Receiver Operating Characteristic (ROC) Curve: A Tool for Describing and Comparing Continuous Diagnostic Tests

Ivana Horová

with cooperation of Jiří Zelinka

Department of Mathematics and Statistics, Faculty of Science

Masaryk University, Czech Republic

Contents

- Introduction

- Measure of Diagnostic Accuracy

- Receiver Operating Characteristic Curve

- Estimates of ROC curves

  Parametric methods

  Nonparametric methods

  Simulations

- Summary ROC measures

  Area under curve

  Partial area under curve

  Maximum inprovement of sensitivity

- Application for real data

Introduction

ROC curves and their analysis are based on statistical decision theory, they were originally developed for electronic-detection-signal theory (see Zhou *et al.* 2002 for details).

The concept of ROC curves was introduced in medicine by Lee Lusted in 1971.

Recently, there has been an increased use of ROC methodology in a wide area of different disciplines.

Statistical aspects of ROC analysis: many excellent books and papers are available (e.g. Pepe 2003 or Zhou *et al.* 2002 and list of References).

Aim: To explain the definition, properties and constructions of ROC curves and to make them accessible for the general scientific audience.

Diagnostic tests play an important role in medical care and contribute significantly to health care cost.

A diagnostic test has two purposes:

1. to provide reliable information about the patient's condition

2. to influence the health care provider's plan for managing the patients

A test can serve these purposes only if the health care provider knows how to interpret it.

This information is acquired through an assessment of the **test's diagnostic accuracy, which is the ability of a test to detect correctly a condition when it is actually present and to correctly rule out when it is truly absent**.

Two basic measures of diagnostic accuracy are **sensitivity** and **specificity**.

Measure of Diagnostic Accuracy

- $\mathcal{G}_1$ group of subjects with a condition

- $\mathcal{G}_0$ group of subjects without a condition

- $D = 0, 1$ random variable denotes absence or presence of the condition

- $T = 1$ positive test result

- $T = 0$ negative test result

Test Results (Confusion matrix)

	Positive test, $T = 1$	Negative test, $T = 0$	Total
$\mathcal{G}_1\ (D = 1)$	True positive (a)	False negative (b)	$a + b$
$\mathcal{G}_0\ (D = 0)$	False positive (c)	True negative (d)	$c + d$
Total	$a + c$	$b + d$	

The sensitivity (Se) of the test is its ability to detect the condition when it is present.

$Se = P(T = 1 | D = 1)$ is a probability P that the test result is positive $(T = 1)$, given that the condition is present $(D = 1)$,

$$Se = \frac{a}{a + b}$$

The specificity (Sp) of a test is its ability to exclude the condition when it is absent.

$Sp = P(T = 0 | D = 0)$ is a probability P that the test result is negative $(T = 0)$, given that the condition is absent $(D = 0)$,

$$Sp = \frac{d}{c + d}, \quad FPR = 1 - Sp = \frac{c}{c + d}, \quad FPR = \textbf{false positive rate}$$

$\mathcal{G}_1$	True positive $\quad(a)$	False negative $\quad(b)$	$a + b$
$\mathcal{G}_0$	False positive $\quad(c)$	True negative $\quad(d)$	$c + d$
Total	$a + c$	$b + d$	

Example

The accuracy of screening mammography test results:

- 30 patients with pathology proven breast cancer

- 30 patients without disease

The mammograph was positive if the mammographer recommended additional diagnostic follow-up.

Test results

Cancer status	Positive	Negative	Total
Present	29	1	30
Absent	19	11	30
Total	48	12	60

$$Se = \frac{29}{30} = 0.967, \quad Sp = \frac{11}{30} = 0.367, \quad FPR = \frac{19}{30} = 0.633$$

Mammographer used a different decision treshold

Test results

Cancer status	Positive	Negative	Total
Present	23	7	30
Absent	8	22	30
Total	31	29	60

$$Se = \frac{23}{30} = 0.767, \quad Sp = \frac{22}{30} = 0.733, \quad FPR = \frac{19}{30} = 0.267$$

Receiver Operating Characteristic (ROC) Curve

The accuracy of a medical diagnostic test is often summarized in a **Receiver Operating Characteristic (ROC) Curve**.

The ROC curve is defined as a plot of the probability

    of **false classification (1-Sp)** of subjects from $\mathcal{G}_0$

versus the probability

    of **true classification (Se)** of subjects from $\mathcal{G}_1$

across of all possible values of the given test.

Explicit formula I

- X – the diagnostic test variable (one-dimensional absolutely continuous random variable)

- The subject is classified as $\mathcal{G}_1$ if $X \geq c$ and $\mathcal{G}_0$ otherwise for given cutoff point $c \in \mathbb{R}$

- $F_0(c) = P(X \leq c | \mathcal{G}_0) = \int\limits_{-\infty}^{c} f_0(x)dx$

  $F_1(c) = P(X \leq c | \mathcal{G}_1) = \int\limits_{-\infty}^{c} f_1(x)dx$

  F_0 or F_1 are distribution functions of group $\mathcal{G}_0$ or $\mathcal{G}_1$, respectively, and f_0 and f_1 are corresponding density functions.

- $\quad F_0 \quad - \quad$ the specificity (Sp) of the test

- $1 - F_1 \quad - \quad$ the sensitivity (Se) of the test

- p – the probability of false classification of subject from $\mathcal{G}_0$

- q – the probability of true classification of subject from $\mathcal{G}_1$

$$p = 1 - F_0(c) \Rightarrow c = F_0^{-1}(1-p), \quad 0 \le p \le 1$$
$$q = 1 - F_1(c) = 1 - F_1(F_0^{-1}(1-p)), \quad 0 \le p \le 1$$

$$ROC(p) = R(p) = 1 - F_1(F_0^{-1}(1-p)), \quad 0 \le p \le 1$$

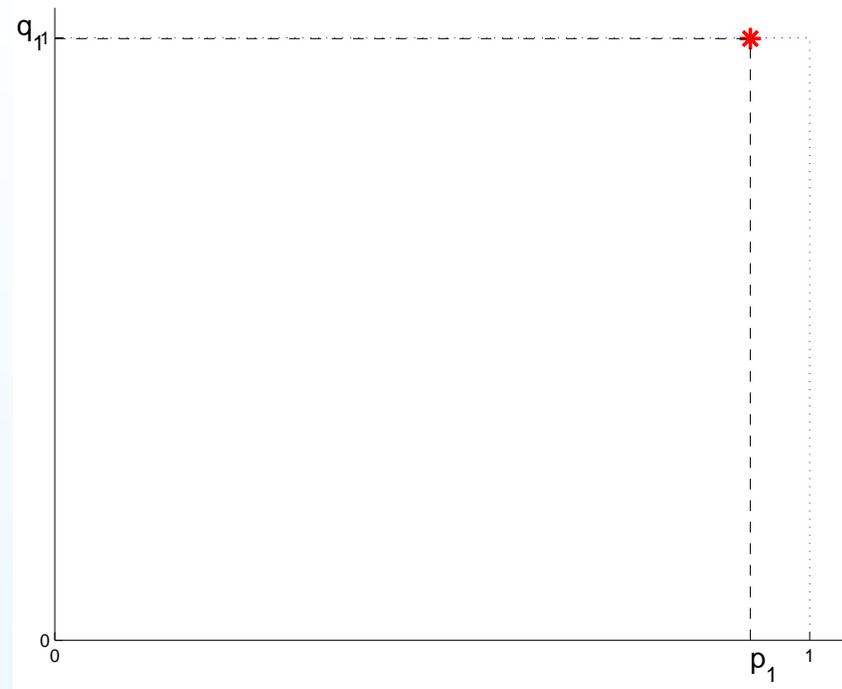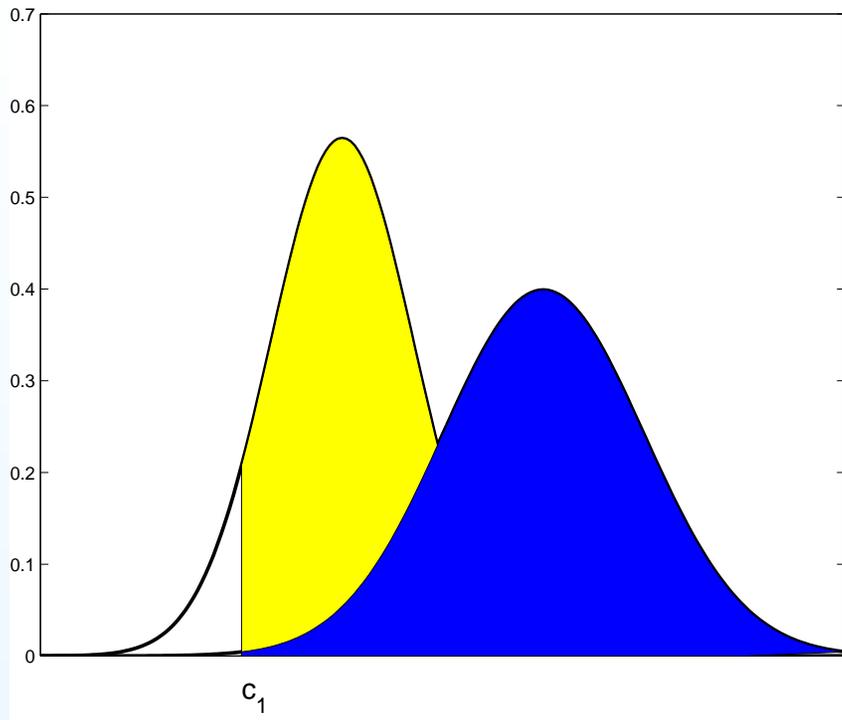**ROC curve is displayed by plotting $1 - F_1(c)$ against $1 - F_0(c)$ for a range of cutoff points $c \in \mathbb{R}$.**

Notation

X_j, $j = 0, 1$ denote random variable X if $D = j$, $j = 0, 1$, X_0, X_1 are independent
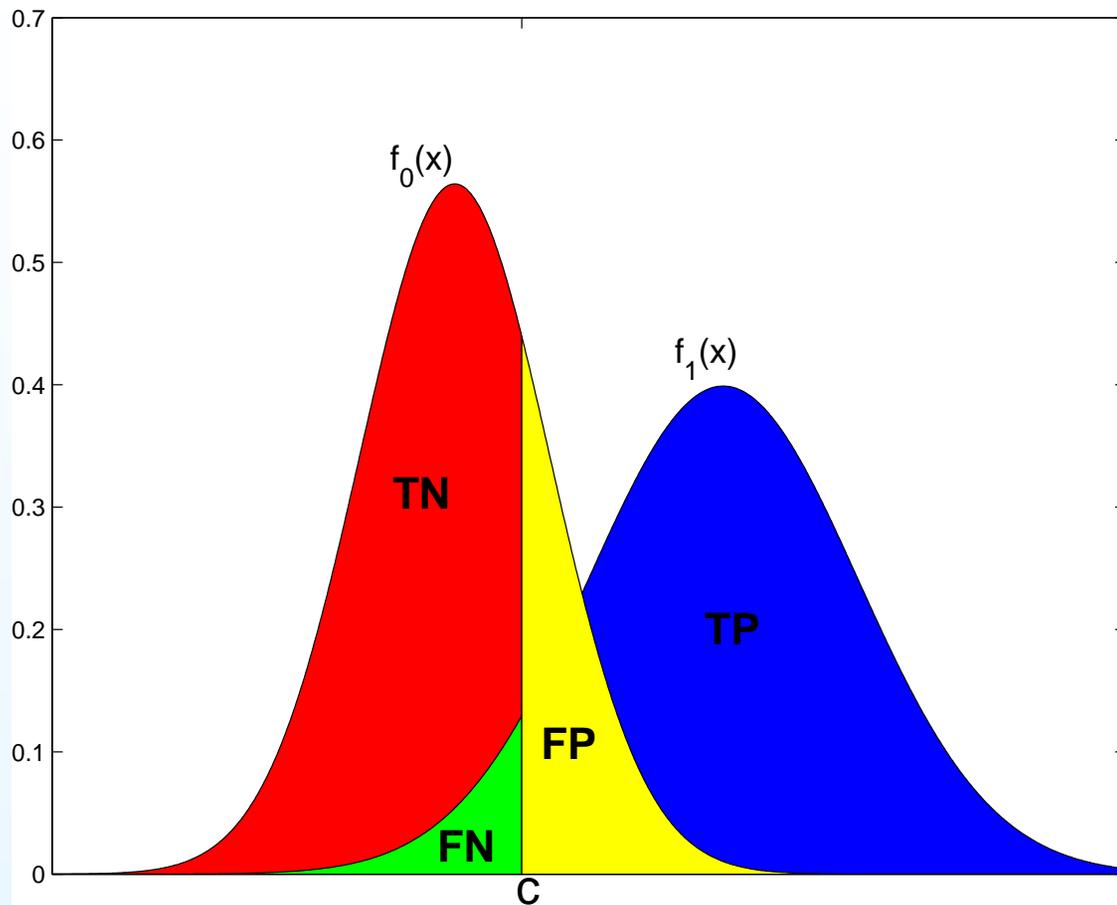
TP – True positive   FP – False positive

FN – False negative   TN – True negative

$$f_0(x) = \frac{1}{\sqrt{\pi}}\, e^{-x^2} \qquad f_1(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(x-2)^2}{2}}$$
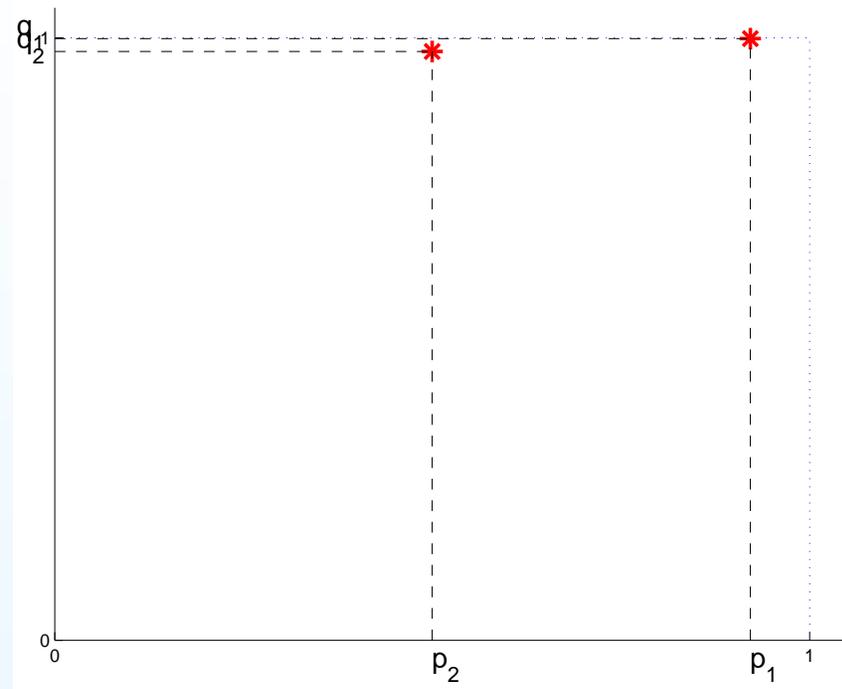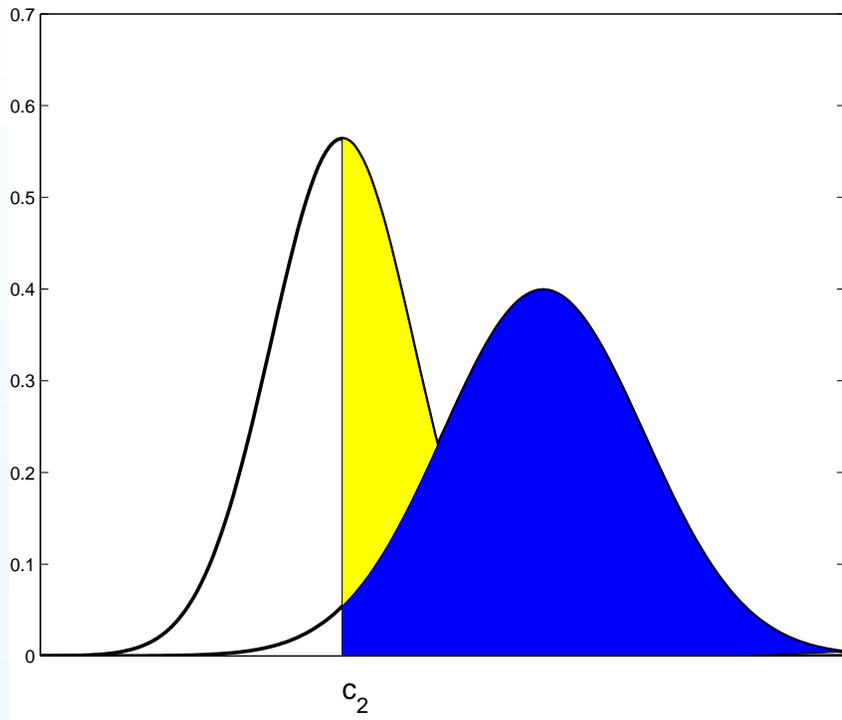
$$c_1 = -1$$

$$
\begin{aligned}
p_1 &= 1 - F_0(c_1) &= \int_{-1}^{\infty} f_0(x)\,dx &= 0.9214 \\
q_1 &= 1 - F_1(c_1) &= \int_{-1}^{\infty} f_1(x)\,dx &= 0.9987
\end{aligned}
$$

$$Sp = 0.0786, \qquad Se = 0.9987$$

$$c_2 = 0$$

$$
\begin{array}{rcccccc}
p_2 & = & 1 - F_0(c_2) & = & \int\limits_0^\infty f_0(x)dx & = & 0.5 \\
q_2 & = & 1 - F_1(c_2) & = & \int\limits_0^\infty f_1(x)dx & = & 0.9772
\end{array}
$$

$$Sp = 0.5, \qquad Se = 0.9772$$

$$c_3 = 1$$
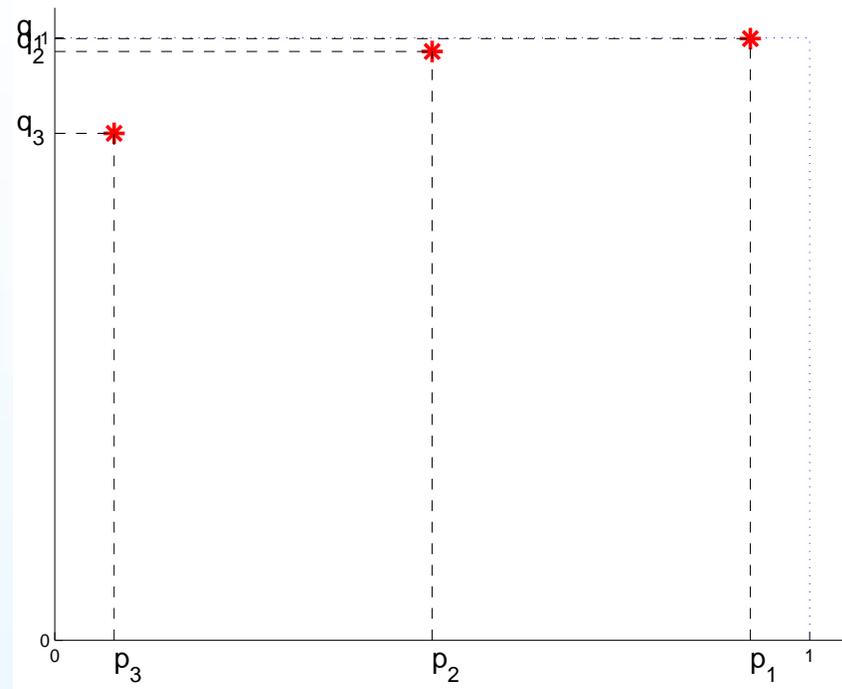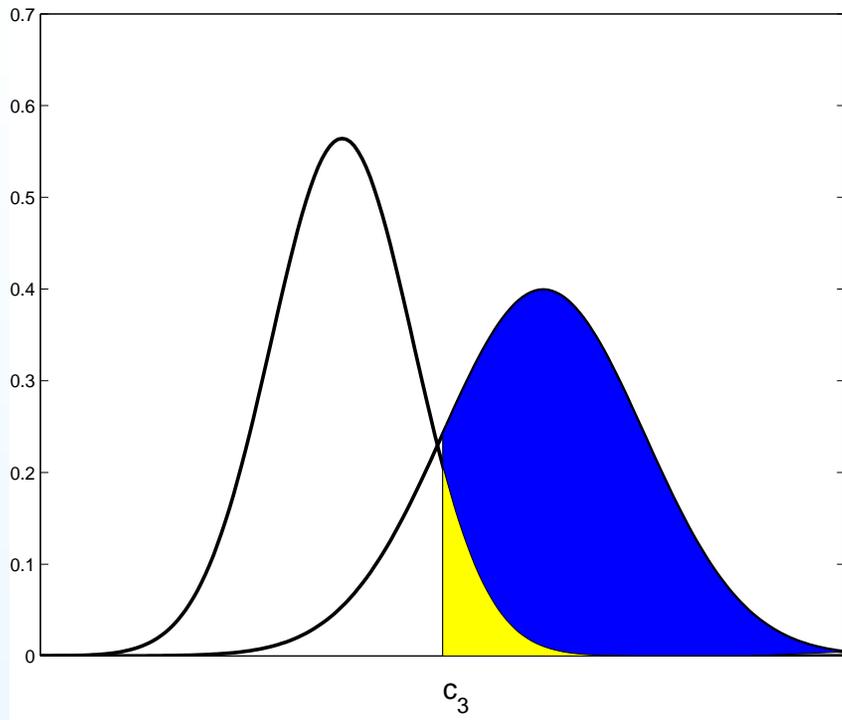
$$
\begin{aligned}
p_3 &= 1 - F_0(c_3) &&= \int_1^\infty f_0(x)dx &&= 0.0786 \\
q_3 &= 1 - F_1(c_3) &&= \int_1^\infty f_1(x)dx &&= 0.8413
\end{aligned}
$$

$$Sp = 0.9214, \qquad Se = 0.8413$$

$$c_4 = 2$$

$$p_4 = 1 - F_0(c_4) = \int_2^\infty f_0(x)dx = 0.0023$$

$$q_4 = 1 - F_1(c_4) = \int_2^\infty f_1(x)dx = 0.5$$
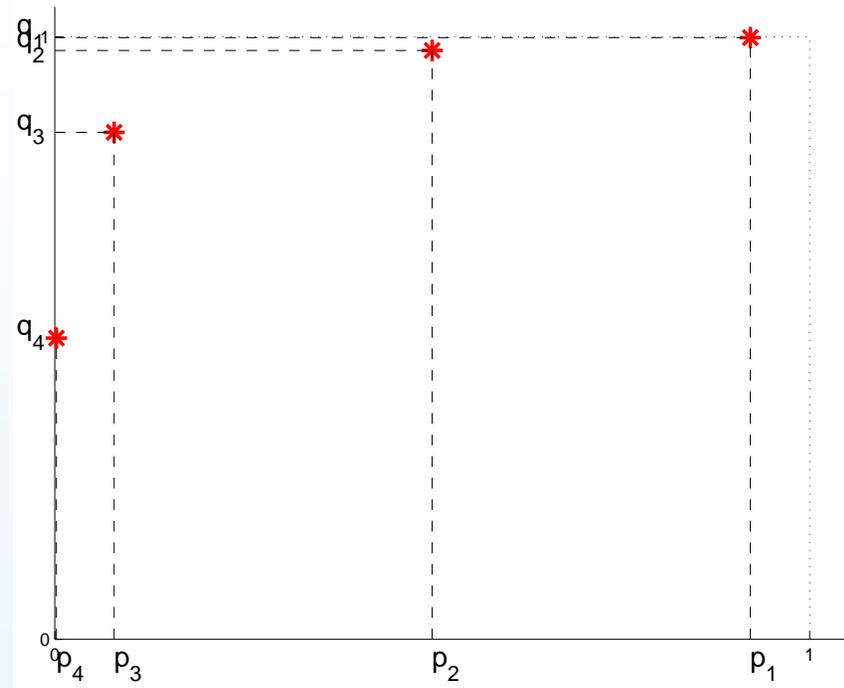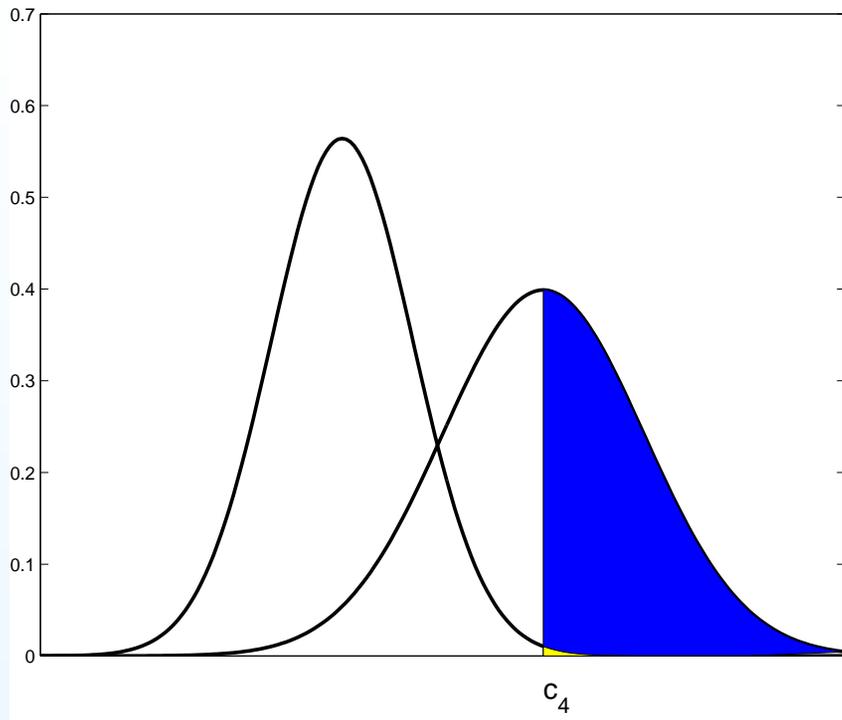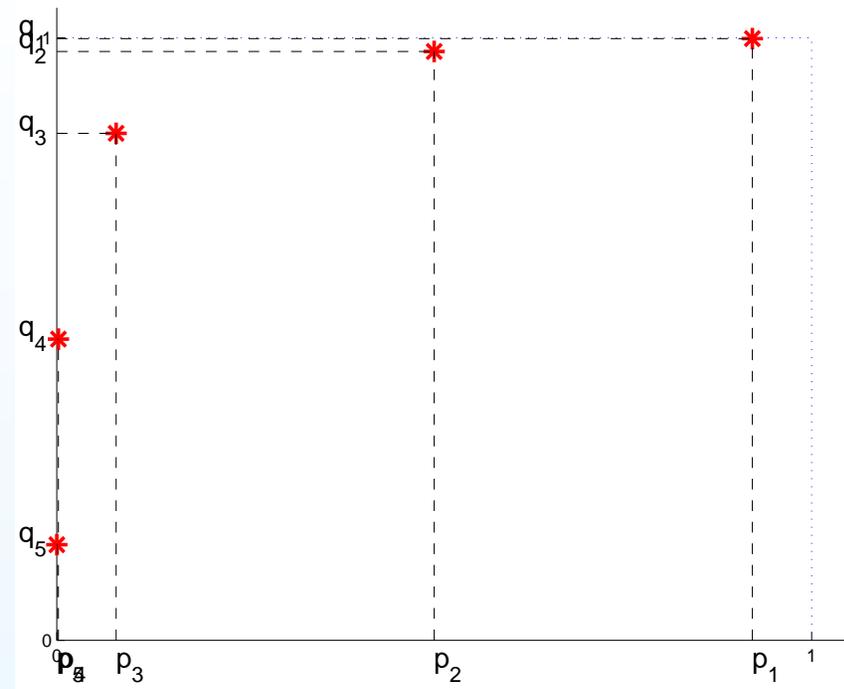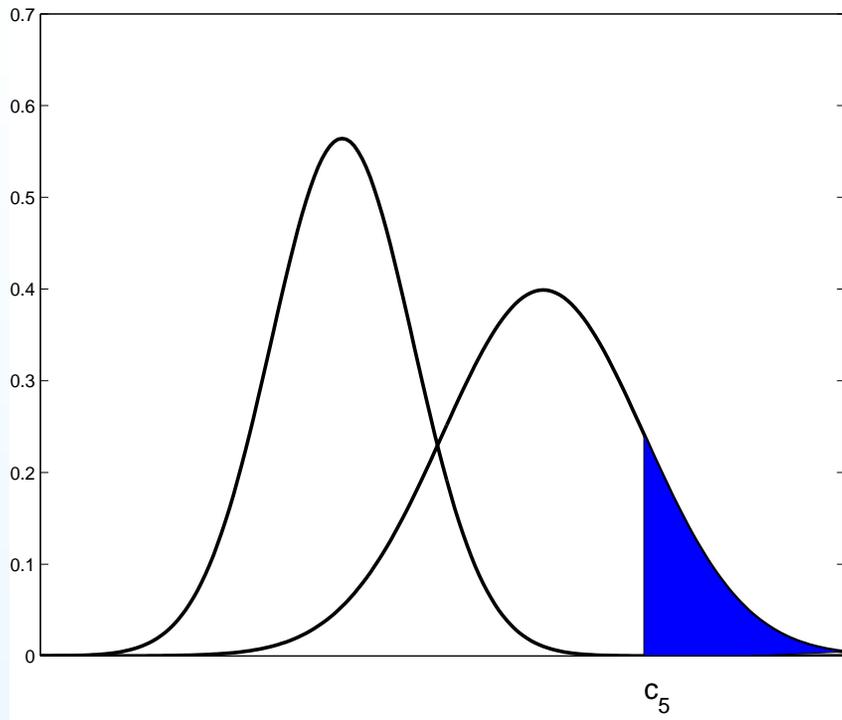
$$Sp = 0.9977, \qquad Se = 0.5$$

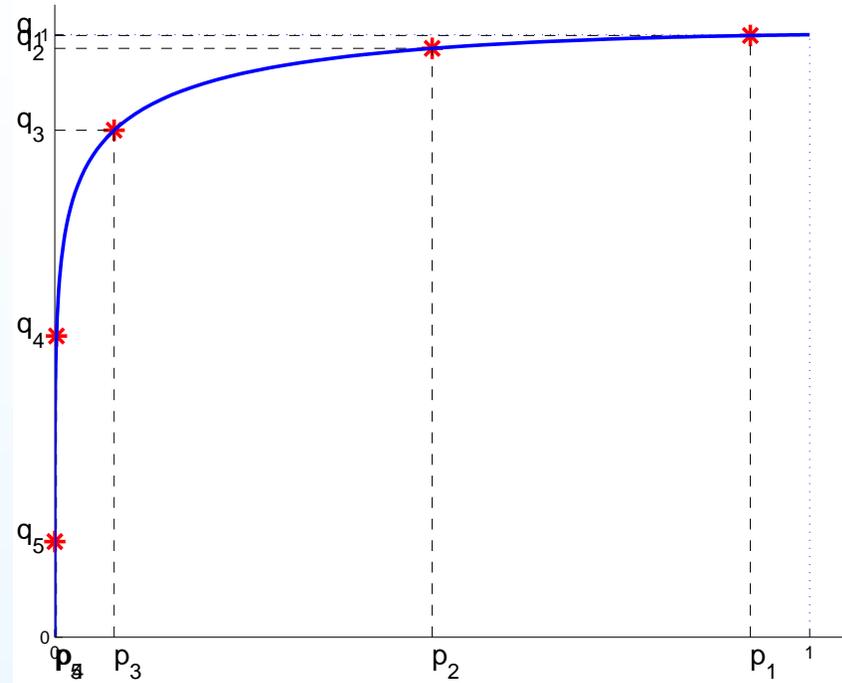$$c_5 = 3$$

$$
\begin{aligned}
p_5 &= 1 - F_0(c_5) &= \int\limits_{3}^{\infty} f_0(x)dx &= 0.00001 \\
q_5 &= 1 - F_1(c_5) &= \int\limits_{3}^{\infty} f_1(x)dx &= 0.1587
\end{aligned}
$$

$$Sp = 0.99999, \qquad Se = 0.1587$$
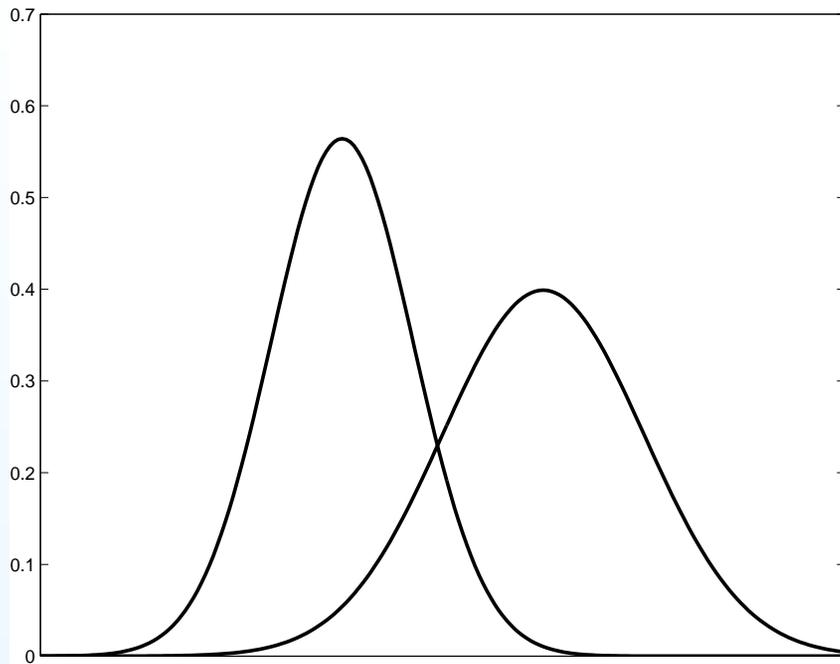
$$p_i = 1 - F_0(c_i) = \int\limits_{c_i}^{\infty} f_0(x)dx \quad q_i = 1 - F_1(c_i) = \int\limits_{c_i}^{\infty} f_1(x)dx, \quad i = 1, \ldots, 5$$
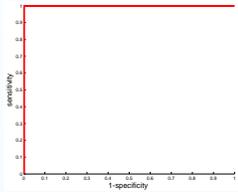
Point $(1, 1)$ –   all subject are classified to be from $\mathcal{G}_1$

Point $(0, 0)$ –   all subject are classified to be from $\mathcal{G}_0$

Extreme cases

A perfectly accurate test because sensitivity is 1.0 when 1-specificity is 0.0
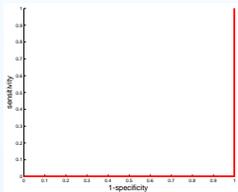
A perfectly inaccurate test, patients with the condition are located incorrectly as negative and patients without condition are located incorrectly as positive
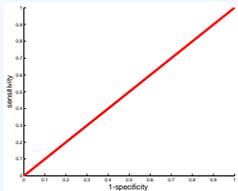
A diagonal – chance diagonal. The test is not usable for separation of the patients.

Diagnostic tests with ROC curves above the chance diagonal have at least some ability do discriminate between patients with and without condition.

ROC curve close to the perfectly accurate one
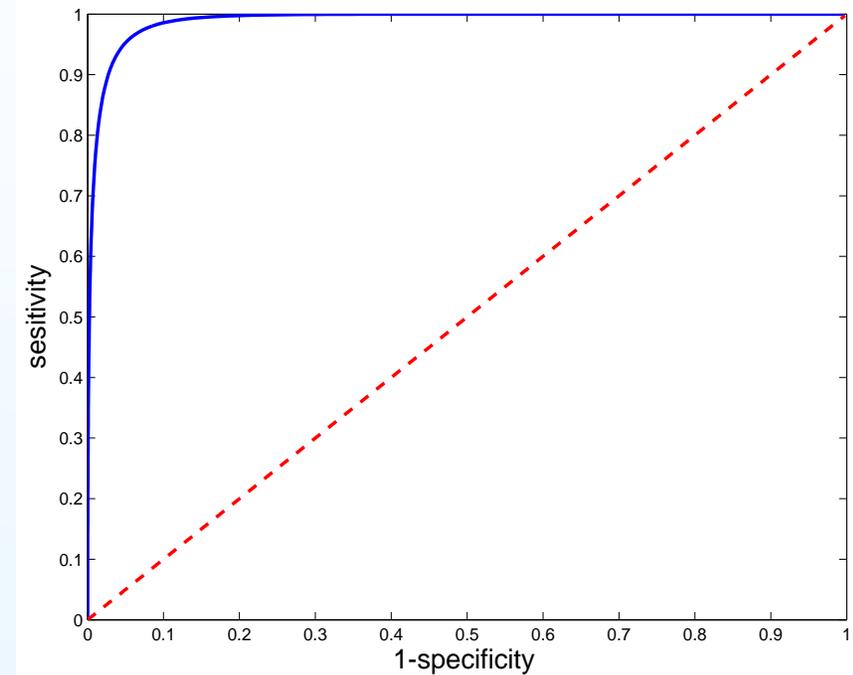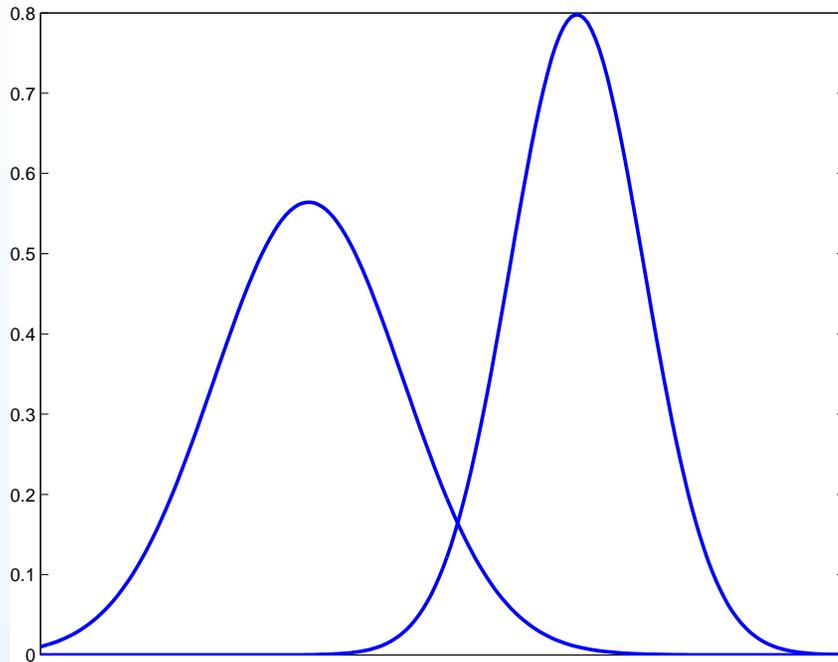
$$f_0(x) = \frac{1}{\sqrt{\pi}}\, e^{-(x+1)^2} \qquad f_1(x) = \frac{1}{0.5\sqrt{2\pi}}\, e^{-\frac{(x-1)^2}{2\cdot 0.5^2}}$$

Explicit formula II

- $ROC(p) = R(p)$ – the distribution function of $1 - F_0(X_1)$, i.e.

- $R(p)$ is the nonzero distribution function of the p-value $1 - F_0(X_1)$ for testing the null hypothesis that an individual comes from $\mathcal{G}_0$

$$V = 1 - F_0(X_1)$$

$$F_V(p) = P(V \leq p) =$$

$$= P(1 - F_0(X_1) \leq p) =$$

$$= P(X_1 \geq F_0^{-1}(1 - p)) =$$

$$= 1 - F_1(F_0^{-1}(1 - p)) = R(p)$$

Estimates of ROC curves

Parametric methods

(see DeLong *et al.* 1988, Zhou *et al.* 2002, Pepe 2003)
X – diagnostic variable $\rightarrow f_X(x) = \alpha_0 f_0(x) + \alpha_1 f_1(x)$, $\alpha_0 + \alpha_1 = 1$, $\alpha_{0,1} \geq 0$.
f_0, f_1 – normal (Gaussian) densities with means μ_0, μ_1 and variances σ_0^2, σ_1^2, respectively,

$$f_j(x) = \frac{1}{\sigma_j \sqrt{2\pi}}\, e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}}, \quad j = 0, 1\,.$$

The ROC curve:

$$R(p) = \Phi\left(a + b\Phi^{-1}(p)\right)$$

$a = \frac{\mu_1 - \mu_0}{\sigma_1}, \quad b = \frac{\sigma_0}{\sigma_1},$

Φ – standard normal distribution function, $\Phi(x) = \int\limits_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$

Nonparametric methods

The empirical ROC curve: F_0 and F_1 are replaced by their cumulative distribution function.

Kernel methods: F_0 and F_1 are estimated by kernel methods (e.g. Azzalini 1981, Lejeune and Sarda 1992, Altman and Léger 1995, Zou, Hall and Shapiro 1997, Bowman *et al.* 1998, Lloyd 1998, Lloyd and Yong 1999, Hall and Hyndman 2002, Zhou *et al.* 2002, Zhou and Harezlak 2002, Peng and Zhou 2004).

Nonparametric estimates of distribution function

Let $Z_1, \ldots, Z_n$ be random sample from random variable Z with distribution function F.

Empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(Z_i \leq x).$$

Kernel estimate of distribution function

$$\hat{F}_h(x) = \frac{1}{n} \sum_{i=1}^{n} W\left(\frac{x - Z_i}{h}\right), \qquad W(x) = \int_{-1}^{x} K(t)\,dt$$

- K – a kernel, a non-negative symmetric function, supported on $[-1, 1]$, integrated to unity
  $K(x) = \frac{15}{16}(1 - x^2)^2 I_{[-1,1]}$, $K(x) = \frac{3}{4}(1 - x^2)I_{[-1,1]}$

- h – a smoothing parameter (bandwidth), $h = h(n)$ – a sequence of nonrandom positive numbers, $h \to 0$, $nh \to \infty$ as $n \to \infty$.

Nonparametric estimates of ROC curve

Notation

- Independent samples $X_{0,1}, \ldots X_{0,n_0}$ from $\mathcal{G}_0$ and $X_{1,1}, \ldots X_{1,n_1}$ from $\mathcal{G}_1$ on, respectively F_0 and F_1 are at hand

Empirical ROC curve

F_0 and F_1 are replaced by their empirical distribution functions

$$F_{n_j}(x) = \frac{1}{n_j} \sum_{i=1}^{n_j} I(X_{i,j} \leq x), \quad j = 0, 1$$

and

$$\hat{R}(p) = 1 - \hat{F}_1(\hat{F}_0^{-1}(1 - p)).$$

Smooth kernel estimate of ROC curve

Estimates of F_0 and F_1:

$$\hat{F}_{j,h_j}(x) = \frac{1}{n_j} \sum_{i=1}^{n_j} W\left(\frac{x - X_{j,i}}{h_j}\right), \quad j = 0, 1$$

Kernel formula I

$$\hat{R}(p) = 1 - \hat{F}_{1,h_1}(\hat{F}_{0,h_0}^{-1}(1-p)), \quad 0 \leq p \leq 1, \quad h_j = O(n_j^{-1/3}), \quad j = 0, 1$$

Estimates of optimal bandwidths for $\hat{F}_j, \; j = 0, 1$ need not to be optimal for $\hat{R}(p)$.

Kernel formula II.

$$\hat{R}(p) = \hat{F}_{V,h_1}(p) = \frac{1}{n_1} \sum_{i=1}^{n_1} W\left(\frac{p - (1 - \hat{F}_0(X_{1,i}))}{\tilde{h}_1}\right)$$

$$\hat{F}_{0,h_0}(X_{1,i}) = \frac{1}{n_0} \sum_{j=1}^{n_0} W\left(\frac{X_{1,i} - X_{0,j}}{h_0}\right), \qquad W(x) = \int_{-1}^{x} K(t)dt$$

Problems with bandwidth selection

$$\hat{F}_h(x) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x - Z_i}{h}\right), \quad W(x) = \int_{-1}^{x} K(t)dt$$

Mean Integrated Square Error (E denotes an expectation)

$$MISE(\hat{F}_h) = \int E(\hat{F}_h(x) - F(x))^2 dx,$$

Optimal bandwidth minimizing $MISE(\hat{F}_h)$ provided that $F \in C^2$:

$$h_{opt} = n^{-1/3}\left(\frac{c_1}{\beta_2^2 \psi_2}\right)^{1/3}$$

$$c_1 = \int_{-1}^{1} W(x)(1 - W(x))dx > 0, \quad \beta_2 = \int_{-1}^{1} x^2 K(x)dx, \quad \psi_2 = \int (F''(x))^2 dx$$

Methods for estimation of the optimal bandwidth:

- Terrell and Scott (1985), Terrell (1990) – maximal smoothing principle

- Sarda (1993) – a cross-validation method

- Altman and Léger (1995), Zhou and Harezlak (2002) – a method of the reference (Gaussian) density

- Lloyd and Yong (1999) – a more complex selection of bandwidth, procedure based on two-stage plug-in method

- Zhou et al. (2002) – the bandwidths optimal for densities estimates

- Hall and Hyndman (2003) – a method allows interaction between distribution for each group

- Peng and Zhou (2004) – a method is based on local linear smoothing

- Horová and Zelinka (2007) – an iterative method

- Horová et al. (2007)

Simulations

Normal data
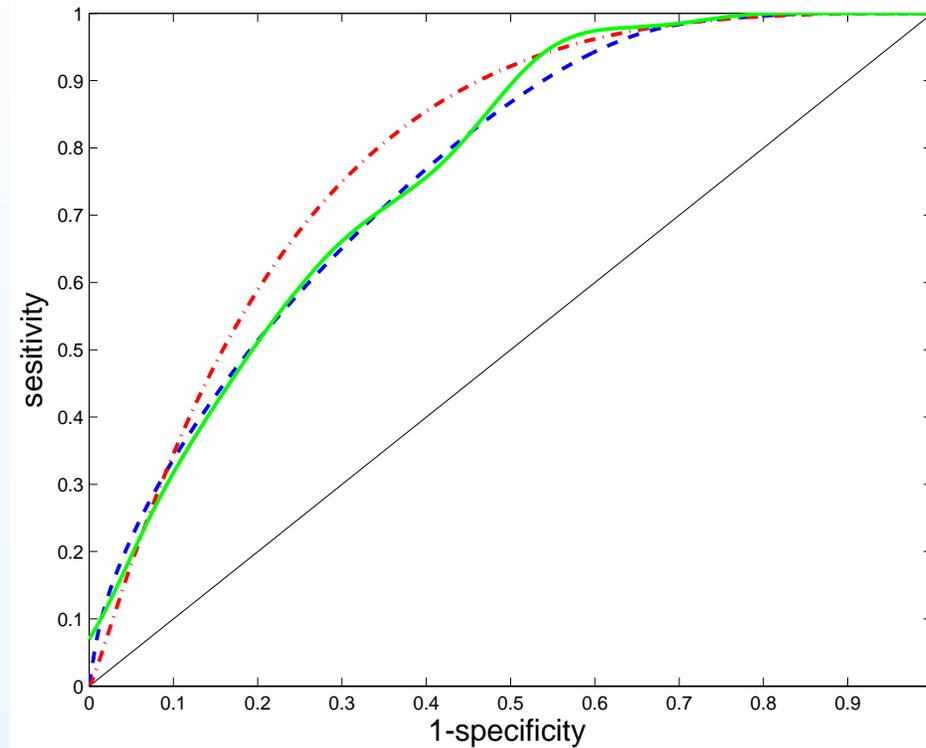
Colour marking:

$- \cdot -$ true ROC curve $f_0(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(x)^2}{2}},\ f_1(x) = \frac{1}{\sqrt{\pi}}\, e^{-(x-1)^2}$

$- - -$ Kernel formula I $h_0 = 0.8595,\ h_1 = 0.6781$

——   Kernel formula II $h_0 = 0.8595,\ \tilde{h}_1 = 0.0982$

Exponential data
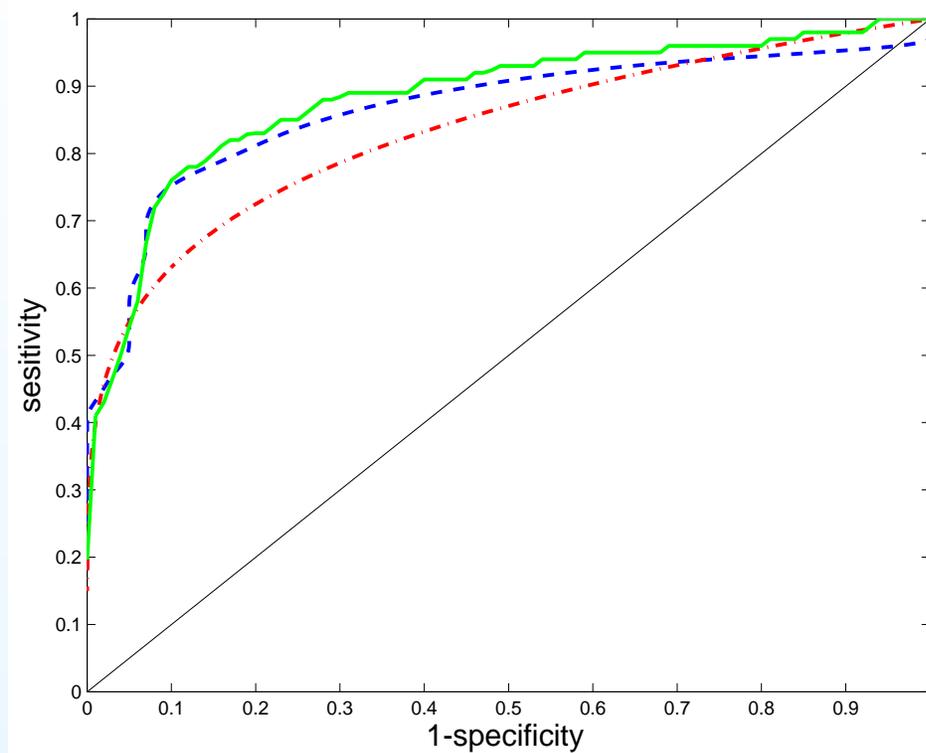
Colour marking:

$- \cdot -$	true ROC curve	$f_0(x) = e^{-x}, \; f_1(x) = \frac{1}{5}\, e^{-\frac{x}{5}}$
$- - -$	Kernel formula I	$h_0 = 0.3424, \; h_1 = 2.2254$
——	Kernel formula II	$h_0 = 0.3424, \; \tilde{h}_1 = 0.0010$

Simulation study

We generated 1000 random samples of normally distributed random variables for testing the quality of kernel estimates of ROC curve: $X_{0,i} \sim N(0,1)$, $X_{1,i} \sim N(1.5, 05)$, $i = 1, \ldots 100$. Following figures present the bounds (yellow area) containing all estimates of ROC curve for the both kernel formulae (dashed blue lines) and the true ROC curve (solid red line), as well.
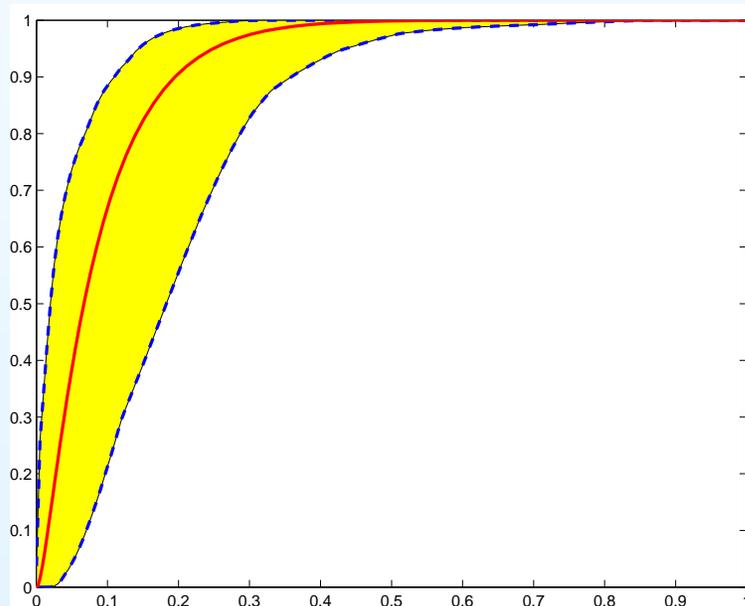
Kernel formula **I**
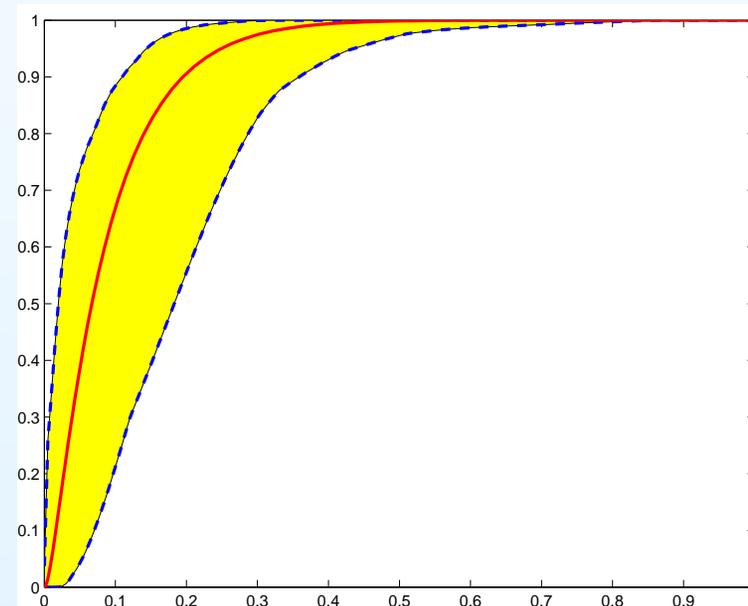
Kernel formula **II**

Summary ROC measures

- Area under ROC curve (AUC)

- Partial area under ROC curve (PAUC)

- Specificity corresponding to maximum improvement of sensitivity (MIS)

Area under curve

- The most common used global index of diagnostic accuracy is the **area under the ROC curve – AUC**

- The area under the ROC curve is the probability that a pair of individuals known to be from different groups will be correctly classified.

$$AUC(R(p)) = \int\limits_{0}^{1} R(p)dp$$

- A simple calculation shows that the area under ROC curve is exactly equal to the probability $P(X_0 < X_1)$:

$$AUC(R(p)) = P(X_0 < X_1)$$

- Values of AUC close to 1.0 indicate that the test has high diagnostic accuracy

Empirical AUC

The **empirical AUC**: calculate the trapezoidal area under each vertical slice of an empirical ROC curve having a straight-line segment as its top; then sum all individual areas.

$$\widehat{AUC}_{emp} = \frac{1}{n_0 n_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} \Psi\left(X_{0j}, X_{1i}\right)$$

where $X_{01},\ldots,X_{0n_0}$ and $X_{11},\ldots,X_{1n_1}$ are independent samples from F_0 and F_1, respectively and

$$\Psi\left(X_{0j}, X_{1i}\right) = \begin{cases} 1 & X_{1i} > X_{0j}, \\ \frac{1}{2} & X_{1i} = X_{0j} \\ 0 & \text{otherwise,} \end{cases} \quad \begin{aligned} & j = 0, 1 \\ & i = 1, \ldots, n_j. \end{aligned}$$

Remark: It is analogous to the Mann-Witney U-statistics.

AUC for binormal model

- X (a diagnostic test variable) $\longrightarrow f_X(x) = \alpha_0 f_0(x) + \alpha_1 f_1(x)$

$$f_j(x) = \frac{1}{\sigma_j \sqrt{2\pi}} \, e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}}, \quad j = 0, 1 \, .$$

$$R(p) = \Phi\left(a + b\Phi^{-1}(p)\right), \quad a = \frac{\mu_1 - \mu_0}{\sigma_1}, \quad b = \frac{\sigma_0}{\sigma_1}$$

$$AUC = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right)$$

- Φ – standard normal distribution function

Nonparametric methods of estimates of AUC

Composite trapezoidal rule

The estimates of F_0 and F_1 are evaluated in some set $\{x_r \in \mathbb{R};\ r = 0 \ldots N\}$, mostly $x_r = x_0 + r\,t$, $t > 0$.

The kernel estimate $\widehat{R}$ of the ROC curve is formed by pairs of points $[p_r, \hat{R}(p_r)]$ where

$$p_r = 1 - \widehat{F}_0(x_r), \quad \widehat{R}(p_r) = 1 - \widehat{F}_1(x_r), \quad r = 0, \ldots, N.$$

p_r is non-increasing in r. The composite trapezoidal rule yields

$$\widehat{AUC} = \sum_{r=1}^{N} \frac{1}{2}(p_{r-1} - p_r)\left(\widehat{R}(p_{r-1}) + \widehat{R}(p_r)\right) =$$

$$= \frac{1}{2}\sum_{r=1}^{N}\left(\widehat{F}_{0,h_0}(x_r) - \widehat{F}_{0,h_0}(x_{r-1})\right)\left(2 - \widehat{F}_{1,h_1}(x_{r-1}) - \widehat{F}_{1,h_1}(x_r)\right)$$

The 1st method of kernel estimation of AUC (Kernel 1)

In terms of distribution function AUC can be expressed as

$$AUC = P(X_0 < X_1) = P(X_0 - X_1 < 0) = F_{X_0 - X_1}(0) = F^c(0),$$

where $F^c = F_{X_0 - X_1}$ is a distribution function of a random variable $Y = X_0 - X_1$. Then a kernel estimate of F^c is

$$\hat{F}^c_{h_0, h_1}(x) = \frac{1}{n_0 n_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} W\left(\frac{x - (X_{0j} - X_{1i})}{\sqrt{h_0^2 + h_1^2}}\right),$$

where h_0 and h_1 are the bandwidths for F_0 and F_1, respectively (Lloyd 1998).

Hence the kernel estimate $\widehat{AUC}_I$ of AUC is given by

$$\widehat{AUC}_I = \hat{F}^c_{h_0, h_1}(0) = \frac{1}{n_0 n_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} W\left(\frac{X_{1i} - X_{0j}}{\sqrt{h_0^2 + h_1^2}}\right)$$

The 2nd method of kernel estimation of AUC (Kernel 2)

An estimate of F^c by means of the **only bandwidth** h, i. e.

$$\hat{F}_h^c(x) = \frac{1}{n_0 n_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} W\left(\frac{x - (X_{0j} - X_{1i})}{h} \right),$$

and

$$h_{opt}^{F^c} = (n_0 n_1)^{-1/3} \left(\frac{c_1}{\beta_2^2 \psi_2^c} \right)^{1/3}, \qquad h_{opt}^{F^c} \approx O((n_0 n_1)^{-1/3}),$$

where

$$\psi_2^c = \int \left(F^{c\prime\prime}(x) \right)^2 dx,$$

$$\widehat{AUC}_{II} = \hat{F}_h^c(0) = \frac{1}{n_0 n_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} W\left(\frac{X_{1i} - X_{0j}}{h} \right).$$

The 3rd method of kernel estimation of AUC (Kernel 3)

This method uses the Kernel formula II for ROC estimate.

We get by direct integration:

$$\widehat{AUC}_{III} = \int_0^1 \hat{R}(p)dp = \int_0^1 \hat{F}_{V,h_1}(p)dp =$$

$$= \frac{1}{n_1}\sum_{i=1}^{n_1}\int_0^1 W\left(\frac{p-(1-\hat{F}_0(X_{1,i}))}{\tilde{h}_1}\right)dp$$

$$\hat{F}_{0,h_0}(X_{1,i}) = \frac{1}{n_0}\sum_{j=1}^{n_0}W\left(\frac{X_{1,i}-X_{0,j}}{h_0}\right), \qquad W(x) = \int_{-1}^{x}K(t)dt$$

This method is usefull for evaluating the Partial Area Under Curve.

Partial area under ROC curve

AUC: the average performance over the entire range of possible sensitivities and specificities.

Problems:

- Two different curves can provide the same area
- Not all regions of the ROC curve have the equal clinical importance
- Clinical relevant sensitivities or specificities are often somewhere away from the ends of the ROC curve
- **PAUC – a partial area under curve**, i.e. area between two specificities or sensitivities

$$PAUC_I = \int_{p_1}^{p_2} R(p)dp, \ p_i = 1 - F_0(c_i), \ i = 1, 2$$

between two specificities

$$PAUC_{II} = \int_{\tilde{p}_1}^{\tilde{p}_2} R(p)dp, \ \tilde{p}_i = 1 - F_0(F_1^{-1}(1 - \tilde{q}_i)), \ i = 1, 2$$

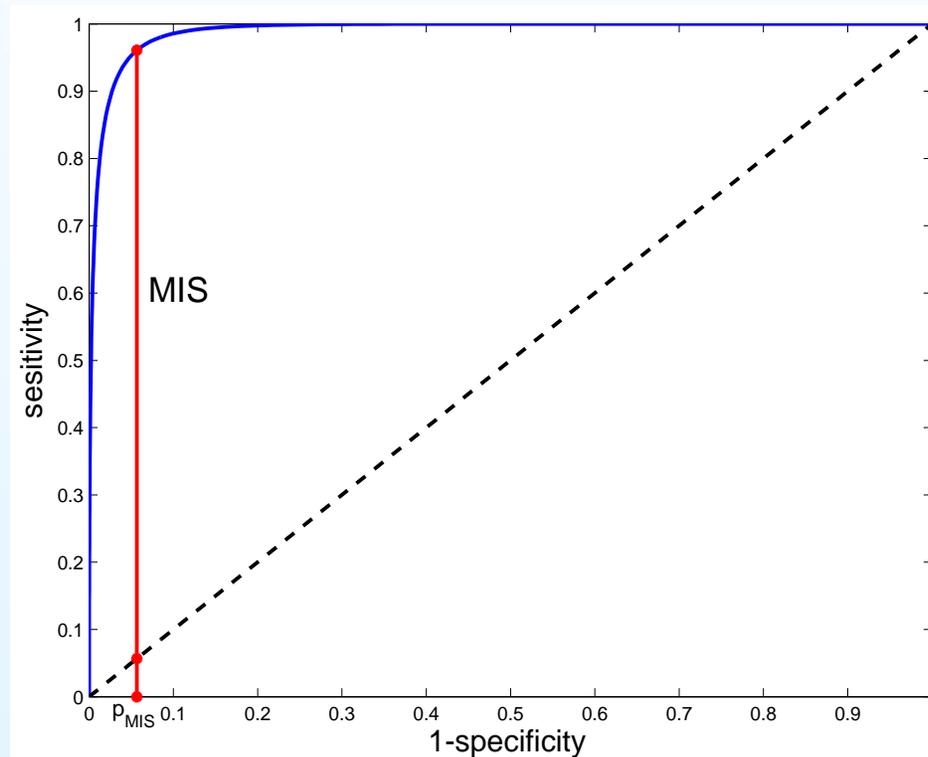between two sensitivities $\tilde{q}_1, \tilde{q}_2$

The choice of the appropriate ranges depends on clinical settings.

Maximum improvement of sensitivity over chance diagonal (MIS)

MIS – the maximum difference in observed sensitivity and sensitivity at chance diagonal over all values of specificity.

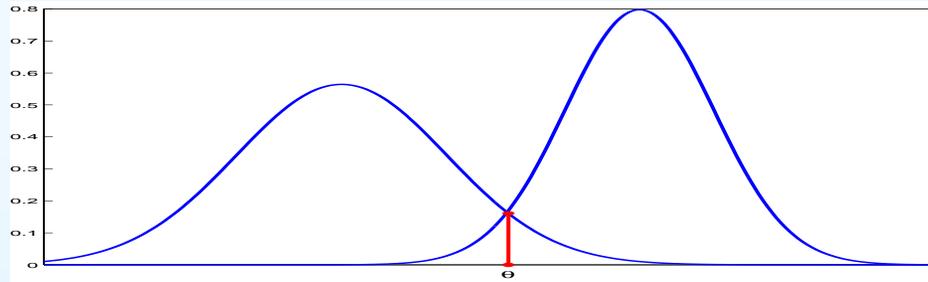The corresponding (1-specificity) is denoted by p_{MIS}

$$MIS = R(p_{MIS}) - p_{MIS}$$

A different point of view:

Assume $R(p)$ is concave. p_{MIS} is defined as argument of maximum of the function $Q(p) = R(p) - p$, i.e. zero of $Q'(p)$:

$$Q'(p) = R'(p) - 1, \quad R'(p) = \frac{f_1(c)}{f_0(c)}, \quad c = F_0^{-1}(1-p)$$

$$Q'(p) = 0: \quad \frac{f_1(\theta)}{f_0(\theta)} = 1 \Rightarrow f_1(\theta) = f_0(\theta)$$

$$\theta = F_0^{-1}(1 - p_{MIS}), \quad p_{MIS} = 1 - F_0^{-1}(\theta)$$

$R''(p) < 0 \Rightarrow p_{MIS}$ realizes the maximum of $Q(p) = R(p) - p$.

Explanation: p_{MIS} is such a point where a tangent to the ROC curve has a slope equal to 1.

Application for real data

Leukaemia data

(Data provided by Faculty Hospital Brno)

Fusion gene (FG) is the most common chromosomal aberration in acute leukaemias. Detectable FG at the end of induction therapy predict relapse with a high probability. However, detection of it with sensitivity of at least one malignant cell among 10 000 normal cells is not successful in all patients.

Wilms Tumour Gene (WT1) is a tumour suppressor gene, expressed in malignant and normal hematopoietic progenitor cells. Because WT1 has been shown to be expressed in the vast majority of patients with acute leukaemias, the relevance of WT1 mRNA expression regarding prognosis and possible prediction of relapse was investigated.

The WT1 expression and FG occurrence was followed in CD34+ peripheral blood progenitor cells collected from 59 leukemic patients in the first remission. 29 patients were in group $\mathcal{G}_0$ (without FG) and 30 in group $\mathcal{G}_1$.

The question: Does higher expression of WT1 indicate FG occurrence?

Leukaemia data
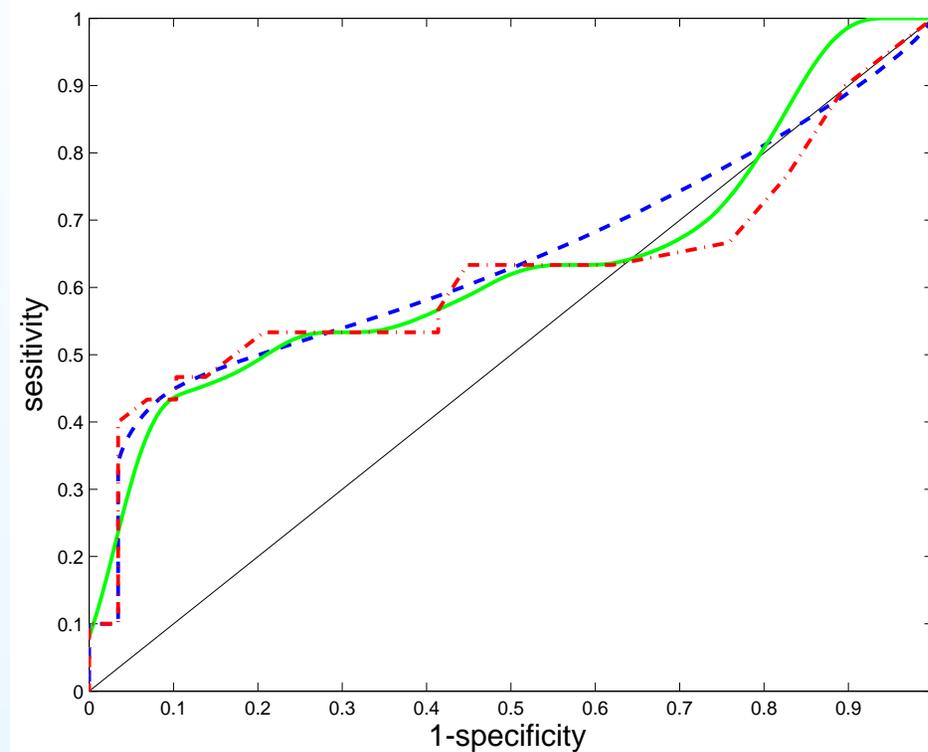
Colour marking, bandwidths and AUC:

$-\cdot-$ empirical ROC curve

$---$ Kernel formula I $h_0 = 3.9740,\ h_1 = 4.5904$ $AUC_I = 0.6397$

——   Kernel formula II $h_0 = 3.9740,\ \tilde{h}_1 = 0.0728$ $AUC_{III} = 0.6314$

Head trauma data

(Source of the data: see Zhou *et al.* 2002)

The bi-normal model and the kernel method were used for processing of the second real data set. We consider the use of CK–BB isoenzyme measured within 24 hours of injury for predicting the outcome of severe head trauma.

We are interested in determining which patients have a poor outcome after suffering a severe head trauma.

60 patients: 19 had moderate to full recovery and 41 eventually had poor or no recovery.

We use the ROC curve to assess the discrimination between patients with and without a poor outcome.

Question: Is CK–BB isoenzyme a good predictor of the outcome?

The data don't satisfy the conditions of normality and the bi-normal model gives worse results in this case. For improvement of them some transformation of the data (logarithmic, Box-Cox) should be used.

Head trauma data
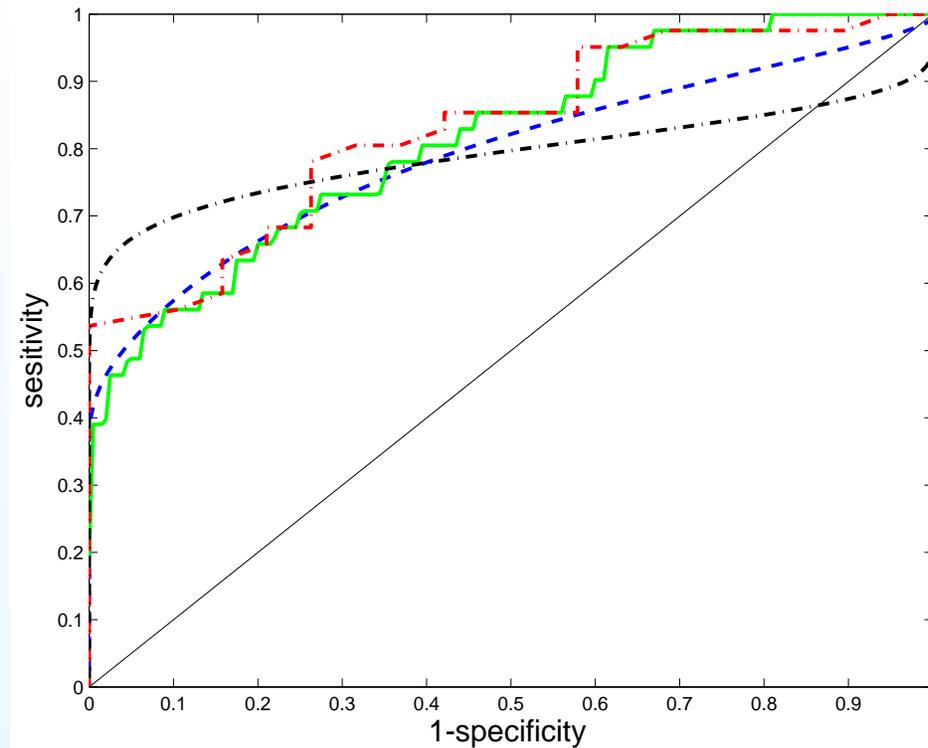
Colour marking, bandwidths and AUC:

$- \cdot -$ empirical ROC curve

$- - -$ Kernel formula I $h_0 = 175.5347,\ h_1 = 300.7402$ $AUC_I = 0.7896$

——   Kernel formula II $h_0 = 175.5347,\ \tilde{h}_1 = 0.0013$ $AUC_{III} = 0.8148$

$- \cdot -$ Binormal model

Pancreatic cancer data

(Source of the data: see Zhou and Hazerlak 2002)

The kernel methods were applied to real data set from Mayo Clinic, where sera from group of 51 'control' patients with pancreatitis and 90 'case' patients with pancreatic cancer were studied with a carbohydrate antigen assay (CA19-9).

We study a relative accuracy of biomarker CA19-9 for 90 patients with condition and 51 patients without condition.

Pancreatic cancer data
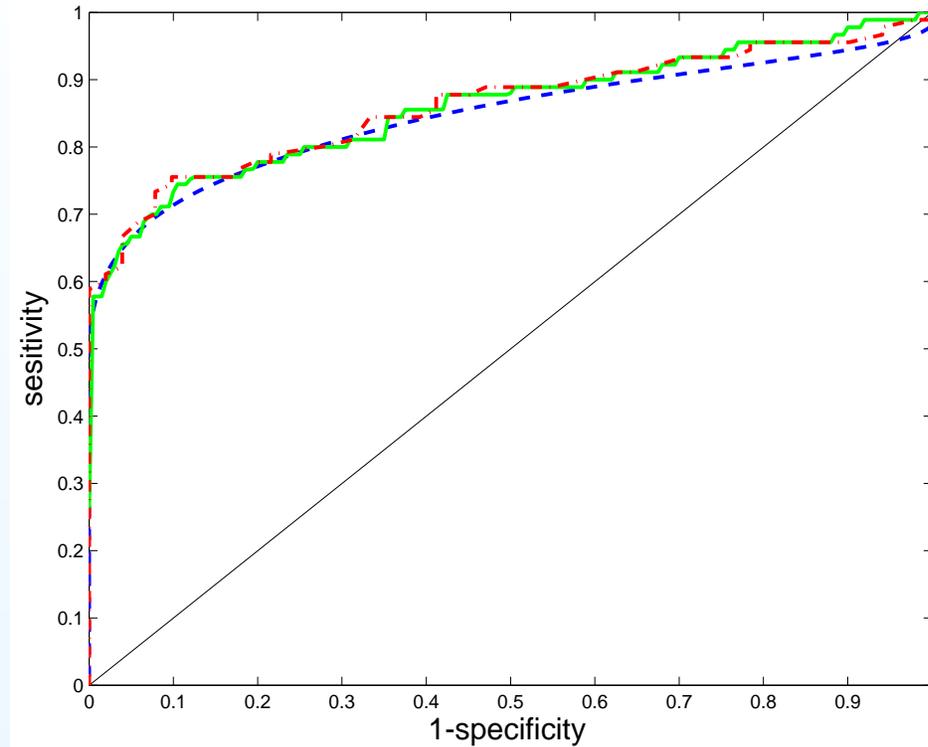
Colour marking, bandwidths and AUC:

$- \cdot -$ empirical ROC curve

$- - -$ Kernel formula I $h_0 = 1.0176,\ h_1 = 2.6694$ $AUC_I = 0.8460$

——   Kernel formula II $h_0 = 1.0176,\ \tilde{h}_1 = 0.000908$ $AUC_{III} = 0.8593$

Salmon data

(Source of the data: see Johnson and Wichern 1992)

The salmons have a remarkable life cycle. They are born in freshwater streams and after a year or two swim into the ocean. After a couple of years in saltwater they return to their, place of birth to spawn and die. At the time they are about to return as mature fish, they are harvested while still in the ocean.

To help regulate catches samples of fish taken during the harvest must be identified, as coming from Alaskan or Canadian waters. The fish carry some information about their birth place in the growth rings on their scales. Typically, the rings associated with freshwater growth are smaller for the Alaskan-born than for the Canadian-born salmon.

X_0: diameter of rings for the first-year freshwater growth for the Alaskan-born salmons (hundredths of an inch)

X_1: diameter of rings for the first-year freshwater growth for the Canadian-born salmons (hundredths of an inch)

Samples of sizes $n_0 = n_1 = 50$

Question: Is the diameter of rings for the first-year freshwater growth suitable indicator of the origin of the salmon?

Salmon data
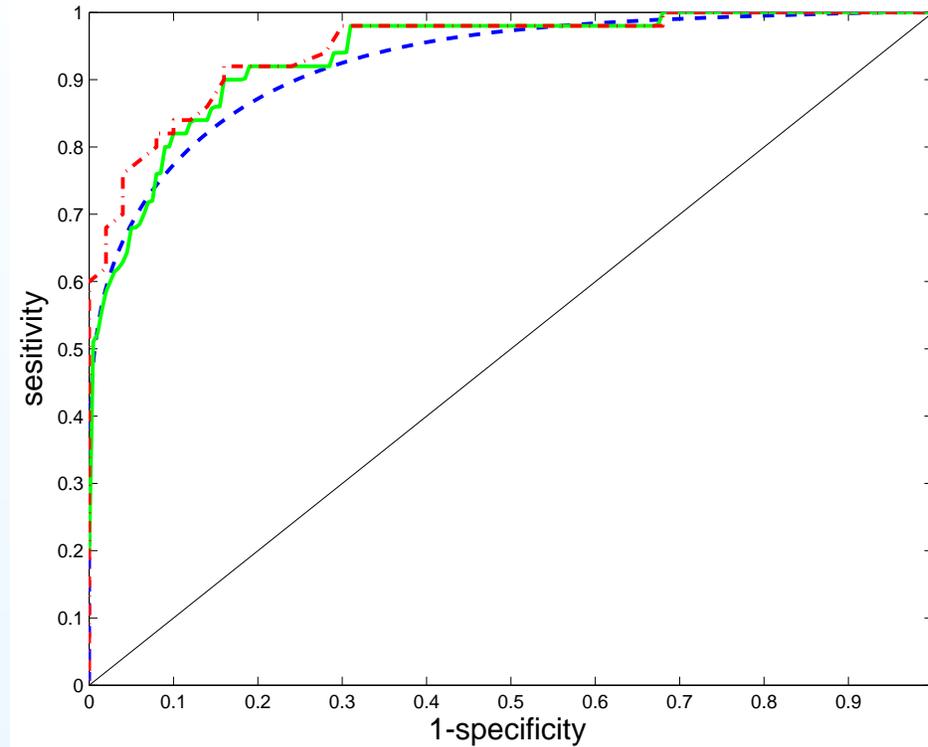
Colour marking, bandwidths and AUC:

$- \cdot -$ empirical ROC curve

$- - -$ Kernel formula I $h_0 = 21.5009,\, h_1 = 24.2317$ $AUC_I = 0.9253$

——   Kernel formula II $h_0 = 21.5009,\, \tilde{h}_1 = 0.0026$ $AUC_{III} = 0.9371$

Hurricanes Data

(Source of the data:

`http://sunsite.univie.ac.at/textbooks/statistics/stclatre.html`)

Suppose you have records of the Longitude and Latitude coordinates at which 37 storms reached hurricane strength for two classifications of hurricanes - Baro hurricanes and Trop hurricanes. The fictitious data were presented for illustrative purposes by Elsner, Lehmiller, and Kimberlain (1996), who investigated the differences between baroclinic and tropical North Atlantic hurricanes.

The Longitude coordinates were taken as the response variable X for the first ROC curve and the Latitude coordinates for the second one.

X_0: Longitude (Latitude) coordinates for Trop hurricanes

X_1: Longitude (Latitude) coordinates for Baro hurricanes

Question: Are Longitude or Latitude coordinates usable for classification of the hurricanes?

Hurricanes Data – Longitude coordinates
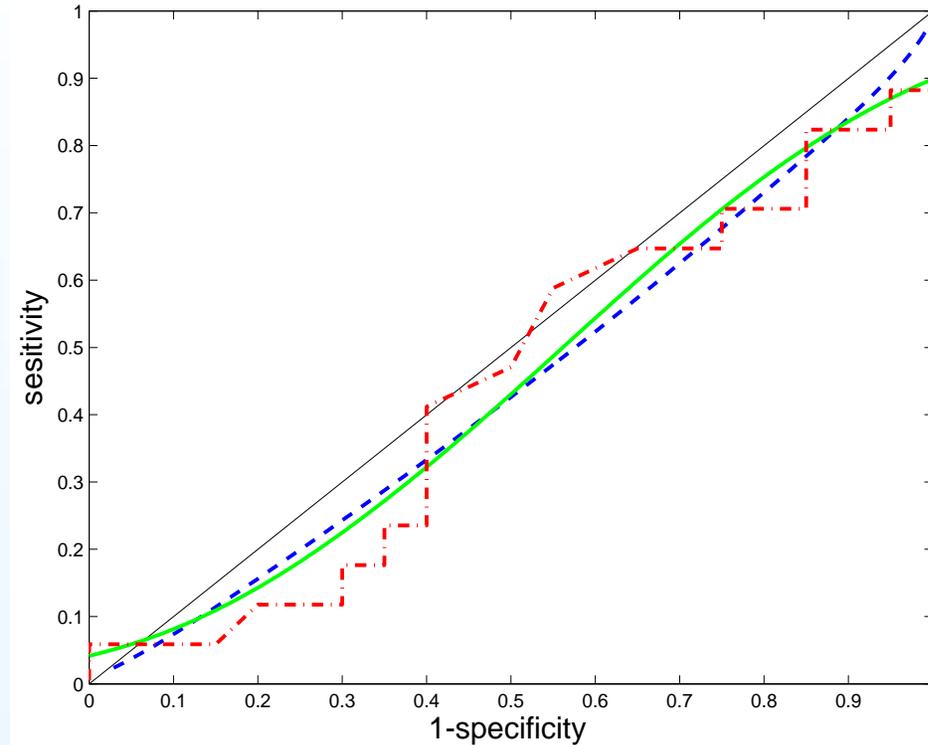
Colour marking, bandwidths and AUC:

$- \cdot -$ empirical ROC curve

$- - -$ Kernel formula I $h_0 = 6.8773, h_1 = 7.5679$ $AUC_I = 0.4440$

———   Kernel formula II $h_0 = 6.8773, \tilde{h}_1 = 0.5611$ $AUC_{III} = 0.4456$

Hurricanes Data – Latitude coordinates
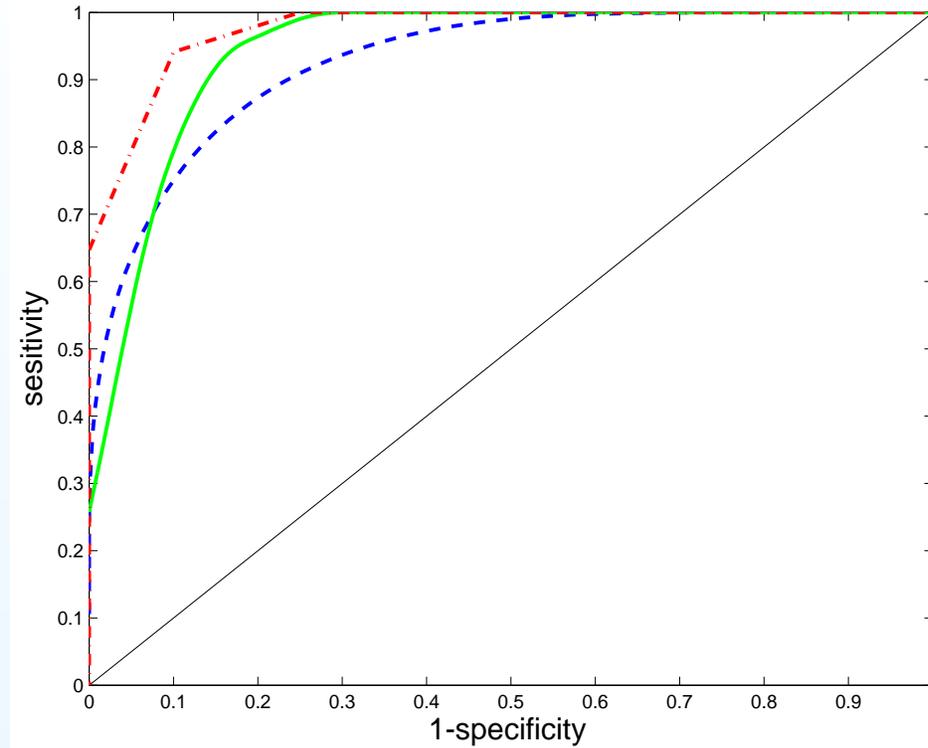
Colour marking, bandwidths and AUC:

$-\cdot-$ empirical ROC curve

$---$ Kernel formula I $h_0 = 3.3794,\ h_1 = 3.5817$ $AUC_I = 0.9258$

$\rule{1.5em}{0.6pt}$ Kernel formula II $h_0 = 3.3794,\ \tilde{h}_1 = 0.0983$ $AUC_{III} = 0.9441$

Conclusion

The ROC curves have found useful application in diagnostic medicine.

Ongoing development in ROC analysis will address more complex types of diagnostic situations and will likely expand the applicability of ROC analysis.

References

[1] Altman, N., Léger, Ch.: *Bandwidth selection for kernel distribution function estimation*, Journal of Stat. Planning and Inference, **46**, pp. 195–214, 1995.

[2] Azzalini, A.: *A note on the estimation of a distribution function and quantiles by a kernel method*. Biometrika, **68**, No. 1, pp. 326–328, 1981.

[3] Bowman, A., Hall, P., Prvan, T.: *Bandwidth selection for the smoothing of distribution functions*. Biometrika, **85**, No. 4, pp. 799–808, 1998.

[4] Bradley A.P.: *The Use of Area Under the ROC Curve in the Ecaluation of Machine Learning Algorithms*, Patern. Recognition, **30**, No. 7, pp. 1145–1159, 1997.

[5] DeLong E. R., DeLong D. M., Clarke-Pearson D. L.: *Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach*, Biometrics, **44**, pp. 837–845, 1988.

References

[6] Elsner, J. B., Lehmiller, G. S. , Kimberlain, T. B.: *Objective classification of Atlantic basin hurricanes*, Journal of Climate, **9**, pp. 2880–2889, 1996.

[7] Eng J.: *Receiver Operating Characteristic Analysis: A Prime*, Academic Radiology, **12**, pp. 909–916, 1999.

[8] Faraggi D., Reiser B.: *Estimation of the area under the ROC curve*, Statistics in Medicine, **21**, pp. 3093–3106, 2002.

[9] Forbelská, M.: *Elliptically Contoured Models in Classification*, Proceedings of Summer School Matlab'06, Masaryk Univerisy, pp. 77–94, 2007.

[10] Forbelská, M.: *Elliptically Contoured Models in ROC Analysis*, S.Co.2007, Fifth Conference, Mantovan P., Pastore A., Tonellato S., (Eds.), Book of Short Papers„ pp. 243–248, 2007.

References

[11] Gu, J., Ghosal, S., Roy A.: *Non-parametric estimation of ROC curve*,
`http://www4.stat.ncsu.edu/ sghosal/papers/ROCBB.pdf`

[12] Hall, P.H., Hyndman R.J.: *Improved methods for bandwidth selection when estimating ROC curves*. Statistics & Probability Letters, **64**, pp 181–189, 2003.

[13] Hanley J. A., McNeil B. J.: *The meaning and use of the area under the receiver operating characteristic (roc) curve*, Radiology, **143**, pp. 29–36, 1982.

[14] Hanley J. A., McNeil B. J.: *A method of comparing the areas under receiver operating characteristic curves derived from the same cases*, Radiology, **148**, pp. 839–843, 1983.

[15] Horová I., Zelinka J.: *Contribution to the bandwidth choice for kernel density estimates*, Computational Statistics, **22**, No. 1, pp. 31–47, 2007.

References

[16] Horová I. Koláček J., Zelinka, J., El-Shaarawi A.: *Smooth Estimates of Distribution Functions with Application in Environmental Studies*, Accepted for MABE'08, 2007.

[17] Jiang Y., Metz C. E., Nishikawa R. M.: *A. receiver. operating characteristic partial area index for highly. sensitive diagnostic tests*, Radiology, **201**, pp. 7745–750, 1996.

[18] Johnson, R. A., Wichern D.W.: Applied multivariate statistical analysis, Prentice-Hall, Inc., 1992.

[19] Lloyd, C.J.: *Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems*. JASA, 93, No. 444, pp. 1356–1364, 1998.

[20] Lloyd, C.J.: *Estimation of a convex ROC curve*, Statistics & probability letters, **59**, pp. 99–111, 2002.

References

[21] Lloyd, C.J., Zhou Yong: *Kernel estimators of the ROC curve are better than empirical*. Statistics and Prob. Letters 44, pp. 221–228, 1999.

[22] Lusted L. B.: *Signal Detectability and Medical Decision-Making*, Science, **171**, pp. 1217-1219, 1971.

[23] O'Malley A.J.1, Zou K.H., Fielding J.R., Tempany C.M.C.: *Bayesian Regression Methodology for Estimating a Receiver Operating Characteristic Curve with Two Radiologic Applications*, Academic Radiology, **8**, No. 8, pp. 713-725, 2001.

[24] Metz C. E., Herman B. A., Shen J-H.: *Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data*, Statistics in Medicine, **17**, pp. 1033–1053, 1998.

[25] Obuchowski N. A.: *Nonparametric Analysis of Clustered ROC Curve Data*, Biometrics, **53**, No. 2, pp. 567–578, 1997.

References

[26] Peng L., Zhou X-H.: *Local linear smoothing of receiver operating characteristic (ROC) curves*, **118**, pp. 129–143, 2004.

[27] Pepe, M. S.: *An Interpretation for the ROC Curve and Inference Using GLM Procedures*, Biometrics, **56**, No. 2, pp. 352–359, 2000.

[28] Pepe, M. S.: The Statistical Evaluation of Medical Tests for Classification and Prediction, Oxford Statistical Science Series, Oxford University Press, 2003.

[29] Pepe, M. S.: *Receiver Operating Characteristic Methodology*, JASA, **95**, No. 449, 2000.

[30] Qin J., Zhang B.: *A goodness-of-fit test for logistic regression models based on case-control data*, Biometrika, **84**, No. 3, pp. 609–618, 1997.

[31] Terrell, G. R.: *The maximal smoothing principle in density estimation*. Journal of the American Statistical Association. **85**, No. 410, pp. 440-447, 1990.

References

[32] Terrel, G. R., Scott, D. W.: *Oversmoothed nonparametric density estimates*. Journal of the American Statistical Association, **80**, 209-214, 1985.

[33] Venkatraman E. S., Begg C. B.: *A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment*, Biometrika, **83**, No. 4, pp. 835–848, 1996.

[34] Wan S., Zhang B.: *Smooth semiparametric receiver operating characteristic curves for continuous diagnostic tests*, Statistics in Medicine, **26**, pp. 2565–2586, 2007.

[35] Wand, I.P., Jones, I.C.: Kernel smoothing. Chapman & Hall, London, 1995.

[36] Wieand S., Gail M. H., James B. R., James K. L.: *A Family of Nonparametric Statistics for Comparing Diagnostic Markers with Paired or Unpaired Data*, Biometrika, **76**, No. 3, pp. 585–592, 1989.

References

[37] Zhang D. D., Zhou X-H., Freeman D. H. Jr., Freeman J. L.: *A non-parametric method for the comparison of partial areas under ROC curves and its application to large health care data sets*, Statistics in Medicine, **21**, pp. 701–715, 2002.

[38] Zhou X.–H., Harezlak J.: *Comparison of bandwidth selection methods for kernel smoothing of ROC curves*, Statistics in Medicine, **21**, 2045–2055, 2002.

[39] Zhou X.–H.,Obuchowski N. A., McClish D. K.: Statistical Methods in Diagnostic Medicine, Wiley–Interscience, 2002.

[40] Zou K. H., Hall W. J., Shapiro D. E.: *Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests*, Statistics in Medicine, **16**, 2143–2156, 1997.