Statistical Inference

Robert L. Wolpert
Institute of Statistics and Decision Sciences
Duke University, Durham, NC, USA
Spring, 2000

February 16, 2000

1. Random Samples & Likelihoods

Much of statistical theory is concerned with inference on the basis of a "sample" $X = (X_1, \cdots, X_n)$ of some number n of independent replicates, all from the same probability distribution with pdf or pmf $f(x \mid \theta)$; interest centers, in part, on what happens as n becomes large. We will pay special attention to the **exponential family** in which each pdf takes the form

$$f(x \mid \theta) = h(x) \exp\left[\sum_{i=1}^{k} \eta_i(\theta) t_i(x) - c(\theta)\right]$$

(surprisingly many of the commonly-considered distributions can be written in this form, for suitable h, c, and $\{\eta_i, t_i\}_{i \le k}$), but will also consider other distributions. The likelihood function for a random sample of size n from the exponential family is

$$L(\theta) = \exp\left[\sum_{i=1}^{k} \eta_i(\theta) \sum_{j=1}^{n} t_i(x_j) - nc(\theta)\right]$$

It is often convenient to reparametrize exponential families to the *natural parameter* $\eta = \eta(\theta) \in \mathbb{R}^k$, leading (with a slight abuse of notation for $c(\cdot)$) to

$$f(x \mid \eta) = h(x) e^{\eta \cdot T(x) - c(\eta)}$$

Since any pdf integrates to unity we have

$$e^{c(\eta)} = \int_{\mathcal{X}} h(x) e^{\eta \cdot T(x)}\, dx$$

and hence can calculate the moment generating function (MGF) for the **natural statistic** $T(x) = \{t_1(x), \cdots, t_k(x)\}$ as

$$\begin{aligned}
M_T(s) &= \mathsf{E}\left[e^{s \cdot T(X)}\right] \\
&= \int_{\mathcal{X}} e^{s \cdot T(x) - c(\eta)}\, h(x) e^{\eta \cdot T(x)}\, dx \\
&= e^{-c(\eta)} \int_{\mathcal{X}} h(x) e^{(\eta + s) \cdot T(x)}\, dx \\
&= e^{c(\eta+s) - c(\eta)},
\end{aligned}$$

so we can find moments for the natural statistic by

$$\begin{aligned}
\mathsf{E}[T] &= \nabla \log M_T(0) &= \nabla c(\eta) \\
\mathsf{V}[T] &= \nabla^2 \log M_T(0) &= \nabla^2 c(\eta).
\end{aligned}$$

2. Leftovers

- Functions of RV's
- Different Reference Measures
- Random Samples
- Exponential Families
- Generating Functions & Moments
- Normal Approximations & the CLT
- Inference for Normals
- Sufficient Statistics, w/ Warning on Robustness
- Likelihood Principle