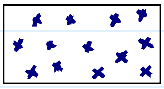
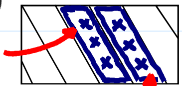
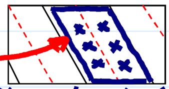
**Question 6.2** check graph in LNp.44 $X_1, \cdots, X_n$

There exist many sufficient statistics. A trivial example is the raw data. However, a large collection of numbers is not as *meaningful* as a few good summary statistics. How can we know whether the data has been reduced *as much as possible* and still keep relelvant information? → often, smaller dimension.

→ If MLE is sufficient, MLE is minimal sufficient (by Thm 6.11, LNp.48)

**Definition 6.14** (minimal sufficient statistic) know $S$ $\rightleftarrows$ know $T$, which contains more information?

A sufficient statistic $S$ for $\theta$ is called **minimal** if $S$ is a function of $T$ for any sufficient statistic $T$ of $\theta$. $T$ contains at least as much information as $S$.

$X_1, \ldots, X_n$ suff  $T = g(X_1, \ldots, X_n)$ suff $t_1$ $t_2$  $S = h(T)$ suff $\{t_1, t_2\} \xrightarrow{h} S$ 

dim(S) ≤ dim(T)

**Example 6.23** (minimal sufficient statistics for i.i.d. Uniform distribution $U(\theta, \theta+1)$)

Let $X_1, \ldots, X_n$ be i.i.d. from uniform distribution $U(\theta, \theta+1)$. Then, the joint pdf is ← statistical modeling → know the data is uniformly distributed in an interval of length one, but don't know where the interval is located.

$\boxed{\theta \le X_{(1)} < X_{(2)} < \cdots < X_{(n)} \le \theta+1 \Rightarrow X_{(n)} - 1 \le \theta \le X_{(1)}}$

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} I_{(\theta,\theta+1)}(x_i) = I_{(x_{(n)}-1, x_{(1)})}(\theta),$$

← indicator function

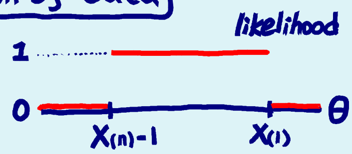
where $I_{(a,b)}(u) = 1$ if $a \le u \le b$ and 0 otherwise. Therefore, $T = (X_{(1)}, X_{(n)})$ is sufficient for $\theta$ by factorization theorem.



Note that function of data

$$x_{(1)} = \sup\{\theta : f(\mathbf{x}|\theta) > 0\}, \text{ and}$$
$$x_{(n)} = 1 + \inf\{\theta : f(\mathbf{x}|\theta) > 0\}.$$

dimension reduction $n \to 2 > 1$ ($\geq$ dimension of parameter)




For a sufficient statistics $T$, by factorization theorem, $f(\mathbf{x}|\theta) = g(t, \theta)h(\mathbf{x})$. Therefore, for $\mathbf{x}$ such that $h(\mathbf{x}) > 0$, $P(T)$ $P(X_1, \cdots, X_n|T)$

$$x_{(1)} = \sup\{\theta : g(t, \theta) > 0\} \text{ and } x_{(n)} = 1 + \inf\{\theta : g(t, \theta) > 0\}.$$

We conclude that $(X_{(1)}, X_{(n)})$ is minimal sufficient. → No 1-dimensional sufficient statistic

**Example 6.24** (cont. Ex. 6.23, ancillary statistics) ← check graph in LNp.44

- In Example 6.23, the pdf of $R = X_{(n)} - X_{(1)}$ is range

  (Ec) — $n(n-1)r^{n-2}(1-r)$, $\quad (\Rightarrow \text{Beta}(n-1, 2))$

  for $0 \le r \le 1$, which is **irrelevant** to $\theta$. (why?)

  Q: Does minimal sufficient statistics still contain useless information?

  – For any $\theta$, the appearance of the values of $R$ follows the same probabilistic pattern. → different θ: the same distribution of $R$ → likelihood of $R$ is a constant over θ (cf. $T = \sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ are i.i.d. from Bernoulli$(\theta)$)

  Known 規律 $n=100$, $T=10 \xleftrightarrow{\text{cf.}} T=80$ $(T \sim B(n,\theta))$

  R carries information about the range, which is assumed known in statistical modeling

  – The observed values of $R$ do not carry information about $\theta$

  – That is, there exists transformation of the minimum sufficient statistics $X_{(1)}$ and $X_{(n)}$ that contains no information about $\theta$.

- Statistics like $R$ are called **ancillary statistics**, which have distributions free of the parameters and seemingly contain *no* information about the parameters. cf. Definition of Suff. Stat. based on $X_1,\cdots,X_n|T$ cf. sufficient statistics

(Ec) other example of ancillary statistics?

LN, CH1~6, P.80 $X_1,\cdots,X_n$, i.i.d. $N(\theta,1)$, why reasonable? $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ is ancillary $\sim \frac{\chi^2_{n-1}}{n-1}$

$X_1,X_2$, i.i.d. $Gamma(1,\theta)$, why reasonable? $Z=\frac{X_1}{X_1+X_2}$ is ancillary $\sim Beta(1,1)$ LN, CH1~6, p.75

**Question 6.3**

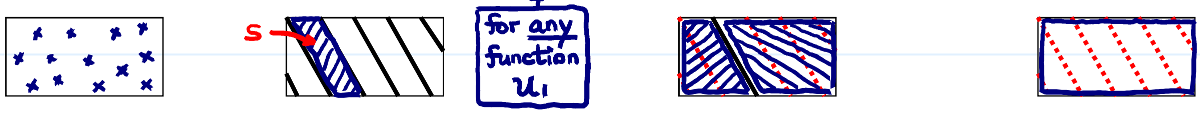
Note that minimal sufficient statistics may still contain ancillary information. What other property can guarantee sufficient statistics containing no ancillary information? 完備

$u^*(S)=u(S)-c$, Then, $E_\theta[u^*(S)]=0 \Leftrightarrow E_\theta[u(S)]=c$, $u^*(S)=0\Leftrightarrow u(S)=c$ irrelevant to $\theta$

**Definition 6.15** (completeness, TBp.310)

Let $f(s|\theta)$, $\theta\in\Omega$, be a family of pdfs or pmfs for a statistic $S=S(X_1,\ldots,X_n)$. The family of probability distributions is called **complete** if $E_\theta[u(S)]=0$ (or $c$: a constant) for all $\theta\in\Omega$, where $u$ is a function of $S$, implies $u(S)=0$ (or $c$) with probability 1 for all $\theta\in\Omega$. Equivalently, $S$ is called a **complete statistics**.

$(X_1,\ldots,X_n)$ $\quad S(X_1,\ldots,X_n)$ $\quad u_1(S)$: non-constant function $\quad u_2(S)$: constant function






- $u(S)=c$, $c$ : a constant, is a trivial ancillary statistic and $E_\theta[u(S)-c]=0$, for any $\theta$. $u^*(S)$ $u^*(S)=0 \Leftrightarrow u(S)=c$ distribution of $u(S)$ irrelevant to $\theta$. cdf 1 0 c

→: constant →: non-constant

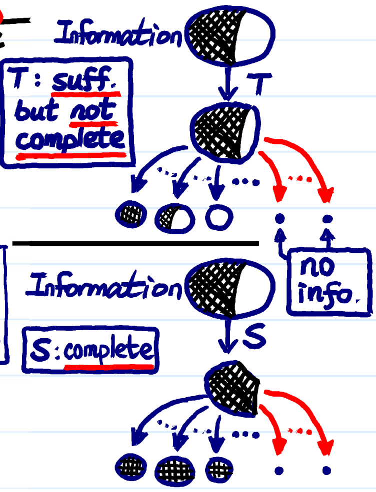
not B ⇒ not A: $S$ is complete $\Leftrightarrow$ $E_\theta[u(S)]$ is a constant for all $\theta$ implies that the trnasformation $u$ is a constant transformation.

A ⇒ B: $S$ is complete $\Leftrightarrow$ any transformations of $S$ (except the constant functions) contains *some* information about $\theta$.

- In Example 6.24, $E_\theta(R)=\frac{n-1}{n+1}$. That is, LNp.51 (Ec) If $u$ not constant, ⇒ $E_\theta[u(S)]$ depends on $\theta$ ⇒ dist. of $u(S)$ depends on $\theta$.

not complete $X_{(n)}-X_{(1)}-\frac{n-1}{n+1}=R-E_\theta(R)$

has mean zero for all $\theta$. ⇒ there is a *nonzero* function of $X_{(1)}$ and $X_{(n)}$ whose expectation is zero for all $\theta$.




**Example 6.25** (sufficient and complete statistics of i.i.d. Uniform distribution $U(0,\theta)$)

Let $X_1,\ldots,X_n$ be i.i.d. from Uniform distribution $U(0,\theta)$, $\theta>0$.

dim: $n\to 1$

- By factorization theorem, $X_{(n)}$, the largest order statistics, is sufficient. joint pdf $=\frac{1}{\theta^n}I_{(X_{(n)},\infty)}(\theta)$ Indicator function
- The pdf of $X_{(n)}$ is $\frac{nx^{n-1}}{\theta^n}I_{(0,\theta)}(x)$. LN, CH1~6, P.35

Let $u$ be a function such that $E[u(X_{(n)})]=0$ for all $\theta$. Then

which implies $\int_0^\theta u(x)x^{n-1}dx=0$, for all $\theta>0$, $\Rightarrow \int_{\theta_1}^{\theta_2}u(x)x^{n-1}dx=0$ $\forall\ 0<\theta_1<\theta_2<\infty$. $\int_0^{\theta_2}-\int_0^{\theta_1}$

$u(x)x^{n-1}=0$, a.s. for $x\in(0,\infty)$ $\Longrightarrow$ $X_{(n)}$ is complete.

**Example 6.26** (sufficient and complete statistic of i.i.d. Poisson distribution)

Suppose $X_1, \ldots, X_n$ is an i.i.d. sample from Poisson distribution $P(\lambda)$. Then

dim: n → 1 $\quad f(x_1, \ldots, x_n; \lambda) = (e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}) / \prod_{i=1}^n x_i!$.

So $S = \sum_{i=1}^n X_i$ is sufficient for $\lambda$ and $S \sim P(n\lambda)$. If $u(S)$ is a function of $S$ s.t.,

$$0 = E[u(S)] = e^{-n\lambda}\sum_{s=0}^{\infty} \frac{u(s)(n)^s}{s!}\lambda^s, \text{ for all } \lambda,$$

consider it as a Taylor expansion of a function of λ

intuition?

then, all coefficients of $\lambda$ are zero and $u(s) = 0$. Hence $S$ is also complete.

**Theorem 6.12**

- A complete and sufficient statistic is minimal sufficient. However, a minimal sufficient statistic is not necessarily complete (e.g., Ex.6.24 in LNp.53).
- If a non-constant function of a sufficient statistic $S = (S_1, \ldots, S_k)$ is ancillary, then $S$ is not complete. complete but not sufficient statistic? → only keep $X_1$ of $X_1, \cdots, X_n \overset{\text{i.i.d}}{\sim} P(\lambda)$

**Definition 6.16** (one-parameter exponential family of probability distributions, TBp.308)

A family of distributions $\{f(x|\theta) : \theta \in \Omega\}$ is a **one-parameter exponential family** if the pdf or pmf is of the form:

support of f

$$f(x|\theta) = \begin{cases} \exp[\underbrace{c(\theta)\,T(x)}_{\text{data and parameter are mixed in the term in this way}} + d(\theta) + S(x)] = e^{c(\theta)T(x)}e^{d(\theta)}e^{S(x)}, & x \in A \\ 0, & x \notin A \end{cases}$$

where the set $A$ does not depend on $\theta$.



**Theorem 6.13** (sufficient and complete statistics, one-parameter exponential family, TBp.309)

Suppose $X_1, X_2, \cdots, X_n$ is an i.i.d. sample from a member of the exponential family, the joint probability function is

$$\begin{aligned} f(x_1, \ldots, x_n|\theta) &= \prod_{i=1}^n \exp[c(\theta)T(x_i) + d(\theta) + S(x_i)]\, I_A(x_i) \\ &= \exp\left[c(\theta)\sum_{i=1}^n T(x_i) + nd(\theta) + \sum_{i=1}^n S(x_i)\right]\prod_{i=1}^n I_A(x_i), \end{aligned}$$

exp(log(□))

where $I_A(x_i) = 1$ if $x_i \in A$ and 0 otherwise. Then $\sum_{i=1}^n T(X_i)$ is a sufficient and complete statistics for $\theta$.

dim = 1 $\quad$ by factorization Thm

**Example 6.27** (some one-parameter exponential families, TBp.309)

- The pmf of the Bernoulli distribution $B(\theta)$ is
$$P(X = x) = \overset{\exp(\log(\cdot))}{\theta^x(1-\theta)^{1-x}} = \exp\left[x \log\left(\frac{\theta}{1-\theta}\right) + \log(1-\theta)\right], \quad x = 0, 1.$$
$c(\theta)$
This is a one-parameter exponential family with $T(x) = x$. For i.i.d. $X_1, \ldots, X_n \sim B(\theta)$, $\sum_{i=1}^n X_i$ is sufficient and complete for $\theta$.
- The pmf of the Binomial distribution $B(m, \theta)$ is: for $x \in \{0, \ldots, m\}$,
$$p(X = x) = \overset{\exp(\log(\cdot))}{\binom{m}{x}}\theta^x(1-\theta)^{m-x} = \exp\left[x \log\frac{\theta}{1-\theta} + m\log(1-\theta)\right]\binom{m}{x}.$$
$c(\theta)$
This is a one-parameter exponential family with $T(x) = x$. For i.i.d. $X_1, \ldots, X_n \sim B(m, p)$, $\sum_{i=1}^n X_i$ is sufficient and complete for $\theta$.

- The pmf of the Poisson distribution $P(\lambda)$ is
$$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}=\exp(x\log\lambda-\underbrace{\lambda}_{c(\theta)}-\log x!),\quad x=0,1,2,\ldots$$
This is a one-parameter exponential family with $T(x)=x$. For i.i.d. $X_1,\ldots,X_n\sim P(\lambda)$, $\sum_{i=1}^n X_i$ is sufficient and complete for $\lambda$.

**Definition 6.17** (regular $k$-parameter exponential family, TBp.309)

A family of distributions $\{f(x|\Theta):\Theta\in\Omega\subset\mathbb{R}^k\}$ is called a **regular $k$-parameter exponential family** if the pdf or pmf is of the form
$$f(x;\Theta)=\begin{cases}\exp\left[\sum_{j=1}^{k} q_j(\Theta)\,T_j(x)\right]\underbrace{c(\Theta)}_{\exp(\log(\cdot))}h(x), & x\in A\\ 0, & \text{otherwise}\end{cases}$$
where $\Theta=(\theta_1,\theta_2,\ldots,\theta_k)$ is $k$-parameter, and the following conditions hold:

1. $A$ does not depend on $\Theta$, and $\Omega$ contains a nonempty, $k$-dimensional open rectangle.
2. $\{(q_1(\Theta),q_2(\Theta),\ldots,q_k(\Theta)):\Theta\in\Omega\}$ is non-degenerate (i.e. dim = k) and $q_j(\Theta)$'s are non-trivial, functionally independent, continuous function of $\Theta$.
3. **(a)** For continuous case, $T_j(x)$'s are linearly independent, continuous functions of $x$ over $A$; **(b)** For discrete case, $T_j(x)$'s are nontrivial functions of $x$, and none is a linear function of the others.



**Theorem 6.14** (sufficient and complete statistics for regular $k$-parameter exponential family)

Let $X_1,X_2,\cdots,X_n$ be an i.i.d. sample from a regular $k$-parameter exponential family, then
$$S_1=\sum_{i=1}^n T_1(X_i),\quad S_2=\sum_{i=1}^n T_2(X_i),\ldots,S_k=\sum_{i=1}^n T_k(X_i)$$
is a minimal set of complete and sufficient statistics for $\theta_1,\theta_2,\ldots,\theta_k$.

**Example 6.28** (some regular $k$-parameter exponential families)

cf. Ex6.22 (LNp.49)

- The pdf of the Normal distribution $N(\mu,\sigma^2)$ is

$(g_1,g_2)\in\mathbb{R}\times(0,\infty)$
$$\begin{aligned}f(x|\mu,\sigma)&=1/(\sqrt{2\pi}\sigma)\exp\left[-(x-\mu)^2/(2\sigma^2)\right]\\&=\exp\left[\underbrace{\frac{\mu}{\sigma^2}}_{g_1(\theta)}x-\underbrace{\frac{1}{2\sigma^2}}_{g_2(\theta)}x^2-\frac{\mu^2}{2\sigma^2}-\log(\sqrt{2\pi}\sigma)\right]\end{aligned}$$
This is a regular two-parameter exponential family with $T_1(x)=x$ and $T_2(x)=x^2$. Consequently, for an i.i.d. sample $X_1,\ldots,X_n$ from $N(\mu,\sigma^2)$, $(S_1=\sum_{i=1}^n X_i,\ S_2=\sum_{i=1}^n X_i^2)$ is a minimal set of sufficient and complete statistics for $(\mu,\sigma^2)$. Since the relations
$$\frac{S_1}{n}=\overline{X}\equiv\hat{\mu}\quad\text{and}\quad\frac{S_2-S_1^2/n}{n-1}=\frac{\overbrace{\sum_{i=1}^n(X_i-\overline{X})^2}^{\sum_{i=1}^n X_i^2-n\overline{X}^2}}{n-1}\equiv\hat{\sigma}^2$$
Same partitions of data space (graph in LNp52)

define a one-to-one transformation, $(\hat{\mu},\hat{\sigma}^2)$ are also sufficient and complete for $(\mu,\ \sigma^2)$. ✱ Also, they offer estimates. Check Thm4.1 (LN.Ch1~6.p.80) check the binomial case (Ex6.27, LNp.55) $(X_1, \cdots, X_r)$

- **Exercise:** Show that Multinomial distribution, $Multinomial(n, p_1, \ldots, p_r)$, is a regular $(r-1)$-parameter exponential family and find its sufficient and complete statistics. [Hint: subsitute $X_r$ by $n - X_1 - X_2 - \cdots - X_{r-1}$ and $p_r$ by $1 - p_1 - p_2 - \cdots - p_{r-1}$]

**Exercise:** Check whether other distributions given in LN, Ch1-6, p.58-85, belong to exponential family.

$\boxed{4/1}$

❖ **Reading**: textbook, 8.8, 8.8.1; **Further reading**: Hogg et al., 7.2, 7.4, 7.5, 7.7, 7.8, 7.9

## • criteria for evaluating estimators

**Question 6.4** (choice among different estimators of the same parameter, TBp.298)

- In most statistical estimation problems, there are a variety of possible parameter estimators. under same statistical modeling
- Among these estimators, how to choose a better ones? $\boxed{\text{criteria}=?}$
- What properties, that we have defined and discussed for estimators, can be used to evaluate estimators?

**1.** unbiased **2.** consistency (large sample criterion)

$E_\theta(\hat{\theta}-\theta)=0 \Leftrightarrow E_\theta(\hat{\theta})=\theta$ $\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta$

Note that the two criteria not directly compare the dispersion of estimator. Any other criteria?