A model for estimating the probability of a fraudulent act per person for a given state

By Joe Ganser

**Abstract:** The purpose of this model is to calculate the probability of fraud per person in a given state, based upon crime records (number of arrests of varying crimes) in each town of the state.

**Assumptions:**

1. The crime data available to use is the number of arrests for each type of crime across each town in each state. Stats concerning the nature of the convicts themselves, the local area income, etc are not present.
2. The number of actual frauds committed is greater than or equal to the number of frauds caught.;
3. Regions with high crime will have higher fraud rates
4. The actual number of frauds that happen in a given town can be estimated by assigning weights to the number of arrests for each crime in that town; i.e.
5. Weights are given by the probability of a particular crime happening and fraud happening for each town;
6. If a crime is not a fraud, the probability of both the crime happening and fraud happening in a given town is the product of the two probabilities happening independently .
7. Crimes can be categorized as two forms – fraud related crimes and non-fraud crimes.

**Derivation**

To derive this model, we begin with Baye’s theorem;

For a given state;

Now,

because a fraud by definition is a crime, so the probability of a crime given the event is fraud must be 1. Hence;

and

And

Therefor;

So the next step is to estimate the number of frauds;

Start with

Where is the probability of there being both fraud and the crime of type C in a given town. is the number of crimes of type c.

Thus, the probability equals 1 if the crime itself is a fraud, and otherwise it’s a product of two probabilities because they’re independent events.

Where P(F) = probability of a fraud being in the town

P(C) = probability of a crime C being in the town

We can then break the summation into two forms; over crimes that are frauds and crimes that aren’t frauds.

Which equals

And since

therefor

And so

Is the estimated number of frauds in a given town.

Now for a given state we add up all the estimated amount of frauds in each town, and divide by the population of the state.

So the probability of fraud per person in a given state is;

Or, more specifically;

Which is our final model.

QED