§3 数据简化的原则

§3.1 充分统计量

在用统计量 T 对 X 压缩简化的过程中，大家自然希望不损失任何有关 θ 的信息。人们称这样的统计量为充分统计量。例如，$X_1,\cdots,X_n$ i.i.d. $N(\theta,1)$，θ 是未知参数。若取$T_1=(X_1,X_2)$ 来压缩数据，直觉上显然是不充分的。而若取 $T_2=\overline{X}$，那它是否损失了有关 θ 的信息就不那么显而易见了。

定义3.1 设X 的分布族为$(\mathscr{X},\mathscr{B}_X,\mathscr{P}=\{P_\theta:\theta\in\Theta\})$，$T=T(X)$ 是一个统计量，其值域空间为$(\mathscr{T},\mathscr{B}_T)$。若对任意$t\in\mathscr{T}$，在给定$T=t$ 时，X 的条件分布都与 θ 无关，则称 T 是 $\mathscr{P}$ 的**充分统计量**。

例3.1 设$X_1,\cdots,X_n$ i.i.d. P_θ.

(1) 若P_θ 为$B(1,\theta),\theta\in(0,1)$，则$T=\sum_{i=1}^n X_i$ 是充分统计量。

证明. 因为$T\sim B(n,\theta)$，所以

$$P_\theta(T=t)=\binom{n}{t}\theta^t(1-\theta)^{n-t}I_{\{0,1,\cdots,n\}}(t).$$

在给定$T=t$ 的条件下，

$$\begin{aligned}&P_\theta(X_1=x_1,\cdots,X_n=x_n|T=t)\\=&\frac{P_\theta(X_1=x_1,\cdots,X_n=x_n,T=t)}{P_\theta(T=t)}\\=&\frac{\prod_{i=1}^n\theta^{x_i}(1-\theta)^{1-x_i}I_{\{0,1\}}(x_i)\cdot I(\sum_{i=1}^n x_i=t)}{\binom{n}{t}\theta^t(1-\theta)^{n-t}I_{\{0,1,\cdots,n\}}(t)}\\=&\binom{n}{t}^{-1}I_{B_t}(x_1,\cdots,x_n),\end{aligned}$$

其中$B_t=\{(x_1,\cdots,x_n):x_i=0,1,\sum_{i=1}^n x_i=t\}$。该条件分布与$\theta$ 无关，对$t=0,1,\cdots,n$ 都成立。因此T是充分的。 □

(2) P_θ 为Poisson$(\theta),\theta>0$，则 $T=\sum_{i=1}^n X_i$ 是充分统计量。

(3) P_θ 为$N(\theta,1),\theta\in\mathbb{R}$，则统计量 $T=\overline{X}$ 是充分的。

(4) P_θ 为$N(0,\sigma^2), \sigma^2>0$，则统计量$T=\sum_{i=1}^n X_i^2$ 是充分的。

若$X=(X_1,\cdots,X_n)$ 具有连续型分布，$T=(T_1,\cdots,T_r)$, $r<n$ 是统计量，且T 仍具有连续型分布，则求条件分布 $X|T$ 的一般步骤为：找一个 $n-r$ 维的统计量 $Y=(Y_1,\cdots,Y_{n-r})$，使得变换

$$x=(x_1,\cdots,x_n)\longleftrightarrow(T_1(x),\cdots,T_r(x),Y_1(x),\cdots,Y_{n-r}(x))$$

在某个合适的区域（包含所有 P_θ 的支撑）上是一一对应的。由 (T,Y) 与 X 两者的jpdf之间的关系式

$$p_\theta^{T,Y}(t,y)=p_\theta^X(x_1(t,y),\cdots,x_n(t,y))\left|\frac{\partial(x_1,\cdots,x_n)}{\partial(t,y)}\right|_+,$$

求出(T,Y) 及T 的jpdf，再计算给定 $T=t$ 时 Y 的条件密度

$$p_\theta^{Y|T}(y|t)=\frac{p_\theta^{T,Y}(t,y)}{p_\theta^T(t)}.$$

若对任意 t，条件密度 $p_\theta^{Y|T}(y|t)$ 都与 θ 无关，则统计量 T 是充分的。用此方法可证上例(3)、(4) 中统计量的充分性。

充分性的严格定义

充分性的定义中需要用到每一个$t\in\mathscr{T}$ 上的 X 的条件概率分布。若 T 是离散型分布的，可直接由条件概率来定义。但当 T 非离散型分布时，条件概率分布就需要用测度论工具来严格定义。

设 X 的概率分布为 $(\mathscr{X},\mathscr{B}_X,P)$。$T$ 是$(\mathscr{X},\mathscr{B}_X)\to(\mathscr{T},\mathscr{B}_T)$ 的可测变换，它的概率分布记为 P^T，即 $P^T(B)=P(T^{-1}(B))$, $\forall B\in\mathscr{B}_T$。对于$\forall A\in\mathscr{B}_X,\forall B\in\mathscr{B}_T$，条件概率$P(A|B)$ 满足

$$P(A\cap T^{-1}(B))=P(A|B)P^T(B).$$

对于固定的$A\in\mathscr{B}_X$，记

$$\nu_A(B)=P(A\cap T^{-1}(B)),\quad \forall B\in\mathscr{B}_T.$$

则 $\nu_A(\cdot)$ 是$(\mathscr{T},\mathscr{B}_T)$ 上的有限正测度，且$\nu_A\ll P^T$。由Radon-Nikodym 定理知，存在非负的$(\mathscr{T},\mathscr{B}_T)$ 可测函数$f_A(t)$ $(a.s.P^T$唯一)，使得

$$\nu_A(B)=\int_B f_A(t)dP^T(t),\quad \forall B\in\mathscr{B}_T.$$

上式中的 $f_A(t)$ 相当于$P(A|T=t)$，它至少在概念上对每一个t 逐点确定了条件概率。且除了某个P^T 零测集$N_A \in \mathscr{B}_T$ 外，$P(A|T=t)$ 是唯一确定的，常记之为$P(A|t)$。

显然，对于$A, B, A_1, A_2, \cdots \in \mathscr{B}_X, \quad a.s. P^T$ 有

(i) $0 \le P(A|t) \le 1, P(\mathscr{X}|t) = 1, P(\phi|t) = 0$；

(ii) 若$A_1, A_2, \cdots$ 互不相交，则$P(\bigcup_{i=1}^{\infty} A_i|t) = \sum_{i=1}^{\infty} P(A_i|t)$；

(iii) 若$A \subset B$，则$P(A|t) \le P(B|t)$。

这样，对于每一个固定的 t，$P(\cdot|t)$ 似乎确定了$(\mathscr{X}, \mathscr{B}_X)$ 上的一个条件概率分布了。但问题在于，对每个A，$P(A|t)$ 是a.s.确定的，例外集 N_A 依赖于A。所以上述三条性质的例外集分别依赖于$A, B, A_1, A_2, \cdots$，这些集合选择不同，例外集也可能不同，所有例外集之并未必是P^T 零测集。因此，条件概率分布仍然未确定，还可能不存在。好在对于定义在欧氏空间上的概率测度，不存在这样的麻烦。

定理3.1 P 是定义在欧氏空间$(\mathscr{X}, \mathscr{B}_X)$ 上的概率测度，T 是$(\mathscr{X}, \mathscr{B}_X) \to (\mathscr{T}, \mathscr{B}_T)$ 的可测函数，则存在**正则条件分布** $P(A|t)$，满足

1) 对任意固定的$A \in \mathscr{B}_X$，$P(A|t)$ 是$\mathscr{B}_T$ 可测函数，且$\int_B P(A|t) dP^T(t) = P(A \cap T^{-1}(B)),\ \forall B \in \mathscr{B}_T$；

2) $\forall t \in \mathscr{T}$，$P(\cdot|t)$ 是$(\mathscr{X}, \mathscr{B}_X)$ 上的概率测度。

证明参见Lehmann *et al.* (2005), TSH, p41/Theorem 2.5.1，或汪嘉冈(2005),《现代概率论基础》, p152/命题4.2.7(Doob)。

定义3.2 (充分性的严格定义) $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ 是$(\mathscr{X}, \mathscr{B}_X)$ 上的分布族，T 是$(\mathscr{X}, \mathscr{B}_X) \to (\mathscr{T}, \mathscr{B}_T)$ 的统计量。若$\forall A \in \mathscr{B}_X$，都存在一个与 θ 无关的 $\mathscr{B}_T$ 可测函数$P(A|t)$，它是所有 $P_\theta,\ \theta \in \Theta$ 在给定 T 时的条件概率函数，则 T 关于$\mathscr{P}$ 充分。

结合定义3.2 与定理3.1，有下列定理。

定理3.2 设$\mathscr{P}$ 是欧氏空间$(\mathscr{X},\mathscr{B}_X)$ 上的分布族，统计量 T 关于$\mathscr{P}$ 充分，则存在 $P(A|t)$，它是所有 P_θ, $\theta\in\Theta$ 在给定 T 时的正则条件分布。

注：

- 若T 是$\mathscr{P}$ 的充分统计量，$\mathscr{P}_0\subset\mathscr{P}$，则$T$ 也是$\mathscr{P}_0$ 的充分统计量；但不一定再是$\mathscr{P}_1\supset\mathscr{P}$ 的充分统计量。
- 统计量的充分性是依赖于模型假定的。T 对于$\mathscr{P}$ 充分，未必对于$\mathscr{P}'$ 充分。若模型假定合适，那么用充分统计量T 代替X 作推断是信息无损的，若不合适，那用错误模型下的充分统计量代替X 作推断是有风险的。

根据定义找充分统计量往往是不方便的。对于可控族，常常可以用下列因子分解定理判断充分统计量。

定理3.3 (因子分解定理) 设$\mathscr{P}=\{P_\theta:\theta\in\Theta\}$ 是$(\mathscr{X},\mathscr{B}_X)$ 上的分布族，关于σ有限测度ν 可控。T 是$(\mathscr{X},\mathscr{B}_X)\to(\mathscr{T},\mathscr{B}_T)$ 的统计量。则T 关于$\mathscr{P}$ 充分的充要条件是：存在一个非负$\mathscr{B}_T$ 可测函数$g_\theta(\cdot)$ 与$\mathscr{B}_X$ 可测函数$h(\cdot)$，使得

$$dP_\theta(x)/d\nu(x)=g_\theta[T(x)]\cdot h(x),\quad \forall\theta\in\Theta. \tag{3.1}$$

证明参见Lehmann *et al.* (2005) p45-46/Theorem 2.6.2 及Corollary 2.6.1，或Shao(2003) p104-106/Lemma 2.1 及Theorem2.2。

推论3.1 若U,T 都是$(\mathscr{X},\mathscr{B}_X,\mathscr{P})$上的统计量，$T$ 关于$\mathscr{P}$ 充分且$T=H(U)\ a.s.\ \mathscr{P}$，则统计量$U$ 关于$\mathscr{P}$ 也充分。

证明参见Shao (2005), Mathematical Statistics: Exercises and Solutions, p59/Exercise 11.

例3.2 (1) 设$X_1,\ldots,X_n$ i.i.d. Poisson(θ), $\theta>0$。因为$(X_1,\cdots,X_n)$ 的分布族关于计数测度可控，jpdf 为

$$P_\theta(X_1=x_1,\ldots,X_n=x_n)=\theta^{\sum_{i=1}^n x_i}e^{-n\theta}\cdot(\prod_{i=1}^n x_i!)^{-1}I_B(x_1,\cdots,x_n),$$

其中$B=\{0,1,2,\cdots\}^n$。因此，$\sum_{i=1}^n X_i$ 是充分统计量。

(2) 设$X_1,\ldots,X_n$ i.i.d. $N(\xi,\sigma^2)$, $\xi\in\mathbb{R},\sigma^2>0$。由因子分解定理知，$(\sum_{i=1}^n X_i,\sum_{i=1}^n X_i^2)$ 是充分统计量。统计量$(\overline{X},S_n^2)$ 与其一一对应，故也充分。

(3) 设$X_1,\ldots,X_n$ i.i.d. $U(0,\theta),\theta>0$。因为$(X_1,\cdots,X_n)$ 关于Lebesgue 测度的jpdf 为

$$p_\theta(x_1,\cdots,x_n)=\theta^{-n}\prod_{i=1}^n I_{(0<x_i<\theta)}=\theta^{-n}I_{(x_{(n)}<\theta)}I_{(x_{(1)}>0)},$$

因此$X_{(n)}$ 是充分统计量。（不用因子分解定理直接用定义3.2 的证明，参见Lehmann *et al.* (1998), p34/Example 6.3。）

例3.3 (指数族的充分统计量) (1) 设X 具有指数族分布，pdf 形如(1.1) 或(1.2)，则由因子分解定理知 $T=(T_1(X),\ldots,T_s(X))$ 是充分统计量。

(2) 设$X_1,\cdots,X_n$ i.i.d.，X_1 的pdf 形如(1.1)或(1.2)，则$(X_1,\cdots,X_n)$ 的联合分布亦是指数族，由因子分解定理知，$T=(\sum_{i=1}^n T_1(X_i),\ldots,\sum_{i=1}^n T_s(X_i))$ 是充分统计量。

例3.4 (次序统计量的充分性) 设$X_1,\ldots,X_n$ i.i.d. F，F 是一维连续型分布，有pdf $f(x)$。则次序统计量$T=(X_{(1)},\ldots,X_{(n)})$ 是充分统计量。（用因子分解定理可证）

一般地，若$X_1,\ldots,X_n$ 是一个i.i.d.样本，X_1 的分布族为$(\mathbb{R},\mathscr{B}_{\mathbb{R}},\mathscr{P}=\{P_\theta:\theta\in\Theta\})$。则次序统计量$T=(X_{(1)},\ldots,X_{(n)})$ 是充分的。

证明. （参见Lehmann *et al.* (2005), TSH, p37/Example 2.4.1 及p53/problem 2.9(i)）需找出与θ 无关的、所有 P_θ 共同的正则条件分布$P(A|t)$。

设 $t=(t_1,\ldots,t_n),t_1\le t_2\le\cdots\le t_n$，对各坐标排序可置换出$n!$ 个点。若其中有相同的坐标，置换出的$n!$ 个点中有些是相同的。对$\forall A\in\mathscr{B}_{\mathbb{R}^n}$ 定义

$$P(A|t)=\frac{(A\text{中包含的由}t\text{置换而得的点数})}{n!}.$$

则可验证，$\forall\theta\in\Theta,P_\theta(A\cap T^{-1}(B))=\int_B P(A|t)dP_\theta^T(t)$。所以$T$ 是充分统计量。

□

最小充分统计量

定义3.3 (最小充分统计量) $\mathscr{P}$ 是$(\mathscr{X},\mathscr{B}_X)$ 上的分布族，T 是$\mathscr{P}$ 的充分统计量，若对$\mathscr{P}$ 的任意一个充分统计量 U，都存在可测函数 $H(\cdot)$，使得 $T = H(U)$ $a.s.\mathscr{P}$，则称 T是$\mathscr{P}$ 的最小充分统计量。

定理3.4 (存在性，Bahadur, 1957，陈希孺《数理统计引论》p75/定理1.5.4) 对于欧氏空间上的分布族 $\mathscr{P}$，在因子分解定理的条件下，存在最小充分统计量。

定理3.5 (最小充分统计量的判断方法) 设$\mathscr{P}$ 是定义在欧氏空间$(\mathscr{X},\mathscr{B}_X)$ 上的分布族，ν 是该空间上的σ 有限测度。

(i) 如果$\mathscr{P}_0 \subset \mathscr{P}$ 且$a.s.\mathscr{P}_0$ 意味着$a.s.\mathscr{P}$。那么若T 是$\mathscr{P}$ 的充分统计量，且是$\mathscr{P}_0$ 的最小充分统计量，则T 也是$\mathscr{P}$ 的最小充分统计量。

(ii) 设$\mathscr{P} = \{P_i : dP_i(x)/d\nu(x) = f_i(x),\ i = 0, 1, \cdots\}$。记$f_\infty(x) = \sum_{i=0}^{\infty} c_i f_i(x)$，其中$c_i > 0$ 为常数，$i = 1, 2\cdots$，且$\sum_{i=1}^{\infty} c_i = 1$。记

$$T_i(x) = f_i(x)/f_\infty(x) I(f_\infty(x) > 0),\ i = 0, 1, \cdots .$$

则$T = (T_0(X), T_1(X), \cdots)$ 是$\mathscr{P}$ 的最小充分统计量。

而且，若$\nu(\{x:\ f_\infty(x) > 0 and\ f_0(x) = 0\}) = 0$，则

$$T = (f_1(X)/f_0(X), f_2(X)/f_0(X), \cdots)$$

是$\mathscr{P}$ 的最小充分统计量。

(iii) 设$\mathscr{P} = \{P_\theta : dP_\theta(x)/d\nu(x) = f_\theta(x), \theta \in \Theta\}$，$T(X)$ 是充分统计量，若

$$f_\theta(x) = f_\theta(y)\phi(x, y), \forall\theta \in \Theta \Longrightarrow T(x) = T(y),$$

则T 是$\mathscr{P}$ 的最小充分统计量。

证明. (i) 若 U 是$\mathscr{P}$ 的充分统计量，则由定义3.1知 U 也是$\mathscr{P}_0$ 的充分统计量。因 T 是$\mathscr{P}_0$ 的最小充分统计量，故存在$\mathscr{B}_X$ 可测函数 $H(\cdot)$，使得 $T = H(U)$ a.s.$\mathscr{P}_0$。再由a.s.$\mathscr{P}_0$ 意味着a.s.$\mathscr{P}$ 即得T 也是$\mathscr{P}$ 的最小充分统计量。

(ii) 由其定义易知，$f_\infty > 0$ a.s.$\mathscr{P}$。记$A = \{x:\ f_\infty(x) > 0\}$，则$P(A) = 1$，$\forall P \in \mathscr{P}$。记$g_i(T(x)) = T_i(x)$，则$dP_i(x)/d\nu(x) = g_i(T(x))f_\infty(x)$，$i = 0, 1, 2, \cdots$。由因子分解定理知，$T$ 是$\mathscr{P}$ 的充分统计量。

若S 是$\mathscr{P}$ 的充分统计量，则由因子分解定理知，存在$\sigma(S)$ 可测函数$\tilde{g}_i(\cdot)$与$\mathscr{B}_X$ 可测函数$h(\cdot)$，使得$dP_i(x)/d\nu(x)=\tilde{g}_i[S(x)]h(x),\ i=0,1,2,\cdots$。因此，

$$T_i(x)=f_i(x)/f_\infty(x)=\tilde{g}_iS[(x)]/\sum_{i=0}^{\infty}c_j\tilde{g}_j[S(x)],\ x\in A\setminus N_i,\ i=0,1,2,\cdots,$$

其中$\nu(N_i)=0$，故在$A\setminus(\cup_{i=1}^{\infty}N_i)$ 上T 是S 的函数，即T 是S 的函数a.s. $\mathscr{P}$。因此T 是$\mathscr{P}$ 的最小充分统计量。

用$f_0(x)$ 替代$f_\infty(x)$ 的情况类似可证。

(iii) 由定理3.4 知，$\mathscr{P}$ 存在最小充分统计量，记之为S。若存在可测函数$\psi(\cdot)$，使得$T(x)=\psi(S(x))$ a.s.$\mathscr{P}$，即说明T 是$\mathscr{P}$ 的最小充分统计量。由因子分解定理知，存在$\sigma(S)$ 可测函数$g_\theta(\cdot)$ 与$\mathscr{B}_X$ 可测函数$h(\cdot)$，使得

$$f_\theta(x)=g_\theta[S(x)]h(x)\quad \text{a.e.}\nu,\quad \forall\theta\in\Theta.$$

显然，$h(x)>0$ a.s. $\mathscr{P}$。记$A=\{x:\ h(x)>0\}$，对于$x,y\in A$ 使得$S(x)=S(y)$，则有

$$f_\theta(x)=g_\theta[S(x)]h(x)=g_\theta[S(y)]h(y)h(x)/h(y)=f_\theta(y)h(x)/h(y),\quad\forall\theta\in\Theta.$$

这意味着$T(x)=T(y)$，即证得了存在函数$\psi(\cdot)$ 使得$T(x)=\psi[S(x)],\ \forall x\in A$。余下只须再证$\psi(\cdot)$ 的可测性。由定义3.3 知，存在可测函数$g(\cdot)$，使得$S(x)=g[T(x)]$ a.s. $\mathscr{P}$。因此，$g(\cdot)$ 是一一对应的，$\psi=g^{-1}$，故ψ 可测。 □

例3.5 设$X_1,\ldots,X_n$ i.i.d. P_θ。

1. 正态总体。P_θ 为$N(\theta,1),\ \theta\in\mathbb{R}$。$(X_1,\ldots,X_n)$ 的jpdf为

   $$p_\theta(x)=(2\pi)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}x_i^2+n\bar{x}\theta-\frac{n\theta^2}{2}\right\},\quad\theta\in\mathbb{R},$$

   该联合分布族有共同支撑$\mathbb{R}^n$。

   取$\mathscr{P}_0=\{p_{\theta_1},p_{\theta_2}\},\theta_1\neq\theta_2$，由定理 3.5，$\mathscr{P}_0$的最小充分统计量为

   $$T(X)=p_{\theta_1}(X)/p_{\theta_2}(X)=\exp\left\{n\bar{X}(\theta_2-\theta_1)-\frac{n}{2}(\theta_2^2-\theta_1^2)\right\},$$

   它与$\bar{X}$ 一一对应。故$\bar{X}$ 关于$\mathscr{P}_0$ 最小充分。又因为$\bar{X}$ 关于$\mathscr{P}$ 充分，所以，$\bar{X}$ 关于$\mathscr{P}$ 最小充分。

2. 指数分布。P_θ 为$E(\theta,1)$，$\theta\in\mathbb{R}$。因为$(X_1,\ldots,X_n)$ 的jpdf 为 $p_\theta(x)=I(x_{(1)}>\theta)\exp(n\theta-\sum_{i=1}^n x_i)$，由因子分解定理知，$X_{(1)}$ 是充分统计量。再由定理 3.5(iii) 知，$X_{(1)}$ 是最小充分统计量。

3. 均匀分布。P_θ 为$U(\theta-\frac{1}{2},\theta+\frac{1}{2}),\theta\in(-\infty,+\infty)$。可证：$(X_{(1)},X_{(n)})$最小充分。(见Casella and Berger (2002), p282/eg 6.2.15 或Shao (2003), p107/Example 2.13)

4. logistic分布。$(X_1,\ldots,X_n)$ 的联合密度函数族为

$$p_\theta(x)=\prod_{i=1}^n\frac{\exp[-(x_i-\theta)]}{\{1+\exp[-(x_i-\theta)]\}^2},\ \theta\in(-\infty,\infty),$$

该分布族有共同支撑$\mathbb{R}^n$。可证：次序统计量就是最小充分统计量。

证明. 取$\mathscr{P}_0=\{p_\theta:\theta=0,\theta_1,\ldots,\theta_k,\ 其中\theta_j互不相同且均非0\},k=n+1$。由定理 3.5 知，$\mathscr{P}_0$ 的最小充分统计量为$T(X)=\Big(T_1(X),\cdots,T_k(X)\Big)$，其中

$$T_j(x)=p_{\theta_j}(x)/p_0(x)=e^{n\theta_j}\prod_{i=1}^n\left(\frac{1+e^{-x_i}}{1+e^{-x_i+\theta_j}}\right)^2,\quad j=1,\ldots,k.$$

再证，$T(X)$ 与次序统计量$(X_{(1)},\ldots,X_{(n)})$ 一一对应，即

$$T(x_1,\cdots,x_n)=T(y_1,\cdots,y_n)\iff(x_{(1)},\ldots,x_{(n)})=(y_{(1)},\ldots,y_{(n)}),$$

其中$(x_1,\cdots,x_n)\in\mathbb{R}^n$，$(x_{(1)},\cdots,x_{(n)})$ 为其从小到大的排列。

"$\Longleftarrow$"显然，只须证"$\Longrightarrow$"。由$T(x_1,\cdots,x_n)=T(y_1,\cdots,y_n)$ 知

$$\prod_{i=1}^n\frac{1+\xi u_i}{1+u_i}=\prod_{i=1}^n\frac{1+\xi v_i}{1+v_i},\ 对\ \xi=\xi_1,\ldots,\xi_k\ 成立,$$

其中 $u_i=e^{-x_i},v_i=e^{-y_i},i=1,\cdots,n$；$\xi_j=e^{\theta_j},j=1,\cdots,k$ 互不相同。再由代数学基本定理知，上面等式两边ξ^r 的系数相同，$r=0,1,\ldots,n$。由 ξ^0 的系数相同，得$\prod_{i=1}^n(1+u_i)=\prod_{i=1}^n(1+v_i)$，因而有

$$\prod_{i=1}^n(1+\xi u_i)=\prod_{i=1}^n(1+\xi v_i),\ 对\ \xi=\xi_1,\ldots,\xi_k成立.$$

令$\eta=1/\xi$，则有

$$\prod_{i=1}^n(\eta+u_i)=\prod_{i=1}^n(\eta+v_i),\ 对\ \eta=1/\xi_1,\ldots,1/\xi_k成立.$$

因此，$(u_1,\ldots,u_n)$ 是$(v_1,\ldots,v_n)$ 的一个排列，即$(x_1,\ldots,x_n)$ 是$(y_1,\ldots,y_n)$ 的一个排列。 □

同理可证，对于Cauchy族 $\left\{p_\theta(x)=\prod_{i=1}^{n}\frac{1}{\pi[1+(x_i-\theta)^2]},\ \theta\in\mathbb{R}\right\}$、双指数族 $\left\{p_\theta(x)=\prod_{i=1}^{n}\frac{1}{2}e^{-|x_i-\theta|},\ \theta\in\mathbb{R}\right\}$，次序统计量是最小充分的。

定理3.6 $\mathscr{P}=\left\{P_\theta:\ dP_\theta(x)/d\nu(x)=B(\theta)\exp[\sum_{j=1}^{s}\eta_j(\theta)T_j(x)]h(x),\theta\in\Theta\right\}$ 为指数族，其中 $(\eta_1(\theta),\cdots,\eta_s(\theta))$、$(T_1(x),\cdots,T_s(x))$ 均线性无关，则 $(T_1(X),\ldots,T_s(X))$ 是最小充分统计量。

证明. 由因子分解定理，充分性显然。下证最小充分。取 $\mathscr{P}_0=\{P_\theta\in\mathscr{P}:\theta=\theta_0,\ldots,\theta_s\in\Theta\}$，记$\eta^k=(\eta_1(\theta_k),\ldots,\eta_s(\theta_k)),k=0,1,\ldots,s$。由定理3.5 知，$\mathscr{P}_0$的最小充分统计量等价于

$$Y=\begin{pmatrix}\sum_{j=1}^{s}[\eta_j(\theta_1)-\eta_j(\theta_0)]T_j(X)\\ \vdots\\ \sum_{j=1}^{s}[\eta_j(\theta_s)-\eta_j(\theta_0)]T_j(X)\end{pmatrix}=AT,\text{其中 }A=\begin{pmatrix}\eta^1-\eta^0\\ \vdots\\ \eta^s-\eta^0\end{pmatrix},T=\begin{pmatrix}T_1\\ \vdots\\ T_s\end{pmatrix}.$$

因为$(\eta_1(\theta),\cdots,\eta_s(\theta))$ 线性无关，故存在$\theta_0,\ldots,\theta_s\in\Theta$ 使A 满秩。即Y 与T 一一对应，因此T 关于$\mathscr{P}_0$ 最小充分。再由$\mathscr{P}$ 有共同支撑知，T 关于$\mathscr{P}$ 最小充分。 □

推论3.2 设$X_i,i=1,\ldots,n$ *i.i.d.* $p_\theta(x)$, $\theta\in\Theta$，则在定理*3.6*条件下，$T=(\sum_{i=1}^{n}T_1(X_i),\ldots,\sum_{i=1}^{n}T_s(X_i))$ 是最小充分统计量。

§3.2 完备性

定义3.4 (附属(Ancillary)) 设X 的分布族为$\mathscr{P}=\{P_\theta:\theta\in\Theta\}$。若统计量$V(X)$ 的分布与 θ 无关，则称 $V(X)$ 是附属的。若$E_\theta V(X)$ 与 θ 无关，则称 $V(X)$ 是一阶附属的。

尽管附属信息不含未知参数的信息，但最小充分统计量中仍可能包含许多附属统计量。例如，logistic分布族$\mathscr{P}=\left\{p_\theta(x)=\prod_{i=1}^n\frac{\exp\{-(x_i-\theta)\}}{[1+\exp(-(x_i-\theta))]^2},\theta\in\mathbb{R}\right\}$，次序统计量$(X_{(1)},\ldots,X_{(n)})$ 是最小充分统计量，但是$X_{(n)}-X_{(i)},(i=1,\ldots,n-1)$ 是附属的。

充分统计量的降维程度，与其所含附属信息量的大小有关。若一个充分统计量的任意非常数函数是非附属的、甚至是非一阶附属的，那它简化数据最成功。

定义3.5 (完备性) 设 X 的分布族为$\mathscr{P}=\{P_\theta:\theta\in\Theta\}$。如果统计量$T$ 满足：对任意$\sigma(T)$ 可测函数 f 有

$$E_\theta f[T(X)]=0,\ \forall\theta\in\Theta\quad\Longrightarrow\quad f[T(x)]=0\ a.s.\ \mathscr{P},$$

则称 T 是完备的(Complete)。

注：

(1) 若定义中限定 $f(t)$ 有界，则称 T 是有界完备统计量。

(2) 若 T 完备，$S=\psi(T)$ a.s.$\mathscr{P}$，$\psi(\cdot)$ 可测，则 S 完备。

(3) 若$\mathscr{P}$ 存在最小充分统计量，则$\mathscr{P}$ 的完备充分统计量一般是最小充分的。反之不真。

(4) 完备性也可对分布族定义：

定义3.6 设X 的分布族为$\mathscr{P}=\{P_\theta:\theta\in\Theta\}$。若对任意可测函数$f(x)$ 有

$$E_\theta f(X)=0,\ \forall\theta\in\Theta\implies f(x)=0,\ \text{a.e.}\mathscr{P},$$

则称$\mathscr{P}$ 完备。

设由统计量T 从$(\mathscr{X},\mathscr{B}_{\mathscr{X}},\mathscr{P})$ 导出$(\mathscr{T},\mathscr{B}_T,\mathscr{P}^T)$。统计量$T$ 完备等价于$\mathscr{P}^T$ 完备。

定理3.7 *若X 的分布服从典则形式的指数族，pdf 形如(1.2)，参数空间$\tilde{\Theta}$ 有内点，则$T = [T_1(X), \ldots, T_s(X)]$ 是充分完备统计量。*

证明. T 的充分性由因子分解定理可知。现证其完备性。由指数族性质1.4 知，存在$(\mathbb{R}^s, \mathscr{B}_{\mathbb{R}^s})$ 上的σ 有限测度λ，使得$dP_\eta^T(t)/d\lambda(t) = A(\eta)\exp(\eta' t), \quad \eta \in \tilde{\Theta}$。因$\eta_0$ 是$\tilde{\Theta}$ 的内点，故有$\varepsilon > 0$，使得$N(\eta_0, \varepsilon) = \{\eta : \|\eta - \eta_0\| < \varepsilon\} \subset \tilde{\Theta}$。设存在可测函数$f(\cdot)$ 使得$E_\eta[f(T)] = 0,\ \forall \eta \in \tilde{\Theta}$，则有

$$\int_{\mathscr{T}} f(t)\exp(\eta' t)A(\eta)d\lambda(t) = 0, \quad \forall \eta \in N(\eta_0, \varepsilon).$$

分别用f^+, f^- 表示函数f 的正部与负部。因$A(\eta) > 0,\ \forall \eta \in \tilde{\Theta}$，故

$$\int_{\mathscr{T}} f^+(t)\exp(\eta_0' t)\exp(s' t)d\lambda(t) = \int_{\mathscr{T}} f^-(t)\exp(\eta_0' t)\exp(s' t)d\lambda(t), \quad \forall s \in N(0, \varepsilon).$$

这表明，非负函数$f^+\exp(\eta_0' t)$ 与$f^-\exp(\eta_0' t)$ 的Laplace 变换在0的一个邻域中相等。因此，$f^+(t) = f^-(t) = 0$ a.e. λ，即$f^+(t) = f^-(t) = 0$ a.s. $\mathscr{P}^T$。故T 是完备的。 □

例3.6 (1) 设X 的分布族为$\{B(n, \theta) : 0 < \theta < 1\}$，则$X$ 充分完备。

(2) 设X 的分布族为$\{\mathrm{Poi}(\lambda) :\ \lambda > 0\}$，则$X$ 充分完备。

(3) 设$X_1, \cdots, X_n$ i.i.d. $N(\xi, \sigma^2),\ \xi \in \mathbb{R}, \sigma^2 > 0$，则$(\overline{X}, S^2)$ 充分完备。

注: 对于弯曲指数族，$T_1, \ldots, T_s$ 往往最小充分，但不完备。如$X_1, \ldots, X_n$ i.i.d. $N(\xi, \xi^2),\ \xi \in \mathbb{R}$，$(\overline{X}, S^2)$ 最小充分，但不完备。

例3.7 (非指数族) (1) 设$X_1, \ldots, X_n$ i.i.d. $U(0, \theta),\ \theta > 0$，则$T = X_{(n)}$ 充分完备。

证明. 充分性由因子分解定理即得。现证完备性。设可测函数$f(t)$ 满足 $E_\theta f(T) = 0,\ \forall \theta > 0$。因 T 关于Lebesgue测度 λ 的pdf 为 $p_\theta(t) = nt^{n-1}/\theta^n I_{(0,\theta)}(t)$，故

$$\int_0^\theta f(t)t^{n-1}d\lambda(t) = 0, \quad \forall \theta > 0.$$

分别用f^+, f^- 表示函数f 的正部与负部，则有

$$\int_A f^+(t)t^{n-1}d\lambda(t) = \int_A f^-(t)t^{n-1}d\lambda(t), \quad \forall A \in \mathscr{B}_{\mathbb{R}^+}.$$

因此，$f^+(t) = f^-(t)$, $t \in \mathbb{R}^+ \setminus N$，其中$\lambda(N) = 0$。故$f(t) = 0$ $a.e.P_\theta$, $\forall\theta > 0$，即T 完备。 □

(2) 设$Y_1, \ldots, Y_n$ i.i.d. $E(\eta, 1)$，其共同的pdf 为$f(y) = e^{-(y-\eta)}I_{(\eta,\infty)}(y)$, $\eta \in \mathbb{R}$。则可证$Y_{(1)}$ 是充分完备统计量。

因为令$X_i = e^{-Y_i}, \theta = e^{-\eta}$，则$X_1, \ldots, X_n$ i.i.d. $U(0, \theta)$, $\theta > 0$。由$X_{(n)}$ 的充分完备性，可证得$Y_{(1)}$ 的充分完备性。

(3) 设$X_1, \ldots, X_n$ i.i.d. $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta \in \mathbb{R}$。则$T = (X_{(1)}, X_{(n)})$ 是最小充分的，但不完备。

因为$E_\theta[X_{(n)} - X_{(1)} - \frac{n-1}{n+1}] = 0$, $\forall\theta \in \mathbb{R}$，但$P_\theta[X_{(n)} - X_{(1)} - \frac{n-1}{n+1} = 0] = 0$, $\forall\theta > 0$。故T 不完备。

定理3.8 (截尾族的完备性) *设常数c_1, c_2 满足$-\infty \le c_1 < c_2 \le \infty$；$\phi(x)$ 是(c_1, c_2) 上的正值Borel可测函数；$X_1, \ldots, X_n$ i.i.d. $f(x;\theta)$，则*

(1) 若$f(x;\theta) = c(\theta)\phi(x)I_{(c_1,\theta)}(x)$, $\theta \in (c_1, c_2)$，其中$c(\theta) = [\int_{c_1}^{\theta}\phi(x)dx]^{-1}$，则$X_{(n)}$ 充分完备。

(2) 若$f(x;\theta) = c(\theta)\phi(x)I_{(\theta,c_2)}(x)$, $\theta \in (c_1, c_2)$，其中$c(\theta) = [\int_{\theta}^{c_2}\phi(x)dx]^{-1}$，则$X_{(1)}$ 充分完备。

(3) (Shao (2003), p149/Exercise 51.) 若$f(x;\theta) = c(\theta)\phi(x)I_{(\theta_1,\theta_2)}(x)$, $c_1 < \theta_1 < \theta_2 < c_2$，其中$\theta = (\theta_1, \theta_2)$, $c(\theta) = [\int_{\theta_1}^{\theta_2}\phi(x)dx]^{-1}$，则$(X_{(1)}, X_{(n)})$ 充分完备。

定理3.9 (Basu) *设对于分布族$\mathscr{P} = \{P_\theta : \theta \in \Theta\}$，$T$是充分完备统计量，$V$是附属统计量。则对于$P_\theta$，$T$与$V$独立，$\forall\theta \in \Theta$。(T 可弱化为：充分、有界完备统计量。)*

证明. 记V 的值域空间为 $(\mathscr{V}, \mathscr{B}_V)$。因 V 是附属统计量，故对$\forall A \in \mathscr{B}_V$，$p_A \hat{=} P(V \in A)$ 与θ 无关。记$\eta_A(t) = P(V \in A|T = t)$，因$T$ 是充分统计量，故$\eta_A(t)$ 与θ 无关，$\eta_A[T(X)]$ 是一个统计量。因$p_A = E_\theta[\eta_A(T)]$, $\forall\theta \in \Theta$，由T的完备性知，$\eta_A[T(X)] - p_A = 0$ $a.s.\mathscr{P}$。这说明，对于$\forall\theta \in \Theta$，条件分布$V|T$ 与T 无关，即T与V独立。 □

例3.8 设$X_1,\ldots,X_n$ i.i.d. $E(a,b)$，即其共同的pdf 为

$$p_{a,b}(x)=\frac{1}{b}\exp\big\{-\frac{x-a}{b}\big\}I_{(a,\infty)}(x),\quad a\in\mathbb{R},b>0.$$

记$T_1=X_{(1)}$，$T_2=\sum_{i=1}^n(X_i-X_{(1)})$，则

(1) (Shao (2003), p144/20, p150/59) $T_1\sim E(a,\frac{b}{n})$，$2T_2/b\sim\chi^2(2n-2)$，且对$\forall a,b$，T_1 与T_2 相互独立。

(2) (T_1,T_2) 充分完备。

证明. (1) T_1,T_2 的分布易由次序统计量的联合密度函数求得。记$\mathscr{P}_b=\{P_{a,b}\in\mathscr{P}:\ b\text{固定},a\in\mathbb{R}\}\subset\mathscr{P}$。对于$\mathscr{P}_b$，由定理3.8知，$T_1$ 充分完备。而T_2 的分布与a 无关，对于$\mathscr{P}_b$ 是附属的。由Basu 定理知，对于固定的b 及$\forall a\in\mathbb{R}$，T_1 与T_2 独立。因而对于$\forall a\in\mathbb{R},b>0$，$T_1$ 与T_2 独立。

(2) (T_1,T_2) 的充分性由因子分解定理即得。现证完备性。设可测函数$f(t_1,t_2)$ 满足 $E_{(a,b)}f(T_1,T_2)=0$, $\forall a\in\mathbb{R},b>0$。记

$$g(t_1,b)=E_bf(t_1,T_2)=\int_0^\infty f(t_1,t_2)p_b(t_2)d\lambda(t_2),$$

其中λ 为Lebesgue测度，$p_b(t_2)$ 是T_2 的pdf。对于固定的$b>0$，由

$$E_ag(T_1,b)=\int_a^\infty g(t_1,b)\frac{n}{b}\exp\{-\frac{n(t_1-a)}{b}\}d\lambda(t_1)=0,\ \forall a\in\mathbb{R},$$

及T_1 对于$\mathscr{P}_b$ 的完备性知，存在N_b，$\lambda(N_b)=0$，$g(t_1,b)=0$, $\forall t_1\in\mathbb{R}\setminus N_b$。由Fubini定理，

$$0=\int_0^\infty\int_{-\infty}^\infty|g(t_1,b)|d\lambda(t_1)d\lambda(b)=\int_{-\infty}^\infty\int_0^\infty|g(t_1,b)|d\lambda(b)d\lambda(t_1),$$

故存在$N_1,\lambda(N_1)=0$，对于固定的$t_1\in\mathbb{R}\setminus N_1$，有$g(t_1,b)=0$, $a.e.\lambda$。因为$\{p_b(t_2):\ b>0\}$ 是指数族，由指数族性质1.3 知，$g(t_1,b)$ 关于b 连续，因此，$g(t_1,b)=0$, $b>0,t_1\in\mathbb{R}\setminus N_1$。再由$T_2$ 的完备性以及$g(t_1,b)=E_bf(t_1,T_2)=0,\forall b>0$ 知，对于固定的$t_1\in\mathbb{R}\setminus N_1$，$f(t_1,t_2)=0,\forall t_2\in\mathbb{R}^+\setminus N_{t_1}$。因此，$f(t_1,t_2)=0$ a.s.$\mathscr{P}$，即(T_1,T_2) 完备。 □

附属统计量提供信息的例子见教材p285/eg6.2.20。

应用Basu定理证明独立性的例子还见教材p288/eg6.2.26，eg6.2.27。

§3.3 统计分析的一些基本原则

充分性原则： 设T 是$\mathscr{P}$ 的充分统计量，若观测x_1, x_2 使得$T(x_1) = T(x_2)$，则由x_1, x_2 作出的推断应相同。

似然函数可视为一种特殊的“统计量”。使用似然函数作推断有多种方法。本节讨论，似然原则及其与其他几条数据简化原则之间的关系。

定义3.7 (似然函数(likelihood function)) 设X 的分布族$(\mathscr{X}, \mathscr{B}_X, \mathscr{P})$ 关于σ 有限测度ν 可控，

$$\mathscr{P} = \{P_\theta : \ dP\theta/d\nu = f(x|\theta), \theta \in \Theta\}.$$

对于给定的$x \in \mathscr{X}$，将$f(x|\theta), \theta \in \Theta$ 看作θ 的函数，则称之为$X = x$ 的似然函数，记作$L(\theta|x)$。

对似然函数的解释

似然原则： 若$x, y \in \mathscr{X}$，使得

$$L(\theta|x) = C(x, y)L(\theta|y), \quad \forall \theta \in \Theta,$$

其中$C(x, y)$ 与θ 无关，那么由x, y 得出的统计推断结论应相同。

Fiducial Inference (Fisher, 1930).

p291/eg6.3.3 正态总体均值的Fiducial Interval.

Evidence Function $E = (X, \theta, \{f(x|\theta)\}$ 表示一个试验，$\mathrm{Ev}(E, x)$ 表示E, x 中包含的关于θ的证据。

正式充分性原则(Formal Sufficiency Principle)：对于E 而言，$T(X)$ 是充分统计量。若$x, y \in \mathscr{X}$ 满足$T(x) = T(y)$，那么$\mathrm{Ev}(E, x) = \mathrm{Ev}(E, y)$。

条件原则(Conditionality Principle)

正式似然原则(Formal Likelihood Principle)

定理3.10 (Birnbaum's Theorem) 正式充分性原则与条件原则可导出正式似然原则，反之亦然。

采用不满足正式似然原则的理由：

1. 充分性依赖于模型；

2. 作模型诊断时往往不采用充分统计量；

3. Birnbaum's Theorem 的证明不是很令人信服。

这些简化数据的原则有数学方面的优点，在有些场合是适用的，但并不适用于所有场合、并不为统计工作者广泛接受。

同变原则

有些统计问题具有某种“对称”结构。根据具有“对称”关系的样本点x, y去作统计推断，若希望得出的结论之间也具有相应的联系，对所使用的推断方法提出的这类要求往往称为“同变性”要求。

例（对称性的含义）： 设用n 次独立的Bernoulli 试验结果$X_1, \ldots, X_n$ 去估计成功概率$p \in (0,1)$。

$X = (X_1, \ldots, X_n)$ 的jpdf 族为

$$\mathscr{P} = \{p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i} : \ p \in (0,1)\}.$$

若把“成功”与“失败”的标签互换一下，那么所得到的“成功”观测为$X_i' = 1 - X_i$，$i = 1, \ldots, n$，所关心的成功概率变为$p' = 1 - p$。而$X' = (X_1', \ldots, X_n')$ 的jpdf 族为

$$\{p'^{\sum_{i=1}^n X_i'}(1-p')^{n-\sum_{i=1}^n X_i'} : \ p' \in (0,1)\}.$$

仍然是$\mathscr{P}$，与原问题完全一样。因此，用X 估计p 与用X' 估计p'，可称为是两个对称的问题。

在原问题中若用$\delta(X)$ 去估计p，那么对p' 有两种自然的估计：

(i) $1 - \delta(X)$ （通常的估计）；

(ii) $\delta(X')$（在对称问题中采用形式对称的估计量）。

在某些应用背景下，要求估计量满足下列条件是合适的：

$$1-\delta(X)=\delta(X').$$

我们称之为同变估计。比如，常用的估计量$\bar{X}$ 就满足上述条件。

需要注意的是，$\delta(X')$ 用作p' 的估计的合理性，完全依赖于数据变化前后两个估计问题的对称性。如果没有对称性，那么用$\delta(X')$ 估计$1-p$ 是不合适的。例如，若将原问题的参数空间$(0,1)$ 改为$(2/3,1)$，则变换后问题的参数空间为$(0,1/3)$，这两个问题不再对称。此时，若$\delta(X)$ 是p 的合理估计，那么应该有$\delta(X)>\frac{2}{3}$。故$\delta(X')>\frac{2}{3}$，若再用它估计$p'(<\frac{1}{3})$ 就显然不合适了。而用$1-\delta(X)$ 作为p' 的估计是自然的，无须建立在问题对称性的基础之上。

同变性的两重含义：度量同变性；形式不变性。

同变性原则：若$Y=g(X)$，Y 与X的模型相同，那么统计推断方法应当同时满足度量同变性与形式不变性。

§4 统计推断概述

假定产生观测数据x 的模型为$(\mathscr{X}, \mathscr{B}_X, \mathscr{P}=\{P_\theta:\ \theta\in\Theta\})$。基本的统计推断问题有

- 估计问题：估计未知参数θ 的某个实函数(或实向量) $\gamma(\theta)$ 的值（点估计、区间估计）；
- 检验问题：对未知参数$\theta\in\Theta_0\subset\Theta$ 或$\theta\overline{\in}\Theta_0$ 作出判断；
- 预测问题等。

数理统计主要研究解决这些问题的原则与方法，以及如何评价方法的优劣。

§4.1 点估计

一般地，可视待估参数(estimand) $\gamma(\theta):\Theta\to\mathbb{R}^k$。通常用一个与之同维的统计量$\delta(X)$ 去估计。用于点估计的统计量常称为估计量(estimator)。将具体观测值x 代入估计量算得的值$\delta(x)$ 称为估计值(estimate)。

构造估计量的方法(estimation)有：矩法、极大似然法、Bayes法、最小二乘法等等。

如何评价估计量的优劣？不能只看一个θ、一个x 处的误差，应看“平均”误差。这涉及到“如何衡量误差”、“如何平均”两个问题。用如下一些量度评价估计量是很自然的：

(1) 偏倚(bias)：$\mathrm{bias}_\theta(\delta)=E_\theta[\delta(X)-\gamma(\theta)]$；

(2) 均方误差(Mean Squared Error)：

$$\mathrm{MSE}_\theta(\delta)=E_\theta[\delta(X)-\gamma(\theta)]^2=\mathrm{bias}_\theta^2(\delta)+\mathrm{Var}_\theta(\delta);$$

(3) $E_\theta|\delta(X)-\gamma(\theta)|$、$E_\theta\big[\frac{\delta(X)-\gamma(\theta)}{\gamma(\theta)}\big]^2$、$P_\theta(|\delta(X)-\gamma(\theta)|>c)$ 等等。

考察估计量的优劣不能只看它在一个θ 处的表现，应该看它在$\theta\in\Theta$ 上的整体表现。

若从均方误差角度评价估计量。设$\delta_1(X), \delta_2(X)$ 是$\gamma(\theta)$ 的两个估计量，若

$$\mathrm{MSE}_\theta(\delta_1) \leq \mathrm{MSE}_\theta(\delta_2), \quad \forall \theta \in \Theta,$$

则称δ_1 不次于δ_2。若还$\exists \theta_0 \in \Theta$，使不等式严格成立，则称$\delta_1$ 优于δ_2。

当然，如果能找到一个估计量不次于任何别的估计量，那最理想了。可是，如果对估计量不加限制的话，这做不到。假如δ^* 是一个均方误差意义下最优的估计量，即对任意一个估计量δ，有

$$\mathrm{MSE}_\theta(\delta^*) \leq \mathrm{MSE}_\theta(\delta), \quad \forall \theta \in \Theta.$$

取$\delta(x) \equiv g(\theta_0),\ \forall x \in \mathscr{X}$，其中$\theta_0$ 是Θ 中的一个元素。这是一个极端的估计量，它用一个常数去估计所有的$\gamma(\theta)$。虽然它在$\theta \neq \theta_0$ 时的表现很差，但是，它在θ_0 处的均方误差为0。δ^* 的最优性就要求$\mathrm{MSE}_{\theta_0}(\delta^*) = 0$，对$\forall \theta_0 \in \Theta$ 都成立，即$\delta^*(X) = \gamma(\theta)$ a.s.$P_\theta,\ \forall \theta \in \Theta$。若存在$\theta_1 \neq \theta_2$ 及$A \in \mathscr{B}_X$ ，使得$\gamma(\theta_1) \neq \gamma(\theta_2)$ 且$P_{\theta_1}(X \in A) P_{\theta_2}(X \in A) > 0$，那么就不存在这样的$\delta^*$。

在应用中，往往倾向于使用无“系统偏差”的估计量，即无偏估计。其优点在于，它在每个θ 处的正负误差在平均意义相抵消，对每个θ 都不存在偏见。

定义4.1 (无偏估计) 若估计量$\delta(X)$ 满足

$$E_\theta \delta(X) = \int_{\mathscr{X}} \delta(x) dP_\theta(x) = \gamma(\theta),\ \forall \theta \in \Theta,$$

则称δ 为$\gamma(\theta)$ 的无偏估计(Unbiased Estimator, U.E.)。

并不是所有的待估参数都存在无偏估计（Shao (2003), p154/84）。无偏估计存在时，往往不唯一，常需要根据其他准则（如均方误差）从中选择较优的估计。无偏估计量的偏倚恒为0，其均方误差恒等于其方差。

例4.1 (无偏估计的例子) (1) Shao (2003), p123/Example 2.26. $X_1, \cdots, X_n$ i.i.d. F，考虑给定t 处$\gamma(F) = 1 - F(t)$ 的无偏估计。若对F 的形式不作假定，那么

$$\hat{\gamma}_1(X) = 1 - F_n(t) = \frac{1}{n} \sum_{i=1}^{n} I_{(t,\infty)}(X_i)$$

是$\gamma(F)$ 的一个自然的U.E.，其中$F_n(\cdot)$ 是样本的经验分布函数。$\hat{\gamma}_1(X)$的均方误差为

$$MSE_F(\hat{\gamma}_1(X)) = \mathrm{Var}_F(\hat{\gamma}_1(X)) = \frac{1}{n} F(t)[1 - F(t)].$$

若假定$X_1 \sim E(0,\theta)$，即$F(t) = [1-\exp(-t/\theta)]I_{(0,\infty)}(t), \theta > 0$，此时$\overline{X}$ 是充分完备统计量，取

$$\hat{\gamma}_2(X) = E[1 - F_n(t)|\overline{X}],$$

则在此参数模型假定下$\hat{\gamma}_2(X)$ 也是$\gamma(F)$ 的U.E.，且方差小于$\mathrm{Var}_F(\hat{\gamma}_1(X))$。这说明缩小模型假定的范围，可能提高参数估计的效率。但是，需要小心的是，缩小模型假定的范围可能导致模型误设的风险。

(2) Shao(2003), p124/Example 2.27. 对有限总体实施不放回简单随机抽样情况下，总体总和的无偏估计及其方差。

在回答实际应用中的点估计问题时，通常需要报告：1. 模型假定，导出估计量的方法与估计量；2. 根据样本观测数据计算得到的估计值；3. 标准误(standard error)——估计量的标准差的估计。

例如，$X_1, \cdots, X_n$ i.i.d.，记$E(X_1) = \mu$, $\mathrm{Var}(X_1) = \sigma^2$。常用$\overline{X}$ 估计μ，它是无偏的，标准差为$\mathrm{Std}(\overline{X}) = \sigma/\sqrt{n}$。若用$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2}$ 估计σ，那就得到了$\overline{X}$ 的一个标准误$\mathrm{se}(\overline{X}) = \hat{\sigma}/\sqrt{n}$。在应用中，将观测数据$x$ 代入$\overline{X}$、$\mathrm{se}(\overline{X})$ 可得μ的估计值及其误差的估计值。

§4.2 假设检验

假设检验(Hypothesis Testing) 是另一类常用的统计推断形式，其理论是Neyman-Pearson 在1920∼ 1930 年间建立起来的。N-P理论的核心是将假设检验问题归结为一个数学优化问题，通过求解这一优化问题获得最优的统计检验方法。该理论同样也可以纳入Wald 的统计决策框架之中。但是，N-P 优化问题仅在有限的场合下才易求解，更多实际的检验问题往往得从其他角度解决。通常的做法是，根据某种直观的想法构造检验法，然后借助于N-P 理论的基本概念，考察其性质。

1. 假设检验问题

实际生活中，经常需要根据不完全的信息对某个命题（表示某种潜在的可能情况）作出是或非的判断。例如，

◇ 通过抽样检验，判断某厂生产的一批产品质量是否符合交货标准；

◇ 通过试验，判断某药物治疗某疾病是否有疗效；等等

这样的问题可归结为假设检验问题。

假设是关于X 的真实分布P 或真实分布参数θ 的一个命题，通常用H_0、H_1 等符号表示。

由一个假设的真或假，可将X 的分布族$\mathscr{P}$ 分割成互斥的两个子集，记为$\mathscr{P}_0$、$\mathscr{P}_1$。相应地，参数空间Θ 也分割成互斥的两个子集Θ_0、Θ_1。

解决假设检验问题，就是要找到一套方法或规则，使得使用者可以根据X 的观测数据，在两个对立的假设H_0、H_1 之间作出判断。其中一个称为原假设(null hypothesis)，常用$H_0 : P \in \mathscr{P}_0$ 或$H_0 : \theta \in \Theta_0$ 表示；其对立假设称为备择假设(alternative hypothesis)，常用$H_1 : P \in \mathscr{P}_1$ 或$H_1 : \theta \in \Theta_1$ 表示。一个假设检验问题常表示为

$$H_0:\ P \in \mathscr{P}_0 \text{ v.s. } H_1:\ P \in \mathscr{P}_1 \quad \text{或者} \quad H_0:\ \theta \in \Theta_0 \text{ v.s. } H_1:\ \theta \in \Theta_1.$$

一般地，Θ_0、$\Theta_1 \neq \emptyset$，$\Theta_0 \cup \Theta_1 = \Theta$。也允许$\Theta_0 \cup \Theta_1 \neq \Theta$，这时$\Theta_1$ 表示在H_0 不成立时研究者最关心的情况。

2. 检验(test)

检验就是假设检验问题的解，即根据观测数据x，对原假设H_0 作出接受(accept) 或拒绝(reject) 判断的一套规则。

对于任意一个给定的x，若作出的接受或拒绝原假设的判断是明确的，则称检验是非随机化的。一个非随机化检验，相当于将样本空间$\mathscr{X}$ 分割成拒绝域(Rejection Region/Critical Region) W 和接受域$\overline{W}$ 两个区域。当$x \in W$ 时拒绝H_0，否则接受H_0。因此，一个非随机化检验与一个拒绝域W 一一对应。

3. 两类错误

由于不同的P_θ 可能产生相同的x，所以，一般来说每一个检验都难以避免犯错误。错误分两类：

$$\begin{array}{ll} \text{第I 类错误(拒真)：} & \theta \in \Theta_0\text{，但 } x \in W; \\ \text{第II 类错误(受伪)：} & \theta \in \Theta_1\text{，但 } x \in \overline{W}. \end{array}$$

犯两类错误的概率大小是衡量一个检验优劣的自然标准：

$$\begin{array}{ll} \text{犯第I类错误的概率：} & P_\theta(X \in W),\ \theta \in \Theta_0; \\ \text{犯第II类错误的概率：} & P_\theta(X \in \overline{W}) = 1 - P_\theta(X \in W),\ \theta \in \Theta_1. \end{array}$$

犯两类错误的概率都可用如下的*势函数*(power function)表示出来

$$g_W(\theta) = P_\theta(X \in W),\ \theta \in \Theta.$$

4. Neyman-Pearson 原则

理想的检验是使两类错误同时最小化的检验。但在样本容量一定时，一般不可能做到。Neyman 和Pearson 提出，在控制犯第I类错误概率的前提下，最小化犯第II类错误的概率。即设定一个$\alpha \in (0,1)$，在满足

$$P_\theta(X \in W) \le \alpha,\ \forall \theta \in \Theta_0$$

的检验中寻找最小化$P_\theta(X \in \overline{W})$，$\forall \theta \in \Theta_1$，也即最大化$P_\theta(X \in W)$，$\forall \theta \in \Theta_1$ 的检验。其中α 称为*显著性水平*(level of significance)。

根据N-P 原则，显著性水平α 下的最优检验，就是下列优化问题的解

$$\max_{W \in \mathscr{B}_X} P_\theta(X \in W),\ \forall \theta \in \Theta_1$$
$$\text{s.t. } \sup_{\theta \in \Theta_0} P_\theta(X \in W) \le \alpha.$$

一个检验的水平是不唯一的，若α 是其水平，则任意$\alpha' \in [\alpha, 1]$ 都是该检验的水平。称

$$\sup_{\theta \in \Theta_0} P_\theta(X \in W)$$

为检验的*真实水平*(size of the test)。

对于某个$\theta \in \Theta_1$，称$P_\theta(X \in W)$ 为检验在θ 处的*功效*（*势*，power）。

例4.2 (假设检验的基本概念) 参见p126/Example 2.28。

5. 水平的选取与p 值

显著性水平的选取是主观的。没有一个客观的标准保证哪个具体的水平可以接受。水平的选取也因实际应用不同而异。习惯上常取$\alpha = 0.10, 0.05$ 或0.01。α 取得越小，就要求拒绝H_0 时理由越充分。

确定一个合适的检验仅顾及α 不行，还得看功效。要两者兼顾的话，须在设计试验时考虑样本容量的大小。同时限定犯两类错误的概率分别不超过给定水平的假设检验，在质量管理领域有实际应用。

在实际应用中，通常有两种方式报告假设检验的结论：一是给出在某个给定水平下接受或拒绝原假设的结论；二是给出p值。统计软件一般给出p 值。简而言之，p值是指根据当前观测数据x 按某种形式的检验规则能拒绝原假设的最小显著性水平。使用p值至少有两方面的优点。一、在给定检验规则形式的前提下，可以让用户方便地作出不同显著性水平下的拒绝或接受原假设的检验结论，方法为

$$\text{当p 值 } \le \alpha \text{ 时，拒绝原假设；否则接受原假设。} \tag{4.1}$$

二、提供了当前观测数据对原假设支持程度的一种量度。p值越小，说明数据与原假设之间的矛盾程度越大。

我们可视 (4.1) 为一种“标准化”的检验法，而p值就相当于“标准化”的检验统计量。如果解决某假设检验问题的检验法，其拒绝域W_α 与显著性水平α 之间呈严格单调“增”的关系，即当$\alpha < \alpha'$ 时有$W_\alpha \subset W_{\alpha'}$，那么就可能构造出相应的p值，从而得到与之等价的p值检验法。例如，设$X_1, \cdots, X_n$ i.i.d. $N(\mu, \sigma^2)$，其中$\mu \in \mathbb{R}, \sigma > 0$，通常采用t法检验$H_0: \mu \le \mu_0$ v.s. $H_1: \mu > \mu_0$，显著性水平α 的拒绝域为

$$W_\alpha = \left\{ (x_1, \cdots, x_n):\ T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \ge t_{1-\alpha}(n-1) \right\}.$$

该规则的临界值$t_{1-\alpha}(n-1)$ 关于α 严格减，故拒绝域关于α 严增。因$t(n-1)$的分布函数$F_{t(n-1)}(t)$ 严增，将它同时作用于等式两边即得

$$W_\alpha = \left\{ (x_1, \cdots, x_n):\ 1 - F_{t(n-1)}(T) \le \alpha \right\},$$

故t检验法对应的p值为 $p(X) = 1 - F_{t(n-1)}(T)$。其他常用检验法对应的p 值同理可得。

值得注意的是，p值是统计量，其构造与初始的检验法有关，是初始检验统计量的“标准化”。不同检验法对应的p值是不同的，但是采用p 值作检验的规则是一样的，都是 (4.1)。有关p值的论述，参见Lehmann *et al* (2005), p63 或Casella *et al* (2002), p397.

6. 随机化检验

在得到x 之后，若允许进行这样的随机化判断：以概率r_x 拒绝H_0，以概率$1 - r_x$ 接受H_0，$r_x \in [0, 1]$ 与x 有关但与θ 无关。这就是所谓的随机化检验。若用1 表示拒绝H_0，0 表示接受H_0，那么一个随机化检验可用一个取值于$[0, 1]$ 上的检验函数(critical function) $\phi(x) = r_x, x \in \mathscr{X}$ 表示。

非随机化检验可视为随机化检验的特例，即当$x \in W$ 时以概率1 拒绝H_0，否则以概率1 接受H_0，其检验函数$\phi(x) = I_W(x)$ 是示性函数。引进随机化检验便于理论研究，也有一定的应用背景。

引进随机化检验后，势函数推广为

$$g_\phi(\theta) = E_\theta\phi(X) = \int_{\mathscr{X}} \phi(x)dP_\theta(x),\ \forall\theta \in \Theta.$$

在N-P 理论中，势函数包含了检验的全部性质。若两个检验的势函数相同，则称它们等价。

若记$\mathscr{F} = \{\phi:\ \phi$为$(\mathscr{X}, \mathscr{B}_X) \to ([0,1], \mathscr{B}_{[0,1]})$的可测函数$\}$，则按N-P 原则，解决假设检验问题就相当于求解下列优化问题：

$$\begin{aligned} &\max_{\phi\in\mathscr{F}}\ E_\theta\phi(X),\ \forall\theta \in \Theta_1 \\ &\text{s.t.}\ \ \sup_{\theta\in\Theta_0} E_\theta\phi(X) \le \alpha. \end{aligned} \tag{4.2}$$

§4.3 置信区域

设$\gamma(\theta) : \Theta \to \mathbb{R}^k$ 是一个未知的k 维参数，求其置信区域，就是要根据观测数据x，在$\mathbb{R}^k$ 中选用一个合适的区域$C(x)$ 去概括它的可能范围。这也是一类常见的统计推断问题。将$C(X)$ 视为一种方法，通常对其有两方面的要求：1)可靠度，即用$C(X)$ 去概括$\gamma(\theta)$，得有一定的把握；2)精确度，即要求$C(X)$ 的某种几何测度平均而言较小。构造置信区域通常的思路是，先规定一个可靠度的下限（即置信水平），从可靠度达到该置信水平的一类区域中挑选一个尽可能精确的区域。

定义4.2 设$\gamma(\theta) : \Theta \to \mathbb{R}^k$ 是一个未知的k 维参数，对于$\forall x \in \mathscr{X}$，$C(x)$ 是$\mathbb{R}^k$ 中的一个区域，与θ 无关。若$C(X)$ 满足

$$\inf_{\theta\in\Theta} P_\theta\big[\gamma(\theta) \in C(X)\big] \ge 1 - \alpha,$$

则称$C(X)$ 为参数$\gamma(\theta)$ 的置信水平为$1-\alpha$ 的置信区域。若等号成立，则称$1-\alpha$ 为$C(X)$的置信系数。

注意，若区域$C(X)$ 的置信水平为$1-\alpha$，那仅仅意味着对于每一个θ，假如反复抽取x 那么$C(x)$ 包含$\gamma(\theta)$ 的频率不小于$1-\alpha$。对于一个x 而言，$C(x)$ 要么包含$\gamma(\theta)$，要么不包含。

对于一维的未知参数，常用的置信区域是如下三种形式的区间

$$(1)\ [\hat\gamma_L(X), \hat\gamma_U(X)]; \qquad (2)\ (-\infty, \hat\gamma_U(X)]; \qquad (3)\ [\hat\gamma_L(X), \infty).$$

其中的$\hat\gamma_L(X), \hat\gamma_U(X)$ 都是统计量。(1)中的区间称为$\gamma(\theta)$ 的置信区间，一般要求它满足$-\infty < \hat\gamma_L(X) < \hat\gamma_U(X) < \infty$ a.s.$\mathscr{P}$；(2)中的$\hat\gamma_U(X)$ 称为$\gamma(\theta)$ 的置信上限；(3) 中的$\hat\gamma_L(X)$ 称为$\gamma(\theta)$ 的置信下限。对于(1)，常用平均长度$E_\theta[\hat\gamma_U(X) - \hat\gamma_L(X)]$ 度量其精度；对于(2)、(3)，常用$E_\theta[\hat\gamma_U(X) - \gamma(\theta)]$、$E_\theta[\gamma(\theta) - \hat\gamma_L(X)]$ 度量其精度。

例4.3 (正态总体均值的置信区间) 参见Shao,J. p129/Example 2.31。

假设检验与置信区域之间存在密切的联系。我们用一维参数的情形来说明，一般的结果在第7章中讨论。若对$\forall \gamma_0 \in \{\gamma(\theta):\ \theta \in \Theta\}$，检验

$$H_0 : \gamma(\theta) = \gamma_0 \quad \text{v.s.} \quad H_1 : \gamma(\theta) \neq \gamma_0 \tag{4.3}$$

的水平为α 的接受域都可以表示为

$$\{x:\ \hat\gamma_L(x) < \gamma_0 < \hat\gamma_U(x)\},$$

其中$\hat\gamma_L(X), \hat\gamma_U(X)$ 是统计量，则$(\hat\gamma_L(X), \hat\gamma_U(X))$ 为$\gamma(\theta)$ 的置信水平为$1-\alpha$ 的置信区间。反之，若$(\hat\gamma_L(X), \hat\gamma_U(X))$ 为$\gamma(\theta)$ 的置信水平为$1-\alpha$ 的置信区间，那么下列规则就是 (4.3)的一个水平为α的检验

$$\text{当 } \gamma_0 \overline{\in} (\hat\gamma_L(X), \hat\gamma_U(X)) \text{ 时，拒绝} H_0\text{；否则接受} H_0.$$

例4.4 (正态总体均值、方差的置信区域) 参见Shao, J. p130/Example 2.32。

§5 统计决策概述

1940年代，Abraham Wald 提出统计决策理论，把统计推断问题抽象地看作是人和自然的一种“博弈”，把多种形式的统计推断问题（如点估计、区间估计、假设检验、预测等）归结为统计决策问题（也即约束最优化问题），纳入统一的框架下研究。这套理论对统计学的发展起到了一定的推动作用。

1. 统计决策的三要素

(1) 模型（或称样本空间与分布族、统计结构）；

(2) 行动空间；

(3) 损失函数。

行动空间

对一个统计决策问题的回答，因具体的问题而异。例如，对于点估计问题，它应该是一个数或一个向量；对于区间估计问题，它应该是一个区间；对于假设检验问题，它是一个拒绝或接受的决定。一般地，对于统计决策问题的一个具体回答称为一个**行动(action)**。

对于一个给定的统计决策问题应该采取哪种行动，与所研究的问题、所获样本观测及统计推断的方式有关。但可能采取的全部行动，一般事先可以知道。称所有可能行动组成的集合$\mathscr{A}$ 为**行动空间**。出于数学处理的需要，有时也需要假定行动空间具有某种可测结构$(\mathscr{A},\mathscr{B}_A)$。

例5.1 在例1.1中，

- 若要估计b，则可取$\mathscr{A}=\mathbb{R}$ 或$\mathscr{A}=\mathbb{R}^+$;
- 若要求b 的区间估计，则可取$\mathscr{A}=\{[a_1,a_2]:\ a_1\le a_2\}$;
- 若要同时估计b,σ, 可取$\mathscr{A}=\{(a_1,a_2):-\infty<a_1<\infty,a_2>0\}$;
- 对于假设检验问题$H_0:b\ge b_0$ v.s. $H_1:b<b_0$，一般取$\mathscr{A}=\{0,1\}$，0表示接受H_0，1 表示拒绝H_0。

损失函数

对于一个统计决策问题采取任何行动，都会导致一定的后果（经济的或其他的)。统计决策理论的一个基本假定：行动的得失最终都可以量化，可采用**损失函数(Loss function)**来度量。

损失函数常用$L(\theta,a)$ 表示，它是一个$\Theta\times A\to[0,\infty)$ 的二元非负实函数，描述的是当参数为θ 时采取行动a 所导致的损失。对损失函数最基本的数学要求有：

(1) 非负。

(2) $\forall\theta\in\Theta,\exists$ 唯一$a\in A$，使得$L(\theta,a)=0$。

(3) 可测性：给定θ，$L(\theta,\cdot)$ $\mathscr{B}_A$可测；$\mathscr{B}_\Theta\times\mathscr{B}_A$可测。

(4) 有些问题中要求：给定θ，L 关于a 是凸函数。

例5.2 例1.1中，

- 对于估计b，常用的损失函数有$L(\theta,a)=(a-b)^2$，$|a-b|$, 或$\left(\frac{a-b}{b}\right)^2$ 等等。更一般地，可考虑$L(\theta,a)=c(\theta)\rho(a-b)$，其中$c(\theta)>0$，$\rho(t)$ 是偶函数，在$t\geq 0$ 上非降。
- 对于检验问题$H_0:b\geq b_0$ v.s. $H_1:b<b_0$，可取

$$L(\theta,a)=\begin{cases} c_1, & \text{if } b\geq b_0, a=a_1,\\ c_2, & \text{if } b<b_0, a=a_0,\\ 0, & \text{otherwise.}\end{cases}$$

- 对于求b 的区间估计的决策问题，自然地损失应包括两个方面：

  区间$a=[a_1,a_2]$ 是否包含b？b离a多远？（可靠性）

  $a=[a_1,a_2]$的长度（精确性)。

  可取$L(\theta,a)=c[1-I_{[a_1,a_2]}(b)]+(a_2-a_1)$ 或$|a_1-b|+|a_2-b|$ 等等。

从应用统计的角度看，损失函数很可能是三要素中最难确定的一个要素。原因有：

(1) 很难确切地评价一个行动的长期或短期后果。

(2) 与研究的具体问题有关。对于纯科学的推断问题，结论所应用的目的不一定明确，往往只能追求精确性；对于实际的决策问题，如对市场占有率、股票价格、病人病情等等问题的判断，损失往往比较具体。

(3) 同一问题的损失有多个方面。

等等。

2. 决策函数

根据具体观测数据x 确定行动a 的一个规则（不含任何未知参数），称为统计决策问题的一个**决策函数**或**决策(decision)**。常记作δ。例如，对于点估计问题，一个决策实质就是一个估计量；对于假设检验问题，一个决策就是一个检验规则。

非随机化决策： δ 是$(\mathscr{X},\mathscr{B}_X)\to(\mathscr{A},\mathscr{B}_A)$ 的可测变换。

随机化决策：在得到观测数据x后，由δ 还不能唯一地确定行动a，只能确定一个$(\mathscr{A},\mathscr{B}_A)$上的一个概率分布$\delta(\cdot|x)$。随机化决策$\delta(\cdot|\cdot)$可视为$\mathscr{B}_A\times\mathscr{X}\to[0,1]$的函数。随机化决策是理论研究的重要工具。实用意义有限。

非随机化决策可视为随机化决策的特例。

3. 风险函数(risk function)

如何衡量一个决策δ 的表现？——重复性原则，即应当从长期使用的、平均的角度去评价。

非随机化决策的风险函数

$$R(\theta,\delta)=E_\theta L(\theta,\delta(X))=\int_{\mathscr{X}} L(\theta,\delta(x))dP_\theta(x),\quad \theta\in\Theta. \tag{5.1}$$

随机化决策的风险函数

$$R(\theta,\delta)=\int_{\mathscr{X}}\left[\int_{\mathscr{A}} L(\theta,a)d\delta(a|x)\right]dP_\theta(x),\quad \theta\in\Theta. \tag{5.2}$$

例5.3 (1) p114/Example 2.19 点估计量的风险函数，是统计推断情况下“平均误差”概念的推广。若用平方损失函数$L(\theta,a)=[\gamma(\theta)-a]^2$ 衡量估计一维参数$\gamma(\theta)$ 的损失，那么由 (5.1)知，非随机化估计量$\delta(X)$ 的风险函数为

$$R(\theta,\delta)=E_\theta[\delta(X)-\gamma(\theta)]^2=MSE_\theta(\delta),\quad \theta\in\Theta.$$

(2) Example 2.20 检验的风险函数。设用损失函数

$$L(\theta,a)=c_1I(\theta\in\Theta_0,a=1)+c_2I(\theta\in\Theta_1,a=0)$$

衡量检验$H_0:\theta\in\Theta_0$ v.s. $H_1:\theta\in\Theta_1$ 的损失；$\phi(\cdot)$ 是一个检验函数，它对应于如下决策（即给定x 时采用0-1分布$B(1,\phi(x))$ 随机选择行动）：

$$\delta(\{1\}|x)=\phi(x)=1-\delta(\{0\}|x),\quad x\in\mathscr{X}.$$

那么由 (5.2)，δ 的风险函数为

$$\begin{aligned}
R(\theta,\delta)&=\int_{\mathscr{X}}\left[\int_{\mathscr{A}}L(\theta,a)d\delta(a|x)\right]dP_\theta(x)\\
&=\int_{\mathscr{X}}\left[L(\theta,1)\phi(x)+L(\theta,0)(1-\phi(x))\right]dP_\theta(x)\\
&=\begin{cases} c_1\displaystyle\int_{\mathscr{X}}\phi(x)dP_\theta(x), & \text{if } \theta\in\Theta_0,\\ c_2\displaystyle\int_{\mathscr{X}}[1-\phi(x)]dP_\theta(x), & \text{if } \theta\in\Theta_1,\end{cases}\\
&=c_1E_\theta\phi(X)I_{\Theta_0}(\theta)+c_2[1-E_\theta\phi(X)]I_{\Theta_1}(\theta),\quad \theta\in\Theta.
\end{aligned}$$

(3) 一般问题中的风险函数。

对于一般的损失函数、决策函数，风险函数未必有显式表达。

4. 决策函数的优良性

通常，对一个统计决策问题可以采取多种决策。那么，怎样才算是好的决策呢？对于给定的分布族$\mathscr{P}=\{P_\theta:\theta\in\Theta\}$ 及损失函数L，风险函数反映着决策的整体表现。因此，比较不同决策δ_1,δ_2 的优劣，应当对风险函数$R(\theta,\delta_1)$, $R(\theta,\delta_2)$ 进行整体比较。

定义5.1 δ_1,δ_2是两个决策函数，如果$R(\theta,\delta_1)\le R(\theta,\delta_2),\forall\theta\in\Theta$，称$\delta_1$（一致）不次于$\delta_2$；若还$\exists\theta_0\in\Theta$，使不等号严格成立，称$\delta_1$（一致）优于$\delta_2$；若$R(\theta,\delta_1)=R(\theta,\delta_2),\forall\theta\in\Theta$，称$\delta_1$ 等价于δ_2。设$\mathscr{D}$ 是一个决策函数类，若$\delta^*\in\mathscr{D}$ 不次于$\mathscr{D}$ 中任何一个决策函数，则称δ^* 是$\mathscr{D}$ 最优的。

两个决策相比，除了一个一致优于另一个的情况之外，更多的是两者各有优劣的情况。若不加限制，绝大多数有意义的统计决策问题不存在一致最优的决策。

定义5.2 (容许性(admissibility)) 设$\mathscr{D}$ 是一个决策函数类，对于$\delta^*\in\mathscr{D}$，若不存在$\delta\in\mathscr{D}$ 一致优于δ^*，则称δ^* 是$\mathscr{D}$可容许的；否则称δ^* 是不可容许的。

一般而言，一个不可容许的决策肯定是不好的决策，应当被排除。但是，一个容许决策未必是好的决策。例如，对于参数$\gamma(\theta)$ 的点估计问题，$\delta(x)\equiv\gamma(\theta_0)$ 是一个可容许的决策，但显然不好。

容许性与最优性的关系： 若δ^* 是$\mathscr{D}$最优的，则它$\mathscr{D}$可容许；若δ^* 是$\mathscr{D}$最优的，δ_0 是$\mathscr{D}$可容许的，则δ^* 与δ_0等价；若$\mathscr{D}$ 中有两个不等价且都可容许的决策函数，则$\mathscr{D}$ 中无最优的决策函数。

定义5.3 (完全类与本质完全类) 设$\mathscr{D}$ 是一个决策类，$\mathscr{D}^*\subset\mathscr{D}$，若对$\forall\delta'\in\mathscr{D}-\mathscr{D}^*$，$\exists\delta\in\mathscr{D}^*$，$\ni\delta$一致优于$\delta'$，则称$\mathscr{D}^*$ 是$\mathscr{D}$ 的完全类。若对$\forall\delta'\in\mathscr{D}-\mathscr{D}^*$，$\exists\delta\in\mathscr{D}^*$，$\ni\delta$一致不次于$\delta'$，则称$\mathscr{D}^*$ 是$\mathscr{D}$ 的本质完全类。

显然，本质完全类$\subset$完全类，而最小本质完全类是对最优决策的搜索范围的最大程度的“无损”压缩。

充分性原则在统计决策中的解释：（参见p117/proposition 2.2）

若某项统计决策原来是基于原始样本X 设计的，决策函数记为$\delta(X)$。设T是充分统计量。假如在观测过程中只记录了T 而抛弃了原始数据X，那么能否找到一个基于T的决策函数$\delta^*(T)$ 使之与δ等价呢？取$\delta^*:t\to\delta(X')$。X' 按$P(\cdot|t)$抽取。则

$$R(\theta,\delta^*)=E_\theta^T E^{X'|T}L(\theta,\delta(X'))=E_\theta^{X'}L(\theta,\delta(X'))=R(\theta,\delta).$$

这说明，在允许使用随机化决策的情况下，对任一决策都能找到一个与之等价的、由充分统计量构造的决策函数。即由充分统计量构造的决策函数类是一个本质完全类。

p117/Theorem 2.5 阐述的是点估计问题中在使用凸损失函数情况下，对任一决策都能找到一个与之等价的、由充分统计量构造的非随机化决策函数。即由充分统计量构造的非随机化决策函数类是一个本质完全类。我们在下一章详细讨论。

统计研究中，经常希望找到最优决策。然而，当考虑的决策函数类$\mathscr{D}$ 太大或太小时，都可能不存在最优决策。

例5.4 (最优决策存在与否的例子) (1) p118/Example 2.22. 设$X_1, \cdots, X_n$ i.i.d. F，$E_F X_1 = \mu$，$\mathrm{Var}_F X_1 = \sigma^2$ 存在但未知。用平方损失函数$L(F, a) = (a-\mu)^2$ 衡量估计μ 的损失。则在一般的线性估计类

$$\mathscr{D}_1 = \Big\{ \sum_{i=1}^n c_i X_i : \ (c_1, \cdots, c_n) \in \mathbb{R}^n \Big\}$$

中不存在最优决策。但在有约束的线性估计类

$$\mathscr{D}_2 = \Big\{ \sum_{i=1}^n c_i X_i : \ (c_1, \cdots, c_n) \in \mathbb{R}^n, \sum_{i=1}^n c_i = 1 \Big\}$$

中，$\overline{X}$ 是最优决策。若将估计类限制得再小些，比如$\mathscr{D}_2 \setminus \{\overline{X}\}$，那么最优决策又不存在了。

(2) Example 2.23. 设$X \sim B(n, \theta)$，$\theta \in (0, 1)$ 是未知参数，$n > 1$ 给定。若用0-1 损失函数

$$L(\theta, a) = I(\theta \in \Theta_0, a = 1) + I(\theta \in \Theta_1, a = 0)$$

衡量检验$H_0 : \theta \in (0, \theta_0]$ v.s. $H_1 : \theta \in (\theta_0, 1)$ 的损失，则在下列非随机化决策函数类

$$\mathscr{D} = \{\delta_j(x) = I_{\{j+1, \cdots, n\}}(x) : \ j = 0, 1, \cdots, n-1\}$$

中不存在最优决策。

找最优决策的思路主要有两条：

(1) 限制挑选决策函数的范围，将显然不合理的决策排除在外（找条件最优决策）。

(2) 将对风险函数的“点点比较”改为对其某些特征进行比较。

无偏性是一种将那些只注重对个别θ推断的准确性而无视其他θ的明显不合理决策排除在外的常用限制条件。将点估计问题中估计量的无偏性概念推广，即有以下概念。

定义5.4 (风险无偏) 若决策δ满足

$$E_\theta L(\theta', \delta(X)) \geq E_\theta L(\theta, \delta(X)), \quad \forall \theta, \theta' \in \Theta,$$

则称δ在损失函数$L(\theta, a)$下风险无偏。

例5.5 (1) 若取平方损失函数$L(\theta, a) = [a - \gamma(\theta)]^2$衡量估计一维参数$\gamma(\theta)$的损失。则决策$\delta$风险无偏等价于$\delta$是$\gamma(\theta)$的U.E.。

(2) 若取损失函数

$$L(\theta, a) = c_1 I(\theta \in \Theta_0, a = 1) + c_2 I(\theta \in \Theta_1, a = 0)$$

衡量检验$H_0 : \theta \in \Theta_0$ v.s. $H_1 : \theta \in \Theta_1$的损失。则$\delta$风险无偏等价于，$\delta$对应的检验函数$\phi(x)$满足

$$\begin{array}{ll} c_1 E_\theta \phi(X) \leq c_2 E_\theta [1 - \phi(X)], & \text{when } \theta \in \Theta_0, \\ c_1 E_\theta \phi(X) \geq c_2 E_\theta [1 - \phi(X)], & \text{when } \theta \in \Theta_1. \end{array}$$

若取$c_1 = c_2 = 1$，则上式意味着做出正确决策的概率应大于做错误决策的概率。上式可简化为

$$\begin{array}{ll} E_\theta \phi(X) \leq \frac{c_2}{c_1 + c_2}, & \text{when } \theta \in \Theta_0, \\ E_\theta \phi(X) \geq \frac{c_2}{c_1 + c_2}, & \text{when } \theta \in \Theta_1. \end{array}$$

同变性(equivariance)是在统计问题满足一定的“对称性”时常用的一种限制条件，它要求决策具有相应的“对称性”。这里我们通过一个例子来说明在一个具体场合下对决策同变性的要求。

例5.6 设$X_{i1}, i = 1, \cdots, n$ i.i.d. $N(\xi_1, \sigma^2)$，$X_{i2}, i = 1, \cdots, n$ i.i.d. $N(\xi_2, \sigma^2)$，且合样本相互独立。统计问题有如下三个行动：

$$a_0: \ |\xi_2 - \xi_1| \leq \Delta, \qquad a_1: \ \xi_2 > \xi_1 + \Delta, \qquad a_2: \ \xi_1 > \xi_2 + \Delta.$$

设在正确行动为a_i时选择行动a_j导致的损失为w_{ij}，$i, j = 0, 1, 2$. 若无其他考虑，则对本问题加上如下对称性要求是很自然的：

i) 损失的对称性： $w_{01} = w_{02}$， $w_{10} = w_{20}$， $w_{12} = w_{21}$.

ii) 若将各X_{ij}的第二个下标的1与2互换，同时a, w, ξ 相应的下标1与2也互换，那么得到的新问题的本质不变。

对这样一个统计问题，合理的决策δ 自然应满足：

$$\begin{aligned} \text{若}\delta(x_{11}, \cdots, x_{n1}, x_{12}, \cdots, x_{n2}) &= a_0,\ a_1,\ \text{或}a_2,\ \text{则应有} \\ \delta(x_{12}, \cdots, x_{n2}, x_{11}, \cdots, x_{n1}) &= a_0,\ a_2,\ \text{或}a_1. \end{aligned}$$

Bayes准则与Minimax 准则是两种比较风险函数整体最优性的常用观点。一般地，要从某角度整体比较风险函数，就是对风险函数集$\{R(\theta, \delta) : \delta \in \mathscr{D}\}$，引进某种“序”，使得任意$\delta_1, \delta_2 \in \mathscr{D}$只有$\prec, =, \succ$ 三种关系。

定义5.5 (Minimax 决策) 设$\mathscr{D}$ 是一个决策函数类，记$M(\delta) = \max_{\theta\in\Theta} R(\theta, \delta)$。若$\delta^* \in \mathscr{D}$ 满足

$$M(\delta^*) = \min_{\delta\in\mathscr{D}} M(\delta),$$

则称δ^* 为$\mathscr{D}$ *Minimax*决策。

定义5.6 (Bayes 决策) 设想研究者对参数空间Θ 中的θ 的关注程度可用权函数$\rho(\theta)$ 来衡量。称

$$r_\rho(\delta) = \int_\Theta R(\theta, \delta)\rho(\theta)d\theta,$$

为决策δ 的*Bayes*风险。设$\mathscr{D}$ 是一个决策函数类，若$\delta^* \in \mathscr{D}$ 满足

$$r_\rho(\delta^*) = \min_{\delta\in\mathscr{D}} r_\rho(\delta),$$

则称δ^* 为$\mathscr{D}$ *Bayes* 决策。

例5.7 Shao, J. p120/Example2.25. 设$X_1, \cdots, X_n$ i.i.d. $N(\mu, \sigma^2)$，求参数μ的Bayes决策与Minimax 决策。