SD 203
Linear Model

François Portier, Joseph Salmon
http://josephsalmon.eu
Télécom Paristech, Institut Mines-Télécom

Outline

Outline

General principle

Context

- We observe $X_1, \ldots, X_n$ from a common distribution $\mathcal{P}$
- We are interested in $\theta \in \Theta$, a parameter of $\mathcal{P}$

Goal

To decide whether an assumption on θ is likely (or not)

$$\mathcal{H}_0 = \{\theta \in \Theta_0\}$$

against some alternative

$$\mathcal{H}_1 = \{\theta \in \Theta_1\}$$

Call $\mathcal{H}_0$ the null hypothesis, $\mathcal{H}_1$: the alternative

General principle

Means

Determine a test statistics $T(X_1, \ldots, X_n)$ and a region R such that if

$$T(X_1, \ldots, X_n) \in R \;\Rightarrow\; \text{we reject } \mathcal{H}_0$$

In words : The observed data discriminates between H_0 and H_1

Hypothesis testing for “ heads or tails”

When flipping a coin the model is a Bernoulli distribution with parameter p, $\mathcal{B}(p)$.

Is the coin fair ?

$$\mathcal{H}_0 = \{p = 0.5\} \quad \text{against} \quad \mathcal{H}_1 = \{p \neq 0.5\}$$

Is the coin possibly unfair ?

$$\mathcal{H}_0 = \{0.45 \leqslant p \leqslant 0.55\} \quad \text{against} \quad \mathcal{H}_1 = \{p \notin [0.45, 0.55]\}$$

Do we reject or do we accept ?

In most practical situations, $\mathcal{H}_0$ is simple, i.e.,

$$\Theta_0 = \{\theta_0\}$$

and $\Theta_1 = \Theta \backslash \Theta_0$ is large

($\mathcal{H}_0$ is often an hypothesis on which we care particularly, e.g., something acknowledged to be true, easy to formulate)

We only reject $\mathcal{H}_0$

If $\mathcal{H}_0$ is not rejected we cannot conclude $\mathcal{H}_0$ is true because $\mathcal{H}_1$ is too general

e.g., $\{p \in [0, 0.5[\cup]0.5, 1]\}$ can not be rejected !

2 types of error

	$\mathcal{H}_0$	$\mathcal{H}_1$
$\mathcal{H}_0$ is not rejected	Correct	Wrong (False negative)
$\mathcal{H}_0$ is rejected	Wrong (False positive)	Correct

- Type I : probability of a wrong reject

$$\mathbb{P}(T(X_1, \ldots, X_n) \in R \mid \mathcal{H}_0)$$

- Type II : probability of wrong non-reject

$$\mathbb{P}(T(X_1, \ldots, X_n) \notin R \mid \mathcal{H}_1)$$

Significance level and power

Significance level α if

$$\limsup_{n\to+\infty} \mathbb{P}(T(X_1,\dots,X_n) \in R \mid \mathcal{H}_0) \leqslant \alpha$$

(We speak of 95%-test when α is 0.05%)

Consistency

A test statistics (given by $T(X_1,\dots,X_n)$ and a region R) is said to be α-consistent if the significant level is α and if the power goes to one, i.e.,

$$\limsup_{n\to+\infty} \mathbb{P}(T(X_1,\dots,X_n) \in R \mid \mathcal{H}_0) \leqslant \alpha$$

$$\lim_{n\to\infty} \mathbb{P}(T(X_1,\dots,X_n) \in R \mid \mathcal{H}_1) = 1$$

Test statistic and reject region

Goal : to build a α-consistent test

(1) Define the test statistic $T(X_1, \ldots, X_n)$ and the level α you wish

(2) Do some maths to determine a reject region R that achieve a significance level α

(3) Prove the consistency

(4) Rule decision : reject whenever $\boxed{T_n(X_1, \ldots, X_n) \in R}$

Famous tests

- Test of the equality of the mean for 1 sample
- Test of the equality of the means between 2 samples
- Chi-square test for the variance
- Chi-square test of independence
- Regression coefficient non-effects test

Examples : " heads or tails"

- Model : $\Theta = [0,1]$, $\mathbb{P}_\theta = \mathcal{B}(\theta)$
- Observe $(X_1, \ldots, X_n)$ i.i.d. from this model
- Null hypothesis $\mathcal{H}_0 : \{\theta = 0.5\}$
- Define $T_n(X_1, \ldots, X_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - 0.5)$
- Critical region for T_n ? Gaussian quantile : Show that

$$\lim_{n\to\infty} \mathbb{P}(T_n \in [-1.96, 1.96] \mid \mathcal{H}_0) \to 0.95$$

- Take $R =]-\infty, -1.96[\cup]1.96, +\infty[$

Exo :

Specify the procedure for an arbitrary significance level α

Example : Gaussian mean

- Model : $\Theta = \mathbb{R}$, $\mathbb{P}_\theta = \mathcal{N}(\theta, 1)$
- Observe $(X_1, \dots, X_n)$ i.i.d. from this model
- Null hypothesis : $\mathcal{H}_0 : \{\theta = 0\}$
- Under $\mathcal{H}_0$, $\quad T_n(X_1, \dots, X_n) = \frac{1}{\sqrt{n}} \sum_i X_i \sim \mathcal{N}(0, 1)$
- Critical region for T_n ? Gaussian quantile :

$$\mathbb{P}(T_n \in [-1.96, 1.96] \mid \mathcal{H}_0) = 0.95$$

- Take $R =]-\infty, -1.96[\cup]1.96, +\infty[$.
- **Numerical example** : If $T_n = 1.5$, we do **not** reject $\mathcal{H}_0$ at level 95%

Outline

Test of no-effect : Gaussian case

Gaussian Model

$$y_i = \theta_0^\star + \sum_{k=1}^{p} \theta_k^\star x_{i,k} + \varepsilon_i$$

$$x_i^\top = (1, x_{i,1}, \ldots, x_{i,p}) \in \mathbb{R}^{p+1} \text{ (deterministic)}$$

$$\varepsilon_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2), \text{ for } i = 1, \ldots, n$$

Theorem

Let $X = (x_1, \ldots, x_n)^\top \in \mathbb{R}^{n \times (p+1)}$ of full rank, and $\hat{\sigma}^2 = \|\mathbf{y} - X\hat{\boldsymbol{\theta}}\|_2^2/(n-(p+1))$, then

$$\hat{T}_j = \frac{\hat{\theta}_j - \theta_j^*}{\hat{\sigma}\sqrt{(X^\top X)^{-1}_{j,j}}} \sim \mathcal{T}_{n-(p+1)}$$

where $\mathcal{T}_{n-p}$ est une loi dite de Student (de degré $n-(p+1)$)

Test of no-effect : Gaussian case

Null hypothesis

Aim is to test

$$\mathcal{H}_0 : \theta_j^* = 0$$

equivalently, $\Theta_0 = \{\theta \in \mathbb{R}^p : \theta_j = 0\}$

Under $\mathcal{H}_0$, we know the value of $\hat{T}_j$:

$$T_j := \frac{\hat{\theta}_j}{\hat{\sigma}\sqrt{(X^\top X)^{-1}_{j,j}}} \sim \mathcal{T}_{n-(p+1)}$$

Choosing $R = [-t_{1-\alpha/2}, t_{1-\alpha/2}]^c$ with $t_{1-\alpha/2}$ the $1-\alpha/2$-quantile of $\mathcal{T}_{n-(p+1)}$), we decide to reject $\mathcal{H}_0$ whenever

$$|\hat{T}_j| > t_{1-\alpha/2}$$

Test of no-effect : Random-design case

Random design Model

$$
\begin{aligned}
y_i &= \theta_0^\star + \sum_{k=1}^{p} \theta_k^\star \mathbf{x}_{i,k} + \varepsilon_i \\
\mathbf{x}_i^\top &= (1, \mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,p}) \in \mathbb{R}^{p+1} \\
(\varepsilon_i, \mathbf{x}_i) &\overset{i.i.d}{\sim} (\varepsilon, \mathbf{x}), \text{ for } i = 1, \dots, n \\
\mathbb{E}(\varepsilon|\mathbf{x}) &= 0, \ \mathrm{Var}(\epsilon|\mathbf{x}) = \sigma^2
\end{aligned}
$$

Theorem

If $\mathrm{var}(\mathbf{x})$ has full rank, then

$$
\hat{T}_j = \frac{\hat{\theta}_j - \theta_j^*}{\hat{\sigma}\sqrt{(X^\top X)^{-1}_{j,j}}} \xrightarrow{\mathsf{d}} \mathcal{N}(0,1)
$$

Test of no-effect : Random design case

Null hypothesis

Aim is to test

$$\mathcal{H}_0 : \theta_j^* = 0$$

equivalently, $\Theta_0 = \{\theta \in \mathbb{R}^p : \theta_j = 0\}$

Under $\mathcal{H}_0$, we know the value of $\hat{T}_j$:

$$T_j := \frac{\hat{\theta}_j}{\hat{\sigma}\sqrt{(X^\top X)^{-1}_{j,j}}} \xrightarrow{\text{d}} \mathcal{N}(0,1)$$

Choosing $R = [-z_{1-\alpha/2}, z_{1-\alpha/2}]^c$ with $z_{1-\alpha/2}$ the $1-\alpha/2$-quantile of $\mathcal{N}(0,1)$), we decide to reject $\mathcal{H}_0$ whenever

$$|\hat{T}_j| > z_{1-\alpha/2}$$

Link between IC and test

Rappel (modèle gaussien) :

$$IC_\alpha := \left[\hat{\theta}_j - t_{1-\alpha/2}\hat{\sigma}\sqrt{(X^\top X)^{-1}_{j,j}}, \hat{\theta}_j + t_{1-\alpha/2}\hat{\sigma}\sqrt{(X^\top X)^{-1}_{j,j}}\right]$$

est un IC de niveau α pour θ^*_j. Dire que "$0 \in IC_\alpha$" signifie que

$$|\hat{\theta}_j| \leqslant t_{1-\alpha/2}\hat{\sigma}\sqrt{(X^\top X)^{-1}_{j,j}} \quad \Leftrightarrow \quad \frac{|\hat{\theta}_j|}{\hat{\sigma}\sqrt{(X^\top X)^{-1}_{j,j}}} \leqslant t_{1-\alpha/2}$$

Cela est donc équivalent à accepter l'hypothèse $\theta^*_j = 0$ au niveau α. Le α le plus petit telle que $0 \in IC_\alpha$ est appelé la p-**value**.

Rem: On sait que si l'on prend α très proche de zéro un IC_α va recouvrir l'espace entier, on peut donc trouver (par continuité) un α qui assure l'égalité dans les équations ci-dessus.

Outline

Contexte médical

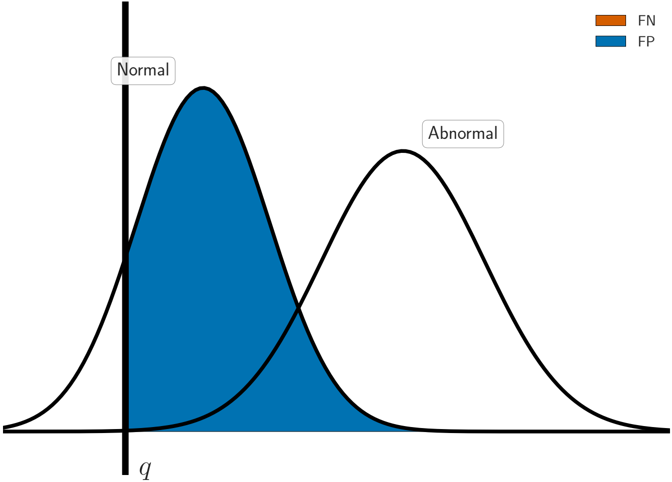
- Un groupe de patients $i = 1, \dots, n$ est suivi pour un dépistage.
- Pour chaque individu, le test se base sur une variable aléatoire $X_i \in \mathbb{R}$ et un seuil $q \in \mathbb{R}$

dès lors que	$X_i > q$	le test est **positif**
sinon		le test est **négatif**

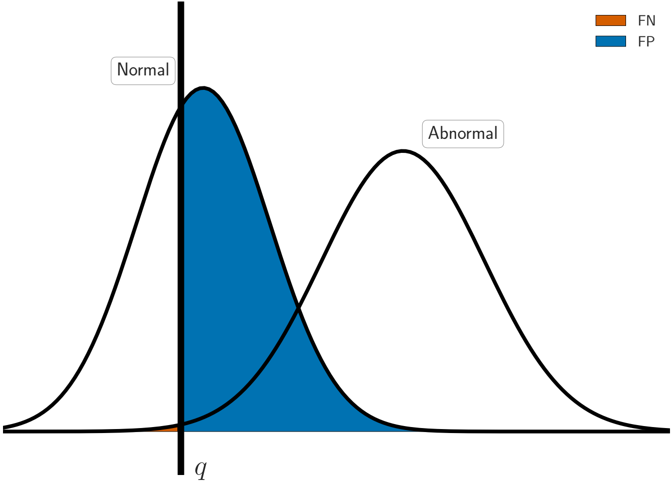
Ensemble des configurations possibles

	Normal H_0	Atteint H_1
négatif	vrai négatif	faux négatif
positif	faux positif	vrai positif

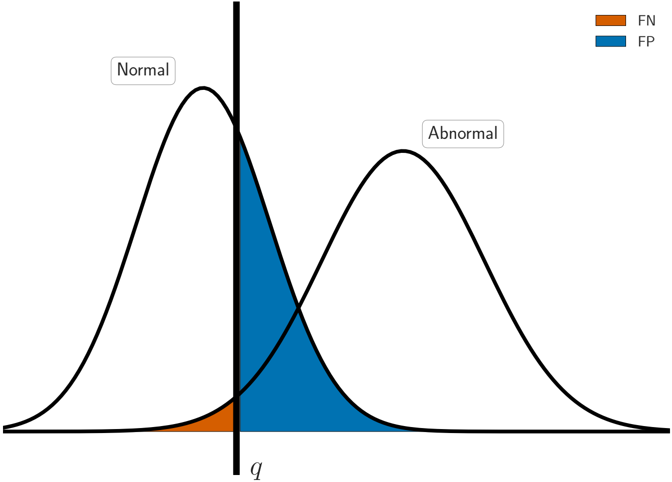
Faux positif vs faux négatif



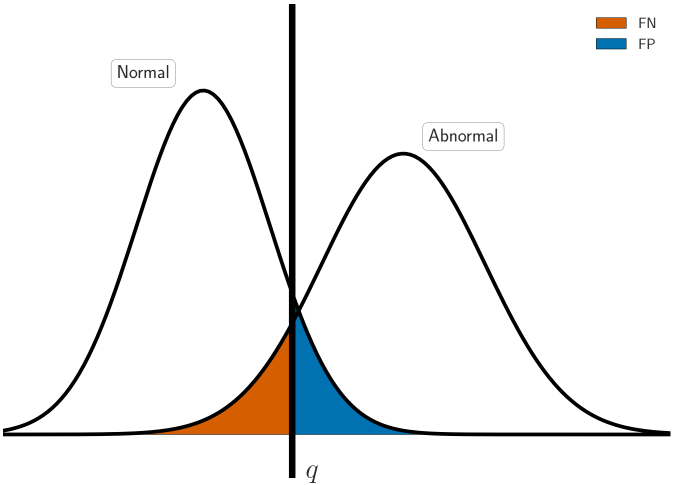
Faux positif vs faux négatif



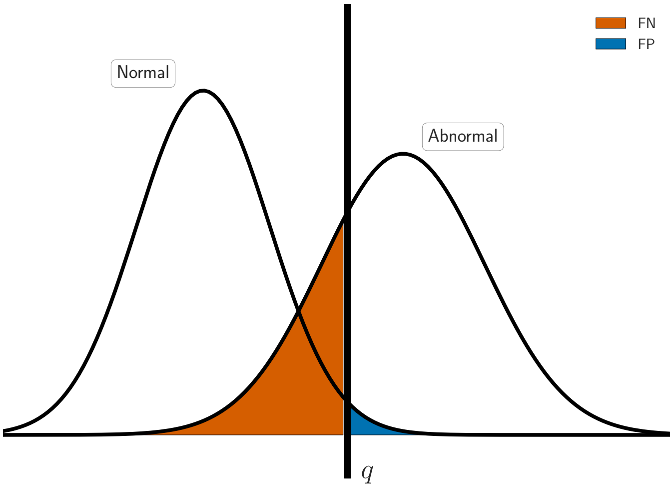
Faux positif vs faux négatif



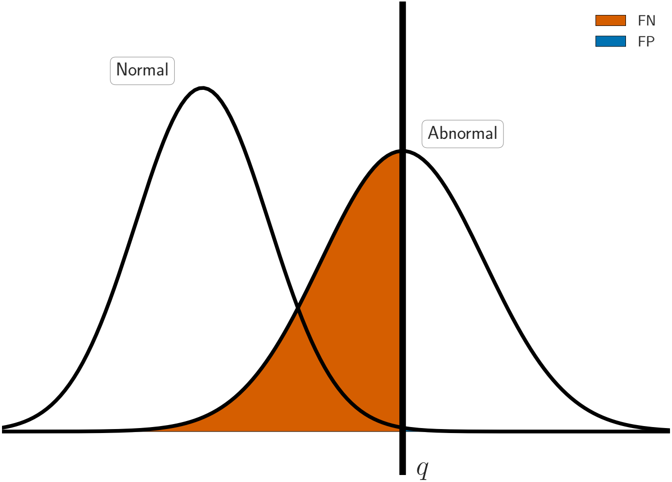
Faux positif vs faux négatif



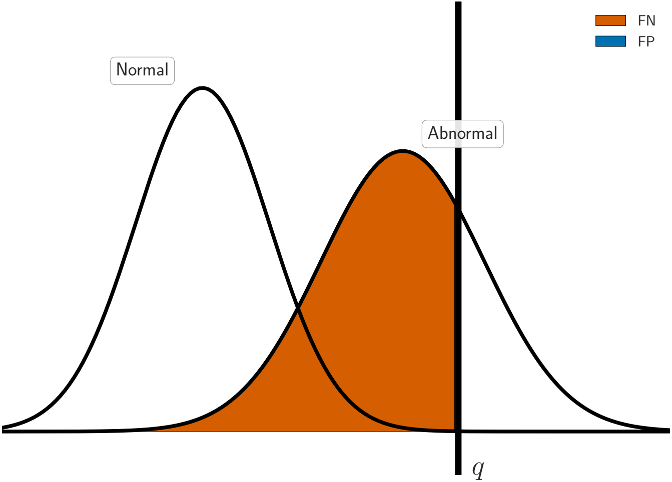
Faux positif vs faux négatif



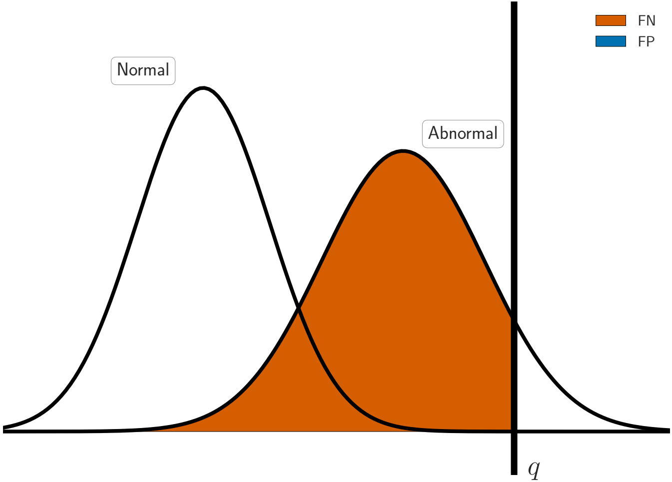
Faux positif vs faux négatif



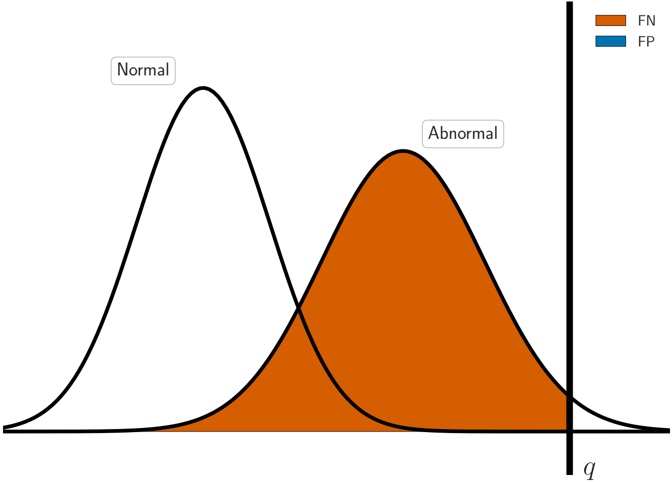
Faux positif vs faux négatif



Faux positif vs faux négatif



Faux positif vs faux négatif



Sensibilité - Spécificité

- On suppose que les individus normaux ont la même fonction de répartition F
- On suppose que les individus atteints ont la même fonction de répartition G

Définition

- Sensibilité : $\mathrm{Se}(q) = 1 - G(q)$ (1$-$ risque de 2^{nde} espèce)
- Spécificité : $\mathrm{Sp}(q) = F(q)$ (1$-$ risque de 1^{re} espèce)

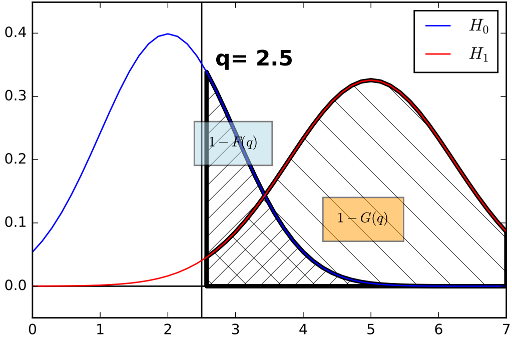
Courbe ROC

Définition

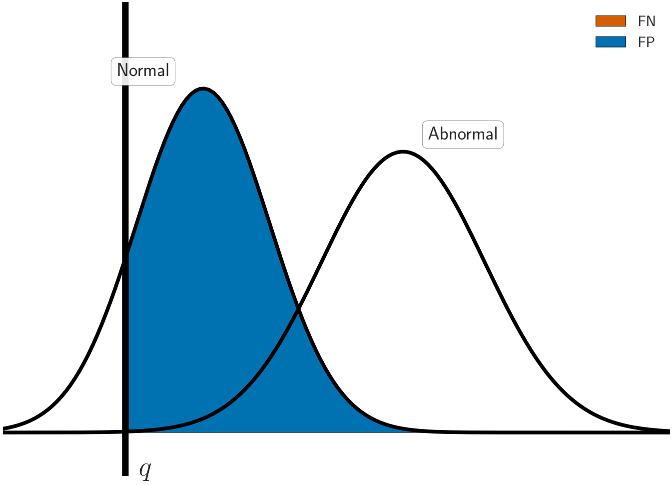
La courbe ROC est la courbe décrit par $(1 - \mathrm{Sp}(q), \mathrm{Se}(q))$, quand $q \in \mathbb{R}$. C'est donc la fonction $[0,1] \to [0,1]$

$$\mathrm{ROC}(t) = 1 - G(F^{-}(1-t))$$

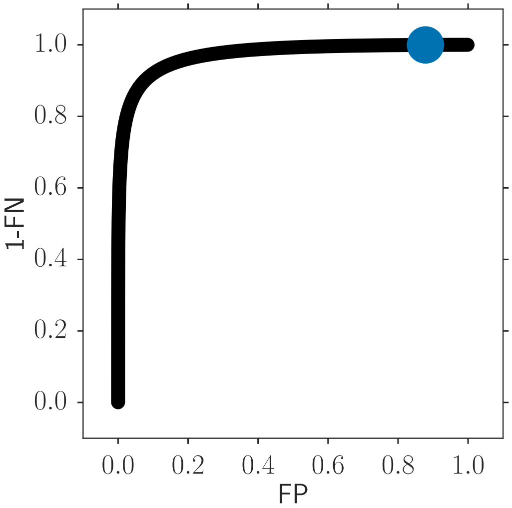
où $F^{-}(1-t) = \inf\{x \in \mathbb{R} \; : \; F(x) \geqslant 1 - t\}$.



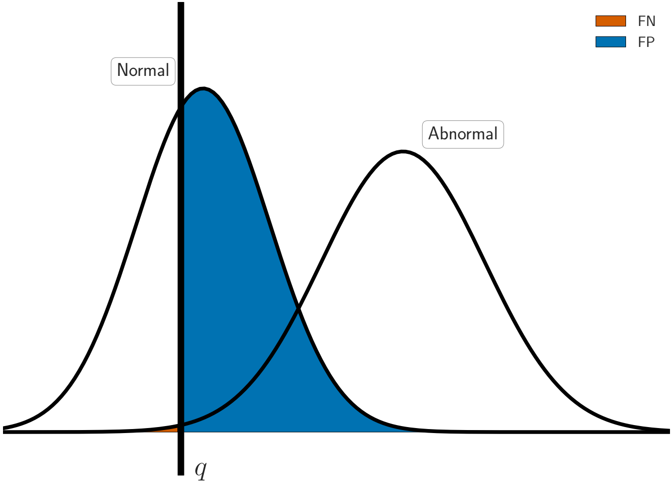
Courbe ROC



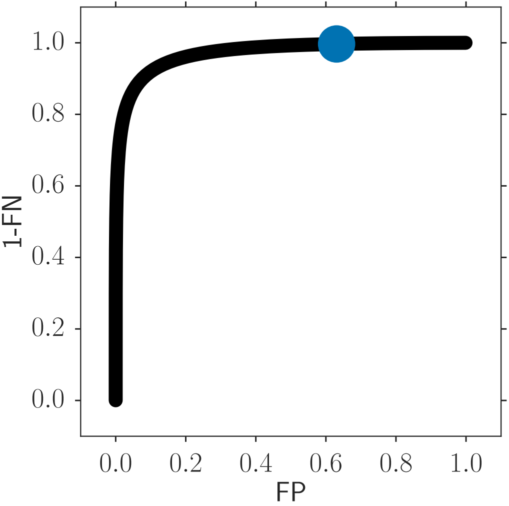
Courbe ROC



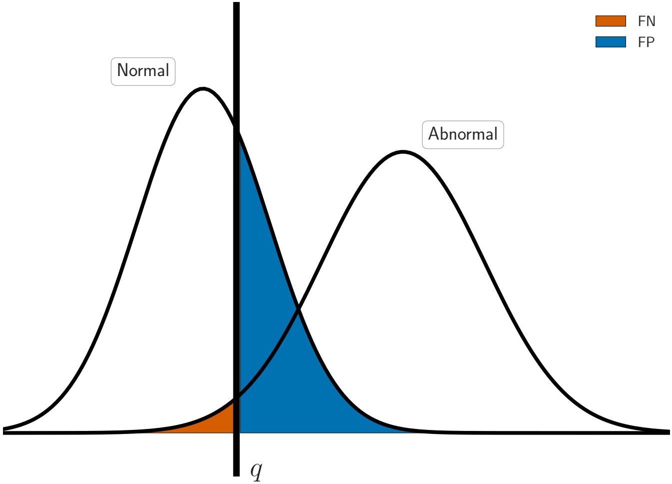
Courbe ROC



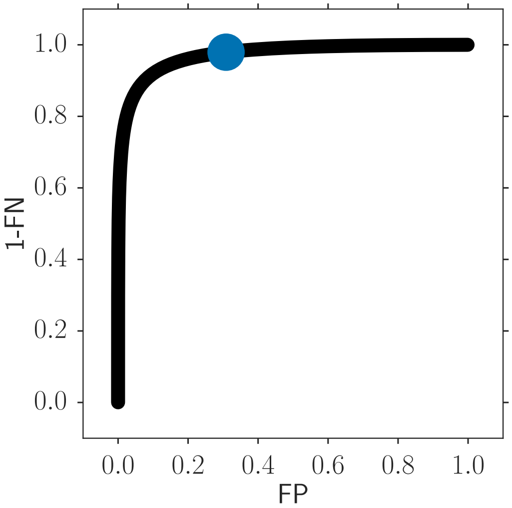
Courbe ROC



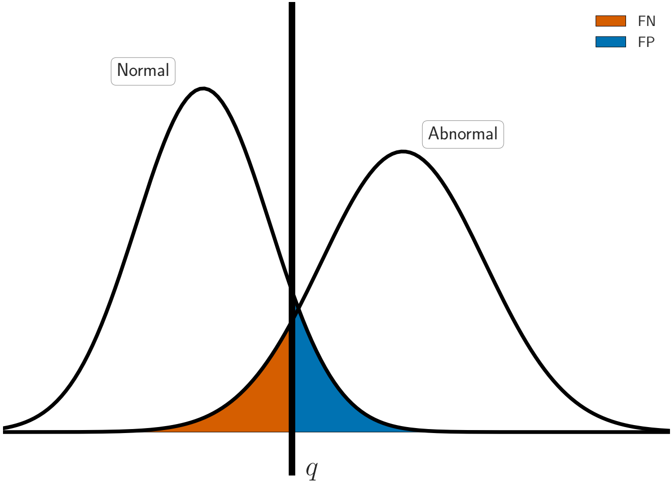
Courbe ROC



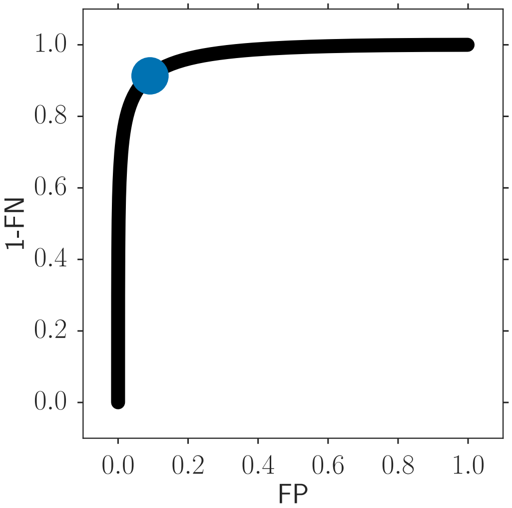
Courbe ROC



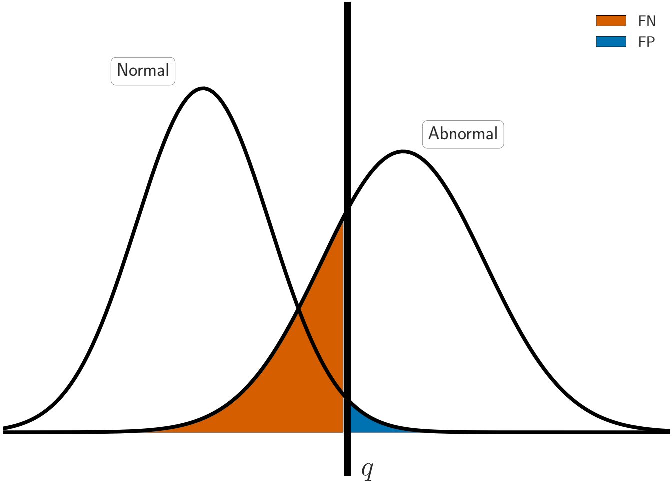
Courbe ROC



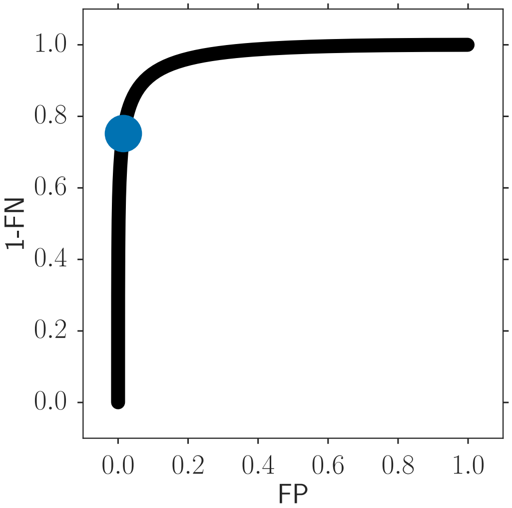
Courbe ROC



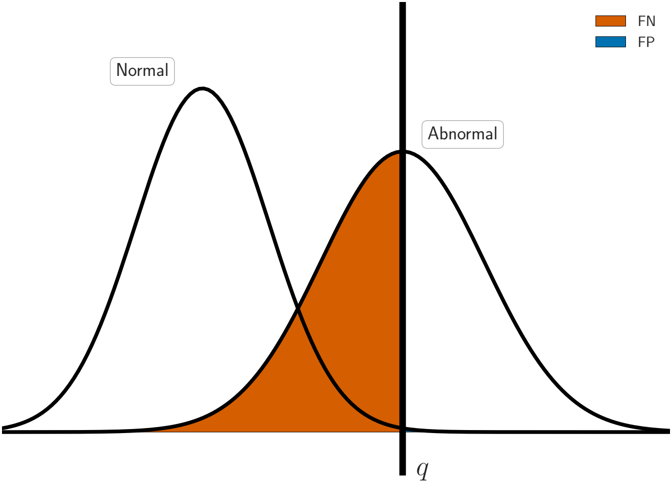
Courbe ROC



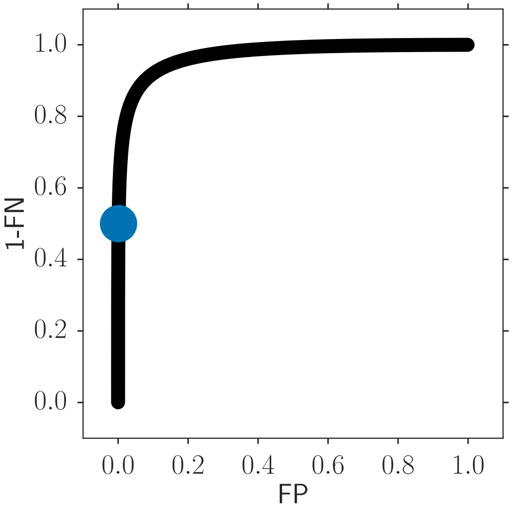
Courbe ROC



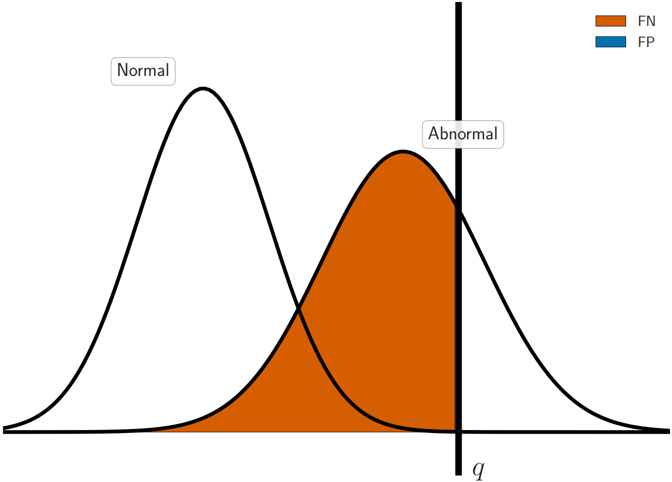
Courbe ROC



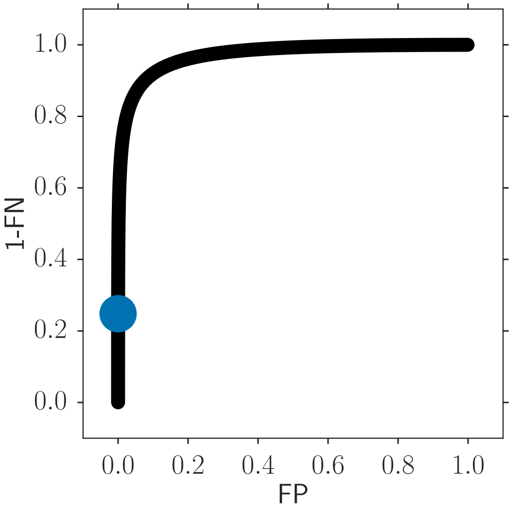
Courbe ROC



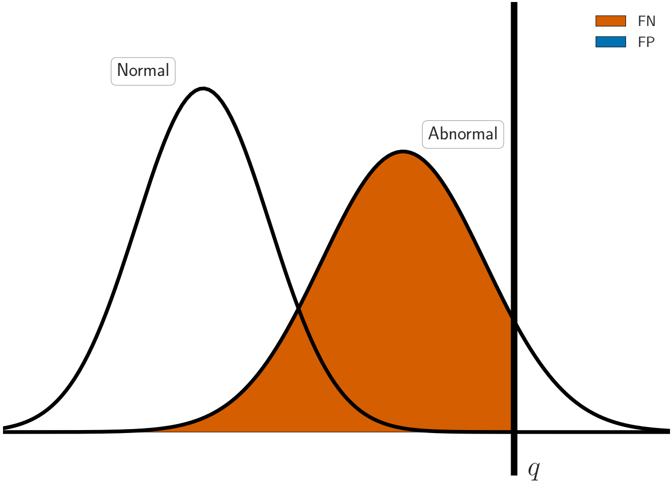
Courbe ROC



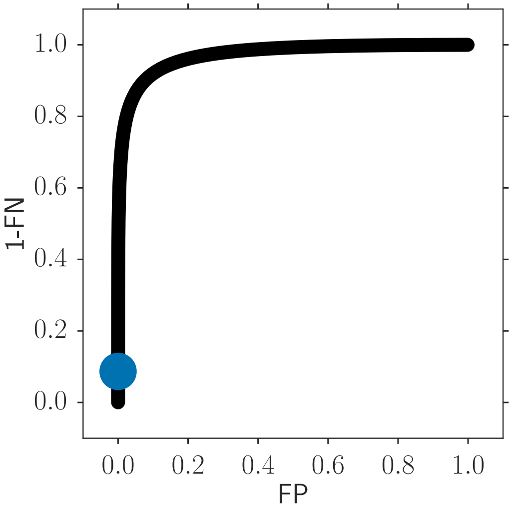
Courbe ROC



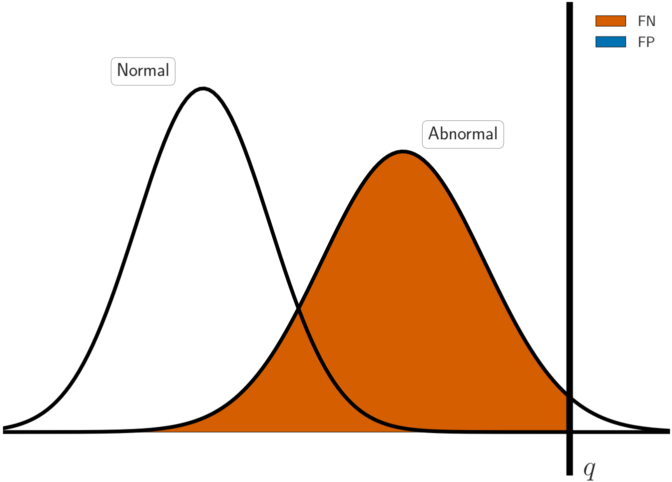
Courbe ROC



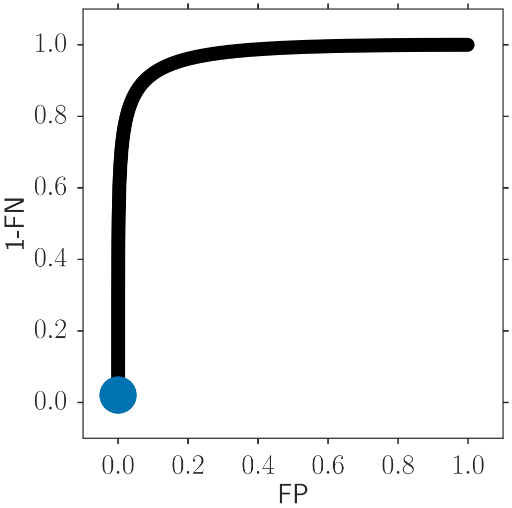
Courbe ROC



Courbe ROC



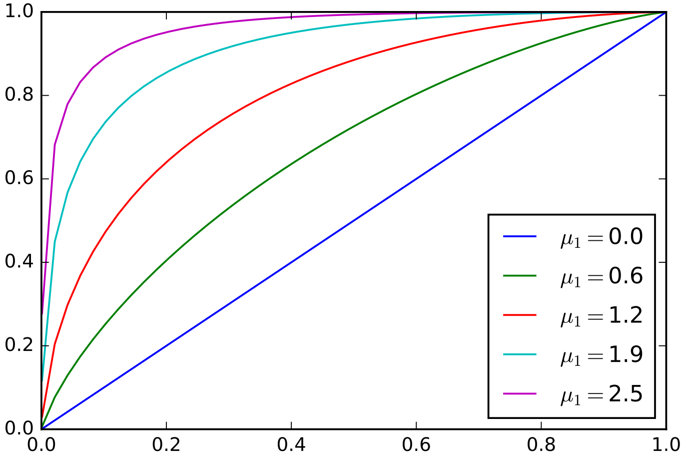
Courbe ROC



Outline

La courbe ROC dans le cas bi-normal

- F et G sont des Gaussiennes de paramètres μ_0, σ_0 et μ_1, σ_1, respectivement.
- On spécifie $\mu_0 = 0$, $\sigma_0 = \sigma_1 = 1$, on fait varier μ_1



Estimation–application

Estimation de la courbe ROC

- Maximum de vraisemblance
- Non-paramétrique
- Bayésien avec variable d'état latente
- Estimation de l'aire sous la courbe ROC

Application

- Pour comparer différents tests statistiques
- Pour comparer différents algorithmes d'apprentissage supervisé
- Pour comparer des méthodes de sélection de support du Lasso

nb : ROC = Receiver Operating Characteristic

Références