上节提要

随机变量

离散型随机变量及其分布律

几种离散型随机变量

上节提要

1. (0-1)分布(或两点分布)分布律为

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

2. 二项分布 b(n, p) 分布律为

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

3. 泊松分布 π (λ) 分布律为

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

清楚上述几种分布之间的关系!

§ 3 随机变量的分布函数

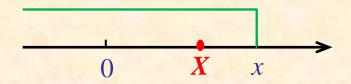
对于非离散型RV,其可能的取值不能一一列出,因此不方便像离散型RV那样用分布律来描述。另外,工程上感兴趣的常是RV的取值落在某个区间内的概率。为了能从数学上对离散型RV和非离散型RV统一研究,于是引入了RV分布函数的定义。

一、分布函数的定义

设X是RV, x是任意实数, 称函数

$$F(x) = P\{X \leq x\}, x \in R$$

为X的分布函数 Cumulative Distribution Function, 简记为CDF。



几何解释 如果将X 看作<u>数轴上随机点的坐标</u>,那么分布函数 F(x) 在x 处的值就表示 X 落在区间 $(-\infty, x]$ 上的概率。

说明

- ① 分布函数是一个普通函数;
- ②分布函数完整地描述了RV的统计规律。

例如:对任意实数a, $x_1 < x_2$ 有

$$\mathbf{P}\{X \leqslant a\} = F(a)$$

$$P\{X > a\} = 1 - P\{X \le a\} = 1 - F(a)$$

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

③ RV在开闭区间内概率,对连续型RV是不严格的,但对离散型RV是严格的。

二、CDF的性质

1. 不减性 F(x) 是一个不减函数;

即对任意实数 $x_1 < x_2$ 有,

$$F(x_2) - F(x_1) = P\{x_1 < X \le x_2\} \ge 0$$

2. 有界性

$$0 \leq F(x) \leq 1$$
, 且 $F(-\infty) = 0$ 和 $F(+\infty) = 1$;

利用有界性可以确定CDF中的待定常数。

3.F(x)右连续,即F(x+0) = F(x)。

性质1~3是鉴别一个函数是否为某个RV的CDF的充要条件。

三、离散型RV分布函数的求法

一般地,设离散型RV的分布律为 $P\{X=x_k\}=p_k, k=1,2,...$

则其CDF为
$$F(x) = P\{X \le x\} = P\left\{\bigcup_{x_i \le x} (X = x_i)\right\}$$

$$\mathbb{P} F(x) = \sum_{x_k \le x} p_k$$

因此,分布函数F(x)是X取 $\leq x$ 的所有值 x_k 的概率之和,且 F(x)在 $X = x_k$ 处有跳跃,其跳 跃值为 p_k 。

具体求解时,先根据X的取值情况将分布函数的定义域($-\infty$, $+\infty$)分为若干个区间,再在每个区间上讨论F(x)的取值。

例1 设X的分布律为

X
 -1
 1
 2
 求X的CDF,

 P
 1/3
 1/2
 1/6

$$P(0 < X \le 2)$$
。

解:
$$F(x) = \sum_{x_k \le x} p_k$$
 — CDF — PMF — PMF

$$P\{0 < X \le 2\} = F(2) - F(0) = 1 - 1/3 = 2/3$$

= $P(X = 1) + P(X = 2) \perp \ddagger$

§4 连续型RV及其概率密度

一、概率密度的定义

如果对于RV X 的分布函数F(x),存在非负可积函数 f(x), $\forall x \in R$ 有

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

连续型RV 的分布函数 是连续函数

则称X为连续型随机变量 Continuous RV,其中 f(x)称为X的概率密度函数 Probability Density Function,简称概率密度,简记为PDF。

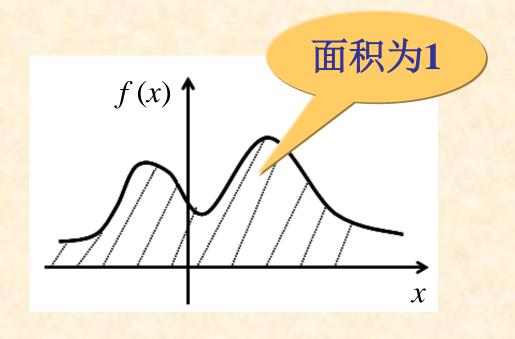
二、PDF的性质

1. 非负性

$$f(x) \geqslant 0;$$

2. 归一性

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



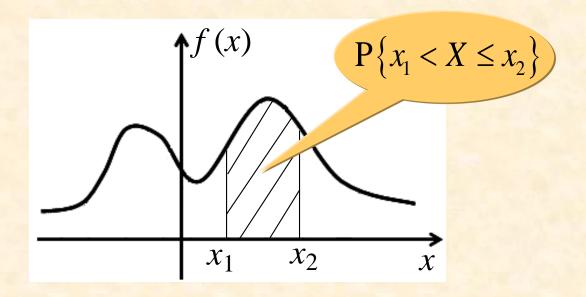
利用归一性可以确定PDF中所含的待定常数。

利用性质1和2可以检验一个函数能否作为连续型RV的PDF。(如P59第32题)

3. 对于任意实数 x_1 , x_2 $(x_1 \leq x_2)$, 有

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

$$= \int_{x_1}^{x_2} f(x) dx \circ$$



4. 若f(x)在点x处连续,则有 F'(x) = f(x)。

进一步理解f(x):若f(x)在点x处连续,由性质4和导数定义知

$$f(x) = F'(x) = \lim_{\Delta x \to 0^{+}} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0^{+}} \frac{P\{x < X \le x + \Delta x\}}{\Delta x}$$
 维密度

上式若不计高阶无穷小,有 $P\{x < X \leq x + \Delta x\} \approx f(x) \Delta x$

据此可知: f(x) 描述了RVX 在点x 附近<u>单位长度区</u> 间内取值的概率。

注意
$$f(x) \neq P\{X = x\}$$

但它的大小反映出X在点x附近取值的概率大小。

连续型RV的一个重要特点:

连续型RV取任一指定实数值a的概率均为0,即

$$P\{ X = a \} = 0$$

据此可知:

①连续型RV取值落在某一区间内的概率是不受区间 开闭的影响。即:

$$P\{a < X \le b\} = P\{a \le X < b\} = P\{a \le X \le b\} = P\{a < X < b\}$$

②虽然 $P\{X=a\}=0$,但事件 $\{X=a\}$ 并非不可能事件.

$$A = \emptyset \Rightarrow P(A)=0; \quad A = S \Rightarrow P(A)=1$$

反之不一定成立

例1 设RV X的PDF为
$$f(x) = \begin{cases} ke^{-3x}, & x > 0 \\ 0, & else \end{cases}$$

求(1) k值; (2) F(x); (3) $P\{X>0.1\}$ 。

解: (1)
$$\int_{-\infty}^{+\infty} f(x) dx = 1 \implies \int_{0}^{\infty} k e^{-3x} dx = 1 \implies k = 3$$

(2)
$$F(x) = \int_{-\infty}^{x} f(t) dt = \begin{cases} 0, & x \le 0 \\ \int_{0}^{x} 3e^{-3x} dt, & x > 0 \end{cases} = \begin{cases} 0, & x \le 0 \\ 1 - e^{-3x}, & x > 0 \end{cases}$$

三、几种常见连续型RV

1、均匀分布 Uniform Distribution

若RVX具有PDF



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

称X在区间(a, b)上服从均匀分布,记为 $X \sim U(a, b)$ 。

易知
$$f(x) \ge 0$$
 和 $\int_{-\infty}^{+\infty} f(x) dx = 1$ 。

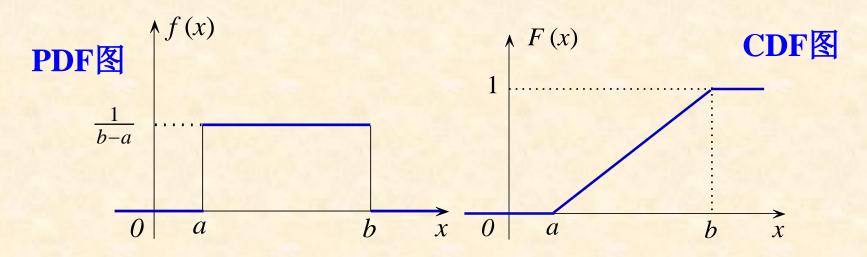
$若X\sim U(a,b)$

(1).
$$\forall a \le x_1 < x_2 \le b$$
,
 $\text{IIP}\{x_1 < X \le x_2\} = \int_{x_1}^{x_2} f(x) dx = \frac{x_2 - x_1}{b - a}$

由此可知: X 落在(a,b)区间中任一子区间内的概率 只依赖于子区间的长度而与子区间的位置无关,即X 落在区间(a,b)中任意等长度的子区间内的可能性相 同,所谓的均匀就是指这种等可能性。

(2). X的CDF为
$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le X < b \\ 1, & X \ge b \end{cases}$$

注:分段定义的连续型RV的CDF F(x),由于F(x)连续,其定义域中各子区间的端点,只要求表达清楚,属于哪个区间无所谓,也没必要与PDF f(x)一致。



均匀分布常见于下列情形:

如在数值计算中,由于四舍五入,小数点后某一位小数引入的误差;

公交线路上两辆公共汽车前后通过某汽车停车 站的时间,即乘客的候车时间等。

2、指数分布 Exponential Distribution

若RV X具有PDF

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x > 0\\ 0, & else \end{cases}$$

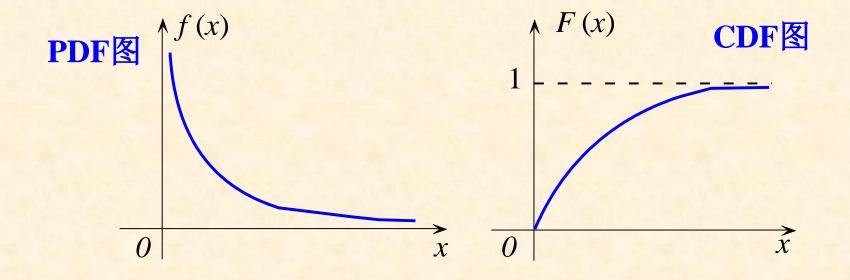
其中 $\theta > 0$ 为常数,则称X 服从参数为 θ 的指数分布,记为 $X \sim Exp(\theta)$ 。

注:有文献将指数分布写为 $X\sim Exp(\lambda)$,此时 $\lambda=1/\theta$ 。

易知
$$f(x) \ge 0$$
 和 $\int_{-\infty}^{+\infty} f(x) dx = 1$ 。

若RV X~Exp(
$$\lambda$$
), $f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x > 0 \\ 0, & else \end{cases}$

则X的CDF为
$$F(x) = \begin{cases} 1-e^{-x/\theta}, & x > 0 \\ 0, & else \end{cases}$$



指数分布的特性: 无记忆性 (又称遗失记忆性)

即
$$\forall s,t>0$$
,条件概率 $P\{X>s+t|X>s\}=P\{X>t\}$

$$P\{X > s+t | X > s\} = \frac{\$件概率}{P\{(X > s+t)\cap(X > s)\}}$$

$$= \frac{P\{X > s + t\}}{P\{X > s\}} = \frac{1 - F(s + t)}{1 - F(s)} = \frac{e^{-(s + t)/\theta}}{e^{-s/\theta}} = e^{-t/\theta} = P\{X > t\}$$

如:某元件的寿命服从指数分布,已知元件已使用了s小时,则它总共能使用至少 s+t 小时的条件概率,与元件从开始使用到至少能用 t 小时的概率相等,也就是说元件对它已使用 s小时无记忆。

指数分布的应用:

- ①电子元器件可靠性研究中,用于描述所发生的缺陷或系统故障数的测量结果;
- ②独立随机事件发生的间隔事件:如旅客进机场的时间间隔;
- ③大型复杂系统平均故障间隔时间(MTBF)的失效分布。

3、正态分布 Normal/Gauss Distribution

若RVX具有PDF

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

其中 μ , σ (σ >0) 为常数,则称X 服从参数为 μ , σ 的正态 / 高斯分布,记为 $X\sim N$ (μ , σ^2)。

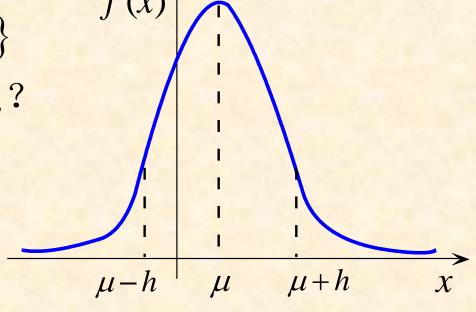
可知
$$f(x) \ge 0$$
 和 $\int_{-\infty}^{+\infty} f(x) dx = 1$ 。

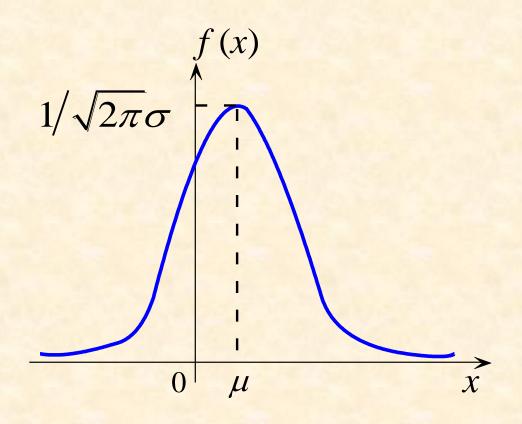
参数 μ 和 σ 的意义将在第4章中介绍。

① PDF曲线关于 $x = \mu$ 对称;

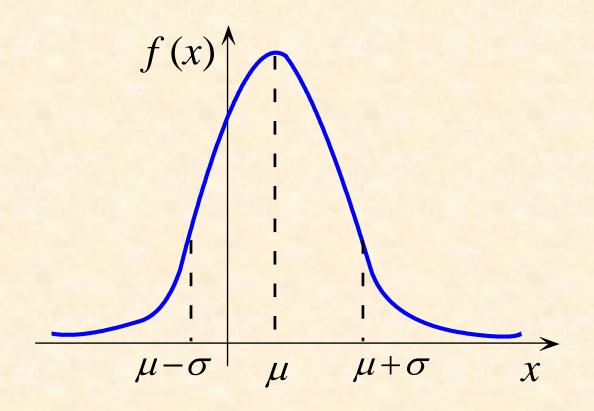
$$\forall h > 0$$
, $P\{\mu - h < X \le \mu\}$
与 $P\{\mu < X \le \mu + h\}$ 关系?

相等

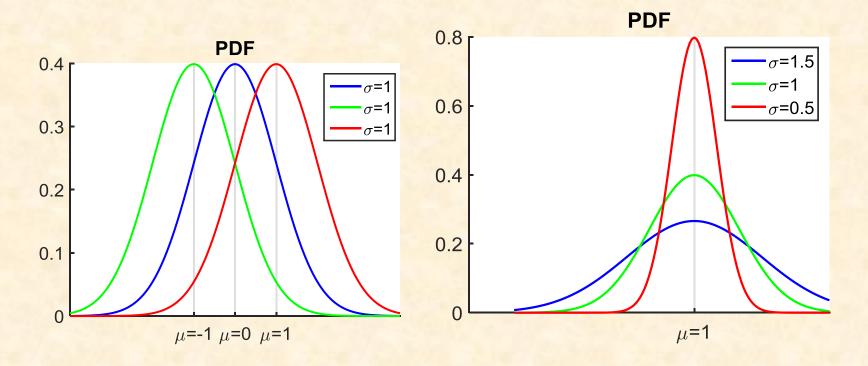




③ f(x) 曲线在 $x = \mu \pm \sigma$ 处有拐点,且以Ox轴为渐近线,即 $x \to \infty$ 时, $f(x) \to 0$;



④ μ为位置参数 (决定对称轴位置) σ为形状参数 (决定曲线分散性)



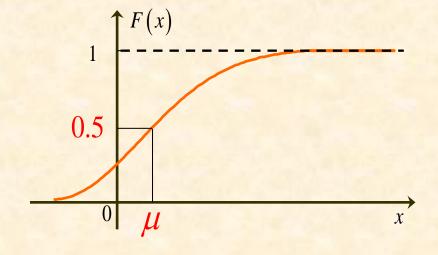
图由Matlab软件绘制,程序附后

正态分布的CDF

若 $X \sim N(\mu, \sigma^2)$,则X的CDF为

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





综上可知:

- ▶ 正态分布的PDF曲线呈现钟型曲线;
- ightharpoonup 正态分布可以由它的两个参数 μ 和 σ <u>唯一</u>确定, 当 μ 和 σ 不同时,是不同的正态分布。

下面我们介绍一种最重要的正态分布

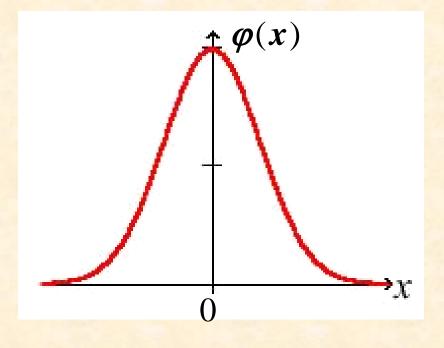
标准正态分布 Standard Normal Distribution

 $\mu=0$, $\sigma=1$ 的正态分布称为标准正态分布,记为N(0,1)。

标准正态分布的PDF用 $\varphi(x)$ 表示,即

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

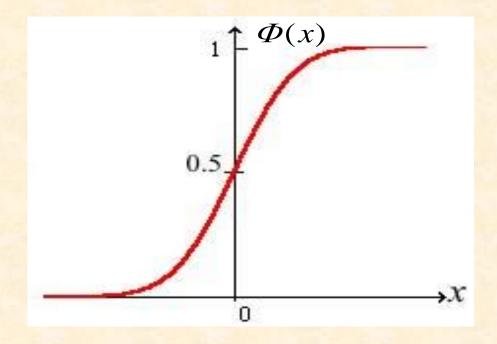
$$(-\infty < x < \infty)$$



标准正态分布

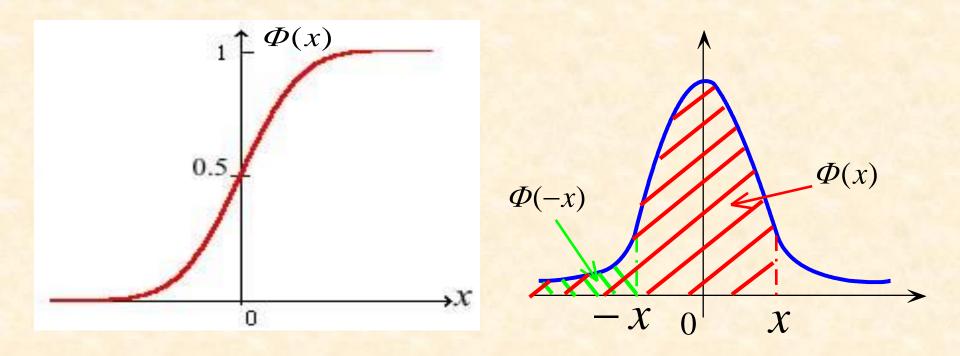
标准正态分布的CDF用 $\phi(x)$ 表示。

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



$\Phi(x)$ 性质

$$\forall x \in \mathbb{R}, \ \Phi(-x) = 1 - \Phi(x)$$



引理

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

证:
$$F_Z(x) \stackrel{\text{CDF定义}}{=\!=\!=\!=} P\{Z \le x\} = P\{\frac{X - \mu}{\sigma} \le x\}$$

$$= P\left\{X \le \mu + \sigma x\right\} = \int_{-\infty}^{\mu + \sigma x} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

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

引理说明

- ▶标准正态分布的重要性在于:任何一个一般的 正态分布都可以通过线性变换转化为标准正态分布, 称为标准化。
- ▶可以用 Φ(x) 来解决一般正态分布的概率计算问题。

具体如下:

若
$$X \sim N(\mu, \sigma^2)$$
,则
$$F_X(x) = P\{X \le x\} = P\left\{\frac{X - \mu}{\sigma}\right\} \le \frac{x - \mu}{\sigma}$$

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

$$\forall$$
区间 $(x_1, x_2]$,则

$$P\{x_1 < X \le x_2\} = \mathcal{D}\left(\frac{x_2 - \mu}{\sigma}\right) - \mathcal{D}\left(\frac{x_1 - \mu}{\sigma}\right)$$

标准正态分布表

P382附表2附有标准正态分布CDF $\phi(x)$ 数值表,利用它可以解决一般正态分布的概率计算查表。

P382附表2中给的是x>0时, $\Phi(x)$ 的值。

当x<0时, $\Phi(-x)=1-\Phi(x)$ 。

例 设一元件的尺寸为RV $X \sim N(15, 0.3^2)$,求: P(14.5 < X < 15.2)。

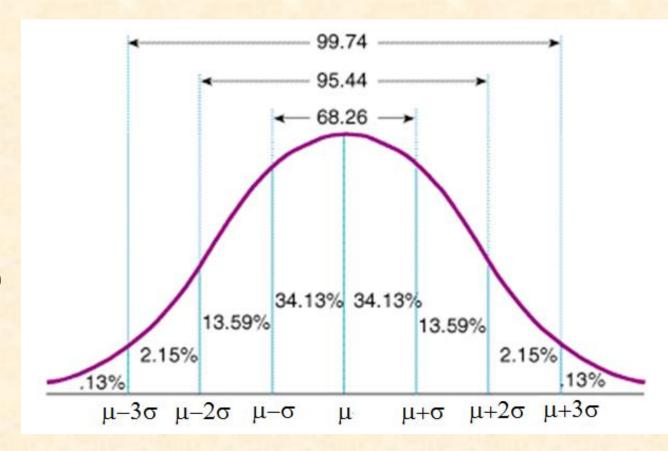
解:
$$P{14.5 < X < 15.2} = \Phi\left(\frac{15.2 - 15}{0.3}\right) - \Phi\left(\frac{14.5 - 15}{0.3}\right)$$

$$= \Phi(0.67) - \Phi(-1.67)$$

$$=0.7011$$

30准则

若 $X\sim N(\mu, \sigma^2)$ 则



$$P\{\mu - \sigma < X < \mu + \sigma\} = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6826$$

$$P\{\mu - 2\sigma < X < \mu + 2\sigma\} = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 0.9544$$

$$P\{\mu-3\sigma < X < \mu+3\sigma\} = \Phi(3)-\Phi(-3) = 2\Phi(3)-1 = 0.9974$$

30准则

如果 $X \sim N(\mu, \sigma^2)$,则X 取值落在 $(\mu \pm 3\sigma)$ 区间内的概率是0.9974,超出这个范围的可能性仅占不到0.3%,也就是说 X 取值落在 $(\mu \pm 3\sigma)$ 区间内几乎是肯定的,这在统计学上称作 3σ 准则。

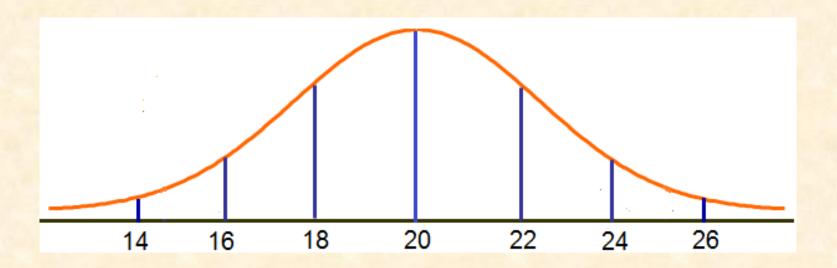
30准则可帮助我们快速计算一些区间概率问题,如

例 RV $X \sim N(20, 2^2)$, 估计以下区间概率:

$$P(16 < X < 24) = 0.9544$$

$$P(X>22) = 0.1578$$

$$P(X < 14) = 0.0013$$



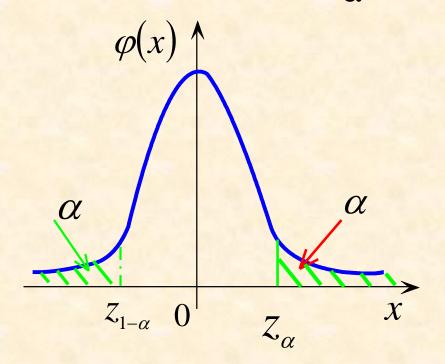
为了便于在数理统计中的应用,对于标准正态分布引入如下分位点定义:

设 $X \sim N(0,1)$, 若 Z_{α} 满足

$$P\{X>z_{\alpha}\}=\alpha, \quad 0<\alpha<1,$$

则称 z_{α} 为标准正态分布的上 α 分位点。(或记 u_{α})

易知: $z_{1-\alpha} = -z_{\alpha}$ α 越小, z_{α} 越大 $z_{0.5} = 0$



正态分布的应用:

正态分布是最常见最重要的一种分布,大量随机变量都服从或近似服从正态分布,如

- > 测量误差;
- >人的生理特征尺寸(身高、体重等);
- > 某地的降水量;
- > 生产条件不变的情况下,产品的长度、重量、强度;
- > 半导体元器件中的热噪声、电流和电压等等。

第5章的中心极限定理将给出正态分布的理论应用背景。

小结

▶介绍了RV的分布函数;

▶介绍了连续型RV及其PDF;

>3种常见连续型分布。

作业

Pages 57, 58: 第17, 19, 21, 24, 25, 27题

标准正态分布表

		9
$^{2}0.05$	_	•
0.05		

Z	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0			0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0	.64	5	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0			0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545

P382附表2

$$z_{0.025} = ?$$

$$z_{0.04} = ?$$

标准正态分布表

Z	0	1	2	3	4	5	6	7	8	9
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.3515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9772	0.9778	0.978			0.9798	0.9803		0.	0.9817
2.1	0.9821	0.9826	0.9830	17	51	0.9842	0.9846		96	0.9858
2.2	0.9861	0.9864	0.986	1.01		0.9878	0.9881			0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

标准正态分布常用分位点值

α	.001	.005	.01	.025	.05	.10
z_{α}	3.090	2.576	2.326	1.960	1.645	1.282

附录 PPT P27 Matlab程序

```
%%绘制第二章正态分布曲线
%%位置参数改变
x=[-4:.05:4]; % 以 0.05 间隔产生从-4 到 4 的散点
y1=normpdf(x,0,1); % 利用正态分布的 PDF 计算 x 点处对应的 PDF 值
figure % 建立一个图
N1=plot(x,y1,'b-','LineWidth',2);%在图中绘制一条曲线
hold on %在一个图中绘制多条曲线
y2=normpdf(x,-1,1);
N2=plot(x,y2,'g-','LineWidth',2);
hold on
y3=normpdf(x,1,1);
N3=plot(x,y3,'r-','LineWidth',2);
hold off
box off
set(gca, 'XTick',[-1,0,1],'xticklabel',{'\mu=-1','\mu=0','\mu=1'})%设置X轴坐标刻度对应处要显示的文字表达
set(gca, 'XGrid','on')% 打开 x 坐标轴网格线
set(gca,'LineWid',1.5)%设置坐标轴线及网格线宽度
```

附录 PPT P27 Matlab程序

```
%%绘制第二章正态分布曲线
%%形状参数改变
x=[-3:.05:4];%以 0.05间隔产生从-4 到 4 的散点
y1=normpdf(x,1,1.5); % 利用正态分布的 PDF 计算 x 点处对应的 PDF 值
figure % 建立一个图
N1=plot(x,y1,'b-','LineWidth',2);%在图中绘制一条曲线
hold on%在一个图中绘制多条曲线
y2=normpdf(x,1,1);
N2=plot(x,y2,'g-','LineWidth',2);
hold on
y3=normpdf(x,1,0.5);
N3=plot(x,y3,'r-','LineWidth',2);
hold off
box off
set(gca, 'XTick',[1], 'xticklabel', {'\mu=1'})%设置 X 轴坐标刻度对应处要显示的文字表达
set(gca, 'XGrid','on')% 打开 x 坐标轴网格线
set(gca,'LineWid',1.5)%设置坐标轴线及网格线宽度
```