概率论与数理统计

眠云跂石整理

第 1 章 事件的概率

组合公式

由 $(a+b)^n=\sum\limits_{i=0}^{n}\binom{n}{i}a^ib^{-ni}$ 知

$$\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{n}=2^n$$
$$\binom{n}{0}-\binom{n}{1}-\cdots+(-1)^{-1}\binom{n}{n}=0$$

由 $(1+x)^{m+n}=(1+x)^m(1+x)^n$ 知

$$\binom{m+n}{k}=\sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i}$$

特别地, 当 $m=k=n$ 时,

$$\binom{2n}{n}=\sum_{i=0}^{n}\binom{n}{i}^2$$

第一式还可以写作下式, 并可以从直观上理解.

$$\sum_{k_1+k_2=k}\binom{n_1}{k_1}\binom{n_2}{k_2}=\binom{n_1+n_2}{k_1+k_2}$$

此外还有下式, 也可以从直观上理解.

$$\sum_{n_1+n_2=n}\binom{n_1}{k_1}\binom{n_2}{k_2}=\binom{n_1+n_2+1}{k_1+k_2+1}$$

多项式系数: $\dfrac{n!}{r_1!\cdots r_k!}$.

利用第一式 (杨辉恒等式) 数归得第二式 (或直观理解)

$$\binom{n+m}{m}+\binom{n+m}{m+1}=\binom{n+m+1}{m+1}$$

$$\sum_{r=0}^{m}\binom{n-1+r}{r}=\binom{n+m}{m}$$

由负指数二项展开式 $(1-x)^{-r}=\sum_{i=0}^{\infty}\binom{-r}{i}(-x)^i=\sum_{i=0}^{\infty}\binom{i+r-1}{r-1}x^i$ 知

$$\sum_{i=0}^{\infty}\binom{i+r-1}{r-1}=0$$

$$\sum_{i=0}^{\infty}\binom{i+r-1}{r-1}(-1)^i=2^{-r}$$

$$p^{-r}=\sum_{i=0}^{\infty}\binom{i+r-1}{r-1}p^r(1-p)^i$$

原式两边求导, 并令 $x=1-p$, 得

$$rp^{-r-1}=\sum_{i=0}^{\infty}i\binom{i+r-1}{r-1}(1-p)^{i-1}$$

[Stirling 数, 拆分数, 装箱问题, Burnside 定理与 Polya 定理]

事件的运算

- 记号
  - $A+B\equiv A\cup B$.
  - $AB\equiv A\cap B$.
  - $A-B\equiv A\overline{B}$.
- 加法
  - 交换律: $A+B=B+A$.
  - 结合律: $(A+B)+C=A+(B+C)$.

    于是可定义 $A+B+C=(A+B)+C$.
  - 自加: $A+A=A$.
- 乘法
  - 交换律: $AB=BA$.
  - 结合律: $(AB)C=A(BC)$.

    于是可定义 $ABC=(AB)C$.
  - 自乘: $AA=A$.
- 分配律
  - 加法与乘法: $(A+B)C=AC+BC$.
  - 减法与乘法: $(A-B)C=AC-BC$.

本质为 $ABC=(AC)(BC)$, 这个式子在推导中是有用的.

- 减法
  - $A-B=A\overline{B}\neq A+(-B)$.
  - $A\subseteq B\Leftrightarrow A-B=\varnothing$.
  - $A=B\Leftrightarrow A-B=B-A=\varnothing$.
- 无消去律
  - $A+B=A+C\nRightarrow B=C$.
  - $A-B=A-C\nRightarrow B=C$.
- 混合运算
  - $(A+B)-C\neq A+(B-C)$. (因为减法本质上是乘法)

    因此 $A+B-C$ 没有意义, 除非定义运算顺序或优先级.
  - $A-(B+C)=A-B-C=(A-C)-(B-C)$.
  - $A-(B-C)=(A-B)C=AC-BC$.
- 负号 (补集)
  - $-(A+B)=(-A)(-B)$.
  - $-(A-B)=B-A=(-A)B$.
  - $-(AB)=(-A)+(-B)$.

    理论上可以这么写, 实际上用 $\overline{A}$ 的符号会更方便.
- 互斥
  - A 与 B 互斥 $\quad\Leftrightarrow\quad AB=\varnothing\quad\Leftrightarrow\quad P(AB)=0$.
  - $AC=BC\quad\Leftrightarrow\quad A-B$ 与 $B-A$ 均与 C 互斥 $\quad\Leftrightarrow\quad P(\overline{A}BC)=P(A\overline{B}C)=0$.

    当且仅当 $C=\Omega$ 时, 可由此推出 $A=B$.
  - A 与 B 互斥 $\quad\Rightarrow\quad AC$ 与 BC 互斥 $\quad\Rightarrow\quad P(C(A+B))=P(AC)+P(BC)$.
- 对立
  - A 与 B 对立 $\quad\Leftrightarrow\quad AB=\varnothing$ 且 $A+B=\Omega$.
  - $\overline{A_1A_2\cdots A_n}=\overline{A}_1+\overline{A}_2+\cdots+\overline{A}_n$.
  - $\overline{A_1+A_2+\cdots+A_n}=\overline{A}_1\overline{A}_2\cdots\overline{A}_n$.
- 条件概率
  - 定义：$P(A\mid B)=P(AB)/P(B)$.
  - 全概率公式：若两两互斥的 B_i 之交为必然事件, 则 $P(A)=P(B_1)P(A\mid B_1)+P(B_2)P(A\mid B_2)+\cdots$.
  - 贝叶斯公式:
    - $P(B\mid A)=\dfrac{P(B)P(A\mid B)}{P(A)}$.
    - $P(B_i\mid A)=\dfrac{P(AB_i)}{P(A)}=\dfrac{P(B_i)P(A\mid B_i)}{\sum P(B_j)P(A\mid B_j)}$.
- 几率
  - 定义：$O(A)=\dfrac{P(A)}{1-P(A)}=\dfrac{P(A)}{P(\overline{A})}$.
  - 贝叶斯公式：$O(B\mid A)=\dfrac{P(B\mid A)}{P(\overline{B}\mid A)}=\dfrac{P(B)P(A\mid B)}{P(A)P(\overline{B}\mid A)}=\dfrac{P(B)P(A\mid B)}{P(\overline{B})P(A\mid \overline{B})}=O(B)\dfrac{P(A\mid B)}{P(A\mid \overline{B})}$.
  - 贝叶斯因子：$\mathrm{BF}=\dfrac{P(A\mid B)}{P(A\mid \overline{B})}$, 故 $O(B\mid A)=\mathrm{BF}\cdot O(B)$.
- 促进作用的性质
  - 具有对称性：A 促进 B, 则 B 促进 A, 即

    $P(A\mid B)>P(A)\quad\Leftrightarrow\quad P(B\mid A)>P(B)$.
  - 不具有传递性：B 促进 A 且 C 促进 B 不能推出 C 促进 A, 即

    $P(A\mid B)>P(A),\ P(B\mid C)>P(B)\quad\nRightarrow\quad P(A\mid C)>P(A)$.
  - 若 B 和 C 都促进 A, 则 $B+C$ 一定促进 A, 但 BC 和 $B-C$ 不一定促进 A, 即

    $P(A\mid B)>P(A),\ P(A\mid C)>P(A)\quad\Leftrightarrow\quad P(A\mid B+C)>P(A)$.
  - 若 B 促进 A, 则 $\overline{B}$ 抑制 A, B 抑制 $\overline{A}$, $\overline{B}$ 促进 $\overline{A}$, 即

$$P(A\mid B)>P(A)\quad\Leftrightarrow\quad P(A\mid\overline{B})<P(A)\quad\Leftrightarrow\quad P(\overline{A}\mid B)<P(\overline{A})\quad\Leftrightarrow\quad P(\overline{A}\mid\overline{B})>P(\overline{A}).$$
$$P(A\mid B)=P(A)\quad\Leftrightarrow\quad P(A\mid\overline{B})=P(A)\quad\Leftrightarrow\quad P(\overline{A}\mid B)=P(\overline{A})\quad\Leftrightarrow\quad P(\overline{A}\mid\overline{B})=P(\overline{A}).$$

- 独立
  - A 与 B 独立 $\quad\Leftrightarrow\quad P(AB)=P(A)P(B)\quad\Rightarrow\quad P(A\mid B)=P(A).$
  - 两两独立: $\forall i,j(1\le i,j\le n,i\ne j):P(A_iA_j)=P(A_i)P(A_j).$
  - 相互独立: $\forall 1<k\le n,\ 1\le i_1<i_2<\cdots<i_k\le n:P(A_{i_1}A_{i_2}\cdots A_{i_k})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k}).$
  - 相互独立 $\Rightarrow$ 两两独立, 反之不一定成立.
  - 独立事件的任一部分也独立.
  - 若 $A_1,A_2,\cdots,A_n$ 独立, $B_i=A_i$ 或 $\overline{A}_i$, 则 $B_1,B_2,\cdots,B_n$ 也独立.
- 独立事件的概率
  - 乘法: $P(\prod E_i)=\prod P(E_i).$
  - 加法: $P(\sum E_i)=1-P(\prod\overline{E_i})=1-\prod P(\overline{E_i}).$
  - 实例:

  $$\begin{aligned}P(E_0+E_1E_2)&=1-P(\overline{E}_0\overline{E_1E_2})=1-(1-P(E_0))(1-P(E_0E_1))\\&=P(E_0)+P(E_1)P(E_2)-P(E_0)P(E_1)P(E_2).\end{aligned}$$
- 运算定理
  - 加法定理 (并集): $P(A+B)=P(A)+P(B)\quad\Leftrightarrow\quad AB=\varnothing\quad\Leftrightarrow\quad P(AB)=0.$
  - 减法定理 (差集): $P(A-B)=P(A)-P(B)\quad\Leftrightarrow\quad A\supseteq B\quad\Leftrightarrow\quad P(B-A)=0.$
  - 乘法定理 (交集): $P(AB)=P(A)P(B)\quad\Leftrightarrow\quad A$ 与 B 独立.
  - 加法推论 (补集): $P(A^c)=P(\overline{A})=1-P(A)$ 恒成立.
- 表为互斥事件
  - $\sum A_i=A_1+\overline{A}_1A_2+\cdots+\overline{A}_1\overline{A}_2\cdots\overline{A}_{n-1}A_n.$
    - $A+B=A+(B-A).$
    - $A+B+C=A+(B-A)+(C-B-A).$
  - 设 $f(m)=\sum\limits_{k,l}\left(\prod\limits_{i=1}^{m}A_{k_i}\prod\limits_{j=1}^{n-m}\overline{A}_{l_j}\right)$, 则 $\sum\limits_{i=1}^{n}A_i=\sum\limits_{m=1}^{n}f(m).$
    - $A+B=(B-A)+(A-B)+AB.$
    - $A+B+C=ABC+(BC-A)+(AC-B)+(AB-C)+(A-B-C)+(B-A-C)+(C-A-B)$.
  - 综合应用
    - $A+B=A+(B-A)=(A-B)+B=(B-A)+(A-B)+AB,$

      即 $A+B=A+\overline{A}B=\overline{A}B+A\overline{B}+AB.$
    - $P(A+B)=P(A)+P(\overline{A}B)=P(\overline{A}B)+P(A\overline{B})+P(AB).$
- 容斥原理
  - $P(A+B)=P(A)+P(B)-P(AB)$, 或 $P(AB)=P(A)+P(B)-P(A+B).$
  - $P(\overline{AB})=P(\overline{A}+\overline{B}),\ P(\overline{AB})+P(\overline{A}\,\overline{B})=P(\overline{A})+P(\overline{B}).$
  - $P(A+B+C)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(CA)+P(ABC).$
- 恒等式
  - 化简含括号的运算
    - $(A+B)+(A-B)=A+B.$
    - $(A+B)-(A-B)=B.$
    - $(A-B)+(B-A)=(A+B)-AB.$
    - $(A-B)-(B-A)=A-B.$
  - 有用的概率恒等式
    - $A-B=A-AB$, 或 $A\overline{B}=A\overline{AB}.$
    - $P(A-B)=P(A)-P(AB)$. (利用减法定理)

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

- 欧拉装错信封：$P_n=\sum\limits_{k=0}^{n}\frac{(-1)^k}{k!}$.
- 先胜 n 局者为胜, 甲 a 胜 b 负, 则甲胜的概率为：$P_n(a,b)=\sum\limits_{i=1}^{n-b}p^{n-a-1+i}(1-p)^{n-b-i}\binom{2n-a-b-1}{n-a-1+i}$.
- 设 n 个独立事件 $A_1,A_2,\cdots,A_n$ 的概率分别为 $p_1,p_2,\cdots,p_n$, 记 $p=p_1+p_2+\cdots+p_n$, 则
  - $A_1,A_2,\cdots,A_n$ 都不发生的概率小于 e^{-p}.
  - $A_1,A_2,\cdots,A_n$ 中至少发生 k 个的概率小于 $p^k/k!$.
- 蒲丰投针问题：$p=\frac{2l}{\pi a}$.

概率的公理化定义

1. 非负性.
2. 规范性.
3. 可列可加性.

定理 1 独立事件的交与并

> 若 A 和 B 均与 C 独立, 则 AB 与 C 独立 $\Leftrightarrow A+B$ 与 C 独立.

证明

- 法一

1. 必要性

$$\begin{aligned}P((A+B)C)&=P(AC+BC)\\&=P(AC)+P(BC)-P(ABC^2)\\&=P(C)(P(A)+P(B)-P(AB))\\&=P(C)P(A+B).\end{aligned}$$

2. 充分性

$$\begin{aligned}P(ABC)&=P((AC)(BC))\\&=P(AC)+P(BC)-P((A+B)C)\\&=P(C)(P(A)+P(B)-P(A+B))\\&=P(C)P(AB).\end{aligned}$$

- 法二

1. 必要性

$$\begin{aligned}P((A+B)C)&=P((A-B)C)+P((B-A)C)+P(ABC)\\&=P(AC-ABC)+P(BC-ABC)+P(ABC)\\&=P(A)P(C)-P(AB)P(C)+P(B)P(C)-P(AB)P(C)+P(AB)P(C)\\&=P(C)(P(A)+P(B)-P(AB))\\&=P(C)P(A+B).\end{aligned}$$

2. 充分性

$$\begin{aligned}P(ABC)&=P((A+B)C)-P((A-B)C)-P((B-A)C)\\&=P(A+B)P(C)-P(AC-ABC)-P(BC-ABC)\\&=P(A+B)P(C)-P(A)P(C)+P(ABC)-P(B)P(C)+P(ABC)\\&=P(AC)+P(BC)-P((A+B)C)\\&=P(C)(P(A)+P(B)-P(A+B))\\&=P(C)P(AB).\end{aligned}$$

推论　增加互斥条件的充分条件

> 若 A 和 B 均与 C 独立, 且 A 与 B 互斥, 则 AB 与 $A+B$ 均与 C 独立.

定理 2 相互独立的充要条件

> 设 $0<P(A)<1$, 则 $P(B\mid A)=P(B\mid\overline{A})$ 是事件 A,B 相互独立的充要条件.

证明

1. 必要性

$$\begin{aligned}P(B)&=P(AB)+P(\overline{A}B)\\&=P(A)P(B\mid A)+P(\overline{A})P(B\mid\overline{A})\\&=P(B\mid A)=P(AB)/P(A).\end{aligned}$$

2. 充分性

$$\begin{aligned}P(B\mid A)&=P(AB)/P(A)=P(B)\\&=P(AB)+P(\overline{A}B)\\&=P(A)P(B\mid A)+P(\overline{A})P(B\mid\overline{A})\\&=P(B\mid\overline{A}).\end{aligned}$$

推论　相互独立的充要条件

> 设 $0<P(A)<1$, 则 $P(A\mid B)=P(\overline{A}\mid B)$ 是事件 A,B 相互独立的充要条件.

第 2 章 随机变量及概率分布

2.1 一维随机变量

常见分布更详细的信息请见笔记附录 (源代码 或 PDF)

注意各分布的定义在不同资料上可能不同, 请注意区分.

2.1.1 离散型随机变量

概率函数, 分布表,

分布函数是一个右连续的不减函数.

1 二项分布

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

理解: 事件发生的概率为 p, 则重复 n 次试验, 事件发生的次数为 x.

概率分布：$P(X=i)=b(i;n,p)=\binom{n}{i}p^i(1-p)^{n-i}$.

最可能数：$x=\lfloor(n+1)p\rfloor$.

2 泊松分布

$X\sim P(\lambda)$.

理解: 单位时间内事件平均发生 λ 次, 则某一段单位时间内发生的次数为 x.

概率分布：$P(X=i)=\lim\limits_{n\to\infty}b(i;n,\frac{\lambda}{n})=\frac{\mathrm{e}^{-\lambda}\lambda^i}{i!}$.

当二项分布满足 $n>50,p<0.1,np<5$ 时, 用泊松分布近似效果较好.

3 超几何分布

$X\sim H(N,n,M)$.

理解: N 件产品中有 M 件次品, 从总体中抽 n 件时次品的数量 m.

概率分布：$P(X=m)=\binom{M}{m}\binom{N-M}{n-m}\bigg/\binom{N}{n}$.

4 负二项分布

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

理解: 合格率为 p, 抽取到 r 个合格产品时, 抽到的不合格产品的个数 x.

概率分布：$P(X=i)=d(i;r,p)=\binom{i+r-1}{r-1}p^r(1-p)^i$.

5 几何分布

$X \sim GE(p)$.

理解: 合格率为 p, 抽取到第一个合格产品时, 抽到的不合格产品的个数 x.

概率分布：$P(X=i)=p(1-p)^i$.

几何分布具有无记忆性.

5' 几何分布

$X \sim G(p)$.

理解: 合格率为 p, 抽取到第一个合格产品时, 抽到的总产品的个数 x.

概率分布: $P(X=i)=p(1-p)^{i-1}$.

几何分布具有无记忆性.

2.1.2 连续型随机变量

概率分布函数, 概率密度函数

注：以下偏度系数定义为 $\beta_1=\mu_3/\mu_2^{3/2}$，峰度系数定义为 $\beta_2=\mu_4/\mu^2$.

1 正态分布

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

概率密度函数：$f(x)=(\sqrt{2\pi}\sigma)^{-1}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

标准正态分布：$Y=(X-\mu)/\sigma \sim N(0,1)$.

3σ 原则：0.6826，09544，9.9974.

上 α 分位数：$\mathit{\Phi}(z_\alpha)=1-\alpha$.

2 指数分布

$X \sim E(\lambda)$.

概率密度函数：$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x>0,\\ 0, & x\le 0.\end{cases}$

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

指数分布具有无记忆性，即 $P(X>m+t \mid X>m)=P(X>t)$.

3 威布尔分布

概率密度函数：$f(x)=\begin{cases}\lambda\alpha x^{\alpha-1}\mathrm{e}^{-\lambda x^\alpha}, & x>0,\\ 0, & x\le 0.\end{cases}$

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

4 均匀分布

$X \sim R(a,b)$.

概率密度函数：$f(x)=\begin{cases}1/(b-a), & a\le x\le b,\\ 0, & x<a \text{ 或 } x>b.\end{cases}$

分布函数：$F(x)=\begin{cases}0, & x\le a,\\ (x-a)/(b-a), & a<x<b,\\ 1, & x\ge b.\end{cases}$

5 对数正态分布

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

概率密度函数：$f(x,\mu,\sigma) = \begin{cases} \left(x\sqrt{2\pi}\sigma\right) \exp\left[-\dfrac{(\ln x-\mu)^2}{2\sigma^2}\right], & x>0, \\ 0, & x \le 0. \end{cases}$

6 柯西分布

$X \sim C(\gamma, x_0)$.

概率密度函数：$f(x; x_0, \gamma) = \dfrac{1}{\pi} \cdot \dfrac{\gamma}{(x-x_0)^2+\gamma^2}\ (-\infty < x < +\infty)$.

7 拉普拉斯分布

$X \sim \mathrm{La}(\mu, b)$.

概率密度函数: $f(x) = \dfrac{1}{2\lambda} \mathrm{e}^{-\frac{|x-\mu|}{\lambda}}$.

2.2 多维随机变量

常见分布更详细的信息请见笔记附录 (源代码 或 PDF)

2.2.1 离散性随机向量

1 多项分布

$X = (X_1, \cdots, X_n) \sim M(N; p_1, \cdots, p_n)$.

$$P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n) = \frac{N!}{k_1!k_2!\cdots k_n!} p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}.$$

多项分布的边缘分布是二项分布.

$(X_1, X_2, \cdots, X_n) \sim M(N; p_1, p_2, \cdots, p_n) \quad \Rightarrow \quad X_1 + X_2 \sim B(N; p_1+p_2)$.

2.2.2 连续型随机向量

1 矩形均匀分布

2 二维正态分布

$X = (X_1, X_2) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$.

$$f(x_1, x_2) = (2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-a)^2}{\sigma_1^2} - \frac{2\rho(x_1-a)(x_2-b)}{\sigma_1\sigma_2} + \frac{(x_2-b)^2}{\sigma_2^2}\right)\right].$$

当且仅当 $\rho = 0$ 时, X_1 和 X_2 独立.

其它性质

- 二维正态分布的边缘分布是正态分布.
- 二维正态分布的条件分布是正态分布.

  若 $(X, Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$, 则给定 $X = x$ 时 Y 的条件分布为

  $$N(b + \rho\sigma_2\sigma_1^{-1}(x-a),\ \sigma_2^2(1-\rho^2)).$$

- 二维正态分布的边缘分布的和仍为正态分布

  若 $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, 则 $Y = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$.

- 独立的正态分布的联合分布是正态分布.

  正态分布的联合分布不一定是二维正态分布.

- 相互独立的正态分布的和仍为正态分布

  若 $X_i \sim N(\mu_i, \sigma_i^2)$, 则 $X_1 + \cdots + X_n \sim N(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2)$.

- 若 $Y = X_1 + X_2$ 服从正态分布, X_1, X_2 独立, 则 X_1, X_2 也是正态分布. ⭐

3 多元正态分布

设 $(X_1, X_2, \cdots, X_n)$ 为 n 元随机变量, 令

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{nn} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & n_{nn} \end{pmatrix},$$

其中 $\boldsymbol{C}$ 为协方差矩阵. 如果 $(X_1, X_2, \cdots, X_n)$ 的概率密度函数为

$$f(x_1, x_2, \cdots, x_n) = \frac{\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}{(2\pi)^{\frac{n}{2}}|\boldsymbol{C}|^{\frac{1}{2}}}$$

则称 $(X_1, X_2, \cdots, X_n)$ 是参数为 $\boldsymbol{\mu}$, $\boldsymbol{C}$ 的 n 元正态变量.

其它性质

- n 维正态分布的边缘分布是正态分布.
- n 维正态分布的条件分布是正态分布.
- n 维正态分布的边缘分布的和是正态分布.
- n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布的充要条件是:

$$\forall l_i \in \mathbb{R}\ (i = 1, 2, \cdots, n) : l_1X_1 + l_2X_2 + \cdots + l_nX_n \sim N(\mu, \sigma^2).$$

- 若 $Y_1, Y_2, \cdots, Y_m$ 都是 n 维正态分布分量 $X_i\ (i = 1, 2, \cdots, n)$ 的线性函数, 则 $(Y_1, Y_2, \cdots, Y_m)$ 服从 m 维正态分布.
- n 维正态分布各分量相互对立充要条件是它们两两不相关.

2.2.3 边缘分布

1 概念解释

- 随机向量的分布可以决定其任一分量的边缘分布, 但反之不亦然.
- 随机向量也叫作其边缘分布的 **联合分布**.
- 类似的有二维的边缘分布.

2 多项分布

$(X_1, \cdots, X_n) \sim M(N; p_1, \cdots, p_n)$ 关于 X_1 的边缘分布为 $M(N, p_1)$.

3 二维正态分布

$(X_1, X_2) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$ 关于 X_1 和 X_2 的边缘分布分别是 $N(a, \sigma_1^2)$ 和 $N(b, \sigma_2^2)$.

2.3 条件概率分布与随机变量的独立性

2.3.1 条件概率分布的概念

2.3.2 离散性随机变量的条件概率分布

1 多项分布

在给定 $X_2 = k_2$ 的条件下, X_1 的条件分布为 $B(N - k_2, p_1/(1 - p_2))$.

2.3.3 连续性随机变量的条件概率分布

$$f_1(x_1 \mid a \le X_2 \le b) = \int_a^b f(x_1, t_2)\,\mathrm{d}t_2 \Big/ \int_a^b f_2(t_2)\,\mathrm{d}t_2$$

$$f(x_1, x_2) = f_2(x_2) f_1(x_1 \mid x_2)$$
$$f(x_1, \cdots, x_n) = g(x_1, \cdots x_k) h(x_{k+1}, \cdots, x_n \mid x_1, \cdots, x_k)$$

$$f_1(x_1) = \int_{-\infty}^{+\infty} f_2(x_2) f_1(x_1 \mid x_2) \mathrm{d}x_2$$

正态变量的条件分布仍为正态. 正态分布条件分布的中心位置是

$$m(x_1) = b + \rho\sigma_2\sigma_1^{-1}(x_1 - a).$$

2.3.4 随机变量的独立性

两个变量的独立 $\quad\Leftrightarrow\quad f_1(x_1) = f_1(x_1 \mid x_2).$

定义 3.1 连续型随机变量的相互独立 (独立)

> $X_1, X_2, \cdots, X_n$ 相互独立 (独立) $\quad\Leftrightarrow\quad f(x_1, \cdots, x_n) = f_1(x_1) \cdots f_n(x_n).$

定理 3.1

> 连续变量独立 $\quad\Leftrightarrow\quad$ 对应的事件独立.

定理 3.2

> 若连续型随机向量 $(X_1, X_2, \cdots, X_n)$ 的概率密度函数 $f(x_1, x_2, \cdots, x_n) = g_1(x_1) g_2(x_2) \cdots g_n(x_n)$, 则 $X_1, X_2, \cdots, X_n$ 相互独立, 且 $f_i(x_i) = C g_i(x_i)$.

定理 3.3

> 若 $X_1, X_2, \cdots, X_n$ 相互独立,
>
> $$Y_1 = g_1(X_1, X_2, \cdots, X_m),\ Y_2 = g_2(X_{m+1}, X_{m+2}, \cdots, X_n),$$
>
> 则 Y_1 和 Y_2 独立.

定义 3.2 离散性随机变量的相互独立

> $X_1, X_2, \cdots, X_n$ 相互独立 (独立) 等价于
>
> $$\forall a_1, a_2, \cdots, a_n\colon\ P(X_1 = a_1, \cdots, X_n = a_n) = P(X_1 = a_1) \cdots P(X_n = a_n).$$

示性函数

$$X = \begin{cases} 1, & \text{当事件 } A \text{ 发生时}, \\ 0, & \text{当事件 } A \text{ 不发生时}. \end{cases}$$

2.4 随机变量的函数的概率分布

2.4.1 离散性分布

1. 多项分布 $(X_1, X_2, \cdots, X_n) \sim M(N; p_1, p_2, \cdots, p_n) \quad\Rightarrow\quad X_1 + X_2 \sim B(N; p_1 + p_2).$
2. 二项分布 $X_1 \sim B(n_1, p),\ X_2 \sim B(n_2, p) \quad\Rightarrow\quad X_1 + X_2 \sim B(n_1 + n_2, p).$
3. 泊松分布 $X_1 \sim P(\lambda_1),\ X_2 \sim P(\lambda_2) \quad\Rightarrow\quad X_1 + X_2 \sim P(\lambda_1 + \lambda_2).$

2.4.2 连续型分布

1 单变量函数

1.1 严格单调

若 X 有密度函数 $f(x)$, $Y = g(X)$ 且该函数严格单调, 令 $X = h(Y)$, 则 Y 的概率密度函数为

$$l(y) = f(h(y))\,|h'(y)|.$$

- $X \sim N(\mu, \sigma^2) \quad\Rightarrow\quad aX + b \sim N(a\mu + b, a^2\sigma^2).$

1.2 幂函数

若 X 有密度函数 $f(x)$, $Y=X^n$, 其中 n 为偶数, 则 Y 的概率密度函数为

$$l(y)=\left|\frac{y^{\frac{1}{n}-1}}{n}\right|\left[f(y^{\frac{1}{n}})+f(-y^{\frac{1}{n}})\right].\quad (n\text{ 是偶数})$$

- 若 $X\sim N(0,1)$, 则 $Y=X^2$ 的密度函数为 $l(y)=\begin{cases}\left(\sqrt{2\pi y}\right)^{-1}\mathrm{e}^{-y/2}, & y>0,\\ 0, & y\le 0.\end{cases}$

2 多变量函数

以两个为例, 多变量是类似的.

$$\begin{cases}Y_1=g_1(X_1,X_2)\\ Y_2=g_2(X_1,X_2)\end{cases}\quad\Rightarrow\quad\begin{cases}X_1=h_1(Y_1,Y_2)\\ X_2=h_2(Y_1,Y_2)\end{cases}$$

则雅可比行列式为

$$J(y_1,y_2)=\begin{vmatrix}\dfrac{\partial h_1}{\partial y_1} & \dfrac{\partial h_1}{\partial y_1}\\ \dfrac{\partial h_2}{\partial y_1} & \dfrac{\partial h_2}{\partial y_1}\end{vmatrix},$$

概率密度函数

$$l(y_1,y_2)=f(h_1(y_1,y_2),h_2(y_1,y_2))\,|J(y_1,y_2)|.$$

2.4.3 随机变量和的密度函数

设 (X_1,X_2) 的联合密度函数为 $f(x_1,x_2)$, 则 $Y=X_1+X_2$ 的密度函数为

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

- 法一：固定 y 后积分得分布函数, 再对 y 求导得上式.
- 法二：补充 $Y_2=X_1$, 利用 2.4.2.2

- 二维正态分布的边缘分布的和仍为正态分布

  若 $(X_1,X_2)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 则 $Y=X_1+X_2\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)$.
- 相互独立的正态分布的和仍为正态分布

  若 $X_i\sim N(\mu_i,\sigma_i^2)$, 则 $X_1+\cdots+X_n\sim N(\mu_1+\cdots+\mu_n,\sigma_1^2+\cdots+\sigma_n^2)$.
- 若 $Y=X_1+X_2$ 服从正态分布, X_1,X_2 独立, 则 X_1,X_2 也是正态分布.

自由度为 n 的皮尔逊卡方密度与 **卡方分布** $X\sim\chi_n^2$

$$k_n(x)=\begin{cases}\dfrac{\mathrm{e}^{-x/2}x^{(n-2)/2}}{\Gamma\left(\dfrac{n}{2}\right)2^{n/2}}, & x>0,\\ 0, & x\le 0.\end{cases}$$

- 若 $X_1,X_2,\cdots,X_n$ 相互独立, 且有公共分布 $N(0,1)$ (**独立同分布 iid**), 则 $Y=X_1^2+X_2^2+\cdots+X_n^2\sim\chi_n^2$.
- 若 $X_1\sim\chi_m^2$ 与 $X_2\sim\chi_n^2$ 独立, 则 $X_1+X_2\sim\chi_{m+n}^2$.
- 若 $X_1,X_2,\cdots,X_n$ 相互独立, 且都服从指数分布 $E(\lambda)$, 则 $X=2\lambda(X_1+X_2+\cdots+X_n)\sim\chi_{2n}^2$.
- $E(\chi_n^2)=n$.
- $E(\chi_n^2)^{-1}=\dfrac{1}{n-2}$.
- $E(\chi_n^2)^k=\dfrac{2^k\Gamma\left(\dfrac{n}{2}+k\right)}{\Gamma\left(\dfrac{n}{2}\right)}\ (k\in\mathbb{Z})$.
- $\mathrm{Var}(\chi_n^2)=2n$.

  注意到方差是均值的两倍，可以以此检验是否为卡方分布. 🌙

2.4.4 随机变量商的密度函数

设 (X_1, X_2) 的联合密度函数为 $f(x_1, x_2)$, 则 $Y = X_2/X_1$ 的密度函数为

$$l(y) = \int_0^{+\infty} x_1 f(x_1, x_1 y) \,\mathrm{d}x_1.$$

- 法一：固定 y 后积分得分布函数, 再对 y 求导得上式.
- 法二：补充 $Y_2 = X_1$, 利用 2.4.2.2

设 X_1, X_2 独立, $X_1 \sim \chi_n^2$, $X_2 \sim N(0,1)$, $Y = X_2/\sqrt{X_1/n}$, 则 Y 的概率函数为

$$t_n(y) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{y^2}{n}\right)^{-\frac{n+1}{2}}.$$

称为自由度为 n 的 t **分布**.

- $E(t_n) = 0 \ (n > 1)$.
- $\operatorname{Var}(t_n) = \dfrac{n}{n-2} \ (n > 2)$.

设 X_1, X_2 独立, $X_1 \sim \chi_n^2$, $X_2 \sim \chi_m^2$, $Y = \dfrac{X_2}{m}\Big/\dfrac{X_1}{n}$, 则 Y 的概率密度函数为

$$f_{m,n}(y) = m^{m/2} n^{n/2} \frac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)} y^{m/2-1}(my+n)^{-(m+n)/2} \quad (y > 0)$$

称为自由度为 (m, n) 的 F **分布**.

- $E(f_{m,n}) = \dfrac{n}{n-2} \ (n > 2)$.
- $\operatorname{Var}(f_{m,n}) = \dfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$.

⭐ 若 $X_1, X_2, \cdots, X_n$ 独立同分布, 且有分布函数 $F(x)$ 和密度函数 $f(x)$, 则

$$Y = \max(X_1, X_2, \cdots, X_n) \sim nF^{n-1}(x) f(x),$$
$$Z = \min(X_1, X_2, \cdots, X_n) \sim n[1-F(x)]^{n-1} f(x).$$

注意事项

- 概率密度函数在某点的取值必为 0, 如果非零, 则不存在这样的概率密度函数. 即混合型随机变量没有概率密度函数.
- 计算随机变量的函数的概率分布时，注意单调性, 值域和值域是否重叠.

第 3 章 随机变量的数字特征

3.1 数学期望与中位数

3.1.1 数学期望的定义

3.1.2 数学期望的性质

定理 1.1 随机变量之和的期望

> 若干个随机变量之和的期望, 等于各变量的期望之和, 即
>
> $$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots E(X_n).$$

定理 1.2 随机变量之积的期望

> 若干个独立随机变量之积的期望, 等于各变量的期望之积, 即
>
> $$E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n).$$

定理 1.3 随机变量函数的期望

> 设随机变量 X 为离散型, 有分布 $P(X=a_i)=p_i, i=1,2,\cdots,n$, 或者为连续型, 有概率密度函数 $f(x)$, 则
>
> $$E(g(X))=\sum_i g(a_i)p_i \text{ 或 } \int_{-\infty}^{+\infty} g(x)f(x)\,\mathrm{d}x \quad (\text{若求和或极限存在.})$$

3.1.3 条件数学期望 (条件均值)

条件期望 $E(Y\mid x)$ 称为 Y 对 X 的回归函数.

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

期望等于条件期望的期望

$$\begin{aligned}E(Y)&=\int_{-\infty}^{+\infty} E(Y\mid x)f_1(x)\,\mathrm{d}x\\&=E[E(Y\mid X)]\end{aligned}$$

3.1.4 中位数

- 中位数总是存在，均值则不然.
- 中位数可以不唯一.

3.2 方差与矩

3.2.1 方差和标准差

1. $\operatorname{Var}(X)=E(X-EX)^2=E(X^2)-(EX)^2$.
2. $\operatorname{Var}(aX+b)=a^2\operatorname{Var}(X)$.
3. 独立随机变量: $\operatorname{Var}(X_1+\cdots+X_n)=\operatorname{Var}(X_1)+\cdots+\operatorname{Var}(X_n)$.
4. $E[(X-c)^2]=\operatorname{Var}(X)+(EX-c)^2$.

   $\operatorname{Var}(X)=\min_c\left\{E[(X-c)^2]\right\}$, 当且仅当 $c=EX$ 时取等.

   注: 令 $c=0$ 即得 1. 式.

注意 比较 $\operatorname{Var}(nX_1)=n^2\operatorname{Var}(X_1)$, 而独立同分布时 $\operatorname{Var}(X_1+\cdots+X_n)=n\operatorname{Var}(X_1)$.

3.2.2 矩

定义

1. X 关于 c 点的 k 阶矩: $E[(X-c)^k]$.
2. k 阶原点矩: $\alpha_k=E(X^k)$.
3. k 阶中心距: $\mu_k=E[(X-EX)^k]$.

特例

1. $\alpha_1=E(X)$.
2. $\mu_1=0$.
3. $\mu_2=\operatorname{Var}(X)$.

偏度系数 $\beta_1=\dfrac{\mu_3}{\mu_2^{3/2}}$.

峰度系数 $\beta_2=\dfrac{\mu_4}{\mu_2^2}$.

正态分布 $N(\mu,\sigma^2)$ 的峰度系数为 3，故有时定义峰度系数为 $\mu_4\,/\mu_2^2-3$.

3.3 协方差与相关系数

3.3.1 协方差

1. $\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]$.
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
3. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.
4. $\text{Cov}(a_1 X + b_1, a_2 Y + b_2) = a_1 a_2 \text{Cov}(X, Y)$.
5. $2\,\text{Cov}(X, Y) = \text{Var}(X + Y) - \text{Var}(X) - \text{Var}(Y)$. ⭐
6. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$. ⭐
7. 若 X, Y 独立，则 $\text{Cov}(X, Y) = 0$.
8. $[\text{Cov}(X, Y)]^2 \le \sigma_1^2 \sigma_2^2$，且当且仅当 $Y = a + bX$ 时取等.

施瓦茨不等式 $E(X^2)E(Y^2) \ge [E(XY)]^2$, 当且仅当具有线性关系即 $aX + bY = 0$ 时取等.

- 由 $E[(Y + tX)^2] \ge 0$ 的判别式小于零即得.

协方差矩阵 设 n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 中 X_i 与 X_j 的协方差都存在, 且记作 $c_{ij} = \text{Cov}(X_i, X_j)$, $i, j, = 1, 2, \cdots, n$, 则协方差矩阵为

$$\boldsymbol{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{nn} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & n_{nn} \end{pmatrix}.$$

其中 $c_{ii} = \text{Var}(X_i)$. 多元正态分布中的应用

3.3.2 相关系数

1. $\text{Corr}(X, Y) \equiv \text{Cov}(X, Y)/(\sigma_1 \sigma_2)$.
   - 相关系数也可以记作

$$\rho(X, Y) = E\left[\left(\frac{X - E(X)}{\sqrt{\text{Var}(X)}}\right)\left(\frac{Y - E(Y)}{\sqrt{\text{Var}(Y)}}\right)\right].$$

2. $\text{Corr}(X, Y) = \text{Corr}(Y, X)$.
   - 相关系数矩阵是对称阵.
3. $\text{Corr}(a_1 X + b_1, a_2 Y + b_2) = \text{Corr}(X, Y)$.
   - 相关系数不受单位影响.
4. 若 X, Y 独立，则 $\text{Corr}(X, Y) = 0$.
   - 若 $\text{Corr}(X, Y) = 0$ 则称 X 与 Y **不相关**.
5. $|\text{Corr}(X, Y)| \le 1$，且当且仅当 X 和 Y 有严格线性关系时取等.
   - 相关系数又称为线性相关系数.
6. 最小二乘及其均方误差.

$$\begin{aligned} E[(Y - a - bX)^2] &\equiv E[(Y - m_2) - b(X - m_1) - c]^2 \\ &= \sigma_2^2 + b^2\sigma_1^2 - 2b\,\text{Cov}(X, Y) + c^2 \\ &\ge \sigma_2^2 + b^2\sigma_1^2 - 2b\,\text{Cov}(X, Y) \end{aligned}$$

$$\begin{aligned} b &= \text{Cov}(X, Y)/\sigma_1^2 = \sigma_1^{-1}\sigma_2\,\text{Corr}(X, Y) \equiv \sigma_1^{-1}\sigma_2\rho \\ L(X) &= m_2 - \sigma_1^{-1}\sigma_2\rho m_1 + \sigma_1^{-1}\sigma_2\rho X \end{aligned}$$

$$\begin{aligned} E[(Y - L(X))^2] &= \sigma_2^2 + b^2\sigma_1^2 - 2b\,\text{Cov}(X, Y) \qquad (\text{由最上式}) \\ &= \sigma_2^2(1 - \rho^2) \end{aligned}$$

二维正态分布

若 $(X, Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$，则

1. 即使允许用任何函数 $M(X)$ 逼近 Y，则所得到的最佳逼近仍是 $L(X)$，故只需考虑线性逼近已足够.
2. 对于二维正态分布, $\text{Corr}(X, Y) = \rho$，即 $\text{Corr}(X, Y) = 0$ 可推出二者独立.

3.4 大数定理和中心极限定理

3.4.1 大数定理

依概率收敛 $Y_n \xrightarrow{P} a \quad \Leftrightarrow \quad \lim\limits_{n\to\infty} P\{|Y_n - a| < \varepsilon\} = 1.$

- $X_n \xrightarrow{P} a \quad \Rightarrow \quad g(X_n) \xrightarrow{P} g(a)$. (其中 X_n 可为向量)

马尔科夫不等式 若 Y 为只取非负值的随机变量，则对 $\forall \varepsilon > 0$，有

$$P(Y \geq \varepsilon) \leq E(Y)/\varepsilon.$$

切比雪夫不等式 若 $\operatorname{Var}(Y)$ 存在，则

$$P(|Y - EY| \geq \varepsilon) \leq \operatorname{Var}(Y)/\varepsilon^2.$$

大数定理 设 $X_1, X_2, \cdots, X_n$ 是独立同分布的随机变量，记它们的公共均值为 a，方差存在并记为 σ^2，则对 $\forall \varepsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - a\right| \geq \varepsilon\right) = 0.$$

伯努利大数定理 即大数定理的特例（频率收敛于概率）

$$\lim_{n\to\infty} P(|p_n - p| \geq \varepsilon) = 0.$$

- 大数定理中无需假定 X_i 的方差存在也可以证明 (即 **辛钦大数定理**)，不必同分布，甚至可以不独立.

3.4.2 中心极限定理

应用

林德伯格—莱维定理 设 $X_1, X_2, \cdots, X_n$ 为独立同分布的随机变量，$E(X_i) = a$, $\operatorname{Var}(X_i) = \sigma^2 \ (0 < \sigma^2 < \infty)$，则对任何实数 x，有

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}\sigma}(X_1 + X_2 + \cdots + X_n - na) \leq x\right) = \Phi(x) = \int_{-\infty}^{x} \frac{\mathrm{e}^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}t.$$

- 因此, 任何独立同分布的大量随机变量之和近似服从正态分布.

$$\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \stackrel{\cdot}{\sim} N(0,1) \quad \Rightarrow \quad \overline{X} \stackrel{\cdot}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right).$$

- 误区: 变量本身的分布并未改变. 可以将大量随机变量分为若干组大量随机变量, 分别计算其和, 并验证其符合正态分布. 这个性质可用于参数估计等, 但不可用于检验随机变量的值是否被篡改.

棣莫弗—拉普拉斯定理 上式的特例，当 $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p \ (0 < p < 1)$ 时，对任何实数 x，有

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{np(1-p)}}(X_1 + X_2 + \cdots + X_n - np) \leq x\right) = \Phi(x).$$

- 或者说, 若随机变量 $\eta_n \sim B(n,p)\ (n = 1, 2, \cdots)$, 则对任何实数 x, 有

$$\lim_{n\to\infty} P\left\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x\right\} = \Phi(x).$$

估值公式

$$P(t_1 \leq X_1 + X_2 + \cdots + X_n \leq t_2) \approx \Phi(y_2) - \Phi(y_1).$$

其中 $y_i = (t_i - np) \Big/ \sqrt{np(1-p)}$，或修正为

$$\begin{cases} y_1 = \left(t_i - \dfrac{1}{2} - np\right) \Big/ \sqrt{np(1-p)}, \\ y_2 = \left(t_i + \dfrac{1}{2} - np\right) \Big/ \sqrt{np(1-p)}. \end{cases}$$

3.5 母函数

3.5.1 母函数的定义

整值随机变量 即只取非负整值的随机变量.

若整值随机变量的概率分布为 $P\{X=k\}=p_k,\ k=0,1,2,\cdots$, 则其 **母函数** 为

$$G(s):=\sum_{k=0}^{+\infty}p_ks^k=E(s^X).$$

- $G(1)$ 收敛且为 1 (而不是书中说的因为某种方式的计算结果为 1 而收敛), 且 $G(-1)$ 绝对收敛, 故 $G(s)$ 至少在 $[-1,1]$ 上绝对收敛.

3.5.2 常见分布的母函数

- 对于 $X\sim B(n,p)$, 有

$$G(s)=\sum_{k=0}^{n}\binom{n}{k}p^kq^{n-k}s^k=(ps+q)^n,\quad s\in(-\infty,+\infty).$$

- 对于 $X\sim P(\lambda)$, 有

$$G(s)=\sum_{k=0}^{+\infty}\frac{\lambda^k\mathrm{e}^{-\lambda}}{k!}\cdot s^k=\mathrm{e}^{\lambda(s-1)},\quad s\in(-\infty,+\infty).$$

- 对于 $X\sim G(p)$, 有

$$G(s)=\sum_{k=1}^{+\infty}pq^{k-1}s^k=\frac{ps}{1-qs},\quad s\in\left(-\frac{1}{q},\frac{1}{q}\right).$$

3.5.2 母函数的性质

- $p_k=\dfrac{G^{(k)}(0)}{k!},\ k=0,1,2,\cdots.$
- $E(X)=\sum_{k=1}^{+\infty}kp_k=G'(1).$
- $\mathrm{Var}(X)=G''(1)+G'(1)-[G'(1)]^2.$

若 X 的 k 阶矩存在, 则有

- $G^{(0)}(1)=1.$
- $G^{(1)}(1)=E(X).$
- $G^{(2)}(1)=E(X^2)-E(X).$
- $G^{(3)}(1)=E(X^3)-3E(X^2)+2E(X).$
- $G^{(4)}(1)=E(X^4-6X^3+11X^2-6X).$

3.5.3 独立和的母函数

设整值随机变量 $X\sim P\{X=k\}=a_k$ 和 $Y\sim P\{Y=k\}=b_k$ 相互独立, 且母函数分别为 $A(s),\ B(s)$, 则 $Z=X+Y$ 的母函数为

$$G(s)=A(s)B(s)=\sum_{n=0}^{\infty}\sum_{k=0}^{n}a_kb_{n-k}s^n.$$

若 n 个整值随机变量独立同分布, 则 $G(s)=[G_1(s)]^n$.

3.6 特征函数

3.6.1 特征函数的定义

设 X,Y 为实随机变量, 则称 $Z=X+\mathrm{i}Y$ 为 **复随机变量**.

设 X 是实随机变量, 则 X 的 **(一维) 特征函数** 为

$$
\begin{aligned}
g(t) &= E(\mathrm{e}^{\mathrm{i}tX}) \qquad (-\infty < t < +\infty) \\
&= \begin{cases} \displaystyle\int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}tx} f(x)\,\mathrm{d}x, & \text{连续型随机变量}, \\ \displaystyle\sum p_k \mathrm{e}^{\mathrm{i}tx_k} = G(\mathrm{e}^{\mathrm{i}t}), & \text{离散性随机变量}. \end{cases} \\
&= \begin{cases} \displaystyle\int_{-\infty}^{+\infty} f(x)\cos tx\,\mathrm{d}x + \mathrm{i}\int_{-\infty}^{+\infty} f(x)\sin tx\,\mathrm{d}x, & \text{连续型随机变量}, \\ \displaystyle\sum p_k \cos tx_k + \mathrm{i}\sum p_k \sin tx_k, & \text{离散性随机变量}. \end{cases}
\end{aligned}
$$

- 上述级数与广义积分绝对收敛.

3.6.2 常见分布的特征函数

常见分布	特征函数
$X \sim B(n,p)$	$g(t) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \mathrm{e}^{\mathrm{i}tk} = (p\mathrm{e}^{\mathrm{i}t} + q)^n$
$X \sim P(\lambda)$	$g(t) = \sum_{k=0}^{+\infty} \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} \mathrm{e}^{\mathrm{i}tk} = \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t}-1)}$
$X \sim G(p)$	$g(t) = \sum_{k=1}^{+\infty} pq^{k-1} \mathrm{e}^{\mathrm{i}tk} = \frac{p\mathrm{e}^{\mathrm{i}t}}{1-q\mathrm{e}^{\mathrm{i}t}}$
$X \sim U(a,b)$	$g(t) = \int_a^b \frac{\mathrm{e}^{\mathrm{i}tx}}{b-a}\,\mathrm{d}x = \begin{cases} \frac{\mathrm{e}^{\mathrm{i}bt} - \mathrm{e}^{\mathrm{i}at}}{\mathrm{i}t(b-a)}, & t \neq 0, \\ 1, & t = 0. \end{cases}$
$X \sim E(\lambda)$	$g(t) = \int_0^{+\infty} \mathrm{e}^{\mathrm{i}tx} \lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{\lambda}{\lambda - \mathrm{i}t}$
$X \sim N(\mu,\sigma^2)$	$g(t) = \int_{-\infty}^{+\infty} \frac{\mathrm{e}^{\mathrm{i}tx}}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x = \mathrm{e}^{\mathrm{i}\mu t} = \mathrm{e}^{\mathrm{i}\mu t - \frac{\sigma^2}{2}t^2}$

3.6.3 特征函数的性质

1. $g(0) = 1$.
2. $|g(t)| \leq 1$.
3. $g(-t) = \overline{g(t)}$.
4. $g(t)$ 在 $(-\infty, +\infty)$ 上一致连续.
5. $\forall n \in \mathbb{N}^+$, $\forall t_i \in \mathbb{R}$, $\forall z_i \in \mathbb{C}$:

$$\sum_{j=1}^{n}\sum_{k=1}^{n} g(t_j - t_k) z_j \bar{z}_k \geq 0.$$

6. 如果 X 的 n 阶原点矩存在, 则它的特征函数 的 n 阶导数存在, 且

$$g^{(k)}(0) = \mathrm{i}^k E(X^k), \quad k = 1, 2, \cdots n.$$

7. 若 X 的特征函数为 $g_X(t)$, 且 $Y = aX + b$ $(a, b \in \mathbb{R})$, 则

$$g_Y(t) = \mathrm{e}^{\mathrm{i}bt} g_X(at).$$

8. 如果 X_1 和 X_2 相互独立, 且特征函数分别为 $g_1(t), g_2(t)$, 则 $Y = X_1 + X_2$ 的特征函数为

$$g_Y(t) = g_1(t) g_2(t).$$

类似的, 由性质 6 可得到 k 阶原点矩.

例题

分鞋：例 1.7 和 2.4

第 4 章 参数估计

4.1 数理统计学的基本概念

总体（母体）是概率分布族的一员.

总体分布 离散性（概率函数），连续型（概率密度函数）

单参数分布族

非参数总体

样本大小 (容量)

经验分布函数 (样本分布函数) $F_n(x)=\{X_1,X_2,\cdots,X_n \text{ 中不大于 } x \text{ 的个数}\}/n.$

即将 X 的样本值 $x_1,x_2,\cdots,x_n$ 从小到大重排后, 定义经验分布函数如下.

$$\underbrace{x_{(1)},\cdots,x_{(1)}}_{n_1},\underbrace{x_{(2)},\cdots,x_{(2)}}_{n_2},\cdots\underbrace{x_{(m)},\cdots,x_{(m)}}_{n_m},$$

$$F_n(x)=\begin{cases}0, & x<x_{(1)},\\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{k}n_i, & x_{(k)}\le x<x_{(k+1)}\\ 1, & x\ge x_{(m)}\end{cases}$$

1. $0\le F_n(x)\le 1.$
2. $F_n(x)$ 单调不减.
3. $F_n(-\infty)=0,\ F_n(+\infty)=1.$
4. $F_n(x)$ 右连续.

常用统计图

1. **频率直方图** 以 $\dfrac{f_i}{\Delta t_i}$ 为高. 所有小矩形的面积和为 1.
2. **条形图**　一般用于小样本离散性随机变量总体分布.

格列文科定理 对于任意实数 x, 经验分布函数 $F_n(x)$ 以概率 1 一致收敛于总体分布函数 $F(x)$, 即

$$P\left\{\lim_{n\to\infty}\sup_{-\infty<x<+\infty}|F_n(x)-F(x)|=0\right\}=1.$$

- 经验分布函数不确定, 不唯一, 所以在极限外套一个 P.

统计量　只依赖于样本，而不依赖于其未知参数.

- 样本的统计量为 $g(X_1,X_2,\cdots,X_n)$.
- 统计量的观测值为 $g(x_1,x_2,\cdots,x_n)$.

样本均值 $\overline{X}=a_1=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i.$

- 其观测值记为 $\overline{x}=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i.$

样本离差平方和 $\mathrm{SS}=\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})^2.$

- 其观测值记为 $\mathrm{ss}=\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2.$

样本方差 $S^2=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})^2.$

- 其观测值记为 s^2.
- 标准差又称均方差, 样本标准差的观测值记为 s.

样本原点矩 $a_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$.

- 或使用 A_k 表示样本原点矩, 用 a_k 表示其观测值.

样本中心矩 $m_k = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$.

- 或使用 B_k 表示样本中心距, 用 b_k 表示其观测值.
- 矩称为理论矩，样本矩称为经验矩，即经验分布函数的矩.
- $m_2 = \dfrac{n-1}{n}S^2$.

样本协方差 $S_{XY} = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$.

- 其观测值记为 $s_{_{XY}}$.

样本相关系数 $\rho_{XY} = \dfrac{S_{XY}}{S_X S_Y}$.

- 其中 S_X 和 S_Y 为样本均方差. 其观测值为 $\rho_{XY} = \dfrac{s_{_{XY}}}{s_{_X} s_{_Y}}$.

样本的 **众数** 记为 M_0.

次序统计量 $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$.

样本中位数 $\hat{m} = \begin{cases} X_{((n+1)/2)}, & n \text{ 为奇数}, \\ (X_{(n/2)} + X_{(n/2+1)})/2, & n \text{ 为偶数}. \end{cases}$

4.2 点估计

4.2.1 矩估计法

$$\alpha_m = \int_{-\infty}^{+\infty} x^m f(x;\theta_1,\cdots,\theta_2)\,\mathrm{d}x \approx a_m = \sum_{i=1}^{n} X_i^m / n$$

取 $m = 1, 2, \cdots, k$，联立方程组即得 $\theta_i \approx \hat{\theta}_i(X_1, \cdots, X_n)$.

变异系数 σ/μ.

4.2.2 极大似然估计法

样本 $(X_1, X_2, \cdots, X_n)$ 的总体分布函数为

$$L(x_1,\cdots,x_n,;\theta_1,\cdots,\theta_k) = f(x_1;\theta_1,\cdots,\theta_k)f(x_2;\theta_1,\cdots,\theta_k)\cdots f(x_n;\theta_1,\cdots,\theta_k)$$
$$L(X_1,\cdots,X_n;\theta_1^*,\cdots,\theta_k^*) = \max_{\theta_1,\cdots,\theta_k} L(X_1,\cdots,X_n;\theta_1,\cdots,\theta_k)$$

欲得到极大似然估计，解如下似然方程组

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

4.2.3 贝叶斯估计法

先验分布，先验概率. 允许使用主观概率.

设总体有概率密度 $f(X,\theta)$，抽样本 $X_1, X_2, \cdots, X_n$，则 $(\theta, X_1, \cdots, X_n)$ 的联合密度为

$$h(\theta)f(X_1,\theta)\cdots f(X_n,\theta)$$

由此算出 $(X_1, X_2, \cdots, X_n)$ 的边缘密度为

$$p(X_1,X_2,\cdots,X_n) = \int h(\theta)f(X_1,\theta)\cdots f(X_n,\theta)\mathrm{d}\theta$$

从而得出 θ 在给定 $X_1, X_2, \cdots, X_n$ 的条件密度为

$$h(\theta \mid X_1,\cdots,X_n) = h(\theta)f(X_1,\theta)\cdots f(X_n,\theta)/p(X_1,\cdots,X_n)$$

一般去上式的均值作为估计, 即

$$\tilde{\theta}=E(h(\theta \mid X_1,X_2,\cdots,X_n)).$$

$h(\theta)$ 一般是概率函数，即满足 $h(\theta)\geq 0, \displaystyle\int h(\theta)\mathrm{d}\theta=1.$

但对于积分域为无穷区间，或一些特定的分布，可以采用其它函数，比如 $h(\theta)=1$，或直接取为先验密度等. 此时 $h(\theta)$ 称为 "广义先验密度".

根据 n 次独立试验中事件 A 发生的次数 X 去估计其发生的概率 p，按照 "同等无知" 原则（贝叶斯原则），由上述方法积分得

$$\tilde{p}=\frac{X+1}{n+2}.$$

- 估计正态分布 $N(\mu,\sigma^2)$ 中的 μ 时，取 $h(\mu)=1$;
- 估计正态分布 $N(\mu,\sigma^2)$ 中的 σ 时，取 $h(\sigma)=\sigma^{-1}$;
- 估计指数分布 $E(\lambda)$ 中的 λ 时，取$h(\lambda)=\lambda^{-1}$.

由先验分布 $N(\mu.\sigma^2)$ 估计正态总体 $N(\theta,1)$ 中的 θ 为（取 $h(\theta)$ 为先验密度）

$$\tilde{\theta}=\frac{n}{n+\sigma^{-2}}\overline{X}+\frac{\sigma^{-2}}{n+\sigma^{-2}}\mu.$$

不知道为什么将正态总体分布的方差取为 1，有什么实际应用吗?

4.3 点估计的优良性准则

估计的整体性能

1. 无偏性.
   1. 没有系统性的偏差, 即误差的均值为零.
   2. 各次估计的均值依概率收敛至被估计值.
2. 数量指标 (如均方误差).

4.3.1 估计量的无偏性

无偏估计量 $\hat{g}$ 须满足

$$E_{\theta_1,\cdots,\theta_k}[\hat{g}(X_1,\cdots,X_n)]=g(\theta_1,\cdots,\theta_k).$$

- $m=\overline{X}$ 是 $E(X)$ 的无偏估计.
- 如果总体均值未知, 则 $S^2=\displaystyle\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{n-1}$ 是 $\operatorname{Var}(X)$ 的无偏估计.
- 如果总体均值已知, 则 $m_2=\displaystyle\sum_{n=1}^{\infty}\frac{(X_i-\overline{X})^2}{n}$ 是 $\operatorname{Var}(X)$ 的无偏估计.

由 $\sigma^2=E(S^2)=\operatorname{Var}(S)+(ES)^2$ 知, S 总是 σ 系统性偏低的估计.

- ⭐ (看思路) 对于正态分布总体 $N(\mu,\sigma^2)$, 由 $(n-1)S^2/\sigma^2\sim\chi^2_{n-1}$ 算出

  $$S/\sigma\sim g(s)=\begin{cases}\dfrac{(n-1)^{\frac{n-1}{2}}}{2^{\frac{n-3}{2}}\Gamma\left(\dfrac{n-1}{2}\right)}, & s>0\\ 0 & s\leq 0.\end{cases}$$

  计算 $E(S)=\sigma\displaystyle\int_0^{+\infty}sg(s)\,\mathrm{d}s$, 故 σ 的一个无偏估计是

  $$\tilde{\sigma}=\sqrt{\frac{n-1}{2}}\frac{\Gamma\left(\dfrac{n-1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)}S.$$

- 无偏估计不一定好, 比如 $X\sim P(\lambda)$, 则 $g(\lambda)=\mathrm{e}^{-2\lambda}$ 唯一的无偏估计为

$$\hat{g}(X)=\begin{cases}1, & X \text{ 为偶数},\\ -1, & X \text{ 为奇数}.\end{cases}$$

4.3.2 最小方差无偏估计

1 均方误差

$$\begin{aligned}M_{\hat{\theta}}(\theta)&=E_\theta\left[\hat{\theta}(X_1,\cdots,X_n)-\theta\right]^2\\&=\mathrm{Var}_\theta(\hat{\theta})+\left[E_\theta(\hat{\theta})-\theta\right]^2\end{aligned}$$

2 最小方差无偏估计 (MVU 估计)

注意是最小方差, 而不是最小均方误差. (Minimum Variance Unbiased)

3 克拉美 - 劳不等式

对于单参数情况 $f(x,\theta)$, 为估计 $g(\theta)$, 记 **费歇尔信息量** 为

$$\begin{aligned}I(\theta)&=E\left[\left(\frac{\partial f(x,\theta)}{\partial\theta}\Big/f(x,\theta)\right)^2\right]\\&=\int\left[\left(\frac{\partial f(x,\theta)}{\partial\theta}\right)^2\Big/f(x,\theta)\right]\mathrm{d}x\qquad(\text{连续的总体分布})\\&=\sum_i\left(\frac{\partial f(a_i,\theta)}{\partial\theta}\right)^2\Big/f(a_i,\theta)\qquad(\text{离散的总体分布})\end{aligned}$$

则对任一无偏估计 $\hat{g}=\hat{g}(X_1,X_2,\cdots,X_n)$, 有 **克拉美 - 劳不等式**:

$$\mathrm{Var}_\theta(\hat{\theta})\geq(g'(\theta))^2/(nI(\theta)).$$

- MVU 的均方误差不一定是最小的, 如对于正态分布 $N(\mu,\sigma^2)$, 其中 μ 已知, 则 m_2 是 MVU 估计, 但 $E(m_2-\sigma^2)^2=\dfrac{2\sigma^4}{n}>E(\dfrac{m_2}{n+1}-\sigma^2)^2=\dfrac{2\sigma^4}{n+1}$.
- 若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是 θ 的 MVU 估计, 则 $a\hat{\theta}_1+b\hat{\theta}_2+c$ 是 $(a+b)\theta+c$ 的 MVU 估计. ⭐ ⭐ (利用第三章定理 3.1, 2°)
- 若 $E(X)=\theta,\ \sum_{i=1}^n c_i=1$, 则 $\sum_{i=1}^n c_iX_i$ 是 θ 的无偏估计, 并且当且仅当 $c_i=\dfrac{1}{n}$ 时, 其为 MVU 估计.

4.3.3 估计量的相合性与渐进正态性

1 相合性

如果当样本大小 n 无限增加时, 估计量 $T(X_1,X_2,\cdots,X_n)$ 依概率收敛于被估计值, 则称该估计量是相合估计, 即

$$\forall\varepsilon>0:\lim_{n\to\infty}P_{\theta_1,\cdots,\theta_k}\left(|T(X_1,\cdots,X_n)-g(\theta_1,\cdots,\theta_n)|\geq\varepsilon\right)=0.$$

具有相合性的例子: $m(n),\ m_2(n)$, 绝大多数极大似然估计等等.

2 渐进正态性

- 大样本性质
  - 相合性
  - 渐进正态性
- 小样本性质
  - 无偏性

4.4 区间估计

4.4.1 基本概念

奈曼理论 的原则: 先保证可靠度, 再提升精度.

称区间估计 $[\hat{\theta}_1,\hat{\theta}_2]$ 的 **置信系数** 为 $1-\alpha$, 如果

$$\exists\alpha>0,\forall\theta:P_\theta\left(\hat{\theta}_1(X_1,\cdots,X_n)\leq\theta\leq\hat{\theta}_2(X_1,\cdots,X_n)\right)=1-\alpha.$$

称区间估计 $[\hat\theta_1, \hat\theta_2]$ 的 **置信水平** 为 $1-\alpha$, 如果

$$\exists \alpha > 0,\ \forall\theta : P_\theta\left(\hat\theta_1(X_1,\cdots,X_n) \le \theta \le \hat\theta_2(X_1,\cdots,X_n)\right) \ge 1-\alpha.$$

α 一般取为 0.1, 0.05, 0.01, 0.001.

区间估计的研究对象:

1. 置信系数或置信水平.
2. 区间长度.
3. 区间右端点与左端点之比.

4.4.2 枢轴变量法

上 β 分位点: $F(v_\beta) = 1-\beta$.

下 β 分位点: $F(w_\beta) = \beta$.

上 β 分位点就是下 $1-\beta$ 分位点.

统计三大分布的上 β 分位点记为: $\chi_n^2(\beta),\ t_n(\beta),\ f_{n,m}(\beta)$.

利用上 β 分位点 w_β 寻找区间估计的 **枢轴变量法**:

1. 找一个与被估计参数 $g(\theta)$ 有关的统计量 T.
2. 找 **枢轴变量** $S(T, g(\theta))$, 使其分布 F 与 θ 无关.
3. $a \le S(T, g(\theta)) \le b \quad \Leftrightarrow \quad A \le g(\theta) \le B$.
4. $P(w_{1-\alpha/2} \le S(T, g(\theta)) \le w_{a/2}) = 1-\alpha$.

一样本 t 区间估计, 为保证长度 $2St_{n-1}(\alpha/2)/\sqrt{n} \le L$, 斯泰因提出了 **两阶段抽样** 的方法, 其中追加抽样的次数为

$$m = \begin{cases} 0, & n \le [4t_{n-1}^2(\alpha/2)S^2/L^2]+1, \\ n-1-[4t_{n-1}^2(\alpha/2)S^2/L^2], & n > [4t_{n-1}^2(\alpha/2)S^2/L^2]+1. \end{cases}$$

记两次样本全体的均值为 $\tilde X$, 则区间估计 $[\tilde X - L/2,\ \tilde X + L/2]$ 有置信系数 $1-\alpha$.

4.4.3 大样本法

大样本区间估计: 利用 中心极限定理 与枢轴变量法.

例如, 一般的, 设总体有均值 θ, 方差 σ^2, 并且都位置, 从样本 $X_1, X_2, \cdots, X_n$ 做 θ 的区间估计. 由于样本均方差 S 是 σ 的祥和估计, 利用中心极限定理, 当 n 足够大时, 有

$$\sqrt{n}(\overline{X}-\theta)/S \sim N(0,1).$$

以此为枢轴变量, 于是有区间估计

$$\left[\overline{X} - Su_{\alpha/2}/\sqrt{n},\ \overline{X} + Su_{\alpha/2}/\sqrt{n}\right].$$

对于二项分布, 当 $\alpha = 0.05,\ n \ge 40$ 时, 有区间长度 $\theta_2 - \theta_1 \le 0.3$.

4.4.4 置信界

置信系数 (水平) 为 α 的置信上界 $\overline{\theta}$ 和置信下界 $\underline{\theta}$:

$$\begin{aligned} &\forall\theta : P_\theta(\overline{\theta}(X_1, X_2, \cdots, X_n) \ge \theta) = 1-\alpha \\ &\forall\theta : P_\theta(\underline{\theta}(X_1, X_2, \cdots, X_n) \le \theta) = 1-\alpha \end{aligned}$$

4.4.5 贝叶斯法

即寻找 $\hat\theta_1,\ \hat\theta_2$, 使得

$$\int_{\hat{\theta}_1}^{\hat{\theta}_2} h(\theta \mid X_1, \cdots, X_n)\, \mathrm{d}\theta = 1 - \alpha \qquad (\text{区间估计})$$

$$\int_{-\infty}^{\hat{\theta}} h(\theta \mid X_1, \cdots, X_n)\, \mathrm{d}\theta = 1 - \alpha \qquad (\text{置信上界})$$

$$\int_{\hat{\theta}}^{+\infty} h(\theta \mid X_1, \cdots, X_n)\, \mathrm{d}\theta = 1 - \alpha \qquad (\text{置信下界})$$

区间估计中确定 $\theta_1,\ \theta_2$ 的方法 (原则):

1. 使 $\hat{\theta}_2 - \hat{\theta}_1$ 最小.
2. 使 $\hat{\theta}_2/\hat{\theta}_1$ 最小.
3. 取置信水平为 $\alpha/2$ 的置信上下界.

第 5 章 假设检验

5.1 问题提法和基本概念

5.1.1 例子与问题提法

原假设 (零假设, 解消假设), 对立假设 (备择假设).

检验统计量, 接受域, 否定域 (临界域), 临界值.

简单假设, 复合假设. 赘余参数.

5.1.2 功效函数

设 H_0 为原假设, Φ 是基于样本 $X_1, X_2, \cdots, X_n$ 而对 H_0 做的一个检验, 则其 **功效函数** 是未知参数 $\theta_1, \theta_2, \cdots, \theta_n$ 的函数:

$$\beta_\Phi(\theta_1, \theta_2, \cdots, \theta_n) = P_{\theta_1, \theta_2, \cdots, \theta_n}(\text{在检验 } \Phi \text{ 之下, } H_0 \text{ 被否定})$$

- 当 $(\theta_1, \theta_2, \cdots, \theta_n) \in H_0$ 时, 上式越小越好.
- 当 $(\theta_1, \theta_2, \cdots, \theta_n) \in H_1$ 时, 上式越大越好, 此时称为功效函数.

5.1.3 两类错误, 检验的水平

两类错误

1. H_0 正确, 但被否定.
2. H_0 错误, 但被接受.

$$\alpha_{1\Phi}(\theta_1, \theta_2, \cdots, \theta_k) = \begin{cases} \beta_\Phi(\theta_1, \theta_2, \cdots, \theta_k), & (\theta_1, \theta_2, \cdots, \theta_k) \in H_0, \\ 0 & (\theta_1, \theta_2, \cdots, \theta_k) \in H_1. \end{cases}$$

$$\alpha_{2\Phi}(\theta_1, \theta_2, \cdots, \theta_k) = \begin{cases} 0, & (\theta_1, \theta_2, \cdots, \theta_k) \in H_0, \\ 1 - \beta_\Phi(\theta_1, \theta_2, \cdots, \theta_k), & (\theta_1, \theta_2, \cdots, \theta_k) \in H_1. \end{cases}$$

H_0 的一个水平为 α 的检验 Φ:

$$\beta_\Phi(\theta_1, \theta_2, \cdots, \theta_k) \le \alpha \quad (\text{对任何 } (\theta_1, \theta_2, \cdots, \theta_k) \in H_0).$$

并使 α 仅可能小. 即固定第一类错误概率的原则.

5.1.4 一致最优检验

假设检验问题 $H_0 : H_1$ 的一个水平为 α 的一致最优检验 Φ: 即对任何一个其它水平 α 的检验 g 有

$$\beta_\Phi(\theta_1, \theta_2, \cdots, \theta_k) \ge \beta_g(\theta_1, \theta_2, \cdots, \theta_k) \quad (\text{对任何 } (\theta_1, \theta_2, \cdots, \theta_k) \in H_1).$$

- 在总体分布只依赖于一个参数 θ, 而原假设 H_0 是 $\theta \le \theta_0$ 或 $\theta \ge \theta_0$ 时, 一致最优检验存在.

5.2 重要参数检验

5.2.1 正态总体均值的检验

1 方差 σ^2 已知