正态分布

教学设计与制作：刘琼荪

引例

小麦的亩产量

新建楼盘的销售价格

精密仪器的测量误差

引例

火车的运行速度

炮弹的落点位置

定义：如果连续型随机变量X的密度函数为

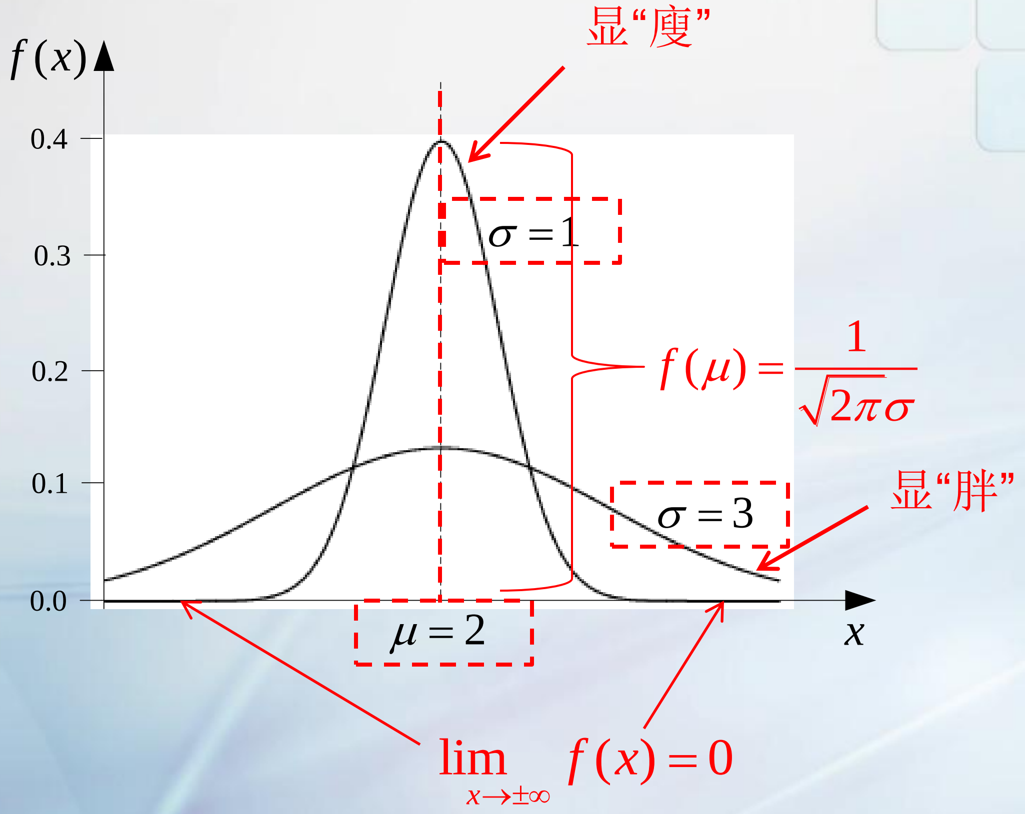
$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},x\in R,$$

德国数学家Gauss

其中μ，σ为参数，称这个随机变量X服从正态分布(高斯分布)，记为 $X\sim N(\mu,\sigma^2)$， $\sigma>0$

密度函数的典型特征：

密度函数曲线图

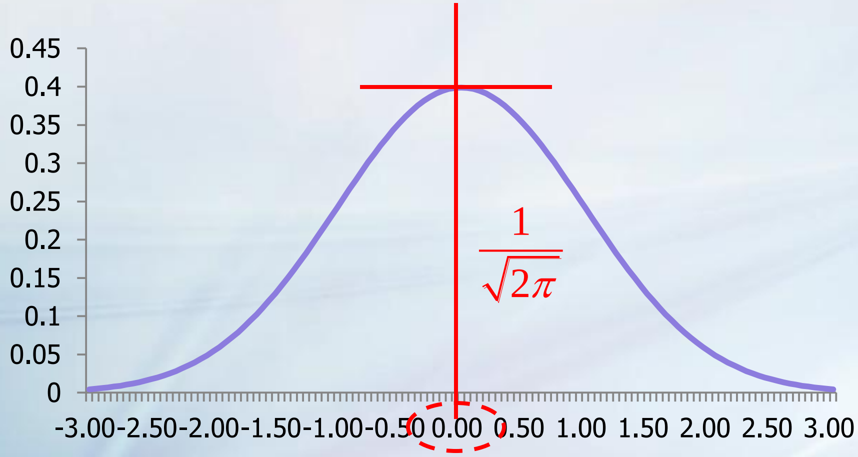


当 $\mu = 0$， $\sigma = 1$ 时，记密度函数为

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

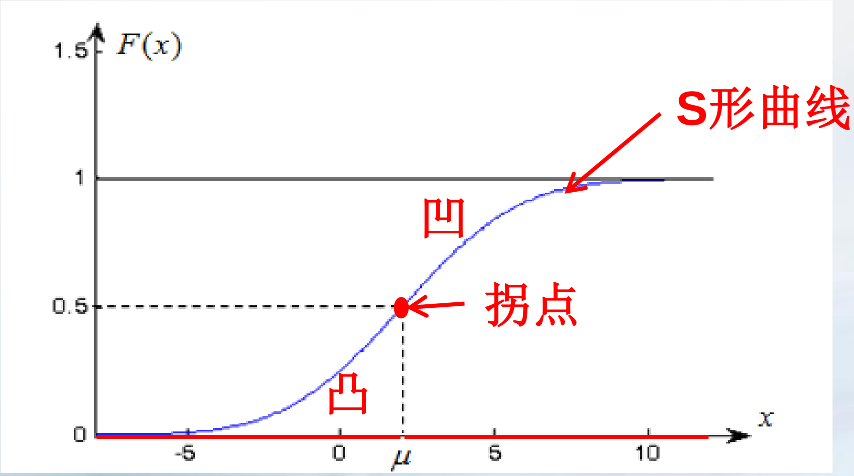
则 $X \sim N(0,1)$. 它是标准正态分布，分布函数记为$\Phi(x)$.

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt$$



密度函数与分布函数具有关系：$F(x)=\int_{-\infty}^{x} f(t)dt$

$F(x)$的几何特征：



计算概率的步骤：

（1）标准化 $X \sim N(\mu,\sigma^2) \xrightarrow{Y=\frac{X-\mu}{\sigma}} Y \sim N(0,1)$

（2）查表计算

$$P\{a < X \le b\} = P\{\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\}$$

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

例

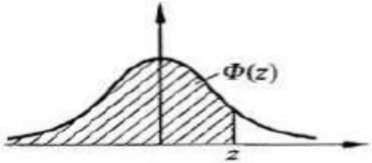
如果$X \sim N(\mu,\sigma^2)$， 求$P\{|X-\mu| \le 2\sigma\}$

第一步，$P\{|X-\mu| \le 2\sigma\}=P\{\frac{|X-\mu|}{\sigma} \le 2\}=2\Phi(2)-1$

第二步，查表 $\Phi(2)=0.9772$， 计算得到概率值0. 9544

$\Phi(0)=0.5$

$$\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du=P\{Z\leqslant z\}$$



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	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	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
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	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	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

$\Phi(2.15)=0.9842$

$\Phi(3.9)\doteq 1$

例

始点 火车站

市区路程短，但交通拥挤，所需时间$X1 \sim N(50,100)$

郊区路程长，但意外阻塞少，所需时间$X2 \sim N(60,16)$

问若有剩余70分钟，应走哪条路线？如果还有剩余65分钟，情况又如何呢？

分析

第一，如果有剩余70分钟

通过**市区**能及时乘火车的概率为

$$P\{0 \le X_1 \le 70\} = \Phi\left(\frac{70-50}{10}\right) - \Phi\left(\frac{0-50}{10}\right) = \Phi(2) - \Phi(-5) \approx \Phi(2) \approx 0.9773$$

通过**郊区**能及时乘火车的概率为

$$P\{0 \le X_2 \le 70\} = \Phi\left(\frac{70-60}{4}\right) - \Phi\left(\frac{0-60}{4}\right) = \Phi(2.5) - \Phi(-12.5) \approx \Phi(2.5) \approx 0.9938$$

分析

第二，如果有剩余65分钟

同理计算

$$P\{0 \le X_1 \le 65\} \approx 0.9332,$$

$$P\{0 \le X_2 \le 65\} \approx 0.8944$$

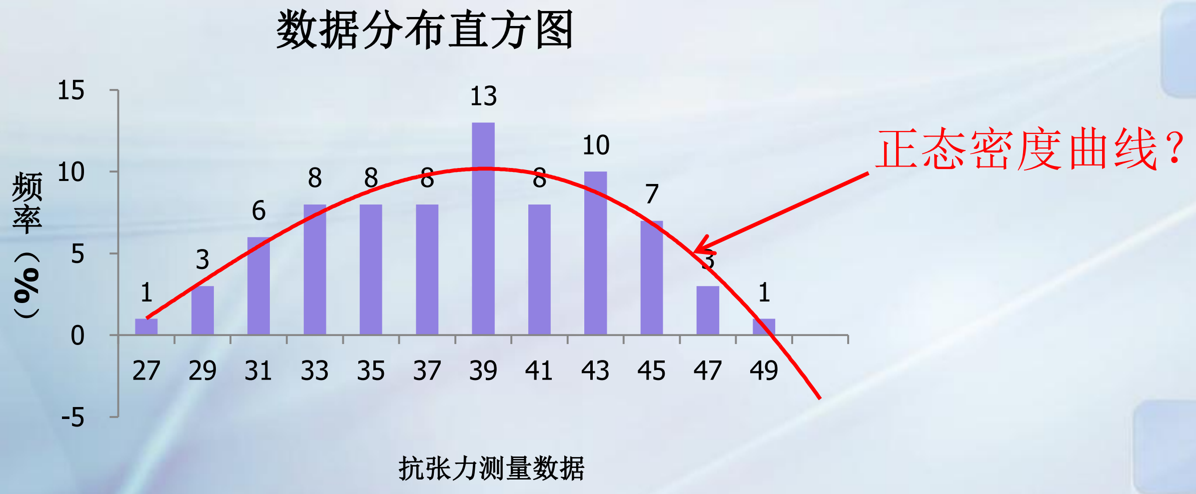
结论：若有70分钟可用，应选择通过**郊区**的路线；如果有65分钟可用应选择通过**市区**的路线。

正态分布的应用非常广泛。

为研究Φ85mm钢材的抗张力，试验测得76个数据 (kg/cm^2)

27	27.5	28.5	29	29.1	29.5	29.6	30	30.5	31	31.2	31.2
31.5	31.8	32	32.5	32.6	33	33.5	33.5	33.8	34	34.5	34.6
34.8	34.8	35.2	35.2	35.5	35.8	36.2	36.5	36.6	37	37.1	37.2
37.3	37.4	37.5	37.5	37.8	38.1	38.3	38.5	38.8	38.8	39	39.5
39.8	40	40.2	40.5	40.6	40.8	41	41.2	41.3	41.3	41.5	41.5
41.8	42	42.5	42.8	42.8	43.6	43.8	44.2	44.2	44.5	44.7	45
45.1	45.5	45.8	48								

数据分布 ⇨



小结

（1）正态分布的密度特征

（2）标准化变换

（3）计算概率