Handout # 8
14.381 Statistics   Fall, 2004

TA: José Tessada (tessada@mit.edu)

The plan for today:

1. Adminstrative stuff: additional office hours? Additional recitation?
2. Sufficient statistics
3. Complete statistics

Sufficient Statistics

The main idea behind the use of statistics is *data reduction*. Keep that in mind as you move on in this part of the class. We do not want, and we do not need, to carry around the whole sample to make inference and gain knowledge about the unferlying distribution.

Definition 1 *A statistic $T(X)$ is a sufficient statistic for θ if the conditional distribution of the sample X given the value of $T(X)$ does not depend on θ.*

The concept of sufficiency seems easy to understand, but trying to prove by definition can be a relatively complicated task. Although *Theorem 6.6.2* provides a shortcut, it still requires the computation of conditional probabilities, not an easy task sometimes.

Theorem 2 *(Factorization Theorem) Let $f(x|\theta)$ denote the joint pdf or pmf of a sample X. A statistic $T(X)$ is a sufficient statistic for θ iff there exist functions $g(t|\theta)$ and $h(\theta)$ sucht that, for all sample points x and all parameters,*

$$f(x|\theta) = g(T(x)|\theta)\, h(x) \tag{1}$$

Example 3 *Normal sufficient statistic, both parameters unknown.*

$$f(\mathbf{x}|\theta) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\left(\sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - \mu)^2\right) / \left(2\sigma^2\right)\right) \tag{2}$$

Equation (2) is the joint pdf of a random sample of size n from a normal distribution. We can factorize the pdf defining $h(\mathbf{x}) = 1$ and noting that it depends on the sample only through $\overline{x}$ and s^2. So, $\left(\overline{x}, s^2\right)$ is a two-dimensional sufficient statistic.

The sufficient statistic can be of a higher dimension than θ. The statistic to be "useful" (that's my definition) needs to be of a dimension smaller than n, the size of the sample. After all, we are talking about data reduction!

Exercise 4 *C&B 6.5*

Summary 5 *A sufficient statistic for a parameter θ is a statistic that, in a certain sense,* ***captures all the information about θ contained in the sample****.*

Any additional information in the sample, besides the value of the sufficient statistic, does not contain any more information about θ.

Completeness

Definition 6 *A statistic $T(\mathbf{X})$ is a statistic complete if $E_\theta\left[g\left(T\right)\right]=0$ implies $\Pr_\theta\left(g\left(T\right)=0\right)=1$.*

Complete statistics are useful because they prove to be an important tool in point estimation. Note that the previous statement says that any unbiased estimator of 0 based on this statistic has to be identically equal to 0 in probability sense.

The following theorem provides us with a very simple way of finding complete statistics in the case of the exponential family of pdf/pmfs.

Theorem 7 *(Complete statistics in the exponential family) Let $X_1,...,X_n$ be iid observations from an exponential family with pdf or pmf of the form*

$$f\left(x|\theta\right)=h\left(x\right)c\left(\theta\right)\exp\left(\sum_{i=1}^{k}w_j\left(\boldsymbol{\theta}\right)t_j\left(x\right)\right), \tag{3}$$

where $\boldsymbol{\theta}=\left(\theta_1,\theta_2,...,\theta_k\right)$. Then the statistic

$$T\left(\mathbf{X}\right)=\left(\sum_{i=1}^{n}t_1\left(X_i\right),\sum_{i=1}^{n}t_2\left(X_i\right),...,\sum_{i=1}^{n}t_k\left(X_i\right)\right) \tag{4}$$

is complete if $\left\{\left(w_1\left(\boldsymbol{\theta}\right),...,w_k\left(\boldsymbol{\theta}\right)\right):\theta\in\Theta\right\}$ contains an open set in $\Re^k$.

Please check your copy of the book contains the right statement. Some old versions have a wrong condition with respect to the open set condition (*Theorem 6.2.25*).

Is the "open set" condition restrictive? Yes, a $N\left(\theta,\theta^2\right)$ can be written as (3), but the open set condition is violated, we have a lower-dimensional curve, a parabola in this case.

Exercise 8 *C&B 6.15. We show that in fact the sufficient statistic is not complete for the $N\left(\theta,a\theta^2\right)$ distribution, a is a known constant.*

Why are complete statistics useful?

Theorem 9 *(Theorem 7.3.20) If $E_\theta\left(W\right)=\tau\left(\theta\right)$, W is the best unbiased estimator of $\tau\left(\theta\right)$ if and only if W is uncorrelated with all unbiased estimators of 0.*

What is the intuition? If we can find an unbiased estimator of 0 that is correlated with W, then we could improve upon W using noise, that should make you think that W is not such a great estimator. Now, think of the definition of a complete statistic, a complete statistic basically tells you that any unbiased estimator of 0 based on it has to be equal to 0 with probability 1. If W is based on a complete statistic, then we know the conditions for *Theorem 9* hold.

Theorem 10 *(Theorem 7.3.23) Let T be a complete sufficient statistic for a parameter θ, and let $\phi\left(T\right)$ be any estimator based only on T. Then $\phi\left(T\right)$ is the unique best unbiased estimator of its expected value.*

Exercise 11 *C&B 7.10 part (a).*
C&B 7.37