Interpreting the GINI

Consider a classifier that outputs a score *s*.

By recording the score when a transaction is Legal or Fraud it is possible to determine the distributions of $P(s\,|\,F)$ and $P(s\,|\,L)$.

For a particular threshold *t* we can then write:

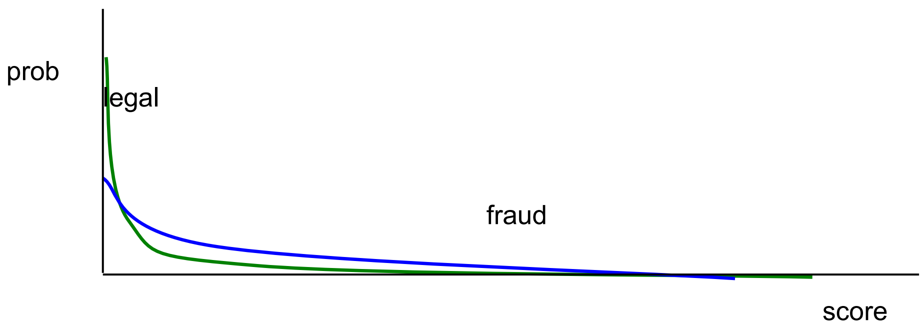
$$TP = N_F \int_t^1 P(s\,|\,F)ds \quad FN = N_F \int_0^t P(s\,|\,F)ds$$
$$FP = N_L \int_t^1 P(s\,|\,L)ds \quad TN = N_L \int_0^t P(s\,|\,L)ds \qquad (2.1)$$

Now consider the True Positive Fraction (the Y-axis of the ROC graph) and the False Positive Fraction (the X-axis of the ROC graph).

$$TPF = \frac{TP}{TP+FN} = \frac{\int_t^1 P(s\,|\,F)ds}{\int_t^1 P(s\,|\,F)ds + \int_0^t P(s\,|\,F)ds} = \frac{\int_t^1 P(s\,|\,F)ds}{\int_0^1 P(s\,|\,F)ds} = \int_t^1 P(s\,|\,F)ds \qquad (2.2)$$

$$FPF = \frac{FP}{FP+TN} = \frac{\int_t^1 P(s\,|\,L)ds}{\int_t^1 P(s\,|\,L)ds + \int_0^t P(s\,|\,L)ds} = \frac{\int_t^1 P(s\,|\,L)ds}{\int_0^1 P(s\,|\,L)ds} = \int_t^1 P(s\,|\,L)ds \qquad (2.3)$$

The graph below shows the form of $P(s\,|\,F)$ and $P(s\,|\,L)$ as output by *Detect*.



Experiments have shown that this data can be closely modeled using a simple power distribution.

The Power PDF

Pdf $$p(x) = ax^{a-1} \quad \text{or} \quad p(x) = \frac{\mu}{1-\mu} x^{\frac{2\mu-1}{1-\mu}} \tag{2.4}$$

Mean $$\mu = E(x) = \int_0^1 x p(x) dx = \left[\frac{\alpha}{\alpha+1} x^{\alpha+1} \right]_0^1 = \frac{\alpha}{\alpha+1} \tag{2.6}$$

Variance $$v = E\left((x-\mu)^2 \right) = \int_0^1 \left(x - \frac{\alpha}{\alpha+1} \right)^2 p(x) dx = \frac{\alpha}{\alpha+2} - \left(\frac{\alpha}{\alpha+1} \right)^2 \tag{2.7}$$

Note: the power is expressed as *a – 1* simply because this greatly simplifies some of the later derived expressions and reveals more clearly some of the symmetries.

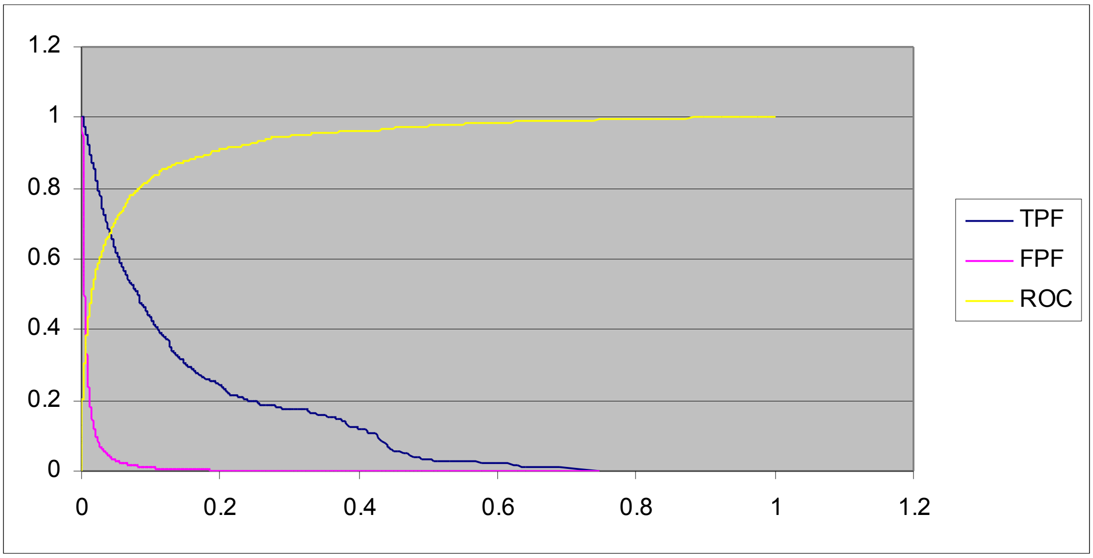
So, using the power pdf defined by 2.4 we can write:

$$P(s \mid F) = \alpha(1-s)^{\alpha-1} \quad \text{and} \quad P(s \mid L) = \beta(1-s)^{\beta-1}$$

And equations 2.2 and 2.3 can now be written as:

$$u(t) = TPF(t) = \int_t^1 P(s \mid F) ds = \int_t^1 \alpha(1-s)^{\alpha-1} ds = (1-t)^{\alpha} \tag{2.8}$$

$$v(t) = FPF(t) = \int_t^1 P(s \mid L) ds = \int_t^1 \beta(1-s)^{\beta-1} ds = (1-t)^{\beta} \tag{2.9}$$



The above three curves for a typical calibration with GINI = 0.871

a. Y axis = TPF X axis = threshold
b. Y axis = FPF X axis = threshold
c. Y axis = TPF X axis = FPF

Using 2.8 and 2.9 the ROC graph can then be written as:

$$u = v^{\frac{\alpha}{\beta}} \tag{2.10}$$

then using the fact that the GINI is twice the area under the ROC minus one, we have:

$$\text{GINI} = g = 2A - 1 = 2\int_0^1 v^{\alpha/\beta} dv - 1 = 2\left[\frac{v^{\frac{\alpha}{\beta}+1}}{\frac{\alpha}{\beta}+1}\right]_0^1 - 1 = \frac{2}{\frac{\alpha}{\beta}+1} - 1 = \frac{\beta-\alpha}{\beta+\alpha} \qquad (2.11)$$

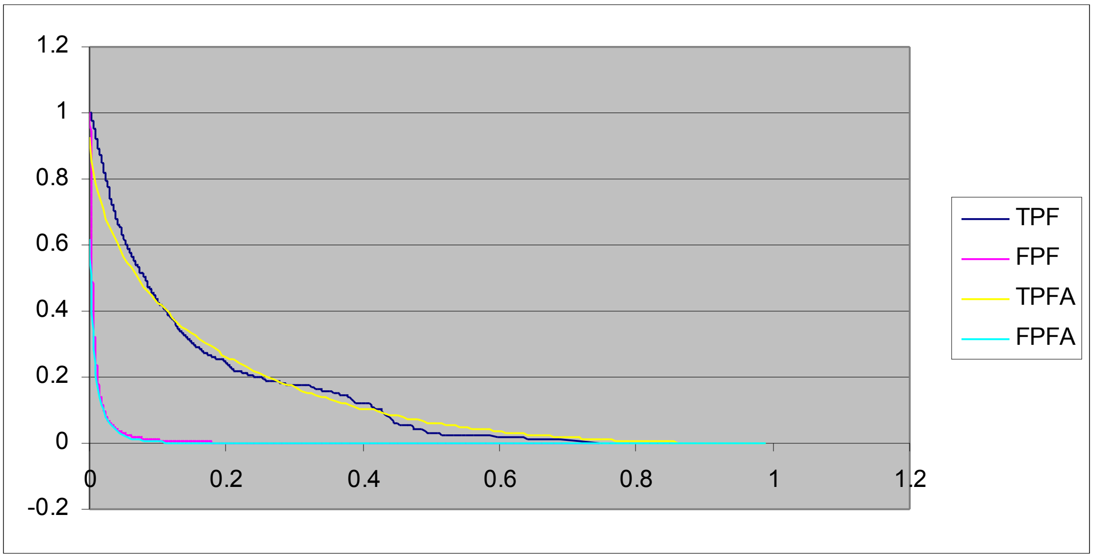
And using the power pdf's mean from equation 2.6 we have:

$$\text{GINI} = g = \frac{\mu_L - \mu_F}{\mu_L + \mu_F - 2\mu_L\mu_F}$$

Also notice that:

$$\frac{\alpha}{\beta} = \frac{1-g}{1+g} \text{ and hence we have } \boxed{u = v^{\frac{1-g}{1+g}}} \qquad (2.12)$$

which is an approximation of the ROC function expressed in terms of the GINI.



The above graph shows how the equations A1 and A2 approximate TPF(t) and FPF(t) with $\alpha = 6.0 \;\; \beta = 90.0$

Using A4 we can immediately compute the GINI = $\frac{90-6}{90+6} = 0.875$ which is very close to the measured GINI of 0.871

The Relationship of the GINI to other measures

$$\frac{v}{u} = F\left(\frac{1}{R} - 1\right)$$

The relationship between GINI and TPF and Fraud Alert Rate (R) can now be calculated as we can show (see Appendix B) that $\frac{v}{u} = F\left(\frac{1}{R} - 1\right)$ so we can write $u^{\frac{2g}{1-g}} = F\left(\frac{1}{R} - 1\right)$

Note: False Positive Ratio (FPR) = $\frac{1}{R} - 1$

and hence $\frac{2g}{1-g}\log u = \log\left(F\left(\frac{1}{R} - 1\right)\right)$

and hence:

$$g = \frac{1}{1 + 2\frac{\log u}{\log\left(F\left(\frac{1}{R} - 1\right)\right)}} \quad \text{or} \quad u = \left(F\left(\frac{1}{R} - 1\right)\right)^{\frac{1-g}{2g}}$$

We can rewrite this equation more succinctly as: $u = e^{(g-1)k/2}$ where $k = \log\left(F\left(\frac{1}{R} - 1\right)\right)$

which shows that by holding the Fraud Alert Rate constant there is an exponential relationship between the TPF and the GINI.

Also, most importantly, the rate of change of TPF increases as the GINI increases as $\frac{(g-1)k}{2} > 0$, as $0 < g < 1$ and $k < 0$. This is borne out by experience and the recent observations that a small increase in the GINI can result in a big increase in other performance measures like the True Positive Fraction when compared at the same Fraud Alert Rate or False Positive Ratio.

This is an important result:

> To achieve the a target performance of 70% fraud detected with false positive ratio of 30:1 we need to have a ***GINI greater than 0.89***
>
> This result closely reflects experience.

Further, if we look at the overall alert-rate *A*:

$$A = \frac{TP + FP}{N} = \frac{TPF.N_F + FPF.N_F}{N} = \frac{uN_F + vN_F}{N}$$

we can re-arrange this:

$$A = (u+v) + (u-v)\frac{N_F - N_L}{N_F + N_L}$$

and as we know that $N_F \ll N_L$ we can write: $\frac{N_F - N_L}{N_F + N_L} \simeq -1$ and $A \simeq (u+v) - (u-v) = 2v$

and hence $u \simeq \left(\frac{A}{2}\right)^{\frac{g+1}{g-1}}$

Appendix B – ROC curves and Fraud Alert Rate

The False Positive Ratio is the ratio of the number of frauds detected to the number of total number of alerts (Fraud Alert Ratio) This can be written as:

$$FAR = R = \frac{TP}{TP+FP} = \frac{1}{1+FP/TP} \qquad \text{(B1)}$$

A ROC curve is a plot of:
$Y = TPF$ (True Positive Fraction) against $X = FPF$ (False Positive Fraction) where

$$TPF = \frac{TP}{TP+FN} \quad \text{and} \quad FPF = \frac{FP}{TN+FP} \qquad \text{(B2)}$$

From the ROC definitions (B2) we can write:

$$\frac{Y}{X} = \frac{TP}{FP}\frac{TN+FP}{TP+FN} = \frac{TP}{FP}\frac{T-(TP+FN)}{TP+FN} = \frac{TP}{FP}\left(\frac{T}{TP+FN}-1\right) \qquad \text{(B3)}$$

where $T = TN+TP+FN+FP$

The term $\frac{T}{TP+FN}$ is the inverse of the overall fraction of fraud, call this *F*

From (A1) we also have :

$$\frac{FP}{TP} = \frac{1}{R}-1$$

(Note: this relates False Alert Ratio to a definition of False Positive Ratio)

So (A3) can be re-written as:

$$\frac{Y}{X} = \left(\frac{1}{F}-1\right)\left(\frac{1}{R}-1\right)^{-1} \qquad \text{(B4)}$$

and

$$R = \left(1+\frac{X}{Y}\left(\frac{1}{F}-1\right)\right)^{-1} \qquad \text{(B5)}$$

Given a value for *F* of about 0.0012 (0.12%) then $\frac{1}{F} \gg 1$ so we can write:

$$R \approx \left(1+\frac{X}{FY}\right)^{-1} \quad \text{and} \quad \frac{Y}{X} \approx \left(F\left(\frac{1}{R}-1\right)\right)^{-1} \qquad \text{(B6)}$$