Sufficiency

So far, we've come up with some unbiased estimators

Ex: $\bar{Y}$, $S^2$ for $\mu$, $\sigma^2$, respectively

Q: In this process of summarizing random samples as estimated parameters $\mu$ and $\sigma^2$, have we retained all the information about $\mu$ and $\sigma^2$, or have we lost some information through this process of reducing the data?

i.e. are $\bar{Y}$ and $S^2$ "sufficient" for $\mu$ and $\sigma^2$? Or can we get more information from other functions of random samples?

Defn: Let $Y_1, \ldots, Y_n$ denote a random sample from a probability distribution with unknown parameter $\theta$. Then the statistic $U = g(Y_1, \ldots, Y_n)$ is said to be sufficient for $\theta$ if the conditional distribution of $Y_1, \ldots, Y_n$ given $U$ does not depend on $\theta$.

Ex 12: Let $Y$ be a binomial RV. Show that $Y$ is sufficient for $p$.

Recall $Y$ can be thought of as $Y = Y_1 + \cdots + Y_n$ where

$Y_i = 1$ if $i$th trial is success $\Leftrightarrow \Pr = p$

$0$ if $i$th trial is failure $\Leftrightarrow \Pr = 1-p$

Let $y = y_1 + \cdots + y_n$

$P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n \mid Y = y) = P(Y_1 = y_1, \ldots, Y_n = y_n \mid Y = y)$

$\frac{P(Y = y)}{C(\theta)^{n-y}}$ ways of having $y$ successes, $n-y$ failures

$\frac{P(Y = y)}{n}$ ways of having $y$ successes, $n-y$ failures

$= \frac{1}{C(n)}$ if $y = y_1 + \cdots + y_n$

Since this does not depend on $p$, then $Y$ is sufficient for $p$.

(Note: this also means $\frac{X}{n}$ is sufficient for $\mu$ in 9.5, show $\frac{X}{n}$ is a "MVE" for $\mu$)

Recall we write joint probability (mass) function as $p(y_1, \ldots, y_n)$ for RVs $Y_1, \ldots, Y_n$ and joint probability density function as $f(y_1, \ldots, y_n)$ for RVs $Y_1, \ldots, Y_n$.

We now want to make this dependence on parameter explicit, so we write $p(y_1, \ldots, y_n \mid \theta)$ and $f(y_1, \ldots, y_n \mid \theta)$. Have single name for both cases.
Defn: let $y_1, \ldots, y_n$ be sample observations taken at corresponding random variables $Y_1, \ldots, Y_n$ whose distribution depends on a parameter $\theta$. Then, if $Y_1, \ldots, Y_n$ are discrete RVs, the **likelihood** of the sample, $L(y_1, \ldots, y_n | \theta)$, is defined to be the joint probability of $y_1, \ldots, y_n$. If $Y_1, \ldots, Y_n$ are continuous RVs, the **likelihood** $L(y_1, \ldots, y_n | \theta)$ is defined to be the joint density evaluated at $y_1, \ldots, y_n$.

Note: Since we will always be considering the case where $Y_1, \ldots, Y_n$ are a random sample, then $Y_1, \ldots, Y_n$ are independent, so

$$p(y_1, \ldots, y_n | \theta) = p(y_1 | \theta) \cdot p(y_2 | \theta) \cdot \ldots \cdot p(y_n | \theta)$$

This makes computation easier!

Thm: Let $U$ be a statistic based on the random sample $Y_1, \ldots, Y_n$. Then $U$ is a **sufficient statistic** for the estimation of a parameter $\theta$ if and only if the likelihood $L(y_1, \ldots, y_n | \theta)$ can be factored into two nonnegative functions

$$L(y_1, \ldots, y_n | \theta) = g(U, \theta) \cdot h(y_1, \ldots, y_n)$$

where $U$ is a function of $\theta$ and only not a function of $\theta$.

Ex 12': Show $Y$ is a sufficient statistic for $p$ using this theorem, where $Y$ is a binomial RV.

Let $Y = Y_1 + \ldots + Y_n$

$$L(y_1, \ldots, y_n | \theta) = p(y_1, \ldots, y_n | \theta) = p(Y_1 | \theta) \cdot p(Y_2 | \theta) \cdot \ldots \cdot p(Y_n | \theta)$$

Recall for Bernoulli RV, $p(Y_i = 1) = p$ & $p(Y_i = 0) = 1 - p$, so $p(y_i) = p^{y_i} (1-p)^{1-y_i}$ for $y_i = 0, 1$

So $p(y_i | \theta) = p^{y_i} (1-p)^{1-y_i}$

So $L(y_1, \ldots, y_n | \theta) = p^{y_1} (1-p)^{1-y_1} \cdot p^{y_2} (1-p)^{1-y_2} \cdot \ldots \cdot p^{y_n} (1-p)^{1-y_n}$

$$= p^{y_1 + \ldots + y_n} (1-p)^{n - (y_1 + \ldots + y_n)}$$

$$= (p^{y_1} (1-p)^{1-y_1}) \cdot (p^{y_2} (1-p)^{1-y_2}) \cdot \ldots \cdot (p^{y_n} (1-p)^{1-y_n})$$

So $U = Y_1 + \ldots + Y_n$ is sufficient for $p$. 


Ex 13: Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$.

a.) If $\mu$ is unknown & $\sigma^2$ is known, show that $\overline{Y}$ is sufficient for $\mu$.

$$L(y_1, \ldots, y_n | \mu) = f(y_1, y_2, \ldots, y_n | \mu) = \frac{1}{\sigma \sqrt{2\pi}} \cdot e^{-\frac{(y_1-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma \sqrt{2\pi}} \cdot e^{-\frac{(y_2-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sigma \sqrt{2\pi}} \cdot e^{-\frac{(y_n-\mu)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^{n} (y_i - \mu)^2 \right)} = \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^{n} y_i^2 - 2\mu \sum_{i=1}^{n} y_i + n\mu^2 \right)}$$

$$= \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \cdot e^{-\frac{n\mu^2}{2\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^{n} y_i^2 \right)} \cdot \left(1 + \frac{1}{\sigma^2} \sum_{i=1}^{n} y_i^2 \right)$$

$$= \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \cdot e^{-\frac{n\mu^2}{2\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^{n} y_i^2 \right) - \frac{n\mu^2}{2\sigma^2}} \cdot \left(1 + \frac{1}{\sigma^2} \sum_{i=1}^{n} y_i^2 \right)$$

$$\left\{ \begin{array}{l}
g(z_{yi} | \mu) \\
h(y_1, \ldots, y_n) \end{array} \right\}$$

So $\overline{Z} Y_i$ is suff for $\mu$.

b.) If $\mu$ is known & $\sigma^2$ is unknown, show that $\sum_{i=1}^{n} (y_i - \mu)^2$ is not suff for $\sigma^2$.

$$L(y_1, \ldots, y_n | \sigma^2) = \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \cdot e^{-\frac{n\mu^2}{2\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^{n} y_i^2 \right) - \frac{n\mu^2}{2\sigma^2}} \cdot \left(1 + \frac{1}{\sigma^2} \sum_{i=1}^{n} y_i^2 \right)$$

$$= \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \cdot e^{-\frac{n\mu^2}{2\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^{n} y_i^2 \right)} \cdot \left(1 + \frac{1}{\sigma^2} \sum_{i=1}^{n} y_i^2 \right)$$

Function of $\sigma^2$!!!

c.) If $\mu$ is known & $\sigma^2$ is unknown, show that $\sum_{i=1}^{n} (y_i - \mu)^2$ is suff for $\sigma^2$.

$$L(y_1, \ldots, y_n | \sigma^2) = \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^{n} (y_i - \mu)^2 \right) - \frac{n\mu^2}{2\sigma^2}}.$$

$$\left\{ \begin{array}{l}
g(z_{yi}^2, \sigma^2) \\
\end{array} \right\}$$

d.) If $\mu$ & $\sigma^2$ are unknown, show that $\overline{Z} Y_i$ & $\sum_{i=1}^{n} (y_i - \mu)^2$ are jointly suff for $\mu$ & $\sigma^2$.

$$L(y_1, \ldots, y_n | \mu, \sigma^2) = \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \cdot e^{-\frac{n\mu^2}{2\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2} \left( \sum_{i=1}^{n} y_i^2 \right) - \frac{n\mu^2}{2\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} y_i^2}$$

$$\left\{ \begin{array}{l}
g(z_{yi}, z_{yi}^2, \mu, \sigma^2) \\ \end{array} \right\}$$
9.5 Minimum-Variance Unbiased Estimators

In last section, we found sufficient statistics. How do we find the "best" sufficient statistic? Clearly, we want minimum variance.

Main Thm of this section (Rao-Blackwell Thm) says (paraphrasing):

- The factorization criterion from 9.4 can be applied to find sufficient statistics that best summarize the information contained in sample data about parameters of interest. Call this U.
- A "minimum-variance unbiased estimator" (MVUE) is some function of U, h(U), such that $E(h(U)) = \theta$.

Ex 12": Find an MVUE for $p$.
From Ex 12': $L(y_1, \ldots, y_n | \theta) = \frac{y_1 \cdots y_n}{p} (1-p)^{n-y}$

So $U = \sum_{i=1}^{n} y_i$ is sufficient for $p$ and best summarizes the information about $p$.

Since $E(U) = np \rightarrow E\left(\frac{U}{n}\right) = \frac{np}{n} = p$

So $\frac{U}{n} = \bar{Y}$ is an unbiased estimator for $p$.

Since $h(U) = \frac{U}{n}$ is a function of U, then $\hat{p} = \frac{U}{n}$ is the MVUE for $p$.

Ex 13': Find the MVUEs for $\mu$ and $\sigma^2$.
From Ex 13: $L(y_1, \ldots, y_n | \theta) \propto \left(\frac{1}{\sigma \sqrt{2\pi}}\right)^n \frac{1}{\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right)$

So $\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n} y_i^2$ are jointly sufficient for $\mu$ and $\sigma^2$, and these best summarize the information about $\mu$ & $\sigma^2$.

Since $E(\bar{Y}) = \mu$ & $E(S^2) = \sigma^2$ and $\bar{Y}$ & $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{Y})^2$

are functions of $\sum_{i=1}^{n} y_i$ & $\sum_{i=1}^{n} y_i^2$, then $\bar{Y}$ & $S^2$ are MVUEs for $\mu$ & $\sigma^2$. 
The Method of Moments

- We derived unbiased point estimators in Sec 9.5 by taking the sufficient statistics found from the Factorization Thm & finding some function of them to be unbiased point estimators.
- But how do we find that function, \( h(W) \), in general? Could be really complicated!
- In Sec 9.6 & 9.7, learn 2 other methods to find unbiased point estimators. May not be minimum variance, but we can try to scale them to make them into MVUEs.

Recall the \( k^{th} \) moment taken about the origin is:

\[
M_k = E(Y^k)
\]

The \( k^{th} \) sample moment is the average:

\[
m_k' = \frac{1}{n} \sum_{i=1}^{n} Y_i^k
\]

Method of Moments: Main idea behind this method is that sample moments should provide good estimators for corresponding population moments, i.e. \( m_k' \) should be a good estimator of \( M_k \).

- Solve \( m_k' = M_k' \) for \( k = 1, 2, \ldots, t \), where \( t = \# \) of parameter to be estimated (We are solving for parameters).

Recall Bernoulli RV: \( p(Y) = p^y(1-p)^{1-y} \) for \( y = 0, 1 \) has 1 parameter: \( p \)

Normal RV: \( f(y) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \) has 2 parameters: \( \mu \& \sigma^2 \)

Ex 12**: Use the method of moments to estimate \( p \).

\[
m_1' = \mu = p
\]

\[
m_1' = \frac{1}{n} \sum_{i=1}^{n} Y_i = \bar{Y}
\]

So \( m_1' = m_1' \Rightarrow p = \bar{Y} \), i.e. \( \hat{p} = \frac{Y_1 + \ldots + Y_n}{n} \) or \( \bar{Y} \)

Ex 13**: Use the method of moments to estimate \( \mu \& \sigma^2 \).

\[
m_1' = \mu
\]

\[
m_1' = \frac{1}{n} \sum_{i=1}^{n} Y_i = \bar{Y} \quad \text{so} \quad m_1' = m_1' \Rightarrow \mu = \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i
\]

\[
m_2' = E(Y^2) = \mu^2 + \sigma^2
\]

\[
m_2' = \frac{1}{n} \sum_{i=1}^{n} Y_i^2 \quad \text{so} \quad m_2' = m_2' \Rightarrow \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} Y_i^2 \Rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \mu^2
\]

So \( \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - (\bar{Y})^2 = \frac{1}{n} \left( \sum_{i=1}^{n} Y_i^2 - n(\bar{Y})^2 \right) \quad \text{Note:} \frac{1}{n} \text{, not} \frac{1}{n-1} \)