目录

1.1 随机事件的概率 3
1.1.1 随机事件 3
1.1.2 随机事件的概率 4
1.1.3 条件概率 6
1.1.4 事件的独立性 8
1.2 一维随机变量及其分布 9
1.2.1 常用离散型分布 10
1.2.2 随机变量的分布函数 12
1.2.3 连续型随机变量及其概率密度函数 12
1.2.4 随机变量函数的分布 14
1.3 多维随机变量及其分布 15
1.3.1 随机向量的联合分布 15
1.3.2 随机向量的边缘分布 17
1.3.3 随机向量的条件分布 17
1.3.4 随机变量的独立性 18
1.3.5 随机向量函数的分布 19
1.3.6 常用的随机向量 20

1.4 随机变量的数字特征 21
1.4.1 数学期望 21
1.4.2 方差与标准差 23
1.4.3 协方差与相关系数 24
1.4.4 矩协方差矩阵 25
1.4.5 n 维正态分布 26
1.5 大数定律和中心极限定理 26
1.5.1 大数定律 26
1.5.2 中心极限定理 27
1.6 样本及抽样分布 28
1.7 参数估计 33
1.7.1 参数估计 33
1.7.2 估计量的评选标准 35

复习

1.1 随机事件的概率

随机事件

主要内容：随机事件的关系与运算, 概率的加法公式, 古典概率模型, 条件概率的定义, 乘法公式, 全概率公式, 贝叶斯公式, 随机事件的独立性.

1.1.1 随机事件

随机事件的关系

关系	记号	概率论含义
包含	$A \subset B$	A 发生则 B 一定发生
相等	$A = B$	A 与 B 必定同时发生
互斥	$A \cap B = \varnothing$	A 与 B 不会同时发生
对立	$A = \overline{B}$	A 与 B 有且仅有一个发生

随机事件的运算

运算	记号	概率论含义
并	$A \cup B$	A 与 B 至少一个发生
积	AB	A 与 B 都发生
差	$A - B$	A 发生但 B 不发生
补	$\overline{A}$	A 不发生

1.1.2 随机事件的概率

古典概率模型

定义：如果一个随机试验具有以下特点：

1. 样本空间只含有限多个样本点；
2. 各样本点出现的可能性相等，

则称此随机试验是古典型的. 此时对每个事件 $A \subset \Omega$, 其概率

$$P(A) = \frac{\text{事件 } A \text{ 的样本点数}}{\text{样本点总数}} = \frac{n(A)}{n(\Omega)}$$

称为事件 A 的古典概率.

几何概型

古典概型是关于试验的结果为有限个，且每个结果出现的可能性相同的概率模型. 一个直接的推广是：保留等可能性，而允许试验具有无限多个结果的.

定义： 如果每个事件发生的概率只与构成该事件区域的长度（面积或体积或度数）成比例，则称这样的概率模型为几何概率模型.

几何概型中，事件 A 的概率：

$$P(A) = \frac{S(A)}{S(\Omega)}.$$

概率的公理化定义

定义 **1.** 设 Ω 是样本空间, 对每个事件 A 定义一个实数 $P(A)$ 与之对应. 若函数 $P(\cdot)$ 满足以下条件:

1. 非负性: 对任意事件 A, 均有 $P(A) \geq 0$;
2. 规范性: $P(\Omega) = 1$;
3. 可列加性: 若事件序列 $\{A_n\}_{n\geq 1}$ 两两互斥, 则

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P_n(A_i)$$

称 $P(A)$ 为事件 A 的概率 *(probability)*.

概率的可加性

概率可加性的常用公式:

1. $P(\varnothing) = 0$.
2. 若 n 个事件 $A_1, A_2, \cdots, A_n$ 两两互斥, 则

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

   特别地, 若两个事件 A, B 互斥, 则

$$P(A \cup B) = P(A) + P(B).$$

概率的可加性

概率可加性的常用公式:

3. 对任意事件 A, 有
$$P(A) = 1 - P\left(\overline{A}\right).$$

4. 若事件 $A \subset B$, 则
$$P(B-A) = P(B) - P(A).$$
特别地, $A \subset B \Longrightarrow P(A) \leq P(B)$.

5. 对任意两个事件 A, B, 有
   - $P(A \cup B) = P(A) + P(B) - P(AB)$.
   - $P(A\overline{B}) = P(A) - P(AB)$.

1.1.3 条件概率

条件概率

定义：设 $P(B) > 0$, 称
$$P(A|B) := \frac{P(AB)}{P(B)}$$

为在事件 B 发生条件下, 事件 A 的条件概率. 在古典概率模型中,
$$P(A|B) = \frac{\text{事件 } AB \text{ 包含的样本点数}}{\text{事件 } B \text{ 包含的样本点数}} = \frac{n(AB)}{n(B)}.$$

乘法公式

由条件概率的定义, 如果 $P(B) > 0$, 则
$$P(AB) = P(B)P(A|B)$$
类似地, 如果 $P(A) > 0$, 则有
$$P(AB) = P(A)P(B|A)$$
以上等式称为乘法公式.

全概率公式

定义：设 Ω 为某试验的样本空间, $B_1, B_2, \cdots$ 为一组事件. 如果以下条件成立：

1. $B_1, B_2, \cdots$ 两两互斥；
2. $\cup_i B_i = \Omega$,

则称 $B_1, B_2, \cdots$ 为样本空间 Ω 的一个划分（分割）, 或称 $B_1, B_2, \cdots$ 为一个完备事件组. 对任意满足 $0 < P(B) < 1$ 的事件 B, B 与 $\overline{B}$ 构成一个完备事件组.

全概率公式

全概率公式：如果 $B_1, B_2, \cdots$ 构成一个完备事件组, 且都有正概率, 则对任意事件 A 有

$$\boxed{P(A) = \sum_i P(B_i) P(A|B_i).}$$

特殊情况：如果事件 B 满足 $0 < P(B) < 1$, 则对事件 A, 有

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

贝叶斯公式

贝叶斯定理：如果 $B_1, B_2, \cdots$ 构成一个完备事件组, 且都有正概率, 则对任意正概率的事件 A 有

$$\boxed{P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)}, \quad i = 1, 2, \cdots}$$

1.1.4 事件的独立性

两个事件的独立性

定义：若两事件 A、B 满足

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

则称事件 A、B 相互独立. 实际意义：若 $P(B) > 0$, 则上式等价于

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

即事件 A 的概率不受事件 B 发生与否的影响.

两个事件的独立性

性质：若事件 A 与 B 相互独立, 则

$$\overline{A} \text{ 与 } B、A \text{ 与 } \overline{B}、\overline{A} \text{ 与 } \overline{B}$$

也是相互独立的.

小注: 若 A 与 B 相互独立, 且 B 与 C 相互独立, 则 A 与 C 未必相互独立.

多个事件的独立性

定义. 称 $n(n \geq 2)$ 个事件 $A_1, A_2, \ldots, A_n$ 相互独立, 如果对任意一组指标

$$1 \leq i_1 < i_2 < \cdots < i_k \leq n \quad (k \geq 2)$$

都有

$$P\left(A_{i_1} A_{i_2} \cdots A_{i_k}\right) = P\left(A_{i_1}\right) P\left(A_{i_2}\right) \cdots P\left(A_{i_k}\right)$$

性质. 设 $n(n \geq 2)$ 个事件 $A_1, A_2, \ldots, A_n$ 相互独立, 则

1. 其中任意 $k(k \geq 2)$ 个事件也是相互独立的.
2. 将若干个 A_i 用 $\overline{A}_i$ 替换后, 得到的新事件集也相互独立.
3. 特别地, 我们有

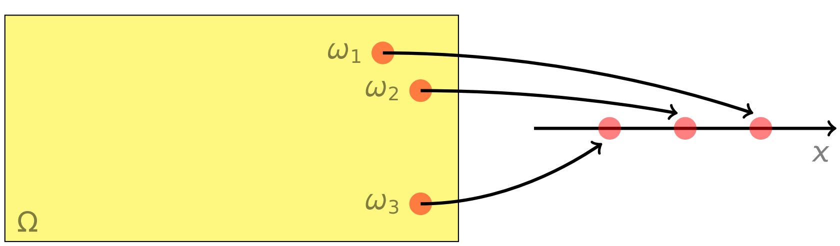
$$
\begin{aligned}
P(A_1 \cup \cdots \cup A_n) &= 1 - P\left(\overline{A}_1\right) \cdots P\left(\overline{A}_n\right) \\
&= 1 - \prod_{1}^{n} [1 - P(A_i)]
\end{aligned}
$$

1.2 一维随机变量及其分布

随机变量

定义 **1.** 设 $X = X(\omega)$ 是定义在样本空间 Ω 上的实值函数, 称 $X = X(\omega)$ 为随机变量.

下图给出样本点 ω 与实数 $X = X(\omega)$ 对应的示意图.



随机变量一般用大写英文字母 X、Y、Z 或小写希腊字母 ξ、η、γ 来表示.

一般地, 若 I 是一个实数集, $\{X \in I\}$ 记为事件 B, 即

$$B = \{\omega | X(\omega) \in I\},$$

则

$$P\{X \in I\} = P(B) = P\{\omega | X(\omega) \in I\}.$$

按照随机变量可能取值的情况，可以把它们分为两类：离散型随机变量和非离散型随机变量，而非离散型随机变量中最重要的是连续型随机变量.

离散型随机变量

定义 2. 如果随机变量的全部可能取的值只有有限个或可列无限多个，则称这种随机变量为离散型随机变量.

一般地，设离散型随机变量 X 所有可能取的值为 x_k $(k = 1, 2, \ldots)$，X 取各个可能值的概率，即事件 $\{X = x_k\}$ 的概率为

$$P\{X = x_k\} = p_k,\ k = 1, 2, \ldots, \tag{1}$$

称 (1) 式为离散型随机变量 X 的分布律或概率分布.

分布律也可以用下面的表格来表示：

X	x_1	$x_2,$	$\cdots$	x_n	$\cdots$
p_k	p_1	$p_2,$	$\cdots$	p_n	$\cdots$

由概率的定义，式 p_k 应满足以下条件：

1. $p_k \geq 0, k = 1, 2 \ldots;$
2. $\sum\limits_{k=1}^{\infty} p_k = 1.$

1.2.1 常用离散型分布

离散型・两点分布

定义：若随机变量 X 只能取 0 或 1, 其概率分布为：

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

则称 X 服从参数为 p 的两点分布, 记为

$$X\sim b(1,p).$$

离散型・二项分布

定义：如果随机变量 X 服从以下分布律

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k},$$

其中 $0<p<1, 0\le k\le n$, 则称 X 服从参数为 n,p 的二项分布. 记为

$$X\sim b(n,p).$$

离散型・泊松分布

定义：如果随机变量 X 服从以下分布律

$$P\{X=m\}=\frac{\lambda^m}{m!}e^{-\lambda}, \quad m=0,1,\cdots$$

其中 $\lambda>0$, 则称 X 服从参数为 λ 的泊松分布, 记为

$$X\sim\pi(\lambda).$$

二项分布的泊松近似

定理 (泊松定理). 在 n 重伯努利试验中, 事件 A 在每次试验中发生概率为 p_n （注意这与实验的次数 n 有关）, 如果 $n\to\infty$ 时, $np_n\to\lambda$ (λ 为常数), 则对任意给定的非负整数 k, 有

$$\lim_{n\to\infty}C_n^k p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}.$$

1.2.2 随机变量的分布函数

随机变量的分布函数

定义：对任何随机变量 X, 称函数

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

为 X 的分布函数. 设 $F(x)$ 为某随机变量的分布函数, 则其有以下性质：

1. 广义单增：对任意实数 $a < b$, 总有 $F(a) \le F(b)$；
2. $0 \le F(x) \le 1$, 且 $F(-\infty) = 0, \quad F(+\infty) = 1.$

离散型随机变量 X 的概率分布 $p_k = P\{X = x_k\}$ 满足

1. $p_k \ge 0, \quad k = 1, 2, \cdots$
2. $\sum\limits_k p_k = 1$
3. $F(x) = \sum\limits_{x_k \le x} p_k$

1.2.3 连续型随机变量及其概率密度函数

连续型随机变量

设 X 为连续性随机变量, 则对每个实数 a, 总有

$$P\{X = a\} = 0.$$

任意区间上概率的计算：由概率密度函数的定义可知,

$$P\{X \in (a, b]\} = F(b) - F(a) = \int_a^b f(x)dx.$$

上式中的区间 $(a, b]$ 改为 $(a, b), [a, b)$ 或 $[a, b]$ 后等式仍成立.

连续型随机变量 X 的概率密度 $f(x)$ 满足

1. $f(x) \geq 0, \quad x \in \mathbb{R}$
2. $\displaystyle\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$
3. $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t$

连续型・均匀分布

定义：若随机变量 X 有概率密度

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}, \qquad (a < b)$$

则称 X 服从区间 $[a,b]$ 上的均匀分布, 记为

$$X \sim U[a,b].$$

连续型・指数分布

定义：如果随机变量 X 有以下概率密度

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases},$$

其中 $\lambda > 0$, 则称 X 服从参数为 λ 的指数分布. 其分布函数为

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

连续型・正态分布

定义：如果随机变量 X 有以下概率密度

$$\varphi_{\mu,\sigma^2}(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

其中 μ,σ 为常数且 $\sigma>0$, 则称 X 服从正态分布. 简记为

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

称 $N(0,1)$ 为标准正态分布, 并简写 $\varphi_{0,1}(x)$ 为 $\varphi(x)$.

正态分布的分布函数为

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

该函数不是初等函数. 标准正态分布的分布函数简记为 $\Phi(x)$.

标准正态分布的分布函数的性质

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

服从正态分布随机变量的标准化：若 $X\sim N(\mu,\sigma^2)$, 则

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

1.2.4 随机变量函数的分布

离散型随机变量函数的分布

设离散型随机变量 X 的概率分布为

$$P\{X=x_k\}=p_k,\quad k=1,2,\cdots$$

令 $Y=g(X)$, 则 Y 也是一个离散型随机变量, 其分布可按如下步骤求得

1. 根据函数关系列出 Y 的所有可能值；
2. 对 Y 的每个可能值 y, $P\{Y=y\}$ 等于所有满足 $g(x_k)=y$ 的 p_k 之和.

连续型随机变量函数的分布

对连续型随机变量 X, 求 $Y = g(X)$ 的密度函数的基本方法是

1. 根据函数关系先求 Y 的分布函数
$$F_Y(y) = P\{Y \le y\} = P\{g(X) \le y\}$$
2. 然后对 $F_Y(y)$ 求导可得 Y 的概率密度.

定理 **1.** 设随机变量 X 具有概率密度 $f_X(x), -\infty < x < +\infty$; 函数 $g(x)$ 处处可导且恒有 $g'(x) > 0$(或恒有 $g'(x) < 0$) 则 $Y = g(X)$ 是连续型随机变量, 其概率密度为

$$f_Y(y) = \begin{cases} f_X[h(y)] \cdot |h'(y)|, & y \in (\alpha, \beta) \\ 0, & otherwise. \end{cases}$$

其中 $h(y)$ 是 $g(x)$ 的反函数,

$$\alpha = \min(g(-\infty), g(+\infty)),\ \beta = \max(g(-\infty), g(+\infty)).$$

设随机变量 $X \sim N(\mu, \sigma^2)$, $Y = aX + b$, 则 $Y \sim N\left(a\mu + b, (a\sigma)^2\right)$.

1.3 多维随机变量及其分布

1.3.1 随机向量的联合分布

二维随机向量的联合分布

定义：设 (X, Y) 为二维随机向量, 称二元函数

$$F(x, y) = P\{X \le x, Y \le y\}$$

为 (X, Y) 的联合分布函数. 联合分布函数的性质：

1. $F(x,y)$ 对每个自变量都是广义单增的；
2. $0 \le F(x,y) \le 1$；
3. $F(x,-\infty) = F(-\infty,y) = 0$, $F(+\infty,+\infty) = 1$；

二维离散型随机向量 (X,Y) 的联合概率分布 $p_{ij} = P\{X = x_i, Y = y_j\}$ 满足

1. $p_{ij} \ge 0, \quad i = 1,2,\cdots, j = 1,2,\cdots$
2. $\sum_i \sum_j p_{ij} = 1$；

二维连续型随机向量 (X,Y) 的联合概率密度 $f(x,y)$ 满足

1. $f(x,y) \ge 0, \quad \forall x,y \in \mathbb{R}$；
2. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$；

二维连续型随机向量

二维连续型随机向量 (X,Y) 的联合分布函数

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\,\mathrm{d}t\,\mathrm{d}s,$$

若联合概率密度 $f(x,y)$ 在点 (x,y) 处连续, 则有

$$\frac{\partial F(x,y)}{\partial x \partial y} = f(x,y)$$

对任意的平面区域 D, 有

$$P\{(X,Y) \in D\} = \iint_{(x,y)\in D} f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

1.3.2　随机向量的边缘分布

边缘分布

二维随机向量 (X,Y) 作为一个整体, 有联合分布函数 $F(x,y)$, 其分量 X 与 Y 都是随机变量, 有各自的分布函数, 分别记成 $F_X(x)$ 和 $F_Y(y)$, 称为 X 和 Y 的边缘分布函数. 边缘分布由联合分布完全确定:

$$F_X(x)=F(x,+\infty),\qquad F_Y(y)=F(+\infty,y).$$

二维离散型随机向量 (X,Y) 的边缘概率分布为

$$p_{i\cdot}=\sum_j p_{ij},\qquad i=1,2,\cdots$$

$$p_{\cdot j}=\sum_i p_{ij},\qquad j=1,2,\cdots$$

二维连续型随机向量 (X,Y) 的边缘概率密度为

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}y,$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}x.$$

1.3.3　随机向量的条件分布

二维离散型随机向量的条件分布

当 $p_{\cdot j} > 0$ 时, $Y = y_j$ 时 X 的条件概率分布为

$$P\{X = x_i | Y = y_j\} = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \cdots$$

当 $p_{i\cdot} > 0$ 时, $X = x_i$ 时 Y 的条件概率分布为

$$P\{Y = y_j | X = x_i\} = \frac{p_{ij}}{p_{i\cdot}}, \quad j = 1, 2, \cdots$$

若 $f_Y(y) > 0$, 在 $Y = y$ 条件下, X 的条件概率密度为

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, \quad x \in \mathbb{R}.$$

若 $f_X(x) > 0$, 在 $X = x$ 条件下, Y 的条件概率密度为

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, \quad x \in \mathbb{R}.$$

二维连续型随机向量的条件分布

定义. 若 $f_Y(y) > 0$, 则称

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(s|y)\,\mathrm{d}s.$$

为 $Y = y$ 条件下, X 的条件分布函数.

定义. $f_X(x) > 0$, 则称

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} f_{Y|X}(t|x)\,\mathrm{d}t.$$

为 $X = x$ 条件下 Y 的条件分布函数.

1.3.4 随机变量的独立性

二维随机向量的独立性

二维离散型随机向量 (X,Y) 相互独立的充要条件为

$$p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$$

对所有的 i, j 都成立.

二维连续型随机向量 (X,Y) 相互独立的充要条件为

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

对几乎所有的实数 x, y 成立.

1.3.5 随机向量函数的分布

二维离散型随机向量函数的分布

设离散型随机变量 (X,Y) 的联合分布为

$$P\{X = x_i, Y = y_j\} = p_{ij}, \quad i,j = 1,2,\cdots$$

令 $Z = g(X,Y)$, 则 Z 也是一个离散型随机变量, 其分布可按如下步骤求得

1. 根据函数关系列出 Z 的所有可能值;
2. 对 Z 的每个可能值 z, $P\{Z = z\}$ 等于所有满足 $g(x_i, y_j) = z$ 的 p_{ij} 之和.

二维连续型随机向量函数的分布

对连续型随机变量 (X,Y), 求 $Z = g(X,Y)$ 的密度函数的基本方法是

1. 根据函数关系先求 Z 的分布函数
$$F_Z(z) = P\{Z \leq z\} = P\{g(X,Y) \leq z\}$$
2. 然后对 $F_Z(z)$ 求导可得 Z 的概率密度.

1.3.6 常用的随机向量

连续型・均匀分布

定义：设 D 是平面上的有界区域, 其面积为 d, 若二维随机向量 (X,Y) 的概率密度为

$$f(x,y)=\begin{cases}\frac{1}{d} & (x,y)\in D\\ 0 & \text{else}\end{cases}$$

则称 (X,Y) 服从 D 上的均匀分布.

连续型・均匀分布

若 (X,Y) 服从 D 上的均匀分布, 则 (X,Y) 落在某一区域 A 内的概率

$$\begin{aligned}P\{(X,Y)\in A\}&=\iint_A f(x,y)dxdy\\&=\iint_{A\cap D}\frac{1}{d}dxdy\\&=\frac{S}{d}\end{aligned}$$

其中 S 为 $A\cap D$ 的面积.

连续型・正态分布

定理：若 $X_1,X_2,\cdots,X_n$ 相互独立, 且

$$X_i\sim N(\mu_i,\sigma_i^2),\ i=1,2,\cdots,n$$

则对不全为零的数 $a_1,a_2,\cdots,a_n$,

$$a_1X_1+\cdots+a_nX_n\sim N\left(\sum_{i=1}^n a_i\mu_i,\sum_{i=1}^n a_i^2\sigma_i^2\right).$$

定理：若 $X\sim N(\mu,\sigma^2)$, 则 $aX+b\sim N(a\mu+b,(a\sigma)^2)$.

1.4 随机变量的数字特征

数字特征

主要内容：期望的定义：离散型、连续型, 随机变量的函数的期望, 期望、方差、协方差的性质, 相关系数, 常见分布的数字特征, 大数定律

1.4.1 数学期望

离散型随机变量的期望

定义：设离散型随机变量 X 的分布律为

$$P\{X = x_k\} = p_k, \qquad k = 1, 2, \cdots.$$

若级数

$$\sum_k x_k p_k$$

绝对收敛, 则称其和为随机变量 X 的数学期望, 记为 $E(X)$.

连续型随机变量的期望

定义：设连续型随机变量 X 的概率密度为 $f(x)$, 若积分

$$\int_{-\infty}^{+\infty} x f(x) dx$$

绝对收敛, 则称此积分为随机变量 X 的数学期望, 记为 $E(X)$.

随机变量函数的数学期望

定理：设 X 为随机变量, $Y = g(X)$, 则

1. 若 X 为离散型随机变量, 分布律为

$$P\{X = x_k\} = p_k, \qquad k = 1, 2, \cdots.$$

则

$$E(Y) = \sum_k g(x_k) p_k.$$

随机变量函数的数学期望

定理（续）：

2. 若 X 为连续型随机变量, 概率密度为 $f(x)$, 则

$$E(Y) = \int_{-\infty}^{+\infty} g(x) f(x) dx.$$

数学期望的性质

设 $X, X_1, X_2, \cdots, X_n$ 为随机变量, c, k 为常数, 则有

1. $E(c) = c$；
2. $E(kX) = kE(X)$；
3. $E(X_1 + X_2) = E(X_1) + E(X_2)$；推论：$E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} E(X_k)$

4. 若 X_1, X_2 相互独立, 则有

$$E(X_1 X_2) = E(X_1) E(X_2).$$

推论：若 $X_1, X_2, \cdots, X_n$ 相互独立, 则有

$$E\left(\prod_{k=1}^{n} X_k\right) = \prod_{k=1}^{n} E(X_k)$$

注：如果没有相互独立这一条件, 上式一般不成立！

1.4.2 方差与标准差

方差

定义：设 X 是一随机变量, 若 $X-E(X)$ 平方的期望存在, 则称该期望为 X 的方差, 记为 $D(X)$（或 $\mathrm{Var}(X)$）, 即

$$D(X) := E[(X-E(X))^2].$$

称 $\sqrt{D(X)}$ 为 X 的标准差. 方差的常用计算公式：

$$D(X) = E(X^2) - [EX]^2.$$

方差的性质

设 $X, X_1, X_2, \cdots, X_n$ 为随机变量, c, k 为常数, 则有

1. $D(c) = 0$, $D(X+c) = D(X)$；
2. $D(X) \geq 0$, 且等式成立当且仅当 X 几乎必然为常数；
3. $D(kX) = k^2 D(X)$；

注：若事件 A 的概率为 1, 则称该事件几乎必然成立.

4. 若 X_1, X_2 相互独立, 则有

$$D(X_1+X_2) = D(X_1) + D(X_2).$$

推论：若 $X_1, X_2, \cdots, X_n$ 相互独立, 则有

$$D\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} D(X_k)$$

注：如果没有相互独立这一条件, 上式一般不成立.

1.4.3 协方差与相关系数

协方差

定义. 定义：对于二维随机向量 (X,Y), 称

$$\operatorname{Cov}(X,Y) := E[(X-E(X))(Y-E(Y))]$$

为 X 与 Y 的协方差*(Covariance).*

由定义直接可得：任意随机变量与其自身的协方差就是该随机变量的方差, 即

$$\operatorname{Cov}(X,X) = D(X).$$

协方差的性质

设 X,Y,Z 为随机变量, a,b,c,d 为常数, 则有

1. $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X)$；
2. $\operatorname{Cov}(aX+b, cY+d) = ac\operatorname{Cov}(X,Y)$；
3. $\operatorname{Cov}(X,Y+Z) = \operatorname{Cov}(X,Y) + \operatorname{Cov}(X,Z)$；
4. $\operatorname{Cov}(X,Y) = E(XY) - E(X)\cdot E(Y)$；
5. $D(X\pm Y) = D(X) + D(Y) \pm 2\operatorname{Cov}(X,Y)$.

推论. 两随机变量相互独立, 则协方差等于零；反之未必成立.

相关系数

定义. 对于二维随机变量 (X,Y), 如果两个变量的方差都不为零, 称

$$\rho_{X,Y} := \frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}} = \mathrm{Cov}(X^*,Y^*)$$

为 X 与 Y 的相关系数(Correlation), 也可以记为 $\rho(X,Y)$.

性质. 相关系数表示随机变量之间的线性相关程度:

1. $|\rho_{X,Y}| \le 1$.
2. $\rho_{X,Y} = -1$ 当且仅当 $Y = aX + b$, $a < 0$.
3. $\rho_{X,Y} = 1$ 当且仅当 $Y = aX + b$, $a > 0$.

定义. 若随机变量 X 与 Y 的相关系数 $\rho_{X,Y} = 0$, 则称 X 与 Y 线性互不相关, 简称不相关.

- $\rho_{X,Y} = -1$ 当且仅当 $Y = aX + b$, $a < 0$;
- $\rho_{X,Y} = 1$ 当且仅当 $Y = aX + b$, $a > 0$.

性质. 相互独立 $\Longrightarrow$ 不相关; 反之未必成立.

1.4.4 矩协方差矩阵

设 X 和 Y 是随机变量, 若

$$\mu_k = E\left(X^k\right), k = 1,2,\cdots$$

存在, 称它为 X 的 k 阶原点矩, 简称 k 阶矩. 若

$$\nu_k = E\left\{[X - E(X)]^k\right|, k = 2,3,\cdots$$

存在, 称它为 X 的 k 阶中心矩. 若

$$E\left(X^k Y^l\right), k,l = 1,2,\cdots$$

存在, 称它为 X 和 Y 的 $k+l$ 阶混合原点矩. 若

$$E\left\{[X - E(X)]^k [Y - E(Y)]^l\right\}, k,l = 1,2,\cdots$$

存在, 称它为 X 和 Y 的 $k+l$ 阶混合中心矩.

1.4.5 n 维正态分布

n 维正态随机变量具有以下四条重要性质:

1. n 维正态变量 $(X_1,X_2,\cdots,X_n)$ 的每一个分量 $X_i, i=1,2,\cdots,n$ 都是正态变量; 反之, 若 $X_1,X_2,\cdots,X_n$ 都是正态变量, 且相互独立, 则 $(X_1,X_2,\cdots,X_n)$ 是 n 维正态变量.
2. n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 服从 n 维正态分布的充要条件是 $X_1, X_2,\cdots,X_n$ 的任意线性组合 $l_1X_1+l_2X_2+\cdots+l_nX_n$ 服从一维正态分布 (其中 $l_1,l_2,\cdots, l_n$ 不全为零).
3. 若 $(X_1,X_2,\cdots,X_n)$ 服从 n 维正态分布, 设 $Y_1,Y_2,\cdots,Y_k$ 是 $X_j(j=1,2,\cdots,, n\ )$ 的线性函数, 则 $(Y_1,Y_2,\cdots,Y_k)$ 也服从多维的正态分布. 这一性质称为正态变量的线性变换不变性.
4. 设 $(X_i,X_2,\cdots,X_n)$ 服从 n 维正态分布, 则"$X_1,X_2,\cdots,X_n$ 相互独立" 与"$X_1, X_2,\cdots,X_n$ 两两不相关" 是等价的.

1.5 大数定律和中心极限定理

主要内容：切比雪夫不等式, 大数定律, 中心极限定理

1.5.1 大数定律

切比雪夫不等式：设随机变量 X 有期望和方差, 则对于任给的 $\varepsilon>0$, 有

$$P(|X-E(X)|\geq\varepsilon)\leq\frac{D(X)}{\varepsilon^2},$$

即

$$P(|X-E(X)|<\varepsilon)\geq 1-\frac{D(X)}{\varepsilon^2}.$$

定义 **1.** 设 $Y_1,Y_2,\cdots,Y_n,\cdots$ 是一个随机变量序列, a 是一个常数, 若对任何正数 ε, 有

$$\lim_{n\to\infty}P\{|Y_n-a|<\varepsilon\}=1$$

则称序列 $Y_1,Y_2,\cdots,Y_n,\cdots$ 依概率收敛于 a, 记为 $Y_n\xrightarrow{P}a$

依概率收敛的序列有如下性质:

- 设 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, 又设 $g(x,y)$ 在点 (a,b) 连续, 则
$$g\left(X_n,Y_n\right) \xrightarrow{P} g(a,b)$$

定理 1 (伯努利大数定律)**.** 设试验 E 是可重复进行的, 事件 A 在每次试验中出现的概率 $P(A)=p\ (0<p<1)$, 将试验独立地进行 n 次, 用 n_A 表示其中事件 A 出现的次数, 则对于任意正数 ε 有

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}=1$$

或者

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|\geq\varepsilon\right\}=0$$

定理 2 (切比雪夫大数定律的特殊情况)**.** 设随机变量 X_1, X_2, $\cdots$, X_n, $\cdots$ 相互独立, 且具有相同的数学期望和方差: $E\left(X_k\right)=\mu$, $D\left(X_k\right)=\sigma^2(k=1,2,\cdots)$ 作前 n 个随机变量的算术平均 $\overline{X}=\frac{1}{n}\sum\limits_{k=1}^{n}X_k$ 则对于任意正数 ε, 有

$$\begin{aligned}&\lim_{n\to\infty} P\{\overline{X}-\mu\mid<\varepsilon\}\\=&\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}E\left(X_k\right)\right|<\varepsilon\right\}=1\end{aligned}$$

定理 3 (辛钦大数定律)**.** 设随机变量 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立, 服从同一分布, 具有数学期望 $E\left(X_k\right)=\mu(k=1,2,\cdots)$, 则对于任意 $\varepsilon>0$ 有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_k-\mu\right|<\varepsilon\right\}=1.$$

1.5.2 中心极限定理

中心极限定理

常用结论：大量的同分布随机变量的和、平均值近似地服从正态分布.

定理 4 (独立同分布情形的中心极限定理)**.** 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 服从同一分布, 且具有数学期望和方差 : $E(X_i)=\mu, D(X_i)=\sigma^2>0(i=1,2,\cdots)$, 则随机变量之和 $\sum_{i=1}^{n} X_i$ 的标准化变量

$$Y_n=\frac{\sum_{i=1}^{n} X_i-E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{D\left(\sum_{i=1}^{n} X_i\right)}}=\frac{\sum_{i=1}^{n} X_i-n\mu}{\sqrt{n}\sigma}$$

的分布函数 $F_n(x)$ 对于任意 x 满足:

$$\lim_{n\to\infty} F_n(x)=\lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i-n\mu}{\sqrt{n}\sigma}\leqslant x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t=\Phi(x)$$

> 该定理表明：n 很大时，Y_n 近似服从标准正态分布.

定理 5 (棣莫弗一拉普拉斯中心极限定理)**.** 设随机变量 $\eta_n(n=1,2,\cdots)$ 服从参数为 $n,p(0<p<1)$ 的二项分布, 则对于任意 x, 有

$$\lim_{n\to\infty} P\left\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\leqslant x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t=\Phi(x)$$

当 n 充分大时, 对任意 $a<b$, 有

$$\begin{aligned}P\{a\leqslant\eta_n\leqslant b\}&=P\left\{\frac{a-np}{\sqrt{np(1-p)}}\leqslant\frac{\eta_u-np}{\sqrt{np(1-p)}}\leqslant\frac{b-np}{\sqrt{np(1-p)}}\right\}\\&\approx\Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right)\end{aligned}$$

1.6 样本及抽样分布

样本与统计量

主要内容：常用统计量, 三大统计分布：χ^2 分布、t 分布、F 分布, 正态分布常用统计量的分布

精确定义

定义：称随机变量 $X_1, X_2, \cdots, X_n$ 构成一个（简单）随机样本, 如果这些随机变量

1. 相互独立；
2. 服从相同的分布.

它们共同服从的分布称为总体分布；样本个数 n 称为样本容量.

常用统计量

定义：对样本 $X_1, X_2, \cdots, X_n$, 称

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

为样本均值.

常用统计量

定义：对样本 $X_1, X_2, \cdots, X_n$, 称

$$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

为样本方差；称

$$S := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

为样本标准差.

常用统计量

样本方差的性质：

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right).$$

χ^2 分布

χ^2 分布的性质：

1. 若 X 服从标准正态分布, 则 X^2 服从 1 个自由度的 χ^2 分布, 即
$$X^2 \sim \chi_1^2.$$
2. 可加性：设 $Y_1 \sim \chi_m^2$, $Y_2 \sim \chi_n^2$, 且两者相互独立, 则
$$Y_1 + Y_2 \sim \chi_{m+n}^2.$$

χ^2 分布

定理：设 $X_1, X_2, \cdots, X_n$ 相互独立, 都服从标准正态分布, 则

$$\chi^2 := \sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

服从 n 个自由度的 χ^2 分布, 即

$$\chi^2 \sim \chi_n^2.$$

t 分布

定理：设两个随机变量 X, Y 相互独立, 并且

$$X \sim N(0,1), \qquad Y \sim \chi_n^2.$$

则

$$T := \frac{X}{\sqrt{Y/n}}$$

服从具有 n 个自由度的 t 分布.

F 分布

定理：设两个随机变量 Y_1, Y_2 相互独立, 并且

$$Y_i \sim \chi^2_{n_i}, \qquad i = 1, 2.$$

则

$$F := \frac{Y_1/n_1}{Y_2/n_2} \sim F_{n_1,n_2}.$$

F 分布

F 分布的性质：

1. 若 $F \sim F_{m,n}$, 则
$$\frac{1}{F} \sim F_{n,m}.$$
2. $F_{m,n}(1-\alpha) = \dfrac{1}{F_{n,m}(\alpha)}$.
3. 若 $X \sim t_n$, 则
$$X^2 \sim F_{1,n}.$$

单个正态总体的统计量的分布

定理 1. 设 $X_1, X_2, \cdots, X_n$ 是取自正态总体 $N(\mu, \sigma^2)$ 的样本. 则 $\overline{X}$ 与 S^2 相互独立, 且有

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1), \qquad \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1},$$

$$\frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{\sigma^2} \sim \chi_n^2, \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

两个正态总体的统计量的分布

定理 2. 设 $X_1, X_2, \cdots, X_m$ 与 $Y_1, Y_2, \cdots, Y_n$ 分别是取自两个相互独立的正态总体

$$N(\mu_1, \sigma^2), \qquad N(\mu_2, \sigma^2)$$

的样本. 则

$$U := \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim N(0,1),$$

其中 $\overline{X}$, $\overline{Y}$ 分别是两个样本各自的均值.

两个正态总体的统计量的分布

定理 3. 设 $X_1, X_2, \cdots, X_m$ 与 $Y_1, Y_2, \cdots, Y_n$ 分别是取自两个相互独立的正态总体

$$N(\mu_1, \sigma^2), \qquad N(\mu_2, \sigma^2)$$

的样本. 则

$$T := \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}} \cdot \sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t_{m+n-2},$$

其中 $\overline{X}$, $\overline{Y}$, S_1^2, S_2^2 分别是两个样本各自的均值及方差.

两个正态总体的统计量的分布

定理 **4.** 设 $X_1,\cdots,X_m$ 与 $Y_1,\cdots,Y_n$ 分别是取自两个相互独立的正态总体

$$N(\mu_1,\sigma_1^2),\qquad N(\mu_2,\sigma_2^2)$$

的样本. 则

$$F:=\frac{\sum_{i=1}^{m}\left(X_i-\mu_1\right)^2\Big/\left(m\sigma_1^2\right)}{\sum_{j=1}^{n}\left(Y_j-\mu_2\right)^2\Big/\left(n\sigma_2^2\right)}\sim F_{m,n}$$

两个正态总体的统计量的分布

定理 **5.** 设 $X_1,\cdots,X_m$ 与 $Y_1,\cdots,Y_n$ 分别是取自两个相互独立的正态总体

$$N(\mu_1,\sigma_1^2),\qquad N(\mu_2,\sigma_2^2)$$

的样本. 则

$$F:=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F_{m-1,n-1}.$$

1.7 参数估计

1.7.1 参数估计

参数估计

主要内容：矩估计, 极大似然估计, 估计量的无偏性和方差的比较

矩估计

设总体 X 的分布函数为 $F(x;\theta_1,\cdots,\theta_k)$, 其中 $\theta_1,\cdots,\theta_k$ 为待估计的 k 个未知参数, 假设 X 的 $1\sim k$ 阶原点矩都存在, 则有

$$\mu_i=E\left(X^i\right)=\mu_i(\theta_1,\cdots,\theta_k)\quad i=1,2,\cdots,k$$

取

$$\hat{\mu}_i = \frac{1}{n}\sum_{j=1}^{n} X_j^i \quad i = 1, 2, \cdots, k$$

得方程组

$$\mu_i(\theta_1, \cdots, \theta_k) = \hat{\mu}_i \quad i = 1, 2, \cdots, k$$

解得

$$\hat{\theta}_i = \hat{\theta}_i(X_1, X_2, \cdots, X_n) \quad i = 1, 2, \cdots, k$$

称 $\hat{\theta}$ 为 θ 的矩法估计量简称矩估计

例子. 当总体中只有一个参数时, 矩估计即是用样本均值估计总体期望.

例子. 当总体中有两个或以上的参数时, 总体期望与方差的矩估计分别为

$$\hat{\mu} = \overline{X}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

极大似然估计

若总体为离散型, 其分布律为

$$P\{X = x\} = p(x, \theta_1, \theta_2, \cdots, \theta_k),$$

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, 其观测值为 $x_1, x_2, \cdots, x_n$ · 记

$$A = \{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\},$$

则事件 A 发生的概率为

$$L(\theta_1, \theta_2, \cdots, \theta_k) = \prod_{i=1}^{n} p(x_i, \theta_1, \theta_2, \cdots, \theta_k)$$

设连续型总体 X 的概率密度函数为

$$f(x;\ \theta_1, \theta_2, \cdots, \theta_k).$$

则样本 $X_1, X_2, \cdots, X_n$ 在观测值 $x_1, x_2, \cdots, x_n$ 处的联合概率密度为

$$L(x_1, \cdots, x_n;\ \theta_1, \cdots, \theta_k) = \prod_{i=1}^{n} f(x_i;\ \theta_1, \cdots, \theta_k).$$

此时 $\theta_1,\theta_2,\cdots,\theta_k$ 被看做固定但是未知的参数.

如果将观测值 $x_1,x_2,\cdots,x_n$ 看成固定的, 将

$$L(x_1,\cdots,x_n;\ \theta_1,\cdots,\theta_k)$$

看做 $\theta_1,\theta_2,\cdots,\theta_k$ 的函数, 则该函数被称为似然函数.

如果似然函数

$$L(x_1,x_2,\cdots,x_n;\ \theta_1,\theta_2,\cdots,\theta_k)$$

在

$$(\theta_1^*,\theta_2^*,\cdots,\theta_k^*)$$

处达到最大值, 则称上述参数为未知参数的极大似然估计.

求极大似然估计的一般方法：

1. 写出似然函数 L；
2. 求似然函数的对数 $\ln L$；
3. 对 $\ln L$ 求导（偏导）并令导数等于零, 得到似然方程组；
4. 解方程组得到 $\ln L$ 的驻点, 判断该驻点是否最大值点；
5. 将最大值点表达式中的 x_i 换为 X_i, 就得到参数的极大似然估计.

1.7.2 估计量的评选标准

估计量的评选标准：

- 无偏性, 即 $E(\hat{\theta})=\theta$
- 有效性, 即方差越小的估计量越有效
- 相合性, 即 $\hat{\theta}$ 依概率收敛于 θ