Mathematical Statistics 2

Chapter 7: Sufficiency and Comparison

Byeong U. Park

Department of Statistics, Seoul National University
7.1 Optimality

7.2 Sufficiency and Completeness

7.3 Exponential Family
Comparison of Estimators

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf $f(\cdot; \theta)$, $\theta \in \Theta \subset \mathbb{R}^d$. Below, we write $X = (X_1, \ldots, X_n)$. How to compare different estimators of a parameter of interest, say $g(\theta) \in \mathbb{R}$?

- **Loss function:** $L(\cdot, \cdot) : \Theta \times A \to \mathbb{R}_+$, where $A \subset \mathbb{R}$ is an *action space* for the estimation of $g(\theta)$ such that $L(\theta, g(\theta)) = 0$. For example,

$$L(\theta, a) = (a - g(\theta))^2, \quad L(\theta, a) = |a - g(\theta)|.$$

- **Risk function:** For an estimator $\delta(X)$ of $g(\theta)$,

$$R(\theta, \delta) = E_\theta L(\theta, \delta(X)).$$

In the case of the *squared error loss*, the risk $R(\theta, \delta)$ is the mean squared error of $\delta(X)$, i.e., $R(\theta, \delta) = E_\theta (\delta(X) - g(\theta))^2$. 
Difficulty with Uniform Comparison

▶ One would prefer $\delta_1$ to $\delta_2$ if

$$R(\theta, \delta_1) \leq R(\theta, \delta_2) \text{ for all } \theta \in \Theta, \text{ and } R(\theta, \delta_1) < R(\theta, \delta_2) \text{ for some } \theta \in \Theta.$$ 

▶ The difficulty is that there exists no estimator that is best in this sense, i.e., there exists no $\delta_0$ such that

$$R(\theta, \delta_0) \leq R(\theta, \delta) \text{ for all } \delta \text{ and for all } \theta \in \Theta.$$ 

▶ Why? If it exists, then $R(\theta, \delta_0)^{\theta \in \Theta} 0$, which leads to a contradiction.
Optimal Estimation

▶ Restricted class of estimators: One may find an estimator $\delta_0$ in the class of unbiased estimators of $g(\theta)$ such that

$$E_\theta(\delta_0(X) - g(\theta))^2 \leq E_\theta(\delta(X) - g(\theta))^2 \text{ for all } \theta \in \Theta$$

for any unbiased estimator $\delta(X)$. If exists, such an estimator is called \textit{UMVUE} (Uniformly Minimum Variance Unbiased Estimator).

▶ Global measures of performance: The estimator that minimizes the maximum risk $r(\delta) = \max_{\theta \in \Theta} R(\theta, \delta)$ is called a \textit{minimax estimator}. The estimator that minimizes the average risk

$$r(\delta; \pi) = \int_{\Theta} R(\theta, \delta) \pi(\theta) \, d\mu(\theta)$$

for a weight function $\pi$ is called a \textit{Bayes estimator}. 
Example: Minimax Estimation

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$. Consider the loss function $L(\theta, a) = (a/\sigma^2 - 1)^2$ for the estimation of $g(\theta) = \sigma^2$. Suppose we want to find an estimator that minimizes the maximum risk among all estimators in the class $\{\delta_c : \delta_c(x) = c \sum_{i=1}^n (x_i - \bar{x})^2, c > 0\}$.

Let $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$. Since $S^2/\sigma^2 \overset{d}{=} \chi^2(n-1)$, it follows that

\[
R(\theta, \delta_c) = E_\theta(cS^2/\sigma^2 - 1)^2
= c^2 \text{Var}(\chi^2(n-1)) + (cE(\chi^2(n-1)) - 1)^2
= c^2 \cdot 2(n-1) + (c(n-1) - 1)^2
\]

\[\theta \equiv \max_{\theta} R(\theta, \delta_c).\]

Since $\arg \min_{c>0} \max_{\theta} R(\theta, \delta_c) = 1/(n+1)$, we get

\[\hat{\sigma}^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n + 1)\]

as the minimax estimator among the class.
7.1 Optimality

7.2 Sufficiency and Completeness

7.3 Exponential Family
The Idea of Sufficiency

Suppose that we observe $X_1$ and $X_2$ that are i.i.d. Bernoulli($\theta$) random variables, where $0 < \theta < 1$.

- The distribution of $Y = X_1 + X_2$:

  $$P_\theta(Y = 0) = (1 - \theta)^2, \quad P_\theta(Y = 1) = 2\theta(1 - \theta), \quad P_\theta(Y = 2) = \theta^2.$$  

- When we observe $Y$, we may produce $(X_1^*, X_2^*)$, without knowledge of the true $\theta$, that has the same distribution of the original $(X_1, X_2)$, as follows:
  (i) Put $(X_1^*, X_2^*) = (0, 0)$ when $Y = 0$; (ii) put $(X_1^*, X_2^*) = (1, 1)$ when $Y = 2$; (iii) conduct a randomized experiment and put $(X_1^*, X_2^*) = (1, 0)$ and $(X_1^*, X_2^*) = (0, 1)$, each with probability $1/2$ when $Y = 1$.

- Thus, for any estimator $\delta(X_1, X_2)$ of $g(\theta)$ one may find an estimator that depends only on $Y$ rather than $(X_1, X_2)$ but has the same risk as $\delta(X_1, X_2)$. 

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf $f(\cdot; \theta)$, $\theta \in \Theta \subset \mathbb{R}^d$. We write $X = (X_1, \ldots, X_n)$ below.

▶ Sufficient statistic for $\theta \in \Theta$: A statistic $Y = u(X)$ is called a \textit{sufficient statistic} if the conditional distribution of $X$ given $Y$ does not depend on $\theta \in \Theta$, i.e., if

$$P_{\theta_1}(X \in A | Y = y) \stackrel{\theta_1, \theta_2 \in \Theta}{=} P_{\theta_2}(X \in A | Y = y)$$

for all $A$ and for all $y$.

▶ This definition is more general than those in the textbook.

▶ Read Remark 7.2.1 on p.391.
Factorization Theorem

A statistic \( Y = u(X) \) is a sufficient statistic for \( \theta \in \Theta \) if and only if there exist functions \( f_1 \) and \( f_2 \) such that

\[
\prod_{i=1}^{n} f(x_i; \theta) = f_1(u(x), \theta) \cdot f_2(x) \quad \text{for all } x \text{ and for all } \theta \in \Theta.
\]

**Proof.** We give a proof for the case of discrete \( X_i \). For the necessity part,

\[
\prod_{i=1}^{n} f(x_i; \theta) = P_{\theta}(X = x, Y = u(x))
\]

\[
= P(X = x | Y = u(x)) \cdot P_{\theta}(Y = u(x)).
\]

The first component on the RHS of the second equation does not depend on \( \theta \in \Theta \). For the sufficiency part,

\[
P_{\theta}(X = x | Y = y) = \frac{P_{\theta}(X = x, u(X) = y)}{P_{\theta}(u(X) = y)} = \frac{P_{\theta}(X = x) I(u(x) = y)}{\sum_{z:u(z)=y} P_{\theta}(X = z)}
\]

\[
= \frac{f_2(x) I(u(x) = y)}{\sum_{z:u(z)=y} f_2(z)}.
\]

\( \square \)
Sufficient Statistic: Examples

- Bernoulli($\theta$), $\theta \in (0, 1)$:

$$\prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum x_i} (1 - \theta)^{n-\sum x_i},$$

so that $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic for $\theta \in (0, 1)$.

- Gamma($2, \theta$), $\theta > 0$:

$$\prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \theta^{-2} \cdot x_i \cdot e^{-x_i/\theta} \cdot I_{(0, \infty)}(x_i)$$

$$= \theta^{-2n} \exp \left( - \sum_{i=1}^{n} x_i/\theta \right) \cdot \left( \prod_{i=1}^{n} x_i \right) I_{(0, \infty)}(\min x_i),$$

so that $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic for $\theta > 0$. 
Sufficient Statistic: Examples

▶ \textbf{Gamma}(\alpha, \beta), \ \theta = (\alpha, \beta) \in \mathbb{R}_+ \times \mathbb{R}_+:

\[
\prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \Gamma(\alpha)^{-1} \beta^{-2} \cdot x_i^{\alpha-1} \cdot e^{-x_i/\beta} \cdot I_{(0,\infty)}(x_i)
\]

\[
= \Gamma(\alpha)^{-n} \beta^{-2n} \left( \prod_{i=1}^{n} x_i \right)^{\alpha-1} \exp \left( - \sum_{i=1}^{n} x_i/\beta \right) 
\times I_{(0,\infty)}(\min x_i),
\]

so that \( Y = (\prod_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i) \) is a sufficient statistic for \( \theta \in \mathbb{R}_+ \times \mathbb{R}_+ \).

▶ \textbf{Uniform}(\theta_1 - \theta_2, \theta_1 + \theta_2), \ \theta = (\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R}_+\): Assume \( n \geq 2 \).

\[
\prod_{i=1}^{n} f(x_i; \theta) = (2\theta_2)^{-n} I_{(\theta_1-\theta_2, \infty)}(x(1))I_{(-\infty, \theta_1+\theta_2)}(x(n)),
\]

so that \( Y = (X_{(1)}, X_{(n)}) \) is a sufficient statistic for \( \theta \in \mathbb{R} \times \mathbb{R}_+ \).

▶ \textbf{Exp}(\theta, 1), \ \theta \in \mathbb{R}: \ Y = X_{(1)} \) is a sufficient statistic for \( \theta \in \mathbb{R} \).
Minimal Sufficient Statistic

There are many sufficient statistics for a given model. As an extreme example, $X = (X_1, \ldots, X_n)$ itself is also a sufficient statistic. As another example, consider the case where $X_1, X_2, X_3$ are i.i.d. Bernoulli($\theta$) random variables and $\theta \in (0, 1)$. For latter model, sufficient statistics include (i) $(X_1, X_2, X_3)$; (ii) $(X_1 + X_2, X_3)$; (iii) $(X_1 + X_3, X_2)$; (iv) $(X_1, X_2 + X_3)$; (v) $X_1 + X_2 + X_3$.

Minimal sufficient statistic: A sufficient statistic is called a *minimal sufficient statistic* (MSS) if it is a function of any sufficient statistic.
Properties of Sufficient Statistic

- Any statistic that is a 1-1 function of a sufficient statistic for $\theta \in \Theta$ is also a sufficient statistic for $\theta \in \Theta$.

**Proof:** Let $Y$ be a sufficient statistic for $\theta \in \Theta$ and let $W = g(Y)$ for a known 1-1 function $g$. Then, the event $(W = w)$ equals $(Y = g^{-1}(w))$ for any $w$, so that for any $A$, $w$ and $\theta \in \Theta$ it holds that

$$P_\theta(X \in A|W = w) = P(X \in A|Y = g^{-1}(w)),$$

the latter not depending on $\theta \in \Theta$. □

- Any statistic that is a 1-1 function of MSS is also an MSS.

- MLE and SS: The unique MLE of $\theta$, when it exists, is a function of any sufficient statistic for $\theta \in \Theta$. 


Properties of Sufficient Statistic

Proof: Let \( u(X) \) be a sufficient statistic for \( \theta \in \Theta \). Then, for the unique MLE \( \hat{\theta} \),

\[
\hat{\theta} = \arg \max_{\theta \in \Theta} \prod_{i=1}^{n} f(x_i; \theta)
\]

\[
= \arg \max_{\theta \in \Theta} f_1(\theta, u(x)) f_2(x)
\]

\[
= \arg \max_{\theta \in \Theta} f_1(\theta, u(x)) = (\text{function of } u(x)). \quad \square
\]

Existence of MSS: If the MLE of \( \theta \in \Theta \) is unique and it is a sufficient statistic for \( \theta \in \Theta \), then it is a minimal sufficient statistic for \( \theta \in \Theta \).

Proof: Immediate from the second property.

Suppose that there exists \( \theta_0 \in \Theta \) such that \( \text{supp}(f(\cdot; \theta)) \subset \text{supp}(f(\cdot; \theta_0)) \) for all \( \theta \in \Theta \). Then, the statistic \( T(X) \), as a (random) function defined on \( \Theta \) in such a way that \( T(X)(\theta) = \prod_{i=1}^{n} \left( f(X_i, \theta) / f(X_i; \theta_0) \right) \), is an MSS.
MSS: An Example

Let $X_1, \ldots, X_n$ ($n \geq 2$) be a random sample from Uniform($\theta_1 - \theta_2, \theta_1 + \theta_2$), $\theta = (\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R}_+$. For this model, $Y = (X_{(1)}, X_{(n)})$ is an MSS.

Proof: The sufficiency was established before. Let the model is reparametrized by $\eta = (\eta_1, \eta_2)$, where $\eta_1 = \theta_1 - \theta_2$ and $\eta_2 = \theta_1 + \theta_2$. Then,

$$\prod_{i=1}^{n} f(x_i; \eta) = (\eta_2 - \eta_1)^{-n} I_{(-\infty,x_{(1)})}(\eta_1) I_{[x_{(n)},\infty)}(\eta_2).$$

Clearly, $(\hat{\eta}_1, \hat{\eta}_2) = (X_{(1)}, X_{(n)})$ is the unique MLE, so that $(\hat{\theta}_1, \hat{\theta}_2)$ defined by

$$\hat{\theta}_1 = (X_{(1)} + X_{(n)})/2, \quad \hat{\theta}_2 = (X_{(n)} - X_{(1)})/2$$

is the unique MLE of $(\theta_1, \theta_2) = ((\eta_1 + \eta_2)/2, (\eta_2 - \eta_1)/2)$. Since $(\hat{\theta}_1, \hat{\theta}_2)$ is a 1-1 function of the SS $(X_{(1)}, X_{(n)})$, it is also an SS and thus an MSS. This establishes that $(X_{(1)}, X_{(n)})$ is an MSS since it is a 1-1 function of $(\hat{\theta}_1, \hat{\theta}_2)$. □
Rao-Blackwell Theorem

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf $f(\cdot; \theta)$, $\theta \in \Theta \subset \mathbb{R}^d$. Let $Y = u(X)$ be a sufficient statistic. Then, for any estimator of $\hat{\eta}(X)$ of $\eta = g(\theta)$ with finite second moment,

$$\hat{\eta}^*(Y) = E(\hat{\eta}(X)|Y)$$

is a statistic with the properties that (i) $E_\theta(\hat{\eta}^*) = E_\theta(\hat{\eta})$; (ii) $\text{var}_\theta(\hat{\eta}^*) \leq \text{var}_\theta(\hat{\eta})$; (iii) $\text{MSE}_\theta(\hat{\eta}^*) \leq \text{MSE}_\theta(\hat{\eta})$.

The Rao-Blackwell theorem tells that one may have a better estimator by conditioning on a sufficient statistic, in terms of MSE. By the theorem, if $\hat{\eta}$ is an unbiased estimator of $\eta$, then $\hat{\eta}^*$ is also an unbiased estimator but with a smaller variance.

Proof of the theorem: $\text{var}(W) = \text{var}(E(W|V)) + E(\text{var}(W|V))$. 
Rao-Blackwell theorem

space of random variables

$E(W)$

$E(W|V)$

space of constants

space of functions of $V$
Example: Rao-Blackwellization

Let \( X_1, \ldots, X_n \ (n \geq 2) \) be a random sample from \( U[0, \theta], \theta > 0 \). Take \( \hat{\theta} = 2\bar{X} \) as an unbiased estimator of \( \theta \). We know that \( X_{(n)} \) is a sufficient statistic for \( \theta > 0 \). By Rao-Blackwell theorem, \( \hat{\theta}^* \equiv E(2\bar{X}|X_{(n)}) \) is an UE of \( \theta \) with variance less than or equal to that of \( \hat{\theta} \). For \( 1 \leq r \leq n - 1 \), we note that

\[
\text{pdf}_{X_{(r)}|X_{(n)}}(x|y) = \frac{(n-1)!}{(r-1)!(n-r-1)!} \left( \frac{x}{y} \right)^{r-1} \left( 1 - \frac{x}{y} \right)^{n-r-1} \frac{1}{y} I_{(0,y)}(x)
\]

and that, recalling the pdf of Beta\((r + 1, n - r + 1)\),

\[
\int_0^y x \cdot \frac{(n-1)!}{(r-1)!(n-r-1)!} \left( \frac{x}{y} \right)^{r-1} \left( 1 - \frac{x}{y} \right)^{n-r-1} \frac{1}{y} \, dx = \frac{ry}{n} \int_0^1 \frac{n!}{r!(n-r-1)!} t^r (1-t)^{n-r-1} \, dt = (r/n)y.
\]
Example: Rao-Blackwellization

Thus, we get

\[
2E(\bar{X}|X_{(n)} = y) = 2n^{-1} \left( y + \sum_{r=1}^{n-1} E(X_r|X_{(n)} = y) \right) \\
= 2n^{-1} \left( y + \sum_{r=1}^{n-1} \frac{r}{n} \cdot y \right) \\
= \frac{n+1}{n} y.
\]

Indeed, \( \hat{\theta}^* = (n+1)X_{(n)}/n \) and

\[
\text{var}_\theta(\hat{\theta}^*) = \left( \frac{n+1}{n} \right)^2 \cdot \text{var}_\theta(X_{(n)}) \\
= \frac{1}{n(n+2)} \theta^2 \\
< \frac{1}{3n} \theta^2 = \text{var}_\theta(\hat{\theta}).
\]
We have seen that taking conditional expectation, on a sufficient statistic, of a given unbiased estimator always improves the estimator in terms of variance.

UMVUE: An estimator $\hat{\eta}$ of $\eta = g(\theta)$ is called the uniformly minimum variance unbiased estimator if it itself is unbiased and $\text{var}_\theta(\hat{\eta}) \leq \text{var}_\theta(\tilde{\eta})$ for all $\theta \in \Theta$ and for any unbiased estimator $\tilde{\eta}$ of $\eta$.

Uniqueness of UMVUE: If there exists an unbiased estimator with finite variance, then UMVUE is unique: Let $\hat{\eta}_1$ and $\hat{\eta}_2$ be UMVUE of $\eta$. Since $\text{var}_\theta(\hat{\eta}_1) = \text{var}_\theta(\hat{\eta}_2) \leq \text{var}_\theta((\hat{\eta}_1 + \hat{\eta}_2)/2) < \infty$ for all $\theta \in \Theta$, we get

$$\text{var}_\theta(\hat{\eta}_1 - \hat{\eta}_2) \leq 0 \text{ (and thus } = 0) \text{ for all } \theta \in \Theta.$$  

This implies $P_\theta(\hat{\eta}_1 - \hat{\eta}_2 = c) = 1$ for all $\theta \in \Theta$ for some constant $c$. That constant equals zero since both $\hat{\eta}_1$ and $\hat{\eta}_2$ are unbiased.
Complete Statistic

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf $f(\cdot; \theta)$, $\theta \in \Theta \subset \mathbb{R}^d$. The following notion of *completeness* facilitates the derivation of UMVUE.

- Complete statistic for $\theta \in \Theta$: A statistic $Y = u(X)$ is called a *complete statistic* for $\theta \in \Theta$ and \{pdf$_Y(\cdot; \theta) : \theta \in \Theta$\} is called a *complete family of distributions* if

  $$E_\theta \varphi(Y) = 0 \text{ for all } \theta \in \Theta \implies P_\theta(\varphi(Y) = 0) = 1 \text{ for all } \theta \in \Theta.$$

- A complete statistic $Y$ is “complete” in the sense that any non-constant function of $Y$ has a non-constant expected value (as a function of $\theta$).

- Complete sufficient statistic: A statistic is called a *complete sufficient statistic* (CSS) for $\theta \in \Theta$ if it is sufficient and complete for $\theta \in \Theta$. 


Rao-Blackwell-Lehmann-Scheffé Theorem

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf $f(\cdot; \theta)$, $\theta \in \Theta \subset \mathbb{R}^d$. Let $Y = u(X)$ be a CSS for $\theta \in \Theta$. Assume that there exists an unbiased estimator with finite variance. Then, (i) for any unbiased estimator $\hat{\eta}_0$ of $\eta = g(\theta)$ with finite variance, $\hat{\eta} = E(\hat{\eta}_0|Y)$ is the UMVUE; (ii) any function of $Y$, say $\varphi(Y)$, is the UMVUE if it is unbiased.

Proof: (i) Clearly, $\hat{\eta}$ is an UE. For an arbitrary unbiased estimator $\tilde{\eta}$ with finite variance,

$$\text{var}_\theta(\tilde{\eta}) \geq \text{var}_\theta(E(\tilde{\eta}|Y)) \equiv \text{var}_\theta(\hat{\eta}),$$

the inequality holding due to the Rao-Blackwell theorem and the equality holding due to the completeness of $Y$ [$P_\theta(\hat{\eta} = E(\tilde{\eta}|Y)) \equiv 1$].

(ii) Replace $\hat{\eta}$ by $\varphi(Y)$ in the proof of (i).

Remark: UE that is a function of a CSS is unique and it is the UMVUE.
Methods of Finding U MVUE

- Method 1: Rao-Blackwellization with a CSS.
  1. Find a CSS $Y = u(X)$.
  2. Find an ‘easy’ UE $\hat{\theta}_0$.
  3. Compute the conditional expectation $E(\hat{\theta}_0|Y)$.

- Method 2: Trial and error.
  1. Find a CSS $Y = u(X)$.
  2. Solve $E_\theta \varphi(Y) \equiv g(\theta)$ with respect to $\varphi$,
     or try some $\varphi(Y)$ and check unbiasedness.
CSS and UMVUE: Bernoulli Model

Let $X_1, \ldots, X_n$ be a random sample from Bernoulli($\theta$), $\theta \in [0, 1]$.

- Finding a CSS: By Factorization theorem, $Y = \sum_{i=1}^{n} X_i$ is an SS. To check whether it is also complete for $\theta \in [0, 1]$, suppose that $E_\theta \varphi(Y) = 0$ for all $\theta \in [0, 1]$ for a function $\varphi$. Then, since $Y \overset{d}{=} \text{Binomial}(n, \theta)$, we have

$$\sum_{y=0}^{n} \varphi(y) \binom{n}{y} \theta^y (1 - \theta)^{n-y} \overset{\theta}{=} 0$$

$$\Rightarrow \sum_{y=0}^{n} \varphi(y) \binom{n}{y} \lambda^y \lambda \overset{\lambda \geq 0}{=} 0$$

$$\Rightarrow \varphi(y) \binom{n}{y} \overset{y}{=} 0 \quad \text{(uniqueness of polynomial coefficients)}$$

$$\Rightarrow \varphi(y) \overset{y}{=} 0.$$ 

- Estimation of the mean: Since $E_\theta (Y/n) \overset{\theta}{=} \theta$, we conclude that $\hat{\eta} = \bar{X}$ is the UMVUE of $\theta$. 
CSS and UMVUE: Bernoulli Model

- Estimation of the variance (Method 1): For the estimation of 
  \( \eta = \theta(1 - \theta) \), consider \( \hat{\eta}_0 = X_1(1 - X_2) \) as an UE of \( \eta \). Then, we get

  \[
  E(\hat{\eta}_0|Y = y) = P\left(X_1 = 1, X_2 = 0 \mid \sum_{i=1}^{n} X_i = y\right)
  \]

  \[
  = \binom{n-2}{y-1} /\binom{n}{y} = y(n - y)/n(n - 1),
  \]

  so that \( \hat{\eta} = n\bar{X}(1 - \bar{X})/(n - 1) \) is the UMVUE.

- Estimation of the variance (Method 2): Try the MLE of \( \eta \):
  \( \hat{\eta}_{\text{MLE}} = \bar{X}(1 - \bar{X}) \). We get

  \[
  E_\theta(\bar{X}(1 - \bar{X})) = E_\theta \bar{X} - \left[\text{var}(\bar{X}) + (E_\theta(\bar{X}))^2\right] = (n - 1)\theta(1 - \theta)/n.
  \]

  Thus, \( \hat{\eta} = n\bar{X}(1 - \bar{X})/(n - 1) \) is an UE that is a function of CSS, concluding that it is the UMVUE.
CSS and UMVUE: Poisson Model

Let $X_1, \ldots, X_n$ be a random sample from $\text{Poisson}(\theta)$, $\theta > 0$.

- **Finding a CSS:** By Factorization theorem, $Y = \sum_{i=1}^{n} X_i$ is an SS. To check whether it is also complete for $\theta > 0$, suppose that $E_{\theta} \varphi(Y) = 0$ for all $\theta > 0$ for a function $\varphi$. Then, since $Y \overset{d}{=} \text{Poisson}(n\theta)$, we have

$$
\sum_{y=0}^{\infty} \varphi(y) \frac{(n\theta)^y e^{-n\theta}}{y!} \overset{\theta > 0}{=} 0
$$

$$
\Rightarrow \sum_{y=0}^{\infty} \frac{\varphi(y)}{y!} \lambda^y \overset{\lambda \geq 0}{=} 0
$$

$$
\Rightarrow \frac{\varphi(y)}{y!} y \overset{y}{=} 0 \text{ (uniqueness of power series coefficients)}
$$

$$
\Rightarrow \varphi(y) \overset{y}{=} 0.
$$

- **Estimation of the mean:** Since $E_{\theta}(Y/n) \overset{\theta}{=} \theta$, we conclude that $\hat{\eta} = \bar{X}$ is the UMVUE of $\theta$. 
CSS and UMVUE: Poisson Model

- Estimation of $\eta = e^{-2\theta}$: We solve $E_\theta \varphi(Y)^\theta e^{-2\theta}$ with respect to $\varphi$.

  Note that

  $$\sum_{y=0}^{\infty} \varphi(y) \frac{(n\theta)^y e^{-n\theta}}{y!} \equiv e^{-2\theta}$$

  $\iff$  $$\sum_{y=0}^{\infty} \varphi(y) \left( \frac{n}{n-2} \right)^y \frac{(n-2\theta)^y}{y!} \equiv e^{(n-2)\theta}$$

  $\iff$  $$\sum_{y=0}^{\infty} \varphi(y) \left( \frac{n}{n-2} \right)^y \frac{(n-2\theta)^y}{y!} \equiv \sum_{y=0}^{\infty} \frac{(n-2\theta)^y}{y!}.$$  

  This gives $\varphi(y) = (1 - 2/n)^y$, so that the UMVUE of $\eta$ is given by

  $$\hat{\eta} = \left( \frac{n-2}{n} \right)^{n\bar{X}}.$$  

- Remark: As $n \uparrow \infty$, the UMVUE $\hat{\eta}$ is approximated by the MLE $e^{-2\bar{X}}$, so one may expect that the UMVUE would behave nicely for large $n$. 
CSS and UMVUE: Exponential Model

Let $X_1, \ldots, X_n$ be a random sample from Exponential($\theta$), $\theta > 0$ with pdf $f(x_1; \theta) = \theta^{-1}e^{-x_1/\theta}$. By Factorization theorem, $Y = \sum_{i=1}^n X_i$ is an SS. To check whether it is also complete for $\theta > 0$, suppose that $E_{\theta} \varphi(Y) = 0$ for all $\theta > 0$ for a function $\varphi$. Then, since $Y \overset{d}{=} \text{Gamma}(n, \theta)$, we have

$$
\int_0^\infty \varphi(y) \frac{y^{n-1}}{\Gamma(n)\theta^n} e^{-y/\theta} \, dy \, \theta \geq 0 \equiv 0
$$

$$
\Rightarrow \int_0^\infty \varphi(y) y^{n-1} e^{-\lambda y} \, dy \, \lambda \geq 0 \equiv 0
$$

$$
\Rightarrow \varphi(y) y^{n-1} \overset{y > 0}{=} 0 \quad \text{(uniqueness of Laplace transform)}
$$

$$
\Rightarrow \varphi(y) \overset{y > 0}{=} 0.
$$

Since $E_{\theta}(Y/n) \overset{\theta}{=} \theta$, we conclude that $\hat{\eta} = \bar{X}$ is the UMVUE of $\theta$. 
CSS and UMVUE: Uniform\([0, \theta]\) Model

Let \(X_1, \ldots, X_n\) be a random sample from Uniform\([0, \theta]\), \(\theta > 0\). By Factorization theorem, it can be verified that \(Y = X_{(n)}\) is an SS. To check whether it is also complete for \(\theta > 0\), suppose that \(E_\theta \varphi(Y) = 0\) for all \(\theta > 0\) for a function \(\varphi\). Then, since the pdf of \(Y\) is given by

\[
\text{pdf}_Y(y; \theta) = n\theta^{-n}y^{n-1}I_{[0, \theta]}(y),
\]

we have

\[
\int_0^\theta \varphi(y) n\theta^{-n}y^{n-1} \, dy \equiv 0 \quad \theta > 0
\]

\[
\Rightarrow \int_0^\theta \varphi(y) y^{n-1} \, dy \equiv 0
\]

\[
\Rightarrow \varphi(\theta) \theta^{n-1} \theta > 0 \equiv 0 \quad \text{(Fundamental Theorem of Calculus)}
\]

\[
\Rightarrow \varphi(y) y > 0 \equiv 0.
\]

Since \(E_\theta(Y) \equiv n\theta/(n + 1)\), we conclude that \(\hat{\eta} = (n+1)X_{(n)}/n\) is the UMVUE of \(\theta\).
CSS and UMVUE: Uniform\([-\theta, \theta]\) Model

Let \(X_1, \ldots, X_n (n \geq 2)\) be a random sample from Uniform\([-\theta, \theta]\), \(\theta > 0\).

- By Factorization theorem, it can be verified that \((X_1, X_n)\) is an SS. But, it is not a complete statistic for \(\theta > 0\). One may find a non-trivial function of \((X_1, X_n)\) such that \(E_{\theta} \varphi(X_1, X_n)\) equals identically zero. For example, \(\varphi(X_1, X_n) = X_n/X_1 - a_n\), where \(a_n = E(U[n]/U_1)\) for a random sample \((U_i : 1 \leq i \leq n)\) from Uniform\([-1, 1]\).

- In fact, for this model, \(Y = \max_{1 \leq i \leq n} |X_i|\) is a CSS and 
  \(\hat{\eta} = (n + 1) \max_{1 \leq i \leq n} |X_i|/n\) is the UMVUE of \(\theta\). To see this, note that the joint density of \(X\) at \(x\) equals \((2\theta)^{-n} I_{[0, \theta]}(\max_{1 \leq i \leq n} |x_i|)\), which tells that \(Y\) is an SS. Since \(|X_i|\) are i.i.d. Uniform\([0, \theta]\), one may then show that \(Y\) is also complete and thus \(\hat{\eta}\) is the UMVUE.
Ancillary Statistic

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf $f(\cdot; \theta)$, $\theta \in \Theta \subset \mathbb{R}^d$.

- **Ancillary statistic**: A statistic $Z = v(X)$ is called an ancillary statistic for $\theta \in \Theta$ if $P_\theta(Z \in A)$ does not depend on $\theta \in \Theta$ for all $A$, i.e.,

  $$P_{\theta_1}(Z \in A) \equiv_{\theta_1, \theta_2 \in \Theta} P_{\theta_2}(Z \in A)$$

  for all $A$.

- **Basu’s Theorem (Independence of CSS and AS)**: If $Y = u(X)$ is a CSS and $Z = v(X)$ is an AS for $\theta \in \Theta$, then $Y$ and $Z$ are independent under $P_\theta$ for all $\theta \in \Theta$, i.e.,

  $$P_{\theta}(Y \in A, Z \in B) \equiv_{A, B} P_{\theta}(Y \in A)P(Z \in B)$$

  for all $\theta \in \Theta$. 

Proof of Basu’s Theorem

Let $A$ and $B$ be arbitrary sets. Since $Y$ is a sufficient statistic for $\theta \in \Theta$, 
$\varphi(y) \equiv P(Z \in B|Y = y)$ does not depend on $\theta \in \Theta$ for all $y$. Also, we note 
that $E_\theta(\varphi(Y)) = P(Z \in B)$, which does not depend on $\theta$ either because of the 
ancillarity of $Z$. Thus,

$$E_\theta (\varphi(Y) - P(Z \in B)) \equiv 0.$$  

Due to the completeness of $Y$, this entails $\varphi(y) \equiv P(Z \in B)$, so that

$$P_\theta(Y \in A, Z \in B) = \int_A P(Z \in B|Y = y) \cdot \text{pdf}_Y(y; \theta) \, d\mu(y)$$

$$= P(Z \in B) \cdot P_\theta(Y \in A)$$

for all $\theta \in \Theta$. 

\[\square\]
Ancillary Statistic: Examples

- \( N(\theta, 1), \theta \in \mathbb{R} \):
  \[
  (X_1 - \bar{X}, \ldots, X_n - \bar{X}) \overset{d}{=} (Z_1 - \bar{Z}, \ldots, Z_n - \bar{Z})
  \]
  for \( Z_i \) being i.i.d. from \( N(0, 1) \).

- \( \text{Exp}(\theta, 1), \theta \in \mathbb{R} \):
  \[
  (X_1 - X_{(1)}, \ldots, X_n - X_{(n)}) \overset{d}{=} (Z_1 - Z_{(1)}, \ldots, Z_n - Z_{(n)})
  \]
  for \( Z_i \) being i.i.d. from \( \text{Exp}(0, 1) \).

- \( \text{Gamma}(\alpha, \beta), \beta > 0 \) with \( \alpha \) known:
  \[
  \left( \frac{X_1}{\sum_{i=1}^{n+1} X_i}, \ldots, \frac{X_n}{\sum_{i=1}^{n+1} X_i} \right) \overset{d}{=} \left( \frac{Z_1}{\sum_{i=1}^{n+1} Z_i}, \ldots, \frac{Z_n}{\sum_{i=1}^{n+1} Z_i} \right)
  \]
  for \( Z_i \) being i.i.d. from \( \text{Gamma}(\alpha, 1) \).
Ancillary Statistic: Examples

▶ Uniform(0, θ), θ > 0: \( X(n)/X_1 \overset{d}{=} Z(n)/Z_1 \) for \( Z_i \) being i.i.d. from Uniform(0, 1).

▶ \( \mathcal{N}(\mu, \sigma^2), (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ : \\
\left( \frac{X_1 - \bar{X}}{S_X}, \ldots, \frac{X_n - \bar{X}}{S_X} \right) \overset{d}{=} \left( \frac{Z_1 - \bar{Z}}{S_Z}, \ldots, \frac{Z_n - \bar{Z}}{S_Z} \right) \)

for \( Z_i \) being i.i.d. from \( \mathcal{N}(0, 1) \), where \( S_X^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1} \) and \( S_Z^2 = \frac{\sum_{i=1}^{n} (Z_i - \bar{Z})^2}{n - 1} \).

▶ \( \mathcal{E}(\mu, \sigma), (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+ : \\
\frac{X_1 - X_1}{\sum_{i=1}^{n} (X_i - X_1)} \overset{d}{=} \frac{Z_1 - Z_1}{\sum_{i=1}^{n} (Z_i - Z_1)} \)

for \( Z_i \) being i.i.d. from \( \mathcal{E}(0, 1) \).
Basu’s Theorem: Examples

Let $X_1, \ldots, X_n$ be a random sample from $\text{Exp}(\mu, \sigma), \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$. Then, $X_{(1)}$ and $\sum_{i=1}^{n} (X_i - X_{(1)})$ are independent.

**Proof:** Fix $\sigma = \sigma_0$. Then, $X_{(1)}$ is a CSS for $\mu \in \mathbb{R}$ and $\sum_{i=1}^{n} (X_i - X_{(1)})$ is an AS for $\mu \in \mathbb{R}$. Thus, they are independent under $P_{\mu, \sigma_0}$ for all $\mu \in \mathbb{R}$. Since this holds for any choice of $\sigma_0$, they are independent under the entire model.

Let $X_1, \ldots, X_{n+1}$ be a random sample from $\text{Gamma}(\alpha, \beta), \alpha > 0, \beta > 0$. Then, $\sum_{i=1}^{n+1} X_i$ is independent of

$$Z \equiv \left( \frac{X_1}{X_1 + \cdots + X_{n+1}}, \cdots, \frac{X_n}{X_1 + \cdots + X_{n+1}} \right).$$

**Proof:** Fix $\alpha = \alpha_0$. Then, $\sum_{i=1}^{n+1} X_i$ is a CSS for $\beta > 0$ (Check this!). Since $Z$ is an AS for $\beta > 0$, the two statistics are independent under $P_{\alpha_0, \beta}$ for all $\beta > 0$, and thus under $P_{\alpha, \beta}$ for all $\alpha > 0$ and $\beta > 0$. 
Use Of Ancillary Statistic To Find UMVUE

Let \( X_1, \ldots, X_n \) be a random sample from \( \text{Exp}(\theta), \theta > 0 \). For a given \( a > 0 \) we want to find the UMVUE of (the system reliability) \( \eta = P_\theta(X_1 > a) = e^{-a/\theta} \).

We have seen that \( Y = \sum_{i=1}^{n} X_i \) is a CSS for \( \theta > 0 \). An easy UE of \( \eta \) is \( \hat{\eta}_0 = I_{(a, \infty)}(X_1) \). Thus, \( \hat{\eta} = E(\hat{\eta}_0|Y) \) is the UMVUE. Now, we note that \( X_1/Y \overset{d}{=} \text{Beta}(1, n-1) \) is an AS, so that it is independent of \( Y \). From this we get that, for \( 0 < a/y < 1 \),

\[
E(\hat{\eta}_0 | Y = y) = P(X_1 > a | Y = y) = P(X_1/Y > a/y | Y = y) = P(X_1/Y > a/y)
\]

\[
= \int_{a/y}^{1} \frac{\Gamma(n)}{\Gamma(1)\Gamma(n-1)} z^0 (1-z)^{n-2} dz
\]

\[
= (1 - a/y)^{n-1}.
\]

In the case where \( a/y \geq 1 \), we have \( P(X_1/Y > a/y) = 0 \). Thus \( \hat{\eta} = (1 - a/Y)^{n-1} I_{(a, \infty)}(Y) \) is the UMVUE.
7.1 Optimality

7.2 Sufficiency and Completeness

7.3 Exponential Family
Exponential Family

A family of distributions \( \{ f(\cdot; \theta) : \theta \in \Theta \} \) for \( \Theta \subset \mathbb{R}^d \) is called exponential family if

(i) the support of the density \( f(\cdot; \theta) \) does not depend on \( \theta \in \Theta \);

(ii) the density has the following form:

\[
f(x; \theta) = \exp(\eta(\theta) \top T(x) - B(\theta)) \cdot h(x)
\]

for some known functions \( \eta = (\eta_1, \ldots, \eta_k) \top, T = (T_1, \ldots, T_k) \top, B \) and \( h \).

An exponential family is called \( k \)-parameter regular exponential family if

(iii) \( \eta(\Theta) \equiv \{ \eta(\theta) : \theta \in \Theta \} \subset \mathbb{R}^k \) contains a \( k \)-dimensional open rectangle.
The $d$-variate real valued function $B$ depends on $\theta$ only through $\eta \equiv \eta(\theta)$, i.e., $B(\theta) = A(\eta(\theta))$ for some $k$-variate real valued function $A$:

$$
1 = \int_X \exp(\eta(\theta)^\top T(x) - B(\theta))h(x) \, d\mu(x)
= e^{-B(\theta)} \int_X \exp(\eta(\theta)^\top T(x))h(x) \, d\mu(x)
\overset{\text{let}}{=} e^{-B(\theta)} \cdot e^{A(\eta(\theta))}.
$$

This means that the density $f(\cdot; \theta)$ actually depends on $\eta$:

$$
f(x; \eta) = \exp(\eta^\top T(x) - A(\eta))h(x).
$$

The parameter $\eta$ is called a \textit{natural parameter}.

\textit{Natural parameter space:} The maximum possible set of $\eta$ is given by

$$
\left\{ \eta \in \mathbb{R}^k : \int_X \exp(\eta^\top T(x))h(x) \, d\mu(x) < \infty \right\}.
$$
Exponential Family: Examples

- **Bernoulli**$(\theta)$, $0 < \theta < 1$: For $x \in \{0, 1\}$,

  \[ f(x; \theta) = \exp \left( x \log \left( \frac{\theta}{1 - \theta} \right) + \log(1 - \theta) \right) = \exp(\eta x - A(\eta)), \]

  where $\eta = \log(\theta/(1 - \theta))$ and $A(\eta) = \log(1 + e^\eta)$. Thus, it is a 1-parameter regular exponential family.

- **$N(\mu, \sigma^2)$**, $\theta \equiv (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$: For $x \in \mathbb{R}$,

  \[ f(x; \theta) = \exp \left( -\frac{1}{2\sigma^2} x^2 + \frac{\mu}{\sigma^2} x - \frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2) \right) \]

  \[ = \exp (\eta_1 T_1(x) + \eta_2 T_2(x) - A(\eta_1, \eta_2)) \cdot (2\pi)^{-1/2}, \]

  where $\eta_1 = -1/(2\sigma^2)$, $\eta_2 = \mu/\sigma^2$, $A(\eta_1, \eta_2) = -\eta_1 \eta_2^2/4 + \log(-\eta_1/2)/2$ and $T_1(x) = x^2$, $T_2(x) = x$. Thus, it is a 2-parameter regular exponential family.
Exponential Family: Examples

▶ Two independent normal populations with a common mean: Let \( X \) and \( Y \) be independent and have \( N(\mu, \sigma_1^2) \) and \( N(\mu, \sigma_2^2) \) distributions, respectively, for \( \theta \equiv (\mu, \sigma_1^2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R}^2_+ \). In this case, for \( (x, y) \in \mathbb{R}^2 \),

\[
f(x, y; \theta) = \exp \left( -\frac{1}{2\sigma_1^2}x^2 + \frac{\mu}{\sigma_1^2}x - \frac{1}{2\sigma_2^2}y^2 + \frac{\mu}{\sigma_2^2}y \\
- \frac{\mu^2}{2\sigma_1^2} - \frac{\mu^2}{2\sigma_2^2} - \frac{1}{2} \log(2\pi\sigma_1^2) - \frac{1}{2} \log(2\pi\sigma_2^2) \right)
\]

\[
= \exp (\eta_1 T_1(x, y) + \eta_2 T_2(x, y)x + \eta_3 T_3(x, y) \\
+ \eta_4 T_4(x, y) - A(\eta_1, \ldots, \eta_4)) \cdot h(x, y),
\]

where \( \eta_1 = -1/(2\sigma_1^2) \), \( \eta_3 = -1/(2\sigma_2^2) \), \( \eta_2 = \mu/\sigma_1^2 \) and \( \eta_4 = \mu/\sigma_2^2 \). In fact, this is not a 4-parameter regular exponential family since \( \eta_4 = (\eta_3/\eta_1) \cdot \eta_2 \).
Random Sample from Exponential Family

Let $X_1, \ldots, X_n$ be a random sample from an exponential family of pdf’s $f(\cdot, \theta)$, $\theta \in \Theta$ with $f(x; \theta) = \exp(\eta(\theta)^\top T(x) - B(\theta)) \cdot h(x)$, where $\eta(\theta)$ is a $k$-vector. Let $\mathcal{X}$ denote the common support of $f(\cdot; \theta)$. Then,

1. the joint densities of $(X_1, \ldots, X_n)$ also form an exponential family with $\mathcal{X}^n$ as the common support and

$$
\prod_{i=1}^n f(x_i; \theta) = \exp \left( \eta(\theta)^\top \sum_{i=1}^n T(x_i) - nB(\theta) \right) \cdot \prod_{i=1}^n h(x_i)
$$

as the joint density;

2. If $\eta(\Theta)$ contains a $k$-dimensional open rectangle, then $Y = \sum_{i=1}^n T(X_i)$ is a CSS for $\theta \in \Theta$. 
Proof of (2)

A proof is given only for the case where $X_i$ are discrete random variables.
Clearly, $Y$ is an SS. Let $\eta = \eta(\theta)$. Then, with $\sum_y^*$ denoting the sum over all $(x_1, \ldots, x_n)$ with $\sum_{i=1}^n T(x_i) = y$, we get

$$pdf_Y(y; \eta) = \sum_y^* \exp \left( \eta^\top \sum_{i=1}^n T(x_i) - nA(\eta) \right) \cdot \prod_{i=1}^n h(x_i)$$

$$= \exp \left( \eta^\top y - nA(\eta) \right) \cdot \sum_y^* \prod_{i=1}^n h(x_i)$$

$$\overset{\text{let}}{=} \exp \left( \eta^\top y - A^*(\eta) \right) h^*(y).$$

Thus, it follows that $E_{\theta} \varphi(Y) \overset{\theta}{=} 0$ implies

$$\sum_{\text{all } y} \varphi(y) h^*(y) \cdot e^{\eta^\top y} \overset{\eta}{=} 0$$

$$\Rightarrow \varphi(y) = 0 \text{ for all } y \text{ with } h^*(y) > 0,$$

where the second implication is from the uniqueness of Laplace transform. \qed
MGF/CGF of $T(X)$ in Exponential Family

Let $X$ be a random variable having a pdf $f(\cdot, \eta)$, $\eta \in \mathcal{N} \subset \mathbb{R}^k$. Assume $f(x; \eta) = \exp(\eta^\top T(x) - A(\eta)) \cdot h(x)$ and that $\mathcal{N}$ contains a $k$-dimensional open rectangle. Then,

(3) the cumulant generating function of $T(X)$ is given by

$$
\text{cgf}_{T(X)}(u; \eta) \equiv \log E \eta e^{u^\top T(X)} = A(\eta + u) - A(\eta)
$$

for all $\eta \in \text{Int}(\mathcal{N})$;

(4) the mean and variance of $T(X)$ under $P_\eta$ with $\eta \in \text{Int}(\mathcal{N})$ are then given by

$$
E_\eta T(X) = \dot{A}(\eta), \quad \text{var}_\eta(T(X)) = \ddot{A}(\eta).
$$
Proofs of (3) and (4)

To prove (3), let \( \eta \in \text{Int}(\mathcal{N}) \). We can find a small \( \epsilon \) such that \( \eta + u \) for all \( u \) with \( \|u\| \leq \epsilon \) belong to \( \mathcal{N} \). For such \( u \), we get

\[
E_{\eta}e^{u^\top T(X)} = \int e^{(\eta+u)^\top T(x)-A(\eta)} h(x) \, d\mu(x)
\]

\[
= e^{A(\eta+u)-A(\eta)} \int e^{(\eta+u)^\top T(x)-A(\eta+u)} h(x) \, d\mu(x)
\]

\[
= e^{A(\eta+u)-A(\eta)}.
\]

The fact (4) is immediate from the fact that

\[
E_{\eta} T(X) = \left. \frac{\partial}{\partial u} \text{cgf}_{T(X)}(u; \eta) \right|_{u=0},
\]

\[
\text{var}_{\eta}(T(X)) = \left. \frac{\partial^2}{\partial uu^\top} \text{cgf}_{T(X)}(u; \eta) \right|_{u=0}.
\]
Differentiation Under Integral Sign

Let $X$ be a random variable having a pdf $f(\cdot, \eta)$, $\eta \in \mathcal{N} \subset \mathbb{R}^k$. Assume $f(x; \eta) = \exp(\eta^\top T(x) - A(\eta)) \cdot h(x)$ and that $\mathcal{N}$ contains a $k$-dimensional open rectangle. Suppose that $E_\eta \phi(X)$ exists for all $\eta \in \mathcal{N}$. Then,

(5) $E_\eta \phi(X)$ as a function of $\eta$ is infinitely many times differentiable at $\eta \in \text{Int}(\mathcal{N})$, and the differentiation can be made under the integral sign.

For example,

$$\frac{\partial}{\partial \eta} E_\eta \phi(X) = \int_X \phi(x) \frac{\partial}{\partial \eta} f(x; \eta) \, d\mu(x).$$

Remark: Applying the theorem to $\phi \equiv 1$ gives

$$0 = E_\eta \left(T(X) - \dot{A}(\eta)\right) \quad \text{and} \quad 0 = E_\eta \left((T(X) - \dot{A}(\eta))^2 - \ddot{A}(\eta)\right),$$

so that $E_\eta T(X) = \dot{A}(\eta)$ and $\text{var}_\eta(T(X)) = \ddot{A}(\eta)$. 

MLE and Exponential Family

Let $X_1, \ldots, X_n$ be a random sample from an exponential family of pdf's $f(\cdot, \eta), \eta \in \mathcal{N} \subset \mathbb{R}^k$ with $f(x; \eta) = \exp(\eta^\top T(x) - A(\eta)) \cdot h(x)$. Assume that $\mathcal{N}$ contains a $k$-dimensional open rectangle. Then,

(6) (i) the log-likelihood is strictly concave, and the unique MLE of $\eta$ is determined by the likelihood equation

$$n^{-1} \sum_{i=1}^{n} T(x_i) = \dot{A}(\eta),$$

provided that it has a solution $\hat{\eta} \in \mathcal{N}$;

(ii) the Fisher information is given by

$$I_1(\eta) = \ddot{A}(\eta).$$
Proof of (6): Positive Definiteness of $\bar{A}(\eta)$

Suppose that there exists $c \neq 0$ in $\mathbb{R}^k$ such that $c^T \bar{A}(\eta_0)c = 0$ for some $\eta_0 \in \mathcal{N}$. Then, $P_{\eta_0}(c^T T(X) = c^T E_{\eta_0} T(X)) = 1$. This implies that

$$c^T T(x) = c^T E_{\eta_0} T(X) \text{ a.e. } [\mu] \text{ for } x \in \mathcal{X} = \{x : h(x) > 0\},$$

i.e.

$$\mu \left( \{x \in \mathcal{X} : c^T T(x) \neq c^T E_{\eta_0} T(X) \} \right) = 0.$$  

Then, there exists a linear map $\ell : \mathbb{R}^k \rightarrow \mathbb{R}^{k-1}$ such that the integral

$$\int_{\mathcal{X}} \exp(\eta^T T(x)) h(x) \, d\mu(x)$$

is actually a function of $\ell(\eta)$. This means that the natural parameter space does not contain an open rectangle in $\mathbb{R}^k$, so that $\mathcal{N}$, a subset of the natural parameter space, does not contain an open rectangle, either. Contradiction.
Multinomial Experiments

Let $X_i = (X_{i,1}, \cdots, X_{i,k-1})^\top$ be i.i.d. Multinomial$(1, p)$,

$p \equiv (p_1, \ldots, p_{k-1})^\top$, $p_j > 0$, $p_1 + \cdots + p_{k-1} < 1$.

Let $p_k = 1 - p_1 - \cdots - p_{k-1}$. Then, the common density of $X_i$ is given by

$$f(x; p) = \exp \left( x_1 \log(p_1/p_k) + \cdots + x_{k-1} \log(p_{k-1}/p_k) + \log p_k \right),$$

so that the distributions of $X_i$ form a $(k-1)$-parameter regular exponential family.

$Y = \sum_{i=1}^n X_i = (\sum_{i=1}^n X_{i,1}, \ldots, \sum_{i=1}^n X_{i,k-1})^\top$ is a CSS for $p$.

MLE of $\eta$: The MLE of $\eta \equiv (\log(p_1/p_k), \ldots, \log(p_{k-1}/p_k))^\top \let h(p)$ solves the equation

$$Y/n = E_\eta X_1, \quad \text{i.e.,} \quad Y/n = h^{-1}(\eta).$$

Thus, the MLE of $\eta$ is given by $\hat{\eta} = h(Y/n)$. 
Multinomial Experiments

- **MLE of \( p \):** The MLE of \( p \) is then \( \hat{p} = h^{-1}(\hat{\eta}) = Y/n. \)

- **UMVUE of \( p \):** \( Y/n \) is an UE of \( p \) and is a function of the CSS \( Y \), so that \( \hat{p} = Y/n \) is also the UMVUE of \( p \).

- **UMVUE of \( \Sigma \equiv \text{diag}(p) - pp^\top \):** Here, an estimator \( \hat{\Sigma} \) of \( \Sigma \) is called the UMVUE of \( \Sigma \) if \( \text{var}_p(\hat{\Sigma}^{UE}) - \text{var}_p(\hat{\Sigma}) \) is nonnegative definite for all \( \hat{\Sigma}^{UE} \) and for all \( p \), with \( \hat{\Sigma} \) and \( \hat{\Sigma}^{UE} \) being the vectorized versions. Note that

\[
\hat{\Sigma}^{MLE} = \text{diag}(Y/n) - (Y/n)(Y/n)^\top.
\]

Computing the expected value of \( \hat{\Sigma}^{MLE} \), we get

\[
E_p(\hat{\Sigma}^{MLE}) = \text{diag}(p) - \text{var}_p(Y/n) - E_p(Y/n)E_p(Y/n)^\top
= \text{diag}(p) - n^{-1}\Sigma - pp^\top = (1 - 1/n)\Sigma.
\]

Thus, \( \hat{\Sigma}^{UMVUE} = n \cdot \hat{\Sigma}^{MLE} / (n - 1). \)
Multivariate Normal Population

Let $X_i = (X_{i,1}, \cdots, X_{i,k})^\top \ (n \geq 2)$ be i.i.d. Normal($\mu, \Sigma$), $\mu \in \mathbb{R}^k$ and $\Sigma$ in the set of $k \times k$ positive definite matrices.

- With $\theta \equiv (\mu, \Sigma) \in \mathbb{R}^d$ for $d = k + k(k + 1)/2$,
  \[
  f(x; \theta) = \det(2\pi \Sigma)^{-1/2} \exp \left( -\frac{(x - \mu)^t \Sigma^{-1} (x - \mu)}{2} \right)
  = \exp \left( -\text{tr} (\Sigma^{-1} xx^\top) / 2 + \mu^\top \Sigma^{-1} x 
  - \mu^\top \Sigma^{-1} \mu / 2 - \log(\det \pi \Sigma) \right).
  \]

- $Y = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i X_i^\top)$ is a CSS for $\theta$, where $\sum_{i=1}^n X_i X_i^\top$ is understood to be a $k(k + 1)/2$-vector.

- MLE of $\eta \equiv (\Sigma^{-1} \mu, \Sigma^{-1})$: It is the solution of
  \[
  \frac{Y}{n} = E_\eta(X_1, X_1 X_1^\top), \quad \text{i.e.,} \quad \frac{Y}{n} = (\mu, \Sigma + \mu \mu^\top) \overset{\text{let}}{=} g(\mu, \Sigma).
  \]

Let $h(\mu, \Sigma) = (\Sigma^{-1} \mu, \Sigma^{-1})$. Then, $\hat{\eta}^{\text{MLE}} = h \circ g^{-1}(Y/n)$. 

Multivariate Normal Population

- **MLE of \(\mu\) and \(\Sigma\):** The MLE of \((\mu, \Sigma) = h^{-1}(\eta)\) is then

\[
(\hat{\mu}^{\text{MLE}}, \hat{\Sigma}^{\text{MLE}}) = h^{-1} \circ h \circ g^{-1}(Y/n) = g^{-1}(Y/n).
\]

By the definition of the function \(g : \mathbb{R}^d \to \mathbb{R}^d\), this means

\[
n^{-1} \sum_{i=1}^{n} X_i = \hat{\mu}^{\text{MLE}} \quad \text{and} \quad n^{-1} \sum_{i=1}^{n} X_i X_i^\top = \hat{\Sigma}^{\text{MLE}} + \hat{\mu} \hat{\mu}^{\text{MLE} \top},
\]

so that \(\hat{\mu}^{\text{MLE}} = \bar{X}\) and \(\hat{\Sigma}^{\text{MLE}} = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^\top\).

- **UMVUE of \(\mu\) and \(\Sigma\):** Since \((\bar{X}, (n - 1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^\top)\) is a 1-1 function of \(Y\), it is also a CSS for \((\mu, \Sigma)\). Since \(\bar{X}\) and \((n - 1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^\top\) are UEs of \(\mu\) and \(\Sigma\), respectively, they are the UMVUE of the respective parameters.
UMVUE of Normal Probabilities

Let \( X_1, \ldots, X_n \ (n \geq 2) \) be a random sample from \( N(\theta, 1) \), \( \theta \in \mathbb{R} \). We want to find the UMVUE of the normal probabilities \( \eta = P_\theta(X_1 \leq c) = \Phi(c - \theta) \). Note that \( \bar{X} \) is a CSS for this model. An easy UE of \( \eta \) is \( I_{(-\infty, c]}(X_1) \). Thus, by R-B-L-S Theorem, the UMVUE of \( \eta \) is given by

\[
\hat{\eta} = E(I_{(-\infty, c]}(X_1) \mid \bar{X}) = P(X_1 \leq c \mid \bar{X}).
\]

We compute \( P(X_1 \leq c \mid \bar{X} = y) \) below. Recall that \( (X_1 - \bar{X}, \ldots, X_n - \bar{X}) \) is an ancillary statistic and thus independent of \( \bar{X} \) (Basu Theorem).

\[
P(X_1 \leq c \mid \bar{X} = y) = P(X_1 - \bar{X} \leq c - y \mid \bar{X} = y)
\]

\[
= P(X_1 - \bar{X} \leq c - y)
\]

\[
= \Phi \left( \sqrt{\frac{n}{n-1}}(c - \bar{X}) \right).
\]

Thus, the UMVUE of \( \eta = \Phi(c - \theta) \) is given by

\[
\hat{\eta} = \Phi \left( \sqrt{\frac{n}{n-1}}(c - \bar{X}) \right).
\]
Independence of Normal Sample Mean and Variance

Let \( X_1, \ldots, X_n \) \((n \geq 2)\) be a random sample from \( N(\mu, \sigma^2) \), \( \theta \equiv (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ \). Then, \( \bar{X} \) and \( \sum_{i=1}^{n} (X_i - \bar{X})^2 \) are independent under \( P_\theta \) for all \( \theta \in \mathbb{R} \times \mathbb{R}_+ \).

**Proof:** Fix \( \sigma^2 = \sigma_0^2 \). Then, \( \bar{X} \) is a CSS and \((X_1 - \bar{X}, \ldots, X_n - \bar{X})\) is an ancillary statistic for the submodel \( \Theta(\sigma_0^2) \equiv \{(\mu, \sigma_0^2) : \mu \in \mathbb{R}\} \). Thus, they are independent under \( P_\theta \) for all \( \theta \in \Theta(\sigma_0^2) \), due to the Basu’s theorem. Since the choice \( \sigma_0^2 \) is arbitrary within \( \mathbb{R}_+ \), this implies that \( \bar{X} \) and \((X_1 - \bar{X}, \ldots, X_n - \bar{X})\) are independent under \( P_\theta \) for all

\[
\theta \in \bigcup_{\sigma_0^2 \in \mathbb{R}_+} \Theta(\sigma_0^2) = \mathbb{R} \times \mathbb{R}_+.
\]

The independence of \( \bar{X} \) and \( \sum_{i=1}^{n} (X_i - \bar{X})^2 \) is immediate from this. \( \square \)