ESE 524: More Applications of Sufficient Statistics

January 24, 2019

"Something something Statistics, Something Something Death Star, ... Something Something COMPLETE"

- Let $\mathbf{x} = [x[0], ..., x[N-1]]$ be a set of samples from a probability distribution $p(\mathbf{x}; \theta)$ with unknown parameter θ.
- A sufficient statistic $T(\mathbf{x})$ is Complete if for all functions g the following statement holds:

$$\mathbb{E}[g(T); \theta] = 0 \implies p(g(T(\mathbf{x})) = 0; \theta) = 1 \forall \theta$$

- This condition ensures that for different values of θ, the probability distributions in the model are different from each other.
- For example, normal RV's with different means are distinct from each other, and $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$ is complete and sufficient to estimate the means.

Example: Bernoulli Trials

- Let $\mathbf{x}$ be i.i.d. samples of a bernoulli distribution with probability θ. So $x[n] \in \{0, 1\}$
- The probability distribution of individual $x[n]$ is $p(x[n]; \theta) = \theta^{x[n]}(1-\theta)^{1-x[n]}$
- Then the joint probability distribution is found by multiplying the individual distributions:
$$p(\mathbf{x}; \theta) = \theta^{\sum_{n=0}^{N-1} x[n]}(1-\theta)^{N-\sum_{n=0}^{N-1} x[n]}$$
- From the Factorization Theorem, $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$ is a sufficient statistic.
- Is it complete?

Example: Bernoulli Trials Cont.

- What is the distribution of a sum of Bernoulli Random Variables?

Example: Bernoulli Trials Cont.

- What is the distribution of a sum of Bernoulli Random Variables?
- The distribution of T is:

$$p(T(\mathbf{x});\theta) = \binom{N}{T}\theta^T(1-\theta)^{N-T}$$

- Check the condition for completeness by assuming $\mathbb{E}[g(T);\theta] = 0$ (see board for details).

Example: Bernoulli Trials Cont.

- What is the distribution of a sum of Bernoulli Random Variables?
- The distribution of T is:

$$p(T(\mathbf{x});\theta) = \binom{N}{T}\theta^T(1-\theta)^{N-T}$$

- Check the condition for completeness by assuming $\mathbb{E}[g(T);\theta] = 0$ (see board for details).
- So T is a complete statistic. Why is that good?

Using Complete Statistics as Estimators

- Let $\hat{\theta}$ be an unbiased estimator you have found.
- Let $T(\mathbf{x})$ be a sufficient statistic.

Rao-Blackwell-Lehmann-Scheffe Theorem.
The new estimator $\hat{\theta}_T = \mathbb{E}[\hat{\theta}(\mathbf{x})|T(\mathbf{x})]$ is also an unbiased estimator of θ, but $\mathrm{var}(\hat{\theta}_T) \leq \mathrm{var}(\hat{\theta})$.
If T is complete, $\hat{\theta}_T$ is the Minimum Variance Unbiased Estimator.

Using Complete Statistics as Estimators

- Let $\hat{\theta}$ be an unbiased estimator you have found.
- Let $T(\mathbf{x})$ be a sufficient statistic.

Rao-Blackwell-Lehmann-Scheffe Theorem.
The new estimator $\hat{\theta}_T = \mathbb{E}[\hat{\theta}(\mathbf{x})|T(\mathbf{x})]$ is also an unbiased estimator of θ, but $\mathrm{var}(\hat{\theta}_T) \leq \mathrm{var}(\hat{\theta})$.
If T is complete, $\hat{\theta}_T$ is the Minimum Variance Unbiased Estimator.

- For a proof, see appendix 5A in Kay.
- In the case where we can find a complete sufficient statistic that is an unbiased estimator (e.g. $\frac{1}{N}\sum_{n=0}^{N-1} x[n]$ for gaussian means), we know that it is the MVU estimator.

Real World Application: Wilderness Search and Rescue

- You are hiking in Elephant Rock State Park, and become lost in the woods.
- How can the park rangers find you?
- Teams first build a probabilistic model centered around your last known location (LKP).
- The International Search and Rescue Incident Database has a compiled set of incident reports.
- People are grouped by age, gender, where they were lost (forest, mountain, desert,...), etc.

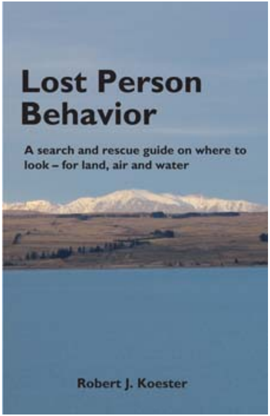

Figure 1: Book of compiled statistics for many different categories of lost subjects.

Discrete Sufficient Statistics

- The data samples we have are the locations where people were found in latitude and longitude, which we will denote as $\mathrm{FP}[n]$ for subject n.
- The simplest possible search model creates concentric rings outward from the last known location using distance as the data point:

$$x[n] = \mathrm{haversine}(\mathrm{LKP[n]}, \mathrm{FP}[n])$$

  where $\mathrm{haversine}(\cdot)$ is the haversine distance used to convert map coordinates to physical distance.
- Our model is a combination of uniform distributions given by:

$$p(\mathbf{x} = x; \theta_1, \theta_2, \theta_3, \theta_4) = \frac{0.25}{\mathrm{Area}_1} I_{[0,\theta_1]}(x) + \frac{0.25}{\mathrm{Area}_2} I_{[\theta_1,\theta_2]}(x) + \frac{.25}{\mathrm{Area}_3} I_{[\theta_2,\theta_3]}(x) +$$

$$\frac{0.2}{\mathrm{Area}_4} I_{[\theta_3,\theta_4]}(x) + \frac{0.05}{\mathrm{Area}_5} I_{\theta_4,\infty}(x)$$

  where Area_i are the areas of each concentric ring.

Discrete Sufficient Statistics Cont.

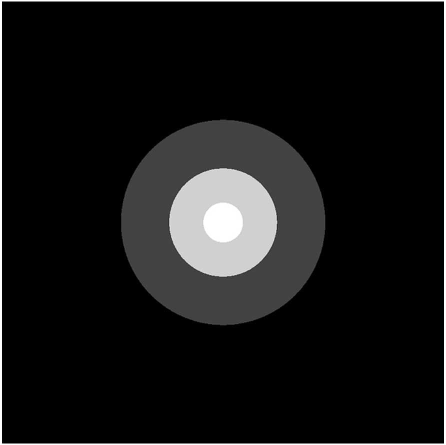

Figure 2: Probability Map Illustrating the "Distance Ring" model of Lost Person Behavior. Lighter colored pixels have higher probability than darker pixels. Roughly 25 % of hikers will be found in each concentric ring, with 5 % given to the rest of the world.

Discrete Sufficient Statistics Cont.

- How can we estimate the radius of each ring?
- Recall the Order Statistics, which organize $\mathbf{x}$ from lowest ($x^{(1)} = \min(\mathbf{x})$) to highest ($x^{(N)} = \max(\mathbf{x})$).
- For a normal uniform distribution, we know that $x^{(N)}$ is sufficient to estimate the length of the distribution, so for this combination we can use

$$\mathbf{T} = [x^{(\frac{N}{4})}, x^{(\frac{N}{2})}, x^{(\frac{3N}{4})}, x^{(\frac{95N}{100})}]$$

  to estimate the boundary radii of each ring.
- Note: If N is not divisible we round up or down as necessary.
- These statistics have been computed for a plethora of categories in Bob's book.

A Continuous Distance Model of Lost Person Behavior

- A more complicated model seeks to fit a continuous distribution to the distance data.
- The Lognormal distribution is a long tailed distribution which looked like a good fit on histograms of distance data.
- With mean μ and variance σ^2, the joint lognormal pdf is given as:

$$\frac{1}{\Pi_{n=0}^{N-1} x[n](2\pi\sigma^2)^{N/2}} \exp(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(\ln(x[n]) - \mu)^2)$$

What are the Sufficient Statistics for μ and σ^2?

Lognormal Models of Distance Behavior

- This distribution works well, but does not capture that 10% of people are found at $x[n] = 0$.
- So I modified the individual distribution:

$$p(x[n]; \mu, \sigma^2) = 0.1\delta(x[n]) + 0.9\frac{1}{x\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}(\ln(x[n]) - \mu)^2)$$

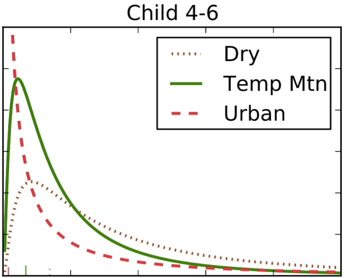

Figure 3: Lognormal Distributions fitted for Children Ages 4-6 lost in Dry, Urban, and Temperate environments. Most of the time, the lognormal distribution misses the 10% probability to be found at the LKP.

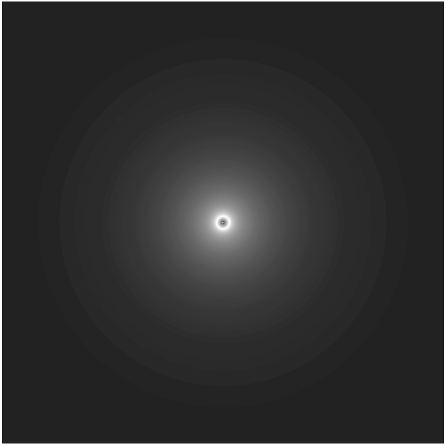

Figure 4: Augmented Lognormal Distribution for a hiker lost in Arizona. Note that if you find compare this to the discrete map, approximately the same amount of probability is distributed in the areas bounded by the rings, but this model applies more probability to the center of each ring.