则称(X,Y)为**二维连续型随机变量**, $f(x,y)$ 为(X,Y) 的联合概率密度函数,简称概率密度函数,简记*pdf* (joint PDF)

联合密度的性质

(1) $f(x,y)\geq 0$， 可以在某些点大于1

(2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dydx=1$

(3) 在$f(x,y)$的连续点处$\frac{\partial^2F(x,y)}{\partial x\partial y}=f(x,y)$

(4) 若G 是平面上的区域，则$P((X,Y)\in G)=\iint_G f(x,y)dxdy$

常用的二维连续型随机变量

（1） 二维均匀分布

设G是平面上的有界区域，面积为A. 若随机变量(X,Y)的概率密度为

$$f(x,y)=\begin{cases} 1/A, & (x,y)\in G \\ 0, & \text{其他} \end{cases}$$

则称(X,Y)服从区域G上的均匀分布，记作$(X,Y)\sim U(G)$

$\forall\, G_1 \subseteq G$，$G_1$的面积为$A_1$，$P((X,Y)\in G_1)=\frac{A_1}{A}$

几何概型：面积之比

（2） 二维正态分布(bivariate Gaussian distribution)

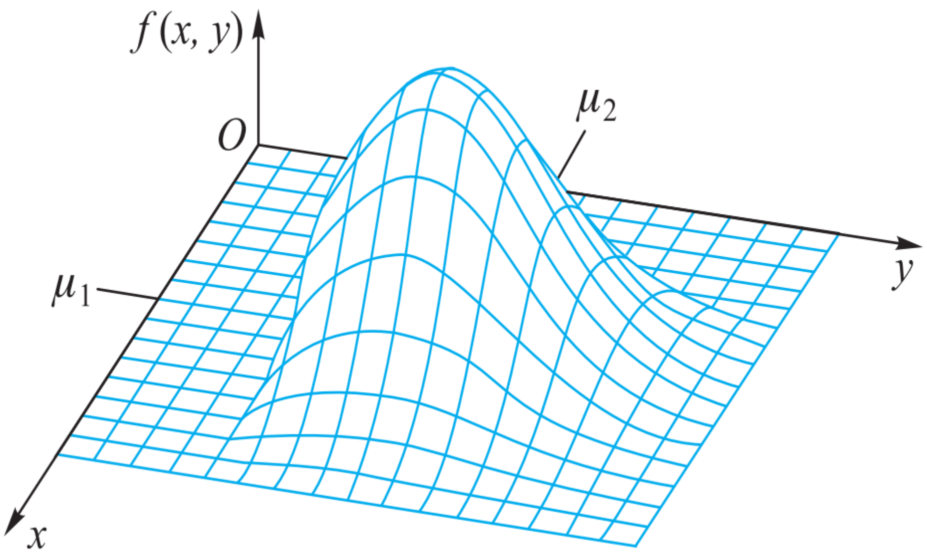
若随机变量(X, Y)的联合pdf为

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

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

则称(X, Y)服从参数为$\mu_1,\sigma_1^2,\mu_2,\sigma_2^2,\rho$的正态分布, 记作

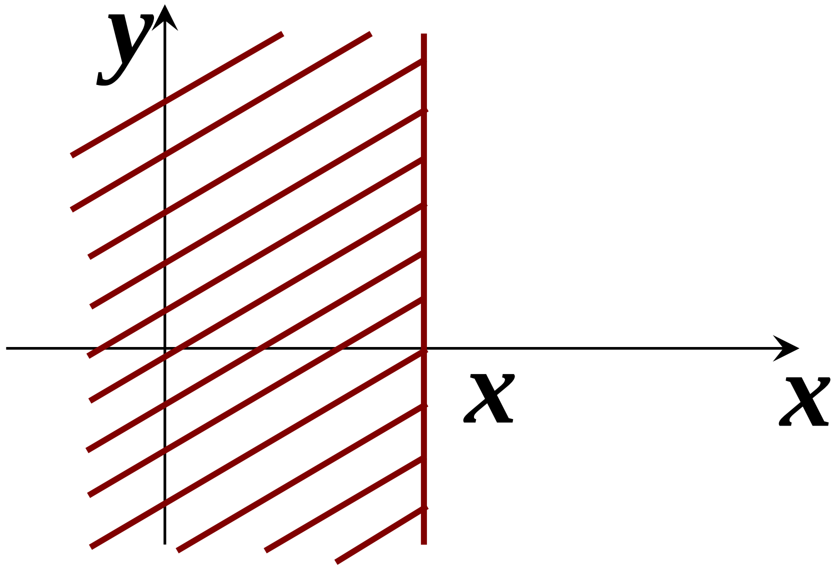
$(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$， 其中$\sigma_1,\sigma_2>0$, $-1<\rho<1$.



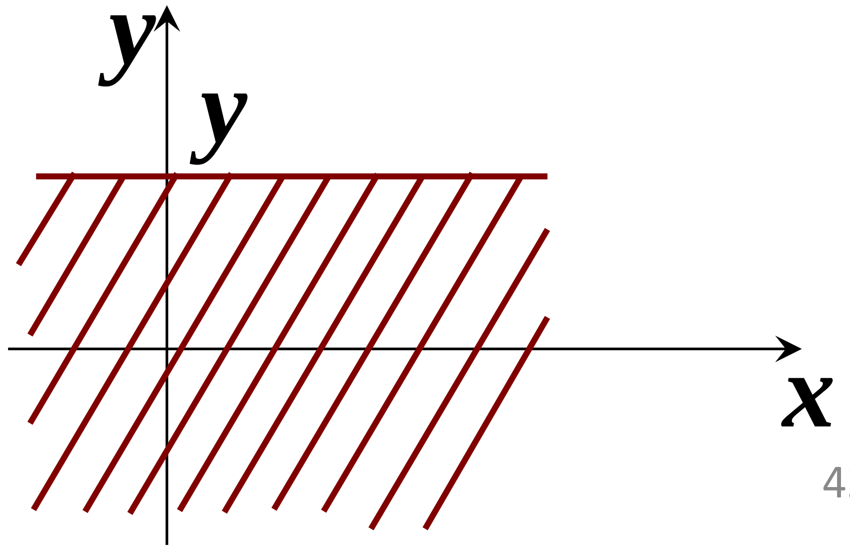
二维随机变量的边缘分布函数

由联合分布函数可以得到边缘分布函数；反过来一般不成立.

$$F_X(x)=P(X\le x)$$
$$=P(X\le x, Y<+\infty)$$
$$=F(x,+\infty)$$



$$F_Y(y)=P(Y\le y)$$
$$=P(X<+\infty, Y\le y)$$
$$=F(+\infty,y)$$



二维离散型随机变量的边缘分布

$$P(X=x_i)=\sum_{j=1}^{+\infty}p_{ij}\overset{\text{记作}}{=}p_{i\bullet},\qquad i=1,2,\cdots$$

$$P(Y=y_j)=\sum_{i=1}^{+\infty}p_{ij}\overset{\text{记作}}{=}p_{\bullet j},\qquad j=1,2,\cdots$$

由联合概率分布可确定边缘概率分布（全概率公式）

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	$P(X=x_i)$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
$P(Y=y_j)$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

二维连续型随机变量的边缘分布

$$F_X(x)=\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty} f(u,y)dy\right]du$$

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

$$F_Y(y)=\int_{-\infty}^{y}\left[\int_{-\infty}^{+\infty} f(x,v)dx\right]dv$$

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

已知联合密度可以求得边缘密度

设二维随机变量 $(X, Y) \sim N\left(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho\right)$

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \qquad (-\infty < x < +\infty)$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \qquad (-\infty < y < +\infty)$$

这表明，$X \sim N\left(\mu_1, \sigma_1^2\right)$，$Y \sim N\left(\mu_2, \sigma_2^2\right)$

二维正态分布的两个边缘分布是一维正态分布,都与参数ρ无关.

$$\text{联合分布} \underset{\not\longleftarrow}{\longrightarrow} \text{边缘分布}$$

由边缘分布函数一般不能完全确定联合分布函数.

离散型随机变量的条件分布律

设(X,Y)是二维离散型随机变量,

对于固定的 j, 若$P(Y=y_j)>0$, 则称

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

为在$Y=y_j$条件下随机变量 X 的条件分布律.

对于固定的 i, 若$P(X=x_i)>0$, 则称

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

为在 $X=x_i$条件下随机变量Y的条件分布律.

条件分布是一种概率分布，它具有概率分布的一切性质.

例如：$1\geq P\left(X=x_i\middle|Y=y_j\right)\geq 0,\sum_{i=1}^{\infty}P(X=x_i|Y=y_j)=1$

连续型随机变量的条件分布

给定 y，设对于任意固定的正数ε，

$P(y-\varepsilon<Y\leq y+\varepsilon)>0$, 若对于任意实数 x，

$$\lim_{\varepsilon\to 0^+} P(X\leq x \mid y-\varepsilon<Y\leq y+\varepsilon)$$

$$=\lim_{\varepsilon\to 0^+}\frac{P(X\leq x, y-\varepsilon<Y\leq y+\varepsilon)}{P(y-\varepsilon<Y\leq y+\varepsilon)}$$

存在，则称其为在条件 $Y=y$ 下 X 的**条件分布函数**，记为 $F_{X|Y}(x|y)=P(X\leq x|Y=y)$.

在条件$Y=y$下X的条件分布函数

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

$f_Y(y)>0$时，X 在$Y=y$条件下的条件密度函数为

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

$f_X(x)>0$时，Y 在 $X=x$条件下的条件密度函数为

$$\boxed{f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}}$$

例：已知 $(X,Y)\sim N\left(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho\right)$

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

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

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

$$X|Y=y\sim N\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\sigma_1^2(1-\rho^2)\right)$$

同理，$Y|X=x\sim N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2)\right)$

随机变量的独立性

设***X*,*Y***是两个随机变量，若对任意的***x*,*y***,有

$$F(x,y) = F_X(x)F_Y(y) \quad 即 \quad P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$$

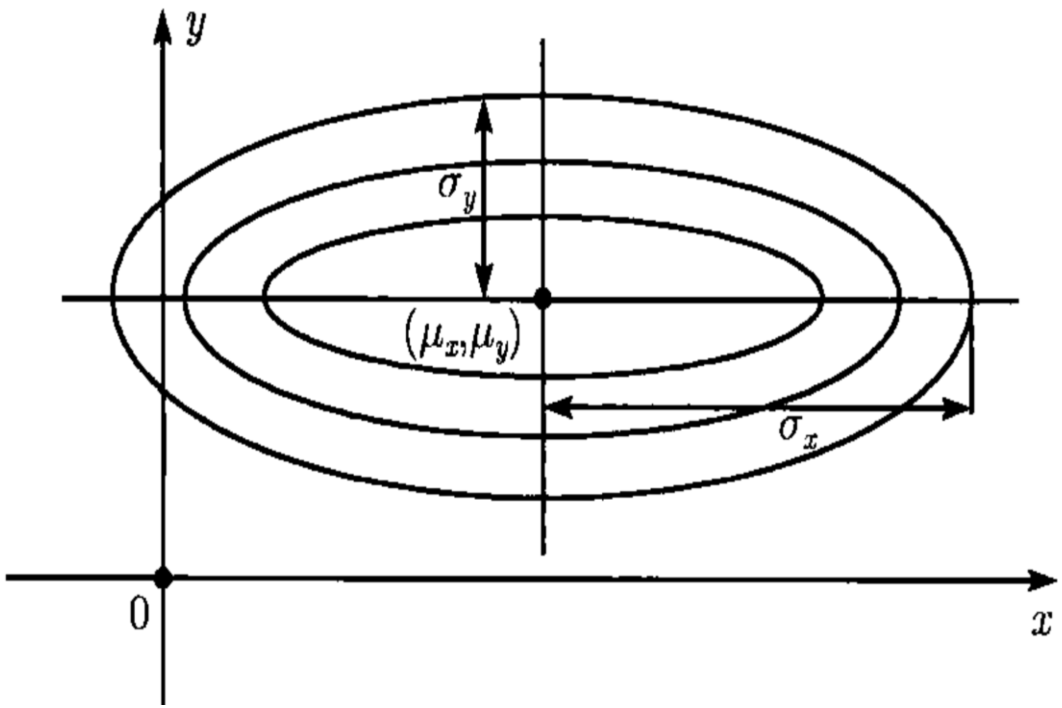
则称***X*,*Y***相互独立.

两个随机变量相互独立时，其联合分布函数等于两个边缘分布函数的乘积，边缘分布完全确定联合分布.

离散型 ***X***与***Y*** 独立 ⟺ 对一切 ***i*** , ***j*** 有$p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$

连续型 ***X***与***Y*** 独立 ⟺ 对任何x ,y 有$f(x,y) = f_X(x)f_Y(y)$

定理 $(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$ X与Y相互独立 $\Longleftrightarrow$ $\rho=0$



相互独立的正态随机变量 X 和 Y 的联合密度函数的等高线

独立的二维正态分布联合密度函数

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_x^2}-\frac{(y-\mu_y)^2}{2\sigma_y^2}\right\}$$

离散型二维随机变量的函数

当(X,Y)为离散型随机变量时，$Z=g(X,Y)$也为离散型随机变量

$$P(Z=z_k)=P(g(X,Y)=z_k)$$

$$=\sum_{g(x_i,y_j)=z_k} P(X=x_i,Y=y_j)=\sum_{g(x_i,y_j)=z_k} p_{ij}$$

$$k=1,2,\cdots$$

若 X 和 Y 是独立的离散型随机变量，则 $Z=X+Y$ 的概率分布（**离散卷积公式**）

$$P(Z=k)=\sum_{i=-\infty}^{+\infty} P(X=i)\cdot P(Y=k-i)$$

二维连续型随机变量函数的分布

(1) 和$Z = X + Y$的分布

设(X, Y)的联合密度为$f(x,y)$, 则

$$\boxed{f_Z(z)=\int_{-\infty}^{+\infty} f(z-y,y)dy}$$

$$\boxed{f_Z(z)=\int_{-\infty}^{+\infty} f(x,z-x)dx}$$

若X, Y相互独立，则

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

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

称为函数$f_X(z)$与$f_Y(z)$的卷积（convolution）

正态随机变量的结论

- 一般地，如果随机变量 X 与 Y 相互独立，且$X \sim N(\mu_1, \sigma_1^2)$,$Y \sim N(\mu_2, \sigma_2^2)$，则 $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

- 更一般地，如果随机变量$X_1, X_2, \cdots, X_n$ 相互独立,

  $X_i \sim N(\mu_i, \sigma_i^2)$,($i$=1, 2, ⋯, n), $a_1, a_2, \cdots, a_n$ 为实常数，则

  $$Z = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

- 独立同分布(independent and identically distributed, i.i.d.)

  若 $X_i \sim N(\mu, \sigma^2)(i = 1,2, \cdots, n)$, 并且 $X_1, X_2, \cdots, X_n$ 相互独立,

  则 $Z = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$.

(2) 一般公式

设$(X，Y)$是二维连续型随机变量, 其联合密度函数为$f(x,y)$

$Z=g(X, Y)$， (X, Y)到(Z, Y)或(Z, X)是一一对应变换

$Y=h(X, Z)$

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, h(x,z)) \left|\frac{\partial h}{\partial z}\right| \mathrm{d}\,x$$

$X=h(Y, Z)$

$$f_Z(z) = \int_{-\infty}^{\infty} f(h(y,z), y) \left|\frac{\partial h}{\partial z}\right| \mathrm{d}\,y$$

设二维随机变量 (X,Y) 的联合密度函数为$f(x,y)$, 如果函数

$$\begin{cases} u = g_1(x,y), \\ v = g_2(x,y) \end{cases}$$

有连续偏导数, 且存在唯一的反函数$\begin{cases} x = x(u,v), \\ y = y(u,v), \end{cases}$

其变换的雅可比行列式

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} \neq 0.$$

若 $\begin{cases} U = g_1(X,Y), \\ V = g_2(X,Y), \end{cases}$ 则 (U,V) 的联合密度函数为

$$f(u,v) = f(x(u,v), y(u,v))|J|.$$

(4) $M=\max(X,Y)$及$N=\min(X,Y)$的分布

设连续型随机变量X , Y 相互独立, X , Y的分布函数分别为 $F_X(x)$, $F_Y(y)$, $M=\max\{X,Y\}$, $N=\min\{X,Y\}$, 求 M ,N 的分布函数.

$$F_M(z)=P(\max\{X,Y\}\le z)=P(X\le z,Y\le z)$$

$$=P(X\le z)P(Y\le z)=F_X(z)F_Y(z)$$

$$F_N(z)=P(\min\{X,Y\}\le z)=1-P(\min\{X,Y\}>z)$$

$$=1-P(X>z,Y>z)=1-P(X>z)P(Y>z)$$

$$=1-[1-F_X(z)][1-F_Y(z)]$$

推广到n维随机变量

设$X_1,\ldots,X_n$是n个相互独立的随机变量,它们的分布函数分别为$F_{X_i}(x)$，$i=1,\ \ldots,n$

- $M=\max(X_1,\ldots,X_n)$的分布函数: $F_M(z) = F_{X_1}(z)F_{X_2}(z)\cdots F_{X_n}(z)$
- $N=\min(X_1,\ldots,X_n)$的分布函数:

$$F_N(z) = 1 - \left[1 - F_{X_1}(z)\right]\left[1 - F_{X_2}(z)\right]\cdots\left[1 - F_{X_n}(z)\right]$$

- 当$X_1,\ldots,X_n$相互独立且具有相同分布函数$F(x)$时，

$$\boxed{F_M(z)=[F(z)]^n} \qquad \boxed{F_N(z)=1-[1-F(z)]^n}$$

密度函数:

$$\boxed{\begin{aligned} f_{max}(z) &= n\left[F(z)\right]^{n-1} f(z) \\ f_{min}(z) &= n\left[1-F(z)\right]^{n-1} f(z) \end{aligned}}$$

三、随机变量的数字特征

数学期望(Expectation)

数学期望又称均值(mean value)

设 离散 r.v. X 的分布列为$P(X=x_k)=p_k,\quad k=1,2,\cdots,$
若无穷级数 $\sum_{k=1}^{+\infty} x_k p_k$ 绝对收敛, 则称其和为 X 的数学期望，记作 $E(X)$， 即

$$\boxed{E(X)=\sum_{k=1}^{+\infty} x_k p_k}$$

设连续 r.v. X 的 d.f. 为$f(x)$，若广义积分$\int_{-\infty}^{+\infty} xf(x)dx$ 绝对收敛, 则称此积分为 X 的数学期望，记作 $E(X)$, 即

$$\boxed{E(X)=\int_{-\infty}^{+\infty} xf(x)dx}$$

注：连续型随机变量X的数学期望E(X)就是概率分布曲线y=f(x)与x轴之间的平面图形的重心的横坐标。

随机变量函数的数学期望

(1) $Y=g(X)$ 的数学期望

- 设离散 r.v. X 的概率分布为

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

若无穷级数 $\sum_{i=1}^{\infty} g(x_i)p_i$ 绝对收敛，则

$$\boxed{E(Y)=\sum_{i=1}^{\infty} g(x_i)p_i}$$

- 设连续 r.v. X 的 pdf 为 $f(x)$

若广义积分 $\int_{-\infty}^{+\infty} g(x)f(x)dx$ 绝对收敛, 则

$$\boxed{E(Y)=\int_{-\infty}^{+\infty} g(x)f(x)dx}$$

(2) $Z=g(X,Y)$的数学期望

❑ 设离散 r.v. (X,Y) 的概率分布为

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

若级数 $\sum\limits_{i,j=1}^{\infty} g(x_i,y_j)p_{ij}$ 绝对收敛，则

$$\boxed{E(Z)=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x_i,y_j)p_{ij}}$$

❑ 设连续 r.v. (X,Y)的联合 pdf 为$f(x,y)$

若广义积分 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy$

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

数学期望的性质

连续型随机变量与离散型随机变量的数学期望的性质类似.

(1) 设C是常数，则$E(C)=C$;

(2) 若C是常数，则$E(CX)=CE(X)$;

(3) $E(X+Y) = E(X)+E(Y)$;

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E(X_i)$$

$$E(c_0 + \sum_{i=1}^{n} c_i X_i) = c_0 + \sum_{i=1}^{n} c_i E(X_i)$$

注意:由$E(XY)=E(X)E(Y)$不一定能推出X,Y独立

(4) 设X、Y独立，则 $E(XY)=E(X)E(Y)$.

当各 X_i独立时成立，$E[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} E(X_i)$

不等式

- If $X \geq 0$, then $\mathrm{E}(X) \geq 0$
- If $a \leq X \leq b$, then $a \leq \mathrm{E}(X) \leq b$
- 设X、Y是两个随机变量，如果$E(X^2)$、$E(Y^2)$都存在，则

$$[E(XY)]^2 <= E(X^2)\, E(Y^2)$$

(柯西-施瓦茨不等式, Cauchy-Schwarz)

方差

若$E[X-E(X)]^2$ 存在, 则称其为随机变量 X 的**方差**(Variance),记为$D(X)$ 或 $V(X)$， $Var(X)$.

$$\boxed{D(X)=E[X-E(X)]^2=E(X^2)-E^2(X)}$$

$\sqrt{D(X)}$为 X 的**标准差**(standard deviation).

方差$D(X)$是随机变量X的函数$g(X)=[X-E(X)]^2$的数学期望 ,描述了随机变量 X 的取值偏离平均值的平均偏离程度**.**

若 X 为离散型 **r.v.**，分布律为$P(X=x_k)=p_k$， $k=1,2,\cdots$

$$\boxed{D(X)=\sum_{k=1}^{+\infty}(x_k-E(X))^2 p_k}$$

若 X 为连续型**r.v.**， 概率密度为 $f(x)$

$$\boxed{D(X)=\int_{-\infty}^{+\infty}(x-E(X))^2 f(x)dx}$$

方差的性质

(1) $D(X)\geq 0$。 $D(X)=0$ 当且仅当 X=常数C 。

(2) 若C是常数,则$D(CX)=C^2 D(X)$

(3) 若X与Y 独立，则$D(X\pm Y)= D(X)+D(Y)$

推广：若$X_1,X_2,\ldots,X_n$相互独立,则

$$D[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} D(X_i)$$

$$D[C_0 + \sum_{i=1}^{n} C_i X_i] = \sum_{i=1}^{n} C_i^2 D(X_i)$$

常见分布的数学期望与方差

分布	期望	方差
$X\sim \mathrm{Bern}(p)$. 0-1分布 $P\{X=k\}=p^k(1-p)^{(1-k)}$，其中 $k=0,1$	p	$p(1-p)$
$B(n,p)$ $P(X=k)=C_n^k p^k(1-p)^{n-k}$,	np	$np(1-p)$
$P(\lambda)$ $P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$,	λ	λ
$f(x)=\begin{cases}\frac{1}{b-a}, & a\le x\le b\\ 0 & 其他\end{cases}$ (a,b)上的均匀分布	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$E(\lambda)$ $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & 其他\end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$N(\mu,\sigma^2)$ $f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
超几何分布	nM/N	$\frac{nM(N-n)(N-M)}{N^2(N-1)}$
$X\sim G(p)$. 几何分布 $P\{X=k\}=(1-p)^{k-1}p$,	$\frac{1}{p}$	$\frac{1-p}{p^2}$

协方差和相关系数

$X\ ,Y$ 的协方差(covariance)：

$$\boxed{\operatorname{cov}(X,Y)=E([X-E(X)][Y-E(Y)])=\sigma_{xy}}$$

X ,Y 的协方差反映了一起波动的程度。

协方差 $Cov(X,Y)$ 反映了随机变量 X 与 Y 的线性相关性：

- **当 $Cov(X,Y)>0$ 时，称 X 与 Y 正相关；**
- **当 $Cov(X,Y)<0$ 时，称 X 与 Y 负相关；**
- **当 $Cov(X,Y)=0$ 时，称 X 与 Y 不相关.** (uncorrelated)

若 X 与 Y 独立， $Cov(X,Y)=0$.

$$\boxed{Cov(X,Y)=E(XY)-E(X)E(Y)}$$

若$D(X)>0, D(Y)>0$,称

$$\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

为X, Y的相关系数(correlation coefficient)，没有量纲.

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

若$\rho_{XY}=0$, 称X, Y不相关(uncorrelated).

$\rho_{XX}=1$ maximally positively correlated

$\rho_{X(-X)}=-1$ maximally negatively correlated

2. 协方差的性质

(1) $\operatorname{cov}(X,Y)=\operatorname{cov}(Y,X)$

(2) $\operatorname{cov}(X,C)=0$

(3) $\operatorname{cov}(a_1X+b_1,a_2Y+b_2)=a_1a_2\operatorname{cov}(X,Y)$

(4) $\operatorname{cov}(X+Y,Z)=\operatorname{cov}(X,Z)+\operatorname{cov}(Y,Z)$

(5) $\operatorname{cov}(X,X)=D(X)$

(6) $D(X\pm Y)=D(X)+D(Y)\pm 2\operatorname{cov}(X,Y)$

$$\boxed{\begin{aligned}&D\left(\sum_{i=1}^{n}X_i\right)=cov\left(\sum_{i=1}^{n}X_i,\sum_{j=1}^{n}X_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{n}cov(X_i,X_j)\\&=\sum_{i=1}^{n}D(X_i)+\sum_{1\le i\ne j\le n}cov(X_i,X_j)=\sum_{i=1}^{n}D(X_i)+2\sum_{1\le i<j\le n}cov(X_i,X_j)\end{aligned}}$$

相关系数的性质

(1) $|\rho_{XY}| \leq 1$

(2) $|\rho_{XY}| = 1$ $\Longleftrightarrow$ 存在常数***a*, *b*(*a*≠0)**，使***P*(*Y*=*aX*+*b*)=1**

即***X***和***Y***以概率**1**线性相关**.**

$$|\rho| = 1 \;\Leftrightarrow\; (X - \mathbf{E}[X]) = c(Y - \mathbf{E}[Y]) \quad \text{(linearly related)}$$

(3) $\rho_{(aX+b)Y} = sign(a)\rho_{XY}$

X,Y 相互独立 $\underset{\not\Leftarrow}{\Rightarrow}$ X,Y 不相关

X,Y 不相关 $\Longleftrightarrow$ $\rho_{XY}=0$

$\Longleftrightarrow$ $\mathrm{cov}(X,Y)=0$

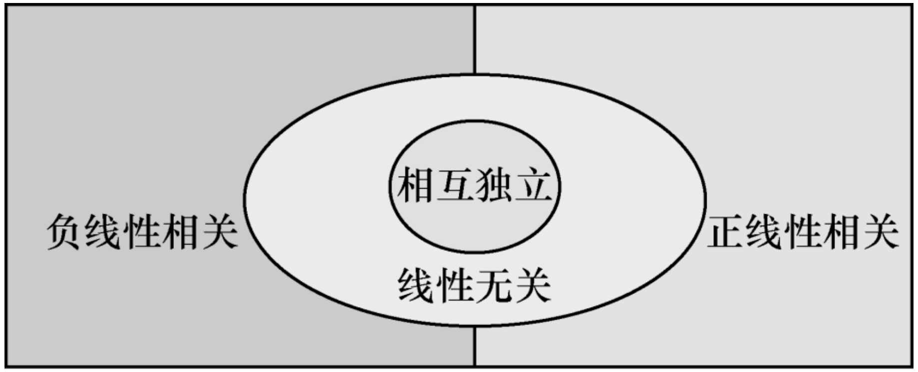
$\Longleftrightarrow$ $\boldsymbol{E(XY)=E(X)E(Y)}$

- 不相关的随机变量间不存在**线性**关系。
- 相互独立的随机变量间不存在**任何**关系。
- 没有线性关系，可以有别的关系，因而不相关不一定独立。

若 $(X,Y)\sim N(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2,\rho)$,

则 X,Y 相互独立 $\Longleftrightarrow$ X,Y 不相关

$\Longleftrightarrow$ $\rho=0$



矩和协方差矩阵

- $E(X^k)$： X 的 k 阶原点矩
- $E((X-E(X))^k)$： X 的 k 阶中心矩
- $E(X^kY^l)$： X,Y 的 $k+l$ 阶混合原点矩
- $E\left((X-E(X))^k(Y-E(Y))^l\right)$： X,Y 的 $k+l$ 阶混合中心矩
- 若$c_{ij}=\mathrm{Cov}\left(X_i,X_j\right), i,j=1,2,\cdots,n$ 都存在, 则称矩阵

$$\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

为 n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的**协方差矩阵**.

例　二维随机变量 $(X,Y)\sim N\left(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho\right)$, 则其协方差矩阵为

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

当X,Y独立时,

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}.$$

当X,Y线性相关时,

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

多维正态分布(Multivariate Gaussian distribution)

利用 均值向量 和 协方差矩阵, 可以表示为:

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)$$

其中，$\boldsymbol{x} = (x_1, x_2, \dots, x_n)^T$，$x_i$均为一维正态随机变量，

x_i 的期望为 μ_i，$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$，

Σ 是 $\boldsymbol{x}$ 的covariance matrix:

$$\Sigma = \begin{bmatrix} \mathrm{cov}(x_1, x_1) & \mathrm{cov}(x_1, x_2) & \dots & \mathrm{cov}(x_1, x_n) \\ \mathrm{cov}(x_2, x_1) & \mathrm{cov}(x_2, x_2) & \dots & \mathrm{cov}(x_2, x_n) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{cov}(x_n, x_1) & \mathrm{cov}(x_n, x_2) & \dots & \mathrm{cov}(x_n, x_n) \end{bmatrix}$$

多维正态分布的性质

性质1 多维正态分布的边缘分布也是多维 (或一维) 正态分布.

设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{pmatrix}_{p-q}^{q}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}_{p-q}^{q}, \boldsymbol{\Sigma} = \overbrace{\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}}^{q \quad p-q} \begin{matrix} q \\ p-q \end{matrix}$, 则

$$\boldsymbol{X}^{(1)} \sim N_q\left(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}\right), \boldsymbol{X}^{(2)} \sim N_{p-q}\left(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}\right)$$

性质2 如果 $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{b}$是 m 维向量, A 是 $m \times n$ 矩阵, 则

$$\boldsymbol{Y} = \boldsymbol{b} + A\boldsymbol{X} \sim N\left(\boldsymbol{b} + A\boldsymbol{\mu}, A\Sigma A^{\mathrm{T}}\right).$$

特别地，对向量 $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)^{\mathrm{T}}$ 和常数 b, 有

$$b + \sum_{j=1}^{n} a_j X_j = b + \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} \sim N\left(b + \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mu}, \boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a}\right).$$

性质3 设 $\boldsymbol{X} = \left(X_1, X_2, \cdots, X_p\right)^{\mathrm{T}} \sim N_p(\mu, \boldsymbol{\Sigma})$, 且 $\boldsymbol{\Sigma}$ 是对角矩阵, 则 $X_1, X_2, \cdots, X_p$ 相互独立且都服从正态分布.

四、大数定律与中心极限定理

切比雪夫不等式(Chebyshev Inequality)

设随机变量 X 的期望$E(X)$与方差 $D(X)$ 存在，则对于任意实数 $\varepsilon > 0$,

$$\boxed{P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}}$$

或 $$\boxed{P(|X - E(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}}$$

理论价值：证明大数定律等

实用价值：估计概率

切比雪夫(Chebyshev)大数定律

设随机变量序列$X_1, X_2, \cdots, X_n, \cdots$相互独立，
且具有相同的数学期望和方差(不一定同分布)

$$E(X_k) = \mu,\ D(X_k) = \sigma^2, \quad k = 1,2,\cdots$$

则 $\forall \varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| \geq \varepsilon\right) = 0$$

或

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right) = 1$$

辛钦大数定律(Wiener-Khinchin law of large Numbers)

设$X_1, X_2, \cdots$是相互独立、**服从同一分布的随机变量序列,且具有数学期望：**$\boldsymbol{E(X_k) = \mu,}$

$\boldsymbol{(k = 1, 2, \cdots),}$ 取前 n 个变量的算术平均 $\frac{1}{n}\sum_{k=1}^{n} X_k$,

则对于任意$\boldsymbol{\varepsilon > 0}$**, 有**

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} = 1.$$

即 $\overline{X}_n \xrightarrow{P} \mu.$

贝努里(Bernoulli)大数定律

设 n_A 是 n 次独立重复试验中事件 A 发生的次数, p (0<p<1)是每次试验中 A 发生的概率, 则

$\forall \varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right) = 0$$

或

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1$$

频率依概率收敛于概率p 即 $\frac{n_A}{n} \xrightarrow{P} p$, 当$n \to +\infty$

林德伯格-列维(Lindburg-Levy)中心极限定理

[独立同分布的中心极限定理] (Central Limit Theorem)

设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布,

且有期望和方差, $E(X_k) = \mu$, $D(X_k) = \sigma^2 > 0$, $k = 1,2,\cdots$,

则对于任意实数 x ,

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \le x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt = \Phi(x)$$

$$\boxed{\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似}}{\sim} N(0,1)}$$

$$\boxed{\sum_{k=1}^{n} X_k \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2)}$$

$$\boxed{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{近似}}{\sim} N(0,1)}$$

$$\boxed{\bar{X} \overset{\text{近似}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)}$$

当n充分大时，n个具有期望和方差的独立同分布的随机变量之和或者平均值近似服从正态分布.

棣莫弗-拉普拉斯(De Moivre-Laplace)中心极限定理

[二项分布以正态分布为极限分布]

设 $Y_n \sim B(n, p)$，$0 < p < 1$，$n = 1,2,\ldots$

则对任一实数 x，有

$$\lim_{n\to\infty} P\left(\frac{Y_n - np}{\sqrt{np(1-p)}} \le x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt = \Phi(x)$$

即 n 足够大时，$Y_n \overset{\text{近似}}{\sim} N(np, np(1-p))$

填空题 1分

设置

设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立, $S_n = X_1 + X_2 + \cdots + X_n$, 则根据列维一林德伯格 (Levy-Lindberg) 中心极限定理, 当 n 充分大时, S_n 近似服从正态分布, 只要 $X_1, X_2, \cdots, X_n$

(A) 有相同的数学期望 (B) 有相同的方差

(C) 服从同一指数分布 (D)有相同的数学期望和方差

答 ([填空1]) (填ABCD之一)

[独立同分布的中心极限定理] (Central Limit Theorem)

设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布, 且有期望和方差, $E(X_k) = \mu$, $D(X_k) = \sigma^2 > 0$, $k = 1,2,\cdots,$ 则对于任意实数 x,

D不能保证同分布

作答