2025 春　概率论与数理统计　课程笔记

作者：招财鱼
组织：清华大学
时间：2025 Spring

If you' re dissatisfied with the world, change yourself. -- Kusanagi Motoko

目录

第一部分

概率部分

第 1 节 概率基本概念及事件运算

1.1 古典概型

定义 1.1 (古典概型)

如果每个基本事件都等可能出现，此时某一事件的概率为：

$$P(A) = \frac{\text{事件}A\text{包含的基本事件数}}{\text{全部可能的基本事件数}} \quad \text{或} \quad P(A) = \frac{\text{事件}A\text{所占区域大小}}{\text{样本空间所占区域大小}}$$

称为古典概型，也叫等可能概型。 ♣

例 1.1 生日问题 求 n $(n \leq 365)$ 个人的生日互不相同的概率 p_n

解：我们暂时只考虑一年 365 天的情况，则每个人的生日在这 365 天中等可能，故

$$p_n = \frac{365 \cdot (365-1) \cdots \big(365-(n-1)\big)}{365^n} = \prod_{k=0}^{n-1}\Big(1 - \frac{k}{365}\Big).$$

利用

$$\ln(1-x) \approx -x \quad (x \ll 1)$$

可得

$$\ln p_n = \sum_{k=0}^{n-1} \ln\Big(1 - \frac{k}{365}\Big) \approx -\sum_{k=0}^{n-1} \frac{k}{365} = -\frac{1+2+\cdots+(n-1)}{365} = -\frac{n(n-1)}{2 \cdot 365} = -\frac{n(n-1)}{730}.$$

于是定义近似值

$$\bar{p}_n = \exp\Big(-\frac{n(n-1)}{730}\Big),$$

则 $\bar{p}_n \approx p_n$。例如

$$\bar{p}_{10} \approx 0.884, \quad p_{10} \approx 0.883; \qquad \bar{p}_{30} \approx 0.3037, \quad p_{30} \approx 0.2937.$$

n 越大则估计的精度越低。

1.2 样本空间与随机事件

定义 1.2 (样本和样本空间)

随机事件一切可能的基本结果组成的集合成为样本空间，$\Omega = \{\omega\}$，每种可能的基本结果称为样本（点） ♣

- 抛掷 1 枚骰子：$\Omega = \{1, 2, 3, 4, 5, 6\}$
- 1 个人的出生月份：$\Omega = \{1, 2, \ldots, 12\}$
- 会面问题中 2 个人的会面时间：$\Omega = \{(x, y) \mid 0 \leq x \leq 60,\ 0 \leq y \leq 60\}$

定义 1.3 (随机事件)

（随机）事件即样本空间的某一子集。设集合 $A \subset \Omega$，则对任意样本点 $\omega \in \Omega$，有 $\begin{cases} \omega \in A, & \text{发生;} \\ \omega \notin A, & \text{未发生.} \end{cases}$ ♣

- 将一枚硬币抛掷两次，至少出现一次正面。令

$$\Omega = \{\text{正正, 正反, 反正, 反反}\}, \qquad A = \{\text{正正, 正反, 反正}\}, \quad A \subset \Omega.$$

- 会面问题中，两个人的到达时间相差不超过 20 分钟。令

$$\Omega = \{(x,y) \mid 0 \le x \le 60,\ 0 \le y \le 60\}, \qquad A = \{(x,y) \in \Omega \mid |x-y| \le 20\}.$$

定义 1.4 (示性函数)

事件 A 的示性函数或标志函数 (indicator) 定义为 $I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$ ♣

定义 (事件的关系)

1. 包含关系：若事件 A, B 满足 $A \subset B$，则称事件 B 包含事件 A，用示性函数表示为
$$I_A(\omega) \le I_B(\omega), \quad \forall \omega \in \Omega.$$
2. 相等关系：若 $A \subset B$ 且 $B \subset A$，即 $A = B$，则称事件 A 与事件 B 相等，用示性函数表示为
$$I_A(\omega) = I_B(\omega), \quad \forall \omega \in \Omega.$$
3. 互不相容关系（又称互斥关系）：对于事件 A, B，如果不可能同时发生，即 $A \cap B = \varnothing$，则称 A, B 互不相容，用示性函数表示为
$$I_A(\omega)\, I_B(\omega) = 0, \quad \forall \omega \in \Omega.$$
4. 对立关系（互逆关系）：如果事件 B 定义为"A 不发生"，即 $B = \bar{A}$，则称 A 与 B 具有对立关系，用示性函数表示为
$$I_A(\omega) + I_B(\omega) = 1, \quad \forall \omega \in \Omega.$$

♣

性质：概率的基本性质

$$P(\Phi) = 1, P(\bar{A}) = 1 - P(A), P(A-B) = P(A) - P(AB)$$

注：事件的运算（集合的运算）

- 交换律
$$A \cup B = B \cup A, \quad AB = BA\,(\text{即} A \cap B = B \cap A)$$
- 分配律
$$(A \cup B) \cap C = AC \cup BC, \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$
- 结合律
$$(A \cup B) \cup C = A \cup (B \cup C), \quad (AB)C = A(BC)$$
- 对偶律
$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \quad \overline{A \cap B} = \overline{A} \cup \overline{B}$$

例 1.2 错排问题 n 封写给不同人的信随机构入 n 个写好收信人姓名的信封，求所有信件都装错信封的概率 p_0。

解：将信和信封分别编号为 $1, 2, \ldots, n$，令

$$A_i = \{\text{第 } i \text{ 封信装入了编号为 } i \text{ 的信封}\}, \quad i = 1, 2, \ldots, n.$$

则"没有一封信装对"的事件就是

$$\bigcap_{i=1}^{n} \overline{A_i} = \overline{\bigcup_{i=1}^{n} A_i},$$

于是

$$p_0 = P\left(\cap_{i=1}^{n} \overline{A_i}\right) = 1 - P\left(\cup_{i=1}^{n} A_i\right).$$

由容斥原理，

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_iA_j) + \sum_{1\le i<j<k\le n} P(A_iA_jA_k) - \cdots + (-1)^{n-1} P(A_1A_2\cdots A_n).$$

而对于任意 $1 \le i_1 < \cdots < i_k \le n$，

$$P(A_{i_1}A_{i_2}\cdots A_{i_k}) = \frac{1}{n(n-1)\cdots(n-k+1)} = \frac{1}{k!}.$$

故

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!} \approx 1 - e^{-1},$$

从而

$$p_0 = 1 - P\Big(\bigcup_{i=1}^{n} A_i\Big) \approx e^{-1} \approx 0.37, \quad (n \ge 4 \text{ 时误差 } < 1\%).$$

第 2 节 条件概率与独立性

2.1 条件概率

2.1.1 条件概率的定义

定义 2.1 (条件概率)

设 A, B 是两个事件，且 $P(B) > 0$，则称

$$P(A \mid B) \;=\; \frac{P(A \cap B)}{P(B)}$$

为在事件 B 发生的条件下事件 A 发生的条件概率，简称条件概率。♣

2.1.2 三个重要公式

定理 2.1 (乘法公式)

1. 设 A, B 是两个事件，且 $P(B) > 0$，则
$$P(A \cap B) \;=\; P(B)\, P(A \mid B).$$
2. 设 $A_1, A_2, \ldots, A_n$ 为 n 个事件，且 $P\big(A_1 \cap A_2 \cap \cdots \cap A_{n-1}\big) > 0$，则
$$P\big(A_1 \cap A_2 \cap \cdots \cap A_n\big) = P(A_1)\, P\big(A_2 \mid A_1\big)\, P\big(A_3 \mid A_1 \cap A_2\big) \cdots P\big(A_n \mid A_1 \cap \cdots \cap A_{n-1}\big).$$
♡

例 2.1 设箱中有 a 个白球、b 个黑球，不放回地连续抽取 3 次，求三次都抽到白球的概率。

解：令事件 A_i 表示第 i 次抽到白球。由乘法公式

$$P(A_1A_2A_3) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1A_2),$$

且

$$P(A_1) = \frac{a}{a+b}, \quad P(A_2 \mid A_1) = \frac{a-1}{a+b-1}, \quad P(A_3 \mid A_1A_2) = \frac{a-2}{a+b-2}.$$

于是

$$P(A_1A_2A_3) = \frac{a}{a+b} \times \frac{a-1}{a+b-1} \times \frac{a-2}{a+b-2}.$$

定理 2.2 (全概率公式)

设 $B_1, B_2, \ldots, B_n$ 是样本空间 Ω 的一组划分，即

$$B_i \cap B_j = \varnothing \; (i \neq j), \qquad \bigcup_{i=1}^{n} B_i = \Omega,$$

且 $P(B_i) > 0 \; (i = 1, 2, \ldots, n)$。则对任意事件 A，有

$$P(A) \;=\; \sum_{i=1}^{n} P(B_i)\, P\big(A \mid B_i\big).$$
♡

例 2.2 设甲箱中有 a 个白球、b 个黑球（$a>0, b>0$）；乙箱中有 c 个白球、d 个黑球。先从甲箱中任取一球放入乙箱，然后再从乙箱中任取一球，求最后由乙箱取出白球的概率。

解：令事件

$$W=\{\text{从甲箱取出的是白球}\}, \qquad A=\{\text{最后从乙箱取出的是白球}\}.$$

由全概率公式，

$$P(A)=P(W)\,P(A\mid W)\ +\ P(\overline{W})\,P(A\mid \overline{W}).$$

显然

$$P(W)=\frac{a}{a+b},\quad P(A\mid W)=\frac{c+1}{c+d+1},\quad P(\overline{W})=\frac{b}{a+b},\quad P(A\mid \overline{W})=\frac{c}{c+d+1}.$$

于是

$$P(A)=\frac{a}{a+b}\cdot\frac{c+1}{c+d+1}+\frac{b}{a+b}\cdot\frac{c}{c+d+1}=\frac{a(c+1)+bc}{(a+b)(c+d+1)}=\frac{ac+bc+a}{(a+b)(c+d+1)}.$$

定理 2.3 (贝叶斯公式)

设 $B_1, B_2, \ldots, B_n$ 是样本空间 Ω 的一组划分，即

$$B_i\cap B_j=\varnothing\ (i\neq j),\qquad \bigcup_{i=1}^{n}B_i=\Omega,$$

且 $P(A)>0$，$P(B_i)>0\ (i=1,2,\ldots,n)$。则对任意事件 A，有

$$P(B_i\mid A)=\frac{P(B_i)\,P(A\mid B_i)}{\sum_{j=1}^{n}P(B_j)\,P(A\mid B_j)}.$$

♡

例 2.3 某地区某疾病的发病率为 $P(B)=0.0004$。现在有一检验方法，对真正患病者，化验结果 99% 呈阳性；对未患病者，则 0.1% 假阳性（即 99.9% 呈阴性）。求

1. 检查结果呈阳性，但实际上没有患病的概率；
2. 检查结果呈阴性，但实际上是患了病的概率。

解 令

$$A=\{\text{化验呈阳性}\},\quad B=\{\text{患病}\}.$$

由贝叶斯公式

$$P(\overline{B}\mid A)=\frac{P(\overline{B})\,P(A\mid\overline{B})}{P(B)\,P(A\mid B)+P(\overline{B})\,P(A\mid\overline{B})},=\frac{0.9996\times 0.001}{0.0004\times 0.99+0.9996\times 0.001}\approx 0.716.$$

同理，对“漏诊”事件

$$P(B\mid\overline{A})=\frac{P(B)\,P(\overline{A}\mid B)}{P(B)\,P(\overline{A}\mid B)+P(\overline{B})\,P(\overline{A}\mid\overline{B})}=\frac{0.0004\times 0.01}{0.0004\times 0.01+0.9996\times 0.999}\approx 4\times 10^{-6}.$$

2.2 事件的独立性

定义 2.2 (事件独立)

对于事件 A, B，若

$$P(A\cap B) = P(A)\,P(B),$$

则称 A 与 B 相互独立，简称 $A\perp B$，否则称 A 与 B 不独立或相关。

结论 若事件 A 与 B 独立，则

$$\begin{aligned} P\big(A\cap\overline{B}\big) &= P(A) - P(A\cap B) = P(A) - P(A)P(B) \\ &= P(A)\big(1-P(B)\big) = P(A)\,P(\overline{B}), \end{aligned}$$

故 A 与 $\overline{B}$ 独立。同理可证 $\overline{A}$ 与 B 独立，$\overline{A}$ 与 $\overline{B}$ 独立。

例 2.4 甲、乙二人轮流抛掷两枚骰子，取点数之和，若甲先掷得 6 点即为甲胜，若乙先掷得 7 点即为乙胜。求甲、乙各自的获胜概率。

解：记甲第 k 轮掷出 6 点为事件 A_k，乙第 k 轮掷出 7 点为事件 B_k，则：

$$\begin{aligned} P(\text{甲获胜}) &= P(A_1) + P(\overline{A_1}B_1A_2) + P(\overline{A_1}B_1\overline{A_2}B_2A_3) + P(\overline{A_1}B_1\cdots\overline{A_k}B_kA_{k+1}) + \cdots \\ &= P(A_1) + P(\overline{A_1})\,P(B_1)\,P(A_2) + P(\overline{A_1})\,P(B_1)\,P(\overline{A_2})\,P(B_2)\,P(A_3) + \cdots \\ &= \frac{5}{36}\left(1+\frac{31}{36}\cdot\frac{5}{6}+\left(\frac{31}{36}\cdot\frac{5}{6}\right)^2+\cdots\right) = \frac{30}{61}. \end{aligned}$$

定义 2.3 (事件的相容性)

对于两个事件 A, B，若 $P(AB) = 0$，则称事件 A, B 是互不相容的。

定义 2.4 (事件相关系数)

设 A, B 是两个事件，且 $0 < P(A), P(B) < 1$。定义它们的相关系数

$$\tau(A,B) = \frac{P(AB) - P(A)\,P(B)}{\sqrt{P(A)\big(1-P(A)\big)\,P(B)\big(1-P(B)\big)}}.$$

相关系数具有以下性质：

1. 若 A 与 B 相互独立，则 $\tau(A,B) = 0$；
2. 若 A 与 B 相等，则 $\tau(A,B) = 1$；
3. $\big|\tau(A,B)\big| \le 1$。

第 3 节 离散型随机变量

3.1 随机变量的定义

定义 3.1 (随机变量)

定义在样本空间 Ω 上的函数 $X:\Omega \to \mathbb{R}$ 称为随机变量。通常用大写字母 X, Y, Z 等表示随机变量，小写字母 x, y, z 等表示其取值：$x = X(\omega)$, $\omega \in \Omega$。随机变量 $X(\omega)$ 一般简记为 X。♣

例 3.1 掷硬币，正面记 H，背面记 T。

1）掷一次，$\Omega = \{H, T\}$，定义随机变量 X 为：

$$X(H) = 0, \quad X(T) = 4.$$

其取值表为：

ω	H	T
$X(\omega)$	0	4

概率分布为：

$$P(X=0) = \frac{1}{2}, \quad P(X=4) = \frac{1}{2}.$$

例 3.2 设 $(\Omega, \mathcal{F}, P)$ 是概率空间，$A \in \mathcal{F}$，定义

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

称 $I_A(\omega)$ 为事件 A 的示性函数。

定义 3.2 (随机变量的分布函数)

设 X 是一个随机变量，对任意实数 x，定义

$$F(x) = P\big(X \le x\big).$$

称 $F(x)$ 为随机变量 X 的分布函数，记为

$$X \sim F(x),$$

有时也记作 $F_X(x)$。♣

性质 根据定义有：

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1.$$

且对于任意实数 $a < b$，有

$$P\big(a < X \le b\big) = F(b) - F(a), \qquad P\big(a \le X \le b\big) = F(b) - F(a) + P\big(X = a\big).$$

3.2 离散型随机变量

定义 3.3 (离散型随机变量)

如果随机变量 X 的所有可能取值是有限个或可列多个，则其分布可表示为

$$\begin{array}{c|ccccc} X & x_1 & x_2 & \cdots & x_n & \cdots \\ \hline P & p_1 & p_2 & \cdots & p_n & \cdots \end{array} \quad \text{或} \quad X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}.$$

这种表示称为分布列。其中

$$p_i = P(X = x_i) \geq 0, \quad i = 1, 2, \ldots, n, \ldots,$$

且满足

$$\sum_{i=1}^{n} p_i = 1 \quad \text{或} \quad \sum_{i=1}^{\infty} p_i = 1.$$

也可记作

$$X \sim \big(x_1,\ x_2,\ \ldots;\ P(X = x_1),\ P(X = x_2),\ \ldots\big).$$

♣

定义 (两点分布)

如果随机变量 X 的所有可能取值只有两个——a_0 和 a_1，且

$$P(X = a_0) = p, \quad P(X = a_1) = q, \qquad p + q = 1,\ p, q \geq 0,$$

则称 X 服从两点分布，记作

$$X \sim \begin{pmatrix} a_0 & a_1 \\ p & q \end{pmatrix}.$$

♣

其分布函数

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < a_0, \\ p, & a_0 \leq x < a_1, \\ 1, & x \geq a_1. \end{cases}$$

最常用的两点分布是 $0-1$ 分布，也称伯努利分布，此时

$$X \sim \begin{pmatrix} 0 & 1 \\ p & q \end{pmatrix}, \quad q = 1 - p.$$

定义 (伯努利 (Bernoulli) 试验)

一随机试验有两个基本结果，记为事件 A 和 $\overline{A}$，且

$$P(A) = p, \quad P(\overline{A}) = q, \quad p + q = 1.$$

♣

定义（二项分布）

将伯努利试验独立地重复 n 次，例如连续投掷 n 次硬币、连续 n 次射击等，基本结果（过程）有 2^n 种。设随机变量 X 表示事件 A 出现的次数，则 X 的分布为：

$$\begin{pmatrix} 0 & 1 & 2 & \cdots & k & \cdots & n \\ p_0 & p_1 & p_2 & \cdots & p_k & \cdots & p_n \end{pmatrix},$$

其中

$$p_k = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n.$$

此分布称为二项分布，记作：

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

p_k 为 $(p+q)^n$ 的二项展开式中的系数。

♣

例 3.3 假设每一台自动车床在一段运行过程中需要维修的概率均为 $p = 0.01$，并且各机床的运行是相互独立的，求：(1) 如果每名维修工人负责看管 20 台机床，不能及时维修的概率；

(2) 如果 3 名维修工人负责看管 80 台机床，不能及时维修的概率。

解 设随机变量 X 为一段时间内出故障的机床数目，则：

(1) $X \sim b(20, 0.01)$，不能及时维修的概率为

$$P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - 0.99^{20} - 20 \times 0.01 \times 0.99^{19} \approx 0.017.$$

(2) $X \sim b(80, 0.01)$，不能及时维修的概率为

$$P(X > 3) = 1 - \sum_{k=0}^{3} P(X = k) \approx 0.0087.$$

定义（泊松（Poisson）分布）

泊松分布用于描述稀有事件的发生次数，记作

$$X \sim P(\lambda),$$

其中 $\lambda > 0$ 是给定的常数。其概率分布为：

$$\begin{pmatrix} 0 & 1 & 2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \end{pmatrix}, \quad 其中 p_k = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

概率的总和为：

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

♣

注 泊松分布与二项分布 $b(n,p)$ 有如下关系：当 n 很大且 p 很小时，$b(n,p)$ 与 $P(np)$ 非常接近。

若 $X \sim P(\lambda), \quad Y \sim b\left(n, \frac{\lambda}{n}\right)$，则当 n 很大时，有：

$$P(X = k) \approx P(Y = k).$$

例 3.4 一个业余选手射箭中靶心的概率是 0.0001。虽然单支箭射中靶心的概率非常之低，但只要重复足够多次，有几次射中靶心的概率还是很大的。例如重复 30000 次，射中靶心 3 次的概率大约为

$$b_3(30000, 0.0001) \approx p_3(3) = 0.224.$$

定理 3.1 (泊松定理)

在 n 重伯努利试验中，设事件 A 在一次试验中发生的概率为 p_n（此概率与试验次数 n 有关），如果当 $n \to \infty$ 时，有 $np_n \to \lambda$，则：

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

♡

定义（几何分布）

考虑伯努利试验，设事件 A 发生的概率为 $P(A) = p$。若随机变量 X 表示第一次成功出现所需的试验次数，则称 X 服从参数为 p 的几何分布，记作

$$X \sim Ge(p).$$

其概率分布为：

$$P(X=k) = (1-p)^{k-1} \cdot p, \quad k = 1, 2, 3, \ldots$$

♣

性质 几何分布的无记忆性： 如果随机变量 $X \sim Ge(p)$，则对任意非负整数 s, t，有：

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

命题

取正整数值的随机变量 X，如果满足 $P(X > s+t \mid X > s) = P(X > t)$，则 X 一定服从几何分布。

♠

连续不断且独立地重复进行一个参数为 p 的伯努利试验，设随机变量 X 表示第 r 次“成功”出现时所需的试验次数，则事件 $\{X = k\}$ 表示第 k 次试验“成功”，且在前 $k-1$ 次试验中恰好有 $r-1$ 次成功。

由试验的独立性及二项分布的性质可得：

$$P(X=k) = p \cdot P\left(B(k-1,p) = r-1\right) = p \cdot \binom{k-1}{r-1} p^{r-1} q^{k-r} = \binom{k-1}{r-1} p^r q^{k-r},$$

其中 $q = 1-p, \quad k = r, r+1, r+2, \ldots$

定义（负二项分布）

若随机变量 X 的分布律为

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

则称 X 服从参数为 r, p 的负二项分布，记作

$$X \sim NB(r,p).$$

♣

注 若随机变量 $X \sim NB(r,p)$，则：

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

其中 $q = 1-p, \quad k = 0, 1, 2, \ldots$

满足归一化条件：

$$\sum_{k=0}^{\infty} P(X=r+k) = p^r \sum_{k=0}^{\infty} \binom{k}{-r} (-q)^k = p^r (1-q)^{-r} = 1.$$

其中 $\binom{a}{k}$ 是整数组合数的推广，定义为

$$\binom{a}{k} = \frac{a(a-1)\cdots(a-k+1)}{k!}.$$

第 4 节 连续型随机变量

定理 4.1 (分布函数的性质)

任一分布函数 $F(x)$ 都具有如下三条基本性质：

1. 单调性：

   $F(x)$ 是 $(-\infty,+\infty)$ 上的单调非减函数，即
   $$\forall x_1 < x_2, \quad F(x_1) \le F(x_2);$$

2. 有界性：

   $\forall x \in \mathbb{R}$，有 $0 \le F(x) \le 1$，且
   $$F(-\infty) = \lim_{x\to-\infty} F(x) = 0, \quad F(+\infty) = \lim_{x\to+\infty} F(x) = 1;$$

3. 右连续性：

   $F_X(x)$ 是关于 x 的右连续函数，即
   $$\forall x_0 \in \mathbb{R}, \quad \lim_{x\to x_0^+} F(x) = F(x_0), \quad F(x_0+0) = F(x_0).$$

♡

此外，还具有以下重要关系式：

$$P(a < X \le b) = F(b) - F(a), \qquad P(a \le X \le b) = F(b) - F(a-0),$$
$$P(X \ge b) = 1 - F(b-0), \qquad P(X = a) = F(a) - F(a-0).$$

4.1 连续型随机变量

定义 4.1 (连续型随机变量)

设随机变量 X 的分布函数为 $F(x)$，如果存在非负可积函数 $p(x)$ （$x \in \mathbb{R}$），使得
$$\forall x \in \mathbb{R}, \quad F(x) = \int_{-\infty}^{x} p(t)\,dt,$$
则称 X 为连续型随机变量，$p(x)$ 称为 X 的概率密度函数，并且满足：
$$\int_{-\infty}^{+\infty} p(x)\,dx = 1, \qquad F'(x) = p(x).$$

♣

定义 (指数分布)

若随机变量 X 的概率密度函数为
$$p(x) = \begin{cases} 0, & x \le 0, \\ \lambda e^{-\lambda x}, & x > 0, \end{cases} \quad \text{其中}\lambda > 0,$$
则称 X 服从参数为 λ 的指数分布，记作
$$X \sim \mathrm{Exp}(\lambda).$$

其分布函数为：
$$F(x) = \begin{cases} 0, & x \le 0, \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

♣

性质 无记忆性：指数分布具有无记忆性，即：

$$\forall s>0,\, t>0, \quad P(Y>s+t \mid Y>s)=P(Y>t).$$

定理（指数分布与几何分布的联系）

设随机变量 $X\sim \text{Exp}(\lambda)$, $\lambda>0$，则满足以下分布律：

$$P(Y=k)=P(k-1\le X\le k), \quad k=1,2,\ldots$$

随机变量 Y 服从参数为 $1-e^{-\lambda}$ 的几何分布。♡

证明 我们有：

$$P(Y=k)=\int_{k-1}^{k}\lambda e^{-\lambda x}dx=-e^{-\lambda x}\Big|_{k-1}^{k}=e^{-\lambda(k-1)}-e^{-\lambda k}=(1-e^{-\lambda})\cdot e^{-\lambda(k-1)}.$$

也即：

$$P(k-1\le X\le k)=F_X(k)-F_X(k-1)=\left(1-e^{-\lambda k}\right)-\left(1-e^{-\lambda(k-1)}\right).$$

定义（正态分布）

标准正态分布：若随机变量 $X\sim N(0,1)$，则其概率密度函数为

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \quad x\in\mathbb{R},$$

其分布函数为

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

一般正态分布：若 $X\sim N(\mu,\sigma^2)$，则其概率密度函数为

$$p(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x\in\mathbb{R},\ \text{满足}\ \frac{X-\mu}{\sigma}\sim N(0,1)$$

♣

注：σ 表示分散程度，σ 越大则数据分布越散开，越小则数据分布越集中.

4.2 随机变量函数的分布

已知随机变量 X 的分布，求 $Y=g(X)$ 的分布。

1. 离散随机变量函数的分布

设 X 是离散型随机变量，其分布列为：

$$\begin{array}{c|cccc} X & x_1 & x_2 & \cdots & x_n \\ \hline P & p_1 & p_2 & \cdots & p_n \end{array} \Rightarrow Y=g(X)\Rightarrow \begin{array}{c|cccc} Y & g(x_1) & g(x_2) & \cdots & g(x_n) \\ \hline P & p_1 & p_2 & \cdots & p_n \end{array}$$

2. 连续随机变量函数的分布

若 $Y=g(X)$，则其分布函数为：

$$F_Y(y)=P(Y\le y)=P(g(X)\le y).$$

例 4.1 设 $X \sim U(-1,1)$，求 $Y = X^2$ 的分布。

解 考虑分布函数 $F_Y(y) = P(Y \le y) = P(X^2 \le y)$。

- 当 $y < 0$ 时，$F_Y(y) = 0$；
- 当 $y > 1$ 时，$F_Y(y) = 1$；
- 当 $0 \le y \le 1$ 时，

$$F_Y(y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \sqrt{y}.$$

综上，$F_Y(y)$ 为：

$$F_Y(y) = \begin{cases} 0, & y < 0, \\ \sqrt{y}, & 0 \le y \le 1, \\ 1, & y > 1. \end{cases} \quad \Rightarrow \quad p_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} \frac{1}{2\sqrt{y}}, & 0 < y \le 1, \\ 0, & \text{其他}. \end{cases}$$

例 4.2 设 $X \sim N(0,1)$，求 $Y = X^2$ 的分布。

解 分布函数：

$$F_Y(y) = P(Y \le y) = P(X^2 \le y).$$

- 当 $y < 0$，$F_Y(y) = 0$；
- 当 $y \ge 0$，

$$F_Y(y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1.$$

概率密度函数：

$$p_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = \frac{\mathrm{d}}{\mathrm{d}y}\left[2\Phi(\sqrt{y}) - 1\right] = 2 \cdot \varphi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{\varphi(\sqrt{y})}{\sqrt{y}}.$$

所以

$$p_Y(y) = \begin{cases} \dfrac{\varphi(\sqrt{y})}{\sqrt{y}}, & y > 0, \\ 0, & y < 0. \end{cases}$$

定理 4.2 (连续型随机变量函数的密度)

设连续随机变量 X 的密度函数为 $p_X(x)$，$Y = g(X)$。

1. 若 $y = g(x)$ 严格单调，且其反函数 $h(y)$ 存在并具有连续导函数，则 $Y = g(X)$ 的密度函数为：

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

   其中

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

2. 若 $y = g(x)$ 在若干不相重叠的区间 $I_1, I_2, \ldots$ 上逐段严格单调，且各段反函数 $h_1(y), h_2(y), \ldots$ 具有连续导数，则 $Y = g(X)$ 的密度函数为：

$$p_Y(y) = \sum_i p_X(h_i(y))\,|h_i'(y)|.$$

♡

定理（从均匀分布构造一般分布）

设随机变量 U 服从区间 $[0,1]$ 上的均匀分布，函数 F 为定义在实数集 $\mathbb{R}$ 上的连续单调递增函数，且满足：

$$F(-\infty)=0,\quad 0\le F(x)\le 1,\quad F(+\infty)=1.$$

则随机变量

$$X=F^{-1}(U)$$

的概率分布函数为 $F(x)$，即 X 服从分布函数 $F(x)$ 所确定的分布。

练习 4.1 利用 $U(0,1)$ 分布的随机数生成服从期望为 $\frac{1}{\lambda}$ 的指数分布随机数。

解 指数分布的分布函数为：

$$F(x)=1-e^{-\lambda x}\quad\Rightarrow\quad F^{-1}(x)=-\frac{\ln(1-x)}{\lambda}.$$

若 $X\sim U(0,1)$，则 $1-X\sim U(0,1)$，故令：

$$Y=-\frac{\ln X}{\lambda},$$

即可使 Y 服从参数为 λ 的指数分布。

验证：当 $x\in(0,1)$ 时，有 $-\frac{\ln x}{\lambda}>0$，因此：

- 当 $y\le 0$ 时，

$$F_Y(y)=P(Y\le y)=0;$$

- 当 $y>0$ 时，

$$F_Y(y)=P\left(-\frac{\ln X}{\lambda}\le y\right)=P\left(\ln X\ge -\lambda y\right)=P\left(X\ge e^{-\lambda y}\right)=1-e^{-\lambda y}.$$

因此，$Y\sim \mathrm{Exp}(\lambda)$。

第 5 节 期望方差

5.1 随机变量的数字特征

5.1.1 随机变量的数学期望

定义 5.1 (随机变量的数学期望)

设随机变量 X，其数学期望（Expectation）定义如下：

$$E(X)=\begin{cases}\sum_{i=1}^{\infty}x_ip(X=x_i), & X \text{ 为离散随机变量;}\\ \int_{-\infty}^{+\infty}xp(x),dx, & X \text{ 为连续随机变量.}\end{cases}$$

♣

注 数学期望存在的充分必要条件如下：

$$\sum_{i=1}^{\infty}|x_i|p(X=x_i)<\infty,\qquad \int_{-\infty}^{+\infty}|x|p(x),dx<\infty.$$

性质 数学期望的性质：

1. 期望的常数性：

$$E(c)=c,\ (c\text{为常数，常值分布})$$

2. 期望的线性性：

$$E(cX)=cE(X),\ E(X+Y)=E(X)+E(Y)E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)$$

定理（随机变量函数的期望）

设随机变量 X，其函数 $g(X)$ 的期望为：

$$E\big(g(X)\big)=\begin{cases}\sum_{i=1}^{\infty}g(x_i)p(X=x_i), & X \text{ 为离散随机变量;}\\ \int_{-\infty}^{+\infty}g(x)p_X(x),dx, & X \text{ 为连续随机变量.}\end{cases}$$

♡

5.1.2 随机变量的方差

定义 5.2 (随机变量的方差)

随机变量偏离期望的程度（随机变量分布的分散程度）。

- 定义式：随机变量 X 的方差 $Var(X)$，或写作 $D(X)$，定义为：

$$Var(X)=E\left((X-E(X))^2\right)=E(X^2)-E(X)^2$$

- 方差的性质（a,b 为常数）：

$$Var(aX+b)=a^2Var(X)$$

- 标准差：随机变量 X 的标准差 $\sigma(X)$ (也记作 σ_X) 定义为方差的算术平方根：

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

♣

5.2 常见随机变量的期望和方差

表 5.1: 常用概率分布及其数学期望和方差

分布	分布列 p_k 或分布密度 $p(x)$	期望	方差
0-1 分布	$p_k = p^k(1-p)^{1-k}, \quad k = 0, 1$	p	$p(1-p)$
二项分布 $b(n, p)$	$p_k = \binom{n}{k} p^k(1-p)^{n-k}, \quad k = 0, 1, \ldots, n$	np	$np(1-p)$
泊松分布 $P(\lambda)$	$p_k = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, \ldots$	λ	λ
超几何分布 $h(n, N, M)$	$p_k = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \ldots, r; r = \min\{M, n\}$	$\frac{nM}{N}$	$\frac{nM(N-M)(N-n)}{N^2(N-1)}$
几何分布 $Ge(p)$	$p_k = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
负二项分布 $Nb(r, p)$	$p_k = \binom{k-1}{r-1}(1-p)^{k-r}p^r, \quad k = r, r+1, \ldots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
正态分布 $N(\mu, \sigma^2)$	$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty$	μ	σ^2
均匀分布 $U(a, b)$	$p(x) = \frac{1}{b-a}, \quad a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布 $Exp(\lambda)$	$p(x) = \lambda e^{-\lambda x}, \quad x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
伽马分布 $Ga(\alpha, \lambda)$	$p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, \quad x \geq 0$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$\chi^2(n)$ 分布	$p(x) = \frac{x^{n/2-1}e^{-x/2}}{\Gamma(n/2)2^{n/2}}, \quad x \geq 0$	n	$2n$
贝塔分布 $Be(a, b)$	$p(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$
对数正态分布 $LN(\mu, \sigma^2)$	$p(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\}, \quad x > 0$	$e^{\mu+\sigma^2/2}$	$e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$

注 某些随机变量的数学期望或方差可能不存在。一个典型的例子是柯西分布。

例 5.1 柯西 (Cauchy) 分布，若随机变量 $X \sim c(\lambda, \mu)$，则其概率密度函数为：

$$p(x) = \frac{\lambda}{\pi(\lambda^2 + (x-\mu)^2)}$$

其中 $\lambda > 0$，$x \in (-\infty, +\infty)$。

证明 考虑计算 $X \sim c(1,0)$ (即标准柯西分布，$\lambda = 1, \mu = 0$) 的期望。标准柯西分布的概率密度函数为 $p(x) = \frac{1}{\pi(1+x^2)}$。根据期望存在的条件，需要绝对一阶矩收敛，即 $\int_{-\infty}^{+\infty} |x| p(x) dx < \infty$。计算该积分：

$$\int_{-\infty}^{+\infty} |x| p(x) dx = \int_{-\infty}^{+\infty} |x| \frac{1}{\pi(1+x^2)} dx = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{|x|}{1+x^2} dx = \frac{2}{\pi} \int_{0}^{+\infty} \frac{x}{1+x^2} dx$$

计算这个瑕积分：

$$\int_{0}^{+\infty} \frac{x}{1+x^2} dx = \lim_{b \to +\infty} \int_0^b \frac{x}{1+x^2} dx = \lim_{b \to +\infty} \left[\frac{1}{2} \ln(1+x^2) \right]_0^b = \lim_{b \to +\infty} \left(\frac{1}{2} \ln(1+b^2) - 0 \right) \to \infty$$

由于该积分发散到无穷大，所以 $E(X)$ 不存在。

同理，由于 $E(X^2) = \int_{-\infty}^{+\infty} x^2 p(x) dx = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x^2}{1+x^2} dx$ 也发散到无穷大（因为当 $x \to \pm\infty$ 时，$\frac{x^2}{1+x^2} \to 1$，积分区间为无限)，所以 $E(X^2)$ 也不存在。因此，方差 $Var(X) = E(X^2) - (E(X))^2$（在期望存在的前提下）自然也不存在。

5.3 期望和方差的几点基本性质

定理 5.1 (期望和方差的基本性质)

以下是关于随机变量期望和方差的一些基本性质：

1. 期望的最小二乘性质：设 X 为一随机变量，对于任意常数 c，均方误差 $E((X-c)^2)$ 在 $c = E(X)$ 时达到其最小值。这个最小值就是随机变量 X 的方差 $Var(X)$。

$$\min_c E((X-c)^2) = E((X - E(X))^2) = Var(X)$$

2. 切比雪夫 **(Chebyshev)** 不等式：
设随机变量 X 具有数学期望 $E(X)$ 和方差 $Var(X)$。则对于任意正数 $\epsilon > 0$，有：

$$P(|X - E(X)| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$$

或者，其等价形式（常用于描述随机变量落在期望值某个范围内的概率）为：

$$P(|X - E(X)| < \epsilon) \geq 1 - \frac{Var(X)}{\epsilon^2}$$

(此等价形式仅在 $1 - \frac{Var(X)}{\epsilon^2} > 0$ 时有意义。)

♡

定义 5.3 (原点矩与中心矩)

对于随机变量 X，我们定义以下两种类型的矩：

- k 阶原点矩：随机变量 X 的 k 阶原点矩定义为 X^k 的数学期望，记作 $E(X^k)$，其中 k 为正整数。
    - 当 $k = 1$ 时，1 阶原点矩就是随机变量 X 的数学期望：
- k 阶中心矩：随机变量 X 的 k 阶中心矩定义为 $(X - E(X))^k$ 的数学期望，记作 $E((X - E(X))^k)$，其中 k 为正整数。
    - 当 $k = 2$ 时，2 阶中心矩就是随机变量 X 的方差：

    $$E((X - E(X))^2)$$

    这通常也记为 $Var(X)$ 或 σ^2。

♣

5.4 特征函数

定义 5.4 (特征函数)

X 是一随机变量，称 $\phi(t)=E(e^{itX}),(t\in\mathbb{R})$, 为 X 的特征函数。

因为 $|e^{itX}|=1$，所以 $E(e^{itX})$ 总是存在的，即任一随机变量的特征函数总是存在的。

$$\phi(t)=E(e^{itX})=\begin{cases}\sum_{k=1}^{\infty}e^{itx_k}P(X=x_k) & (\text{对于离散型随机变量})\\ \int_{-\infty}^{+\infty}e^{itx}p(x)\,dx & (\text{对于连续型随机变量，}p(x)\text{ 为其密度函数})\end{cases}$$

分布函数与特征函数一一对应。

♣

定理 5.2 (特征函数的连续性)

随机变量的特征函数 $\phi(t)$ 在实数轴上一致连续。

♡

5.4.1 几种常见分布的特征函数

1. **0-1 分布**，$X\sim b(1,p)$

$$\phi_X(t)=E(e^{itX})=p\cdot e^{it\cdot 1}+(1-p)\cdot e^{it\cdot 0}=p\cdot e^{it}+q$$

2. **几何分布**，$X\sim \mathbf{Ge}(p)$

$$\phi_X(t)=\sum_{k=1}^{+\infty}e^{itk}pq^{k-1}=p\cdot e^{it}\sum_{k=0}^{+\infty}(q\cdot e^{it})^k=\frac{p\cdot e^{it}}{1-q\cdot e^{it}}$$

3. **泊松分布**，$X\sim P(\lambda)$

$$\phi_X(t)=\sum_{k=0}^{+\infty}e^{itk}\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\cdot\sum_{k=0}^{+\infty}\frac{(\lambda e^{it})^k}{k!}=e^{\lambda(e^{it}-1)}$$

4. **均匀分布**，$X\sim U(-1,1)$

$$\phi_X(t)=E(e^{itX})=\int_{-1}^{1}e^{itx}\cdot\frac{1}{2}dx=\int_{0}^{1}\cos(tx)dx=\frac{\sin t}{t}$$

5. **标准正态分布**，$X\sim N(0,1)$

$$\phi_X(t)=\int_{-\infty}^{+\infty}e^{itx}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=2\int_0^{+\infty}\cos(tx)\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

$$\begin{aligned}\frac{\mathrm{d}\phi_X(t)}{\mathrm{d}t}&=-2\int_0^{+\infty}x\cdot\sin(tx)\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=2\int_0^{+\infty}\sin(tx)\cdot d\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)\\&=-2t\cdot\int_0^{+\infty}\cos(tx)\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=-t\cdot\phi_X(t),\quad \phi_X(0)=1\end{aligned}$$

$$\Longrightarrow\ \phi_X(t)=e^{-\frac{t^2}{2}}$$

5.4.2 特征函数的性质

1. $|\phi(t)| \leq \phi(0) = 1$
2. $\phi(-t) = \overline{\phi(t)}$
3. 若 $Y = aX + b$，则 $\phi_Y(t) = e^{ibt}\phi_X(at)$
4. 若 X, Y 相互独立，则 $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$
   卷积公式：
   $$p_Z(z) = \int_{-\infty}^{\infty} p_{XY}(z-w, w)dw = \int_{-\infty}^{\infty} p_X(z-w)p_Y(w)dw \qquad \text{卷积运算} \to \text{乘法运算}$$
5. 若 $E(X^m)$ 存在（即 m 阶矩存在），则特征函数 $\phi(t)$ 对于 t m 阶可导，并且对于任意整数 $1 \leq k \leq m$，有：
   $$\phi^{(k)}(0) = i^k E(X^k)$$
   (其中 $\phi^{(k)}(0)$ 表示 $\phi(t)$ 在 $t=0$ 处的 k 阶导数，i 是虚数单位。) (积分运算 (各阶矩) $\to$ 微分运算)

例 5.2 正态分布 $Y \sim N(\mu, \sigma^2)$ 的特征函数

解

$$Y = \sigma \cdot X + \mu, \quad X \sim N(0,1)$$

$$\phi_Y(t) = e^{i\mu t}\phi_X(\sigma \cdot t) = e^{i\mu t - \dfrac{\sigma^2 t^2}{2}}$$

例 5.3 二项分布 $Y \sim b(n,p)$ 的特征函数，$Y = X_1 + X_2 + \cdots + X_n$, 其中 $X_k \sim b(1,p)$ 且相互独立

解 $\phi_{X_k}(t) = p \cdot e^{it} + q$

$$\phi_Y(t) = \prod_{k=1}^{n} \phi_{X_k}(t) = (p \cdot e^{it} + q)^n$$

例 5.4 $X \sim N(0,1)$，计算 $E(X^k)$

解 $\phi_X(t)$ 的泰勒展开：

$$\phi_X(t) = e^{-\frac{t^2}{2}} = \sum_{k=0}^{\infty} \frac{1}{k!}\left(-\frac{t^2}{2}\right)^k = \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!} t^{2k}$$

由泰勒级数系数可知 $\phi_X(t)$ 在 $t=0$ 处的各阶导数：

$$\phi_X^{(2k+1)}(0) = 0, \quad \phi_X^{(2k)}(0) = \frac{(-1)^k (2k)!}{2^k k!}$$

利用特征函数与矩的关系 $E(X^m) = \dfrac{\phi_X^{(m)}(0)}{i^m}$：对于偶数阶矩 $m = 2k$ $(k \geq 0)$：

$$E(X^{2k}) = \frac{1}{i^{2k}}\phi_X^{(2k)}(0) = \frac{1}{(-1)^k}\frac{(-1)^k(2k)!}{2^k k!} = \frac{(2k)!}{2^k k!} = (2k-1)!!$$

定理 5.3 (傅里叶变换)

对于连续型随机变量 X，其特征函数 $\phi(t)$ 和密度函数 $p(x)$ 满足：

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} p(x)\, dx, \quad p(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\phi(t)\, dt$$

♡

定理 5.4 (唯一性定理)

两个分布函数 $F_1(x)$ 与 $F_2(x)$ 相等的充分必要条件是它们的特征函数 $\phi_1(t)$ 与 $\phi_2(t)$ 恒等。

♡

定理 5.5 (逆转公式)

设 $F(x)$ 与 $\phi(t)$ 分别为随机变量 X 的分布函数和特征函数，则对任意 $x_1 < x_2$：

$$\frac{F(x_2+0)+F(x_2-0)}{2} - \frac{F(x_1+0)+F(x_1-0)}{2} = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}\phi(t)\,dt$$

若 $F(x)$ 在 x_1, x_2 连续，则：

$$F(x_2)-F(x_1) = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}\phi(t)\,dt$$

例 5.5 林德伯格-勒维中心极限定理的证明:

$$\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \xrightarrow{n\to\infty} N(0,1)$$

解 证明：

设 $X_k-\mu$ 的特征函数是 $\varphi(t)$，则

$$Y_n = \frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}} = \frac{(X_1-\mu)+(X_2-\mu)+\cdots+(X_n-\mu)}{\sigma\sqrt{n}}$$

的特征函数为

$$\varphi_{Y_n}(t) = \left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n$$

而 $E(X_k-\mu)=0, \mathrm{Var}(X_k-\mu)=\sigma^2$，则 $X_k-\mu$ 的特征函数 $\varphi(t)$ 展开为：

$$\varphi(t) = \varphi(0)+\varphi'(0)t+\varphi''(0)\frac{t^2}{2}+o(t^2) = 1-\frac{1}{2}\sigma^2t^2+o(t^2)$$

所以，

$$\varphi_{Y_n}(t) = \left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n = \left(1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right)^n \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}}$$

第 6 节 多元随机变量

6.1 多维随机变量

定义 6.1 (n 维随机变量)

如果 $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$ 是定义在同一样本空间 $\Omega = \{\omega\}$ 上的 n 个随机变量，则称

$$X(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$$

为 n 维随机变量，或称为随机向量。

♣

定义 6.2 (联合分布函数)

对于 n 维随机变量 $X = (X_1, X_2, \ldots, X_n)$，和任意 n 个实数 $x_1, x_2, \ldots, x_n$，事件 $\{X_1 \leq x_1\}, \{X_2 \leq x_2\}, \ldots, \{X_n \leq x_n\}$ 同时发生的概率，定义为 n 维随机变量 X 的联合分布函数 (Joint Distribution Function)，记为 $F(x_1, x_2, \ldots, x_n)$：

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

♣

定理 6.1 (二维联合分布函数 $F(x, y)$ 的性质)

二维联合分布函数 $F(x, y)$ 具有如下性质：

1. 单调性：$F(x, y)$ 对其每一个变量都是单调不减的函数，即：

$$若 x_1 < x_2，则 F(x_1, y) \leq F(x_2, y)$$

$$若 y_1 < y_2，则 F(x, y_1) \leq F(x, y_2)$$

2. 有界性：对任意的 x 和 y，有 $0 \leq F(x, y) \leq 1$，且：

$$F(-\infty, y) = \lim_{x \to -\infty} F(x, y) = 0$$

$$F(x, -\infty) = \lim_{y \to -\infty} F(x, y) = 0$$

$$F(+\infty, +\infty) = \lim_{x \to +\infty, y \to +\infty} F(x, y) = 1$$

3. 右连续性：$F(x, y)$ 关于每个变量都是右连续的，即：

$$F(x + 0, y) = \lim_{\epsilon \to 0^+} F(x + \epsilon, y) = F(x, y)$$

$$F(x, y + 0) = \lim_{\delta \to 0^+} F(x, y + \delta) = F(x, y)$$

4. 非负性：对任意实数 a, b, c, d 满足 $a < b$ 和 $c < d$，有：$P(a < X \leq b, c < Y \leq d) \geq 0$

$$F(b, d) - F(a, d) - F(b, c) + F(a, c) \geq 0$$

♡

6.2 二维随机变量

6.2.1 二维离散随机变量

定义 6.3 (二维离散型随机变量及其联合分布列)

若二维随机变量 (X,Y) 的所有可能取值是有限对或可列无限多对数对 (x_i,y_j)（其中 $i=1,2,\dots;\ j=1,2,\dots$），则称 (X,Y) 为二维离散型随机变量。

称

$$p_{ij}=P(X=x_i,Y=y_j),\quad (i=1,2,\dots;j=1,2,\dots)$$

为二维离散型随机变量 (X,Y) 的联合分布列。

联合分布列可以用下表形式表示：

$X\setminus Y$	y_1	y_2	$\cdots$	y_j	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\cdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

联合分布列 p_{ij} 具有以下性质：

1. 非负性：$p_{ij}\geq 0$ 对所有的 i,j 成立。
2. 归一性：$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}p_{ij}=1$。

♣

定义 6.4 (边缘分布列)

设 (X,Y) 为一个二维离散型随机变量，其联合分布列为 $p_{ij}=P(X=x_i,Y=y_j)$。随机变量 X 和 Y 各自的概率分布称为边缘分布列。

- X 的边缘分布列：随机变量 X 的边缘分布列是 X 取各个可能值 x_i 的概率 $P(X=x_i)$。它可以表示为：

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_i & \cdots \\ P(X=x_1) & P(X=x_2) & \cdots & P(X=x_i) & \cdots \end{pmatrix}$$

  其中，对于任意 i，

$$P(X=x_i)=\sum_{j=1}^{\infty}P(X=x_i,Y=y_j)=\sum_{j=1}^{\infty}p_{ij}$$

- Y 的边缘分布列：随机变量 Y 的边缘分布列是 Y 取各个可能值 y_j 的概率 $P(Y=y_j)$。它可以表示为：

$$\begin{pmatrix} y_1 & y_2 & \cdots & y_j & \cdots \\ P(Y=y_1) & P(Y=y_2) & \cdots & P(Y=y_j) & \cdots \end{pmatrix}$$

  其中，对于任意 j，

$$P(Y=y_j)=\sum_{i=1}^{\infty}P(X=x_i,Y=y_j)=\sum_{i=1}^{\infty}p_{ij}$$

♣

6.2.2 二维连续型随机变量

定义 6.5 (二维连续型随机变量及其相关函数)

- 二维连续型随机变量与联合概率密度函数：如果对于二维随机变量 (X, Y) 的分布函数 $F(x, y)$，存在一个非负的二元函数 $p(x, y)$，使得对于任意实数 x, y，有：

$$F(x, y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} p(u, v)\, dv \right) du$$

则称 (X, Y) 为二维连续型随机变量，$p(x, y)$ 称为 (X, Y) 的联合概率密度函数。$p(x, y)$ 具有以下性质：

1. 非负性：$p(x, y) \geq 0$ 对所有 x, y 成立。
2. 归一性：$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y)\, dx\, dy = 1$。

如果 $F(x, y)$ 关于 x 和 y 的二阶混合偏导数存在，则有：

$$p(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

- 边缘分布函数：
  - 随机变量 X 的边缘分布函数 $F_X(x)$ 定义为：

$$F_X(x) = P(X \leq x) = F(x, +\infty) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} p(u, v)\, dv \right) du$$

  - 随机变量 Y 的边缘分布函数 $F_Y(y)$ 定义为：

$$F_Y(y) = P(Y \leq y) = F(+\infty, y) = \int_{-\infty}^{y} \left(\int_{-\infty}^{+\infty} p(u, v)\, du \right) dv$$

- 边缘概率密度函数：
  - 随机变量 X 的边缘概率密度函数 $p_X(x)$ 定义为：

$$p_X(x) = \frac{dF_X(x)}{dx} = \int_{-\infty}^{+\infty} p(x, y)\, dy$$

  因此，$F_X(x) = \int_{-\infty}^{x} p_X(u)\, du$。
  - 随机变量 Y 的边缘概率密度函数 $p_Y(y)$ 定义为：

$$p_Y(y) = \frac{dF_Y(y)}{dy} = \int_{-\infty}^{+\infty} p(x, y)\, dx$$

  因此，$F_Y(y) = \int_{-\infty}^{y} p_Y(v)\, dv$。

♣

例 6.1 设随机变量 (X, Y) 的联合概率密度函数为：

$$p(x, y) = \begin{cases} 6e^{-2x-3y}, & x > 0, y > 0 \\ 0, & \text{其它 (otherwise)} \end{cases}$$

求：(1) $P(X \leq 1, Y > 1)$ (2) $P(X > Y)$ (3) X 的边缘概率密度函数 $p_X(x)$ (4) Y 的边缘概率密度函数 $p_Y(y)$

解 **(1)** 计算 $P(X \leq 1, Y > 1)$：

$$\begin{aligned} P(X \leq 1, Y > 1) &= \iint_{x \leq 1, y > 1} p(x, y)\, dx\, dy = \int_0^1 \left(\int_1^{+\infty} 6e^{-2x-3y}\, dy \right) dx \\ &= \int_0^1 2e^{-2x}\, dx \cdot \int_1^{+\infty} 3e^{-3y}\, dy \\ &= F_X(1) \cdot (1 - F_Y(1)) = (1 - e^{-2})e^{-3} \quad (Y \sim Exp(3)) \end{aligned}$$

(2) 计算 $P(X>Y)$:

$$\begin{aligned}P(X>Y)&=\iint_{x>y}p(x,y)\,dx\,dy=\int_0^{+\infty}dx\int_0^x 6e^{-2x-3y}\,dy\\&=\int_0^{+\infty}2e^{-2x}(1-e^{-3x})\,dx=\int_0^{+\infty}(2e^{-2x}-2e^{-5x})\,dx\\&=\int_0^{+\infty}(2e^{-2x}-\frac{2}{5}\cdot 5e^{-5x})\,dx=1-\frac{2}{5}=\frac{3}{5}\end{aligned}$$

两种不同的计算方式:

$$P(X>Y)=\iint_{x>y}p(x,y)\,dx\,dy=\int_0^{+\infty}dx\int_0^x 6e^{-2x-3y}\,dy=\int_0^{+\infty}\left(\int_0^x 6e^{-2x-3y}\,dy\right)dx$$

$$P(X>Y)=\iint_{x>y}p(x,y)\,dx\,dy=\int_0^{+\infty}dy\int_y^{+\infty}6e^{-2x-3y}\,dx$$

(3) 计算 X 的边缘概率密度函数 $p_X(x)$: $6e^{-2x}e^{-3y}=2e^{-2x}\cdot 3e^{-3y}$

$$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)\,dy=\begin{cases}\int_0^{+\infty}6e^{-2x}e^{-3y}\,dy=2e^{-2x}, & x>0\\ 0, & x\le 0\end{cases}$$

(4) 计算 Y 的边缘概率密度函数 $p_Y(y)$:

$$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)\,dx=\begin{cases}\int_0^{+\infty}6e^{-2x}e^{-3y}\,dx=3e^{-3y}, & y>0\\ 0, & y\le 0\end{cases}$$

6.3 随机变量的独立性

定义 6.6 (随机变量的独立性)

设 n 维随机变量 $X=(X_1,X_2,\ldots,X_n)$ 的联合分布函数为 $F(x_1,x_2,\ldots,x_n)$,$F_{X_i}(x_i)$ 为 X_i 的边缘分布函数 $(i=1,2,\ldots,n)$。

如果对于任意 n 个实数 $x_1,x_2,\ldots,x_n$,有:

$$F(x_1,x_2,\ldots,x_n)=\prod_{i=1}^{n}F_{X_i}(x_i)$$

则称随机变量 $X_1,X_2,\ldots,X_n$ 相互独立。

上述定义等价于:

$$P(X_1\le x_1,X_2\le x_2,\ldots,X_n\le x_n)=\prod_{i=1}^{n}P(X_i\le x_i)$$

对于特定类型的随机变量,独立性有如下等价定义:

- 离散型随机变量的等价定义:
  如果 $X_1,X_2,\ldots,X_n$ 为离散型随机变量,它们相互独立的充要条件是,对于其联合概率质量函数 $P(X_1=x_1,\ldots,X_n=x_n)$ 和各边缘概率质量函数 $P(X_i=x_i)$,有:
  $$P(X_1=x_1,X_2=x_2,\ldots,X_n=x_n)=\prod_{i=1}^{n}P(X_i=x_i)$$
  对于所有可能的 $x_1,x_2,\ldots,x_n$ 均成立。
- 连续型随机变量的等价定义:
  如果 $X_1,X_2,\ldots,X_n$ 为连续型随机变量,它们相互独立的充要条件是,对于其联合概率密度函数

$p(x_1,\ldots,x_n)$ 和各边缘概率密度函数 $p_{X_i}(x_i)$，有：

$$p(x_1,x_2,\ldots,x_n)=\prod_{i=1}^{n}p_{X_i}(x_i)$$

在几乎所有点 $(x_1,x_2,\ldots,x_n)$ 处成立。

6.4 随机变量函数的期望

定义 6.7 (随机变量函数的期望)

对于 n 元随机变量 $(X_1,X_2,\ldots,X_n)$，若 $Z=g(X_1,X_2,\ldots,X_n)$ 是这些随机变量的函数，则 Z 的数学期望 $E(Z)$ 定义如下：

- 离散情形：如果 $(X_1,X_2,\ldots,X_n)$ 是 n 维离散型随机变量，其联合密度函数为 $P(X_1=x_1,\ldots,X_n=x_n)$，则 $Z=g(X_1,X_2,\ldots,X_n)$ 的数学期望为：
$$E(Z)=\sum_{x_1}\cdots\sum_{x_n}g(x_1,\ldots,x_n)P(X_1=x_1,\ldots,X_n=x_n)$$
- 连续情形：如果 $(X_1,X_2,\ldots,X_n)$ 是 n 维连续型随机变量，其联合概率密度函数为 $p(x_1,x_2,\ldots,x_n)$，则 $Z=g(X_1,X_2,\ldots,X_n)$ 的数学期望为：
$$E(Z)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(x_1,x_2,\ldots,x_n)p(x_1,x_2,\ldots,x_n)\,dx_1\cdots dx_n$$

定理 6.2 (随机变量和的期望与方差的性质)

令 $X_1,X_2,\ldots,X_n$ 为随机变量，$c_1,c_2,\ldots,c_n$ 为常数。

1. 期望的线性性质：
$$E(c_1X_1+c_2X_2+\cdots+c_nX_n)=c_1E(X_1)+c_2E(X_2)+\cdots+c_nE(X_n)$$
2. 独立随机变量乘积的期望：若 $X_1,X_2,\ldots,X_n$ 相互独立，则：
$$E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n)$$
3. 独立随机变量线性组合的方差：若 $X_1,X_2,\ldots,X_n$ 相互独立，则：
$$\mathrm{Var}(c_1X_1+c_2X_2+\cdots+c_nX_n)=c_1^2\mathrm{Var}(X_1)+c_2^2\mathrm{Var}(X_2)+\cdots+c_n^2\mathrm{Var}(X_n)$$

第 7 节 相关系数与条件分布

7.1 二元随机变量的协方差

定义 7.1 (协方差)

设 (X,Y) 是二元随机变量，则 $E[(X-E(X))(Y-E(Y))]$ 称为 X,Y 的协方差，记为 $\text{Cov}(X,Y)$。

$$\begin{aligned}\text{Cov}(X,Y) &= E[(X-E(X))(Y-E(Y))] \\ &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

♣

性质 协方差的性质：

$$\begin{aligned}\text{Cov}(X,a) &= 0 \\ \text{Cov}(X,Y) &= \text{Cov}(Y,X) \\ \text{Cov}(X+Y,Z) &= \text{Cov}(X,Z)+\text{Cov}(Y,Z) \\ \text{Cov}(c_1X+a,c_2Y+b) &= c_1c_2\cdot\text{Cov}(X,Y) \\ \text{Var}(X\pm Y) &= \text{Var}(X)+\text{Var}(Y)\pm 2\text{Cov}(X,Y) \\ [\text{Cov}(X,Y)]^2 &\le \text{Var}(X)\cdot\text{Var}(Y)=\sigma_X^2\cdot\sigma_Y^2\end{aligned}$$

7.2 二元随机变量的相关系数

定义 7.2 (相关系数)

相关系数是将随机变量做方差为 1 的标准化后的协方差，即：

$$\text{Corr}(X,Y) = \text{Cov}\left(\frac{X}{\sigma_X},\frac{Y}{\sigma_Y}\right)$$

对于二元随机变量 (X,Y)，若 $\text{Var}(X)\cdot\text{Var}(Y)>0$，则 X,Y 的相关系数定义为：

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\cdot\sqrt{\text{Var}(Y)}} = \frac{\text{Cov}(X,Y)}{\sigma_X\cdot\sigma_Y}$$

♣

性质 相关系数的性质与判断：

1. 取值范围：由

$$[\text{Cov}(X,Y)]^2 \le \text{Var}(X)\cdot\text{Var}(Y)=\sigma_X^2\cdot\sigma_Y^2$$

   可得：

$$-1\le\text{Corr}(X,Y)\le 1$$

2. 根据相关系数的符号判断相关性：
   - $\text{Corr}(X,Y)>0 \implies$ 正相关
   - $\text{Corr}(X,Y)<0 \implies$ 负相关
   - $\text{Corr}(X,Y)=0 \implies$ 不相关

相关系数反映的是随机变量之间在线性关系意义下的相关程度，所以也称（线性）相关系数。不相关指的是不存在线性相关的关系，相关系数并不能充分地表达非线性的相关关系。

命题

$\mathrm{Corr}(X, Y) = \pm 1$ 的充分必要条件是 X, Y 之间几乎处处有线性关系，即存在常数 a, b，使得 $P(Y = aX + b) = 1$。

7.3 常见多元分布

定义 7.3 (二维正态分布)

若二维随机变量 (X, Y) 服从参数为 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ 的二维正态分布，记为 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则其联合概率密度函数 $p(x, y)$ 为：

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

其中参数满足：$\mu_1, \mu_2 \in \mathbb{R}$，$\sigma_1 > 0, \sigma_2 > 0$，且 $|\rho| \leq 1$。

性质 若 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则由对称性可知其边缘分布：$X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$。也即二维正态分布的边缘分布即为一元正态分布。

定理 7.1 (二维正态分布密度函数的向量表示)

二维正态分布的密度函数：

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

令：

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

从而得到以下形式：

$$p(x_1, x_2) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

定义 7.4 (多元正态分布)

对于一般的 n 元正态随机变量 $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$，其联合密度函数为：

$$p(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

记为：$(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，或 $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$。

性质：多元正态分布的性质

1. 多元正态分布的边缘分布为一元正态分布。但是逆命题不成立，即边缘密度均为正态，联合分布未必是多元正态。
2. 多元正态分布 $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ 相互独立 $\iff$ $\mathbf{X}_1, \ldots, \mathbf{X}_n$ 两两不相关。

定理 7.2 (线性组合的性质与二元正态分布)

设随机变量 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则：

1. $(a_1X + b_1Y, a_2X + b_2Y)$ 也服从二元正态分布；
2. $aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)$。

定理 7.3 (独立正态随机变量和的分布)

设随机变量 $X_1, X_2, \ldots, X_n$ 相互独立，且 $X_k \sim N(\mu_k, \sigma_k^2)$，$k = 1, 2, \ldots, n$，则它们的和

$$X_1 + X_2 + \cdots + X_n \sim N\left(\sum_{k=1}^{n} \mu_k, \sum_{k=1}^{n} \sigma_k^2\right)$$

即和的期望等于期望的和，和的方差等于方差的和（在独立条件下）。♡

7.4 条件分布

7.4.1 离散随机变量的条件分布

定义 7.5 (离散随机变量的条件分布)

对一切使 $P(Y = y_j) = p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij} > 0$ 的 y_j，称

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

为给定 $Y = y_j$ 条件下 X 的分布列。在 $Y = y_j$ 条件下 X 的分布函数为

$$F(x|y_j) = \sum_{x_i \leq x} P(X = x_i | Y = y_j).$$

♣

例 7.1 设 X, Y 相互独立，$X \sim P(\lambda_1)$，$Y \sim P(\lambda_2)$，求 $X + Y = n$ 条件下 X 的分布。

解

$$P(X = k | X + Y = n) = \frac{P(X = k, Y = n - k)}{P(X + Y = n)} = \frac{\frac{\lambda_1^k}{k!} e^{-\lambda_1} \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2}}{\frac{(\lambda_1 + \lambda_2)^n}{n!} e^{-(\lambda_1 + \lambda_2)}}$$

$$= \binom{n}{k} \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n} = \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(1 - \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{n-k}$$

因此，

$$X | X + Y = n \sim b\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$$

命题 7.1 (泊松分布和二项分布的可加性)

- 二项分布的可加性：若 $X_1 \sim b(n_1, p), X_2 \sim b(n_2, p), \ldots, X_m \sim b(n_m, p)$，且 $X_1, X_2, \ldots, X_m$ 相互独立，则 $X_1 + X_2 + \cdots + X_m \sim b(n_1 + n_2 + \cdots + n_m, p)$。
- 泊松分布的可加性：若 $X_1 \sim P(\lambda_1), X_2 \sim P(\lambda_2), \ldots, X_m \sim P(\lambda_m)$，且 $X_1, X_2, \ldots, X_m$ 相互独立，则 $X_1 + X_2 + \cdots + X_m \sim P(\lambda_1 + \lambda_2 + \cdots + \lambda_m)$。

♠

7.4.2 连续随机变量的条件分布

定义 7.6 (连续随机变量的条件分布)

对一切使 $p_Y(y)>0$ 的 y，给定 $Y=y$ 条件下的条件分布函数与条件概率密度函数定义如下：

条件分布函数 $F(x|y)$ 的推导：

$$
\begin{aligned}
F(x|y) &= P(X\le x|Y=y) = \lim_{h\to 0} P(X\le x|y\le Y\le y+h) \\
&= \lim_{h\to 0}\frac{P(X\le x, y\le Y\le y+h)}{P(y\le Y\le y+h)} = \lim_{h\to 0}\frac{\int_{-\infty}^{x}\int_{y}^{y+h} p(u,v)\,dv\,du}{\int_{y}^{y+h} p_Y(v)\,dv} \\
&= \lim_{h\to 0}\frac{\int_{-\infty}^{x} p(u,y+c_2h)\cdot h\,du}{p_Y(y+c_1h)\cdot h} \quad (\text{由积分中值定理，其中} 0<c_1,c_2<1) \\
&= \frac{\int_{-\infty}^{x} p(u,y)\,du}{p_Y(y)} = \int_{-\infty}^{x}\frac{p(u,y)}{p_Y(y)}\,du
\end{aligned}
$$

从而条件概率密度函数 $p(x|y)$（或 $p_{X|Y}(x|y)$）：

$$
p(x|y) = \frac{p(x,y)}{p_Y(y)} \quad \text{或} \quad p_{X|Y}(x|y)
$$

♣

例 7.2 设 (X,Y) 服从 $G=\{(x,y);x^2+y^2\le 1\}$ 上的均匀分布，求 $p(x|y)$。

解 联合概率密度函数 $p(x,y)$ 为：

$$
p(x,y) = \begin{cases} \frac{1}{\pi}, & x^2+y^2\le 1 \\ 0, & \text{otherwise} \end{cases}
$$

边缘概率密度函数 $p_Y(y)$ 的计算：

$$
p_Y(y) = \int_{-\infty}^{+\infty} p(x,y)\,dx = \int_{-\sqrt{1-y^2}}^{+\sqrt{1-y^2}} \frac{1}{\pi}\,dx = \frac{1}{\pi}[x]_{-\sqrt{1-y^2}}^{+\sqrt{1-y^2}} = \frac{2}{\pi}\sqrt{1-y^2}\cdot I_{-1\le y\le 1}
$$

条件概率密度函数 $p(x|y)$ 的计算：

$$
p(x|y) = \frac{p(x,y)}{p_Y(y)} = \frac{1}{2\sqrt{1-y^2}}, \quad |x|\le\sqrt{1-y^2}, |y|<1
$$

定理 7.4 (连续随机变量的条件分布计算公式)

1. 连续场合下的全概率公式：由 $p(x,y)=p_X(x)p(y|x)$ 可推导出边缘概率密度函数 $p_Y(y)$：
$$
p_Y(y) = \int_{-\infty}^{+\infty} p_X(x)p(y|x)\,dx
$$
2. 连续场合下的贝叶斯公式：由条件概率密度定义 $p(x|y)=\frac{p(x,y)}{p_Y(y)}$ 可推导出：
$$
p(x|y) = \frac{p_X(x)p(y|x)}{\int_{-\infty}^{+\infty} p_X(x)p(y|x)\,dx}
$$

♡

第 8 节 条件期望与随机变量函数的分布

8.1 条件期望

定义 8.1 (条件数学期望)

首先，条件分布 $X|Y=y$ 定义为：

$$X|Y=y \sim \begin{cases} x_i \to P(X=x_i|Y=y), & (X,Y) \text{ 为离散随机变量} \\ x \to p(x|y), & (X,Y) \text{ 为连续随机变量} \end{cases}$$

基于此，条件数学期望 $E(X|Y=y)$ 定义为：

$$E(X|Y=y) = \begin{cases} \sum_i x_i P(X=x_i|Y=y), & (X,Y) \text{ 为离散随机变量} \\ \int_{-\infty}^{+\infty} xp(x|y)\,dx, & (X,Y) \text{ 为连续随机变量} \end{cases}$$

♣

例 8.1 $X \sim Ge\left(\frac{1}{4}\right)$，求：$E(X|X>3)$

解：

$$\begin{aligned} E(X|X>3) &= \sum_{k=1}^{+\infty}(k+3)\cdot P(X=k+3|X>3) = \sum_{k=1}^{+\infty}(k+3)\cdot P(X=k) \quad (\text{利用几何分布的无记忆性}) \\ &= \sum_{k=1}^{+\infty} k\cdot P(X=k) + \sum_{k=1}^{+\infty} 3\cdot P(X=k) = E(X) + 3\cdot\sum_{k=1}^{+\infty} P(X=k) = E(X)+3 = 4+3 = 7 \end{aligned}$$

注 $E(X|Y)$ 是一个随机变量：$E(X|Y)=g(Y)$

$$\text{离散型的情形：} E(X|Y) \sim \begin{pmatrix} E(X|Y=y_1) & E(X|Y=y_2) & \dots & E(X|Y=y_n) & \dots \\ P(Y=y_1) & P(Y=y_2) & \dots & P(Y=y_n) & \dots \end{pmatrix}$$

连续型的情形：$\mathbf{Z}=E(X|Y)\sim z=E(X|Y=y)$ 时，$p_Y(y)\to p_Z(z)$

定理 8.1 (重期望公式)

$$E(X) = E(E(X|Y))$$

其中，$E(E(X|Y))$ 的计算方式为：

$$E(E(X|Y)) = \begin{cases} \sum_{k=1}^{\infty} E(X|Y=y_k)\cdot P(Y=y_k), & \text{若} Y \text{ 为离散型随机变量} \\ \int_{-\infty}^{+\infty} E(X|Y=y)\cdot p_Y(y)\,dy, & \text{若} Y \text{ 为连续型随机变量} \end{cases}$$

♡

例 8.2 口袋里有编号 $1,2,\dots,n$ 的 n 个球，任取 1 个，若为 1 号则得 1 分停止，若为 i 号 $(i\geq 2)$，则得到 i 分，放回继续摸球，求总得分的期望。

解 设 X 为总得分，Y 为第一次抽到的号码（得分）。则：

$$P(Y=k)=\frac{1}{n}, \quad (k=1,\dots,n)$$

条件期望为：$E(X|Y=1)=1, E(X|Y=k,k\neq 1)=k+E(X)$

根据全期望公式：

$$\begin{aligned} E(X) = E(E(X|Y)) &= \sum_{k=1}^{n} E(X|Y=k)P(Y=k) \\ &= E(X|Y=1)P(Y=1) + \sum_{k=2}^{n} E(X|Y=k)P(Y=k) \\ &= 1 \cdot \frac{1}{n} + \sum_{k=2}^{n}(k+E(X))\frac{1}{n} = \frac{1}{n} + \frac{1}{n}\sum_{k=2}^{n}(k+E(X)) \qquad \Rightarrow E(X) = \frac{n(n+1)}{2} \end{aligned}$$

命题 8.1 (条件期望与最佳预报)

- 最小二乘：常数预报

$$\min_c E(Y-c)^2 \implies c = E(Y)$$

- 最小二乘：线性预报
若要最小化 $E(Y-(aX+b))^2$，则最佳的 a 和 b 为：

$$a = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}, \quad b = E(Y) - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}E(X)$$

(即 $b = E(Y) - aE(X)$)

- 非线性的最小二乘最佳预报
对于二元随机变量 (X,Y)，假设 $\mathrm{Var}(X)$、$\mathrm{Var}(Y)$ 存在。令 $\phi(X) = E(Y|X)$，则 $\phi(X)$ 是 Y 的最佳预报，即它在所有函数 $\psi(X)$ 中使得均方误差 $E(Y-\psi(X))^2$ 最小：

$$E(Y-\phi(X))^2 = \min_\psi E(Y-\psi(X))^2$$

8.2 随机变量函数的分布

定理 8.2 (连续随机变量的函数的分布)

设 n 维随机变量 $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ 的密度函数为 $p_X(x_1,x_2,\ldots,x_n)$，n 元函数 $\{g_i(x_1,x_2,\ldots,x_n)\}_{i=1}^n$ 满足条件：

1. 存在唯一的反函数 $x_i(y_1,y_2,\ldots,y_n)$，即存在方程组 $g_i(x_1,x_2,\ldots,x_n) = y_i$ 的唯一实数解。
2. $g_i(x_1,x_2,\ldots,x_n)$ 与 $x_i(y_1,y_2,\ldots,y_n)$ 都连续。
3. 存在连续的偏导数 $\dfrac{\partial x_i}{\partial y_j}$，$\dfrac{\partial g_i}{\partial x_j}$，记 $J = \det\begin{pmatrix} \dfrac{\partial x_1}{\partial y_1} & \cdots & \dfrac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial x_n}{\partial y_1} & \cdots & \dfrac{\partial x_n}{\partial y_n} \end{pmatrix}$。(注：$J$ 是反函数变换的雅可比行列式。)

则 n 维随机向量 $\mathbf{Y} = (Y_1,\ldots,Y_n)^T = (g_1(X_1,\ldots,X_n),\ldots,g_n(X_1,\ldots,X_n))^T$ 的密度函数为

$$p_Y(y_1,\ldots,y_n) = \begin{cases} p_X(x_1(y_1,\ldots,y_n),\ldots,x_n(y_1,\ldots,y_n)) \cdot |J|, & \text{当}(y_1,\ldots,y_n)\text{ 使得}\mathbf{y} = g(\mathbf{x})\text{ 有解时} \\ 0, & \text{其他} \end{cases}$$

注

1. 如果反函数不唯一，即方程组 $g_i(x_1,x_2,\ldots,x_n) = y_i \quad (i=1,2,\ldots,n)$ 有多个解 $x_i^{(s)}(y_1,\ldots,y_n), s = 1,2,\ldots$，则此时 y-空间中一个点对应于 x-空间中多个点。将 x-空间分成若干部分，使 y-空间与 x-空间的每部分成一一对应，于是 Y 取值于 y-空间中某集合的概率就等于 X 取值于 x-空间每部分对应集合的概率之和。这时，$p_Y(y_1,\ldots,y_n)$ 的计算公式修改为：
$$p_Y(y_1,\ldots,y_n) = \begin{cases} \sum_s p_X(x_1^{(s)}(y_1,\ldots,y_n),\ldots,x_n^{(s)}(y_1,\ldots,y_n)) \cdot |J^{(s)}|, & \text{当}(y_1,\ldots,y_n)\text{ 使得}\mathbf{y}=g_i(\mathbf{x})\text{ 有解时} \\ 0, & \text{其他} \end{cases}$$
其中 $J^{(s)}$ 是对应于第 s 个反函数解 $x_i^{(s)}$ 的雅可比行列式。
2. 如果只有 $m<n$ 个函数 $g_i(x_1,x_2,\ldots,x_n) \quad (i=1,2,\ldots,m)$，则可增补定义 $g_j(x_1,x_2,\ldots,x_n)=x_j \quad (j=m+1,\ldots,n)$，得到 n 维随机向量
$$\mathbf{Y}_n = (g_1(x_1,\ldots,x_n),\ldots,g_m(x_1,\ldots,x_n),x_{m+1},\ldots,x_n)^T$$
的密度函数 $p_{\mathbf{Y}_n}(y_1,\ldots,y_n)$。
然后，所求的 m 维随机向量 $\mathbf{Y}=(g_1,\ldots,g_m)^T$ 的密度函数 $p_{\mathbf{Y}}(y_1,\ldots,y_m)$ 可以通过对 $p_{\mathbf{Y}_n}$ 积分得到（即求边缘密度）：
$$p_{\mathbf{Y}}(y_1,\ldots,y_m) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} p_{\mathbf{Y}_n}(y_1,\ldots,y_m,y_{m+1},\ldots,y_n)\,dy_{m+1}\ldots dy_n$$

例 8.3 X 和 Y 相互独立，且均服从参数为 1 的指数分布，求 $U=X+Y, V=\dfrac{X}{X+Y}$ 的联合密度函数。

解 设：
$$U=X+Y,\quad V=\frac{X}{X+Y} \implies \begin{cases} x=uv \\ y=u-uv \end{cases}$$
则雅可比矩阵为：
$$J=\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -u$$
所以，联合密度函数为：
$$p_{UV}(u,v)=p_{XY}(uv,u-uv)\cdot|J| = p_X(uv)\cdot p_Y(u-uv)\cdot|u| = e^{-uv}\cdot e^{-(u-uv)}\cdot u = ue^{-u}$$

例 8.4 设 (X,Y) 的密度函数为 $p_{XY}(x,y)$，$U=\sqrt{X^2+Y^2}$，$V=\dfrac{X}{Y}$，求 $p_{UV}(u,v)$。

解 设：
$$U=\sqrt{X^2+Y^2},\quad V=\frac{X}{Y}$$
则
$$\begin{cases} x^{(1)} = \dfrac{uv}{\sqrt{1+v^2}} \\ y^{(1)} = \dfrac{u}{\sqrt{1+v^2}} \end{cases} \Rightarrow \begin{cases} x^{(2)} = -x^{(1)} \\ y^{(2)} = -y^{(1)} \end{cases}$$
雅可比行列式为：
$$|J^{(1)}| = |J^{(2)}| = \frac{u}{1+v^2}$$
所以联合密度函数为：
$$p_{UV}(u,v) = \begin{cases} \dfrac{u}{1+v^2}\left[p_{XY}\left(\dfrac{uv}{\sqrt{1+v^2}},\dfrac{u}{\sqrt{1+v^2}}\right) + p_{XY}\left(\dfrac{-uv}{\sqrt{1+v^2}},\dfrac{-u}{\sqrt{1+v^2}}\right)\right], & u\ge 0 \\ 0, & u<0 \end{cases}$$

例 8.5 设 (X,Y) 的密度函数为 $p_{XY}(x,y)$，求 $Z=X+Y$ 的分布。

解 设:

$$Z=X+Y,\quad W=Y \quad \Rightarrow \quad \begin{cases} x=z-w \\ y=w \end{cases}$$

雅可比矩阵为:

$$J=\frac{\partial(x,y)}{\partial(z,w)}=\begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix}=1$$

所以联合密度函数为:

$$p_{ZW}(z,w)=p_{XY}(z-w,w),\quad p_Z(z)=\int_{-\infty}^{+\infty}p_{ZW}(z,w)dw=\int_{-\infty}^{+\infty}p_{XY}(z-w,w)dw$$

若 X,Y 相互独立，则有:

$$\int_{-\infty}^{+\infty}p_{XY}(z-w,w)dw=\int_{-\infty}^{+\infty}p_X(z-w)\cdot p_Y(w)dw \quad \text{(卷积公式)}$$

第 9 节 极限定理

9.1 大数定律

定理（切比雪夫不等式）

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

定理 9.1 (伯努利大数定律)

设 μ_n 为 n 重伯努利试验中事件 A 发生的次数，p 为每次试验中 A 发生的概率。则对于任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty}\Pr\left(\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right)=1.$$

定理 9.2 (大数定律的一般形式)

设有一个随机变量序列 $\{X_n\}$，如果它具有如下性质：

$$\lim_{n\to\infty}\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}E(X_k)\right|<\varepsilon=1,$$

则称该随机变量序列 $\{X_n\}$ 服从大数定律。

定义 9.1 (依概率收敛)

设 $\{X_n\}$ 为一个随机变量序列，X 为一个随机变量，如果对任意的 $\varepsilon>0$，有：

$$\lim_{n\to\infty}\Pr(|X_n-X|<\varepsilon)=1,$$

则称 $\{X_n\}$ 依概率收敛于 X，记作 $X_n\xrightarrow{P}X$。

定理（切比雪夫大数定律）

设 $X_1,X_2,\ldots,X_k,\ldots$ 是两两不相关的随机变量序列，方差有界，则对于任意给定的正数 ε，有：

$$\lim_{n\to\infty}\Pr\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}E(X_k)\right|<\varepsilon\right)=1.$$

定理（马尔科夫大数定律）

若 $\frac{1}{n^2}\mathrm{Var}\left(\sum_{k=1}^{n}X_k\right)\to 0$ （马尔科夫条件），则 $\{X_k\}$ 服从大数定律，

$$\lim_{n\to\infty}\Pr\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}E(X_k)\right|<\varepsilon\right)=1.$$

定理（辛钦大数定律）

设 $\{X_n\}$ 是独立同分布的随机变量序列。如果其期望

$$E(X_1)=\mu,$$

则 $\forall\varepsilon>0$，当 $n\to\infty$ 时，都会有：

$$P\left(\left|\frac{X_1+X_2+\cdots+X_n}{n}-\mu\right|<\varepsilon\right)\to 1.$$

9.2 中心极限定理

定理 9.3 (林德伯格—勒维中心极限定理)

设 $\{X_n\}$ 是独立同分布的随机变量序列。如果其期望 $E(X_1)=\mu$，方差 $Var(X_1)=\sigma^2$，则对于每一个固定的 y 有：

$$\lim_{n\to\infty} \Pr\left(\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\cdot\sqrt{n}}\le y\right)=\Phi(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}e^{-t^2/2}\,dt.$$

♡

定理 9.4 (李雅普诺夫中心极限定理)

设 $\{X_n\}$ 为独立随机变量序列，若存在 $\delta>0$，满足：

$$\lim_{n\to\infty}\frac{1}{\sqrt{\sum_{k=1}^{n}\mathrm{Var}(X_1+\cdots+X_n)}^{2+\delta}}\sum_{k=1}^{n}E\left(|X_k-E(X_k)|^{2+\delta}\right)=0,$$

则：

$$\frac{\sum_{k=1}^{n}(X_k-E(X_k))}{\sqrt{\sum_{k=1}^{n}\mathrm{Var}(X_k)}}\xrightarrow{n\to\infty}N(0,1),$$

即：

$$X_1+X_2+\cdots+X_n\sim N\left(\sum_{k=1}^{n}E(X_k),\sum_{k=1}^{n}\mathrm{Var}(X_k)\right).$$

由

$$\sum_{k=1}^{n}E(X_k)=n\mu,\quad \sum_{k=1}^{n}Var(X_k)=n\sigma^2.\ \Rightarrow\ X_1+X_2+\cdots+X_n\sim N(n\mu,n\sigma^2).$$

♡

例 9.1 设系统由 100 个相互独立的部件组成，运行时间每个部件损坏的概率为 0.1，至少有 85 个部件完成好是系统才能正常工作，求系统正常工作的概率。

解 正常工作部件的数目 X 服从二项分布 $b(100,0.9)$，

$$E(X)=90,\quad Var(X)=100\times 0.9\times 0.1=9,$$

所以 $X\sim N(90,9)$，

$$\frac{X-90}{3}\sim N(0,1).$$

求系统正常工作概率：

$$P(\text{正常工作})=P(X\ge 85)=P\left(\frac{X-90}{3}\ge\frac{85-90}{3}\right)=1-\Phi\left(\frac{-5}{3}\right)=\Phi\left(\frac{5}{3}\right).$$

例 9.2 设系统由一些相互独立的部件组成，运行时间每个部件损坏的概率为 0.1，至少有 80 个部件完成是系统才能正常工作，问部件数 n 至少为多少才使得系统正常工作的概率不小于 0.95。

解 正常工作部件的数目 X 服从二项分布 $b(n,0.9)$，

$$X\sim N(0.9n,0.09n).$$

求系统正常工作概率：

$$P(X\ge 0.8n)=P\left(\frac{X-0.9n}{0.3\sqrt{n}}\ge\frac{0.8n-0.9n}{0.3\sqrt{n}}\right)\approx 1-\Phi\left(\frac{-0.1n}{0.3\sqrt{n}}\right)=\Phi\left(\frac{\sqrt{n}}{3}\right).$$

因此：

$$\frac{\sqrt{n}}{3}\ge u_{0.05}=1.645\Rightarrow n\ge 25.$$

第二部分

统计部分

第 10 节 统计学基本概念

10.1 总体与样本

定义 10.1 (总体和样本)

- 总体：一个统计问题研究对象的整体。构成总体的每个成员称为个体。简单说总体即为分布。
- 样本：从总体中随机抽样的部分个体组成的集合为样本，样本中的个体称为样品，样品的个数称为样本容量或样本量。

  简单随机抽样：(1) 样本具有随机性。(2) 样本之间相互独立。

10.2 统计量

定义 10.2 (统计量)

设 $X_1, X_2, \ldots, X_n$ 为从某总体现象中获取的样本，不含有任何未知参数的样本函数，样本函数的统计量 $T = T(X_1, X_2, \ldots, X_n)$，即称为统计量。

定义 10.3 (样本均值)

设 $x_1, x_2, \ldots, x_n$ 为某总体的简单随机样本，样本均值为

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n},$$

称为样本均值。

定理 10.1 (样本均值的期望与方差)

1. 若总体分布为 $N(\mu, \sigma^2)$，则样本均值 $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$。
2. 若总体分布不是正态分布或未知，$E(X) = \mu, \mathrm{Var}(X) = \sigma^2$，则当 n 较大时，$\bar{x}$ 的近似分布为 $N(\mu, \frac{\sigma^2}{n})$，通常记作 $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$。

定义 10.4 (样本方差)

样本方差为：

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

定理 10.2 (样本方差的期望)

设总体 X 具有二阶矩，即 $E(X) = \mu, Var(X) = \sigma^2 < +\infty$，$x_1, x_2, \ldots, x_n$ 为该总体得到的样本，$\bar{x}$ 和 s^2 分别为样本均值和样本方差，则：

$$E(\bar{x}) = \mu, \quad Var(\bar{x}) = \frac{\sigma^2}{n}, \quad E(s^2) = \sigma^2. \quad \text{是无偏估计。}$$

定义（样本原点矩和中心矩）

样本 k 阶原点矩：

$$a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k,$$

样本 k 阶中心矩：

$$b_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k.$$

♣

10.2.1 次序统计量

例 10.1 $X_1, X_2, \ldots, X_n$ 独立同分布，分布函数 $F(x)$，密度函数 $p(x)$，将这 n 个随机变量做升序排列 $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$。求 $X_{(k)}$ 的分布，$F_k(x), p_k(x)$。

解 考虑 $X_{(k)}$ 落在区间 $[x, x+\Delta x]$ 的概率

$$F_k(x+\Delta x) - F_k(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1} \cdot [F(x+\Delta x) - F(x)] \cdot [1 - F(x+\Delta x)]^{n-k}$$

令 $\Delta x \to 0$，则得到 $X_{(k)}$ 的概率密度函数 $p_k(x)$ 为：

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1} \cdot p(x) \cdot [1 - F(x)]^{n-k}$$

注 $X_{(k)}$ 被称为（第 k 个）次序统计量，最小次序统计量 $x_{(1)}$，最大次序统计量 $x_{(n)}$。

10.3 三大统计分布

定义 10.5 (三大统计分布)

1. 卡方分布：设 $X_1, X_2, \ldots, X_n$ 独立同分布，服从 $N(0,1)$ 则
   $$Y = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n),$$
   称为自由度为 n 的卡方分布。
2. **t** 分布：设 $X_1 \sim N(0,1), X_2 \sim \chi_n^2, X_1, X_2$ 相互独立，则
   $$t = \frac{X_1}{\sqrt{X_2/n}} \sim t(n),$$
   称为自由度为 n 的 t 分布。
3. **F** 分布：设 X_1, X_2 相互独立，$X_1 \sim \chi_m^2, X_2 \sim \chi_n^2$，则
   $$F = \frac{X_1/m}{X_2/n} \sim F(m,n),$$
   称为自由度为 m 与 n 的 F 分布。

♣

定理 10.3 (统计抽样定理)

设 $x_1, x_2, \ldots, x_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为：

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

则有：

1. $\bar{x}$ 与 s^2 相互独立；
2. $\bar{x} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$；

3. $\dfrac{(n-1)\cdot s^2}{\sigma^2} \sim \chi^2(n-1)$。

♡

第 11 节 参数点估计

11.1 参数点估计

定义 11.1 (参数点估计)

设 $X_1, X_2, \ldots, X_n$ 是来自某总体的样本，利用这些样本估计总体分布的参数 θ。构造适当的统计量

$$\hat{\theta} = \theta(X_1, X_2, \ldots, X_n),$$

称为参数 θ 的点估计。♣

11.1.1 矩估计

定义 11.2 (矩估计法)

矩的理论表达式为参数的函数。以样本矩的某一函数代替总体矩的同一函数来构造估计量的方法称为矩估计法。

这里的“矩”可以是原点矩、中心矩，以及样本方差等，为了计算简单，尽可能用低阶矩。

期望表达式为：

$$E(X^k) = f_k(\theta_1, \theta_2, \ldots, \theta_m) \Rightarrow f_k(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m) = \frac{x_1^k + x_2^k + \cdots + x_n^k}{n}, \quad k = 1, \ldots, m.$$

♣

例 11.1 对于均匀分布总体 $U(a, b)$，估计参数 a 和 b。

解 总体期望与方差分别为

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

从而有

$$\frac{a+b}{2} = \bar{x}, \quad \frac{(b-a)^2}{12} = s^2.$$

解得

$$\hat{a} = \bar{x} - \sqrt{3}s, \quad \hat{b} = \bar{x} + \sqrt{3}s.$$

得到参数 a, b 的估计量分别为

$$\hat{a} = \bar{x} - \sqrt{3}s, \quad \hat{b} = \bar{x} + \sqrt{3}s.$$

例 11.2 设随机变量 $X_1, X_2, \ldots, X_m$ 是来自二项分布总体 $X \sim b(n, p)$ 的样本，用矩估计法估计参数 n, p。

解 总体的期望与方差为：

$$E(X) = np, \quad Var(X) = np(1-p).$$

样本均值和方差分别为：

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_m}{m}, \quad S^2 = \frac{1}{m-1}\sum_{k=1}^{m}(X_k - \bar{X})^2.$$

根据矩估计法，得到：

$$\bar{X} = np, \quad S^2 = np(1-p).$$

因此：

$$\hat{p} = 1 - \frac{S^2}{\bar{X}}, \quad \hat{n} = \frac{\bar{X}^2}{\bar{X} - S^2}.$$

例 11.3 $X_1, X_2, \ldots, X_n$ 是来自均匀总体 $X \sim U(-a, a)$ 的样本，用矩估计法估计参数 a。

解 由样本方差可得到矩估计量：

$$\mathrm{Var}(X) = \frac{a^2}{3}, \frac{a^2}{3} = s^2 \implies \hat{a} = \sqrt{3}s$$

$$E(s^2) = \mathrm{Var}(X) = \frac{a^2}{3} \to E(3s^2) = a^2 \overset{?}{\implies} E(\sqrt{3s^2}) = a$$

我们有：

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 > 0 \implies E(X) < \sqrt{E(X^2)}$$

从而

$$E(3s^2) > E(\sqrt{3s^2}) = E(\sqrt{3}s) \implies E(\hat{a}) = E(\sqrt{3}s) < \sqrt{E(3s^2)} = a$$

因此这不是一个无偏估计。

11.1.2 极大似然估计

定义 11.3 (极大似然估计)

设总体的概率函数为 $p(x;\theta), \theta \in \Theta$，其中 θ 是一组未知参数，Θ 称为参数空间，即参数 θ 可能取值的集合。$x_1, x_2, \ldots, x_n$ 是来自该总体的样本，则样本的联合概率函数是关于 θ 的函数，用 $L(\theta; x_1, x_2, \ldots, x_n)$ 表示，简记为 $L(\theta)$。

$$L(\theta) = L(\theta; x_1, x_2, \ldots, x_n) = \prod_{k=1}^{n} p(x_k; \theta),$$

称为样本的似然函数，如上式所示。

若某统计量 $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ 满足

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(\theta),$$

则称 $\hat{\theta}$ 为 θ 的极大似然估计。

对数似然函数 $\ln(L(\theta))$：由于 $\ln x$ 为单调函数，使得 $\ln(L(\theta))$ 与 $L(\theta)$ 达到最大的 θ 相同，常常利用对数似然函数求解极大似然估计。 ♣

例 11.4 总样本来自指数分布 $Exp(\lambda)$，设 $x_1, x_2, \ldots, x_n$ 为来自该总体的样本，求参数 λ 的最大似然估计。

解 似然函数 $L(\lambda; x_1, x_2, \ldots, x_n)$ 为：

$$L(\lambda; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda(x_1 + x_2 + \cdots + x_n)}.$$

对似然函数取对数：

$$\ln L(\lambda) = n \ln \lambda - \lambda(x_1 + x_2 + \cdots + x_n).$$

最大化 $\ln L(\lambda)$：

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \ln L(\lambda) = \frac{n}{\lambda} - (x_1 + x_2 + \cdots + x_n) = 0.$$

解得：

$$\hat{\lambda} = \frac{1}{(x_1 + x_2 + \cdots + x_n)/n} = \frac{1}{\bar{x}}.$$

因此，得到参数 λ 的极大似然估计量 $\hat{\lambda} = \dfrac{1}{\bar{x}}$。

例 11.5 总样本来自泊松分布 $P(\lambda)$，设 $x_1, x_2, \ldots, x_n$ 为来自该总体的样本，求参数 λ 的最大似然估计。

解 似然函数 $L(\lambda; x_1, x_2, \ldots, x_n)$ 为：

$$L(\lambda; x_1, x_2, \ldots, x_n) = \prod_{k=1}^{n} \frac{\lambda^{x_k} e^{-\lambda}}{x_k!} = \frac{1}{x_1! x_2! \ldots x_n!} \lambda^{x_1+x_2+\cdots+x_n} e^{-n\lambda}.$$

对似然函数取对数：

$$\ln L(\lambda) = C + (x_1 + x_2 + \cdots + x_n) \ln \lambda - n\lambda.$$

最大化 $\ln L(\lambda)$：

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \ln L(\lambda) = \frac{1}{\lambda}(x_1 + x_2 + \cdots + x_n) - n = 0.$$

解得：

$$\hat{\lambda} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

因此，得到参数 λ 的最大似然估计量为：

$$\hat{\lambda} = \bar{x}.$$

例 11.6 设元件的寿命从指数分布 $p(x) = \lambda e^{-\lambda x}$ $(x > 0)$。指定来一个时刻 $T > 0$，抽样试验进行到元件失败或时刻 T，求参数 λ 的最大似然估计。

解 若 n 个样本值中有 r 个是元件的实际寿命，则设寿命小于 T 的 r 个观测值为 $x_1, x_2, \ldots, x_r$，对应的 r 个 y_k 取值为 T，每一个发生概率为 $P(X \geq T) = e^{-\lambda T}$。似然函数 $L(\lambda; y_1, y_2, \ldots, y_n)$ 为：

$$L(\lambda; y_1, y_2, \ldots, y_n) = \lambda^r e^{-\lambda(x_1+x_2+\cdots+x_r)} e^{-\lambda(n-r)T}.$$

对似然函数取对数：

$$\ln L(\lambda) = r \ln \lambda - \lambda(x_1 + x_2 + \cdots + x_r + (n-r)T).$$

对 $\ln L(\lambda)$ 求导：

$$\frac{d \ln L(\lambda)}{d\lambda} = \frac{r}{\lambda} - (x_1 + x_2 + \cdots + x_r + (n-r)T) = 0.$$

因此，得到参数 λ 的最大似然估计量为：

$$\hat{\lambda} = \frac{r}{x_1 + x_2 + \cdots + x_r + (n-r)T}.$$

例 11.7 设 $X_1, X_2, \ldots, X_n$ 为来自均匀分布总体 $U(a,b)$ 的样本，试利用极大似然估计给出参数 a, b 的估计量。

解 设 $x_1, x_2, \ldots, x_n$ 为样本观测值，总体分布的密度函数为：

$$f(x; a, b) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

似然函数为：

$$L(a,b) = \prod_{k=1}^{n} f(x_k; a, b) = \left(\frac{1}{b-a}\right)^n, \quad a \leq x_k \leq b, k = 1, 2, \ldots, n.$$

显然，$L(a,b)$ 关于 a 是单调递增函数，关于 b 是单调递减函数。为了使 $L(a,b)$ 最大，必须使 $b-a$ 达到最小，即使 b 尽可能小，a 尽可能大。

考虑到 $a \leq x_1, x_2, \ldots, x_n \leq b$，则

$$a = \min\{x_1, x_2, \ldots, x_n\}, \quad b = \max\{x_1, x_2, \ldots, x_n\}.$$

因此，参数 a, b 的极大似然估计量为：

$$\hat{a} = \min(X_1, X_2, \ldots, X_n), \quad \hat{b} = \max(X_1, X_2, \ldots, X_n).$$

11.2 点估计量的评价标准

定义 11.4 (点估计量的评价标准)

1. 相合性：估计量随着样本量的增加而逼近参数的真实值
   设 $\theta \in \Theta$ 为未知参数，$\hat{\theta}_n = \hat{\theta}_n(x_1, x_2, \ldots, x_n)$ 是由 n 个样本得到的 θ 的一个估计量，n 是样本容量，若对任意 $\epsilon > 0$，有
   $$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0$$
   则称 $\hat{\theta}_n$ 为参数 θ 的相合估计。
2. 无偏性：保证没有系统偏差
   设 $\theta \in \Theta$ 为未知参数，$\hat{\theta}_n = \hat{\theta}_n(x_1, x_2, \ldots, x_n)$ 是 θ 的一个估计量，若对任意 $\theta \in \Theta$ 有
   $$E(\hat{\theta}_n) = \theta$$
   则称 $\hat{\theta}_n$ 为参数 θ 的无偏估计。
3. 有效性：希望估计围绕参数的波动越小越好
   设 $\hat{\theta}_1, \hat{\theta}_2$ 是参数 θ 的无偏估计，若对任意 $\theta \in \Theta$ 有
   $$Var(\hat{\theta}_1) \le Var(\hat{\theta}_2)$$
   且至少有一个 $\theta \in \Theta$ 使得上式不等号成立，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效。

♣

例 11.8 设总体 X 在区间 $[0, \theta]$ 上服从均匀分布，其中 $\theta > 0$ 未知。从总体中抽取样本 X_1, X_2, X_3, X_4，则下列统计量均为 θ 的无偏估计量，比较它们的有效性：

$$\begin{aligned}
\hat{\theta}_1 &= \frac{1}{2}(X_1 + X_2 + X_3 + X_4) \\
\hat{\theta}_2 &= 2X_1 + X_2 - X_3 \\
\hat{\theta}_3 &= X_1 + X_2 + X_3 - X_4
\end{aligned}$$

解

$$\mathrm{Var}(\hat{\theta}_1) = \mathrm{Var}\left(\frac{1}{2}(X_1 + X_2 + X_3 + X_4)\right) = \frac{1}{4}\cdot(\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3) + \mathrm{Var}(X_4)) = \frac{\theta^2}{12}$$

$$\mathrm{Var}(\hat{\theta}_2) = \mathrm{Var}(2X_1 + X_2 - X_3) = 4\cdot\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3) = \frac{6\theta^2}{12} = \frac{\theta^2}{2}$$

$$\mathrm{Var}(\hat{\theta}_3) = \mathrm{Var}(X_1 + X_2 + X_3 - X_4) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3) + \mathrm{Var}(X_4) = \frac{4\theta^2}{12} = \frac{\theta^2}{3}$$

例 11.9 $X_1, X_2, \ldots, X_n$ 来自均匀总体 $U(0,\theta)$ 的样本，参数 θ 的矩估计量 $\hat{\theta}_1 = 2\bar{X}$，极大似然估计量 $\tilde{\theta} = X_{(n)}$，考查无偏性。

解 $E(\hat{\theta}_1) = E(2\bar{X}) = 2E(\bar{X}) = 2 \cdot \dfrac{\theta}{2} = \theta$，所以 $\hat{\theta}_1 = 2\bar{X}$ 是参数 θ 的无偏估计。计算 $\tilde{\theta} = X_{(n)}$ 的分布函数，当 $0 \le y \le \theta$ 时，由样本的独立同分布性质，可知其分布函数为

$$F_{\tilde{\theta}}(y) = P(\tilde{\theta} \le y) = P\left(X_{(n)} \le y\right) = \prod_{k=1}^{n} P(X_k \le y) = \left(\frac{y}{\theta}\right)^n$$

于是 $\tilde{\theta}$ 的概率密度为 $f_{\tilde{\theta}}(y) = \begin{cases} \dfrac{n}{\theta^n} y^{n-1}, & y \in [0,\theta] \\ 0, & \text{其它} \end{cases}$，因此，我们有

$$E(\tilde{\theta}) = \int_0^{+\infty} y f_{\tilde{\theta}}(y)\,dy = \int_0^{\theta} y \cdot \frac{n}{\theta^n} y^{n-1}\,dy = \frac{n}{\theta^n}\int_0^{\theta} y^n\,dy = \frac{n}{\theta^n}\left[\frac{y^{n+1}}{n+1}\right]_0^{\theta} = \frac{n}{n+1}\theta$$

$E(\tilde{\theta}) = E\left(X_{(n)}\right) = \dfrac{n}{n+1}\theta$，因此统计量 $X_{(n)}$ 不是参数 θ 的无偏估计，而 $\dfrac{n+1}{n}X_{(n)}$ 则为无偏估计，称为无偏校正。

例 11.10 $X_1, X_2, \ldots, X_n$ 是来自均匀总体 $U(0,\theta)$ 的样本，$\hat{\theta}_1 = 2\bar{X}$ 和 $\hat{\theta}_2 = \dfrac{n+1}{n}X_{(n)}$ 都是参数 θ 的无偏估计，比较它们的有效性。

解 均匀总体 $U(0,\theta)$ 的期望、方差分别为 $E(X) = \dfrac{\theta}{2}, \mathrm{Var}(X) = \dfrac{\theta^2}{12}$。

$$\mathrm{Var}(\hat{\theta}_1) = \mathrm{Var}(2\bar{X}) = 4\mathrm{Var}(\bar{X}) = 4\frac{\mathrm{Var}(X)}{n} = 4\frac{\theta^2/12}{n} = \frac{\theta^2}{3n}$$

对于 $\hat{\theta}_2 = \dfrac{n+1}{n}X_{(n)}$，计算 $\tilde{\theta} = X_{(n)}$ 的方差，$\tilde{\theta}$ 的概率密度为 $f_{\tilde{\theta}}(y) = \begin{cases} \dfrac{n}{\theta^n} y^{n-1}, & y \in [0,\theta] \\ 0, & \text{其它} \end{cases}$。

$$E(\tilde{\theta}^2) = \int_0^{\theta} y^2 f_{\tilde{\theta}}(y)\,dy = \int_0^{\theta} y^2 \frac{n}{\theta^n} y^{n-1}\,dy = \frac{n}{\theta^n}\int_0^{\theta} y^{n+1}\,dy = \frac{n}{\theta^n}\left[\frac{y^{n+2}}{n+2}\right]_0^{\theta} = \frac{n}{n+2}\theta^2$$

$$\mathrm{Var}(\tilde{\theta}) = E(\tilde{\theta}^2) - (E(\tilde{\theta}))^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2$$

因此

$$\mathrm{Var}(\hat{\theta}_2) = \mathrm{Var}\left(\frac{n+1}{n}\tilde{\theta}\right) = \left(\frac{n+1}{n}\right)^2 \mathrm{Var}(\tilde{\theta}) = \frac{(n+1)^2}{n^2}\frac{n}{(n+1)^2(n+2)}\theta^2 = \frac{1}{n(n+2)}\theta^2$$

当 $n > 1$ 时，$\dfrac{1}{n(n+2)}\theta^2 < \dfrac{1}{3n}\theta^2$，所以 $\hat{\theta}_2 = \dfrac{n+1}{n}X_{(n)}$ 比 $\hat{\theta}_1 = 2\bar{X}$ 更有效。

11.2.1 均方误差

定义 11.5 (均方误差 (Mean Square Error, MSE))

对于参数 θ 的一个估计量 $\hat{\theta}$，其均方误差定义为 $E\left((\hat{\theta}-\theta)^2\right)$。

$$E\left((\hat{\theta}-\theta)^2\right) = E\left(\left[(\hat{\theta}-E(\hat{\theta})) + (E(\hat{\theta})-\theta)\right]^2\right) = \mathrm{Var}(\hat{\theta}) + (E(\hat{\theta})-\theta)^2$$

♣

11.2.2 最小方差无偏估计

定义 11.6 (最小方差无偏估计 (MVU 估计))

设 $\hat{\theta}$ 为 $g(\theta)$ 的无偏估计，若对 $g(\theta)$ 的任何无偏估计 $\hat{\theta}_1$，不等式

$$\mathrm{Var}_\theta(\hat{\theta}) \leq \mathrm{Var}_\theta(\hat{\theta}_1)$$

对 θ 的任何可能的值都成立，则称 $\hat{\theta}$ 为 $g(\theta)$ 的一个最小方差无偏估计 (Minimum Variance Unbiased, MV)。♣

定理 11.1 (克拉美-劳不等式 (Cramer-Rao Inequality))

在满足一定正则条件下，若 $T = T(X_1, X_2, \ldots, X_n)$ 是 $g(\theta)$ 的任一无偏估计，则其方差满足：

$$\mathrm{Var}(T) \geq \frac{(g'(\theta))^2}{nI(\theta)}$$

其中，n 是样本容量，$I(\theta)$ 称为 Fisher 信息量，定义为：

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\ln p(X;\theta)\right)^2\right]$$

♡

例 11.11 设样本 $X_1, \ldots, X_n$ 来自指数总体 $\mathrm{Exp}(\lambda)$，估计其期望 $g(\lambda) = \frac{1}{\lambda}$。

解 考虑使用样本均值 $\bar{X}$ 作为 $g(\lambda) = \frac{1}{\lambda}$ 的估计量。

首先，计算 $\bar{X}$ 的期望和方差。由于 $X_k \sim \mathrm{Exp}(\lambda)$，我们知道 $E(X_k) = \frac{1}{\lambda}$ 且 $\mathrm{Var}(X_k) = \frac{1}{\lambda^2}$。

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \frac{1}{n}\cdot n\cdot\frac{1}{\lambda} = \frac{1}{\lambda}$$

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}(X_k) = \frac{1}{n^2}\cdot n\cdot\frac{1}{\lambda^2} = \frac{1}{n\lambda^2}$$

由于 $E(\bar{X}) = g(\lambda) = \frac{1}{\lambda}$，所以 $\bar{X}$ 是 $g(\lambda)$ 的无偏估计。

接下来，计算克拉美-劳下界。被估计的函数为 $g(\lambda) = \frac{1}{\lambda}$，其导数的平方为：

$$(g'(\lambda))^2 = \left(-\frac{1}{\lambda^2}\right)^2 = \frac{1}{\lambda^4}$$

单个观测的 Fisher 信息量 $I(\lambda)$ 为：

$$I(\lambda) = E\left[\left(\frac{\partial}{\partial\lambda}\ln p(X;\lambda)\right)^2\right] = E\left[\left(\frac{\partial}{\partial\lambda}\ln(\lambda e^{-\lambda X})\right)^2\right] = E\left[\left(\frac{1}{\lambda} - X\right)^2\right] = \frac{1}{\lambda^2}$$

因此，克拉美-劳下界为：

$$\frac{(g'(\lambda))^2}{nI(\lambda)} = \frac{1/\lambda^4}{n(1/\lambda^2)} = \frac{1}{n\lambda^2} = \mathrm{Var}(\bar{X})$$

我们发现 $\mathrm{Var}(\bar{X}) = \frac{1}{n\lambda^2}$ 正好等于克拉美-劳下界。因此，$\bar{X}$ 是 $g(\lambda) = \frac{1}{\lambda}$ 的一个有效估计量（也是最小方差无偏估计量）。

11.2.3 无偏估计的改善

定理 11.2

设 X, Y 是两个随机变量, $E(X) = \mu, \mathrm{Var}(X) > 0$。令 $\varphi(y) = E(X|Y = y)$，则有

$$E(\varphi(Y)) = \mu, \quad \mathrm{Var}(\varphi(Y)) \le \mathrm{Var}(X),$$

其中等号成立的充分必要条件是 X 与 $\varphi(Y)$ 几乎处处相等。

♡

定理 11.3 (Rao-Blackwell 定理)

设总体的概率（密度）函数为 $p(x;\theta), X_1, X_2, \ldots, X_n$ 是其样本。$T = T(X_1, \ldots, X_n)$ 是 θ 的一个充分统计量。若 $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ 是 θ 的任意无偏估计，则通过条件期望构造的新估计量

$$\tilde{\theta} = E(\hat{\theta}|T)$$

也是 θ 的一个无偏估计，并且其方差不劣于原估计量，即

$$\mathrm{Var}(\tilde{\theta}) \le \mathrm{Var}(\hat{\theta})$$

♡

第 12 节 参数区间估计

12.1 参数区间估计的定义

点估计是用一个点（即一个数或随机变量）估计未知参数，而区间估计就是用一个区间估计未知参数。

例 12.1 设样本 X_1, X_2, X_3, X_4 来自正态总体 $N(\mu, 1)$，样本均值 $\bar{x}$ 是参数 μ 的一个点估计。

则有：

$$P(\bar{x} = \mu) = 0, \quad \bar{x} - \mu \sim N(0, \frac{1}{4})$$

$$\begin{aligned} P(\mu \in [\bar{x} - 1, \bar{x} + 1]) &= P(\bar{x} - 1 \leq \mu \leq \bar{x} + 1) = P(-1 \leq \bar{x} - \mu \leq 1) \\ &= P(-2 \leq 2(\bar{x} - \mu) \leq 2) = \Phi(2) - \Phi(-2) \approx 0.9544 \end{aligned}$$

$[\bar{x} - 1, \bar{x} + 1]$ 是 μ 的一个置信水平 0.9544 的区间估计，也可以说是置信水平 0.9 或 0.8 的区间估计。给出的估计区间称为置信区间。置信水平可以取到的最大值称为置信系数。

定义 12.1 (参数区间估计)

设 $x_1, x_2, \ldots, x_n$ 是来自总体 $X \sim F(x; \theta)$ 的样本，$\theta \in \Theta \subset \mathbb{R}^1$ 为未知参数。$I(x_1, x_2, \ldots, x_n)$ 是一个随机区间，由样本值完全确定。称该区间为参数 θ 的一个置信水平为 $1 - \alpha$ $(0 < \alpha < 1)$ 的区间估计，是指

$$P_\theta(\theta \in I(x_1, x_2, \ldots, x_n)) \geq 1 - \alpha, \quad 0 < \alpha < 1, \quad \forall \theta \in \Theta.$$

其中：

$$I = [\hat{\theta}_L(x_1, x_2, \ldots, x_n), \hat{\theta}_U(x_1, x_2, \ldots, x_n)] \quad \text{（双侧区间估计）}$$

$$I = [\hat{\theta}_L(x_1, x_2, \ldots, x_n), +\infty) \quad \hat{\theta}_L：\text{置信下界（单侧区间估计）}$$

$$I = (-\infty, \hat{\theta}_U(x_1, x_2, \ldots, x_n)] \quad \hat{\theta}_U：\text{置信上界（单侧区间估计）}$$

♣

注 区间估计的两个基本要求：

1. 未知参数 θ 要以尽可能大的概率落在区间 $I(x_1, x_2, \ldots, x_n)$ 中；
2. 估计的精度要尽可能高。例如，在达到一定的置信水平的前提下，要求区间的长度尽可能小，或某种能体现这个要求的其他准则。

12.2 区间估计的构造方法

定义 12.2 (枢轴量法)

以双侧置信区间为例：

1. 构造“枢轴量”，$G(x_1, x_2, \ldots, x_n, \theta)$ 的值完全由样本值和未知参数确定，G 的分布不依赖于未知参数。
2. 选取两个常数 c, d，对给定的 α $(0 < \alpha < 1)$，有

$$P(c \leq G \leq d) \geq 1 - \alpha$$

3. 求解 $c \leq G(x_1, x_2, \ldots, x_n, \theta) \leq d$，得到

$$\hat{\theta}_1(x_1, x_2, \ldots, x_n) \leq \theta \leq \hat{\theta}_2(x_1, x_2, \ldots, x_n)$$

♣

假设枢轴量的分布函数是 $F(x)$, 密度函数为 $p(x)$。则得到优化问题:

$$\min\ d-c$$

约束条件:

$$F(d)-F(c)=1-\alpha$$

由 Lagrange 乘子法:

$$L(c,d;\lambda)=d-c+\lambda\left(F(d)-F(c)-1+\alpha\right)$$

$$\begin{cases}\dfrac{\partial L(c,d;\lambda)}{\partial c}=-1-\lambda p(c)=0\\ \dfrac{\partial L(c,d;\lambda)}{\partial d}=-1-\lambda p(d)=0 \qquad \Rightarrow \qquad p(c)=p(d)\\ \dfrac{\partial L(c,d;\lambda)}{\partial \lambda}=F(d)-F(c)-1+\alpha=0\end{cases}$$

例 12.2 区间估计公式：总体现象 $X\sim N(\mu,\sigma^2),\sigma^2$ 已知，μ 未知，$x_1,x_2,\ldots,x_n$ 是简单随机样本，求 μ 的 $1-\alpha$ 置信区间。

解

$$\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\quad \text{可作为枢轴量,}\qquad Z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)$$

则有:

$$\Phi(d)-\Phi(c)=P\left(c\le \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}\le d\right)=1-\alpha$$

取定 c,d 使得区间尽可能短，则:

$$p(c)=p(d),\quad \Phi(d)-\Phi(c)=1-\alpha$$
$$\Phi(d)=1-\frac{\alpha}{2},\quad \Phi(c)=\frac{\alpha}{2}\Rightarrow d=\Phi^{-1}\left(1-\frac{\alpha}{2}\right),\quad c=\Phi^{-1}\left(\frac{\alpha}{2}\right)$$

c,d 是标准正态分布的 $1-\dfrac{\alpha}{2}$ 分位数和 $\dfrac{\alpha}{2}$ 分位数:

$$d=u_{1-\frac{\alpha}{2}},\quad c=u_{\frac{\alpha}{2}}=-u_{1-\frac{\alpha}{2}}$$

例 12.3 设总体 $X\sim N(\mu,\sigma^2)$，σ^2 已知，μ 未知，样本为简单随机样本，求 μ 的 $1-\alpha$ 置信区间。

解 首先，样本均值 $\bar{x}$ 和样本标准差 s^2 的分布分别为:

$$\bar{x}\sim N\left(\mu,\frac{\sigma^2}{n}\right),\quad \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

而 $\bar{x}$ 与 s^2 相互独立，且

$$\frac{(n-1)\cdot s^2}{\sigma^2}\sim\chi^2(n-1)$$

根据 χ^2 分布的性质:

$$\frac{\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)\cdot s^2}{\sigma^2}/(n-1)}}=\frac{\overline{X}-\mu}{s/\sqrt{n}}\sim t(n-1)\quad\Rightarrow\quad P\left(t_{\alpha/2}(n-1)\le\frac{\overline{X}-\mu}{s/\sqrt{n}}\le t_{1-\alpha/2}(n-1)\right)=1-\alpha$$

通过推导可以得到 μ 的置信区间:

$$\bar{x}-\frac{s}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)\le\mu\le\bar{x}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)$$

例 12.4 总体 $X \sim N(\mu, \sigma^2)$，参数 μ 和 σ^2 均未知，求 σ^2 的 $1-\alpha$ 置信区间。

解 $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$ 可作为枢轴量。根据 χ^2 分布，我们有：

$$P\left(\chi^2_{\frac{\alpha}{2}} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

对不等式进行变换以得到 σ^2 的范围：

$$P\left(\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}\right) = 1-\alpha$$

因此，σ^2 的 $1-\alpha$ 置信区间为：

$$\sigma^2 \in \left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}\right]$$

例 12.5 $x_1, \ldots, x_m$ 来自正态总体 $N(\mu_1, \sigma_1^2)$，$y_1, \ldots, y_n$ 来自正态总体 $N(\mu_2, \sigma_2^2)$，μ_1, μ_2 未知，σ_1^2, σ_2^2 未知，求 σ_1^2/σ_2^2 的区间估计。

解 我们已知（假设 s_1^2 和 s_2^2 分别是两个样本的样本方差）：

$$\frac{(m-1)s_1^2}{\sigma_1^2} \sim \chi^2(m-1), \quad \frac{(n-1)s_2^2}{\sigma_2^2} \sim \chi^2(n-1)$$

假设两个样本相互独立，则可以构造 F 分布的枢轴量：

$$F = \frac{\dfrac{(m-1)s_1^2/\sigma_1^2}{m-1}}{\dfrac{(n-1)s_2^2/\sigma_2^2}{n-1}} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F(m-1, n-1)$$

例 12.6 $x_1, \ldots, x_n$ 来自均匀总体 $U(0,\theta)$，求参数 θ 的 $1-\alpha$ 置信区间，参数 θ 的极大似然估计量为 $\hat{\theta} = \max\{x_1, x_2, \ldots, x_n\} = x_{(n)}$。

解 随机变量 $X_{(n)} = \max\{X_1, \ldots, X_n\}$ 的分布函数为 $F_{X_{(n)}}(x;\theta) = \left(\dfrac{x}{\theta}\right)^n$。令 $Y = \dfrac{X_{(n)}}{\theta}$，则 Y 的分布函数为：

$$F_Y(y) = P(Y \le y) = P\left(\frac{X_{(n)}}{\theta} \le y\right) = P(X_{(n)} \le y\theta) = \left(\frac{y\theta}{\theta}\right)^n = y^n, \quad \text{for } 0 \le y \le 1$$

$\dfrac{X_{(n)}}{\theta}$ 可作为枢轴量。我们寻找常数 c 和 d 使得：

$$P\left(c \le \frac{X_{(n)}}{\theta} \le d\right) = F_Y(d) - F_Y(c) = d^n - c^n = 1-\alpha$$

从上述概率不等式可以得到 θ 的范围：

$$\frac{X_{(n)}}{d} \le \theta \le \frac{X_{(n)}}{c}$$

可取 $d = 1, c = \sqrt[n]{\alpha}$，因此，$\theta$ 的一个 $1-\alpha$ 置信区间为：

$$\theta \in \left[X_{(n)}, \frac{X_{(n)}}{\sqrt[n]{\alpha}}\right]$$

例 12.7 样本 $X_1, \ldots, X_n$ 来自两点分布总体 $b(1,p)$，求 p 的区间估计。

解 样本均值的期望、方差分别为

$$E(\bar{X}) = p, \mathrm{Var}(\bar{X}) = \frac{p(1-p)}{n}$$

根据中心极限定理当 n 较大时，有近似分布

$$\bar{X} \approx N\left(p, \frac{p(1-p)}{n}\right),$$

标准化后得到枢轴量

$$\frac{\bar{X}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \approx N(0,1),$$

因此，

$$P\left(\left|\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}}\right| \le u_{1-\frac{\alpha}{2}}\right) \approx 1-\alpha, \quad \text{即} (\bar{X}-p)^2 \le u_{1-\frac{\alpha}{2}}^2 \frac{p(1-p)}{n}$$

解得：

$$p \in \left[\frac{1}{1+c}\left(\bar{X}+\frac{c}{2}-u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}+\frac{c^2}{4}}\right), \frac{1}{1+c}\left(\bar{X}+\frac{c}{2}+u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}+\frac{c^2}{4}}\right)\right]$$

其中 $c=\frac{u_{1-\frac{\alpha}{2}}^2}{n}$。当 n 较大时，c 的值很小可略去，得到参数 p 的 $1-\alpha$ 置信水平的近似估计区间：

$$\left[\bar{X}-u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}, \bar{X}+u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right]$$

第 13 节 假设检验

13.1 假设检验的基本步骤

定义 13.1 (假设检验的步骤)

1. 建立假设

   建立原假设 (null hypothesis) H_0 与备择假设 (alternative hypothesis) H_1。

   $$H_0:\theta\in\Theta_0 \quad \text{VS} \quad H_1:\theta\in\Theta_1$$

2. 选择检验统计量，给出拒绝域的形式

   所谓拒绝域 **(Rejection Region) W** 是指使原假设被拒绝的样本观测值所在的区域，有时也将 $\overline{\mathbf{W}}$ 称为接受域 **(Acceptance Region)**。

3. 选择显著性水平 α (具体异常到什么程度拒绝原假设)

   其定义为犯第一类错误的概率的最大值：

   $$\alpha=\max\{P(\text{拒绝}H_0|H_0\text{ 为真})\}=\max_{\theta\in\Theta_0}\{P_\theta(\mathbf{X}\in\mathbf{W})\}$$

   根据选定的 α 值，最终确定拒绝域的具体范围。

♣

第一类错误的概率：

$$\alpha(\theta)=\mathbf{P}(\text{拒绝}\mathbf{H}_0\,|\,\mathbf{H}_0\text{ 为真})=\mathbf{P}_\theta(\mathbf{X}\in\mathbf{W}),\quad \theta\in\Theta_0$$

第二类错误的概率：

$$\beta(\theta)=\mathbf{P}(\text{接受}\mathbf{H}_0\,|\,\mathbf{H}_1\text{ 为真})=\mathbf{P}_\theta(\mathbf{X}\in\overline{\mathbf{W}}),\quad \theta\in\Theta_1$$

例 13.1 某工厂生产一种标准长度 35mm 的螺钉，实际生产的产品长度服从正态分布 $N(\mu,3^2)$。做假设检验，样本容量 $n=36$，$H_0:\mu=35$，$H_1:\mu\neq 35$，拒绝域为 $W=\{\bar{x}:|\bar{x}-35|>1\}$。

1. 犯第一类错误的概率。
2. $\mu=36$ 时，犯第二类错误的概率。

解 检验统计量 $\bar{X}$ 的分布为 $\bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)=N\left(\mu,\frac{3^2}{36}\right)=N\left(\mu,\left(\frac{1}{2}\right)^2\right)$。

第一类错误的概率为

$$\begin{aligned}\alpha&=P(|\bar{X}-35|>1|\mu=35)=1-P(|\bar{X}-35|\le 1|\mu=35)\\&=1-P\left(-2\le\frac{\bar{X}-35}{1/2}\le 2\middle|\mu=35\right)\\&=1-(\Phi(2)-\Phi(-2))=1-\Phi(2)+(1-\Phi(2))=2-2\Phi(2)=0.0455\end{aligned}$$

第二类错误的概率为

$$\begin{aligned}\beta&=P(|\bar{X}-35|\le 1|\mu=36)=P(-1\le\bar{X}-35\le 1|\mu=36)\\&=P\left(\frac{-2}{1/2}\le\frac{\bar{X}-36}{1/2}\le\frac{0}{1/2}\middle|\mu=36\right)\\&=\Phi(0)-\Phi(-4)=\Phi(0)-(1-\Phi(4))=\Phi(0)+\Phi(4)-1=0.5\end{aligned}$$

13.1.1 随机化检验

例 13.2 为检验一位女士品茶的能力（p 为品尝正确的概率），建立假设 $H_0: p \le 1/2$（无特殊能力）vs $H_1: p > 1/2$。她从 8 杯（4 杯 A 类，4 杯 B 类）中选出 4 杯 A 类。令 X 为她正确选出的 A 类杯数。在 $p = 1/2$（随机猜测）时，X 服从超几何分布，我们关心的是如何在 $\alpha = 0.05$ 的水平下进行检验。其概率分布的关键值为：

$$P(X=4|p=\frac{1}{2}) = \frac{1}{70}, \quad P(X=3|p=\frac{1}{2}) = \frac{16}{70}$$

解 我们希望当 X 足够大时拒绝 H_0。

- 若拒绝域为 $\{X=4\}$，则显著性水平为 $\alpha = \frac{1}{70} \approx 0.014$，远小于我们期望的 0.05。
- 若拒绝域为 $\{X \ge 3\}$，则显著性水平为 $\alpha = \frac{1}{70} + \frac{16}{70} = \frac{17}{70} \approx 0.243$，又太大了。

由于 X 的离散性，无法通过一个固定的临界值使 α 恰好等于 0.05。为了精确地使用给定的显著性水平，我们采用随机化检验。

该检验通过一个检验函数 $\phi(x)$ 来定义，$\phi(x)$ 是当观测到 $X = x$ 时拒绝 H_0 的概率：

$$\phi(x) = \begin{cases} 1, & x = 4 \quad (\text{当} X = 4 \text{时，总是拒绝}) \\ r, & x = 3 \quad (\text{当} X = 3 \text{时，以概率} r \text{拒绝}) \\ 0, & x < 3 \quad (\text{当} X < 3 \text{时，从不拒绝}) \end{cases}$$

检验的显著性水平 α 就是 $\phi(X)$ 在 H_0 $(p = 1/2)$ 下的期望值。我们令其等于 0.05 来求解概率 r：

$$\begin{aligned} \alpha &= E_{p=1/2}[\phi(X)] = \phi(4)P(X=4) + \phi(3)P(X=3) \\ &= 1 \cdot \frac{1}{70} + r \cdot \frac{16}{70} = 0.05 \end{aligned}$$

解得：

$$1 + 16r = 70 \times 0.05 = 3.5 \implies 16r = 2.5 \implies r = 0.15625$$

因此，在 $\alpha = 0.05$ 水平下的随机化检验规则为：

- 若 $X = 4$，拒绝 H_0。
- 若 $X = 3$，则做一个成功概率为 0.156 的伯努利试验，如果试验成功，则拒绝原假设，否则接受原假设
- 若 $X < 3$，接受 H_0。

13.2 正态总体均值的假设检验

设 $X_1, X_2, \ldots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本，μ_0 是一个已知的常数。我们考虑如下三种关于参数 μ 的检验问题：

1. $$H_0: \mu \le \mu_0 \quad \text{vs} \quad H_1: \mu > \mu_0 \quad (\text{右单侧检验})$$

2. $$H_0: \mu \ge \mu_0 \quad \text{vs} \quad H_1: \mu < \mu_0 \quad (\text{左单侧检验})$$

3. $$H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu \ne \mu_0 \quad (\text{双侧检验})$$

根据总体方差 σ^2 是否已知，我们选择不同的检验方法。

定理 (U 检验 (方差 σ^2 已知))

条件：正态总体 $N(\mu,\sigma^2)$，μ 未知，σ^2 已知。

检验统计量：

$$u=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma}\sim N(0,1)$$

在显著性水平 α 下，对于不同的假设，拒绝域形式如下：

- 双侧检验 $H_0:\mu=\mu_0$ vs $H_1:\mu\neq\mu_0$
  拒绝域为
  $$|\bar{x}-\mu_0|>\frac{\sigma}{\sqrt{n}}u_{1-\alpha/2}$$
- 左单侧检验 $H_0:\mu\geq\mu_0$ vs $H_1:\mu<\mu_0$
  拒绝域为
  $$\bar{x}<\mu_0-\frac{\sigma}{\sqrt{n}}u_{1-\alpha}$$
- 右单侧检验 $H_0:\mu\leq\mu_0$ vs $H_1:\mu>\mu_0$
  拒绝域为
  $$\bar{x}>\mu_0+\frac{\sigma}{\sqrt{n}}u_{1-\alpha}$$

(其中 u_q 是标准正态分布的 q 分位数。)

♡

定理 (t 检验 (方差 σ^2 未知))

条件：正态总体 $N(\mu,\sigma^2)$，μ 未知，σ^2 未知。

检验统计量：

$$t=\frac{\bar{X}-\mu_0}{s/\sqrt{n}}=\frac{\sqrt{n}(\bar{X}-\mu_0)}{s}\sim t(n-1)$$

其中 s 是样本标准差。

在显著性水平 α 下，对于不同的假设，拒绝域形式如下：

- 双侧检验: $H_0:\mu=\mu_0$ vs $H_1:\mu\neq\mu_0$
  拒绝域为:
  $$|\bar{x}-\mu_0|>\frac{s}{\sqrt{n}}t_{1-\alpha/2}(n-1)$$
- 左单侧检验: $H_0:\mu\geq\mu_0$ vs $H_1:\mu<\mu_0$
  拒绝域为):
  $$\bar{x}<\mu_0-\frac{s}{\sqrt{n}}t_{1-\alpha}(n-1)$$
- 右单侧检验: $H_0:\mu\leq\mu_0$ vs $H_1:\mu>\mu_0$
  拒绝域为:
  $$\bar{x}>\mu_0+\frac{s}{\sqrt{n}}t_{1-\alpha}(n-1)$$

(其中 $t_q(n-1)$ 是自由度为 $n-1$ 的 t 分布的 q 分位数。)

♡

定义 13.2 (检验的 p 值 (p-value))

在一个假设检验问题中，利用样本观测值能够做出拒绝原假设的最小显著性水平；即原假设成立条件下，检验统计量出现在比观测值更异常的范围的概率的最大值。♣

例 13.3 总体服从 $N(\mu, 3^2)$。做假设检验，样本容量 $n = 36$，$H_0 : \mu = 35$，$H_1 : \mu \neq 35$。现得到样本均值的观测值为 36.5mm，求其 p 值。

解

$$\begin{aligned} p &= P(|\bar{X} - 35| \geq 1.5 | \mu = 35) \\ &= 1 - P(|\bar{X} - 35| < 1.5 | \mu = 35) \\ &= 1 - P\left\{-3 < \frac{\bar{X} - 35}{1/2} < 3 \middle| \mu = 35\right\} \\ &= 1 - (\Phi(3) - \Phi(-3)) = 0.0027 \end{aligned}$$

第 14 节 拟合优度检验

14.1 拟合优度检验

定义 14.1 (χ^2 检验 (拟合优度检验))

设总体服从一个有 k 个可能取值的离散分布，其概率分布为：

X	x_1	x_2	$\dots$	x_k
P	p_1	p_2	$\dots$	p_k

我们进行 n 次独立观测，得到各个值 x_i 出现的频次分别为 N_i（其中 $i = 1, 2, \dots, k$，且 $\sum N_i = n$）。检验统计量定义为：

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}$$

其中，N_i 是观测频次，np_i 是在原假设成立下的期望频次。该统计量近似服从自由度为 $\mathbf{k-1}$ 的 χ^2 分布。

检验统计量的推导：

$$N_i \sim b(n, p_i), \qquad \frac{N_i - np_i}{\sqrt{np_i(1-p_i)}} \sim N(0,1)$$

$$\sum_{i=1}^{k} \left(\frac{N_i - np_i}{\sqrt{np_i(1-p_i)}} \right)^2 \approx \sum_{i=1}^{k} \left(\frac{N_i - np_i}{\sqrt{np_i}} \right)^2 \sim \chi^2(k-1)$$

♣

当参数未知时： 若理论分布中的概率 p_i 依赖于 s 个未知参数，需要通过样本来估计这 s 个参数，则检验统计量

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

（其中 $\hat{p}_i$ 是利用估计出的参数计算得到的概率）近似服从自由度为 $\mathbf{k-s-1}$ 的 χ^2 分布。

例 14.1 给定卢瑟福和盖革观测的放射性衰变数据（n_k 为观测到 k 个粒子的次数，总观测次数 $n = 2608$）：

k	0	1	2	3	4	5	6	7	8	9	≥ 10
n_k	57	203	383	525	532	408	273	139	45	27	16

检验这组数据是否来自泊松分布。

解 1. 估计参数与计算期望频次：原假设为数据服从泊松分布 $P(\lambda)$。由于参数 λ 未知，我们用样本均值作为其估计值：

$$\hat{\lambda} = \frac{\text{总粒子数}}{\text{总观察次数}} = \frac{10094}{2608} \approx 3.87$$

2. 计算 χ^2 统计量：检验统计量的值为：

$$Y = \sum_{i=0}^{10} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = 12.88$$

3. 决策：我们有 $k = 11$ 个类别，估计了 $s = 1$ 个参数 (λ)，所以卡方分布的自由度为 $df = k - 1 - s = 9$。

$$Y \sim \chi^2(9)$$

查表或使用软件可知，该统计量对应的 p 值为 $p = P(Y > 12.88) > 0.1$。由于 p 值大于常用的显著性水平（如 0.05），我们没有充分的理由拒绝原假设，故可以接受这组数据来自泊松分布。

14.2 独立性检验

定理 14.1 (列联表独立性检验)

$A \setminus B$	1	2	$\cdots$	j	$\cdots$	t	行合计
1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1t}	c_1
2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2t}	c_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{it}	c_i
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
s	n_{s1}	n_{s2}	$\cdots$	n_{sj}	$\cdots$	n_{st}	c_s
列合计	d_1	d_2	$\cdots$	d_j	$\cdots$	d_t	n

为了检验两个分类变量 A 和 B 是否相互独立，我们构造如下的检验统计量：

$$Y=\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{\left(n_{ij}-n\cdot\dfrac{c_i}{n}\cdot\dfrac{d_j}{n}\right)^2}{n\cdot\dfrac{c_i}{n}\cdot\dfrac{d_j}{n}}=\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{(nn_{ij}-c_id_j)^2}{nc_id_j}$$

该统计量近似服从自由度为 $(s-1)(t-1)$ 的 χ^2 分布。

例 14.2 一项关于是否应提高小学生计算机课程比例的调查，其结果按年龄和回答交叉分类如下：

年龄	同意	不同意	不知道	总计
55 岁以上	32	28	14	74
36 ～ 55 岁	44	21	17	82
15 ～ 35 岁	47	12	13	72
总计	123	61	44	228

问：年龄因素是否影响了对问题的回答 ($\alpha=0.05$)?

解 待检验的假设为 H_0：年龄因素对问题的回答无关联。统计表示如下：

$$H_0: p_{ij}=p_{i\cdot}p_{\cdot j},\quad i=1,2,3, j=1,2,3$$

在原假设成立下，我们用样本的边际频率来估计边际概率：

$$\hat{p}_{\cdot1}=\frac{74}{228}=0.3246,\qquad \hat{p}_{\cdot2}=\frac{82}{228}=0.3596,\qquad \hat{p}_{\cdot3}=\frac{72}{228}=0.3158$$
$$\hat{p}_{1\cdot}=\frac{123}{228}=0.5395,\qquad \hat{p}_{2\cdot}=\frac{61}{228}=0.2675,\qquad \hat{p}_{3\cdot}=\frac{44}{228}=0.1930$$

进而利用 $E_{ij}=n\hat{p}_{i\cdot}\hat{p}_{\cdot j}$ 得到期望频次 $\hat{E}_{ij}$：

$$n\hat{p}_{11}=39.9277,\qquad n\hat{p}_{12}=19.7974,\qquad n\hat{p}_{13}=14.2837$$
$$n\hat{p}_{21}=44.2330,\qquad n\hat{p}_{22}=21.9320,\qquad n\hat{p}_{23}=15.8238$$
$$n\hat{p}_{31}=38.8453,\qquad n\hat{p}_{32}=19.2606,\qquad n\hat{p}_{33}=13.8965$$

由此上结果可计算出检验的统计量：

$$\chi^2=\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(n_{ij}-n\hat{p}_{ij})^2}{n\hat{p}_{ij}}=9.6133$$

此处行数 $r=3$, 列数 $c=3$，自由度为 $df=(r-1)(c-1)=(3-1)(3-1)=4$。$\chi^2_{0.95}(4)=9.4877$。$\chi^2=9.6133>9.4877$，所以它落在了拒绝域内。故拒绝原假设，即认为年龄因素与问题的回答有关联。

此处的 p 值为 $p=P(\chi^2(4)\geq 9.6133)=0.0475$。