The Rao Blackwell Theorem

Lecture #5: Tuesday, 18 January 2005
Lecturer: Prof. Charles Elkan
Scribe: Taylor Sittler

1 Rao-Blackwell Theorem

In general, the Rao-Blackwell theorem is used to construct and prove minimally variant unbiased estimators. In order to apply the Rao-Blackwell theorem, there are four assumptions,

- Let P_θ be a family of distributions on sample space X, $\theta \in \Theta$
- Let $\hat{g} : X \rightarrow \Re$ be an unbiased estimator of $g : \theta \rightarrow \Re$
- Let $t : X {\rightarrow} Y$ be a sufficient statistic for θ
- Define $\hat{g} : X \rightarrow \Re$, $\hat{g}(x) = E[\tilde{g}(x')|t(x') = a]$ where $a = t(x)]$

Given these assumptions, the Rao-Blackwell theorem makes the following claims,

1. $\hat{g}(a) = \hat{g}(t(x))$ is not a function of θ
2. $E[\hat{g}(x)] = g(\theta)$ for every θ, i.e., $\hat{g}$ is unbiased
3. $Var[\tilde{g}(x)] {\geq} Var[\hat{g}(x)]$ for every θ

Proof of Claim 1:
We present only a summary of the proof here. The expanded proof was given in the previous lecture.
$\hat{g}(x) = E_\theta[\tilde{g}(x') : f(x') = t(x)] = \sum_{x':t(x')=t(x)} \tilde{g}(x')p[x')|t(x') = t(x)]$
Because t(x) is sufficient, $\hat{g}$ is a function of x only (not θ)

Proof of Claim 2:
This result is contigent upon the lemma on nested expectations. Dividing x into subsets, each summarized by t(x), and averaging over those subsets, the lemma asserts we obtain the same expectation as that of the original function $E[\tilde{g}(x)]$. This lemma was also discussed in more detail in a previous lecture. Therefore, if the function has an event $A = x' : t(x') = t(x)$:
$E[E[\tilde{g}(x)|A]] = E[\hat{g}(x)] = g(\theta)$ for all θ

Proof of Claim 3:
Define $c(u) = (u - g(\theta))^2$ for any function u: $X \rightarrow \Re$
Note 1: if u(x) is an unbiased estimator, then $E[c(u)] = E[(u(x) - g(\theta))^2] = var[u]$
Note 2: c is a different function for each θ
Note 3: c(u) is a convex function

Example 1.1. Convex Functions

Because of the shape of a convex function, the average value of the function at any two points is greater than the function of the average of those two points.

e.g:$mean(f(x, y)) \geq f(mean(x, y))$

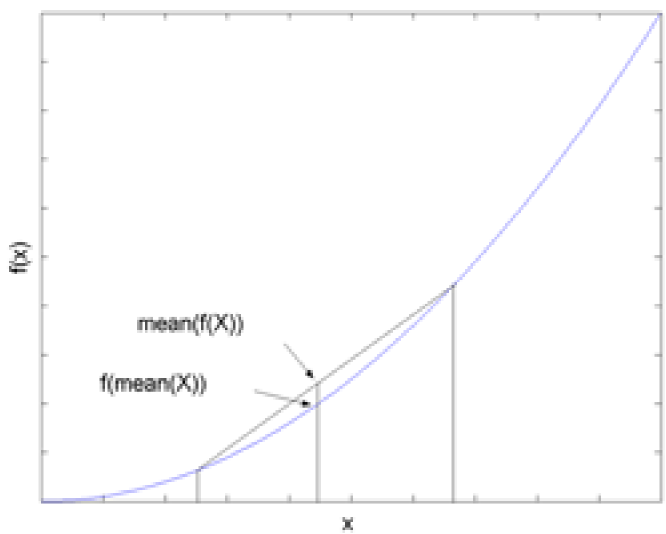

Figure 1: General Convex Function

This can be observed schematically by drawing a line between any two points in the graph. Observe that the y-value of the function is greater than the value of the function between those two points, which including the mean. So we use Jensen's inequality, which is independent of the underlying probability distribution, to prove this result in the theorem. So in this context, given $A = x' : t(x') = t(x)$, $u(x') = \tilde{g}(x')$:

$$E[c(u)] \geq c(E[u]) \tag{1}$$

$$E[c(u)|A] \geq c(E[u|A]) \tag{2}$$

Then:

$$E[c(\tilde{g}(x))|A] \geq c(E[\tilde{g}(x')|A]) \tag{3}$$

Remembering the assumptions and taking the expectation of both sides:

$$E[c(\tilde{g}(x))|A] \geq c(\hat{g}(x)]) \tag{4}$$

$$E[E[c(\tilde{g}(x))|A]] \geq E[c(\hat{g}(x)])] \tag{5}$$

Then by definition of c():

$$E[c(\hat{g}(x))] = E_{\tilde{x}P_\theta}[(\hat{g}(x) - g(\theta))^2] = var[\hat{g}(x)] \tag{6}$$

$$E[E[c(\tilde{g}(x))|A]] \geq var[\hat{g}(x)] \tag{7}$$

And again using nested expectation $E[E[c(\tilde{g}(x')|A)]] = E[c(\tilde{g}(x))]$:

$$E[c(\tilde{g}(x))] \geq var[\hat{g}(x)] \tag{8}$$

$$var[\tilde{g}(x)] \geq var[\hat{g}(x)] \tag{9}$$

This proves that the new function $\hat{g}(x)$, the expectation of $\tilde{g}(x)$ conditioned on a complete,sufficient statistic, has an <u>equivalent or better</u> variance than the original function $\tilde{g}(x)$.