Goodness of Fit

Shyamalkumar, N. D.

Dept. of Statistics and Actuarial Science, The University of Iowa, Iowa City, USA.

The Goodness of Fit Problem

- We discussed a small subset of count distributions
- None of them may be a good/useful model for the data at hand
  - *Goodness of Fit Problem* - determining if one is adequate
- Need for a method to make sound decision on the fit
  - Introducing one such is the goal of the section
- We do so via an example

Example: Singapore Automobile Data

- A 1993 portfolio from a major insurance company in Singapore
- $7,483$ automobile insurance policies
- Policy level data - count variable is the number of accidents
- Maximum of 3 accidents per policy observed
- Average of 69.89 accidents per $1,000$ policies ($\overline{N} = 0.06989$)

Observed Accident Counts per Policy

Count (k)	Observed (m_k)
0	$6,996$
1	455
2	28
3	4
Total	$7,483$

Fitting a Poisson

- With the Poisson distribution
  - The MLE of λ is $\widehat{\lambda} = \overline{N}$.
  - Fitted probabilities $\widehat{p}_k$ below use $\widehat{\lambda}$
  - Fitted counts are 7,483 times the fitted probabilities
- Created a cell for counts ≥ 4
  - To account for remaining fitted probability

Table. Comparison of Observed to Fitted Counts

Count (k)	Observed (m_k)	Fitted Counts using the Poisson Distribution($n\widehat{p}_k$)
0	6,996	6,977.86
1	455	487.70
2	28	17.04
3	4	0.40
≥ 4	0	0.01
Total	7,483	7,483.00

Adequacy of the Poisson Model

- For goodness of fit, consider *Pearson's chi-square statistic*

$$\sum_k \frac{(m_k - n\widehat{p}_k)^2}{n\widehat{p}_k}.$$

- Has an asymptotic chi-square distribution
  - If the Poisson distribution is the correct model
- The degrees of freedom (df) equals
  - the number of cells minus one minus the number of estimated parameters.
- For the Singapore data
  - $df = 5 - 1 - 1 = 3$; 99-th %ile equals 11.34487
  - The Pearson's statistic equals 41.98 (> 11.34487)
  - The basic Poisson model is **inadequate**
- In the exercise below, you will fit a zero-inflated Poisson