ECE 7251: Signal Detection and Estimation

Spring 2002

Prof. Aaron Lanterman

Georgia Institute of Technology

Lecture 29, 3/22/02:

Generalized Likelihood Ratio Tests and Model Order Selection Criteria

The Setup

- Usual parametric data model $p(y;q)$
- In previous lecture on LMP tests, we assumed specials structures like:

$$H_0: q = q_0,\ H_1: q > q_0$$

$$\text{or } H_0: q = q_0,\ H_1: q \neq q_0$$

- What to do if we have a more general structure like:

$$H_0: q \in \mathcal{S}_0,\ H_1: q \in \mathcal{S}_1$$

- Often, we do something a bit ad hoc!

The GLRT

- Find parameter estimates $\hat{q}_0$ and $\hat{q}_1$ under H_0 and H_1
- Substituting estimates into likelihood ratio yields a generalized loglikelihood ratio test:

$$\Lambda_{GLR}(y) = \frac{p(y;\hat{q}_1)}{p(y;\hat{q}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} l$$

- If convenient, use ML estimates:

$$\frac{\max_{q \in \mathcal{S}_1} p(y;q)}{\max_{q \in \mathcal{S}_0} p(y;q)} \underset{H_0}{\overset{H_1}{\gtrless}} l$$

Two-Sided Gaussian Mean Example

$$y_i \in \mathcal{N}(q, s^2),\ H_0: q = 0,\ H_1: q \neq 0$$

$$\ln \frac{p(y;\hat{q})}{p(y;0)} = -\sum_{i=1}^{n} \frac{\Big(y_i - \overbrace{\frac{1}{n}\sum_{j=1}^{n} y_j}^{\hat{q}}\Big)^2}{2s^2} + \sum_{i=1}^{n} \frac{y_i^2}{2s^2}$$

$$= \frac{2\sum_{i-1}^{n} y_i \frac{1}{n}\sum_{j=1}^{n} y_j - n\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2}{2s^2}$$

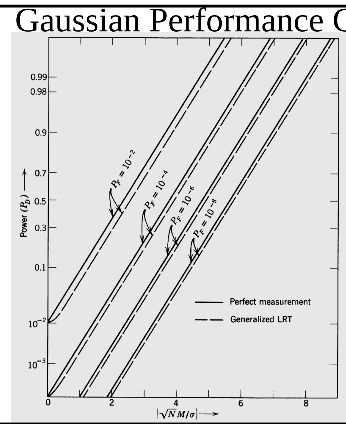
Two-Sided Gaussian Example Con't

$$\frac{2n\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2}{2s^2} = \frac{n}{2s^2}\bar{y}^2 \underset{H_0}{\overset{H_1}{\gtrless}} l$$

Same as the LMPU test from last lecture! $\left\{ |\bar{y}| \underset{H_0}{\overset{H_1}{\gtrless}} g \right.$

- Chapter 9 of Hero derives and analyzes the GLRT for every conceivable Gaussian problem – a fantastic reference!

Gaussian Performance Comparison



We take a performance hit from not knowing the true mean

(Graph from p. 95 of Van Trees Vol. I)

Some Gaussian Examples

- Single population:
  - Tests on mean, with unknown variance yield "T tests"
  - Statistic has a Student-T distribution
  - Asymptotically Gaussian
- Two populations:
  - Tests on equality of variances, with unknown means yields a "Fisher F test"
  - Statistic has a Fisher-F distribution
  - Asymptotically Chi-Square
- See Chapter 9 of Hero

Asymptotics to the Rescue

- Suppose $n \to \infty$. Since the ML estimates are asymptotically consistent, the GLRT is asymptotically UMP
- If the GLRT is hard to analyze directly, sometimes asymptotic results can help
- Assume a partition

$$q = (j_1, \ldots, j_p, \underbrace{x_1, \ldots, x_q}_{\text{(nuisance parameters)}})$$

Asymptotics Con't

- Consider GLRT for a two-sided problem

$$H_0 : j = j_0, \; H_1 : j \neq j_0$$

  where x is unknown, but we don't care what it is
- When the density *p(y;q)* is smooth under H_0, it can be shown that for large *n*

$$2 \ln \Lambda_{GLR}(Y) = 2 \ln \frac{p(Y;\hat{q})}{p(Y;q_0)} \sim c_p \Bigg\} \text{(Chi-square with } p \text{ degrees of freedom)}$$

- Recall $E[c_p] = p, \;\; \mathrm{var}(c_p) = 2p$

A Strange Link to Bayesianland

- Remember if we had a prior *p(q)*, we could handle composite hypothesis tests by integrating and reducing things to a simple hypothesis test

$$p(y) = \int_{R^p} p(y \mid q) p(q) dq$$

- If *p(q)* varies slowly compared to *p(y|q)* around the MAP estimate, we can approx.

$$p(y) \approx p(q) \int_{R^p} \exp[L(q)] dq$$

Laplace's Approximation

- Do a Taylor series expansion

$$\int_{R^p} \exp[L(\hat{q}_{ML}) - \frac{(q - \hat{q}_{ML})^T F(y;q)(q - \hat{q}_{ML})^T}{2}] dq$$

$$= e^{L(\hat{q}_{ML})} \int_{R^p} \exp\left[- \frac{(q - \hat{q}_{ML})^T F(y;\hat{q}_{ML})(q - \hat{q}_{ML})^T}{2} \right] dq$$

where $F(y;\hat{q}_{ML}) = \left[- \frac{d^2 L(q)}{dq_r dq_c} \bigg|_{q = \hat{q}_{ML}} \right] \Bigg\}$ *Empirical Fisher info*

Laplace's Approximation Con't

- Recognize quadradic form of the Gaussian:

$$\int_{R^p} \exp\left[- \frac{(q - \hat{q}_{ML})^T F(y;\hat{q}_{ML})(q - \hat{q}_{ML})^T}{2} \right] dq$$

$$= \frac{(2p)^{p/2}}{\sqrt{\det F(y;\hat{q}_{ML})}}$$

- So $p(y) = p(\hat{q}_{ML}) p(y \mid \hat{q}_{ML}) \dfrac{(2p)^{p/2}}{\sqrt{\det F(y;\hat{q}_{ML})}}$

Large Sample Sizes

- Consider the logdensity:

$$\ln p(y) \approx \ln p(\hat{q}_{ML}) + \ln p(y|\hat{q}_{ML})$$
$$+\frac{p}{2}\ln 2p - \frac{1}{2}\ln\det F(y;\hat{q}_{ML})$$

- Suppose we have n i.i.d. samples. By the law of large numbers:

$$\ln\det F(y|\hat{q}_{ML}) \approx \ln\det F(\hat{q}_{ML}) = \ln\det nF_1(\hat{q}_{ML})$$
$$= \ln\det[nI \cdot F_1(\hat{q}_{ML})] = \ln\det[nI] + \ln\det F_1(\hat{q}_{ML})$$
$$= \ln n^p + \ln\det F_1(\hat{q}_{ML}) = p\ln n + \ln\det F_1(\hat{q}_{ML})$$

Schwarz's Result

- As n gets big

$$\ln p(y) \approx \ln p(\hat{q}_{ML}) + L(\hat{q}_{ML})$$
$$+\frac{p}{2}\ln 2p - \frac{1}{2}p\ln n - \frac{1}{2}\ln\det F_1(\hat{q}_{ML})$$
$$\approx L(\hat{q}_{ML}) - \frac{p}{2}\ln n$$

- Called Bayesian Information Criterion (BIC) or Schwarz Information Criterion (SIC)
- Often used in model selection; second term is a penalty on model complexity

Minimum Description Length

- BIC is related to Rissanen's Minimum Description Length criterion; *(p/2) ln(n)* is viewed as the optimum number of "nats" (like bits, but different base) used to encode the ML parameter estimate with limited precision
- Data is encoded with a string of length

$$\text{description length} = \underbrace{-L(\hat{q}_{ML})}_{\text{nats used to encode data given ML est.}} + \frac{p}{2}\ln n$$

- Choose model which describes the data using the smallest number of bits

References

- A.R. Barron, J. Rissanen, B. Yu, The Minimum Description Length Principle in Coding and Modeling, IEEE Trans. Info. Theory, Vol. 44, No. 6, Oct. 1998, pp. 2743-2760.
- A.D. Lanterman, *Schwarz, Wallace, and Rissanen: Intertwining Themes in Theories of Model Order Estimation*, International Statistical Review, Vol. 69, No. 2, August 2001, pp. 185-212.
- Special Issues:
  - Statistics and Computing (Vol. 10, No. 1, 2000)
  - The Computer Journal (Vol. 42, No. 4, 1999)