Finding Distributions

2-18-2011

Opening Discussion

- Do you have any questions about the quiz?
- What did we talk about last class?
- Minute essay comments:
  - Motivation for distributions. Similarity of distributions.

Determining Independence

- One of the big things we want to know with distributions is whether the values are independent.
- Plotting estimate of ρ_j vs j.

$$\hat{\rho}_j = \frac{\hat{C}_j}{S^2(n)}, \hat{C}_j = \frac{\sum_{i-1}^{n-j} (X_i - \bar{X}(n))(X_{i+j} - \bar{X}(n))}{n-j}$$

- Scatter plot of X_i vs. X_{i+1}.

Finding the Right Distribution

- The book lists three different approaches to determining a family of distributions.
- Summary Stats
  - Min and max
  - Mean
  - Median
  - Variance
  - Coefficient of variation (continuous only)

$$cv = \frac{\sqrt{\sigma^2}}{\mu}, \hat{cv}(n) = \frac{\sqrt{S^2(n)}}{\bar{X}(n)}$$

Continued

- Lexis ratio (discrete only)

$$\tau = \frac{\sigma^2}{\mu}, \hat{\tau}(n) = \frac{S^2(n)}{\bar{X}(n)}$$

- Skewness

$$\nu = \frac{E\left[(X-\mu)^3\right]}{(\sigma^2)^{3/2}}, \hat{\nu}(n) = \frac{\sum_{i=1}^{n}\left[X_i - \bar{X}(n)\right]^3 / n}{(S^2(n))^{3/2}}$$

Histograms

- Make a histogram of the values.
- Have to be careful when picking bin size.
- Let's look at an example of this in SwiftVis.

Linear Combination

- It is possible that you have values that come not from a single distribution, but from two or more summed together.
- That might be evidenced by multiple peaks in the distribution.

Quartile Summary

- You can use a type of plot called a box plot to look at quartile information.
- This gives you a feel for the shape of a distribution without some of the challenges associated with histograms.

Getting Parameters

- Once we have a family of distributions we can use X_i to estimate the parameters.
- This is where the maximum-likelihood estimators (MLEs) come into play. They are an alternative to things like a least-squares fit.
- MLE maximizes the likelihood function.

$$L(\theta) = p_\theta(X_1)\, p_\theta(X_2) \ldots p_\theta(X_n)$$
$$L(\theta) = f_\theta(X_1) f_\theta(X_2) \ldots f_\theta(X_n)$$

Maximization Techniques

- Ideally you can do the algebra and solve for the derivative being zero.
- In theory you could plot the function and find the maximum value that way when the algebra won't work, but the book gives algebra for all the distributions in this chapter.
- Can do sensitivity analysis to determine if you have found the value to sufficient accuracy.

Minute Essay

- The midterm is a week from today. Do you have any thoughts on it?