Pages 293-306: To Catch a Thief. Use of graphics for time series. General Strategy of Estimation. Probability. Histogram. Moving Average. Cumulation.

Cell phone activity is monitored by the cell phone company to try to detect and eliminate fraudulent use of the phones, and to try to reduce the lost revenue from such activity.   Because of the large volume of such activity (legitimate and fraudulent), this monitoring needs to be done electronically, at least at first.

Each cell phone account has a call signature defined by its typical use. This is based on the typical use of the cell phone which is defined separately for each account. This signature is based on the proportion of calls of certain types (e.g. local or roam),or on the distribution of call durations, and possibly other characteristics.

The phone company has determined signatures for fraudulent use as well as for each separate typical legal use. Each subsequent call is scored for its probability under the account signature and also for its probability under one or more fraud signatures, and the ratio of the two scores is used to identify which calls are likely legitimate and which fraudulent.

There are a few technical details in this process:

Q1: Does the coloring of the graph in Figure 1 represent a finding of fraudulent calls or merely a characteristic of the calls?

Q2: Ignoring the possible difference between local and roamed calls, what would the account signature look like (see example in Fig 2.) before March 27? And after March 27?

Q3: Is the fraud obvious from Figure 1, and if so, why is a more complicated probabilistic method required?

Q4: What is wrong with an account signature based on a very small number of account characteristics, or a very large number of account characteristics?

Q5: The account signature is usually described by the "probability distribution of the account". What is meant by "the probability distribution of the account"?

Q6: How is the account signature updated when a new call is made?

Q7: Explain why the ratio of the typical score to fraudulent score is used to monitor an account. Why use the logarithm of this?

Q8: Is a fraudulent call used to update the account signature?

Q9: The fraud detection system described is typical of a decision making process?
Explain. (Hint: p 303).

Q10: What additional utility does the cumulation chart in Figure 3 provide?