Statistics

Luc Veldhuis

December 2016

Optimality theory

Sufficient statistics

Definition

A statistic $V = V(X)$ is sufficient for the data/observation X with probability distribution p_θ if if functions g_θ and h exist such that for all x:

$$p_\theta(x) = g_\theta(x)h(x)$$

V generally is a simple low dimensional statistic derived from the data.

ML-estimation with sufficient statistics

For Maximum Likelihood estimation of θ it suffices to consider only the first factor $g_\theta(V(X))$. This involves only $V(X)$. Hence, $V(X)$ is sufficient to estimate θ. In fact, it yields the same estimate as the one based on X.

Example

$X_1, \ldots, X_n$, $X \sim N(\mu, \sigma^2)$ with μ, σ^2 unknown, $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}$. Apply the factorization definition to find a sufficient statistic.

$$
\begin{aligned}
p_\theta(X) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i-\mu)}{2\sigma^2}} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2)} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2}
\end{aligned}
$$

Optimality theory

Lemma

Let $V = V(X)$ be sufficient and $V^* = f(V)$ with f a $1-1$ (invertable) function. Then V^* is also sufficient.

Proof

$V^* = f(x)$ and $V = f^{-1}(V^*)$. As V is sufficient,
$p_\theta(X) = g_\theta(V(X))h(X)$
$= g_\theta(f^{-1}(V(X)))h(X) = \hat{g}_\theta(V^*(X))h(X)$ with $\hat{g}_\theta = g_\theta \circ f^{-1}$
Thus V^* is sufficient.

Optimality theory

Estimation theory

Performance measure of an estimator $T(X)$ is the MSE.
An estimator T_0 for $g(\theta)$ is the absolute best if

$$MSE(\theta; T_0) \leq MSE(\theta; T)$$

for all θ and T.
Impossible, it requires $MSE(\theta; T_0) = 0$ as the estimator $T = g(\theta)$ then $MSE(g(\theta); T) = 0$ for $g(\theta)$ nonzero if true parameter is unequal to $g(\theta)$.
Consider an alternative criterion.

Definition

An estimator T of $g(\theta)$ is uniformly minimum variance unbiased (UMVU) if T is unbiased for $g(\theta)$ and $\mathbb{V}_\theta(S)$ for all θ and all other unbiased estimators S.

Optimality theory

Theorem (Rao-Blackwell)

Let $V = V(X)$ be a sufficient statistic and $T = T(X)$ is an estimator of $g(\theta)$. Then, there exists an estimator $T^* = T(V)$ that only depeds on V such that $\mathbb{E}_\theta(T^*) = \mathbb{E}(T)$ and $\mathbb{V}_\theta(T^*) \leq \mathbb{V}_\theta(T)$ for all θ. in particular, $MSE(\theta; T^*) \leq MSE(\theta; T)$ for all θ

Consequence of theorem

Constrain ourselves to unbiased estimates based on sufficient statistic when searching for a UMVU-estimator. If for a given sufficient statistic V there exists only one $(i = 1)$ estimator $T = T(V)$ then automatically T is UMVU. V is then called compete.

Optimality theory

Definition

A statistic V is complete if $\mathbb{E}_\theta(f(V)) = 0$ for all θ is only feasible for function f such that $p_\theta(f(V) = 0) = 1$ for all θ.
Then if V is complete, then there is only 1 estimator $T(V)$ that is unbiased.

Proof

Suppose not. $T(V)$ and $S(V)$ unbiased. Then:

$$\mathbb{E}_\theta(T(V) - S(V)) = \mathbb{E}_\theta(T(V)) - \mathbb{E}_\theta(S(V)) = 0$$
$$\Rightarrow$$
$$p_\theta((T(V) - S(V)) = 0) = 1$$

for all θ. Thus $T(V) = S(V)$
Completeness assures $T(V)$ is the <u>unique</u> unbiased estimator.

Optimality theory

Theorem

If V is sufficient and complete and $T = T(V)$ is as unbiased estimator of $g(\theta)$. Then T is UMVU.

Proof

Let $S = S(X)$ be a different unbiased estimator of $g(\theta)$.
With Rao-Blackwell it follows that $S^*(V)$ exists, also unbiased, and $\mathbb{V}_\theta(S^*) \leq \mathbb{V}_\theta(S)$.
Then $S^* - T$ only depends on V and $\mathbb{E}_\theta(S^* - T) = 0$ for all θ.
As V is complete, $\mathbb{E}_\theta(S^* - T) = 0$ implies $P_\theta((S^* - T) = 0) = 1$ for all θ. Thus $T = S^*$ with probability 1 and
$\mathbb{V}_\theta(T) = \mathbb{V}_\theta(S^*) \leq \mathbb{V}_\theta(X)$. for all θ.
Thus, T is better than S.

Example

$X_1, \ldots, X_n \sim U[0, \theta]$, θ unknown with $\theta > 0$

$p_\theta(X) = \prod_{i=1}^{n} \frac{1}{\theta} 1_{0 \le x_i \le \theta}(x_i) = \theta^{-n} 1_{[X_{(n)}, \infty)}(\theta)$

The factorisation definition applies: $X_{(n)}$ is sufficient.

The density of $X_{(n)}$ is $\theta^{-n} n x^{n-1}$ (excercises week 1 or 2).

$X_{(n)}$ is also complete as $\mathbb{E}_\theta(f(X_{(n)})) = \int_0^\theta f(x)\theta^{-n} n x^{n-1} dx = 0$ for all θ implies that $f \equiv 0$.

Can be seen through differentiation of $\int_0^\theta f(x) x^{n-1} dx$ with respect to $\theta \Rightarrow f(\theta)\theta^{n-1} = 0$ for all $\theta > 0$

$\mathbb{E}_\theta(X_{(n)}) = \frac{n}{n+1}$ thus $\frac{n+1}{n} X_{(n)}$ is unbiased and UMVU.

Optimality theory

To find complete statistics, we study the exponential family.

Definition

A group of probability distribution $p_\theta(x)$ is called the k-dimentional exponential family if functions c, h, Q, V exists such that

$$p_\theta(X) = c(\theta)h(X)e^{\sum_{j=1}^{k} Q(\theta)V(X)}$$

In particular, $V = (V_1(X), \ldots, V_n(X))$ is sufficient.

Theorem

For a given exponential family the statistic $V = (V_1(X), \ldots, V_n(X))$ is sufficient and complete if the set $\{(Q_1(\theta), \ldots, Q_n(\theta)) : \theta \in \Theta\}$ has an interior point.
I.e this set has volume in $\mathbb{R}^n$

Optimality theort

Example

$X_1, \ldots, X_n \sim Poisson(\theta)$ with unknown $\theta > 0$

$$
\begin{aligned}
p_\theta(x) &= \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} \\
&= e^{-n\theta} \left(\prod_{i=1}^{n} \frac{1}{x_i!}\right) e^{log(\theta) \sum_{j=1}^{n} x_i} \\
&= c(\theta) h(x) e^{Q_1(\theta) V_1(X)}
\end{aligned}
$$

with $c(\theta) = e^{-n\theta}$, $h(x) = \prod_{i=1}^{n} \frac{1}{x_i!}$, $Q_1(\theta) = log(\theta)$, $V_1(X) = \sum_{j=1}^{n} x_i$

As $\{log(\theta) : \theta > 0\} = \mathbb{R}$ does contain an interior point, $\sum_{j=1}^{n} x_i$ is sufficient and complete.
Furthermore, $\overline{X}$ is an UMVU estimator of θ.

Optimality theory

Example

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, σ^2 unknown.

$$
\begin{aligned}
p_\theta(X) &= \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} e^{\frac{1}{2\sigma^2}(x_i-\mu)^2} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + n\mu^2)} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{n\mu^2}{2\sigma^2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{2\mu}{2\sigma^2}\sum_{i=1}^{n} x_i} \\
&= c(\mu, \sigma^2) e^{Q_1(\mu,\sigma^2)V_1(X) + Q_2(\mu,\sigma^2)V_2(X)}
\end{aligned}
$$

As $\{(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma}) : \mu \in \mathbb{R}, \sigma^2 > 0\} = \mathbb{R}_{<0} \times \mathbb{R} \subset \mathbb{R}^2$ has an interior point. Thus $(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2)$ is sufficient and complete. $\overline{X}$ and S_x^2 are UMVU estimator of μ and σ

Optimality theory

Example (bended normal)

$X_1, \dots, X_n \sim N(\theta, \theta^2)$ with $\theta \in \mathbb{R}$ unknown.

$$\begin{aligned} P_\theta(X) &= \prod_{i=1}^{n} \frac{1}{2\pi\theta^2} e^{\frac{1}{2\theta^2}(x_i-\theta)^2} \\ &= (2\pi\theta^2)^{-\frac{n}{2}} e^{-\frac{1}{2\theta^2}(\sum\limits_{i=1}^{n} x_i^2 - 2\theta \sum\limits_{i=1}^{n} x_i + n\theta^2)} \\ &= (2\pi\theta^2)^{-\frac{n}{2}} e^{-\frac{n}{2}} e^{-\frac{1}{2\theta^2}\sum\limits_{i=1}^{n} x_i^2 + \frac{1}{\theta}\sum\limits_{i=1}^{n} x_i} \end{aligned}$$

Sufficient statistics again $(\sum\limits_{i=1}^{n} x_i, \sum\limits_{i=1}^{n} x_i^2)$ and $Q_1(\theta) = -\frac{1}{2\sigma^2}$ and $Q_2(\theta) = \frac{1}{\theta}$. As $\{(-\frac{1}{2\sigma^2}, \frac{1}{\theta}) : \theta \in \mathbb{R}\}$ is a 1-dimentional curve in $\mathbb{R}^2$ and hence does not contain an interior point. Statistic is sufficient but not complete.

Optimality theory

Remark

- UMVU estimators do not always exist
- UMVU estimators are not uniformly best estimators.
  Recall: $MSE = (Bias)^2 + Var$
  Trade off between bias and variance.