ESE 524: Sufficient Statistics

January 22, 2019

Finding Sufficient Statistics from the Definition

- Let $\mathbf{x} = [x[0], x[1], x[2], ..., x[N-1]]$ be a set of samples, θ be an unknown parameter, and $p(\mathbf{x}; \theta)$ be the corresponding probability density function.
- Statistic: Function $T(\mathbf{x})$ of the samples, intended to capture important information from the sample data.
- Sufficient Statistics: T is sufficient if $p(x|T, \theta)$ is not a function of theta.
- One way to check sufficiency is using the definition of conditional densities:

$$p(x|T, \theta) = \frac{p(x, T|\theta)}{p(T|\theta)}$$

- Instead of the factorization theorem, let's use this definition with Exampls 5.1 from Kay.

Normal PDF - Hard Version

- Let $\mathbf{x}$ consist of i.i.d. samples from a normal distribution with mean θ and known variance σ^2.
- Let $T(\mathbf{x}) = \sum_{i=0}^{N-1} x[n]$.
- T is the sum of independent gaussian random variables so it is a Gaussian RV with mean $N\theta$ and variance $N\sigma^2$.
- Next, we need the joint distribution of $\mathbf{x}$ and T. Note that $\mathbf{x}$ and T are linked.
- Let $T_0(\mathbf{x}) = \sum_{i=0}^{N-1} x[n]$. Then $\mathrm{Prob}(X = \mathbf{x}, T! = T_0(\mathbf{x})) = 0$, and $\mathrm{Prob}(X = \mathbf{x}, T = T_0(\mathbf{x})) = p(\mathbf{x}; \theta)$.
- Therefore a functional form is $p(\mathbf{x}, T; \theta) = p(\mathbf{x}; \theta)\delta(T - T_0(\mathbf{x}))$. (continued on next slide)

NormalPDF - Hard Version Cont.

- The joint pdf is:

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-\theta)^2)\delta(T-T_0(\mathbf{x})) =$$

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]^2)-2\theta T_0(\mathbf{x})+N\theta^2)\delta(T-T_0(\mathbf{x}))$$

- And the conditional PDF is:

$$\frac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]^2)-2\theta T_0(\mathbf{x})+N\theta^2)\delta(T-T_0(\mathbf{x}))}{\frac{1}{\sqrt{2\pi N\sigma^2}}\exp(-\frac{1}{2N\sigma^2}(T-N\theta)^2)}$$

- Now we have to simplify to see that the terms with θ cancel out (see board).
- This was much more complicated than the factorization theorem - often times figuring out a sufficient statistic from this method is impossible.

Two Jointly Sufficient Statistics

- Given a known amplitude A, frequency f, and white noise $w[n] \sim N(0, \sigma^2)$ consider the model:
$$x[n] = A\cos(2\pi f n + \theta) + w[n]$$
- What is the distribution of $x[n]$?
- What is the joint distribution of $\mathbf{x}$?
- Let's use the factorization theorem to find sufficient statistics that can be used to estimate θ (see board for details).

Two Jointly Sufficient Statistics

- Given a known amplitude A, frequency f, and white noise $w[n] \sim N(0, \sigma^2)$ consider the model:
$$x[n] = A\cos(2\pi fn + \theta) + w[n]$$
- What is the distribution of $x[n]$?
- What is the joint distribution of $\mathbf{x}$?
- Let's use the factorization theorem to find sufficient statistics that can be used to estimate θ (see board for details).
- What are $T_1(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]\cos(2\pi fn)$ and $T_2(X) = \sum_{n=0}^{N-1} x[n]\sin(2\pi fn)$?

Minimal Sufficient Statistics

- Minimal Sufficient Statistic: A sufficient statistic $M(\mathbf{x})$ is minimally sufficient if any other sufficient statistic, T, can be expressed as a function $f(M(\mathbf{x}))$.
- Intuitively, they take up the least memory compared to any other sufficient statistic.
- M is minimally sufficient if the following statement holds for all possible $\mathbf{x}_1$ and $\mathbf{x}_2$:

$$\frac{p(\mathbf{x}_1;\theta)}{p(\mathbf{x}_2;\theta)} \text{ is not a function of } \theta \iff M(\mathbf{x}_1) = M(\mathbf{x}_2)$$

- For most "nice" probability distributions, there is at least one minimal sufficient statistic.

Example of Minimal Sufficient Statistics

- Let Y be a binomial random variable with known size N and unkown probability p.
- Let X be a random variable with a binomial conditional density with size y and unknown probability q, i.e. $f_{X|Y}(X|Y=y) \sim Bin(y,q)$
- Find a minimal sufficient statistic for (p,q).
- First set up the probabiity distribution:

$$f(X,Y;p,q) = f_Y(Y;p)f_{X|Y}(X|Y;q) = \begin{pmatrix} N \\ y \end{pmatrix} \pi^y (1-\pi)^{N-y} \begin{pmatrix} y \\ x \end{pmatrix} p^x (1-p)^{y-x}$$

$$\text{for } x \in \{0,1,...y\}, y \in \{0,1,...N\}$$

- Let $T_1 = x$ and $T_2 = y$, $h(x,y) = \binom{N}{y}\binom{y}{x}$ and $g(x,y;p,q) = p^y(1-p)^{N-y}q^x(1-q)^{y-x}$. By the factorization theorem (x,y) are sufficient statistics.

Example of Minimal Sufficient Statistics cont.

- Let (x_1, y_1) and (x_2, y_2) be two samples from $f(X, Y; p, q)$.
- First show that $(T_1(x_1) = x_1 = x_2 = T_2(x_2)\ and\ T_2(y_1) = y_1 = y_2 = T_2(y_2)) \rightarrow$ $(\frac{p(\mathbf{x}_1;\theta)}{p(\mathbf{x}_2;\theta)}$ is not a function of $\theta)$:

$$\frac{f(x_1, y_1; p, q)}{f(x_2, y_2; p, q)} = \frac{\binom{N}{y_1} p^{y_1}(1-p)^{N-y_1}\binom{y_1}{x_1} q^{x_1}(1-q)^{y_1-x_1}}{\binom{N}{y_2} p^{y_2}(1-p)^{N-y_2}\binom{y_2}{x_2} q^{x_2}(1-q)^{y_2-x_2}} = 1,$$

  which is not a function of p or q.
- In fact, we can see above that to cancel out p and q from the equation, $x_1 = x_2$ and $y_1 = y_2$.
- Therefore, (x, y) are minimal sufficient statistics of p an q.

Why bother with sufficient statistics?

- Sufficient statistics save on memory, which is useful for hardware/microchip applications or when dealing with big data.
- In future lectures, we will be optimizing $p(\mathbf{x};\theta) = h(\mathbf{x})g(T(\mathbf{x}),\theta)$ with respect to θ. The factorization theorem allows us to remove $h(\mathbf{x})$.
- There are many performance results related to sufficient statistics, for example when using certain types of probability distributions we can use the sufficient statistic to find the best estimator.