Visualization of the significance of Receiver Operating Characteristics based on confidence ellipses

Sarlis, N.V. and Christopoulos, S-R.

Original citation:
Sarlis, N.V. and Christopoulos, S-R. (2013) Visualization of the significance of Receiver Operating Characteristics based on confidence ellipses. Computer Physics Communications, volume 185 (3): 1172–1176. DOI: 10.1016/j.cpc.2013.12.009

http://dx.doi.org/10.1016/j.cpc.2013.12.009

Elsevier

Visualization of the significance of Receiver Operating Characteristics based on confidence ellipses

Nicholas V. Sarlis[a,*], Stavros-Richard G. Christopoulos[a]

[a] *Solid State Section and Solid Earth Physics Institute, Physics Department, University of Athens, Panepistimiopolis, Zografos 157 84, Athens, Greece*

Abstract

The Receiver Operating Characteristics (ROC) is used for the evaluation of prediction methods in various disciplines like meteorology, geophysics, complex system physics, medicine etc. The estimation of the significance of a binary prediction method, however, remains a cumbersome task and is usually done by repeating the calculations by Monte Carlo. The FORTRAN code provided here simplifies this problem by evaluating the significance of binary predictions for a family of ellipses which are based on confidence ellipses and cover the whole ROC space.

Keywords: Receiver Operating Characteristics (ROC), complex systems, systems obeying power laws, significance level, p-value

---

PROGRAM SUMMARY

Manuscript Title: Visualization of the significance of Receiver Operating Characteristics based on confidence ellipses
Authors: N.V. Sarlis and S.-R. G. Christopoulos
Program Title: VISROC.f
Journal Reference:
Catalogue identifier:
Licensing provisions: none
Programming language: FORTRAN
Computer: Any computer supporting a GNU FORTRAN compiler
Operating system: Linux, MacOS, Windows
RAM: 1Mbyte

---

*Corresponding author.
E-mail address: nsarlis@phys.uoa.gr

Number of processors used:
Supplementary material:
Keywords: Receiver Operating Characteristics (ROC), significance level, p-value
Classification: 4.13 Statistical Methods, 9 Databases, Data Compilation and Information Retrieval, 14 Graphics.
Nature of problem: The Receiver Operating Characteristics (ROC) is used for the evaluation of prediction methods in various disciplines like meteorology, geophysics, complex system physics, medicine etc. The estimation of the significance of a binary prediction method, however, remains a cumbersome task and is usually done by repeating the calculations by Monte Carlo. The FORTRAN code provided here simplifies this problem by evaluating the significance of binary predictions for a family of ellipses which are based on confidence ellipses and cover the whole ROC space.

Solution method: Using the statistics of random binary predictions for a given value of the predictor threshold ϵ_t, one can construct the corresponding confidence ellipses. The envelope of these corresponding confidence ellipses is estimated when ϵ_t varies from 0 to 1. This way a new family of ellipses is obtained, named k-ellipses, which covers the whole ROC plane and leads to a well defined Area Under the Curve (AUC). For the latter quantity, Mason and Graham [1] has shown that it follows the Mann-Whitney U-statistics[2] which can be applied[3] for the estimation of the statistical significance of each k-ellipse. As the transformation is invertible, any point on the ROC plane corresponds to a unique value of k, thus to a unique p-value to obtain this point by chance. The present FORTRAN code provides this p-value field on the ROC plane as well as the k-ellipses corresponding to the (p=)10%, 5% and 1% significance levels using as input the number of the positive (P) and negative (Q) cases to be predicted.

Restrictions:

Unusual features: In some machines, the compiler directive `-O2` *or* `-O3` *should be used to avoid* `NaN`*'s in some points of the p-field along the diagonal*

Additional comments:

Running time: Depending on the application, e.g., 4s for an `Intel(R) Core(TM)2 CPU E7600 at 3.06GHz` *with* `2GB RAM` for the examples presented here

[1] S. J. Mason, N. E. Graham, Quart. J. R. Meteor. Soc. 128 (2002) 2145.

[2] H. B. Mann, D. R. Whitney, Ann. Math. Statist. 18 (1947) 50. Program Summary section.

[3] L. C. Dinneen, B. C. Blakesley, J. R. Stat. Soc. Ser. C Appl. Stat. 22 (1973) 269. the end of

1. Introduction

Receiver Operating Characteristics (ROC) graphs is a technique currently used[1, 2, 3, 4, 5, 6, 7], for estimating the predictability of various complex systems and has already found useful applications in various fields like medicine, e.g. see [8], meteorology[9, 10], etc. As suggested by Fawcett [11] ROC graphs depict the trade off between hit rates and false alarm rates with a conceptually simple way which also applies to the case of skewed class distributions which is usually the case in the physics of complex systems.

We limit ourselves in the case of binary predictions. In this case, there are two classes of events **p** or **n**; for example in the case of extreme events in complex systems an event is classified as **p** if its magnitude exceeds a given threshold M_t otherwise it is considered **n**. Before the occurrence of each event a predictor ϵ is assigned based on some prediction algorithm and a hypothesized class **P** or **N** for the forthcoming event is decided on the basis of whether ϵ *exceeds* or not, respectively, the predictor threshold value ϵ_t. If the hypothesized class is **P** and the event is **p** we have a successful prediction called True Positive (TP). Similarly, if the hypothesized class is **N** and the event is **n** we again have a successful prediction called True Negative (TN). If, however, the hypothesized class is **P** and the event is **n** we have a False Positive (FP) unsuccessful prediction. Finally, if the hypothesized class is **N** and the event is **p** we have a False Negative (FN) unsuccessful prediction. Assuming that in total we examine P events of the **p** class and Q events of the **n** class, one defines[11] the True Positive rate (TPr) -or hit rate (H)- as the ratio of the totality of TP's over P,

$$H \equiv \frac{|TP|}{P} = \frac{|TP|}{|TP| + |FN|}, \tag{1}$$

and the False Positive rate (FPr) -or false alarm rate (F)- as the ratio of the totality of FP's over Q,

$$F \equiv \frac{|FP|}{Q} = \frac{|FP|}{|FP| + |TN|}. \tag{2}$$

The ROC curve is obtained as we vary ϵ_t and plot H as a function of F, e.g. see Fig.1.

In the example of Fig. 1, we depict the ROC curves for three different values of M_t which result[6] when considering as predictor the number ϵ_k of successive extrema of the aftershock magnitude time series m_{k+1} in the case of the 1999 Hector Mine earthquake. As we vary the threshold ϵ_t an ROC curve is obtained, e.g. see the results depicted by the red squares which correspond to a target magnitude M_t=5.0. In this particular example, the hypothesized class is either **P** or **N** depending on whether ϵ_k is *smaller* or larger, respectively, than ϵ_t. We observe that the diagonal (black solid line) corresponds to random predictions and would have been obtained as an ROC curve for a random predictor if both P and Q had tended to infinity. If we repeat the experiment using the definition of the hypothesized classes of the previous paragraph, i.e., interchanging **P** with **N**, we obtain the ROC curves that lie in the lower triangle which are also depicted with thick lines without symbols in Fig.1 which are symmetric images of the originals with respect to the center $(F_c, H_c) = (1/2, 1/2)$. Thus, the statistical significance of an ROC curve depends on its deviation from the $H = F$ diagonal and by negating the condition used for the construction of the hypothesized classes we can 'reflect' an ROC curve with respect to the center of the ROC space. Moreover, in the present example, as we change M_t we drastically affect the number P of **p** events, see Table 1, because earthquake magnitudes follow the Gutenberg-Richter law[12]. The corresponding ROC curves may become more distant from the diagonal, but this may be misleading as their statistical significance varies when P (and Q) change. It is the aim of the present paper to present a plausible visualization method for the statistical significance of an ROC curve together with a FORTRAN code that generates the corresponding significance intervals. In Section 2, we will present the proposed k-ellipses that cover the ROC space and obtain the related statistics. Section 3 discusses the implementation of the method in FORTRAN and Section 4 summarizes the results.

2. The *k*-ellipses family

In order the estimate the statistical significance of an ROC curve, we need to compare it with similar results obtained when using a random predictor. Without loss of generality, we assume that the predictor threshold ϵ_t varies in the range [0,1]. Then, as a random predictor we can consider a uniformly distributed random number u_i in the same interval. Under these assumptions, the conditional probabilities $P(\mathbf{P}|\mathbf{p})$ and $P(\mathbf{P}|\mathbf{n})$ to obtain the hypothesized class $\mathbf{P}$ under the assumption that the event is either $\mathbf{p}$ or $\mathbf{n}$ are both equal to ϵ_t. Thus, the number l of TP's as well as the number m of FP's follow the binomial distribution with attempt probability ϵ_t for P and Q attempts, respectively. The mean value of the hit rate H and the false alarm rate F result in

$$\langle H \rangle = \frac{\langle l \rangle}{P} = \epsilon_t, \tag{3}$$

$$\langle F \rangle = \frac{\langle m \rangle}{Q} = \epsilon_t, \tag{4}$$

with variances

$$\sigma_H^2 = \langle (H - \langle H \rangle)^2 \rangle = \frac{\left(\langle l^2 \rangle - \langle l \rangle^2\right)}{P^2} = \frac{\epsilon_t(1-\epsilon_t)}{P}, \tag{5}$$

$$\sigma_F^2 = \langle (F - \langle F \rangle)^2 \rangle = \frac{\left(\langle m^2 \rangle - \langle m \rangle^2\right)}{Q^2} = \frac{\epsilon_t(1-\epsilon_t)}{Q}, \tag{6}$$

respectively. When P and Q are sufficiently large[13] the binomial distributions of l and m can be approximated by Gaussian ones leading to a two dimensional Gaussian distribution for H and F

$$f(F,H) = \frac{\sqrt{PQ}}{2\pi\epsilon_t(1-\epsilon_t)} \exp\left[-\frac{Q(F-\epsilon_t)^2 + P(H-\epsilon_t)^2}{2\epsilon_t(1-\epsilon_t)}\right]. \tag{7}$$

The confidence regions in this case[14] are the confidence ellipses

$$Q(F-\epsilon_t)^2 + P(H-\epsilon_t)^2 = k\epsilon_t(1-\epsilon_t), \tag{8}$$

with center $(F,H) = (\epsilon_t, \epsilon_t)$ -on the diagonal of the ROC space- and k is a positive parameter signifying the confidence level with $p_0 = \exp(-k/2)$. *If* we vary ϵ_t within the region [0,1], the aforementioned center moves along the diagonal and we obtain different ellipses for a given p_0-value. The envelope of these ellipses can be determined by the maximization (minimization) of H under the condition (8) for given values of F and k. Such an analysis results in the following expressions for $H_{\max}$ and $H_{\min}$

$$H_{\max} = \frac{1}{2} + \left(\frac{Q}{Q+k}\right)\left(F - \frac{1}{2}\right) + \frac{\sqrt{k(Q+k+P)[k+4Q(F-F^2)]}}{2(Q+k)\sqrt{P}}, \tag{9}$$

$$H_{\min} = \frac{1}{2} + \left(\frac{Q}{Q+k}\right)\left(F - \frac{1}{2}\right) - \frac{\sqrt{k(Q+k+P)[k+4Q(F-F^2)]}}{2(Q+k)\sqrt{P}}, \quad (10)$$

respectively. By substituting $\mathcal{Y} = H - \frac{1}{2}$ and $\mathcal{X} = F - \frac{1}{2}$, Eqs.(9) and (10) turn to

$$[\underbrace{4Q(k+P)}_{A_0}]\mathcal{X}^2 + [\underbrace{-8PQ}_{B_0}]\mathcal{X}\mathcal{Y} + [\underbrace{4P(k+Q)}_{C_0}]\mathcal{Y}^2 + [\underbrace{-k(k+P+Q)}_{F_0}] = 0 \quad (11)$$

which describes the envelope. Equation (11) is[15] an ellipse since $B_0^2 - 4A_0C_0 = -64PQk(k+Q+P) < 0$ with center at $(F,H) = (1/2, 1/2)$ and inclination θ with respect to the x-axis given by

$$\cot(2\theta) = \frac{A_0 - C_0}{B_0} = \frac{k}{2}\left(\frac{1}{Q} - \frac{1}{P}\right). \quad (12)$$

Let us call the family of the ellipses described either by Eq.(11) or by Eqs.(9) and (10) as the k-ellipses family. The k-ellipses family covers the whole ROC space and the k value corresponding to an arbitrary ROC point (F,H) is given by the positive root of Eq.(11) which equals

$$\begin{aligned} k(F,H) &= 2P(H^2 - H) + 2Q(F^2 - F) \\ &+ 2\sqrt{[P(H^2 - H) + Q(F^2 - F)]^2 + PQ(F-H)^2}. \end{aligned} \quad (13)$$

We observe that for the ideal predictor, $(F,H) = (0,1)$, we have $k(0,1) = 2\sqrt{PQ}$. Morover, Eq.(13) exhibits the aforementioned (see Section 1) symmetry of the ROC graphs: A reflection of the ROC curve with respect to the ROC center $(1/2, 1/2)$ leading to the transformation $F' = 1 - F$ and $H' = 1 - H$ -which corresponds to negating the condition for the attribution of the hypothesized classes- results in the same value of k, i.e., $k(F', H') = k(F,H)$. Thus, the k-ellipses family may be used for the estimation of the statistical significance of ROC curves.

Although each k-ellipse has been obtained for a given value of p_0 [$= \exp(-k/2)$] for every value of ϵ_t, the probability to obtain by chance an ROC curve enclosed by this envelope may significantly differ from that p_0 value as it is highly improbable for a given random realization of the prediction method to keep a constant value of k for all ϵ_t. In order to estimate the statistical significance of each k-ellipse, we have to resort to the statistics of the Area Under the Curve (AUC), labelled by A, in the ROC plane which has been the subject of the study by Mason and Graham[10]. Mason and Graham[10] have shown that

$$A = 1 - \frac{U}{PQ}, \tag{14}$$

where U follows the Mann-Whitney U-statistics[16], i.e., it equals to the sum of the number of cases a number u_k, $k = 1, 2, ..., P$ is larger than a number u'_k, $k = 1, 2, ..., Q$ when both $\{u_k\}$ and $\{u'_k\}$ originate from the same continuous distribution. For large values of P and Q, i.e., when P or Q are greater or equal to 30 and $P + Q \geq 40$, the U-statistics can be[10] approximated by a Gaussian distribution having a mean

$$\mu_U = \frac{PQ}{2} \tag{15}$$

and a standard deviation

$$\sigma_U = \sqrt{\frac{PQ(P+Q+1)}{12}}. \tag{16}$$

Thus, if we calculate the AUC corresponding to Eq.(9) we can estimate the correct significance level p of a k-ellipse. By direct integration of the AUC $A(k)$ of Eq.(9), we obtain that

$$\begin{aligned} A(k) = {} & \left(1 - \frac{x_1}{2}\right) + \left(\frac{Q}{Q+k}\right) \frac{x_1}{2} (x_1 - 1) \\ & + \frac{1}{2} \left(\frac{1}{Q+k}\right) \sqrt{\frac{k(Q+k+P)}{P}} \left\{ \sqrt{Q} \left[\left(x_1 - \frac{1}{2}\right) \sqrt{\frac{Q+k}{4Q} - \left(x_1 - \frac{1}{2}\right)^2} \right. \right. \\ & \left. + \left(\frac{Q+k}{4Q}\right) \arcsin\left(2\left(x_1 - \frac{1}{2}\right)\sqrt{\frac{Q}{Q+k}}\right) \right] \\ & \left. + \frac{1}{4\sqrt{Q}} \left[\sqrt{kQ} + (Q+k) \arcsin\left(\sqrt{\frac{Q}{Q+k}}\right) \right] \right\} \end{aligned} \tag{17}$$

where

$$x_1 = \frac{1}{2} + \frac{PQ - k\sqrt{Q(k+Q+P)}}{2Q(k+P)} \tag{18}$$

is the F value corresponding to $H_{\max} = 1$ in Eq.(9).

3. Discussion and Implementation

Given the values of P and Q, in order to estimate the statistical significance p in the ROC plane, we proceed as follows: For each point (F, H), we estimate using Eq.(13) the k value of the k-ellipse that passes through this point. Then, we calculate the AUC $A(k)$ using Eq.(17) and estimate the corresponding statistical significance p using either Mann-Whitney U-statistics or its Gaussian approximation where applicable. In the former case, the exact numerical solution[17] given by the Algorithm AS62 was incorporated in the FORTRAN code, whereas in the latter the GNU FORTRAN intrinsic error function `ERF(x)` was used. The table of the p-values as a function of the corresponding points (F, H) is given in the output file `outfield.dat`.

As the AUC statistics in both cases is available three k-ellipses with p-values equal to 10%, 5% and 1% are also calculated by solving

$$\mathrm{AUC}(p) = A(k) \tag{19}$$

(cf. Eq.(17)) using Newton's method. The results are written to the output file `outCL.dat`.

As the user might like to know the statistical significance of an already calculated AUC a fourth entry in the input file `input.dat` was reserved for this real number which should be smaller than unity. The resulting statistical significance is exported in the file `outp.dat` together with the integer values of P and Q . The values P, Q and N constitute the first three entries of the input file `input.dat`. The value of N denotes the number of segments in which the interval [0,1] is divided for the calculation of the p-values in the ROC plane. The final entry in the file `outp.dat` is an error code which should be zero for correct execution and becomes unity if some input parameter is unreasonable.

Typical results for the examples presented in Fig.1, are shown in Figs.2(a), (b) and (c) for M_t=3.5, 4.5 and 5.0, respectively. As mentioned in the Introduction due to Gutenberg-Richter law the values of P vary significantly in each case and are given in Table 1 together with the values of N used for the construction of Fig.2. The latter figure was prepared in `Gnuplot 4.6` by plotting `outfield.dat` using image. We observe that in all three cases statistically significant predictions are obtained since the estimated p-values for the AUC's found by numerical integration of the experimental ROC curves in each case (see the last column of Table 1) are well below 1%. We also observe that although in Fig.1 the ROC curve for $M_t = 5.0$ lies markedly higher than that for $M_t = 3.5$, the latter is by orders of magnitude more statistically significant.

Finally, we note that if in some specific case only one point (F_1, H_1) in the ROC plane is available (see Fig.3), then by using the fact that the ROC curve $H = H(F)$ is a non-decreasing function of F, e.g. see Figs.1 and 2, the unknown AUC, labeled A_1, passing through this point lies between the limits

$$H_1(1 - F_1) \leq A_1 \leq H_1 F_1 + 1 - F_1 \tag{20}$$

which correspond to the extreme ROC curves depicted with the thick green and blue lines in Fig.3, respectively. Selecting as $k_1 = k(F_1, H_1)$ from Eq.(13), the AUC $A(k_1)$ corresponding to the k_1-ellipse passing through the point (F_1, H_1), see the red ROC curve in Fig.3, also satisfies inequality (20) since

$$H_1(1 - F_1) \leq A(k_1) \leq H_1 F_1 + 1 - F_1. \tag{21}$$

This fact may additionally justify the selection of the p-values of the k-ellipses for the approximation of the statistical significance of the points on the ROC plane in Fig.2.

4. Summary

A FORTRAN code for the visualization of the confidence intervals in the ROC plane has been presented. The code provides an estimate of the statistical significance p for each point on the ROC plane based on the family of the k-ellipses introduced here. The code also provides the k-ellipses with p-values 10%, 5% and 1% which also may be of interest for researchers using ROC curves in various fields. The statistical significance p of a user defined area under the ROC curve can be also estimated.

[1] A. Garber, S. Hallerberg, H. Kantz, Predicting extreme avalanches in self-organized critical sandpiles, Phys. Rev. E 80 (2009) 026124.

[2] N. V. Sarlis, E. S. Skordas, P. A. Varotsos, Order parameter fluctuations of seismicity in natural time before and after mainshocks, EPL 91 (2010) 59001.

[3] P. A. Varotsos, N. V. Sarlis, E. S. Skordas, Natural Time Analysis: The new view of time. Precursory Seismic Electric Signals, Earthquakes and other Complex Time-Series, Springer-Verlag, Berlin Heidelberg, 2011.

[4] F. Caruso, H. Kantz, Prediction of extreme events in the OFC model on a small world network, Eur. Phys. J. B 79 (2011) 7 – 11.

[5] N. Sarlis, E. Skordas, P. Varotsos, The change of the entropy in natural time under time-reversal in the Olami-Feder-Christensen earthquake model, Tectonophysics 513 (2011) 49 – 53.

[6] N. V. Sarlis, S.-R. G. Christopoulos, Predictability of the coherent-noise model and its applications, Phys. Rev. E 85 (2012) 051136.

[7] P. A. Varotsos, N. V. Sarlis, E. S. Skordas, Order parameter fluctuations in natural time and b-value variation before large earthquakes, Natural Hazards and Earth System Science 12 (2012) 3473–3481.

[8] J. Swets, Measuring the accuracy of diagnostic systems, Science 240 (1988) 1285–1293.

[9] I. Mason, A model for assessment of weather forecasts, Aust. Met. Mag. 30 (1982) 291–303.

[10] S. J. Mason, N. E. Graham, Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical signififcance and interpretation, Quart. J. R. Meteor. Soc. 128 (2002) 2145–2166.

[11] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (2006) 861–874.

[12] B. Gutenberg, C. F. Richter, Magnitude and energy of earthquakes, Ann. Geophys. 9 (1956) 1–15.

[13] W. Feller, An Introduction to Probability Theory and Its Applications, Vol. I, Wiley, New York, 1968.

[14] J. Beringer, *et al.*, Particle Data Group, Review of particle physics, Phys. Rev. D 86 (2012) 010001.

[15] I. N. Bronshtein, K. A. Semendyayev, G. Musiol, H. Muehlig, Handbook of Mathematics, 5th Edition, Springer-Verlag, New York, 2007.

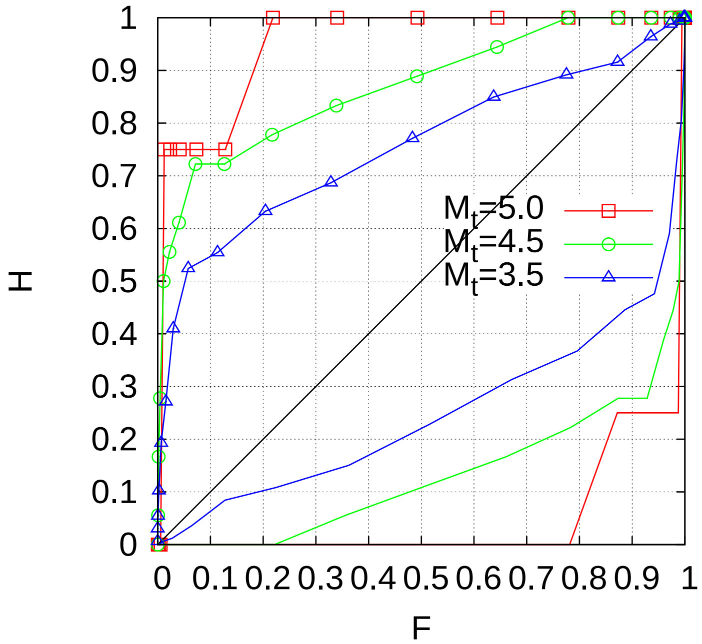

Figure 1: (color online) Receiver Operating Characteristics when using ϵ_k as a predictor for the aftershock sequence of Hector Mine earthquake for target earthquake magnitudes $M_t = 3.5$ (blue triangles), 4.5 (green circles) and 5.0 (red squares). The lines with the corresponding colors without symbols refer to the results obtained with negating the condition for the attribution of hypothesized classes.

[16] H. B. Mann, D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Statist. 18 (1947) 50–60.

[17] L. C. Dinneen, B. C. Blakesley, Algorithm AS 62: A Generator for the Sampling Distribution of the Mann-Whitney U Statistic, J. R. Stat. Soc. Ser. C Appl. Stat. 22 (1973) pp. 269–273.

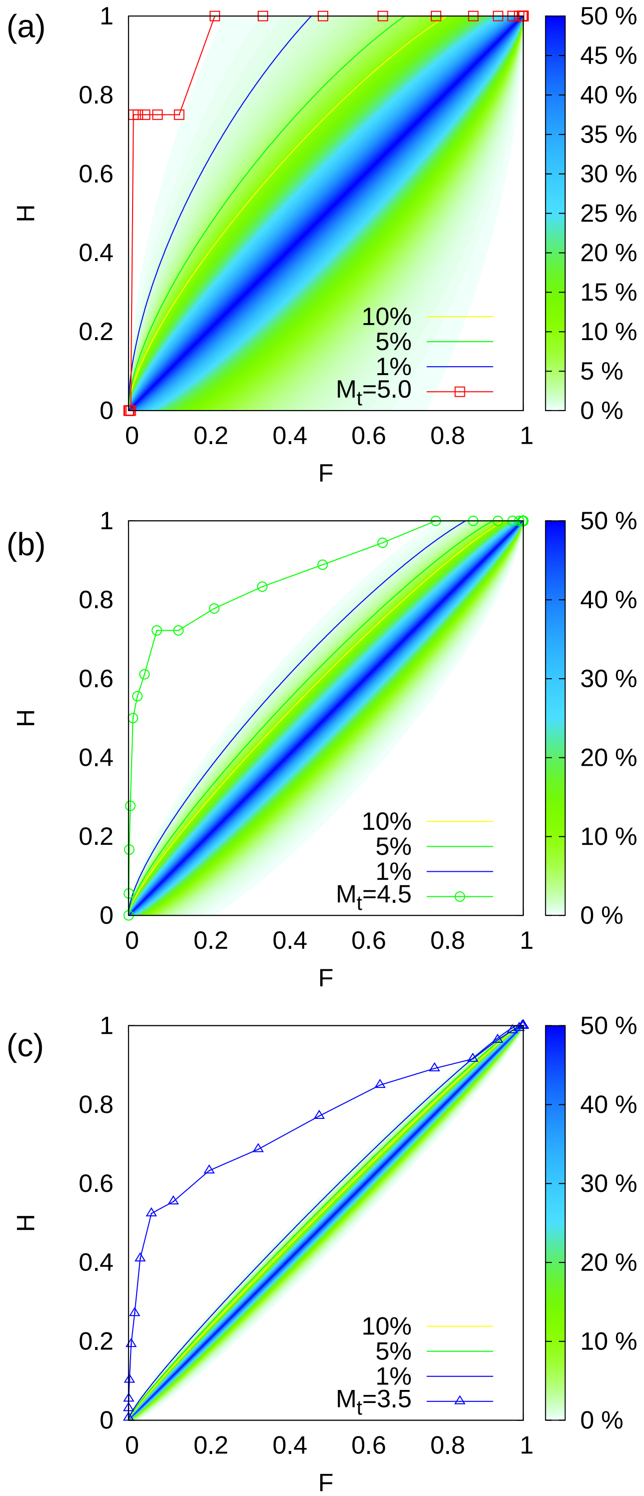

Figure 2: (color online) Receiver Operating Characteristics when using ϵ_k as a predictor for the aftershock sequence of Hector Mine earthquake for target earthquake magnitudes $M_t = 3.5$ (a), 4.5 (b) and 5.0 (c). Using the FORTRAN code presented here and the values of the second and the third column of Table 1 for each M_t, we also plot the p-values (see the color table on the left) as well the k-ellipses corresponding to 10%, 5% and 1% significance level.

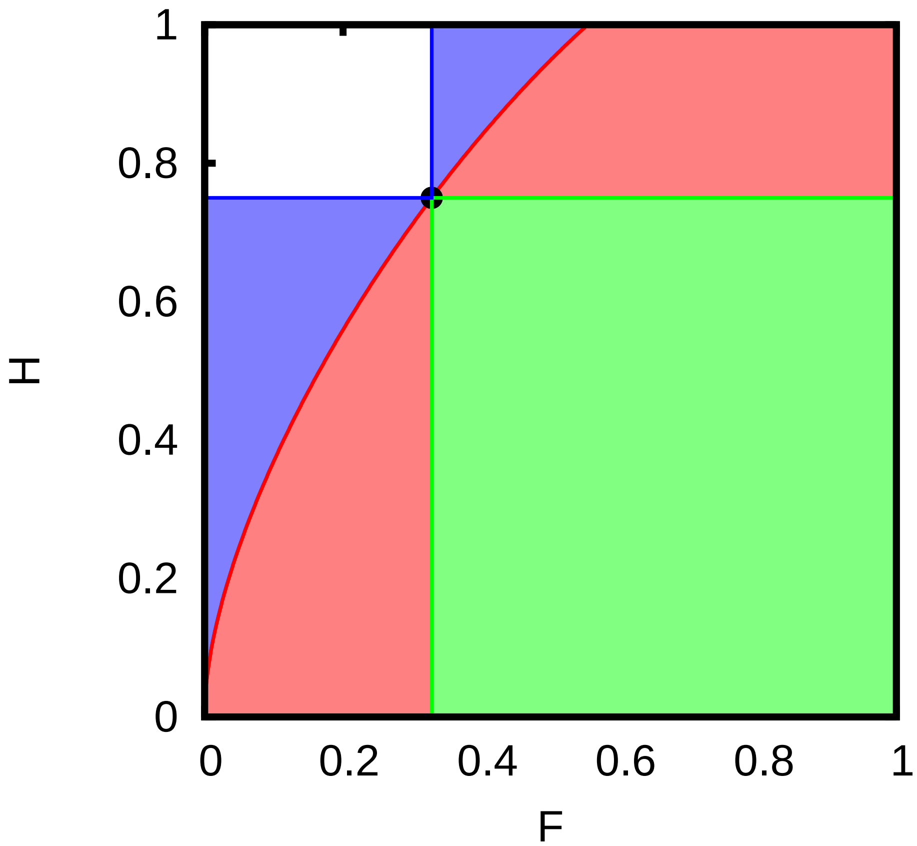

Figure 3: (color online) A schematic Receiver Operating Characteristics showing a point (F_1, H_1) (black circle) together with the k-ellipse (red thick line) passing through this point. The two extreme ROC curves which may include (F_1, H_1) are also depicted by the thick blue and thick green line. The areas under the upper extreme ROC curve (blue) and under the lower extreme ROC curve (green) bound the AUC of any ROC curve passing through (F_1, H_1), see inequalities (20) and (21).

Table 1: The values of P, Q for the various target magnitudes M_t together with the AUC of the ROC curves depicted Fig.1 and the corresponding p-values. The values of N used for the construction of Fig.2 are also inserted.

M_t	P	Q	N	AUC	p
5.0	4	4763	1000	0.950	0.09%
4.5	18	4749	1000	0.870	3×10^{-6}%
3.5	166	4601	1000	0.755	$< 10^{-6}$%