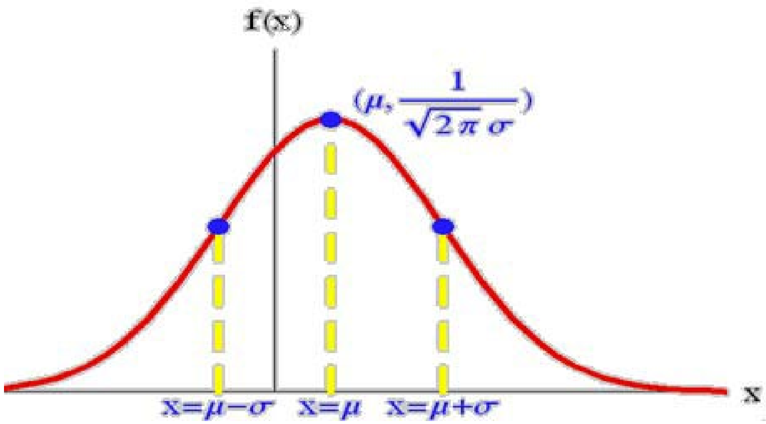
第十四讲 正态分布

◆ 正态分布的定义：

若X的概率密度函数为 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$，$-\infty < x < +\infty$

其中，$-\infty<\mu<+\infty, \sigma>0$,就称$X$服从参数为$\mu,\sigma$的正态分布(或高斯分布)，记为 $X\sim N(\mu,\sigma^2)$.

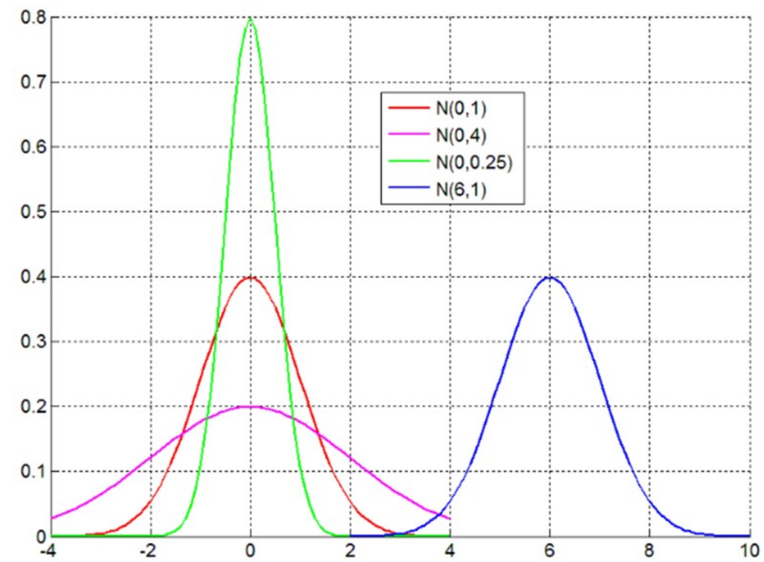


特征:

- $f(x)$=关于$x=\mu$对称;
- 当$x\le\mu$时,$f(x)$是严格单调递增函数;
- $f_{\max}=f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$;
- $\lim\limits_{|x-\mu|\to\infty} f(x)=0$

◆两个参数的含义:

(1)当固定σ,改变μ的大小时，$f(x)$图形的 形状不变,只是沿着x轴作平移变换； **μ称为位置参数(决定对称轴位置).**



(2)当固定μ,改变σ的大小时,$f(x)$图的对称轴不变，而形状在改变, σ越小，图形 越高越瘦, σ越大，图形越矮越胖.

σ称为尺度参数(决定曲线分散程度).

◆正态分布的用途:

(1)自然界和人类社会中很多现象可以看做正态分布

如: 人的生理尺寸(身高、体重);

医学检验指标(红细胞数、血小板);

测量误差;等等

(2)多个随机变量的和可以用正态分布来近似

如: 某位同学完成所有作业的时间;

二项分布; 等等

(By 中心极限定理)

◆ 正态分布的概率计算

若$X\sim N(\mu,\sigma^2)$,对实数x,

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

积分算不出来~~~~

方法一:用EXCEL、MATLAB、R等软件来计算;

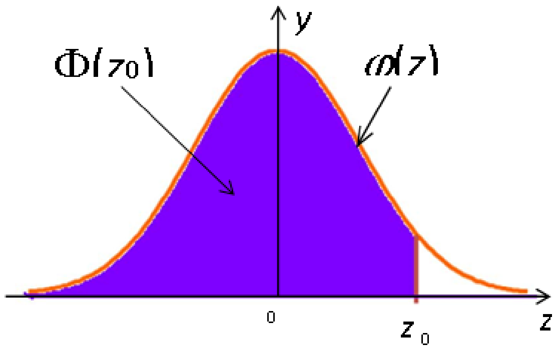
方法二:用数值积分法;

方法三:转化为标准正态,然后利用标准正态分布表来求.

◆标准正态分布

若$Z\sim N(0,1)$,称Z服从标准正态分布.

Z的概率密度函数: $\varphi(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$



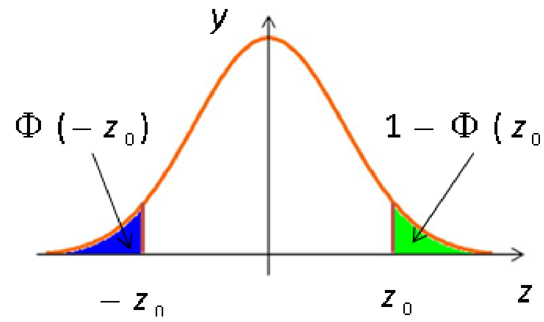
Z的分布函数: $\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.$

标准正态分布函数($\Phi(z)$)(http://en.wikipedia.org/wiki/Standard_normal_table)

z	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53980	0.54380	0.54776	0.55172	0.55567	0.55966	0.56360	0.56749	0.57142	0.57535
0.2	0.57930	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900

$$\Phi(1.23) = 0.89065$$

注意到 $\varphi(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$关于 轴的对称性,

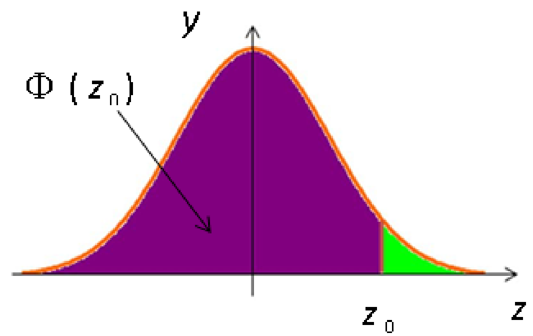


则标准正态分布的分布函数有一个

重要性质:

$$\Phi(-z_0)=1-\Phi(z_0),$$

对于任意的实数z_0都成立.



性质：当$X \sim N(\mu, \sigma^2)$时，$\frac{X-\mu}{\sigma} \sim N(0,1)$.

证明：对于任意实数z，

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

令 $s = \frac{t-\mu}{\sigma}$ $\longrightarrow$ $= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds = \Phi(Z)$

则 $\frac{(t-\mu)^2}{2\sigma^2} = \frac{s^2}{2}$

$ds = \frac{1}{\sigma}\mathrm{dt}$

由此可知，当$X \sim N(\mu, \sigma^2)$时，对于任意实数a，有

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

◆ 例1：

一批钢材(线材)长度(cm)$X\sim N(\mu,\sigma^2)$, $\mu=100, \sigma=2$,

求:(1) 这批钢材长度小于97.8的概率;

(2) 这批钢材长度落在区间(97.8,103)的概率.

(1) $$P(X<97.8)=P\left(\frac{X-\mu}{\sigma}<\frac{97.8-\mu}{\sigma}\right)=\Phi\left(\frac{97.8-\mu}{\sigma}\right)=\Phi\left(\frac{97.8-100}{2}\right)$$
$$=\Phi(-1.1)=1-\Phi(1.1)=1-0.86433=0.13576$$

查表

(2) $$P(97.8<X<103)=P\left(\frac{97.8-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{103-\mu}{\sigma}\right)=\Phi\left(\frac{103-100}{2}\right)-$$
$$\Phi\left(\frac{97.8-100}{2}\right)=\Phi(1.5)-\Phi(-1.1)=0.93319-0.13576=0.79743.$$

查表

◆ 例2：

用天平称一实际重量为μ的物体,天平的读数记为随机变量

$X \sim N(\mu,\sigma^2)$, 求读数与μ的偏差在3σ范围之内的概率;

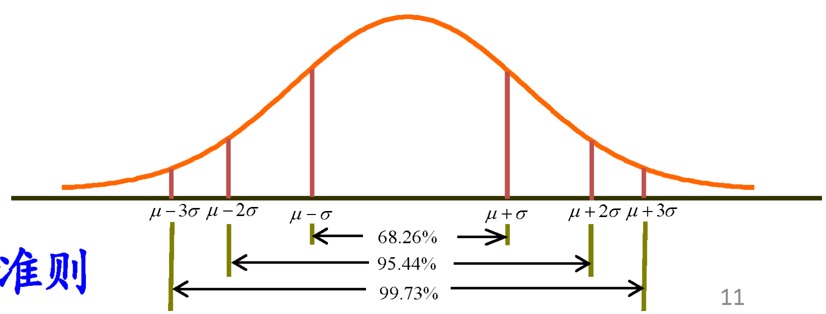
解： 由题意知,要求的是

$$P(|X-\mu|<3\sigma)=P(-3\sigma<X-\mu<3\sigma)=P\left(-\frac{3\sigma}{\sigma}<\frac{X-\mu}{\sigma}<\frac{3\sigma}{\sigma}\right)$$

$$=P\left(-3<\frac{X-\mu}{\sigma}<3\right)=\Phi(3)-\Phi(-3)=\Phi(3)-\{1-\Phi(3)\}\underset{\text{查表}}{=}2\Phi(3)-1$$

$$=2*0.99865-1=0.9973$$

3σ准则



THE END

第十五讲 随机变量函数的分布

- 若要得到一个圆的面积Y ,总是测量其半径, 半径的测量值可看作随机变量X，若 $X\sim N(\mu,\sigma^2)$, 则$Y=\pi X^2$的分布是什么?
- 若已知体重$W\,(kg)$均服从正态分布, 在身高L（m）确定的情形下,则体质指数$BMI= W/L^2$服从什么分布？

问题：已知随机变量X的分布,$\mathbf{Y = g(X)}$,

函数$\mathbf{g(\cdot)}$已知,

求Y的分布.

◆ **例1:**

设随机变量X的概率分布律为$Y=X^2$,

求Y的概率分布律.

X	-1	0	1
P	0.1	0.6	0.3

解: 因为X的可能取值为$-1, 0$和1, 而$Y=X^2$,

故可知Y的可能取值为0和1.

又因为$Y=X^2$, 从而

$$\{Y=0\}=\{X=0\},\{Y=1\}=\{(X=1)\cup(X=-1)\}$$

因此 $P(Y=0)=P(X=0)=0.6$,

$$P(Y=1)=P\{(X=1)\cup(X=-1)\}$$
$$=P\{(X=1)+(X=-1)\}=0.4$$

Y	0	1
P	0.6	0.4

◆ 例2：

设随机变量 的概率密度函数为

$$f_x(x)=\begin{cases}\dfrac{x}{8} & 0<x<4\\ 0 & \text{其他}\end{cases}$$

求$Y=X^2$的概率密度函数.

解： 由题意知$P(0<X<4)=1$, 从而$P(0<Y<16)=1$.

故$f_Y(y)=0$, 当$y\notin(0,16)$时.

当$y\in(0,16)$时, 先考察Y的分布函数:

$$F_Y(y)=P\{Y\le y\}=P\{X^2\le y\}=P\{-\sqrt{y}\le X\le\sqrt{y}\}$$

$\because P\{-\sqrt{y}\le X<0\}=0$

$$=P\{0\le X\le\sqrt{y}\}=\int_0^{\sqrt{y}}\frac{t}{8}dt=\frac{y}{16}$$

故$f_Y(y)=F_Y'(y)=\frac{1}{16}$。即$Y$服从均匀分布$U$（0，16）

或者：

当 $y\in(0,16)$时，先考察Y的分布函数：

$$F_Y(y)=P\{Y\le y\}=P\{X^2\le y\}=P\{-\sqrt{y}\le X\le\sqrt{y}\}$$

$$=P\{X\le\sqrt{y}\}$$

$$=F_X(\sqrt{y})$$

$P\{X<-\sqrt{y}\}=0$

$$f_X(x)=\begin{cases}\dfrac{x}{8} & 0<x<4\\ 0 & \text{其他}\end{cases}$$

故$f_Y(y)=F'_X(\sqrt{y}))=f_x(\sqrt{y})\cdot\dfrac{1}{2\sqrt{y}}=\dfrac{\sqrt{y}}{8}\cdot\dfrac{1}{2\sqrt{y}}=\dfrac{1}{16}$

即Y服从均匀分布$U(0,16)$.

一般，若已知X的概率分布，$Y=g(X)$，求Y的概率分布的过程为：先给出Y的可能取值；再利用等价事件来给出概率分布.

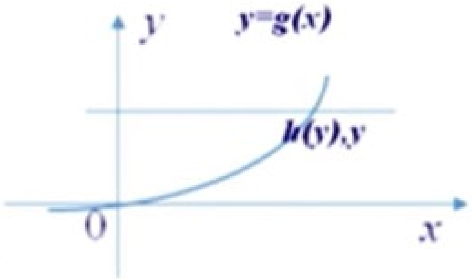
➢ 若X为离散型随机变量，则先写出Y的可能取值,y_1，$y_2\cdots y_j\cdots$；再找出$\{Y=y_j\}$的等价事件$\{X\in D\}$,得$P(Y=y_j)=P(X\in D)$;

➢若X为连续型随机变量,先根据X的取值范围,给出Y的取值范围;然后写出Y的概率分布函数：$F_Y(\mathrm{y})=\mathrm{P}(Y\le y)$，找出$\{Y\le y\}$的等价事件$\{X\in D\}$,得$F_Y(\mathrm{y})=\mathrm{P}(X\in D)$；再求出Y的概率密度函数$f_Y(\mathrm{y})$.

◆ 定理：

设随机变量$X \sim f_X(x), -\infty < x < +\infty, Y = g(X)$,

$g'(x) > 0$ (或$g'(x) < 0$),则Y具有概率密度为：



反函数

$$f_Y(x) = \begin{cases} f_X(h(y)) \cdot |h'(y)| & \alpha < y < \beta \\ 0 & \text{其他} \end{cases}$$

注意：

- **这里**(α, β)是Y的取值范围，其中：$\alpha = g(-\infty), \beta = g(+\infty)$
当$g'(x) < 0$时$\alpha = g(+\infty)$，$\beta = g(-\infty)$}.

- h**是**g**的反函数，即**$h(y) = x \Leftrightarrow y = g(x)$.

◆ **例3：**

设$X\sim N(\mu,\sigma^2)$, $Y=aX+b(a\neq 0)$,求Y的概率密度$f_Y(y)$.

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

解： $y=g(x)=ax+b,\ g'(x)=a\neq 0,$

$$x=h(y)=(y-b)/a,\ h'(y)=1/a,$$

$$f_Y(y)=f_X\left(\frac{y-b}{a}\right)\cdot\frac{1}{|a|}=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{((y-b)/(a-\mu)^2}{2\sigma^2}\right]\cdot\frac{1}{|a|}$$

$$=\frac{1}{\sqrt{2\pi}|a|\sigma}e^{-\frac{[y-(a\mu+b)]^2}{2a^2\sigma^2}}\quad\Rightarrow Y\sim N(a\mu+b,\ a^2\sigma^2)$$

一般地，若随机变量$X \sim N(\mu, \sigma^2)$，

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

$$X \sim N(1,3), Y = 3 - 2X$$

$$\Rightarrow Y \sim N(1,12).$$

$$\because a\mu + b = -2 \times 1 + 3 = 1$$
$$a^2\sigma^2 = (-2)^2 \times 3 = 12$$

THE END