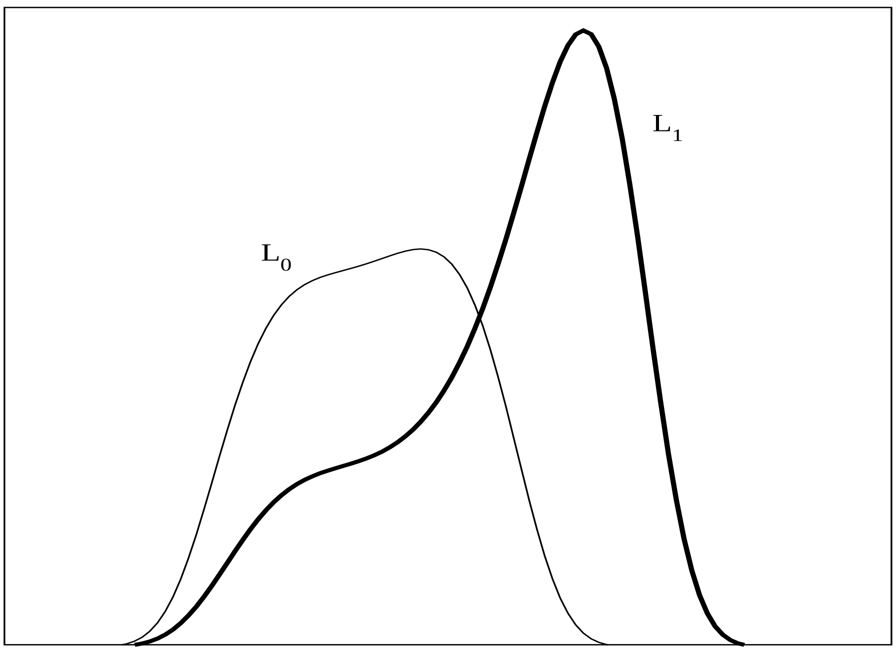
On the ROC Curve and the Area Thereunder

Caren Marzban, Univ. of Oklahoma and Washington
http://www.nhn.ou.edu/~marzban

Performance is a multifaceted thing.
ROC plot is a multidimensional thing.

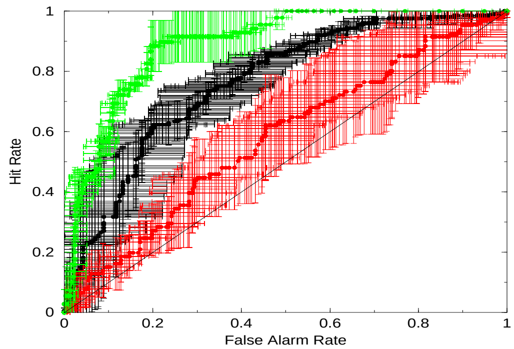


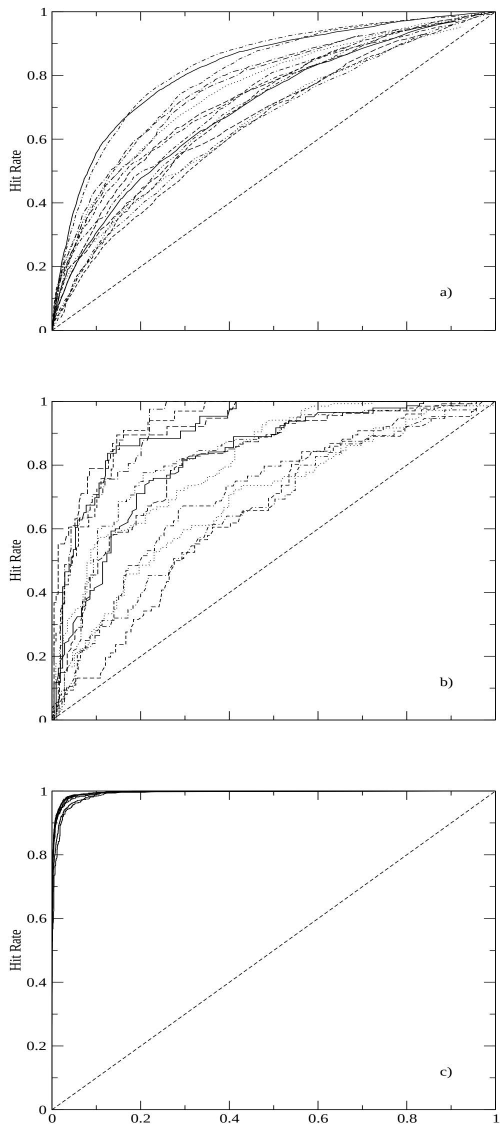
$$H = \int_t^{x_{max}} L_1(x)\, dx \quad , \quad F = \int_t^{x_{max}} L_0(x)\, dx$$
$$t = \text{decision threshold}, x_{max} = \infty \text{ or } 1.$$

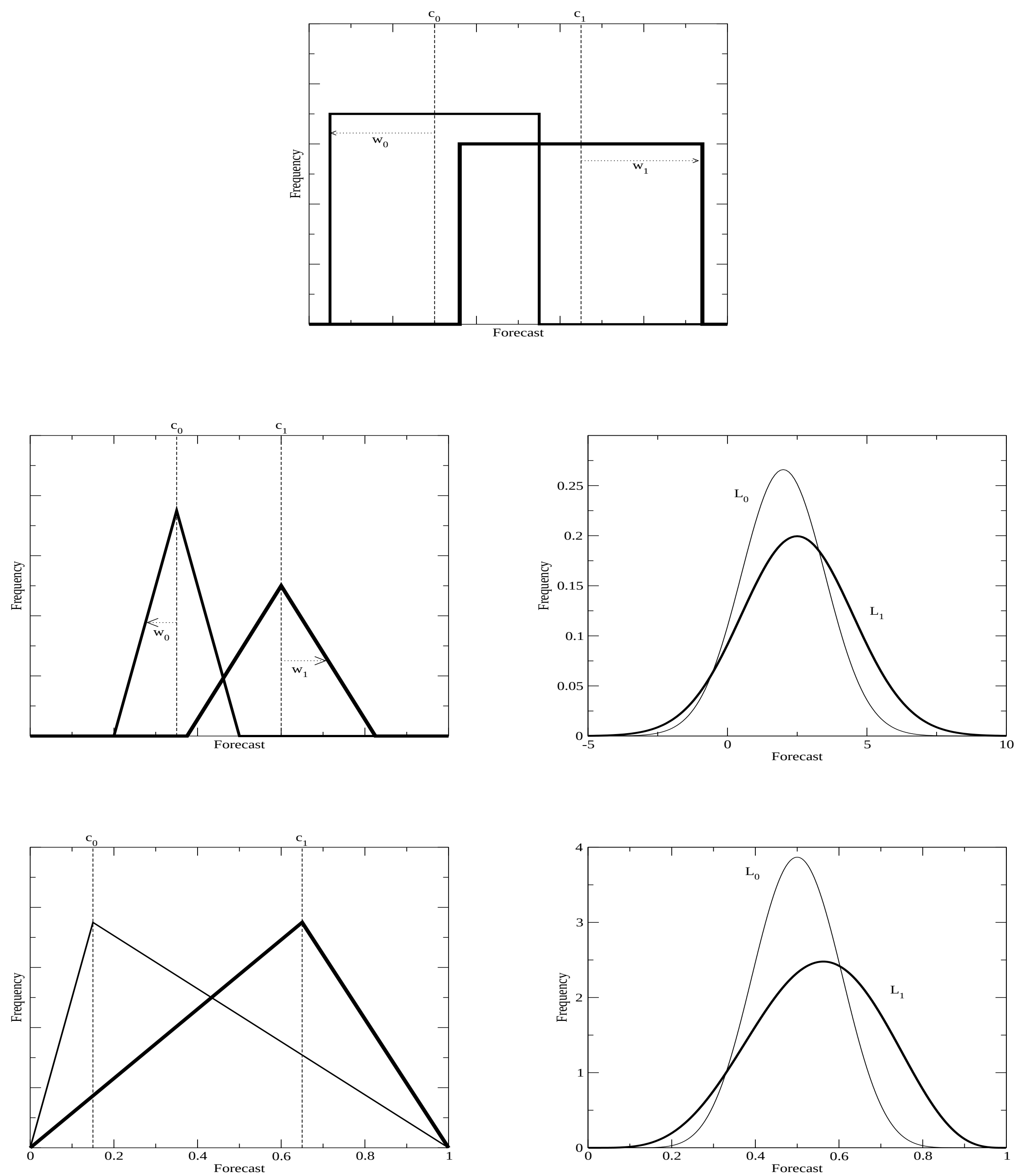
Q: Shape of the ROC $\leftrightarrow$ underlying distributions?
Q: AUC ?

Back of the envelope calculation.

Real-world Examples







c0
c1
w0
w1
Frequency
Forecast
L0
L1
0.25
0.2
0.15
0.1
0.05
0
-5
0
5
10
4
3
2
1
0
0.2
0.4
0.6
0.8
1

Uniform:

$$F = \frac{c_0 + w_0 - t}{2w_0}, \quad H = \frac{c_1 + w_1 - t}{2w_1},$$

$$\begin{aligned} H &= \frac{w_0}{w_1} \; F \; + \; \frac{\delta c + \delta w}{2w_1} \; , \\ \delta c &= c_1 - c_0 \;\text{ and }\; \delta w = w_1 - w_0 \end{aligned}$$

$$\begin{aligned} AUC &= 1 - \frac{1}{8}\left(\frac{\Delta}{\sqrt{w_0 w_1}}\right)^2 \\ \Delta &= \delta c - (w_0 + w_1). \end{aligned}$$

AUC selects for narrow-width and well-separated L's.

Triangular with unconstrained support:

$$F = \frac{1}{2}(\frac{c_0 + w_0 - t}{w_0})^2, \quad H = 1 - \frac{1}{2}(\frac{t - c_1 + w_1}{w_1})^2 \quad .$$

$$H = 1 - \frac{1}{2}(\frac{\Delta - w_0\sqrt{2F}}{w_1})^2 \quad ,$$

$$AUC = 1 - \frac{1}{8}(\frac{\Delta}{\sqrt{w_0 w_1}})^4$$

Gaussian:

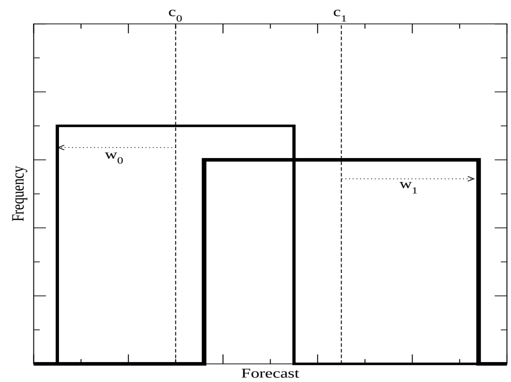
$$F = \Phi(\frac{c_0 - t}{w_0}) \ , \quad H = \Phi(\frac{c_1 - t}{w_1}),$$

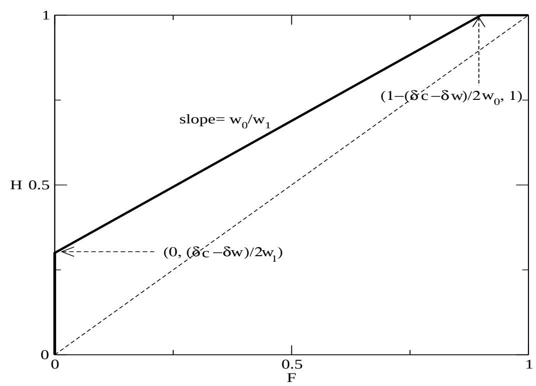
$$H = \Phi[\frac{\delta c}{w_1} - \frac{w_0}{w_1}\Phi^{-1}(F)],$$

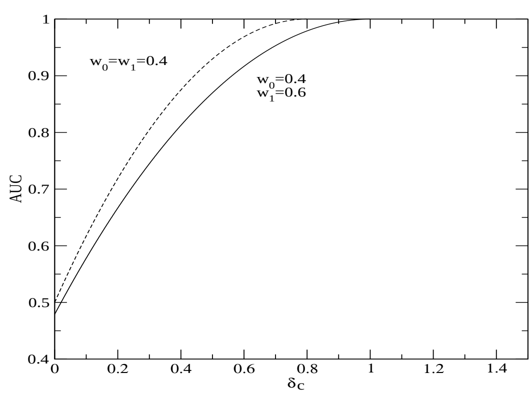
ROC curve can even cross diagonal!

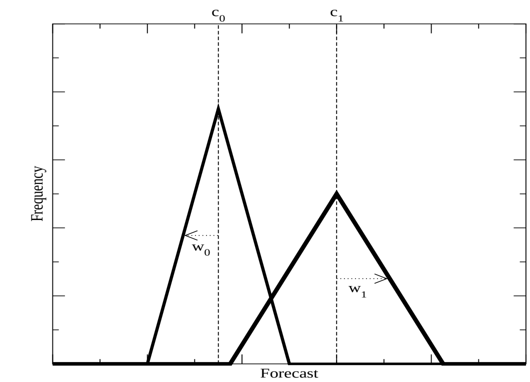
$$AUC = \Phi(\frac{\delta c}{\sqrt{w_0^2 + w_1^2}}) \ .$$

Etc. Etc.

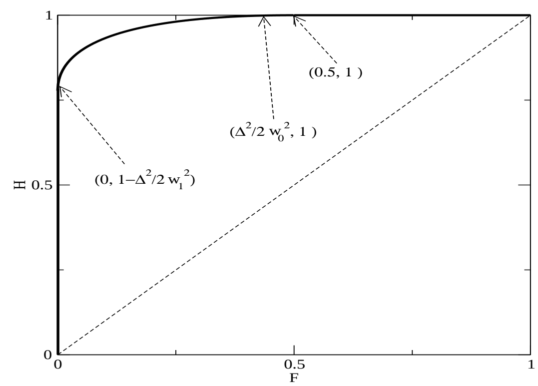

c0
c1
w0
w1
Frequency
Forecast


slope= w0/w1
(1−(δc−δw)/2w0, 1)
(0, (δc−δw)/2w1)
H
F
0
0.5
1

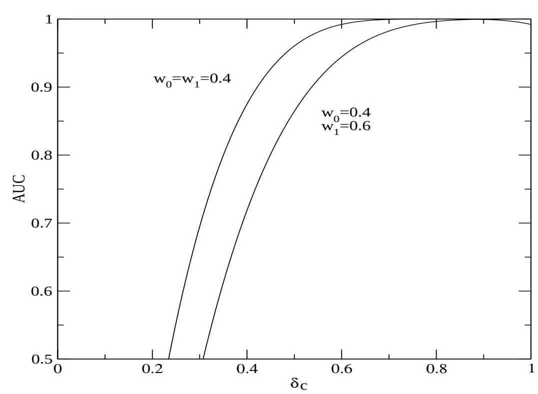

w0=w1=0.4
w0=0.4
w1=0.6
AUC
δc
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1
1.2
1.4



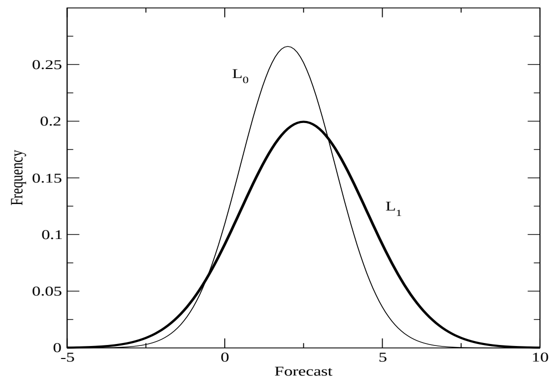
c_0
c_1
Frequency
w_0
w_1
Forecast



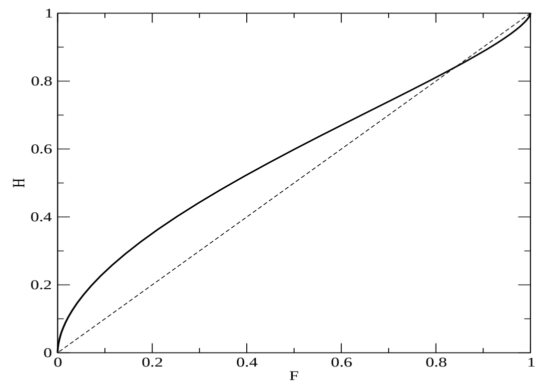
1
0.5
0
H
F
(0.5, 1)
($\Delta^2/2\ w_0^2$, 1)
(0, $1-\Delta^2/2\ w_1^2$)



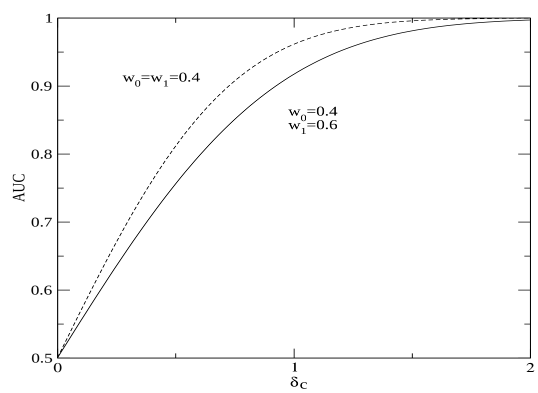
AUC
δ_c
$w_0=w_1=0.4$
$w_0=0.4$
$w_1=0.6$
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8



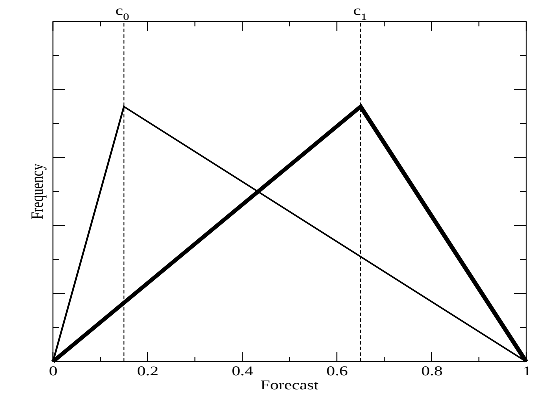
Frequency
Forecast
L_0
L_1
0.25
0.2
0.15
0.1
0.05
0
-5
0
5
10



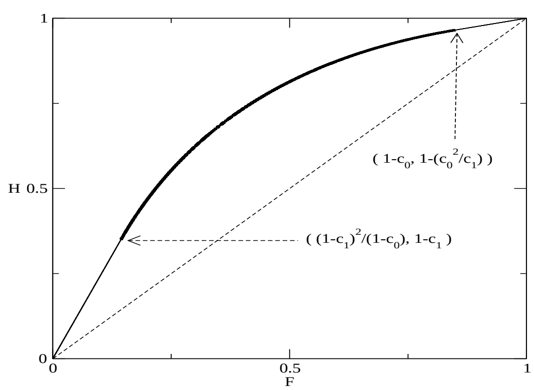
H
F
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1



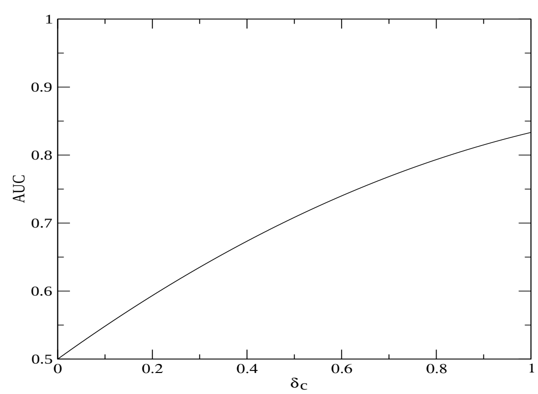
AUC
δ_c
$w_0=w_1=0.4$
$w_0=0.4$
$w_1=0.6$
1
0.9
0.8
0.7
0.6
0.5
0
1
2



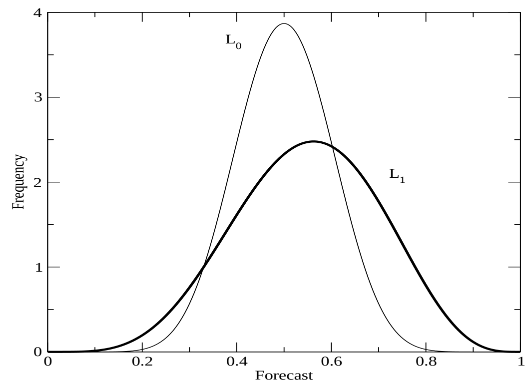
c0
c1
Frequency
0
0.2
0.4
0.6
0.8
1
Forecast



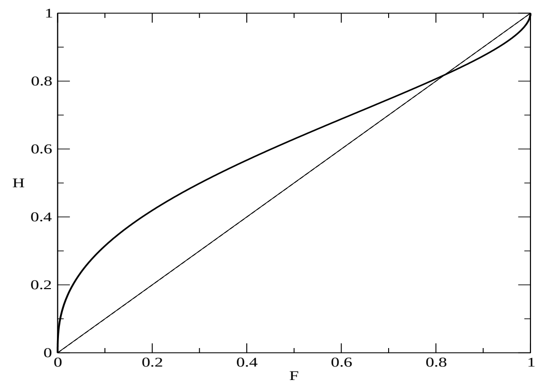
1
H 0.5
0
0
0.5
1
F
(1-c0, 1-(c0^2/c1))
((1-c1)^2/(1-c0), 1-c1)



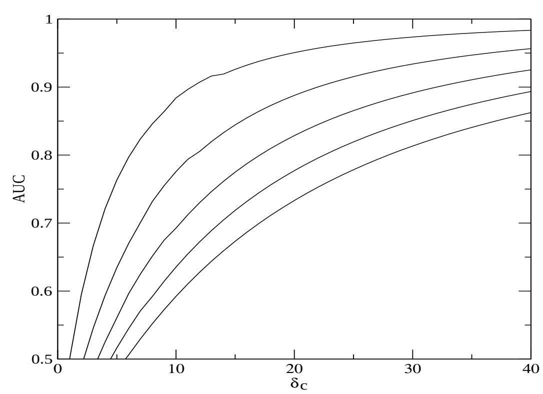
1
0.9
0.8
0.7
0.6
0.5
AUC
0
0.2
0.4
0.6
0.8
1
δc



L0
L1
Frequency
Forecast



H
F



AUC
δ_c

Conclusions

For unbounded forecasts

Asymmetric ROC: unequal widths.
Overlap with axes: difference in widths.
Crossing diagonal: $|\frac{\delta c}{\delta w}| < 2$ (gaussian)

For bounded forecasts

Symmetry: $\frac{c_1(1-c_1)}{w_1^2} = \frac{c_0(1-c_0)}{w_0^2}$
No overlap with axes.

For both, AUC flattens off.