03 充分统计量与完备性(补充)-教学辅导

一、【内容提要】

1．充分统计量（sufficient statistic）

1）定义 5.5.1 ： 设 $X_1, X_2, \cdots, X_n$ 是来自某个总体的样本，总体分布函数为 $F(x;\theta)$，统计量 $T = T(X_1, X_2, \cdots, X_n)$ 称为 θ 的充分统计量，如果在给定 T 的取值后，$X_1, X_2, \cdots, X_n$ 的条件与 θ 无关.

即不包含关于参数的信息

2）定理 5.5.1（因子分解定理 Factorization Theorem）：设总体概率函数为 $f(x;\theta)$，$X_1, X_2, \cdots, X_n$ 为样本，则 $T = T(X_1, X_2, \cdots, X_n)$ 为充分统计量得**充分必要条件**是：存在两个函数 $g(t,\theta)$ 和 $h(X_1, X_2, \cdots, X_n)$ 使得对任意的 θ 和任意组观测值 $X_1, X_2, \cdots, X_n$，有 $f(X_1, X_2, \cdots, X_n;\theta) = g(T(X_1, X_2, \cdots, X_n),\theta)h(X_1, X_2, \cdots, X_n)$，其中是通过统计量的取值而依赖于样本的.

证明：一般性结果的证明超出本课程范围，此处我们将给出离散型随机变量下的证明，此时，$f\left(x_1,\cdots,x_n;\theta\right) = P\left(X_1 = x_1,\cdots X_n = x_n;\theta\right)$.

先证必要性.设 T 使充分统计量,则在 $T=t$ 下, $P\left(X_1 = x_1,\cdots,X_n = x_n \middle| T=t\right)$ 与 θ 无关,记为 $h\left(x_1,\cdots x_n\right)$ 或 $h\left(X\right)$,令 $A\left(t\right)=\left\{X: T\left(X\right)=t\right\}$,当 $x \in A\left(t\right)$ 时有

$$\left\{T=t\right\} \supset \left\{X_1 = x_1,\cdots,X_n = x_n\right\},$$

故

$$\begin{aligned}
&P\left(X_1 = x_1,\cdots,X_n = x_n;\theta\right) = P\left(X_1 = x_1,\cdots,X_n = x_n, T=t;\theta\right)\\
&= P\left(X_1 = x_1,\cdots,X_n = x_n \middle| T=t\right)P\left(T=t;\theta\right)\\
&= h\left(x_1,\cdots,x_n\right)g\left(t,\theta\right),
\end{aligned}$$

其中 $g\left(t,\theta\right)=P\left(T=t;\theta\right)$, 而 $h\left(X\right)=P\left(X_1 = x_1,\cdots,X_n = x_n \middle| T=t\right)$ 与 θ 无关,必要性得证.

对充分性,由于

$$\begin{aligned}
P\left(T=t;\theta\right) &= \sum\nolimits_{\left\{\left(x_1,\cdots x_n\right): T\left(x_1,\cdots x_n\right)=t\right\}} P\left(X_1 = x_1,\cdots X_n = x_n;\theta\right)\\
&= \sum\nolimits_{\left\{\left(x_1,\cdots x_n\right): T\left(x_1,\cdots x_n\right)=t\right\}} g\left(t,\theta\right)h\left(x_1,\cdots x_n\right),
\end{aligned}$$

对任给 $X=\left(x_1,\cdots x_n\right)$ 和 t,满足 $X \in A\left(t\right)$,有

$$
\begin{aligned}
&P\left(X_1=x_1,\cdots,X_n=x_n \middle| T=t\right)\\
&=\frac{P\left(X_1=x_1,\cdots,X_n=x_n,T=t;\theta\right)}{P\left(T=t;\theta\right)}\\
&=\frac{P\left(X_1=x_1,\cdots,X_n=x_n;\theta\right)}{P\left(T=t;\theta\right)}\\
&=\frac{g\left(t,\theta\right)h\left(x_1,\cdots,x_n\right)}{g\left(t,\theta\right)\sum_{\left\{\left(y_1,\cdots y_n\right):T\left(y_1,\cdots y_n\right)=t\right\}}h\left(y_1,\cdots y_n\right)}\\
&=\frac{h\left(x_1,\cdots,x_n\right)}{\sum_{\left\{\left(y_1,\cdots y_n\right):T\left(y_1,\cdots y_n\right)=t\right\}}h\left(y_1,\cdots y_n\right)},
\end{aligned}
$$

该分布与θ无关,这证明了充分性.

3）充分性判别法则

定理 4.1 设样本分布密度函数族（连续或离散）为$F=\left\{f\left(x,\theta\right):\theta\in\Theta\right\},T=T\left(X\right)$为统计量.则：$T$为充分统计量的充分必要条件为：存在关于$t$的可测函数$g_\theta\left(t\right)$与关于$x$的非负可测函数$h\left(x\right)$，使得

$$f\left(x,\theta\right)=g_\theta\left(T\left(x\right)\right)h\left(x\right) \tag{0.1}$$

对每一$\theta\in\Theta$与$x\in X$成立.

注：$h\left(x\right)$不依赖于θ.

证：只对离散型情况给出证明.这时,

$$f\left(x,\theta\right)=P_\theta\left[X=x\right]$$

对于$T\left(X\right)$的值域中任意固定的t,定义集合

$$A(t)=\left\{x:T\left(x\right)=t\right\}.$$

充分性 设$f_\theta\left(x,\theta\right)$使因子分解式（1.1）成立.则对任意的$x\in A\left(t\right)$, $T\left(x\right)=t$成立,条件概率

$$
\begin{aligned}
P_\theta\left[X=x \middle| T(X)=t\right] &= \frac{P\left[X=x,T(X)=t\right]}{P_\theta\left[T(X)=t\right]} \\
&= \frac{P_\theta\left[X=x\right]}{P_\theta\left[T=t\right]} = \frac{f(x,\theta)}{\sum\limits_{u\in A(t)} f(u,\theta)} = \frac{g_\theta(T(x))h(x)}{\sum\limits_{u\in A(t)} g_\theta(T(u))h(u)} \\
&= \frac{g_\theta(t)h(x)}{\sum\limits_{u\in A(t)} g_\theta(t)h(u)} \doteq \frac{h(x)}{\sum\limits_{u\in A(t)} h(u)},
\end{aligned}
$$

它与参数θ无关.又若$x\notin A(t)$,则$T(x)\neq t$,

$$
\begin{aligned}
P\left[X=x \middle| T(X)=t\right] &= \frac{P_\theta\left[X=x,T(X)=t\right]}{P_\theta\left[T(X)=t\right]} \\
&= \frac{P_\theta\left[\varnothing\right]}{P_\theta\left[T=t\right]} = 0.
\end{aligned}
$$

也与θ无关.因此,条件分布$f_\theta(x|t)=f(x|t)$与θ无关, 即$T(X)$是θ的充分统计量.

必要性　设$T(X)$是θ的充分统计量,由充分统计量的定义,$P_\theta\left[X=x \middle| T(X)=t\right]$与参数$\theta$无关,它是$x$的函数,记为$h(x)$. 于是,对任意固定的$t$,当$x\in A(t)$时,$T(x)=t$成立;这时

$$
\begin{aligned}
T(x,\theta) &= P_\theta\left[X=x\right] = P_\theta\left[X=x,T(X)=t\right] \\
&= P_\theta\left[T(X)=t\right] P_\theta\left[X=x \middle| T(X)=t\right] \\
&= P_\theta\left[T(X)=t\right] h(x) = g_\theta(t)h(x) \\
&= g_\theta(T(x))h(x),
\end{aligned}
$$

式中$g_\theta(t)=P_\theta\left[T(X)=t\right]$.因而（1.1）成立.

由因子分解定理,若样本的密度函数$f(x,\theta)$能分解成两个因子的乘积,其中一个为$T(X)$的函数,而另一个仅为x的函数,与参数θ无关,则$T(X)$是θ的充分统计量.

2．完备性

1）定义：$F=\{p(x;\theta),\theta\in\Theta\}$，设$g(x)$是定义在样本空间$\Omega$上的一个实函数，一般来说，积分（如果存在）$E[g(x)]=\int_\Omega g(x)p(x;\theta)dx$ （$\theta\in\Theta$），因此上述积分（数学期望）可以看作一个变换，且是一对一的变换**.**

即对$\forall\theta\in\Theta$， $g_1(x)=g_2(x)=1 \Leftrightarrow E_g[g_1(x)]=E_g[g_2(x)]$

$g = g_1 - g_2 = 0$， $E_g(g_1 - g_2) = 0$ ，则$\{p_\theta g(x) = 0\} = 1 \Leftrightarrow E_g[g(x)] = 0$

英文注释：Definition (Complete Statistic) : Let be a family of pdfs of pmfs for a statistic. The family of probability distributions is called complete if for all implies for all. Equivalently, is called a complete statistic.

2）分布族的完备性：

定义：$F = \{p(x;\theta), \theta \in \Theta\}$

对于任何一个可测函数$g(x)$，由$E_g[g(x)] = \varphi(\theta) = \int_\pi g(x)p(x;\theta)dx = 0$

对 $\forall\theta$ 有$\{p_g\{g(x)\} = 0\} = 1$ or $g(x) = 0(a.e.p)$ 等价的， $E_\theta[g_1(x)] = E_g[g_2(x)]$ 对 $\forall\theta \in \Theta$ 成立，可推出 $p_\theta\{g_1(x) = g_2(x)\} = 1$

3）完备性意义：

① 积分变换（数学期望）的唯一性.

② 常用的积分变换.

a. 傅里叶变换 $f(x) \to \int_{-\infty}^{+\infty} e^{itx} f(x)dx$ 特征函数，它在$t \in (-\infty, +\infty)$上 都存在且有唯一性.

b. laplas 变换 $f(x) \to \int_{-\infty}^{+\infty} e^{-sx} f(x)dx$，该式在 s=0 存在至少在 s=0 某个领域内有定义，则有唯一性.

4）完备充分统计量（**complete sufficient statistic**）

定义：设 $p(x;\theta)$ 是一概率密度函数且是指数族的正规案，设 $X_1, \ldots, X_n$ 是具有 p.d.f $p(x;\theta)$ 的分配的随机样本.则统计量$T = \sum_{i=1}^{n} X_i$ 是θ的完备充分统计量.

5) 某些完全性

定理（指数族的完全性）：设 X 的样本空间为 (x, β_x)，分布族为指数族，对$\theta \in \Theta$，有 $\theta \in \Theta$ $dp_\theta(x) = c(\theta)\exp[\sum_{i=1}^{k} \theta_i T_i(x)]du(x)$ ，此处Θ为R_k之一子集，若Θ（作为R_k的子集）由内点，则统计量$t(x) = (T_1(x), \cdots, T_k(x))$是完全统计量.

定理（次序统计量的完全性）： 设分布族 f 满足以下两个条件：

（a）若$F_1 \in f, F_2 \in f$ ，则对任何$P_1 > 0, P_2 > 0, P_1 + P_2 = 1,$ 有 $P_1F_1 + P_2F_2 \in f$.

（b）若 $F \in f, S = [a,b), -\infty < a < b < +\infty$ ，而 $F(s) \neq 0$，则 $F_B \in f$ ，则次序统计量

$X_{(1)},\dots,X_{(n)}$ 是完全的（对任何自然数 n）.

引理 ：设分布族满足上面的条件（a），$f(X_1,\dots,X_n)$ 为 Bore（可测得对称函数），满足条件 $\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(X_1,\dots,X_n)dF(X_1)\cdots dF(X_n)=0 f(X_{(1)},\dots,X_{(n)})$，对任何 $F\in f$，则对 F 中的任意 n 个分布 $F_1,\cdots,F_n$，必有

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(X_1,\dots,X_n)dF_1(X_1)\cdots dF_n(X_n)=0 f(X_{(1)},\dots,X_{(n)})$$

定义（有界完全性）：设变量 X 的样本空间为 (x,β_x)，分布族为 $\{p_\theta,\theta\in\Theta\}$，$t(x)$ 为定义于 X 取值于 (f,β_f) 的统计量，其分布族为 $\{p_\theta^T,\theta\in\Theta\}$，若对任何满足条件" $\int_x f(x)dp_\theta(x)=0$，对一切 $\theta\in\Theta$ "的有界 β_x 可测函数 $f(x)$，必有 $p_\theta\{X=f(x)\neq 0\}=0$，对一切 $\theta\in\Theta$，则称分布族 $\{p_\theta,\theta\in\Theta\}$ 为有界完全的.若 $\{p_\theta^T,\theta\in\Theta\}$ 为有界完全的，则称 t 为有界完全统计量.

3．极小充分统计量（minimal sufficient statistic）

1）定义： 设 $t(x)$ 为 (x,β_x) 上的一个充分统计量，取值于 (f,β_f)，β 上的分布族为 $\{p_\theta,\theta\in\Theta\}$.若对任何定义于 x，取值于某可测空间 (S,β_s) 的充分统计量.必存在由 (S,β_s) 到 (f,β_f) 的可测变换 $t=q(s)$，以及 $A\in\beta_x$，满足条件 $p_\theta(A)=0$ 对任何 $\theta\in\Theta$，致 $t(x)=q(S(x))$，对任何 $x\in A$，则称唯一极小充分统计量.

2）定理（极小充分统计量的存在定理）： 假定分解定理中的条件成立，且样本空间为欧式的，则极小充分统计量存在.

3）要求：①信息损失越少越好　　　　②统计量越简化越好

4．指数族：

1）定义：设 $(\mathrm{X},\mathrm{B}\mid p_\theta:\theta\in\Theta|)$ 是可控参数统计结构，加入其密度函数可表示为如下形式：$p_\theta(x)=c(\theta)\exp\{\sum_{i=1}^{k}c_j(\theta)T_j(x)\}h(x)$

并且它的支撑 $\{x:p_\theta(x)>0\}$ 不依赖于 θ，则称此结构为指数型的统计结构，简称指数结构，其中的分布族为指数族，这里的 $0<c(\theta),c_1(\theta),\cdots,c_k(\theta)<\infty,T_j(x)$ 都与 θ

无关，且取有限值的B可测函数，k 为正整数，$h(x)>0$.

2）定理：

① 自然参数空间Ω为凸集

② $\Phi(x)$ 是 X 上的 B 可测函数，且对一切 $w=(w_1,\ldots,w_k)\in\Omega$ 有

$$\int|\Phi(x)|\exp\{\sum_{j=1}^{k}w_jT_j(x)\}d\mu<\infty$$

③ 设 $X=(X_1,\ldots,X_n)$ 是来自指数型分布标准形式的一个样本，则有统计量

$(T_1(X),\ldots,T_k(X))=(\sum_{i=1}^{n}T_1(x_i),\cdots,\sum_{i=1}^{n}T_k(x_i))$ 是指数型分布族的充分统计量.

3）常见指数分布族

二项分布族：

$$p_\theta(x)=\binom{n}{x}\theta^x(1-\theta)^{n-x}=\binom{n}{x}(1-\theta)^n e^{x\ln\frac{\theta}{1-\theta}}$$

$$=c(\theta)\exp\{c_1(\theta)x\}h(x),x=0,1,\cdots,n$$

其中$c(\theta)=(1-\theta)^n,c_1(\theta)=x\ln\dfrac{\theta}{1-\theta},h(x)=\dbinom{n}{x}$

二元正态分布族：

$$p_{u,\sigma}(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{\mu^2}{2\sigma^2}\}\exp\{-\frac{x^2}{2\sigma^2}+\frac{\mu x}{\sigma^2}\}$$

其中$c(\mu,\sigma)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\{-\dfrac{\mu^2}{2\sigma^2}\},c_1(\mu,\sigma)=\dfrac{\mu}{\sigma^2},c_2(\mu,\sigma)=\dfrac{1}{2\sigma^2}$

$$h(x)=1,T_1(x)=x,T_2(x)=x^2$$

伽玛分布族：

$$p_{\alpha,\lambda}(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$$

$$=\frac{\lambda^\alpha}{\Gamma(\alpha)}\exp\{-\lambda x+(\alpha-1)\ln x\}$$

$$=c(\alpha,\lambda)\exp\{c_1(\alpha,\lambda)x+c_2(\alpha,\lambda)\ln x\},x<0$$

其中$c(\alpha,\lambda)=\dfrac{\lambda^\alpha}{\Gamma(\alpha)},c_1(\alpha,\lambda)=-\lambda,c_2(\alpha,\lambda)=(\alpha-1)$

注：如果 Gammar 分布中引入第三个参数——门限参数μ，其密度函数为

$$p_{\mu,\alpha,\lambda}(x)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}(x-\mu)^{\alpha-1}e^{-\lambda(x-\mu)},x>\mu$$

5．辅助统计量（ancillary statistic）：

1）定义： 设 $X\sim\{f(x;\theta),\theta\in\Theta\}$，若统计量 $A=A(X)$ 的分布与 θ 无关，则称 $A(X)$ 为辅助统计量（即 $A(X)$ 中不包含关于 θ 的信息）

英文注释：Definition (Ancillary Statistic) : A statistic $S(X)$ whose distribution does not depend on the parameter θ is called an ancillary statistic**.**

Alone, an ancillary statistic contains no information about θ**.**

An ancillary statistic is an observation on a random variable whose distribution is fixed and known, unrelated to θ**.**

Paradoxically, an ancillary statistic, when used in conjunction with other statistics, sometimes does contain valuable information for inferences about θ**.**

6．常见的充分统计量

分布	分布列或密度函数	参数	充分统计量
二项分布 b(1,p)	$P\left(X=x\right)=p^{x}\left(1-p\right)^{1-x},x=0,1,\cdots$	p	$T=x_1+\cdots+x_n$
泊松分布 $P\left(\lambda\right)$	$P(X=x)=\frac{\lambda^{x}}{x!}\mathrm{e}^{-\lambda},x=0,1,2,\cdots$	λ	$T=x_1+\cdots+x_n$
几何分布 $Ge(\theta)$	$P\left(X=x\right)=\left(1-\theta\right)^{x-1}\theta,x=1,2,\cdots$	θ	$T=x_1+\cdots+x_n$
指数分布 Exp(λ)	p(x)= $\lambda e^{-\lambda x},x>0$	λ	$T=x_1+\cdots+x_n$
均匀分布 $U(0,\theta)$	$p\left(x\right)=\frac{1}{\theta},0<x<\theta$	θ	$T=\max(x_1,\cdots,x_n)$ 即 $T=x_{(n)}$
均匀分布 $U(\theta_1,\theta_2)$	$p\left(x\right)=\frac{1}{\theta_2-\theta_1},\theta_1<x<\theta_2$	θ_1,θ_2	$T_1=x_{(1)},T_2=x_{(n)}$
均匀分布 $U(\theta,2\theta)$	$p\left(x\right)=\frac{1}{\theta},\theta<x<2\theta$	θ	$T_1=x_{(1)},T_2=x_{(n)}$

正态分布 $N\left(\mu,\sigma^2\right)$	$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ,σ^2	$\overline{x}$与$\sum_{i=1}^{n}(x_i-\overline{x})^2$
幂分布	$p(x;\theta)=\theta x^{\theta-1},0<x<1$	θ	$T=\prod_{i=1}^{n}x_i$或$T=\sum_{i=1}^{n}\ln x_i$
双参数指数分布	$p(x;\theta,\mu)=\frac{1}{\theta}e^{-\frac{x-\mu}{\theta}},x>\mu$	μ,θ	$T_1=x_{(1)},T_2=\sum_{i=1}^{n}x_i$
伽玛分布 Ga(α，λ)	$p(x;\alpha,\lambda)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{a-1}e^{-\lambda x}$,x>0	α,λ	$T_1=\sum_{i=1}^{n}x_i,T_2=\prod_{i=1}^{n}x_i$
对数正态分布 $LN\left(\mu,\sigma^2\right)$	$p(x)=\frac{1}{\sqrt{2\pi}\sigma x}\mathrm{e}^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$	μ,σ^2	$T_1=\sum_{i=1}^{n}\ln x_i,T_2=\sum_{i=1}^{n}(\ln x_i)^2$
贝塔分布 Be(a,b)	$p(x)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}$,0<x<1	a,b	$T_1=\sum_{i=1}^{n}\ln x_i,T_2=\sum_{i=1}^{n}\ln(1-x_i)$

二、【释疑解难】

1．对上述充分统计量的证明

***对于指数分布族直接找出充分统计量，以下为一些例子**

二项分布：b(1, p)

设 $X_1,X_2,\cdots,X_n$ 使来自二点分布 $b\left(1,p\right)$ 的一个样本，其中 $0<p<1,n>2$，现在我们来考察如下两个统计量：

$$T_1=\sum_{i=1}^{n}X_i,\qquad T_2=X_1+X_2.$$

我们知道，样本 $\left(X_1,X_2,\cdots,X_n\right)$ 的联合分布是

$$P\left(X_1=x_1,X_2=x_2,\cdots,X_n=x_n\right)=p^{\sum_{i=1}^{n}x_i}\left(1-p\right)^{n-\sum_{i=1}^{n}x_i},$$

其中，诸 x_i 非 0 即 1.而统计量 $T_1=\sum_{i=1}^{n}X_i$ 的分布为二项分布 $b\left(n,p\right)$，即

$$P\left(T=t\right)=\binom{n}{t}p^t\left(1-p\right)^{n-t},\qquad t=0,1,\cdots,n.$$

而在给定 $T_1=t$ 下，样本的条件分布为

$$
\begin{aligned}
&P\left(X_1=x_1,X_2=x_2,\cdots,X_n=x_n\middle|T_1=t\right)\\
&=\frac{P\left(X_1=x_1,X_2=x_2,\cdots,X_n=x_n,T_1=t\right)}{P\left(T_1=t\right)}\\
&=\frac{P\left(X_1=x_1,\cdots,X_{n-1}=x_{n-1},X_n=t-\sum_{i=1}^{n-1}x_i\right)}{P\left(T_1=t\right)}\\
&=\frac{p^t\left(1-p\right)^{n-t}}{\binom{n}{t}p^t\left(1-p\right)^{n-t}}=\binom{n}{t}^{-1}.
\end{aligned}
$$

计算结果表明，这个条件分布与参数 p 无关.它已不含有参数 p 的有关信息了.样本中有关 p 的信息都含在统计量 T_1 中.

另外，统计量 $T_2=X_1+X_2$ 的分布仍是二次分布 $b\left(2,p\right)$，即

$$
P\left(T_2=t\right)=\binom{2}{t}p^t\left(1-p\right)^{2-t},t=0,1,2.
$$

于是在给定 $T_2=t$ 下，样本的条件分布为

$$
\begin{aligned}
&P\left(X_1=x_1,X_2=x_2,\cdots,X_n=x_n\middle|T_2=t\right)\\
&=\frac{P\left(X_1=x_1,X_2=t-x_1,X_3=x_3\cdots,X_n=x_n\right)}{P\left(T_2=t\right)}\\
&=\frac{p^{t+\sum_{i=3}^{n}x_i}\left(1-p\right)^{n-t-\sum_{i=3}^{n}x_i}}{\binom{2}{t}p^t\left(1-p\right)^{2-t}}\\
&=\binom{2}{t}^{-1}p^{\sum_{i=3}^{n}x_i}\left(1-p\right)^{n-2-\sum_{i=3}^{n}x_i}.
\end{aligned}
$$

可见，这个条件分布与参数 p 有关.这意味着，这个条件分布还含有参数 p 的信息，而样本中有关 p 的信息没有完全包含在统计量 T_2 之中.

注：从上例可以直观地看出，用条件分布与参数无关来表示不损失样本中有价值的信息室妥当的.一般的充分统计量的定义也正是这样给出的.
(数理统计_茆诗松王静龙 P46/Ex 1.6.2)

泊松分布：$P\left(\lambda\right)$

（书 P283/Ex5.5.2)：设 $x_1,\cdots,x_n$ 是来自泊松分布 $P\left(\lambda\right)$ 的样本，则 $T=\sum_{i=1}^{n}x_i$ 是充分

统计量

解：$p(x;\lambda)=\dfrac{\lambda^x}{x!}e^{-\lambda}=\exp\{(\ln\theta)x-\ln(x!)-\theta\}, x=0,1,\cdots$

且 $X_1,\ldots,X_n$ 独立同分布，根据充分完备统计量定义可得，$T=x_1+\cdots+x_n$ 为其充分统计量.

令解：由泊松分布性质知，$T\sim P(n\lambda)$

在给定 T 的取值后，对任意的一组 $x_1,\cdots,x_n\left(\sum_{i=1}^{n}x_i=t\right)$，有

$$P\left(X_1=x_1,\cdots,X_n=x_n \middle| T=t\right)=\frac{P\left(X_1=x_1,\cdots,X_{n-1}=x_{n-1},X_n-t=\sum_{i=1}^{n-1}x_i\right)}{P\left(\sum_{i=1}^{n}x_i=t\right)}$$

$$=\frac{\prod_{i=1}^{n-1}P\left(X_i=x_i\right)\bullet P\left(X_n=t-\sum_{i=1}^{n-1}x_i\right)}{\dfrac{(n\lambda)^t}{t!}\mathrm{e}^{-n\lambda}}$$

$$=\frac{\prod_{i=1}^{n-1}\dfrac{\lambda^{x_i}}{x_i!}\mathrm{e}^{-\lambda}\bullet\dfrac{\lambda^{t-\sum_{i=1}^{n-1}x_i}}{\left(t-\sum_{i=1}^{n-1}x_i\right)!}\mathrm{e}^{-\lambda}}{\dfrac{(n\lambda)^t}{t!}\mathrm{e}^{-n\lambda}}$$

$$=\frac{\dfrac{\lambda^t}{\prod_{i=1}^{n}x_i!}\mathrm{e}^{-n\lambda}}{\dfrac{(n\lambda)^t}{t!}\mathrm{e}^{-n\lambda}}$$

$$=\frac{t!}{n^t\cdot\prod_{i=1}^{n}x_i!}$$ 与 λ 无关，是充分统计量.

几何分布：$Ge(\theta)$

（书 P283/Ex5. 5. 1)：设 $x_1,\cdots,x_n$ 是来自几何分布 $P(X=x)=\theta(1-\theta)^x, x=0,1,2,\cdots$

的样本，则$T=\sum_{i=1}^{n}x_i$是充分统计量.

解：$P\left(X=x\right)=\theta\left(1-\theta\right)^x=\exp\{\ln\theta+x\ln(1-\theta)\},x=0,1,2,\cdots$

且$X_1,\ldots,X_n$独立同分布，则由充分完备统计量定义得，$T=x_1+\cdots+x_n$为其充分统计量.

令解：由几何分布性质知， $T\sim Nb\left(n,\theta\right)$

其分布列为$P\left(T=t\right)=\binom{n+t-1}{t}\theta^n\left(1-\theta\right)^t,t=0,1,2,\cdots$

在给定T的取值后，对任意的一组$x_1,\cdots,x_n\left(\sum_{i=1}^{n}x_i=t\right)$，有

$$P\left(X_1=x_1,\cdots,X_n=x_n\middle|T=t\right)=\frac{P\left(X_1=x_1,\cdots,X_{n-1}=x_{n-1},X_n-t=\sum_{i=1}^{n-1}x_i\right)}{P\left(\sum_{i=1}^{n}x_i=t\right)}$$

$$=\frac{\prod_{i=1}^{n-1}P\left(X_i=x_i\right)\bullet P\left(X_n=t-\sum_{i=1}^{n-1}x_i\right)}{\binom{n+t-1}{t}\theta^n\left(1-\theta\right)^t}$$

$$=\frac{\left(\prod_{i=1}^{n-1}\theta\left(1-\theta\right)^{x_i}\right)\cdot\theta\left(1-\theta\right)^{t-\sum_{i=1}^{n-1}x_i}}{\binom{n+t-1}{t}\theta^n\left(1-\theta\right)^t}$$

$$=\frac{\theta^n\left(1-\theta\right)^t}{\binom{n+t-1}{t}\theta^n\left(1-\theta\right)^t}$$

$$=\frac{1}{\binom{n+t-1}{t}}$$ 与θ无关，是充分统计量.

$\boxed{\text{指数分布}}$：Exp(λ)

设$x_1,\cdots,x_n$是来自指数分布Exp(λ)的样本，则$T=\sum_{i=1}^{n}x_i$是充分统计量.

解： $p(x;\lambda)=\lambda e^{-\lambda x}=\exp\{\ln\lambda-\lambda x\},x=0,1,\cdots$

且 $X_1,\ldots,X_n$ 独立同分布，则由充分完备统计量定义得，$T=x_1+\cdots+x_n$ 为其充分统计量.

令解：由泊松分布性质知，$T\sim Ga(n,\lambda)$

其分布函数为 $p(t;\lambda)=\dfrac{\lambda^n}{\Gamma(n)}t^{n-1}e^{-\lambda t}=\dfrac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}$

在给定 T 的取值后，对任意的一组 $x_1,\cdots,x_n\left(\sum_{i=1}^{n}x_i=t\right)$，有

$$P\left(X_1=x_1,\cdots,X_n=x_n\middle|T=t\right)=\frac{P\left(X_1=x_1,\cdots,X_{n-1}=x_{n-1},X_n-t=\sum_{i=1}^{n-1}x_i\right)}{P\left(\sum_{i=1}^{n}x_i=t\right)}$$

$$=\frac{\prod_{i=1}^{n-1}P(X_i=x_i)\bullet P\left(X_n=t-\sum_{i=1}^{n-1}x_i\right)}{\frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}}$$

$$=\frac{\left(\prod_{i=1}^{n-1}\lambda e^{-\lambda x_i}\right)\cdot\lambda e^{-\lambda\left(t-\sum_{i=1}^{n-1}x_i\right)}}{\frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}}$$

$$=\frac{\lambda^n e^{-\lambda t}}{\frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}}$$

$=\dfrac{(n-1)!}{t^{n-1}}$ 与 λ 无关，是充分统计量.

***对于非指数族用其因子分解定理来求充分统计量，以下就是典型的例子**

均匀分布：$U(0,\theta)$

（书 P282/Eg5.5.4）：设 $x_1,\cdots,x_n$ 是取自总体 $U(0,\theta)$ 的样本，即总体的密度函数为

$$p(x;\theta)=\begin{cases}\dfrac{1}{\theta},0<x<\theta\\ 0\ldots\ldots\text{其他}\end{cases}$$

解：于是样本的联合密度函数为

$$p(x_1;\theta)\cdots p(x_n;\theta)=\begin{cases}\left(\dfrac{1}{\theta}\right)^n, 0<\min\{x_i\}\le\max\{x_i\}<\theta\\ 0\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{其他}\end{cases}$$

由 于 诸 $x_i>0$ ， 所 以 我 们 可 将 上 式 改 写 为

$$p(x_1;\theta)\cdots\ p(x_n;\theta)=\left(\frac{1}{\theta}\right)^n I_{\{x_{(n)}<\theta\}},$$

取$T=x_{(n)}$, 并令$g(t,\theta)=\left(\dfrac{1}{\theta}\right)^n I_{\{t<\theta\}}, h(X)=1,$

由因子分解定理知，$T=x_{(n)}$是θ的充分统计量.

均匀分布：$U(\theta_1,\theta_2)$

（书 P283/Ex5.5.１0）：设$x_1,\cdots,x_n$是来自均匀分布$U(\theta_1,\theta_2)$的样本，试给出一个充分统计量.

解：总体的密度函数为 $p(x;\theta_1,\theta_2)=\begin{cases}\dfrac{1}{\theta_2-\theta_1},\theta_1<x<\theta_2\\ 0\ldots\ldots\ldots\ldots\text{其他}\end{cases}$

于是样本的联合密度函数为

$$p(x_1;\theta_1,\theta_2)\cdots p(x_n;\theta_1,\theta_2)=\begin{cases}\left(\dfrac{1}{\theta_2-\theta_1}\right)^n, 0<\min\{x_i\}\le\max\{x_i\}<\theta\\ 0\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{其他}\end{cases}$$

由于诸$x_i>0$，所以我们可将上式改写为

$$p(x_1;\theta_1,\theta_2)\cdots\ p(x_n;\theta_1,\theta_2)=\left(\frac{1}{\theta_2-\theta_1}\right)^n I_{\{\theta_1<x_{(1)}\le x_{(n)}<\theta_2\}}$$

取$t_1=x_{(1)},t_2=x_{(n)}$，并令$g(t_1,t_2,\theta_1,\theta_2)=\left(\dfrac{1}{\theta_2-\theta_1}\right)^n I_{\{\theta_1<t_1\le t_2<\theta_2\}}, h(X)=1,$

由因子分解定理知，$T=(t_1,t_2)=\left(x_{(1)},x_{(n)}\right)$是$\theta_1,\theta_2$的充分统计量.

均匀分布：$U(\theta,2\theta)$

（书 P283/Ex5.5.１1）：设$x_1,\cdots,x_n$是来自均匀分布$U(\theta,2\theta)$的样本，试给出一个充分统计量.

解：总体的密度函数为 $p\left(x;\theta\right)=\begin{cases}\dfrac{1}{\theta},\theta<x<2\theta\\ 0\dots\dots\dots\dots\text{其他}\end{cases}$

于是样本的联合密度函数为

$$p\left(x_1;\theta\right)\cdots p\left(x_n;\theta\right)=\begin{cases}\left(\dfrac{1}{\theta}\right)^n,\theta<\min\left\{x_i\right\}\le\max\left\{x_i\right\}<2\theta\\ 0\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\text{其他}\end{cases}$$

由于诸 $x_i>0$，所以我们可将上式改写为

$$p\left(x_1;\theta\right)\cdots\ p\left(x_n;\theta\right)=\left(\frac{1}{\theta}\right)^n I_{\left\{\theta<x_{(1)}\le x_{(n)}<2\theta\right\}}$$

取 $t_1=x_{(1)},t_2=x_{(n)}$，并令 $g\left(t_1,t_2,\theta\right)=\left(\dfrac{1}{\theta}\right)^n I_{\left\{\theta<t_1\le t_2<2\theta\right\}},h\left(X\right)=1$

由因子分解定理知，$T=\left(t_1,t_2\right)=\left(x_{(1)},x_{(n)}\right)$是 θ 的充分统计量.

<u>***均匀分布族不是指数型分布族**</u>

正态分布：$N\left(\mu,\sigma^2\right)$

（书 P282/Eg5.5.5）：设 $x_1,\cdots,x_n$ 是取自总体 $N\left(\mu,\sigma^2\right)$ 的样本，$\theta=\left(\mu,\sigma^2\right)$ 是未知的，

解：联合密度函数为 $p\left(x_1,\cdots,x_n;\theta\right)=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp(-\dfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x-\mu)^2)$

$$=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp(-\frac{n\mu^2}{2\sigma^2})\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}x^2+2\mu\sum_{i=1}^{n}x_i\right)\right\}$$

取 $t_1=\sum_{i=1}^{n}x_i,t_2=\sum_{i=1}^{n}x_i^2$,

并令 $g(t_1,t_2,\theta)=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp(-\dfrac{n\mu^2}{2\sigma^2})\exp\left\{-\dfrac{1}{2\sigma^2}\left(t_2+2\mu t_1\right)\right\},h(X)=1$

由因子分解定理知，$T=\left(t_1,t_2\right)=\left(\sum_{i=1}^{n}x_i,\sum_{i=1}^{n}x_i^2\right)$是充分统计量.

进一步，我们指出这个统计量与$\left(\overline{x},s^2\right)$是一一对应的，

这明在正态总体场合常用的$\left(\overline{x},s^2\right)$是充分统计量.

幂分布：

解 1：样本联合密度函数为 $p\left(x_1,\cdots,x_n;\theta\right)=\prod_{i=1}^{n}\theta x_i^{\theta-1}=\theta^n\left(\prod_{i=1}^{n}x_i\right)^{\theta-1}$

取 $t=\prod_{i=1}^{n}x_i$，并令 $g\left(t;\theta\right)=\theta^n t^{\theta-1},h\left(X\right)=1$

由因子分解定理知，$T=\prod_{i=1}^{n}x_i$ 是充分统计量.

解 2：样本联合密度函数为 $p\left(x_1,\cdots,x_n;\theta\right)=\prod_{i=1}^{n}\theta x_i^{\theta-1}=\theta^n\left(\prod_{i=1}^{n}x_i\right)^{\theta-1}$

$$=\theta^n\exp\left\{\ln\left[\left(\prod_{i=1}^{n}x_i\right)^{\theta-1}\right]\right\}$$

$$=\theta^n\exp\left\{\sum_{i=1}^{n}\ln x_i\cdot\left(\theta-1\right)\right\}$$

$$=\theta^n e^{\theta-1}\,\mathrm{e}^{\sum_{i=1}^{n}\ln x_i}$$

取 $t=\sum_{i=1}^{n}\ln x_i$，并令 $g\left(t;\theta\right)=\theta^n e^{\theta-1}\,\mathrm{e}^t,h\left(X\right)=1$

由因子分解定理知，$T=\sum_{i=1}^{n}\ln x_i$ 是充分统计量.

$\boxed{\text{双参数指数分布}}$:

（书 P284/x5.5.１2）：设 $x_1,\cdots,x_n$ 是来自双参数指数分布

$p\left(x;\theta,\mu\right)=\frac{1}{\theta}e^{-\frac{x-\mu}{\theta}},x>\mu,\theta>0$ 的样本，证明 $\left(\bar{x},x_{(1)}\right)$ 是分统计量.

解：样本联合密度函数为 $p\left(x_1,\cdots,x_n;\theta,\mu\right)=\prod_{i=1}^{n}\frac{1}{\theta}e^{-\frac{x_i-\mu}{\theta}}I_{\left\{x_{(1)}>\mu\right\}}$

$$=\left(\frac{1}{\theta}\right)^n e^{-\frac{1}{\theta}\sum_{i=1}^{n}\left(x_i-\mu\right)}I_{\left\{x_{(1)}>\mu\right\}}$$

$$=\left(\frac{1}{\theta}\right)^n e^{-\frac{1}{\theta}\left(n\bar{x}-n\mu\right)}I_{\left\{x_{(1)}>\mu\right\}}$$

取 $t_1=\bar{x},t_2=x_{(1)}$

并令 $g(t_1,t_2;\theta,\mu)=\left(\frac{1}{\theta}\right)^n e^{-\frac{1}{\theta}(n\bar{x}-n\mu)} I_{\{x_{(1)}>\mu\}}, h(X)=1$

由因子分解定理知，$T=(t_1,t_2)=(\bar{x},x_{(1)})$ 是充分统计量.

伽玛分布：Ga(α，λ)

解：样本联合密度函数为 $p(x_1,\cdots,x_n;\alpha,\lambda)=\prod_{i=1}^{n}\frac{\lambda^\alpha}{\Gamma(\alpha)}x_i^{a-1}e^{-\lambda x_i}$

$$=\frac{\lambda^\alpha}{\Gamma(\alpha)}\prod_{i=1}^{n}x_i^{a-1}e^{-\lambda x_i}$$

$$=\frac{\lambda^\alpha}{\Gamma(\alpha)}\left(\prod_{i=1}^{n}x_i\right)^{a-1}e^{-\lambda\sum_{i=1}^{n}x_i}$$

取 $t_1=\sum_{i=1}^{n}x_i, t_2=\prod_{i=1}^{n}x_i$，并令 $g(t_1,t_2;\alpha,\lambda)=\frac{\lambda^\alpha}{\Gamma(\alpha)}(t_2)^{a-1}e^{-\lambda t_1}, h(X)=1$

由因子分解定理知，$T=(t_1,t_2)=\left(\sum_{i=1}^{n}x_i,\prod_{i=1}^{n}x_i\right)$ 是充分统计量.

对数正态分布：$LN(\mu,\sigma^2)$

设 $x_1,\cdots,x_n$ 是取自总体 $LN(\mu,\sigma^2)$ 的样本，$\theta=(\mu,\sigma^2)$ 是未知的，

解：联合密度函数为 $p(x_1,\cdots,x_n;\theta)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln x-\mu)^2)$

$$=(2\pi\sigma^2)^{-\frac{n}{2}}\exp(-\frac{n\mu^2}{2\sigma^2})\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(\ln x_i)^2+2\mu\sum_{i=1}^{n}\ln x_i\right)\right\}$$

取 $t_1=\sum_{i=1}^{n}\ln x_i, t_2=\sum_{i=1}^{n}(\ln x_i)^2$,

并令 $g(t_1,t_2,\theta)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp(-\frac{n\mu^2}{2\sigma^2})\exp\left\{-\frac{1}{2\sigma^2}(t_2+2\mu t_1)\right\}, h(X)=1$

由因子分解定理知，$T=(t_1,t_2)=\left(\sum_{i=1}^{n}\ln x_i,\sum_{i=1}^{n}(\ln x_i)^2\right)$ 是充分统计量.

贝塔分布：Be(a, b)

解：样本联合密度函数为 $p(x_1,\cdots,x_n;a,b)=\prod_{i=1}^{n}\frac{1}{B(a,b)}x_i^{a-1}(1-x_i)^{b-1}$

$$=\frac{1}{B(a,b)}\prod_{i=1}^{n} x_i^{a-1}(1-x_i)^{b-1}$$

$$=\frac{1}{B(a,b)}\exp\left\{\ln\left(\prod_{i=1}^{n} x_i^{a-1}(1-x_i)^{b-1}\right)\right\}$$

$$=\frac{1}{B(a,b)}\exp\left\{\ln\left(\prod_{i=1}^{n} x_i^{a-1}\right)+\ln\left(\prod_{i=1}^{n}(1-x_i)^{b-1}\right)\right\}$$

$$=\frac{1}{B(a,b)}\exp\left\{\ln\left(\prod_{i=1}^{n} x_i\right)(a-1)+\ln\left(\prod_{i=1}^{n}(1-x_i)\right)(b-1)\right\}$$

$$=\frac{1}{B(a,b)}\exp\left\{\sum_{i=1}^{n}\ln x_i\,(a-1)+\sum_{i=1}^{n}\ln(1-x_i)(b-1)\right\}$$

取 $t_1=\sum_{i=1}^{n}\ln x_i, t_2=\sum_{i=1}^{n}\ln(1-x_i)$，

并令 $g(t_1,t_2;a,b)=\dfrac{1}{B(a,b)}\exp\{t_1(a-1)+t_2(b-1)\},h(X)=1$

由因子分解定理知，$T=(t_1,t_2)=\left(\sum_{i=1}^{n}\ln x_i,\sum_{i=1}^{n}\ln(1-x_i)\right)$ 是充分统计量.

2．常用分布族的完备性

Γ 分布族 $F=\{\Gamma(\lambda,v)\}$ 的完备性

若有 $\varphi(\lambda,v)=\int_0^\infty h(x)\dfrac{\lambda^v}{\Gamma(v)}e^{-\lambda x}x^{v-1}dx=0,\forall(\lambda,v)$，

则对任何 (λ,v) 有 $\int_0^\infty h(x)e^{-\lambda x}x^{v-1}dx=0$；该式左端可视为 $h(x)x^{v-1}$ 的拉氏变换，因此有拉氏变换的唯一性，可以推出 $h(x)x^{v-1}=0(a.e.)$，$x^{v-1}\neq 0$，即得 $h(x)\neq 0(a.e.)$.类似的，分布族 $F_0=\{\Gamma(\lambda,v_0\}$ 也完备.

正态分布族 $F_0=\{N(\mu,\sigma^2)\}$ 的完备性

1） $F_1=\{N(\mu,1),\mu\in(-\infty,+\infty)\}$ 完备.

解：因为对任何 μ，由 $E_\mu[h(X)]=\int\limits_{-\infty}^{\infty}h(x)\dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2}dx=0$

可以推得$\varphi(\mu)=\int_{-\infty}^{\infty}h(x)e^{-\frac{1}{2}x^2}e^{x\mu}dx=0$

有拉氏变换唯一性可知：$h(x)e^{-\frac{1}{2}x^2}=0(a.e.)$，即可得$h(x)=0(a.e.)$.

2）$F_2=\{N(\mu,\sigma_0),\mu\in(-\infty,+\infty)\}$完备.与1）类似.

3）$F_3=\{N(0,\sigma^2),\sigma^2>0\}$不完备.因为$h(x)=x,E_\sigma[h(X)]=0$，但$h(x)\neq 0(a.e.)$

4）$F_4=\{N(\mu_0,\sigma^2),\sigma^2>0\}$不完备.与3）类似.

5）$F=\{N(\mu,\sigma^2),\forall\mu,\sigma\}$完备.因为若对任何$(\mu,\sigma)$有$\int h(x)\varphi_{\mu,\sigma}(x)dx=0$，其中$\varphi_{\mu,\sigma}(x)$为正态分布$N(\mu,\sigma^2)$的密度函数，必有$\int h(x)\varphi_{\mu,1}(x)dx=0$，$\forall\mu$，由（1）知$h(x)=0(a.e.)$.

二项分布族$F=\{b(n,\theta),\theta\in(0,1)\}$的完备性

若对任何θ有 $\varphi(\theta)=E_\theta[h(X)]=0$，即$\varphi(\theta)=\sum_{x=0}^{n}h(x)\binom{n}{x}\theta^x(1-\theta)^{n-x}=0$

由此可推出$\sum_{x=0}^{n}h(x)\binom{n}{x}(\frac{\theta}{1-\theta})^x=0$，该式为$y=\frac{\theta}{1-\theta}$的$n$次多项式，它对一切$y>0$为零，则其系数必为零，即$h(x)\binom{n}{x}=0$，所以$h(x)=0,x=0,1,\cdots,n$.

均匀分布族$F=\{R(0,\theta),\theta>0\}$的完备性

若对任何θ有$E_\theta[h(X)]=\int_0^\theta h(x)\theta^{-1}dx=0$，则$\varphi(\theta)=\int_0^\theta h(x)dx=0$.由于$h(x)$可测，其不连续点为零测集，在$h(x)$的连续点处，$\varphi(\theta)$可导，因此对任何$h(x)$的连续点$\theta$处有$\varphi'(\theta)=h(x)=0$，即$h(\theta)=0$，$\forall\theta(a.e.)$，因此有$h(x)=(a.e.)$

3. 因子分解定理中的是不是向量统计量？

答： 假如存在充分统计量$T(X)$，那么样本分布$f_\theta(x)$一定可以分解为两个因子的乘积，其中一个因子与θ无关，仅与样本有关，另一个因子与θ有关，但与样本的关系可以通过充分统计量$T(X)$表现出来.所以，应该指出，这个定理中的$T(X)$可以是向量统计量.

4．用指数族去解决问题的完全性有多大的作用？

答：我们通过学习，可以总结出指数族的三个<u>优点</u>：

1）是它包含了很多常见的分布.

2）其次是它有良好的分析性质.

3）是它有（在定理条件下）完全充分统计量.

这后两条性质决定了许多问题在这个族中有满意的解决，因此，指数族的重要性就可想而知了.

5． 分布族要有怎样的性质，才能使次序统计量有完全性？

答：先引进若干有关的记号，设 $F_1, F_2, \cdots F_\gamma$ 为 γ 个一维概率测度，$P_i \ge 0, i = 1, 2, \cdots, \gamma$ 而 $P_1 + \cdots + P_\gamma = 1$，则 $F = \sum_{i=1}^{\gamma} P_i F_i$ 理解为一概率测度，定义为 $F(S) = \sum_{i=1}^{\gamma} P_i F_i(S)$，对任何 $S \in \beta_1$，又若 F 为一概率测度，$S \in \beta_1$ 而 $F(S) \ne 0$，则记号 F_S 表示一个概率测度定义为 $F_S(A) = F(S \cap A) / F(S), A \in \beta_1$.

6． 充分统计量的函数是不是充分统计量？

答：设 $X \sim N(\mu, \sigma^2)$，σ^2 已知，$X = (X_1, \ldots, X_n)$ 是抽自 X 的 iid 样本，则依因子分解可知 $\overline{X}$ 是 μ 的充分统计量，但 $\overline{X^2}$ 不是 μ 的充分统计量，事实上

$$f(X_1, \ldots, X_n \mid \overline{X^2} = t) = \frac{(\frac{1}{\sqrt{2\pi\sigma}})^n Exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\}}{f(t)} = \frac{(\frac{1}{\sqrt{2\pi\sigma}})^n Exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\}}{\frac{\sqrt{n}}{\sqrt{2\pi}\sigma}[Exp\{-\frac{n(-\sqrt{t} - \mu)^2}{2\sigma^2} + Exp\{-\frac{n(\overline{X} - \mu)^2}{2\sigma^2}\}]}$$

与 μ 有关，故 $\overline{X^2}$ 不是 μ 的充分统计量

因为 $\overline{X} \sim \varphi(t) = N(\mu, \frac{\sigma^2}{n}) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} Exp\{-\frac{n(\sqrt{t} - \mu)^2}{2\sigma^2}\}$

$$\overline{X^2} \sim f(t) = \varphi(-\sqrt{t})\left|\frac{1}{-2\sqrt{t}}\right| + \varphi(t)\left|\frac{1}{2\sqrt{t}}\right| = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} Exp\{-\frac{n(-\sqrt{t} - \mu)^2}{2\sigma^2}\}\frac{1}{2\sqrt{t}} + \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} Exp\{\frac{-n(\sqrt{t} - \mu)^2}{2\sigma^2}\}$$

7. 指数族分布表达式中的是不是充分完全统计量？

答：我们可以给出一个指数族分布，其中并不是参数的充分完全统计量.

设 $X \sim N(\mu, \sigma^2)$，其中 $\sigma > 0$ 是未知参数，$X = (X_1, \cdots X_n)$ 是抽自其中的 iid 样本，

则 $T(X) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ 是 σ 的充分统计量.但不是 σ 的完全充分统计量，事实上，

因为 $X \sim N(\mu, \sigma^2)$，所以，子样 $X = (X_1, \cdots X_n)$ 的联合分布为

$$L(x;\theta)=\prod_{i=1}^{n} f(x_i;\theta)=(\frac{1}{\sqrt{2\pi}\sigma})^n Exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\sigma)^2\}=)^n Exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^{\ 2}+\frac{1}{\sigma}\sum_{i=1}^{n}x_i-\frac{n}{2}\}$$

令 $h(X)=1$ $\quad g(T(X),\theta)=(\frac{1}{\sqrt{2\pi}\sigma})^n Exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}x_i-\frac{n}{2}\}$

则 $L(X;\theta)=g(T(X),\theta)h(X)$

所以，根据 Fisher-Neyman 因子分解定理得 $T(X)=(\sum_{i=1}^{n}x_i,\sum_{i=1}^{n}x_i^2)$ 是 σ 的充分统计量，记 $T(X)=(\sum_{i=1}^{n}x_i,\sum_{i=1}^{n}x_i^2)=(u,v)$

则 $E(u)=\sum_{i=1}^{n}E(X)=n\sigma$ $\quad Var(u)=\sum_{i=1}^{n}Var(x_i)=n\sigma^2$

所以, $E(u^2)=Var(u)+(E^2(u))=n(n+1)\sigma^2$

因为， $E(v)=\sum_{i=1}^{n}E(x_i^2)=\sum_{i=1}^{n}(Var(x_i)+E^2(x_i))=2n\sigma^2$

令 $g(T)=2u^2-(n+1)v$ $\quad E(g(T))=2n(n+1)\sigma^2-(n+1)2n\sigma^2=0$

但 $n\geq 2$ 时， $g(T)\neq 0$，故 $\sigma>0$ 不是 σ 的完全统计量

那么，为什么会出现这种情况呢？原因是上述定理中 $\theta=(\theta_1,\theta_2)=(-\frac{1}{2\sigma^2},\frac{1}{\sigma})$ 的值域应包含一个二维开集的条件得不到满足

进一步的讨论，得

注：设 $X_1,\ldots,X_n$ 是抽自 $N(\alpha\sigma,\sigma^2)$ 分布族的 iid 样本，其中 α 是已知的实数，$\sigma>0$ 是未知参数，则同样可知， $T(X)=(\sum_{i=1}^{n}x_i,\sum_{i=1}^{n}x_i^2)$ 是 σ 的充分统计量，但不是 σ 的完全统计量

注：设 $X\sim N(\mu,\sigma_1^2)$ ， $Y\sim N(\mu,\sigma_2^2)$ 其中 $(\mu,\sigma_1^2,\sigma_2^2)$ 是未知参数

又设 $(X_1,X_2\ldots,X_m)$， $(Y_1,Y_2\ldots,Y_m)$ 是分别从 X,Y 母体中抽取的 iid 样本，可证明

$$T(X,Y)=(\sum_{i=1}^{m}X_i,\sum_{j=1}^{n}Y_i,\sum_{i=1}^{m}X_i^2,\sum_{j=1}^{n}Y_i^2)$$

是 $(u,\sigma_1^2,\sigma_2^2)$ 的充分统计量，但不是完全统计量，原因是上述定理中

$Q=(Q_1,Q_2,Q_3,Q_4)=(\frac{1}{\sigma_1^2},\frac{1}{\sigma_2^2},\frac{1}{-2\sigma_1^2},\frac{1}{-2\sigma_2^2})$ 的值域应包含一个四维井集的条件得不到满足.

三、【典型例题和研究生试题】

1.（充分统计量）

（1） 设 $x_1,\cdots,x_n$ 是来自

$$p(x;\theta)=\theta\cdot x^{\theta-1},0<x<1,\theta>0$$

的样本，试给出一个充分统计量.

解： 样本的联合密度函数为

$$p(x_1,x_2,\cdots,x_n;\theta)=\theta^n(x_1x_2\cdots x_n)^{\theta-1}=\theta^n\left(\prod_{i=1}^n x_i\right)^{\theta-1},$$

令 $T=\prod_{i=1}^n x_i$，取 $g(t;\theta)=t^{\theta-1}\theta^n,h(x_1,\cdots,x_n)=1$，由因子分解定理，$T=\prod_{i=1}^n x_i$ 为 θ 的充分统计量.另外，T 的一一变换得到的统计量，如 $x_1,\cdots,x_n$ 的几何平均 $(x_1\cdots x_n)^{1/n}$ 或其对数 $\frac{1}{n}\sum_{i=1}^n \ln x_i$，都是 θ 的充分统计量.

注：（概率论与数理统计教程 习题与解答--茆诗松 程依明 濮晓龙）P257~261

（2） 设 $x_1,\cdots,x_n$ 是来自正态分布 $N(\mu,\sigma^2)$ 的样本.

（1） 在 μ 已知时给出 σ^2 的一个充分统计量；

（2） 在 σ^2 已知时给出 μ 的一个充分统计量.

解：（1）在 μ 已知时，样本联合密度函数为

$$p(x_1,x_2,\cdots,x_n;\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2\right\}.$$

令 $T=\sum_{i=1}^n(x_i-\mu)^2$，取 $g(t;\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{t}{2\sigma^2}\right\},h(x)=1$，由因子分解定理，

$T=\sum_{i=1}^n(x_i-\mu)^2$ 为 σ^2 的充分统计量.

（2）在 σ^2 已知时，样本联合密度函数为

$$p\left(x_1,\cdots,x_n;\mu\right)=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right\}$$
$$=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^{\ 2}\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(n\mu^2-2\mu\sum_{i=1}^{n}x_i\right)\right\}.$$

令 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ ，取

$$g\left(\bar{x};\mu\right)=\exp\left\{-\frac{1}{2\sigma^2}\left(n\mu^2-2n\mu\bar{x}\right)\right\},h\left(x\right)=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^{\ 2}\right\},$$

由因子分解定理，$\bar{x}$ 为 μ 的充分统计量.

注：（概率论与数理统计教程 习题与解答--茆诗松 程依明 濮晓龙）P257~261

（**3**）设 $Y_i\sim N\left(\beta_0+\beta_1x_i,\sigma^2\right),i=1,2,\cdots,n,$ 诸 Y_i 独立， $x_1,\cdots,x_n$ 是已知常数，证明 $\left(\sum_{i=1}^{n}Y_i,\sum_{i=1}^{n}x_iY_i,\sum_{i=1}^{n}Y_i^2\right)$ 是充分统计量.

证：$Y_1,\cdots,Y_n$ 的联合密度函数为

$$p\left(y_1,\cdots,y_n\right)=\prod_{i=1}^{n}\left\{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left(y_i-\beta_0-\beta_1x_i\right)^2\right\}\right\}$$
$$=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\beta_0-\beta_1x_i\right)^2\right\}$$
$$=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}y_i^2+n\beta_0^2+\beta_1^2\sum_{i=1}^{n}x_i^2-2\beta_0\sum_{i=1}^{n}y_i-2\beta_1\sum_{i=1}^{n}x_iy_i+2\beta_0\beta_1\sum_{i=1}^{n}x_i\right)\right\}.$$

注意到 $x_1,\cdots,x_n$ 是已知常数，令 $t=\left(t_1,t_2,t_3\right)=\left(\sum_{i=1}^{n}y_i,\sum_{i=1}^{n}x_iy_i,\sum_{i=1}^{n}y_i^{\ 2}\right)$，取

$$g\left(t,\sigma^2,\beta_0,\beta_1\right)=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left(n\beta_0^2+\beta_1^2\sum_{i=1}^{n}x_i^2+2\beta_0\beta_1\sum_{i=1}^{n}x_i\right)\right\}\cdot$$
$$\exp\left\{-\frac{1}{2\sigma^2}\left(t_3-2\beta_0t_1-2\beta_1t_2\right)\right\},$$

$h\left(y_1,\cdots,y_n\right)=1.$

由因子分解定理，$\left(\sum_{i=1}^{n}Y_i,\sum_{i=1}^{n}x_iY_i,\sum_{i=1}^{n}Y_i^2\right)$ 是 $\left(\beta_0,\beta_1,\sigma^2\right)$ 的充分统计量.

注：（概率论与数理统计教程 习题与解答--茆诗松 程依明 濮晓龙）P257~261

（**4**）设 $x_1,\cdots,x_n$ 是来自正态总体 $N\left(\mu,\sigma_1^2\right)$ 的样本， $y_1,\cdots,y_n$ 是来自另一正态总体 $N\left(\mu,\sigma_2^2\right)$ 的样本，这两个样本相互独立，试给出 $\left(\mu,\sigma_1^2,\sigma_2^2\right)$ 的充分统计量.

解：样本 $x_1,\cdots,x_n$ ，的联合密度函数为

$$p\left(x_1,\cdots,x_n,y_1,\cdots,y_m\right)=\prod_{i=1}^{n}\left\{\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2\sigma_1^2}\left(x_i-\mu\right)^2}\right\}\prod_{i=1}^{m}\left\{\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2\sigma_2^2}\left(y_i-\mu\right)^2}\right\}$$

$$=\left(2\pi\right)^{-(n+m)/2}\sigma_1^{-n}\sigma_2^{-m}e^{-\left\{\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}x_i^2+\frac{1}{2\sigma_2^2}\sum_{i=1}^{m}y_i^2-\left(\frac{n\overline{x}}{\sigma_1^2}+\frac{m\overline{y}}{\sigma_2^2}\right)\mu+\left(\frac{n}{2\sigma_1^2}+\frac{m}{2\sigma_2^2}\right)\mu^2\right\}}$$

其中 $\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\overline{y}=\frac{1}{m}\sum_{i=1}^{m}y_i$ ， 令 $t=\left(t_1,t_2,t_3,t_4\right)=\left(\overline{x},\overline{y},\sum_{i=1}^{n}x_i^2,\sum_{i=1}^{m}y_i^2\right)$，取

$$g\left(t,\mu,\sigma_1^2,\sigma_2^2\right)=\left(2\pi\right)^{-(n+m)/2}\sigma_1^{-n}\sigma_2^{-m}e^{-\left\{\frac{1}{2\sigma_1^2}t_3+\frac{1}{2\sigma_2^2}t_4-\left(\frac{n}{\sigma_1^2}t_1+\frac{m}{\sigma_2^2}t_2\right)\mu+\left(\frac{n}{2\sigma_1^2}+\frac{m}{2\sigma_2^2}\right)\mu^2\right\}},$$

$$h\left(x_1,\cdots,x_n,y_1,\cdots,y_m\right)=1,$$

由因子分解定理，$t=\left(t_1,t_2,t_3,t_4\right)=\left(\overline{x},\overline{y},\sum_{i=1}^{n}x_i^2,\sum_{i=1}^{m}y_i^2\right)$ 是 $\left(\mu,\sigma_1^2,\sigma_2^2\right)$ 的充分统计量.

注：（概率论与数理统计教程 习题与解答--茆诗松 程依明 濮晓龙）P257~261

2.（完全性）

证 $Ga(\alpha,\lambda)$ 具有完全性**.**

证：$\int_0^{+\infty}g(x)\frac{\lambda^\alpha}{\gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}dx=\varphi(\alpha,\lambda)=0$，对 $\forall(\alpha,\lambda)$，有 $\int_0^{+\infty}g(x)x^{\alpha-1}e^{-\lambda x}dx=0$

有拉式变换唯一性知 $g(x)x^{\alpha-1}=0\Rightarrow g(x)=0$**.**

3.（辅助统计量）

设 $X_1,\ldots,X_n$ 独立同分布， $X_1\sim N(\theta,1)$，则 $S=\sum_{i=1}^{n}(X_i-\overline{X})^2$ 为辅助统计量**.**

解：因为 $S\sim\chi^2(n-1)$ 与 θ 无关， $T=X_{(n)}-X_{(1)}$ 亦为辅助统计量，因为 $T=(X_{(n)}-\theta)-(X_{(1)}-\theta)=Y_{(n)}-Y_{(1)}$，而 $Y_i=X_i-\theta\sim N(0,1)$，其分布与 θ 无关

4.（极小充分统计量）

设 $X_1,\ldots,X_n$ 为相互独立的样本，且 $X_j \sim N(0,\sigma^2)$ 一切 $j \neq i$，求完备的极小充分统计量.

（1） 若 $X_i \sim N(\gamma,\sigma^2)$

（2） 若 $X_i \sim N(0,\omega^{-1}\sigma^2)$ 其中 $\gamma \in R$，$\omega > 0$，σ^2 都是未知参数

证：**（1）** $X = (X_1,\ldots,X_n)^T$ 的联合分布可表示为

$$f(x,\theta) = Exp\{\frac{1}{-2\sigma^2}\sum_{j=1}^{n} X_j^2 + \frac{\gamma^2}{2\sigma^2}X_i - \frac{\gamma^2}{2\sigma^2} - \frac{n}{2}\log(2\pi\sigma^2)\}$$

其中 $T_1(X) = \sum_{i=1}^{n} X_j^2$，$Q_1(\theta) = -\dfrac{1}{2\sigma^2}$；$T_2(X) = X_i$，$Q_2(X) == \dfrac{\gamma}{\sigma^2}$；

$b(\theta) = \dfrac{\gamma^2}{2\sigma^2} + n\log\sqrt{2\pi\sigma^2}$，因此，为 $(\sum_{j=1}^{n} X_j^2, X_i)$ 完备的极小充分统计量

（2） 类似的，$X = (X_1,\ldots,X_n)^T$ 的联合分布可表示为

$$f(x,\theta) = Exp\{\frac{1}{-2\sigma^2}\sum_{j\neq i} X_j^2 - \frac{\omega}{2\sigma^2}X_i^2 - \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(\omega)\}$$

其中，$T_1(X) = \sum_{j\neq i} X_j^2$，$Q_1(X) = -\dfrac{1}{2\sigma^2}$；$T_2(X) = X_i^2$，$Q_2(X) = -\dfrac{\omega}{2\sigma^2}$

$b(\theta) = n\log\sqrt{2\pi\sigma^2} - \dfrac{1}{2}\log\omega$，因此，$(\sum_{j\neq i}^{n} X_j^2, X_i)$ 为完备的极小充分统计量.

5.（完备性的极小充分统计量）

$X_1,\ldots,X_n$ 独立同分布，$X_1 \sim R(0,\theta)$，则 $X_{(n)}$ 为完备的极小充分统计量.

证：$\dfrac{X_{(n)}}{\theta} \sim Be(n,1)$，故有 $X_{(n)} \sim f(t,\theta) = n(\dfrac{t}{\theta})^{n-1}\theta^{-1}I_{\{0\le t\le\theta\}}$

若 $E_\theta[h(T)] = 0$，则 $\int_0^\theta h(t)t^{n-1}dt = 0$，该式在 $h(t)$ 的连续处对 θ，求导可得

$h(\theta)\theta^{n-1} = 0$，所以 $h(\theta) = 0$，即 $h(t) = 0$.因此，$T = X_{(n)}$ 为完备的极小充分统计量

注：若 $X_1,\ldots,X_n$ 独立同分布，$X_1 \sim R(\theta_1,\theta_2)$，则 $T = (X_{(1)}, X_{(n)})$ 为完备的极小充分

统计量.

6.（完备性）

设总体 X 在 (θ_1,θ_2) 上服从均匀分布，概率密度函数为

$$f(x,\theta_1,\theta_2)=\begin{cases}\dfrac{1}{\theta_2-\theta_1} & \theta_1<x<\theta_2\\ 0 & \text{其它}\end{cases}$$

$X_1,\ldots,X_n$ 为子样，顺序统计量为 $X_{(1)}\le\ldots\le X_{(n)}$，当 $\theta_1=\theta>0,\theta_2=2\theta$ 时，试证 $X_{(1)}$ 及 $X_{(n)}$ 都为充分统计量，但并非完备统计量.

证：子样的联合密度函数为：$f(X_1,\ldots,X_n,\theta)=\theta^{-n}I_{(0<X_{(1)}<X_{(n)}<2\theta)}$

可见 $X_{(1)}$ 及 $X_{(n)}$ 都为的充分统计量.

总体 X 的概率密度函数及分布函数为

$$f(x,\theta)=\begin{cases}\dfrac{1}{\theta} & \theta<x<2\theta\\ 0 & \text{其它}\end{cases}\qquad \mathrm{F}(x,\theta)=\begin{cases}0 & x\le 0\\ \dfrac{x}{\theta} & 0<x<2\theta\\ 1 & x\ge 2\theta\end{cases}$$

$X_{(1)}$ 的分布函数为

$$F_1(v)=P(X_{(1)}<v)=1-P(X_{(1)}\ge v)=1-\prod_{i=1}^{n}[1-F(v)]=1-(2-\frac{v}{\theta})^n,$$

$0<v<2\theta$

密度函数为 $f_1(v)=n(2\theta-v)^{n-1}/\theta^n$，$0<v<2\theta$

次序统计量 $p_k(x)$

$=\dfrac{n!}{(k-1)!(n-k)!}\bullet$

$(F(x))^{k-1}(1-F(x))^{n-k}p(x)$

$$E(X_1)=\int_{\theta}^{2\theta}vf_1(v)=\theta[1+2(-1)^n+\frac{1+(-1)^{n-1}}{n+1}]$$

$X_{(n)}$ 的分布函数为

$$F_1(u)=P(X_{(n)}<u)=\prod_{i=1}^{n}P(X_i<u)=(\frac{u}{\theta}-1)^n,\quad 0<u<2\theta$$

密度函数为 $f_n(u)=n(u-\theta)^{n-1}/\theta^n$，$0<u<2\theta$

$$E(X_n)=\int_{\theta}^{2\theta}uf_n(u)=\theta[1+2(-1)^n+\frac{1+(-1)^{n-1}}{n+1}]$$

记$T=(X_{(1)},X_{(n)})$

又$g(T)=X_{(1)}[1+2(-1)^n+\frac{1+(-1)^{n-1}}{n+1}]^{-1}-X_{(n)}[1+2(-1)^n+\frac{1+(-1)^{n-1}}{n+1}]^{-1}$

则$E[g(T)]=0$，但$g(T)\neq 0$，于是由$X_{(1)}$及$X_{(n)}$组成的统计量是非完备的.

7.（英文习题）

(1).

6.1 Let X be one observation from a $\mathrm{n}(0,\sigma^2)$ population. Is $|X|$ a sufficient statistic?

Let X be one observation from a $n\left(0,\sigma^2\right)$ population. Is $|X|$ a sufficient statistic?

解：By the Factorization Theorem, $|X|$ is sufficient because the pdf of X is

$$f\left(x|\sigma^2\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/2\sigma^2}=\frac{1}{\sqrt{2\pi}\sigma}e^{-|x|^2/2\sigma^2}=g\left(|x|\|\sigma^2\right)\cdot\underbrace{1}_{h(x)}$$

中文注释：随机变量X是来自总体$n\left(0,\sigma^2\right)$的一个样本.$|X|$是一个充分统计量吗？

解：由因子分解定理知，X的密度函数为

$$f\left(x|\sigma^2\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/2\sigma^2}=\frac{1}{\sqrt{2\pi}\sigma}e^{-|x|^2/2\sigma^2}=g\left(|x|\|\sigma^2\right)\cdot\underbrace{1}_{h(x)}$$

所以，$|X|$是充分统计量.

注：**E02 Statistical Inference(2ed) by George Casella**

(2).

6.2 Let $X_1,\ldots,X_n$ be independent random variables with densities

$$f_{X_i}(x|\theta)=\begin{cases}e^{i\theta-x} & x\geq i\theta\\ 0 & x<i\theta.\end{cases}$$

Prove that $T=\min_i(X_i/i)$ is a sufficient statistic for θ.

Let $X_1,\cdots,X_n$ be independent random wariables with densities

$$fx_i\left(x|\theta\right)=\begin{cases}e^{i\theta-x} & x\geq i\theta\\ 0 & x<i\theta\end{cases}$$

Prove that $T=\min_i\left(X_i/i\right)$ is a sufficient statistic for θ.

解：By the Factorization Theorem, $T(X)=\min_i(X_i/i)$ is sufficient because the joint pdf is

$$f(x_1,\cdots,x_n|\theta)=\prod_{i=1}^{n}e^{i\theta-x_i}I_{(i\theta,+\infty)}(x_i)=\underbrace{e^{in\theta}I_{(\theta,+\infty)}(T(x))}_{g(T(x)|\theta)}\cdot\underbrace{e^{-\sum_i x_i}}_{h(x)}.$$

Notice ,we use the fact that $i>0$,and the fact that all $x_i s>i\theta$ if and only if $\min_i(x_i/i)>\theta.$

中文注释：随机变量的密度函数为 $fx_i(x|\theta)=\begin{cases}e^{i\theta-x} & x\geq i\theta\\ 0 & x<i\theta\end{cases}$

证明$T=\min_i(X_i/i)$是θ的充分统计量。

解：由因子分解定理知，联合密度函数为

$$f(x_1,\cdots,x_n|\theta)=\prod_{i=1}^{n}e^{i\theta-x_i}I_{(i\theta,+\infty)}(x_i)=\underbrace{e^{in\theta}I_{(\theta,+\infty)}(T(x))}_{g(T(x)|\theta)}\cdot\underbrace{e^{-\sum_i x_i}}_{h(x)}.$$

所以，$T(X)=\min_i(X_i/i)$是充分统计量.

注意：我们这里用了$i>0$，事实上当且仅当$\min_i(x_i/i)>\theta.$时，$x_i s>i\theta$.

注：E02 Statistical Inference(2ed) by George Casella

(3).

6.6 Let $X_1,\ldots,X_n$ be a random sample from a gamma(α,β) population. Find a two-dimensional sufficient statistic for (α,β).

Let $X_1,\cdots,X_n$ be a random sample from a $gamma(\alpha,\beta)$ population.Find a two-dimensional sufficient statistic for (α,β).

解：The joint pdf is given by

$$f(x_1,\cdots,x_n|\alpha,\beta)=\prod_{i=1}^{n}\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x_i^{\alpha-1}e^{-x_i/\beta}=\left(\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right)^n\left(\prod_{i=1}^{n}x_i\right)^{\alpha-1}e^{-\sum_i x_i/\beta}.$$

By the Factorization Theorem, $\left(\prod_{i=1}^{n}X_i,\sum_{i=1}^{n}X_i\right)$ is sufficient for (α,β).

中文注释：$X_1,\cdots,X_n$是来自总体$gamma(\alpha,\beta)$的样本.试找到一个(α,β)的二维的充分统计量.

解：由题得联合密度函数为

$$f\left(x_1,\cdots,x_n\middle|\alpha,\beta\right)=\prod_{i=1}^{n}\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x_i^{\alpha-1}e^{-x_i/\beta}=\left(\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right)^n\left(\prod_{i=1}^{n}x_i\right)^{\alpha-1}e^{-\sum_i x_i/\beta}.$$

由因子分解定理知，$\left(\prod_{i=1}^{n}X_i,\sum_{i=1}^{n}X_i\right)$是$(\alpha,\beta)$的充分统计量.

注：E02 Statistical Inference(2ed) by George Casella

(4).

Let $X_1,\ldots,X_n$ be iid with geometric distribution**.**

$$P_\theta(X=x)=\theta(1-\theta)^{x-1},x=1,2,\cdots,\ 0<\theta<1$$

show that $\sum X_i$ is sufficient for θ,and find the family of distribution of $\sum X_i$,Is the family complete?

中文注释： $X_1,\ldots,X_n$ 互相独立，且服从同一几何分布

$$P_\theta(X=x)=\theta(1-\theta)^{x-1},x=1,2,\cdots,\ 0<\theta<1$$

证明$\sum X_i$对θ是充分的，并找出$\sum X_i$的分布族，这个分布族是完备的吗？

证明：因为 $X_i\sim Ge(\theta)$，设$T=\sum_{i=1}^{n}X_i\sim Nb(n,\theta)$

所以有 $P(T=t)=\binom{t+n-2}{n-1}\theta^n(1-\theta)^{t-1},t=0,1,2,\cdots$

当$T=t$时，样本的条件分布为

$$P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n\,|\,T=t)$$

$$=\frac{P(X_1=x_1,X_2=x_2,\cdots,X_n=t-\sum_{i=1}^{n-1}X_i)}{P(T=t)}$$

$$=\frac{\prod_{i=1}^{n-1}P(X_i=x_i)P(X_n=t-\sum_{i=1}^{n-1}X_i)}{\binom{t+n-2}{n-1}\theta^n(1-\theta)^{t-1}}=\frac{\prod_{i=1}^{n-1}\theta^n(1-\theta)^{x_i-1}\theta^n(1-\theta)^{t-\sum_{i=1}^{n-1}X_i)}}{\binom{t+n-2}{n-1}\theta^n(1-\theta)^{t-1}}$$

$$=\frac{\theta^n(1-\theta)^{(n-1)x_i-(n-1)+t-1-(n-1)x_i}}{\binom{t+n-2}{n-1}\theta^n(1-\theta)^{t-1}}=\frac{(1-\theta)^{t-n}}{\binom{t+n-2}{n-1}(1-\theta)^{t-1}}$$

$p(x \mid \theta) = \theta(1-\theta)^{z-1} = \dfrac{\theta}{1-\theta} e^{\log(1-\theta)x}$, an exponential family with $t(x) = x$,Thus, $\sum_i X_i$ is a complete, sufficient statistic. (此密度函数是一个关于 $t(x) = x$ 的指数族，由释难解疑第一题证明 $\sum_i X_i$ 是一个充分完备统计量)

(5).

6.9 For each of the following distributions let $X_1, \ldots, X_n$ be a random sample. Find a minimal sufficient statistic for θ.

(a) $f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty$ (normal)

(b) $f(x|\theta) = e^{-(x-\theta)}, \quad \theta < x \leq \infty, \quad -\infty < \theta < \infty$ (location exponential)

(c) $f(x|\theta) = \frac{e^{-(x-\theta)}}{\left(1+e^{-(x-\theta)}\right)^2}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty$ (logistic)

(d) $f(x|\theta) = \frac{1}{\pi[1+(x-\theta)^2]}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty$ (Cauchy)

(e) $f(x|\theta) = \frac{1}{2} e^{-|x-\theta|}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty$ (double exponential)

For each of the following distributions let $X_1, \cdots, X_n$ be a random sample. Find a minimal sufficient statistic for θ.

(a) $f\left(x|\theta\right) = \dfrac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}, -\infty < x < \infty, -\infty < \theta < \infty$ (normal)

(b) $f\left(x|\theta\right) = e^{-(x-\theta)}, \theta < x < \infty, -\infty < \theta < \infty$ (location exponential)

(c) $f\left(x|\theta\right) = \dfrac{e^{-(x-\theta)}}{\left(1+e^{-(x-\theta)}\right)^2}, -\infty < x < \infty, -\infty < \theta < \infty$ (logistic)

(d) $f\left(x|\theta\right) = \dfrac{1}{\pi\left[1+\left(x-\theta\right)^2\right]}, -\infty < x < \infty, -\infty < \theta < \infty$ (Cauchy)

(e) $f\left(x|\theta\right) = \dfrac{1}{2} e^{-|x-\theta|}, -\infty < x < \infty, -\infty < \theta < \infty$ (double exponential)

解：a.

$$\frac{f\left(x|\theta\right)}{f\left(y|\theta\right)} = \frac{\left(2\pi\right)^{-n/2} e^{-\sum_i (x_i-\theta)^2/2}}{\left(2\pi\right)^{-n/2} e^{-\sum_i (y_i-\theta)^2/2}} = \exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2\right) + 2\theta n\left(\bar{y}-\bar{x}\right)\right]\right\}.$$

This is constant as a function of θ if and only if $\bar{y} = \bar{x}$;therefore $\bar{X}$ is a minimal sufficient statistic for θ.

b.Note, for $X \sim \text{location exponential}\left(\theta\right)$, the range depends on the parameter. Now

$$\frac{f(x|\theta)}{f(y|\theta)}=\frac{\prod_{i=1}^{n}\left(e^{-(x_i-\theta)}I_{(\theta,\infty)}(x_i)\right)}{\prod_{i=1}^{n}\left(e^{-(y_i-\theta)}I_{(\theta,\infty)}(y_i)\right)}$$

$$=\frac{e^{n\theta}e^{-\sum_i x_i}\prod_{i=1}^{n}I_{(\theta,\infty)}(x_i)}{e^{n\theta}e^{-\sum_i y_i}\prod_{i=1}^{n}I_{(\theta,\infty)}(y_i)}=\frac{e^{-\sum_i x_i}I_{(\theta,\infty)}(\min x_i)}{e^{-\sum_i y_i}I_{(\theta,\infty)}(\min y_i)}.$$

To make the ratio independent of θ we need the ratio of indicator functions independent of θ. This will be the case if and only if $\min\{x_1,\cdots,x_n\}=\min\{y_1,\cdots,y_n\}$.So $T(X)=\min\{X_1,\cdots,X_n\}$ is a minimal sufficient statistic.

c.

$$\frac{f(x|\theta)}{f(y|\theta)}=\frac{e^{-\sum_i(x_i-\theta)}}{\prod_{i=1}^{n}\left(1+e^{-(x_i-\theta)}\right)^2}\cdot\frac{\prod_{i=1}^{n}\left(1+e^{-(y_i-\theta)}\right)^2}{e^{-\sum_i(y_i-\theta)}}$$

$$=e^{-\sum_i(y_i-x_i)}\left(\frac{\prod_{i=1}^{n}\left(1+e^{-(y_i-\theta)}\right)}{\prod_{i=1}^{n}\left(1+e^{-(x_i-\theta)}\right)}\right)^2.$$

This is constant as a function of θ if and only if x and y have the same order statistics.

Therefore, the order statistics are minimal sufficient for θ.

d.This is difficult problem.The order statistics are a minimal sufficient statistic.

e.Fix sample points x and y.Define $A(\theta)=\{i:x_i\le\theta\}$, $B(\theta)=\{i:y_i\le\theta\}$, $a(\theta)=$ the number of elements in $A(\theta)$ and $b(\theta)=$ the number of elements in $B(\theta)$.Then the function $f(x|\theta)/f(y|\theta)$ depends on θ only through the function

$$\begin{aligned}
&\sum_{i=1}^{n}|x_i-\theta|-\sum_{i=1}^{n}|y_i-\theta| \\
&=\sum_{i\in A(\theta)}(\theta-x_i)+\sum_{i\in A(\theta)^c}(x_i-\theta)-\sum_{i\in B(\theta)}(\theta-y_i)-\sum_{i\in B(\theta)^c}(y_i-\theta) \\
&=\left(a(\theta)-\left[n-a(\theta)\right]-b(\theta)+\left[n-b(\theta)\right]\right)\theta \\
&+\left(-\sum_{i\in A(\theta)}x_i+\sum_{i\in A(\theta)^c}x_i+\sum_{i\in B(\theta)}y_i-\sum_{i\in B(\theta)^c}y_i\right) \\
&=2\left(a(\theta)-b(\theta)\right)\theta+\left(-\sum_{i\in A(\theta)}x_i+\sum_{i\in A(\theta)^c}x_i+\sum_{i\in B(\theta)}y_i-\sum_{i\in B(\theta)^c}y_i\right).
\end{aligned}$$

Consider an interbal of θs that does not contain any $x_i s$ or $y_i s$ .The second term is contant on such an interbal. The first term will be constant, on the interval if and only if $a(\theta)=b(\theta)$.This will be true for all such interbals if and only if the order statistics for x are the same as the order statistics for y. Therefore, the order statistics are a minimal sufficient statistic.

中文注释：设 $X_1,\cdots,X_n$ 是下面各个分布的样本.试找到一个 θ 的最小充分统计量.

(a) $f(x|\theta)=\dfrac{1}{\sqrt{2\pi}}e^{-(x-\theta)^2/2},-\infty<x<\infty,-\infty<\theta<\infty$

(b) $f(x|\theta)=e^{-(x-\theta)},\theta<x<\infty,-\infty<\theta<\infty$

(c) $f(x|\theta)=\dfrac{e^{-(x-\theta)}}{\left(1+e^{-(x-\theta)}\right)^2},-\infty<x<\infty,-\infty<\theta<\infty$

(d) $f(x|\theta)=\dfrac{1}{\pi\left[1+(x-\theta)^2\right]},-\infty<x<\infty,-\infty<\theta<\infty$

(e) $f(x|\theta)=\dfrac{1}{2}e^{-|x-\theta|},-\infty<x<\infty,-\infty<\theta<\infty$

解：1.

$$\frac{f(x|\theta)}{f(y|\theta)}=\frac{(2\pi)^{-n/2}e^{-\sum_i(x_i-\theta)^2/2}}{(2\pi)^{-n/2}e^{-\sum_i(y_i-\theta)^2/2}}=\exp\left\{-\frac{1}{2}\left[\left(\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}y_i^2\right)+2\theta n(\bar{y}-\bar{x})\right]\right\}.$$

这是一个 θ 的函数，当且仅当 $\bar{y}=\bar{x}$；所以 $\bar{X}$ 是 θ 的一个最小充分统计量.

2.注意：因为 $X\sim\exp(\theta)$，范围随参数改变.现在，

$$\frac{f(x|\theta)}{f(y|\theta)}=\frac{\prod_{i=1}^{n}\left(e^{-(x_i-\theta)}I_{(\theta,\infty)}(x_i)\right)}{\prod_{i=1}^{n}\left(e^{-(y_i-\theta)}I_{(\theta,\infty)}(y_i)\right)}$$

$$=\frac{e^{n\theta}e^{-\sum_i x_i}\prod_{i=1}^{n}I_{(\theta,\infty)}(x_i)}{e^{n\theta}e^{-\sum_i y_i}\prod_{i=1}^{n}I_{(\theta,\infty)}(y_i)}=\frac{e^{-\sum_i x_i}I_{(\theta,\infty)}(\min x_i)}{e^{-\sum_i y_i}I_{(\theta,\infty)}(\min y_i)}.$$

为了使它的比不依赖于θ，我们需要指示函数的比不依赖于θ.这就需要当且仅当$\min\{x_1,\cdots,x_n\}=\min\{y_1,\cdots,y_n\}$，所以$T(X)=\min\{X_1,\cdots,X_n\}$是最小充分统计量.

3.

$$\frac{f(x|\theta)}{f(y|\theta)}=\frac{e^{-\sum_i(x_i-\theta)}}{\prod_{i=1}^{n}\left(1+e^{-(x_i-\theta)}\right)^2}\cdot\frac{\prod_{i=1}^{n}\left(1+e^{-(y_i-\theta)}\right)^2}{e^{-\sum_i(y_i-\theta)}}$$

$$=e^{-\sum_i(y_i-x_i)}\left(\frac{\prod_{i=1}^{n}\left(1+e^{-(y_i-\theta)}\right)}{\prod_{i=1}^{n}\left(1+e^{-(x_i-\theta)}\right)}\right)^2.$$

这是一个θ的函数，当且仅当x和y有相同的次序统计量.

因此，这个次序统计量是θ的最小充分统计量.

4.这是一个困难的问题.这个次序统计量是一个最小充分统计量.

5.确定x和y的样本点，定义$A(\theta)=\{i:x_i\le\theta\}$, $B(\theta)=\{i:y_i\le\theta\}$, $a(\theta)$是$A(\theta)$的基本量，$b(\theta)$是$B(\theta)$的基本量.于是，函数$f(x|\theta)/f(y|\theta)$仅仅通过函数

$$\begin{aligned}&\sum_{i=1}^{n}|x_i-\theta|-\sum_{i=1}^{n}|y_i-\theta|\\&=\sum_{i\in A(\theta)}(\theta-x_i)+\sum_{i\in A(\theta)^c}(x_i-\theta)-\sum_{i\in B(\theta)}(\theta-y_i)-\sum_{i\in B(\theta)^c}(y_i-\theta)\\&=\left(a(\theta)-[n-a(\theta)]-b(\theta)+[n-b(\theta)]\right)\theta\\&+\left(-\sum_{i\in A(\theta)}x_i+\sum_{i\in A(\theta)^c}x_i+\sum_{i\in B(\theta)}y_i-\sum_{i\in B(\theta)^c}y_i\right)\\&=2\left(a(\theta)-b(\theta)\right)\theta+\left(-\sum_{i\in A(\theta)}x_i+\sum_{i\in A(\theta)^c}x_i+\sum_{i\in B(\theta)}y_i-\sum_{i\in B(\theta)^c}y_i\right).\end{aligned}$$

依赖于θ.

考虑到一个θs的 interbal 不包含任何$x_i s$或$y_i s$.那么第二关系是建立在这么一个

interbal 上.第一关系将会不变，在这个 interbal 上，当且仅当 $a(\theta)=b(\theta)$.这将对于所有此类 interbal 都是正确的当且仅当 x 的次序统计量和 y 的次序统计量是相同的.因此，这个次序统计量是最小充分统计量.

注：**E02 Statistical Inference(2ed) by George Casella**

(6).

.15 Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\theta, a\theta^2)$, where a is a known constant and $\theta > 0$.

(a) Show that the parameter space does not contain a two-dimensional open set.

(b) Show that the statistic $T = (\bar{X}, S^2)$ is a sufficient statistic for θ, but the family of distributions is not complete.

Let $X_1,\cdots,X_n$ be iid $n(\theta,a\theta^2)$,where a is a known constant and $\theta>0$.

(a) Show that the parameter space does not contain a two-dimensional open set.

(b) Show that the statistic $T=(\bar{X},S^2)$ is a sufficient statistic for θ,but the family of distributions is not complete.

解：a. The parameter space consists only of the points (θ,ν) on the graph of the function $\nu=a\theta^2$ .This quadratic graph is a line and does not contain a two-dimensional open set.

b. $E(S^2)=a\theta^2$ and $E(\bar{X}^2)=Var\bar{X}+(E\bar{X})^2=a\theta^2/n+\theta^2=(a+n)\theta^2/n$. Therefore,

$$E\left(\frac{n}{a+n}\bar{X}^2-\frac{S^2}{a}\right)=\left(\frac{n}{a+n}\right)\left(\frac{n}{a+n}\theta^2\right)-\frac{1}{a}a\theta^2=0,\ \text{for all}\ \theta\,.$$

Thus $g(\bar{X},S^2)=\dfrac{n}{a+n}\bar{X}^2-\dfrac{S^2}{a}$ has zero expectation so $(\bar{X},S^2)$ not complete.

中文注释：$X_1,\cdots,X_n$ 是总体 $n(\theta,a\theta^2)$ 的样本，在 a 处连续且 $\theta>0$.

（1） 证明参数空间不包含一个二维 open set

（2） 证明统计量 $T=(\bar{X},S^2)$ 是 θ 的一个充分统计量，但是分布族不完备.

解：1.参数空间由函数 $\nu=a\theta^2$ 曲线上的点 (θ,ν) 构成.这个二维曲线图是一条线并且不包含一个 two-dimensional open set.

2.由题得知，$E(S^2)=a\theta^2$ 且 $E(\bar{X}^2)=Var\bar{X}+(E\bar{X})^2=a\theta^2/n+\theta^2=(a+n)\theta^2/n$，因

此，$E\left(\frac{n}{a+n}\bar{X}^2-\frac{S^2}{a}\right)=\left(\frac{n}{a+n}\right)\left(\frac{n}{a+n}\theta^2\right)-\frac{1}{a}a\theta^2=0,\ for\ \text{all}\,\theta$.

所以，$g\left(\bar{X},S^2\right)=\frac{n}{a+n}\bar{X}^2-\frac{S^2}{a}$ 没有期望，所以 $\left(\bar{X},S^2\right)$ 不是完备统计量.

注：**E02 Statistical Inference(2ed) by George Casella**

(7).

6.21 Let X be one observation from the pdf

$$f(x|\theta)=\left(\frac{\theta}{2}\right)^{|x|}(1-\theta)^{1-|x|},\quad x=-1,0,1,\quad 0\le\theta\le 1.$$

(a) Is X a complete sufficient statistic?
(b) Is $|X|$ a complete sufficient statistic?
(c) Does $f(x|\theta)$ belong to the exponential class?

Let X be one obserbation from the pdf

$$f\left(x|\theta\right)=\left(\frac{\theta}{2}\right)^{|x|}\left(1-\theta\right)^{1-|x|},x=-1,0,1,\quad 0\le\theta\le 1.$$

(a)Is X a complete sufficient statistic?

(b)Is $|X|$ a complete sufficient statistic?

(c)Does $f\left(x|\theta\right)$ belong to the exponential class?

解：a. X is sufficient because it is the data. To check completeness, calulate

$$Eg\left(X\right)=\frac{\theta}{2}g\left(-1\right)+\left(1-\theta\right)g\left(0\right)+\frac{\theta}{2}g\left(1\right).$$

If $g\left(-1\right)=g\left(1\right)$ and $g\left(0\right)=0$,then $Eg\left(X\right)=0$ for all θ ,but $g\left(x\right)$ need not be identically 0 .So the family is not complete.

b.$|X|$ is sufficient by Theorem 6.2.6, because $f\left(x|\theta\right)$ depends on x only through the value of $|x|$.The distribution of $|X|$ is Bernoulli, bucause $P\left(|X|=0\right)=1-\theta$ and $P\left(|X|=1\right)=\theta$. By Example 6.2.22,a binomial family (Bernoulli is special case) is complete.

c.Yes, $f\left(x|\theta\right)=\left(1-\theta\right)\left[\theta/2\left(1-\theta\right)\right]^{|x|}=\left(1-\theta\right)e^{|x|\log\left[\theta/2(1-\theta)\right]}$,the form of an exponential family.

中文注释：随机变量 X 的密度函数为 $f\left(x|\theta\right)=\left(\frac{\theta}{2}\right)^{|x|}\left(1-\theta\right)^{1-|x|},x=-1,0,1,\quad 0\le\theta\le 1.$

（1） X 是否充分完备统计量？

（2）$|X|$是否充分完备统计量？

（3） $f(x|\theta)$是否是指数族？

解：1.已知 X 是充分的，所以只需检验完备性.由计算得，

$$Eg(X)=\frac{\theta}{2}g(-1)+(1-\theta)g(0)+\frac{\theta}{2}g(1).$$

如果$g(-1)=g(1)$且$g(0)=0$，于是对于任何θ，$Eg(X)=0$，但是$g(x)$不能被证明是 0.所以这个分布族不完备.

2.由定理 6.2.6 知$|X|$是充分统计量，因为仅仅通过$|x|$的值可知$f(x|\theta)$依赖于 x . $|X|$的分布式伯努利分布，因为$P(|X|=0)=1-\theta$以及$P(|X|=1)=\theta$.由例 6.2.22，一个二项分布族（伯努利分布是一个特例）是完备的.

3.是的，由指数分布族的构成得$f(x|\theta)=(1-\theta)\left(\theta/2(1-\theta)\right)^{|x|}=(1-\theta)e^{|x|\log[\theta/2(1-\theta)]}$.

E02 Statistical Inference(2ed) by George Casella

(8).

6.23 Let $X_1,\ldots,X_n$ be a random sample from a uniform distribution on the interval $(\theta,2\theta)$, $\theta>0$. Find a minimal sufficient statistic for θ. Is the statistic complete?

Let $X_1,\cdots,X_n$ be a random sample from a uniform distribution on the interval $(\theta,2\theta)$, $\theta>0$ .Find a minimal sufficient statistic for θ .Is the statistic complete?

解：Use Theorem 6.2.13 The ratio

$$\frac{f(x|\theta)}{f(y|\theta)}=\frac{\theta^{-n}I_{(x(n)/2,x_{(1)})}(\theta)}{\theta^{-n}I_{(y(n)/2,y_{(1)})}(\theta)}$$

In constant (in fact ,one) if and only if $x_{(1)}=y_{(1)}$ and $x_{(n)}=y_{(n)}$.So $(X_{(1)},X_{(n)})$is a minimal sufficient statistic for θ.From Exercise6.10,we know that if a function of the sufficient statistics is ancillary ,then the sufficient statistic is not complete.The uniform$(\theta,2\theta)$family is a scale family, with standard pdf $f(z)\sim uniform(1,2)$.So if $Z_1,\cdots,Z_n$ is a random sample from a uniform $(1,2)$ population, then $X_1=\theta Z_1,\cdots,X_n=\theta Z_n$ is a random sample from a uniform $(\theta,2\theta)$ population, and $X_1=\theta Z_1$ and $X_n=\theta Z_n$. So $X_{(1)}/X_{(n)}=Z_{(1)}/Z_{(n)}$,a statistic whose distribution does not depend on θ.Thus as in Exercise 6.10 , $\left(X_{(1)},X_{(n)}\right)$ is not complete.

中文注释：随机变量$X_1,\cdots,X_n$是$(\theta,2\theta)$,$\theta>0$上的均匀分布的一个样本.试找到一个θ的最小充分统计量.这个统计量是否完备？

解：运用定理 6.2.13，得

$$\frac{f(x|\theta)}{f(y|\theta)}=\frac{\theta^{-n}I_{\left(x(n)/2,x_{(1)}\right)}(\theta)}{\theta^{-n}I_{\left(y(n)/2,y_{(1)}\right)}(\theta)}$$

是连续的，当且仅当$x_{(1)}=y_{(1)}$ 且 $x_{(n)}=y_{(n)}$.所以$\left(X_{(1)},X_{(n)}\right)$是$\theta$的一个最小充分统计量.从习题 6.10，我们得知，如果一个函数的充分统计量是辅助的，则这个充分统计量不完备.均匀分布族$(\theta,2\theta)$是一个范围的分布族，随着$f(z)$的密度函数函数服从$U(1,2)$.所以，如果$Z_1,\cdots,Z_n$是一个来自$U(1,2)$随机抽样样本，则$X_1=\theta Z_1,\cdots,X_n=\theta Z_n$是来自$U(\theta,2\theta)$的一个随机抽样样本，且$X_1=\theta Z_1$，$X_n=\theta Z_n$.所以，$X_{(1)}/X_{(n)}=Z_{(1)}/Z_{(n)}$是一个分布不依赖于$\theta$的统计量.正如习题 6.10 中，$\left(X_{(1)},X_{(n)}\right)$不完备.

注：E02 Statistical Inference(2ed) by George Casella

(9).

6.40 Let $X_1,\ldots,X_n$ be iid observations from a location–scale family. Let $T_1(X_1,\ldots,X_n)$ and $T_2(X_1,\ldots,X_n)$ be two statistics that both satisfy

$$T_i(ax_1+b,\ldots,ax_n+b)=aT_i(x_1,\ldots,x_n)$$

for all values of $x_1,\ldots,x_n$ and b and for any $a>0$.

(a) Show that T_1/T_2 is an ancillary statistic.

(b) Let R be the sample range and S be the sample standard deviation. Verify that R and S satisfy the above condition so that R/S is an ancillary statistic.

Let $X_1,\cdots,X_n$ be iid obserbations from a location-scale family. Let $T_1(X_1,\cdots,X_n)$ and $T_2(X_1,\cdots,X_n)$ be two statistics that both satisfy

$$T_i(ax_1+b,\cdots,ax_n+b)=aT_i(x_1,\cdots,x_n)$$

For all values of $x_1,\cdots,x_n$ and b and for any $a>0$.

(a) Show that T_1/T_2 is an ancillary statistic.

(b) Let R be the sample rangs and S be the sample standard debiation.Verify that R and S satisfy the above condition so that R/S is an ancillary statistic.

解：a.Because $X_1,\cdots,X_n$ is from a location scale family ,by Theorem 3.5.6,we can write $X_i=\sigma Z_i+\mu$, where $Z_1,\cdots,Z_n$ is a random sample from the standard pdf $f(z)$.

Then, $\dfrac{T_1(X_1,\cdots,X_n)}{T_2(X_1,\cdots,X_n)}=\dfrac{T_1(\sigma Z_1+\mu,\cdots,\sigma Z_n+\mu)}{T_2(\sigma Z_1+\mu,\cdots,\sigma Z_n+\mu)}=\dfrac{\sigma T_1(Z_1,\cdots,Z_n)}{\sigma T_2(Z_1,\cdots,Z_n)}=\dfrac{T_1(Z_1,\cdots,Z_n)}{T_2(Z_1,\cdots,Z_n)}$.

Because T_1/T_2 is a function of only $Z_1,\cdots,Z_n$, the distribution of T_1/T_2 does not depand on μ or σ; that is , T_1/T_2 is an ancillary atatistic.

b. $R(x_1,\cdots,x_n)=x_{(n)}-x_{(1)}$. Because $a>0$, $\max\{ax_1+b,\cdots,ax_n+b\}=ax_{(n)}+b$ and $\min\{ax_1+b,\cdots,ax_n+b\}=ax_{(1)}+b$.Thus,

$$R(ax_1+b,\cdots,ax_n+b)=(ax_{(n)}+b)-(ax_{(1)}+b)=a(x_{(n)}-x_{(1)})=aR(x_1,\cdots,x_n).$$

For the sample variance we have

$$S^2(ax_1+b,\cdots,ax_n+b)=\frac{1}{n-1}\sum\left((ax_i+b)-(a\bar{x}+b)\right)^2$$
$$=a^2\frac{1}{n-1}\sum(x_i-\bar{x})^2=a^2S^2(x_1,\cdots,x_n).$$

Thus, $S(ax_1+b,\cdots,ax_n+b)=aS(x_1,\cdots,x_n)$ .Therefore, R and S both satisfy the above condition,and R/S is ancillary by a).

中文注释：随机变量$X_1,\cdots,X_n$是 iid obserbations from a location-scale family.随机变量$T_1(X_1,\cdots,X_n)$和$T_2(X_1,\cdots,X_n)$是满足$T_i(ax_1+b,\cdots,ax_n+b)=aT_i(x_1,\cdots,x_n)$的两个统计量，对于所有$x_1,\cdots,x_n$和$b$的值，并且任意$a>0$.

（1） 证明T_1/T_2是一个辅助统计量.

（2） 随机变量R是样本范围，S是样本水平 debiation.证明R和S满足以上条件让R/S是一个辅助统计量.

解：1.因为 $X_1,\cdots,X_n$ 来自一个 location scale family，由定理 3.5.6，我们能写出 $X_i=\sigma Z_i+\mu$，这里 $Z_1,\cdots,Z_n$ 是一个来自密度函数 $f(z)$ 随机抽样样本.于是，

$$\frac{T_1(X_1,\cdots,X_n)}{T_2(X_1,\cdots,X_n)}=\frac{T_1(\sigma Z_1+\mu,\cdots,\sigma Z_n+\mu)}{T_2(\sigma Z_1+\mu,\cdots,\sigma Z_n+\mu)}=\frac{\sigma T_1(Z_1,\cdots,Z_n)}{\sigma T_2(Z_1,\cdots,Z_n)}=\frac{T_1(Z_1,\cdots,Z_n)}{T_2(Z_1,\cdots,Z_n)}$$

因为 T_1/T_2 只是 $Z_1,\cdots,Z_n$ 的函数，它的分布不依赖于 μ 或者 σ，所以 T_1/T_2 是一个辅助统计量.

2. $R(x_1,\cdots,x_n)=x_{(n)}-x_{(1)}$，因为 $a>0$，$\max\{ax_1+b,\cdots,ax_n+b\}=ax_{(n)}+b$ 且 $\min\{ax_1+b,\cdots,ax_n+b\}=ax_{(1)}+b$.所以，

$$R(ax_1+b,\cdots,ax_n+b)=(ax_{(n)}+b)-(ax_{(1)}+b)=a(x_{(n)}-x_{(1)})=aR(x_1,\cdots,x_n)$$

由于样本差异，我们有

$$S^2(ax_1+b,\cdots,ax_n+b)=\frac{1}{n-1}\sum\left((ax_i+b)-(a\bar{x}+b)\right)^2$$
$$=a^2\frac{1}{n-1}\sum(x_i-\bar{x})^2=a^2S^2(x_1,\cdots,x_n).$$

所以，$S(ax_1+b,\cdots,ax_n+b)=aS(x_1,\cdots,x_n)$.因此，$R$ 和 S 都满足以上条件，且 R/S 是第一小题的辅助.

注：E02 Statistical Inference(2ed) by George Casella

四、【经典和趣味问题】

1.把七个不可分辨的球放到 n 各不同盒子里去的所有可能的方法的总数.

解：记 $X_i^{'}$ 为第个盒子中球的个数，$i=1,2,\cdots,n$，则有 $\sum_{i=1}^{n}X_i^{'}=t$，所以，上述概率实际上

是条件概率，应记为

$$P(X_1^{'}=x_1^{'},\cdots,X_n^{'}=x_n^{'}\mid\sum_{i=1}^{n}X_i^{'}=t)$$

$$=\frac{P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n,T=t)}{P(X=t)}=\frac{P(X_1=x_1)\cdots P(X_{n-1}=x_{n-1})P(X_n=t-\sum_{i=1}^{n-1}X_i)}{P(X=t)}$$

$$=\frac{\theta(1-\theta)^{x_i}\theta(1-\theta)^{x_{n-1}}\theta(1-\theta)^{t-\sum_{i=1}^{n-1}X_i)}}{\binom{t+n-1}{n-1}\theta^n(1-\theta)^t}=\binom{t+n-1}{n-1}^{-1}$$

2.充分而不完备的统计量：

设一事件A在一次试验中出现的概率为θ，$0\le\theta\le1$. X为在一次试验中事件A出现的次数，则X的分布为$P\left(X=x\right)=\theta^x\left(1-\theta\right)^{1-x},x=0,1$，若做$n$次独立的重复试验.令

$$X_i=\begin{cases}1, & \text{第}i\text{次试验中事件}A\text{出现，}\\ 0, & \text{否则。}\end{cases}$$

$i=1,2,\cdots,n.\left(X_1,\cdots,X_n\right)$是从母体$X$中抽取的一个子样.取$n_1,1\le n_1\le n-1$，定义统计量$t\left(X_1,\cdots,X_n\right)=\left(t_1\left(X_1,\cdots,X_n\right),t_2\left(X_1,\cdots,X_n\right)\right)$，其中$t_1\left(X_1,\cdots,X_n\right)=\sum_{i=1}^{n_1}X_i$是前$n_1$次试验中事件$A$出现的次数，容易证明$t\left(X_1,\cdots,X_n\right)$是充分统计量.但是，它并不是完备统计量.

（1）$t\left(X_1,\cdots,X_n\right)=\left(t_1\left(X_1,\cdots,X_n\right),t_2\left(X_1,\cdots,X_n\right)\right)$是充分统计量.因为

$$\begin{aligned}&P\left(X_1=x_1,\cdots,X_n=x_n;\theta\right)=\theta^{\sum_{i=1}^n x_i}\left(1-\theta\right)^{n-\sum_{i=1}^n x_i}\\&=g\left[t\left(x_1,\cdots,x_n\right);\theta\right]h\left(x_1,\cdots,x_n\right)\end{aligned}$$

其中$h\left(x_1,\cdots,x_n\right)=1$，由因子分解定理知，统计量$t\left(x_1,\cdots,x_n\right)$是充分统计量.

（2）$t\left(x_1,\cdots,x_n\right)$不是完备统计量.

因为若令$g\left(t\right)=\frac{t_1}{n_1}-\frac{t_2}{n-n_1}$，则

$$\begin{aligned}&E_\theta\left\{g\left(t\left(x_1,\cdots,x_n\right)\right)\right\}=E_\theta\left(\frac{t_1\left(x_1,\cdots,x_n\right)}{n_1}\right)-E_\theta\left(\frac{t_2\left(x_1,\cdots,x_n\right)}{n-n_1}\right)\\&=\frac{n_1\theta}{n_1}-\frac{\left(n-n_1\right)\theta}{n-n_1}=0\end{aligned}$$

对一切$\theta\in\left[0,1\right]$，但是，

$$\begin{aligned}&P_\theta^T\left(g\left(t\right)\ne0\right)>P_\theta^T\left(t_1=n_1,t_2=0\right)\\&=P_\theta\left(X_1=1,\cdots,X_{n_1}=1,X_{n_1+1}=0,\cdots,X_n=0\right)\\&=\theta^{n_1}\left(1-\theta\right)^{n-n_1}>0,\left(0<\theta<1\right)\end{aligned}$$

所以，$P_\theta^T\left(g(t)=0\right)=1-P_\theta^T\left(g(t)\neq 0\right)<1$，故$t\left(X_1,\cdots,X_n\right)$不是完备统计量.

注：概率论与数理统计中的反例 陈俊雅 王秀花 天津科技 P234

3.完备而不充分的统计量：

设 $X_1,X_2,\cdots,X_n$ 是从母体为 $0-1$ 分布 $b(1,\theta),0<\theta<1$ 中抽取的一个子样，令

$T_n=\sum_{i=1}^{n}X_i$，定义如下的统计量.

$$S\left(X_1,\cdots,X_n\right)=\begin{cases}1, & \text{当}T_n=0\text{或}1,\\ 0, & \text{当}T_n=i\geq 2.\end{cases}$$

则$S\left(X_1,\cdots,X_n\right)$是一个完备而不充分的统计量.

（1）$S\left(X_1,\cdots,X_n\right)$是完备统计量，因为$S\left(X_1,\cdots,X_n\right)$具有$0-1$分布.

$$\begin{aligned}&P_\theta\left(S=1\right)=P_\theta\left(T_n=0\right)+P_\theta\left(T_n=1\right)\\&=\left(1-\theta\right)^n+n\theta\left(1-\theta\right)^{n-1}\triangleq p_n\\&P_\theta\left(S=0\right)=1-P_\theta\left(s=1\right)=1-p_n\end{aligned}$$

若对一切$0\leq p\leq t$，有$E_\theta g\left(s\right)=0$.即

$$g\left(0\right)\left(1-p\right)+g\left(1\right)p=0$$

必有$g\left(0\right)=0,g\left(1\right)=0$，所以

$$P_\theta^T\left(g\left(s\right)=0\right)=P_\theta^T\left(s=0\right)+P_\theta^T\left(s=1\right)=1.$$

（2）$S\left(X_1,\cdots,X_n\right)$不是充分统计量.

因为条件概率

$$\begin{aligned}&P_\theta\left(X_1=0,\cdots,X_n=0\middle|S=1\right)\\&=\frac{P_\theta\left(X_1=0,\cdots,X_n=0,S=1\right)}{P_\theta\left(S=1\right)}\\&=\frac{P_\theta\left(X_1=0,\cdots,X_n=0\right)}{P_\theta\left(S=1\right)}\\&=\frac{\left(1-\theta\right)^n}{\left(1-\theta\right)^n+n\theta\left(1-\theta\right)^{n-1}}\\&=\frac{1-\theta}{1+\theta\left(n-1\right)}\end{aligned}$$

依赖于未知参数θ，所以$S\left(X_1,\cdots,X_n\right)$不是充分统计量.

注：这两道题目主要是让我们了解一下，类似的统计量，如果服从的分布不同，那么其性质也大不相同，一个充分而不完备，一个完备而不充分，所以，分清楚充分性和完备性的定义很重要.

概率论与数理统计中的反例 陈俊雅 王秀花 天津科技 P237

五、【参考书目】

［1］参数统计教程，韦博成 ，《高等教育出版社》，2006.

［2］统计推断导引， 范金城 吴可法，《科学出版社》，2001.

［3］概率统计教学参考书，杨维权 邓集贤，《高等教育出版社》，1996.

［4］(美)卡塞拉(Casella G.) (美)贝耶(Berger R.L.)，统计推断（英文版/原书第 2 版），机械工业出版社，2005.

［5］概率论与数理统计教程 习题与解答，茆诗松 程依明 濮晓龙，高等教育出版社.

［6］概率论与数理统计中的反例 陈俊雅 王秀花 天津科技

通过对充分完备统计量的整理，我们从中可以知道，充分完备统计量包含了许多定理，有很多类型的题目都是需要这些定理来融会贯通的，其中证明题里，有三种方法可以证明一个统计量是不是充分的.一、从定义上证明.二、因子分解定理.三、指数族分布，其中指数族分布是极其重要的，也是证明充分完备样的重要方法.除了这些方法，我们还需要熟练地运用很多分布的密度函数及分布函数.通过对资料的分析，我们可以通过证明充分统计量，从而得出该统计量又是完备统计量.例题中有几道题目中可以看出，对于充分完备统计量，其定义是非常重要的，只有在充分了解其定义的内容，再做题目，才会融会贯通，并且这类题目还涉及到许多其它知识点.例如：点估计、最大似然估计、相合性、无偏性、有效性、均方误差等等，而不仅仅需要这章的知识，这些知识点的前提是要把充分完备统计量求对，而且后面所学的 UMVUE 也正是用到了充分性以及完备性,所以务必注意充分完备统计量的重要性.这章的理论方面是比较多的，所以要学好这章的内容必须把他们都掌握，相较于其他章节是比较困难的，所以更需要努力地学习。