概率论与数理统计

金玲飞

复旦大学计算机学院
Email: lfjin@fudan.edu.cn

2018.10.23

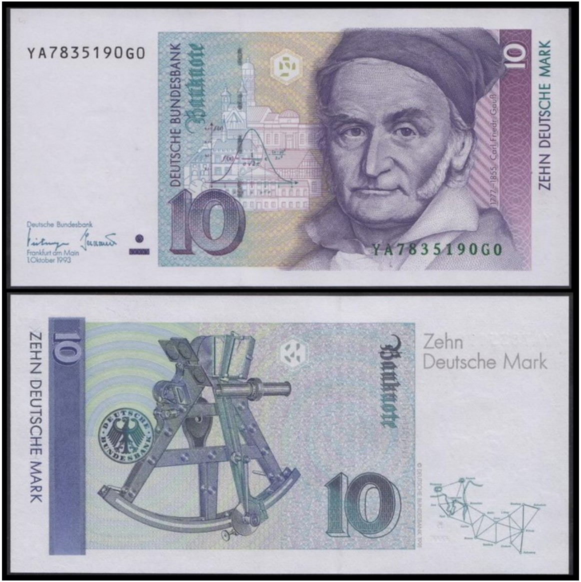
正态分布

定义 (Normal distribution)

设X是连续型随机变量，若X的概率密度函数为

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

（其中$-\infty < \mu < \infty, \sigma > 0$为常数）则称$X$服从参数为$\mu, \sigma^2$的正态分布（或高斯分布），记为 $X \sim N(\mu, \sigma^2)$，相应的概率密度为正态密度。


YA7835190G0
DEUTSCHE BUNDESBANK
Banknote
10
ZEHN DEUTSCHE MARK
1777–1855 Carl Friedr. Gauß
Deutsche Bundesbank
Frankfurt am Main
1.Oktober 1993
YA7835190G0
ZEHN DEUTSCHE MARK
10
Banknote
Zehn
Deutsche Mark
10

正态分布的概率分布函数

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

正态分布的概率分布函数

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

正态分布是概率论中最重要的一个分布：

- 一般说来，若决定某一数量指标的随机因素很多，且相互独立，每个因素起的作用都不大，则这个指标近似服从正态分布（中心极限定理）。
- 理论上，正态分布具有优良的性质，许多分布可用正态分布来逼近，另一些分布可用正态分布来导出。

正态密度函数的性质

- 对称性: 关于$x = \mu$ 对称
- 最大值： $\frac{1}{\sqrt{2\pi}\sigma}$
- 拐点： $\mu \pm \sigma$
- 渐近线：x轴
- μ, σ对函数图形的影响

正态密度函数的性质

- 对称性: 关于$x=\mu$ 对称
- 最大值： $\frac{1}{\sqrt{2\pi}\sigma}$
- 拐点： $\mu \pm \sigma$
- 渐近线：x轴
- μ, σ对函数图形的影响

正态分布函数的性质

$$F(\mu) = 1 - F(\mu) \Rightarrow F(\mu) = \frac{1}{2}$$
$$F(\mu - x) = 1 - F(\mu + x)$$

例子 (2.3.8)

设随机变量$X \sim N(3, \sigma^2)$，且$P(3 < x < 6) = 0.4$。求概率$P(X < 0)$。

正态分布的“3σ原则”

正态分布的“3σ原则”

$$P(-\sigma \leq X-\mu \leq \sigma) \approx 0.683$$
$$P(-2\sigma \leq X-\mu \leq 2\sigma) \approx 0.954$$
$$P(-3\sigma \leq X-\mu \leq 3\sigma) \approx 0.997$$

正态分布的“3σ原则”

正态分布的“3σ原则”

$$P(-\sigma \le X-\mu \le \sigma) \approx 0.683$$
$$P(-2\sigma \le X-\mu \le 2\sigma) \approx 0.954$$
$$P(-3\sigma \le X-\mu \le 3\sigma) \approx 0.997$$

说明$\{|X-\mu| > 3\sigma\}$的概率很小(0.003)，即基本认为

$$-3\sigma+\mu \le X \le 3\sigma+\mu$$

Standard normal distribution

标准正态分布$N(0,1)$

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

$$\Phi(x) = \int_{-\infty}^{x} \varphi(y) dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

Standard normal distribution

标准正态分布$N(0,1)$

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

$$\Phi(x) = \int_{-\infty}^{x} \varphi(y) dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

- $\varphi(-x) = \varphi(x)$
- $\Phi(-x) = 1 - \Phi(x)$
- $P(|X| \leq x) = 2\phi(x) - 1, x > 0$

正态分布标准化

一般的正态变量$X \sim N(\mu, \sigma^2)$，其分布函数有，

正态分布标准化

一般的正态变量$X \sim N(\mu, \sigma^2)$，其分布函数有，

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\
&= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \quad (u = \frac{y-\mu}{\sigma}) \\
&= \Phi(\frac{x-\mu}{\sigma})
\end{aligned}
$$

正态分布标准化

一般的正态变量$X \sim N(\mu, \sigma^2)$，其分布函数有，

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \quad (u = \frac{y-\mu}{\sigma}) \\ &= \Phi(\frac{x-\mu}{\sigma}) \end{aligned}$$

正态概率计算公式

$$P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right), -\infty \leq a < b \leq \infty$$

例子 (2.3.10)

由历史记录知，某地区年总降雨量$X \sim N(600, 150^2)$（单位：*mm*）。求

- 明年降雨量在*400mm*~*700mm*之间的概率
- 明年降雨量至少为*300mm*的概率
- 明年降雨量小于何值的概率为*0.1*

见书P70！

例子 (2.3.11)

某工厂生产的电子管的寿命X（单位：h）服从正态分布$N(1600, \sigma^2)$。如果要求该厂生产的电子管的寿命在*1200h*以上的概率不小于 *0.96*，求σ的值。

见书P70！

上α分位数

定义 (2.3.6 上α分位数)

设$X \sim N(0,1)$，对给定的$\alpha(0<\alpha<1)$，若数z_α满足条件

$$P(X > z_\alpha) = \int_{z_\alpha}^{\infty} \phi(x)dx = \alpha$$

即

$$\Phi(z_\alpha) = 1 - \alpha$$

则称z_α为标准正态分布的上α分位数。

定义 (双侧α分位数)

对于给定的$0 < \alpha < 1$，求出常数c_α使得 $P(|X| > c_\alpha) = \alpha$
称c_α为双侧α分位数。

$$c_\alpha = z_{\frac{\alpha}{2}}$$

Γ分布

定义 (Γ分布)

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0, x \le 0 \end{cases}$$

记$X \sim \Gamma(\alpha, \lambda)$。当$\alpha = n$为自然数时， 又称爱尔兰分布。

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$$

$$\Gamma(1) = 1, \Gamma(1/2) = \sqrt{\pi}, \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

Γ分布

定义 (Γ分布)

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0, x \leq 0 \end{cases}$$

记$X \sim \Gamma(\alpha, \lambda)$。当$\alpha = n$为自然数时， 又称爱尔兰分布。

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$$

$$\Gamma(1) = 1, \Gamma(1/2) = \sqrt{\pi}, \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

- 当$\alpha = 1$时，就是指数分布
- 当$\alpha = \frac{n}{2}, \lambda = \frac{1}{2}$，就是$\chi^2$分布

例子

设随机变量X的绝对值不大于*1*，$P(X = -1) = 1/8, P(X = 1) = 1/4$。在事件$\{-1 < X < 1\}$出现的条件下，$X$在$(-1, 1)$的任一子区间上取值的条件概率与该子区间的长度成正比，求

- X的分布函数$F(x)$。
- $P(X < 0)$。

2.4 随机变量的函数及其分布

随机变量的函数

- 若 X 是概率空间上的随机变量， $g(x)$ 为连续函数或单调函数，则 $Y = g(x)$ 也是随机变量。
- 已知 X 的分布和函数 $g(x)$，求它的函数 $Y = g(X)$ 的分布。

离散型

求离散型随机变量的函数的分布律的方法

若离散随机变量X的分布列为

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

则$Y = g(X)$也是一个离散型随机变量，其分布列可以表示为

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

其中某些$g(x_i)$的取值相等时，将相等的值合并，并将对应的概率相加。

离散型随机变量函数的分布

例子 (2.4.1)

设随机变量X的分布律为

X	-1	0	1	2
P	0.2	0.3	0.1	0.4

求$Y=(X-1)^2$的分布律。

例子 (2.4.2)

设X的分布律为

X	-2	-1	0	1	2
P	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{10}$

求如下两个随机变量函数的分布律

- $Y = 2X + 1$
- $Z = X^2$

连续型随机变量函数的分布

- 若X 是连续型随机变量，则$Y = g(x)$ 不一定是连续型随机变量。

例子 (2.4.3)

设随机变量X有（分段连续）概率密度$f_X(x)$，而$Y = aX + b$，这里a, b是常数且$a \neq 0$，求Y的概率密度$f_Y(y)$。

连续型随机变量函数的分布

1：当$g(X)$为严格单调时

设X是连续随机变量，其密度函数是$f_X(x)$。$Y = g(X)$是另一个随机变量。若$g(X)$ 严格单调，其反函数$h(y)$有连续导函数，则$Y = g(X)$的密度函数是

$$f_Y(y) = \begin{cases} f_X(h(y))|h'(y)|, & a < y < b \\ 0, & \text{其它} \end{cases}$$

其中$a = \min\{g(-\infty), g(\infty)\}, b = \max\{g(-\infty), g(\infty)\}$.

例子 (2.4.3)

设随机变量X有（分段连续）概率密度$f_X(x)$，而$Y = aX + b$，这里a, b是常数且$a \neq 0$，求Y的概率密度$f_Y(y)$。

例子 (2.4.3)

设随机变量X有（分段连续）概率密度$f_X(x)$，而$Y = aX + b$，这里a, b是常数且$a \neq 0$，求Y的概率密度$f_Y(y)$。

- 若$X \sim N(\mu, \sigma^2)$，求$Y = aX + b$的概率密度？

例子 (2.4.3)

设随机变量X有（分段连续）概率密度$f_X(x)$，而$Y = aX + b$，这里a, b是常数且$a \neq 0$，求Y的概率密度$f_Y(y)$。

- 若$X \sim N(\mu, \sigma^2)$，求$Y = aX + b$的概率密度？

$$aX + b \sim N(a\mu + b, a^2\sigma^2)$$

正态随机变量的线性函数仍为正态随机变量。

- 正态分布的标准化:若$X \sim N(\mu, \sigma^2)$，求$Y = \frac{X-\mu}{\sigma}$的概率密度？

Logarithmic normal distribution

例子 (2.4.4 对数正态分布)

设随机变量$X \sim N(\mu, \sigma^2)$，$Y = e^X$。求Y 的概率密度。

Logarithmic normal distribution

例子 (2.4.4 对数正态分布)

设随机变量$X \sim N(\mu, \sigma^2)$，$Y = e^X$。求Y 的概率密度。

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma y} \exp\left\{-\frac{1}{2\sigma^2}(\ln y - \mu)^2\right\}, & y > 0 \\ 0, & y \le 0 \end{cases}$$

Logarithmic normal distribution

例子 (2.4.4 对数正态分布)

设随机变量$X \sim N(\mu, \sigma^2)$，$Y = e^X$。求Y的概率密度。

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma y} \exp\left\{-\frac{1}{2\sigma^2}(\ln y - \mu)^2\right\}, & y > 0 \\ 0, & y \le 0 \end{cases}$$

- 分布函数也可以用标准正态分布表示
- 是一种常用的寿命分布。

例子 (2.4.5)

设随机变量X有（分段连续的）概率密度$f_X(x)$，求$Y = cX^2$（$c > 0$为常常数）的概率密度$f_Y(y)$。

连续型随机变量函数的分布

2:当$g(X)$为其它形式时
分布函数微分法

- 化简$Y = g(X)$的分布函数$F_Y(y) = P(g(X) \leq y)$，直到看出除有限多个点外导数$F_Y'(y)$存在且连续为止
- 求导数$F_Y'(y)$，令

$$f_Y(y) = \begin{cases} F_Y'(y), & \text{如果存在} \\ 0, & \text{否则} \end{cases}$$

- 结论：$Y = g(X)$的概率密度即为$f_Y(y)$。

例子 (2.4.5)

设随机变量X有（分段连续的）概率密度$f_X(x)$，
求$Y = cX^2$（$c > 0$为常数）的概率密度$f_Y(y)$。

例子 (2.4.6)

设随机变量X的密度函数为

$$f_X(x) = \begin{cases} \frac{2x}{\pi^2}, & 0 < x < \pi, \\ 0, & 其它 \end{cases}$$

求随机变量$Y = sinX$的密度函数。

例子

若随机变量X的分布函数$F_X(x)$为严格单调递增的连续函数，其反函数$F_X^{-1}(y)$ 存在，证明$Y = F_X(X)$ 服从$(0, 1)$上的均匀分布$U(0, 1)$。

例子

若随机变量X的分布函数$F_X(x)$为严格单调递增的连续函数，其反函数$F_X^{-1}(y)$ 存在，证明$Y = F_X(X)$ 服从$(0,1)$上的均匀分布$U(0,1)$。

- $y < 0$, $\{F_X(X) \leq y\}$ 是不可能事件，$F_Y(y) = 0$.
- $0 \leq y < 1$, $F_Y(y) = P(F_X(X) \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y$.
- $y \geq 1$, $F_Y(y) = 1$.

$$Y \sim U(0,1)$$

例子

若随机变量X的分布函数$F_X(x)$为严格单调递增的连续函数，其反函数$F_X^{-1}(y)$ 存在，证明$Y = F_X(X)$ 服从$(0,1)$上的均匀分布$U(0,1)$。

- $y < 0$, $\{F_X(X) \leq y\}$ 是不可能事件，$F_Y(y) = 0$.
- $0 \leq y < 1$, $F_Y(y) = P(F_X(X) \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y$.
- $y \geq 1$, $F_Y(y) = 1$.

$$Y \sim U(0,1)$$

任意连续随机变量都可通过其分布函数与均匀分布发生关系。