一、基本概念

• 两类现象 { 确定性现象：结果确定
不确定性现象：结果不确定 < 个别现象
随机现象：个别试验中结果不确定，但大量重复试验中结果有统计规律性.

• 对随机现象的观察、记录、实验统称为随机试验. ：(1). 相同条件可重复 (2). 预知所有可能结果 (3). 每次无法预知哪个结果发生.

样本空间：随机试验的所有可能结果构成的集合，常用 S (或 Ω) 表示.

样本空间S中的每一个元素，即试验的每一个结果称为样本点. 样本空间的任一子集称为随机事件 (事件)，常用字母A,B,C等表示

p5. 只含有一个样本点的事件称为基本事件；当某一事件包含的一个样本点恰好为此次试验出现的结果时，称该事件发生，否则不发生

把样本空间S视为一事件时，S一定会发生，S：必然事件　　ϕ 为不可能事件.

• 相互关系及运算.

(1). 包含：事件B发生一定导致事件A发生，则称B包含于A　　$B\subset A$

(2). 相等：$B\subset A$ 且 $A\subset B$，则事件A与事件B相等　　$A=B$

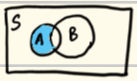
(i) 和事件：$A\cup B=\{x: x\in A 或 x\in B\}$　$\bigcup_{i=1}^{n} A_i$ 为n个事件 $A_1, A_2, \cdots, A_n$ $(n\geq 2)$ 的和事件

(ii) 积事件：$A\cap B=AB=A\cdot B=\{x: x\in A 且 x\in B\}$　$\bigcap_{i=1}^{n} A_i$ 为n个事件 $A_1, A_2, \cdots, A_n$ $(n\geq 2)$ 的积事件

(iii) 互斥 (互不相容)：A,B为两随机事件，$A\cap B=\phi$　任何随机事件其基本事件互斥.

(iv). 逆事件 (对立事件)：$A\cup B=S$ 且 $A\cap B=\phi$　A与B互逆或互为对立事件　$\bar{A}$ 或 A^c 为事件A的逆事件

(v). 差事件：A与 $\bar{B}$ 的共同样本点构成的集合为A对B的差事件　$A-B=A\cap\bar{B}=\{x: x\in A 且 x\notin B\}$



运算：$A\cup B=B\cup A$，$A\cap B=B\cap A$；$A\cup(B\cup C)=(A\cup B)\cup C$，$A\cap(B\cap C)=(A\cap B)\cap C$

$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$，$(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$

• 德摩根定律：$\overline{\bigcup_{j=1}^{n} A_j}=\bigcap_{j=1}^{n}\overline{A_j}$　$\overline{\bigcap_{j=1}^{n} A_j}=\bigcup_{j=1}^{n}\overline{A_j}$

n个元件组成一个系统，若有一个损坏，系统就损坏 ⇒ 串联系统；只要有一个不损坏，系统就不损坏 ⇒ 并联系统

• 频率：相同条件下进行 $n(n\geq 1)$ 次重复试验，若事件A在n次中发生 n_A 次 (频数 $0\leq n_A\leq n$)，则称比值 $\frac{n_A}{n}$ 为事件A在这n次试验中发生的频率，记为 $f_n(A)$，即 $f_n(A)=\frac{n_A}{n}$

性质：(1). $\forall$事件A，$0\leq f_n(A)\leq 1$　(2). $f_n(S)=1$　(3). A与B互斥 ($A\cap B=\phi$) 时，$f_n(A\cup B)=f_n(A)+f_n(B)$

推广：当 $A_1, A_2, \cdots, A_k$ $(k\geq 3)$ 两两不相容时，$f_n(\bigcup_{j=1}^{k} A_j)=\sum_{j=1}^{k} f_n(A_j)$

• 概率：统计性定义：随机试验对应的样本空间为S，对S中的任一事件A，当总试验次数n充分大时，A频率 $f_n(A)$ 稳定值p定义为事件A的概率. $P(A)=p$

公理定义：某一随机试验对应的样本空间为S，对S中的任一事件A，定义实数 $P(A)$，满足：

(1) 非负性：$P(A)\geq 0$　(2). 规范性：$P(S)=1$

(3) 可列可加性：对S中的可列个两两互不相容的事件 $A_1, A_2 \cdots A_n \cdots$ (即 $A_iA_j=\phi$，$i\neq j$，$i,j=1,2\cdots$) 有：$P(\bigcup_{j=1}^{+\infty} A_j)=\sum_{j=1}^{+\infty} P(A_j)$

则称 $P(A)$ 为事件A发生的概率.

·性质：(1). 有限可加性：对有限个两两互斥的事件的和事件，有 $P(\cup_{j=1}^{n}A_j)=\sum_{j=1}^{n}P(A_j)$

(2). $P(A)=1-P(\bar{A})$ 特别地，$\phi\cup S=S$，可得 $P(\phi)=0$

(3). 当 $A\supset B$ 时，$P(A-B)=P(A)-P(B)$，由此可推出 $P(A)\geq P(B)$

$A=B\cup A\bar{B}$ $\therefore P(A)=P(B\cup A\bar{B})=P(B)+P(A\bar{B})\Rightarrow P(A-B)=P(A)-P(B)\geq 0$ $\therefore P(A)\geq P(B)$

ps. $A\cap B=A-A\cap\bar{B}$

即 $A\bar{B}\bar{C}=A\cap\overline{B\cup C}=A-A\cap(B\cup C)$

$=A-(A\cap B)\cup(A\cap C)$

$\therefore P(A\bar{B}\bar{C})=P(A)-P(AB)-P(AC)+P(ABC)$

(4) 概率加法公式：$P(A\cup B)=P(A)+P(B)-P(AB)$

$A\cup B=A\cup B\bar{A}\Rightarrow P(A\cup B)=P(A\cup B\bar{A})=P(A)+P(B\bar{A})$ $A=A\bar{B}\cup AB$ 互斥（有限可加性） $\therefore P(A)=P(A\bar{B})+P(AB)$

$\Rightarrow P(A\cup B)=P(A\bar{B})+P(AB)+P(B\bar{A})$ $A\bar{B}=A\cup B-B$ $P(A\bar{B})=P(A\cup B)-P(B)$ $P(B\bar{A})=P(A\cup B)-P(A)$

$\therefore$ 代入整理：$P(A\cup B)=P(A)+P(B)-P(A\cap B)$

推广：$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC)$

$$P\left(\bigcup_{j=1}^{n}A_j\right)=\sum_{j=1}^{n}P(A_j)-\sum_{i<j}P(A_iA_j)+\sum_{i<j<k}P(A_iA_jA_k)-\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n),\quad n\geqslant 1. \tag{1.2.2}$$

eg.

例 1.2.3 设甲、乙、丙三人去参加某个活动的概率均为 0.4，且甲、乙、丙中至少有两人去参加的概率为 0.3，三人同时参加的概率为 0.05，求甲、乙、丙三人至少有一人参加的概率.

设 $A=\{甲参加\}$ $B=\{乙参加\}$ $C=\{丙参加\}$ 由题：$P(A)=P(B)=P(C)=0.4$

$P(AB\cup BC\cup AC)=0.3$ $P(ABC)=0.05$ $P(A\cup B\cup C)=?$

$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$

$=P(AB)+P(BC)+P(AC)-2P(ABC)$

$\therefore P(\{至少一人参加\})=1.2-0.3-P(ABC)=0.85$

·等可能概型：

一个随机试验，满足：(1). 样本空间中样本点数有限（有限性） (2). 出现每一个样本点的概率相等（等可能性），则称这个试验为等可能概型，又称古典概型. 任一事件A的概率为 $P(A)=\frac{A中包含的样本点数}{S中样本点总数}$

例 1.3.2 (抽签问题) 袋中有编号为 $1,2,\cdots,n(n>1)$ 的 n 个球，其中有 a 个红球，b 个白球 $(a+b=n)$. 从中每次摸一球，无放回，共摸 n 次. 假设各个球的大小和材质相同，求第 $k(1\leqslant k\leqslant n)$ 次摸到红球的概率.

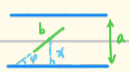
解 设 $A_k=\{第\ k\ 次摸到红球\}$，$k=1,2,\cdots,n$.

解法① 若视 n 次摸出的球的编号的先后排列为一个样本点，则样本空间中共有 $n!$ 个样本点. 由题意知，出现每一个样本点的概率相等，而第 k 次为红球共有 $a(n-1)!$ 个样本点. 所以

$$P(A_k)=\frac{a(n-1)!}{n!}=\frac{a}{n}=\frac{a}{a+b},\quad k=1,2,\cdots,n.$$

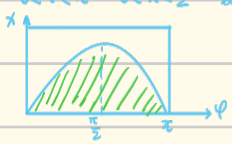
ps. 当样本空间S所含的样本点个数为无穷多个时 ⇒ 几何概型.

蒲丰投针：平行线间距 a，针长 $b(<a)$，相交的概率？



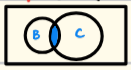
针与最近的线距离为 x，夹角 φ.

$0\leq\varphi\leq\pi$ $0\leq x\leq\frac{a}{2}$ 发生的情况：$0\leq x\leq\frac{b}{2}\sin\varphi$



·条件概率：如果 $P(B)>0$，那么在B发生的条件下A发生的条件概率为 $P(A|B)=\frac{P(AB)}{P(B)}$

当 $P(C)\neq 0$ 时，有：(1) $P(A|C)\geq 0$ (2). $P(S|C)=1$ (3). $P(B|C)=1-P(\bar{B}|C)$



(4). 当 $A\supset B$ 时，$P(A|C)\geq P(B|C)$

(5). $P(A\cup B|C)=P(A|C)+P(B|C)-P(AB|C)$ 特别地，若 $AB=\phi$，则 $P(A\cup B|C)=P(A|C)+P(B|C)$

ps. 当 $B\neq D$ 时，$P(A|B)+P(C|D)$ 无意义.

eg.

例 1.4.1 掷一颗骰子两次，已知两次点数之和大于 8，求第一次是 6 点的概率.

设 $A=\{第一次是6点\}$ $B=\{两次点数之和大于8\}$. $P(B)=\frac{4\times2+2}{6\times6}=\frac{5}{18}$ $P(AB)=\frac{1\times4}{6\times6}=\frac{2}{18}$

$\therefore P(A|B)=\frac{P(AB)}{P(B)}=\frac{2}{5}$.

例 1.4.2　一袋中有 5 个红球, 4 个白球, 从中每次摸一球, 无放回抽样, 抽 4 次.

(1) 已知前两次中至少有一次摸到红球, 求前两次中恰有一次摸到红球的概率;

(2) 已知第 4 次摸到的是红球, 求第 1 次和第 2 次摸到的都是红球的概率.

(1). B={前两次中至少有一次红球}　C={前两次中恰有一次红球}

$P(B)=1-\frac{C_4^2}{C_9^2}=1-\frac{4\times3}{9\times8}=\frac{5}{6}$　$P(BC)=\frac{C_5^1C_4^1}{C_9^2}=\frac{5}{9}$　$\therefore P(C|B)=\frac{P(BC)}{P(B)}=\frac{2}{3}$

5红 4白

(2). 设 A_i={第i次摸到红球}　$P(A_1A_2|A_4)=\frac{P(A_1A_2A_4)}{P(A_4)}$　$P(A_4)=\frac{5\cdot C_8^3\cdot3!}{C_9^4\cdot4!}=\frac{5}{9}$

_ _ _ 红.

$P(A_1A_2A_4)=\frac{C_5^3\cdot3!\cdot C_6^1}{C_9^4\cdot4!}=\frac{5}{42}$　红红 _ 红　5红4白　$\therefore P(A_1A_2|A_4)=\frac{3}{14}$

•乘法公式: 由条件概率定义知: 当 $P(A)\neq0$, $P(B)\neq0$ 时, $P(AB)=P(A)\cdot P(B|A)=P(B)\cdot P(A|B)$

推广: 当 $P(AB)\neq0$ 时, $P(ABC)=P(A)P(B|A)P(C|AB)$

一般地, 当 $P(A_1A_2\cdots A_{n-1})\neq0$ $(n\geq3)$ 时, 有 $P(A_1A_2\cdots A_n)=P(A_1)\cdot P(A_2|A_1)\cdot P(A_3|A_1A_2)\cdots P(A_n|AA_2\cdots A_{n-1})$

eg.

例 1.4.3　设一社区 "三口之家" 占了 70%, 已知该社区有 40% 的家庭至少有一人职业为教师, 且在 "三口之家" 中有 30% 的家庭至少有一人职业为教师. 在该社区随机找一户.

(1) 求这一户既是 "三口之家" 又 "至少有一人职业为教师" 的概率;

(2) 求这一户既不是 "三口之家" 又 "没有人职业为教师" 的概率;

(3) 已知这一户 "没有人职业为教师", 求这一户是 "三口之家" 的概率.

解: (1). 设A={这一户是"三口之家"}　B={这一户"有人职业为教师"}　由题: $P(A)=70\%$　$P(B)=40\%$　$P(B|A)=30\%$

$\therefore P(AB)=P(B|A)\cdot P(A)=0.7\times0.3=0.21$

(2). $P(\bar{A}\bar{B})=P(\overline{A\cup B})=1-P(A\cup B)=1-P(A)-P(B)+P(AB)=1-0.7-0.4+0.21=0.11$

(3). $P(\bar{B})=1-P(B)=0.6$　$P(A|\bar{B})=1-P(\bar{A}|\bar{B})=1-\frac{P(\bar{A}\bar{B})}{P(\bar{B})}=1-\frac{0.11}{0.6}=\frac{49}{60}$

例 1.4.4　某人参加某种技能考试, 已知第 1 次考合格的概率为 60%; 若第 1 次没有合格, 第 2 次能考合格的概率为 70%; 若第 1 次和第 2 次均不合格, 第 3 次能考合格的概率为 80%, 求此人最多 3 次能考合格的概率.

解: ① 设 A_i={第i次合格}, $i=1,2,3$　B={最多3次能考合格}　由题: $P(A_1)=0.6$　$P(A_2|\bar{A}_1)=0.7$　$P(A_3|\bar{A}_1\bar{A}_2)=0.8$　B={最多3次能考合格}

$B=A_1\cup A_2\bar{A}_1\cup A_3\bar{A}_1\bar{A}_2$　$\therefore P(B)=P(A_1)+P(A_2\bar{A}_1)+P(A_3\bar{A}_1\bar{A}_2)=P(A_1)+P(\bar{A}_1)\cdot P(A_2|\bar{A}_1)+P(\bar{A}_1)\cdot P(\bar{A}_2|\bar{A}_1)\cdot P(A_3|\bar{A}_1\bar{A}_2)=0.6+0.4\times0.7+0.4\times0.3\times0.8=0.976$

② $\bar{B}=\bar{A}_1\bar{A}_2\bar{A}_3$　$P(B)=1-P(\bar{B})=1-P(\bar{A}_1\bar{A}_2\bar{A}_3)=1-P(\bar{A}_1)P(\bar{A}_2|\bar{A}_1)P(\bar{A}_3|\bar{A}_1\bar{A}_2)=1-0.4\times0.3\times0.2=0.976$

•全概率公式、贝叶斯公式

假设A,B为随机事件, 那么总有 $A=AB\cup A\bar{B}$ $[=(A\cap B)\cup(A\cap\bar{B})=A\cup(B\cap\bar{B})=A]$. 又AB与$A\bar{B}$互不相容, 利用有限可加性: $P(A)=P(AB)+P(A\bar{B})$

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

定义: 设S为某一随机试验的样本空间, $B_1,B_2\cdots B_n$为该试验的一组事件且满足: (1). $B_iB_j=\phi$　$i,j=1,2,\cdots,n$　$i\neq j$　(2). $B_1\cup B_2\cup\cdots\cup B_n=S$, 则称 $B_1,B_2,\cdots,B_n$ 为S的一个划分, 或称为S的一个完备事件组

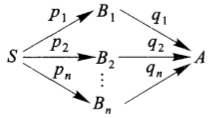
上述 $B_1, B_2, \cdots, B_n$ 两两同时发生是不可能的，而至少有一个发生又是必然的. 任一个事件B与其逆事件$\bar{B}$就是S的一个划分.

若A为S中的一个事件，$B_1, B_2, \cdots B_n$ 是S的一个划分，则类似的：$A = AS = A\cap(B_1\cup B_2\cup\cdots\cup B_n) = AB_1\cup AB_2\cup AB_3\cup\cdots\cup AB_n = \bigcup_{j=1}^{n} AB_j$

即：
$$P(A) = P(\cup_{j=1}^{n} AB_j) = \sum_{j=1}^{n} P(AB_j)$$

定理：设S为某一试验的样本空间，若 $B_1, B_2, \cdots, B_n$ 为S的一个划分，且 $P(B_j)>0$，$j=1,2\cdots n$，则对任一事件A，有 $P(A) = \sum_{j=1}^{n} P(B_j)P(A|B_j)$ 称为概率的全概率公式.

运用公式 (1.4.7) 的关键是找到一组合适的划分. 设 $P(B_j) = p_j$, $P(A|B_j) = q_j$, $j = 1, 2, \cdots, n$, 则 $P(A) = \sum_{j=1}^{n} p_j q_j$. 全概率公式的图示分析如图 1.4.2 所示.



定理：设S为某一试验的样本空间，若 $B_1, B_2, \cdots, B_n$ 为S的一个划分，且 $P(B_j)>0$，$j=1,2\cdots n$，则对任一事件A，$P(A)\neq 0$，有 $P(B_k|A) = \frac{P(B_kA)}{P(A)} = \frac{P(B_k)\cdot P(A|B_k)}{\sum_{j=1}^{n} P(B_j)\cdot P(A|B_j)}$ $k=1,2,\cdots,n$ 称为概率的贝叶斯公式或逆概公式.

在利用贝叶斯公式时，其中的 $P(B_j)(j = 1, 2, \cdots, n)$ 的概率往往是已知或事先假设 (或者根据以往的资料或经验的累积) 的，常称 $P(B_j)$ 为先验概率 (prior probability); 而当事件 A 发生后，对 B_j 发生的概率重新进行推断 (或修正)，常称 $P(B_j|A)$ 为后验概率 (posterior probability).

▶ 案例 贝叶斯方法

eg.

例 1.4.5　一小学举办家长开放日，欢迎家长参加. 某同学母亲参加的概率为 0.80, 如果母亲参加，那么父亲参加的概率为 0.30; 如果母亲不参加，那么父亲参加的概率为 0.90.

(1) 求该同学父母都参加的概率;

(2) 求该同学父亲参加的概率;

(3) 在已知父亲参加的条件下，求母亲也参加的概率.

设 $A=\{母亲参加\}$ $B=\{父亲参加\}$

由题：$P(A)=0.8$　$P(B|A)=0.3$　$P(B|\bar{A})=0.9$

(1). $P(AB) = P(A)\cdot P(B|A) = 0.24$

(2). $P(B) = P(A)\cdot P(B|A) + P(\bar{A})\cdot P(B|\bar{A}) = 0.42$

(3). $P(A|B) = \frac{P(AB)}{P(B)} = \frac{0.24}{0.42} = \frac{4}{7} \approx 0.57$

例 1.4.6　某保险公司的某险种有高、中、低三种类型风险客户，他们的年度索赔概率分别为 0.02, 0.01, 0.002. 设三类客户所占份额分别是 0.05, 0.15, 0.80, 在其中任选一位客户，

(1) 求该客户的年度索赔概率;

(2) 若该客户在这一年中发生了索赔，求他是高风险客户的概率.

$A=\{发生了索赔\}$ $B_1=\{高风险类型\}$ $B_2=\{中\cdots\}$ $B_3=\{低\cdots\}$

$P(B_1)=0.05$, $P(B_2)=0.15$, $P(B_3)=0.80$

$P(A|B_1)=0.02$　$P(A|B_2)=0.01$　$P(A|B_3)=0.002$

(1) $\therefore P(A) = \sum_{i=1}^{3} P(B_i)\cdot P(A|B_i) = 0.0041$

(2). $P(B_1|A) = \frac{P(B_1A)}{P(A)} = \frac{P(B_1)\cdot P(A|B_1)}{P(A)} = \frac{10}{41}$

事件的独立性与独立试验

定义：设A，B为两随机事件，当 $P(AB)=P(A)\cdot P(B)$ 时，称事件A，B相互独立.

当 $P(A)\cdot P(B)\neq 0$ 时，"事件A与事件B相互独立"等价于"条件概率等于无条件概率"即 $P(B|A)=P(B)$ 或 $P(A|B)=P(A)$

定理：当事件A与事件B相互独立时，A与$\bar{B}$，$\bar{A}$与B，$\bar{A}$与$\bar{B}$均相互独立　$P(A)\cdot P(B)=P(AB)$　$P(\bar{A})\cdot P(B)=P(B)-P(A)\cdot P(B)=P(AB)+P(\bar{A}B)-P(AB)=P(\bar{A}B)$ 即证

! 描述两个随机事件A，B的关系除包含、不相容外，还有独立关系

例 1.5.1　甲、乙两人同时向一目标进行射击，若甲的击中率为 0.8, 乙的击中率为 0.7, 求目标被击中的概率.

$A=\{甲击中\}$　$B=\{乙击中\}$　$C=\{目标被击中\}$.

$P(C)=P(A\cup B)=P(A)+P(B)-P(AB)$

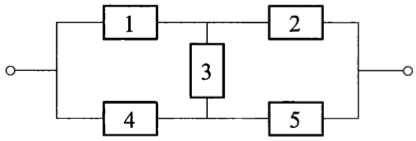
$=0.8+0.7-0.8\times 0.7=0.94$

定义：设A,B,C为三个随机事件，当 $P(AB)=P(A)P(B)$，$P(AC)=P(A)P(C)$，$P(BC)=P(B)P(C)$ 都成立，称事件A,B,C两两独立，若同时还满足 $P(ABC)=P(A)P(B)P(C)$ 则称事件A,B,C相互独立.（ps. 相互独立 $\Rightarrow$ 两两独立，两两独立 $\not\Rightarrow$ 相互独立）

定义：设n个事件 $A_1, A_2, \cdots A_n$ $(n\geq 2)$，若对其中任意k个事件 $A_{i1}, A_{i2}, \cdots, A_{ik}$ $(2\leq k\leq n)$，都有 $P(A_{i1}A_{i2}\cdots A_{ik})=\prod_{j=1}^{k}P(A_{ij})$ 成立，则称事件 $A_1, A_2, \cdots A_n$ 相互独立

（对可列个事件，若其中任意有限个事件相互独立，则称这可列个事件相互独立）

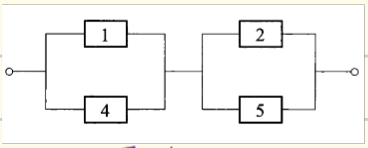
例 1.5.3　由 5 个独立元件组成的系统（如图 1.5.1 所示），设每个元件正常工作的概率为 $p, 0<p<1$，求系统正常工作的概率.



解：设 $A_i=\{$第i个元件正常工作$\}$ $i=1,2,3,4,5$. 由题知 $A_1, A_2 \cdots A_5$ 相互独立，设 $A=\{$系统正常工作$\}$

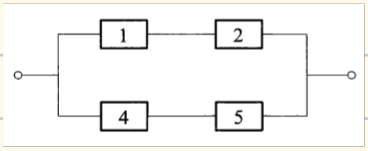
则 $A=AA_3\cup A\bar{A}_3$　$P(A)=P(AA_3\cup A\bar{A}_3)=P(A_3)\cdot P(A|A_3)+P(\bar{A}_3)\cdot P(A|\bar{A}_3)$.

①.记 $P(A|A_3)$ 为 p_1.　　$A_1\cup A_4$ 与 $A_2\cup A_5$ 相互独立



因此 $p_1=P((A_1\cup A_4)\cdot(A_2\cup A_5))=P(A_1\cup A_4)\cdot P(A_2\cup A_5)=(p+p-p^2)^2=(2p-p^2)^2$

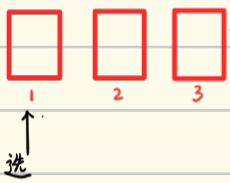
②.记 $P(A|\bar{A}_3)$ 为 p_2.　　$P(A_1A_2)=p^2=P(A_4A_5)$



$p_2=P(A_1A_2\cup A_4A_5)=P(A_1A_2)+P(A_4A_5)-P(A_1A_2A_4A_5)=2p^2-p^4$

$\therefore P(A)=p(2p-p^2)^2+(1-p)(2p^2-p^4)=2p^2+2p^3-5p^4+2p^5$

·补条件概率：eg. Monty Hall Problem：3扇门，2扇山羊，1扇汽车，先选1扇门，主持人选一扇打开为山羊，问是否换门？



选择门1 < A:选的门后为汽车　B:选的门后为山羊　C:主持人选门3且为山羊

$\Rightarrow P(C|A)=\frac{1}{2}$　$P(C|B)=\frac{1}{2}$　$P(A)=\frac{1}{3}$　$P(B)=\frac{2}{3}$

因此：全概率公式：$P(C)=\frac{1}{6}+\frac{1}{3}=\frac{1}{2}$　$P(A|C)=\frac{P(AC)}{P(C)}=\frac{1}{3}$　$P(B|C)=\frac{P(BC)}{P(C)}=\frac{2}{3}$.　$\therefore$ 换门

例5.6　一袋中有编号为1,2,3,4共4个球，采用放回抽样，每次取一球，共取2次，记录号码之和，这样独立重复进行试验，求“和等于3”出现在“和等于5”之前的概率。

$A=\{$和=3在和=5之前$\}$　$B=\{$第一次和为3$\}$　$C=\{$第一次和为5$\}$　$D=\{$第一次和既不为3也不为5$\}$

$P(B)=\frac{2}{16}$　$P(C)=\frac{4}{16}$　$P(D)=\frac{10}{16}$

$$P(A)=P(B)P(A|B)+P(A|C)P(C)+P(A|D)P(D)$$
$$=\frac{2}{16}\times 1+\frac{4}{16}\times 0+\frac{10}{16}\times P(A|D)$$
$$\Rightarrow\quad P(A)=\frac{1}{3}.$$

$P(A|D)=P(A)$

在第一次和不等于3或5的情况下求A的条件概率，相当于重新考虑A的概率。

二、随机变量及其概率分布

随机事件：$A=\{e: X(e)\in I\}=\{X\in I\}$，$I\subset R$

·定义：随机试验的样本空间为S，若 $X=X(e)$ 为定义在样本空间S上的实值单值函数，$e\in S$，则称 $X=X(e)$ 为随机变量.

若随机变量的所有可能取值为有限个或可列个，则称此随机变量为离散型随机变量.

设X为离散型随机变量，可能取值：$x_1, x_2, \cdots x_k, \cdots$，则称 $P\{X=x_k\}=p_k$，$k=1,2,\cdots$ 为X的概率分布列(律)

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

满足：(1) $p_k\geq 0$，$k=1,2,\cdots$ (2) $\sum_{k=1}^{+\infty} p_k=1$

eg. 例 2.2.1 设随机变量 X 的概率分布律为 ★

$$P\{X=k\}=\frac{c\cdot\lambda^k}{k!},\quad \lambda>0,\quad k=0,1,2,\cdots,$$

求常数 c 的值.

★ $e^x=\sum_{n=0}^{+\infty}\frac{x^n}{n!}$

$\therefore \sum_{k=0}^{\infty} P(X=k)=c\cdot e^{\lambda}=1 \Rightarrow c=e^{-\lambda}$

· 0-1 分布　$0<p<1$　X服从参数为p的0-1分布(两点分布)

X	0	1
P	1-p	p

$X\sim 0\text{-}1(p)$ / $B(1,p)$

·二项分布：随机变量X的概率分布律为 $P\{X=k\}=C_n^k p^k(1-p)^{n-k}$，$k=0,1,2,\cdots,n$.　$0<p<1$，$n\geq 1$.　X服从参数为(n,p)的二项分布　$X\sim B(n,p)$

定义 2.2.1 设在 n 次独立重复试验中，每次试验都只有两个结果：$A,\overline{A}$，且每次试验中 A 发生的概率不变，记 $P(A)=p, 0<p<1$，称这一系列试验为 n 重伯努利 (Bernoulli) 试验.

·泊松分布：随机变量X的概率分布律为 $P\{X=k\}=\frac{e^{-\lambda}\lambda^k}{k!}$，$k=0,1,2\cdots$　$\lambda>0$，称X服从参数为λ的泊松分布，记为 $X\sim P(\lambda)$

当n足够大，p充分小(一般要求 $p<0.1$)，且np保持适当大小时，参数为(n,p)的二项分布也可用泊松分布近似.

设 $X\sim B(n,p)$ 记 $\lambda=np$ 则 $P\{X=k\}=C_n^k p^k(1-p)^{n-k}=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}=\frac{n!}{k!(n-k)!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k}=\frac{n(n-1)\cdots(n-k+1)}{n^k}\cdot\frac{\lambda^k}{k!}\cdot\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}$　$\Rightarrow$ n充分大：$(1-\frac{\lambda}{n})^n=e^{n\ln(1-\frac{\lambda}{n})}\doteq e^{n[-\frac{\lambda}{n}-\frac{1}{2}\frac{\lambda^2}{n^2}]}\approx e^{-\lambda}$

$(1-\frac{\lambda}{n})^k\approx 1$；$\frac{n(n-1)\cdots(n-k+1)}{n^k}\approx 1$　$\therefore P\{X=k\}=\frac{e^{-\lambda}\cdot\lambda^k}{k!}$　n充分大，p足够小时.

eg. 例 2.2.5 设某公共汽车站单位时间内候车人数 X 服从参数为 4.8 的泊松分布，求：

(1) 随机观察 1 个单位时间，至少有 3 人候车的概率；

(2) 随机地独立观察 5 个单位时间，恰有 4 个单位时间至少有 3 人候车的概率.

(1). 由题 $X\sim P(\lambda)$，其中 $\lambda=4.8$ 则. $P\{X\geq 3\}=1-P\{X=0\}-P\{X=1\}-P\{X=2\}=1-e^{-4.8}-4.8e^{-4.8}-\frac{4.8^2}{2!}e^{-4.8}=0.8575$

(2). 被观察的5个单位时间内有Y个单位时间是"至少3人候车"，则 $Y\sim B(5,p)$.　$p=P\{X\geq 3\}$

$\Rightarrow P\{Y\geq 4\}=C_5^4 p^4\cdot(1-p)=0.3852$.

·其他：超几何分布：

例 2.2.7 一袋中共有 N 个球，其中有 a 个白球与 b 个红球 $(a+b=N)$. 从中无放回地取 $n(n\leqslant N)$ 个球，设每次取到各球的概率相等. 若其中有 X 个白球，试写出 X 的概率分布律.

(解) 这是一个等可能概型，

$$P\{X=k\}=\frac{C_a^k C_b^{n-k}}{C_N^n},\quad k=l_1,l_1+1,\cdots,l_2, \tag{2.2.5}$$

其中 $l_1=\max\{0,n-b\}$，$l_2=\min\{a,n\}$.

如果随机变量 X 具有形如 (2.2.5) 的概率分布律，就称 X 服从超几何分布 (hypergeometric distribution)，记为 $H(n,a,N)$. 图 2.2.4 是超几何分布的概率分布律柱形图.

·几何分布：

例 2.2.8 设独立重复试验中每次试验有两个结果：$A,\overline{A}$，且每次试验中 A 出现的概率不变，记 $P(A)=p\ (0<p<1)$. 又设直至 A 首次发生时所需的试验次数为 X，求 X 的概率分布律.

(解) 设 $A_i=\{$第 i 次试验 A 发生$\}$，$i=1,2,\cdots$，由题意知，$A_1,A_2,\cdots,A_n,\cdots$ 相互独立，且 $P(A_i)=p, i=1,2,\cdots$，$P\{X=k\}=P(\overline{A}_1\overline{A}_2\cdots\overline{A}_{k-1}A_k)$. 所以

$$P\{X=k\}=p(1-p)^{k-1},\quad k=1,2,\cdots. \tag{2.2.6}$$

如果随机变量 X 具有形如 (2.2.6) 的概率分布律，就称 X 服从参数为 p 的几何分布 (geometric distribution). 图 2.2.5 是几何分布的概率分布律折线图.

·负二项分布.

例 2.2.9 设独立重复试验中每次试验有两个结果：$A,\overline{A}$，且每次试验中 A 出现的概率不变，记 $P(A)=p(0<p<1)$. 设直至 A 发生 $r(r\geqslant 1)$ 次时所需的试验次数为 X，求 X 的概率分布律.

(解) 设 $A_i=\{$第 i 次试验 A 发生$\}$，$i=1,2,\cdots$. 由题意知，$A_1,A_2,\cdots$ 相互独立，且 $P(A_i)=p(i=1,2,\cdots)$. 因此

$$\begin{aligned}P\{X=k\}&=P\{\text{前 } k-1 \text{ 次中恰有 } r-1 \text{ 次 } A \text{ 发生，且第 } k \text{ 次 } A \text{ 发生}\}\\&=C_{k-1}^{r-1}p^r(1-p)^{k-r},\quad k=r,r+1,r+2,\cdots.\end{aligned} \tag{2.2.7}$$

如果随机变量 X 具有形如 (2.2.7) 的概率分布律，就称 X 服从参数为 (r,p) 的帕斯卡分布 (Pascal distribution)，也称为负二项分布 (negative binomial distribution)，记为 $NB(r,p)$. 几何分布即为 $r=1$ 的帕斯卡分布. 图 2.2.6 是帕斯卡分布的概率分布律折线图.

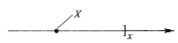
·随机变量的概率分布函数.

注意等号 $\Rightarrow F(x)-F(x-0)=P(X=x)$

·定义：X为一随机变量，x为任意实数，函数 $F(x)=P\{X\leq x\}$ 称为随机变量X的概率分布函数

对 $\forall x_1,x_2\in R\ (x_1<x_2)$ 有 $P\{x_1<X\leq x_2\}=P\{X\leq x_2\}-P\{X\leq x_1\}=F(x_2)-F(x_1)$

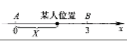
从几何上看分布函数，将 X 设想成一随机点，那么 X 落在区间 $(-\infty,x]$ 上的概率即为 $F(x)$ (如图 2.3.1 所示).



◀图 2.3.1

性质：(1). $F(x)$ 单调不减　(2) $0\leq F(x)\leq 1$ 且 $\lim_{a\to-\infty}F(a)=0$　$\lim_{b\to+\infty}F(b)=1$.　(3). $F(x+0)=F(x)$，即 $F(x)$ 为右连续函数.

eg. 例 2.3.3 设某人在 A,B 两点间移动，A,B 两点间的距离为 3 个单位. 设他离 A 点的距离为 X (如图 2.3.4 所示)，且处于 A,B 两点间任一区间的概率与区间长度成正比，求 X 的分布函数.



◀图 2.3.4

分段的分布函数范围：左闭右开

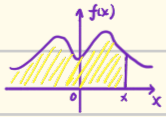
$P\{0\leq X\leq 3\}=1$. 比例常数为 k　$P\{X\}=kX$　$\therefore P\{0\leq X\leq 3\}=3k=1 \Rightarrow k=\frac{1}{3}$.

$\therefore x<0$ 时，$F(x)=0$　$x\geq 3$ 时，$F(x)=1$.　$0\leq x<3$ 时，$F(x)=\frac{x}{3}$　$F(x)=\begin{cases}0, & x<0\\ \frac{x}{3}, & 0\leq x<3\\ 1, & x\geq 3\end{cases}$

·连续型随机变量

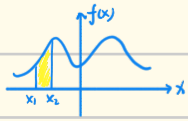
·定义：对随机变量X，分布函数为F(x)，若存在一个非负的实值函数f(x)，$-\infty<x<+\infty$，使得对任意实数x，有 $F(x)=\int_{-\infty}^{x}f(t)\,dt$，则称X为连续型随机变量，f(x)为X的概率

密度函数



(1). $f(x)\geq 0$　(2). $\int_{-\infty}^{+\infty}f(x)\,dx=1$

(3). $\forall x_1,x_2\ (x_1<x_2)$　$P\{x_1<X\leq x_2\}=F(x_2)-F(x_1)=\int_{x_1}^{x_2}f(t)\,dt$



连续型随机变量取任一定值的概率为0

(4). f(x)的连续点x处，$F'(x)=f(x)$　$P\{x<X\leq x+\Delta x\}\approx f(x)\Delta x$

eg. 例 2.4.1　随机变量 X 的密度函数为

$$f(x)=\begin{cases}k(1-x), & 0<x<1,\\ \frac{x}{4}, & 1\leqslant x\leqslant 2,\\ 0, & \text{其他}.\end{cases}$$

(1) 求常数 k; (2) 求分布函数 $F(x)$; (3) 求 $P\{X<1.5\}$.

(1). 由密度函数性质：$\int_{-\infty}^{+\infty}f(x)\,dx=\int_0^1 k(1-x)\,dx+\int_1^2\frac{x}{4}\,dx=\frac{k}{2}+\frac{3}{8}=1\Rightarrow k=\frac{5}{4}$

(2). $F(x)=\int_{-\infty}^{x}f(t)\,dt$　∴ 当 $x\leq 0$ 时，$F(x)=0$　当 $x>2$ 时，$F(x)=1$　当 $0<x<1$ 时，$F(x)=\int_0^x\frac{5}{4}(1-t)\,dt=\frac{5}{4}(x-\frac{x^2}{2})$

当 $1\leq x\leq 2$ 时，$F(x)=\int_0^1\frac{5}{4}(1-t)\,dt+\int_1^x\frac{t}{4}\,dt=\frac{x^2}{8}+\frac{1}{2}$

即：$F(x)=\begin{cases}0, & x\leq 0\\ \frac{5}{4}(x-\frac{x^2}{2}), & 0<x<1\\ \frac{x^2}{8}+\frac{1}{2}, & 1\leq x\leq 2\\ 1, & x>2.\end{cases}$

(3). $P\{X<1.5\}=F(1.5)=\frac{25}{32}$.

某人进了银行，且计划稍后还要去办另一件事，故打算先等待，如果 15 min 后还是没有等到服务就离开银行，设此人在银行实际等待时间为 Y.

(1) 求 Y 的分布函数;

(2) 问 Y 是离散型随机变量吗? Y 是连续型随机变量吗?

解　(1) 由分布函数的定义知，Y 的分布函数为

$$F(y)=P\{Y\leqslant y\}.$$

显然，当 $y\leqslant 0$ 时，$F(y)=0$; 当 $y\geqslant 15$ 时，$F(y)=1$; 当 $0<y<15$ 时，

$$F(y)=P\{Y\leqslant y\}=P\{X\leqslant y\}=\int_{-\infty}^{y}f(x)\mathrm{d}x=1-\mathrm{e}^{-y/10}.$$

即

$$F(y)=\begin{cases}0, & y\leqslant 0,\\ 1-\mathrm{e}^{-y/10}, & 0<y<15,\\ 1, & y\geqslant 15.\end{cases}$$

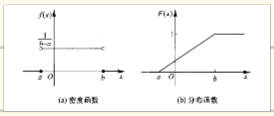
(2) 因为 Y 的取值范围为 $[0,15]$，Y 的取值不可数，所以 Y 不是离散型随机变量. 又

$$P\{Y=15\}=F(15)-P\{Y<15\}=1-(1-\mathrm{e}^{-15/10})=\mathrm{e}^{-1.5}\neq 0,$$

即 Y 取定值 15 的概率不为零，所以 Y 亦不是连续型随机变量. 因此，Y 是既非离散型又非连续型的随机变量.

·**均匀分布**：随机变量有密度函数 $f(x)=\begin{cases}\frac{1}{b-a}, & x\in(a,b)\\ 0, & \text{其他}\end{cases}$，则称X服从区间(a,b)上均匀分布，记为 $X\sim U(a,b)$

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



实数c，l满足 $a\leq c<c+l\leq b$ 则 $P\{c<X<c+l\}=\int_c^{c+l}\frac{1}{b-a}\,dx=\frac{l}{b-a}$

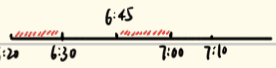
X落在(a,b)内任一长为l的子区间的概率与位置无关

eg. 例 2.4.4　杭州某长途汽车站每天从早上 6:00 (第一班车) 开始，每隔 30 min 有一班车发往上海. 设王先生在早上 6:20 以后 X min 到站，又设 X 服从 (0, 50) 上均匀分布.

(1) 求王先生候车时间不到 15 min 的概率;

(2) 如果王先生一个月中有两次按此方式独立地去候车，求他有一次候车时间不到 15 min、另一次候车时间大于 10 min 的概率.

(1). $X\sim U(0,50)$



$P\{$候车时间小于15 min$\}=\frac{25}{50}=\frac{1}{2}$.

(2). $P\{$候车时间大于10 min$\}=\frac{30}{50}=\frac{3}{5}$　$A_i=\{$第i次候车小于15 min$\}$　$B_i=\{$第i次候车大于10 min$\}$

$P(A_1B_1)=P(A_2B_2)=\frac{1}{10}$　∴ $P\{A_1B_2\cup A_2B_1\}=P(A_1B_2)+P(B_1A_2)-P(A_1B_1A_2B_2)=2\times\frac{1}{2}\times\frac{3}{5}-\frac{1}{100}=\frac{59}{100}$

·**正态分布**

定义 2.4.3　设随机变量 X 具有密度函数

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)},\quad -\infty<x<+\infty,\tag{2.4.4}$$

其中 $-\infty<\mu<+\infty$, $\sigma>0$, 则称 X 服从参数为 (μ,σ) 的正态分布 (normal distribution), 简称 X 为正态变量, 记为 $X\sim N(\mu,\sigma^2)$.

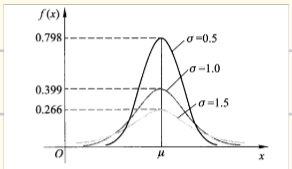
其相应的分布函数为

$$F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(t-\mu)^2/(2\sigma^2)}\mathrm{d}t.$$

证 $\int_{-\infty}^{+\infty}f(x)\,dx=1\Rightarrow I=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}dx$　令 $\frac{x-\mu}{\sigma}=t$，则 $I=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$　$I^2=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy$

$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{2\pi}e^{-(x^2+y^2)/2}dx\,dy\xlongequal[y=r\sin\theta]{x=r\cos\theta}\int_0^{2\pi}d\theta\int_0^{+\infty}r\cdot\frac{1}{2\pi}\cdot e^{-r^2/2}dr=1\Rightarrow I=1$

性质：f(x)关于 $x=\mu$ 对称；$f(x)_{max}=f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$；$\lim_{|x-\mu|\to+\infty}f(x)=0$



μ：位置参数　对称轴位置

σ：尺度参数　分散程度

$\mu=0,\sigma=1$；$Z\sim N(0,1)$ 标准正态分布

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

分布函数：$\Phi(z)=\int_{-\infty}^{+\infty}\varphi(t)\,dt$

$\Phi(-z_0)=1-\Phi(z_0)$

同时，当 $X\sim N(\mu,\sigma^2)$ 时，对于任意实数 $a,b(a<b)$，有

$$P\{a<X<b\}=\int_a^b\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\mathrm{d}x.$$

作积分变量变换，令 $t=\frac{x-\mu}{\sigma}$，得

$$P\{a<X<b\}=\int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-t^2/2}\mathrm{d}t.$$

此时的被积函数为标准正态分布的密度函数，故

$$P\{a<X<b\}=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right).\tag{2.4.7}$$

这样将正态变量的概率计算归结为标准正态分布函数值的计算问题. 而这些数值可查本书附表 2.

若 $X\sim N(\mu,\sigma^2)$，由 (2.4.5), (2.4.6) 可得

$$P\{|X-\mu|<k\sigma\}=\Phi(k)-\Phi(-k)=2\Phi(k)-1.$$

例 2.4.5　用天平称一实际重量为 a 的物体，天平的读数为随机变量 $X, X\sim N(a,\sigma^2)$. 当 $\sigma=0.01$ 时，

(1) 求读数与 a 的误差小于 0.005 的概率;

(2) 求读数至少比 a 多 0.008 5 的概率;

(3) 若重复称 3 次，求 3 次中恰有 1 次 "读数与 a 的误差小于 0.005" 的概率.

(1). $P\{|X-a|<0.005\}=\Phi(\frac{0.005}{\sigma})-\Phi(-\frac{0.005}{\sigma})=2\Phi(0.5)-1=0.3830$

(2). $P\{X-a\geq 0.0085\}=1-\Phi(0.85)=1-0.8023=0.1977$

(3). 3次中有Y次"读数与a的误差小于0.005"　$Y\sim B(3,0.3830)$.

$X\sim N(\mu,\sigma)$ 若 $Y=aX+b\Rightarrow Y\sim N(a\mu+b,a^2\sigma^2)$

∴ $P\{Y=1\}=C_3^1\times 0.3830\times 0.6170^2\approx 0.4374$

·**指数分布**：随机变量X有密度函数 $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\leq 0\end{cases}$　$\lambda>0$，称X服从参数为λ的指数分布，$X\sim E(\lambda)$.

$\int_{-\infty}^{+\infty}f(x)\,dx=\int_0^{+\infty}\lambda e^{-\lambda x}dx=-e^{-\lambda x}\Big|_0^{+\infty}=1$

分布函数：$F(x)=\int_{-\infty}^{x}f(t)\,dt=\int_0^x\lambda e^{-\lambda t}dt=-\int_0^x e^{-\lambda t}d(-\lambda t)=-e^{-\lambda t}\Big|_0^x=1-e^{-\lambda x}\Rightarrow F(x)=\begin{cases}1-e^{-\lambda x}, & x>0\\ 0, & x\leq 0\end{cases}$

当 $X\sim E(\lambda)$，$\forall t>0, t_0>0$ 有：$P\{X>t_0+t\mid X>t_0\}=\frac{P\{X>t_0+t\}}{P\{X>t_0\}}=\frac{1-F(t_0+t)}{1-F(t_0)}=e^{-\lambda t}=P\{X>t\}\Rightarrow$ 理解！

连续型中只有指数分布有无记忆性.

若将 X 看成某电子产品的寿命 (单位: h)，则 (2.4.10) 可解释为: "在已知产品用了 t_0 h 没有坏的条件下，再用 t h 不坏" 的条件概率等于此产品 "最初使用 t h 不坏" 的概率. 形象地说此产品 "忘却" 了 "已使用 t_0 h". 所以常将 (2.4.10) 形象地称作指数分布的 "无记忆性".

·其他连续型随机变量.

Γ分布:

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

$\alpha>0,\lambda>0$,称X服从参数为(α,λ)的Γ分布,记$X\sim\Gamma(\alpha,\lambda)$

$\Gamma(\alpha)=\int_0^{+\infty}e^{-x}x^{\alpha-1}dx$,且有$\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$; $n\in N^+$ $\Gamma(n)=(n-1)!$

威布尔分布

$$f(x)=\begin{cases}\dfrac{\gamma}{\alpha}\left(\dfrac{x}{\alpha}\right)^{\gamma-1}e^{-(x/\alpha)^{\gamma}}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$$

$\alpha>0,\gamma>0$,称X服从参数为(α,γ)的二参数威布尔分布

定义 2.4.7 记随机变量 X 具有密度函数

$$f(x)=\begin{cases}\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, & 0<x<1,\\ 0, & \text{其他},\end{cases}\tag{2.4.14}$$

其中参数 $a>0,b>0$, 则称 X 服从参数为 (a,b) 的 β 分布 (Beta distribution), 记为 $X\sim\beta(a,b)$.

·随机变量函数的分布.

已知X的分布, Y=g(X)的分布?

eg. X的分布律

X	-1	0	1	2
P	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$

知: $Y=2X+1$ $Z=X^2$.

∴ $Y=-1,1,3,5$ ⇒

Y	-1	1	3	5
P	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$

$Z=\underbrace{1,0,1}_{\text{合并}:\frac{1}{3}+\frac{1}{6}=\frac{1}{2}},4$ ⇒

Z	0	1	4
P	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

例 2.5.4 设 X 的密度函数为 $f_X(x)$, $Y=|X|$, $Z=X^2$, 分别求 Y 与 Z 的密度函数 $f_Y(y)$, $f_Z(t)$.

设Y与Z的分布函数分别为$F_Y(y)$, $F_Z(t)$. 当$y\leq 0$,显然$F_Y(y)=0$ 当$y>0$, $F_Y(y)=P\{Y\leq y\}=P\{|X|\leq y\}=F_X(y)-F_X(-y)$ ⇒ $f_Y(y)=\begin{cases}f_X(y)+f_X(-y), & y>0\\ 0, & y\leq 0\end{cases}$

当$t\leq 0$时,显然$F_Z(t)=0$; $t>0$时, $F_Z(t)=P\{Z\leq t\}=P\{X^2\leq t\}=F_X(\sqrt{t})-F_X(-\sqrt{t})$ ⇒ $f_Z(t)=\begin{cases}\frac{1}{2\sqrt{t}}[f_X(\sqrt{t})+f_X(-\sqrt{t})], & t>0\\ 0, & t\leq 0\end{cases}$

·特别地,当函数$y=g(x)$具有严格单调性时,有:

·定理:设X为一连续型随机变量,其密度函数为$f_X(x)$,随机变量$Y=g(X)$,若函数$y=g(x)$为一处处可导的严格单调增/减函数,记$y=g(x)$的反函数为$x=h(y)$

则Y的密度函数为: $f_Y(y)=\begin{cases}f_X(h(y))\cdot|h'(y)|, & y\in D\\ 0, & y\notin D\end{cases}$

证明 先设 $y=g(x)$ 为一严格单调增函数, 即 $g'(x)>0$, 注意此时的 $h'(y)>0$, 那么当 $y\in D$ 时, Y 的分布函数为

$$\begin{aligned}F_Y(y)&=P\{Y\leqslant y\}=P\{g(X)\leqslant y\}\\&=P\{X\leqslant h(y)\}=\int_{-\infty}^{h(y)}f_X(x)\mathrm{d}x.\end{aligned}$$

从而

$$f_Y(y)=F_Y'(y)=f_X(h(y))\cdot h'(y)=f_X(h(y))\cdot|h'(y)|.$$

若 $y=g(x)$ 为一严格单调减函数, 即 $g'(x)<0$, 此时有 $h'(y)<0$, 那么当 $y\in D$ 时, Y 的分布函数为

$$\begin{aligned}F_Y(y)&=P\{Y\leqslant y\}=P\{g(X)\leqslant y\}\\&=P\{X\geqslant h(y)\}=\int_{h(y)}^{+\infty}f_X(x)\mathrm{d}x.\end{aligned}$$

从而

$$f_Y(y)=F_Y'(y)=-f_X(h(y))\cdot h'(y)=f_X(h(y))\cdot|h'(y)|.$$

而当 $y\notin D$ 时, $f_Y(y)=0$. 于是就证明了 (2.5.1) 式.

☆

更一般地, 我们有如下结论 (见书本例 2.5.6).

Theorem 2 (Probability integral transformation)

Let X have continuous c.d.f. $F_X(x)$ and define the random variable Y as $Y=F_X(X)$. Then Y is uniformly distributed on $(0,1)$.

连续的分布函数

9.多维随机变量及其分布

·二维离散型随机变量：若二维随机变量 (X,Y) 的取值有限或可列，则称 (X,Y) 为二维离散型随机变量

二维离散型随机变量的联合分布：设 (X,Y) 的可能取值为 (x_i,y_j)，$i,j=1,2,\cdots$，称 $P\{X=x_i,Y=y_j\}=p_{ij}$，$i,j=1,2\cdots$ 为 (X,Y) 的联合概率分布律（联合分布律）

(1) $p_{ij}\geq 0$ $i,j=1,2\cdots$ (2). $\sum_i\sum_j p_{ij}=1$

eg.

例 3.1.1 一袋中有 7 个球，其中 4 个白球，1 个红球和 2 个黑球. 每次摸 1 球，无放回抽样 3 次. 设 3 次中有 X 次摸到白球，Y 次摸到红球，求 (X,Y) 的联合分布律.

由题知：X可能值为0,1,2,3 Y可能为0,1 记 $p(i,j)=P\{X=i,Y=j\}$，则

$P(0,0)=0$ $P(0,1)=\frac{C_4^1C_2^2}{C_7^3}=\frac{1}{35}$ $P(1,0)=\frac{C_4^1\cdot C_2^2}{C_7^3}=\frac{4}{35}$ $P(1,1)=\frac{C_4^1C_2^1}{C_7^3}=\frac{8}{35}$ $P(2,0)=\frac{C_4^2\cdot C_2^1}{C_7^3}=\frac{12}{35}$

$P(2,1)=\frac{C_4^2}{C_7^3}=\frac{6}{35}$ $P(3,0)=\frac{C_4^3}{C_7^3}=\frac{4}{35}$ $P(3,1)=0$

显然有 $p_{i\cdot}\geqslant 0$, $p_{\cdot j}\geqslant 0$, $\sum_i p_{i\cdot}=1$, $\sum_j p_{\cdot j}=1$, 即 $p_{i\cdot}, i=1,2,\cdots$ 与 $p_{\cdot j}, j=1,2,\cdots$ 满足概率分布律的性质, 它们分别是随机变量 X 与 Y 的概率分布律, 称为 X 和 Y 的边际分布律 (marginal distribution law) 或边缘分布律. 下面用列表的方法来表示联合分布律及边际分布律, 并由此更清晰地理解它们之间的关系:

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

上表内第 i 行 (或第 j 列) 累计后记作 $p_{i\cdot}$ (或 $p_{\cdot j}$). 上表中最右侧的一列 (或最下面一行) 是 X (或 Y) 的概率分布律, 故称其为边际分布律.

二维离散型随机变量的边际分布：二维离散型随机变量 (X,Y) 的联合分布律：$P\{X=x_i,Y=y_j\}=p_{ij}$，$i,j=1,2,\cdots$，

$P\{X=x_i\}=P(\bigcup_{j=1}^{+\infty}\{X=x_i,Y=y_j\})=\sum_{j=1}^{+\infty}p_{ij}\overset{def}{=\!=}p_{i\cdot}$，$i=1,2,\cdots$ （$\overset{def}{=\!=}$ 表示"记为"） 同理：$P(Y=y_j)=\sum_{i=1}^{+\infty}p_{ij}\overset{def}{=\!=}p_{\cdot j}$，$j=1,2\cdots$

eg.

例 3.1.2 设一群体中 80% 的人不吸烟, 15% 的人少量吸烟, 5% 的人吸烟较多, 且已知近期他们患呼吸道疾病 (以下简称患病) 的概率分别为 5%, 25%, 70%. 记

$$X=\begin{cases}0, & \text{不吸烟},\\ 1, & \text{少量吸烟},\\ 2, & \text{吸烟较多}.\end{cases}\qquad Y=\begin{cases}1, & \text{患病},\\ 0, & \text{不患病}.\end{cases}$$

(1) 求 (X,Y) 的联合分布律与边际分布律;

(2) 求患者中吸烟的概率.

(1). $P(X=0)=0.80$ $P(X=1)=0.15$ $P(X=2)=0.05$

$P(Y=1|X=0)=0.05$ $P(Y=1|X=1)=0.25$ $P(Y=1|X=2)=0.70$

$P(0,1)=0.04$ $P(1,1)=0.0375$ $P(2,1)=0.035$ $P(0,0)=0.76$ $P(1,0)=0.1125$

$P(2,0)=0.015$

可得联合分布律及边际分布律：

X	Y		$P\{X=i\}$
	0	1	
0	0.76	0.04	0.80
1	0.112 5	0.037 5	0.15
2	0.015	0.035	0.05
$P\{Y=j\}$	0.887 5	0.112 5	1

(2). $P\{\{X=1\}\cup\{X=2\}|Y=1\}=\frac{P(1,1)+P(2,1)}{P\{Y=1\}}=\frac{0.0725}{0.1125}\approx 0.6444$

·二维离散型随机变量的条件分布

设二维离散型随机变量 (X,Y) 的联合分布律为 $P\{X=x_i,Y=y_j\}=p_{ij}$, $i,j=1,2,\cdots$, 则当 $P\{Y=y_j\}\neq 0$ 时,

$$P\{X=x_i|Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\cdot j}},\quad i=1,2,\cdots. \tag{3.1.4}$$

同理可得当 $P\{X=x_i\}\neq 0$ 时,

$$P\{Y=y_j|X=x_i\}=\frac{P\{X=x_i,Y=y_j\}}{P\{X=x_i\}}=\frac{p_{ij}}{p_{i\cdot}},\quad j=1,2,\cdots. \tag{3.1.5}$$

称 (3.1.4)(或 (3.1.5)) 为给定 $\{Y=y_j\}$ (或 $\{X=x_i\}$) 的条件下 X (或 Y) 的条件分布律 (conditional distribution law).

(3.1.4) 中显然有 $\frac{p_{ij}}{p_{\cdot j}}\geqslant 0$ 且 $\sum_{i=1}^{+\infty}\frac{p_{ij}}{p_{\cdot j}}=1$, 同样 (3.1.5) 中有 $\frac{p_{ij}}{p_{i\cdot}}\geqslant 0$ 且 $\sum_{j=1}^{+\infty}\frac{p_{ij}}{p_{i\cdot}}=1$, 亦即 (3.1.4) 及 (3.1.5) 满足概率分布律的性质.

eg.

例 3.1.4 设某客车乘车人数 X 服从参数为 λ 的泊松分布, 每位乘车人的下车行为相互独立, 且每位乘车人在中途下车 (没有坐到终点站) 的概率为 $p(0<p<1)$. 设中途只下不上, 并记中途下车的人数为 Y.

(1) 求 (X,Y) 的联合分布律;

(2) 求 Y 的边际分布律;

(3).求给定 $\{Y=5\}$ 的条件下X的条件分布律

(1). 知 $P(X=m)=\frac{e^{-\lambda}\cdot\lambda^m}{m!}$，$m=0,1,\cdots$ 当 $m=0,1,\cdots$ 时，$P(Y=n|X=m)=C_m^n\cdot p^n(1-p)^{m-n}$，$n=0,1,\cdots,m$

$P\{X=m,Y=n\}=P\{X=m\}\cdot P\{Y=n|X=m\}=\frac{e^{-\lambda}\cdot\lambda^m}{m!}C_m^np^n(1-p)^{m-n}$，$m=0,1,\cdots$ $n=0,1,\cdots,m$

(2). $P\{Y=n\}=\sum_{m=0}^{+\infty}P\{X=m,Y=n\}=\sum_{m=0}^{n-1}0+\sum_{m=n}^{+\infty}\frac{e^{-\lambda}\cdot\lambda^m}{m!}\cdot\frac{m!}{n!(m-n)!}p^n(1-p)^{m-n}=\frac{(\lambda p)^ne^{-\lambda}}{n!}\sum_{j=0}^{+\infty}\frac{[\lambda(1-p)]^j}{j!}$

$=\frac{(\lambda p)^ne^{-\lambda}}{n!}\cdot e^{\lambda(1-p)}=\frac{e^{-\lambda p}(\lambda p)^n}{n!}$，$n=0,1,\cdots$，即 $Y\sim P(\lambda p)$

(3). $P\{X=m|Y=5\}=\frac{P\{X=m,Y=5\}}{P\{Y=5\}}=\frac{\frac{e^{-\lambda}\lambda^m}{m!}C_m^5p^5(1-p)^{m-5}}{\frac{e^{-\lambda p}(\lambda p)^5}{5!}}=\frac{e^{-\lambda(1-p)}[\lambda(1-p)]^{m-5}}{(m-5)!}$，$m=5,6\cdots$

二维随机变量的分布函数

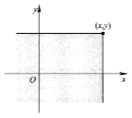
·联合分布函数

定义 3.2.1　设二维随机变量 (X,Y), 对于任意的实数 x,y, 称函数

$$F(x,y)=P\{X\leqslant x,Y\leqslant y\} \tag{3.2.1}$$

为 (X,Y) 的联合概率分布函数, 简称联合分布函数 (joint distribution function).

若将 (X,Y) 看作随机点的坐标, 则分布函数 $F(x,y)$ 即为 (X,Y) 落在图 3.2.1 阴影部分区域的概率.



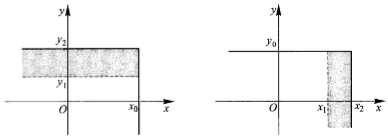
与一维随机变量的分布函数相类似, $F(x,y)$ 具有以下性质:

(1) 当给定 $x=x_0$ 时, $F(x_0,y)$ 关于 y 单调不减; 当给定 $y=y_0$ 时, $F(x,y_0)$ 关于 x 单调不减, 如图 3.2.2 所示.

(2) $0\leqslant F(x,y)\leqslant 1$, 且 $F(x,-\infty)=F(-\infty,y)=F(-\infty,-\infty)=0, F(+\infty,+\infty)=1$.

(3) $F(x,y)=F(x+0,y)$, $F(x,y)=F(x,y+0)$, 即 $F(x,y)$ 关于 x 右连续, 关于 y 右连续 (证略).

◀图 3.2.2



(4) 当实数 $x_2>x_1, y_2>y_1$ 时,

$$\begin{aligned}&F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)\\&=P\{x_1<X\leqslant x_2,y_1<Y\leqslant y_2\}\geqslant 0,\end{aligned}$$

·边际分布函数

记二维随机变量 (X,Y) 的联合分布函数为 $F(x,y)$, X,Y 的边际分布函数为 $F_X(x),F_Y(y)$, 则

$$F_X(x)=P\{X\leqslant x\}=P\{X\leqslant x,Y<+\infty\}=F(x,+\infty). \tag{3.2.3}$$

同理, $F_Y(y)=F(+\infty,y)$. 即, 二维随机变量的边际分布函数是联合分布函数当另一个变量趋于 $+\infty$ 时的极限函数.

以后我们不再一一说明, 常用 $F_X(x),F_Y(y)$ 表示 X,Y 的边际分布函数, 用 $F(x,y)$ 表示 (X,Y) 的联合分布函数.

·条件分布函数

设 (X,Y) 为二维离散型随机变量, 当 $P\{X=x_i\}\neq 0$ 时, 称函数

$$F_{Y|X}(y|x_i)=P\{Y\leqslant y|X=x_i\} \tag{3.2.4}$$

为给定 $\{X=x_i\}$ 的条件下 Y 的条件概率分布函数, 简称条件分布函数.

设: 设 (X,Y) 为二维连续型随机变量 (下一节介绍), 对任意给定的实数 x, 若 $P\{x<X\leqslant x+\delta\}>0$, 其中 $\delta>0$, 对任意实数 y, 称函数

$$F_{Y|X}(y|x)=\lim_{\delta\to 0^+}P\{Y\leqslant y|x<X\leqslant x+\delta\} \tag{3.2.5}$$

为给定 $\{X=x\}$ 的条件下 Y 的条件分布函数.

一般地, 设 (X,Y) 为二维随机变量, 对给定的实数 x, 若极限

$$\lim_{\delta\to 0^+}P\{Y\leqslant y|x-\delta<X\leqslant x+\delta\}=\lim_{\delta\to 0^+}P\{Y\leqslant y,x-\delta<X\leqslant x+\delta\}$$

对任何实数 y 均存在, 则称函数

$$F_{Y|X}(y|x)=\lim_{\delta\to 0^+}P\{Y\leqslant y|x-\delta<X\leqslant x+\delta\} \tag{3.2.6}$$

为给定 $\{X=x\}$ 的条件下 Y 的条件分布函数.

·二维连续型随机变量

(一) 二维连续型随机变量的联合分布

定义 3.3.1　设二维随机变量 (X,Y) 的联合分布函数为 $F(x,y)$, 若存在二元非负函数 $f(x,y)$, 使对任意的实数 x,y 有

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)\mathrm{d}u\mathrm{d}v, \tag{3.3.1}$$

则称 (X,Y) 为二维连续型随机变量 (bivariate continuous random variable), 称 $f(x,y)$ 为 (X,Y) 的联合概率密度函数 (joint probability density function), 简称联合密度函数.

$f(x,y)$ 具有以下性质 (其中 $F(x,y)$ 为 (X,Y) 的联合分布函数):

(1) $f(x,y)\geqslant 0$;

(2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y=F(+\infty,+\infty)=1$;

(3) 在 $f(x,y)$ 的连续点处有

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

(4) (X,Y) 落入 xOy 平面任一区域 D 的概率为

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

例3.1 设二元随机变量 (X,Y) 的概率密度函数

$$f(x,y)=\begin{cases}ke^{-(2x+3y)}, & x>0,\ y>0\\ 0, & 其他\end{cases}$$

(1)求常数 k;
(2)求联合分布函数 $F(x,y)$;
(3)求 $P(Y\leq X)$.

(1). $1=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\,dx\,dy=k\int_0^{\infty}e^{-2x}dx\int_0^{\infty}e^{-3y}dy=\frac{k}{6}=1\Rightarrow k=6$

(2). $F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)\,dv\,du$　　$G=\{(u,v):u\leq x, v\leq y\}$

$=\iint_G f(u,v)\,du\,dv=\iint_{G\cap D}f(u,v)\,du\,dv+\iint_{G\cap\bar{D}}f(u,v)\,du\,dv=\begin{cases}(1-e^{-2x})(1-e^{-3y}), & x>0,\ y>0\\ 0, & 其他\end{cases}$

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

$G=\{(x,y):y\leq x\}$　$=\iint_{G\cap D}f(x,y)\,dx\,dy=$　$=\frac{3}{5}$

(二) 二维连续型随机变量的边际分布

设 (X,Y) 为二维连续型随机变量, $F(x,y),f(x,y)$ 分别为 (X,Y) 的联合分布函数及联合密度函数, 称单个随机变量 X (或 Y) 的密度函数为 X (或 Y) 的边际概率密度函数 (marginal probability density function), 简称边际密度函数, 且常分别用 $f_X(x),f_Y(y)$ 表示. 由于

$$\begin{aligned}F_X(x)&=P\{X\leqslant x\}=P\{X\leqslant x,Y\in(-\infty,+\infty)\}\\&=\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y\right]\mathrm{d}x,\end{aligned}$$

由连续型随机变量的定义知 X 为连续型随机变量, 且 X 的边际密度函数为

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y. \tag{3.3.2}$$

同理可得

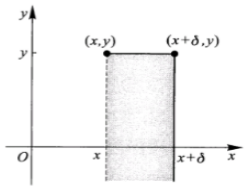
$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x. \tag{3.3.3}$$

即边际密度函数为联合密度函数关于另一个变量在 $(-\infty,+\infty)$ 上的积分.

(三) 二维连续型随机变量的条件分布

在下文中, 一般用 $f(x,y)$ 表示二维连续型随机变量 (X,Y) 的联合密度函数, 用 $f_X(x), f_Y(y)$ 分别表示 X,Y 的边际密度函数.

设 (X,Y) 为二维连续型随机变量, 由条件分布函数的定义 (3.2.5) 和图 3.3.3 知



◀图 3.3.3

$$
\begin{aligned}
F_{Y|X}(y|x) &= \lim_{\delta\to 0^+} P\{Y\leqslant y | x<X\leqslant x+\delta\} \\
&= \lim_{\delta\to 0^+} \frac{P\{x<X\leqslant x+\delta, Y\leqslant y\}}{P\{x<X\leqslant x+\delta\}} \\
&= \lim_{\delta\to 0^+} \frac{F(x+\delta,y)-F(x,y)}{F_X(x+\delta)-F_X(x)} \\
&= \lim_{\delta\to 0^+} \frac{(F(x+\delta,y)-F(x,y))/\delta}{(F_X(x+\delta)-F_X(x))/\delta}.
\end{aligned}
$$

由于

$$\lim_{\delta\to 0^+} \frac{F_X(x+\delta)-F_X(x)}{\delta} = f_X(x),$$

且

$$
\begin{aligned}
\lim_{\delta\to 0^+} \frac{F(x+\delta,y)-F(x,y)}{\delta} &= \frac{\partial}{\partial x}F(x,y) = \frac{\partial}{\partial x}\int_{-\infty}^{x}\left[\int_{-\infty}^{y} f(u,v)\mathrm{d}v\right]\mathrm{d}u \\
&= \int_{-\infty}^{y} f(x,v)\mathrm{d}v,
\end{aligned}
$$

即此时有

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} \frac{f(x,v)}{f_X(x)}\mathrm{d}v.$$

所以有下面的定义.

定义 3.3.2 设 (X,Y) 为二维连续型随机变量, $f(x,y)$ 为 (X,Y) 的联合密度函数, $f_X(x), f_Y(y)$ 为 X,Y 的边际密度函数. 给定 $\{X=x\}(f_X(x)\neq 0)$ 的条件下 Y 的条件概率密度函数 (conditional probability density function), 简称条件密度函数, 为

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \quad -\infty<y<+\infty. \tag{3.3.4}$$

同样, 给定 $\{Y=y\}(f_Y(y)\neq 0)$ 的条件下 X 的条件密度函数为

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad -\infty<x<+\infty. \tag{3.3.5}$$

条件密度函数性质(以 $f_{X|Y}(x|y)$ 为例):

1. $f_{X|Y}(x|y)\geq 0$,
2. $\int_{-\infty}^{+\infty} f_{X|Y}(x|y)dx = 1$,
3. $P\{a<X<b|Y=y\} = \int_a^b f_{X|Y}(x|y)dx$,
4. 在 $f_{X|Y}(x|y)$ 的连续点 x, 有 $\dfrac{\mathrm{d}F_{X|Y}(x|y)}{\mathrm{d}x} = f_{X|Y}(x|y)$,
5. $f(x,y) = f_{X|Y}(x|y)f_Y(y) = f_X(x)f_{Y|X}(y|x)$.

· 二元均匀分布和二元正态分布

定义 3.3.3 设二维随机变量 (X,Y) 在二维有界区域 D 上取值, 且具有联合密度函数

$$f(x,y) = \begin{cases} \dfrac{1}{D\text{ 的面积}}, & (x,y)\in D, \\ 0, & \text{其他}, \end{cases} \tag{3.3.6}$$

则称 (X,Y) 服从 D 上均匀分布.

eg. 例3.5 设二元随机变量(X,Y)在区域 $\{(x,y):|y|<x<1\}$ 内均匀分布，求条件密度函数 $f_{X|Y}(x|y)$ 及 $P(X>\frac{2}{3}|Y=\frac{1}{2})$.

$f(x,y)=\begin{cases}1, & |y|<x<1\\ 0, & \text{其他}\end{cases}$

$\therefore$ Y的边际概率密度函数：$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx$

$=\begin{cases}\int_{|y|}^{1}dx=1-|y|, & -1<y<1\\ 0, & \text{其他}\end{cases}$

$\therefore f_{X|Y}(x|y)=\begin{cases}\frac{1}{1-|y|}, & |y|<x<1\\ 0, & \text{其他}\end{cases}$

二元均匀分布的条件分布仍为均匀分布

$$P(X>\tfrac{2}{3}|Y=\tfrac{1}{2}) = \int_{2/3}^{\infty} f_{X|Y}(x|\tfrac{1}{2})dx = \int_{2/3}^{1} 2dx = \frac{2}{3}$$

定义 3.3.4 设二维随机变量 (X,Y) 具有联合密度函数

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}, \tag{3.3.7}$$

其中 $-\infty<\mu_1<+\infty, -\infty<\mu_2<+\infty, \sigma_1>0, \sigma_2>0, |\rho|<1$, 则称 (X,Y) 服从参数为 $(\mu_1,\mu_2;\sigma_1,\sigma_2;\rho)$ 的二元正态分布 (bivariate normal distribution), 记为 $(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$.

同理 $f_Y(y) = \dfrac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}, \quad -\infty<y<+\infty$

$Y\sim N(\mu_2,\sigma_2^2)$.

即二元正态分布的边际分布是正态分布，并且都不依赖于参数ρ.

二维正态随机变量 (X,Y)

X与Y相互独立的充要条件是 $\rho=0$

思考题：若随机变量（X，Y）的密度函数如下所示，问哪些密度函数对应的X与Y是相互独立的？

(1) $f(x,y)=\begin{cases}e^{-2x}, & x>0, 0<y<2,\\ 0, & \text{其他}.\end{cases}$

(2) $f(x,y)=\begin{cases}xy/2, & 0<x<y<2,\\ 0, & \text{其他}.\end{cases}$

(3) $f(x,y)=\begin{cases}x+y, & 0<x<1, 0<y<1,\\ 0, & \text{其他}.\end{cases}$

(4) $f(x,y)=\begin{cases}xy, & 0<x<2, 0<y<1,\\ 0, & \text{其他}.\end{cases}$

答：（1），（4）。

eg

定义 3.4.1 对任意两个实数集合 D_1, D_2, 若

$$P\{X\in D_1, Y\in D_2\} = P\{X\in D_1\}\cdot P\{Y\in D_2\}, \tag{3.4.1}$$

则称随机变量 X,Y 相互独立, 简称 X,Y 独立.

定义 3.4.1 也可写为: 当且仅当对任意实数 x,y, 有

$$P\{X\leqslant x, Y\leqslant y\} = P\{X\leqslant x\}\cdot P\{Y\leqslant y\} \tag{3.4.2}$$

成立, 即 $F(x,y) = F_X(x)\cdot F_Y(y)$, X,Y 相互独立.

若(X,Y)是离散型随机变量，则X,Y相互独立的条件等价于：
$P(X=x_i, Y=y_j) = P(X=x_i)P(Y=y_j)$
即 $\boxed{p_{ij}=p_{i\bullet}p_{\bullet j}}$ 对一切i,j都成立.

若(X,Y)是连续型随机变量，$f(x,y), f_X(x), f_Y(y)$分别是(X,Y)的联合密度函数和边缘密度函数，则X,Y相互独立的条件等价于：
$\boxed{f(x,y)=f_X(x)f_Y(y)}$ 几乎处处成立；
即平面上除去零“面积”集以外，处处成立.

eg.

例4.6 设甲乙两种元件的寿命X,Y相互独立服从同一分布，其密度函数为：

$$f(x)=\begin{cases}\frac{1}{2}e^{-\frac{x}{2}}, x>0,\\ 0, \quad 其它.\end{cases}$$

求甲元件寿命不大于乙元件寿命2倍的概率.

(X,Y)的联合密度函数：$f(x,y)=f_X(x)\cdot f_Y(y)=\begin{cases}\frac{1}{4}e^{-\frac{x+y}{2}}, & x>0,y>0\\ 0, & 其他\end{cases}$

$P(X\le 2Y)=\int_0^{+\infty}dx\int_{\frac{x}{2}}^{+\infty}\frac{1}{4}e^{-\frac{x+y}{2}}dy=\frac{2}{3}$

定理

设$(X_1,X_2,\cdots,X_m)$与$(Y_1,Y_2,\cdots,Y_n)$相互独立，

则$X_i\,(i=1,2,\cdots,m)$与$Y_j\,(j=1,2,\cdots,n)$相互独立；

若$h(x_1,x_2,\cdots,x_m)$和$g(y_1,y_2,\cdots,y_n)$是连续函数，

则$h(X_1,X_2,\cdots,X_m)$和$g(Y_1,Y_2,\cdots,Y_n)$相互独立。

(一) $Z=X+Y$的分布

设(X,Y)为离散型随机变量，分布律为

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

设Z的可能取值为$z_1,z_2,\ldots,z_k,\ldots$，则

$Z=X+Y$的分布律为

$$P(Z=z_k)=P(X+Y=z_k)$$
$$=\sum_{i=1}^{+\infty}P(X=x_i,Y=z_k-x_i),k=1,2,\ldots$$

或 $P(Z=z_k)=\sum_{j=1}^{+\infty}P(X=z_k-y_j,Y=y_j),k=1,2,\ldots$

特别地，当X与Y相互独立时，

$$P(Z=z_k)=\sum_{i=1}^{+\infty}P(X=x_i)P(Y=z_k-x_i),k=1,2,\ldots$$

或 $P(Z=z_k)=\sum_{j=1}^{+\infty}P(X=z_k-y_j)P(Y=y_j),k=1,2,\ldots$

eg. $X\sim P(\lambda_1)$，$Y\sim P(\lambda_2)$ 且X,Y相互独立。若$Z=X+Y$，求Z的概率分布律.

$P(X=i)=\frac{\lambda_1^i e^{-\lambda_1}}{i!}$ $i=0,1,2,\cdots$

$P(Y=j)=\frac{\lambda_2^j e^{-\lambda_2}}{j!}$ $j=0,1,2\cdots$

$P(Z=k)=\sum_{i=0}^{k}P(X=i)P(Y=k-i)=\sum_{i=0}^{k}\frac{\lambda_1^i e^{-\lambda_1}}{i!}\cdot\frac{\lambda_2^{k-i}e^{-\lambda_2}}{(k-i)!}=\frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\lambda_1^i\lambda_2^{k-i}=\frac{(\lambda_1+\lambda_2)^k e^{-(\lambda_1+\lambda_2)}}{k!}$，$k=0,1,2,\cdots$

即 $Z=X+Y\sim P(\lambda_1+\lambda_2)$

泊松分布有可加性.

设连续型随机变量(X,Y)的密度函数为 $f(x,y)$，

则$Z=X+Y$的分布函数为：

$$F_Z(z)=P(Z\le z)=\iint_{x+y\le z}f(x,y)dxdy$$
$$=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{z-y}f(x,y)dx\right]dy$$

固定z,y 令$u=x+y$

$$=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{z}f(u-y,y)du\right]dy$$
$$=\int_{-\infty}^{z}\left[\int_{-\infty}^{+\infty}f(u-y,y)dy\right]du=\int_{-\infty}^{z}f_Z(u)du$$

故Z的密度函数为$f_Z(z)=\int_{-\infty}^{+\infty}f(z-y,y)dy$

由X,Y 的对称性，$f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx.$

当X与Y相互独立时，

$$\boxed{f_Z(z)=\int_{-\infty}^{+\infty}f_X(z-y)f_Y(y)dy=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)dx}$$

称为卷积公式.

正态，二项分布也是

符合卷积公式

Corollary 1

设 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，且 X,Y 相互独立，则

$$X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2).$$

更一般地，有

$$aX+bY+c\sim N(a\mu_1+b\mu_2+c,a^2\sigma_1^2+b^2\sigma_2^2).$$

线性变换后还是正态

Example 27

设 X 和 Y 相互独立同分布，均服从 $[0,1]$ 上的均匀分布，求 $Z=X+Y$ 的密度函数.

Solution

由于 X 与 Y 相互独立，故利用卷积公式可得

$$f_Z(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)\mathrm{d}x.$$

分范围积分.

上述积分的被积函数不为零当且仅当

$$\begin{cases}0\le x\le 1,\\ 0\le z-x\le 1\end{cases}\Leftrightarrow\begin{cases}0\le x\le 1,\\ z-1\le x\le z\end{cases}$$

Solution(Cont.)

于是

$$f_Z(z)=\begin{cases}\int_0^z 1\mathrm{d}x=z, & 0<z<1;\\ \int_{z-1}^1 1\mathrm{d}x=2-z, & 1\le z<2;\\ 0, & 其他.\end{cases}$$

Example 28

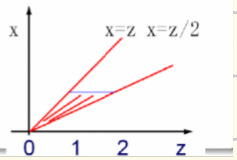
设随机变量 (X,Y) 的联合概率密度为 $f(x,y)=\begin{cases}3x, & 0<y<x<1,\\ 0, & 其他.\end{cases}$ 求 $Z=X+Y$ 的概率密度.

Solution

利用公式 $f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)\mathrm{d}x,$

(不可以用卷积公式，因为 X 与 Y 不独立) 其中

$$f(x,z-x)=\begin{cases}3x, & 0<z-x<x<1,\\ 0, & 其他.\end{cases}$$



Solution (Cont.)

于是

$$f_Z(z)=\begin{cases}\int_{\frac{z}{2}}^{z}3x\mathrm{d}x=\frac{9}{8}z^2, & 0<z<1,\\ \int_{\frac{z}{2}}^{1}3x\mathrm{d}x=\frac{3}{2}\left(1-\frac{z^2}{4}\right), & 1\le z<2,\\ 0, & 其他.\end{cases}$$

$M=\max(X,Y)$, $N=\min(X,Y)$ 的分布

设 X,Y 是两个相互独立的随机变量，它们的分布函数分别为 $F_X(x)$ 和 $F_Y(y)$. 于是 M 的分布函数为

$$\begin{aligned}F_{\max}(z)&=\mathsf{P}(M\le z)=\mathsf{P}(X\le z,Y\le z)(\text{这里无需独立})\\&=\mathsf{P}(X\le z)\mathsf{P}(Y\le z)\\&=F_X(z)F_Y(z).\end{aligned}$$

N 的分布函数为

$$\begin{aligned}F_{\min}(z)&=\mathsf{P}(N\le z)=1-\mathsf{P}(N>z)\\&=1-\mathsf{P}(X>z,Y>z)(\text{这里无需独立})\\&=1-\mathsf{P}(X>z)\mathsf{P}(Y>z)\\&=1-[1-F_X(z)][1-F_Y(z)].\end{aligned}$$

推广到 n 个相互独立的随机变量情形

设 $X_1,X_2,\dots,X_n$ 是 n 个相互独立的随机变量，它们的分布函数分别为 $F_{X_i}(x_i)$ $i=1,2,\dots,n$. 于是 $M=\max\limits_{1\le i\le n}X_i$ 的分布函数为

$$F_{\max}(z)=F_{X_1}(z)F_{X_2}(z)\cdots F_{X_n}(z).$$

$N=\min\limits_{1\le i\le n}X_i$ 的分布函数为

$$F_{\min}(z)=1-[1-F_{X_1}(z)][1-F_{X_2}(z)]\cdots[1-F_{X_n}(z)].$$

特别地，当 $X_1,X_2,\dots,X_n$ 相互独立且具有相同分布函数 $F(x)$ 时，

$$F_{\max}(z)=[F(z)]^n,\quad F_{\min}(z)=1-[1-F(z)]^n.$$

eg.

Example 30

X 与 Y 独立同分布，均服从 $U(0,1)$，求 $M=\max(X,Y)$，$N=\min(X,Y)$ 的概率密度函数.

从分布函数 ⇒ 密度函数.

由题意，X,Y 的分布函数均为 $F(x)=\begin{cases}0, & x<0,\\ x, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}$ 于是 $M=\max(X,Y)$ 的分布函数为

$$F_M(x)=[F(x)]^2=\begin{cases}0, & x<0,\\ x^2, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}\Rightarrow f_M(x)=\begin{cases}2x, & 0<x<1,\\ 0, & \text{其他}.\end{cases}$$

$N=\min(X,Y)$ 的分布函数为

$$F_N(x)=1-[1-F(x)]^2=\begin{cases}0, & x<0,\\ 1-(1-x)^2, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}$$

$$\Rightarrow\quad f_N(x)=\begin{cases}2(1-)x, & 0<x<1,\\ 0, & \text{其他}.\end{cases}$$

5. 随机变量的数字特征

•数学期望:

定义：设离散型随机变量X的分布律为
$$P(X=x_k)=p_k \quad k=1,2,\cdots$$
若级数 $\sum_{k=1}^{+\infty}|x_k|p_k<\infty$，则称级数 $\sum_{k=1}^{+\infty}x_kp_k$ 的值为X的数学期望，记为$E(X)$，即
$$\boxed{E(X)=\sum_{k=1}^{+\infty}x_kp_k}$$

定义：设连续型随机变量X的概率密度函数为$f(x)$，若积分 $\int_{-\infty}^{+\infty}|x|f(x)dx<\infty$，则称积分 $\int_{-\infty}^{+\infty}xf(x)dx$ 的值为X的数学期望，记为$E(X)$，即
$$\boxed{E(X)=\int_{-\infty}^{+\infty}xf(x)dx}$$
数学期望简称期望，又称均值。

例1.3 设随机变量X的概率密度函数为
$$f(x)=\frac{1}{\pi(1+x^2)},-\infty<x<\infty,$$
证明X不存在数学期望.

$\int_{-\infty}^{+\infty}|x|f(x)dx=\int_{-\infty}^{+\infty}|x|\frac{1}{\pi(1+x^2)}dx=\int_0^{\infty}\frac{2x}{\pi(1+x^2)}dx=\frac{1}{\pi}\ln(1+x^2)\Big|_0^{\infty}=\infty$ 不存在数学期望.

两点分布: $E(X)=p$.　　二项分布: $E(X)=np$.　　泊松分布: $E(X)=\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}e^{-\lambda}=\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}=\lambda$

几何分布:
$$P(X=k)=p(1-p)^{k-1}\Rightarrow E(X)=\sum_{k=1}^{\infty}kp(1-p)^{k-1}=p\sum_{k=1}^{\infty}\left((1-p)^k\right)'=p\cdot\left(\frac{x}{1-x}\right)'\Big|_{x=1-p}=p\cdot\frac{1}{p^2}=\frac{1}{p}$$

均匀分布 $U(a,b)$.
$$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & 其他\end{cases}\qquad E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_a^b\frac{x}{b-a}dx=\frac{a+b}{2}$$

指数分布 $E(\lambda)$　　正态分布 $N(\mu,\sigma^2)$.　$E(X)=\mu$
$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}\quad 则\ E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^{+\infty}x\lambda e^{-\lambda x}dx=\frac{1}{\lambda}$$

eg.

设 X 与 Y 独立同分布，均服从 $E(\lambda),\lambda>0$. 令 $N=\min(X,Y),M=\max(X,Y)$，求 $\mathsf{E}(N),\mathsf{E}(M)$.

由题意可知 X 与 Y 的分布函数均为 $F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x>0,\\ 0, & x\le 0.\end{cases}$ 故 N 的分布函数为 $F_N(x)=1-[1-F(x)]^2$，因此 N 的概率密度函数为
$$f_N(x)=\begin{cases}2\lambda\mathrm{e}^{-2\lambda x}, & x>0,\\ 0, & x\le 0.\end{cases}$$
即 $N\sim E(2\lambda)$，所以 $\mathsf{E}(N)=\dfrac{1}{2\lambda}$.

M 的分布函数为 $F_M(x)=[F(x)]^2$，因此 M 的概率密度函数为
$$f_M(x)=\begin{cases}2\lambda\mathrm{e}^{-\lambda x}-2\lambda\mathrm{e}^{-2\lambda x}, & x>0,\\ 0, & x\le 0.\end{cases}$$
所以
$$\mathsf{E}(M)=\int_{-\infty}^{+\infty}xf_M(x)\mathrm{d}x=2\int_0^{+\infty}\lambda x\mathrm{e}^{-\lambda x}\mathrm{d}x-\int_0^{+\infty}2\lambda x\mathrm{e}^{-2\lambda x}\mathrm{d}x=\frac{2}{\lambda}-\frac{1}{2\lambda}=\frac{3}{2\lambda}.$$

Theorem 1

设 Y 是随机变量 X 的函数: $Y=g(X)$，其中 $g(\cdot)$ 是连续函数.

- *X 是离散型随机变量，它的分布律为 $P(X=x_k)=p_k,k=1,2,\ldots$. 若 $\sum_{k=1}^{+\infty}g(x_k)p_k$ 绝对收敛，则有*
$$E(Y)=E[g(X)]=\sum_{k=1}^{+\infty}g(x_k)p_k.$$
- *X 是连续型随机变量，它的概率密度为 $f(x)$. 若 $\int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x$ 绝对收敛，则有*
$$E(Y)=E[g(X)]=\int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x.$$

Theorem 2

设 Z 是随机变量 X, Y 的函数: $Z = h(X, Y)$, 其中 $h(\cdot,\cdot)$ 是连续函数.

- 若二维离散型随机变量 (X, Y) 的分布律为 $P(X=x_i, Y=y_j)=p_{ij}, i,j=1,2,\ldots$, 则有

$$E(Z) = E[h(X, Y)] = \sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty} h(x_i, y_j)p_{ij},$$

这里设上式右边的级数绝对收敛.

Theorem 2 (Cont.)

- 若二维连续型随机变量 (X, Y) 的概率密度为 $f(x, y)$, 则有

$$\mathrm{E}(Z) = \mathrm{E}[h(X, Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x, y)f(x, y)\mathrm{d}x\mathrm{d}y,$$

这里设上式右边的积分绝对收敛.

特别地,

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x, y)\mathrm{d}x\mathrm{d}y, \quad \mathrm{E}(Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x, y)\mathrm{d}x\mathrm{d}y.$$

Example 20

设随机变量 (X, Y) 的概率密度为 $f(x, y) = \begin{cases} \dfrac{3}{2x^3y^2}, & \dfrac{1}{x} < y < x, x > 1, \\ 0, & \text{其他}. \end{cases}$ 求数学期望 $\mathrm{E}(Y)$ 和 $\mathrm{E}\left(\dfrac{1}{XY}\right)$.

$$E(Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y f(x,y)\,dx\,dy = \int_1^{\infty}\left[\int_{\frac{1}{x}}^{x} \frac{3}{2x^3y}\,dy\right]dx = \int_1^{+\infty}\frac{3\ln x}{x^3}\,dx = \frac{3}{4}$$

$$E\left(\frac{1}{XY}\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{xy} f(x,y)\,dx\,dy = \int_1^{+\infty}\left[\int_{\frac{1}{x}}^{x}\frac{3}{2x^4y^3}\,dy\right]dx = \int_1^{+\infty}\frac{3}{2x^4}\left(-\frac{1}{2y^2}\right)\Big|_{\frac{1}{x}}^{x}dx = \frac{3}{4}\int_1^{+\infty}\left(-\frac{1}{x^6}+\frac{1}{x^2}\right)dx = \frac{3}{5}$$

Example 21

某商店经销某种商品, 每周进货量 X 与需求量 Y 是相互独立的随机变量, 且都服从区间 $[10, 20]$ 上的均匀分布. 商店每售出一单位商品可获利 1000 元; 若需求量超过进货量, 商店可从他处调剂供应, 这时每单位商品可获利 500 元. 试计算出此商店经销该种商品每周所获得利润的数学期望.

Solution

设 Z 表示此商店经销该种商品每周所获得利润, 则

$$Z = g(X, Y) = \begin{cases} 1000Y, & Y \le X, \\ 1000X + 500(Y - X) = 500(X + Y), & Y > X. \end{cases}$$

又 X 和 Y 相互独立, 因此 (X, Y) 的联合概率密度函数为

$$f(x, y) = f_X(x)f_Y(y) = \begin{cases} 1/100, & 10 \le x, y \le 20, \\ 0, & \text{其他}. \end{cases}$$

Solution (Cont.)

于是

$$\begin{aligned}
\mathrm{E}(Z) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x, y)f(x, y)\mathrm{d}x\mathrm{d}y \\
&= \int_{10}^{20}\left(\int_{10}^{x} 10y\mathrm{d}y\right)\mathrm{d}x + \int_{10}^{20}\left(\int_{x}^{20} 5(x+y)\mathrm{d}y\right)\mathrm{d}x \\
&= \int_{10}^{20}\left[5x^2 - 500 + 5x(20 - x) + \frac{5}{2}(20^2 - x^2)\right]\mathrm{d}x \\
&= \int_{10}^{20}\left(-\frac{5}{2}x^2 + 100x + 500\right)\mathrm{d}x = \frac{42500}{3}(\text{元}) \approx 14166.67(\text{元}).
\end{aligned}$$

性质：

(1) 设 C 是常数，则有 $\mathsf{E}(C)=C$.

Proof

因为 C 为常数，随机变量 X 满足 $\mathsf{P}(X=C)=1$，则 $\mathsf{E}(X)=\mathsf{E}(C)=1\times C=C$.

(2) 设 X 是一个随机变量，C 是常数，则有 $\mathsf{E}(CX)=C\cdot\mathsf{E}(X)$.

Proof (以连续型随机变量为例)

$$\mathsf{E}(CX)=\int_{-\infty}^{+\infty}Cxf(x)\mathrm{d}x=C\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x=C\cdot\mathsf{E}(X).$$

(3) 设 X,Y 是两个随机变量，则有 $\mathsf{E}(X+Y)=\mathsf{E}(X)+\mathsf{E}(Y)$.

Proof (以连续型随机变量为例)

$$\begin{aligned}\mathsf{E}(X+Y)&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)\mathrm{d}x\mathrm{d}y+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)\mathrm{d}x\mathrm{d}y=\mathsf{E}(X)+\mathsf{E}(Y)\end{aligned}$$

综合上面三条可以推广得到任意有限个随机变量线性组合的情形.

Proposition 1

若 n 个随机变量 $X_1,X_2,\dots,X_n(n\geq 1)$ 的数学期望都存在，则对任意 $n+1$ 个实数 $c_0,c_1,c_2,\dots,c_n$，随机变量 $c_0+\sum_{i=1}^{n}c_iX_i$ 的数学期望也存在，且

$$E\left(c_0+\sum_{i=1}^{n}c_iX_i\right)=c_0+\sum_{i=1}^{n}c_iE(X_i).$$

(4) 设 X,Y 是两个相互独立的随机变量，则有 $\mathsf{E}(XY)=\mathsf{E}(X)\mathsf{E}(Y)$.

Proof (以连续型随机变量为例)

$$\begin{aligned}\mathsf{E}(XY)&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{+\infty}xf_X(x)\mathrm{d}x\int_{-\infty}^{+\infty}yf_Y(y)\mathrm{d}y=\mathsf{E}(X)\mathsf{E}(Y)\end{aligned}$$

随机变量 X 的期望 (均值) 为 $\mathsf{E}(X)$

随机变量 X 对于期望 (均值)$\mathsf{E}(X)$ 的离差：$X-\mathsf{E}(X)$ (仍然随机)

随机变量 X 对于期望的平均离差：$\mathsf{E}(X-\mathsf{E}(X))=0$
(因为出现了正负离差的抵消，故用此值衡量偏离程度无意义)

常常用$\mathsf{E}[X-\mathsf{E}(X)]^2$来衡量随机变量的波动性.
(这一数字特征称为方差)

Definition 4 (方差)

设 X 是一个随机变量，若 $\mathsf{E}\{[X-\mathsf{E}(X)]^2\}$ 存在，则称其为 X 的方差 (variance)，记为 $\mathsf{Var}(X)$ 或 $D(X)$，即

$$\mathsf{Var}(X)=D(X)=\mathsf{E}\{[X-\mathsf{E}(X)]^2\}.$$

将 $\sqrt{\mathsf{Var}(X)}$ 记为 $\sigma(X)$ 或 $SD(X)$，称为 X 的标准差 (standard deviation) 或均方差，它是与随机变量 X 具有相同量纲的量.

方差 $\mathsf{Var}(X)$ 和标准差 $SD(X)$ 均可刻画 X 取值的分散程度，是衡量 X 取值分散程度的常用标准.
若 X 的取值比较集中，则 $\mathsf{Var}(X)(SD(X))$ 较小；若 X 的取值比较分散，则 $\mathsf{Var}(X)(SD(X))$ 较大.

方差的计算可以看成随机变量的函数 $g(X)$ 的数学期望，其中函数

$$g(x)=[x-\mathsf{E}(X)]^2.$$

因此由定理1，可得

- 对于**离散型**随机变量 X，其分布律为 $\mathsf{P}(X=x_i)=p_i,i=1,2,\dots$，当其方差存在时，有
$$\mathsf{Var}(X)=\sum_{i=1}^{+\infty}[x_i-\mathsf{E}(X)]^2p_i.$$
- 对于**连续型**随机变量 X，其概率密度函数为 $f(x)$，当其方差存在时，有
$$\mathsf{Var}(X)=\int_{-\infty}^{+\infty}[x-\mathsf{E}(X)]^2f(x)\mathrm{d}x.$$

在实际操作中，我们常常采用下面的公式来计算方差.

Theorem 4

若随机变量 X 的方差存在，则

$$\mathsf{Var}(X)=\mathsf{E}(X^2)-[\mathsf{E}(X)]^2.\quad\text{——方差的常用计算公式}$$

Proof

由方差的定义和期望的性质，可得

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

2E(X)为定值

Example 31

设随机变量 X 具有数学期望 $\mathsf{E}(X)=\mu$，方差 $\mathsf{Var}(X)=\sigma^2\neq 0$，记 $X^*=\dfrac{X-\mu}{\sigma}$ (称为 X 的标准化变量). 证明：$\mathsf{E}(X^*)=0$，$\mathsf{Var}(X^*)=1$.

Proof

由期望的性质，可得

$$\mathsf{E}(X^*)=\mathsf{E}\left(\frac{X-\mu}{\sigma}\right)=\frac{1}{\sigma}\mathsf{E}(X-\mu)=\frac{1}{\sigma}(\mathsf{E}(X)-\mu)=0.$$

由方差的计算公式和期望的性质，可得

$$\mathsf{Var}(X^*)=\mathsf{E}(X^{*2})-[\mathsf{E}(X^*)]^2=\mathsf{E}\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]-0=\frac{1}{\sigma^2}\mathsf{E}[(X-\mu)^2]=\frac{\sigma^2}{\sigma^2}=1.$$

Example 32

设随机变量 X 服从 $0-1(p)$ 分布 $(0<p<1)$，其分布律为 $\mathsf{P}(X=0)=1-p,\mathsf{P}(X=1)=p$，求方差 $\mathsf{Var}(X)$.

Solution

之前已知期望 $\mathsf{E}(X)=p$.

(法一)

$$\mathsf{Var}(X)=\mathsf{E}\{[X-\mathsf{E}(X)]^2\}=\mathsf{P}(X=0)\cdot(0-p)^2+\mathsf{P}(X=1)\cdot(1-p)^2=p(1-p).$$

(法二) 由于 $\mathsf{E}(X^2)=0^2\cdot(1-p)+1^2\cdot p=p$，故利用方差的计算公式，可得

$$\mathsf{Var}(X)=\mathsf{E}(X^2)-[\mathsf{E}(X)]^2=p-p^2=p(1-p).$$

Example 33

设随机变量 X 服从泊松分布 $P(\lambda)(\lambda>0)$，求方差 $\mathsf{Var}(X)$.

Solution

泊松分布 $P(\lambda)$ 的分布律为：$\mathsf{P}(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$，$k=0,1,\dots$. 之前也已得到期望 $\mathsf{E}(X)=\lambda$，利用方差的计算公式，先计算 $\mathsf{E}(X^2)$.

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

$E(X^2)=Var(X)+(E(X))^2$

泊松分布的均值和方差相等，都等于参数 λ.

从泊松分布方差的计算过程中不难推广得到如下结论.

Proposition 3

设随机变量 X 服从泊松分布 $P(\lambda)(\lambda>0)$，则

$$\mathsf{E}[X(X-1)\cdots(X-m)]=\lambda^{m+1},\quad m\in\mathbb{N}.$$

Proof

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

Example 35

设随机变量 X 具有均匀分布 $U(a,b)(a<b)$，求方差 $\mathsf{Var}(X)$.

Solution

之前已得期望 $\mathsf{E}(X)=\dfrac{a+b}{2}$，而

$$\mathsf{E}(X^2)=\int_{-\infty}^{+\infty}x^2f(x)\mathrm{d}x=\int_a^b\frac{x^2}{b-a}\mathrm{d}x=\frac{b^3-a^3}{3(b-a)}=\frac{a^2+b^2+ab}{3}.$$

$$\Rightarrow\ \mathsf{Var}(X)=\mathsf{E}(X^2)-[\mathsf{E}(X)]^2=\frac{a^2+b^2+ab}{3}-\frac{a^2+b^2+2ab}{4}=\frac{(b-a)^2}{12}.$$

Example 36

设随机变量 X 服从指数分布 $E(\lambda)(\lambda>0)$，求方差 $\mathsf{Var}(X)$.

Solution

之前已得期望 $\mathsf{E}(X)=1/\lambda$，而

$$\mathsf{E}(X^2)=\int_{-\infty}^{+\infty}x^2f(x)\mathrm{d}x=\int_0^{+\infty}x^2\lambda\mathrm{e}^{-\lambda x}\mathrm{d}x=-x^2\mathrm{e}^{-\lambda x}\Big|_0^{+\infty}+\int_0^{+\infty}2x\mathrm{e}^{-\lambda x}\mathrm{d}x=\frac{2}{\lambda^2}.$$

$$\Rightarrow\ \mathsf{Var}(X)=\mathsf{E}(X^2)-[\mathsf{E}(X)]^2=\frac{1}{\lambda^2}.$$

指数分布的方差是其期望的平方.

方差的性质

(1) 设 C 是常数，则有 $\mathsf{Var}(C)=0$.

Proof

由定义，$\mathsf{Var}(C)=\mathsf{E}\{[C-\mathsf{E}(C)]^2\}=\mathsf{E}[(C-C)^2]=0$.

(2) 设 X 是一个随机变量，C 是常数，则有 $\mathsf{Var}(CX)=C^2\mathsf{Var}(X)$.

Proof

由方差的计算公式和期望的性质，可得

$$\begin{aligned}\mathsf{Var}(CX)&=\mathsf{E}(CX)^2-[\mathsf{E}(CX)]^2\\&=C^2\mathsf{E}(X^2)-C^2[\mathsf{E}(X)]^2\\&=C^2\{\mathsf{E}(X^2)-[\mathsf{E}(X)]^2\}=C^2\mathsf{Var}(X).\end{aligned}$$

(3) 设 X,Y 是两个随机变量，则有

$$\mathsf{Var}(X+Y)=\mathsf{Var}(X)+\mathsf{Var}(Y)+\underline{2\mathsf{E}\{[X-\mathsf{E}(X)][Y-\mathsf{E}(Y)]\}}.$$

特别地，若 X,Y 相互独立，则有 $\mathsf{Var}(X+Y)=\mathsf{Var}(X)+\mathsf{Var}(Y)$.

Proof

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

特别地，当 X,Y 相互独立时，$X-\mathsf{E}(X)$ 与 $Y-\mathsf{E}(Y)$ 相互独立，此时第三项

$$\mathsf{E}\{[X-\mathsf{E}(X)][Y-\mathsf{E}(Y)]\}=\mathsf{E}[X-\mathsf{E}(X)]\mathsf{E}[Y-\mathsf{E}(Y)]=0.$$

综合上面三条可以推广得到任意有限个**独立**随机变量线性组合的情形.

Proposition 4

若 n 个随机变量 $X_1,X_2,\ldots,X_n(n\geq 2)$ 相互独立，且方差都存在. 对任意 $n+1$ 个有限实数 $c_0,c_1,c_2,\ldots,c_n$，则 $c_0+\sum_{i=1}^n c_iX_i$ 的方差也存在，且

$$\mathsf{Var}\left(c_0+\sum_{i=1}^n c_iX_i\right)=\sum_{i=1}^n c_i^2\mathsf{Var}(X_i).$$ 注：没有 c_0

(4) $\mathsf{Var}(X)=0\ \Leftrightarrow\ \mathsf{P}(X=C)=1$ 且 $C=\mathsf{E}(X)$.

必要性显然，但充分性的证明需要用到 Chebyshev 不等式 (Chapter 5)，此处从略.

变异系数 $C_v(X)=\dfrac{\sqrt{\mathsf{Var}(X)}}{\mathsf{E}(X)}$

Table: 几种常见分布的均值与方差

分布	分布律或密度函数	期望	方差
0-1 分布 $0-1(p)$	$\mathsf{P}(X=k)=p^k(1-p)^{1-k},k=0,1$	p	$p(1-p)$
二项分布 $B(n,p)$	$\mathsf{P}(X=k)=\mathsf{C}_n^kp^k(1-p)^{n-k},k=0,1,\ldots,n$	np	$np(1-p)$
泊松分布 $P(\lambda)$	$\mathsf{P}(X=k)=\lambda^k\mathrm{e}^{-\lambda}/k!,k=0,1,\ldots$	λ	λ
均匀分布 $U(a,b)$	$f(x)=\begin{cases}1/(b-a), & a<x<b,\\ 0, & \text{其他}\end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
指数分布 $E(\lambda)$	$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x>0,\\ 0, & x\leq 0\end{cases}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
正态分布 $N(\mu,\sigma^2)$	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$	μ	σ^2

4.3 协方差与相关系数

定义 $E\{[X-E(X)][Y-E(Y)]\}$

称为随机变量X与Y的协方差，记为：$Cov(X,Y)$，即

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

- **协方差的计算公式：**

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

- **方差性质的补充：**

$$Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$$

协方差的性质：

1. $Cov(X,Y)=Cov(Y,X)$;
2. $Cov(X,X)=Var(X)$;
3. $Cov(aX,bY)=ab\cdot Cov(X,Y)$,其中$a,b$为两个实数;
4. $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$;
5. 当$Var(X)Var(Y)\neq 0$时，有

$$(Cov(X,Y))^2\le Var(X)Var(Y),$$

其中等号当且仅当X与Y之间有严格的线性关系，即存在常数a,b，使$P(Y=a+bX)=1$.

定义 $\boxed{\rho_{XY}=\dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}}$

称为X与Y的相关系数. 它是无量纲的量.

$$\rho_{XY}=Cov\left(\underbrace{\frac{X-E(X)}{\sqrt{Var(X)}}}_{X^*},\underbrace{\frac{Y-E(Y)}{\sqrt{Var(Y)}}}_{Y^*}\right)$$

X,Y标准化变量

相关系数的性质：

1. $|\rho_{XY}|\le 1$

$[Cov(X,Y)]^2\le Var(X)\,Var(Y)$

2. $|\rho_{XY}|=1\Leftrightarrow$ 存在常数a,b，使$P(Y=a+bX)=1$

特别的，$\rho_{XY}=1$时，$b>0$；$\rho_{XY}=-1$时，$b<0$

定义：$\boxed{\rho_{XY}=0,}$ 称X与Y不相关或零相关.

随机变量X与Y不相关，即$\rho_{XY}=0$的等价条件有：

1. $Cov(X,Y)=0$
2. $E(XY)=E(X)E(Y)$
3. $Var(X+Y)=Var(X)+Var(Y)$

从而可知，当X与Y相互独立$\Rightarrow$ X与Y一定不相关

反之，若X与Y不相关，X与Y却不一定相互独立

即二元正态变量(X,Y)的概率密度中的参数ρ就是X,Y的相关系数，因而二元正态变量的分布完全可由X,Y各自的均值、方差以及它们的相关系数所确定。

若(X,Y)服从二元正态分布，那么

X和Y相互独立$\Leftrightarrow\rho=0$

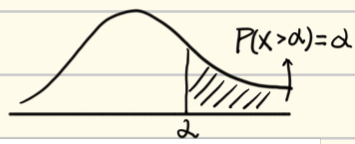
现在知道，$\rho_{XY}=\rho$，从而知：

对于二元正态变量(X,Y)来说，

X和Y不相关$\Leftrightarrow$ X与Y相互独立

独立$\Rightarrow$不相关

不相关$\not\Rightarrow$独立.



4.4 其它数字特征

定义：设X和Y是随机变量

若$E(X^k)$ $k=1,2,\cdots$存在，

则称它为X的k阶(原点)矩；

若$E\{[X-E(X)]^k\}$ $k=1,2,\cdots$存在，

则称它为X的k阶中心矩；

若$E\{X^kY^l\}$存在 $k,l=1,2,\cdots$存在，

则称它为X和Y的$k+l$阶混合(原点)矩；

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

则称它为X,Y的$k+l$阶混合中心矩；

定义：X为连续型随机变量，其分布函数和概率密度函数分别为$F(x)$和$f(x)$,称满足条件

$$P\{X>x_\alpha\}=1-F(x_\alpha)=\int_{x_\alpha}^{+\infty}f(x)dx=\alpha$$

的实数x_α为随机变量X(或此分布）的上（侧）α分位数.

定义：设n元随机变量$\mathbf{X}=(X_1,X_2,\cdots X_n)^T$，若其每一分量的数学期望都存在，则称

$$E(\mathbf{X})=(E(X_1),E(X_2),\cdots E(X_n))^T$$

为n元随机变量$\mathbf{X}$的数学期望(向量).

定义：设二元随机变量(X_1,X_2)的四个二阶中心矩存在，

将它们排成矩阵：$\begin{pmatrix}Var(X_1) & Cov(X_1,X_2)\\ Cov(X_2,X_1) & Var(X_2)\end{pmatrix}$,

称为(X_1,X_2)的协方差矩阵.

n元正态变量具有以下四条重要性质：

1. n元正态变量$(X_1,X_2,\cdots,X_n)^T$中的任意子向量$(X_{i_1},X_{i_2},\cdots,X_{i_k})^T$ $(1\le k\le n)$也服从k元正态分布. 特别地，每一个分量$X_i, i=1,2,\cdots n$都是正态变量；反之，若$X_1,X_2,\cdots,X_n$都是正态变量，且相互独立，则$(X_1,X_2,\cdots,X_n)$是n元正态变量；

2. n元随机变量$(X_1,X_2,\cdots,X_n)$服从n元正态分布 $\Leftrightarrow$ $X_1,X_2,\cdots,X_n$的任意线性组合$l_1X_1+l_2X_2+\cdots+l_nX_n$服从一元正态分布 其中$l_1,l_2,\cdots,l_n$不全为零

3. 若$(X_1,X_2,\cdots,X_n)$服从n元正态分布，设$Y_1,Y_2,\cdots,Y_k$是$X_j(j=1,2,\cdots,n)$的线性函数，则$(Y_1,Y_2,\cdots,Y_k)$也服从多元正态分布；
这一性质称为正态变量的线性变换不变性

4. 设$(X_1,X_2,\cdots,X_n)$服从n元正态分布，
则$X_1,X_2,\cdots,X_n$相互独立 $\Leftrightarrow$ $X_1,X_2,\cdots,X_n$两两不相关
$\Leftrightarrow$ 协方差矩阵为对角矩阵.

Definition 15 (协方差矩阵)

设 n 维随机变量 $(X_1,X_2,\ldots,X_n)$ 的任意二者协方差 $Cov(X_i,X_j)$ 存在, $i,j=1,2,\ldots,n$, 称矩阵

$$(Cov(X_i,X_j))_{n\times n}=\begin{pmatrix} Var(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & Var(X_2) & \cdots & Cov(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n,X_1) & Cov(X_n,X_2) & \cdots & Var(X_n) \end{pmatrix}$$

为 n 维随机变量 $(X_1,X_2,\ldots,X_n)$ 的协方差矩阵.

协方差矩阵是一个对称的非负定矩阵.

Example 45

设二元随机变量 (X,Y) 服从二元正态分布, $X\sim N(0,1)$, $Y\sim N(1,4)$, X 与 Y 的相关系数 $\rho=-\frac{1}{2}$. 求:

1. $Var(2X-Y)$;
2. $P(2X>Y)$;
3. $(Z_1,Z_2)=(X+Y,X-Y)$ 的分布.

(1)

由题: $E(X)=0$ $Var(X)=1$; $E(Y)=1$, $Var(Y)=4$ $\quad Var(2X-Y)=Var(2X)+Var(-Y)+2Cov(2X,-Y)$

$Cov(X,Y)=\rho_{XY}\sqrt{Var(X)Var(Y)}=-1$ $\quad =4Var(X)+Var(Y)-4Cov(X,Y)=4+4+4=12$

(2). $P(2X>Y)=P(2X-Y>0)$ $\quad Z=2X-Y$ $\quad E(Z)=2E(X)-E(Y)=-1$ $\quad Var(Z)=Var(2X-Y)=12$

$\therefore Z\sim N(-1,12)$. $\quad P(Z>0)=1-\Phi(\frac{x-\mu}{\sigma})=1-\Phi(\frac{0-(-1)}{\sqrt{12}})=1-\Phi(\frac{1}{\sqrt{12}})=1-\Phi(\frac{1}{2\sqrt{3}})$

(3). 由正态变量的线性变换不变性，(Z_1,Z_2)也符合二维正态分布

$E(Z_1)=E(X+Y)=1$ $\quad E(Z_2)=-1$

$Var(Z_1)=1+4+2Cov(X,Y)=3$ $\quad Var(Z_2)=1+4-2Cov(X,Y)=7$.

$Cov(Z_1,Z_2)=Cov(X+Y,X-Y)=Var(X)-Var(Y)=-3$ $\quad \rho_{Z_1Z_2}=\frac{Cov(Z_1,Z_2)}{\sqrt{Var(Z_1)}\sqrt{Var(Z_2)}}=\frac{-3}{\sqrt{3}\cdot\sqrt{7}}=-\sqrt{\frac{3}{7}}$

$\therefore (Z_1,Z_2)\sim N(1,-1;3,7;-\sqrt{\frac{3}{7}})$

·大数定律

(一). 依概率收敛

定义 5.1.1　设 $\{Y_n, n \geqslant 1\}$ 为一随机变量序列, c 为一常数. 若对任意的 $\varepsilon > 0$, 都有

$$\lim_{n\to+\infty} P\{|Y_n - c| \geqslant \varepsilon\} = 0 \tag{5.1.1}$$

成立, 则称 $\{Y_n, n \geqslant 1\}$ 依概率收敛 (convergence in probability) 于 c, 记为 $Y_n \xrightarrow{P} c$, $n \to +\infty$.

显然, (5.1.1) 可以等价表示为

$$\lim_{n\to+\infty} P\{|Y_n - c| < \varepsilon\} = 1. \tag{5.1.2}$$

下面不加证明地给出依概率收敛的一个重要性质:

设 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, $n \to +\infty$, 其中 a, b 为两个常数, 若二元函数 $g(x, y)$ 在点 (a, b) 处连续, 则有

$$g(X_n, Y_n) \xrightarrow{P} g(a, b), \quad n \to +\infty. \tag{5.1.3}$$

(二). Markov 不等式 & Chebyshev 不等式

· Markov 不等式: $E(Y^k)$ 存在.

定理 5.1.1　(马尔可夫 (Markov) 不等式) 若随机变量 Y 的 k 阶 (原点) 矩存在 $(k \geqslant 1)$, 则对任意的 $\varepsilon > 0$, 有

$$P\{|Y| \geqslant \varepsilon\} \leqslant \frac{E(|Y|^k)}{\varepsilon^k} \left(\text{或写为 } P\{|Y| < \varepsilon\} \geqslant 1 - \frac{E(|Y|^k)}{\varepsilon^k}\right). \tag{5.1.4}$$

特别地, 当 Y 为取非负值的随机变量且它的 k 阶矩存在时, 则有

$$P\{Y \geqslant \varepsilon\} \leqslant \frac{E(Y^k)}{\varepsilon^k}. \tag{5.1.5}$$

例 5.1.1　某城市一周内发生交通事故的次数记为随机变量 X, 显然 $P\{X \geqslant 0\} = 1$. 已知 $E(X) = 75$, 求一周内发生事故的次数不少于 100 的概率上界.

$P\{X \geq 0\} = 1$ 非负 X.

$\therefore P\{X \geq 100\} \leq \frac{E(X)}{100} = 75\%$

· Chebyshev 不等式. $E(X) = \mu$　$Var(X) = \sigma^2$ 存在

定理 5.1.2　(切比雪夫 (Chebyshev) 不等式) 设随机变量 X 的数学期望和方差存在, 分别记为 μ, σ^2, 则对任意的 $\varepsilon > 0$, 有

$$P\{|X - \mu| \geqslant \varepsilon\} \leqslant \frac{\sigma^2}{\varepsilon^2} \left(\text{或写为 } P\{|X - \mu| < \varepsilon\} \geqslant 1 - \frac{\sigma^2}{\varepsilon^2}\right). \tag{5.1.6}$$

从 (5.1.6) 可以看出, X 的方差越小, 对于同一个 $\varepsilon > 0$, $P\{|X - \mu| \geqslant \varepsilon\}$ 的上界就越小, 即 X 落入区域 $(-\infty, \mu - \varepsilon] \cup [\mu + \varepsilon, +\infty)$ 的可能性就越小, 落入 $(\mu - \varepsilon, \mu + \varepsilon)$ 这个 μ 附近区域的可能性就越大. 这也进一步说明了方差这个数字特征的确刻画了 X 的概率分布偏离其中心位置 (数学期望) 的离散程度.

例 5.1.2　证明: 设随机变量 X 的方差存在, 若 $\mathrm{Var}(X) = 0$, 则 $P\{X = c\} = 1$, 其中 $c = E(X)$.

$Var(X) = \sigma^2$ 存在　$E(X)$ 存在　存在 $n \in \mathbb{N}^+$, $|X - E(X)| \geq \frac{1}{n}$　$\{|X - E(X)| > 0\} = \bigcup_{n=1}^{+\infty} \{|X - E(X)| \geq \frac{1}{n}\}$

$\therefore 0 \leq P\{|X - E(X)| > 0\} = P\{\bigcup_{n=1}^{+\infty} (|X - E(X)| \geq \frac{1}{n})\} \leq \sum_{n=1}^{\infty} P\{|X - E(X)| \geq \frac{1}{n}\} \leq \sum_{n=1}^{+\infty} \frac{Var(X)}{(1/n)^2} = 0$

$\Rightarrow P\{|X - E(X)| > 0\} = 0 \Rightarrow P\{X = E(X)\} = 1.$

(三). 大数定律

定义 5.1.2　设 $\{X_i, i \geqslant 1\}$ 为一随机变量序列, 若存在常数序列 $\{c_n, n \geqslant 1\}$, 使得对任意的 $\varepsilon > 0$, 有

$$\lim_{n\to+\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - c_n\right| \geqslant \varepsilon\right\} = 0 \left(\text{或 } \lim_{n\to+\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - c_n\right| < \varepsilon\right\} = 1\right) \tag{5.1.7}$$

成立, 即当 $n \to +\infty$ 时, 有 $\frac{1}{n}\sum_{i=1}^{n} X_i - c_n \xrightarrow{P} 0$, 则称随机变量序列 $\{X_i, i \geqslant 1\}$ 服从弱大数定律 (weak law of large numbers), 简称服从大数定律.

特别地, 当 $c_n = c (n = 1, 2, \cdots)$ 时, 可写为

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} c, \quad n \to +\infty.$$

· Bernoulli 大数定律

定理 5.1.3　(伯努利 (Bernoulli) 大数定律) 设 n_A 为 n 重伯努利试验中事件 A 发生的次数, $p (0 < p < 1)$ 为事件 A 在每次试验中发生的概率, 即 $P(A) = p$, 则对任意的 $\varepsilon > 0$, 有

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

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 次试验中事件 } A \text{ 发生}, \\ 0, & \text{第 } i \text{ 次试验中事件 } A \text{ 不发生}, \end{cases} \quad i = 1, 2, \cdots, n.$$

易见 $n_A = \sum_{i=1}^{n} X_i$, $X_1, X_2, \cdots, X_n$ 相互独立, 且都服从参数为 p 的 $0-1$ 分布. 从而

$$E(X_i) = p, \quad \mathrm{Var}(X_i) = p(1-p), \quad i = 1, 2, \cdots, n.$$

故 $E\left(\frac{n_A}{n}\right) = p$, $\mathrm{Var}\left(\frac{n_A}{n}\right) = \frac{p(1-p)}{n}$. 利用切比雪夫不等式, 可得

$$\begin{aligned} P\left\{\left|\frac{n_A}{n} - p\right| \geqslant \varepsilon\right\} &= P\left\{\left|\frac{n_A}{n} - E\left(\frac{n_A}{n}\right)\right| \geqslant \varepsilon\right\} \\ &\leqslant \frac{\mathrm{Var}\left(\frac{n_A}{n}\right)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \to 0, \quad n \to +\infty. \end{aligned}$$

再结合 $P\left\{\left|\frac{n_A}{n} - p\right| \geqslant \varepsilon\right\} \geqslant 0$, 定理得证.

定理 5.1.4 (辛钦 (Khinchin) 大数定律) 设 $\{X_i, i \geqslant 1\}$ 为独立同分布的随机变量序列, 且数学期望存在, 记为 μ, 则对任意的 $\varepsilon > 0$, 有

$$\lim_{n\to+\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geqslant \varepsilon\right\} = 0, \tag{5.1.8}$$

即 $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu\ (n \to +\infty)$, 并认为此时随机变量序列 $\{X_i, i \geqslant 1\}$ 服从大数定律.

Corollary 1

设 $\{X_i, i \geq 1\}$ 为独立同分布的随机变量序列, 若 $h(x)$ 为一连续函数, 且 $E(|h(X_1)|) < +\infty$, 则对 $\forall \varepsilon > 0$, 有

$$\lim_{n\to+\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} h(X_i) - a\right| \geq \varepsilon\right\} = 0$$

成立, 其中 $a = E(h(X_1))$, 即 $\frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{P} a = E(h(X_1))$, 当 $n \to +\infty$ 时. 即随机变量序列 $\{h(X_i), i \geq 1\}$ 也服从大数定律.

Example 6

设随机变量 $X_1, \ldots, X_n, \ldots$ 相互独立同分布, $X_1 \sim U(-1, 1)$. 请问: (1) $\frac{1}{n}\sum_{k=1}^{n} X_k$, (2) $\frac{1}{n}\sum_{k=1}^{n} |X_k|$, (3) $\frac{1}{n}\sum_{k=1}^{n} X_k^2$ 分别依概率收敛吗? 如果依概率收敛, 分别收敛于什么?

Solution

注意到 $X_1, \ldots, X_n, \ldots$ 相互独立同分布, $\mathsf{E}(X_i)$ 存在;
$|X_1|, \ldots, |X_n|, \ldots$ 相互独立同分布, $\mathsf{E}(|X_i|)$ 存在;
$X_1^2, \ldots, X_n^2, \ldots$ 相互独立同分布, $\mathsf{E}(X_i^2)$ 存在.

因此利用辛钦大数定律, 可知 $\frac{1}{n}\sum_{k=1}^{n} X_k$, $\frac{1}{n}\sum_{k=1}^{n} |X_k|$, $\frac{1}{n}\sum_{k=1}^{n} X_k^2$ 均依概率收敛.

Solution (Cont.)

计算得

$$\mathsf{E}(X_1) = 0, \quad \mathsf{E}(|X_1|) = \int_{-1}^{1} \frac{|x|}{2}\mathrm{d}x = \frac{1}{2}, \quad \mathsf{E}(X_1^2) = \int_{-1}^{1} \frac{x^2}{2}\mathrm{d}x = \frac{1}{3}.$$

因此

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\mathsf{P}} 0, \quad \frac{1}{n}\sum_{k=1}^{n} |X_k| \xrightarrow{\mathsf{P}} \frac{1}{2}, \quad \frac{1}{n}\sum_{k=1}^{n} X_k^2 \xrightarrow{\mathsf{P}} \frac{1}{3},$$

当 $n \to +\infty$ 时.

Example 7

设随机变量 $X_1, \ldots, X_n, \ldots$ 相互独立同分布, $X_1 \sim U(0, 1)$, 请问: 当 $n \to +\infty$ 时, $\sqrt[n]{X_1 X_2 \cdots X_n}$ 依概率收敛吗? 如果依概率收敛, 收敛于什么?

Solution

记 $Y_n = \sqrt[n]{X_1 X_2 \cdots X_n}$, 令

$$Z_n = \ln Y_n = \frac{1}{n}(\ln X_1 + \ln X_2 + \cdots + \ln X_n).$$

由 $X_1, \ldots, X_n, \ldots$ 相互独立同分布, 可知 $\ln X_1, \ldots, \ln X_n, \ldots$ 相互独立同分布, 且 $\mathsf{E}(\ln X_1) = \int_0^1 \ln x \mathrm{d}x = -1$, 那么利用辛钦大数定律知 $Z_n \xrightarrow{\mathsf{P}} -1$, 当 $n \to +\infty$.

进一步利用依概率收敛的性质, 可知 $Y_n = \mathrm{e}^{Z_n} \xrightarrow{\mathsf{P}} \mathrm{e}^{-1}$, 当 $n \to +\infty$.

Theorem 6 (*Chebyshev 大数定律, 仅需了解)

设 $\{X_i, i \geq 1\}$ 为相互独立的随机变量序列, 若存在常数 C, 使得 $\mathsf{Var}(X_i) \leq C$, $i = 1, 2, \ldots$, 即所有的 X_i 的方差有共同的上界, 则对 $\forall \varepsilon > 0$, 有

$$\lim_{n\to+\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| \geq \varepsilon\right\} = 0$$

成立, 即随机变量序列 $\{X_i, i \geq 1\}$ 服从大数定律.

Proof

由于 $\frac{1}{n^2}\mathsf{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{Var}(X_i) \leq \frac{nC}{n^2} = \frac{C}{n} \to 0$, 当 $n \to +\infty$ 时.

因此由 Markov 大数定律知结论成立.

下面这个结论较为常用.

特别地, $\{X_i, i \geq 1\}$ 为相互独立的随机变量序列, 具有相同的期望 μ, 相同的方差 σ^2, 那么

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathsf{P}} \mu, \quad \text{当 } n \to +\infty \text{ 时}.$$

Example 8

设随机变量 $X_1, \ldots, X_n, \ldots$ 相互独立, 且它们的分布律为

$$\mathsf{P}(X_i = \sqrt{i}) = \mathsf{P}(X_i = -\sqrt{i}) = \frac{1}{2i}, \ \mathsf{P}(X_i = 0) = 1 - \frac{1}{i}, \quad i = 1, 2, \ldots.$$

试判断 $\{X_i, i \geq 1\}$ 是否服从大数定律?

Solution

由于对任意的 $i \geq 1$, 有

$$\mathsf{E}(X_i) = 0, \quad \mathsf{Var}(X_i) = \mathsf{E}(X_i^2) = 0 + (\sqrt{i})^2 \cdot \frac{1}{2i} + (-\sqrt{i})^2 \cdot \frac{1}{2i} = 1,$$

所以 $\{X_i, i \geq 1\}$ 相互独立, 期望存在并且相同, 方差存在并且相同, 由 Chebyshev 大数定律的推论知满足大数定律, 且

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathsf{P}} 0, \quad \text{当 } n \to +\infty \text{ 时}.$$

中心极限定理

Theorem 7 (Lindeberg-Lévy 中心极限定理，亦称为独立同分布的中心极限定理)

设随机变量 $X_1, X_2, \ldots, X_n, \ldots$ 相互独立且同分布，满足 $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, $i = 1, 2, \ldots$，则对于充分大的 n 有

$$\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2).$$

此时

$$P\Big(a < \sum_{i=1}^{n} X_i \le b\Big) \approx \Phi\Big(\frac{b - n\mu}{\sqrt{n}\sigma}\Big) - \Phi\Big(\frac{a - n\mu}{\sqrt{n}\sigma}\Big).$$

类似地，有 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 的近似分布是 $N\Big(\mu, \frac{\sigma^2}{n}\Big)$.

注意到：当随机变量 $X_1, X_2, \ldots, X_n, \ldots$ 相互独立同分布，满足 $\mathsf{E}(X_i) = \mu$, $\mathsf{Var}(X_i) = \sigma^2$, $i = 1, 2, \ldots$ 时，

$$\mathsf{E}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathsf{E}(X_i) = n\mu, \quad \mathsf{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathsf{Var}(X_i) = n\sigma^2.$$

$$\mathsf{E}(\overline{X}) = \mathsf{E}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}(X_i) = \frac{1}{n} \cdot n\mu = \mu,$$

$$\mathsf{Var}(\overline{X}) = \mathsf{Var}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

据 CLT，当 n 充分大时，有 $\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2)$, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N\Big(\mu, \frac{\sigma^2}{n}\Big)$

故 CLT 仅仅是分布类型上的一种近似.（"万物归一"）

$X \sim B(n,p)$ 当 n 充分大时，$X \overset{\text{近似}}{\sim} N(np, np(1-p))$

Example 10

某校 1500 名学生选修"概率论与数理统计"课程，共有 10 名教师主讲此课，假定每位学生可以随意地选择一位教师（即选择任意一位教师的可能性均为 1/10），而且学生之间的选择是相互独立的. 问：每位教师的上课教室应该设有多少座位才能保证因没有座位而使学生离开的概率小于 5%.

Solution

由于每位学生可以随意地选择一位老师，因此我们只需要考虑某个老师甲的上课教室的座位即可. 引入随机变量

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 个学生选择教师甲}, \\ 0, & \text{其他}, \end{cases} \quad i = 1, 2, \ldots, 1500,$$

则 X_i 独立同分布，$X_i \sim B(1, 1/10)$.

记 $Y = \sum_{i=1}^{1500} X_i$，则 $Y \sim B(1500, 1/10)$，由 De Moivre-Laplace 中心极限定理，

$$\frac{Y - 1500 \times 1/10}{\sqrt{1500 \times 1/10 \times 9/10}} \overset{\text{近似}}{\sim} N(0, 1).$$

设教室需要设 a 个座位，则由题意，a 需要满足

$$\Phi(1.645) = 95\% \le \mathsf{P}\{Y \le a\} \approx \Phi\Big(\frac{a - 1500 \times 1/10}{\sqrt{1500 \times 1/10 \times 9/10}}\Big) = \Phi\Big(\frac{a - 150}{\sqrt{135}}\Big),$$

即 $\frac{a - 150}{\sqrt{135}} \ge 1.645 \Rightarrow a \ge 169.11$. 故每位教师的上课教室应该至少设 170 个座位才能保证因没有座位而使学生离开的概率小于 5%.

某保险公司的老年人寿保险有 1 万人参加（这些人的健康状况相互独立），每人每年交 200 元，若老人在该年内死亡，公司付给受益人 1 万元. 设老年人死亡率为 0.017，试求保险公司在一年内这项保险亏本的概率约为多少？

Solution

设 X 为一年中投保老人的死亡数，则 $X \sim B(10000, 0.017)$. 现求 $\mathsf{P}(X > 200)$. 由 De Moivre-Laplace 中心极限定理，

$$\frac{X - 10000 \times 0.017}{\sqrt{10000 \times 0.017 \times 0.983}} \overset{\text{近似}}{\sim} N(0, 1).$$

故保险公司亏本的概率为

$$\mathsf{P}(X > 200) \approx 1 - \Phi\Big(\frac{200 - 10000 \times 0.017}{\sqrt{10000 \times 0.017 \times 0.983}}\Big) = 1 - \Phi(2.321) = 0.01.$$

Example 10 (Cont.)

求保险公司盈利 10 万元以上的概率.

Solution (Cont.)

类似地，保险公司盈利 10 万元以上的概率

$$\begin{aligned} & \mathsf{P}(10000 \times 200 - 10000X > 100000) \\ = \ & \mathsf{P}(X < 190) = \mathsf{P}(X \le 189) = \mathsf{P}(0 \le X \le 189) \\ \approx \ & \Phi\Big(\frac{189 - 10000 \times 0.017}{\sqrt{10000 \times 0.017 \times 0.983}}\Big) - \Phi\Big(\frac{0 - 10000 \times 0.017}{\sqrt{10000 \times 0.017 \times 0.983}}\Big) \\ = \ & \Phi(1.47) - \Phi(-13.15) = 0.9292 \end{aligned}$$

Example 12

一颗骰子投掷了 11520 次，出现"1 点"的次数超过了 2160 次，由此可否推断该骰子是不均匀的呢？

Solution (Cont.)

考虑用近似计算，利用 De Moivre-Laplace 中心极限定理，

$$\frac{n_A - 11520 \times 1/6}{\sqrt{11520 \times 1/6 \times 5/6}} \overset{\text{近似}}{\sim} N(0, 1).$$

所以

$$\mathsf{P}\{n_A > 2160\} \approx 1 - \Phi\Big(\frac{2160 - 11520 \times 1/6}{\sqrt{11520 \times 1/6 \times 5/6}}\Big) = 1 - \Phi(6) = 0.$$

因此可以认为该骰子是不均匀的.

Example 14

设随机变量 $X_1, X_2, \ldots, X_{50}$ 相互独立同分布，$X_1 \sim U(-1, 1)$，分别求 (1) $\frac{1}{50}\sum_{k=1}^{50} X_k$, (2) $\frac{1}{50}\sum_{k=1}^{50} |X_k|$, (3) $\frac{1}{50}\sum_{k=1}^{50} X_k^2$ 的近似分布.

Solution

由中心极限定理，$\frac{1}{50}\sum_{k=1}^{50} X_k$, $\frac{1}{50}\sum_{k=1}^{50} |X_k|$, $\frac{1}{50}\sum_{k=1}^{50} X_k^2$ 均近似服从正态分布.

因为，$\mathsf{E}(X_1) = 0$, $\mathsf{Var}(X_1) = \frac{1}{3}$，所以 $\frac{1}{50}\sum_{k=1}^{50} X_k \overset{\text{近似}}{\sim} N\Big(0, \frac{1}{150}\Big)$.

Solution (Cont.)

因为

$$\mathsf{E}(|X_1|) = \frac{1}{2}, \quad \mathsf{Var}(|X_1|) = \mathsf{E}(X_1^2) - [\mathsf{E}(|X_1|)]^2 = \frac{1}{12},$$

所以 $\frac{1}{50}\sum_{k=1}^{50} |X_k| \overset{\text{近似}}{\sim} N\Big(\frac{1}{2}, \frac{1}{600}\Big)$.

因为

$$\mathsf{E}(X_1^2) = \frac{1}{3}, \quad \mathsf{Var}(X_1^2) = \mathsf{E}(X_1^4) - [\mathsf{E}(X_1^2)]^2 = \frac{1}{5} - \frac{1}{9} = \frac{4}{45},$$

所以 $\frac{1}{50}\sum_{k=1}^{50} X_k^2 \overset{\text{近似}}{\sim} N\Big(\frac{1}{3}, \frac{2}{1125}\Big)$.

样本 $\tilde{X}=(X_1, X_2, \cdots, X_n)$

总体中每个个体，个个机会访问，成为样本中的元素.

- 随机样本：从总体中随机地取n个个体，称为一个随机样本。
- 简单随机样本：满足以下两个条件的随机样本$(X_1,X_2,\ldots,X_n)$称为容量是n的简单随机样本。

1. 代表性：每个X_i与X同分布；
2. 独立性：$X_1,X_2,\ldots,X_n$是相互独立的随机变量。

[说明]：后面提到的样本均指简单随机样本。

- 统计量：样本的不含任何未知参数的函数。
- 常用统计量：设（$X_1,X_2,\ldots,X_n$）为取自总体X的简单随机样本。常用的统计量如下：

1. 样本均值 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$
2. 样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$，$S$为样本标准差
3. 样本矩 k阶矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \quad (k=1,2,\cdots)$ 反映总体K阶原点矩

   k阶中心矩：$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k \quad (k=1,2,\cdots)$

Review: 三种常见连续型分布的概率密度函数

- 均匀分布 $U(a,b)$: $f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b,\\ 0, & 其他.\end{cases}$
- 指数分布 $Exp(\lambda)$: $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0,\\ 0, & 其他.\end{cases}$
- 正态分布 $N(\mu,\sigma^2)$: $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty<x<+\infty.$

Review: 三种常见离散型分布的概率分布律 (列)

- Poisson 分布 $P(\lambda)$: $\mathrm{P}(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,\ldots,$

  $\Longrightarrow f(x)=\mathrm{P}(X=x)=\begin{cases}\frac{\lambda^x e^{-\lambda}}{x!}, & x=0,1,\ldots,\\ 0, & 其他.\end{cases}$
- 二项分布 $B(n,p)$: $\mathrm{P}(X=k)=\mathrm{C}_n^k p^k(1-p)^{n-k}, k=0,1,\ldots,n,$

  $\Longrightarrow f(x)=\mathrm{P}(X=x)=\begin{cases}\mathrm{C}_n^x p^x(1-p)^{n-x}, & x=0,1,\ldots,n,\\ 0, & 其他.\end{cases}$
- 0-1 分布 $B(1,p)$: $\mathrm{P}(X=k)=p^k(1-p)^{1-k}, k=0,1,$

  $\Longrightarrow f(x)=\mathrm{P}(X=x)=\begin{cases}p^x(1-p)^{1-x}, & x=0,1,\\ 0, & 其他.\end{cases}$

Table: 样本与总体的各阶矩对比表

	样本的矩 (随机变量)	总体 X 的矩 (数)
均值	$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$	$\mu=\mathrm{E}(X)$
方差	$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$	$\sigma^2=\mathrm{Var}(X)=\mathrm{E}\{(X-\mathrm{E}(X))^2\}$
均方差/标准差	$S=\sqrt{S^2}$	$\sigma=\sqrt{\sigma^2}$
样本 k 阶 (原点) 矩	$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$	$\mu_k=\mathrm{E}(X^k)$
样本 k 阶中心矩	$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$	$\nu_k=\mathrm{E}\{(X-\mathrm{E}(X))^k\}$

§6.2 Sampling Distributions ★★★

Example 4

(性质:) 设 $(X_1,X_2,\ldots,X_n)$ 是总体 X 的样本, 若 $\mathrm{E}(X)=\mu$, $\mathrm{Var}(X)=\sigma^2$, 则

$\mathrm{E}(\overline{X})=$ $\underline{\mu}$, $\mathrm{Var}(\overline{X})=$ $\underline{\sigma^2/n}$, $\mathrm{E}(S^2)=$ $\underline{\sigma^2}$.

Example 5

设总体 X 的概率密度为 $f(x)=\begin{cases}2(1-x), & 0<x<1,\\ 0, & 其他.\end{cases}$ $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)$ 是来自总体 X 的样本, 则 $E(X)=\int_0^1 2x(1-x)\mathrm{d}x=\frac{1}{3}$, $E(X^2)=\int_0^1 2x^2(1-x)\mathrm{d}x=\frac{1}{6}$,

$Var(X)=\frac{1}{18}$. 于是 $E(\overline{X})=$ $\underline{\frac{1}{3}}$, $E(\overline{X}^2)=$ $\underline{\frac{1}{18n}+\frac{1}{9}}$. $E(S^2)=$ $\underline{\frac{1}{18}}$.

三大分布:

(1) χ^2 分布

Definition 6 (χ^2 分布)

设随机变量 $X_1,X_2,\ldots,X_n$ 相互独立, $X_i\sim N(0,1)$, $i=1,2,\ldots,n$, 则称

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

服从自由度为 n 的 χ^2 分布 (chi-square distribution), 记为 $\chi_n^2\sim\chi^2(n)$.

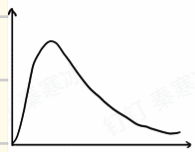
说明: 若随机变量 $K\sim\chi^2(n)$, 那么一定存在相互独立的随机变量 $X_1,X_2,\ldots,X_n$, $X_i\sim N(0,1)$, $i=1,2,\ldots,n$, 使得 K 可以表示为

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

$\chi^2(n)$ 分布的概率密度为

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

其中 $\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}\mathrm{d}x$ 为伽马函数.



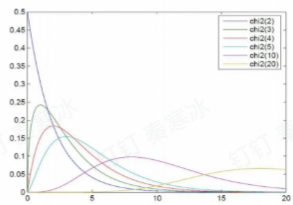


Figure: 不同自由度的 χ^2 分布概率密度函数图像比较

χ^2 分布的性质

1. 设随机变量 $X \sim \chi^2(n)$, 则有 $\mathsf{E}(X) = n$, $\mathsf{Var}(X) = 2n$.
2. 设 $Y_1 \sim \chi^2(n_1)$, $Y_2 \sim \chi^2(n_2)$, 且 Y_1, Y_2 相互独立, 则有 $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$.

该性质称为 χ^2 分布的可加性. 可推广至任意有限个的情形:

设 $Y_i \sim \chi^2(n_i)$, $i = 1, 2, \ldots, m$, 且 $Y_1, Y_2, \ldots, Y_m$ 相互独立, 则

$$\sum_{i=1}^{m} Y_i \sim \chi^2\Big(\sum_{i=1}^{m} n_i\Big).$$

χ^2 分布的性质 1 的证明

由于 $X \sim \chi^2(n)$, 故存在 n 个相互独立的标准正态变量 $X_1, X_2, \ldots, X_n$, 使得 $X = \sum_{i=1}^{n} X_i^2$. 因此

$$\mathsf{E}(X) = \mathsf{E}\Big(\sum_{i=1}^{n} X_i^2\Big) = n\mathsf{E}(X_1^2) = n\{\mathsf{Var}(X_1) + [\mathsf{E}(X_1)]^2\} = n(1 + 0^2) = n,$$

$$\mathsf{Var}(X) = \mathsf{Var}\Big(\sum_{i=1}^{n} X_i^2\Big) = n\mathsf{Var}(X_1^2) = n\{\mathsf{E}(X_1^4) - [\mathsf{E}(X_1^2)]^2\} = n(3 - 1^2) = 2n,$$

其中 $\mathsf{E}(X_1^4) = \int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 3$.

χ^2 分布的性质 2 的证明

由于 $Y_1 \sim \chi^2(n_1)$, $Y_2 \sim \chi^2(n_2)$, 且 Y_1, Y_2 相互独立, 故 Y_1, Y_2 可以表示为 $Y_1 = \sum_{i=1}^{n_1} X_{1i}^2$, $Y_2 = \sum_{j=1}^{n_2} X_{2j}^2$, 其中诸标准正态随机变量 X_{1i}, X_{2j} 相互独立, $i = 1, 2, \ldots, n_1, j = 1, 2, \ldots, n_2$. 所以

$$Y_1 + Y_2 = \sum_{i=1}^{n_1} X_{1i}^2 + \sum_{j=1}^{n_2} X_{2j}^2.$$

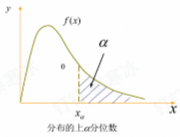
利用 χ^2 分布的定义可知 $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$.

上 α 分位数.

对给定的正数 $0 < \alpha < 1$, 若连续型随机变量 X 的分布函数和密度函数分别为 $F(x)$ 与 $f(x)$, 称满足条件

$$\mathsf{P}(X > x_\alpha) = 1 - F(x_\alpha) = \int_{x_\alpha}^{+\infty} f(x)\mathrm{d}x = \alpha$$

的实数 x_α 为随机变量 X (或此分布) 的上 (侧)α 分位数(或上 (侧)α 分位点).

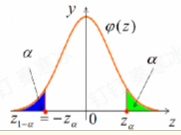


设 $X \sim N(0,1)$, 若 z_α 满足条件

$$\mathsf{P}(X > z_\alpha) = \alpha, \quad 0 < \alpha < 1,$$

则称点 z_α 为标准正态分布的上 (侧)α 分位数(或上 (侧)α 分位点). (Excel 实现: NORMSINV)

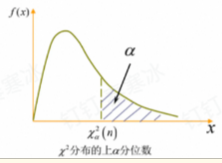
- $z_{1/2} = 0$.
- $z_{1-\alpha} = -z_\alpha$.



Definition 7 (χ^2 分布的上 α 分位数)

对给定的正数 $\alpha(0 < \alpha < 1)$, 称满足条件 $\int_{\chi_\alpha^2(n)}^{+\infty} f_n(y)\mathrm{d}y = \alpha$ 的点 $\chi_\alpha^2(n)$ 为 $\chi^2(n)$ 分布的上 (侧)α 分位数(或上 (侧)α 分位点), 其中 $f_n(y)$ 为 $\chi^2(n)$ 分布的概率密度函数.

上 α 分位数 $\chi_\alpha^2(n)$ 的值可查 χ^2 分布表 (附表 4) 或者利用各类软件得到.



Example 7

设总体 $X \sim N(\mu, \sigma^2)$, $-\infty < \mu < +\infty$, $\sigma > 0$. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 是取自总体 X 的 (简单随机) 样本.

1. 求随机变量 $\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$ 的分布;
2. 设 $n = 5$, 若 $a(X_1 - X_2)^2 + b(2X_3 - X_4 - X_5)^2 \sim \chi^2(k)$, 则 a, b, k 各为多少?

Solution

1. 作变换 $Y_i = \frac{X_i - \mu}{\sigma}$, $i = 1, \ldots, n$, 则 $Y_1, Y_2, \ldots, Y_n$ 相互独立, 且 $Y_i \sim N(0,1)$, 于是

$$\chi^2 = \sum_{i=1}^{n}\Big(\frac{X_i - \mu}{\sigma}\Big)^2 = \sum_{i=1}^{n} Y_i^2 \sim \chi^2(n).$$

2. 由于 $(X_1, X_2, \ldots, X_n)$ 是来自总体的简单随机样本, 利用正态分布的性质,

$$X_1 - X_2 \sim N(0, 2\sigma^2) \Rightarrow \Big(\frac{X_1 - X_2}{\sqrt{2}\sigma}\Big)^2 \sim \chi^2(1),$$

$$2X_3 - X_4 - X_5 \sim N(0, 6\sigma^2) \Rightarrow \Big(\frac{2X_3 - X_4 - X_5}{\sqrt{6}\sigma}\Big)^2 \sim \chi^2(1),$$

且 $X_1 - X_2$ 与 $2X_3 - X_4 - X_5$ 相互独立, 故由卡方分布的性质, 可得

$$\frac{(X_1 - X_2)^2}{2\sigma^2} + \frac{(2X_3 - X_4 - X_5)^2}{6\sigma^2} \sim \chi^2(2),$$

即 $a = \frac{1}{2\sigma^2}, b = \frac{1}{6\sigma^2}, k = 2$.

(2) t 分布

Definition 8 (t 分布)

设 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 并且 X, Y 相互独立, 则称随机变量

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

服从自由度为 n 的 t 分布 (t-distribution), 记为 $T \sim t(n)$. 也称为学生氏 (Student) 分布.

说明: 若随机变量 $T \sim t(n)$, 那么一定存在随机变量 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 并且 X, Y 相互独立, 使得 T 可以表示为

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

$t(n)$ 分布的概率密度为

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

其中 $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x$ 为伽马函数.

峰始终低于标准正态.

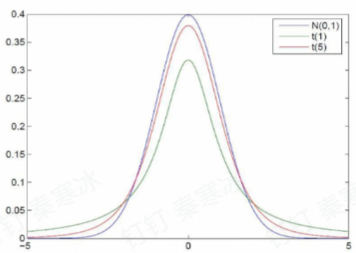


Figure: 不同自由度的 t 分布与标准正态分布密度函数图像比较

t 分布的性质

1. 设 $T \sim t(n)$, 当 $n \geq 2$ 时, 期望 $\mathsf{E}(T) = 0$,
   当 $n \geq 3$ 时, 方差 $\mathsf{Var}(T) = \frac{n}{n-2}$.
2. 当 n 足够大时, t 分布近似于标准正态分布 $N(0,1)$.
3. 若 $T \sim t(n)$, $N \sim N(0,1)$, 则对任意的 $n \geq 1$, 都存在 $a_0 > 0$, 使得

$$\mathsf{P}(|T| \geq a_0) \geq \mathsf{P}(|N| \geq a_0).$$

   性质 3 告诉我们, t 分布的概率密度函数图像尾部比标准正态分布厚重.
4. 对于任意 $0 < \alpha < 1$, $t_{1-\alpha}(n) = -t_\alpha(n)$, $n \geq 1$.

(3) F 分布

Definition 10 (F 分布)

设 $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, 并且 U, V 相互独立, 则称随机变量

$$F = \frac{U/n_1}{V/n_2}$$

服从自由度为 (n_1, n_2) 的 F 分布 (F-distribution), 记为 $F \sim F(n_1, n_2)$; 其中 n_1 称为第一自由度, n_2 称为第二自由度.

说明: 若随机变量 $F \sim F(n_1, n_2)$, 那么一定存在随机变量 $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, 并且 U, V 相互独立, 使得 F 可以表示为

$$F = \frac{U/n_1}{V/n_2}.$$

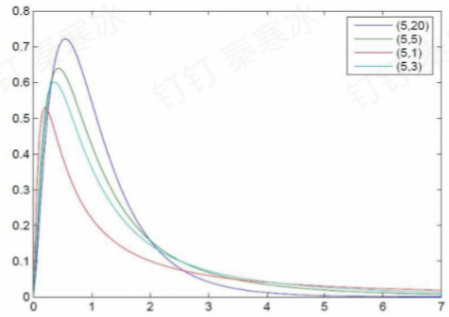
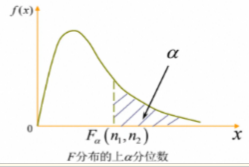


Figure: 不同自由度的 F 分布概率密度函数图像比较

F 分布的性质

1. 设 $F \sim F(n_1, n_2)$, 则 $\frac{1}{F} \sim F(n_2, n_1)$;
2. 对任意的 $0 < \alpha < 1$, $F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}$.
3. 若 $\xi \sim t(n)$, 则 $\xi^2 \sim F(1, n)$.

- $F_{1-\alpha}(n_1, n_2) = \dfrac{1}{F_\alpha(n_2, n_1)}$



Proof

设 $F \sim F(n_1, n_2)$, 由于 $\mathsf{P}(F > 0) = 1$, 且结合上 α 分位数的定义知

$$1 - \alpha = \mathsf{P}\big(F > F_{1-\alpha}(n_1, n_2)\big) = \mathsf{P}\Big(\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1, n_2)}\Big).$$

且注意到 F 为连续型的随机变量, 所以

$$\alpha = 1 - \mathsf{P}\Big(\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1, n_2)}\Big) = \mathsf{P}\Big(\frac{1}{F} \geq \frac{1}{F_{1-\alpha}(n_1, n_2)}\Big) = \mathsf{P}\Big(\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\Big),$$

由 F 分布的性质可知 $\frac{1}{F} \sim F(n_2, n_1)$, 故 $\frac{1}{F_{1-\alpha}(n_1, n_2)}$ 为 $F(n_2, n_1)$ 的上 α 分位数, 即 $\frac{1}{F_{1-\alpha}(n_1, n_2)} = F_\alpha(n_2, n_1)$, 因此

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}.$$

Proposition 1

设 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 是来自总体 X 的样本, 若 $\mathsf{E}(X) = \mu$, $\mathsf{Var}(X) = \sigma^2$, 则

$$\mathsf{E}(\overline{X}) = \mu, \quad \mathsf{Var}(\overline{X}) = \frac{\sigma^2}{n},$$

$$\mathsf{E}(S^2) = \sigma^2, \quad \mathsf{E}(B_2) = \frac{n-1}{n}\sigma^2.$$

Theorem 1

设 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 为来自正态总体 $N(\mu, \sigma^2)$ 的简单随机样本, $\overline{X}$ 是样本均值, S^2 是样本方差, 则有

1. $\overline{X} \sim N\Big(\mu, \dfrac{\sigma^2}{n}\Big)$, 即, $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$;
2. $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n-1)$;

   注意与 $\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2(n)$ 作比较
3. $\overline{X}$ 与 S^2 相互独立.

Theorem 2

设 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 为来自正态总体 $N(\mu, \sigma^2)$ 的简单随机样本, $\overline{X}$ 是样本均值, S^2 是样本方差, 则有

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

注意与 $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \dfrac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \sim N(0, 1)$ 作比较.

Proof

由 Theorem 1 知

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

且二者独立, 由 t 分布定义得

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{(n-1)S^2}{\sigma^2} \Big/ (n-1)} = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Theorem 3

设样本 $\boldsymbol{X} = (X_1, X_2, \ldots, X_{n_1})$ 和 $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_{n_2})$ 分别来自正态总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$, 并且它们相互独立. 记 $\overline{X}, \overline{Y}$ 分别是两个样本的样本均值, S_1^2, S_2^2 分别是两个样本的样本方差, 则有

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

   其中 $S_w^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \dfrac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{j=1}^{n_2}(Y_j - \overline{Y})^2}{n_1 + n_2 - 2}$, $S_w = \sqrt{S_w^2}$.

1. $F = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \dfrac{S_1^2}{S_2^2} \Big/ \dfrac{\sigma_1^2}{\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$

由 Theorem 1 知,

$$\frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1), \quad \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1),$$

且二者独立, 由 F 分布的定义, 有

$$\frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1 - 1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2 - 1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1).$$

2 $\dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$

由 Theorem 1 知,

$$\overline{X} \sim N\Big(\mu_1, \frac{\sigma_1^2}{n_1}\Big), \quad \overline{Y} \sim N\Big(\mu_2, \frac{\sigma_2^2}{n_2}\Big)$$

且 $\overline{X}$ 与 $\overline{Y}$ 相互独立, 所以

$$\overline{X} - \overline{Y} \sim N\Big(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\Big),$$

即

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1).$$

当 $\sigma_1 = \sigma_2 = \sigma$ 时. $S_w = \sqrt{S_w^2}$

$$S_w^2 = \frac{1}{n_1+n_2-2}\Big(\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{j=1}^{n_2}(Y_j - \overline{Y})^2\Big) = \frac{1}{n_1+n_2-2}\big((n_1-1)S_1^2 + (n_2-1)S_2^2\big).$$

Proof (Cont.)

3 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时, $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$

当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时, 由 (2) 得

$$U:=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1).$$

又由给定条件知

$$\frac{(n_1-1)S_1^2}{\sigma^2}\sim\chi^2(n_1-1),\quad \frac{(n_2-1)S_2^2}{\sigma^2}\sim\chi^2(n_2-1),$$

且它们相互独立.

Proof (Cont.)

3 故由 χ^2 分布的可加性知

$$V:=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2}\sim\chi^2(n_1+n_2-2),$$

且 U 与 V 相互独立. 于是据 t 分布的定义可知

$$\frac{U}{\sqrt{V/(n_1+n_2-2)}}=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2).$$

通常把 S_w^2 称作合样本方差.

Example 8

设总体 $X\sim N(\mu,\sigma^2)$, $\boldsymbol{X}=(X_1,\ldots,X_4)$ 与 $\boldsymbol{Y}=(Y_1,\ldots,Y_9)$ 是取自总体 X 的两个独立样本, $\overline{X},\overline{Y}$ 和 S_1^2,S_2^2 分别为样本均值和样本方差.

1. 求 $\mathsf{Var}(S_1^2-S_2^2)$;
2. 若 $a\dfrac{\overline{X}-\overline{Y}}{S_1}\sim t(k)$, 求 a,k;
3. 若 $b\displaystyle\sum_{i=1}^4(X_i-\mu)^2\Big/S_2^2\sim F(n_1,n_2)$, 求 b,n_1,n_2.

1 求 $\mathsf{Var}(S_1^2-S_2^2)$

在正态总体下, 有

$$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)\Rightarrow\mathsf{Var}\Big(\frac{(n-1)S^2}{\sigma^2}\Big)=2(n-1)\Rightarrow\mathsf{Var}(S^2)=\frac{2\sigma^4}{n-1}.$$

所以

$$\mathsf{Var}(S_1^2)=\frac{2\sigma^4}{3},\quad \mathsf{Var}(S_2^2)=\frac{\sigma^4}{4}.$$

由于 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 独立, 故 S_1^2 与 S_2^2 独立, 因此

$$\mathsf{Var}(S_1^2-S_2^2)=\mathsf{Var}(S_1^2)+\mathsf{Var}(S_2^2)=\frac{11\sigma^4}{12}.$$

Solution (Cont.)

2 若 $a\dfrac{\overline{X}-\overline{Y}}{S_1}\sim t(k)$, 求 a,k

由于 $\overline{X}\sim N\Big(\mu,\dfrac{\sigma^2}{4}\Big)$, $\overline{Y}\sim N\Big(\mu,\dfrac{\sigma^2}{9}\Big)$, 且 $\overline{X}$ 与 $\overline{Y}$ 相互独立, 故

$$\overline{X}-\overline{Y}\sim N\Big(0,\frac{13\sigma^2}{36}\Big)\Rightarrow\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{13\sigma^2}{36}}}\sim N(0,1),$$

又 $\dfrac{3S_1^2}{\sigma^2}\sim\chi^2(3)$, 且 $\overline{X}-\overline{Y}$ 与 S_1^2 相互独立, 所以

$$\frac{\frac{6}{\sqrt{13}}\frac{\overline{X}-\overline{Y}}{\sigma}}{\sqrt{\frac{3S_1^2}{\sigma^2}/3}}=\frac{6\sqrt{13}}{13}\frac{\overline{X}-\overline{Y}}{S_1}\sim t(3)\Rightarrow a=\frac{6\sqrt{13}}{13},k=3.$$

3 若 $b\displaystyle\sum_{i=1}^4(X_i-\mu)^2\Big/S_2^2\sim F(n_1,n_2)$, 求 b,n_1,n_2

由于

$$\frac{1}{\sigma^2}\sum_{i=1}^4(X_i-\mu)^2\sim\chi^2(4),\quad \frac{8S_2^2}{\sigma^2}\sim\chi^2(8),$$

且 $\displaystyle\sum_{i=1}^4(X_i-\mu)^2$ 与 S_2^2 相互独立, 故

$$\frac{\frac{1}{\sigma^2}\sum_{i=1}^4(X_i-\mu)^2/4}{\frac{8S_2^2}{\sigma^2}/8}=\frac{1}{4}\sum_{i=1}^4(X_i-\mu)^2\Big/S_2^2\sim F(4,8),$$

所以 $b=\dfrac{1}{4},(n_1,n_2)=(4,8)$.

Example 9

设总体 $X\sim\chi^2(n)$, $n\geq1$. 取样本 $\boldsymbol{X}=(X_1,X_2,\ldots,X_{16})$, 求 $\mathsf{E}(\overline{X})$, $\mathsf{Var}(\overline{X})$, $\mathsf{E}(B_2)$ 及 $\mathsf{P}\Big(\displaystyle\sum_{i=1}^8X_i\Big/\sum_{i=9}^{16}X_i\geq1\Big)$.

Solution

由题意知, $\mathsf{E}(X)=n,\mathsf{Var}(X)=2n$, 利用 Proposition 1, 可得

$$\mathsf{E}(\overline{X})=\mathsf{E}(X)=n,\quad \mathsf{Var}(\overline{X})=\frac{\mathsf{Var}(X)}{16}=\frac{n}{8},\quad \mathsf{E}(B_2)=\frac{15}{16}\mathsf{Var}(X)=\frac{15}{8}n.$$

注意: 此题中样本容量为 16, n 不表示样本容量.

Solution (Cont.)

因为 $\displaystyle\sum_{i=1}^8X_i\sim\chi^2(8n)$, $\displaystyle\sum_{i=9}^{16}X_i\sim\chi^2(8n)$ 且相互独立, 所以

$$F:=\sum_{i=1}^8X_i\Big/\sum_{i=9}^{16}X_i\sim F(8n,8n)\ \text{且}\ \frac{1}{F}\sim F(8n,8n).$$

因此

$$\gamma=\mathsf{P}(F\geq1)=\mathsf{P}\Big(\frac{1}{F}\leq1\Big)=1-\mathsf{P}\Big(\frac{1}{F}>1\Big)=1-\gamma\Rightarrow\gamma=0.5,$$

其中还利用了 F 分布的连续性, $\mathsf{P}\Big(\displaystyle\sum_{i=1}^8X_i\Big/\sum_{i=9}^{16}X_i=1\Big)=0$

Example 10

设总体 $X \sim N(\mu, \sigma^2)$, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)(n \geq 2)$ 是取自总体 X 的样本, $\overline{X}, S^2$ 分别为样本均值和样本方差. 求 $\mathsf{E}(\overline{X}S^2)$, $\mathsf{Var}(\overline{X}S^2)$.

Solution

注意到在正态总体下, $\overline{X}$ 和 S^2 是独立的, 故 $\mathsf{E}(\overline{X}S^2) = \mathsf{E}(\overline{X})\mathsf{E}(S^2) = \mu\sigma^2$, 而

$$\begin{aligned}
\mathsf{Var}(\overline{X}S^2) &= \mathsf{E}(\overline{X}^2 S^4) - [\mathsf{E}(\overline{X}S^2)]^2 \\
&= \mathsf{E}(\overline{X}^2)\mathsf{E}(S^4) - \mu^2\sigma^4 \\
&= \{\mathsf{Var}(\overline{X}) + [\mathsf{E}(\overline{X})]^2\}\{\mathsf{Var}(S^2) + [\mathsf{E}(S^2)]^2\} - \mu^2\sigma^4 \\
&= \left(\frac{\sigma^2}{n} + \mu^2\right)\left(\frac{2\sigma^4}{n-1} + \sigma^4\right) - \mu^2\sigma^4.
\end{aligned}$$

•参数估计

参数估计 (Parameter Estimation)

The rationale behind parameter estimation is quite simple. When sampling is from a population described by a d.f. $F(x;\theta)$, knowledge of θ yields knowledge of the entire population.

Example 1

若浙江大学学生的身高 X 服从正态分布, 其概率密度函数为

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty.$$

但参数 μ, σ^2 的值未知, 要求估计 μ, σ^2, 有时还希望以一定的可靠性来估计 μ 的值是在某个范围内或者不低于某个数.

参数估计就是通过样本估计总体分布所包含的未知参数.

参数估计有点估计和区间估计两种形式.

$$\text{Estimation}\begin{cases}\text{Point Estimation}\\ \text{Interval Estimation}\end{cases}$$

- 点估计 ——用一个的具体数值 (样本观察值的函数) 去估计一个未知参数
- 区间估计 ——把参数估计在某两个界限 (都是样本观察值的函数) 之间

§7.1 点估计 (Point Estimation)

点估计

设总体 $X \sim F(x;\boldsymbol{\theta})$, 分布函数 $F(\cdot)$ 形式已知, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ 未知, 为待估参数. 根据样本 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, 对每一个未知参数 $\theta_i (i = 1, 2, \ldots, k)$, 构造出统计量

$$\widehat{\theta}_i = \widehat{\theta}_i(\boldsymbol{X}) = \widehat{\theta}_i(X_1, X_2, \ldots, X_n)$$

作为参数 θ_i 的估计, 称 $\widehat{\theta}_i$ 为 θ_i 的 (点) 估计量 (point estimator). 若得到样本的观察值为 $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, 则 $\widehat{\theta}_i = \widehat{\theta}_i(\boldsymbol{x}) = \widehat{\theta}_i(x_1, x_2, \ldots, x_n)$ 为 θ_i 的 (点) 估计值.

说明: 估计量是一个估计参数的规则 (方案), 没有样本观察值时也可构造估计量. 当有了样本观察值, 就可以根据这个规则得出 θ 的估计值.

矩法 (Method of Moments)

思想: 用样本矩去估计相应的总体矩.

总体矩: 在总体 X 的下列数学期望存在的情况下,

总体 k 阶 (原点) 矩: $\mu_k = \mathsf{E}(X^k)$ $(k = 1, 2, \ldots)$;

总体 k 阶中心矩: $\nu_k = \mathsf{E}\{[X - \mathsf{E}(X)]^k\}$ $(k = 2, 3, \ldots)$.

样本矩: 样本 k 阶 (原点) 矩: $A_k = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^k$ $(k = 1, 2, \ldots)$;

样本 k 阶中心矩: $B_k = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^k$ $(k = 2, 3, \ldots)$.

简而言之, 就是用 A_k 去估计 μ_k, 用 B_k 去估计 ν_k.

☆☆☆

基本步骤:

1. 写出总体 X 的前 k 阶矩 (关于 k 个待估参数 $\theta_1, \theta_2, \ldots, \theta_k$ 的函数)
   (此步的目的是为了下一步, 有时这一步可以省略)
   $$\mu_v = \mathsf{E}(X^v) = h_v(\theta_1, \theta_2, \ldots, \theta_k), \quad v = 1, 2, \ldots, k.$$
2. 写出待估参数关于总体矩的函数表达式
   $$\theta_i = g_i(\mu_1, \mu_2, \ldots, \mu_k), \quad i = 1, 2, \ldots, k.$$
3. 将第二步中出现的总体矩用相应的样本矩 "替换", 即可得待估参数的矩法估计量
   $$\widehat{\theta}_i = g_i(A_1, A_2, \ldots, A_k), \quad i = 1, 2, \ldots, k.$$

Example 3

设总体 X 的均值 μ 和方差 σ^2 都存在, 且 $\sigma > 0$, 但 μ, σ^2 均未知, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 是取自总体 X 的一个样本, 试求 μ, σ^2 的矩法估计量.

Solution (解法一)

先求总体矩

$$\begin{cases}\mu_1 = \mathsf{E}(X) = \mu \\ \mu_2 = \mathsf{E}(X^2) = \mathsf{Var}(X) + [\mathsf{E}(X)]^2 = \sigma^2 + \mu^2\end{cases}$$

从而得到待估参数

$$\begin{cases}\mu = \mu_1 \\ \sigma^2 = \mu_2 - \mu_1^2\end{cases} \qquad (1)$$

Solution (解法一, Cont.)

从而得到待估参数

$$\begin{cases}\mu = \mu_1 \\ \sigma^2 = \mu_2 - \mu_1^2\end{cases} \qquad (1)$$

注意到样本 1 阶矩和 2 阶矩分别为

$$A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}, \quad A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

在 (1) 式中分别用 A_1 和 A_2 代替总体 1 阶矩 μ_1 和 2 阶矩 μ_2, 则可得到 μ, σ^2 的矩法估计量为

$$\begin{cases}\widehat{\mu} = A_1 = \overline{X}, \\ \widehat{\sigma^2} = A_2 - A_1^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2.\end{cases}$$

Solution (解法二)

先求总体矩, 再解待估参数

$$\begin{cases}\mu_1 = \mathsf{E}(X) = \mu \\ \nu_2 = \mathsf{Var}(X) = \sigma^2\end{cases} \Rightarrow \begin{cases}\mu = \mu_1 \\ \sigma^2 = \nu_2\end{cases} \qquad (2)$$

注意到样本 1 阶矩和 2 阶中心矩分别为 $A_1 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i = \overline{X}$, $B_2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$, 用此分别替换 (2) 式中的 μ_1 和 ν_2, 则可得 μ, σ^2 的矩法估计量为

$$\begin{cases}\widehat{\mu} = A_1 = \overline{X} \\ \widehat{\sigma^2} = B_2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2\end{cases}$$

Example 4

设总体的概率密度函数为 $f(x) = \begin{cases}\sqrt{\theta}x^{\sqrt{\theta}-1}, & 0 \leq x \leq 1, \\ 0, & \text{其他},\end{cases}$ 其中 $\theta > 0$ 为未知参数. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 是取自总体 X 的样本, 求 θ 的矩法估计量.

Solution

因为

$$\mu_1 = \mathsf{E}(X) = \int_0^1 \sqrt{\theta}x^{\sqrt{\theta}}\mathrm{d}x = \frac{\sqrt{\theta}}{\sqrt{\theta}+1} \Rightarrow \theta = \left(\frac{\mu_1}{1-\mu_1}\right)^2,$$

而 $A_1 = \overline{X}$, 所以 θ 的矩估计量为 $\widehat{\theta} = \left(\dfrac{\overline{X}}{1-\overline{X}}\right)^2$.

Example 5

设总体 $X \sim N(\mu, \sigma^2), -\infty < \mu < +\infty, \sigma > 0$, μ, σ 均未知. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 是取自 X 的一个样本, $n \geq 2$.

1. 求 $\mathsf{P}(X > 1)$ 的矩法估计量;
2. 若抽取了容量为 6 的样本, 样本观察值为 $\boldsymbol{x} = (2, 5, -3, -3, -1, 0)$, 求 $\mathsf{P}(X > 1)$ 的矩法估计值.

Solution

1. 因为

$$\mathsf{P}(X > 1) = \mathsf{P}\Big(\frac{X-\mu}{\sigma} > \frac{1-\mu}{\sigma}\Big) = 1 - \Phi\Big(\frac{1-\mu}{\sigma}\Big) = 1 - \Phi\Big(\frac{1-\mu_1}{\sqrt{\nu_2}}\Big),$$

Solution (Cont.)

1. 所以

$$\widehat{\mathsf{P}(X > 1)} = 1 - \Phi\Big(\frac{1-A_1}{\sqrt{B_2}}\Big) = 1 - \Phi\Big(\frac{1-\overline{X}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}}\Big).$$

2. 由样本观察值计算得 $\overline{x} = 0$, $b_2 = \frac{1}{6}\sum_{i=1}^{6}(x_i - \overline{x})^2 = 8$. 所以 $\mathsf{P}(X > 1)$ 的矩估计值为

$$\widehat{\mathsf{P}(X > 1)} = 1 - \Phi\Big(\frac{1-\overline{x}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}}\Big) = 1 - \Phi\Big(\frac{1}{\sqrt{8}}\Big) \approx 1 - \Phi(0.35) = 0.3632.$$

Example 6

设 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 是来自对数级分布的一组样本 $(n \geq 2)$, 总体 X 的分布律为

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

其中 $0 < p < 1$ 未知, 求参数 p 的矩法估计量.

Solution

注意到总体 X 的 1 阶矩为

$$\mu_1 = \mathsf{E}(X) = \sum_{k=1}^{+\infty} k \cdot \Big(-\frac{1}{\ln(1-p)} \cdot \frac{p^k}{k}\Big) = -\frac{1}{\ln(1-p)} \sum_{k=1}^{+\infty} p^k = -\frac{p}{(1-p)\ln(1-p)}$$

从中无法反解出待估参数 p, 进而继续考虑 2 阶矩.

Solution (Cont.)

而总体 X 的 2 阶 (原点) 矩

$$\mu_2 = \mathsf{E}(X^2) = \sum_{k=1}^{+\infty} k^2 \cdot \Big(-\frac{1}{\ln(1-p)} \cdot \frac{p^k}{k}\Big) = -\frac{p}{\ln(1-p)} \sum_{k=1}^{+\infty} k p^{k-1}$$

$$= -\frac{p}{\ln(1-p)} \cdot \Big(\frac{x}{1-x}\Big)'\Big|_{x=p} = -\frac{p}{\ln(1-p)} \cdot \frac{1}{(1-p)^2}.$$

因此 $\frac{\mu_1}{\mu_2} = 1 - p$, 故可得 $p = 1 - \frac{\mu_1}{\mu_2}$. 所以 p 的矩估计量为

$$\widehat{p} = 1 - \frac{A_1}{A_2} = 1 - \frac{\overline{X}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2} = 1 - \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2}.$$

矩法的理论依据

矩法的主要理论依据是大数定律. 因为若样本为简单随机样本时, 当总体 X 的 k 阶原点矩 $\mathsf{E}(X^k)$ 存在时, 根据辛钦大数律有

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{\mathsf{P}} \mu_k, \quad 当 \quad n \to +\infty, \quad k = 1, 2, \ldots.$$

更进一步, 由依概率收敛的性质可知, 对连续函数 $g(\cdot)$, 有

$$g(A_1, A_2, \ldots, A_k) \xrightarrow{\mathsf{P}} g(\mu_1, \mu_2, \ldots, \mu_k), \quad 当 \quad n \to +\infty.$$

Example 8

在一罐中放着大量的黑球和白球, 并假定已知两种球的数目之比是 1 : 3, 但不知道哪种颜色的球多. 如果用放回抽样方法从罐中取 5 个球, 观察结果是: 黑、白、黑、黑、黑. 由此来估计罐中的球况 (即估计随机取一球, 取到黑球的概率 p).

Solution

从罐中随机取一球, 考察该球的颜色. 设总体为 $X = \begin{cases} 1, & 取到的是黑球, \\ 0, & 取到的是白球, \end{cases}$

则 $X \sim 0-1(p)$, 这里 p 未知. 由题意可知, $p \in \{1/4, 3/4\} \triangleq \Theta$ (参数空间).

Θ(参数空间).

Solution (Cont.)

现抽取了容量为 5 的一组样本 $\boldsymbol{X}$, 得到的样本观察值 $\boldsymbol{x} = (1, 0, 1, 1, 1)$, 则出现本次观察结果的概率为

$$\mathsf{P}(\boldsymbol{X} = \boldsymbol{x}) = \mathsf{P}(X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1) = (1-p)p^4.$$

(为参数 p 的函数)

当 $p = \frac{1}{4}$ 时, 出现本次观察结果的概率为 $\frac{3}{4} \times \Big(\frac{1}{4}\Big)^4 = \frac{3}{1024}$.

当 $p = \frac{3}{4}$ 时, 出现本次观察结果的概率为 $\frac{1}{4} \times \Big(\frac{3}{4}\Big)^4 = \frac{81}{1024}$.

由于 $\frac{3}{1024} < \frac{81}{1024}$, 因此认为 $p = \frac{3}{4}$ 比 $p = \frac{1}{4}$ 更有可能, 故认为取 $\widehat{p} = \frac{3}{4}$ 更合理.

似然函数

一般地, 设离散型总体 $X \sim f(x;\theta) = \mathsf{P}\{X = x\}$, θ 未知, $\theta \in$ 参数空间 Θ. 从总体中取得样本 $\boldsymbol{X} = (X_1, \ldots, X_n)$, 其样本观察值为 $\boldsymbol{x} = (x_1, \ldots, x_n)$, 则事件 $\{\boldsymbol{X} = \boldsymbol{x}\}$ 发生的概率为

离散型可能常用! 求极大似然估计值!

$$\mathsf{P}_\theta\{\boldsymbol{X} = \boldsymbol{x}\} = \mathsf{P}_\theta\{X_1 = x_1, \ldots, X_n = x_n\} = \prod_{i=1}^{n} f(x_i;\theta), \quad \theta \in \Theta.$$

它是参数 θ 的函数, 记为似然函数 (likelihood function), 即似然函数为

$$L(\theta) = L(\theta; \boldsymbol{x}) = L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta), \quad \theta \in \Theta.$$

Definition 1 (极大似然估计)

称

$$\widehat{\theta}(\boldsymbol{x}) = \arg\max_{\theta \in \Theta} L(\theta; \boldsymbol{x})$$

为 θ 的极大似然估计值, 即 $\widehat{\theta}(\boldsymbol{x})$ 为满足

$$L(\widehat{\theta}(\boldsymbol{x}); \boldsymbol{x}) = \max_{\theta \in \Theta} L(\theta; \boldsymbol{x}).$$

相应的统计量

$$\widehat{\theta}(\boldsymbol{X}) = \arg\max_{\theta \in \Theta} L(\theta; \boldsymbol{X})$$

即为 θ 的极大似然估计量 (maximum likelihood estimator).

似然函数 (Cont.)

设连续型总体 X 的概率密度为 $f(x;\theta)$, θ 未知, $\theta \in$ 参数空间 Θ. 从总体中取得样本 $\boldsymbol{X} = (X_1, \ldots, X_n)$, 其样本观察值为 $\boldsymbol{x} = (x_1, \ldots, x_n)$, 则似然函数为

$$L(\theta) = L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i;\theta), \quad \theta \in \Theta.$$

极大似然估计 (MLE). (总体 $X \sim f(x)$) ①. 写出似然函数 $L(\theta)$. $\theta \in \Theta$ ②. 找 $\hat{\theta}$.

- (定义法) 若 $L(\theta)$ 关于某个 θ_i 是单调增 (减) 函数, 此时 θ_i 的极大似然估计在其边界取得;
- (不变原则) 若 $\widehat{\theta}$ 是 θ 的极大似然估计, $g(\theta)$ 是 θ 的连续函数, 则 $g(\widehat{\theta})$ 是 $g(\theta)$ 的极大似然估计.

说明:

- 未知参数可能不是一个, 一般设为 $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$;
- (微分法) 在求 $L(\theta; \boldsymbol{x})$ 的最大值时, 当 $L(\theta)$ 关于 θ 可微时, 通常转化为求似然方程

$$\frac{\partial L(\theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \ldots, k.$$

有时, 还可以转换为求对数似然函数 (log likelihood function)

$$\ln L(\theta) = \ln L(\theta; \boldsymbol{x}) = \sum_{i=1}^{n} \ln f(x_i;\theta)$$

的最大值, 当 $\ln L(\theta)$ 关于 θ 可微时, 通常转化为求对数似然方程

$$\frac{\partial \ln L(\theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \ldots, k.$$

Example 9

设总体的概率密度函数为 $f(x)=\begin{cases}\sqrt{\theta}x^{\sqrt{\theta}-1}, & 0\le x\le 1,\\ 0, & \text{其他},\end{cases}$ 其中 $\theta>0$ 为未知参数.

$\boldsymbol{X}=(X_1,X_2,\dots,X_n)$ 是取自总体 X 的样本, 求 θ 的极大似然估计量.

Solution

似然函数

$$L(\theta)=\prod_{i=1}^n f(x_i;\theta)=\theta^{\frac{n}{2}}\Big(\prod_{i=1}^n x_i\Big)^{\sqrt{\theta}-1},\quad \underline{\theta>0, 0\le x_i\le 1, i=1,2,\dots,n.}$$

$$\Rightarrow\quad \ln L(\theta)=\frac{n}{2}\ln\theta+(\sqrt{\theta}-1)\sum_{i=1}^n\ln x_i.$$

Solution (Cont.)

令

$$\frac{\partial\ln L(\theta)}{\partial\theta}=\frac{n}{2}\cdot\frac{1}{\theta}+\frac{1}{2\sqrt{\theta}}\sum_{i=1}^n\ln x_i=0,\quad\Longrightarrow\quad \theta=\frac{n^2}{\Big(\sum\limits_{i=1}^n\ln x_i\Big)^2}$$

经微分法检验, θ 的极大似然估计量为

$$\widehat{\theta}_{\mathsf{MLE}}=\frac{n^2}{\Big(\sum\limits_{i=1}^n\ln X_i\Big)^2}.$$

微分法求解 MLE:

- 写出似然函数 $L(\theta)=L(\theta;x_1,x_2,\dots,x_n)=\prod\limits_{i=1}^n f(x_i;\theta),\ \theta\in\Theta.$

  (有时需要写出对数似然函数

  $$\ln L(\theta)=\ln L(\theta;x_1,x_2,\dots,x_n)=\sum_{i=1}^n\ln f(x_i;\theta),\ \theta\in\Theta.)$$

- 列出似然方程 $\dfrac{\partial}{\partial\theta}L(\theta)=0$

  (有时需要列出对数似然方程 $\dfrac{\partial}{\partial\theta}\ln L(\theta)=0.$)
- 解 (对数) 似然方程, 得 $\theta=h(x_1,x_2,\dots,x_n)$.
- 经微分法检验, 得 θ 的 MLE 为 $\widehat{\theta}=h(X_1,X_2,\dots,X_n)$.

Example 10

设总体 X 服从 $[0,\theta]$ 上的均匀分布, $\theta>0$ 未知, 试由样本 $\boldsymbol{X}=(X_1,X_2,\dots,X_n)$ 求出 θ 的矩法估计量和极大似然估计量.

Solution

(矩估计) 由 $\mu_1=\mathsf{E}(X)=\dfrac{\theta}{2}$ 得 θ 的矩估计量为 $\widehat{\theta}=2\overline{X}$.

(极大似然估计) 因 X 的概率密度为 $f(x;\theta)=\begin{cases}\dfrac{1}{\theta}, & 0\le x\le\theta,\\ 0, & \text{其他},\end{cases}$ 故参数 θ 的似然函数为

$$L(\theta)=\begin{cases}\dfrac{1}{\theta^n}, & \underline{0\le x_1,x_2,\dots,x_n\le\theta},\\ 0, & \text{其他},\end{cases}$$

Solution (Cont.)

故对数似然函数为

$$\ln L(\theta)=\begin{cases}-n\ln\theta, & 0\le x_1,x_2,\dots,x_n\le\theta,\\ 0, & \text{其他},\end{cases}$$

由于 $\dfrac{\partial\ln L(\theta)}{\partial\theta}=-\dfrac{n}{\theta}\ne 0$, 不能用微分法求 $\widehat{\theta}_{\mathsf{MLE}}$.

下面从定义出发求之.

注意到参数 θ 的似然函数为

$$L(\theta)=\begin{cases}\dfrac{1}{\theta^n}, & \underline{0\le x_1,x_2,\dots,x_n\le\theta},\\ 0, & \text{其他},\end{cases}$$

因为 $L(\theta)>0$ 需 $0\le x_i\le\theta, i=1,2,\dots,n$, 即此时 θ 的取值范围是 $\theta\in[x_{(n)},+\infty)$, 其中 $x_{(n)}=\max\{x_1,x_2,\dots,x_n\}$. 又 $L(\theta)=\dfrac{1}{\theta^n}$ 对 $\theta\ge x_{(n)}\ge 0$ 是减函数, θ 越小, $L(\theta)$ 越大, 故 $x_{(n)}$ 为似然函数的极大值点, 因此 θ 的极大似然估计量为

$$\widehat{\theta}_{\mathsf{MLE}}=X_{(n)},$$

其中 $X_{(n)}=\max\{X_1,X_2,\dots,X_n\}$.

Example 11

设总体 X 的概率密度函数为 $f(x)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-\frac{x-\mu}{\theta}}, & x\ge\mu,\\ 0, & \text{其他},\end{cases}$ 其中 $\theta>0,\theta,\mu$ 是未知参数.

$\boldsymbol{X}=(X_1,X_2,\dots,X_n)$ 为来自总体 X 的样本, 求 θ,μ 的矩估计与极大似然估计.

Solution

(矩估计)

$$\mu_1=\mathsf{E}(X)=\int_\mu^{+\infty}\frac{x}{\theta}\mathrm{e}^{-\frac{x-\mu}{\theta}}\mathrm{d}x=\mu+\theta,$$

$$\mu_2=\mathsf{E}(X^2)=\int_\mu^{+\infty}\frac{x^2}{\theta}\mathrm{e}^{-\frac{x-\mu}{\theta}}\mathrm{d}x=\mu^2+2\mu\theta+2\theta^2,$$

$$\nu_2=\mathsf{Var}(X)=\mathsf{E}(X^2)-[\mathsf{E}(X)]^2=\theta^2.$$

Solution (Cont.)

从而反解得 $\begin{cases}\theta=\sqrt{\nu_2},\\ \mu=\mu_1-\sqrt{\nu_2},\end{cases}$ 因此 θ,μ 的矩估计量为

$$\begin{cases}\widehat{\theta}=\sqrt{B_2}=\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^n(X_i-\overline{X})^2},\\ \widehat{\mu}=A_1-\sqrt{B_2}=\overline{X}-\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^n(X_i-\overline{X})^2}.\end{cases}$$

Solution (Cont.)

(极大似然估计) 写出似然函数

$$L(\theta,\mu)=\prod_{i=1}^n\frac{1}{\theta}\mathrm{e}^{-\frac{x_i-\mu}{\theta}}=\frac{1}{\theta^n}\mathrm{e}^{-\frac{1}{\theta}\sum\limits_{i=1}^n(x_i-\mu)},\quad \underline{\theta>0, x_i\ge\mu, i=1,2,\dots,n.}$$

(不能通过求偏导数获得 μ 的 MLE.)

对数似然函数为

$$\ln L(\theta,\mu)=-n\ln\theta-\frac{1}{\theta}\sum_{i=1}^n x_i+\frac{n\mu}{\theta}.$$

因此对数似然方程为

$$\begin{cases}\dfrac{\partial\ln L(\theta,\mu)}{\partial\theta}=-\dfrac{n}{\theta}+\dfrac{1}{\theta^2}\sum\limits_{i=1}^n(x_i-\mu)=0, & \Rightarrow\theta=\overline{x}-\mu\\ \dfrac{\partial\ln L(\theta,\mu)}{\partial\mu}=\dfrac{n}{\theta}=0. & \text{无解}\end{cases}$$

考虑到 $L(\theta,\mu)>0$ 需 $x_i\ge\mu, i=1,2,\dots,n$, 即 $\mu\le x_{(1)}=\min\{x_1,x_2,\dots,x_n\}$.
而 $L(\theta,\mu)$ 关于 μ 是增函数, 所以 μ 的极大似然估计为

$$\widehat{\mu}_{\mathsf{MLE}}=X_{(1)}=\min\{X_1,X_2,\dots,X_n\},$$

从而 θ 的极大似然估计为 $\widehat{\theta}_{\mathsf{MLE}}=\overline{X}-X_{(1)}$.

PS. 总体密度函数的支撑范围依赖参数, 则一般考虑增/减性?

离散型极大似然估计

Example 13

设总体 X 的概率分布律为 $\begin{pmatrix} 1 & 2 & 3 \\ \theta & \frac{\theta}{2} & 1-\frac{3\theta}{2} \end{pmatrix}$，其中 $0<\theta<\frac{2}{3}$ 未知. 现得到样本观察值 $\boldsymbol{x}=(2,3,2,1,3)$，求 θ 的矩估计值和极大似然估计值.

Solution

(矩估计)

$$\mu_1=\mathsf{E}(X)=\theta+2\times\frac{\theta}{2}+3\times\left(1-\frac{3\theta}{2}\right)=3-\frac{5\theta}{2}\implies\theta=\frac{2}{5}(3-\mu_1).$$

故 θ 的矩估计量为 $\widehat{\theta}=\frac{2}{5}(3-\overline{X})$. 现有 $\overline{x}=2.2$，所以 θ 的矩估计值为 $\widehat{\theta}=\frac{2}{5}(3-\overline{x})=0.32$.

Solution (Cont.)

(极大似然估计) 似然函数

$$L(\theta)=\mathsf{P}\{\boldsymbol{X}=\boldsymbol{x}\}=\frac{\theta}{2}\cdot\left(1-\frac{3\theta}{2}\right)\cdot\frac{\theta}{2}\cdot\theta\cdot\left(1-\frac{3\theta}{2}\right)=\frac{1}{16}\theta^3(2-3\theta)^2.$$

对数似然函数

$$\ln L(\theta)=-\ln 16+3\ln\theta+2\ln(2-3\theta).$$

由 $\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=\frac{3}{\theta}-\frac{6}{2-3\theta}=0$ 得 θ 的极大似然估计值为

$$\widehat{\theta}_{\mathsf{MLE}}=0.4.$$

Example 15

设总体 $X\sim N(\mu,\sigma^2),-\infty<\mu<+\infty,\sigma>0$，$\mu,\sigma$ 均未知. $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)$ 是取自 X 的一个样本.

1. 求 $\mathsf{P}(X>1)$ 的极大似然估计量;
2. 若抽取了容量为 6 的样本，样本观察值为 $\boldsymbol{x}=(2,5,-3,-3,-1,0)$，求 $\mathsf{P}(X>1)$ 的极大似然估计值.

Solution

1. 由书本例 7.1.7 的 (3)，知 μ,σ^2 的极大似然估计量为

正态.
→与矩估计一样

$$\widehat{\mu}_{\mathsf{MLE}}=\overline{X},\quad \widehat{\sigma^2}_{\mathsf{MLE}}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$$ (注：不是 S^2).

Solution (Cont.)

1. 而

$$\mathsf{P}(X>1)=\mathsf{P}\left(\frac{X-\mu}{\sigma}>\frac{1-\mu}{\sigma}\right)=1-\Phi\left(\frac{1-\mu}{\sigma}\right)$$

是 μ,σ^2 的连续函数，由不变原则，得

$$\widehat{\mathsf{P}(X>1)}_{\mathsf{MLE}}=1-\Phi\left(\frac{1-\widehat{\mu}_{\mathsf{MLE}}}{\widehat{\sigma}_{\mathsf{MLE}}}\right)=1-\Phi\left(\frac{1-\overline{X}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}}\right).$$

2. 由于 $\widehat{\mu}_{\mathsf{MLE}}=\overline{x}=0$，$\widehat{\sigma}_{\mathsf{MLE}}=\sqrt{8}$. 所以 $\mathsf{P}(X>1)$ 的极大似然估计值为

$$\widehat{\mathsf{P}(X>1)}=1-\Phi\left(\frac{1-\widehat{\mu}_{\mathsf{MLE}}}{\widehat{\sigma}_{\mathsf{MLE}}}\right)=1-\Phi\left(\frac{1}{\sqrt{8}}\right)\approx 1-\Phi(0.35)=0.3632.$$

Table 1. 两种估计方法所得结果对比

Example	矩估计量	极大似然估计量
Example 4 Example 9	$\widehat{\theta}=\left(\frac{\overline{X}}{1-\overline{X}}\right)^2$	$\widehat{\theta}_{\mathsf{MLE}}=\frac{n^2}{\left(\sum_{i=1}^{n}\ln X_i\right)^2}$
Example 10	$\widehat{\theta}=2\overline{X}$	$\widehat{\theta}_{\mathsf{MLE}}=X_{(n)}$
Example 11	$\widehat{\theta}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}$ $\widehat{\mu}=\overline{X}-\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}$	$\widehat{\mu}_{\mathsf{MLE}}=X_{(1)}$ $\widehat{\theta}_{\mathsf{MLE}}=\overline{X}-X_{(1)}$

Example 16

设总体 X 的一阶和二阶矩存在，分布是任意的，记 $\mathsf{E}(X)=\mu$，$\mathsf{Var}(X)=\sigma^2$，证明：样本均值 $\overline{X}$ 和样本方差 S^2 分别是 μ 和 σ^2 的无偏估计量.

Proof

因 $X_1,X_2,\ldots,X_n$ 与 X 同分布，故有

$$\mathsf{E}(\overline{X})=\mathsf{E}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}\mathsf{E}(X_i)=\frac{1}{n}\cdot n\mu=\mu.$$

故 $\overline{X}$ 是 μ 的无偏估计量.

Proof (Cont.)

对于样本方差，有

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

故 S^2 是 σ^2 的无偏估计量.

标准检验.

(1) 无偏性准则

Definition 2 (无偏估计量)

若参数 θ 的估计量 $\widehat{\theta}=\widehat{\theta}(\boldsymbol{X})$ 的数学期望存在，且满足

$$\mathsf{E}(\widehat{\theta})=\theta,\quad\forall\theta\in\Theta,$$

则称 $\widehat{\theta}$ 是 θ 的一个无偏估计量 (unbiased estimator).

若 $\mathsf{E}(\widehat{\theta})\neq\theta$，则 $\widehat{\theta}$ 是 θ 的一个有偏估计量，并称 $|\mathsf{E}(\widehat{\theta})-\theta|$ 为估计量 $\widehat{\theta}$ 的偏差 (bias).

若 $\mathsf{E}(\widehat{\theta})\neq\theta$，但满足 $\lim\limits_{n\to+\infty}\mathsf{E}(\widehat{\theta})=\theta$，则称 $\widehat{\theta}$ 是 θ 的渐近无偏估计量 (asymptotic unbiased estimator).

Example 17

检验 Example 10 中矩估计量 $\widehat{\theta}=2\overline{X}$ 与极大似然估计量 $\widehat{\theta}_{\mathsf{MLE}}=X_{(n)}$ 的无偏性.

Solution

对均匀分布总体 $X\sim U[0,\theta]$，有 $\mathsf{E}(X)=\frac{\theta}{2}$. 由于 $X_1,\ldots,X_n$ 与 X 同分布，所以

$$\mathsf{E}(\widehat{\theta})=\mathsf{E}(2\overline{X})=\frac{2}{n}\sum_{i=1}^{n}\mathsf{E}(X_i)=\frac{2}{n}\cdot n\cdot\frac{\theta}{2}=\theta.$$

因此 $\widehat{\theta}=2\overline{X}$ 是 θ 的无偏估计.

Solution (Cont.)

为考察 $\widehat{\theta}_{\mathsf{MLE}}=X_{(n)}$ 的无偏性，先求 $X_{(n)}$ 的分布. 由最大值的分布结论可知，

$$F_{X_{(n)}}(x)=[F(x)]^n=\left(\frac{x}{\theta}\right)^n,\quad 0\le x\le\theta.$$

$$\Rightarrow\quad f_{X_{(n)}}(x)=\frac{nx^{n-1}}{\theta^n},\quad 0\le x\le\theta.$$

因此有

$$\mathsf{E}(\widehat{\theta}_{\mathsf{MLE}})=\mathsf{E}(X_{(n)})=\int_0^\theta\frac{x\cdot nx^{n-1}}{\theta^n}\mathrm{d}x=\frac{n}{n+1}\theta\neq\theta.$$

所以 $\widehat{\theta}_{\mathsf{MLE}}=X_{(n)}$ 是有偏的.

Proposition 1 (纠偏方法)

如果 $\mathsf{E}(\widehat{\theta})=a\theta+b$，$\theta\in\Theta$，其中 a,b 是常数，且 $a\neq 0$. 则 $\frac{1}{a}(\widehat{\theta}-b)$ 是 θ 的无偏估计.

例如，在 Example 17 中，取 $X_{(n)}^*=\frac{n+1}{n}X_{(n)}$，则 $X_{(n)}^*$ 是 θ 的无偏估计.

无偏性是对估计量的一个最常见的重要要求，是"好"估计的标准之一.
无偏性的统计意义是指在大量重复试验下，由 $\widehat{\theta}(\boldsymbol{X})$ 所作的估计值的平均恰是 θ，从而无偏性保证了 $\widehat{\theta}$ 没有系统误差.

(2) 有效性准则

为比较两个或多个无偏估计量，一般会考察估计量的波动性，即估计量的方差. 无偏估计量的方差越小，该估计量的取值越集中在参数真值的附近，这就是有效性准则.

Definition 3

设 $\widehat{\theta}_1 = \widehat{\theta}_1(\boldsymbol{X})$ 和 $\widehat{\theta}_2 = \widehat{\theta}_2(\boldsymbol{X})$ 都是参数 θ 的无偏估计，若

$$Var(\widehat{\theta}_1) \le Var(\widehat{\theta}_2), \quad \forall \theta \in \Theta$$

且至少有一个 $\theta \in \Theta$ 使不等号成立，则称 $\widehat{\theta}_1$ 比 $\widehat{\theta}_2$ 有效 (effective).

Example 19

设样本 $\boldsymbol{X} = (X_1, \ldots, X_{n_1})$ 和 $\boldsymbol{Y} = (Y_1, \ldots, Y_{n_2})$ 分别来自总体 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$ 并且它们相互独立，其样本方差分别为 S_1^2, S_2^2，令 $T = aS_1^2 + bS_2^2$.

1. 问 T 是 σ^2 的无偏估计的条件?
2. 在这些无偏估计中哪个最有效? (当 n_1, n_2 均大于 1 时)

Solution

1. 注意到
$$\mathsf{E}(T) = a\mathsf{E}(S_1^2) + b\mathsf{E}(S_2^2) = (a+b)\sigma^2,$$
所以 T 是 σ^2 的无偏估计 $\Leftrightarrow a + b = 1$.

Solution (Cont.)

2 计算
$$\mathsf{Var}(T) = a^2\frac{2\sigma^4}{n_1-1} + b^2\frac{2\sigma^4}{n_2-1} = \frac{2\sigma^4}{n_1-1}a^2 + \frac{2\sigma^4}{n_2-1}(1-a)^2.$$
由
$$\frac{\partial \mathsf{Var}(T)}{\partial a} = \frac{4\sigma^4}{n_1-1}a - \frac{4\sigma^4}{n_2-1}(1-a) = 4\sigma^4\Big[\frac{n_1+n_2-2}{(n_1-1)(n_2-1)}a - \frac{1}{n_2-1}\Big] = 0$$
解得 $a = \dfrac{n_1-1}{n_1+n_2-2}$，进而 $b = \dfrac{n_2-1}{n_1+n_2-2}$，且当 n_1, n_2 均大于 1 时，$\dfrac{\partial^2 \mathsf{Var}(T)}{\partial a^2} > 0$. 即当 $T = S_w^2$ 时最有效.

我们自然希望估计量的均方误差越小越好，即有如下准则.

(3) 均方误差准则

Definition 4 (均方误差, Mean Squared Error, MSE)

设 $\widehat{\theta} = \widehat{\theta}(\boldsymbol{X})$ 是参数 θ 的估计量，其方差存在，称
$$Mse(\widehat{\theta}) = E[(\widehat{\theta} - \theta)^2]$$
为估计量 $\widehat{\theta}$ 的均方误差 (mean squared error, MSE).

若 $\widehat{\theta}$ 是 θ 的无偏估计，则有 $\mathsf{Mse}(\widehat{\theta}) = \mathsf{Var}(\widehat{\theta})$.

Definition 5

设 $\widehat{\theta}_1$ 与 $\widehat{\theta}_2$ 都是 θ 的估计量，若
$$Mse(\widehat{\theta}_1) \le Mse(\widehat{\theta}_2), \quad \forall \theta \in \Theta$$
且至少有一个 $\theta \in \Theta$ 使不等号成立，则称 $\widehat{\theta}_1$ 优于 $\widehat{\theta}_2$.

在实际应用中，有时人们认为均方误差准则比无偏性准则更为重要.

另外，均方误差有如下分解式:
$$\mathsf{Mse}(\widehat{\theta}) = \mathsf{Var}(\widehat{\theta}) + \big[\mathsf{E}(\widehat{\theta}) - \theta\big]^2.$$

Example 20

试利用均方误差准则，对用样本方差 S^2 和样本二阶中心矩 B_2 分别估计正态总体方差 σ^2 时进行评价 (样本容量 $n \ge 2$).

Solution

因为 S^2 是 σ^2 的无偏估计，则 $\mathsf{Mse}(S^2) = \mathsf{Var}(S^2) = \dfrac{2\sigma^4}{n-1}$? (利用定理 6.3.2，并结合卡方分布的方差). 而
$$\begin{aligned}\mathsf{Mse}(B_2) &= \mathsf{Var}(B_2) + \big[\mathsf{E}(B_2) - \sigma^2\big]^2 \\ &= \mathsf{Var}\Big(\frac{n-1}{n}S^2\Big) + \Big[\mathsf{E}\Big(\frac{n-1}{n}S^2\Big) - \sigma^2\Big]^2 = \frac{2n-1}{n^2}\sigma^4 < \frac{2\sigma^4}{n-1}.\end{aligned}$$
因此在均方误差准则下，B_2 优于 S^2.

(4) 相合性准则

一个好的估计量，当样本容量 n 越大时，该估计应该越精确越可靠，特别是当 $n \to +\infty$ 时，估计量的取值与参数真值应几乎完全一致，这就是相合性准则.

Definition 6

设 $\widehat{\theta}_n = \widehat{\theta}(\boldsymbol{X})$ 是参数 θ 的估计量，若对 $\forall \theta \in \Theta$，当 $n \to +\infty$ 时，$\widehat{\theta}_n$ 依概率收敛于 θ，即 $\widehat{\theta}_n \xrightarrow{P} \theta$，也就是
$$\lim_{n\to+\infty} P(|\widehat{\theta}_n - \theta| < \varepsilon) = 1, \quad \forall \varepsilon > 0,$$
则称 $\widehat{\theta}_n$ 是 θ 的(弱) 相合估计量 (consistent estimator) 或一致估计量.

一般用 Chebyshev 不等式或大数律验证，有时还需结合依概率收敛的性质.

Example 22

设总体 X 的 k 阶矩 $\mathsf{E}(X^k) = \mu_k (k \ge 1)$ 存在，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 是取自总体 X 的样本. 证明:

1. $\overline{X}$ 是 $\mu_1 = \mathsf{E}(X)$ 的相合估计;
2. $A_l = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i^l$, $l = 2, 3, \ldots, k$ 是 μ_l, $l = 2, 3, \ldots, k$ 的相合估计;
3. $B_2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^n (X_i - \overline{X})^2$, S^2 是 $\mathsf{Var}(X) = \sigma^2$ 的相合估计;
4. S 是 σ 的相合估计.

Proof

由 Khinchin 大数定律知，$A_l = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i^l$ 依概率收敛到 $\mu_l = \mathsf{E}(X^l)$，$l = 1, 2, \ldots, k$.

因此 (1), (2) 成立.

根据依概率收敛的性质，因为 $\mathsf{Var}(X) = \sigma^2 = \mu_2 - \mu_1^2$，所以
$$B_2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 = A_2 - \overline{X}^2$$
是 σ^2 的相合估计.

注意到 $S^2 = \dfrac{n}{n-1}B_2$，因此 S^2 也是 σ^2 的相合估计，进而 S 是 σ 的相合估计.

置信区间 (Confidence Interval)

Definition 7 (置信区间)

设总体 X 的分布函数 $F(x;\theta)$ 含有一个未知参数 θ, $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)$ 是来自总体 X 的样本, 对给定的值 $\alpha(0<\alpha<1)$, 如果有两个统计量 $\widehat{\theta}_L=\widehat{\theta}_L(\boldsymbol{X})$ 和 $\widehat{\theta}_U=\widehat{\theta}_U(\boldsymbol{X})$ 满足 $\widehat{\theta}_L<\widehat{\theta}_U$, 使得

$$P(\widehat{\theta}_L<\theta<\widehat{\theta}_U)\geq 1-\alpha,\quad \forall\theta\in\Theta,$$

则称随机区间 $(\widehat{\theta}_L,\widehat{\theta}_U)$ 是 θ 的置信水平为 $1-\alpha$ 的 (双侧) 置信区间 (confidence interval); $\widehat{\theta}_L$ 和 $\widehat{\theta}_U$ 分别称为 θ 的置信水平为 $1-\alpha$ 的 (双侧) 置信下限 (lower confidence limit) 和 (双侧) 置信上限 (upper confidence limit); 称 $1-\alpha$ 为置信水平 (confidence level) 或置信度.

Definition 8 (单侧置信区间)

对给定的 $\alpha\in(0,1)$, 如果统计量 $\widehat{\theta}_L=\widehat{\theta}_L(\boldsymbol{X})$ 满足

$$P(\widehat{\theta}_L<\theta)\geq 1-\alpha,\quad \theta\in\Theta,$$

则称 $\widehat{\theta}_L$ 是参数 θ 的置信水平为 $1-\alpha$ 的单侧置信下限, 随机区间 $(\widehat{\theta}_L,+\infty)$ 是 θ 的置信水平为 $1-\alpha$ 的单侧置信区间.

对给定的 $\alpha\in(0,1)$, 如果统计量 $\widehat{\theta}_U=\widehat{\theta}_U(\boldsymbol{X})$ 满足

$$P(\theta<\widehat{\theta}_U)\geq 1-\alpha,\quad \theta\in\Theta,$$

则称 $\widehat{\theta}_U$ 是参数 θ 的置信水平为 $1-\alpha$ 的单侧置信上限, 随机区间 $(-\infty,\widehat{\theta}_U)$ 是 θ 的置信水平为 $1-\alpha$ 的单侧置信区间.

评价区间估计的原则

- 置信度原则
  希望随机区间 $(\widehat{\theta}_L,\widehat{\theta}_U)$ 包含真值 θ 的概率越大越好.
- 精确度原则
  我们称随机区间 $(\widehat{\theta}_L,\widehat{\theta}_U)$ 的平均长度, 即

$$\mathsf{E}(\widehat{\theta}_U-\widehat{\theta}_L)$$

  为其精确度, 希望其越短越好. 并称二分之一区间的平均长度为置信区间的误差限.

当 n 固定时, 置信度原则和精确度原则常常是相互制约的.
实际中, 我们采用 Neyman 原则, 即在保证置信度的前提下, 尽可能提高精确度.

寻求区间估计的常用方法——枢轴量法 (Pivotal Quantity Method)

枢轴量

枢轴量是样本 $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)$ 和待估参数 θ 的函数, 即 不能含待估参数外的其他参数.

$$G=G(\boldsymbol{X},\theta)=G(X_1,X_2,\ldots,X_n;\theta)$$

一般要求 G 的分布已知, 即其分布不依赖于任何未知的参数.

注意:

- 用枢轴量法构造区间估计, 最关键在于选取一个合适的枢轴量.
  一般会根据合理的思想, 得到 θ 的点估计量 $\widehat{\theta}$ (如: 矩估计量、极大似然估计量等), 在此基础上进行加工, 以构造枢轴量.
- 若第 2 步中 a 和 b 的解不唯一, 通常用如下两种方案处理:
  1. 根据 Neyman 原则, 求 a 和 b 使得区间平均长度 $\mathsf{E}(\widehat{\theta}_U-\widehat{\theta}_L)$ 最短;
     (这样得到的置信区间称为最优置信区间.)
  2. 如果最优解不存在或求解比较复杂, 为应用的方便, 常取 a 和 b 满足

$$\mathsf{P}(G(\boldsymbol{X};\theta)\leq a)=\mathsf{P}(G(\boldsymbol{X};\theta)\geq b)=\alpha/2.$$

两边各留 $\frac{\alpha}{2}$

     (这样得到的置信区间称为等尾置信区间.)

基本步骤:

1. 构造枢轴量 $G(\boldsymbol{X};\theta)$, 其形式上是样本 $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)$ 和待估参数 θ 的函数, 且其分布已知, 不依赖于任何未知参数;
2. 对给定的置信水平 $1-\alpha$, 选择两个常数 a 和 b, 使得

$$\mathsf{P}(a<G(\boldsymbol{X};\theta)<b)\geq 1-\alpha,\quad \forall\theta\in\Theta;$$

   (当总体 X 为连续型总体时, 一般选择两个常数 a 和 b, 使得

$$\mathsf{P}(a<G(\boldsymbol{X};\theta)<b)=1-\alpha\quad \forall\theta\in\Theta;)$$

3. 假如能从 $a<G(\boldsymbol{X};\theta)<b$ 得到等价的不等式

$$\widehat{\theta}_L(\boldsymbol{X})<\theta<\widehat{\theta}_U(\boldsymbol{X}),$$

   那么 $(\widehat{\theta}_L,\widehat{\theta}_U)$ 就是 θ 的置信水平为 $1-\alpha$ 的置信区间.
   (若第二步中是取等号的, 那么得到的置信区间为同等置信区间)

Example 24

设某产品的寿命 X 服从参数为 λ 的指数分布, $\boldsymbol{X}=(X_1,X_2,\ldots,X_7)$ 是来自该总体的样本, 若已经证明 $2\lambda\sum_{i=1}^{7}X_i=2n\overline{X}\lambda=14\overline{X}\lambda\sim\chi^2(2n)$, 即 $\chi^2(14)$ 并已知 $\overline{x}=125$ 小时, 试利用枢轴量法推断该产品平均寿命 $\frac{1}{\lambda}$ 的范围. (置信水平为 0.90.)

Solution

由于 $14\overline{X}\lambda\sim\chi^2(14)$, 形式上依赖于样本和待估参数 $1/\lambda$, 且分布已知, 与参数 λ 无关, 因此可取枢轴量为 $14\lambda\overline{X}\sim\chi^2(14)$, 且有

$$\mathsf{P}\{\chi^2_{0.95}(14)<14\lambda\overline{X}<\chi^2_{0.05}(14)\}=0.90.$$

Solution (Cont.)

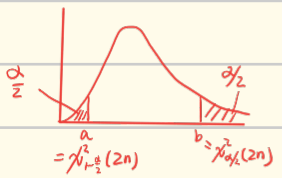
因此该产品平均寿命 $\frac{1}{\lambda}$ 的置信水平为 0.90 的置信区间为

$$\left(\frac{14\overline{X}}{\chi^2_{0.05}(14)},\frac{14\overline{X}}{\chi^2_{0.95}(14)}\right).$$

由 Excel 或查表得 $\chi^2_{0.05}(14)=23.685$, $\chi^2_{0.95}(14)=6.571$, 并将样本资料 $\overline{x}=125$ 代入上式得

$$(73.89,\quad 266.32)$$

为平均寿命 $\frac{1}{\lambda}$ 的置信水平为 0.90 的 (双侧) 置信区间.



(一) 单个正态总体 $N(\mu,\sigma^2)$ 的情形

设 $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)$ 是取自正态总体 $N(\mu,\sigma^2)$ 的一个样本, $\overline{X}$ 和 S^2 分别为样本均值和样本方差, 置信水平为 $1-\alpha$.

1. 均值 μ 的置信区间

(1) σ^2 已知时

$\overline{X}$ 是 μ 的一个常用估计量, 且 $\overline{X}\sim N(\mu,\sigma^2/n)$. 取枢轴量为

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

因为 $z_{1-\alpha/2}=-z_{\alpha/2}$, 所以有

$$\mathsf{P}\left\{-z_{\alpha/2}<\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<z_{\alpha/2}\right\}=1-\alpha,$$

即

$$\mathsf{P}\left\{\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right\}=1-\alpha.$$

因此 μ 的置信区间为

$$\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right),$$

此时区间的平均长度为 $\mathsf{E}\left[\overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}-\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)\right]=\frac{2\sigma}{\sqrt{n}}z_{\alpha/2}.$

下面考虑 μ 的置信水平为 $1-\alpha$ 的单侧置信下限.

因为仍然属于估计 μ(方差 σ^2 已知), 故同样取枢轴量为

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

现要求 μ 的单侧置信下限, 就是要求 $\mathrm{P}\{\mu>?\}=1-\alpha$ 中的 ?, 注意到

$$\mathrm{P}\left\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<z_\alpha\right\}=1-\alpha,\iff \mathrm{P}\left\{\mu>\overline{X}-\frac{\sigma}{\sqrt{n}}z_\alpha\right\}=1-\alpha.$$

那么 μ 的置信水平为 $1-\alpha$ 的单侧置信下限为 $\overline{X}-\frac{\sigma}{\sqrt{n}}z_\alpha$, 即 μ 的置信水平为 $1-\alpha$ 的单侧置信区间为 $\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_\alpha,+\infty\right)$.

思考: μ 的置信度为 $1-\alpha$ 的单侧置信上限是 ??? $\overline{X}+\frac{\sigma}{\sqrt{n}}z_\alpha$

(2) σ^2 未知时

此时, (1) 中的 $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ 不能取为枢轴量, 因为其形式上含有未知参数 σ(非待估参数). 一个自然的想法, 对于 $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ 中的未知的总体方差 σ^2, 用其无偏估计量——样本方差 S^2 代替, 故可取枢轴量为

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

注意到 t 分布的单峰对称性, 所以有

$$\mathrm{P}\left\{-t_{\alpha/2}(n-1)<\frac{\overline{X}-\mu}{S/\sqrt{n}}<t_{\alpha/2}(n-1)\right\}=1-\alpha,$$

即

$$\mathrm{P}\left\{\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)<\mu<\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right\}=1-\alpha.$$

因此 μ 的置信区间为 $\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1),\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right)$.

Example 25

设某种植物的高度 X (cm) 服从正态分布 $N(\mu,\sigma^2)$, 随机选取 36 棵, 其平均高度为 15 cm. 就以下两种情形, 求 μ 的置信水平为 95% 的双侧置信区间.

1. $\sigma^2=16$;
2. σ^2 未知, $s^2=16$.

Solution

1. 已知 $n=36$, $\overline{x}=15$, $\sigma=4$, 估计 μ, 此时取 $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$ 作为枢轴量. 那么

$$\mathrm{P}\left\{-z_{\alpha/2}<\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<z_{\alpha/2}\right\}=1-\alpha,$$

故 $\mathrm{P}\left\{\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right\}=1-\alpha$, 即 μ 的置信水平为 $1-\alpha$ 的置信区间为

$$\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right).$$

而 $\alpha=0.05$, 查表得 $z_{\alpha/2}=z_{0.025}=1.96$, 代入数据得

$$\overline{x}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}=15-\frac{1.96\times4}{\sqrt{36}}=13.693,\quad \overline{x}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}=15+\frac{1.96\times4}{\sqrt{36}}=16.307,$$

即 μ 的置信水平为 95% 的置信区间为 $(13.693,\quad 16.307)$.

2 已知 $n=36$, $\overline{x}=15$, $s^2=16$, 但 σ 未知, 估计 μ, 此时取 $\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$ 作为枢轴量. 那么 $\mathrm{P}\left\{\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)<\mu<\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right\}=1-\alpha$, 即 μ 的置信水平为 $1-\alpha$ 的置信区间为

$$\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1),\quad \overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right).$$

而 $\alpha=0.05$, $n=36$, 查表得 $t_{\alpha/2}(n-1)=t_{0.025}(35)=2.0301$, 代入数据得

$$15-\frac{2.0301\times4}{\sqrt{36}}=13.647,\quad 15+\frac{2.0301\times4}{\sqrt{36}}=16.353,$$

即 μ 的置信水平为 95% 的置信区间为 $(13.647,\quad 16.353)$.

(3) 成对数据情形

例如, 为考察某种降压药的降压效果, 测试了 n 个高血压病人在服药前后的血压(收缩压) 分别为 $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$.

由于个人体质的差异, $X_1,X_2,\ldots,X_n$ 不能看成来自同一个正态总体的样本, 即 $X_1,X_2,\ldots,X_n$ 是相互独立但不同分布的样本, $Y_1,Y_2,\ldots,Y_n$ 也是.

此外, 对同一个个体, X_i 和 Y_i 也是不独立的.

对于此类问题, 通常作差, 得

$$D_i=X_i-Y_i,\ i=1,2,\ldots,n,$$

这就取消了个体的差异, 仅与降压药的作用有关, 因此可以将 $D_1,D_2,\ldots,D_n$ 看成来自同一正态总体 $N(\mu_D,\sigma_D^2)$ 的样本, 且相互独立.

考察降压药的降压效果, 就可转化为参数 μ_D 的估计问题. 此时取枢轴量为 $T(\boldsymbol{D};\mu_D)=\frac{\overline{D}-\mu_D}{S_D/\sqrt{n}}\sim t(n-1)$, 则 μ_D 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\overline{D}-\frac{S_D}{\sqrt{n}}t_{\alpha/2}(n-1),\overline{D}+\frac{S_D}{\sqrt{n}}t_{\alpha/2}(n-1)\right),$$

其中 $\overline{D}=\overline{X}-\overline{Y}$, $S_D=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(D_i-\overline{D})^2}$.

Example 26

A、B 两种小麦品种分别播种在面积相等的 8 块试验田中, 每块田地 A、B 品种各播种一半, 收获后 8 块试验田的产量如下:

品种 A: 140, 137, 136, 140, 145, 148, 140, 135
品种 B: 135, 118, 115, 140, 128, 131, 130, 115

假设两品种产量的差服从正态分布, 求两品种产量差的期望的置信区间 ($\alpha=0.05$).

Solution

这是成对数据问题, 令 $D_i=X_i-Y_i$, $i=1,2,\ldots,n$, 由题意知, 可将此看成来自 $N(\mu_D,\sigma_D^2)$ 的样本. 取枢轴量为

$$T(\boldsymbol{D};\mu_d)=\frac{\overline{D}-\mu_D}{S_D/\sqrt{n}}\sim t(n-1),$$

则 μ_D 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\overline{D}-\frac{S_D}{\sqrt{n}}t_{\alpha/2}(n-1),\overline{D}+\frac{S_D}{\sqrt{n}}t_{\alpha/2}(n-1)\right).$$

由已知数据, 计算得

$$d_i=x_i-y_i:\ 5,\ 19,\ 21,\ 0,\ 17,\ 17,\ 10,\ 20$$

则 $n=8$, $\overline{d}=13.625$, $s_D=7.745$. 查表得 $t_{\alpha/2}(n-1)=t_{0.025}(7)=2.365$, 代入相关数据得所求置信区间为

$$(7.149,\quad 20.101).$$

σ^2 的置信区间

(2) μ 未知时

注意到 $\frac{(n-1)S^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2\sim\chi^2(n-1)$, 取枢轴量为

$$G(\boldsymbol{X};\mu)=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1).$$

由于

$$\mathrm{P}\left\{\chi^2_{1-\alpha/2}(n-1)<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{\alpha/2}(n-1)\right\}=1-\alpha,$$

即

$$\mathrm{P}\left\{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}<\sigma^2<\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right\}=1-\alpha.$$

因此 σ^2 的置信区间为

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

区间的平均长度为

$$\mathrm{E}\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}-\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right]=(n-1)\sigma^2\left(\frac{1}{\chi^2_{1-\alpha/2}(n-1)}-\frac{1}{\chi^2_{\alpha/2}(n-1)}\right).$$

进一步可知 σ 的置信区间为

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}},\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}\right).$$

注 如上得到的置信区间也不是最优置信区间.

Example 27

一个园艺科学家正在培养一个新品种的苹果，这种苹果除了口感好和颜色鲜艳以外，另一个重要特征是单个重量差异不大. 为了评估新苹果，她随机挑选了 25 个测试重量 (单位: 克). 已知重量服从 $N(\mu,\sigma^2)$，其样本方差为 $s^2=4.25$. 试求 σ^2 的置信水平为 95% 和 99% 的置信区间.

Solution

取枢轴量为 $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，则有

$$\mathrm{P}\left\{\chi^2_{1-\alpha/2}(n-1)<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{\alpha/2}(n-1)\right\}=1-\alpha$$

Solution (Cont.)

因此 σ^2 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

现 $n=25$, $s^2=4.25$, 查表得 $\chi^2_{0.025}(24)=39.4$, $\chi^2_{0.975}(24)=12.4$, $\chi^2_{0.005}(24)=45.6$, $\chi^2_{0.995}(24)=9.89$, 代入得

$$\frac{24\times4.25}{39.4}=2.59,\ \frac{24\times4.25}{12.4}=8.23,\ \frac{24\times4.25}{45.6}=2.24,\ \frac{24\times4.25}{9.89}=10.31,$$

所以 σ^2 的置信度为 95% 的置信区间为 $(2.59, 8.23)$,
σ^2 的置信度为 99% 的置信区间为 $(2.24, 10.31)$.

(二) 两个正态总体 $N(\mu_1,\sigma_1^2)$, $N(\mu_2,\sigma_2^2)$ 的情形

设 $\boldsymbol{X}=(X_1,X_2,\ldots,X_{n_1})$ 是来自 $N(\mu_1,\sigma_1^2)$ 的样本, $\boldsymbol{Y}=(Y_1,Y_2,\ldots,Y_{n_2})$ 是来自 $N(\mu_2,\sigma_2^2)$ 的样本, 且两样本相互独立. $\overline{X},\overline{Y}$ 和 S_1^2,S_2^2 分别是两个总体的样本均值和样本方差, 置信度为 $1-\alpha$.

考虑以下两个总体比较问题:

- 比较均值 (估计 $\mu_1-\mu_2$)
- 比较方差 (估计 $\dfrac{\sigma_1^2}{\sigma_2^2}$)

1. $\mu_1-\mu_2$ 的置信区间 (Behrens-Fisher 问题)

(1) σ_1^2,σ_2^2 已知时

由 $\overline{X}-\overline{Y}\sim N\left(\mu_1-\mu_2,\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)$, 取枢轴量为

$$G(\boldsymbol{X},\boldsymbol{Y};\mu_1,\mu_2)=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1).$$

可得置信区间为

$$\left(\overline{X}-\overline{Y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ \overline{X}-\overline{Y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right).$$

(2) σ_1^2,σ_2^2 未知时

此时, 自然的想法是用它们的无偏估计 S_1^2 和 S_2^2 代替, 得

$$G'(\boldsymbol{X},\boldsymbol{Y};\mu_1,\mu_2)=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}.$$

遗憾的是, 一般情况下, G' 的分布依赖于未知参数, 它不能成为枢轴量.

下面就几种特殊情形进行分析.

(i) $\sigma_1^2=\sigma_2^2=\sigma^2$, 但 σ^2 未知

此时由第六章定理知枢轴量

$$T(\boldsymbol{X},\boldsymbol{Y};\mu_1,\mu_2)=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2).$$

因此置信区间为

$$\left(\overline{X}-\overline{Y}-t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \overline{X}-\overline{Y}+t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right),$$

其中 $S_w^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$, $S_w=\sqrt{S_w^2}$.

(ii) $\sigma_1^2\neq\sigma_2^2$ 且未知 (仅作参考)

当样本量 n_1 和 n_2 都充分大时 (一般要大于 50), 根据中心极限定理得

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}\overset{\text{渐近}}{\sim}N(0,1).$$

取之为枢轴量, 则 $\mu_1-\mu_2$ 的近似置信区间为

$$\left(\overline{X}-\overline{Y}-z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}},\overline{X}-\overline{Y}+z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right).$$

说明 实际操作中, 若采样得 $\boldsymbol{x},\boldsymbol{y}$, 可得区间估计的一个实现

$$(\widehat{\mu_1-\mu_2}_L(\boldsymbol{x},\ \boldsymbol{y}),\widehat{\mu_1-\mu_2}_U(\boldsymbol{x},\boldsymbol{y})),$$

若

- 此区间中含有 0, 则大致可以认为 μ_1 与 μ_2 差不多;
- 此区间的下限大于 0, 则大致可以认为 μ_1 比 μ_2 大;
- 此区间的上限小于 0, 则大致可以认为 μ_1 比 μ_2 小.

(2) μ_1,μ_2 未知时

注意到 $\dfrac{(n_1-1)S_1^2}{\sigma_1^2}\sim\chi^2(n_1-1)$, $\dfrac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2(n_2-1)$, 且相互独立. 取枢轴量为

$$\frac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}\Big/(n_1-1)}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2}\Big/(n_2-1)}=\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1).$$

从而

$$\mathrm{P}\left\{F_{1-\alpha/2}(n_1-1,n_2-1)<\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}<F_{\alpha/2}(n_1-1,n_2-1)\right\}=1-\alpha,$$

即

$$\mathrm{P}\left\{\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right\}=1-\alpha.$$

由此可以等尾地构造出置信区间

$$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\ \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right).$$

注意到 F 分布分位数的"三变性质", 该置信区间也可以写作

$$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\ \frac{S_1^2}{S_2^2}F_{\alpha/2}(n_2-1,n_1-1)\right).$$

说明 实际操作中, 若采样得 $\boldsymbol{x},\boldsymbol{y}$, 可得区间估计的一个实现

$$\left(\widehat{\frac{\sigma_1^2}{\sigma_2^2}}_L(\boldsymbol{x},\boldsymbol{y}),\widehat{\frac{\sigma_1^2}{\sigma_2^2}}_U(\boldsymbol{x},\boldsymbol{y})\right),$$

若

- 此区间中含有 1, 则大致可以认为 σ_1^2 与 σ_2^2 差不多;
- 此区间的下限大于 1, 则大致可以认为 σ_1^2 比 σ_2^2 大;
- 此区间的上限小于 1, 则大致可以认为 σ_1^2 比 σ_2^2 小.

Example 28

两台机床生产同一个型号的滚珠，从甲机床生产的滚珠中抽取 8 个，从乙机床生产的滚珠中抽取 9 个，测得这些滚珠的直径 (单位: mm) 如下:

甲机床: 15.0 14.8 15.2 15.4 14.9 15.1 15.2 14.8
乙机床: 15.2 15.0 14.8 15.1 14.6 14.8 15.1 14.5 15.0

设两机床生产的滚珠直径分别为 X, Y，且 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$.

1. $\sigma_1 = 0.18$, $\sigma_2 = 0.24$，求 $\mu_1 - \mu_2$ 的置信水平为 0.90 的置信区间;
2. $\sigma_1 = \sigma_2 = \sigma$ 未知，求 $\mu_1 - \mu_2$ 的置信水平为 0.90 的置信区间;
3. $\sigma_1 \neq \sigma_2$ 未知，求 $\mu_1 - \mu_2$ 的置信水平为 0.90 的近似置信区间 (仅作参考);
4. μ_1, μ_2 未知，求 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信水平为 0.90 的置信区间.

1 当 $\sigma_1 = 0.18$, $\sigma_2 = 0.24$ 时，取枢轴量为

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0, 1),$$

则 $\mu_1 - \mu_2$ 的置信水平为 $1 - \alpha$ 的置信区间为

$$\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\quad \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

现 $\alpha = 0.1$，查表得 $z_{0.05} = 1.645$，从而所求区间为 $(-0.018,\ \ 0.318)$.

2 当 $\sigma_1 = \sigma_2 = \sigma$ 未知时，取枢轴量为

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_w\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

则 $\mu_1 - \mu_2$ 的置信水平为 $1 - \alpha$ 的置信区间为

$$\left(\overline{X} - \overline{Y} \pm t_{\alpha/2}(n_1 + n_2 - 2) S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right),$$

现 $\alpha = 0.1$，且 $n_1 = 8$, $n_2 = 9$，查表得 $t_{0.05}(15) = 1.7531$，并计算得 $s_w = 0.228$，从而所求区间为 $(-0.044,\ \ 0.344)$.

3 当 $\sigma_1 \neq \sigma_2$ 未知时，取枢轴量为

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \overset{\text{近似}}{\sim} t(k),$$

其中自由度 k 取 $\min(n_1 - 1, n_2 - 1) = 7$，则 $\mu_1 - \mu_2$ 的置信水平为 $1 - \alpha$ 的置信区间为

$$\left(\overline{X} - \overline{Y} \pm t_{\alpha/2}(k)\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right),$$

现 $\alpha = 0.1$，查表得 $t_{0.05}(7) = 1.895$，从而所求区间为 $(-0.058,\ \ 0.358)$.

4 当 μ_1, μ_2 未知时，取枢轴量为

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1),$$

则 σ_1^2/σ_2^2 的置信水平为 $1 - \alpha$ 的置信区间为

$$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1 - 1, n_2 - 1)},\ \ \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)}\right),$$

现 $\alpha = 0.1$，查表得 $F_{0.05}(7, 8) = 3.50$，$F_{0.95}(7, 8) = \dfrac{1}{F_{0.05}(8, 7)} = \dfrac{1}{3.73} = 0.268$，从而所求区间为 $(0.227,\ \ 2.965)$.

正态总体均值、方差的置信区间与单侧置信限 (置信度为 $1 - \alpha$)

	待估参数	其他参数	枢轴量 G 及其分布
单个正态总体	μ	σ^2 已知	$\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$
	μ	σ^2 未知	$\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n - 1)$
	σ^2	μ 未知	$\dfrac{(n - 1)S^2}{\sigma^2} \sim \chi^2(n - 1)$
两个正态总体	$\mu_1 - \mu_2$	σ_1^2, σ_2^2 已知	$\dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0, 1)$
	$\mu_1 - \mu_2$	$\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知	$\dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_w\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$ 其中 $S_w^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$
	$\mu_1 - \mu_2$	$\sigma_1^2 \neq \sigma_2^2$ 未知	$\dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \begin{cases} \sim N(0, 1), & \text{当 } n_1, n_2 > 50, \\ \sim t(k), & k = \min(n_1 - 1, n_2 - 1) \end{cases}$
	$\dfrac{\sigma_1^2}{\sigma_2^2}$	μ_1, μ_2 未知	$\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$

§7.5 Interval Estimation for Non-normal Population

前面我们均假设总体 X 服从正态分布，若分布不是正态分布或者根本未知呢?

Example 30

设总体 $X \sim F(x)$，分布未知，方差为 4. 样本 $\boldsymbol{X} = (X_1, X_2, \ldots, X_{100})$ 来自此总体，且已知 $\overline{x} = 5$. 求 X 的期望 μ 的置信水平为 95% 的近似置信区间.

Solution

已知 $\mathsf{E}(X) = \mu$, $\mathsf{Var}(X) = 4$，而 $\widehat{\mu} = \overline{X}$，现从 $\overline{X}$ 出发寻找枢轴量.
注意到 $n = 100$ 足够大，X_i 独立同分布，且方差存在，故由中心极限定理知

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{近似}}{\sim} N(0, 1).$$

Solution (Cont.)

取之为枢轴量，可导出均值 μ 的置信度为 $1 - \alpha$ 的近似置信区间为

$$\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

现 $\sigma = 2$, $\overline{x} = 5$, $z_{0.025} = 1.96$，则均值 μ 的置信水平为 95% 的近似置信区间为

$$\left(5 - 1.96 \times \frac{2}{\sqrt{100}}, 5 + 1.96 \times \frac{2}{\sqrt{100}}\right) = (4.608, 5.392).$$

注　当 σ^2 未知时，以样本方差 S^2 代入，得 μ 的置信水平为 $1 - \alpha$ 的近似置信区间为

$$\left(\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right).$$

Example 31

设总体 $X \sim B(1,p)$, 参数 p 未知, $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)$ 是来自该总体的样本. 当样本容量 $n>50$ 时, 求 X 的期望 p 的置信水平为 $1-\alpha$ 的近似置信区间.

Solution

由中心极限定理, 取枢轴量为

$$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} = \frac{\overline{X}-p}{\sqrt{p(1-p)/n}} \overset{\text{近似}}{\sim} N(0,1),$$

则有

$$\mathrm{P}\left\{-z_{\alpha/2} < \frac{\overline{X}-p}{\sqrt{p(1-p)/n}} < z_{\alpha/2}\right\} \approx 1-\alpha.$$

Solution (Cont.)

即

$$\mathrm{P}\{(n+z_{\alpha/2}^2)p^2-(2n\overline{X}+z_{\alpha/2}^2)p+n\overline{X}^2<0\} \approx 1-\alpha.$$

令 $a=n+z_{\alpha/2}^2$, $b=-(2n\overline{X}+z_{\alpha/2}^2)$, $c=n\overline{X}^2$, 得参数 p 的置信度为 $1-\alpha$ 的近似置信区间为

$$\left(\frac{-b-\sqrt{b^2-4ac}}{2a}, \frac{-b+\sqrt{b^2-4ac}}{2a}\right) = (\widehat{p}_L, \widehat{p}_U).$$

或取 $p(1-p)$ 的估计量为 $\widehat{p}(1-\widehat{p})$, 得 p 的置信度为 $1-\alpha$ 的近似置信区间为

$$\left(\overline{X}-z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \overline{X}+z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right).$$

Chapter 8. 假设检验

§8.1 Basic Ideas of Hypothesis Testing

" 反证法"

给定假设 ⇒ 在假设下推出与已知条件相矛盾的结果 ⇒ 否定假设

在统计中

给定假设 ⇒ 在假设下若发生了不太合理的现象 ⇒ 否定假设

假设检验:

给定假设 ⇒ 在假设下若发生了小概率事件 ⇒ 拒绝假设 (否则就保留假设)

几个概念

(一) 假设

假设: 指的是关于总体分布的命题.

在一个假设检验问题中, 常常涉及两个假设:
所要检验的假设称为原假设 (null hypothesis) 或零假设, 记为 H_0.
与 H_0 不相容的假设称为备择假设 (alternative hypothesis) 或对立假设, 记为 H_1.

给定了 H_0 与 H_1, 相当于给定了一个检验问题, 有时也记为检验问题 (H_0, H_1).

- 参数假设是指对总体分布中的未知参数作假设, 如: $H_0: \theta \in \Theta_0 \leftrightarrow H_1: \theta \in \Theta_1$; 除此以外的其它假设均称为非参数假设.
- 对于参数假设检验, 分为简单假设 (假设空间是单元素集) 和复杂假设.

注:

1. 原假设 H_0 和备择假设 H_1 一定不相容, 但不一定互补;
2. 原假设 H_0 是不易被否定的, 原假设和备择假设对于一定立场下不可以互换:
   - 原假设: 研究者想收集证据予以反对的假设;
   - 备择假设: 研究者想收集证据予以支持的假设;
3. (不成文规定) 参数假设检验问题中, 原假设 H_0 中一般含有 "=".

疑问(样本所产生的)观点常作为 H_1

(二) 检验

检验: 给出一个规则, 凭此规则, 在有了样本观测值之后, 就可以作出接受 (保留) 还是拒绝原假设 H_0 的判断, 这样的规则称为检验.

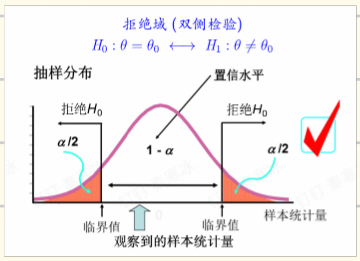
检验的实质: 给出了样本空间 Ω 的一个分划, 即将样本空间分成了两个互补的集合 W 和 A:

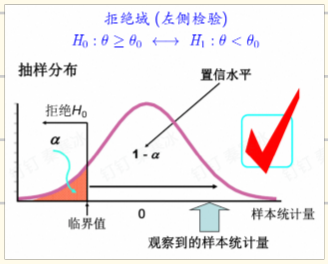
$$W \cup A = \Omega \quad (W \cap A = \varnothing)$$

当样本 $\boldsymbol{x} \in W$ 时, 就拒绝 H_0, 称 W 为拒绝域 (或否定域);
当样本 $\boldsymbol{x} \in A$ 时, 就接受 H_0, 称 A 为接受域. 这就是检验的规则.

为了确定拒绝域, 需要构造一个统计量. 称能从样本空间中划分出拒绝域的统计量为检验统计量.

给定一个检验
⇕
给出一个拒绝域
⇑
给出检验统计量





	H_0为真	H_0为假
$\tilde{x} \in W$	× I(弃真)	✓
$\tilde{x} \notin W$ ($\tilde{x} \in A$)	✓	× II(取伪).

(三) 两类错误

由于样本的随机性, 任一检验规则在应用时, 都有可能发生错误的判断.

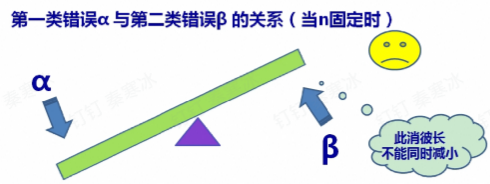
第 I 类错误 (弃真): 拒绝真实的原假设. 该错误发生的概率称为犯第 I 类错误的概率, 或弃真概率, 常记为 $\alpha(\theta)$, 即

$$\alpha(\theta) = \mathrm{P}\{\text{第 } I \text{ 类错误}\} = \mathrm{P}\{\text{拒绝 } H_0 \mid H_0 \text{ 为真}\} = \mathrm{P}\{\boldsymbol{X} \in W \mid H_0 \text{ 为真}\}.$$

第 II 类错误 (取伪/存伪): 接受错误的原假设. 该错误发生的概率称为犯第 II 类错误的概率, 或取伪概率, 常记为 $\beta(\theta)$, 即

$$\beta(\theta) = \mathrm{P}\{\text{第 } II \text{ 类错误}\} = \mathrm{P}\{\text{接受 } H_0 \mid H_0 \text{ 为假}\} = \mathrm{P}\{\boldsymbol{X} \in A \mid H_0 \text{ 为假}\}.$$

两类错误的关系: 当 n 固定时, α 越小, β 就越大; α 越大, β 就越小.



采用 Neyman-Pearson(N-P) 原则 (简称 NP 原则) 进行 "妥协".

(四) 显著性水平为 α 的检验

控制犯第 I 类错误的概率不超过给定的一个数 $\alpha(0<\alpha<1)$, 即要求

$$\sup_{\theta \in \Theta_0} \alpha(\theta) \leq \alpha.$$

为了兼顾犯第 II 类错误的概率, 通常取检验的临界点使得

$$\sup_{\theta \in \Theta_0} \alpha(\theta) = \alpha \text{ 或尽量地接近 } \alpha.$$

控制犯第 I 类错误的概率不超过 α 的检验, 称为显著性水平为 α 的检验.

注:

- 显著性水平 (significance level) 是事先由研究者指定的犯第 I 类错误的最大允许值, 通常记为 α, $0<\alpha<1$.
- 常用的 α 值有 0.01, 0.05, 0.10;
- 拒绝原假设, 表明检验的结果是显著的;
  不拒绝原假设, 表明检验的结果是不显著的.

处理假设检验问题的基本步骤 (一)

1. 根据问题提出原假设 H_0 和备择假设 H_1;
2. 确定检验统计量 $T(\boldsymbol{X})$, 并根据原假设和备择假设确定拒绝域的形式 (W 的形式主要依赖于备择假设, 且 W 中一般含有 "="):
   - 单侧拒绝域: $W=\{\boldsymbol{x}: T(\boldsymbol{x})\le c\}$ 或 $W=\{\boldsymbol{x}: T(\boldsymbol{x})\ge c\}$;
   - 双侧拒绝域: $W=\{\boldsymbol{x}: c_1\le T(\boldsymbol{x})\le c_2\}$ 或 $W=\{\boldsymbol{x}: |T(\boldsymbol{x})|\le c\}$ 或 $W=\{\boldsymbol{x}: T(\boldsymbol{x})\le c_1$ 或 $T(\boldsymbol{x})\ge c_2\}$ 或 $W=\{\boldsymbol{x}: |T(\boldsymbol{x})|\ge c\}$;
3. 选取适当的显著性水平 α, 根据 N-P 原则求出拒绝域的临界值. 一般求 c, 使得
$$\sup_{H_0\text{ 下}} \mathsf{P}(\boldsymbol{X}\in W\,|\,H_0\text{ 为真})\le\alpha \text{ 并尽可能地接近 } \alpha.$$
   在总体为连续型随机变量时, 往往要使得 $\sup\limits_{H_0\text{ 下}} \mathsf{P}(\boldsymbol{X}\in W\,|\,H_0\text{ 为真})=\alpha$;
4. 根据样本的观测值 $\boldsymbol{x}$ 算出检验统计量 $T(\boldsymbol{X})$ 的值 $T(\boldsymbol{x})$, 并与临界点进行比较. 若观察值 $\boldsymbol{x}$ 落入拒绝域 W, 则拒绝原假设, 否则接受原假设.

1、正态总体均值 μ 的假设检验

(A) σ^2 已知 (1). $H_0: \mu=\mu_0 \longleftrightarrow H_1: \mu>\mu_0$ $\bar{X}$ 是 μ 很好的估计量.

拒绝域 $W=\{x: \bar{x}\ge C\}=\{(x_1,x_2,\cdots,x_n): \bar{x}\ge C\}$ C为待定的临界值 $\begin{cases} P(I)=P(\bar{x}\ge C\mid \mu=\mu_0)=1-\Phi(\frac{C-\mu_0}{\sigma/\sqrt{n}}) \\ =P(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\ge C'\mid \mu=\mu_0) \end{cases}$ $\frac{C-\mu_0}{\sigma/\sqrt{n}}=Z_\alpha$ $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\ge Z_\alpha$ 即 $\bar{x}\ge\mu_0+\frac{\sigma}{\sqrt{n}}Z_\alpha$

$=\alpha$

$W=\{\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\ge C'\}$ $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\sim N(\frac{\mu-\mu_0}{\sigma/\sqrt{n}},1)$.

取 $Z=Z(x)=\underbrace{\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}}_{\text{取的检验统计量}}\overset{\mu=\mu_0}{\sim}N(0,1)$. 拒绝域: $W=\{Z(x)\ge Z_\alpha\}=(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\ge Z_\alpha)$

(2). $H_0: \mu=\mu_0 \longleftrightarrow H_1: \mu<\mu_0$ 时, 检验是否小于 μ_0, 当 $\bar{x}<<\mu_0$ 时才拒绝 H_0 $\therefore$ $Z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ $W=\{Z\le C\}$ $\alpha=P(Z\le C\mid H_0\text{为真})=\Phi(C)$.

(3). $H_0: \mu=\mu_0 \longleftrightarrow H_1: \mu\ne\mu_0$ 时, 此时拒绝域为 $W=\{|Z|\ge C\}$, $P_\mu(|Z|\ge C\mid \mu=\mu_0)=P_{\mu_0}(|Z|\ge C)=2(1-\Phi(C))$.

eg. **例:** 某洗涤剂厂有一台瓶装洗洁精的灌装机, 在生产正常时, 每瓶洗洁精的净重服从正态分布, 均值为 454g, 标准差为 12g. 为检查近期机器工作是否正常, 从中抽出 16 瓶, 称得其净重的平均值为 $\bar{x}=456.64$g, 试对机器工作正常与否作出判断 (取 $\alpha=0.01$, 并假定分布类型和 σ 不变).

解一: 设目前工作状态下每瓶洗洁精的净重服从 $N(\mu,12^2)$, 则这里需检验的假设为:
$$H_0:\mu=454\longleftrightarrow H_1:\mu\ne454$$
取检验统计量为
$$Z=\frac{\overline{X}-454}{12/\sqrt{16}}.$$
则拒接域为 $W=\{|z|\ge Z_{\alpha/2}\}$. 在 $\alpha=0.01$ 时, $Z_{\alpha/2}=Z_{0.005}=2.575$, 从而拒绝域为
$$W=\{|z|\ge 2.575\}.$$
现在由样本求得检验统计量的值
$$z=\frac{456.64-454}{12/4}=0.88,$$
由于 $|z|<2.58$, 所以不能拒绝 H_0, 从而认定机器工作是正常的.

补: 若已知 $\mu=450$, 求上述检验犯第二类错误的概率.
$$\mathsf{P}(\text{取伪})=\mathsf{P}(\boldsymbol{X}\in A|H_0\text{为假})=\mathsf{P}(|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}|<2.575|\mu=450)$$
注意到在 $\mu=450$ 时, $\overline{X}\overset{\mu=450\text{时}}{\sim}N(450,\frac{\sigma^2}{n})$, 故
$$\begin{aligned}\mathsf{P}(\text{取伪})&=\mathsf{P}(-2.575<\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}<2.575|\mu=450)\\&=\mathsf{P}(-2.575+\frac{\mu_0-450}{\sigma/\sqrt{n}}<\frac{\overline{X}-450}{\sigma/\sqrt{n}}<2.575+\frac{\mu_0-450}{\sigma/\sqrt{n}}|\mu=450)\\&=\Phi(2.575+\frac{454-450}{12/\sqrt{16}})-\Phi(-2.575+\frac{454-450}{12/\sqrt{16}})\\&\approx1-\Phi(-1.242)=\Phi(1.242)=0.893.\end{aligned}$$

假设检验问题:
$$\begin{aligned}H_0:\mu\ge\mu_0&\longleftrightarrow H_1:\mu<\mu_0, &(1)\\H_0:\mu\le\mu_0&\longleftrightarrow H_1:\mu>\mu_0, &(2)\\H_0:\mu=\mu_0&\longleftrightarrow H_1:\mu\ne\mu_0, &(3)\end{aligned}$$
(1) → 与 $H_0: \mu=\mu_0 \leftrightarrow H_1: \mu<\mu_0$ 的 W 相同
(2) → 与 $H_0: \mu=\mu_0 \leftrightarrow H_1: \mu>\mu_0$ 的 W 相同

其中 μ_0 为一个已知常数.
(1) 和 (2) 是单边假设检验问题, 分别为左边检验和右边检验; (3) 是双边假设检验问题.

正态总体均值 μ 的假设检验 (σ^2 已知时): 采用 Z 检验
取 $Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 为检验统计量.
(1) $H_0:\mu\le\mu_0\longleftrightarrow H_1:\mu>\mu_0$ ($H_0:\mu=\mu_0\longleftrightarrow H_1:\mu>\mu_0$) 的水平为 α 的检验的拒绝域为 $W=\{z\ge Z_\alpha\}$.
(2) $H_0:\mu\ge\mu_0\longleftrightarrow H_1:\mu<\mu_0$ ($H_0:\mu=\mu_0\longleftrightarrow H_1:\mu<\mu_0$) 的水平为 α 的检验的拒绝域为 $W=\{z\le -Z_\alpha\}$.
(3) $H_0:\mu=\mu_0\longleftrightarrow H_1:\mu\ne\mu_0$ 的水平为 α 的检验的拒绝域为 $W=\{|z|\ge Z_{\alpha/2}\}$.

(B). σ^2 未知时

ps. σ^2 已知时, 检验统计量为 $Z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ 思考: σ^2 未知时, 用 σ^2 的无偏估计: $S^2=\frac{1}{n-1}\sum_{i=1}^n(x_i-\bar{x})^2$

代替 σ^2 得检验统计量为: $t=t(x)=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ (因 $\frac{\bar{x}-\mu}{s/\sqrt{n}}\sim t(n-1)$)

$\therefore$ $H_0: \mu\le\mu_0 \longleftrightarrow H_1: \mu>\mu_0$ 的水平为 α 的检验的拒绝域为 $W=\{t\ge t_\alpha(n-1)\}$

$H_0: \mu=\mu_0 \longleftrightarrow H_1: \mu\ne\mu_0$ 的水平为 α 的检验拒绝域为 $W=\{|t|\ge t_{\alpha/2}(n-1)\}$ t-检验法

2、正态总体方差σ^2的假设检验

假设检验问题:

$$H_0:\sigma^2\ge\sigma_0^2\longleftrightarrow H_1:\sigma^2<\sigma_0^2, \quad (1)$$
$$H_0:\sigma^2\le\sigma_0^2\longleftrightarrow H_1:\sigma^2>\sigma_0^2, \quad (2)$$
$$H_0:\sigma^2=\sigma_0^2\longleftrightarrow H_1:\sigma^2\ne\sigma_0^2, \quad (3)$$

其中 σ_0 为一个已知正常数.

(A)、μ已知 : $H_0:\sigma^2\ge\sigma_0^2\longleftrightarrow H_1:\sigma^2<\sigma_0^2$

取 $C=\frac{\sigma_0^2\chi^2_{1-\alpha}(n)}{n}$, 得问题 (1) 的显著性水平为 α 的检验的拒绝域为

$$W=\{\boldsymbol{x}:\sum_{i=1}^n(x_i-\mu)^2/\sigma_0^2\le\chi^2_{1-\alpha}(n)\}.$$

也可取

$$\chi^2=\chi^2(\boldsymbol{X})=\frac{\sum_{i=1}^n(X_i-\mu)^2}{\sigma_0^2}$$

作为检验统计量, 那么拒绝域为

$$W=\{\boldsymbol{x}:\chi^2(\boldsymbol{x})\le\chi^2_{1-\alpha}(n)\},$$

简记为 $\{\chi^2\le\chi^2_{1-\alpha}(n)\}$.

类似地, 假设检验问题

$$H_0:\sigma^2\le\sigma_0^2\longleftrightarrow H_1:\sigma^2>\sigma_0^2 \quad (2)$$

的显著性水平为 α 的检验的拒绝域为

$$W=\{\chi^2\ge\chi^2_\alpha(n)\}.$$

假设检验问题

$$H_0:\sigma^2=\sigma_0^2\longleftrightarrow H_1:\sigma^2\ne\sigma_0^2 \quad (3)$$

的显著性水平为 α 的检验的拒绝域为

$$W=\{\chi^2\le\chi^2_{1-\alpha/2}(n)\ \text{或}\ \chi^2\ge\chi^2_{\alpha/2}(n)\}.$$

这些检验称为χ^2-检验.

(B). μ未知时 : 检验统计量 : $\chi^2=\chi^2(X)=\frac{\sum_{i=1}^n(X_i-\mu)^2}{\sigma_0^2}$ (μ未知, 用无偏估计$\bar{X}$代入)

$\chi^2=\chi^2(X)=\frac{(n-1)S^2}{\sigma_0^2}$

类似地, 假设检验问题

$$H_0:\sigma^2\le\sigma_0^2\longleftrightarrow H_1:\sigma^2>\sigma_0^2 \quad (2)$$

的显著性水平为 α 的检验的拒绝域为

$$W=\{\chi^2\ge\chi^2_\alpha(n-1)\}.$$

假设检验问题

$$H_0:\sigma^2=\sigma_0^2\longleftrightarrow H_1:\sigma^2\ne\sigma_0^2 \quad (3)$$

的显著性水平为 α 的检验的拒绝域为

$$W=\{\chi^2\ge\chi^2_{\alpha/2}(n-1)\ \text{或}\ \chi^2\le\chi^2_{1-\alpha/2}(n-1)\}.$$

这些检验也称为χ^2-检验.

二、两个正态总体

$X_1,X_2,\cdots,X_m$ 是来自正态总体 $N(\mu_X,\sigma_X^2)$ 的样本, $Y_1,Y_2,\cdots,Y_n$ 是来自正态总体 $N(\mu_Y,\sigma_Y^2)$ 的样本, 且两个样本相互独立. 并记

$$\overline{X}=\frac{1}{m}\sum_{i=1}^m X_i,\quad \overline{Y}=\frac{1}{n}\sum_{j=1}^n Y_j.$$

$$S_X{}^2=\frac{1}{m-1}\sum_{i=1}^m(X_i-\overline{X})^2,\quad S_Y{}^2=\frac{1}{n-1}\sum_{j=1}^n(Y_j-\overline{Y})^2.$$

1、比较均值的假设检验

假设检验问题:

$$H_0:\mu_X\ge\mu_Y\longleftrightarrow H_1:\mu_X<\mu_Y, \quad (1)$$
$$H_0:\mu_X\le\mu_Y\longleftrightarrow H_1:\mu_X>\mu_Y, \quad (2)$$
$$H_0:\mu_X=\mu_Y\longleftrightarrow H_1:\mu_X\ne\mu_Y. \quad (3)$$

(A). σ_X^2与σ_Y^2均已知

$\bar{X}-\bar{Y}\sim N(\mu_X-\mu_Y,\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n})$. 取 $Z=\frac{\bar{X}-\bar{Y}}{\sqrt{\sigma_X^2/m+\sigma_Y^2/n}}$

$$H_0:\mu_X\ge\mu_Y\longleftrightarrow H_1:\mu_X<\mu_Y, \quad (1)$$
$$W=\{(\boldsymbol{x},\boldsymbol{y}):z\le-z_\alpha\}$$

$$H_0:\mu_X\le\mu_Y\longleftrightarrow H_1:\mu_X>\mu_Y, \quad (2)$$
$$W=\{(\boldsymbol{x},\boldsymbol{y}):z\ge z_\alpha\}$$

$$H_0:\mu_X=\mu_Y\longleftrightarrow H_1:\mu_X\ne\mu_Y, \quad (3)$$
$$W=\{(\boldsymbol{x},\boldsymbol{y}):z\le-z_{\alpha/2}\ \text{或}\ z\ge z_{\alpha/2}\}$$

(B)、σ_X^2与σ_Y^2均未知: $\sigma_X^2=\sigma_Y^2=\sigma^2$时. $t=\frac{\bar{X}-\bar{Y}}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}$ 且 $\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(m+n-2)$

取 t 为检验统计量, 得

$$H_0:\mu_X\ge\mu_Y\longleftrightarrow H_1:\mu_X<\mu_Y, \quad (1)$$
$$W=\{(\boldsymbol{x},\boldsymbol{y}):t\le-t_\alpha(n+m-2)\}$$

$$H_0:\mu_X\le\mu_Y\longleftrightarrow H_1:\mu_X>\mu_Y, \quad (2)$$
$$W=\{(\boldsymbol{x},\boldsymbol{y}):t\ge t_\alpha(n+m-2)\}$$

$$H_0:\mu_X=\mu_Y\longleftrightarrow H_1:\mu_X\ne\mu_Y, \quad (3)$$
$$W=\{(\boldsymbol{x},\boldsymbol{y}):t\le-t_{\alpha/2}(n+m-2)\ \text{或}\ t\ge t_{\alpha/2}(n+m-2)\}$$

2、比较方差的假设检验

假设检验问题:

$$H_0: \sigma_X^2 \geq \sigma_Y^2 \longleftrightarrow H_1: \sigma_X^2 < \sigma_Y^2,$$
$$H_0: \sigma_X^2 \leq \sigma_Y^2 \longleftrightarrow H_1: \sigma_X^2 > \sigma_Y^2,$$
$$H_0: \sigma_X^2 = \sigma_Y^2 \longleftrightarrow H_1: \sigma_X^2 \neq \sigma_Y^2.$$

μ_X 和 μ_Y 都未知.

这时 σ_X^2 和 σ_Y^2 的无偏估计量分别为

$$S_X{}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})^2, \quad S_Y{}^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y})^2.$$

取

$$F = \frac{S_X{}^2}{S_Y{}^2} = \frac{\sum_{i=1}^{m}(X_i - \overline{X})^2/(m-1)}{\sum_{j=1}^{n}(Y_i - \overline{Y})^2/(n-1)}$$

作检验统计量. 且

$$\frac{F}{\sigma_X^2/\sigma_Y^2} = \frac{S_X{}^2/\sigma_X^2}{S_Y{}^2/\sigma_Y^2} \sim F(m-1, n-1).$$

$$H_0: \sigma_X^2 \geq \sigma_Y^2 \longleftrightarrow H_1: \sigma_X^2 < \sigma_Y^2 \tag{1}$$
$$\{F \leq F_{1-\alpha}(m-1, n-1)\};$$

$$H_0: \sigma_X^2 \leq \sigma_Y^2 \longleftrightarrow H_1: \sigma_X^2 > \sigma_Y^2 \tag{2}$$
$$\{F \geq F_{\alpha}(m-1, n-1)\};$$

$$H_0: \sigma_X^2 = \sigma_Y^2 \longleftrightarrow H_1: \sigma_X^2 \neq \sigma_Y^2 \tag{3}$$
$$\{F \leq F_{1-\alpha/2}(m-1, n-1) \text{ 或 } F \geq F_{\alpha/2}(m-1, n-1)\}.$$

这些检验均称为F-检验法.

检验的 p 值 (p-value)

假设检验的结论通常是: 在给定的显著性水平下, 不是拒绝原假设就是保留原假设.

然而有时也会出现这样的情况: 在一个较大的显著性水平 ($\alpha = 0.05$) 下得到拒绝原假设的结论, 而在一个较小的显著性水平 ($\alpha = 0.01$) 下却会得到相反的结论.

这种情况在理论上很容易解释:
因为显著性水平变小后会导致检验的拒绝域变小, 于是原来落在拒绝域中的观测值就可能落入接受域.

这种情况在应用中会带来一些麻烦: 假如这时一个人主张选择显著性水平 $\alpha = 0.05$, 而另一个人主张选显著性水平 $\alpha = 0.01$, 则对于同一组样本, 可能第一个人的结论是拒绝 H_0, 而后一个人的结论是接受 H_0.

引进检验的 P_ 值有明显的好处:

- 它比较客观, 避免了事先确定显著性水平;
- 由检验的 P_ 值与人们心中的显著性水平 α 进行比较, 可以很容易作出检验的结论:
  如果 $p \leq \alpha$, 则在显著性水平 α 下拒绝 H_0;
  如果 $p > \alpha$, 则在显著性水平 α 下保留 H_0.

P_值.

$W = \{T(\bar{X}) \geq C\}$

服从的分布.

检验统计量的值.

Example 1

例: 一支香烟中的尼古丁含量 X 服从正态分布 $N(\mu, 1)$, 质量标准规定 μ 不能超过 1.5 毫克. 现从某厂生产的香烟中随机抽取 20 支, 测得其中平均每支香烟的尼古丁含量为 $\overline{x} = 1.97$ 毫克, 试问该厂生产的香烟尼古丁含量是否超过了质量标准的规定?

解: 这是一个假设检验问题

①. $$H_0: \mu \leq 1.5 \leftrightarrow H_1: \mu > 1.5.$$

现在 $\sigma = 1$ 已知, 故采用 Z 检验, 取检验统计量为

$$Z(\boldsymbol{X}) = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X} - 1.5}{1/\sqrt{20}},$$

②.

拒绝域为

$$W = \{Z(\boldsymbol{x}) \geq z_\alpha\} = \{\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \geq z_\alpha\} =: \{z \geq z_\alpha\}.$$

计算得, $z = Z(\boldsymbol{x}) = \frac{\overline{x}-1.5}{1/\sqrt{20}} = \frac{1.97-1.5}{1/\sqrt{20}} \approx 2.10.$

对一些不同的显著性水平, 下表列出了相应的拒绝域和检验结论.

显著性水平	拒绝域	$z = 2.10$ 对应的结论
$\alpha = 0.05$	$z \geq 1.645$	拒绝 H_0
$\alpha = 0.025$	$z \geq 1.96$	拒绝 H_0
$\alpha = 0.01$	$z \geq 2.33$	接受 H_0
$\alpha = 0.005$	$z \geq 2.58$	接受 H_0

我们看到, 对于同一个样本进行分析, 不同的显著性水平 α 可能得到不同的结论.

现在换一个角度来看, 当 $\mu = 1.5$ 时, $Z(\boldsymbol{X})$ 的分布是 $N(0, 1)$. 此时可算得,

$$\mathsf{P}_{\mu=1.5}(Z(\boldsymbol{X}) \geq 2.10) = \mathsf{P}_{\mu=1.5}(\frac{\overline{X} - 1.5}{1/\sqrt{20}} \geq 2.10) = \mathsf{P}(N(0, 1) \geq 2.10) = 1 - \Phi(2.10) = 0.0179.$$

若以 0.0179 为基准来看上述检验问题, 可得

- 当 $\alpha < 0.0179$ 时, $z_\alpha > 2.10$. 于是样本就不在 $\{z \geq z_\alpha\}$ 中, 此时应接受原假设 H_0;
- 当 $\alpha \geq 0.0179$ 时, $z_\alpha \leq 2.10$. 于是样本就落入 $\{z \geq z_\alpha\}$ 中, 此时应拒绝原假设 H_0;

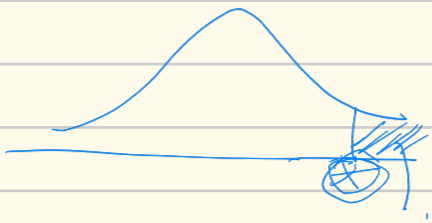
由此可以看出,0.0179 是能用此组样本 (检验统计量的值为 2.10) 做出" 拒绝 H_0" 的最小的显著性水平, 这就是P_ 值.

处理假设检验问题的基本步骤 (二)

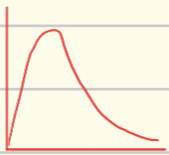
1. 根据问题提出原假设 H_0 和备择假设 H_1;
2. 确定检验统计量 $T(\boldsymbol{X})$, 并根据原假设和备择假设确定拒绝域的形式 (W 的形式主要依赖于备择假设, 且 W 中一般含有 "="):
   - 单侧拒绝域: $W = \{\boldsymbol{x}: T(\boldsymbol{x}) \leq c\}$ 或 $W = \{\boldsymbol{x}: T(\boldsymbol{x}) \geq c\}$;
   - 双侧拒绝域: $W = \{\boldsymbol{x}: c_1 \leq T(\boldsymbol{x}) \leq c_2\}$ 或 $W = \{\boldsymbol{x}: |T(\boldsymbol{x})| \leq c\}$ 或 $W = \{\boldsymbol{x}: T(\boldsymbol{x}) \leq c_1$ 或 $T(\boldsymbol{x}) \geq c_2\}$ 或 $W = \{\boldsymbol{x}: |T(\boldsymbol{x})| \geq c\}$;
3. 计算检验统计量的观测值与 P_ 值;
4. 根据给定的显著性水平 α 作出判断: 若 P_ 值小于等于 α, 则拒绝 H_0, 称检验结果在显著性水平 α 下是显著的; 若 P_ 值大于 α, 则保留 H_0.

ps. 对于双边检验: ① Z / t 检验时 P_值: $2P(\begin{matrix} N(0,1) \\ t(*) \end{matrix} \geq |值|)$.

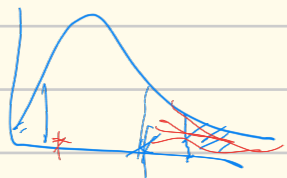
2倍!



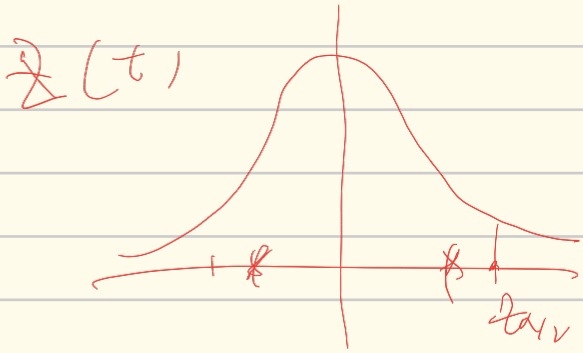
② χ^2 / F 检验时



P_值 $= 2 \times \min\{P(\begin{matrix} \chi^2(*) \\ F(*) \end{matrix} \geq 值), P(\begin{matrix} \chi^2(*) \\ F \end{matrix} \leq 值)\}$



$\chi^2_{1-\alpha/2}(n-1)$ $\chi^2_{\alpha/2}(n-1)$



P值 $\tilde{x}$ ⇒ P值 ⇒ P值 Vs. α

临界: α ⇒ C_α ⇒ 拒绝域 C_α