Everything you've ever wanted to know about Receiver Operating Characteristic Curves but were afraid to ask

Jim Muirhead

Sept. 29, 2008

Outline

- Historical context and uses of Receiver Operating Characteristic curves (ROC)
- Empirical case study: step-by-step evaluation of ROC characteristics
- Analytical and numerical evaluation of ROC for uniform and normal distribution of forecast probabilities

Historical use of Receiver Operating Characteristic Curves

- Originally developed for radar-signal detection methodology (signal-to-noise), hence “Radar Receiver Operator Characteristic”)
- Used extensively in medical and psychological test evaluation
- More recently in atmospheric science
- Draws on the “power” of statistical tests

Primary uses

- Used to compare probabilistic forecasts to *events* or *non-events*
- Assess the probability of being able to distinguish a *hit* from a *miss*
- Classify forecast probabilities into binary categories (0,1) based on probabilistic thresholds
- Compare detection ability of different experimental methods

Definitions of hit rate, false alarm rate

		Observed	
		Non Event (0)	Event (1)
Predicted	Non Event (0)	a) Correct negative	b) Miss
	Event (1)	c) False Alarm	d) Hit

Hit rate (*H*): d/(b+d)
False alarm rate (*F*): c/(a+c)

Empirical case study

- Example from Mason and Graham (2002) *Q. J. Meterol. Soc* **128**: 2145-2166
- Data describes March-May precipitation over North-East Brazil for 1981-1995
- Arranged in decreasing probability

- n = total number of cases
- e = number of events (1)
- e' = n-e = number of non-events (0)
- FP = Forecast Probabilities

n=15, e=7,e'=8

Year	Observed event (1) or non-event (0)	Forecast Probability (FP)
1994	1	0.984
1995	1	0.952
1984	1	0.944
1981	0	0.928
1985	1	0.832
1986	1	0.816
1988	1	0.584
1982	0	0.576
1991	0	0.28
1987	0	0.136
1989	1	0.032
1992	0	0.024
1990	0	0.016
1983	0	0.008
1993	0	0

Classified predictions at different thresholds

Vary Threshold (t) from 0 - 1

Hit

False alarm

Miss

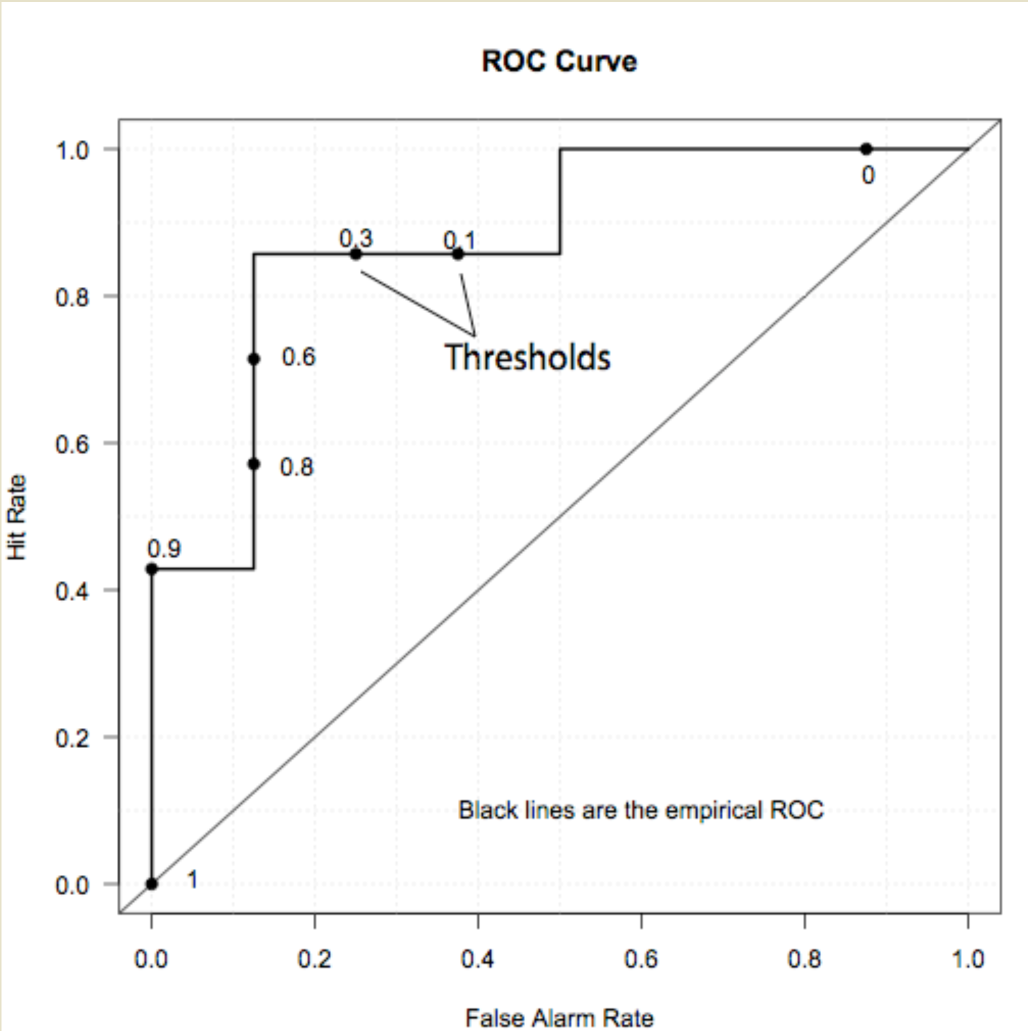
Correct negative

Year	Observed	Forecast Probability	Prediction t=0.1	t=0.5	t=0.8
1994	1	0.984	1	1	1
1995	1	0.952	1	1	1
1984	**1**	0.944	**1**	1	1
1981	**0**	0.928	**1**	1	1
1985	1	0.832	1	1	1
1986	1	0.816	1	1	1
1988	1	0.584	1	1	0
1982	0	0.576	1	1	0
1991	0	0.28	1	0	0
1987	0	0.136	1	0	0
1989	**1**	0.032	**0**	0	0
1992	0	0.024	0	0	0
1990	**0**	0.016	**0**	0	0
1983	0	0.008	0	0	0
1993	0	0	0	0	0

ROC curve developed over range of thresholds

- Hit rates and false alarm rates vary with changing thresholds

- Curve will be stepped there are no ties in forecast probabilities and each forecast is considered in turn



Relationship between thresholds, hit and false alarm rates

Threshold is low (t=0.2)

		Observed	
		0	**1**
Predicted	**0**	3	1
	1	5	6
Total		8	7

Hit rate (H)	0.857
False alarm rate (F)	0.625
Overall	0.6

Threshold is high (t=0.8)

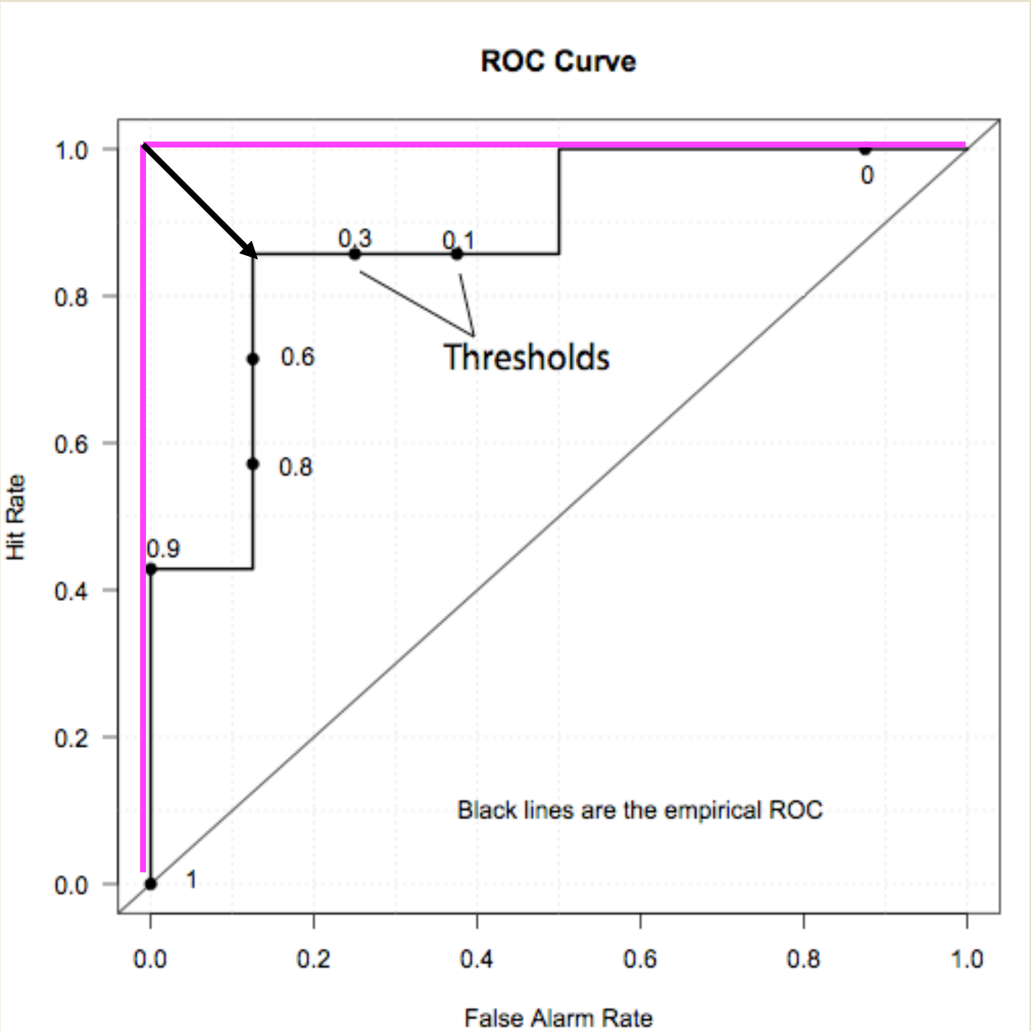
		Observed	
		0	**1**
Predicted	**0**	7	2
	1	1	5
Total		8	7

Hit rate (H)	0.714
False alarm rate (F)	0.125
Overall	0.8

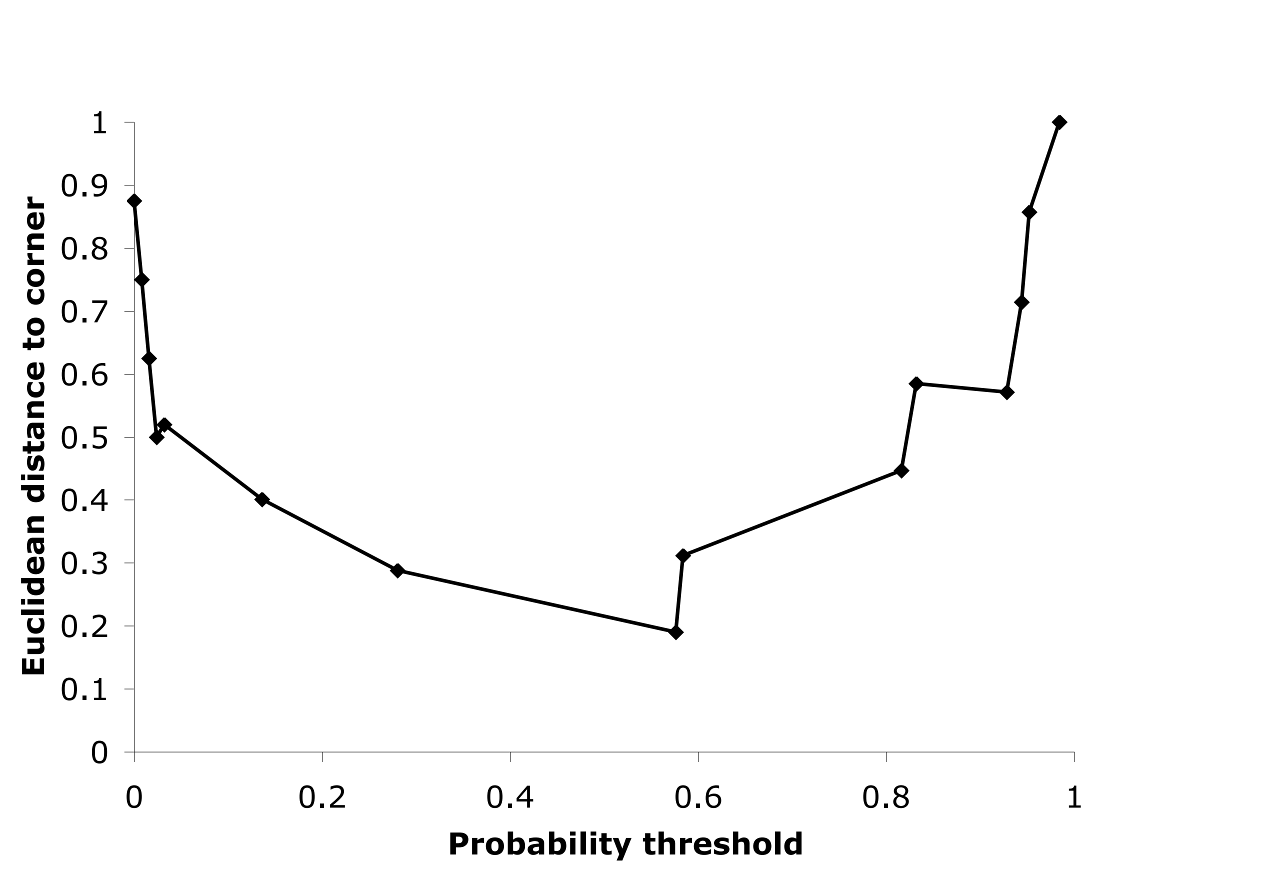
Optimum choice of threshold

- Perfect model: 100% Hit Rate, 0% False Alarm Rate

- Optimal threshold on curve chosen by Euclidean distance away from perfect model



Optimal threshold and hit/false alarm rates



Optimal threshold (t) = 0.576, corresponds to hit rate = 0.857 and false alarm rate of 0.25

Calculation of Area under the Curve (AUC)

- Empirical curve
  - Area under the curve is gained when a hit has higher associated forecast probability than any false alarms
- No area is gained when a false alarm occurs

Calculation of Area under the curve

For each hit, f_i is the number of misses with FP greater than the current hit

e = number of events (1)

e' = n-e = number of non-events (0)

FP = Forecast Probabilities

$$\text{area gained} = \frac{(e' - f)}{e'e}$$

Total ROC area

$$A = \frac{1}{e'e}\sum_{i=1}^{e}(e' - f_i)$$

A=0.875

Year	Observed	Probability	f	Area gained
1994	1	0.984	0	0.142857143
1995	1	0.952	0	0.142857143
1984	1	0.944	0	0.142857143
1981	**0**	**0.928**		
1985	1	0.832	1	0.125
1986	1	0.816	1	0.125
1988	1	0.584	1	0.125
1982	0	0.576		
1991	0	0.28		
1987	0	0.136		
1989	1	0.032	4	0.071428571
1992	0	0.024		
1990	0	0.016		
1983	0	0.008		
1993	0	0		

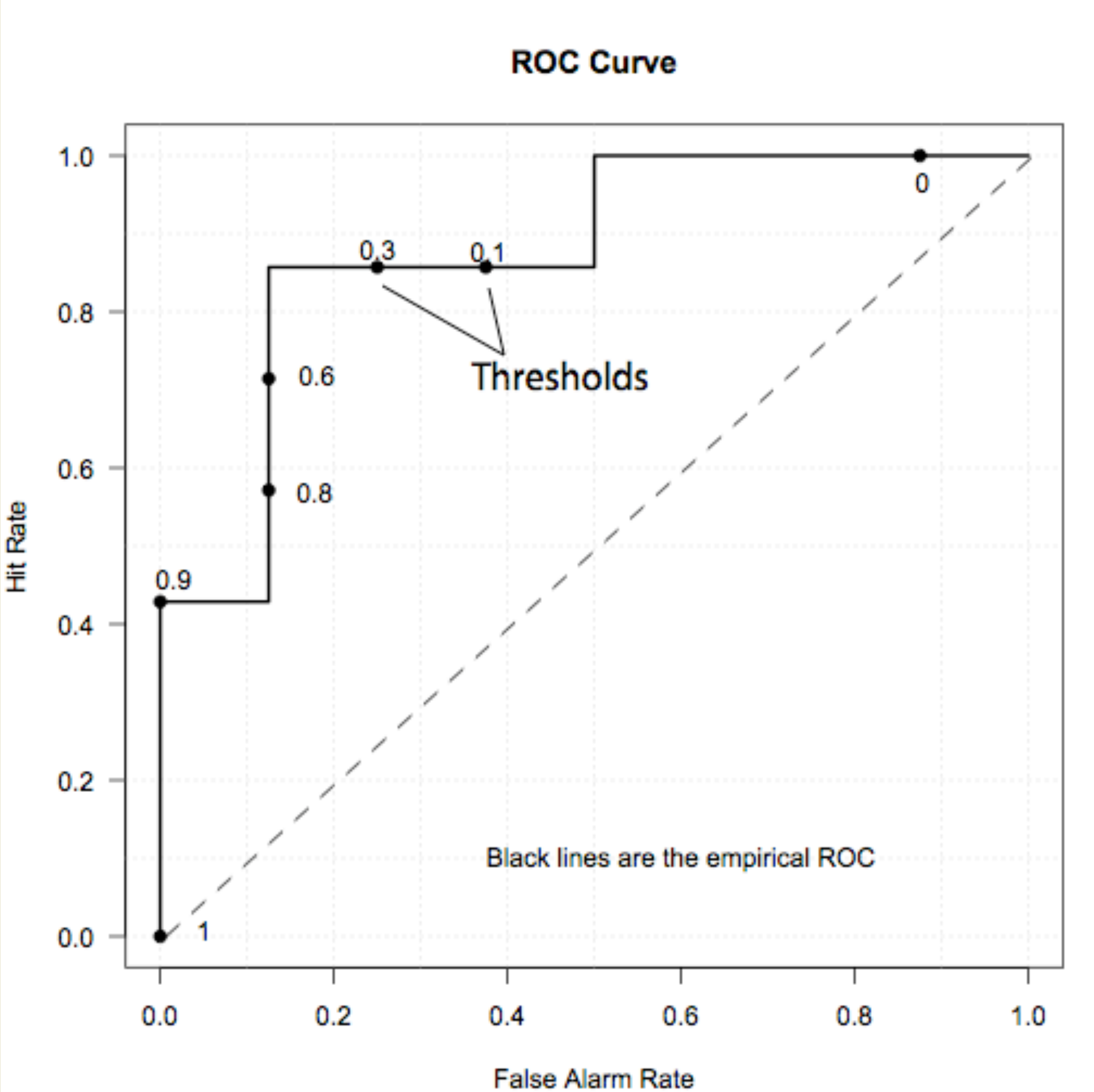
Total 0.875

Hypothesis testing of AUC

- The AUC is the probability of being able to distinguish a *hit (e)* from a *miss (e') (AUC*=0.875*)*

- Dashed line indicates forecasting skill is no better than random (0.5)

- Is AUC significantly greater than 0.5?



Significance testing for AUC

- Mann-Whitney U test

$$U = \sum_{i=1}^{e} r_{ei} - \frac{e(e+1)}{2}$$

U = (15+14+13+11+10+9+5)-(7*8)/2 = 49

p = 0.007 in our example

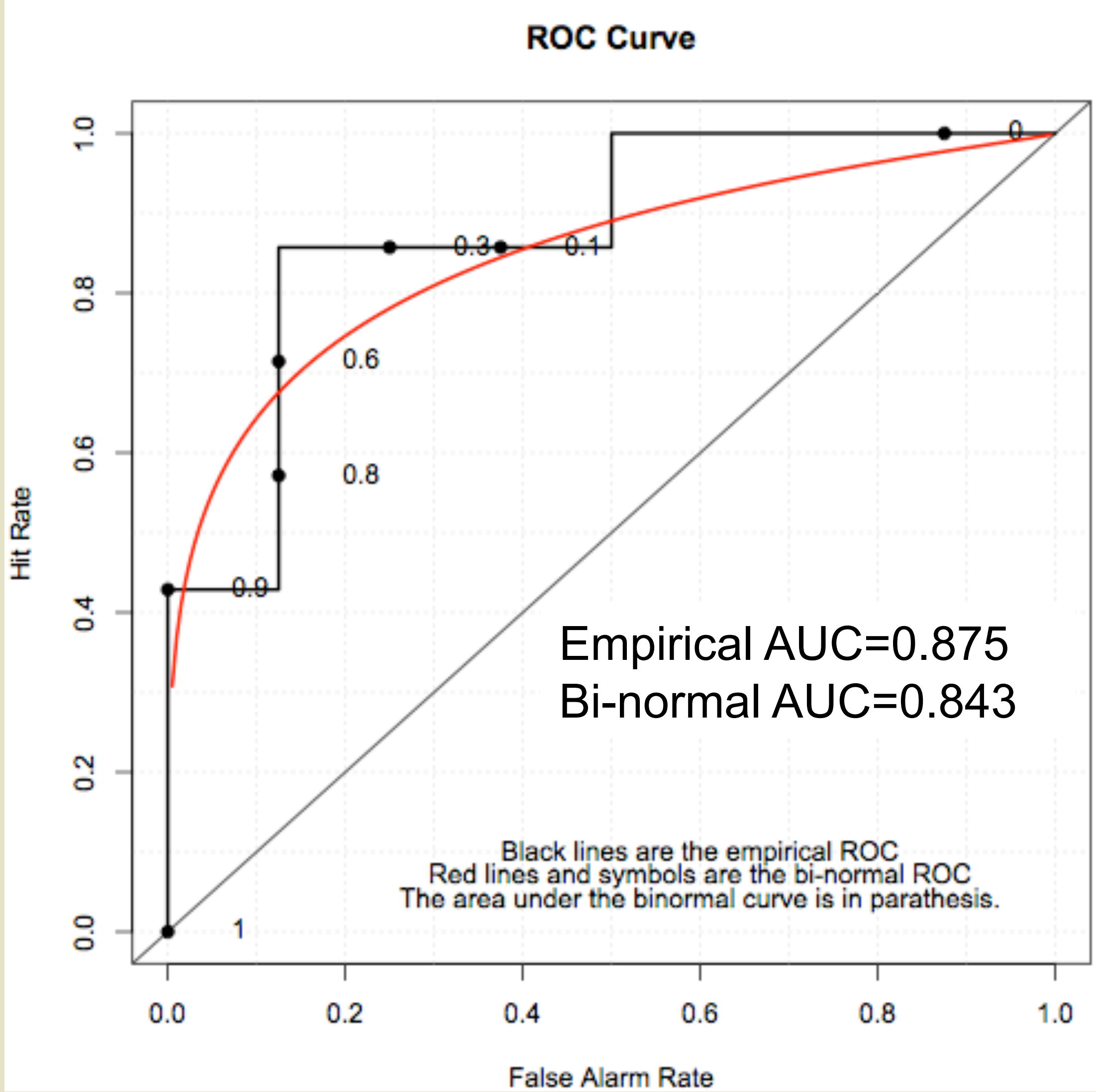
The relationship between *U* and *AUC*

$$U = e'e\left(1 - A\right)$$

Year	Observed	Probability	Rank
1994	1	0.984	15
1995	1	0.952	14
1984	1	0.944	13
1985	1	0.832	11
1986	1	0.816	10
1988	1	0.584	9
1989	1	0.032	5
1981	0	0.928	12
1982	0	0.576	8
1991	0	0.28	8
1987	0	0.136	6
1992	0	0.024	4
1990	0	0.016	3
1983	0	0.008	2
1993	0	0	1

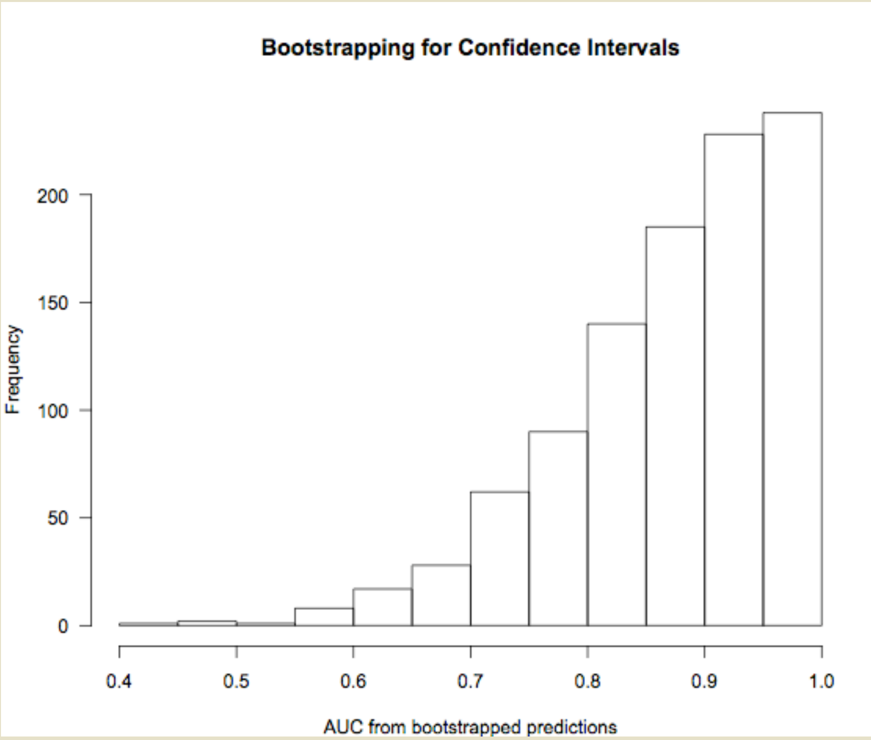
Normal transformation of Hit and False Alarm rates

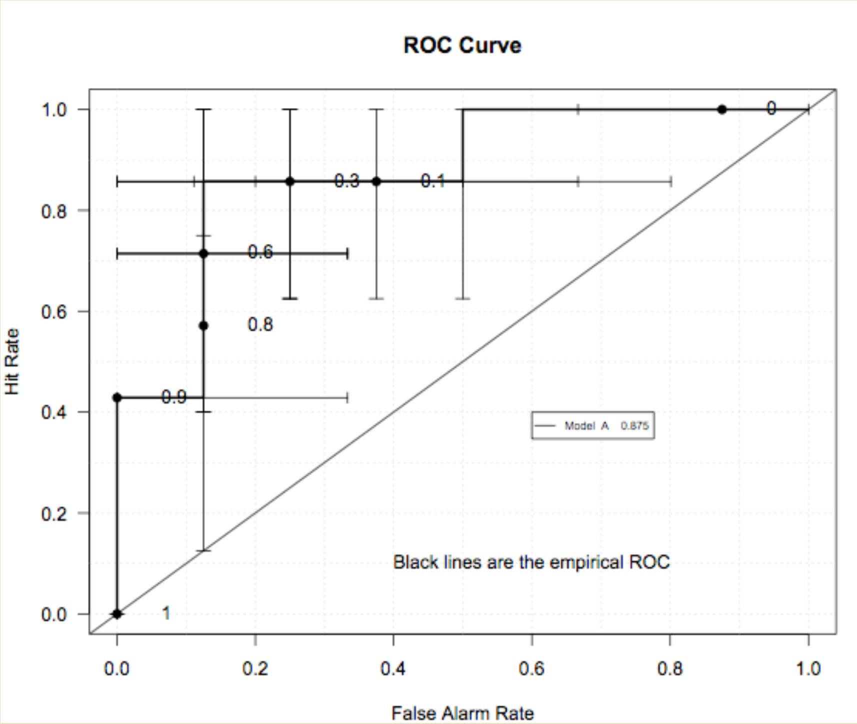
- Hit and False alarm rates transformed to bi-normal distribution useful for comparing differences in AUC for competing models.
- AUC under bi-normal ROC is not as sensitive to the number of points as the empirical ROC
- Important to distinguish transforming axes (*H* and *F*) from transforming forecasting probabilities.



Confidence Intervals for AUC, Hit and False Alarm rates

- Significance can also be tested with permuting or bootstrapping data





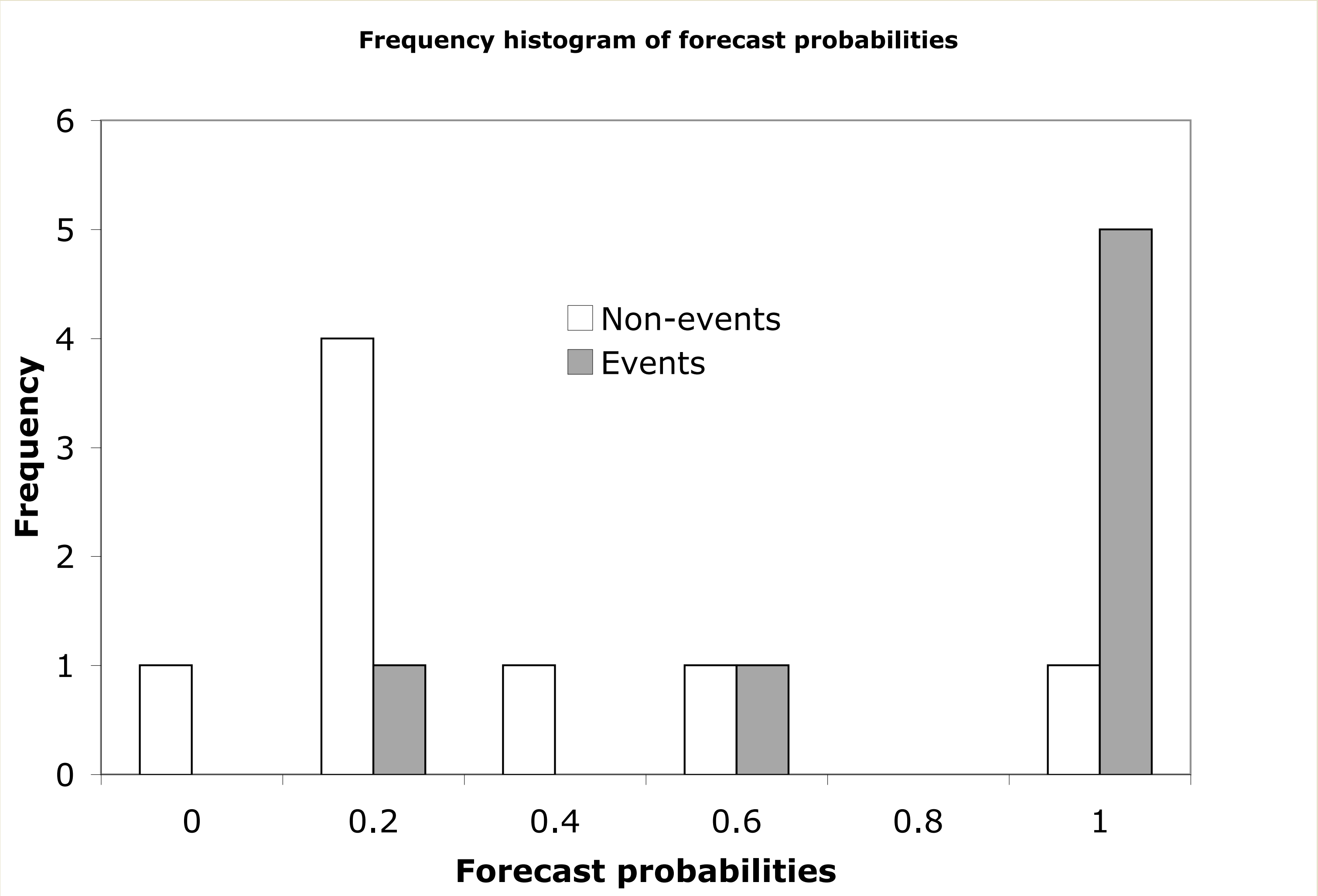
95% CI for AUC=0.643 - 1.00
Note: Does not include 0.5

95% CI for Hit and False alarm rates

Effects of assuming parametric distributions of forecast probabilities

- Previous example was empirically derived ROC
- What are the effects of assuming a uniform and normal distribution of forecast probabilities?

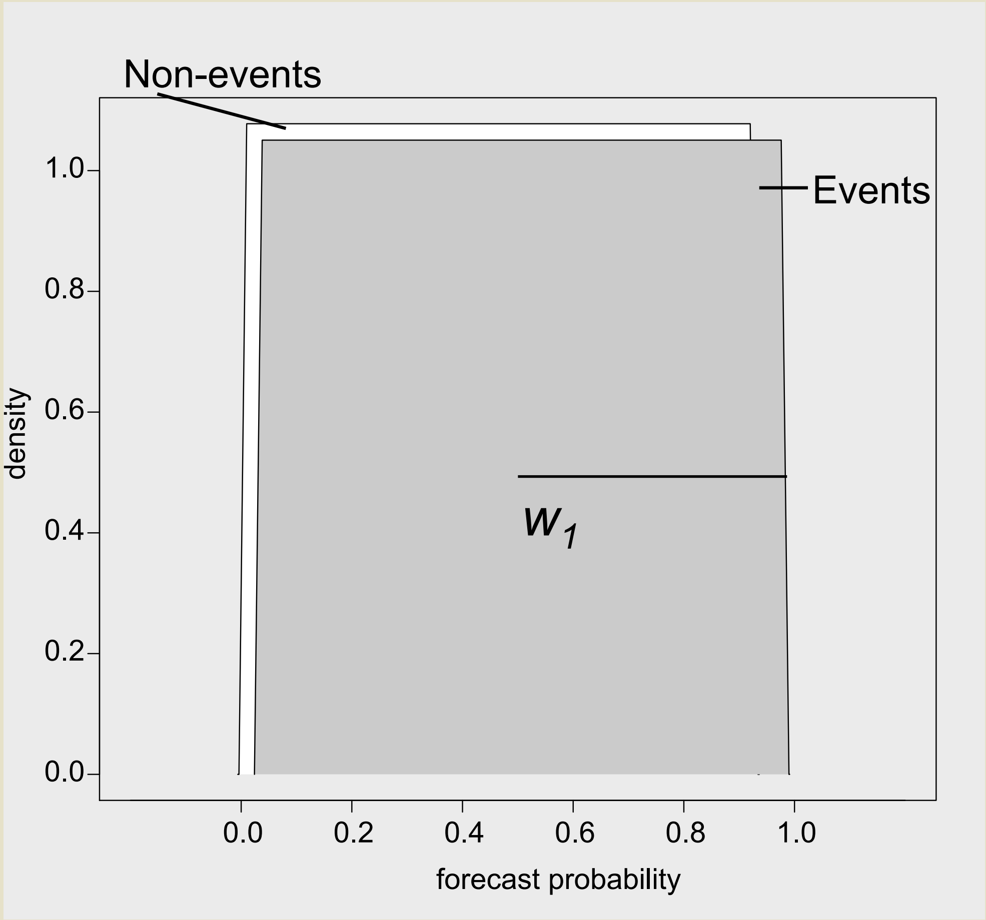
Forecast probabilities for rain events from Mason and Graham 2002



Uniform distribution

- 4 parameters needed, means c_0 and c_1 and half-widths w_0 and w_1 for distribution of negative and positive forecasts, respectively

- For uniform distribution, w_1 is simply the half range of probabilities associated with positive forecasts



Data parameterized from Mason and Graham 2002

Uniform distribution

From Marzban (2004)

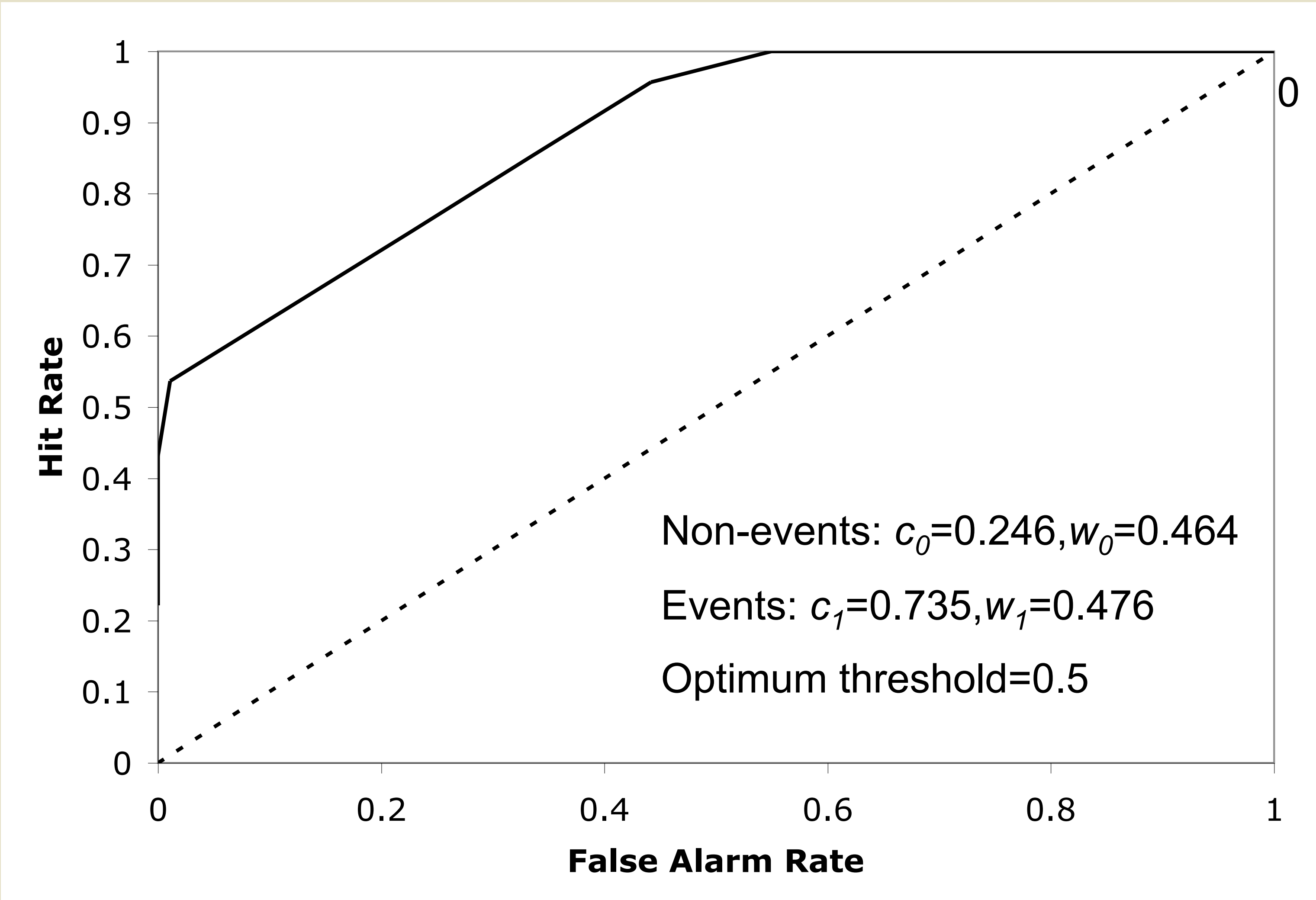
- Hit and False Alarm rates calculated as:

$$H = \frac{c_1 + w_1 - t}{2w_1}, \quad F = \frac{c_0 + w_0 - t}{2w_0}$$, where *t* is the threshold

The Area under the curve is calculated as:

$$AUC = 1 - \frac{1}{8}\left(\frac{(c_1 - c_0) - (w_1 + w_0)}{\sqrt{w_0 w_1}}\right)^2$$

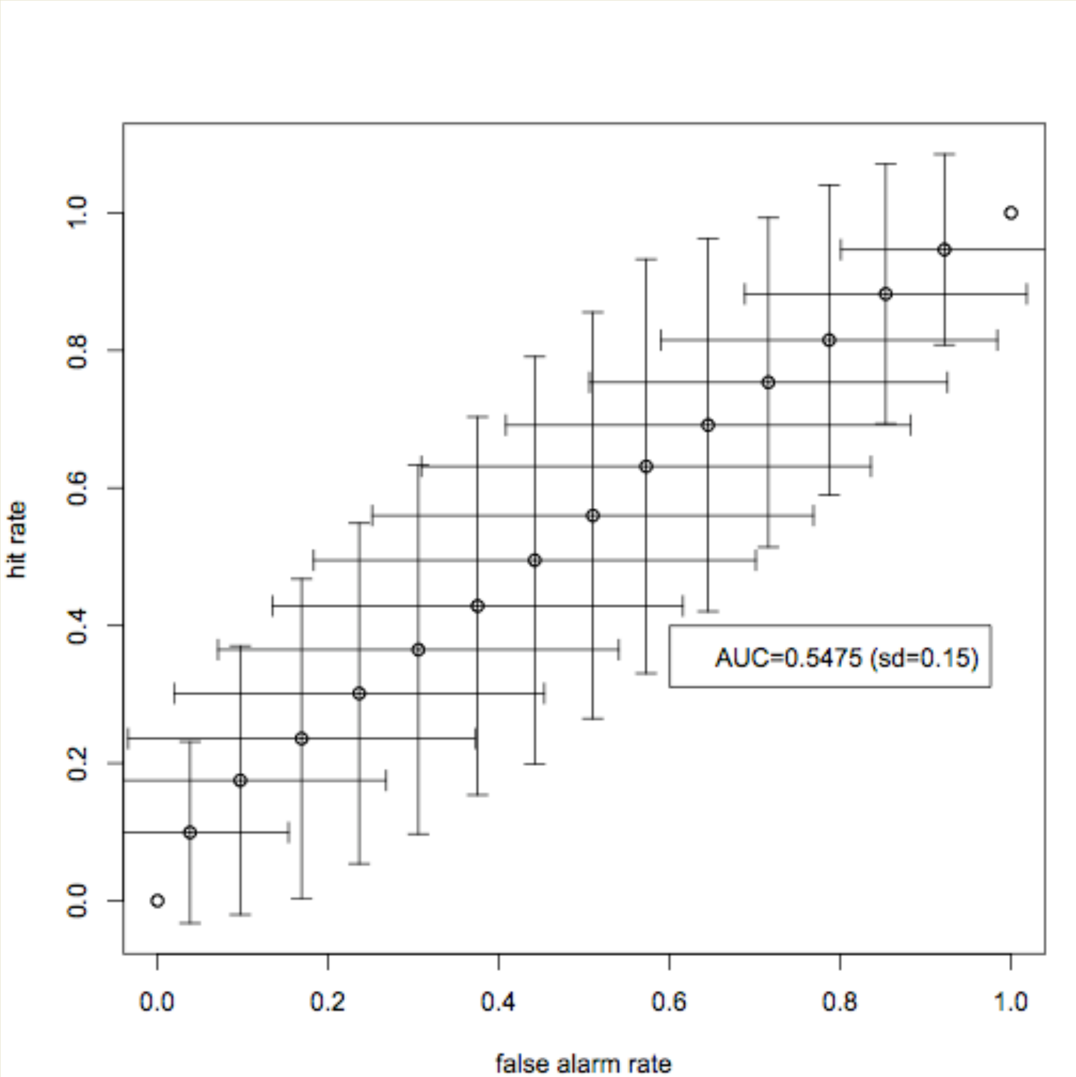
ROC of uniformly distributed forecast probabilities



Numerical simulation of uniformly distributed forecast probabilities

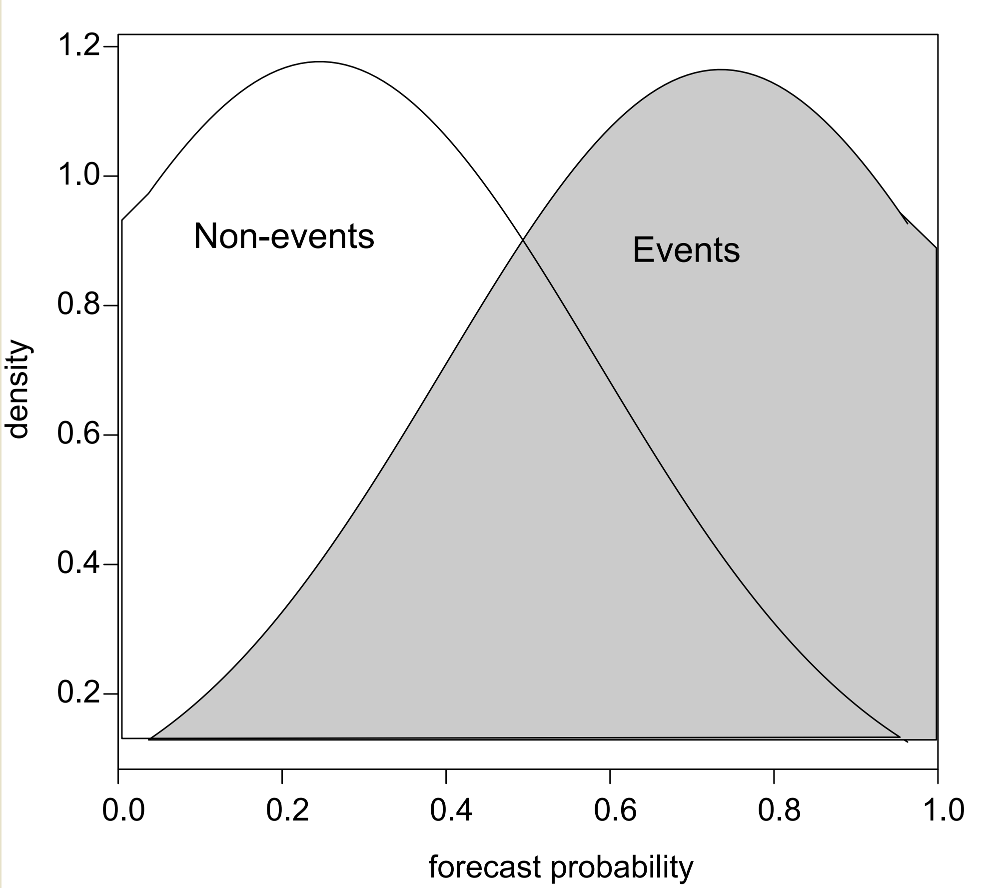
- Generated uniform deviates with min. and max. from Mason and Graham 2002 data.

- n = 200 iterations



Normal distribution of forecast probabilities

- For the normal distribution, c_0 and c_1 are means for non-events and events, and w_0 and w_1 are standard



Non-events: $c_0=\overline{x}_0=0.246$,

$w_0=\sigma_0=0.339$

Events: $c_1= \overline{x}_1=0.735$,

$w_1= \sigma_1=0.338$

Normal distribution of forecast probabilities

• False alarm rates *(F)* and hit rates *(H)* calculated as:

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

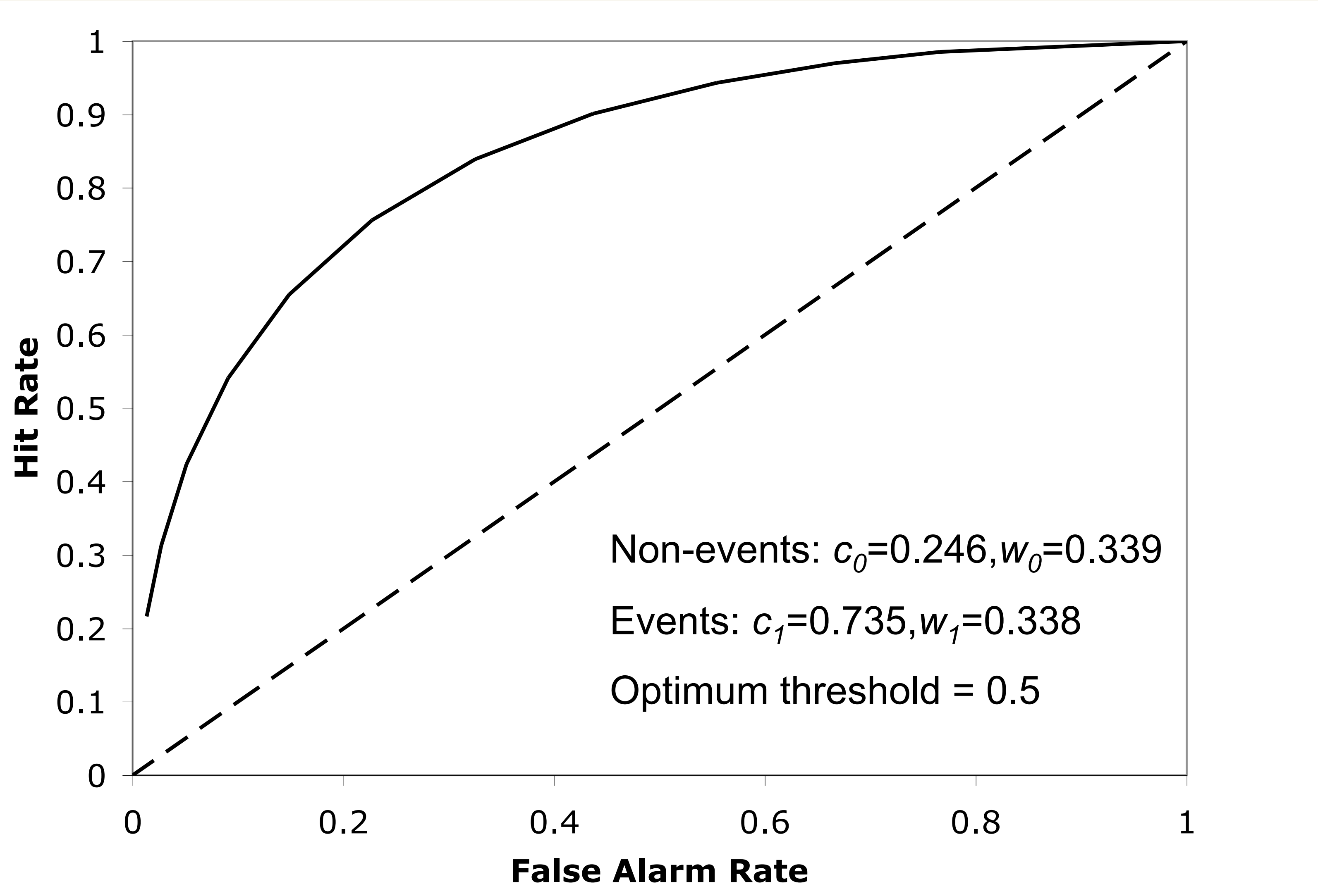
(Marzban 2004)

where $\Phi(x)$ is the standard normal cumulative distribution

Area is calculated as:

$$AUC = \Phi\left(\frac{c_1 - c_0}{\sqrt{w_0^2 + w_1^2}}\right)$$

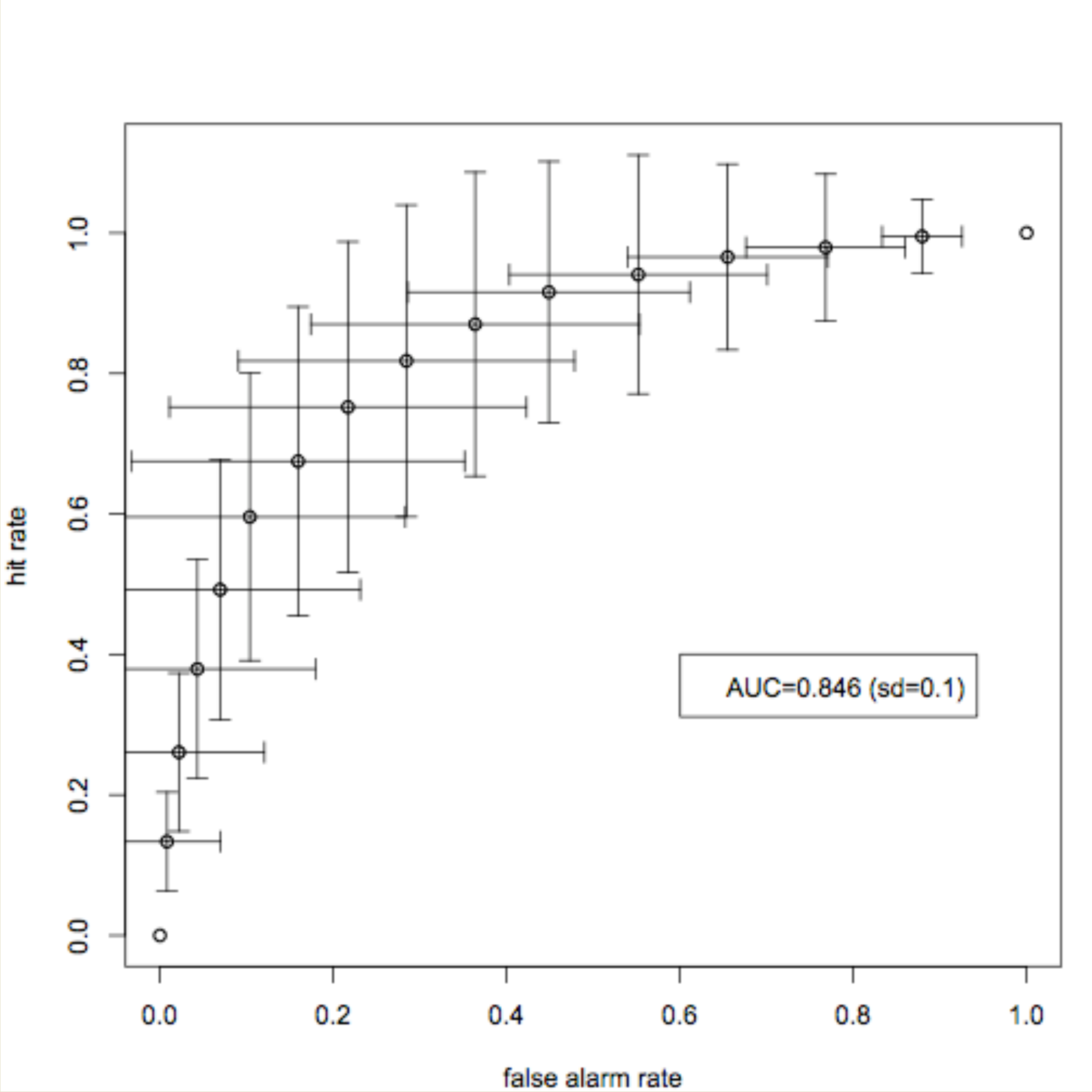
ROC of normally distributed forecast probabilities



Numerical simulation of normally distributed forecast probabilities

- Generated gaussian deviates with mean and sd from Mason and Graham 2002 data.

- n = 200 iterations



Summary

- Empirical ROC may result in overestimated AUC relative to bi-normal distribution of hit and false alarm rates
- Similar results in AUC for normalizing either hit/false alarm rates (0.843), analytical solution of normally distributed forecast probabilities (0.846) or numerical simulations (avg.=0.846)
- AUC from numerical simulations for uniform forecast probabilities not significant (avg.=0.547) unlike analytical approach (0.88).

Summary

- Recommendation:

1) Examine distribution of forecast probabilities from data
2) Do not assume uniform distribution if using the analytical approach, especially for low sample sizes