Lecture 5: Functions of sufficient and complete statistics

Approach 3: using functions of sufficient and complete statistics

If there is a complete and sufficient statistic T, there are two typical ways to derive a UMVUE using Lehmann- Scheffé Theorem.

Solving for the function ψ

The first one is solving for ψ when the distribution of T is available. The following are two typical examples.

Example (uniform family).

Let $X_1,...,X_n$ be iid from *uniform*$(0,\theta)$ with unknown $\theta > 0$.
Suppose that g is a differentiable function on $(0,\infty)$.
Since the sufficient and complete statistic $T = X_{(n)}$ has pdf $n\theta^{-n}x^{n-1}$, $0 < x < \theta$, an unbiased estimator $\psi(X_{(n)})$ of $g(\theta)$ must satisfy

$$\theta^n g(\theta) = n\int_0^\theta \psi(x)x^{n-1}dx \qquad \text{for all } \theta > 0$$

Differentiating both sizes of the previous equation leads to

$$n\theta^{n-1}g(\theta) + \theta^n g'(\theta) = n\psi(\theta)\theta^{n-1}$$

Hence, the UMVUE of $g(\theta)$ is $\psi(X_{(n)}) = g(X_{(n)}) + n^{-1}X_{(n)}g'(X_{(n)})$.

Example (Poisson family).

Let $X_1, ..., X_n$ be iid from $Poisson(\theta)$ with unknown $\theta > 0$.
Then $T = \sum_{i=1}^n X_i$ is sufficient and complete for θ and has $Poisson(n\theta)$ distribution.
Suppose that g is a smooth function such that $g(x) = \sum_{j=0}^\infty a_j x^j$, $x > 0$.
An unbiased estimator $\psi(T)$ of $g(\theta)$ must satisfy

$$\sum_{t=0}^\infty \frac{\psi(t)n^t}{t!}\theta^t = e^{n\theta}g(\theta) = \sum_{k=0}^\infty \frac{n^k}{k!}\theta^k \sum_{j=0}^\infty a_j\theta^j$$

$$= \sum_{t=0}^\infty \left(\sum_{j,k:j+k=t} \frac{n^k a_j}{k!} \right) \theta^t \qquad \theta > 0$$

Thus, a comparison of coefficients in front of θ^t leads to

$$\psi(t) = \frac{t!}{n^t} \sum_{j,k:j+k=t} \frac{n^k a_j}{k!},$$

i.e., $\psi(T)$ is the UMVUE of $g(\theta)$.
In particular, if $g(\theta) = \theta^r$ for some fixed integer $r \geq 1$, then $a_r = 1$ and $a_k = 0$ if $k \neq r$ and

$$\phi(t) = \begin{cases} 0 & t < r \\ \frac{t!}{n^r (t-r)!} & t \geq r. \end{cases}$$

Example (normal family).

Let $X_1, ..., X_n$ be iid $N(\mu, \sigma^2)$ with unknown $\theta = (\mu, \sigma^2) \in \mathscr{R} \times (0, \infty)$.
$T = (\bar{X}, S^2)$ is sufficient and complete for θ and $\bar{X}$ and $(n-1)S^2/\sigma^2$ are independent and have the $N(\mu, \sigma^2/n)$ and chi-square distribution with degrees of freedom $n-1$, respectively.
Using the method of solving for ψ directly, we find that the UMVUE for μ is $\bar{X}$; the UMVUE of μ^2 is $\bar{X}^2 - S^2/n$; the UMVUE for σ^r with $r > 1-n$ is $k_{n-1,r} S^r$, where

$$k_{n,r} = \frac{n^{r/2} \Gamma(n/2)}{2^{r/2} \Gamma\left(\frac{n+r}{2}\right)}$$

(in particular, the UMVUE of σ^2 is S^2, which is a conclusion we cannot get in Example 7.3.14), and the UMVUE of μ/σ is $k_{n-1,-1}\bar{X}/S$.

Suppose that $g(\theta)$ satisfies $P(X_1 \le g(\theta)) = p$ with a fixed $p \in (0,1)$.

Let Φ be the cdf of the standard normal distribution.

Then $g(\theta) = \mu + \sigma\Phi^{-1}(p)$ and its UMVUE is $\bar{X} + k_{n-1,1}S\Phi^{-1}(p)$.

Conditioning

The second method of deriving a UMVUE when there is a sufficient and complete statistic T is conditioning on T, i.e., if W is any unbiased estimator of $g(\theta)$, then $E(W|T)$ is the UMVUE of $g(\theta)$.

To apply this method, we do not need the distribution of T, but need to work out the conditional expectation $E(W|T)$.

From the uniqueness of the UMVUE, it does not matter which W is used and, thus, we should choose W so as to make the calculation of $E(W|T)$ as easy as possible.

Example 7.3.24 (binomial family)

Let $X_1, ..., X_n$ be iid from *binomial*(k, θ) with known k and unknown $\theta \in (0, 1)$.

We want to estimate $g(\theta) = P_\theta(X_1 = 1) = k\theta(1-\theta)^{k-1}$.

Note that $T = \sum_{i=1}^n X_i \sim binomial(kn, \theta)$ is the sufficient and complete statistic for θ.

But no unbiased estimator based on it is immediately evident.

To apply conditioning, we take the simple unbiased estimator of $P_\theta(X_1 = 1)$, the indicator function $I(X_1 = 1)$.

By Theorem 7.3.23, the UMVUE of $g(\theta)$ is

$$\psi(T) = E[I(X_1 = 1)|T) = P(X_1 = 1|T)$$

We need to simply $\psi(T)$ and obtain an explicit form.

For $t = 1, ..., kn$,

$$\begin{aligned}\psi(t) &= P(X_1 = 1|T = t) = \frac{P_\theta(X_1 = 1, \sum_{i=1}^n X_i = t)}{P_\theta(\sum_{i=1}^n X_i = t)} \\ &= \frac{P_\theta(X_1 = 1, \sum_{i=2}^n X_i = t-1)}{P_\theta(\sum_{i=1}^n X_i = t)}\end{aligned}$$

$$= \frac{P_\theta(X_1 = 1) P_\theta(\sum_{i=2}^n X_i = t-1)}{P_\theta(\sum_{i=1}^n X_i = t)}$$

$$= \frac{k\theta(1-\theta)^{k-1}\left[\binom{k(n-1)}{t-1}\theta^{t-1}(1-\theta)^{k(n-1)-(t-1)}\right]}{\binom{kn}{t}\theta^t(1-\theta)^{kn-t}}$$

$$= \frac{k\binom{k(n-1)}{t-1}}{\binom{kn}{t}}$$

When $T = 0$, $P(X_1 = 1 | T = 0) = 0$.
Hence, the UMVUE of $g(\theta) = k\theta(1-\theta)^{k-1}$ is

$$\psi(T) = \begin{cases} \frac{k\binom{k(n-1)}{T-1}}{\binom{kn}{T}} & T = 1, ..., kn \\ 0 & T = 0 \end{cases}$$

Example (exponential distribution family)

Let $X_1, ..., X_n$ be iid with pdf $\theta^{-1}e^{-x\theta}$, $x > 0$, where $\theta > 0$ is unknown.
Let $t > 0$ and the parameter of interest to be $g(\theta) = P_\theta(X_1 > t)$.

Since $\bar{X}$ is sufficient and complete for $\theta > 0$ and the indicator $I(X_1 > t)$ is unbiased for $g(\theta)$,

$$\psi(\bar{X}) = E[I(X_1 > t)|\bar{X}] = P(X_1 > t|\bar{X})$$

is the UMVUE of $g(\theta)$.

If the conditional distribution of X_1 given $\bar{X}$ is available, then we can calculate $P(X_1 > t|\bar{X})$ directly.

But the following technique can be applied to avoid the derivation of conditional distributions.

By Basu's theorem, $X_1/\bar{X}$ and $\bar{X}$ are independent.

Then

$$\begin{aligned} P(X_1 > t|\bar{X} = \bar{x}) &= P(X_1/\bar{X} > t/\bar{X}|\bar{X} = \bar{x}) \\ &= P(X_1/\bar{X} > t/\bar{x}|\bar{X} = \bar{x}) \\ &= P(X_1/\bar{X} > t/\bar{x}) \end{aligned}$$

To compute this unconditional probability, we need the distribution of

$$X_1 \Big/ \sum_{i=1}^n X_i = X_1 \Big/ \left(X_1 + \sum_{i=2}^n X_i \right)$$

Using the transformation technique discussed earlier and the fact that $\sum_{i=2}^{n} X_i$ is independent of X_1 and has a gamma distribution, we obtain that $X_1 / \sum_{i=1}^{n} X_i$ has pdf $(n-1)(1-x)^{n-2}$, $0 < x < 1$ (a beta pdf).

$$P(X_1 > t | \bar{X} = \bar{x}) = (n-1) \int_{t/(n\bar{x})}^{1} (1-x)^{n-2} dx = \left(1 - \frac{t}{n\bar{x}}\right)^{n-1}$$

Hence the UMVUE of $g(\theta)$ is

$$T(X) = \left(1 - \frac{t}{n\bar{X}}\right)^{n-1}$$

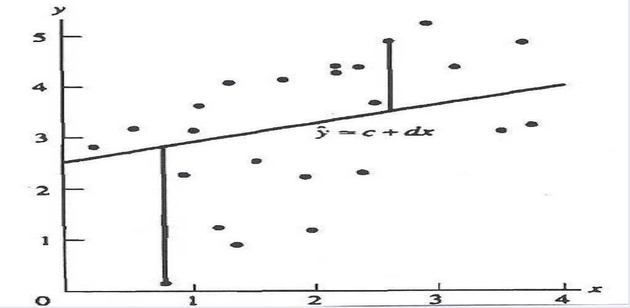
Simple linear regression

Simple linear regression studies the relationship between a variable of interest Y_i (often called response or dependent variable) and a univariate covariate X_i (also called auxiliary variable, explanatory variable, or independent variable), when the following simple linear regression model is assumed:

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \qquad i = 1, ..., n,$$

where Y_i is a random response, x_i is the univariate covariate, which is either a deterministic value or the observed value of a random variable, in which case our analysis is conditional on $x_1, ..., x_n$, $\alpha \in \mathscr{R}$ and $\beta \in \mathscr{R}$ are unknown intercept and slope, respectively, ε_i's are measurement errors and are independent random variables with mean 0 and a finite common unknown variance $\sigma^2 > 0$, and $n \geq 3$ is the sample size.

An example of a set of observed (y_i, x_i)'s is shown in the next figure.



- Data are from Table 11.3.1.
- Points should be on a line if there is no error ($\varepsilon_i = 0$ for all i).
- $\widehat{y} = c + dx$ is an estimate of $y = \alpha + \beta x$ that generates the data.

The MLE and UMVUE

Under the additional assumption that ε_i's are iid $N(0,\sigma^2)$, the likelihood function is

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\sum_{i=1}^{n} \frac{(y_i-\alpha-\beta x_i)^2}{2\sigma^2}\right)$$

Maximizing this likelihood is equivalent to minimizing

$$\psi(a,b) = \sum_{i=1}^{n}(y_i - a - bx_i)^2 \qquad \text{over } a \text{ and } b$$

Consider $(y_1,x_1),...,(y_n,x_n)$ as n pairs of numbers plotted in a scatterplot as in the previous figure.

Think of drawing through this cloud of points a straight line that comes “as close as possible” to all the points, measured by the vertical distances from the points to the straight line.

For any line $y = a + bx$, the squared distances is $\psi(a,b)$.

The MLE $(\widehat{\alpha},\widehat{\beta})$ is the point that minimizes $\psi(a,b)$ over a and b.

Because of this, the estimator $(\widehat{\alpha},\widehat{\beta})$ is also called the least squares estimator (LSE).

Consider

$$\frac{\partial \psi(a,b)}{\partial a} = -2\sum_{i=1}^{n}(y_i - a - bx_i) = 0$$
$$\frac{\partial \psi(a,b)}{\partial b} = -2\sum_{i=1}^{n}x_i(y_i - a - bx_i) = 0$$

The first equation is

$$\bar{y} - a - b\bar{x} = 0 \quad \text{iff} \quad a = \bar{y} - b\bar{x}$$

Substituting a in the second equation by $\bar{y} - b\bar{x}$ results in

$$\sum_{i=1}^{n}x_i(y_i - \bar{y}) - b\sum_{i=1}^{n}x_i(x_i - \bar{x}) = 0$$

This equation is the same as $S_{xy} = bS_{xx}$, where

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \quad S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

Therefore, replacing y_i by the random variable Y_i for all i (and we still use S_{xy} when y_i is replaced by Y_i), we obtain the MLE or LSE as

$$\widehat{\beta} = \frac{S_{xy}}{S_{xx}}, \qquad \widehat{\alpha} = \bar{Y} - \widehat{\beta}\bar{x} = \bar{Y} - \frac{S_{xy}}{S_{xx}}\bar{x}$$

We can always assume that $S_{xx} > 0$, since $S_{xx} = 0$ is the trivial case of identical x_i's.

We now show that $\widehat{\alpha}$ and $\widehat{\beta}$ are UMVUE's of α and β, respectively.

First, we show that they are unbiased estimators.

$$E(S_{xy}) = \sum_{i=1}^{n}(x_i - \bar{x})E(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})\beta(x_i - \bar{x}) = \beta S_{xx}$$

$\widehat{\beta}$ is unbiased for β and

$$E(\widehat{\alpha}) = E(\bar{Y}) - E(\widehat{\beta})\bar{x} = \alpha + \beta\bar{x} - \beta\bar{x} = \alpha$$

The likelihood function is

$$\begin{aligned} & \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2} \right) \\ = & \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\sum_{i=1}^{n} \frac{[(y_i - \bar{y}) - (\alpha - \widehat{\alpha}) - \beta(x_i - \bar{x})]^2}{2\sigma^2} \right) \\ = & \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{S_{yy} + n(\alpha - \widehat{\alpha})^2 + \beta^2 S_{xx} - 2\beta S_{xy}}{2\sigma^2} \right) \end{aligned}$$

where

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

We still use the notation S_{yy} when y_i is replaced by Y_i.

From the properties of the exponential family, a complete and sufficient statistic for $\theta = (\alpha, \beta, \sigma^2)$ is $(\widehat{\alpha}, \widehat{\beta}, S_{yy})$.

Since $\widehat{\alpha}$ and $\widehat{\beta}$ are unbiased estimators and functions of the sufficient and complete statistic, they are UMVUE's.

The best Linear unbiased estimator (BLUE)

What if we remove the normality assumption?

$\widehat{\alpha}$ and $\widehat{\beta}$ are still LSE, but not MLE.

A statistical property of the LSE is that it is the best linear unbiased estimator (BLUE) in the sense that $\widehat{\beta}$ (or $\widehat{\alpha}$) has the smallest variance within the class of linear unbiased estimators of β (or α) of the form

$$\sum_{i=1}^{n} d_i Y_i, \qquad d_i\text{'s are known constants}$$

If the estimator of this form is unbiased for β, then

$$\beta = E\left(\sum_{i=1}^{n} d_i Y_i\right) = \sum_{i=1}^{n} d_i E(Y_i) = \sum_{i=1}^{n} d_i(\alpha + \beta x_i) = \alpha \sum_{i=1}^{n} d_i + \beta \sum_{i=1}^{n} d_i x_i$$

holds for all α and β, which implies that

$$\sum_{i=1}^{n} d_i = 0, \qquad \sum_{i=1}^{n} d_i x_i = 1$$

A geometric description of the BLUE of β is given in the next figure.

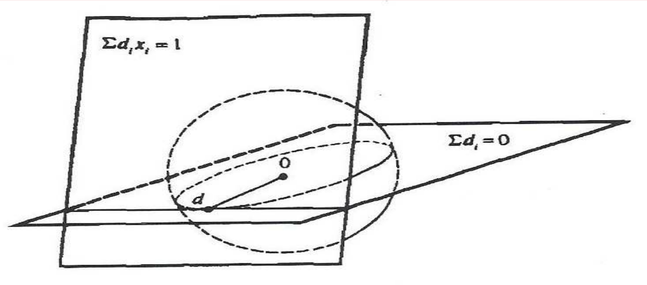


Figure: Geometric description of the BLUE of β

Proof: the LSE $\widehat{\beta}$ is BLUE

Since

$$\mathrm{Var}\left(\sum_{i=1}^n d_i Y_i\right) = \sum_{i=1}^n d_i^2 \mathrm{Var}(Y_i) = \sigma^2 \sum_{i=1}^n d_i^2$$

the BLUE of β must be a solution of

$$\min \sum_{i=1}^n d_i^2 \quad \text{subject to} \quad \sum_{i=1}^n d_i = 0, \quad \sum_{i=1}^n d_i x_i = 1$$

Consider the Lagrange multiplier method by minimizing

$$g(d_1, ..., d_n, \lambda_1, \lambda_2) = \sum_{i=1}^n d_i^2 + \lambda_1 \sum_{i=1}^n d_i + \lambda_2 \left(\sum_{i=1}^n d_i x_i - 1\right)$$

Taking derivatives, we obtain that

$$0 = \frac{\partial g}{\partial d_i} = 2d_i + \lambda_1 + \lambda_2 x_i, \quad i = 1, ..., n$$

Then

$$0 = \sum_{i=1}^n (2d_i + \lambda_1 + \lambda_2 x_i) = \lambda_1 n + \lambda_2 \sum_{i=1}^n x_i$$

which gives $\lambda_1 = -\lambda_2 \bar{x}$ and, hence

$$0 = 2d_i + \lambda_2(x_i - \bar{x})$$

Then

$$0 = \sum_{i=1}^{n} (x_i - \bar{x})[2d_i + \lambda_2(x_i - \bar{x})] = 2 + \lambda_2 S_{xx}$$

which gives $\lambda_2 = -2/S_{xx}$.

Then

$$d_i = -(\lambda_1 + \lambda_2 x_i)/2 = -\lambda_2(x_i - \bar{x})/2 = (x_i - \bar{x})/S_{xx}$$

and the BLUE of β is

$$\sum_{i=1}^{n} d_i Y_i = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{S_{xx}} Y_i = \frac{S_{xy}}{S_{xx}} = \widehat{\beta}$$

Since

$$\widehat{\beta} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(\beta x_i + \varepsilon_i)}{S_{xx}} = \beta + \sum_{i=1}^{n} d_i \varepsilon_i$$

where $d_i = (x_i - \bar{x})/S_{xx}$, we obtain that

$$\mathrm{Var}(\widehat{\beta}) = \sum_{i=1}^{n} d_i^2 \mathrm{Var}(\varepsilon_i) = \frac{\sigma^2}{S_{xx}}$$