西北工业大学
NORTHWESTERN POLYTECHNICAL UNIVERSITY

数学与统计学院
School of Mathematics and Statistics, NPU

概率论与数理统计

唐亚宁

西北工业大学数学与统计学院

统计学（statistics）---收集、处理、分析、解释数据并从数据中得出结论的科学。是一门综合性科学。应用范围几乎覆盖了社会科学和自然科学的各个领域。

统计学是一门很古老的科学，一般认为其学理研究始于古希腊的亚里士多德时代，迄今已有2300多年的历史。它起源于研究社会经济问题，在两千多年的发展过程中，统计学至少经历了“城邦政情”、“政治算术”和“统计分析科学(十九世纪末兴起）”三个发展阶段。

数理统计---探讨随机现象统计规律性的一门学科，它以概率论为理论基础，研究如何以有效的方式收集，整理和分析受到随机因素影响的数据，从而对所考察的问题做出推理和预测，直至为采取某种决策提供依据和建议。

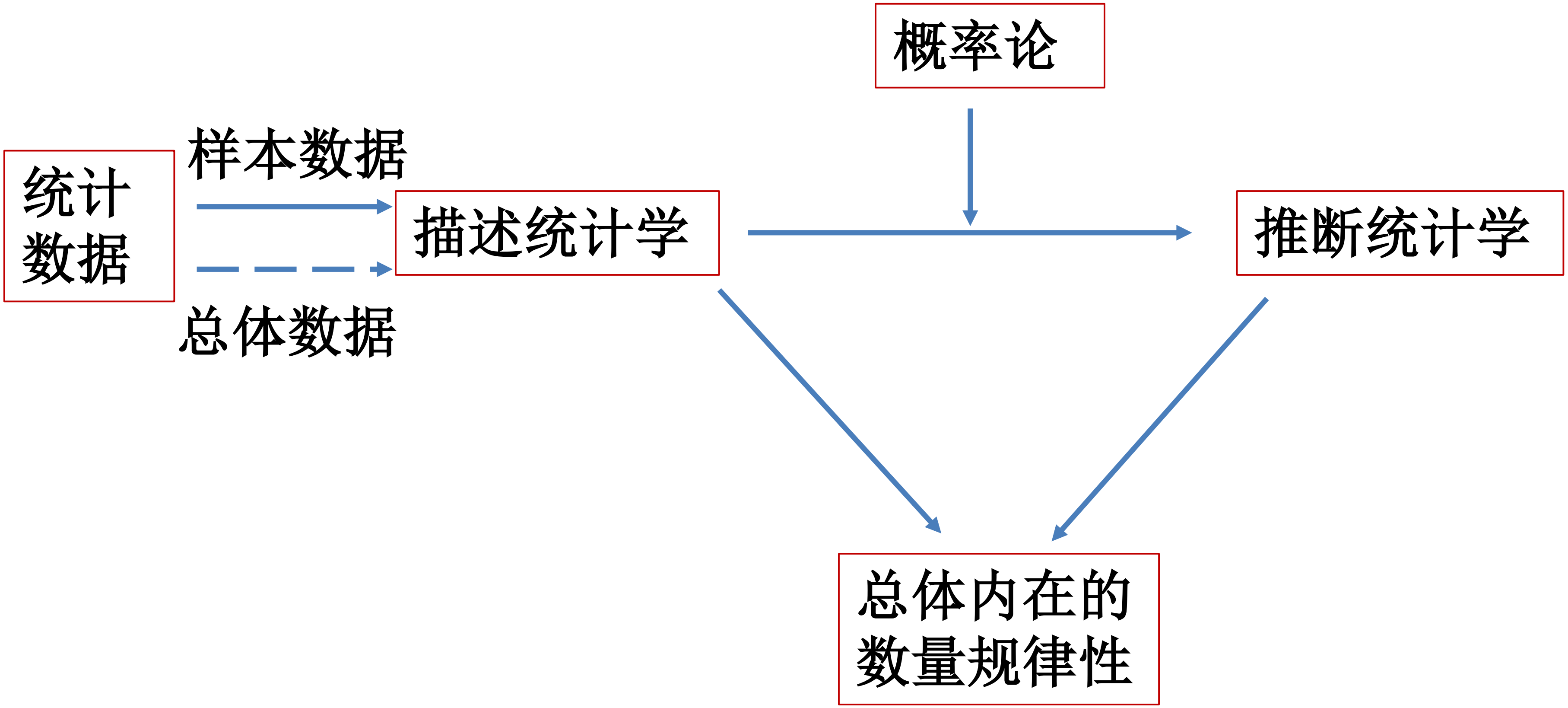
“数理统计”并非独立于统计学的新学科，确切地说，它是统计学在第三个发展阶段所形成的所有收集和分析数据的新方法的一个综合性名词。概率论是数理统计方法的理论基础，但是它不属于统计学的范畴，而是属于数学的范畴。

<table>
<tr><th></th><th>社会经济统计学</th><th>数理统计学</th></tr>
<tr><td rowspan="2">研究方法（共通性）</td><td colspan="2">两者都是统计学的重要分支，都是对事物的统计规律进行研究，都是利用归纳推理的研究方法而不是演绎推理的研究方法</td></tr>
<tr><td>多用描述统计学</td><td>主要用推断统计学</td></tr>
<tr><td>研究范围（不同）</td><td>研究社会经济现象。社会经济现象除了具有随机现象以外，还有确定性现象</td><td>自然现象，社会现象中的随机现象</td></tr>
</table>

统计学的研究方法

描述统计学：数据收集，处理，汇总，图表描述，概括与分析等。图表展示，有均值，方差，众数，中位数，极差，偏度峰度等术语描述。

推断统计学：是研究如何根据样本数据去推断总体数量特征的方法，它是在对样本数据进行描述的基础上，对统计总体的未知数量特征做出以概率形式表述的推断。参数估计，假设检验，方差分析，回归分析等


统计数据
样本数据
总体数据
描述统计学
概率论
推断统计学
总体内在的数量规律性

例如 某厂生产一型号的合金材料，随机的选取100个样品进行强度测试，需要解决下列几个问题：

1、估计这批合金材料的强度均值是多少?（参数的点估计）

2、强度均值在什么范围内？（参数的区间估计）

3、若规定强度均值不小于某个定值为合格，那么这批材料是否合格?(参数的假设检验问题)

4、这批合金的强度是否服从正态分布?（分布检验问题）

5、若这批材料是由两种不同工艺生产的，那么不同的工艺对合金强度有否影响？若有影响，哪一种工艺生产的强度较好？（方差分析问题）

6、若这批合金由几种原料用不同的比例合成，那么如何表达这批合金的强度与原料比例之间的关系？（回归分析问题）

本书我们依次讨论参数的点估计、区间估计、假设检验、方差分析、回归分析

统计就是研究数据的学问

世界三大谎言

谎言lies,

糟糕透顶的谎言damned lies,

<u>统计数据 statistics.</u>

---英国前首相本杰明·迪斯雷利

马克·吐温

数据来源

必须真实

可靠！！

https://v.qq.com/x/page/t0876yw5pnu.html

华为总裁

大数据---数据科学---机器学习---人工智能

强调：

统计推断

数据可视化

试验设计

领域知识

沟通

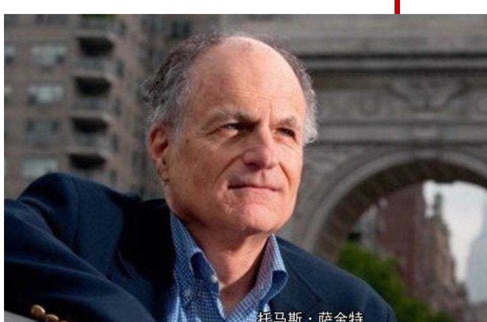
利用：

预测（回归分析）

聚类

贝叶斯理论

是统计学！



2011年诺贝尔经济学奖

概率论补充知识

一. 随机事件的概率

二. 随机变量及其分布

三. 随机向量及其分布

四. 数字特征

一、随机事件的概率

1. 定义 设 Ω 是给定的样本空间，F是 Ω 中的一个事件域，$P(A)(A\in F)$是定义在 F上一个实值集函数，如果它满足条件：

(1) 对任一事件$A\in F$，有$P(A)\geq 0$;　（非负性）

（2）$P(\Omega)=1$;　（规范性）

(3) 若$A_1\in F, i=1,2,\cdots$，且 $A_iA_j=\phi, i\neq j$，则有

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

(可列可加性)

则称$P(A)$是事件A的概率（简称为概率）.

2. 概率的公理化定义的主要性质

- 设A_1，$A_2, \cdots, A_m$两两互斥，则

$$P(A_1 + A_2 + \cdots + A_m)$$
$$= P(A_1) + P(A_2) + \cdots + P(A_m).$$

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

- 对于任意两个事件A, B，有

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

- 若$A \supset B$，则 $P(A - B) = P(A) - P(B).$

且有 $P(A) \geq P(B).$

注：一般地，对任意两个事件A, B, 有

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

3. 条件概率

定义 设$\boldsymbol{A},\boldsymbol{B}$是两个随机事件, 且 $P(A)>0$ 则称

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

为在事件A发生的条件下, 事件$\boldsymbol{B}$发生的条件概率.

性质

1）对于任一事件$\boldsymbol{B}$, $\Rightarrow P(B \mid A) \geq 0;$

2）$\boldsymbol{P(\Omega \mid A)=1;}$

3） $\boldsymbol{P(\overline{B} \mid A)=1-P(B \mid A).}$

4.　事件的相互独立性

由条件概率可知，一般地

$$P(B \mid A) \neq P(B) \quad (P(B) > 0)$$

但

1）掷一颗均匀的骰子两次,

A={第一次掷出6 点}　B={第二次掷出6点}

可知 $P(B|A) = P(B) = \dfrac{1}{6}$

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

定义 设 A, B是两个事件, 如果如下等式成立

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

则称事件 A, B 相互独立。

例1 从一副不含大小王的扑克牌中任取一张，记 $A=$ {抽到 K}, $B =$ {抽到的牌是黑色的}

问事件A, B是否相互独立？

解 $P(A) = \dfrac{4}{52}$; $P(B) = \dfrac{26}{52}$; $P(AB) = \dfrac{2}{52}$

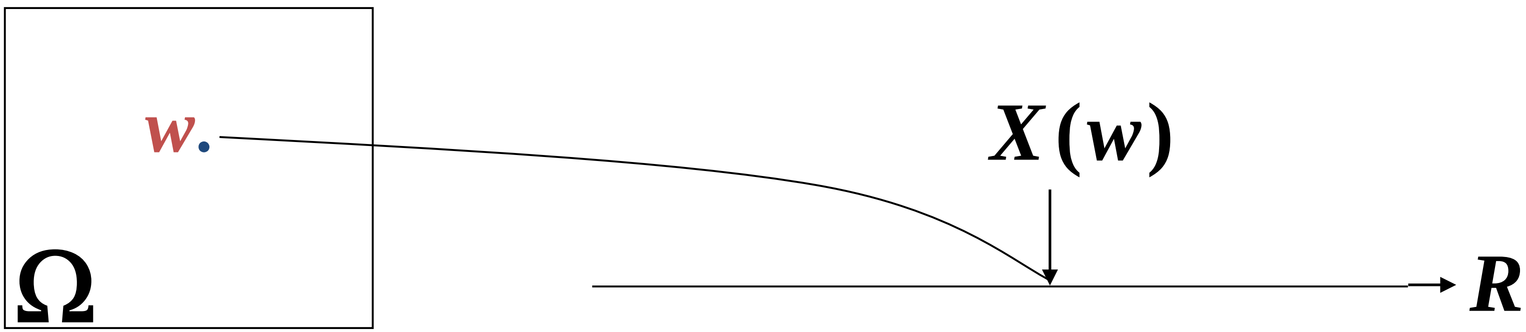
$$\Rightarrow P(AB) = P(A)P(B)$$

即事件A, B相互独立

二　随机变量及其分布

一、随机变量的定义

定义 设随机试验的样本空间 $\Omega=\{\omega\}$, $X=X(\omega)$ 是定义在样本空间Ω上的单值实函数，称$X=X(\omega)$为随机变量。



例1 抛一枚均匀硬币，观察正反面情况。$\Omega=\{H,T\}$

设

$$X=X(\omega)=\begin{cases}1, & \omega=H,\\ 0, & \omega=T.\end{cases}$$ X为随机变量

试验结果的出现是随机的，故 X 的取值也是随机的。

$$P\{X=1\}=P\{\omega=H\}=0.5$$

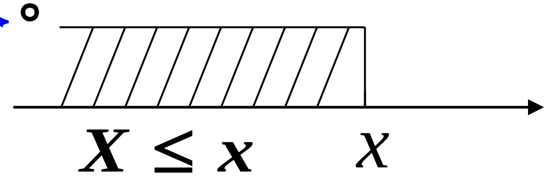
$$P\{X=0\}=P\{\omega=T\}=0.5$$

二、分布函数的概念

定义 设 X 是一个随机变量，x 是任意实数，称函数

$$F(x)=P\{X\le x\}\quad(-\infty<x<+\infty)$$

为 X 的**分布函数**。



分布函数 $F(x)$ 的值就表示 X 落在区间 $(-\infty,x]$ 上的概率.

$$P\{x_1<X\le x_2\}=P\{X\le x_2\}-P\{X\le x_1\}$$
$$=F(x_2)-F(x_1)$$

分布函数的性质

(1) 单调不减性：若 $x_1 < x_2$，则 $F(x_1) \le F(x_2)$

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

$$F(+\infty) = \lim_{x\to+\infty} F(x) = 1;$$

(3) 右连续性： $F(x+0) = \lim\limits_{y\to x+0} F(y) = F(x).$

上述三条性质，也为判别函数是否是分布函数的充要条件。

（1）分布函数完整描述了随机变量的统计规律性

（2）分布函数是一个普通实值函数

三. 随机变量的分类

所有取值可以逐个一 一列举

1 离散型随机变量

例如：“抽验一批产品中次品的个数”，
“电话交换台在一定时间内收到的呼叫次数”等

全部可能取值有无穷多，
充满一个或几个区间

2 连续型随机变量

例如：“电视机的寿命”，

实际中常遇到的“测量误差”等.

1、离散型随机变量及其分布

定义　若随机变量X的全部可能取值是**有限个**或**可列无穷多个**，则称此随机变量是**离散型随机变量**。

例　(1) 扔一均匀硬币三次，出现正面的次数

$$X(w) = X = \{0,1,2,3\}$$ 离散型随机变量

(2) 某一时间段进入商场的人数

$$X = \{0,1,2,3,\cdots\}$$ 离散型随机变量

灯泡的寿命　$X = \{t \mid t \geq 0\}$　非离散型随机变量

定义 设随机变量X的所有可能取值为 x_k，$k=(1,2,\cdots)$

X取各个可能值的概率为 $P\{X=x_k\}=p_k$，

p_k 满足

$(1)\ p_k \geq 0;$ $\qquad (2)\sum_{k=1}^{\infty} p_k = 1;$

判断分布律的条件

则称p_k为离散型随机变量X的概率分布或分布律。

分布律也可用如下表格的形式表示

X	x_1	x_2	$\cdots$	x_k
p_k	p_1	p_2	$\cdots$	p_k

二、常用的离散型随机变量

1. (0—1)分布

定义 若随机变量X的分布律为

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

则称X服从以p为参数的（$0-1$）分布。

X只能取两个值：0, 1

（0—1）分布的分布律也可写成

X	0	1
p_k	$1-p$	p

$$X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

2 二项分布

定义：如果X的分布律为

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k} \quad k=0,1,2,\cdots,n$$

$0<p<1, q=1-p$, 称X服从参数为n, p的二项分布

记为$X \sim B(n,p)$

$X \sim B(n, p)$

二项分布描述的是 n 重贝努里试验中出现

“成功”次数X 的概率分布.

特别$n=1$，二项分布为

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

这就是$(0-1)$分布，常记为 $X \sim B(1, p)$

3. 泊松分布

若随机变量 X 的分布律

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

其中 $\lambda>0$ 是常数, 称 X 服从参数为 λ 的泊松分布,

记为 $X \sim p(\lambda)$

常见的离散型分布：

分布名称	记号	分布律	背景
两点分布 (或 0–1分布)	$X \sim B(1,p)$ $(0<p<1)$	$P\{X=k\}$ $=p^k(1-p)^{1-k}$ $(k=0,1)$	伯努利事件
二项分布	$X \sim B(n,p)$ $(0<p<1)$	$P\{X=k\}$ $=C_n^k p^k(1-p)^{1-k}$ $(k=0,1,\dots,n)$	n重伯努利概型中，事件发生k次的概率
泊松分布	$X \sim P(\lambda)$ $(\lambda>0)$	$P\{X=k\}=\dfrac{\lambda^k}{k!}e^{-\lambda}$ $(k=0,1,2,\cdots)$	稀有事件发生k次的概率
几何分布	$X \sim Ge(p)$ $(0<p<1)$	$P\{X=k\}$ $=(1-p)^{k-1}p$ $(k=1,2,\cdots)$	独立伯努利试验中，A首次发生的试验次数为X.

2 连续型随机变量及其分布

一、定义

如果随机变量 X 的分布函数为

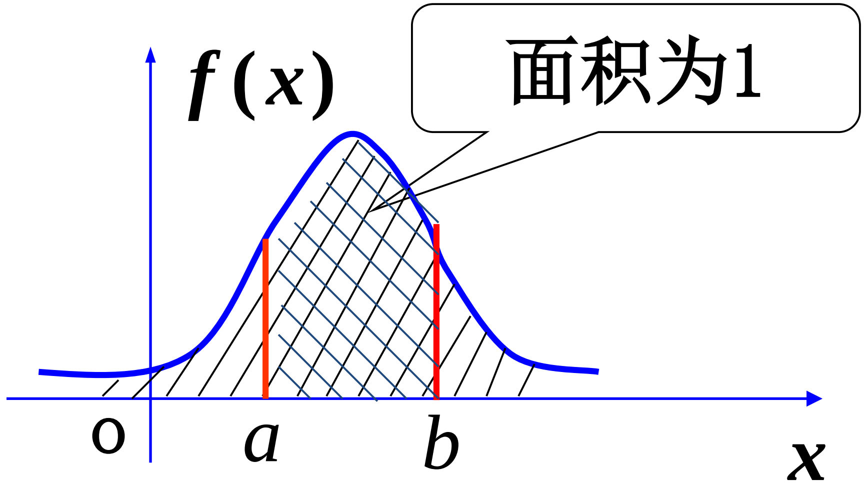
$$F(x)=\int_{-\infty}^{x} f(t)dt$$

其中被积函数 $f(t)\geq 0$，则称 X 为连续型随机变量

称 $f(t)$ 为概率密度函数 或 概率密度。

二. 概率密度的性质

1. $f(x) \geq 0$



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

3. $P\{a < X \leq b\} = F(b) - F(a) = \int_a^b f(x)dx$

$\Rightarrow P\{X = b\} = 0$

4.在 $f(x)$ 的连续点 x 处，有 $F'(x) = f(x)$

5．X 在任意区间G（G可以是开区间,也可以是闭区间，或半开半闭区间；可以是有限区间，也可以是无穷区间）上取值的概率为，

$$P\left\{X \in G\right\} = \int_G f\left(x\right)dx$$

对连续型 $r.v$ X,有

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

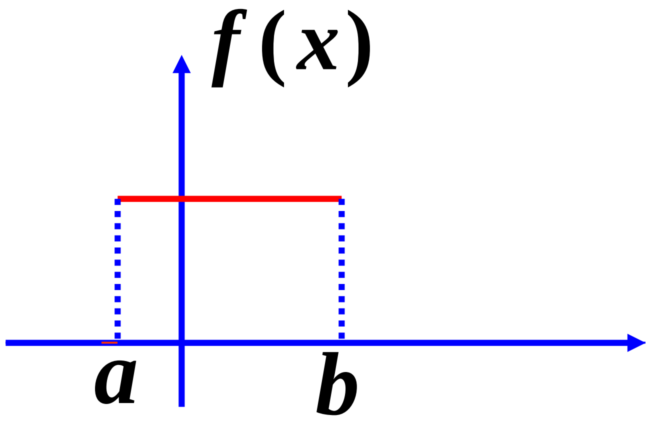
$$= P(a \le X < b) = P(a < X < b)$$

几种常见的分布

一、均匀分布

1. 若X的概率密度为

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



则称X服从(a,b)上的均匀分布，记作 $X\sim U(a,b)$

分布函数为:

$$F(x)=\int_{-\infty}^{x}f(t)dt=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\le x<b,\\ 1, & x\ge b.\end{cases}$$

二、指数分布

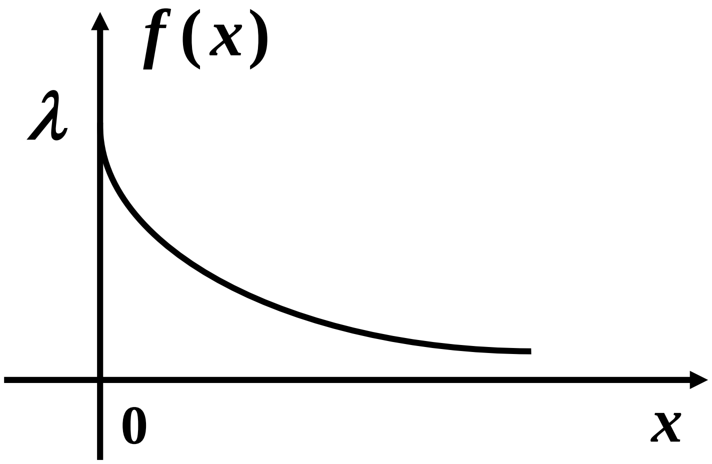
若 随机变量 X 具有概率密度

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

则称 X 服从参数为 λ 的指数分布. 记为 $X\sim E(\lambda)$

X 的分布函数

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



三、正态分布

若X具有概率密度

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

其中μ和σ为常数，且$\sigma>0$，则称X服从参数为μ和σ 的正态分布，或高斯分布.

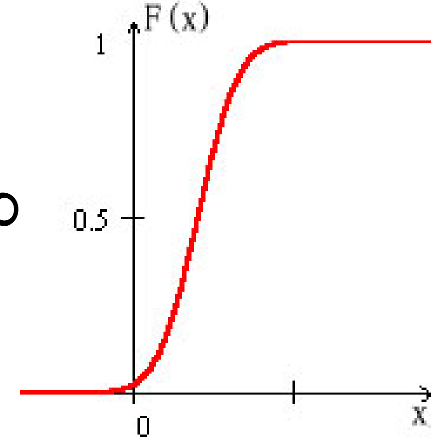
记为 $X\sim N(\mu,\sigma^2)$

$f(x)$ 所确定的曲线称为正态曲线

正态分布的分布函数

若 $X \sim N(\mu,\sigma^2)$, X 的分布函数是

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

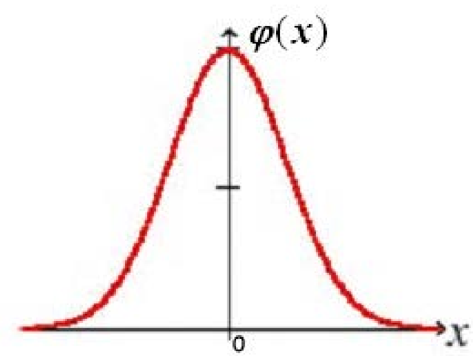


标准正态分布

$\mu=0,\sigma=1$的正态分布称为标准正态分布.

其密度函数和分布函数常用 $\varphi(x)$和 $\Phi(x)$表示

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad -\infty<x<\infty$$



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

若$X \sim N(\mu,\sigma^2)$,

$$\Rightarrow P\{a<X<b\}=P\left\{\frac{a-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{b-\mu}{\sigma}\right\}$$

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

常见的连续型分布

(1) 均匀分布　　$X \sim U[a,b]$

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

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

(2) 指数分布　　$X \sim Exp(\lambda).$

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

$$\Rightarrow F(x)=\begin{cases}1-e^{-\lambda x}, & x>0,\\ 0, & x\le 0.\end{cases}$$

(3) 正态分布*** $X \sim N(\mu, \sigma^2)$

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

$-\infty < x < \infty.$

$$\Rightarrow F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t$$

标准正态分布 $X \sim N(0,1)$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

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

$$\Phi(-x) = 1 - \Phi(x)$$

三、随机向量及其分布

随机向量及其分布

定义2 称n元函数

$$F(x_1,x_2,\cdots,x_n)=P\{X_1\le x_1,X_2\le x_2,\cdots,X_n\le x_n\}$$

为随机向量$X=(X_1,X_2,\cdots,X_n)$的(联合)分布函数。

二维随机变量的联合分布

定义3 设(X,Y)是二维随机变量，对于任意实数x,y，称 $F(x,y)=P\{X\le x,Y\le y\}$ 为(X,Y)的分布函数。

分布函数的性质

$$F(x,y)=P\{X\le x, Y\le y\}$$

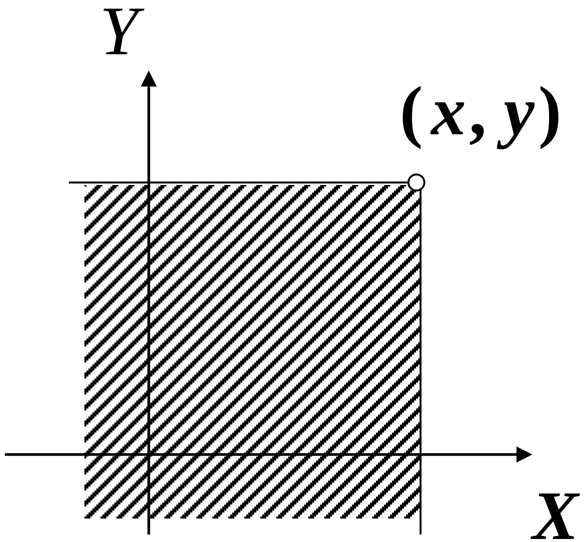
1. $F(x,y)$ 关于x 和y **单调不减**

对于任意固定的 y，当$x_1<x_2$时，

$$F(x_1,y)\le F(x_2,y)$$

对于任意固定的 x，当$y_1<y_2$时，

$$F(x,y_1)\le F(x,y_2)$$



2. $0 \le F(x, y) \le 1$

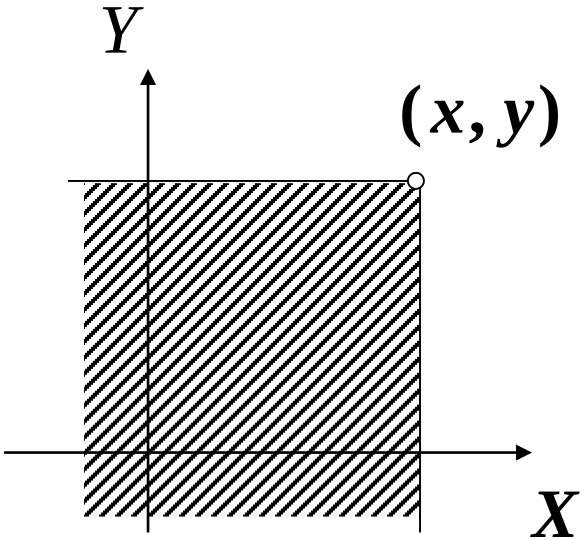
$$F(+\infty,+\infty) = \lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1$$

$$F(-\infty,-\infty) = \lim_{\substack{x \to -\infty \\ y \to -\infty}} F(x, y) = 0$$

$\forall\ x, y,$有

$$F(x,-\infty) = \lim_{y \to -\infty} F(x, y) = 0$$

$$F(-\infty, y) = \lim_{x \to -\infty} F(x, y) = 0$$

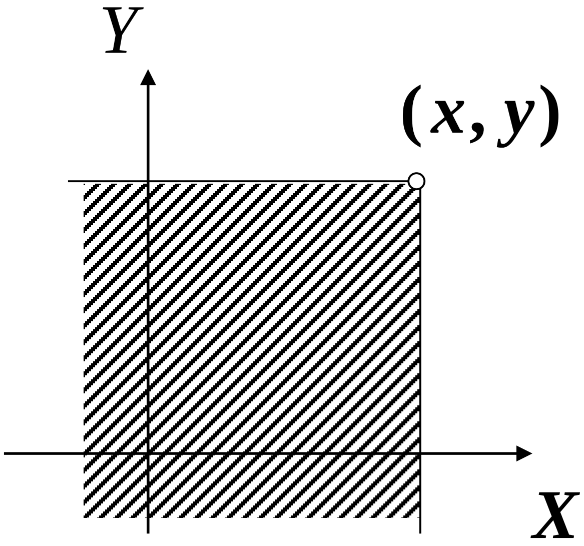


3. $F(x,y)=F(x+0,y)$

即 $F(x,y)$ 关于x右连续

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

即$F(x,y)$ 关于y右连续



二维离散型随机变量

定义4 **若二维随机变量 (X, Y) 的所有可能取值是有限多对或可列无限多对,则称(X, Y)为二维离散型随机变量.**

定义5 **设 (X, Y) 所有可能取值为 (x_i, y_j)，则称**

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

为随机变量 (X, Y) 的分布律。

性质： 1) $p_{ij} \geq 0$ 2) $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij} = 1$

分布律的表格表示

$X \backslash Y$	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}	$\cdots$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

二维连续型随机变量

定义 **设二维随机变量** (X,Y) **的分布函数** $F(x,y)$ **若存在非负函数** $f(x,y)$ **对于任意的** x,y **，有**

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

则称 $f(x,y)$**为**(X,Y)**的概率密度**。

概率密度的性质

1) $f(x,y) \geq 0$

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

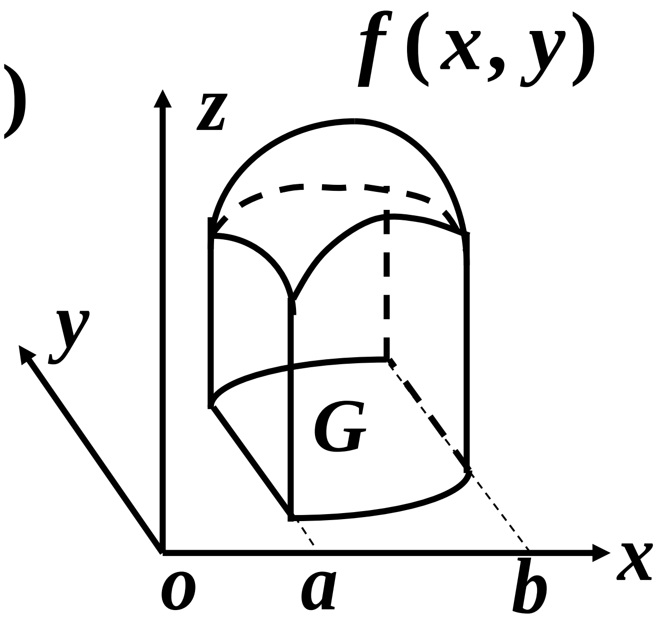
3) 若 $f(x,y)$ 在点 (x,y) 处连续, 则有

$$\frac{\partial^2 F(x,y)}{\partial x\,\partial y} = f(x,y)$$

4） 设 G 是 xoy 平面上的任意一个区域，则有

$$\boxed{P\{(X,Y)\in G\}=\iint\limits_{G} f(x,y)dxdy}$$

(表示以 G 为底，以曲面 $f(x,y)$ 为顶面的曲顶柱体的体积)



二维正态分布

(X,Y) 的概率密度为

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

其中 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 都是常数, 且$\sigma_1>0$ ， $\sigma_2>0$ ， $-1<\rho<1$，则称(X,Y)服从参数为$\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$的二维正态分布，记为

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

边缘分布

（二） 边缘分布律（离散型）

设 (X,Y)的分布律为

$$P\{X=x_i,Y=y_j\}=p_{ij}$$

则 (X,Y)关于 X 的边缘分布律为 $(i,j=1,2,\cdots)$

$$P\{X=x_i\}=\sum_{j=1}^{\infty}P\{X=x_i,Y=y_j\}$$

$$=\sum_{j=1}^{\infty}p_{ij}\ \text{记为}\ p_{i\bullet}$$

则有：$p_{i\bullet}\geq 0;\quad \sum_{i=1}^{\infty}p_{i\bullet}=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}p_{ij}=1$

同理 $P\{Y=y_j\}=\sum_{i=1}^{\infty} p_{ij}$ 记为 $p_{\bullet j}$

称为 (X,Y)关于 Y 的边缘分布律

则有: $p_{\bullet j} \geq 0;\quad \sum_{i=1}^{\infty} p_{\bullet j} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij} = 1$

通常用以下表格表示 (X,Y)的分布律和边缘分布律

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

（三） 边缘概率密度（连续型）

若 (X,Y) 是二维连续型随机变量，其概率密度为 $f(x,y)$，则

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

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

同理

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

随机变量的独立性

设X_i的分布函数为$F_i(x)$,它们的联合分布函数为$F(x_1,x_2,\cdots,x_n)$,则上式等价于

$$\boldsymbol{F(x_1,x_2,\cdots,x_n)=F_1(x_1)\ F_2(x_2)\cdots F_n(x_n),}$$

可以证明：

(1) 若随机变量X_1，$X_2,\cdots$，X_n相互独立，则其中任意$m(2\le m\le n)$个随机变量也相互独立；

(2) 若随机变量X_1，$X_2,\cdots$，X_n相互独立，且它们的**函数$\boldsymbol{f_i(X_i)(i=1,2,\cdots,n)}$也是独立的随机变量。**

二维随机变量的相互独立性

定义　若二维随机变量(X,Y)对任意实数x, y, 都有

$$P\{X \le x, Y \le y\} = P\{X \le x\} \cdot P\{Y \le y\}$$

即

$$F(x,y) = F_X(x) \cdot F_Y(y)$$

成立，则称随机变量 X 与 Y 是相互独立的。

当X,Y独立时,由X,Y的边缘分布可以唯一决定(X,Y)的联合分布。

1) 对于离散型随机变量

$$X与Y独立 \Leftrightarrow$$

$$P\{X = x_i, Y = y_j\} = P\{X = x_i\} \cdot P\{Y = y_j\}$$

$$\Leftrightarrow p_{ij} = p_{i\bullet} \times p_{\bullet j} \qquad \forall\, i, j$$

2) 对于连续型的随机变量

X与Y独立 $\Leftrightarrow f(x, y) = f_X(x) f_Y(y)$ 几乎处处成立

可直接推广至两个以上随机变量的相互独立性

➢ 正态分布性质

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

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

$$\Rightarrow X/y\sim N(\mu_2+\rho\sigma_2(x-\mu_1)/\sigma_1,\sqrt{1-\rho^2}\sigma_2)$$

$$Y/x\sim N(\mu_1+\rho\sigma_1(y-\mu_2)/\sigma_2,\sqrt{1-\rho^2}\sigma_1)$$

$\Rightarrow$ X和Y独立 $\Leftrightarrow \rho=0$

(2) $X \sim N(\mu, \sigma^2)$,

$$\Rightarrow \quad \frac{X-\mu}{\sigma} \sim N(0,1).$$

$$\Rightarrow \quad Y = aX + b \sim N(a\mu + b, (a\sigma)^2).$$

(3) $X_1, X_2, \cdots, X_n$ 相互独立，且

$X_i \sim N(\mu_i, \sigma_i^2), \quad i = 1, 2, \cdots, n$

则有

$$a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + b$$

$$\sim N(a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n + b, \sum_{i=1}^{n} a_i^2 \sigma_i^2).$$

➢ **分布函数法** $F_Y(y) = P\{Y \le y\} = P\{f(X) \le y\}$

$$= \int_{f(x)\le y} p_X(x)\mathrm{d}x \qquad (-\infty < x < \infty)$$

$$p_Y(y) = F_Y'(y)$$

➢ **公式法**

1. **判断** $Y = f(X)$是否是严格单调函数,

2. **反解** 由$y = f(x)$得$x = f^{-1}(y)$,

3. **求导** $[f^{-1}(y)]'$

4. **代公式**

$$p_Y(y) = \begin{cases} p_X[f^{-1}(y)] \cdot \left|[f^{-1}(y)]'\right|, & \alpha < y < \beta, \\ 0, & \text{其它}. \end{cases}$$

其中 $\alpha = \min\{f(a), f(b)\}$, $\beta = \max\{f(a), f(b)\}$.

四、 数字特征

一、随机变量的数学期望

定义1　设离散型随机变量的分布律为

$$P\{X = x_k\} = p_k, \quad k = 1, 2, 3\cdots$$

如果级数 $\sum_{k=1}^{\infty} x_k p_k$ <u>绝对收敛</u>，则级数 $\sum_{k=1}^{\infty} x_k p_k$ 的和称为随机变量X的数学期望，简称期望或均值。

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

若 $\sum_{k=1}^{\infty} x_k p_k$<u>不绝对收敛</u>，则$X$的数学期望不存在。

定义2　设连续型随机变量X 的概率密度为 $f(x)$,

若积分 $\int_{-\infty}^{+\infty} x f(x)dx$ 绝对收敛，则称该积分值为随机变量X 的数学期望或平均值，简称期望或均值

记为 $E(X)$ 即$E(X)=\int_{-\infty}^{+\infty} x f(x)dx$

数学期望的性质

(1). 设C是常数，则$E(C)=C$；

(2). 若C是常数，则$E(CX)=CE(X)$;

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

推广： $E[\prod_{i=1}^{n} X_i]=\prod_{i=1}^{n} E(X_i)$ （当X_i独立时）

(4). 设X、Y独立，则 $E(XY)=E(X)E(Y)$;

注意：由$E(XY)=E(X)E(Y)$不一定能推出X,Y独立

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

随机变量函数的数学期望

定理 设随机变量Y 是随机变量X 的函数，$Y = g(X)$

1) 设X 为离散型随机变量，其分布律为

$$P\{X = x_k\} = p_k, \quad k = 1, 2, \cdots$$

若级数 $\sum_{k=1}^{\infty} g(x_k)\, p_k$ 绝对收敛,则有

$$E(Y) = E[g(X)] = \sum_{k=1}^{\infty} g(x_k)\, p_k$$

2) 设X 为连续型随机变量，其概率密度为 $f(x)$,

若积分 $\int_{-\infty}^{+\infty} g(x)f(x)dx$绝对收敛,则有

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

$$E[g(X)]=\int_{-\infty}^{+\infty} g(x)\,dF(x)$$

例3设X 服从 N (0，1) 分布，求E (X^2), E (X^3), E (X^4)

解：

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

$$E(X^2)=\int_{-\infty}^{\infty}\frac{x^2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=-\int_{-\infty}^{\infty}\frac{x}{\sqrt{2\pi}}de^{-\frac{x^2}{2}}=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

$$=1$$

$$E(X^3)=\int_{-\infty}^{\infty}\frac{x^3}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=0$$

$$E(X^4)=\int_{-\infty}^{\infty}\frac{x^4}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=-\int_{-\infty}^{\infty}\frac{x^3}{\sqrt{2\pi}}de^{-\frac{x^2}{2}}$$

$$=3$$

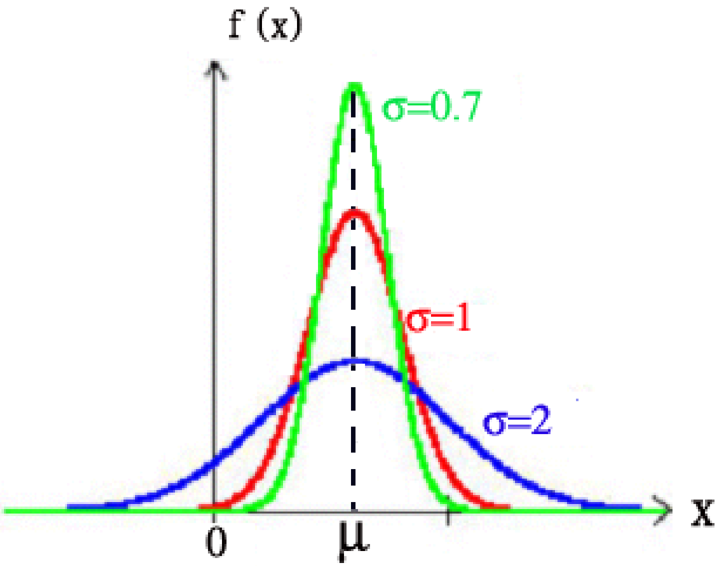
二、方差的概念

定义 设X是一个随机变量，若$E\{[X-E(X)]^2\}$存在，则称$D(X)=E\{[X-E(X)]^2\}$为X的方差。

方差的算术平方根$\sqrt{D(X)}=\sigma(X)$称为均方差或标准差。

方差刻划了随机变量的取值对于其数学期望的离散程度

若X的取值比较集中，则方差较小；若X的取值比较分散则方差较大．



$$D(X)=\begin{cases}\sum_{k=1}^{\infty}[x_k-E(X)]^2 p_k\ , \\ \int_{-\infty}^{\infty}[x-E(X)]^2 f(x)dx\ .\end{cases}$$

(2)方差描述了随机变量*X*的取值与其均值的偏离程度。

计算方差的简便公式：$D(X)=E(X^2)-[E(X)]^2$

方差的性质

性质1 设 C 是常数, 则有 $D(C)=0.$

证 $D(C)=E(C^2)-[E(C)]^2=C^2-C^2=0.$

性质2 设 X 是一个随机变量, k是常数, 则有

$$D(kX)=k^2D(X).$$

证 $D(kX)=E[kX-E(kX)]^2$

$$=k^2E[X-E(X)]^2=k^2D(X).$$

性质3 设随机变量 X, Y 相互独立，且$D(X), D(Y)$ 存在，则 $D(X \pm Y) = D(X) + D(Y)$.

证 $D(X \pm Y) = E[(X \pm Y) - E(X \pm Y)]^2$

$$= E\{[X - E(X)] \pm [Y - E(Y)]\}^2$$

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

$$\pm 2E\{[X - E(X)][Y - E(Y)]\}$$

$$= D(X) + D(Y).$$

推广 若 $X_1, X_2, \cdots, X_n$ 相互独立,则有

$$D\left(a_1X_1 + a_2X_2 + \cdots + a_nX_n\right)$$

$$= a_1^2 D(X_1) + a_2^2 D(X_2) + \cdots + a_n^2 D(X_n).$$

6．常见分布的方差

(1)．（0–1）分布　　参数为p

X	0	1
p	$1-p$	p

$$E(X)=p$$

$$DX=E(X^2)-[EX]^2=p-p^2=p(1-p)$$

$$EX=p$$
$$DX=p(1-p)$$

(2). 二项分布 $X \sim b(n, p)$

$$X = X_1 + X_2 + \cdots + X_n$$

其中 $X_i \sim (0-1)$分布，且 $X_1, X_2, \cdots, X_n$ 相互独立。

$$P(X_i = 1) = p \qquad P(X_i = 0) = 1 - p$$

则由方差的性质可得

$$D(X) = D(X_1) + D(X_2) + \cdots + D(X_n)$$

$$= nD(X_i)$$

$$= np(1-p)$$

$$E(X) = np$$

$$D(X) = np(1-p)$$

(3)．泊松分布 $X \sim \pi(\lambda)$ 参数为 λ

分布律为 $P\{X=k\}=\dfrac{\lambda^k e^{-\lambda}}{k!} \quad (k=0,1,2,\cdots)$

$$E(X)=\lambda$$

$$E(X^2)=\sum_{k=1}^{\infty} k^2 \cdot \frac{\lambda^k e^{-\lambda}}{k!}=\sum_{k=1}^{\infty}(k-1+1)\cdot\frac{\lambda^k e^{-\lambda}}{(k-1)!}$$

$$=\sum_{k=1}^{\infty}\frac{\lambda^k e^{-\lambda}}{(k-2)!}+\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

$$=\lambda^2 e^{-\lambda}\cdot e^{\lambda}+\lambda e^{-\lambda}\cdot e^{\lambda}=\lambda^2+\lambda$$

$$D(X)=E(X^2)-E^2(X)$$

$$=\lambda^2+\lambda-\lambda^2=\lambda$$

$E(X)=\lambda$

$D(X)=\lambda$

（4）．均匀分布 $X \sim U(a,b)$ 参数为 a,b

密度函数 $$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b, \\ 0, & others.\end{cases}$$

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

（5） 指数分布 参数为θ

密度函数 $f(x)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}, & x>0,\\ 0, & x\le 0.\end{cases}$

$$E(X)=\theta$$

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

$$=\int_{-\infty}^{+\infty}(x-\theta)^2\frac{1}{\theta}e^{-x/\theta}dx=\cdots=\theta^2$$

$$E(X)=\frac{1}{\lambda}\ or\ \theta,\ D(X)=\frac{1}{\lambda^2}\ or\ \theta^2$$

（6）正态分布 $X \sim N(\mu,\sigma^2)$ 参数为 $\boldsymbol{\mu},\boldsymbol{\sigma}^2$

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

$$E(X)=\mu,\ D(X)=\sigma^2$$

特别，当 $X \sim N(0,1)$ 时 $E(X)=0,\ D(X)=1$

注：服从正态分布的随机变量完全由它的数学期望和方差所决定。

注：为了方便计算，常对X进行标准化。即当X的期望和方差都存在时，考虑它的标准化。

$$Y=\frac{X-EX}{\sqrt{DX}}$$

EX, DX均为常数。

$$EY=\frac{1}{\sqrt{DX}}E(X-EX)=\frac{1}{\sqrt{DX}}(EX-EX)=0$$

$$DY=\frac{1}{DX}D(X-EX)=\frac{1}{DX}DX=1$$

称Y是随机变量X的标准化了的随机变量。

$$X\sim N(\mu,\sigma^2)\Rightarrow Y=\frac{X-EX}{\sqrt{DX}}\sim N(0,1)$$

常见离散型分布对应的数学期望与方差

分布	分布律	$E(X)$	$D(X)$
0-1分布 $X \sim B(1, p)$	$P\{X=k\}=p^k(1-p)^{1-k}$ $k=0,1$	p	$p(1-p)$
二项分布 $X \sim B(n, p)$	$P\{X=k\}=C_n^k p^k(1-p)^{n-k}$ $k=0,1,2,\ldots,n$	np	$np(1-p)$
泊松分布 $X \sim P(\lambda)$	$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$ $k=0,1,2,\ldots$	λ	λ
几何分布	$P\{X=k\}=(1-p)^{k-1}p$ $k=1,2,\ldots$	$\frac{1}{p}$	$\frac{q}{p^2}$

常见连续型分布的数学期望与方差

分布	分布律或分布密度	$E(X)$	$D(X)$
均匀分布 $X\sim U[a,b]$	$p(x)=\begin{cases}\dfrac{1}{b-a}, & x\in[a,b]\\ 0, & 其他\end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
正态分布 $X\sim N(\mu,\sigma^2)$	$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
指数分布 $X\sim E(\lambda)$	$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & 其他\end{cases}$ $(\lambda>0)$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$

➢ 协方差，相关系数

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

$$=E(XY)-E(X)E(Y).$$

$$\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$$

$\operatorname{cov}(aX, bY)=ab\operatorname{cov}(X,Y)$, a, b为常数.

$$\operatorname{cov}(X_1+X_2,Y)=\operatorname{cov}(X_1,Y)+\operatorname{cov}(X_2,Y).$$

随机变量的矩

- K阶原点矩

$$\alpha_k = E(X^k), \quad k = 1, 2, \cdots, n$$

- **K阶中心矩**

$$\beta_k = E\{(X - EX)^k\}, \quad k = 2, 3, \cdots, n$$

林德贝格-列维中心极限定理

设随机变量$X_1, X_2, \ldots, X_n$相互独立同分布,

且具有数学期望与方差

$$E(X_i) = \mu, \quad D(X_i) = \sigma^2 \neq 0 \quad (i=1, 2, \ldots, n)$$

则随机变量 $$Y_n^* = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

的分布函数$F_n(x)$ 对于任意 x 满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{Y_n^* \leq x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t$$

西北工業大學
NORTHWESTERN POLYTECHNICAL UNIVERSITY

数学与统计学院
School of Mathematics and Statistics, NPU

Thank You !