Lec-6. 分布函数、连续型随机变量

主讲教师：吴利苏 (**wulisu@sdust.edu.cn**)
主　　页：wulisu.cn

目录

1. 分布函数
2. 连续型随机变量
3. 典型连续型随机变量
   - 均匀分布
   - 指数分布
   - 正态分布

随机变量的分布函数

- 对于离散型随机变量，可以用分布律来刻画，比如：表格和概率公式.
- 如何刻画非连续型随机变量?

随机变量的分布函数

- 对于离散型随机变量，可以用分布律来刻画，比如：表格和概率公式.
- 如何刻画非连续型随机变量？$\Rightarrow$ 分布函数.

分布函数

定义

设 X 是一个随机变量，函数

$$F(x) = P(X \le x), \qquad x \in \mathbb{R},$$

为 X 的概率分布函数，简称**分布函数**.

注

(1) $F(X)$ 是增函数，定义域是 $\mathbb{R}$;

注

(1) $F(X)$ 是增函数，定义域是 $\mathbb{R}$;

(2) 任何随机变量都有相应的分布函数;

注

(1) $F(X)$ 是增函数，定义域是 $\mathbb{R}$;

(2) 任何随机变量都有相应的分布函数;

(3) 不同的随机变量, 它们的分布函数可以相同.

注

(1) $F(X)$ 是增函数，定义域是 $\mathbb{R}$;

(2) 任何随机变量都有相应的分布函数;

(3) 不同的随机变量, 它们的分布函数可以相同.

(4) 分布函数的几何意义: $F(x)$ 的值表示 X 落在 $(-\infty, x]$ 上的概率;

注

(1) $F(X)$ 是增函数，定义域是 $\mathbb{R}$;

(2) 任何随机变量都有相应的分布函数;

(3) 不同的随机变量，它们的分布函数可以相同.

(4) 分布函数的几何意义: $F(x)$ 的值表示 X 落在 $(-\infty, x]$ 上的概率;

(5) 分布函数的用途: 可以给出随机变量 X 落入任意一个范围的可能性.

X 的分布函数为 $F(x)$, $a, b \in \mathbb{R}$, $a < b$,
则

- $$\begin{aligned} P\{a < X \leq b\} &= P\{X \leq b\} - P\{X \leq a\} \\ &= F(b) - F(a). \end{aligned}$$

X 的分布函数为 $F(x)$, $a, b \in \mathbb{R}$, $a < b$,
则

- $$\begin{aligned} P\{a < X \le b\} &= P\{X \le b\} - P\{X \le a\} \\ &= F(b) - F(a). \end{aligned}$$
- $P\{X > a\} = 1 - P\{X \le a\} = 1 - F(a)$.

X 的分布函数为 $F(x)$, $a, b \in \mathbb{R}$, $a < b$,
则

- $$\begin{aligned} P\{a < X \leq b\} &= P\{X \leq b\} - P\{X \leq a\} \\ &= F(b) - F(a). \end{aligned}$$
- $P\{X > a\} = 1 - P\{X \leq a\} = 1 - F(a)$.
- $$\begin{aligned} P\{a < X < b\} &= P\{a < X \leq b\} - P\{X = b\} \\ &= F(b) - F(a) - P\{X = b\}. \end{aligned}$$

X 的分布函数为 $F(x)$, $a, b \in \mathbb{R}$, $a < b$,
则

- $$\begin{aligned} P\{a < X \leq b\} &= P\{X \leq b\} - P\{X \leq a\} \\ &= F(b) - F(a). \end{aligned}$$
- $P\{X > a\} = 1 - P\{X \leq a\} = 1 - F(a).$
- $$\begin{aligned} P\{a < X < b\} &= P\{a < X \leq b\} - P\{X = b\} \\ &= F(b) - F(a) - P\{X = b\}. \end{aligned}$$
- $$\begin{aligned} P\{a \leq X \leq b\} &= P\{a < X \leq b\} + P\{X = a\} \\ &= F(b) - F(a) + P\{X = a\}. \end{aligned}$$

例

设 X 的分布律为

X	-1	2	3
P_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

求

(1) X 的分布函数,

(2) $P\{X \leq \frac{1}{2}\}$, $P\{\frac{3}{2} < X \leq \frac{5}{2}\}$, $P\{2 \leq X \leq 3\}$.

解:(1) 当 $x < -1$ 时,

$$F(x) = P\{X \leq x < -1\} = 0$$

当 $-1 \leq x < 2$ 时,

$$F(x) = P\{X \leq x\} = P\{X = -1\} = \frac{1}{4}$$

当 $2 \leq x < 3$ 时,

$$F(x) = P\{X \leq x\} = P\{X = -1\} + P\{X = 2\} = \frac{3}{4}$$

当 $x \geq 3$ 时,

$F(x) = P\{X \leq x\} = P\{X = -1\}+P\{X = 2\}+P\{X = 3\} = 1$

故 $F(x) = \begin{cases} 0 & x < -1; \\ \frac{1}{4} & -1 \leq x < 2; \\ \frac{3}{4} & 2 \leq x < 3; \\ 1 & x \geq 3. \end{cases}$

解: (2) $P\{X \leq \frac{1}{2}\} = F(\frac{1}{2}) = \frac{1}{4}$,
$P\{\frac{3}{2} < X \leq \frac{5}{2}\} = F(\frac{5}{2}) - F(\frac{3}{2}) = \frac{1}{2}$,
$P\{2 \leq X \leq 3\} = F(3) - F(2) + P\{X = 2\} = \frac{3}{4}$. □

离散型随机变量的分布函数

一般, 设离散型 X 的分布律为 $P\{X = x_k\} = p_k$,
则 X 的分布函数

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

分布函数在 $x = x_k$ 处有跳跃, 跳跃值为
$p_k = P\{X = x_k\}$.

例

设随机变量 X 的分布函数

$$F(x)=\begin{cases}0 & x<-1;\\ 0.2 & -1\leq x<3;\\ 0.6 & 3\leq x<4;\\ 1 & x\geq 4.\end{cases}$$

求 X 的分布律.

解: $F(x)$ 只在 $-1, 3, 4$ 跳跃, 跳的幅度分别是 $0.2, 0.4, 0.4$. 所以, 分布律为

X	-1	3	4
P	0.2	0.4	0.4

离散型随机变量的分布函数的性质

性质

设 X 是离散型随机变量，$F(X)$ 是 X 的分布函数. 则

(1) $0 \leq F(x) \leq 1$;

离散型随机变量的分布函数的性质

性质

设 X 是离散型随机变量，$F(X)$ 是 X 的分布函数. 则

(1) $0 \leq F(x) \leq 1$;

(2) $F(x)$ 单调不减，即对任意 $x_1 < x_2$, 有
$F(x_2) - F(x_1) = P\{x_1 < X \leq x_2\} \geq 0$;

离散型随机变量的分布函数的性质

性质

设 X 是离散型随机变量，$F(X)$ 是 X 的分布函数. 则

(1) $0 \le F(x) \le 1$;

(2) $F(x)$ 单调不减, 即对任意 $x_1 < x_2$, 有

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

(3) $F(-\infty) = 0$, $F(+\infty) = 1$;

离散型随机变量的分布函数的性质

性质

设 X 是离散型随机变量，$F(X)$ 是 X 的分布函数. 则

(1) $0 \le F(x) \le 1$;

(2) $F(x)$ 单调不减，即对任意 $x_1 < x_2$，有 $F(x_2) - F(x_1) = P\{x_1 < X \le x_2\} \ge 0$;

(3) $F(-\infty) = 0$, $F(+\infty) = 1$;

(4) $F(x)$ 是右连续函数，即 $F(x - 0) = F(x)$. 但不一定是连续函数.

例

一个靶子是半径为 $2m$ 的圆盘, 设击中靶上任一同心圆盘上的点的概率与该圆盘的面积成正比, 并设射击都能中靶, 以 X 表示弹着点与圆心的距离, 求 X 的分布函数.

解: 当 $x < 0$ 时, $\{X \leq x\}$ 是不可能事件, 于是

$$F(x) = P\{X \leq x\} = 0,$$

当 $0 \leq x \leq 2$ 时,

$$P\{0 \leq X \leq x\} = kx^2.$$

取 $x = 2$, 有 $P\{0 \leq X \leq 2\} = k \cdot 2^2$, 而
$P\{0 \leq X \leq 2\} = 1$, 则 $k = \frac{1}{4}$, 即

$$P\{0 \leq X \leq x\} = \frac{x^2}{4}.$$

从而

$$F(x) = P\{X \leq x\} = P\{X < 0\} + P\{0 \leq X \leq x\} = \frac{x^2}{4}.$$

当 $x \geq 2$ 时,

$$F(x) = P\{X \leq x\} = 1.$$

故 $F(x)=\begin{cases}0 & x<0;\\ \frac{x^2}{4} & 0\le x<2;\\ 1 & x\ge 4.\end{cases}$

图像为一连续曲线. $\square$

本题中，分布函数 $F(x)$ 可写为

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

其中 $f(t)=\begin{cases}\frac{t}{2} & 0<t<2;\\ 0 & \text{其他}.\end{cases}$

即 $F(x)$ 是非负函数 $f(t)$ 在区间 $(-\infty,x]$ 上的积分. 此时, 称 X 为连续型随机变量. （使得 $\underline{\ F(X)\ 连续\ }$ 的随机变量）

连续型随机变量及其概率密度

定义

若对于随机变量 X 的分布函数 $F(x)$, 存在非负可积函数 $f(x)$, 使得对于任意实数 x 有

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

则称 X 为连续型随机变量. $f(x)$ 称为 X 的概率密度函数, 简称概率密度.

连续型随机变量及其概率密度

定义

若对于随机变量 X 的分布函数 $F(x)$, 存在非负可积函数 $f(x)$, 使得对于任意实数 x 有

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

则称 X 为**连续型随机变量**. $f(x)$ 称为 X 的概率密度函数, 简称**概率密度**.

改变概率密度 $f(x)$ 的个别点的取值不影响 $F(x)$ 的取值. (见注3)

性质

设 $f(x)$ 是连续型随机变量 X 的概率密度，则

(1) $f(x) \geq 0$;

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

(3) $\forall x_1, x_2 \in \mathbb{R},\ x_1 \leq x_2,$

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)\,dt.$$

- $P\{X \leq a\} = F(a) = \int_{-\infty}^{a} f(x)\,dx$;

- $P\{X \leq a\} = F(a) = \int_{-\infty}^{a} f(x)\,dx;$
- $P\{X > a\} = 1 - P\{X \leq a\} = \int_{a}^{+\infty} f(x)\,dx;$

- $P\{X \leq a\} = F(a) = \int_{-\infty}^{a} f(x)\,dx$;
- $P\{X > a\} = 1 - P\{X \leq a\} = \int_{a}^{+\infty} f(x)\,dx$;
- $\forall a \in \mathbb{R}$, $P\{X = a\} = 0$, 且

$$P\{x_1 < X \leq x_2\} = P\{x_1 < X < x_2\};$$

- $P\{X \le a\} = F(a) = \int_{-\infty}^{a} f(x)\,dx;$
- $P\{X > a\} = 1 - P\{X \le a\} = \int_{a}^{+\infty} f(x)\,dx;$
- $\forall a \in \mathbb{R}$, $P\{X = a\} = 0$, 且

$$P\{x_1 < X \le x_2\} = P\{x_1 < X < x_2\};$$

- $\{X = a\}$ 是不可能事件 $\Rightarrow P\{X = a\} = 0$;
  而 $P\{X = a\} = 0 \nRightarrow \{X = a\}$ 是不可能事件;

- $P\{X \le a\} = F(a) = \int_{-\infty}^{a} f(x)\,dx$;
- $P\{X > a\} = 1 - P\{X \le a\} = \int_{a}^{+\infty} f(x)\,dx$;
- $\forall a \in \mathbb{R}$, $P\{X = a\} = 0$, 且

$$P\{x_1 < X \le x_2\} = P\{x_1 < X < x_2\};$$

- $\{X = a\}$ 是不可能事件 $\Rightarrow P\{X = a\} = 0$; 而 $P\{X = a\} = 0 \not\Rightarrow \{X = a\}$ 是不可能事件;
- 对于连续型随机变量 X,

$$P\{X \in D\} = \int_D f(x)\,dx, \qquad \forall D \subset \mathbb{R}.$$

性质

设 $f(x)$ 是连续型随机变量 X 的概率密度，则

(4) 在 $f(x)$ 的连续点 x, $F'(x) = f(x)$, 即

$$f(x) = F'(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{P\{x < X \leq x + \Delta x\}}{\Delta x}.$$

其中 $P\{x < X \leq x + \Delta x\} \approx f(x) \cdot \Delta x$ 表示 X 落在 x 附近 $(x, x + \Delta x]$ 上的概率近似等于 $f(x)\Delta x$.

注

(1) $f(X_2) > f(x_1)$ 表示落在 x_2 附近的概率大于落在 x_1 附近的概率，而不是取 x_2 的概率大于取 x_1 的概率；

(2) $f(x)$ 的值是可以大于 1.

(3) $f(x) \underset{\frac{d}{dx}F(x)}{\overset{\int_{-\infty}^{x} f(t)\,dt}{\rightleftharpoons}} F(x).$

例

设随机变量 X 具有概率密度

$$f(x)=\begin{cases} kx & 0\le x<3;\\ 2-\frac{x}{2} & 3\le x\le 4;\\ 0 & \text{其他}. \end{cases}$$

(1) 确定常数 k.

(2) 求 X 的分布函数 $F(x)$.

(3) 求 $P\{1<X\le \frac{7}{2}\}$.

解: (1) 由 $\int_{-\infty}^{+\infty} f(x)\,dx = 1$, 得 $k = \frac{1}{6}$.

(2)

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$
$$= \begin{cases} 0 & x < 0; \\ \frac{x^2}{12} & 0 \le x < 3; \\ -3 + 2x - \frac{x^2}{4} & 3 \le x < 4; \\ 1 & x \ge 4. \end{cases}$$

(3)$P\{1 < X \le \frac{7}{2}\} = F(\frac{7}{2}) - F(1) = \frac{41}{48}$. □

例

设连续型随机变量 X 的分布函数为

$$F(x) = \begin{cases} 0 & x < 0; \\ \frac{x}{3} & 0 \le x < 3; \\ 1 & x \ge 3. \end{cases}$$

求 X 的概率密度 $f(x)$.

解: $f(x) = F'(x) = \begin{cases} \frac{1}{3} & 0 < x < 3; \\ 0 & \text{其他}. \end{cases}$ $\square$

均匀分布

定义

若 X 的概率密度为

$$f(x)=\begin{cases}\frac{1}{b-a} & a<x<b;\\ 0 & \text{其他}.\end{cases}$$

则称 X 在 (a,b) 上服从**均匀分布**, 记为 $X\sim U(a,b)$.

性质

均匀分布具有等可能性.

即 $\forall a < k < k + l < b$, 均有

$$P\{k < X < k + l\} = \int_k^{k+l} \frac{1}{b-a} dt = \frac{l}{b-a},$$

与 k 无关, 仅与 l 有关. 即服从 $U(a, b)$ 上的均匀分布的 X 落入 (a, b) 中任意子区间上的概率只与区间长度有关, 与区间所处位置无关. 即 X 落入 (a, b) 中的等长度的任意子区间上是等可能的.

若 $X \sim \cup(a, b)$, 则 $P\{a < X < b\} = 1$,
且分布函数为

$$F(x) = \begin{cases} 0 & x < a; \\ \frac{x-a}{b-a} & a \leq x < b; \\ 1 & x \geq b. \end{cases}$$

当 $a \leq x \leq b$ 时,

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{a}^{x} \frac{1}{b-a} = \frac{x-a}{b-a}.$$

例

在区间 $(-1,2)$ 上随机取一数 X, 求

(1) 写出 X 的概率密度函数;

(2) 该数在 $(-0.5,1)$ 中的概率;

(3) 该数为正数的概率.

例

在区间 $(-1,2)$ 上随机取一数 X, 求

(1) 写出 X 的概率密度函数;

(2) 该数在 $(-0.5,1)$ 中的概率;

(3) 该数为正数的概率.

解:(1) X 在 $(-1,2)$ 上服从均匀分布, 故

$$f(x)=\begin{cases}\frac{1}{3} & -1<x<2;\\ 0 & \text{其他}.\end{cases}$$

(2)$P\{-0.5<X<1\}=\int_{-0.5}^{1} f(x)\,dx=\frac{1}{2}$.

(3)$P\{X>0\}=\int_{0}^{+\infty} f(x)\,dx=\frac{2}{3}$. □

均匀分布的概率计算

若 $X \sim \mathrm{U}(a, b)$, 则 $\forall I \subset \mathbb{R}$, 有

法一:

$$P\{X \in I\} = \int_I f(x)\,dx,$$

法二:

$$P\{X \in I\} = \frac{I \cap (a, b) \text{ 的长度}}{(a, b) \text{ 的长度}}.$$

指数分布

定义

若 X 的概率密度函数为

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0; \\ 0 & x \le 0. \end{cases}$$

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

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\theta}} & x > 0; \\ 0 & x \le 0. \end{cases}$$

性质

指数分布具有无记忆性. 即 $\forall s, t > 0$, 有

$$P\{X > s + t | X > s\} = P\{X > t\}.$$

性质

指数分布具有无记忆性. 即 $\forall s, t > 0$, 有

$$P\{X > s+t | X > s\} = P\{X > t\}.$$

证明:

$$\begin{aligned} P\{X > s+t | X > s\} &= \frac{P\{\{X > s+t\} \cap \{X > s\}\}}{P\{X > s\}} \\ &= \frac{P\{X > s+t\}}{P\{X > s\}} = \frac{1 - F(s+t)}{1 - F(s)} \\ &= e^{-\frac{t}{\theta}} = P\{X > t\} \qquad \square \end{aligned}$$

若把 X 记为一元件的寿命. 已知元件使用了 s 小时, 总共能使用至少 $s+t$ 小时的概率与从开始使用时算起它至少能使用 t 小时的概率相等. 元件对它已使用过 s 小时没有记忆.

例

设某人电话通话时间 X (分钟) 服从指数分布, 概率密度

$$f(x)=\begin{cases}\frac{1}{15}e^{-\frac{x}{15}} & x>0;\\ 0 & x\leq 0.\end{cases}$$

求

(1) 他通话时间在 $10\sim 20$ 分钟之间的概率.

(2) 若他已打了 10 分钟, 求他继续通话超过 15 分钟的概率.(即, 若他已打了 10 分钟, 求他总共通话超过 25 分钟的概率).

解: (1)$P\{10 < X < 20\} = \int_{10}^{20} f(x)dx = e^{-\frac{2}{3}} - e^{-\frac{4}{3}}$.

利用分布函数 $F(x) = \begin{cases} 1 - e^{-\frac{x}{15}} & x > 0; \\ 0 & x \le 0. \end{cases}$

$P\{10 < X < 20\} = F(20) - F(10) = e^{-\frac{2}{3}} - e^{-\frac{4}{3}}$.

(2) 指数分布无记忆性.

$$\begin{aligned} P\{X > 25 | X > 10\} &= P\{X > 15\} \\ &= \int_{15}^{+\infty} f(x)dx = e^{-1} \end{aligned} \qquad \square$$

例

设一地段相邻两次交通事故的间隔时间 (小时) X 服从参数为 $\frac{13}{2}$ 的指数分布. 求已知在过去的 13 小时中没有发生交通事故, 那么在未来的 2 小时内不发生事故的概率.

例

设一地段相邻两次交通事故的间隔时间 (小时) X 服从参数为 $\frac{13}{2}$ 的指数分布. 求已知在过去的 13 小时中没有发生交通事故, 那么在未来的 2 小时内不发生事故的概率.

解: $X \sim E(\theta)$, $\theta = \frac{13}{2}$,

$$f(x) = \begin{cases} \frac{2}{13} e^{-\frac{2x}{13}} & x > 0; \\ 0 & x \leq 0. \end{cases}$$

所以 $P\{X > 15 | X > 13\} = P\{X > 2\}$

$$= \int_2^{+\infty} f(x)\,dx = 1 - F(2) = e^{-\frac{4}{13}}. \qquad \square$$

指数分布的用途

- 表示独立随机事件发生的间隔, 比如旅客进机场的时间间隔、维基百科新条目出现的时间间隔等等;.
- 在排队登记中, 一个顾客接受服务的时间长短也可用指数分布近似.
- 无记忆性的现象 (连续时).

正态分布 (高斯分布 Gauss)

例

若连续型随机变量 X 的概率密度为

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

其中 $\mu,\sigma(\sigma>0)$ 为常数, 则称 X 服从参数为 μ,σ 的正态分布或高斯分布. 记 $X\sim N(\mu,\sigma^2)$.

- 非负性：$f(x) \geq 0$.
- 规范性：$\int_{-\infty}^{+\infty} f(x)\,dx = 1$

- 非负性：$f(x) \geq 0$.
- 规范性：$\int_{-\infty}^{+\infty} f(x)\,dx = 1$

证：令 $(x-\mu)/\sigma = t$, 则

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

记 $I = \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\,dt$, 则

$$I^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{t^2+u^2}{2}}\,dt du.$$

取极坐标变换, 令 $t = r\cos\theta$, $u = r\sin\theta$, 则

$$I^2 = \int_{0}^{2\pi} \int_{0}^{+\infty} r e^{-\frac{r^2}{2}}\,dr d\theta = 2\pi.$$

$I > 0$, 则 $I = \sqrt{2\pi}$, 故 $\int_{-\infty}^{+\infty} f(x)\,dx = 1$.

性质 (正态分布的性质)

(1) $f(x)$ 关于 $x=\mu$ 对称, 即 $\forall h>0$,

$P\{\mu-h<X\leq\mu\}=P\{\mu<X\leq\mu+h\}$.

(2) 当 $x\leq\mu$ 时, $f(x)$ 严格单增.

(3) $f_{\max})=f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$, 表明 x 离 μ 越远, $f(x)$ 值越小, 同样长度的区间, 当区间离 μ 远时, X 落在这个区间上的概率越小.

性质 (正态分布的性质)

(4) 在 $x = \boldsymbol{\mu} \pm \boldsymbol{\sigma}$ 处有拐点, 以 x 轴为渐近线.

(5) $\lim\limits_{x \to \pm\infty} f(x) = 0.$

(6) 当 $\boldsymbol{\sigma}$ 固定, 改变 $\boldsymbol{\mu}$ 的大小时, $f(x)$ 的图像形状不变, 沿 x 轴平移. $\boldsymbol{\mu}$ 为位置参数, 决定对称轴的位置.

(7) 当 $\boldsymbol{\mu}$ 固定, 改变 $\boldsymbol{\sigma}$ 的大小时, $f(x)$ 的图像形状变, $\boldsymbol{\sigma}$ 越小, 图像越高越瘦; $\boldsymbol{\sigma}$ 越大, 图像越胖. $\boldsymbol{\sigma}$ 为尺度参数, 决定曲线分散程度.

正态分布的用途

- 自然界和人类社会中很多现象可以看成正态分布, 比如, 人的身高, 体重, 医学检验指标, 测量误差, 半导体器件中的热噪声电流或电压.
- 正态分布时最常见的一种分布, 一个变量如果受到大量微小的, 独立的随机因素的影响, 则一般是正态随机变量.
- 二项分布, 泊松分布的极限分布是正态分布.

正态分布的分布函数

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

正态分布的概率计算

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

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

用 Matlab, excel, R 等软件; 数值积分; 转为标准正态分布, 利用标准正态分布表求.

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 $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, 关于 y 轴对称.

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

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

性质

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

证: $Z = \frac{X-\mu}{\sigma}$ 的分布函数为

$P\{Z \le x\} = P\{\frac{X-\mu}{\sigma} \le x\} = P\{X \le \mu + \sigma x\} =$

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

令 $\frac{t-\mu}{\sigma} = \mu$, 得

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

故 $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

性质

若 $X \sim N(\mu, \sigma^2)$, 则 $F_X(x) = P\{X \leq x\} = P\{\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\} = \Phi(\frac{x-\mu}{\sigma})$. 对 $\forall (x_1, x_2]$, $P\{x_1 < X \leq x_2\} = P\{\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{x_2-\mu}{\sigma}\} = \Phi(\frac{x_2-\mu}{\sigma}) - \Phi(\frac{x_1-\mu}{\sigma})$.

证: $P\{x_1 < X \leq x_2\} = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{\frac{x_1-\mu}{\sigma}}^{\frac{x_2-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{\frac{x_2-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt - \int_{-\infty}^{\frac{x_1-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(\frac{x_2-\mu}{\sigma}) - \Phi(\frac{x_1-\mu}{\sigma})$.

例

用天平称一实际重量为 μ 得物体, 天平得读数为随机变量 X, 若 $X \sim N(\mu, \sigma^2)$, 求读数与 μ 的偏差在 3σ 范围内的概率.

解: $P\{|X-\mu|<3\sigma\} = P\{-3\sigma < X-\mu < 3\sigma\} = P\{-3 < \frac{X-\mu}{\sigma} < 3\} = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 0.9973.$

在 $(\mu-3\sigma, \mu+3\sigma)$ 内几乎是肯定的概率. 3σ 法则.

例

将一温度调节器放置在储存着某种液体的容器内, 调节器调整在 $d^\circ C$, 液体的温度 X (以 $^\circ C$ 计) 是一个随机变量, 且 $X \sim N(d, 0.5^2)$.

(1) 若 $d = 90^\circ C$, 求 X 小于 $89^\circ C$ 的概率.

(2) 若要求保持液体的温度至少为 $80^\circ C$ 的概率不低于 0.99, 问 d 至少为多少?

解: (1) $P\{X<89\}=P\{\frac{X-90}{0.5}<\frac{89-90}{0.5}\}=$
$\Phi(-2)=1-\Phi(2)=1-0.9772=0.0228$.
(2)$P\{X\geq 80\}\geq 0.99=P\{\frac{X-d}{0.5}\geq\frac{80-d}{0.5}\}=1-$
$P\{\frac{X-d}{0.5}<\frac{80-d}{0.5}\}=1-\Phi(\frac{80-d}{0.5})\geq 0.99=\Phi(2.327)$.
$\Phi(\frac{80-d}{0.5})\leq 0.01$.
即 $d\geq 81.1635$