概率论与数理统计

Weijie Li

Winter 2025

目录

Chapter 1

概率模型

1.1 随机事件

下面是对一些基本概念的定义.

【定义 1.1】(试验) 通过观察、测量进行的实验称为试验.

【定义 1.2】(随机试验) 在相同条件下可重复的、结果不止一个的、无法预测实验结果的试验, 记作E.

【定义 1.3】(事件) 每种实验的结果称为事件.

【定义 1.4】(随机事件) 随机发生的实验结果称为随机事件, 一般用大写字母表示.

【定义 1.5】(基本事件) 相对于实验目的不能(必)再分的事件称为基本事件.

【定义 1.6】(复合事件) 由基本事件复合而成的事件称为复合事件.

【定义 1.7】(必然事件) 一定发生的事件称为必然事件, 即全集, 记作Ω.

【定义 1.8】(不可能事件) 一定不发生的事件称为不可能事件, 即空集, 记作Φ.

【定义 1.9】(样本空间) 所有基本事件组成的集合称为样本空间, 记作Ω.

【定义 1.10】(样本点) 样本空间中的元素称为样本点, 记作ω.

【例题 1.11】 求空面内某一质子的坐标所在样本空间.

<u>**解.**</u>

$$\Omega = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}.$$

□

下面对事件的基本运算做一些说明, 事实上它们等价于集合论中的一切运算.

【定义 1.12】(包含) 事件A发生必导致事件B发生, 则称A包含于B, 记作$A \subset B$.
直接推论: $\forall A$, 均有$\Phi \subset A \subset \Omega$.

【定义 1.13】(加和) 设事件C表示事件A与事件B中至少有一个发生, 则有$C = A \cup B$, 同时记作$C = A + B$.

【定义 1.14】(乘积) 设事件C表示事件A与事件B同时发生, 则有$C = A \cap B$, 同时记作$C = AB$.

【定义 1.15】(差) 设事件C表示事件A发生而事件B不发生, 记作$C = A - B$.

【定义 1.16】(互不相容事件) 若事件A与B不同时发生, 即$AB = \Phi$, 则称A和B互不相容.

【定义 1.17】(对立事件) 若A和B互不相容, 且$A \cup B = \Omega$, 则称A和B互为对立事件.

【定义 1.18】(完备事件组) 若$A_1, A_2, ..., A_n$两两互不相容, 且满足$\bigcup\limits_{i=1}^{n} A_i = \Omega$, 则称这组事件为完备事件组.

1.2 事件的概率

【定义 1.19】(古典概型) 若研究的样本空间仅包含有限个样本点,且每个样本点出现的可能性相等,则称这样的模型为古典概型.计算公式为

$$P(A)=\frac{A\text{的有利样本点}}{\Omega\text{中样本点的总数}}=\frac{A\text{中包含的基本事件数}}{\text{基本事件总数}}.$$

【例题 1.20】 给定a个白球b个黑球,从中接连取出m个,求第m次抽出白球的概率.

解. 考虑对所有球做全排列,则目标样本点对应于在第m个位置上分别取定a个白球,对剩余球做全排列,即

$$P(A)=\frac{a(a+b-1)!}{(a+b)!}=\frac{a}{a+b}.$$

□

古典概型具有一些性质如下

1. 非负性:$0\le P(A)\le 1$.
2. 规范性:$P(\Omega)=1,P(\Phi)=0$.
3. 有限可加性:设$A_1,A_2,...,A_n$互不相容,则$P(\sum\limits_{i=1}^{n}A_i)=\sum\limits_{i=1}^{n}P(A_i)$.

【定义 1.21】(几何概型) 若研究的样本空间包含无限个样本点,且每个样本点出现的可能性相等,则称这样的模型为几何概型.计算公式为

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

其中$\mu(G)$表示度量.

【例题 1.22】 (蒲丰投针问题)向相距d的两条平行线之间随机投下一根长为l的针,其中$l<d$,求针与直线相交的概率.

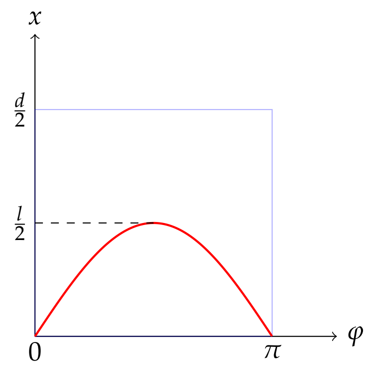
解. 设x表示针的中点到最近线的距离,则显然有$0\le x\le\frac{d}{2}$.又设针与直线的夹角为φ,则$0\le\varphi\le\pi$.于是

$$\Omega=\{(\varphi,x)\mid 0\le\varphi\le\pi,0\le x\le\frac{d}{2}\}.$$

$$G=\{(\varphi,x)\mid 0\le\varphi\le\pi,0\le x\le\frac{l}{2}\sin\varphi\}.$$

$x-\varphi$关系图像如右图,故

$$P(A)=\frac{\int_0^{\pi}\frac{l}{2}\sin\varphi d\varphi}{\pi\times\frac{d}{2}}=\frac{2l}{\pi d}.$$



□

几何概型与古典概型的不同之处在于其具有完全可加性:设$A_1,A_2,...$互不相容,则$P(\sum\limits_{i=1}^{\infty}A_i)=\sum\limits_{i=1}^{\infty}P(A_i)$.

【定义 1.23】(频率) 在n次试验中,A共发生了m次,则称$\frac{m}{n}$为A发生的频率,记作$\omega_n(A)$.

频率具有与古典概型极其相似的性质:

1. 非负性:$0\le\omega_n(A)\le 1$.
2. 规范性:$\omega_n(\Omega)=1,\omega_n(\Phi)=0$.
3. 有限可加性:设$A_1,A_2,...,A_m$互不相容,则$\omega_n(\sum\limits_{i=1}^{m}A_i)=\sum\limits_{i=1}^{m}\omega_n(A_i)$.

下面我们将对概率模型进行公理化定义. 一般地, 当概率模型满足以下公理时, 我们称实数$P(A)$为事件A的概率:

1. 非负性:$0 \le P(A) \le 1$.
2. 规范性:$P(\Omega) = 1$.
3. 完全可加性:设$A_1, A_2, ...$互不相容, 则$P(\sum\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} P(A_i)$.

根据这三条公理, 我们可以推出$P(A)$所具有的一些性质.

【定理 1.24】 (性质1) $P(\Phi) = 0$.

证明. 由于$\Omega = \Omega + \sum\limits_{i=1}^{\infty} \Phi$, 于是

$$P(\Omega) = P(\Omega + \sum_{i=1}^{\infty} \Phi) = P(\Omega) + \sum_{i=1}^{\infty} P(\Phi),$$

故$\sum\limits_{i=1}^{\infty} P(\Phi) = 0$.又$P(\Phi) \ge 0$, 故$P(\Phi) = 0$. □

【定理 1.25】 (性质2) 有限可加性:设$A_1, A_2, ..., A_n$互不相容, 则$P(\sum\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{n} P(A_i)$.

证明. 用可列个Φ将n个事件扩充为可列个事件, 于是由完全可加性有

$$P(\sum_{i=1}^{n} A_i) = P(\sum_{i=1}^{n} A_i + \sum_{i=n+1}^{\infty} \Phi) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\Phi) = \sum_{i=1}^{n} P(A_i).$$

□

【定理 1.26】 (性质3) $P(\overline{A}) = 1 - P(A)$.

证明. 注意到$A, \overline{A}$互不相容, 由有限可加性即有

$$1 = P(\Omega) = P(A + \overline{A}) = P(A) + P(\overline{A}).$$

□

一条推论直接由此性质推出:设$A_1, A_2, ..., A_n$是一完备事件组, 则$\sum\limits_{i=1}^{n} P(A_i) = 1$.

【定理 1.27】 (性质4) $P(A - B) = P(A) - P(AB)$. 特别地, 当$B \subset A$时, 有$P(A - B) = P(A) - P(B)$.

证明. 由于$A = (A - B) \cup AB$, 且$A - B$与AB互不相容.故由有限可加性,有

$$P(A) = P(A - B) + P(AB).$$

□

【定理 1.28】 (性质5) $P(A + B) = P(A) + P(B) - P(AB)$.

证明. 由于$A + B = A + (B - AB)$, 且A与$B - AB$互不相容.故由有限可加性,有

$$P(A + B) = P(A) + P(B - AB) = P(A) + P(B) - P(AB).$$

□

与此性质的定理完全同理, 可知概率运算满足容斥原理.

【例题 1.29】 已知$P(A) = 0.4, P(B) = 0.3, P(A + B) = 0.6$, 求$P(A\overline{B})$.

解. 由于$P(A + B) = P(A) + P(B) - P(AB) \Rightarrow P(AB) = P(A) + P(B) - P(A + B) = 0.1$,故

$$P(A\overline{B}) = P(A - B) = P(A) - P(AB) = 0.3.$$

□

1.3 条件概率

【定义 1.30】 **(条件概率)** 在样本空间Ω中, 存在A,B两个事件, 其中$P(B)>0$. 我们称在B已经发生的条件下A发生的概率为A对B的条件概率, 记作$P(A\mid B)$.计算公式为

$$P(A\mid B)=\frac{n_{AB}}{n_B}=\frac{P(AB)}{P(B)}.$$

事实上, 条件概率$P(A\mid B)$相当于在样本空间$\Omega_B=B$下的无条件概率.条件概率满足的性质有:

1. 非负性:$0\leq P(A\mid B)\leq 1$.
2. 规范性:$P(\Omega\mid B)=1$.
3. 完全可加性:设$A_1,A_2,...$互不相容, 则$P(\sum\limits_{i=1}^{\infty}A_i\mid B)=\sum\limits_{i=1}^{\infty}P(A_i\mid B)$.

【定理 1.31】 **(乘法公式)** $P(\prod\limits_{i=1}^{n}A_i)=\prod\limits_{i=1}^{n}P(A_i\mid\prod\limits_{j=1}^{i-1}A_j)$.

【例题 1.32】 (传染病模型) 暗箱中有a个红球, b个黑球. 每次抽取球并放回后都再放入c个颜色相同的球, 求三次抽取都得到红球的概率.

<u>解.</u> 设A_1,A_2,A_3分别表示第1, 2, 3次抽取到红球, 则

$$P(A_1A_2A_3)=P(A_1)P(A_2\mid A_1)P(A_3\mid A_1A_2)=\frac{a}{a+b}\cdot\frac{a+c}{a+b+c}\cdot\frac{a+2c}{a+b+2c}.$$

当$c=-1$时, 表示不放回抽取试验; 当$c=0$时, 表示放回抽取试验; 当$c>0$时, 即表示不采取隔离措施的传染病模型. □

【定理 1.33】 **(全概率公式)** 设$A_1,A_2,...,A_n$是E的完备事件组, 且$P(A_i)>0$. 则$P(B)=\sum\limits_{i=1}^{n}P(A_i)P(B\mid A_i)$.

【例题 1.34】 现有10件产品, 其中等可能地有次品0, 1, 2件. 在检验产品时, 正品的检验正确率为0.98, 次品的检验正确率为0.95. 从产品中随机抽出一件进行检验, 求产品通过验证的可能性.

<u>解.</u> 设事件A_0,A_1,A_2分别表示产品中有0, 1, 2件次品, 则$P(A_1)=P(A_2)=P(A_3)=\frac{1}{3}$. 设事件$B_1$表示抽出的产品为正品, 则$P(B_1\mid A_0)=1,P(B_1\mid A_1)=\frac{9}{10},P(B_1\mid A_2)=\frac{1}{5}$. 故

$$P(B_1)=P(A_0)P(B_1\mid A_0)+P(A_1)P(B_1\mid A_1)+P(A_2)P(B_1\mid A_2)=0.9,P(\overline{B_1})=0.1.$$

于是

$$P(B)=P(B_1)P(B\mid B_1)+P(\overline{B_1})P(B\mid\overline{B_1})=0.9\times0.98+0.1\times0.05=0.887.$$

□

【定理 1.35】 **(贝叶斯公式)** 设$A_1,A_2,...,A_n$是完备事件组, 满足$P(A_i)>0$. 另有一事件B, 满足$P(B)>0$. 则

$$P(A_k\mid B)=\frac{P(A_kB)}{P(B)}=\frac{P(A_k)P(B\mid A_k)}{\sum\limits_{i=1}^{n}P(A_i)P(B\mid A_i)}.$$

其中我们称$P(A_i)$为先验概率, $P(A_i\mid B)$为后验概率.

【例题 1.36】 现有一发病率为0.0004的疾病, 医院的误诊率为1%. 若某人已被确诊, 求此人确实患病的概率.

<u>解.</u> 设事件A表示此人为患者, B表示此人被确诊, 则所求概率即$P(A\mid B)$. 由贝叶斯公式, 有

$$P(A\mid B)=\frac{P(B\mid A)P(A)}{P(B\mid A)P(A)+P(B\mid\overline{A})P(\overline{A})}=\frac{0.99\times0.0004}{0.99\times0.0004+0.01\times0.9996}\approx0.0.381.$$

□

1.4 事件的独立性

【定义 1.37】(事件的独立性) 当事件A的概率不受B发生与否的影响, 则称A与B互相独立. 此时显然满足性质$P(A \mid B) = P(A)$.

【定理 1.38】 设$P(A) > 0, P(B) > 0$. 则A, B相互独立当且仅当$P(AB) = P(A)P(B)$.

证明. 在$P(A) > 0, P(B) > 0$下, 我们有

$$P(AB) = P(A)P(B) \Leftrightarrow P(A \mid B) = \frac{P(AB)}{P(B)} = P(A).$$

□

据此我们有两个事件独立的另一定义:

【定义 1.39】 当A, B满足$P(AB) = P(A)P(B)$时, 称A, B独立.

【定理 1.40】 若A, B独立, 则A与$\overline{B}$, $\overline{A}$与B, $\overline{A}$与$\overline{B}$均独立.

证明. $P(A\overline{B}) = P(A - B) = P(A) - P(AB) = P(A)(1 - P(B)) = P(A)P(\overline{B})$.同理可证另外两条. □

【定理 1.41】 若A满足$P(A) = 0$或$P(A) = 1$, 则A与任意事件独立.

证明. 当$P(A) = 0$时, 有

$$AB \subset A \Rightarrow 0 \leq P(AB) \leq P(A) = 0,$$

于是得到$P(AB) = 0 = P(A)P(B)$, 故A, B独立.同理可证$P(A) = 1$的情况. □

【定理 1.42】 设$P(A) > 0, P(B) > 0$, 则A, B独立与A, B互不相容无法同时成立.

证明. 若A, B独立, 则$P(AB) = P(A)P(B) > 0$, 这与A, B互不相容的定义矛盾. 反之同理可证. □

【例题 1.43】 已知$P(A + B) = 0.9, P(A) = 0.4$, 在下列条件下分别求$P(B)$. (1) A, B互不相容; (2) A, B独立.

解. (1) 由于A, B互不相容, 故$P(AB) = 0$. 从而$P(A + B) = P(A) + P(B) \Rightarrow P(B) = 0.5$..
(2) 由于A, B独立, 故$P(AB) = P(A)P(B)$. 从而$P(A + B) = P(A) + P(B) - P(A)P(B) \Rightarrow P(B) = \frac{5}{6}$. □

【例题 1.44】 设$0 < P(A) < 1, 0 < P(B) < 1$, 且满足$P(A \mid B) + P(\overline{A} \mid \overline{B}) = 1$. 证明:$A, B$独立.

证明. 由于基本事实$P(A \mid \overline{B}) + P(\overline{A} \mid \overline{B}) = 1$, 故

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

展开即得

$$\frac{P(AB)}{P(B)} = \frac{P(A\overline{B})}{P(\overline{B})} = \frac{P(A) - P(AB)}{1 - P(B)} \Rightarrow P(AB) = P(A)P(B).$$

□

1.5 Bernoulli模型

【定义 1.45】(独立实验序列) n个实验$\mathrm{E}_1, \mathrm{E}_2, ..., \mathrm{E}_n$相互独立, 称为独立实验序列.

【定义 1.46】(n重独立实验) 同一实验E重复做n次,实验之间相互独立, 称为n重独立实验. 记作E^n.

【定义 1.47】(Bernoulli实验) 称结果只有两种的实验为Bernoulli实验.

【定义 1.48】(n重Bernoulli实验) 即重复做n次Bernoulli实验.

【定理 1.49】 设$P(A) = p(0 < p < 1)$, 则在n重Bernoulli实验中A发生k次的概率为

$$P_n(k) = C_n^k p^k (1 - p)^{n-k}.$$

这一公式被称为二项概率公式.

Chapter 2

一维随机变量及其分布

【定义 2.1】 **(随机变量)** 在样本空间Ω上定义实值函数$X = X(\omega)$, 称该实值函数为样本空间Ω的随机变量.

2.1 离散型随机变量及其概率分布

【定义 2.2】 **(概率函数)** 设X有可列个取值, 分别为$x_k(k = 1, 2, ...)$, 对应样本点的概率$P\{X = x_k\} = p_k$, 称为该离散型随机变量的概率函数(分布).

离散型随机变量的概率函数应当满足以下性质:

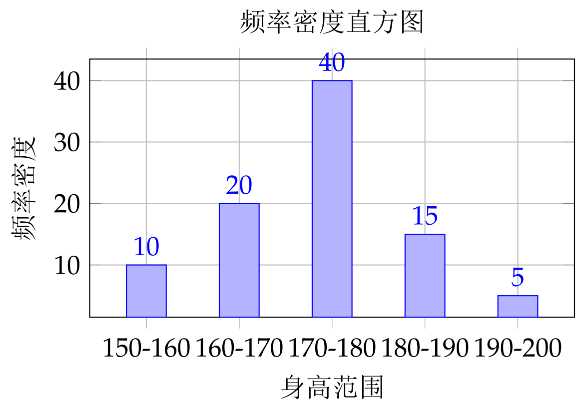
1. $p_k \geq 0$.
2. $\sum p_k = 1$.

2.2 连续型随机变量及其概率密度函数

对于连续型数据(例如身高), 我们进行适当的分组, 定义频率密度=频率/组距, 可以绘出频率密度直方图(例如下图).

身高范围	人数
150-160	10
160-170	20
170-180	40
180-190	15
190-200	5

图 2.1: 身高数据表



容易证明频率密度直方图满足以下性质:

1. 每个长方形的面积等于该组的频率.
2. 所有长方形的面积之和等于1.
3. 介于$x = a$和$x = b$之间的面积近似于$(a, b]$的频率.

当所作的适当分组的组距趋于0时, 我们将得到一条光滑曲线, 我们称该曲线$y = f(x)$为连续型随机变量的概率密度函数.

【定义 2.3】 **(概率密度函数(pmf/pdf))** 设X的取值为某个区间上的全体实数, 若存在非负可积函数$f(x)$, 使得$\forall a, b \in X (a < b)$, 都有

$$P\{a < x \le b\} = \int_a^b f(x)dx,$$

则称$f(x)$为该连续型随机变量的概率分布密度函数, 记作$X \sim f(x)$.

由定义可以直接得到概率密度函数满足以下性质:

1. $f(x) \ge 0$.

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

稍加推理还可得到下面基本性质.

【定理 2.4】 连续变量取个别值的概率为0.

证明. 由于

$$0 \le P\{X = x_0\} \le P\{x_0 - \Delta x < X \le x_0\} = \int_{x_0-\Delta x}^{x_0} f(x)dx \to 0,$$

由夹逼性即得$P\{X = x_0\} = 0$. □

注意: 概率密度函数$f(x)$在$x = x_0$处的函数值表示$\varepsilon \cdot P\{X = \delta(x_0, \varepsilon)\}$, 而非$\varepsilon \cdot P\{X = x_0\}$.形式化地, 即

$$f(x_0) = \lim_{\varepsilon \to 0} \frac{\int_{x_0-\varepsilon}^{x_0+\varepsilon} f(x)dx}{\varepsilon} = \frac{P\{x_0 - \varepsilon < X < x_0 + \varepsilon\}}{\varepsilon}.$$

【例题 2.5】 设概率密度函数

$$f(x) = \begin{cases} kx + 1 & ,0 \le x \le 2 \\ 0 & ,\text{其它} \end{cases},$$

求k的值.

解.

$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^2 (kx+1)dx = 2k + 2 = 1 \Rightarrow k = -\frac{1}{2}.$$

□

2.3 分布函数

【定义 2.6】 **(分布函数(cdf))** 随机变量X的取值不超过x的概率称为分布函数. 形式化地, 定义分布函数

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

由定义可以直接得到分布函数满足以下性质:

1. $F(x)$的定义域为$(-\infty, +\infty)$, 值域为$[0, 1]$.

2. $F(x)$单调不减. 形式化地, 即$\forall x_1 < x_2, F(x_1) \le F(x_2)$.

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

4. 对于离散型随机变量, $F(x)$是右连续的($\lim_{x \to x_0^+} F(x) = F(x_0)$), 且至多有可列个间断点.

5. 对于连续型随机变量, $F(x)$是连续的.

【例题 2.7】 设分布函数

$$F(x)=\begin{cases} a-e^{-\lambda x} & ,x>0 \\ 0 & ,x\le 0 \end{cases},$$

其中$\lambda>0$, 求a的值.

解. $\lim\limits_{x\to+\infty}F(x)=a=1.$ □

【例题 2.8】 随机变量X服从的分布列如下表

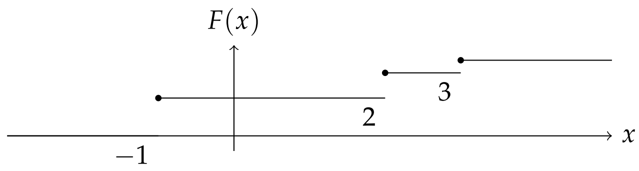
X	-1	2	3
P	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

求X的分布函数$F(x)$,并绘制$F(x)$的函数图像.

解. 由定义直接得到

$$F(x)=\begin{cases} 0 & ,x<-1 \\ \frac{1}{2} & ,-1\le x<2 \\ \frac{5}{6} & ,2\le x<3 \\ 1 & ,x\ge 3 \end{cases}.$$

函数图像为:



□

【例题 2.9】 设连续型随机变量X的概率分布密度函数为

$$f(x)=\begin{cases} -\frac{1}{2}x+1 & ,0\le x\le 2 \\ 0 & ,\text{其它} \end{cases},$$

求X的分布函数$F(x)$.

解. 由于

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

故

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

□

【例题 2.10】 设连续型随机变量X的分布函数为

$$F(x)=\begin{cases} 0 & ,x<0 \\ Ax^2 & ,0\le x<1 \\ 1 & ,x\ge 1 \end{cases},$$

求参数A.

解. 由$F(x)$在$x=1$处的连续性, 即得$A=1$. □

2.4 离散型随机变量的常见分布

【定义 2.11】(0-1分布) 分布列为

X	1	0
P	p	$1-p$

的随机变量即服从0-1分布. 该分布实际上是二项分布的在$n=1$时的特例.

【定义 2.12】(几何分布) 设$P(A)=p$. 某次试验中, 第k次实验才首次发生A, 即

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

则称此随机变量服从几何分布, 记作$X\sim G(p)$.

【定义 2.13】(二项分布) 设$P(A)=p$. 在n次实验中, A共发生了k次, 即

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

则称此随机变量服从二项分布, 记作$X\sim B(n,p)$.

【定理 2.14】(二项分布的最可能值) 设$X\sim B(n,p)$. 那么:

1. 若$(n+1)p$不为整数, 则P在$[(n+1)p]$处取得最大值.
2. 若$(n+1)p$为整数, 则P在$(n+1)p$和$(n+1)p-1$处取得最大值.

【例题 2.15】 现有一报警器, 遇到危险时报警的概率为0.8. 若希望在遇到危险时有不低于99%的概率正确报警, 则至少应安装多少台报警器?

解. 设随机变量X表示发生危险时正确报警的报警器台数, n表示安装台数. 则$X\sim B(n,p)$. 我们希望

$$P\{X\geq 1\}\geq 0.99\Rightarrow P\{X=0\}\leq 0.01.$$

也即$C_n^0 0.2^n\leq 0.01\Rightarrow n\geq\frac{\ln 0.01}{\ln 0.2}$. 取整数即$n\geq 3$. □

【定义 2.16】(Poisson分布) 概率函数为

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda},\lambda>0$$

的随机变量X服从Poisson分布, 记作$X\sim p(\lambda)$.

Poisson分布常用于近似二项分布. 当$X\sim B(n,p)$满足$n\geq 100, np\leq 10$时, 此二项分布可以用Poisson分布近似.

【例题 2.17】 证券营业部有1000个账户, 每户有存款10万元. 每户来提取20%的存款的概率是0.006. 问应当准备多少现金, 以有95%以上的把握满足用户的提款需求.

解. 设随机变量X表示提款用户数, 则$X\sim B(1000,0.006)$. 又设应当准备现金x万元, 则我们希望有

$$P\{2X\leq x\}\geq 0.95.$$

采用Poisson分布近似, 其中$\lambda=np=6$, 有

$$P\{X\leq\frac{x}{2}\}=\sum_{k=0}^{\frac{x}{2}}\frac{6^k}{k!}e^{-6}\geq 0.95.$$

查表得$\frac{x}{2}\geq 10$, 即得$x\geq 20$. □

【例题 2.18】 某疾病发病率为0.001, 某单位共有5000人, 求此单位至少两人得病的概率.

解. 设随机变量X表示患病人数, 则$X \sim B(5000, 0.001)$. 于是

$$P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\}.$$

采用Poisson分布近似, 其中$\lambda = np = 5$, 查表即得

$$P\{X \geq 2\} = 1 - \frac{5^0}{0!}e^{-5} - \frac{5^1}{1!}e^{-5} = 0.959572.$$

□

【定义 2.19】(超几何分布) 设有N个元素, 其中N_1个属于第一类, N_2个属于第二类. 从中取n个, 其中有k个属于第一类的概率为

$$P\{X = k\} = \frac{C_{N_1}^k C_{N_2}^{n-k}}{C_N^n}.$$

则称此随机变量服从超几何分布.

当$N \gg n$时, 不放回抽样对概率的影响可以近似忽略, 则可以用二项分布近似超几何分布.

【例题 2.20】 现有10000粒发芽率为99%的种子. 从中取200粒, 求其中至多1粒不发芽的概率.

解.

$$\begin{aligned}
P\{X \leq 1\} = P\{X = 0\} + P\{X = 1\} &= \frac{C_{100}^0 C_{9900}^{200}}{C_{10000}^{200}} + \frac{C_{100}^1 C_{9900}^{199}}{C_{10000}^{200}} \\
&= C_{200}^0 \times 0.01^0 \times 0.99^2 00 + C_{200}^1 \times 0.01 \times 0.99^1 99 \\
&= \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} \\
&= 0.406.
\end{aligned}$$

□

2.5 连续型随机变量的常见分布

【定义 2.21】(均匀分布) 称概率密度函数为

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

的连续型随机变量服从均匀分布, 记作$X \sim U[a, b]$, 其分布函数为

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

【例题 2.22】 已知$X \sim U[a, b]$, 并取区间$[c, d] \subset [a, b]$. 求$P\{c \leq x \leq d\}$.

解.

$$P\{c \leq x \leq d\} = \int_c^d \frac{1}{b-a}dt = \frac{d-c}{b-a}.$$

□

【定义 2.23】(指数分布) 称概率密度函数为

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

的连续型随机变量服从指数分布, 记作$X\sim Exp(\lambda)$, 其分布函数为

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

【例题 2.24】 已知$X\sim Exp(\lambda)$. 设$s>0,t>0$, 证明:$P\{X>s+t\mid X>s\}=P\{X>t\}$.

证明.

$$\begin{aligned}P\{X>s+t\mid X>s\}&=\frac{P\{X>s+t\}}{P\{X>s\}}\\&=\frac{1-F(s+t)}{1-F(s)}\\&=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}\\&=e^{-\lambda t}=1-F(t)=P\{X>t\}.\end{aligned}$$

这一例题表明指数分布具有无记忆性. □

【定义 2.25】(正态分布) 称概率密度函数为

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

的连续型随机变量服从正态分布, 记作$X\sim N(\mu,\sigma^2)$, 其分布函数为

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

特别地, 若$X\sim N(0,1)$, 则称X服从标准正态分布, 对应概率密度函数和分布函数分别为$\phi_0(x)$和$\Phi_0(x)$.注意到

$$\phi(x)=\frac{1}{\sigma}\phi_0(\frac{x-\mu}{\sigma}),\qquad\qquad \Phi(x)=\Phi_0(\frac{x-\mu}{\sigma}),$$

于是任意正态分布可以被化为标准正态分布.

【例题 2.26】 已知$X\sim N(1,4)$, 求$P\{0<X<1.6\}$.

解. 化为标准正态分布

$$P\{0<X<1.6\}=P\{\frac{0-1}{2}<\frac{X-1}{2}<\frac{1.6-1}{2}\},$$

于是$P\{0<X<1.6\}=\Phi_0(0.3)-\Phi_0(-0.5)=\Phi_0(0.3)+\Phi_0(0.5)-1$,查表可得结果. □

【定义 2.27】(上α分位数) 已知$X\sim N(0,1)$. 给定$\alpha\in(0,1)$, 称满足

$$P\{X>\mu_\alpha\}=\alpha$$

的μ_α为上α分位数. 常见的上α分位数有: $\mu_{0.05}=1.645$, $\mu_{0.025}=1.96$, $\mu_{0.01}=2.33$.

2.6 随机变量函数的分布

本节将考虑:在给定随机变量X的分布的前提下, 求$Y = f(X)$的分布.

【例题 2.28】 离散型随机变量X服从分布列:

X	-2	-1	0	1	2
P	0.2	0.1	0.5	0.1	0.1

求$Y = X^4 - 1$的分布列.

解.

Y	-1	0	15
P	0.5	0.2	0.3

□

【例题 2.29】 已知$X \sim N(\mu, \sigma^2)$, 求$Y = ax + b(a \neq 0)$的概率密度函数.

解. 当$a > 0$时,

$$F_Y(x) = P\{Y \leq x\} = P\{X \leq \frac{x-b}{a}\} = \Phi(\frac{x-b}{a}).$$

两边同时求导即得

$$f_Y(x) = \frac{1}{a}\phi(\frac{x-b}{a}) = \frac{1}{\sqrt{2\pi}a\sigma}e^{-\frac{(x-(b+a\mu))^2}{2a^2\sigma^2}}.$$

即$Y \sim N(a\mu + b, a^2\sigma^2)$.$a < 0$的情况完全同理可得. □

【定理 2.30】 已知X的密度函数为$f_X(x)$, $Y = kx + b(k \neq 0)$, 则$f_Y(x) = \frac{1}{|k|}f_X(\frac{x-b}{k})$.

证明. 当$k < 0$时,

$$F_Y(x) = P\{Y \leq x\} = P\{X \geq \frac{x-b}{k}\} = 1 - F_X(\frac{x-b}{k}).$$

两边同时求导即得

$$f_Y(x) = \frac{1}{|k|}f_X(\frac{x-b}{k}).$$

□

容易证明, 若$X \sim U(a, b)$, 则$Y = kX + c$服从相应区间上的均匀分布.

【例题 2.31】 已知$X \sim N(0, 1)$, 求$Y = X^2$的概率密度函数.

解. 当$x < 0$时,

$$F_Y(x) = P\{Y \leq x\} = P\{X^2 \leq x\} = 0.$$

当$x \geq 0$时,

$$F_Y(x) = P\{Y \leq x\} = P\{-\sqrt{x} \leq X \leq \sqrt{x}\} = 2\int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.$$

两边同时求导即得, 最终整理得

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}}e^{-\frac{x}{2}}x^{-\frac{1}{2}} & ,x > 0 \\ 0 & ,x \leq 0 \end{cases}.$$

特别地, 我们称Y服从卡方分布, 记作$Y \sim \mathcal{X}^2$. □

Chapter 3

多维随机变量及其分布

3.1 二维随机变量

【定义 3.1】(二维随机变量) 设随机试验E的样本空间为Ω,X,Y是定义在Ω的两个随机变量,则称向量(X,Y)为Ω上的二维随机变量.

【定义 3.2】(联合分布) 定义二元函数

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

为X和Y的联合分布函数.

由定义可以直接得到联合分布函数满足以下性质:

1. $0\leq F(x,y)\leq 1$.
2. $F(x)$单调不减. 形式化地, 即$\forall x_1<x_2,F(x_1,y)\leq F(x_2,y);\forall y_1<y_2,F(x,y_1)\leq F(x,y_2)$.
3. $\lim\limits_{\substack{x\to+\infty\\y\to+\infty}}F(x,y)=1,\ \lim\limits_{\substack{x\to-\infty\\y\to-\infty}}F(x,y)=0$.
4. $F(x,y)$分别关于x和y右连续.
5. $F(x,y)$是在二维平面上的前缀和. 形式化地, $\forall x_1<x_2,y_1<y_2$, 有

$$P\{x_1\leq X\leq x_2,y_1\leq Y\leq y_2\}=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1).$$

【定义 3.3】(边缘分布) 称分布函数

$$F_X(x)=\lim_{y\to+\infty}F(x,y),$$
$$F_Y(y)=\lim_{x\to+\infty}F(x,y).$$

为边缘分布函数.

3.2 二维离散型随机变量的联合分布及边缘分布

【定义 3.4】(分布表) 称形如

X \ Y	Y_1	Y_2	Y_3
X_1	p_{11}	p_{12}	p_{13}
X_2	p_{21}	p_{22}	p_{23}
X_3	p_{31}	p_{32}	p_{33}

的表格为二维离散型随机变量的分布表.显然它满足性质:

1. $p_{ij} \geq 0$.
2. $\sum\limits_i \sum\limits_j p_{ij} = 1$.

二维离散型随机变量的联合分布及边缘分布可由分布表直接得出.

3.3 二维连续型随机变量的联合密度及边缘密度

【定义 3.5】(联合密度) 定义二元函数$f(x,y)$, 使得

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

则称这样的f为二维连续型随机变量的联合密度.显然它满足性质:

1. $f(x,y) \geq 0$.
2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dxdy = 1$.
3. $\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$.
4. 设G是xy平面内的一个区域, 则$P\{(X,Y) \in G\} = \iint\limits_G f(x,y)dxdy$.

【例题 3.6】 已知二维连续型随机变量的联合密度为

$$f(x,y) = \begin{cases} e^{-(x+y)} & ,x > 0, y > 0 \\ 0 & ,\text{else} \end{cases},$$

求(1)X,Y的分布函数$F(x,y)$. (2)设$G = \{(x,y) \mid x > 0, y > 0, x + y < 1\}$, 求$P\{(X,Y) \in G\}$. (3) 求$F_X(x)$.

<u>**解.**</u> (1) 当$x > 0$且$y > 0$时,

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)dsdt = \int_0^x e^{-s}ds \int_0^y e^{-t}dt = (1 - e^{-x})(1 - e^{-y}).$$

故

$$F(x,y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}) & ,x > 0, y > 0 \\ 0 & ,\text{else} \end{cases}.$$

(2)

$$P\{(X,Y) \in G\} = \iint\limits_G f(x,y)dxdy = 1 - 2e^{-1}.$$

(3)

$$F_X(x) = \lim_{y \to +\infty} F(x,y) = \begin{cases} 1 - e^{-x} & ,x > 0, y > 0 \\ 0 & ,\text{else} \end{cases}$$

□

【定义 3.7】(边缘密度) 对边缘分布函数

$$F_X(x)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}f(s,t)dsdt,$$

两边同时求导, 即得边缘密度函数

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

特别地, 二维正态分布满足以下性质:

1. 二维正态分布的边缘分布也是正态分布.
2. 边缘分布均为正态分布无法保证二维随机变量服从正态分布.

【例题 3.8】 已知联合密度

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

求$f_X(x)$.

解. 当$|x|\le r$时,

$$f_X(x)=\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}}\frac{1}{\pi r^2}dy=\frac{2\sqrt{r^2-x^2}}{\pi r^2}.$$

当$|x|>r$时, $f_X(x)=0$.综上,

$$f_X(x)=\begin{cases}\frac{2\sqrt{r^2-x^2}}{\pi r^2} & ,|x|\le r\\ 0 & ,|x|>r\end{cases},$$

□

3.4 条件分布

【定义 3.9】(条件分布) 称函数

$$F(x\mid A)=P\{X\le x\mid A\}$$

为条件分布函数.

【例题 3.10】 求$f(x)=\frac{1}{\pi(1+x^2)}$在$x>1$的条件下的条件分布.

解.

$$F(x\mid X>1)=P\{X\le x\mid X>1\}=\frac{P\{X\le x,X>1\}}{P\{X>1\}}.$$

当$x\le 1$时,

$$F(x\mid X>1)=0,$$

当$x>1$时,

$$F(x\mid X>1)=\frac{P\{1<X\le x\}}{P\{X>1\}}=\frac{\int_1^x\frac{1}{\pi(1+t^2)}dt}{\int_1^{+\infty}\frac{1}{\pi(1+t^2)}dt}=\frac{4}{\pi}\arctan x-1.$$

综上,

$$F(x\mid X>1)=\begin{cases}0 & ,x\le 1\\ \frac{4}{\pi}\arctan x-1 & ,x>1\end{cases}.$$

□

对于离散型随机变量的条件分布, 直接根据分布表写出即可.

对于连续型随机变量的条件分布, 设已知其联合密度$f(x,y)$和边缘密度$f_X(x),f_Y(y)$, 则在$Y=y$的条件下, 有

$$F(x\mid y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du,\qquad f(x\mid y)=\frac{f(x,y)}{f_Y(y)}.$$

3.5 随机变量的独立性

【定理 3.11】 对于二维随机变量(X,Y), 以下命题等价:

1. X,Y独立.
2. $f(x,y)=f_X(x)f_Y(y)$.
3. $F(x,y)=F_X(x)F_Y(y)$.
4. $P\{X\in S_X,Y\in S_Y\}=P\{X\in S_X\}P\{Y\in S_Y\}$.

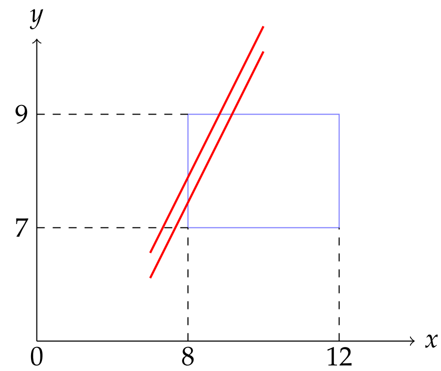
【例题 3.12】 已知甲乙两人上午打卡上班的时间均为均匀分布, 甲将于$8\sim 12$时到岗, 乙将于$7\sim 9$时到岗, 且二人到岗时间独立. 求两人到岗时间不超过5分钟的概率.

解. 依题意, 有

$$f_X(x)=\begin{cases}\frac{1}{4} & ,8<x<12\\ 0 & ,\text{else}\end{cases}\qquad f_Y(y)=\begin{cases}\frac{1}{2} & ,7<x<9\\ 0 & ,\text{else}\end{cases}$$

由于X,Y独立,故

$$f(x,y)=f_X(x)f_Y(y)=\begin{cases}\frac{1}{8} & ,8<x<12,7<y<9\\ 0 & ,\text{else}\end{cases}$$



如右图, 设直线$y=x-\frac{1}{12}$和直线$y=x+\frac{1}{12}$与矩形围成的区域为G, 则

$$P\{|X-Y|\le\frac{1}{12}\}=\iint\limits_G f(x,y)dxdy=\frac{1}{48}.$$

□

【定理 3.13】 若X,Y独立, 则其随机变量函数$g(X),g(Y)$同样独立.

证明. 联合矩生成函数为

$$E(e^{tf(X)+sg(Y)})=\iint\limits_{X,Y} e^{tf(X)+sg(Y)}\prod_X P_{x_i}(x_i)\prod_Y P_{y_i}(y_i)dxdy.$$

边缘矩生成函数为

$$E(e^{tf(X)})E(e^{sg(Y)})=\int\limits_X e^{tf(X)}\prod_X P_{x_i}(x_i)dx\int\limits_Y e^{sg(Y)}\prod_Y P_{y_i}(y_i)dy.$$

二者显然相等. □

3.6 随机变量函数的分布

二维离散型随机变量的分布直接将分布表对应元素相乘即可, 下面考虑二维连续型随机变量的分布.

已知二维随机变量(X, Y), 并已知其密度函数$f(x, y)$, 设随机变量$Z = g(X, Y)$, 则

$$F_Z(\mathfrak{z}) = P\{Z \le \mathfrak{z}\} = \iint\limits_{D_\mathfrak{z}} f(x, y)dxdy,$$

其中$D_\mathfrak{z} = \{(x, y) \mid g(x, y) \le \mathfrak{z}\}$.

【例题 3.14】 已知(X, Y)独立, 它们的密度函数为$f(x, y)$. 设随机变量$Z = X + Y$, 求$F_Z(\mathfrak{z})$和$f_Z(\mathfrak{z})$.

<u>**解.**</u> 设$D_\mathfrak{z} = \{(x, y) \mid x + y \le \mathfrak{z}\}$, 则

$$F_Z(\mathfrak{z}) = \iint\limits_{D_\mathfrak{z}} f(x, y)dxdy = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{\mathfrak{z}-x} f(x, y)dy.$$

做换元$t = x + y$, 则

$$F_Z(\mathfrak{z}) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{\mathfrak{z}} f(x, t - x)dt = \int_{-\infty}^{\mathfrak{z}} \int_{-\infty}^{+\infty} f(x, t - x)dxdt.$$

求导, 并基于变量的独立性将联合密度展开为边缘密度, 即得

$$f_Z(\mathfrak{z}) = \int_{-\infty}^{+\infty} f(x, \mathfrak{z} - x)dx = \int_{-\infty}^{+\infty} f_X(x)f_Y(\mathfrak{z} - x)dx.$$

这一公式被称为卷积公式. □

【例题 3.15】 已知(X, Y)的边缘密度函数为$F_X(x), F_Y(y)$. 设随机变量$M = \max\{X, Y\}, N = \min\{X, Y\}$. 求$F_M(\mathfrak{z}), F_N(\mathfrak{z})$.

<u>**解.**</u> 由于$\{M \le \mathfrak{z}\} = \{X \le \mathfrak{z}, Y \le \mathfrak{z}\}$, 故

$$F_M(\mathfrak{z}) = P\{X \le \mathfrak{z}, Y \le \mathfrak{z}\} = P\{X \le \mathfrak{z}\}P\{Y \le \mathfrak{z}\} = F_X(\mathfrak{z})F_Y(\mathfrak{z}).$$

由于$\{N > \mathfrak{z}\} = \{X > \mathfrak{z}, Y > \mathfrak{z}\}$, 故

$$F_N(\mathfrak{z}) = 1 - P(N > \mathfrak{z}) = 1 - P\{X > \mathfrak{z}, Y > \mathfrak{z}\} = 1 - P\{X > \mathfrak{z}\}P\{Y > \mathfrak{z}\} = 1 - (1 - F_X(\mathfrak{z}))(1 - F_Y(\mathfrak{z})).$$

□

Chapter 4

数值特征

4.1 数学期望

【定义 4.1】(离散型的期望) 设离散型随机变量X的概率密度为$P\{X=x_k\}=p_k$. 若$\sum\limits_{k=1}^{\infty} x_k p_k$绝对收敛, 则称

$$E(X)=\sum_{k=1}^{\infty} x_k p_k$$

为X的数学期望.

【定义 4.2】(连续型的期望) 设连续型随机变量X的概率密度为$f(x)$. 若$\int_{-\infty}^{+\infty} x f(x) dx$绝对收敛, 则称

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

为X的数学期望.

【定义 4.3】(离散型随机变量函数的期望) 设离散型随机变量X的概率密度为$P\{X=x_k\}=p_k$. 设随机变量$Y=g(X)$. 若$\sum\limits_{k=1}^{\infty} g(x_k) p_k$绝对收敛, 则称

$$E(Y)=\sum_{k=1}^{\infty} g(x_k) p_k$$

为Y的数学期望.

【定义 4.4】(连续型随机变量函数的期望) 设连续型随机变量X的概率密度为$f(x)$. 设随机变量$Y=g(X)$. 若$\int_{-\infty}^{+\infty} g(x) f(x) dx$绝对收敛, 则称

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

为Y的数学期望.

【定义 4.5】(二维离散型变量函数的期望) 设$Z=g(X,Y)$, 其中X,Y均为离散型随机变量, 则称

$$E(Z)=\sum_{i}\sum_{j} g(x_i, y_j) p_{ij}$$

为Z的数学期望.

【定义 4.6】(二维连续型变量函数的期望) 设$Z=g(X,Y)$, 其中X,Y均为连续型随机变量, 则称

$$E(Z)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f(x,y) dx dy$$

为Z的数学期望.

注意: 高维随机变量的期望维度取决于变量函数Z的维度.

【例题 4.7】 设某货物的需求量$X \sim U[2000,4000]$. 每出口10吨货物可以获利3万, 每10吨货物滞销将损失1万. 求达到最大获利的出口量安排.

解. 设出口量为y吨, 随机变量Y为获利额. 则

$$Y = g(X) = \begin{cases} 3y & ,x \geq y \\ 3x-(y-x) & ,x < y \end{cases},$$

又概率密度函数

$$f(x) = \begin{cases} \frac{1}{2000} & ,2000 \leq x \leq 4000 \\ 0 & ,\text{else} \end{cases}.$$

故

$$E(Y) = \int_{-\infty}^{+\infty} g(x)f(x)dx = \frac{1}{2000}(\int_{2000}^{y}(4x-y)dx + \int_{y}^{4000} 3ydx) = \frac{1}{1000}(-y^2+7000y-4\times 10^6),$$

当$y = 3500$时上式取得最大值. □

数学期望满足下列性质(其中C是常数):

1. $E(C) = C$.
2. $E(X+C) = E(X)+C$.
3. $E(CX) = C \cdot E(X)$.
4. $E(X \pm Y) = E(X) \pm E(Y)$.
5. 若X,Y独立, 则$E(XY) = E(X)E(Y)$.

下面证明上面的4,5两条性质.

证明. 性质4:

$$\begin{aligned} E(X \pm Y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x \pm y)f(x,y)dxdy \\ &= \int_{-\infty}^{+\infty} xdx \int_{-\infty}^{+\infty} f(x,y)dy \pm \int_{-\infty}^{+\infty} ydy \int_{-\infty}^{+\infty} f(x,y)dx \\ &= \int_{-\infty}^{+\infty} xf_X(x)dx \pm \int_{-\infty}^{+\infty} yf_Y(y)dy \\ &= E(X) \pm E(Y). \end{aligned}$$

性质5:

$$E(XY) = \int_{-\infty}^{+\infty} xyf(x,y)dxdy = \int_{-\infty}^{+\infty} xf_X(x)dx \int_{-\infty}^{+\infty} yf_Y(y)dy = E(x)E(Y).$$

□

【例题 4.8】 现有N件产品, 其中混有M件次品. 从中任取n件($n \leq M \leq N$), 求取出次品个数的期望.

解. 对次品标号为$1 \sim M$, 其中$x_i = \begin{cases} 1 & ,\text{取到第}i\text{件次品} \\ 0 & ,\text{未取到第}i\text{件次品} \end{cases}$.设$X$表示取到次品的个数, 则$X = \sum\limits_{i=1}^{M} x_i$. 由于

$$P\{x_i = 1\} = \frac{C_{N-1}^{n-1}}{C_N^n} = \frac{n}{N}.$$

故

$$E(X_i) = \frac{n}{N} \Rightarrow E(X) = ME(X_i) = \frac{nM}{N}.$$

□

【**定义 4.9**】**(条件期望)** 称在一个变量取某值的条件下，另一个变量的期望为条件期望.
对于离散型,

$$E(X \mid Y = y_j) = \sum x_i P\{X = x_i \mid Y = y_j\}.$$

对于连续型,

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

事实上，条件期望满足数学期望的一切性质.

【**定理 4.10**】**(law of total expectation)** 设离散型二维随机变量(X_1, X_2)，则$E(X_1) = E[E(X_1 \mid X_2)]$.

证明.

$$\begin{aligned}
right &= \sum_{x_2} \sum_{x_1} x_1 P_{X_1|X_2}(x_1 \mid x_2) \\
&= \sum_{x_2} \sum_{x_1} x_1 \frac{P(X_1, X_2)}{P_{X_2} x_2} P_{X_2}(x_2) \\
&= \sum_{x_1} x_1 \sum_{x_2} P(X_1, X_2) \\
&= \sum_{x_1} x_1 P_{X_1}(x_1) \\
&= E(X_1) = left.
\end{aligned}$$

□

4.2 方差

【**定义 4.11**】**(方差)** 使用方差来描述数据的偏离程度，其定义为

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

【**定义 4.12**】**(标准差)** 为了统一量纲，我们定义标准差为

$$\sigma(X) = \sqrt{D(X)}.$$

【例题 4.13】 已知随机变量X的密度函数为

$$f(x) = \begin{cases} ax^2 + bx + c & ,0 < x < 1 \\ 0 & ,\text{else} \end{cases},$$

且已知$E(X) = 0.5, D(X) = 0.15$.确定参数a, b, c.

解. 由于

$$E(X) = \int_0^1 x(ax^2 + bx + c) dx \qquad E(X^2) = \int_0^1 x^2(ax^2 + bx + c) dx = \frac{a}{5} + \frac{b}{4} + \frac{c}{3}.$$

故

$$\begin{cases} \int_0^1 (ax^2 + bx + c) dx = 1 \\ E(X) = 0.5 \\ E(X^2) - E^2(X) = 0.15 \end{cases} \Rightarrow \begin{cases} a = 12 \\ b = -12 \\ c = 3 \end{cases}.$$

□

方差满足下列性质(其中C是常数):

1. $D(C) = 0$.

2. $D(X+C)=D(X)$.

3. $D(CX)=C^2\cdot D(X)$.

4. $D(X\pm Y)=D(X)+D(Y)\pm 2\text{Cov}(X,Y)$.特别地, 若$X,Y$独立, 则$D(X\pm Y)=D(X)+D(Y)$.

5. $D(X)=0$的充要条件是$P\{X=E(X)\}=1$.

下面证明第4条性质:

证明.

$$\begin{aligned}D(X\pm Y)&=E[(X\pm Y)^2]-E^2(X\pm Y)\\&=E(X^2)+E(Y^2)\pm 2E(XY)-[E(X)\pm E(Y)]^2\\&=E(X^2)-E^2(X)+E(Y^2)-E^2(Y)\pm 2(E(XY)-E(X)E(Y))\\&=D(X)+D(Y)\pm \text{Cov}(X,Y).\end{aligned}$$

□

常见分布的期望与方差如下表:

分布	$E(X)$	$D(X)$
0-1分布	p	$p(1-p)$
二项分布$B(n,p)$	np	$np(1-p)$
几何分布$G(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson分布$p(\lambda)$	λ	λ
均匀分布$U[a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布$Exp(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态分布$N(\mu,\sigma^2)$	μ	σ^2

4.3 协方差

【定义 4.14】(协方差) 定义协方差

$$\text{Cov}(X,Y)=E[(X-E(X))(Y-E(Y))].$$

进一步地, 可以推出

$$\text{Cov}(X,Y)=E(XY)-E(X)E(Y).$$

注意, X,Y独立仅为$\text{Cov}(X,Y)=0$的充分不必要条件.存在反例:

X \ Y	-1	0	1	
-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
0	$\frac{1}{8}$	0	$\frac{1}{8}$	$\frac{1}{4}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	

显然X,Y不独立, 而由于边缘分布

X	-1	0	1
P	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$

Y	-1	0	1
P	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$

故$E(X)=0, E(Y)=0, E(X+Y)=0$. 从而$\text{Cov}(X,Y)=0$.

【例题 4.15】 已知概率密度函数

$$f(x,y)=\begin{cases} x+y & ,0\le x\le 1,0\le y\le 1 \\ 0 & ,\text{else} \end{cases},$$

求$\mathrm{Cov}(X,Y)$.

解. 由于

$$f_X(x)=\begin{cases} x+\frac{1}{2} & ,0\le x\le 1 \\ 0 & ,\text{else} \end{cases} \qquad f_Y(y)=\begin{cases} y+\frac{1}{2} & ,0\le y\le 1 \\ 0 & ,\text{else} \end{cases}$$

故$E(X)=\int_0^1 x(x+\frac{1}{2})dx=\frac{7}{12}$, $E(X)=\int_0^1 y(y+\frac{1}{2})dy=\frac{7}{12}$, 且 $E(XY)=\int_0^1\int_0^1 xy(x+y)dxdy=\frac{1}{3}$, 故

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=-\frac{1}{144}.$$

□

协方差满足下列性质:

1. $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$.
2. $\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$.
3. $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$.
4. $\mathrm{Cov}(C,X)=0$.
5. 若X,Y独立, 则$\mathrm{Cov}(X,Y)=0$.
6. 若$\mathrm{Cov}(X,Y)=0$, 则X,Y不相关.

4.4 相关系数

为了消除度量单位对协方差的影响, 我们引入标准化随机变量$X^*=\frac{X-E(X)}{\sqrt{D(X)}}$, 其满足

$$E(X^*)=0 \qquad D(X^*)=1.$$

此时

$$\mathrm{Cov}(X^*,Y^*)=E(X^*Y^*)-E(X^*)E(Y^*)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}.$$

【定义 4.16】(相关系数) 对于随机变量(X,Y), 定义相关系数

$$\rho=\mathrm{Cov}(X^*,Y^*)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}.$$

【定理 4.17】 任意随机变量的相关系数ρ均满足

$$|\rho|\le 1.$$

证明. 由Cauchy-Schwarz不等式, 有

$$E^2(XY)\le E(X^2)E(Y^2).$$

令$X_1=X-E(X), Y_1=Y-E(Y)$, 则(X,Y)的相关系数ρ满足

$$\rho^2=\frac{E^2[(X-E(X))(Y-E(Y))]}{D(X)D(Y)}=\frac{E^2(X_1,Y_1)}{E(X_1^2)E(Y_1^2)}\le 1.$$

也即$|\rho|\le 1$. □

【定理 4.18】 $|\rho| = 1$ 等价于 X, Y成线性关系.具体来说:

1. $\rho = 1$时, 称X, Y完全正相关.

2. $\rho = -1$时, 称X, Y完全负相关.

3. $|\rho|$接近0时, 表明X, Y线性关系弱.

4. $\rho = 0$时, 表明X, Y不存在线性关系.

【例题 4.19】 已知$X \sim N(0, 1)$. 设随机变量函数$Y = X^2$. 求(X, Y)的相关系数ρ.

解. 由于$E(X) = 0$, $E(X^2) = 1$, $E(X^3) = 0$, 故

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X^3) - E(X)E(X^2) = 0.$$

于是$\rho = 0$. □

本例表明$\rho = 0$只保证X, Y无线性关系, 但不能说明X, Y独立. 事实上, 仅对于二维正态分布(X, Y), $\rho = 0$与两变量独立等价.

4.5 中心矩与原点矩

【定义 4.20】 **(原点矩)** 定义原点矩为

$$A_k = E(X^k),$$

特别地, 称期望$E(X)$为一阶原点矩.

【定义 4.21】 **(中心矩)** 定义中心矩为

$$B_k = E[(X - E(X))^k],$$

特别地, 一阶中心矩$E(X - E(X)) = 0$, 二阶中心矩$E[(X - E(X))^2] = D(X)$.

实际运用中, 一般不会用到超过4阶的原点矩或中心矩.

4.6 矩生成函数(MGF)

【定义 4.22】 **(矩生成函数)** 对于随机变量X, 其MGF为

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx.$$

一个重要的性质是, MGF与分布函数一一对应, 也即MGF可以唯一确定分布类型.

另外一个性质是, 可以通过对MGF求t的偏导, 求出任意原点矩, 即

$$\frac{\partial^k M_X(t)}{\partial t^k} \mid_{t=0} = (E(e^{tX} X^k)) \mid_{t=0} = E(X^k).$$

【定义 4.23】 **(联合矩生成函数)** 设n维随机变量X, 其联合矩生成函数为

$$M_X(t) = E(e^{t^T X}).$$

Chapter 5

大数定律与中心极限定理

5.1 大数定律

大数定律是用于描述大量重复试验的平均结果的稳定性的一系列相关定律.

【定理 5.1】 **(Chebyshev不等式)** 对于随机变量X, 若$E(X)$和$D(X)$存在. 则$\forall \varepsilon > 0$, 有

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

<u>**证明.**</u> 当X是连续型随机变量时,

$$\int\limits_{|X-E(X)|\geq\varepsilon} f(x)dx \leq \int\limits_{|X-E(X)|\geq\varepsilon} \frac{(x-E(X))^2}{\varepsilon^2} f(x)dx \leq \int_{-\infty}^{+\infty} \frac{(x_k-E(X))^2}{\varepsilon^2} f(x)dx = \frac{D(X)}{\varepsilon^2}.$$

当X是离散型随机变量时,

$$\sum_{|X-E(X)|\geq\varepsilon} p_k \leq \sum_{|X-E(X)|\geq\varepsilon} \frac{(x_k-E(X))^2}{\varepsilon^2} p_k \leq \sum_{k=1}^{+\infty} \frac{(x_k-E(X))^2}{\varepsilon^2} p_k = \frac{D(X)}{\varepsilon^2}.$$

□

【例题 5.2】 已知$X \sim N(\mu, \sigma^2)$, 用Chebyshev不等式估计$P\{|X-\mu| \geq 3\sigma\}$的值, 并与确切值作比较.

<u>**解.**</u> 直接套用Chebyshev不等式, 有

$$P\{|X-\mu| \geq 3\sigma\} \leq \frac{D(X)}{(3\sigma)^2} = \frac{1}{9}.$$

计算精确值:

$$P\{|X-\mu| \geq 3\sigma\} = 1 - P\{-3 < \frac{X-\mu}{\sigma} < 3\} = 2 - 2\Phi_0(3) \approx 0.0027.$$

□

【定理 5.3】 **(Bernoulli大数定律)** 设有n重Bernoulli实验, 其中事件A发生了m_n次. 设概率$P(A) = p$, 以及事件A发生的频率$\frac{m_n}{n}$. 则有

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

<u>**证明.**</u> 由于$m_n \sim B(n,p)$, 则$E(\frac{m_n}{n}) = p, D(\frac{m_n}{n}) = p(1-p)$.由Chebyshev不等式, 有

$$1 \geq \lim_{n\to\infty} P\{\left|\frac{m_n}{n} - p\right| < \varepsilon\} \geq \lim_{n\to\infty}(1 - \frac{p(1-p)}{n\varepsilon^2}) = 1.$$

由夹逼性即得证. □

【定理 5.4】 (Chebyshev大数定律) 设随机变量$X_1, X_2, ..., X_n$两两不相关. 其中$E(X_i), D(X_i)$都存在且满足$D(X_i) \le M$. 则$\forall \varepsilon > 0$, 有

$$\lim_{n\to\infty} P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\} = 1.$$

证明. 由于$X_1, X_2, ..., X_n$两两不相关, 故$\forall i, j \in [1, n]$, 均有

$$\mathrm{Cov}(X_i, X_j) = 0.$$

故

$$D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) \le \frac{M}{n}.$$

又$E(\frac{1}{n}\sum\limits_{i=1}^{n} X_i) = \frac{1}{n}\sum\limits_{i=1}^{n} E(X_i)$,故

$$1 \ge \lim_{n\to\infty} P\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\} \ge \lim_{n\to\infty}(1 - \frac{D(\frac{1}{n}\sum\limits_{i=1}^{n} X_i)}{\varepsilon^2}) \ge \lim_{n\to\infty}(1 - \frac{M}{n\varepsilon^2}) = 1.$$

由夹逼性即得证. □

【定理 5.5】 (辛钦大数定律) 设随机变量$X_1, X_2, ..., X_n$独立同分布. 其中$E(X_i) = \mu$. 则

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

5.2 中心极限定理

【定理 5.6】 (中心极限定理) 设随机变量$X_1, X_2, ..., X_n$独立同分布. 其中$E(X_i) = \mu, D(X_i) = \sigma^2$, 其中$0 < \sigma^2 < +\infty$.则

$$\lim_{n\to\infty} P\{\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\} = \Phi_0(x).$$

中心极限定理表明大量独立同分布的变量和的极限分布是正态分布.也即

$$\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

可以被近似为$N(0, 1)$. 或$\sum\limits_{i=1}^{n} X_i$可以被近似为$N(n\mu, n\sigma^2)$.

【例题 5.7】 某商店每天接待顾客100人, 每个顾客的消费额度为$U[0, 60]$, 且每个消费者消费的金额独立. 求日销售额超3500元的概率.

解. 设随机变量X_i表示第i人的消费数额, 则$E(X_i) = 30$, $D(X_i) = \frac{60^2}{12} = 300$. 由中心极限定理,

$$\frac{\sum\limits_{i=1}^{100} X_i - 3000}{100\sqrt{3}}$$

可以被近似为$N(0, 1)$. 于是

$$P\{\sum_{i=1}^{100} X_i > 3500\} = 1 - P\{\frac{\sum\limits_{i=1}^{100} X_i - 3000}{100\sqrt{3}} \le \frac{3500 - 3000}{100\sqrt{3}}\} \approx 1 - \Phi_0(2.887) \approx 0.002.$$

□

【定理 5.8】 **(De Moivre - Laplace Theorem)** 设$Y_n \sim B(n,p)$. 则

$$\lim_{n\to\infty} P\{\frac{Y_n - np}{\sqrt{np(1-p)}} \le x\} = \Phi_0(x).$$

<u>**证明.**</u> 设

$$Y_n = \sum_{i=1}^{n} x_i,$$

其中$x_i = \begin{cases} 1 & ,第i个事件发生 \\ 0 & ,第i个事件不发生 \end{cases}$.于是由中心极限定理直接得证. □

De Moivre - Laplace定理表明二项分布可用正态分布近似.也即

$$\frac{Y_n - np}{\sqrt{np(1-p)}}$$

可以被近似为$N(0,1)$. 或Y_n可以被近似为$N(np, np(1-p))$.

【例题 5.9】 已知某保险的参保人群死亡率为0.005. 求在1万个参保人中, (1)最终死亡人数不超过70的概率. (2)最终死亡人数为k的概率.

<u>**解.**</u> (1) 设随机变量X表示死亡人数, 则$X \sim B(10^4, 0.005)$. 由De Moivre - Laplace定理, 有

$$P\{X \le 70\} = P\{\frac{X - np}{\sqrt{np(1-p)}} \le \frac{70 - np}{\sqrt{np(1-p)}}\} \approx \Phi_0(2.84) \approx 0.9977.$$

(2)

$$\begin{aligned} P\{X = k\} = P\{k - \frac{1}{2} < X < k + \frac{1}{2}\} &= P\{\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} < \frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\} \\ &= \Phi_0(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}) - \Phi_0(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}). \end{aligned}$$

□

Chapter 6

样本及抽样分布

6.1 总体与样本

统计学中, 总体指研究对象的全部, 样本指从总体中随机抽取的部分元素.

在研究过程中, 称想要获得的数据为变量, 记作X_i. 当某一变量的具体值被确定, 则称该值为该变量的观测值, 记作x_i.

若变量X服从同分布, 且相互独立, 则可以进行简单随机抽样. 显然, 简单随机抽样是在理想状态的下的抽样方法, 一般情况下无法适用.

【定理 6.1】(抽样分布) 对于简单随机抽样产生的样本$(X_1, X_2, ..., X_n)$, 由于样本之间相互独立, 故有

$$F(X_1, ..., X_n) = F(\prod_{i=1}^{n} X_i)$$
$$f(X_1, ..., X_n) = f(\prod_{i=1}^{n} X_i)$$
$$P\{X_1 = x_1, ..., X_n = x_n\} = \prod_{i=1}^{n} P\{X_i = x_i\}.$$

【例题 6.2】 已知X服从$0-1$分布, 求样本$(X_1, ..., X_n)$的联合概率分布.

解. 由于X服从$0-1$分布, 故

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

故

$$\begin{aligned} P\{X_1 = x_1, ..., X_n = x_n\} &= \prod_{i=1}^{n} P\{X_i = x_i\} \\ &= p^{x_1+...+x_n}(1-p)^{n-(x_1+...+x_n)}. \end{aligned}$$

□

6.2 统计量

【定义 6.3】**(统计量)** 称不含任何未知参数的由样本构成的函数为统计量.

常见统计量如下表

统计量	定义式
样本均值	$\overline{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$
未修正的样本方差	$s_0^2=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$
样本方差	$S^2=\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$
样本标准差	$S=\sqrt{S^2}$
样本k阶原点矩	$A_k=\frac{1}{n}\sum\limits_{i=1}^{n}X_i^k$
样本k阶中心矩	$B_k=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^k$

【例题 6.4】 已知$X\sim p(\lambda)$, 求样本$(X_1,...,X_n)$均值的期望和方差.

<u>**解.**</u> 由于$E(X)=D(X)=\lambda$, 故

$$E(\overline{X})=E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\lambda.$$

以及

$$D(\overline{X})=D(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}D(\sum_{i=1}^{n}X_i)=\frac{\lambda}{n}.$$

□

【定理 6.5】 设总体X的均值为$E(X)=\mu$, 方差为$D(X)=\sigma^2$. 则其中的样本$(X_1,...,X_n)$满足

1. $E(\overline{X})=\mu$.
2. $D(\overline{X})=\frac{\sigma^2}{n}$.
3. $E(S^2)=\sigma^2$.

性质1, 2的证明过程见例题6.4. 下面证明性质3:$E(S^2)=\sigma^2$.

<u>**证明.**</u> 由于

$$\begin{aligned}S^2&=\frac{1}{n-1}(\sum_{i=1}^{n}(X_i-\mu)^2+n(\overline{X}-\mu)^2-2(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu))\\&=\frac{1}{n-1}(\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2).\end{aligned}$$

故

$$\begin{aligned}E(S^2)&=\frac{1}{n-1}[\sum_{i=1}^{n}E((X_i-\mu)^2)-nE((\overline{X}-\mu)^2)]\\&=\frac{1}{n-1}(\sum_{i=1}^{n}D(X_i)-nD(\overline{X}))\\&=\frac{1}{n-1}(n\sigma^2-n\frac{\sigma^2}{n})=\sigma^2.\end{aligned}$$

□

6.3 抽样分布

【定义 6.6】(抽样分布) 考察统计量的分布即为抽样分布. 常见抽样分布有:正态分布、$\mathcal{X}^2$分布、t分布、F分布.

【定义 6.7】(卡方分布) 若n个相互独立的随机变量$X_1, X_2, ..., X_n$均服从标准正态分布. 则称随机变量函数

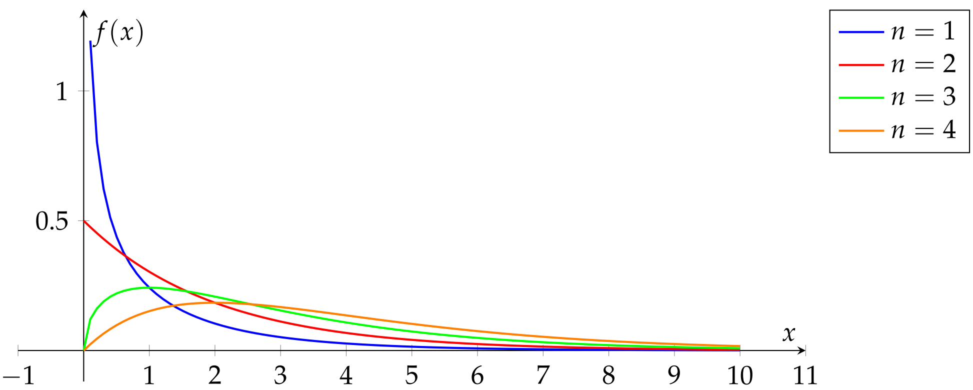
$$Y = \sum_{i=1}^{n} X_i^2$$

服从卡方分布, 记作$Y \sim \mathcal{X}^2(n)$. 其概率密度函数为

$$f(y) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} & ,y > 0 \\ 0 & ,y \leq 0 \end{cases},$$

其中$\Gamma(z) = \int_0^{+\infty} x^{z-1}e^{-x}dx(z > 0)$.部分图像如下:

χ^2 Distribution for $n = 1, 2, 3, 4$



由定义知卡方分布满足以下性质:

1. $\mathcal{X}^2(2)$是一个$\lambda = \frac{1}{2}$的指数分布.
2. 卡方分布的图像是单峰函数, 峰值在$x = n - 2$处取到.
3. 当n很大时, 卡方分布可用正态分布近似. 由中心极限定理知$\frac{Y-n}{\sqrt{2n}} \sim N(0, 1)$.

【命题 6.8】 设$Y \sim \mathcal{X}^2(n)$, 则$E(Y) = n, D(Y) = 2n$.

<u>**证明.**</u>

$$\begin{aligned} E(X_i^4) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} [(-xe^{-\frac{x^2}{2}})|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx] \\ &= 1. \end{aligned}$$

同理通过分部积分, 得

$$E(X_i^2) = E(X_i^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}} dx = 3.$$

故

$$\begin{aligned} E(Y) &= nE(X_i^2) = n. \\ D(Y) &= nD(X_i^2) = n[E(X_i^4) - E^2(X_i^2)] = 2n. \end{aligned}$$

□

【定理 6.9】 设$X \sim \mathcal{X}^2(n), Y \sim \mathcal{X}^2(m)$. 且已知$X, Y$独立. 则$X+Y \sim \mathcal{X}^2(m+n)$.

【例题 6.10】 设随机变量$X_1, ..., X_6$均服从$N(0, 2^2)$. 求$P\{\sum_{i=1}^{6} X_i^2 > 6.54\}$.

解. 标准化处理得$\frac{X_1}{2}, ..., \frac{X_6}{2}$均服从$N(0,1)$, 故

$$\sum_{i=1}^{6}(\frac{X_i}{2})^2 \sim \mathcal{X}^2(6).$$

故$P\{\frac{\sum_{i=1}^{6} X_i^2}{4} > \frac{6.54}{4}\} = \alpha$, 其中$\alpha$满足

$$\mathcal{X}_\alpha^2(6) = \frac{6.54}{4} = 1.635.$$

查表得$\alpha = 0.95$(其中$\mathcal{X}_\alpha^2(n)$表示$\mathcal{X}^2(n)$的上α分位数). □

【定义 6.11】(*t*分布) 设随机变量$X \sim N(0,1), Y \sim \mathcal{X}^2(n)$, 且$X, Y$独立. 则随机变量函数

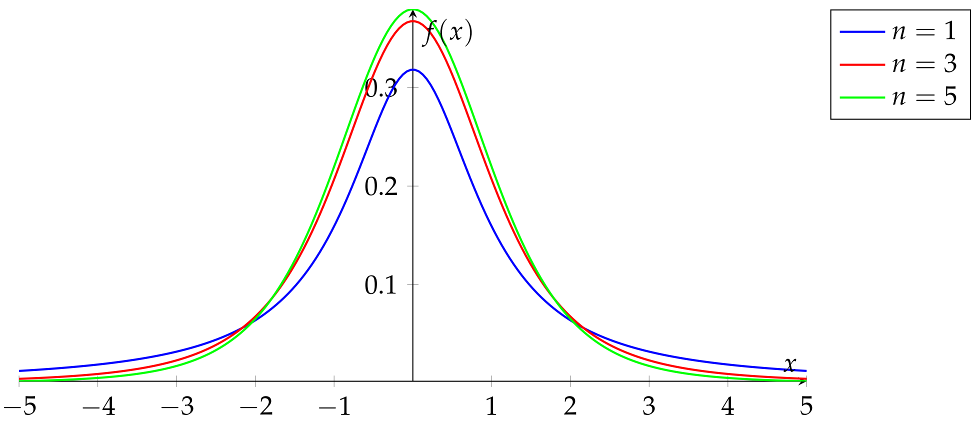
$$Z = \frac{X}{\sqrt{\frac{Y}{n}}}$$

服从t分布, 记作$Z \sim t(n)$. 其概率密度函数为

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}},$$

其中$\Gamma(z) = \int_0^{+\infty} x^{z-1}e^{-x}dx (z > 0)$. 部分图像如下:

t Distribution for $n = 1, 3, 5$



由定义知t分布满足以下性质:

1. t分布的概率密度函数是偶函数, 有推论$t_{1-\alpha}(n) = -t_\alpha(n)$.
2. 当n很大($n \geq 30$)时, t分布可用正态分布近似.

【例题 6.12】 已知$X \sim N(2,1), Y_1, ..., Y_4$均服从$N(0,4)$, 且随机变量之间相互独立. 设随机变量函数

$$T = \frac{4(X-2)}{\sqrt{\sum_{i=1}^{4} Y_i^2}},$$

求T的分布.

解. 标准化处理得$(X-2) \sim N(0,1)$, $\frac{Y_i}{2} \sim N(0,1)$. 于是$\sum_{i=1}^{4}(\frac{Y_i}{2})^2 \sim \mathcal{X}^2(4)$.故

$$T = \frac{X-2}{\sqrt{\frac{\sum_{i=1}^{4}(\frac{Y_i}{2})^2}{4}}} \sim t(4).$$

□

【定义 6.13】(F分布) 设随机变量$X\sim\mathcal{X}^2(n_1)$, $Y\sim\mathcal{X}^2(n_2)$, 且X,Y相互独立. 则随机变量函数

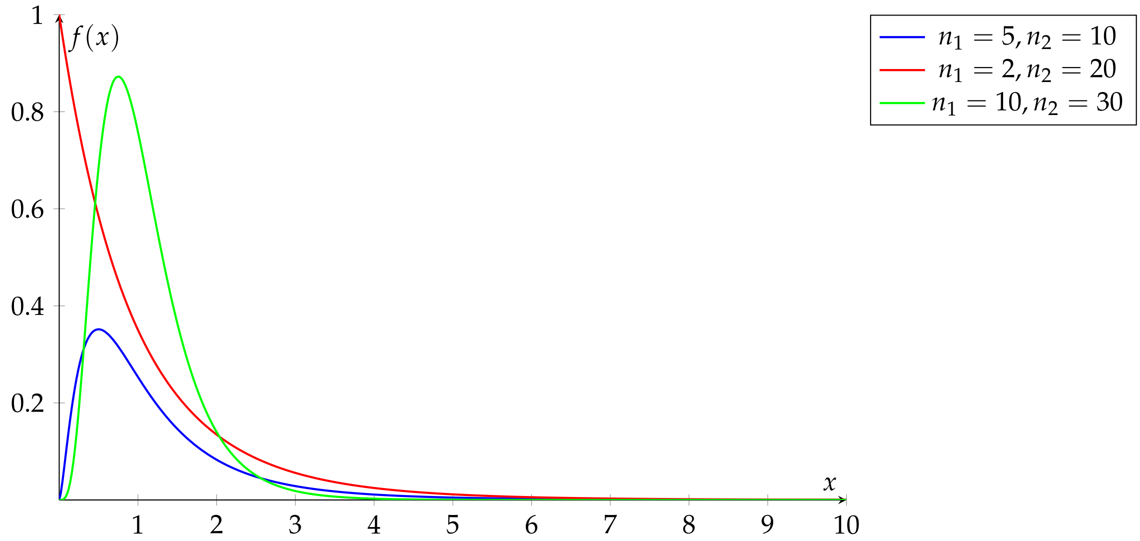
$$Z=\frac{\frac{X}{n_1}}{\frac{Y}{n_2}}$$

服从F分布, 记作$Z\sim F(n_1,n_2)$. 其概率密度函数为

$$f(x)=\begin{cases}\frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}(\frac{n_1}{n_2})^{\frac{n_1}{2}}x^{\frac{n_1}{2}-1}(1+\frac{n_1}{n_2}x)^{-\frac{n_1+n_2}{2}} & ,x>0\\ 0 & ,x\le 0\end{cases},$$

其中$\Gamma(z)=\int_0^{+\infty}x^{z-1}e^{-x}dx(z>0)$. 部分概率密度函数图像如下:

F Distribution



由定义知F分布满足以下性质:

1. 若$F\sim F(n_1,n_2)$, 则$\frac{1}{F}\sim F(n_2,n_1)$.
2. $F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}$.

【例题 6.14】 已知$X_1,...,X_6$均服从$N(0,\sigma^2)$. 设随机变量函数

$$F=\frac{2(X_1^2+X_2^2)}{X_3^2+X_4^2+X_5^2+X_6^2},$$

求F的分布.

解. 标准化处理后, 我们有$\frac{X_1^2}{\sigma^2}+\frac{X_2^2}{\sigma^2}\sim\mathcal{X}^2(2)$, $\sum_{i=3}^6\frac{X_i^2}{\sigma^2}\sim\mathcal{X}^2(4)$.故

$$F=\frac{\frac{\frac{X_1^2}{\sigma^2}+\frac{X_2^2}{\sigma^2}}{2}}{\frac{\sum_{i=3}^6\frac{X_i^2}{\sigma^2}}{4}}\sim F(2,4).$$

□

【例题 6.15】 已知$F\sim F(10,15)$, 求λ, 使得$P\{F\le\lambda\}=0.01$.

解.

$$P\{F\le\lambda\}=P\{\frac{1}{F}\ge\frac{1}{\lambda}\}=0.01,$$

其中$\frac{1}{F}\sim F(15,10)$.故$\frac{1}{\lambda}\sim F_{0.01}(15,10)\Rightarrow\lambda=0.293$. □

6.4 正态总体下的抽样分布

总体是正态分布, 从中抽取样本构造统计量. 在本节中, 我们将考虑这些统计量服从的分布.

【定理 6.16】 设$X \sim N(\mu,\sigma^2)$, $\{X_1,...,X_n\}$是样本. 则

1. $\overline{X} \sim N(\mu,\frac{\sigma^2}{n})$. 也即$\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$.
2. $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2 \sim \mathcal{X}^2(n-1)$.
3. $\overline{X}$与S^2独立.

【定理 6.17】 设$X \sim N(\mu,\sigma^2)$, $\{X_1,...,X_n\}$是样本. 则

1. $\frac{\mathcal{X}-\mu}{S}\sqrt{n} = \frac{\frac{\overline{X}-\mu}{\sigma}\sqrt{n}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} \sim t(n-1)$.
2. $\frac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\mu) = \sum\limits_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2 \sim \mathcal{X}^2(n)$.

【定理 6.18】 设$X \sim N(\mu_1,\sigma_1^2), Y \sim N(\mu_2,\sigma_2^2)$. $\{X_1,...,X_{n_1}\},\{Y_1,...,Y_{n_2}\}$是样本. 则

1. $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$.
2. $\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{\frac{(n-1)S_1^2}{\sigma_1^2(n-1)}}{\frac{(n-1)S_2^2}{\sigma_2^2(n-1)}} \sim F(n_1-1,n_2-1)$.
3. 当$\sigma_1^2=\sigma_2^2$时, $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$.

【例题 6.19】 设$X \sim N(\mu,\sigma^2)$. 若样本容量为16, 求k使得$P\{\overline{X} > \mu+kS\} = 0.95$.

解. 由于

$$\frac{\overline{X}-\mu}{S}\sqrt{n} = \frac{4(\overline{X}-\mu)}{S} \sim t(15).$$

故

$$P\{\overline{X} > \mu+kS\} = P\{\frac{4(\overline{X})-\mu}{S} > 4k\} = 0.95 \Rightarrow k = -0.438.$$

□

【例题 6.20】 设$X \sim N(\mu,\sigma^2)$, $\{X_1,...,X_{n+1}\}$是样本. 设$\overline{X_n}, S_n^2$是$\{X_1,...,X_n\}$的均值和方差, 求

$$T = \frac{X_{n+1}-\overline{X_n}}{S_n}\sqrt{\frac{n}{n+1}}$$

的分布.

解. 由于$\overline{X_n} \sim N(\mu,\frac{\sigma^2}{n}), U = X_{n+1} \sim N(\mu,\sigma^2)$.则

$$\frac{X_{n+1}-\overline{X_n}}{\sqrt{1+\frac{1}{n}\sigma}} \sim N(0,1).$$

又$\frac{(n-1)S_n^2}{\sigma^2} \sim \mathcal{X}^2(n-1)$, 故

$$T = \frac{U}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2(n-1)}}} \sim t(n-1).$$

□

Chapter 7

参数估计

本章将通过构造统计量来对总体分布的参数(如正态分布中的μ, σ, Poisson分布中的λ等)进行估计.

7.1 点估计

【定义 7.1】 **(矩估计法)** 用样本的原点矩近似总体的原点矩即为矩估计法.即

1. 用$\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$估计$E(X)$.
2. 用$A_2 = \frac{1}{n}\sum\limits_{i=1}^{n} x_i^2$估计$E(X^2)$.

【例题 7.2】 设$X \sim N(\mu, \sigma^2)$, $\{X_1, ..., X_n\}$是样本, 求μ, σ^2的矩估计.

解.

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

$$\hat{\sigma}^2 = A_2 - \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n^2}(\sum_{i=1}^{n} x_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 = B_2.$$

□

【例题 7.3】 设$X \sim p(\lambda)$, $\{X_1, ..., X_n\}$是样本, 求λ的矩估计.

解. 由于$\lambda = E(X)$, 故

$$\hat{\lambda} = \overline{X}.$$

又$\lambda = D(X)$, 故

$$\hat{\lambda} = B_2.$$

一般来说, 低阶矩的估计效果优于高阶矩. 故在这里我们选取$\hat{\lambda} = \overline{X}$进行估计. □

【例题 7.4】 设$X \sim U[\theta_1, \theta_2]$, $\{X_1, ..., X_n\}$是样本, 求θ_1, θ_2的矩估计.

解. 由于$E(X) = \frac{1}{2}(\theta_1 + \theta_2)$, $D(X) = \frac{1}{12}(\theta_2 - \theta_1)^2$. 故有方程组

$$\begin{cases} E(X) = \frac{1}{2}(\theta_1 + \theta_2) = \overline{X} \\ E(X^2) = \frac{(\theta_2 - \theta_1)^2}{12} + \frac{(\theta_1 + \theta_2)^2}{4} = A_2 \end{cases},$$

解方程即得

$$\begin{cases} \hat{\theta}_1 = \overline{X} - \sqrt{3B_2} \\ \hat{\theta}_2 = \overline{X} + \sqrt{3B_2} \end{cases}.$$

□

【定义 7.5】(极大似然估计法) 由于概率大的事件比概率小的事件更容易发生, 故将使$P(A)$最大的参数值作为估计值是可行的, 这被称为极大似然估计.具体来说, 我们要求的参数估计在似然函数

$$L=\prod_{i=1}^{n}P\{X=x_i\}$$

的最值处取到.

【例题 7.6】 设$X\sim p(\lambda)$, $\{X_1,...,X_n\}$为样本. 求λ的极大似然估计.

解. 总体X的概率密度函数为

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda},$$

则λ的似然函数为

$$L(\lambda)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}.$$

两边同时取对数, 即得

$$\ln L(\lambda)=-\ln\prod_{i=1}^{n}x_i!+\sum_{i=1}^{n}x_i\ln\lambda-n\lambda.$$

两边同时求导, 有

$$\frac{d\ln L(\lambda)}{d\lambda}=\frac{1}{\lambda}\sum_{i=1}^{n}x_i-n.$$

$L(\lambda)$取最值时, 有$\frac{d\ln L(\lambda)}{d\lambda}=0$. 解得

$$\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{X}.$$

□

【例题 7.7】 设$X\sim N(\mu,\sigma^2)$, $\{X_1,...,X_n\}$为样本, 求μ,σ^2的极大似然估计.

解. 总体X的概率密度函数为

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

则μ,σ^2的似然函数为

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}.$$

两边同时取对数, 即得

$$\ln L(\mu,\sigma^2)=n\ln\frac{1}{\sqrt{2\pi}}-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

两边同时求偏导, 有

$$\frac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$
$$\frac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

$L(\mu,\sigma^2)$取最值时, 两个偏导数均为0. 解得

$$\begin{cases}\hat{\mu}=\frac{1}{n}\sum\limits_{i=1}^{n}x_i=\overline{X}\\ \hat{\sigma}^2=\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-\mu)^2=B_2\end{cases}.$$

□

特别地, 均匀分布$X\sim U[\theta_1,\theta_2]$中对$\theta_1,\theta_2$的极大似然估计无法使用上面的方法. 此时的估计值为

$$\begin{cases}\hat{\theta}_1=\min\{x_1,...,x_n\}\\ \hat{\theta}_2=\max\{x_1,...,x_n\}\end{cases}.$$

7.2 点估计的优良性准则

【定义 7.8】(无偏性) 若$E(\hat{\theta})=\theta$, 则称对θ的估计$\hat{\theta}$是无偏的.

具体来说, 设总体X满足$E(X)=\mu, D(X)=\sigma^2$, $\{X_1,...,X_n\}$为样本. 则

1. $\overline{X}$是μ的无偏估计(由于$E(\overline{X})=\mu$).
2. S^2是σ^2的无偏估计(由于$E(S^2)=\sigma^2$).
3. 未修正方差S_0^2是σ^2的有偏估计.

注意: $\hat{\theta}$是θ的无偏估计, 无法得到$g(\hat{\theta})$是$g(\theta)$的无偏估计.

【例题 7.9】 设总体X满足$E(X)=\mu, D(X)=\sigma^2$, $\{X_1,...,X_n\}$为样本. 证明: S不是σ的无偏估计.

证明.

$$D(S)=E(S^2)-E^2(S)=\sigma^2-E^2(S)\Rightarrow E(S)=\sqrt{\sigma^2-D(S)}\le\sigma.$$

由本例可见S^2是σ^2的无偏估计无法保证S是σ的无偏估计. □

【定义 7.10】(有效性) 若$D(\hat{\theta}_1)\le D(\hat{\theta}_2)$, 则表明估计值$\hat{\theta}_1$比$\hat{\theta}_2$更有效.

【例题 7.11】 设总体为X满足$E(X)=\mu$, $D(X)=\sigma^2$, $\{X_1,...,X_n\}$为样本. 确定参数$a_1,...,a_n$(满足$\sum_{i=1}^n a_i=1$), 使得

$$\hat{\mu}=\sum_{i=1}^n a_i x_i$$

的拟合效果最好.

解. 由于

$$E(\hat{\mu})=\sum_{i=1}^n a_i E(X_i)=\mu\sum_{i=1}^n a_i=\mu.$$

故$\hat{\mu}$是无偏的. 于是我们考虑最小化

$$D(\hat{\mu})=\sum_{i=1}^n a_i^2 D(X_i)=\sigma^2\sum_{i=1}^n a_i^2.$$

由Cauchy-Schwarz不等式, 有

$$n\sum_{i=1}^n a_i^2\ge 1=(\sum_{i=1}^n a_i)^2=1,$$

故$D(\hat{\mu})\ge\frac{\sigma^2}{n}$, 当且仅当$a_1=...=a_n=\frac{1}{n}$时取等. □

【定义 7.12】(相合性/一致性) 基于样本$\{X_1,...,X_n\}$的点估计值$\hat{\theta}$应当满足$\forall\varepsilon>0$, 都有

$$\lim_{n\to+\infty}P\{\left|\hat{\theta}-\theta\right|<\varepsilon\}=1.$$

7.3 区间估计

【定义 7.13】(置信度) θ的估计值区间$[\hat{\theta}_1,\hat{\theta}_2]$满足

$$P\{\hat{\theta}_1\le\theta\le\hat{\theta}_2\}=1-\alpha,$$

称$1-\alpha$为该区间估计的置信度.

【定义 7.14】(枢轴变量) 从θ的一个点估计出发, 构造与θ相关的一个函数G, 使得G的分布F是已知的并与θ无关. 则称函数G为枢轴变量.

实际应用中, 给定$1-\alpha$, 构造枢轴变量G, 使得G的分布F的上$\frac{\alpha}{2}$分位数为$\nu_{\frac{\alpha}{2}}$, 上$(1-\frac{\alpha}{2})$分位数为$\nu_{1-\frac{\alpha}{2}}$, 满足

$$P\{\nu_{1-\frac{\alpha}{2}} \leq G \leq \nu_{\frac{\alpha}{2}}\} = 1-\alpha.$$

通过该式进一步求出使得$P\{\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2\} = 1-\alpha$的$\theta_1, \theta_2$即可.

【例题 7.15】 已知总体$X \sim N(\mu, \sigma^2)$, 其中σ^2是已知量, $\{X_1, ..., X_n\}$为样本. 给定置信度$1-\alpha$, 求μ的区间估计.

解. 构造枢轴变量

$$G = \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim N(0,1).$$

查表可知$N(0,1)$的上$\frac{\alpha}{2}$分位数为$\nu_{\frac{\alpha}{2}}$. 则

$$P\{-\nu_{\frac{\alpha}{2}} \leq G \leq \nu_{\frac{\alpha}{2}}\} = 1-\alpha \Rightarrow P\{\overline{X} - \frac{\sigma}{\sqrt{n}}\nu_{\frac{\alpha}{2}} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}}\nu_{\frac{\alpha}{2}}\} = 1-\alpha.$$

即μ的区间估计$[\hat{\mu}_1, \hat{\mu}_2]$为

$$\begin{cases} \hat{\mu}_1 = \overline{X} - \frac{\sigma}{\sqrt{n}}\nu_{\frac{\alpha}{2}} \\ \hat{\mu}_2 = \overline{X} + \frac{\sigma}{\sqrt{n}}\nu_{\frac{\alpha}{2}} \end{cases}$$

□

【例题 7.16】 已知总体$X \sim N(\mu, \sigma^2)$, 其中σ^2是未知量, $\{X_1, ..., X_n\}$为样本. 给定置信度$1-\alpha$, 构造求μ的区间估计的枢轴变量.

解. 构造枢轴变量

$$T = \frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t(n-1).$$

□

【例题 7.17】 已知总体$X \sim N(\mu, \sigma^2)$, 其中μ是已知量, $\{X_1, ..., X_n\}$为样本. 给定置信度$1-\alpha$, 求σ^2的区间估计.

解. 构造枢轴变量

$$G = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \mathcal{X}^2(n).$$

查表可知$\mathcal{X}^2(n)$的上$\frac{\alpha}{2}$分位数为$\mathcal{X}^2_{\frac{\alpha}{2}}$, 上$1-\frac{\alpha}{2}$分位数为$\mathcal{X}^2_{1-\frac{\alpha}{2}}$. 则

$$P\{\mathcal{X}^2_{1-\frac{\alpha}{2}} \leq G \leq \mathcal{X}^2_{\frac{\alpha}{2}}\} = 1-\alpha \Rightarrow P\{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\mathcal{X}^2_{\frac{\alpha}{2}}(n)} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\mathcal{X}^2_{1-\frac{\alpha}{2}}(n)}\} = 1-\alpha.$$

即σ^2的区间估计$[\hat{\sigma}^2_1, \hat{\sigma}^2_2]$为

$$\begin{cases} \hat{\sigma}^2_1 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\mathcal{X}^2_{\frac{\alpha}{2}}(n)} \\ \hat{\sigma}^2_2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\mathcal{X}^2_{1-\frac{\alpha}{2}}(n)} \end{cases}$$

□

【例题 7.18】 已知总体$X \sim N(\mu, \sigma^2)$, 其中μ是未知量, $\{X_1, ..., X_n\}$为样本. 给定置信度$1-\alpha$, 构造求σ^2的区间估计的枢轴变量.

解. 构造枢轴变量

$$G = \frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2(n-1).$$

□

Chapter 8

假设检验

8.1 基本概念

【定义 8.1】(假设) 当总体分布未知时, 对总体做出的推断称为假设. 其中推断总体分布类型的为非参数假设, 推断总体分布的参数的为参数假设.

【定义 8.2】(假设检验) 对于假设的检验即假设检验问题. 根据假设的类型可以分为参数假设检验和非参数假设检验. 本章主要研究参数假设检验问题.

【定义 8.3】(原假设 / 备择假设) 原假设是在一次试验中有绝对优势出现的事件, 而备择假设在一次试验中不易发生(或几乎不可能发生)的事件.因此, 在进行单侧检验时, 最好把原假设取为预想结果的反面, 即把希望证明的命题放在备择假设上.

当假设检验问题中只存在唯一假设H_0时, 称该问题为显著性假设检验问题; 当问题中同时存在假设H_0, H_1时, 称该问题为H_0对H_1假设检验问题. 一般来说, 参数假设检验问题的基本步骤为:

1. 提出H_0对H_1;
2. 假定H_0成立, 构造已知分布的统计量T;
3. 根据给定的α, 通过查表确定拒绝域的临界值;
4. 代入题目数据, 验证其是否落于拒绝域内. 若在拒绝域内, 则拒绝H_0; 若不在拒绝域内, 则接受H_0.

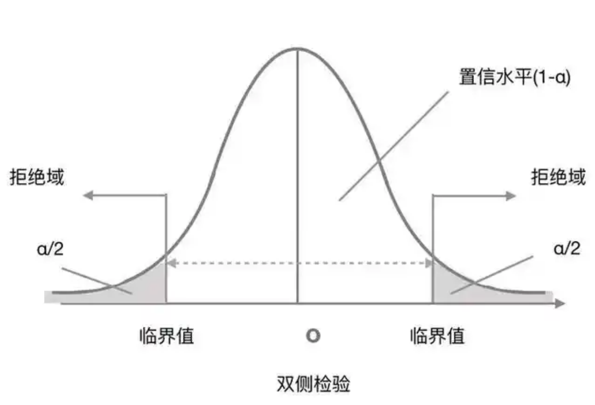


图 8.1: 双侧检验的拒绝域

【例题 8.4】 某化工厂用包装机自动包装洗衣粉. 已知洗衣粉重量(克)$X \sim N(\mu, 4)$. 机器正常工作时, $\mu = 500g$. 某日开工后, 随机取9袋, 其重量分别为:$505, 499, 502, 506, 498, 498, 497, 510, 503$(单位:克).假定$\sigma = 2$不变, 问包装机工作是否正常?

解. 提出原假设H_0: 包装机工作正常, 即$\mu = 500g$. 则相应备择假设H_1: 包装机工作不正常, 即$\mu \neq 500g$.

假定H_0成立, 则$X \sim N(500, 4)$. 故

$$\overline{X} \sim N(500, \frac{4}{9}) \Rightarrow U = \frac{\overline{X} - 500}{\frac{2}{3}} \sim N(0, 1).$$

我们需要考虑原假设是否会与"小概率事件实际不发生"相矛盾, 也即考虑

$$|U| = \frac{502 - 500}{\frac{2}{3}} = 3$$

与$\nu_{\frac{\alpha}{2}}$的大小关系, 其中$\nu_{\frac{\alpha}{2}}$是$N(0, 1)$的上$\frac{\alpha}{2}$分位数. 例如取$\alpha = 0.05$, 则$\nu_{\frac{\alpha}{2}} = 1.96$, 此时$|U| > \nu_{\frac{\alpha}{2}}$, 此时发生矛盾, 故有95%的把握拒绝原假设$H_0$, 接受备择假设$H_1$, 认为包装机工作不正常. □

假设检验存在以下两类错误:

1. 第一类错误-弃真错误:犯错误的概率为$P\{拒绝H_0 \mid H_0为真\} = \alpha$.
2. 第二类错误-纳伪错误:犯错误的概率为$P\{接受H_0 \mid H_0为假\} = \beta$.

在假设检验中, 我们希望首先保证最小化α, 并在此基础上尽可能降低β.

单正态总体的参数假设检验参考之前的枢轴变量即可.

两个正态总体的参数假设检验选取的枢轴变量参考定理6.18.