定义 2.12　正态分布

随机变量 X 的密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},x\in\mathbb{R} \tag{2.23}$$

其中 $\mu\in\mathbb{R},\sigma>0$ 为参数, 则称 X 服从参数为 μ,σ^2 的正态分布 (normal distribution), 记为 $X\sim N(\mu,\sigma^2)$. ♣

具有参数 $\mu=0$, $\sigma=1$ 的正态分布称为标准正态分布. 用 $\Phi(x)$ 和 $\varphi(x)$ 表示标准正态分布 $N(0,1)$ 的分布函数和密度函数.

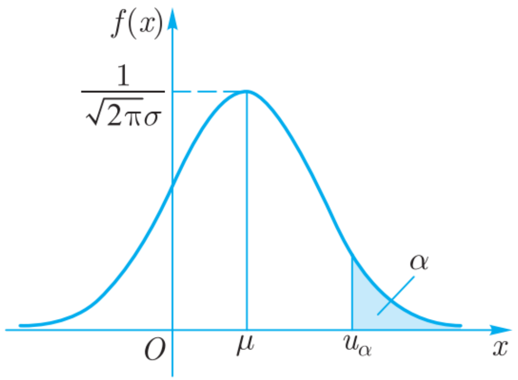
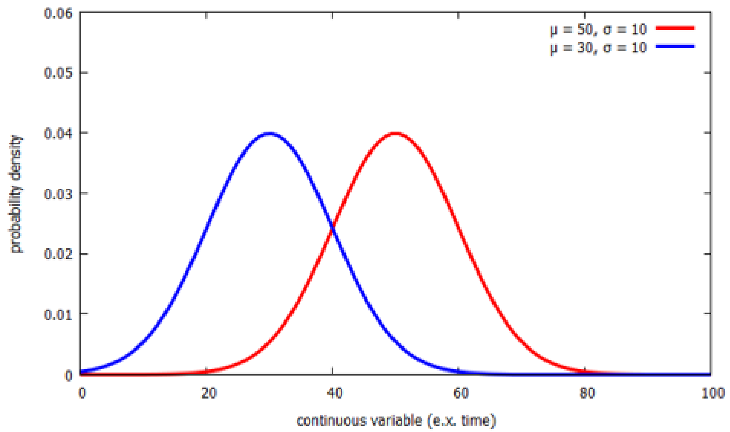
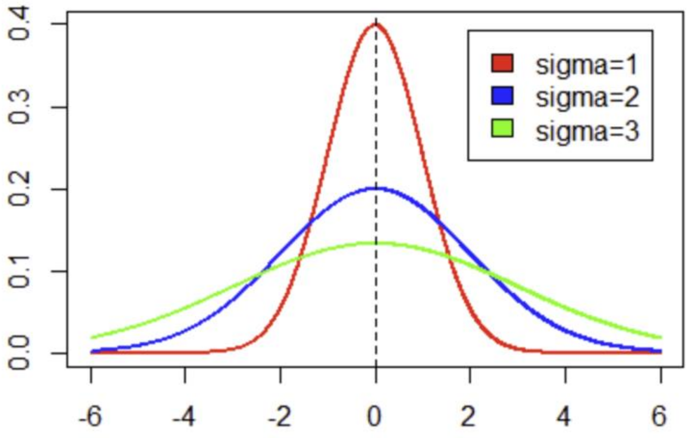


图 2.12　正态分布的密度函数





正态分布的密度函数图形性质

- 位置参数μ (location parameter)：正态分布的密度函数呈“钟形”曲线，两头小，中间大，关于$x = \mu$对称；密度函数在 $x = \mu$ 处达到最大值$\frac{1}{\sqrt{2\pi}\sigma}$，在$(-\infty, \mu)$ 和 $(\mu, +\infty)$ 内严格单调.

$$\varphi(-x) = \varphi(x), \qquad \Phi(-x) = 1 - \Phi(x)$$

- 尺度参数σ (scale parameter)：也称形状参数，σ的大小决定了密度函数的陡峭程度.

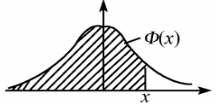
- 若固定σ：μ变化时，图形左、右平移，但是不改变形状.

- 若固定μ：σ越大，图形越平缓，峰值越低；σ越小，图形越陡峭，峰值越高．故对相同的常数d，X 落在 $(\mu - d, \mu + d)$ 中的概率随σ减小而增加.

- 正态分布是最常用的一种数据模型，确实在很多场合下，许多数据可以认为近似服从正态分布. 例如一个人的智商、身高、体重，加工零件的尺寸等.
- 教育统计学统计规律表明，学生的智力水平，包括学习能力，实际动手能力等呈正态分布。因而正常的考试成绩分布应基本服从正态分布.
- 某些医学现象，如同质群体的身高、红细胞数、血红蛋白量，以及实验中的随机误差，呈现为正态或近似正态分布.

标准正态分布表

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$



x	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.	0.9987	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

$$P(|X| \leqslant 1) = 0.6826,$$
$$P(|X| \leqslant 2) = 0.9544,$$
$$P(|X| \leqslant 3) = 0.9974.$$

一个标准正态随机变量绝对值超过 2 的概率小于 0.05, 而它的绝对值超过 3 的概率小于 0.01,

注：表中末行为函数值 $\Phi(3.0), \Phi(3.1), \cdots, \Phi(3.9)$.

一般正态分布的标准化变换

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

变换 $\dfrac{X-\mu}{\sigma}$ 也称为**标准化变换**

重要！

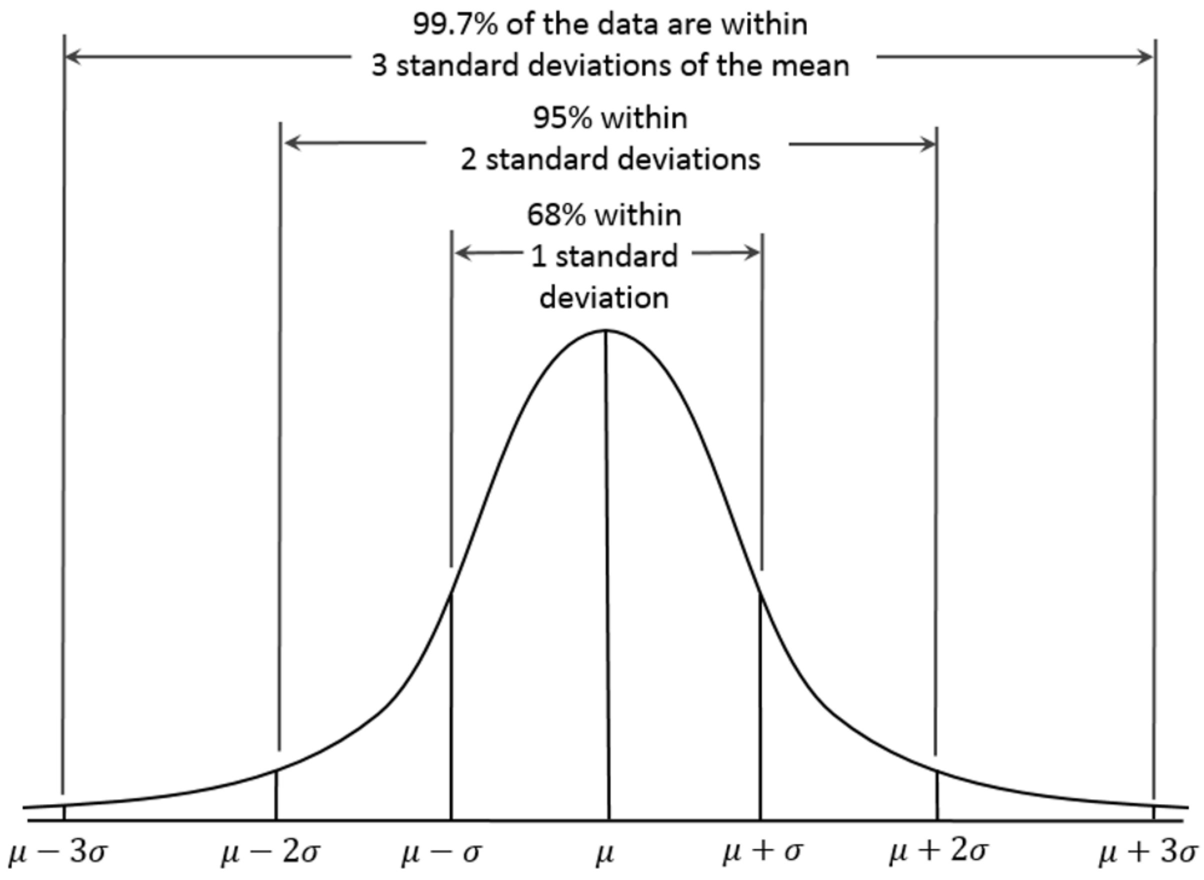
- 以 $F(x)$ 记正态分布 $N(\mu, \sigma^2)$ 的概率分布函数，则恒有

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

- 任一正态分布的概率分布函数都可通过标准正态分布的分布函数计算出来.

设$X \sim N\left(\mu, \sigma^2\right)$,求数$k$使得$P(\mu - k\sigma < X < \mu + k\sigma) = 0.95$.

3σ原则



X落在（$\mu-3\sigma$，$\mu+3\sigma$）以外的概率小于千分之三，在实际问题中常认为相应的事件是不会发生的，基本上可以把区间（$\mu-3\sigma,\mu+3\sigma$）看作是随机变量 X 实际可能的取值区间，这称之为正态分布的“3σ”原则。

数值分布在（$\mu-\sigma$，$\mu+\sigma$）中的概率为0.6826;
数值分布在（$\mu-2\sigma$，$\mu+2\sigma$）中的概率为0.9544;
数值分布在（$\mu-3\sigma$，$\mu+3\sigma$）中的概率为0.9974.

例子

公共汽车车门的高度是按男子与车门碰头的概率小于0.01来设计的. 设男子身高 $X \sim N(172, 36)$（单位：cm），问应该如何选择车门高度?

- 随机变量的分布是一个大概念；

- 当 X 为离散型随机变量时，我们一般用分布律来研究；

- 当 X 为连续型随机变量时，我们一般用密度函数来研究；

- 当 X 为一般随机变量时，我们可以用分布函数来研究.

- 随机变量的分布函数在任何场合都可以用，有时候求密度函数时，可以先考虑求出它的分布函数，然后求导得到密度函数.

4. 随机变量函数的分布

- 感兴趣的量是一些随机变量的函数$Y = g(X)$，如
  - 物体运动的速度为随机变量V， 该物体的动能 $E = \frac{1}{2}V^2$
  - 圆直径为D, 面积 $S = \frac{1}{4}\pi D^2$
- 感兴趣的量不能直接测量得到.
- 通过已知函数关系，以及容易测量变量的分布函数，来得出感兴趣量的分布函数.

• 离散型随机变量函数的分布

设随机变量 $Y = g(X)$, 其中 X 的分布律为

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

其中 $\sum p_k = 1$. 则随机变量 Y 的分布为

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

其中 $g(x_i)$ 相同时把对应的 p_i 相加, 即

$$P(Y = g(x_k)) = \sum_{i:g(x_i)=g(x_k)} p_i.$$

特别地, 如果 g 恒为常数, 那么 Y 取该常数的概率为 1, 此时称 Y 的分布退化.

设 X 的概率函数为

X	-1	0	1	2
P	1/4	1/2	1/8	1/8

试求 $Y = X^2, \quad Z = X^3 + 1$ 的分布律。

• 连续型随机变量函数的分布

命题 2.1　随机变量函数的分布函数

设 $X \sim f(x)$, $Y = g(X)$, 则随机变量 Y 的分布函数 $F_1(y)$ 为

$$F_1(y) = P(Y \leqslant y) = P(g(X) \leqslant y) = \int_{g(x)\leqslant y} f(x)\mathrm{d}x. \tag{2.25}$$

✓ 命题2.1---普适

✓ 推论2.1---特殊：当变换 g 满足某些条件时，Y 仍为连续型随机变量，且其密度函数与 X 的密度函数之间有着特殊的关系.

（注：连续型随机变量的函数未必是连续型随机变量.）

推论 2.1　密度函数变换公式

若 $g(x)$ 是严格单调的且反函数可导时，则随机变量 Y 仍为连续型随机变量，且有概率密度函数 $f_1(y)$

$$f_1(y) = \begin{cases} f(h(y))|h'(y)|, & \alpha < y < \beta, \\ 0, & \text{其他}, \end{cases} \tag{2.26}$$

其中 $h(y)$ 为 $g(x)$ 的反函数，$\alpha = \min\{g(-\infty), g(\infty)\}, \beta = \max\{g(-\infty), g(\infty)\}$.

设随机变量 $X \sim f(x)$, 求 $Y = aX + b$ 的分布, 其中 $a \neq 0$.

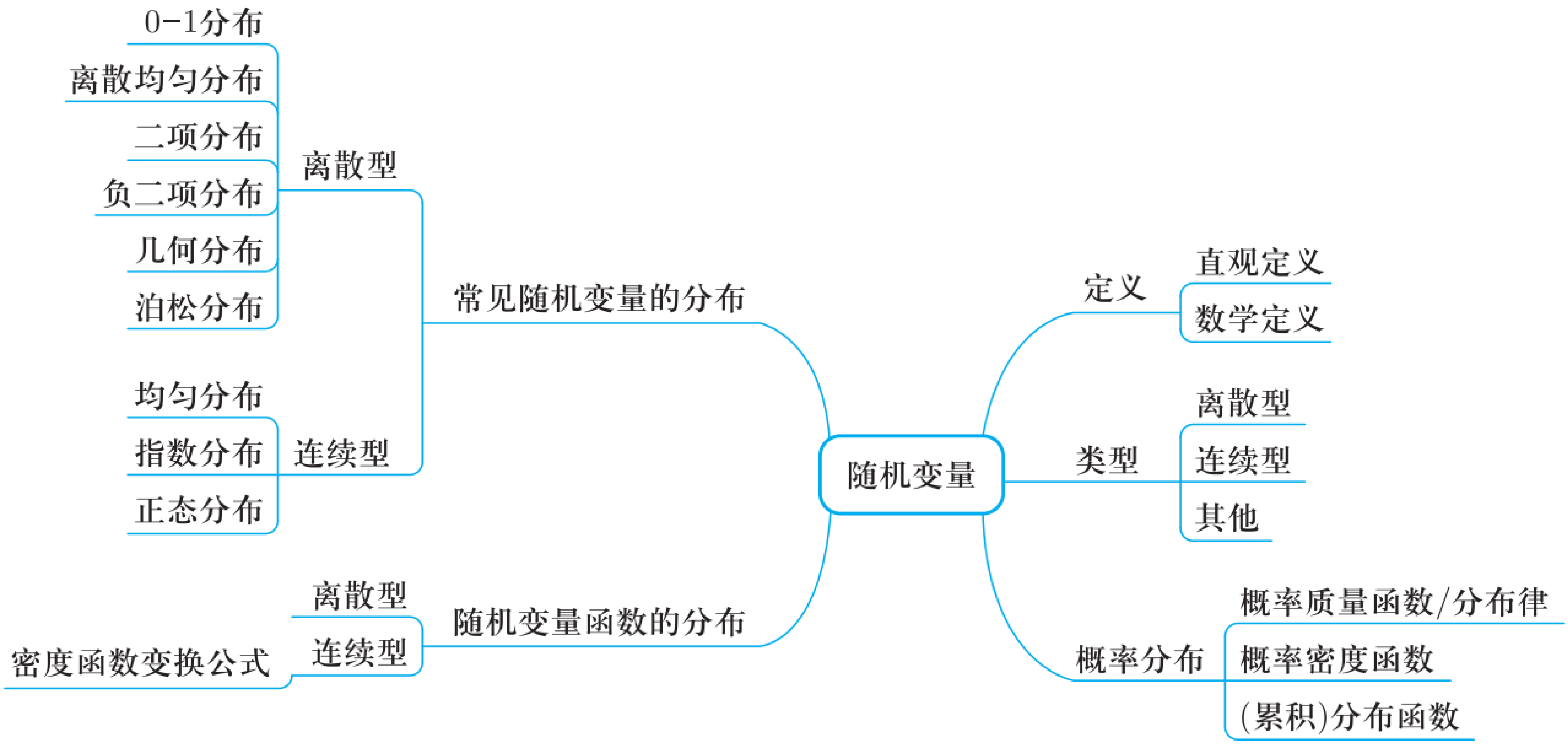
注 当 g 不是在全区间上单调而是逐段单调时，密度变换公式为下面的形式：设随机变量 X 的密度函数为 $f(x)$，$a < x < b$. 如果可以把 (a,b) 分割为一些 (有限个或可列个) 互不重叠的子区间的和 $(a,b) = \bigcup_j I_j$，使得函数 $y = g(x)$，$x \in (a,b)$ 在每个子区间上有唯一的反函数 $h_j(y)$，并且导函数 $h_j^{'}(y)$ 存在连续，那么 $Y = g(X)$ 是连续型随机变量，其密度函数为

$$f_1(y) = \sum_j f(h_j(y))|h_j^{'}(y)|I_j\ .$$

例子

设随机变量 $X \sim f(x), x \in \mathbb{R}$, 求 $Y = X^2$ 的密度函数.

本章总结


0-1分布
离散均匀分布
二项分布
负二项分布
几何分布
泊松分布
离散型
均匀分布
指数分布
正态分布
连续型
常见随机变量的分布
离散型
连续型
密度函数变换公式
随机变量函数的分布
随机变量
定义
直观定义
数学定义
类型
离散型
连续型
其他
概率分布
概率质量函数/分布律
概率密度函数
(累积)分布函数

Thank you!