ECE 7251: Signal Detection and Estimation

Spring 2002
Prof. Aaron Lanterman
Georgia Institute of Technology

Lecture 2, 1/7/02:
Sufficient Statistics and Exponential Families

Sufficient Statistics

- A <u>statistic</u> $T=T(y)$ is any function of the data
- Think of statistics as "compressing" the data
- T is a <u>sufficient statistic</u> for $p(y;q)$ if

$$p(y\,|\,T(y);q) = p(y\,|\,T(y))$$

is not a function of the parameters

- Think of sufficient statistics as not throwing away any useful information about q
- Note if Y is a r.v., $T=T(Y)$ is a r.v.

Minimal Sufficient Statistics

- A sufficient statistic T is minimal for $p(y;q)$ if it is a function of every other sufficient statistic for $p(y;q)$
- Minimal sufficient statistics "compress" the data as much as possible without losing information about q
- Can be hard to find!

Finite Dimensional Statistics

- A sufficient statistic T is <u>fixed</u> (or <u>finite) dimensional</u> if it is not a function of the number of data samples
- Collecting more data does not ultimately require more "storage"
- Minimal sufficient statistics often finite dimensional (but not always!)

Ex: Coin flipping

- q =prob. of heads ; assume indep. flips
- y_i=1 if ith flip heads, 0 if tails

$$p(y;q) = \prod_{i=1}^{n} q^{y_i}(1-q)^{1-y_i}$$

$$= q^T(1-q)^{n-T} \quad \text{where} \quad T=\sum_{i=1}^{n} y_i$$

$$p(y\,|\,T;q) = \frac{p(y,T)}{p(T)} = \frac{q^T(1-q)^{n-T}}{\binom{n}{T} q^T(1-q)^{n-T}} = \frac{1}{\underbrace{\binom{n}{T}}}$$

Not a function of q, hence T is a suf. stat.

Fisher Factorization Theorem

- T is a sufficient statistic for $p(y;q)$ if we can decompose

$$p(y;q) = \underbrace{g(T(y),q)}h(y)$$

Parameters coupled to suff. stat., not raw data

- A useful consequence: density on T is

$$p(t;q) = g(t,q)\underbrace{q(t)}$$

Indep. of q (will come in handy later)

FF Ex. 1: Uniform Density

- $y = (y_1, y_2, \ldots, y_n)$ is i.i.d. uniform$(0, q)$

$$p(y;q) = \frac{1}{q^n} \prod_{i=1}^{n} I_{[0,q]}(y_i)$$

$$= \underbrace{\frac{1}{q^n} I_{[0,q]}(\overbrace{\max\{y\}}^{T(y)})}_{g(T(y),q)} \qquad h(y) = 1$$

- Satisfies FF, so T is a sufficient statistic
- Note T is finite dimensional

FF Ex. 2: Detection Problems

- $\theta \in \{0,1\}$

$$p(y;q) = q\, p(y;1) + (1-q)\, p(y;0)$$

$T(y)$ is the likelihood ratio

$$= \underbrace{\left[q \frac{p(y;1)}{p(y;0)} + (1-q) \right]}_{g(T(y),q)} \underbrace{p(y;0)}_{h(y)}$$

- Satisfies FF, so the likelihood ratio is a sufficient statistic for $p(y;q)$

FF Ex. 3: Cauchy Density

- $y = (y_1, y_2, \ldots, y_n)$ is i.i.d. Cauchy w/median q

$$p(y;q) = \underbrace{\frac{1}{p^n} \prod_{i=1}^{n} \frac{1}{1 + (y_i - q)^2}}_{g(T(y),q)}$$

- Uh oh! Denominator is a 2n-degree polynomial in q; depends on all cross products of data points
- The raw data is a minimal suff. stat.
- No fixed dimensional suff. stat. exists

FF Ex. 4: Exponential Families

- $p(y;q)$ is an exponential family if it can be decomposed as

$$p(y;q) = a(q) b(x) \exp \underbrace{\left[\sum_{l=1}^{m} c_l(q) d_l(y) \right]}$$

Coupling between parameters and data has a very specific form

- If support of $p(y;q)$ depends on q, then $p(y;q)$ can't be an exponential family!
  - Ex: *uniform(0,q)* is not an exponential family!
- Cauchy with unknown median not exp. family

FF for Exponential Families

- Suppose $y = (y_1, y_2, \ldots, y_n)$ i.i.d.

$$p(y;q) = \prod_{i=1}^{n} a(q) \exp\left[\sum_{l=1}^{m} c_l(q) d_l(y_i) \right] b(y_i)$$

$$= \underbrace{a^n(q) \exp\left[\sum_{l=1}^{m} c_l(q) \sum_{i=1}^{n} d_l(y_i) \right]}_{g(T(y),q)} \underbrace{\prod_{i=1}^{n} b(y_i)}_{h(y)}$$

- Sufficient statistic is

$$T = \left(\sum_{i=1}^{n} d_1(y_i), \ldots, \sum_{i=1}^{n} d_m(y_i) \right)$$

- Note T is finite dimensional

Sometimes It's Pretty Obvious...

- Ex: Gaussian distribution, $q = (m, \sigma^2)$

$$p(y;m,s^2) = \frac{1}{\sqrt{2ps^2}} \exp\left(-\frac{(y-m)^2}{2s^2} \right)$$

$$= \underbrace{\frac{1}{\sqrt{2ps^2}} \exp\left(-\frac{m^2}{2s^2} \right)}_{a(q)} \exp\left(\overbrace{\frac{m}{s^2}}^{c_1(q)} \underbrace{y}_{d_1(y)} - \overbrace{\frac{1}{2s^2}}^{c_2(q)} \underbrace{y^2}_{d_2(y)} \right)$$

$$b(y) = 1$$

Sometimes It's Not So Obvious...

- Ex: Beta distribution, $\boldsymbol{q}=(q,r)$

$$p(y;q,r) = \overbrace{\frac{\Gamma(q+r)}{\Gamma(q)\Gamma(r)}}^{a(\boldsymbol{q})} \boxed{y^{q-1}(1-y)^{r-1}} \overbrace{I_{(0,1)}(y)}^{b(y)}$$

$$\boxed{\exp\{\ln[y^{q-1}(1-y)^{r-1}]\}}$$

$$= \exp[\underbrace{(q-1)}_{c_1(\boldsymbol{q})}\underbrace{\ln y}_{d_1(y)} + \underbrace{(r-1)}_{c_2(\boldsymbol{q})}\underbrace{\ln(1-y)}_{d_2(y)}]$$

- So the Beta distribution is a two-parameter exponential family in q and r

Putting it All Together

- If y is i.i.d. Gaussian with $\boldsymbol{q}=(m,\sigma^2)$, then

$$T = \left(\sum_{i=1}^{n} y_i, \sum_{i=1}^{n} y_i^2\right) \text{ or } T = (\bar{y}, s^2) \text{ where } \underbrace{\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,}_{\text{Sample mean}} \quad \underbrace{s^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}_{\text{Sample variance}}$$

- If y is i.i.d. Beta with $\boldsymbol{q}=(q,r)$, then

$$T = \left(\sum_{i=1}^{n} \ln(y_i), \sum_{i=1}^{n} \ln(1-y_i)\right)$$