Exercise H1.1

a) Suppose the data X in a statistical model take values in a countable set $\mathcal{X}$ (i.e.X has a discrete law). In class it was claimed that the data itself are a sufficient statistic (i.e. $T(X) = X$ is sufficient). Write down the argument (it can be a very short paragraph).

b). Suppose T is a sufficient statistic with values in a set $\mathcal{T}$ and $S : \mathcal{T} \mapsto \mathcal{S}$ is a mapping with values in a set $\mathcal{S}$ which is one-to-one (i.e there exists an inverse mapping S^{-1}such that $S^{-1}(S(t)) = t$ for all $t \in \mathcal{T}$) Show that the statistic $S(T(X))$.is sufficient.

c) In example 2.1 handout it was claimed that the statistic $T(X) = (X_1, \bar{X}_n)$ is sufficient in Model I. Prove this claim.

Exercise H1.2. Let $X_1, \dots, X_n$ be independent and identically distributed with Poisson law $\text{Po}(\lambda)$, where $\lambda > 0$ is unknown. Show that the sample mean $\bar{X}_n$ is again a sufficient statistic (**Comment**: the sample mean is a sufficient statistic not only for i.i.d. Bernoulli data, but in a number of statistical models).

Exercise H1.3. Let $X_1, \dots, X_n$ be independent and identically distributed such that X_1 has the uniform law on the set $\{1, \dots, r\}$ for some integer $r > 1$ (i.e. $P(X_1 = k) = 1/r$, $k = 1, \dots, r$). In the statistical model where $r > 1$ is unknown, show that $T(X) = \max_{i=1,\dots,n} X_i$ is a sufficient statistic. **Hint:** as an intermediate step, show that $P(T(X) \le k) = (k/r)^n$ for all $k = 1, \dots, r$. (**Comment**: in this model, the unknown parameter is the largest value that the data can possibly take, i.e. r. It turns out that the largest value which they take in the sample is a sufficient statistic.).

Exercise H1.4. Let $X_1, \dots, X_n$ be independent and identically distributed such that X_1 has the geometric law $\text{Geom}(p)$, i.e.

$$P(X_1 = k) = (1 - p)^{k-1} p, \ k = 1, 2, \dots$$

for some $p \in (0, 1)$. In the statistical model where p is unknown, show that the sample mean $\bar{X}_n$ is a sufficient statistic. **Hint:** it can be used that $n\bar{X}_n = \sum_{i=1}^n X_i$ has the negative binomial distribution with parameters n and p, i.e.

$$P(n\bar{X}_n = k) = \binom{k-1}{k-n} (1 - p)^{k-n} p^n \text{ for } k \ge n$$

Homework Solusion #1

Exercise H1.1.

a) $P_\theta(X \in B | X = x) = \dfrac{P_\theta(X \in B, X = x)}{P_\theta(X = x)} = 1_B(x)$, independent of θ.

b) $P_\theta(X \in B | T(X) = t)$ is independent of θ for any event B and any value t of the set $\mathcal{T}$, since T is a sufficient statistic. This implies $P_\theta(X \in B | S(T(X)) = s) = P_\theta(X \in B | T(X) = S^{-1}(s))$ is independent of θ for any event B and any value s of the set $\mathcal{S}$. Because $S : \mathcal{T} \longrightarrow \mathcal{S}$ is one to one and onto, $S^{-1}(s) \in \mathcal{T}$. for any $s \in \mathcal{S}$.

c) We will show that $P_p(X_1 = x_1, X_2 = x_2, ..., X_n = x_n | X_1 = x_1^{'}, n\overline{X_n} = k)$ is independent of p for any $(x_1, ..., x_n) \in \{0,1\}^n$, $x_1^{'} \in \{0,1\}$, and $0 \leqslant k \leqslant n$. If $x_1 \neq x_1'$,or $k \neq n\overline{x_n}$, the conditional probability is 0. If $x_1 = x_1'$,and $k = n\overline{x_n}$, then

$$\begin{aligned} P_p\left(X_1 = x_1^{'}, n\overline{X_n} = k\right) &= P_p(X_1 = x_1, X_2 + ... + X_n = x_2 + ... + x_n) \\ &= P_p(X_1 = x_1)\ P_p(X_2 + ... + X_n = x_2 + ... + x_n) \\ &= (p^{x_1}\ p^{1-x_1})\ \ \tbinom{n-1}{x_2+...+x_n} p^{x_2+...+x_n}(1-p)^{(n-1)-(x_2+...+x_n)} \\ &= \tbinom{n-1}{x_2+...+x_n} p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, \end{aligned}$$

and,

$$\begin{aligned} P_p\left(X_1 = x_1, ..., X_n = x_n, X_1 = x_1, n\overline{X_n} = k\right) &= P_p(X_1 = x_1, ..., X_n = x_n) \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}. \end{aligned}$$

This implies $P_p\left(X_1 = x_1, ..., X_n = x_n | X_1 = x_1, n\overline{X_n} = n\overline{x_n}\right) = 1/\tbinom{n-1}{x_2+...+x_n}$ is independent of p.

So $\left(X_1, n\overline{X_n}\right)$ is a sufficient stattistic, i.e., $\left(X_1, \overline{X_n}\right)$ is sufficient from (b).

Exercise H1.2.

We need to prove that $\overline{X_n}$ is sufficcient, i.e., $n\overline{X_n}$ is sufficient from(b), i.e., $P_\lambda\left(X_1 = x_1, ..., X_n = x_n | n\overline{X_n} = k\right)$ is independent of λ for any $(X_1, ..., X_n) \in \{0,1,2,...\}^n$, and $k \geqslant 0$. If $k \neq \sum_{i=1}^n x_i$, the conditional probability is 0. If $k = \sum_{i=1}^n x$,we know that the distribution of $\sum_{i=1}^n x_i$ is Po$(n\lambda)$,

then

$$P_\lambda\left(n\overline{X_n} = n\overline{x_n}\right) = e^{-n\lambda}\frac{(n\lambda)^{\sum_{i=1}^n x_i}}{(\sum_{i=1}^n x_i)!},$$

and,

$$\begin{aligned} P_\lambda\left(X_1 = x_1, ..., X_n = x_n, n\overline{X_n} = k\right) &= P_p(X_1 = x_1, ..., X_n = x_n) \\ &= \prod_{i=1}^n \left(e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right) \\ &= e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}. \end{aligned}$$

This implies $P_\lambda\left(X_1 = x_1, ..., X_n = x_n | n\overline{X_n} = n\overline{x_n}\right) = (\sum_{i=1}^n x_i)!/(n^{\sum_{i=1}^n x_i}\prod_{i=1}^n (x_i!))$ is independent of λ. So $\overline{X_n}$ is sufficient.

Exercise H1.3.

We need to show that $P_r(X_1 = x_1, ..., X_n = x_n | T(x) = \max_{i=1,...n} X_i = k)$ is independent of r for any $x_1, ..., x_n$, and k. If $\max_{i=1,...,n} x_i \gneqq k$, the conditional probability is 0. If $\max_{i=1,...,n} x_i \leqslant k$, we have

$$\begin{aligned} P_r(T(x) = \max_{i=1,...,n} x_i = k) &= P_r(T(x) \leqslant k) - P_r(T(x) \leqslant k-1) \\ &= P_r(X_1 \leqslant k, ..., X_n \leqslant k) - P_r(X_1 \leqslant k-1, ..., X_n \leqslant k-1) \\ &= \prod_{i=1,...,n}^n P_r(X_i \leqslant k) - \prod_{i=1,...,n}^n P_r(X_i \leqslant k-1) \\ &= \left(\tfrac{k}{r}\right)^n - \left(\tfrac{k-1}{r}\right)^n, \end{aligned}$$

and,

$$\begin{aligned} P_r(X_1 = x_1, ..., X_n = x_n) &= \prod_{i=1,...,n}^n P_r(X_i = x_i) \\ &= \left(\tfrac{1}{r}\right)^n. \end{aligned}$$

This implies $P_r = (X_1 = x_1, ..., X_n = x_n | T(x) = \max_{i=1,...,n} x_i) = \dfrac{\left(\dfrac{1}{r}\right)^n}{\left(\dfrac{k}{r}\right)^n - \left(\dfrac{k-1}{r}\right)^n} = \dfrac{1}{k^n - (k-1)^n}$ is independent of r.

So $T(x)$ is sufficient.

Exercise H1.4.

We need to show $\overline{X_n}$ is sufficient, i.e., $n\overline{X_n}$ is sufficient, i.e., $P_p\left(X_1 = x_1, ..., X_n = x_n | n\overline{X_n} = k\right)$ is independent of p for any $(x_1, ..., x_n) \in \{1, 2, 3....\}^n$,and $k \geqslant n$. If $k \neq n\overline{x_n}$, then the conditinal probability is 0. If $k = n\overline{x_n}$, then

$$P_p\left(n\overline{X_n} = n\overline{x_n}\right) = \tbinom{n\overline{x_n}-1}{n\overline{x_n}-n}(1-p)^{\,n\overline{x_n}-n}p^n ,$$

and,

$$\begin{aligned} P_p\left(X_1 = x_1, ..., X_n = x_n, n\overline{X_n} = n\overline{x_n}\right) &= P_p(X_1 = x_1, ..., X_n = x_n) \\ &= \prod_{i=1}^n (1-p)^{x_i - 1} p \\ &= (1-p)^{n\overline{x_n}-n} p^n. \end{aligned}$$

This implies $P_p\left(X_1 = x_1, ..., X_n = x_n | n\overline{X_n} = n\overline{x_n}\right) = 1/\tbinom{n\overline{x_n}-1}{n\overline{x_n}-n}$ is independent of r.

So $\overline{X_n}$ is sufficient.