

華東師範大學
EAST CHINA NORMAL UNIVERSITY

概率论 与数理统计 笔记

© syqwq
East China Normal University

目录

1. 随机事件与概率	3
1.1. 随机事件及其运算	3
1.2. 概率的定义及其确定方法	5
1.3. 概率的性质	5
1.4. 条件概率	7
1.5. 独立性	9
2. 随机变量及其分布	10
2.1. 随机变量及其分布	10
2.2. 随机变量的数学期望	12
2.3. 随机变量的方差和标准差	13
2.4. 常用离散分布	15
2.4.1. 二项分布	15
2.4.2. 泊松分布	17
2.4.3. 超几何分布	18
2.4.4. 几何分布	18
2.4.5. 负二项分布	19
2.5. 常用连续分布	19
2.5.1. 正态分布	19
2.5.2. 均匀分布	24
2.5.3. 指数分布	25
2.5.4. 伽马分布	27
2.5.5. 贝塔分布	29
2.5.6. 总结	30
2.6. 随机变量函数的分布	32
2.6.1. 离散随机变量函数的分布	32
2.6.2. 连续随机变量函数的分布	32
I. Definitions	1
II. Theorems	1
III. Figures	1

Chapter 1

随机事件与概率

1.1 随机事件及其运算

定义 1.1.1 样本空间

随机现象的一切可能基本结果组成的集合称为**样本空间**，记作 $\Omega = \{\omega\}$.

定义 1.1.2 样本点

样本空间中的每一个基本结果称为**样本点**，记作 $\omega \in \Omega$.

定义 1.1.3 离散样本空间

如果样本空间 Ω 中的样本点是可数个的，则称样本空间 Ω 为**离散样本空间**.

定义 1.1.4 连续样本空间

如果样本空间 Ω 中的样本点是不可数个的，则称样本空间 Ω 为**连续样本空间**.

定义 1.1.5 随机事件

样本空间 Ω 的任一子集称为**随机事件**，简称**事件**，记作 $A \subseteq \Omega$.

注解 1.1.1 随机事件的分类

- (1) **基本事件**：包含单个样本点的事件，记作 $A = \{\omega\}$.
- (2) **必然事件**：包含样本空间中所有样本点的事件，记作 $A = \Omega$.
- (3) **不可能事件**：不包含任何样本点的事件，记作 $A = \emptyset$.

注解 1.1.2 事件的关系

设 $A, B \subseteq \Omega$ 是样本空间 Ω 上的两个事件，

- (1) **包含关系**：若属于 A 的任一样本点也属于 B ，则称事件 A 包含于事件 B ，记作 $A \subseteq B$.
- (2) **等价关系**：若属于 A 的任一样本点也属于 B ，且属于 B 的任一样本点也属于 A ，则称事件 A 与事件 B 等价，记作 $A = B$.
- (3) **互不相容**：若事件 A 与事件 B 没有共同的样本点，则称事件 A 与事件 B 互不相容，记作 $A \cap B = \emptyset$.

定义 1.1.6 可测函数

设 $(X, \mathcal{X}), (Y, \mathcal{Y})$ 是两个可测空间, 映射 $f: X \rightarrow Y$ 如果满足对任意 $B \in \mathcal{Y}$, 都有 $f^{-1}(B) = \{x \in X \mid f(x) \in B\} \in \mathcal{X}$, 则称映射 f 为**可测函数**.

定义 1.1.7 随机变量

设 (Ω, \mathcal{F}) 是一个可测空间, 如果映射 $X: \Omega \rightarrow \mathbb{R}$ 是从样本空间 Ω 到实数集 \mathbb{R} 的可测函数, 则称映射 X 为**随机变量**.

定义 1.1.8 事件的并

设 $A, B \subseteq \Omega$ 是样本空间 Ω 上的两个事件, 则定义它们的**并**为

$$A \cup B := \{\omega \mid \omega \in A \vee \omega \in B\}$$

定义 1.1.9 事件的交

设 $A, B \subseteq \Omega$ 是样本空间 Ω 上的两个事件, 则定义它们的**交**为

$$A \cap B := \{\omega \mid \omega \in A \wedge \omega \in B\}$$

定义 1.1.10 事件的差

设 $A, B \subseteq \Omega$ 是样本空间 Ω 上的两个事件, 则定义它们的**差**为

$$A - B := \{\omega \mid \omega \in A \wedge \omega \notin B\}$$

定义 1.1.11 对立事件

设 $A \subseteq \Omega$ 是样本空间 Ω 上的一个事件, 则定义它的**对立事件**为

$$\bar{A} := \Omega - A = \{\omega \mid \omega \in \Omega \wedge \omega \notin A\}$$

命题 1.1.1 互为对立事件

事件 $A, B \subseteq \Omega$ 互为对立事件的充分必要条件是 $A \cup B = \Omega$ 且 $A \cap B = \emptyset$.

注解 1.1.3

- (1) 对立事件一定是互不相容的. 但是, 互不相容的事件不一定是对立事件.
- (2) $A - B = A\bar{B}$.

性质 1.1.1 事件的运算性质

设 $A, B, C \subseteq \Omega$ 是样本空间 Ω 上的任意三个事件, 则有:

- (1) 交换律: $A \cup B = B \cup A, A \cap B = B \cap A$.
- (2) 结合律: $(A \cup B) \cup C = A \cup (B \cup C), (A \cap B) \cap C = A \cap (B \cap C)$.
- (3) 分配律: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C), A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
- (4) De Morgan 律: $\overline{A \cup B} = \overline{A} \cap \overline{B}, \overline{A \cap B} = \overline{A} \cup \overline{B}$.

定义 1.1.12 σ -代数

设 Ω 是一个集合, \mathcal{F} 是 Ω 的幂集的子集. 如果 \mathcal{F} 满足:

- (1) $\emptyset \in \mathcal{F}$
- (2) 对任意 $A \in \mathcal{F}$, 有 $\overline{A} \in \mathcal{F}$
- (3) 对任意可列个集合 $A_1, A_2, A_3, \dots \in \mathcal{F}$, 有 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

则称 (Ω, \mathcal{F}) 是一个可测空间, \mathcal{F} 称为 Ω 上的一个 σ -代数.

定义 1.1.13 样本空间的分割

设 Ω 为样本空间, A_1, A_2, \dots, A_n 为样本空间 Ω 上的 n 个事件. 如果 A_1, A_2, \dots, A_n 满足:

- (1) $\bigcup_{i=1}^n A_i = \Omega$
- (2) 对任意 $i, j = 1, 2, \dots, n$ 且 $i \neq j$, 有 $A_i \cap A_j = \emptyset$

则称事件组 $\{A_1, A_2, \dots, A_n\}$ 为样本空间 Ω 的一个分割.

1.2 概率的定义及其确定方法

定义 1.2.1 测度空间

设 (Ω, \mathcal{F}) 是一个可测空间, $\mu: \mathcal{F} \rightarrow [0, +\infty]$ 是一个函数. 如果 μ 满足:

- (1) $\mu(\emptyset) = 0$
- (2) 对任意互不相容的集合 $A_1, A_2, A_3, \dots \in \mathcal{F}$, 有 $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$

则称 $(\Omega, \mathcal{F}, \mu)$ 为一个测度空间, μ 称为 Ω 上的一个测度.

定义 1.2.2 概率空间

设 (Ω, \mathcal{F}, P) 是一个测度空间, 且满足 $P(\Omega) = 1$. 则称 (Ω, \mathcal{F}, P) 为一个概率空间, P 称为 Ω 上的一个概率.

1.3 概率的性质

性质 1.3.1 互补事件的概率

设 $A \subseteq \Omega$ 是样本空间 Ω 上的一个事件, 则有 $P(\bar{A}) = 1 - P(A)$.

性质 1.3.2 概率的单调性

设 $A, B \subseteq \Omega$ 是样本空间 Ω 上的两个事件, 且 $A \subset B$, 则有 $P(B - A) = P(B) - P(A)$.

推論 1.3.1

设 $A, B \subseteq \Omega$ 是样本空间 Ω 上的两个事件, 且 $A \subseteq B$, 则有 $P(A) \leq P(B)$.

性质 1.3.3 概率的加法公式

设 $A, B \subseteq \Omega$ 是样本空间 Ω 上的两个事件, 则有 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

推論 1.3.2 容斥原理

设 $A_1, A_2, \dots, A_n \subseteq \Omega$ 是样本空间 Ω 上的 n 个事件, 则有

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) \\ &\quad + \dots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right) \end{aligned}$$

推論 1.3.3 概率的半可加性

设 $A_1, A_2, A_3, \dots \subseteq \Omega$ 是样本空间 Ω 上的可列个事件, 则有 $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$.

定义 1.3.1 极限事件

(1) 对 \mathcal{F} 中任一单调不减的事件列 $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$, 称可列并 $\bigcup_{i=1}^{\infty} A_i$ 为 $\{A_n\}$ 的极限事件, 记作

$$\lim_{n \rightarrow \infty} A_n := \bigcup_{i=1}^{\infty} A_i$$

(2) 对 \mathcal{F} 中任一单调不增的事件列 $A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$, 称可列交 $\bigcap_{i=1}^{\infty} A_i$ 为 $\{A_n\}$ 的极限事件, 记作

$$\lim_{n \rightarrow \infty} A_n := \bigcap_{i=1}^{\infty} A_i$$

定义 1.3.2 连续性

对 \mathcal{F} 上的一个测度 μ ,

- (1) 如果他对任一单调不减的集合序列 $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$ 满足 $\mu\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n)$, 则称测度 μ 具有**下连续性**.
- (2) 如果他对任一单调不增的集合序列 $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$ 满足 $\mu\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n)$, 则称测度 μ 具有**上连续性**.

性质 1.3.4 概率的连续性

若 P 为可测集 \mathcal{F} 上的概率测度, 则 P 具有上、下连续性.

1.4 条件概率

定义 1.4.1 条件概率

设 (Ω, \mathcal{F}, P) 是一个概率空间, $A, B \subseteq \Omega$ 是样本空间 Ω 上的两个事件, 且 $P(B) > 0$. 定义在事件 B 已经发生的条件下事件 A 发生的概率为**条件概率**, 记作 $P(A|B)$, 其计算公式为

$$P(A|B) = \frac{P(AB)}{P(B)}$$

性质 1.4.1 条件概率是概率

条件概率是概率, 即

- (1) 非负性: 对任意事件 $A \subseteq \Omega$, 有 $P(A|B) \geq 0$.
- (2) 正则性: $P(\Omega|B) = 1$.
- (3) 可列可加性: 对任意互不相容的事件列 $A_1, A_2, A_3, \dots \subseteq \Omega$, 有

$$P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \sum_{i=1}^{\infty} P(A_i | B)$$

性质 1.4.2 乘法公式

- (1) 若 $P(B) > 0$, 则

$$P(AB) = P(B)P(A|B)$$

(2) 若 $P(A_1 A_2 \cdots A_{n-1}) > 0$, 则

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \cdots P(A_n | A_1 A_2 \cdots A_{n-1})$$

定理 1.4.1 全概率公式

设 $B_1, B_2, \dots, B_n \subseteq \Omega$ 是样本空间 Ω 上的一个事件分割, 且对任意 $i = 1, 2, \dots, n$, 有 $P(B_i) > 0$. 则对任意事件 $A \subseteq \Omega$, 有

$$P(A) = \sum_{i=1}^n P(B_i)P(A | B_i)$$

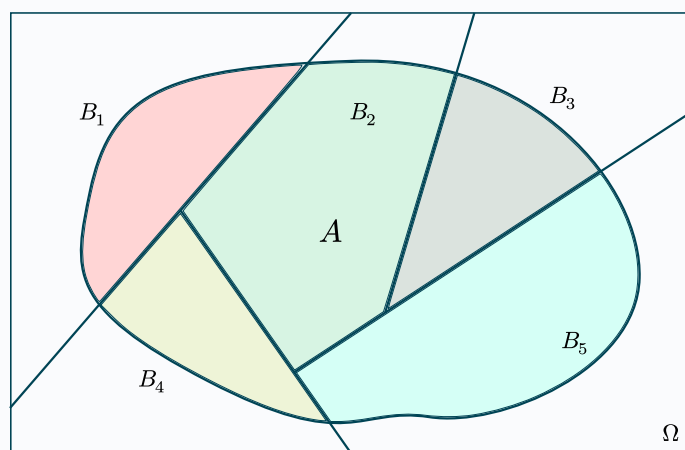


Fig. 1.1 Total Probability Formula

对于全概率公式, 我们要注意以下几点:

(1) 假如 $0 < P(B) < 1$, 则

$$P(A) = P(B)P(A | B) + P(\bar{B})P(A | \bar{B})$$

(2) 条件中 B_1, B_2, \dots, B_n 为样本空间的一个划分可以改成 B_1, B_2, \dots, B_n 互不相容, 且 $A \subset \bigcup_{i=1}^{\infty} B_i$, 全概率公式依然成立.

(3) 对可列无限个事件, 全概率公式依然成立.

定理 1.4.2 Bayes 公式

设 $B_1, B_2, \dots, B_n \subseteq \Omega$ 是样本空间 Ω 上的一个事件分割, 如果 $P(A) > 0, P(B_i) > 0$, 则

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)} \quad i = 1, 2, \dots, n$$

1.5 独立性

定义 1.5.1 两个事件的独立性

设 $A, B \subseteq \Omega$ 是样本空间 Ω 上的两个事件, 如果 $P(AB) = P(A)P(B)$, 则称事件 A 与事件 B 相互独立.

性质 1.5.1

若 A 与 B 独立, 则 A 与 \bar{B} 独立, \bar{A} 与 B 独立, \bar{A} 与 \bar{B} 独立.

定义 1.5.2 多个事件的独立性

设有 n 个事件 $A_1, A_2, \dots, A_n \subseteq \Omega$, 如果对任意的子集 $\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, n\}$, 都有

$$\begin{cases} P(A_{i_1} A_{i_2}) = P(A_{i_1})P(A_{i_2}) \\ P(A_{i_1} A_{i_2} A_{i_3}) = P(A_{i_1})P(A_{i_2})P(A_{i_3}) \\ \dots\dots\dots \\ P(A_{i_1} A_{i_2} \dots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}) \end{cases}$$

则称这 n 个事件 A_1, A_2, \dots, A_n 相互独立.

注解 1.5.1

两两独立 \neq 相互独立.

定义 1.5.3 试验的独立性

设有两个试验 E_1, E_2 , 假如试验 E_1 的任一结果 (事件) 与试验 E_2 的任一结果 (事件) 相互独立, 则称试验 E_1 与试验 E_2 相互独立.

类似的, 可以定义 n 个试验的独立性. 如果试验 E_1, E_2, \dots, E_n 的任一结果都是互相独立的事件, 则称试验 E_1, E_2, \dots, E_n 相互独立.

如果这 n 个试验还是相同的, 则称其为 n 重独立试验.

如果这 n 个试验, 每次试验的结果都只有两个 A, \bar{A} , 则称其为 n 重伯努利试验.

Chapter 2

随机变量及其分布

2.1 随机变量及其分布

定义 1.1.7 随机变量

设 (Ω, \mathcal{F}) 是一个可测空间, 如果映射 $X: \Omega \rightarrow \mathbb{R}$ 是从样本空间 Ω 到实数集 \mathbb{R} 的可测函数, 则称映射 X 为**随机变量**.

定义 2.1.1 离散型随机变量

假如一个随机变量仅可能取有限个或者可列个值, 则称该随机变量为**离散型随机变量**.

定义 2.1.2 连续型随机变量

假如一个随机变量可能取某个区间内的任意值, 则称该随机变量为**连续型随机变量**.

随机变量是从 $\Omega \rightarrow \mathbb{R}$ 的一个可测函数, 若 \mathcal{B} 是某些实数组成的集合, 则 $\{X \in \mathcal{B}\}$ 表示如下集合

$$\{X \in \mathcal{B}\} = \{\omega \in \Omega \mid X(\omega) \in \mathcal{B}\}$$

定义 2.1.3 分布函数

对于随机变量 X , 对于任意 $-\infty < x < \infty$ 称

$$F(x) = P(X \leq x)$$

为随机变量 X 的**分布函数**. 且称 X 服从 $F(x)$, 记作 $X \sim F(x)$

注解 2.1.1

从分布函数的定义可见, 任何随机变量 X 都能诱导出一个分布函数.

性质 2.1.1 分布函数的性质

对于任意随机变量 X , 其分布函数 $F(x)$ 具有如下性质:

- (1) **单调性**: 对任意 $x_1 < x_2$, 有 $F(x_1) \leq F(x_2)$
- (2) **有界性**: $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$, $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$
- (3) **右连续性**: 对任意实数 x_0 , 有 $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$

Proof. 对于右连续性, 取任意实数 x_0 , 对于任意 $\varepsilon > 0$, 令 $\delta = \varepsilon$, 当 $0 < x - x_0 < \delta$ 时, 有 $|F(x) - F(x_0)| = |P(X \leq x) - P(X \leq x_0)| = P(x_0 < X \leq x) \leq P(X \leq x_0 + \delta) - P(X \leq x_0) = F(x_0 + \delta) - F(x_0) < \varepsilon$, 因此 $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$. ■

定义 2.1.4 分布列

对于离散型随机变量 X , 设其可能取的值为 $x_1, x_2, \dots, x_n, \dots$, 则称 x 取 x_i 的概率

$$p_i = p(x_i) = P(X = x_i)$$

为随机变量 X 的**分布列**, 记作 $X \sim \{p_i\}$.

性质 2.1.2 分布列的性质

(1) $p(x_i) \geq 0, i = 1, 2, \dots$

(2) $\sum_{i=1}^{\infty} p(x_i) = 1$

离散型随机变量的分布函数是有限级的阶梯函数.

对于连续型随机变量, 由于单点集是零测集, 因此我们引入概率密度函数.

定义 2.1.5 概率密度函数

设随机变量 X 的分布函数为 $F(x)$, 如果存在非负可积函数 $p(x)$, 使得对于任意实数 x , 都有

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

则称 $p(x)$ 为随机变量 X 的**概率密度函数**. 称 X 为连续型随机变量, 称 $F(x)$ 为连续型分布函数.

可以看出, 在可导的点上有

$$F'(x) = p(x)$$

性质 2.1.3 概率密度函数的性质

对于概率密度函数 $p(x)$, 有如下性质:

(1) $p(x) \geq 0$

(2) $\int_{-\infty}^{\infty} p(x) dx = 1$

例 2.1.1 均匀分布

设随机变量 X 在区间 $[a, b]$ 上服从均匀分布, 记作 $X \sim U(a, b)$, 则其概率密度函数为

$$p(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{else} \end{cases}$$

2.2 随机变量的数学期望

定义 2.2.1 离散型随机变量的数学期望

设离散型随机变量 X 的分布列为 $\{p_i\}$, 如果 $\sum_{i=1}^{\infty} |x_i| p(x_i) < \infty$, 则称

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

为随机变量 X 的**数学期望**.

以上定义中, 要求级数绝对收敛的目的在于使数学期望唯一. 因为随机变量的取值可正可负, 取值次序可先可后, 由无穷级数的理论知道, 如果此无穷级数绝对收敛, 则可保证其和不受次序变动的影响. 由于有限项的和不受次序变动的影响, 故取有限个可能值的随机变量的数学期望总是存在的.

定义 2.2.2 连续型随机变量的数学期望

设连续型随机变量 X 的概率密度函数为 $p(x)$, 如果 $\int_{-\infty}^{\infty} |x| p(x) dx < \infty$, 则称

$$E(X) = \int_{-\infty}^{\infty} x p(x) dx$$

为随机变量 X 的**数学期望**.

定义 2.2.3 数学期望

若 X 是定义在 (Ω, Σ, P) 上的随机变量, 则称 Lebesgue 积分

$$E(X) = \int_{\Omega} X dP$$

为随机变量 X 的**数学期望**. 其中, 概率测度为

$$P(x \in A) = \int_A p(x) dx$$

注解 2.2.1

一般将期望记作 $\mu = \mu_X = E(X)$.

定理 2.2.1 随机变量函数的期望

若随机变量 X 的分布用分布列 $\{p_i\}$ 或者 概率密度函数 $p(x)$ 表示, 则 X 的某一函数 $g(X)$ 的数学期望为

$$E(g(X)) = \begin{cases} \sum_i g(x_i)p(x_i), & \text{离散型随机变量} \\ \int_{-\infty}^{\infty} g(x)p(x) dx, & \text{连续型随机变量} \end{cases}$$

也就是

$$E(g(X)) = \int_{\Omega} g(X) dP$$

性质 2.2.1 期望的性质

- (1) 对于常数 c , 有 $E(c) = c$
- (2) 对于随机变量 X 和常数 c , 有 $E(cX) = cE(X)$
- (3) 对于随机变量 X 和 Y , 有 $E(X + Y) = E(X) + E(Y)$

2.3 随机变量的方差和标准差

定义 2.3.1 方差

若随机变量 X^2 的数学期望 $E(X^2)$ 存在, 则称

$$\text{Var}(X) := E[(X - \mu)^2]$$

为随机变量 X 的方差, 其中 $\mu = E(X)$.

定义 2.3.2 标准差

随机变量 X 的标准差定义为方差的正平方根, 即

$$\sigma = \sigma(X) = \sigma_X = \sqrt{\text{Var}(X)}$$

方差与标准差之间的差别主要在量纲上, 由于标准差与所讨论的随机变量、数学期望有相同的量纲, 其加减 $E(X) \pm k\sigma(X)$ 是有意义的 (k 为正实数), 所以在实际中, 人们比较乐意选用标准差, 但标准差的计算必须通过方差才能算得.

注解 2.3.1

- 如果随机变量 X 的数学期望存在, 其方差不一定存在
- 当 X 的方差存在时, 则 $E(X)$ 必定存在, 其原因在于 $|X| \leq X^2 + 1$ 总是成立的

性质 2.3.1 方差的性质

- (1) $\text{Var}(X) = E(X^2) - E(X)^2$
- (2) 对于常数 c , 有 $\text{Var}(c) = 0$
- (3) 对于随机变量 X 和常数 a, b , 有 $\text{Var}(aX + b) = a^2 \text{Var}(X)$
- (4) 对于随机变量 X 和 Y , 若 X 和 Y 相互独立, 则有 $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Proof. 对第 1 条性质的证明如下

$$\begin{aligned}\text{Var}(X) &= E[(X - \mu)^2] \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2\end{aligned}$$

■

定理 2.3.1 Markov 不等式

若随机变量 X 的数学期望 $E(X)$ 存在, 且 $X \geq 0$, 则对于任意 $a > 0$, 有

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Proof.

$$E(X) = \int_{\Omega} X \, dP \geq \int_{X \geq a} X \, dP \geq \int_{X \geq a} a \, dP = aP(X \geq a)$$

■

定理 2.3.2 Chebyshev 不等式

若随机变量 X 的数学期望 $E(X)$ 和方差 $\text{Var}(X)$ 存在, 则对于任意 $\varepsilon > 0$, 有

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}$$

或

$$P(|X - \mu| < \varepsilon) \geq 1 - \frac{\text{Var}(X)}{\varepsilon^2}$$

Proof. 由 定理 2.3.1 得

$$P(|X - \mu| \geq \varepsilon) = P((X - \mu)^2 \geq \varepsilon^2) \leq \frac{E[(X - \mu)^2]}{\varepsilon^2} = \frac{\text{Var}(X)}{\varepsilon^2}$$

切比雪夫不等式给出大偏差发生概率的上界，这个上界与方差成正比，方差愈大上界也愈大。以下定理进一步说明了方差为 0 就意味着随机变量的取值几乎集中在一点上。

定理 2.3.3 单点分布

若随机变量 X 的方差存在，则

$$\text{Var}(X) = 0 \Leftrightarrow P(X = c) = 1 \Leftrightarrow X = c \text{ a.e.}$$

Proof. 充分性显然，必要性则由切比雪夫不等式得证。对于任意 $\varepsilon > 0$ ，有

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2} = 0$$

因此 $P(|X - \mu| < \varepsilon) = 1$ ，即 X 在 $[\mu - \varepsilon, \mu + \varepsilon]$ 上几乎处处取值。由于 ε 可任意小，故 $X = \mu$ a.e. ■

2.4 常用离散分布

2.4.1 二项分布

定义 2.4.1 二项分布

若随机变量 X 表示在 n 次独立重复的伯努利试验中成功的次数，且每次试验成功的概率为 p ，则称 X 服从参数为 (n, p) 的**二项分布**，记作 $X \sim B(n, p)$ ，其分布列为

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

定义 2.4.2 二点分布

$n = 1$ 时的二项分布称为**二点分布** 或者 **伯努利分布**，记作 $X \sim B(1, p)$ ，其分布列为

$$p(x) = P(X = x) = p^x (1-p)^{1-x}, \quad x = 0, 1$$

假设样本空间只有 A 和 \bar{A} ，由此形成伯努利试验。 n 重伯努利试验由 n 个相同的、相互独立的伯努利试验组成。若将第 i 个伯努利试验中 A 出现的次数记作 X_i 。由于 n 重伯努利试验中，每个伯努利试验是相互独立的，因此 X_i 互相独立，且均服从参数为 $(1, p)$ 的二点分布。设 X 表示 n 重伯努利试验中 A 出现的总次数，则有

$$X = X_1 + X_2 + \dots + X_n$$

就是 n 重伯努利试验中出现的 A 的总次数，服从二项分布 $B(n, p)$ 。

服从二项分布的随机变量总可分解为 n 个独立同为二点分布的随机变量之和。

若 $X \sim B(n, p)$ ，则有

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\ &= np(p+1-p)^{n-1} = np \end{aligned}$$

其中的第二个等号是利用组合数的性质 $k \binom{n}{k} = n \binom{n-1}{k-1}$ 得到的, 可以考虑在 n 个人中选择 k 个人组成小组, 然后再选择一个组长.

定义 2.4.3 原点矩

若随机变量 X 的第 n 阶矩存在, 则称

$$\mu'_n = E(X^n)$$

为随机变量 X 的第 n 阶原点矩.

定义 2.4.4 中心矩

若随机变量 X 的第 n 阶中心矩存在, 则称

$$\mu_n = E[(X - \mu)^n]$$

为随机变量 X 的第 n 阶中心矩, 其中 $\mu = E(X)$.

注解 2.4.1

记 $X^{(r)} := X(X-1)(X-2)\cdots(X-r+1)$

定义 2.4.5 阶乘矩

若随机变量 X 的第 n 阶阶乘矩存在, 则称

$$\mu^{(r)} = E[X^{(r)}]$$

为随机变量 X 的第 n 阶阶乘矩.

若 $X \sim B(n, p)$, 考虑他的阶乘矩

$$\begin{aligned} \mu^{(r)} &= E[X(X-1)(X-2)\cdots(X-r+1)] \\ &= \sum_{k=0}^n k(k-1)(k-2)\cdots(k-r+1) \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=r}^n \frac{k!}{(k-r)!} \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \frac{n!}{(n-r)!} p^r \cdot \sum_{k=r}^n \binom{n-r}{k-r} p^{k-r} (1-p)^{n-k} \\ &= \frac{n!}{(n-r)!} p^r (p+1-p)^{n-r} \end{aligned}$$

$$= n^{(r)} p^r$$

从而 $E(X^2) = E[X(X-1)] + E(X) = n(n-1)p^2 + np = n^2p$.

于是, 若 $X \sim B(n, p)$

$$\text{Var}(X) = E(X^2) - E(X)^2 = n^2p - (np)^2 = np(1-p)$$

注解 2.4.2

阶乘矩是下降幂的形式, 而原点矩是普通幂的形式, 两者可以通过 Stirling 反演相互转换.

2.4.2 泊松分布

定义 2.4.6 泊松分布

若随机变量 X 的分布列为

$$p(k) = P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

则称 X 服从参数为 λ 的泊松分布, 记作 $X \sim P(\lambda)$.

注解 2.4.3

将 $e^{-\lambda}$ 提出后, 其余项是 e^x 在 $x = \lambda$ 处的 Taylor 展开.

若 $X \sim P(\lambda)$, 则

$$E(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

考虑他的阶乘矩

$$\begin{aligned} \mu^{(r)} &= E[X(X-1)(X-2)\cdots(X-r+1)] \\ &= \sum_{k=0}^{\infty} k(k-1)(k-2)\cdots(k-r+1) \frac{\lambda^k e^{-\lambda}}{k!} \\ &= \sum_{k=r}^{\infty} \frac{k!}{(k-r)!} \frac{\lambda^k e^{-\lambda}}{k!} \\ &= \lambda^r e^{-\lambda} \sum_{k=r}^{\infty} \frac{\lambda^{k-r}}{(k-r)!} \\ &= \lambda^r e^{-\lambda} e^{\lambda} \\ &= \lambda^r \end{aligned}$$

因此, 若 $X \sim P(\lambda)$, 则

$$\text{Var}(X) = E(X^2) - E(X)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$$

定理 2.4.1 泊松定理

设随机变量 $X_n \sim B(n, p_n)$, 且当 $n \rightarrow \infty$ 时, $np_n \rightarrow \lambda$, 则对于任意非负整数 k , 有

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

即二项分布 $B(n, p_n)$ 在 $n \rightarrow \infty$ 且 $np_n \rightarrow \lambda$ 时收敛到泊松分布 $P(\lambda)$.

Proof.

$$\begin{aligned} \lim_{n \rightarrow \infty} P(X_n = k) &= \lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} \frac{(np_n)^k}{n^k} (1 - p_n)^n (1 - p_n)^{-k} \\ &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \lim_{n \rightarrow \infty} (1 - p_n)^n \cdot \lim_{n \rightarrow \infty} (1 - p_n)^{-k} \\ &= \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda} \cdot 1 \\ &= \frac{\lambda^k e^{-\lambda}}{k!} \end{aligned}$$

■

2.4.3 超几何分布

定义 2.4.7 超几何分布

设总体中有 N 个个体, 其中有 M 个是某一特征的个体, 从中不放回地随机抽取 n 个个体, 若随机变量 X 表示所抽取的个体中具有该特征的个体数, 则称 X 服从参数为 (N, M, n) 的超几何分布, 记作 $X \sim H(N, M, n)$, 其分布列为

$$p(k) = P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad \max(0, n - N + M) \leq k \leq \min(n, M)$$

2.4.4 几何分布

定义 2.4.8 几何分布

在成功概率为 p 的伯努利试验中, 得到一次成功所需要的试验次数 X 服从参数为 p 的几何分布, 记作 $X \sim \text{Ge}(p)$, 其分布列为

$$p(k) = P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

若 $X \sim \text{Ge}(p)$, 记 $q = 1 - p$, 则

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p \sum_{k=1}^{\infty} kq^{k-1} = p \frac{d}{dq} \left(\sum_{k=0}^{\infty} q^k \right) \\ &= p \frac{d}{dq} \left(\frac{1}{1-q} \right) = p \frac{1}{(1-q)^2} = \frac{1}{p} \end{aligned}$$

定理 2.4.2 几何分布的无记忆性

设随机变量 $X \sim \text{Ge}(p)$, 则对于任意正整数 m, n , 有

$$P(X > m+n \mid X > m) = P(X > n)$$

这个定理表明, 在前 m 次试验 A 没有出现的情况下, 在接下来的 n 次试验中 A 出现的概率与前 m 次试验是否出现 A 无关, 只与 n 有关, 因此称几何分布具有**无记忆性**.

2.4.5 负二项分布

定义 2.4.9 负二项分布

在伯努利试验中, 记事件 A 发生的概率是 p , 如果 X 为事件 A 第 r 次出现时的试验次数, 则 X 服从参数为 (r, p) 的**负二项分布**, 记作 $X \sim \text{Nb}(r, p)$, 其分布列为

$$p(k) = P(X = k) = \binom{k-1}{r-1} (1-p)^{k-r} p^r, \quad k = r, r+1, \dots$$

注解 2.4.4

当 $r = 1$ 时, 负二项分布即为几何分布.

2.5 常用连续分布

2.5.1 正态分布

正态分布源于高斯在 19 世纪初研究天文学和大地测量学时对测量误差的思考.

高斯设定了几个他认为“理所当然”的公理来描述测量误差的概率分布. 设 μ 是一个物理量的“真值”, 我们进行了 n 次独立测量 x_1, x_2, \dots, x_n . 每次测量的误差为 $\varepsilon_i = x_i - \mu$. 我们想要找到误差的概率密度函数 $\varphi(\varepsilon)$.

高斯提出 3 个公理:

- (1) **对称性与最大值**: 误差的概率密度函数 $\varphi(\varepsilon)$ 关于 $\varepsilon = 0$ 对称, 并且在 $\varepsilon = 0$ 处取得最大值.
- (2) **独立性**: 每次测量是相互独立的. 因此, 观测到这一组特定误差的联合概率 (似然) 是它们各自概率的乘积:

$$L(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = \prod_{i=1}^n \varphi(\varepsilon_i)$$

(3) **算数平均公理**: 高斯假设, 对于一组测量值, 其**算术平均值** $\bar{x} = \frac{1}{n} \sum x_i$ 是真值 μ 的“最可几估计值” (Maximum Likelihood Estimate) .

因此, 我们的目标是找到一个概率密度函数 $\varphi(\varepsilon)$, 使得对于任意 n 和任意误差组 $\varepsilon_1, \dots, \varepsilon_n$, 当算术平均值 $\bar{x} = \mu$ 时, 联合概率 $L(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ 取得最大值.

为了计算方便, 我们最大化 L 的对数 $\ln(L)$

$$\ln(L) = \sum_{i=1}^n \ln(\varphi(x_i - \mu))$$

要找到最大值, 我们对 μ 求导并令其为 0

$$\frac{d}{d\mu} \ln(L) = - \sum_{i=1}^n \frac{\varphi'(x_i - \mu)}{\varphi(x_i - \mu)} = 0 \Rightarrow \sum_{i=1}^n \frac{\varphi'(x_i - \mu)}{\varphi(x_i - \mu)} = 0$$

现在, 我们代入高斯的关键公理 (3): 这个等式在 $\mu = \bar{x}$ 时必须成立

$$\sum_{i=1}^n \frac{\varphi'(x_i - \bar{x})}{\varphi(x_i - \bar{x})} = 0$$

定义函数 $g(z) = \frac{\varphi'(z)}{\varphi(z)}$, 则上式变为

$$\sum_{i=1}^n g(x_i - \bar{x}) = 0 \quad (2.5.1)$$

我们还需要一个已知的事实: 算术平均值的定义. 对于任何一组 x_i , 其偏差之和恒为零:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \quad (2.5.2)$$

现在, 我们有两个等式 式 (2.5.1) 式 (2.5.2), 他们必须对任意一组测量值 x_i 成立. 要使这两个和式同时为零, 唯一的 (最简单的) 非平凡解是 $g(x_i - \bar{x})$ 必须与 $x_i - \bar{x}$ 成正比.

于是, 我们得到

$$g(z) = \frac{\varphi'(z)}{\varphi(z)} = kz$$

解这个微分方程得到

$$\ln(\varphi) = \frac{1}{2}kz^2 + C \Rightarrow \varphi(z) = A \exp\left(\frac{1}{2}kz^2\right)$$

现在, 由于公理 (1), $\varphi(z)$ 在 $z = 0$ 处取得最大值, 因此 k 必须是一个负数. 设 $k = -\frac{1}{\sigma^2}$, 则

$$\varphi(\varepsilon) = A \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$$

接下来就是进行归一化, 即

$$\int_{-\infty}^{\infty} A \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) d\varepsilon = 1 = A \cdot \sqrt{2\pi}\sigma$$

因此, $A = \frac{1}{\sqrt{2\pi}\sigma}$, 最终得到

$$\varphi(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$$

如果误差 $\varepsilon = x - \mu$, 那么测量值 x 的概率密度函数就是将这个函数平移 μ 个单位:

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

定义 2.5.1 正态分布

设随机变量 X 的概率密度函数为

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

则称 X 服从参数为 (μ, σ^2) 的正态分布, 记作 $X \sim N(\mu, \sigma^2)$.

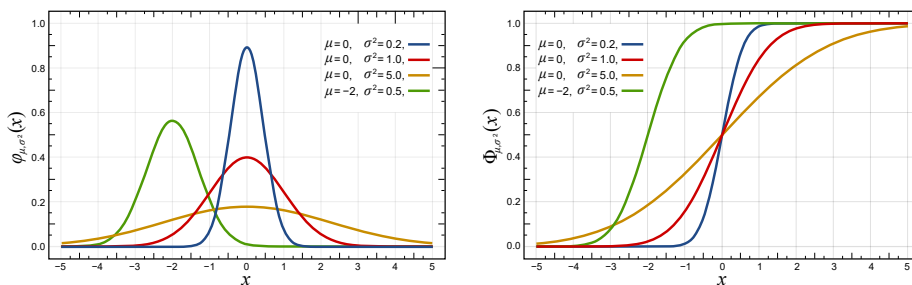


Fig. 2.1 PDF and CDF of Normal Distribution

正态分布的 PDF 是一个钟形曲线, 对称于 $x = \mu$, 在 $x = \mu$ 处取得最大值 $\frac{1}{\sqrt{2\pi}\sigma}$, 并且随着 $|x - \mu|$ 的增大而迅速趋近于 0. $\mu \pm \sigma$ 是图像的拐点.

正态分布 $N(\mu, \sigma^2)$ 的分布函数为

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

定义 2.5.2 标准正态分布

称 $\mu = 0, \sigma = 1$ 的正态分布为**标准正态分布**, 记作 $U \sim N(0, 1)$, 其 PDF 和 CDF 分别为

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u \exp\left(-\frac{t^2}{2}\right) dt$$

性质 2.5.1 常用的计算

- (1) $\Phi(-u) = 1 - \Phi(u)$
- (2) $P(a < U \leq b) = \Phi(b) - \Phi(a)$
- (3) $P(|U| \leq u) = 2\Phi(u) - 1$
- (4) $P(U > u) = 1 - \Phi(u)$

定理 2.5.1 标准化

设随机变量 $X \sim N(\mu, \sigma^2)$, 则随机变量 $U = \frac{X - \mu}{\sigma} \sim N(0, 1)$.

Proof. 考虑分布函数, 对于任意实数 u , 有

$$\begin{aligned} P(U \leq u) &= P\left(\frac{X - \mu}{\sigma} \leq u\right) \\ &= P(X \leq \sigma u + \mu) = F(\sigma u + \mu) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\sigma u + \mu} \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u \exp\left(-\frac{s^2}{2}\right) ds \\ &= \Phi(u) \end{aligned}$$

■

从而, 对于 $X \sim N(\mu, \sigma^2)$, 有

$$P(a \leq x \leq b) = P\left(\frac{a - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

若 $X \sim N(\mu, \sigma^2)$, 由于 $U = \frac{X - \mu}{\sigma} \sim N(0, 1)$, 因此

$$E(U) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \exp\left(-\frac{u^2}{2}\right) du = 0$$

从而, $E(X) = E(U)\sigma + \mu = 0 \cdot \sigma + \mu = \mu$.

同样的

$$\begin{aligned} \text{Var}(U) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 \exp\left(-\frac{u^2}{2}\right) du \\ &= \frac{1}{\sqrt{2\pi}} \cdot 2 \int_0^{\infty} u^2 \exp\left(-\frac{u^2}{2}\right) du \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} 2t \cdot e^{-t} \frac{1}{\sqrt{2t}} dt = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \sqrt{2t} \cdot e^{-t} dt \end{aligned}$$

$$= \sqrt{\frac{2}{\pi}} \cdot \sqrt{2} \cdot \Gamma\left(\frac{3}{2}\right) = \frac{2}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2} = 1$$

从而, $\text{Var}(X) = \text{Var}(U) \cdot \sigma^2 = 1 \cdot \sigma^2 = \sigma^2$.

定义 2.5.3 标准化

对于任意随机变量 X , 如果其数学期望和方差存在, 则称

$$X^* = \frac{X - E(X)}{\sqrt{\text{Var}(X)}}$$

为 X 的标准化随机变量, 且有

$$E(X^*) = 0, \text{Var}(X^*) = 1$$

定义 2.5.4 分位数

设随机变量 X 的分布函数为 $F(x)$, 对于给定的 $0 < p < 1$, 如果实数 x_p 满足

$$F(x_p) = P(X \leq x_p) = p$$

则称 x_p 为随机变量 X 的 p -分位数.

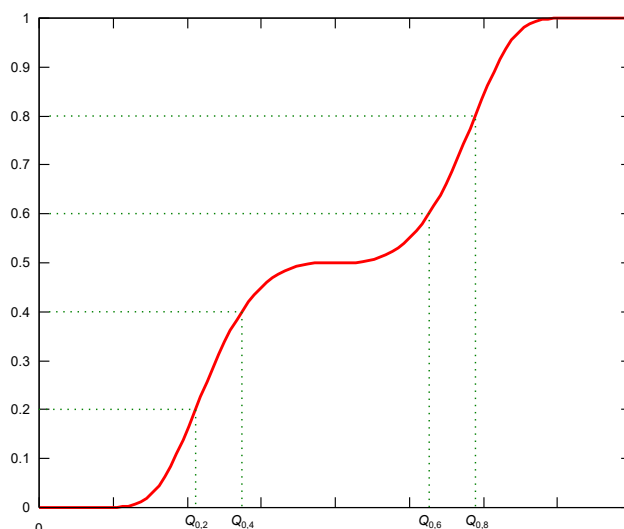


Fig. 2.2 Illustration of quantiles

注解 2.5.1

一般来说, 如果 $X \sim F(x)$ 是连续型随机变量

- (1) 给定 x 可以计算 $p = F(x)$, 也就是 CDF
- (2) 给定 p 可以计算 $x_p = F^{-1}(p)$, 也就是 p -分位数

这里的 F^{-1} 实际上是

$$F^{-1}(p) := \inf\{x \in \mathbb{R} \mid F(x) \geq p\}$$

$F^{-1}(p)$ 是所有满足“累积概率 $F(x)$ 至少达到 p 的 x 值中, 最小的那个 x .”

例 2.5.1 标准正态分布的 p -分位数

若随机变量 $X \sim N(0, 1)$, 则对于 $0 < p < 1$, $u_p := \Phi^{-1}(p)$ 即为 X 的 p -分位数.

问题 2.1 正态分布的 p -分位数

若 $X \sim N(\mu, \sigma^2)$, 则 X 的 p -分位数为多少?

Solution. 设 X 的 p -分位数为 x_p , 则

$$p = P(X \leq x_p) = \Phi\left(\frac{x_p - \mu}{\sigma}\right)$$

因此

$$x_p = \mu + \sigma \cdot u_p = \mu + \sigma \cdot \Phi^{-1}(p)$$

设随机变量 $X \sim N(\mu, \sigma^2)$, 则

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = P\left(\left|\frac{X - \mu}{\sigma}\right| \leq k\right) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$$

当 $k = 1, 2, 3$ 时, 分别有

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 2\Phi(1) - 1 \approx 0.6826$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 2\Phi(2) - 1 \approx 0.9545$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 2\Phi(3) - 1 \approx 0.9973$$

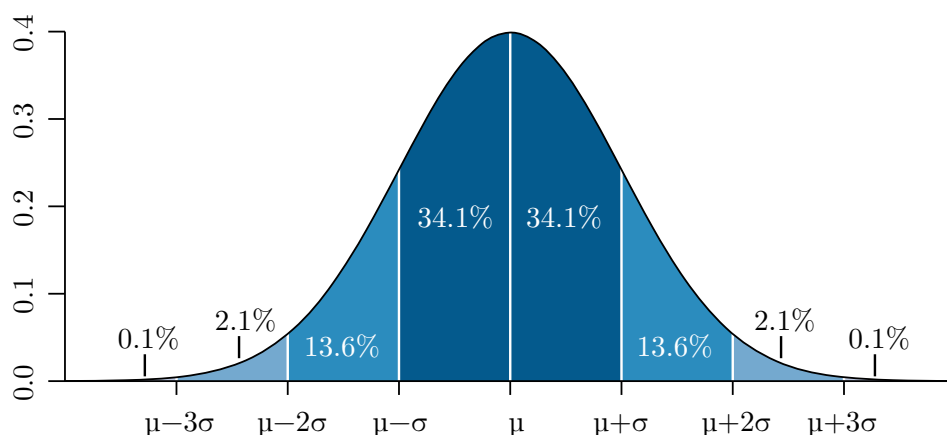


Fig. 2.3 Empirical Rule

2.5.2 均匀分布

定义 2.5.5 均匀分布

若随机变量 X 的概率密度函数为

$$p(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x < a \vee x > b \end{cases}$$

则称 X 服从参数为 (a, b) 的均匀分布, 记作 $X \sim U(a, b)$.

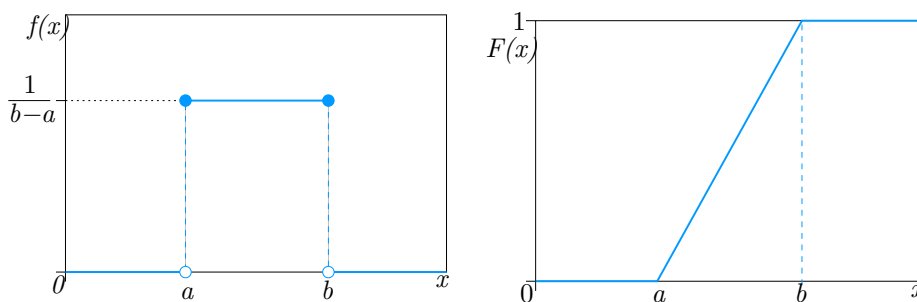


Fig. 2.4 PDF and CDF of Uniform Distribution

$U(a, b)$ 的 CDF 为

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

设 $X \sim U(a, b)$, 则

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

又有

$$E(X^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

因此

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{(b-a)^2}{12}$$

2.5.3 指数分布

定义 2.5.6 指数分布

设随机变量 X 的概率密度函数为

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

则称 X 服从参数为 λ 的指数分布, 记作 $X \sim \text{Exp}(\lambda)$.

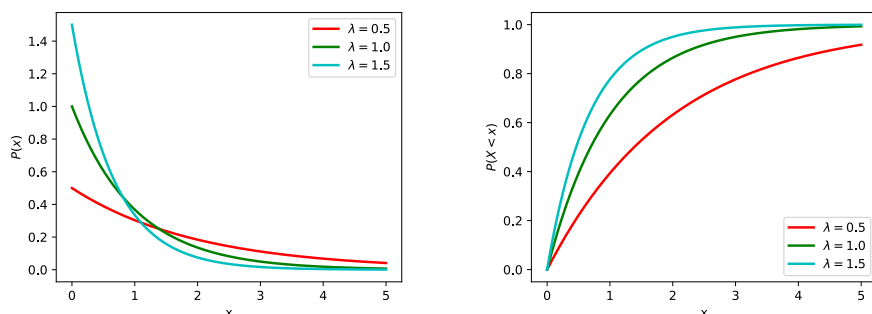


Fig. 2.5 PDF and CDF of Exponential Distribution

$\text{Exp}(\lambda)$ 的 CDF 为

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

设 $X \sim \text{Exp}(\lambda)$, 则

$$E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \int_0^{\infty} x d(-e^{-\lambda x}) = \frac{1}{\lambda}$$

考虑

$$E(X^2) = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \int_0^{\infty} x^2 d(-e^{-\lambda x}) = \frac{2}{\lambda^2}$$

从而

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

注解 2.5.2

对于指数分布 $\text{Exp}(\lambda)$ 有时记 $\theta := \frac{1}{\lambda} = E(X)$.

定理 2.5.2 指数分布的无记忆性

设随机变量 $X \sim \text{Exp}(\lambda)$, 则对于任意 $s, t \geq 0$, 有

$$P(X > s + t \mid X > s) = P(X > t)$$

Proof.

$$P(X > s + t \mid X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t)$$

■

2.5.4 伽马分布

定义 2.5.7 Γ 函数

设复数 z 的实部 $\operatorname{Re}(z) > 0$, 则称函数

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$$

为 Γ -函数.

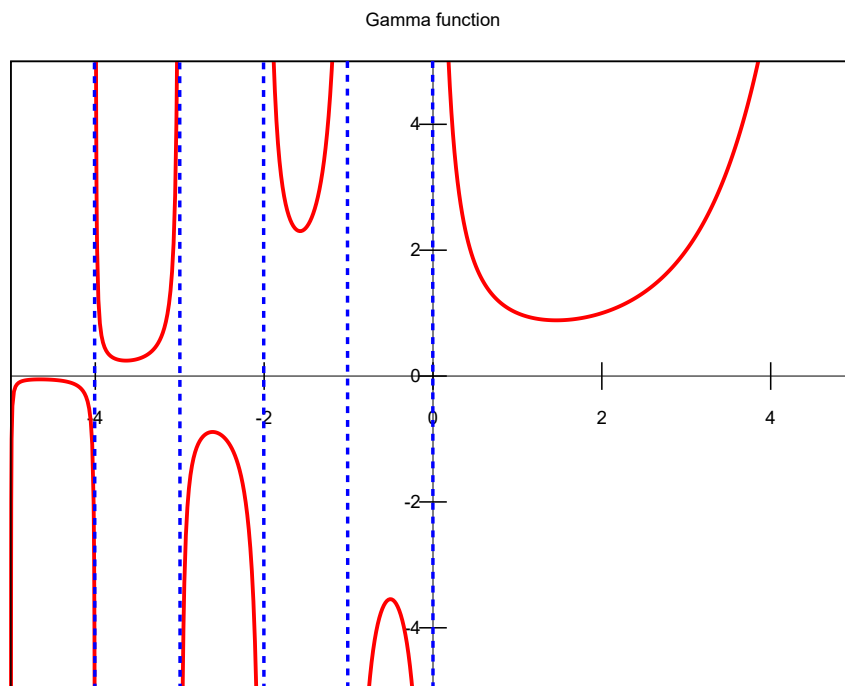


Fig. 2.6 The Γ function along part of the real axis

性质 2.5.2 Gamma 函数的性质

- $\Gamma(z) = (z-1)!$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- $\Gamma(z+1) = z\Gamma(z)$

定义 2.5.8 伽马分布

若随机变量 X 的概率密度函数为

$$p(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

则称 X 服从参数为 (α, λ) 的伽马分布, 记作 $X \sim \Gamma(\alpha, \lambda)$.

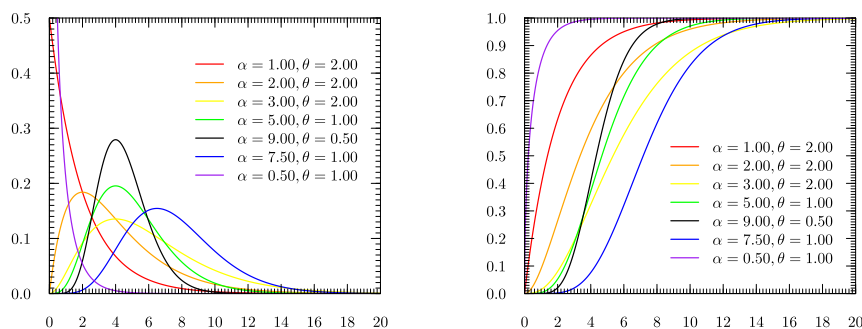


Fig. 2.7 PDF and CDF of Γ -Distribution

设 $X \sim \Gamma(\alpha, \lambda)$, 记 $\theta = \frac{1}{\lambda}$. 考虑 X 的 r 阶矩

$$\begin{aligned} E(X^r) &= \int_0^{\infty} x^r \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{r+\alpha-1} e^{-\lambda x} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{\lambda^{r+\alpha}} \int_0^{\infty} t^{r+\alpha-1} e^{-t} dt \\ &= \theta^r \cdot \frac{\Gamma(r+\alpha)}{\Gamma(\alpha)} \end{aligned}$$

从而,

$$E(X) = \alpha\theta, \quad \text{Var}(X) = \alpha\theta^2$$

例 2.5.2 指数分布

$\alpha = 1$ 时的伽马分布即为指数分布 $\text{Exp}(\lambda)$, 即 $\Gamma(1, \lambda) = \text{Exp}(\lambda)$.

定义 2.5.9 χ^2 分布

$\alpha = \frac{n}{2}, \lambda = \frac{1}{2}$ 时的伽马分布称为 χ^2 分布, 记作 $X \sim \chi^2(n)$, 即

$$\chi^2(n) \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

其概率密度函数为

$$p(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

这里的 n 称为 **自由度**, 可以是正实数, 但一般是正整数.

若 $X \sim \chi^2(n)$, 则

$$E(X) = n, \quad \text{Var}(X) = 2n$$

2.5.5 贝塔分布

定义 2.5.10 B 函数

设 $a, b \in \mathbb{C}$ 且 $\text{Re}(a), \text{Re}(b) > 0$, 则称函数

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$$

为 **B-函数**.

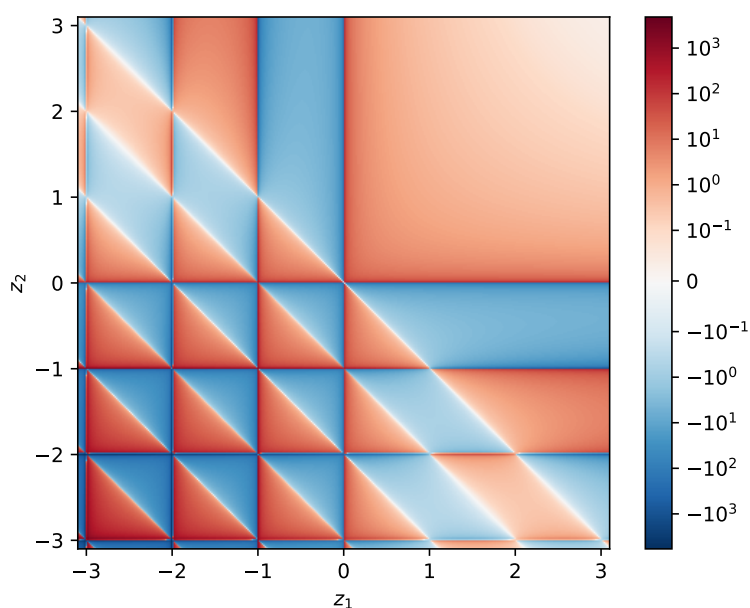


Fig. 2.8 Contour plot of B function

性质 2.5.3 Beta 函数的性质

- $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$
- $B(a, b) = B(b, a)$
- $B(a+1, b) = \frac{a}{a+b} B(a, b)$

定义 2.5.11 贝塔分布

若随机变量 X 的概率密度函数为

$$p(x) = \begin{cases} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

则称 X 服从参数为 (α, β) 的贝塔分布, 记作 $X \sim B(\alpha, \beta)$.

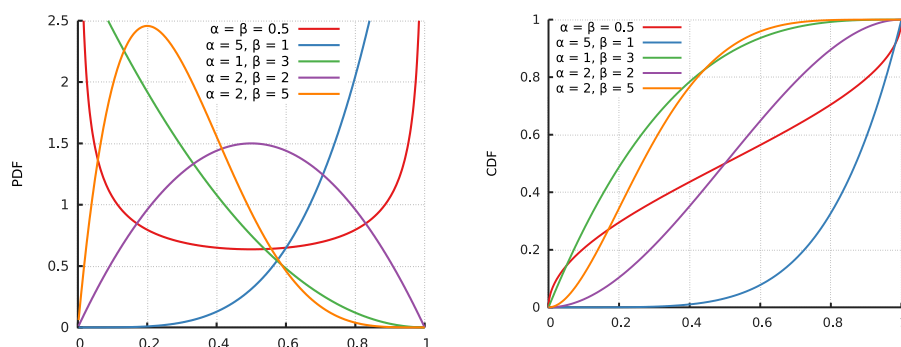


Fig. 2.9 PDF and CDF of B Distribution

设 $X \sim B(\alpha, \beta)$, 则

$$E(X) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^\alpha (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha + 1)\Gamma(\beta)}{\Gamma(\alpha + \beta + 1)} = \frac{\alpha}{\alpha + \beta}$$

又

$$\begin{aligned} E(X^2) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+1} (1-x)^{\beta-1} dx \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha + 2)\Gamma(\beta)}{\Gamma(\alpha + \beta + 2)} \\ &= \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} \end{aligned}$$

从而,

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

2.5.6 总结

我们将常见的概率分布和方差做一下总结

分布	PMF / PDF	$E(X)$	$\text{Var}(X)$
二点分布	$p_k = p^k(1-p)^{1-k}$ $k = 0, 1$	p	$p(1-p)$
二项分布 $b(n, p)$	$p_k = \binom{n}{k} p^k (1-p)^{n-k}$ $k = 0, 1, \dots, n$	np	$np(1-p)$
泊松分布 $P(\lambda)$	$p_k = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \dots$	λ	λ

分布	PMF / PDF	$E(X)$	$\text{Var}(X)$
超几何分布 $H(n, N, M)$	$p_k = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$ $k = 0, 1, \dots, r, r = \min\{M, n\}$	$\frac{nM}{N}$	$\frac{nM(N-M)(N-n)}{N^2(N-1)}$
几何分布 $\text{Ge}(p)$	$p_k = (1-p)^{k-1}p$ $k = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
负二项分布 $\text{Nb}(r, p)$	$p_k = \binom{k-1}{r-1} (1-p)^{k-r} p^r$ $k = r, r+1, \dots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
正态分布 $N(\mu, \sigma^2)$	$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $-\infty < x < \infty$	μ	σ^2
均匀分布 $U(a, b)$	$p(x) = \frac{1}{b-a}, a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布 $\text{Exp}(\lambda)$	$p(x) = \lambda e^{-\lambda x}, x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
伽马分布 $\Gamma(\lambda)$	$p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x \geq 0$	$\alpha \frac{1}{\lambda}$	$\alpha \frac{1}{\lambda^2}$
χ^2 分布 $\chi^2(n)$	$p(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1}$ $x \geq 0$	n	$2n$
贝塔分布 $B(\alpha, \beta)$	$p(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$ $0 \leq x \leq 1$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$
对数正态分布 $\text{LN}(\mu, \sigma^2)$	$p(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$ $x > 0$	$e^{\mu + \frac{1}{2}\sigma^2}$	$(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$
柯西分布 $\text{Cau}(\mu, \lambda)$	$p(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x - \mu)^2}$ $-\infty < x < \infty$	DNE	DNE

分布	PMF / PDF	$E(X)$	$\text{Var}(X)$
韦布尔分布 $\text{Wb}(\eta, m)$	$p(x) = F'(x)$ $F(x) = 1 - \exp\left\{-\left(\frac{x}{\eta}\right)^m\right\}$	$\eta\Gamma\left(1 + \frac{1}{m}\right)$	$\eta^2\left[\Gamma\left(1 + \frac{2}{m}\right) - \Gamma^2\left(1 + \frac{1}{m}\right)\right]$

2.6 随机变量函数的分布

已知随机变量 X 的分布, 求随机变量函数 $Y = g(X)$ 的分布是概率论中的一个重要问题.

2.6.1 离散随机变量函数的分布

设离散型随机变量 X 的取值为 x_1, x_2, \dots , 且对应的概率为 p_1, p_2, \dots . 设 $Y = g(X)$, 则 Y 的取值为 $y_i = g(x_i)$, 且对应的概率为 p_i . 如果存在 $i \neq j$ 使得 $y_i = y_j$, 则将这些 y_i 合并, 其对应的概率为这些 p_i 之和.

2.6.2 连续随机变量函数的分布

离散随机变量的函数一定是离散型随机变量, 但是连续型随机变量的函数不一定是连续的.

当 $Y = g(X)$ 为离散型随机变量时, 可以用离散随机变量函数的分布方法求解.

当 $Y = g(X)$ 为严格单调函数时, 可以用以下定理求解.

定理 2.6.1 连续随机变量函数的分布

设连续型随机变量 X 的概率密度函数为 $p_X(x)$, $Y = g(X)$. 若 $y = g(x)$ 严格单调, 其反函数 $h(y)$ 有连续导函数, 则 $Y = g(X)$ 的密度函数为

$$p_Y(y) = \begin{cases} p_X[h(y)]|h'(y)|, & a < y < b \\ 0, & \text{otherwise} \end{cases}$$

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

定理 2.6.2 正态变量的线性变换

设 $X \sim N(\mu, \sigma^2)$, 若 $a \neq 0$, 则 $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$.

定义 2.6.1 对数正态分布

若随机变量 X 的对数 $\ln(X)$ 服从参数为 (μ, σ^2) 的正态分布, 则称 X 服从参数为 (μ, σ^2) 的对数正态分布, 记作 $X \sim \text{LN}(\mu, \sigma^2)$. 其概率密度函数为

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

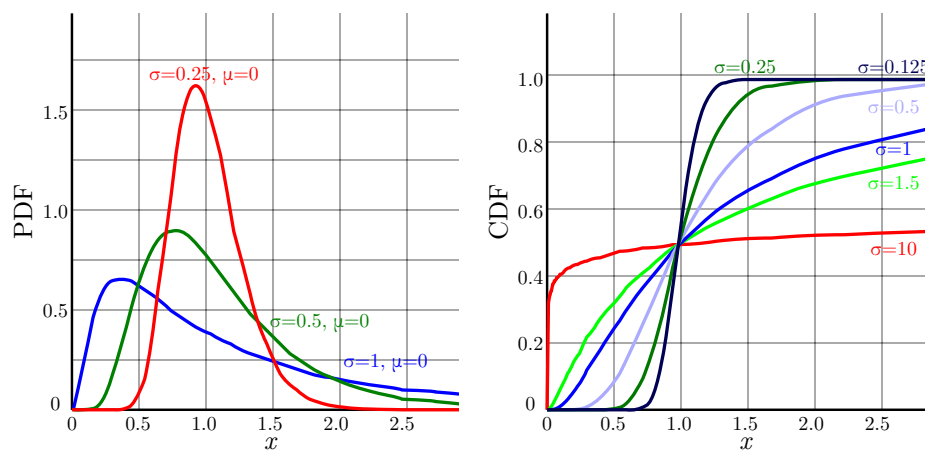


Fig. 2.10 PDF and CDF of Log-Normal Distribution

I. Definitions

定义 1.1.1	样本空间	3
定义 1.1.2	样本点	3
定义 1.1.3	离散样本空间	3
定义 1.1.4	连续样本空间	3
定义 1.1.5	随机事件	3
定义 1.1.6	可测函数	4
定义 1.1.7	随机变量	4
定义 1.1.8	事件的并	4
定义 1.1.9	事件的交	4
定义 1.1.10	事件的差	4
定义 1.1.11	对立事件	4
定义 1.1.12	σ -代数	5
定义 1.1.13	样本空间的分割	5
定义 1.2.1	测度空间	5
定义 1.2.2	概率空间	5
定义 1.3.1	极限事件	6
定义 1.3.2	连续性	7
定义 1.4.1	条件概率	7
定义 1.5.1	两个事件的独立性	9
定义 1.5.2	多个事件的独立性	9
定义 1.5.3	试验的独立性	9
定义 2.1.1	离散型随机变量	10
定义 2.1.2	连续型随机变量	10
定义 2.1.3	分布函数	10
定义 2.1.4	分布列	11
定义 2.1.5	概率密度函数	11
定义 2.2.1	离散型随机变量的数学期望	12
定义 2.2.2	连续型随机变量的数学期望	12
定义 2.2.3	数学期望	12
定义 2.3.1	方差	13
定义 2.3.2	标准差	13
定义 2.4.1	二项分布	15
定义 2.4.2	二点分布	15
定义 2.4.3	原点矩	16
定义 2.4.4	中心矩	16
定义 2.4.5	阶乘矩	16
定义 2.4.6	泊松分布	17
定义 2.4.7	超几何分布	18
定义 2.4.8	几何分布	18
定义 2.4.9	负二项分布	19
定义 2.5.1	正态分布	21
定义 2.5.2	标准正态分布	21
定义 2.5.3	标准化	23
定义 2.5.4	分位数	23
定义 2.5.5	均匀分布	24

定义 2.5.6	指数分布	25
定义 2.5.7	Γ 函数	27
定义 2.5.8	伽马分布	27
定义 2.5.9	χ^2 分布	28
定义 2.5.10	B 函数	29
定义 2.5.11	贝塔分布	29
定义 2.6.1	对数正态分布	32

II. Theorems

定理 1.4.1	全概率公式	8
定理 1.4.2	Bayes 公式	8
定理 2.2.1	随机变量函数的期望	12
定理 2.3.1	Markov 不等式	14
定理 2.3.2	Chebyshev 不等式	14
定理 2.3.3	单点分布	15
定理 2.4.1	泊松定理	17
定理 2.4.2	几何分布的无记忆性	19
定理 2.5.1	标准化	22
定理 2.5.2	指数分布的无记忆性	26
定理 2.6.1	连续随机变量函数的分布	32
定理 2.6.2	正态变量的线性变换	32

III. Figures

Fig. 1.1	Total Probability Formula	8
Fig. 2.1	PDF and CDF of Normal Distribution	21
Fig. 2.2	Illustration of quantiles	23
Fig. 2.3	Empirical Rule	24
Fig. 2.4	PDF and CDF of Uniform Distribution	25
Fig. 2.5	PDF and CDF of Exponential Distribution	26
Fig. 2.6	The Γ function along part of the real axis	27
Fig. 2.7	PDF and CDF of Γ -Distribution	28
Fig. 2.8	Contour plot of B function	29
Fig. 2.9	PDF and CDF of B Distribution	30
Fig. 2.10	PDF and CDF of Log-Normal Distribution	33