

关于统计量 $T = T(X_1, X_2, \dots, X_n)$ 必须满足什么条件才能具有上述意义的充分性的讨论，要说明的是判别充分统计量的方法为因子分解定理。这个定理是奈曼和哈尔姆斯在 20 世纪 40 年代提出并严格证明的。

设样本 $\tilde{X}$ 的联合密度函数或联合分布列为 $p(\tilde{x}; \theta)$ ，其中 θ 为未知参数，则 $T = T(\tilde{X})$ 为关于 θ 的充分统计量当且仅当

$$p(\tilde{x}; \theta) = g(T(\tilde{x}); \theta) h(\tilde{x})$$

直观上看，就是联合密度函数或分布列能够分解为 2 项，其中一个只与样本有关，与参数无关；另外一个与参数和样本都有关系，但是与样本的关系完全可以用统计量 T 来代替。

证明：严格证明需要用到测度论，这里只证明离散情形，这时 $p(\tilde{x}; \theta) = P_\theta(\tilde{X} = \tilde{x})$ ，为了简便利用样本与统计量值得相容性，令 $A(t) = \{\tilde{x} : T(\tilde{x}) = t\}$ 。

充分性证明：如果样本与统计量值不相容，即 $\tilde{x} \notin A(t)$ ，那么条件概率为 0；如果 $\tilde{x} \in A(t)$ ，则：

$$\begin{aligned} P_\theta(\tilde{X} = \tilde{x} | T = t) &= \frac{P_\theta(\tilde{X} = \tilde{x}, T = t)}{P_\theta(T = t)} = \frac{P_\theta(\tilde{X} = \tilde{x})}{P_\theta(T = t)} \\ &= \frac{p(\tilde{x}; \theta)}{\sum_{\tilde{y} \in A(t)} p(\tilde{y}; \theta)} = \frac{g(t; \theta) h(\tilde{x})}{\sum_{\tilde{y} \in A(t)} g(t; \theta) h(\tilde{y})} \\ &= \frac{h(\tilde{x})}{\sum_{\tilde{y} \in A(t)} h(\tilde{y})} \end{aligned}$$

其中第二个等号中，分子利用相容性，可以把关于 T 的去掉，第三个等号中分母利用了全概率公式，且当 $\tilde{y} \notin A(t)$ 时联合概率为 0。第四个等号套用因子分解定理中那个形式，进而上下可以同时约去一个因式。最后的结果说明条件分布与 θ 无关，所以 T 是充分统计量。

必要性证明：如果 $T(\tilde{x})$ 是充分统计量，根据定义应该有条件密度与参数无关。令 $t = T(\tilde{x})$ ，则：

$$\begin{aligned} p(\tilde{x}; \theta) &= P_\theta(\tilde{X} = \tilde{x}) \\ &= P_\theta(\tilde{X} = \tilde{x}, T = t) \\ &= P_\theta(\tilde{X} = \tilde{x} | T = t) P_\theta(T = t) \\ &= h(\tilde{x}) g(t; \theta) \\ &= h(\tilde{x}) g(T(\tilde{x}); \theta) \end{aligned}$$

其中第二个等号反向利用相容性，可以添加关于 T 的那项；最后的结果和因子分解定理中的

形式一样，因此得到了必要性。