

Séminaire final Q'HACCP
ANSES, le 24 Novembre 2011

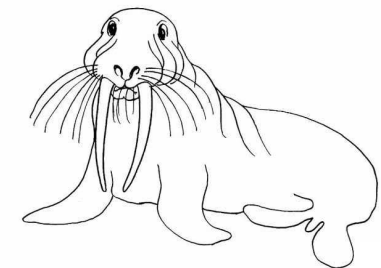


Surveillance et échantillonnage : Approches classique ou décisionnelle?

Eric PARENT, Natalie COMMEAU



AgroParisTech
ENGREF
INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT
PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES

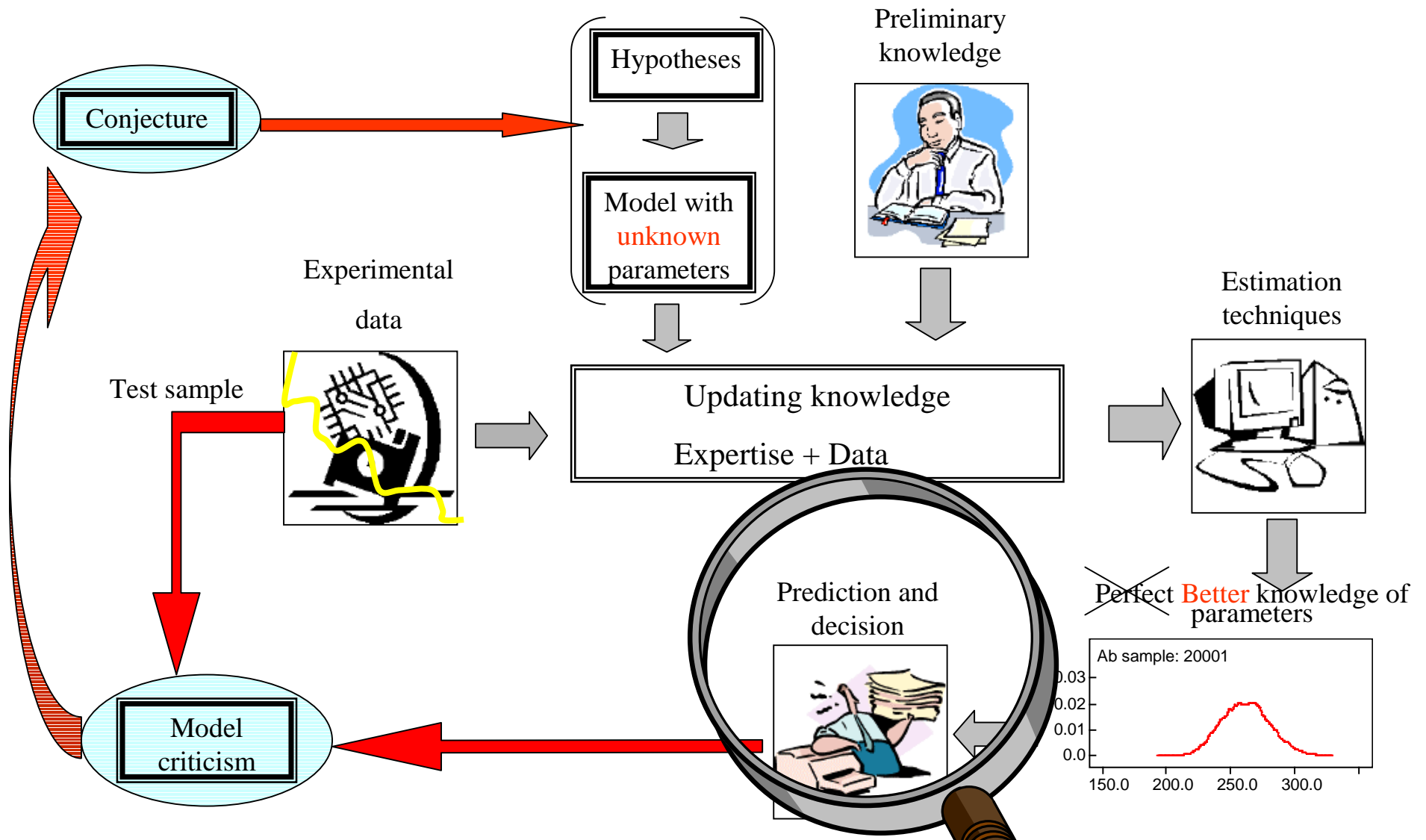


Equipe de **Modélisation**, **Risques**, **Statistique** et **Environnement**
(UMR 518)

AgroParisTech/ Institut National de la Recherche Agronomique
16, rue Claude BERNARD, 750005 Paris

Eric.Parent@agroparistech.fr Natalie.Commeau@agroparistech.fr

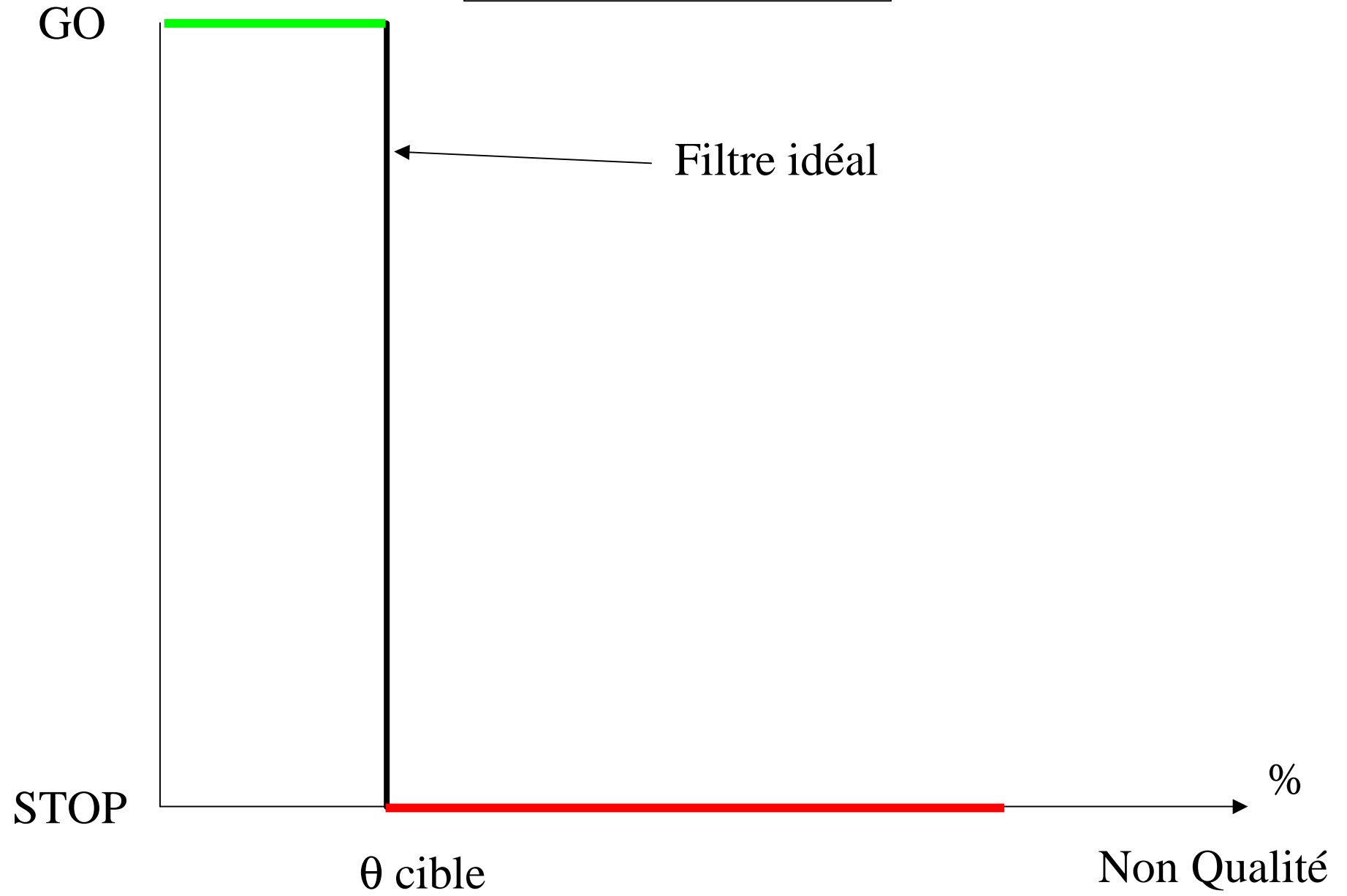
Que fait on après l'inférence?...



1. Approche classique

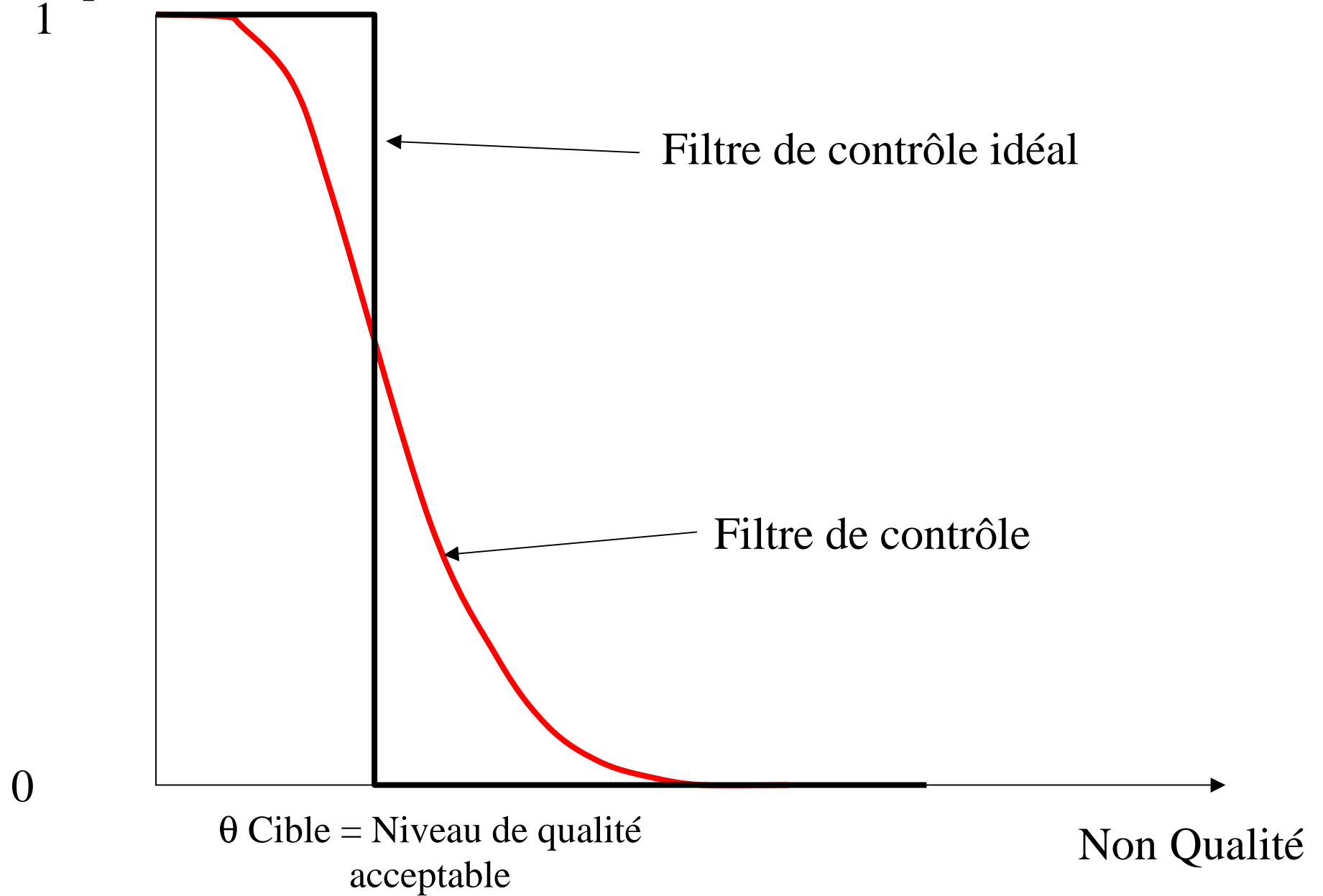
- Filtre
- Filtre imparfait
- Risques type 1 type 2
- Difficultés

Contrôler une qualité



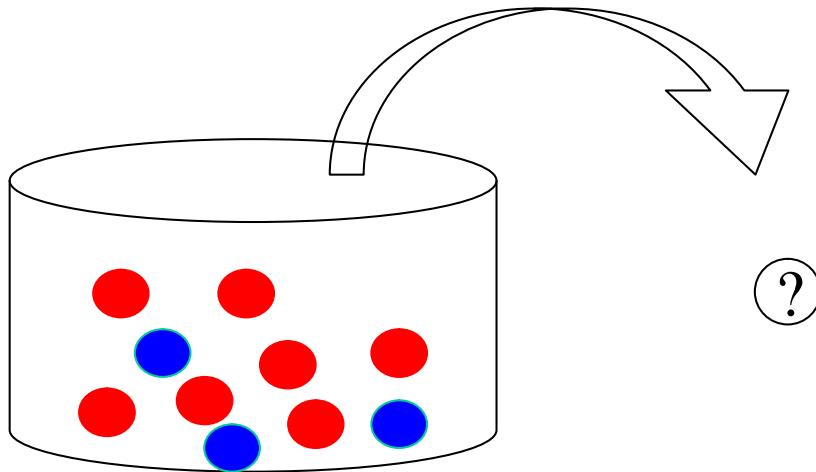
Probabilité
d'acceptation
1

Echantillonnage aléatoire



Principe: Analyse de sensibilité de Y sachant θ

Faire varier la non qualité θ et imaginer les résultats possibles Y



Le nombre d'objets défectueux est une variable aléatoire, directement observable et dont la loi est paramétrisée par l'inconnue θ

De la fréquence asymptotique :

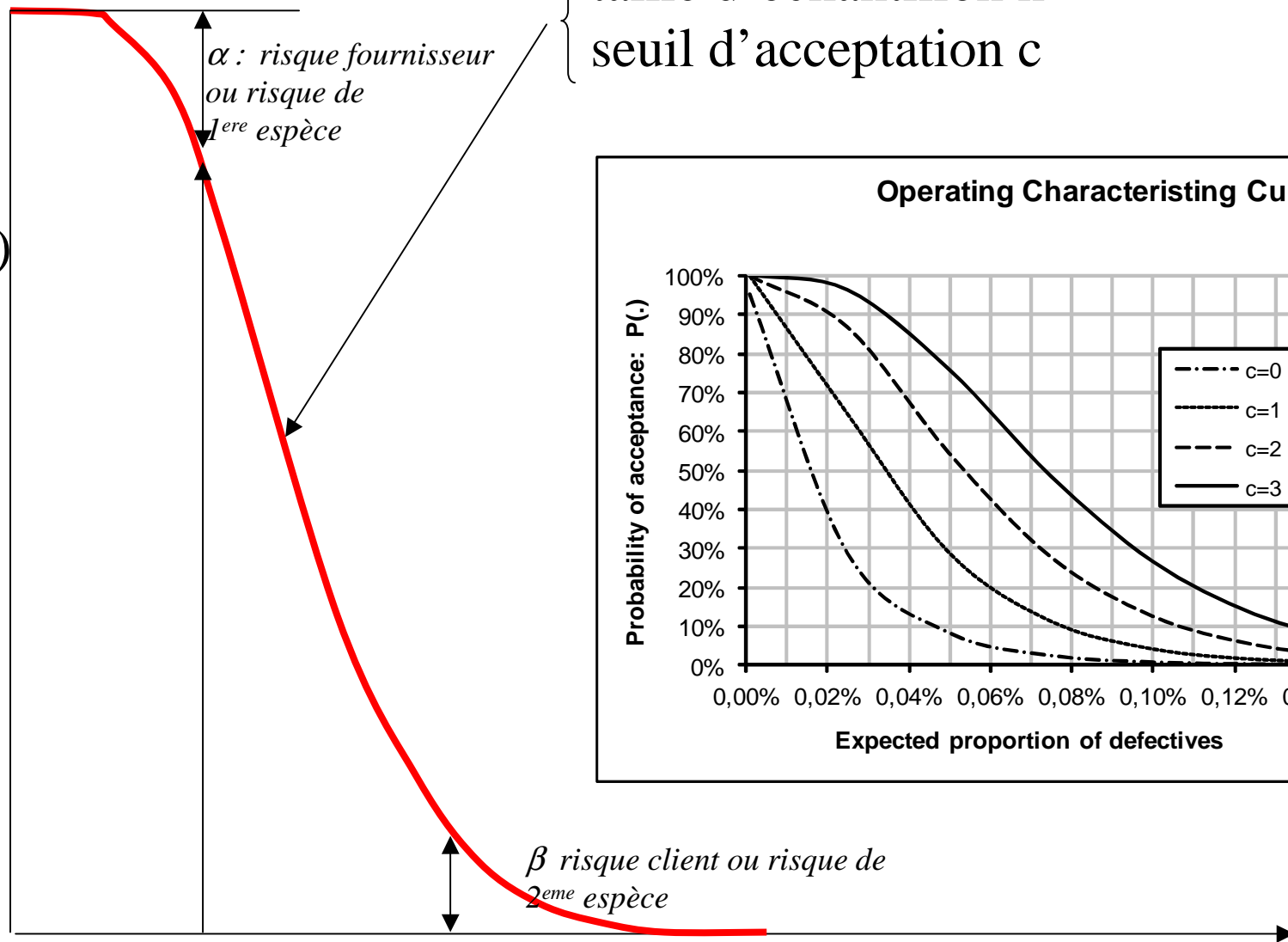
La probabilité p de l'événement incertain A est la limite ($n \rightarrow \infty$) de la fréquence $\sum_{i=1:n} 1_A(i)/n$ lorsqu'on répète les tirages en conditions iid

Epreuve de Bernoulli: $\Pr(\textcircled{?} = \bullet) = \theta$
 $\Pr(\textcircled{?} = \bullet) = 1 - \theta$

Binomial = Somme iid de Bernoulli's

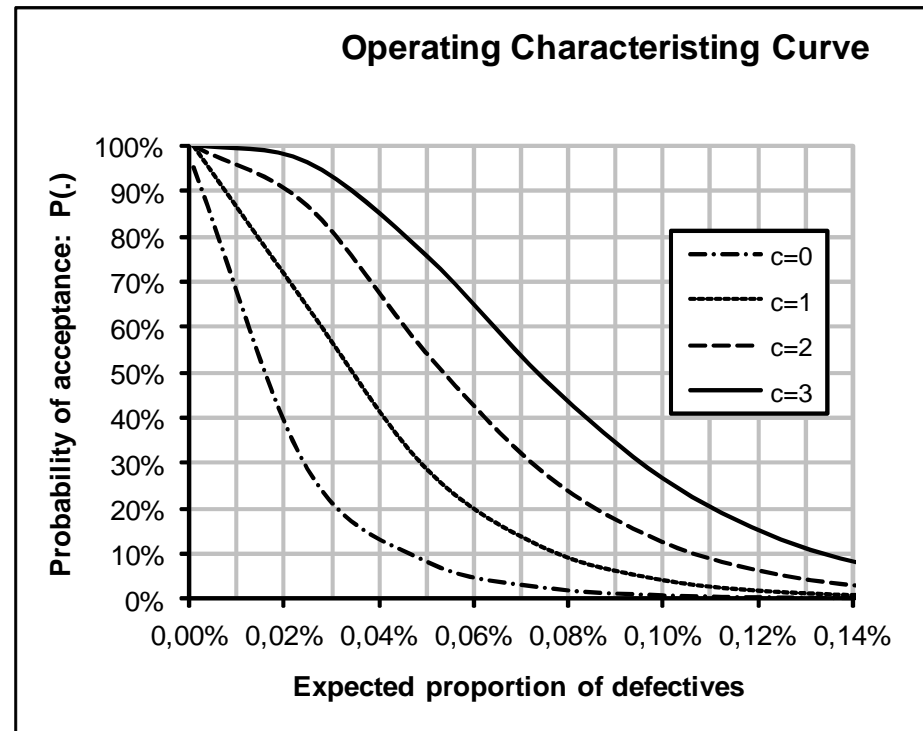
Probabilité
d'accepter
le lot

$$\Pr(Y < c | n, \theta)$$



Approche classique : Courbe d'efficacité

taille d'échantillon n
seuil d'acceptation c



Non Qualité

Des normes et des réglementations



Food industry bodies



Codex Alimentarius

- CCPR
- CCMAS

Codex Committee on Pesticide Residue
Codex Committee on Methods of Analysis and Sampling



Book entitled:
"Microorganisms in foods 2: Sampling for Microbiological Analysis: Principles and Specific Applications"

TC#

TC69 Application of statistical methods

- SC1 Vocabulary and terms
- SC4 Applications of statistical methods in process management
- SC5 Acceptance sampling
- SC6 Measurement methods and results

Soup industry



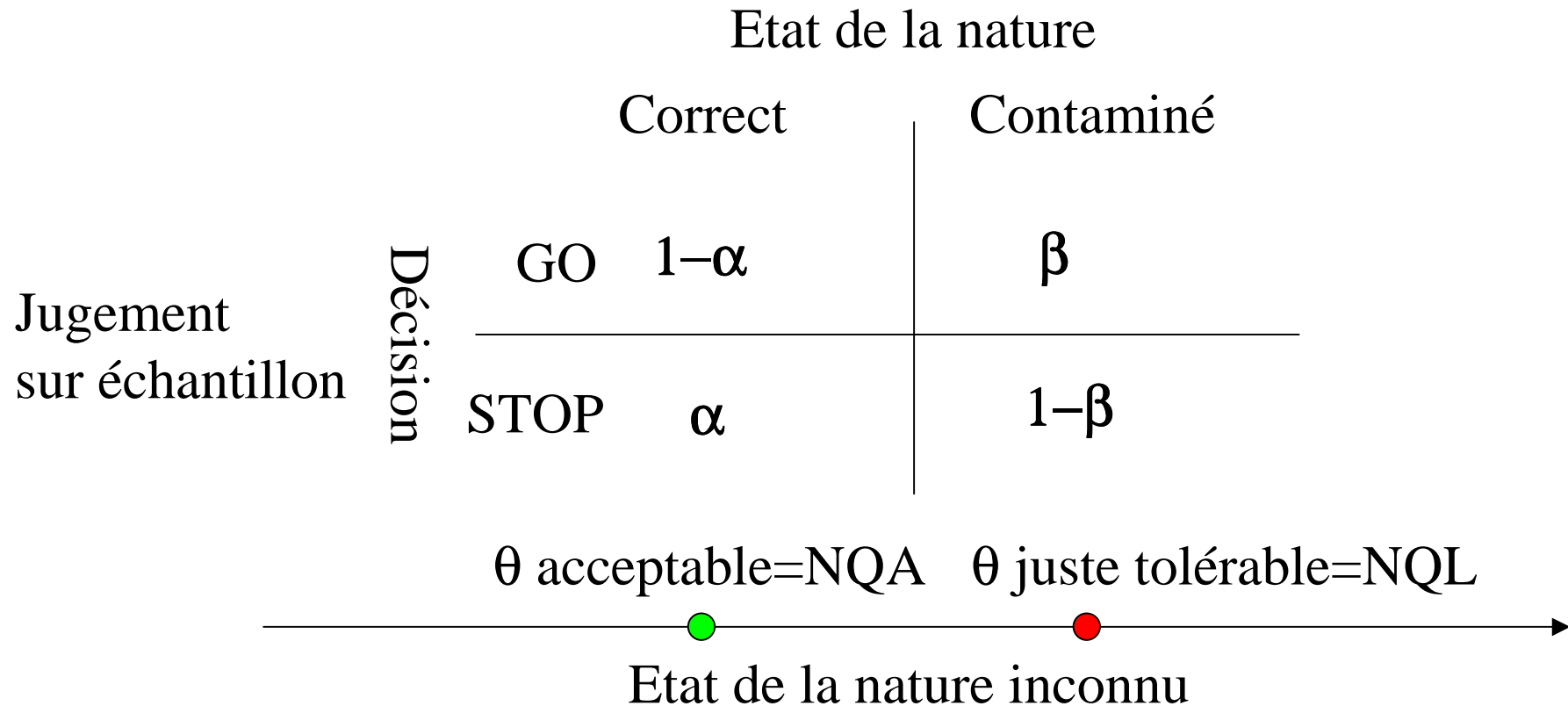
Normes



pour l'échantillonnage

ISO 2859-0:1995	Sampling procedures for inspection by attributes -- Part 0: Introduction to the ISO 2859 attribute sampling system
ISO 2859-1:1999	Sampling procedures for inspection by attributes -- Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection
ISO 2859-1:1999/Cor 1:2001 ISO 2859-2:1985	Sampling procedures for inspection by attributes -- Part 2: Sampling plans indexed by limiting quality (LQ) for isolated lot inspection
ISO 2859-3:1991	Sampling procedures for inspection by attributes -- Part 3: Skip-lot sampling procedures
ISO 2859-4:2002	Sampling procedures for inspection by attributes -- Part 4: Procedures for assessment of declared quality levels
ISO 3951:1989	Sampling procedures and charts for inspection by variables for percent nonconforming
ISO 8422:1991	Sequential sampling plans for inspection by attributes
ISO 8422:1991/Cor 1:1993 ISO 8423:1991	Sequential sampling plans for inspection by variables for percent nonconforming (known standard deviation)
ISO 8423:1991/Cor 1:1993 ISO/TR 8550:1994	Guide for the selection of an acceptance sampling system, scheme or plan for inspection of discrete items in lots
ISO 10725:2000	Acceptance sampling plans and procedures for the inspection of bulk materials
ISO 11648 -1:2003	Statistical aspects of sampling from bulk materials -- Part 1: General principles
ISO 11648 -2:2001	Statistical aspects of sampling from bulk materials -- Part 2: Sampling of particulate materials

Un problème de décision



Vocabulaire: Risques 1^{er} et 2^{eme} esp, sensibilité, spécificité, client, fournisseur

Six grandeurs, deux équations

- $\alpha = \Pr(Y > c | n, \theta_{\text{acceptable}})$
- $\beta = \Pr(Y \leq c | n, \theta_{\text{tolérable}})$

θ_{acc} , θ_{tol} , c , n fixés $\rightarrow \alpha, \beta$

θ_{acc} , θ_{tol} , α , n fixés $\rightarrow c, \beta$

θ_{acc} , θ_{tol} , α , β fixés $\rightarrow c, n$

Avantages et ennuis

- Obéir aux plans classiques procure une garantie contractuelle
- Comment choisir α et β ? On fixe plutôt n et on déduit c et β pour $\alpha=5\%$.
- Contrôler des boulons ou des yaourts de la même façon?
- Prendre en compte d'autres informations ...bricoler les procédures classiques?

2. Approche décisionnelle

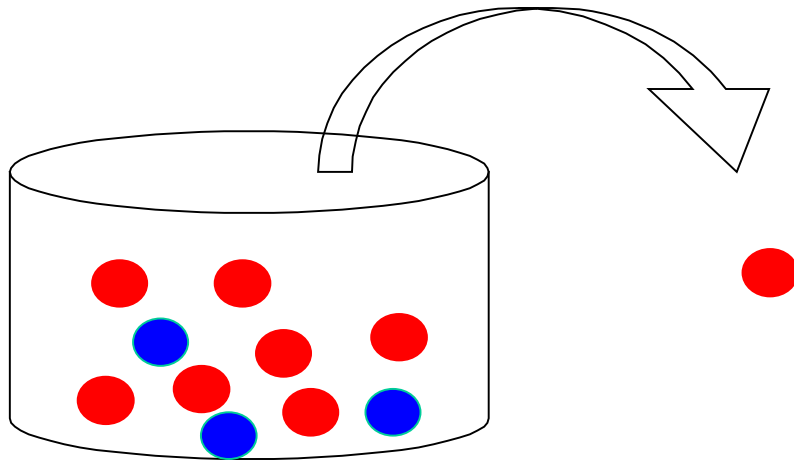
- Apprentissage statistique bayésien
- Formulation décisionnelle
- Approche décisionnelle=classique ?
- Elicitation des éléments du pb décisionnel
- Approche décisionnelle > classique ?

Des effets aux causes



Autre point de vue : Analyse de θ sachant Y

Se mettre en situation prédictive (divers possibles Y) et parier sur θ



La non qualité θ est une inconnue, seulement indirectement observable et unique

Nouvel emploi de la probabilité:

J'attribue la probabilité p à l'événement incertain A si je suis prêt à acheter ou vendre p euros le pari qui rapporte 1 euro si A se produit, 0 sinon.

....conditionnelle à mon état d'information

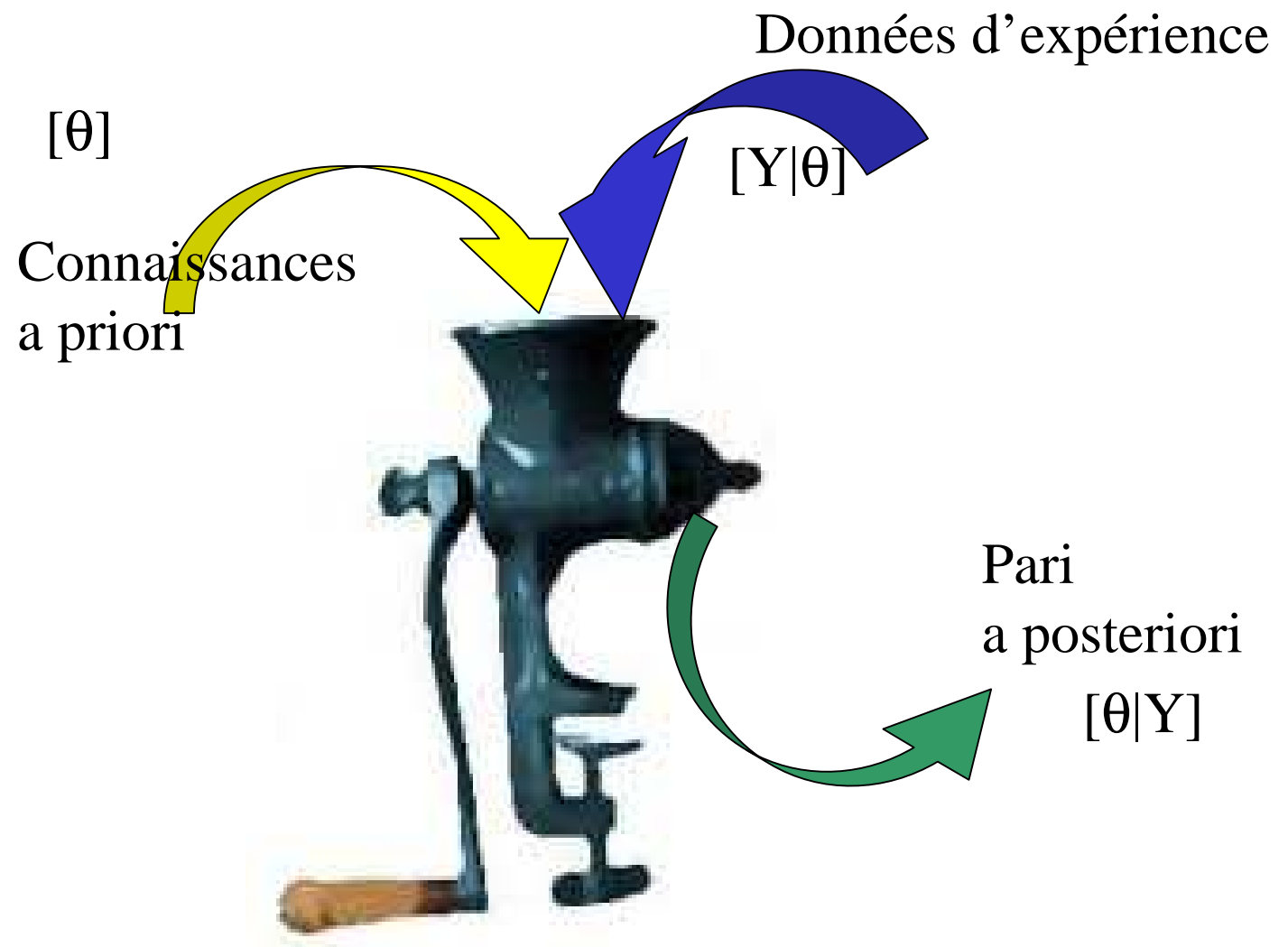
Vraisemblance

$$[y / \theta, n] = C_n^y \theta^y (1 - \theta)^{n-y}$$

Prior

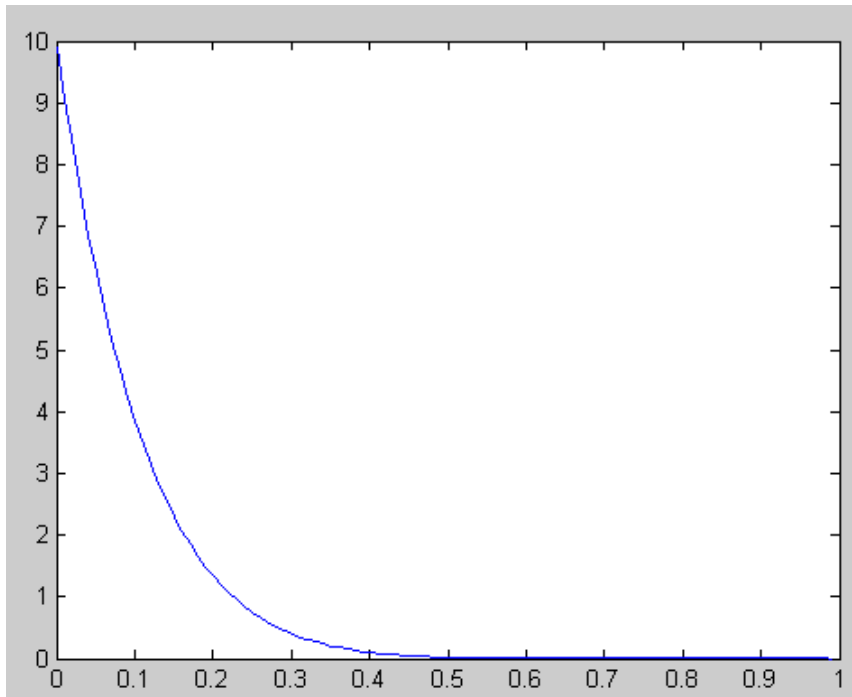
$$[\theta / a, b] = \frac{1}{B(a, b)} \theta^{a-1} (1 - \theta)^{b-1}$$

The « Bayesian crank »



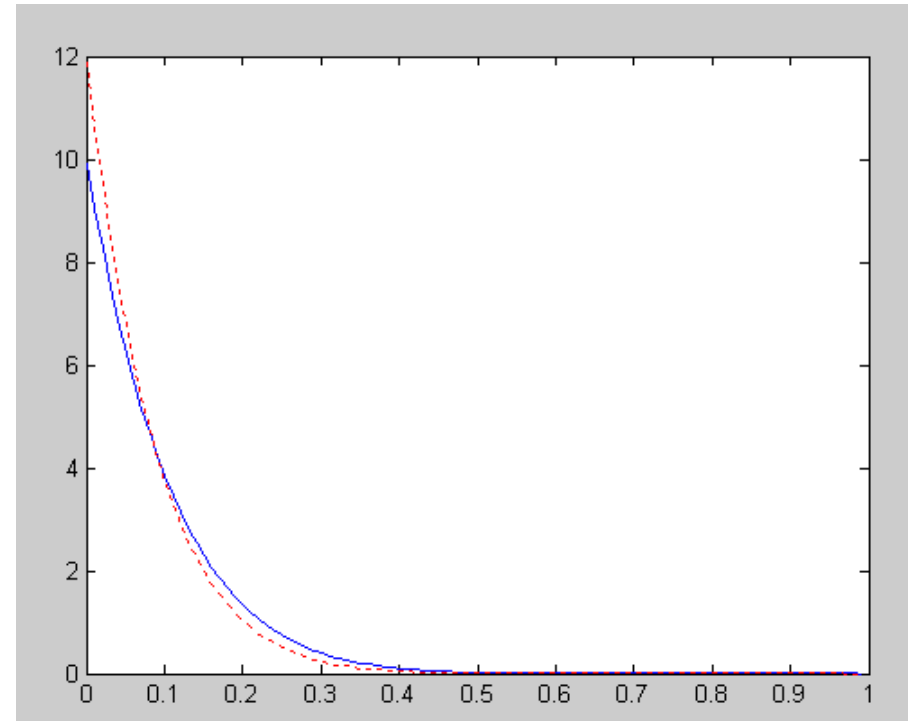
L'apprentissage bayésien(I)

Prior



Beta(1,10)

Posterior



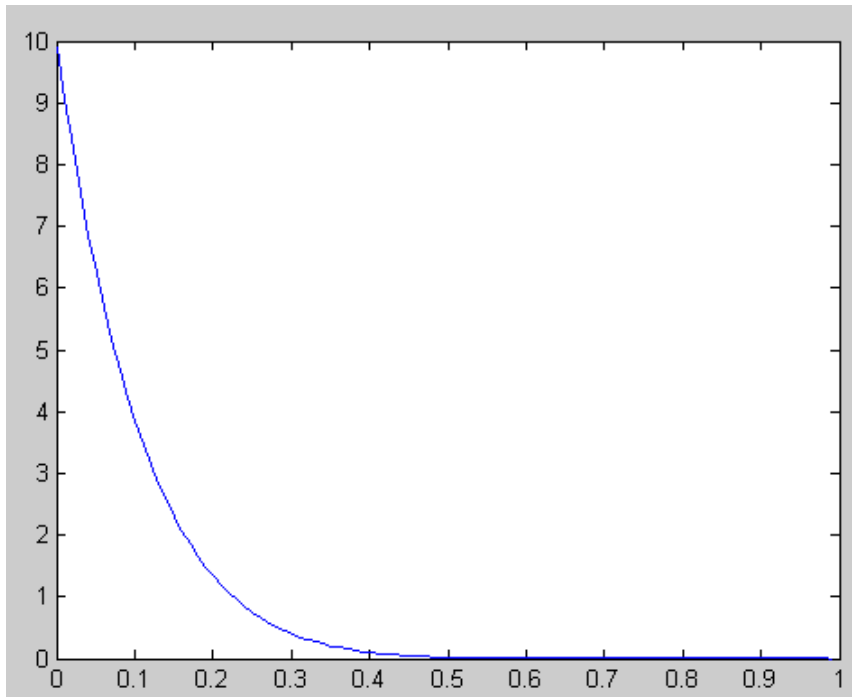
$n=2$

$x=0$ defective

Beta(1,12)

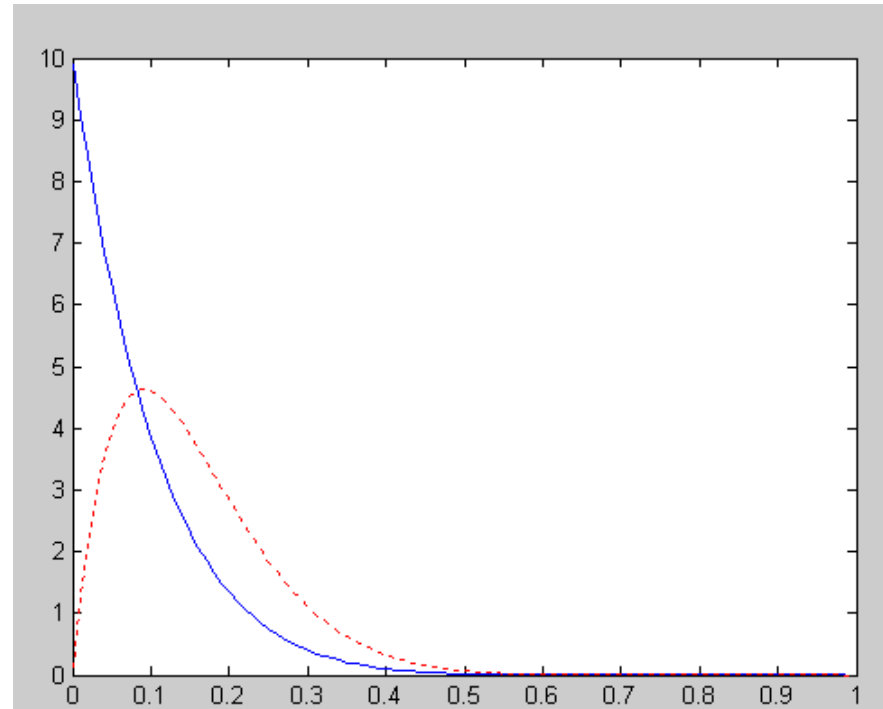
L'apprentissage bayésien(II)

Prior



Beta(1,10)

Posterior



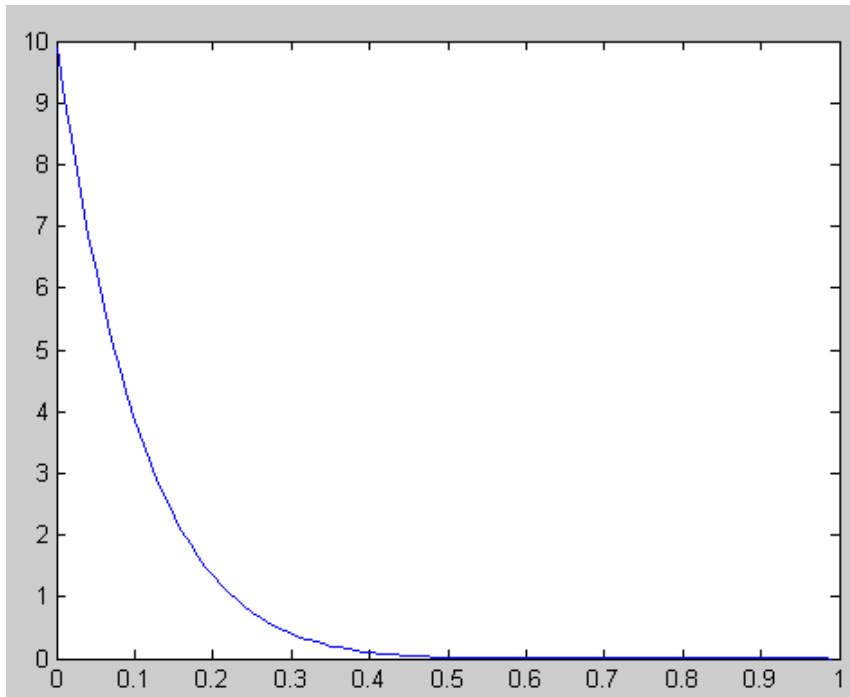
$n=2$

$x=1$ defective

Beta(2,11)

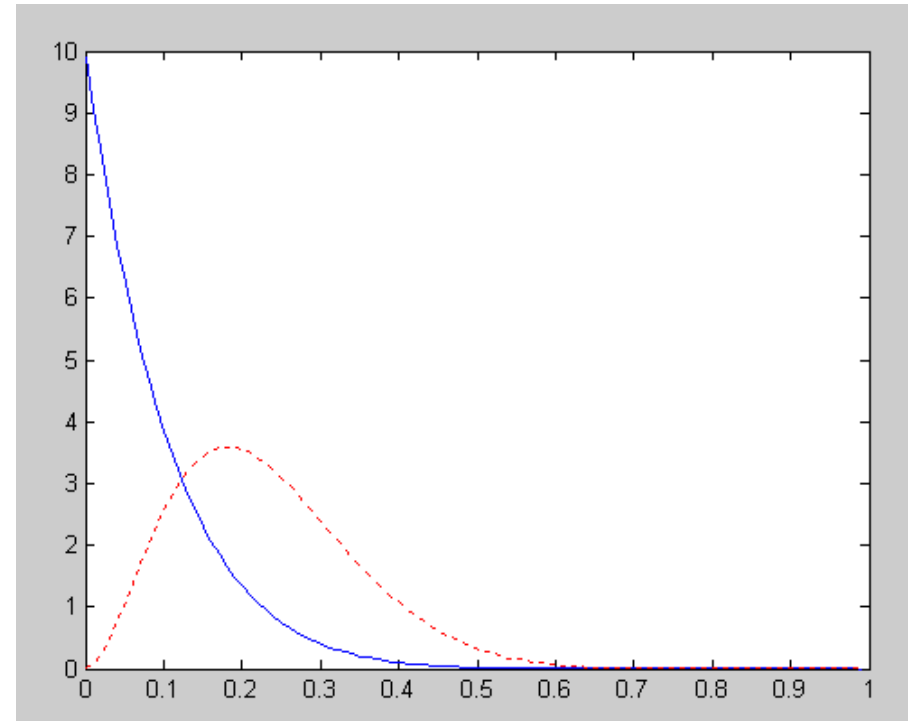
L'apprentissage bayésien(III)

Prior



Beta(1,10)

Posterior



$n=2$

$x=2$ defectives

Beta(3,10)

Eléments décisionnels

Décisions:

d1: GO

d2: STOP

Prior : a succès

pour a+b essais

Coûts type (linéaires):

$$W(d_1, \theta) = W(\text{livraison}, \text{qualité}) = C_1\theta + kn$$

$$W(d_2, \theta) = W(\text{rebut}, \text{qualité}) = C_2 + kn$$

Règles de Décision (à n fixé)

Règles intuitives classiques

$$y \leq c \Rightarrow \delta_c(y) = d_1(\text{livrer})$$

$$y > c \Rightarrow \delta_c(y) = d_2(\text{rebut})$$

Règles bayésiennes

$$E_{\theta/y}(W(\delta(y)=d_1, \theta)) = C_1 E_{\theta/y}(\theta) + kn$$

$$E_{\theta/y}(W(\delta(y)=d_2, \theta)) = C_2 + kn$$

Choisir livrer si

$$C_1 E_{\theta/y}(\theta) < C_2$$

Règles de décisions

$$C_1 E_{\theta/y}(\theta) = C_1 \frac{a+y}{a+b+n} < C_2$$

Choisir livrer d_1 si

$$y_{seuil} = \frac{n+b - \left(\frac{C_1 - C_2}{C_2}\right)a}{\frac{C_1}{C_2}}$$

$$y_{seuil} = \frac{C_2}{C_1}(a+b+n) - a$$

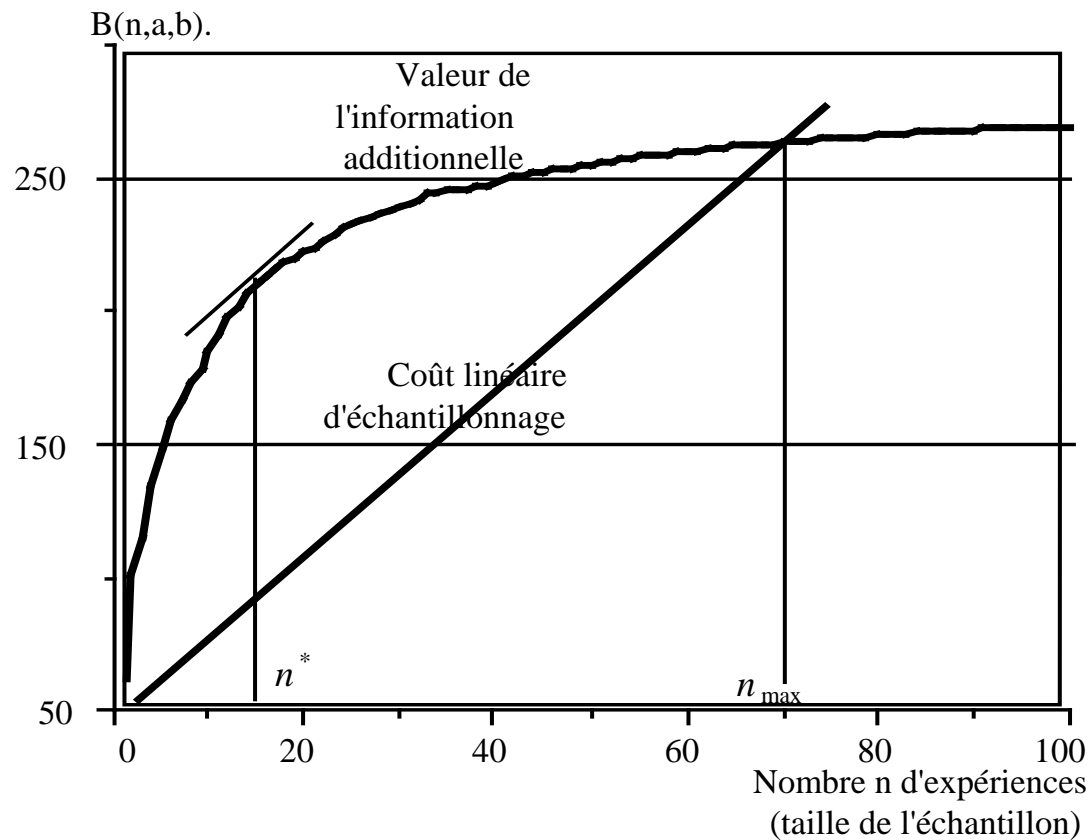
Interprétation économique du critère classique!

a nombre virtuel de défectueux

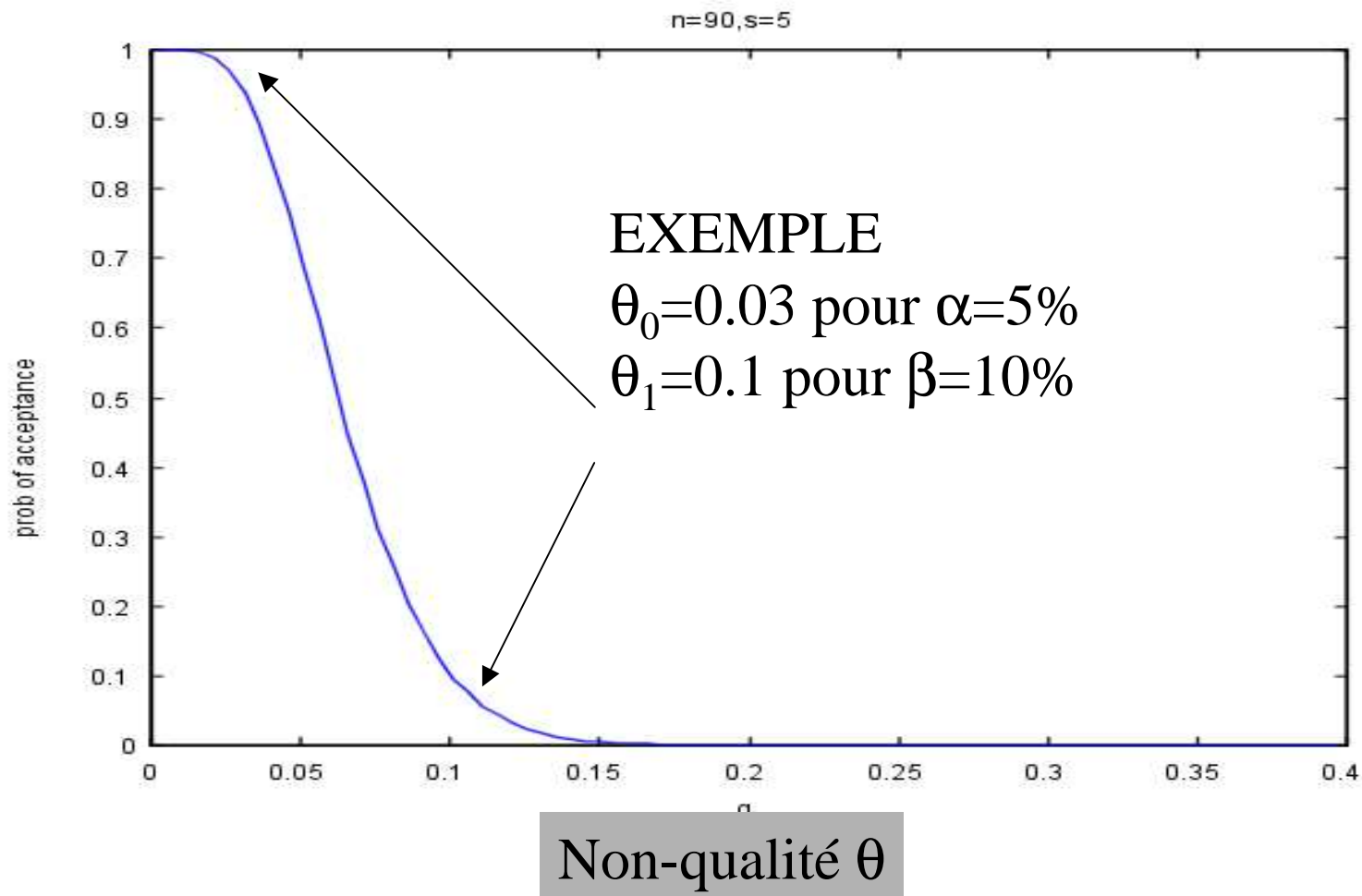
b nombre virtuel d'acceptables

Economie= valeur moyenne de l'information additionnelle, ici:

$$\sum_{y > y_{seuil}} \left(c \frac{a}{a+b} - 1 \right) [y|a,b]$$



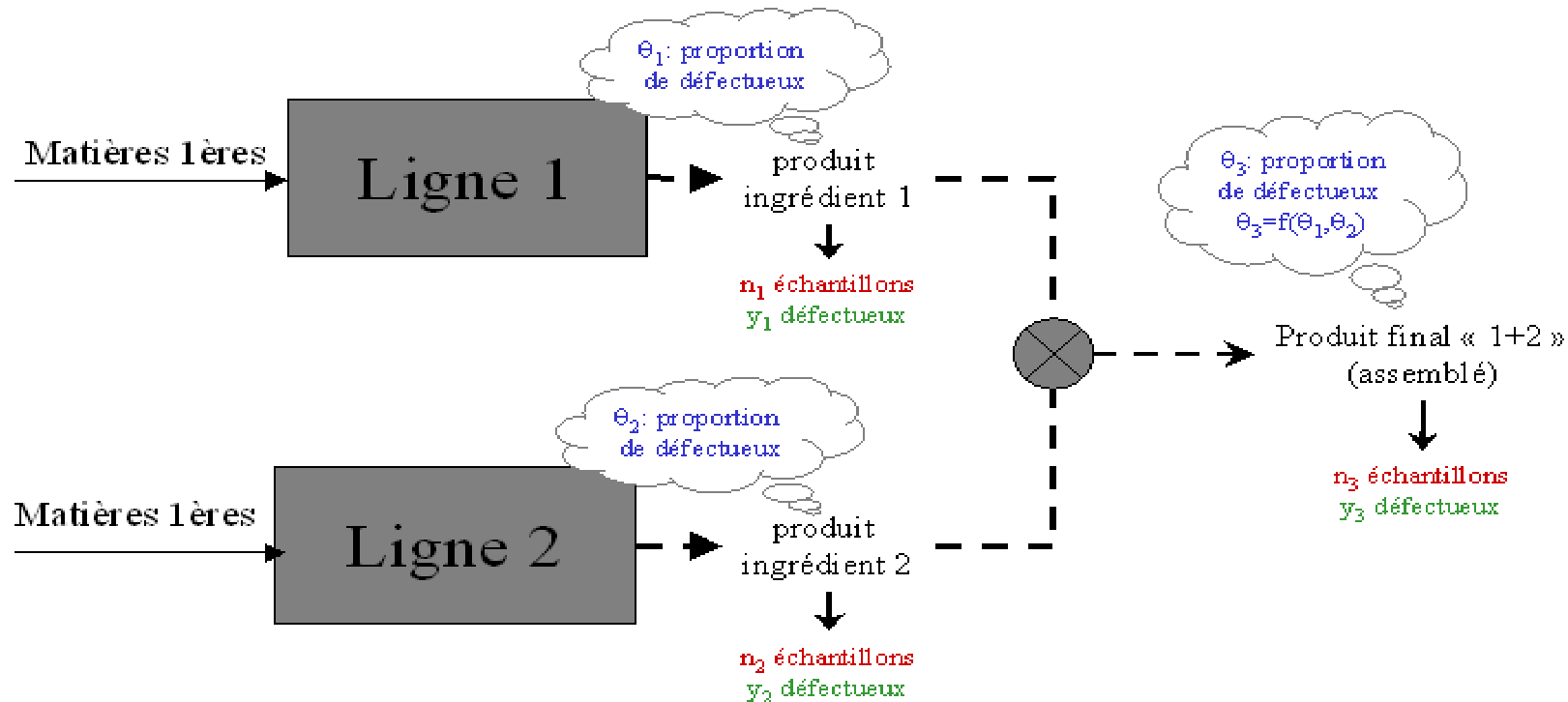
Equivalence classique - décisionnel



Pourquoi sortir du plan classique?

- Une méthode implicite: $n=90$, $c=5$ *classique*
« correspond » à
 $a=2.18, b=38.18$ (*savoir a priori*)
 $C_2/C_1=17$ (*perte image*) et
 $k=C_2/3000$ (*cout d'échantillonnage*)
- Possibilité d'élicitation par morceaux et utilisation dans des problèmes assemblés plus compliqués
- Difficile de faire annoncer des coûts à un responsable qualité, réticence à sortir des outils classique tout faits

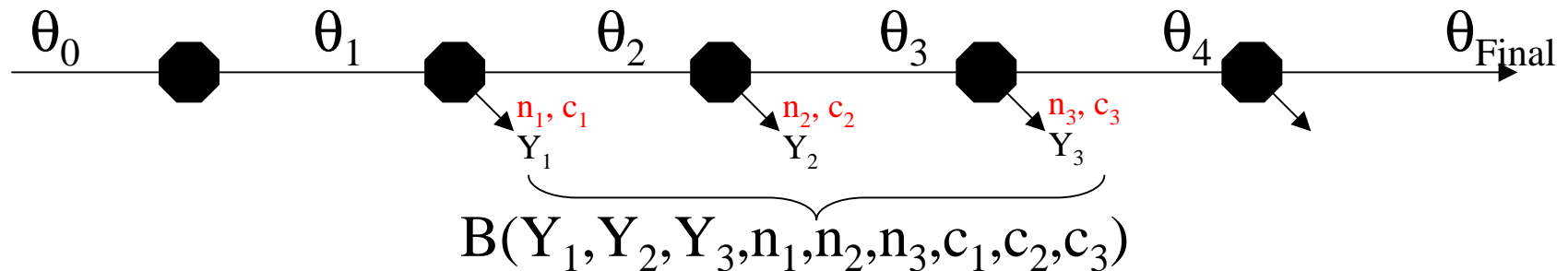
Contrôle multipostede qualité



- Modèle pour chaque observable+lien entre les inconnues
- Sous des contraintes pratiques et budgétaires, à quel(s) poste(s) faut-il contrôler?

3. Conclusions: Le Challenge

- De la fourche à la fourchette : *découper* la chaîne en une suite d'opérations élémentaires



- *Modéliser* les lois conjointes de la contamination à chaque étape
- *Envisager* contraintes, coûts et bénéfices d'un contrôle à chaque étape (\$ ou santé publique?)
- *Additionner* les unités d'utilité et tester la valeur (moyenne en prédictif) de stratégies alternatives de monitoring ($n_1, n_2, n_3, c_1, c_2, c_3$)

Biblio

- Parent E., Chaouche A., Girard P. (1995). Sur l'apport des statistiques bayésiennes au contrôle statistique de la qualité par attribut. Partie 1 : Contrôle simple. *Revue de Statistique Appliquée* XLLIII(4) : 5-18.
- Parent E, Lang G., Girard P. (1996). Apport des statistiques bayésiennes au contrôle statistique de la qualité par attribut. Partie 2 : Contrôle séquentiel tronqué. *Revue de Statistique Appliquée* XLIV(1) : 37-54.
- Chaouche A., Parent E. (1998). Sur l'apport des statistiques bayésiennes au contrôle statistique de la qualité par attribut. Partie 3 : Contrôle rétrospectif par détection de changement dans une série de variables binomiales. *Revue de Statistiques Appliquées*, XLVI: 5-29.
- Parent E., Chaouche A. (2002) Méthodes bayésiennes pour la qualité in *Techniques Mathématiques pour l'Industrie Agroalimentaire* (Daudin JJ., Duby C., eds), **chap 3**: 59-91. Lavoisier, Éditions TEC et DOC

Formalisme général de la décision ?

- Cadre bayésien : lois **informatives** sur les paramètres du modèle.
- On dispose de :
 - \mathbf{d} = décision / plan d'échantillonnage
 - θ = paramètre inconnu \rightarrow **loi a priori** $[\theta]$
 - $y | \theta, \mathbf{d}$ = données prédites \rightarrow **modèle** $[y | \mathbf{d}, \theta]$
 - $u(\mathbf{d}, \theta, y)$ = utilité/critère de l'optimisation (>0 , bornée sur \mathbb{K})
- On veut maximiser l'utilité $u(\mathbf{d}, \theta, y)$ en moyenne :

$$U(d) = \iint_{\theta, y} [y | d, \theta] [\theta] u(d, \theta, y) d\theta dy \quad \longleftarrow \quad \text{Le défi!}$$

Travailler les algorithmes



Reverend
Thomas Bayes
(1702-1761)

Bayes formula

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$$

1762

$$[\theta|y] = \frac{[\theta] \times [y|\theta]}{[y]}$$

- Conjugate prior
- Non informative prior
- Bayesian decision theory

1950's

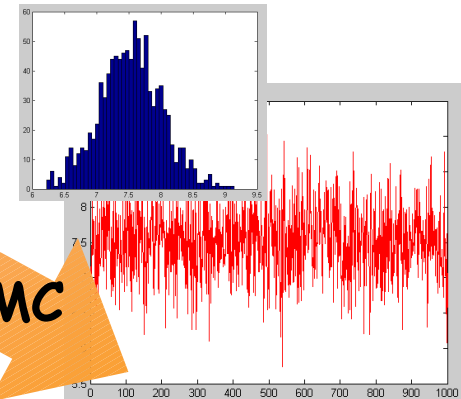
$$[\theta_1|y] = \iiint_{\theta_2} [\theta_1, \theta_2, \dots, \theta_n | y] d\theta_2 \dots d\theta_n$$

IS

1990's

MCMC

- Gibbs sampling
- Hasting-Metropolis
- Importance Sampling
- Interactive MCMC's



2000's

- Population MCMC
- Particles
- ABC
- Variational Bayes