

Distribution Quality

2-23-2011

Opening Discussion

- What did we talk about last class?
- Minute essay comments:
 - How much coding will be on the midterm?
 - How similar will the test be to the quizzes?
 - Review sheet.
 - ODE by hand?

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.

$$L(\theta) = p_{\theta}(X_1) p_{\theta}(X_2) \dots p_{\theta}(X_n)$$

$$L(\theta) = f_{\theta}(X_1) f_{\theta}(X_2) \dots f_{\theta}(X_n)$$

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.

Determining Quality of Fit Distribution

- Now that you have picked a distribution to use to represent the data and have parameters for it, the question is, “Is it good enough?”

Heuristic Procedures

- Density-Histogram plots
 - Over plot the histogram of your data with the distribution.
- Distribution-Function-Differences plots
 - Plot the difference between an empirical $F(x)$ and the fitted $F(x)$.
- Probability plots
 - Q-Q plots
 - P-P plots
 - Both show differences between the $F(x)$ plots.

Goodness of Fit Tests

- Chi-Square tests
 - Break distribution up into chunks and see if the number of points in each chunk matches.

$$\chi^2 = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j}$$

- Kolmogorov-Smirnov tests

$$D_n = \sup_x (|F_n(x) - \hat{F}_n(x)|)$$

ExpertFit Software

- Lots of the things we have been talking about are very tedious to do.
- Your book contains a student version of the ExpertFit software.
- Section 6.7 runs you through a tutorial on how to use it.

Minute Essay

- Should we have a review session and if so, when?
- Assignment #3 is due Friday.