ECE 7251: Signal Detection and Estimation

Spring 2002
Prof. Aaron Lanterman
Georgia Institute of Technology

Lecture 25, 3/13/01:
The Role of Information Theory in Detection Theory

Kullback-Leibler Distances

- The Kullback-Leibler distance (or relative entropy) between two probability densities is given by

$$D(a\|b)=\int a(y)\ln\frac{a(y)}{b(y)}dy$$

- Is zero iff $a=b$ almost everywhere
- Not a "distance" in the usual sense of the word like "Euclidean distance"
  - Not symmetric in general, i.e.

$$D(a\|b)\neq D(b\|a)$$

Setup for Stein's Lemma

- Suppose we fix P_{FA}, and want to study the Neyman-Pearson test's asymptotic P_M as the number of independent samples n gets large
- To keep P_{FA} constant as n increases, we need to make the threshold vary with n; we also note its dependence on the P_{FA} we've chosen

$$g=g_n(P_{FA})$$

- P_M is then, in a roundabout way, a function of n and the chosen P_{FA}

$$P_M=P_M(g_n)=P_M(g_n(P_{FA}))$$

Stein's Lemma

- Formally, Stein's lemma says:

$$\lim_{P_{FA}\to 0}\lim_{n\to\infty}\frac{1}{n}\ln P_M=-D(p_0\|p_1)$$

- Informally, Stein's lemma says that if P_{FA} is small and n is big, then

$$P_M\approx\exp[-nD(p_0\|p_1)]$$

- Conversely, we could let $g=g_n(P_M)$, then

$$\lim_{P_M\to 0}\lim_{n\to\infty}\frac{1}{n}\ln P_{FA}=-D(p_1\|p_0)$$

$$P_{FA}\approx\exp[-nD(p_1\|p_0)]$$

What Are These K-L Distances?

- They involve the conditional expectations of the loglikelihood ratio:

$$-D(p_1\|p_0)=-\int p_1\ln\frac{p_1}{p_0}=-E[L\,|\,H_1]$$

$$-D(p_0\|p_1)=-\int p_0\ln\frac{p_0}{p_1}=\int p_0\ln\frac{p_1}{p_0}=E[L\,|\,H_0]$$

- Again rephrasing:

$$P_{FA}\approx\exp[-nE[L\,|\,H_1]]$$

$$P_M\approx\exp[nE[L\,|\,H_0]]$$

On the Proof of Stein's Lemma

- For discrete data spaces, can prove using the "method of types"
  - Chapter 12 of Cover & Thomas, "Elements of Information Theory"
- For continuous data spaces, the proof is much much much much harder; lies in the land of "large deviations theory"
  - Most texts highly mathematical and obsess on technical details
  - One refreshingly readable text:
    J.A. Bucklew, "Large Deviation Techniques in Decision, Simulation, and Estimation", Wiley, 1990.

Returning to Bayesianland

- Suppose we have prior probabilities, and use the Bayes that minimizes the prob. of error (notice g is now fixed)
- A theorem from p. 312 of Cover & Thomas:

$$\lim_{n\to\infty}\frac{1}{n}\ln P_e = -D^* \equiv -D(p_{s^*}\,\|\,p_1) \underset{(\aleph)}{=} -D(p_{s^*}\,\|\,p_0)$$

$$\text{where } p_s(y) = \frac{p_1^s(y)\,p_0^{1-s}(y)}{\int p_1^s(z)\,p_0^{1-s}(z)\,dz}$$

$$\text{where } s^* = s \text{ such that } (\aleph) \text{ holds}$$

The Chernoff Information

- $D^* \equiv D(p_{s^*}\,\|\,p_1) = D(p_{s^*}\,\|\,p_0)$ is called the Chernoff information
- Why? It turns out there's an equivalent expression for D^* (C&T, p. 314)

$$D^* = -\min_{0\le s\le 1}\log\int p_1^s(y)\,p_0^{1-s}(y)\,dy = -\min_{0\le s\le 1} m(s)$$

- So rephrasing the theorem, we can write:

$$\lim_{n\to\infty}\frac{1}{n}\ln P_e = \min_{0\le s\le 1} m(s)$$

$$P_e \approx \exp[n\min_{0\le s\le 1} m(s)]$$

A Link to Estimation Theory

- Suppose we have a hypothesis testing prob.:

$$H_0 \sim p(y;q),\; H_1 \sim p(y;q+\Delta)$$

- Let's look at:

$$D(p_1\,\|\,p_0) + D(p_0\,\|\,p_1)$$

$$= \int p(y;q+\Delta)\ln\frac{p(y;q+\Delta)}{p(y;q)}dy + \int p(y;q)\ln\frac{p(y;q)}{p(y;q+\Delta)}dy$$

$$= \int [p(y;q+\Delta) - p(y;q)]\ln\frac{p(y;q+\Delta)}{p(y;q)}dy$$

Rediscovering Fisher Information

$$\lim_{\Delta\to 0}\frac{1}{\Delta^2}\int [p(y;q+\Delta) - p(y;q)]\ln\frac{p(y;q+\Delta)}{p(y;q)}dy =$$

$$\int \lim_{\Delta\to 0}\frac{[p(y;q+\Delta) - p(y;q)]}{\Delta}\,\frac{\ln p(y;q+\Delta) - \ln p(y;q)}{\Delta}dy$$

$$= \int \underbrace{\frac{dp(y;q)}{dq}}_{dv}\;\underbrace{\frac{d}{dq}\ln p(y;q)}_{u}\,dy \qquad \boxed{\begin{aligned} v &= p(y;q) \\ du &= \frac{d^2}{dq^2}\ln p(y;q)\,dy \end{aligned}}$$

$$= -\int v\,du = -\int p(y;q)\frac{d^2}{dq^2}\ln p(y;q)\,dy = F(q)$$