Detection of Suspicious Transactions using Bayesian Approach

PRN-21060641054,21060641034,21060641044,21060641030

Study of Background:
Money laundering (ML) is the process by which large amounts of illegally obtained money or other serious crimes, is given the appearance of having originated from a legitimate source. Suspicious activity reporting (SAR) is part of the anti-money laundry. The anomalous behaviours reported by SAR cover almost any activity that is out of the ordinary or that is not consistent with a customer's normal behaviour.

Binomial Distribution
It is a discrete probability distribution of the number of successes in a sequence of 'n' independent experiments, each asking a yes–no question, and each with its own Boolean-valued outcome: success (with probability p) or failure (with probability 1-p)

Beta distribution
It is a family of continuous probability distributions defined on the interval [0, 1] in terms of two positive parameters, denoted by alpha (α) and beta (β). The beta distribution is a suitable model for the random behaviour of percentages and proportions.
In Bayesian inference, the beta distribution is the conjugate prior probability distribution for the Bernoulli, binomial, negative binomial and geometric distributions.

Abstract:
This report presents a Beta-Binomial Bayesian Model approach to analyse customer's transactions in the financial industry and then to find the probable customers most likely to make any suspicious transactions (ST). The probability of a customer doing a ST is taken as p, whose prior distribution is beta with parameters a = 2 and b = 3 The proposed approach will give us a model for accurately predicting whether a customer will do a ST or not by comparing the empirical and theoretical Bayesian estimates under quadratic loss.

Problem Statement and its impact:
The development of an effective mechanism to detect suspicious transactions (STs) is a critical problem for financial institutions in their endeavour to prevent Anti Money laundering (AML) activities. The financial institution has the responsibility to file a report within 30 days regarding any account activity they deem to be suspicious or out of the ordinary. Suspicious activity reporting (SAR) allows law enforcement to detect patterns and trends in organized and personal financial crimes. This way they can anticipate criminal and fraudulent behaviour and counteract it before it escalates. Need for ST detection has increased and the same has found a new interest in the minds of budding data scientists and business analysts.

Proposed solution:
Analysis of the problem:

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

Approaching with Bayes theorem which describes the probability of person making ST based on prior knowledge of conditions that might be related

Where,
H = Hypotheses that the transaction made by the customer is a Suspicious Transaction (ST),
P(H) = The prior probability, is the estimate of H which we know follows beta distribution,
E = The event of transaction being made by the customers,
P(H/E) = probability of the transaction being an ST given that customer has made a transaction,

P(E/H) = probability of the customer making a transaction given that it is ST (equals 1 reason being whether the transaction is an ST or not, transaction being made becomes a sure event), and
P(E) = probability of the transaction being made by the customer (equals 1 because we have generated data assuming that the event of the transaction being made is a sure event).

Need for a Bayesian solution:
Bayesian statistics is applied when we have knowledge about our domain before we see any of the data. Bayesian inference provides a straightforward way to encode that belief into a prior probability distribution. It gives us mathematical tools to update our beliefs about random events in response to new data or evidence about those events. It interprets probability as a measure of belief or confidence that an individual may have in the occurrence of a specific event. For the above problem, we are considering prior information about the probability that a customer is making a suspicious transaction to identify such transactions. This will help financial companies/industries to analyse and make further decisions for such happenings.

Code and implementation with interpretation:
Let's consider the probability of a customer making a suspicious transaction as p, Where p has the prior distribution of beta with parameters α=2 and β=3.

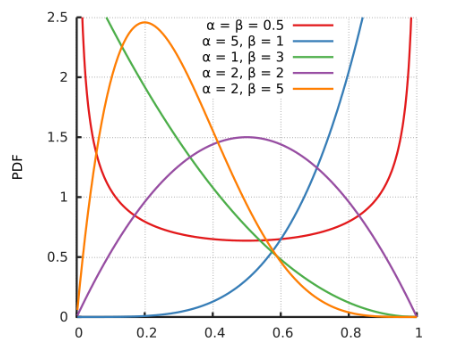
Here, we get pdf of p as

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

```
a=2  #alpha
b=3   #beta
p=rbeta(1,a,b)

#generating random numbers according to the beta
distribution

x=numeric(1000) #vector with 1000 zeros
set.seed(77)
```



For different values of α and β

Using this prior information, we are trying to estimate whether that person is actually doing a suspicious transaction or not using binomial distribution.

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

We are considering likelihood of x|p in order to provide insight into the relative compatibility of an observed suspicious transaction

```
for(i in 1:1000){
 p=rbeta(1,a,b)
 x[i]=rbinom(1,1,p)
}  #estimating p
#under quadratic loss,mean of posterior is Baye's estimator
phat=mean(x)
phat
## [1] 0.382
```

Posterior probability of p|x is obtained using binomial-beta model and the posterior distribution is beta distribution with parameters $(x+\alpha), (\alpha+\beta+1)$
Under squared-error loss, mean of posterior distribution is considered as Bayes estimate.

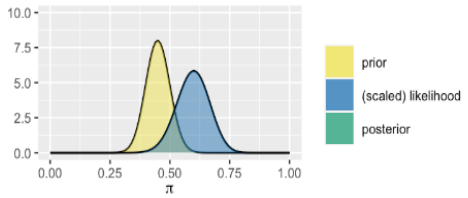
Therefore, Bayesian estimate for p under quadratic loss for a single outcome x is $\frac{x+\alpha}{\alpha+\beta+1}$ (Mean of posterior beta)

```
pm=numeric(1000)
x=numeric(1000)
set.seed(77)
for(i in 1:1000){
 p=rbeta(1,a,b)
 x[i]=rbinom(1,1,p)          #binomial distribution
 pm[i]=(x[i]+a)/(a+b+1)      #beta- binomial model
}
mean(x)                   ## [1] 0.382

e=mean(pm)
e  #empirical mean
## [1] 0.397
t= (a+sum(x))/(a+b+length(x))
t  #theoretical mean
## [1] 0.3820896
(e-t)/t  #comparison
## [1] 0.03902344
#we can see that there's not much difference between empirical and theoretical mean.
```

Visual understanding of Bayesian model

$$Y|\pi \sim \mathrm{Bin}(50, \pi)$$
$$\pi \sim \mathrm{Beta}(45, 55).$$



Now we are considering sample of 20 people, to analyse suspicious transactions and also to test the model

```
set.seed(77)
xp=numeric(1000)
pm=numeric(1000)
n=20  #sample of 20 people suspected for suspicious transaction

for(i in 1:1000){
 p=rbeta(1,a,b)
 xp[i]=rbinom(1,n,p)  #binomial distribution for n people
 pm[i]=((xp[i])+a)/(a+b+n)   #beta-binomial model
}   #under quadratic loss, mean of the posterior distribution is Bayes estimate
em=mean(pm)
em   #empirical

## [1] 0.39768
(a+mean(xp))/(a+b+n)
## [1] 0.39768
tm=(a+sum(xp))/(a+b+n*length(x))
tm   #theoretical
## [1] 0.3971007
(em-tm)/tm #comparison
## [1] 0.001458761
```

This shows that there is not much difference between empirical and theoretical mean so we can use it for estimating suspicious transaction.

CONCLUSION:

Probability of a person making a suspicious transaction is coming out to be almost 0.4, which is according to the parameters we have chosen.

This output changes when worked on actual real data with estimated parameters and hence such probability might be less than 0.1 also out of large no. of transactions.

Using simulated data, we can conclude that our estimation for sample of 20 people is similar to the one person doing a suspicious transaction. So, we can increase the sample size and interpret for 'n' no. of transactions helping financial companies to predict the behaviour of their customers in order to reduce anti-money laundering issue.

Recommendation:
We can use beta-binomial model considering prior information of p being a probability of a person dying from a particular disease and then analyse the sample of size ‘n’ as a health insurance company to reduce cost and anticipating probability of dying.
We can use machine learning models to detect ST depending upon the various factors/ variables provided by the company/ financial industry. Classification Models can be used to differ between normal transactions and ST. Different approaches like rational, collaborative, issue-based can be used.

Sources:

https://www.bayesrulesbook.com/chapter-3.html#the-binomial-data-model-likelihood-function
https://en.wikipedia.org/wiki/Binomial_distribution
https://en.wikipedia.org/wiki/Beta_distribution
A Bayesian approach for suspicious financial activity reporting, January 2013, International Journal of Computers and Applications
https://www.investopedia.com/terms/a/aml.asp#toc-what-is-anti-money-laundering-aml