## Exercise 3.13

The proof follows from the original definition of sufficiency, We have following partition created by $T$:

$$\mathcal{A} = (A_0, A_1, \dots, A_n)$$

where $A_r = \{X \mid \sum_{i=1}^{n} X_i = r\}$ for $r = 0, 1, \dots, n$.

Then $P(X = x \mid X \in A_r) = \dfrac{P((X = x) \cap (X \in A_r))}{P(X \in A_r)}$ - . . . . $(*)$

Note: that $X \in A_r$ means that $\sum_{i=1}^{n} X_i = r$ and we known from basic statistical knowledge that

$$\sum_{i=1}^{n} X_i \sim \text{Binomial}(n, \theta)$$

(i.e the sum of Bernoulli's has a binomial distribution)

Hence, $P(X \in A_r) = \binom{n}{r} \theta^r (1-\theta)^{n-r}$ $r = 0, 1, \dots, n$

and

$$P((X = x) \cap (X \in A_r)) = \begin{cases} 0 & \text{if } \sum_{i=1}^{n} x_i \neq r \\ \theta^{\sum x_i} (1-\theta)^{n - \sum x_i} & \text{if } \sum_{i=1}^{n} x_i = r \end{cases}$$

Therefore the conditional probability in $(*)$ equals

$$P(X = x \mid X \in A_r) = \begin{cases} 0 & \text{if } \sum_{i=1}^{n} x_i \neq r \\ \dfrac{\theta^r (1-\theta)^{n-r}}{\binom{n}{r} \theta^r (1-\theta)^{n-r}} = \dfrac{1}{\binom{n}{r}} & \text{if } \sum_{i=1}^{n} x_i = r \end{cases}$$

The conditional probability does not depend on $\theta$. Hence $T(X) = \sum_{i=1}^{n} X_i$ is sufficent for $\theta$.

Exercise 3.14

(i) For Bernoulli: $T = \sum_{i=1}^{n} X_i$ is sufficient which directly follows from:

$$L(X;\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} = \theta^t(1-\theta)^{n-t}$$
$$= g(t,\theta)$$

which involves the data only via the value of $\sum_{i=1}^{n} x_i = t$ so the whole RHS can be thought of as $g(t,\theta) = \theta^t(1-\theta)^{n-t}$ and simply $h(X) = 1$.

$\bar{x}$

$\bar{X}$ sample mean

$\frac{1}{n}\sum X_i$

(ii) $N(\mu,\sigma^2)$ with $\theta = \begin{pmatrix}\mu\\ \sigma^2\end{pmatrix}$

Then
$$L(X,\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(X_i-\mu)^2\right)$$
$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right)$$

apply fundamental equality

$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2\right]\right)$$

which involves the data via $T = \begin{pmatrix}T_1\\ T_2\end{pmatrix} = \begin{pmatrix}\bar{X}\\ \sum_{i=1}^{n}(X_i-\bar{X})^2\end{pmatrix}$ only.

Hence
$$g(t,\theta) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\left[T_2 + n(T_1-\mu)^2\right]\right)$$
and $h(X) = 1$.

Hence this 2-dim vector statistic is sufficient for $\theta = \begin{pmatrix}\mu\\ \sigma^2\end{pmatrix}$.

Note: every one-to-one transformation of $T$ is also sufficient. In particular
$$\tilde{T} = \begin{pmatrix}\tilde{T}_1\\ \tilde{T}_2\end{pmatrix} = \begin{pmatrix}\sum_{i=1}^{n} X_i\\ \sum_{i=1}^{n} X_i^2\end{pmatrix}$$

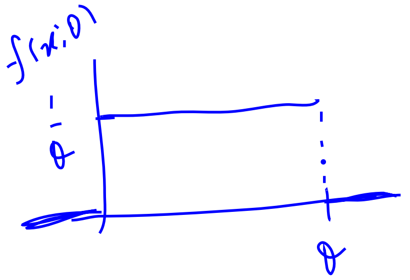
is also sufficient for $\theta$ since knowing $T = \begin{pmatrix}T_1\\ T_2\end{pmatrix}$ we can get $\begin{pmatrix}\tilde{T}_1\\ \tilde{T}_2\end{pmatrix}$ and vice versa.

(iii) $X_1, X_2, \dots, X_n$ iid $U[0,\theta]$.

Claim: $T = X_{(n)}$ - the $n^{th}$ order statistic, equal to the maximal of the $n$ observation. is sufficient for $\theta$.

The density can be expressed using indicator functions

$$f(x;\theta) = \frac{1}{\theta} I_{(x,\infty)}(\theta) \quad \text{where} \quad I_{(x,\infty)}(\theta) = \begin{cases} 1 & \text{if } \theta \in (x,\infty) \\ 0 & \text{otherwise} \end{cases}$$




Then

$$L(X;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{(X_i,\infty)}(\theta)$$

$$= \frac{1}{\theta^n} \prod_{i=1}^{n} I_{(X_i,\infty)}(\theta)$$

$$= \frac{1}{\theta^n} I_{(X_{(n)},\infty)}(\theta) \times 1$$

$$= g(X_{(n)},\theta)\, h(\underline{X})$$

which represents a factorization and $T = X_{(n)}$ is sufficent for $\theta$.

(iv) Multivariate normal. $X_i \sim N_p(\mu, \Sigma)$ $\quad \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots \\ \sigma_{21} & \sigma_{22} & \cdots \\ \vdots & & \ddots \end{pmatrix}$

$$L(X;\mu,\Sigma) = \prod_{i=1}^{n} (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(X_i-\mu)^T \Sigma^{-1}(X_i-\mu)\right)$$

$$= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^T \Sigma^{-1}(X_i-\mu)\right)$$

trace property

$$= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \operatorname{tr}\left(\Sigma^{-1}(X_i-\mu)(X_i-\mu)^T\right)\right)$$

$$= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left[\Sigma^{-1}\left(\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})^T + n(\bar{X}-\mu)(\bar{X}-\mu)^T\right)\right]\right)$$

tr and $\sum_{i=1}^{n}$ are linear operators and their order can be reversed

Therefore $L(X;\mu,\Sigma)$ involves the data only via $T_1 = \bar{X}$ and $T_2 = \sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})^T$. Hence $T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$ is a sufficient statistic for $\mu$ and $\Sigma$.

Exercise 3.15

(i) For Bernoulli. Take two independent samples of n-tuples of data $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$. then

$$\frac{L(Y,\theta)}{L(X,\theta)} = \frac{\theta^{\sum Y_i}(1-\theta)^{\cancel{n}-\sum Y_i}}{\theta^{\sum x_i}(1-\theta)^{\cancel{n}-\sum x_i}} = \left(\frac{\theta}{1-\theta}\right)^{\sum Y_i - \sum x_i}$$

For this not to depend on $\theta$ we require $\sum_{i=1}^n y_i = \sum_{i=1}^n x_i$.

Hence, $T = \sum_{i=1}^n X_i$ is the minimal sufficient statistic.

(ii) $N(\mu, \sigma^2)$ $\theta = \binom{\mu}{\sigma^2}$. We have that:

$$\frac{1}{e^{-x}} = e^x$$

$$\frac{L(Y,\theta)}{L(X,\theta)} = \frac{\cancel{(2\pi\sigma^2)^{-n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n Y_i^2 - 2\mu \sum_{i=1}^n Y_i + \cancel{n\mu^2}\right)\right)}{\cancel{(2\pi\sigma^2)^{-n/2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + \cancel{n\mu^2}\right)\right)}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\underbrace{\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2}_{=0} - 2\mu\left(\underbrace{\sum_{i=1}^n Y_i - \sum_{i=1}^n X_i}_{=0}\right)\right)\right)$$

and for this to not depend on $\theta$ we need:

$\begin{pmatrix} \sum X_i = \sum Y_i \\ \sum X_i^2 = \sum Y_i^2 \end{pmatrix}$ to hold. Hence $T = \begin{pmatrix} \sum X_i \\ \sum X_i^2 \end{pmatrix}$ is minimal sufficient for $\theta = \binom{\mu}{\sigma^2}$. As is any one-to-one transformation of T.

(iii) $X_1, X_2, \dots, X_n \sim U[0,\theta]$

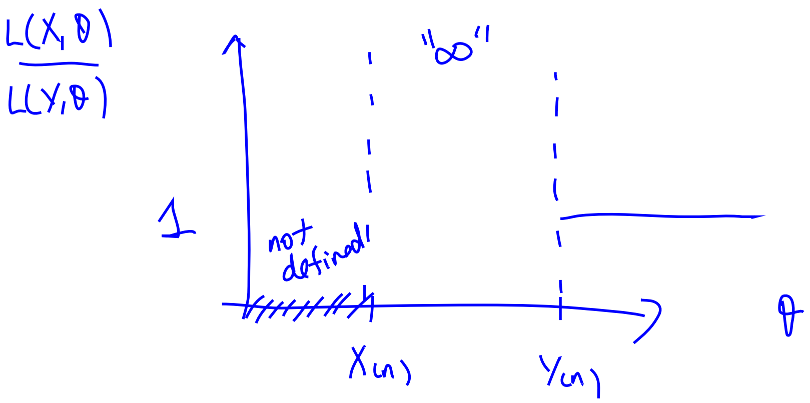
then $$\frac{L(X,\theta)}{L(Y,\theta)} = \frac{\cancel{\frac{1}{\theta^n}} I_{(X_{(n)},\infty)}(\theta)}{\cancel{\frac{1}{\theta^n}} I_{(Y_{(n)},\infty)}(\theta)} = \frac{I_{(X_{(n)},\infty)}(\theta)}{I_{(Y_{(n)},\infty)}(\theta)}$$

This is independent of $\theta$ iff $X_{(n)} = Y_{(n)}$ which implies that $T = X_{(n)}$ is minimal sufficient.

Indeed if $X_{(n)} \neq Y_{(n)}$ we can consider two cases.

① Suppose $X_{(n)} < Y_{(n)}$ then

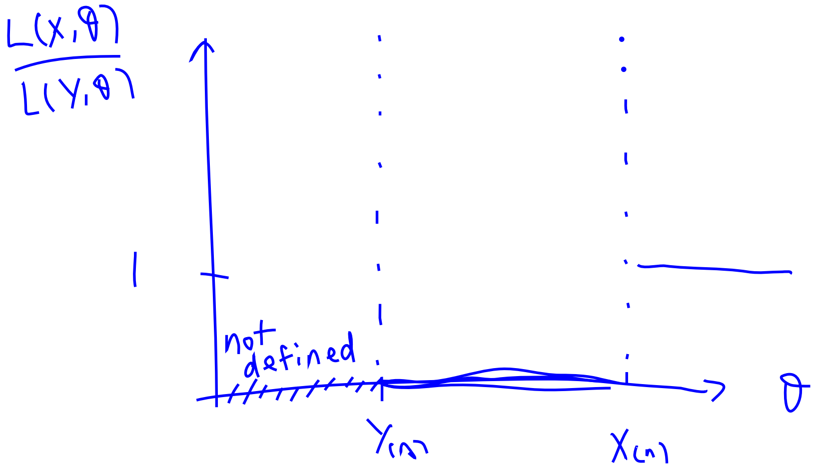
- if $\theta < X_{(n)}$ then $\frac{L(X,\theta)}{L(Y,\theta)} = \text{"}\frac{0}{0}\text{"}$ undefined
- if $X_{(n)} < \theta < Y_{(n)}$ then $\frac{L(X,\theta)}{L(Y,\theta)} = \frac{1}{0}$ "$\infty$"
- if $\theta > Y_{(n)}$ then $\frac{L(X,\theta)}{L(Y,\theta)} = \frac{1}{1} = 1$




Clearly depends on $\theta$.

② Now suppose $Y_{(n)} < X_{(n)}$ then

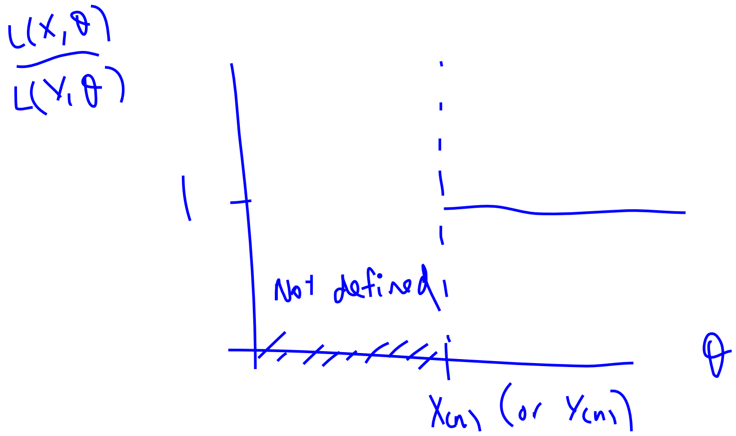
- if $\theta < Y_{(n)}$ then $\frac{L(X,\theta)}{L(Y,\theta)} = \text{"}\frac{0}{0}\text{"}$ undefined
- if $Y_{(n)} < \theta < X_{(n)}$ then $\frac{L(X,\theta)}{L(Y,\theta)} = \frac{0}{1} = 0$
- if $\theta > X_{(n)}$ then $L(X,\theta)/L(Y,\theta) = \frac{1}{1} = 1$




In both cases, when $X_{(n)} \neq Y_{(n)}$ the ratio $\frac{L(X,\theta)}{L(Y,\theta)}$ (where it is defined) depends on the position of $\theta$ i.e. it is not independent of $\theta$.

To have it not depend on $\theta$ we require $X_{(n)} = Y_{(n)}$ to hold since

- if $\theta < X_{(n)} = Y_{(n)}$ then $\frac{L(X,\theta)}{L(Y,\theta)} = \frac{"0"}{0}$ undefined.
- if $\theta > X_{(n)} = Y_{(n)}$ then $\frac{L(X,\theta)}{L(Y,\theta)} = \frac{1}{1} = 1$



(iv). $X_1, X_2, \ldots, X_n$ are iid Cauchy$(\theta)$ with density

$$f(x;\theta) = \frac{1}{\pi(1+(x-\theta)^2)} \qquad -\infty < x < \infty$$

then

$$\frac{L(Y,\theta)}{L(X,\theta)} = \frac{\cancel{\pi} \prod_{i=1}^{n} (1+(X_i-\theta)^2)}{\cancel{\pi^n} \prod_{i=1}^{n} (1+(Y_i-\theta)^2)} \qquad \frac{(1+(X_{(1)}-\theta)^2) \cdots \cdots (1+(X_{(n)}-\theta)^2)}{(1+(Y_{(1)}-\theta)^2) \cdots \cdots (1+(Y_{(n)}-\theta)^2)}$$

and for it to not depend on $\theta$ $\quad \begin{pmatrix} X_{(1)} = Y_{(1)} \\ X_{(2)} = Y_{(2)} \\ \vdots \quad\quad \vdots \\ X_{(n)} = Y_{(n)} \end{pmatrix}$

Hence $T = \begin{pmatrix} X_{(1)} \\ X_{(2)} \\ \vdots \\ X_{(n)} \end{pmatrix}$ is the minimal sufficient in this case.

Note: the dimension is equal to the sample size, so virtually no dimension reduction is possible in this case.

Exercise 3.16

(i) Poisson$(\theta)$:
$$f(x,\theta) = \frac{e^{-\theta}\theta^{x}}{x!} = e^{-\theta}\frac{1}{x!}e^{x\log\theta} \qquad x = 0,1,2,\ldots$$

Hence $a(\theta) = e^{-\theta}$ $\quad b(x) = \frac{1}{x!}$ $\quad c(\theta) = \log\theta$ $\quad d(x) = x$

There, $T(X) = \sum_{i=1}^{n} X_i$ is minimal sufficient for $\theta$.

(ii) Bernoulli$(\theta)$:
$$f(x,\theta) = \theta^{x}(1-\theta)^{1-x} = (1-\theta)\theta^{x}(1-\theta)^{-x}$$
$$= (1-\theta)\left(\frac{\theta}{1-\theta}\right)^{x}$$
$$= (1-\theta)\exp\left(x\log\left(\frac{\theta}{1-\theta}\right)\right)$$

Hence, $a(\theta) = 1-\theta$

$b(x) = 1$

$c(\theta) = \log\left(\frac{\theta}{1-\theta}\right)$

$d(x) = x$

Therefore, $T(X) = \sum_{i=1}^{n} X_i$ is minimal sufficient for $\theta$.

(iii) Normal$(\theta, 1)$:
$$f(x,\theta) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\theta)^2}$$
$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x^2-2x\theta+\theta^2)}$$
$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\theta^2}e^{-\frac{1}{2}x^2}e^{x\theta}$$

$a(\theta) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\theta^2}$ $\quad b(x) = e^{-\frac{1}{2}x^2}$ $\quad c(\theta) = \theta$ $\quad d(x) = x$

therefore, $T =$

(iv) Normal$(0, \theta^2)$: $\quad f(x,\theta) = \dfrac{1}{\sqrt{2\pi\theta^2}} e^{-\frac{1}{2\theta^2} x^2}$

$$a(\theta^2) = \frac{1}{\sqrt{2\pi\theta^2}} \qquad b(x) = 1 \qquad c(\theta^2) = -\frac{1}{2\theta^2} \qquad d(x) = x^2$$

Note: the parameter of interest is $\theta^2$.

Therefore, $T = \sum_{i=1}^{n} X_i^2$ is minimal sufficient for $\theta^2$

Exercise 3.17

$$f(x;\theta_1,\theta_2) = \frac{1}{B(\theta_1,\theta_2)} x^{\theta_1 - 1} (1-x)^{\theta_2 - 1}$$

$$= \frac{1}{B(\theta_1,\theta_2)} \frac{1}{x} \frac{1}{1-x} x^{\theta_1} (1-x)^{\theta_2}$$

$$= \underbrace{\frac{1}{B(\theta_1,\theta_2)}}_{a(\theta)} \underbrace{\frac{1}{x(1-x)}}_{b(x)} e^{\theta_1 \log x + \theta_2 \log(1-x)}$$

$$c_1(\theta_1,\theta_2) = \theta_1 \qquad c_2(\theta_1,\theta_2) = \theta_2$$

$$d_1(x) = \log x \qquad d_2(x) = \log(1-x)$$

Hence, $$T = \begin{pmatrix} \sum_{i=1}^{n} \log X_i \\ \sum_{i=1}^{n} \log(1-X_i) \end{pmatrix} = \begin{pmatrix} \log \prod_{i=1}^{n} X_i \\ \log \prod_{i=1}^{n} 1-X_i \end{pmatrix}$$

is minimal sufficient for $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$

[and so is $\tilde{T} = \begin{pmatrix} \prod_{i=1}^{n} X_i \\ \prod_{i=1}^{n} 1-X_i \end{pmatrix}$ ]

Proof of information properties.

(i) First $L_{(X,Y)}(x,y;\theta) = L_X(x,\theta)\, L_Y(y,\theta)$

this implies
$$\frac{\partial}{\partial\theta}\log L_{(X,Y)}(x,y,\theta) = \frac{\partial}{\partial\theta}\log L_X(x,\theta) + \frac{\partial}{\partial\theta}\log L_Y(y,\theta)$$

Thus giving:
$$I_{(X,Y)}(\theta) = E\left[\frac{\partial}{\partial\theta}\log L_{(X,Y)}(x,y;\theta)^2\right]$$
$$= E\left[\frac{\partial}{\partial\theta}\log L_X(x,\theta)^2\right] + E\left[\frac{\partial}{\partial\theta}\log L_Y(y,\theta)^2\right]$$
$$+ 2E\left[\frac{\partial}{\partial\theta}\log L_X(x,\theta)\cdot\frac{\partial}{\partial\theta}\log L_Y(y,\theta)\right]$$
$$= I_X(\theta) + I_Y(\theta) + 2E\underbrace{[V(X,\theta)\cdot V(Y,\theta)]}_{E(V(X,\theta))\cdot E(V(Y,\theta)) = 0\cdot 0 = 0}$$
$$= I_X(\theta) + I_Y(\theta)$$

$\square$

(iii)
$$\frac{\partial^2}{\partial\theta^2}\log f(x,\theta) = \frac{\partial}{\partial\theta}\left[\frac{\frac{\partial}{\partial\theta}f(x,\theta)}{f(x,\theta)}\right]$$

then by applying the quotient rule $\left(\frac{u}{v}\right)' = \frac{u'v - v'u}{v^2}$

with $u = \frac{\partial}{\partial\theta}f(x,\theta)$ and $v = f(x,\theta)$.

$$\frac{\partial^2}{\partial\theta^2}\log f(x,\theta) = \frac{\frac{\partial^2}{\partial\theta^2}f(x,\theta)\cdot f(x,\theta) - \frac{\partial}{\partial\theta}f(x,\theta)\cdot\frac{\partial}{\partial\theta}f(x,\theta)}{f(x,\theta)^2}$$

$$= \frac{\frac{\partial^2}{\partial\theta^2}f(x,\theta)}{f(x,\theta)} - \Bigg[\underbrace{\frac{\frac{\partial f(x,\theta)}{\partial\theta}}{f(x,\theta)}}_{\frac{\partial}{\partial\theta}\log f(x,\theta)}\Bigg]^2$$

$E(\ \ ) = 0$

(iv) $E\left[\frac{\partial}{\partial\theta}\log L(X,\theta)\cdot\frac{\partial}{\partial\theta}\log f_T(T,\theta)\right]$

$$= E_T\left[E\left[\frac{\partial}{\partial\theta}\log L(X,\theta)\cdot\frac{\partial}{\partial\theta}\log f_T(T,\theta)\,\Big|\,T=t\right]\right]$$

by using (2) with $Y=\frac{\partial}{\partial\theta}\log L(X,\theta)\cdot\frac{\partial}{\partial\theta}\log f_T(T,\theta)$ and $Z=T$. The by using (1) and noting we are conditioning on $T$ we get

$$E_T\left[\frac{\partial}{\partial\theta}\log f_T(T,\theta)\ \underbrace{E\left[\frac{\partial}{\partial\theta}\log L(X,\theta)\,\Big|\,T=t\right]}_{\text{show } = \frac{\partial}{\partial\theta}\log f_T(T,\theta)}\right]$$

$(*)$

$$(*) = \sum_{\text{all } x\in\mathcal{X}}\frac{\partial}{\partial\theta}\log L(x,\theta)\cdot P_\theta(X=x\,|\,T=t) \qquad \frac{P(X=x\cap T=t)}{P(T=t)}$$

$$= \sum_{\text{all } x:\,T(x)=t}\frac{\partial}{\partial\theta}\log L(x,\theta)\cdot\frac{L(x,\theta)}{f_T(t,\theta)}$$

$$= \sum_{\text{all } x:\,T(x)=t}\frac{\frac{\partial}{\partial\theta}L(x,\theta)}{\cancel{L(x,\theta)}}\ \frac{\cancel{L(x,\theta)}}{f_T(t,\theta)} \qquad (\log f)' = \frac{f'}{f}$$

$$= \frac{\partial}{\partial\theta}f_T(t,\theta)\,\frac{1}{f_T(t,\theta)} = \frac{\partial}{\partial\theta}\log f_T(t,\theta)$$