

# Actuariat IARD - ACT2040

## Partie 1 - introduction

Arthur Charpentier

charpentier.arthur@uqam.ca

[http ://freakonometrics.hypotheses.org/](http://freakonometrics.hypotheses.org/)



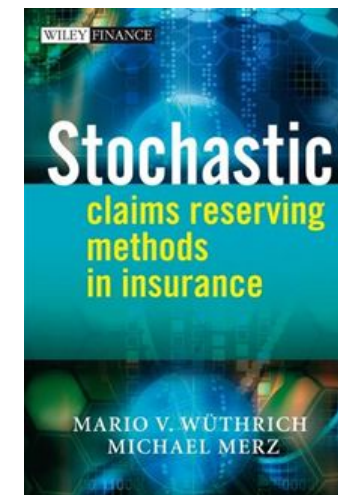
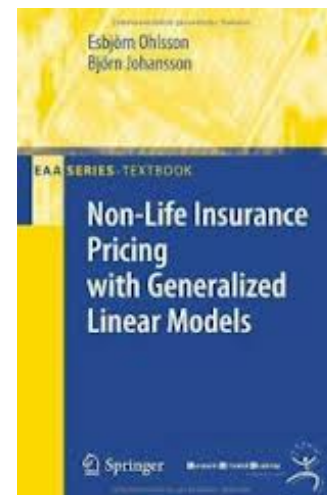
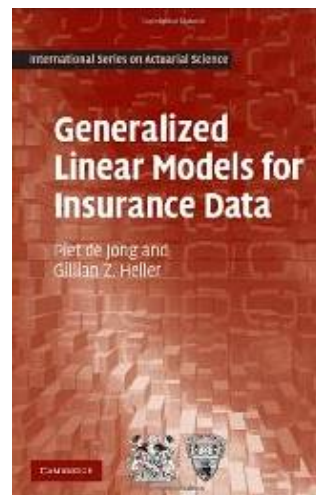
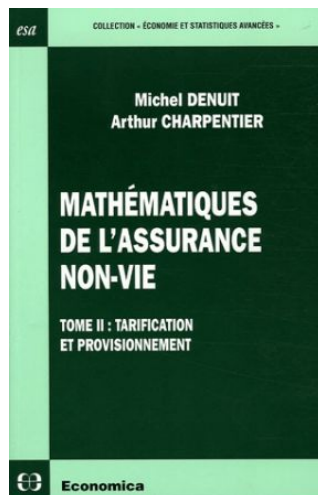
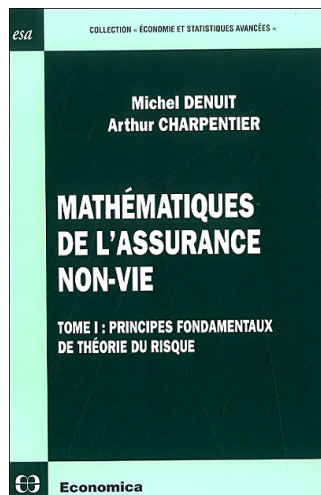
AUTOMNE 2013

## Plan du cours

- **Motivation et introduction : tarification et provisionnement**
- **Modélisation de la survenance d'un sinistre**
  - Régressions logit et probit,  $Y_i | \mathbf{X}'_i \sim \mathcal{B}(\pi[\mathbf{X}'_i \boldsymbol{\beta}])$
  - Les arbres de régression
- **Fréquence de survenance de sinistres**
  - Régression de Poisson  $Y_i | \mathbf{X}'_i \sim \mathcal{P}(\lambda[\mathbf{X}'_i \boldsymbol{\beta}])$
  - Surdispersion : quasi-Poisson, Binomiale Négative & inflation de zéros
- **Coûts individuels de sinistres**
  - Régressions Gamma et lognormale
  - Écurement des grands sinistres, et mutualisation
- **Provisionnement pour sinistres à payer**
  - Les triangles de provisionnement
  - Méthode de Mack
  - Régression de Poisson

## Prérquis et ouvrages de référence

Ce cours suppose acquises les méthodes vues au cours ACT6420, *méthodes de prévision*.



## L'hétérogénéité



## La prime pure

Pour une variable de comptage

$$\mathbb{E}(N) = \sum_{n \in \mathbb{N}} n \mathbb{P}(N = n) = \sum_{n \in \mathbb{N}} \mathbb{P}(N > n)$$

Pour une variable continue positive

$$\mathbb{E}(X) = \int_{x \in \mathbb{R}_+} x f(x) dx = \int_{x \in \mathbb{R}_+} \mathbb{P}(X > x) dx$$

où  $\mathbb{P}(X > x) = \bar{F}(x) = \int_x^\infty f(t) dt.$

## La prime pure

Plus généralement,

$$\mathbb{E}(X) = \int_{x \in \mathbb{R}} x dF(x)$$

où  $F$  admet un nombre fini (ou dénombrable) de discontinuité  $\{d_1 \leq d_2 \leq \dots\}$ .

$$dF(x) = \begin{cases} F(d_n) - F(d_n^-) & \text{si } x = d_n \\ \tilde{f}(x) & \text{sinon} \end{cases}$$

où

$$F(x) = \sum_{d_n \leq x} \mathbb{P}(X = d_n) + \int_{t \leq x} \tilde{f}(t) dt$$

**Définition 1.** *Etant donné un risque  $X$ , la prime pure est  $\pi_X = \mathbb{E}[X]$ .*

## Exemples de calculs d'espérance

Pour les lois discrètes classiques

- si  $N \sim \mathcal{B}(n, p)$ ,  $\mathbb{P}(N = k) = \binom{n}{k} p^k [1 - p]^{n-k}$ , alors  $\mathbb{E}(X) = np$ ,
- si  $N \sim \mathcal{P}(\lambda)$ ,  $\mathbb{P}(N = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ , alors  $\mathbb{E}(X) = \lambda$ ,
- si  $N \sim NB(r, p)$ ,  $\mathbb{P}(N = k) = \frac{\Gamma(r + k)}{k! \Gamma(r)} p^r [1 - p]^k$ , alors  $\mathbb{E}(X) = \frac{r[1 - p]}{p}$ ,

Pour les lois continues classiques

- si  $N \sim \mathcal{N}(\mu, \sigma^2)$ ,  $\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$ , alors  $\mathbb{E}(X) = \mu$ ,
- si  $N \sim LN(\mu, \sigma^2)$ ,  $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{[\ln(x) - \mu]^2}{2\sigma^2}\right)$ , alors  

$$\mathbb{E}(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right),$$
- si  $N \sim \mathcal{G}(\alpha, \beta)$ ,  $f(x) = x^{\alpha-1} \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)}$ , alors  $\mathbb{E}(X) = \frac{\alpha}{\beta}$ ,

## Espérance pour une loi composée

Dans un modèle collectif, on s'intéresse à  $S = \sum_{n=1}^N X_i$  si  $N \geq 1$ , 0 sinon.

Si les  $X_i$  sont i.i.d., indépendants de  $N$ , alors

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}[\mathbb{E}(S|N)] = \sum_{k=1}^{+\infty} \mathbb{P}[N = k] \cdot \mathbb{E} \left[ \sum_{i=1}^k X_i \right] \\ &= \left( \sum_{k=1}^{+\infty} \mathbb{P}[N = k] \cdot k \right) \mathbb{E}[X_1] = \mathbb{E}[N] \cdot \mathbb{E}[X_1].\end{aligned}$$

où

- $\mathbb{E}[N]$  est la **fréquence**
- $\mathbb{E}[X_1]$  est le **coût moyen**



## De l'espérance a la prime pure

Pascal, Fermat ou Condorcet (XVIIIème siècle) proposaient d'évaluer le “*produit scalaire des probabilités et des gains*”,

$$\langle \mathbf{p}, \mathbf{x} \rangle = \sum_{i=1}^n p_i x_i = \mathbb{E}_{\mathbb{P}}(X),$$

selon la “*règle des parties*”.

⇒ garantie un équilibre du système, en moyenne.

L'espérance mathématique est un prix “*juste*” (FELLER (1943)),

– moindres carrés,  $\mathbb{E}(X) = \operatorname{argmin}\{\|X - c\|_{L^2}, c \in \mathbb{R}\}$

– loi des grands nombres,  $\frac{X_1 + \dots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}(X),$

– théorème central limite,  $\frac{X_1 + \dots + X_n}{n} \stackrel{\mathcal{L}}{\approx} \mathcal{N}\left(\mathbb{E}(X), \frac{\operatorname{Var}(X)}{\sqrt{n}}\right),$

– probabilité de ruine,  $\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_1 + \dots + X_n}{n} > \pi\right) = 1$  pour  $\pi < \mathbb{E}(X).$

## De l'espérance a la prime pure

**Remarque :** Paradoxe de Saint Pétersbourg : l'espérance de gain est infinie,

$$\sum_{i=1}^{\infty} \mathbb{P}(\text{arrêt au } i\text{ème coup}) \cdot 2^i = \sum_{i=1}^{\infty} \frac{1}{2^i} \cdot 2^i = \sum_{i=1}^{\infty} 1 = \infty.$$

Peut-être pas forcément la méthode la plus intuitive, pour les grands risques...

## La prime pure - sans segmentation

Pour tarifer, il est raisonnable de choisir une constante  $c$  “*la plus proche*” de la variable aléatoire  $S$ , i.e.  $c$  qui minimise  $\mathbb{E}([S - c]^2)$ . Or

$$\mathbb{E}([S - c]^2) = \mathbb{E}([S - \mathbb{E}(S) + \mathbb{E}(S) - c]^2) = \mathbb{E}([S - \mathbb{E}(S)]^2) + \mathbb{E}([\mathbb{E}(S) - c]^2) + 0$$

i.e.

$$\mathbb{E}([S - c]^2) = [\mathbb{E}(S) - c]^2 + \text{quelque chose indépendant de } c.$$

donc  $\mathbb{E}([S - c]^2)$  est minimal pour  $c = \mathbb{E}(S)$ .

**Remarque :** cf méthode des moindres carrés dans le cours de prévision.

## L'excédant moyen de sinistre

L'excédant moyen de sinistre (encore appelé durée de vie moyenne restante en assurance sur la vie) est défini par

$$\begin{aligned}e_X(x) &= \mathbb{E}[X - x | X > x] \\ &= \frac{1}{\overline{F}_X(x)} \int_x^{+\infty} (s - x) dF_X(s), \quad x \geq 0.\end{aligned}$$

cf. section du cours sur la réassurance

## L'excédant moyen de sinistre

Loi de probabilité	$e_X(x)$
$\mathcal{E}(\lambda)$	$\frac{1}{\theta}$
$\mathcal{G}(\alpha, \beta)$	$\frac{\alpha}{\beta} \frac{1 - \Gamma(\alpha + 1, \beta x)}{1 - \Gamma(\alpha, \beta x)} - x$
$LN(\mu, \sigma^2)$	$\exp\left(\mu + \frac{\sigma^2}{2}\right) \frac{\bar{\Phi}\left(\frac{\ln(x) - \mu - \sigma^2}{\sigma}\right)}{\bar{\Phi}\left(\frac{\ln(x) - \mu}{\sigma}\right)} - x$
$\mathcal{P}(\alpha, x_0)$	$\frac{x_0 + x}{\alpha - 1}$

## Primes stop-loss (contrat avec franchise)

Etant donné un risque  $X$ , la **prime stop-loss** pour une franchise  $d \geq 0$  est définie par

$$\pi_X(d) = \mathbb{E}[(X - d)_+] = e_X(d)\bar{F}(d)..$$

Un traité de réassurance stop-loss (ou excédent de perte) consiste à faire prendre en charge par le réassureur la partie de la charge totale  $S$  des sinistres qui dépasse une certaine somme  $d$ . La portion réassurée, notée  $S^R$ , est donc définie par

$$S^R = (S - d)_+ = \begin{cases} 0, & \text{si } S \leq d, \\ S - d, & \text{si } S > d. \end{cases}$$

La prime pure que la cédante devra verser au réassureur pour un tel contrat, appelée prime stop-loss, est donnée par

$$\mathbb{E}[S^R] = \mathbb{E}[(S - d)_+].$$

**Définition 2.** *Etant donné un risque  $X$ , la prime stop-loss pour une rétention  $t \geq 0$  est définie par*

$$\pi_X(t) = \mathbb{E}[(X - t)_+].$$

*La fonction  $\pi_X$  est encore appelée la transformée stop-loss de la variable aléatoire  $X$ .*

**Proposition 3.** *La transformée stop-loss peut s'exprimer*

$$\pi_X(t) = \int_{x=t}^{+\infty} \bar{F}_X(x) dx.$$

**Proposition 4.** *Supposons  $\mathbb{E}[X] < +\infty$ . La transformée stop-loss  $\pi_X$  possède les propriétés suivantes :*

- (i) elle est décroissante et convexe.*
- (ii)  $\lim_{t \rightarrow +\infty} \pi_X(t) = 0$  et  $\lim_{t \rightarrow -\infty} \{\pi_X(t) + t\} = \mathbb{E}[X]$ .*

## Primes d'Esscher

**Définition 5.** *Etant donné un risque  $X$ , la prime d'Esscher de paramètre  $\alpha$  est définie par*

$$\pi_X(\alpha) = \frac{\mathbb{E}_{\mathbb{P}}(Xe^{\alpha X})}{\mathbb{E}_{\mathbb{P}}(e^{\alpha X})} = \mathbb{E}_{\mathbb{Q}}(X),$$

*où sous  $\mathbb{Q}$ , la fonction de répartition de  $X$  qui s'écrivait auparavant (sous  $\mathbb{P}$ )  $F_X(\cdot)$  devient ici*

$$G_X(\cdot) = \frac{e^{\alpha \cdot} dF_X(\cdot)}{\int_{\mathbb{R}} e^{\alpha t} dF_X(t)}.$$



## Hétérogénéité d'un portefeuille avec information parfaite

Considérons une assurance voyage, avec les indemnités

- pays A, \$ 250 avec probabilité 10%
- pays B, \$ 250 avec probabilité 20%

La dépense sera, conditionnelle à la destination

$$S_A = \begin{cases} 0, & \text{avec la probabilité } 0.9, \\ 250, & \text{avec la probabilité } 0.1 \end{cases}$$

et

$$S_B = \begin{cases} 0, & \text{avec la probabilité } 0.8, \\ 250, & \text{avec la probabilité } 0.2. \end{cases}$$

Si la destination n'est pas observée

$$S_{AB} = \begin{cases} 0, & \text{avec la probabilité } 0.85, \\ 250, & \text{avec la probabilité } 0.15. \end{cases}$$

## Hétérogénéité d'un portefeuille avec information parfaite

La prime pure

- sans segmentation sera  $\mathbb{E}(S_{AB}) = 37.5$
- avec segmentation sera  $\mathbb{E}(S_A) = 25$  à destination de A et  $\mathbb{E}(S_B) = 50$  à destination de B

S'il y a deux compagnies qui se font concurrence, une qui segmente, l'autre pas...

## Hétérogénéité et mélange de loi

Supposons qu'il existe un **effet aléatoire**  $\Theta$  représentant le niveau de risque (inconnu) d'un assuré pris au hasard dans le portefeuille. On suppose que  $X$  sachant  $\Theta = \theta$  admet pour loi  $F_\theta$ . Alors

$$\mathbb{P}(X \leq x) = \mathbb{E}[\mathbb{P}(X \leq x | \Theta)] = \int_{\Omega} F_\theta(x) d\Pi(\theta)$$

(moyenne des  $F_\theta$  pondérée par  $d\Pi$ ).

## Mélange de loi de Poisson et surdispersion

**Définition 6.** *La variable aléatoire de comptage  $N$  est de loi de Poisson mélange de moyenne  $\lambda$  et de niveau de risque relatif  $\Theta$  lorsque*

$$\begin{aligned}\mathbb{P}[N = k] &= \mathbb{E} \left[ \exp(-\lambda\Theta) \frac{(\lambda\Theta)^k}{k!} \right] \\ &= \int_0^{+\infty} \exp(-\lambda\theta) \frac{(\lambda\theta)^k}{k!} dF_{\Theta}(\theta), \quad k \in \mathbb{N},\end{aligned}\tag{1}$$

où  $F_{\Theta}$  est la fonction de répartition de  $\Theta$ , telle que  $\mathbb{E}[\Theta] = 1$ . On notera  $\mathcal{MPoi}(\lambda, F_{\Theta})$  ou  $\mathcal{MPoi}(\lambda, \Theta)$ .

## Mélange de loi de Poisson et surdispersion

Considérons l'exemple classique des bons et mauvais conducteur,

- pour les bons,  $N \sim \mathcal{Poi}(\lambda\theta_1)$
- pour les mauvais,  $N \sim \mathcal{Poi}(\lambda\theta_2)$

avec  $\theta_2 > 1 > \theta_1$ .

Si la proportion de bons risques est  $\varrho$ ,

$$\Theta = \begin{cases} \theta_1, & \text{avec une probabilité } \varrho, \\ \theta_2, & \text{avec une probabilité } 1 - \varrho, \end{cases}$$

où  $\theta_1$ ,  $\theta_2$  et  $\varrho$  sont contraints par

$$\mathbb{E}[\Theta] = \varrho\theta_1 + (1 - \varrho)\theta_2 = 1.$$

## Mélange de loi de Poisson et surdispersion

La probabilité qu'une police (dont on ne sait pas s'il s'agit d'un bon ou d'un mauvais risque) donne lieu à  $k$  sinistres durant la période de référence est alors de

$$\mathbb{P}[N = k] = \varrho \exp(-\lambda\theta_1) \frac{(\lambda\theta_1)^k}{k!} + (1 - \varrho) \exp(-\lambda\theta_2) \frac{(\lambda\theta_2)^k}{k!},$$

Si  $\Theta$  devient une variable aléatoire continue, de densité de probabilité  $f_{\Theta}$  alors

$$\mathbb{P}[N = k] = \int_0^{+\infty} \exp(-\lambda\theta) \frac{(\lambda\theta)^k}{k!} f_{\Theta}(\theta) d\theta, \quad k \in \mathbb{N}. \quad (2)$$

## Mélange de loi de Poisson et surdispersion

Soit  $N \sim \mathcal{MPoi}(\lambda, \Theta)$ , alors

$$\begin{aligned}\mathbb{E}[N] &= \int_0^{+\infty} \left( \sum_{k=0}^{+\infty} k \exp(-\lambda\theta) \frac{(\lambda\theta)^k}{k!} \right) dF_{\Theta}(\theta) \\ &= \lambda \int_0^{+\infty} \theta dF_{\Theta}(\theta) = \lambda.\end{aligned}$$

et

$$\begin{aligned}\text{Var}[N] &= \int_0^{+\infty} \left( \sum_{k=0}^{+\infty} k^2 \exp(-\lambda\theta) \frac{(\lambda\theta)^k}{k!} \right) dF_{\Theta}(\theta) - \lambda^2 \\ &= \int_0^{+\infty} (\lambda\theta + \lambda^2\theta^2) dF_{\Theta}(\theta) - \lambda^2 = \lambda + \lambda^2 \text{Var}[\Theta].\end{aligned}$$

Comme

$$\text{Var}[N] = \mathbb{E}[N] + \lambda^2 \text{Var}[\Theta] > \mathbb{E}[N]$$

pour autant que  $\Theta$  ne soit pas constant.

## Mélange de loi de Poisson et surdispersion

Tout mélange de Poisson implique donc une surdispersion des données. La fonction génératrice des probabilités de  $N \sim \mathcal{MPoi}(\lambda, \Theta)$  et la transformée de Laplace de  $\Theta$  sont liées par la formule

$$M_N(z) = \int_0^{+\infty} \exp(\lambda\theta(z-1)) f_{\Theta}(\theta) d\theta = L_{\Theta}(\lambda(1-z)). \quad (3)$$

Si nous considérons que  $\Theta \sim \mathcal{G}(\alpha, \alpha)$ ,

$$M_N(z) = \left(1 + \frac{\lambda(1-z)}{\alpha}\right)^{-\alpha},$$

qui est la fonction génératrice des probabilités associée à la loi  $BN(\alpha, \alpha/(\alpha + \lambda))$ .

Les mélanges de Poisson sont identifiables, i.e. si  $N_1 \sim \mathcal{MPoi}(\lambda, \Theta_1)$  et  $N_2 \sim \mathcal{MPoi}(\lambda, \Theta_2)$  alors

$$N_1 \stackrel{\mathcal{L}}{=} N_2 \Rightarrow \Theta_1 \stackrel{\mathcal{L}}{=} \Theta_2.$$



## Mélange de loi de Poisson et surdispersion

Les mélanges de Poisson jouissent d'une propriété trs importante établie par SHAKED (1980) et connue comme le "Shaked's Two Crossings Theorem".

**Proposition 7.** *Si  $N \sim \mathcal{MPoi}(\lambda, \Theta)$  alors il existe deux valeurs entières  $0 \leq k_0 < k_1$  telle que*

$$\mathbb{P}[N = k] \geq \exp(-\lambda) \frac{\lambda^k}{k!} \text{ pour } k = 0, 1, \dots, k_0,$$

$$\mathbb{P}[N = k] \leq \exp(-\lambda) \frac{\lambda^k}{k!} \text{ pour } k = k_0 + 1, \dots, k_1,$$

$$\mathbb{P}[N = k] \geq \exp(-\lambda) \frac{\lambda^k}{k!} \text{ pour } k \geq k_1 + 1.$$

## Mélange de loi de Poisson et surdispersion

```
> library(stats4); library(MASS)
> FREQUENCE=c(7840,1317,239,42,14,4,4,1)
> x=rep(0:7,FREQUENCE)
> table(x)
x
  0    1    2    3    4    5    6    7
7840 1317  239   42   14    4    4    1
> mean(x)
[1] 0.2143537
> var(x)
[1] 0.2889314
```

## Mélange de loi de Poisson et surdispersion

```
> fitdistr(x,"poisson")
  lambda
0.21435366
(0.00475989)
> fitdistr(x,"negative binomial")
  size      mu
0.701486138 0.214355746
(0.062786579) (0.005438638)
Messages d'avis :
1: In dnbinom_mu(x, size, mu, log) : production de NaN
2: In dnbinom_mu(x, size, mu, log) : production de NaN
3: In dnbinom_mu(x, size, mu, log) : production de NaN
> library(vcd)
> model.poisson=goodfit(x,type="poisson",method="ML")
```

## Mélange de loi de Poisson et surdispersion

```
> model.poisson
```

```
Observed and fitted values for poisson distribution  
with parameters estimated by 'ML'
```

count	observed	fitted
0	7840	7635.622
1	1317	1636.724
2	239	175.4188
3	42	12.53389
4	14	0.671671
5	4	0.028795
6	4	0.001028
7	1	0.000031

## Mélange de loi de Poisson et surdispersion

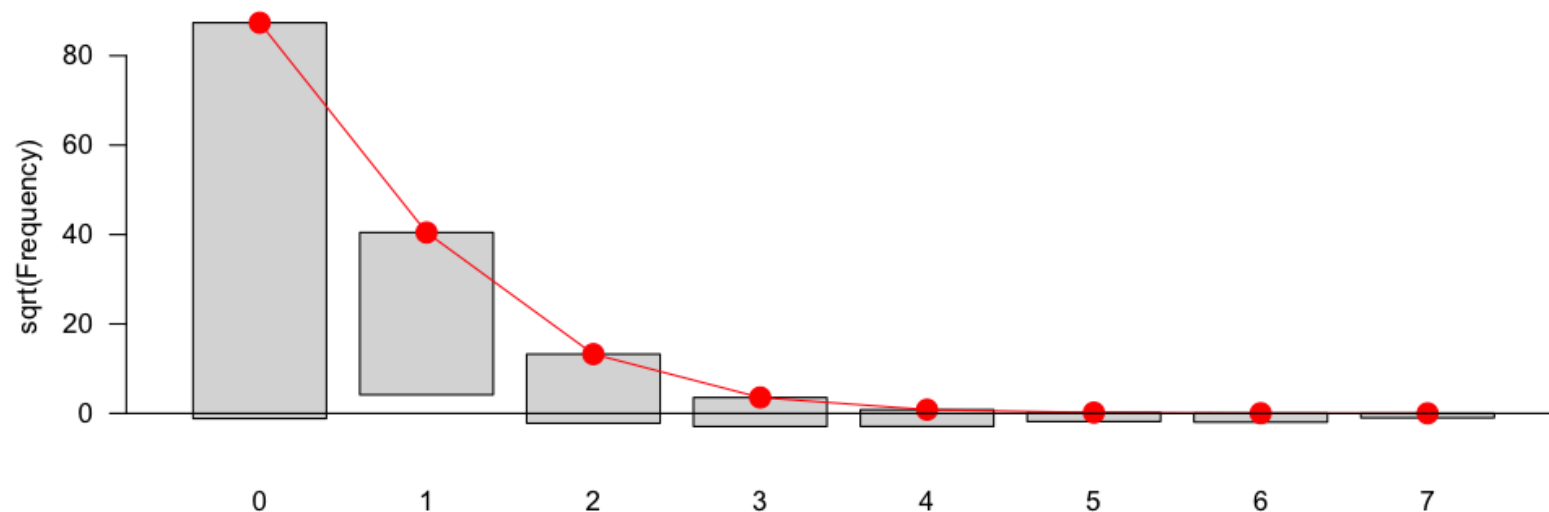
```
> model.poisson
```

```
> summary(model.poisson)
```

Goodness-of-fit test for poisson distribution

	X <sup>2</sup>	df	P(> X <sup>2</sup> )
Likelihood Ratio	302.484	6	2.401523e-62

```
> plot(model.poisson)
```



## Mélange de loi de Poisson et surdispersion

```
> model.nb=goodfit(x,type="nbinomial",method="ML")  
> model.nb
```

Observed and fitted values for nbinomial distribution  
with parameters estimated by 'ML'

count	observed	fitted
0	7840	7847.0055590
1	1317	1288.3613321
2	239	256.5374241
3	42	54.0688144
4	14	11.7105586
5	4	2.5772550
6	4	0.5732035
7	1	0.1284389

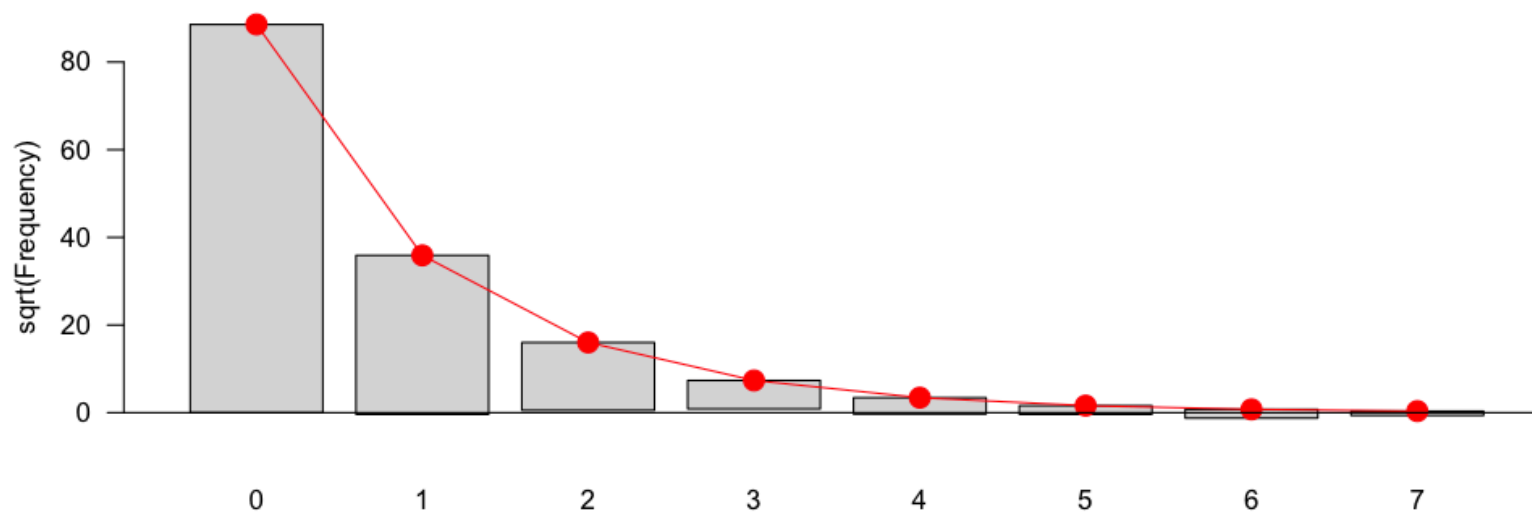
## Mélange de loi de Poisson et surdispersion

```
> summary(model.nb)
```

```
Goodness-of-fit test for nbinomial distribution
```

```
                X^2 df    P(> X^2)  
Likelihood Ratio 17.00285  5 0.00449439
```

```
> plot(model.nb)
```



## Segmentation en assurance

**Définition 8.** *On qualifie de segmentation toute technique que l'assureur utilise pour différencier la prime, et éventuellement aussi la couverture, en fonction d'un certain nombre de caractéristiques spécifiques du risque à assurer, et ce afin de parvenir à une meilleure concordance entre les coûts qu'une personne déterminée met à charge de la collectivité des preneurs d'assurance et la prime que cette personne doit payer pour la couverture offerte. Dans certains cas, cela peut impliquer que l'assureur refuse le risque à assurer.*



## Segmentation et espérance conditionnelle

Si  $N \sim \mathcal{MPoi}(\lambda, \Theta)$ , la fonction de répartition de  $\Theta$  sachant que  $N = n$ , notée  $F_{\Theta}(\cdot|n)$ ,

$$\begin{aligned}
 F_{\Theta}(t|n) &= \frac{\mathbb{P}[\Theta \leq t, N = n]}{\mathbb{P}[N = n]} \\
 &= \frac{\int_0^t \exp(-\lambda\theta) \frac{(\lambda\theta)^n}{n!} dF_{\Theta}(\theta)}{\int_{\mathbb{R}^+} \exp(-\lambda\theta) \frac{(\lambda\theta)^n}{n!} dF_{\Theta}(\theta)} \\
 &= \frac{\int_0^t \exp(-\lambda\theta) \theta^n dF_{\Theta}(\theta)}{\int_{\mathbb{R}^+} \exp(-\lambda\theta) \theta^n dF_{\Theta}(\theta)}.
 \end{aligned}$$

On peut alors en calculer l'espérance :

$$\mathbb{E}[\Theta|N = n] = \int_{\mathbb{R}^+} \theta dF_{\Theta}(\theta|n)$$

## Segmentation et espérance conditionnelle

qui donne finalement

$$\mathbb{E}[\Theta | N = n] = \frac{\int_0^t \exp(-\lambda\theta)\theta^{n+1} dF_{\Theta}(\theta)}{\int_{\theta \in \mathbb{R}^+} \exp(-\lambda\theta)\theta^n dF_{\Theta}(\theta)}.$$

## Segmentation et espérance conditionnelle

On vérifie aisément que l'espérance conditionnelle possède les propriétés suivantes : quelles que soient les variables aléatoires  $X_1$ ,  $X_2$  et  $X_3$ , et la constante réelle  $c$ ,

- (i)  $\mathbb{E}[c|X_1 = x_1] = c$  quel que soit  $x_1 \in \mathbb{R}$ .
- (ii)  $\mathbb{E}[X_1 + X_2|X_3 = x_3] = \mathbb{E}[X_1|X_3 = x_3] + \mathbb{E}[X_2|X_3 = x_3]$  quel que soit  $x_3 \in \mathbb{R}$ .
- (iii)  $\mathbb{E}[cX_1|X_2 = x_2] = c\mathbb{E}[X_1|X_2 = x_2]$ , quel que soit  $x_2 \in \mathbb{R}$ .
- (iv) quelle que soit la fonction  $g$ ,  $\mathbb{E}[g(X_1, X_2)|X_2 = x_2] = \mathbb{E}[g(X_1, x_2)|X_2 = x_2]$  quel que soit  $x_2 \in \mathbb{R}$ .
- (v) si  $X_1$  et  $X_2$  sont indépendantes alors  $\mathbb{E}[X_1|X_2 = x_2] = \mathbb{E}[X_1]$ .

## Segmentation et espérance conditionnelle

A moins que  $X_1$  et  $X_2$  ne soient indépendantes, la loi de  $X_1$  sachant  $X_2 = x_2$  dépend de  $x_2$ . En particulier,  $\mathbb{E}[X_1|X_2 = x_2]$  est une fonction de  $x_2$ , i.e.

$$\mathbb{E}[X_1|X_2 = x_2] = h(x_2).$$

On pourrait ainsi s'intéresser à la variable aléatoire

$$h(X_2) = \mathbb{E}[X_1|X_2].$$

**Proposition 9.** *La variable aléatoire  $\mathbb{E}[X_1|X_2]$  a mme moyenne que  $X_1$  :*

$$\mathbb{E}\left[\mathbb{E}[X_1|X_2]\right] = \mathbb{E}[X_1].$$

## Segmentation et espérance conditionnelle

*Démonstration.* Lorsque  $X_1$  et  $X_2$  sont continues, cette égalité s'établit comme suit :

$$\begin{aligned}
 \mathbb{E}\left[\mathbb{E}[X_1|X_2]\right] &= \int_{x_2 \in \mathbb{R}} \mathbb{E}[X_1|X_2 = x_2] f_2(x_2) dx_2 \\
 &= \int_{x_2 \in \mathbb{R}} \left\{ \int_{x_1 \in \mathbb{R}} x_1 f_{1|2}(x_1|x_2) dx_1 \right\} f_2(x_2) dx_2 \\
 &= \int_{x_2 \in \mathbb{R}} \int_{x_1 \in \mathbb{R}} x_1 f_{\mathbf{X}}(x_1, x_2) dx_1 dx_2 \\
 &= \int_{x_1 \in \mathbb{R}} x_1 f_1(x_1) dx_1 = \mathbb{E}[X_1].
 \end{aligned}$$

Le raisonnement est similaire dans les cas discret et mixte. □

## Segmentation et espérance conditionnelle

Épinglons à présent une caractéristique remarquable de l'espérance conditionnelle, qui peut être prise comme définition générale de ce concept.

**Proposition 10.** *Quelle que soit la fonction  $h : \mathbb{R} \rightarrow \mathbb{R}$ , nous avons*

$$\mathbb{E} \left[ h(X_2) \{ X_1 - \mathbb{E}[X_1 | X_2] \} \right] = 0.$$

*Démonstration.* La Propriété 9 nous permet d'écrire

$$\begin{aligned} & \mathbb{E} \left[ h(X_2) \{ X_1 - \mathbb{E}[X_1 | X_2] \} \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ h(X_2) \{ X_1 - \mathbb{E}[X_1 | X_2] \} \mid X_2 \right] \right] \\ &= \mathbb{E} \left[ h(X_2) \mathbb{E} \left[ \{ X_1 - \mathbb{E}[X_1 | X_2] \} \mid X_2 \right] \right] = 0, \end{aligned}$$

ce qui achève la vérification du résultat annoncé. □

On peut voir  $X_1 - \mathbb{E}[X_1|X_2]$  comme un résidu (c'est-à-dire comme la partie de  $X_1$  que ne parvient pas à expliquer  $X_2$ ).

Ceci garantit que  $\mathbb{E}[X_1|X_2]$  est le meilleur prédicteur de  $X_1$  au sens des moindres carrés, comme l'indique le résultat suivant.

**Proposition 11.** *La variable aléatoire  $h^*(X_2) = \mathbb{E}[X_1|X_2]$  est celle minimisant  $\mathbb{E}[(X_1 - h(X_2))^2]$  sur toutes les fonctions  $h$ .*

*Démonstration.* Ecrivons

$$\begin{aligned} \mathbb{E}[(X_1 - h(X_2))^2] &= \mathbb{E}[(X_1 - \mathbb{E}[X_1|X_2] + \mathbb{E}[X_1|X_2] - h(X_2))^2] \\ &= \underbrace{\mathbb{E}[(X_1 - \mathbb{E}[X_1|X_2])^2]}_{\text{indépendant de } h} + \mathbb{E}[(\mathbb{E}[X_1|X_2] - h(X_2))^2] \\ &\quad + 2 \underbrace{\mathbb{E}[(X_1 - \mathbb{E}[X_1|X_2])(\mathbb{E}[X_1|X_2] - h(X_2))]}_{=0 \text{ par définition de l'espérance conditionnelle}} \end{aligned}$$

sera minimum lorsque  $h(X_2) = \mathbb{E}[X_1|X_2]$ . □

## Segmentation et espérance conditionnelle

Il n'y a bien entendu aucune raison de se limiter à une seule variable conditionnante, et on peut considérer un vecteur  $\mathbf{X}$  de dimension  $n$  et définir  $\mathbb{E}[X_1|X_2, \dots, X_n]$ .

**Définition 12.** *Considérons un vecteur aléatoire  $\mathbf{X}$  de dimension  $n$ .*

*L'espérance conditionnelle  $\mathbb{E}[X_1|X_2, \dots, X_n]$  de  $X_1$  sachant  $X_2, \dots, X_n$  est la variable aléatoire  $h^*(X_2, \dots, X_n)$  telle que l'égalité*

$$\mathbb{E}[h(X_2, \dots, X_n)\{X_1 - h^*(X_2, \dots, X_n)\}] = 0, \quad (4)$$

*est vérifiée pour toute fonction  $h : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ .*



## Transfert de risque sans segmentation

Tentons à présent de formaliser le problème de la segmentation. Considérons un assuré soumis à un risque  $S$ , prélevé au hasard au sein d'un portefeuille d'assurance. Supposons que toutes les caractéristiques de l'assuré influençant le risque soient reprises dans un vecteur aléatoire  $\Omega = (\Omega_1, \Omega_2, \dots)$ . La notation "oméga" rappelle que le vecteur du même nom contient toute l'information à propos de l'assuré, qu'elle soit observable ou non par l'assureur.

On peut imaginer que l'assureur ne tienne compte en aucune manière des caractéristiques  $\Omega$  de l'assuré, et lui réclame donc une prime pure de montant  $\mathbb{E}[S]$ , la même que celle qu'il réclame à tous les assurés du portefeuille. Dans ce cas, la situation est telle que présentée au Tableau 1.

## Transfert de risque sans segmentation

	Assurés	Assureur
Dépense	$\mathbb{E}[S]$	$S - \mathbb{E}[S]$
Dépense moyenne	$\mathbb{E}[S]$	0
Variance	0	$\text{Var}[S]$

TABLE 1 – Situation des assurés et de l'assureur en l'absence de segmentation.

L'assureur prend donc l'entièreté de la variance des sinistres  $\text{Var}[S]$  à sa charge, que celle-ci soit due à l'hétérogénéité du portefeuille, ou à la variabilité intrinsèque des montants des sinistres.

## Segmentation avec information parfaite

A l'autre extrême, supposons que l'assureur incorpore toute l'information  $\Omega$  dans la tarification. On serait alors dans la situation décrite au Tableau 2.

	Assurés	Assureur
Dépense	$\mathbb{E}[S \Omega]$	$S - \mathbb{E}[S \Omega]$
Dépense moyenne	$\mathbb{E}[S]$	0
Variance	$\text{Var}[\mathbb{E}[S \Omega]]$	$\text{Var}[S - \mathbb{E}[S \Omega]]$

TABLE 2 – Situation des assurés et de l'assureur dans le cas où la segmentation est opérée sur base de  $\Omega$ .

Contrairement au cas précédent, la prime payée par un assuré prélevé au hasard dans le portefeuille est à présent une variable aléatoire :  $\mathbb{E}[S|\Omega]$  dépend des caractéristiques  $\Omega$  de cet assuré.

## Segmentation avec information parfaite

Comme la variable aléatoire  $S - \mathbb{E}[S|\Omega]$  est centrée, le risque assumé par l'assureur la variance du résultat financier de l'opération d'assurance, i.e.

$$\begin{aligned} \text{Var}\left[S - \mathbb{E}[S|\Omega]\right] &= \mathbb{E}\left[\left(S - \mathbb{E}[S|\Omega]\right)^2\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left(S - \mathbb{E}[S|\Omega]\right)^2 \mid \Omega\right]\right] \\ &= \mathbb{E}\left[\text{Var}[S|\Omega]\right]. \end{aligned}$$

On assiste dans ce cas à un partage de la variance totale de  $S$  (c'est-à-dire du risque) entre les assurés et l'assureur, matérialisé par la formule

$$\text{Var}[S] = \underbrace{\mathbb{E}\left[\text{Var}[S|\Omega]\right]}_{\rightarrow \text{assureur}} + \underbrace{\text{Var}\left[\mathbb{E}[S|\Omega]\right]}_{\rightarrow \text{assurés}}.$$

## Segmentation avec information parfaite

Ainsi, lorsque toutes les variables pertinentes  $\Omega$  ont été prises en compte, l'intervention de l'assureur se limite à la part des sinistres due exclusivement au hasard ; en effet,  $\text{Var}[S|\Omega]$  représente les fluctuations de  $S$  dues au seul hasard. Dans cette situation idéale, l'assureur mutualise le risque et il n'y a donc aucune solidarité induite entre les assurés du portefeuille : chacun paie en fonction de son propre risque.

## Segmentation avec information imparfaite

Bien entendu, la situation décrite au paragraphe précédent est purement théorique puisque parmi les variables explicatives  $\Omega$  nombreuses sont celles qui ne peuvent pas être observées par l'assureur. En assurance automobile par exemple, l'assureur ne peut pas observer la vitesse à laquelle roule l'assuré, son agressivité au volant, ni le nombre de kilomètres qu'il parcourt chaque année.

Dès lors, l'assureur ne peut utiliser qu'un sous-ensemble  $X$  des variables explicatives contenues dans  $\Omega$ , i.e.  $X \subset \Omega$ . La situation est alors semblable à celle décrite au Tableau 3.

## Segmentation avec information imparfaite

	Assuré	Assureur
Dépense	$\mathbb{E}[S \mathbf{X}]$	$S - \mathbb{E}[S \mathbf{X}]$
Dépense moyenne	$\mathbb{E}[S]$	0
Variance	$\text{Var}[\mathbb{E}[S \mathbf{X}]]$	$\mathbb{E}[\text{Var}[S \mathbf{X}]]$

TABLE 3 – Situation de l'assuré et de l'assureur dans le cas où la segmentation est opérée sur base de  $\mathbf{X} \subset \Omega$ .

Il est intéressant de constater que

$$\begin{aligned}
 \mathbb{E}[\text{Var}[S|\mathbf{X}]] &= \mathbb{E}[\mathbb{E}[\text{Var}[S|\Omega]|\mathbf{X}]] + \mathbb{E}[\text{Var}[\mathbb{E}[S|\Omega]|\mathbf{X}]] \\
 &= \underbrace{\mathbb{E}[\text{Var}[S|\Omega]]}_{\text{mutualisation}} + \underbrace{\mathbb{E}\{\text{Var}[\mathbb{E}[S|\Omega]|\mathbf{X}]\}}_{\text{solidarité}}. \tag{5}
 \end{aligned}$$

## Segmentation avec information imparfaite

On peut interpréter cette décomposition du risque pris en charge par la compagnie comme suit : l'assureur, lorsque tous les facteurs de risque ne sont pas pris en compte, intervient pour réparer les conséquences fâcheuses du hasard (premier terme de (5) traduisant la mutualisation du risque), mais doit aussi prendre en charge les variations de la prime pure exacte  $\mathbb{E}[S|\Omega]$  qui ne sont pas expliquées par les facteurs de risque  $\mathbf{X}$  intégrés au tarif (second terme de (5), traduisant la solidarité induite par une personnalisation imparfaite du montant de la prime). En d'autres mots, en plus de contrer les mauvais coups du sort, l'assureur doit également supporter la variabilité des sinistres due aux caractéristiques des assurés, non prises en compte par le tarif.



## Segmentation avec information imparfaite

Dans une tarification segmentée en fonction de  $\mathbf{X} \subset \Omega$ , le partage de la variance de  $S$  s'effectue comme suit :

$$\begin{aligned}
 \text{Var}[S] &= \mathbb{E} \left[ \text{Var}[S|\mathbf{X}] \right] + \text{Var} \left[ \mathbb{E}[S|\mathbf{X}] \right] \\
 &= \underbrace{\mathbb{E} \left[ \text{Var}[S|\Omega] \right]}_{\text{mutualisation}} + \underbrace{\mathbb{E} \left[ \text{Var} \left[ \mathbb{E}[S|\Omega] \mid \mathbf{X} \right] \right]}_{\text{solidarité}} \\
 &\quad \xrightarrow{\text{assureur}} \\
 &+ \underbrace{\text{Var} \left[ \mathbb{E}[S|\mathbf{X}] \right]}_{\text{assurés}} .
 \end{aligned}$$

## *Tarifications a priori et a posteriori*

Toute l'idée qui sous-tend la tarification d'expérience est que l'historique des sinistres révèle les caractéristiques non observables des assurés. Plus précisément, si on note  $\mathbf{S}$  les informations concernant la sinistralité passée des assurés disponibles pour l'assureur, l'information contenue dans  $(\mathbf{X}, \mathbf{S})$  devient comparable à  $\Omega$ , tant et si bien que  $\mathbb{E}[S|\mathbf{X}, \mathbf{S}]$  devrait converger vers  $\mathbb{E}[S|\Omega]$  (dans un sens à préciser).

## Rappels sur l'inférence statistique

En statistique, on dispose d'observations  $\{y_1, \dots, y_n\}$ , supposées être de réalisations de variables aléatoires i.i.d.  $\{Y_1, \dots, Y_n\}$ . On suppose que les  $Y_i$  sont de loi  $F_\star$ , inconnue, mais que

$$F_\star \in \mathcal{F} = \{F_\theta, \theta \in \Theta\}$$

Aussi,  $F_\star = F_{\theta_\star}$ . On cherche alors à **estimer** le paramètre  $\theta$  de la loi.

Un **estimateur**  $\hat{\theta}$  est une fonction des observations,  $\hat{\theta} = g(y_1, \dots, y_n)$ . Chercher les propriétés d'un estimateur consiste à étudier les propriétés de la variable aléatoire  $\hat{\theta} = g(Y_1, \dots, Y_n)$ .

Par exemple,  $\hat{\theta}$  est un estimateur sans biais de  $\theta$  si

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}[g(Y_1, \dots, Y_n)] = \theta$$

lorsque les variables  $Y_i$  sont i.i.d. de loi  $F_\theta$ .

## Rappels sur l'inférence statistique

La précision d'un estimateur se juge souvent à l'aide de l'erreur quadratique moyenne (*mean squared error*),

$$\text{mse}(\hat{\theta}) = \mathbb{E} \left( (\hat{\theta} - \theta)^2 \right).$$

Un estimateur sera dit **convergent** si  $\hat{\theta}_n = g(Y_1, \dots, Y_n)$  converge (en probabilité) vers  $\theta$  lorsque  $n \rightarrow \infty$ , i.e.

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta}_n - \theta| > \epsilon) = 0, \quad \forall \epsilon > 0.$$

Un estimateur sera dit **asymptotiquement Gaussien** si (à une transformation affine près)  $\hat{\theta}_n = g(Y_1, \dots, Y_n)$  converge (en loi) vers une variable Gaussienne

$$\sqrt{n}[\hat{\theta}_n - \theta] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \star).$$

## Le maximum de vraisemblance

La méthode la plus classique pour estimer  $\theta$  est la méthode du **maximum de vraisemblance**. On appelle **vraisemblance** la fonction

$$\mathcal{L}(\theta; \mathbf{y}) = \prod_{i=1}^n f_{\theta}(y_i)$$

et on peut définir la log-vraisemblance comme

$$\log \mathcal{L}(\theta; \mathbf{y}) = \sum_{i=1}^n \log f_{\theta}(y_i)$$

Le maximum de vraisemblance (s'il existe) est alors

$$\hat{\theta} = \operatorname{argmax} \{ \mathcal{L}(\theta; \mathbf{x}), \theta \in \Theta \} = \operatorname{argmax} \{ \log \mathcal{L}(\theta; \mathbf{x}), \theta \in \Theta \}$$

## Le maximum de vraisemblance

Ce problème d'optimisation se résout (le plus souvent) en résolvant la condition du premier ordre,

$$\left. \frac{\partial \log \mathcal{L}(\theta; \mathbf{y})}{\partial \theta} \right|_{\theta=\hat{\theta}} = \sum_{i=1}^n \left. \frac{\partial \log f_{\theta}(y_i)}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0$$

On notera que (sous certaines conditions)

$$\sqrt{n}[\hat{\theta}_n - \theta] \xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}(\theta)),$$

où  $I(\theta)$  est l'information de Fisher, i.e.

$$I(\theta) = \mathbb{E} \left[ \left( \frac{\partial \log f_{\theta}(X)}{\partial \theta} \right)^2 \right] \text{ où } X \sim F_{\theta}.$$

## Maximum de vraisemblance et tests

**Test de Wald** l'idée est d'étudier la différence entre  $\hat{\theta}$  et  $\theta_*$ . Sous  $H_0 : \theta = \theta_*$ ,

$$T = n \frac{(\hat{\theta} - \theta_*)^2}{I^{-1}(\theta_*)} \xrightarrow{\mathcal{L}} \chi^2(1)$$

**Test du rapport de vraisemblance** l'idée est d'étudier la différence entre  $\log \mathcal{L}(\hat{\theta})$  et  $\log \mathcal{L}(\theta_*)$ . Sous  $H_0$ ,

$$T = 2 \log \left( \frac{\log \mathcal{L}(\theta_*)}{\log \mathcal{L}(\hat{\theta})} \right) \xrightarrow{\mathcal{L}} \chi^2(1)$$

**Test du score** l'idée est d'étudier la différence entre  $\frac{\partial \log \mathcal{L}(\theta_*)}{\partial \theta}$  et 0. Sous  $H_0$ ,

$$T = \left( \frac{1}{n} \sum_{i=1}^n \frac{\partial \log f_{\theta_*}(y_i)}{\partial \pi} \right)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$$

## Mise en oeuvre pratique

Considérons l'échantillon suivant,  $\{y_1, \dots, y_n\}$

```
> set.seed(1)
> n=20
> (Y=sample(0:1,size=n,replace=TRUE))
[1] 0 0 1 1 0 1 1 1 1 0 0 0 1 0 1 0 1 1 0 1
```

Comme  $Y_i \sim \mathcal{B}(\pi)$ , avec  $\pi = \mathbb{E}(Y)$ , l'estimateur de la méthode des moments est  $\hat{\pi} = \bar{y}$ , soit ici

```
> mean(X)
[1] 0.55
```

On peut faire un test de  $H_0 : \pi = \pi_*$  contre  $H_1 : \pi \neq \pi_*$  (par exemple 50%)

On peut utiliser le test de Student,

$$T = \sqrt{n} \frac{\hat{\pi} - \pi_*}{\sqrt{\pi_*(1 - \pi_*)}}$$



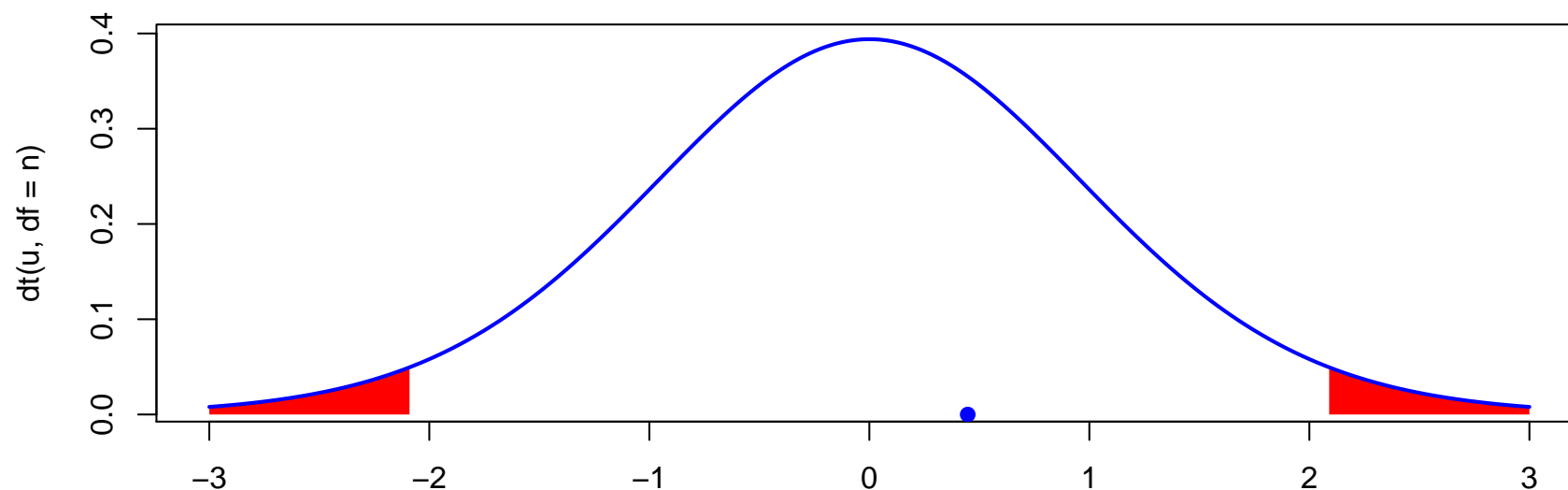
qui suit, sous  $H_0$  une loi de Student à  $n$  degrés de liberté.

```
> (T=sqrt(n)*(pn-p0)/(sqrt(p0*(1-p0))))
```

```
[1] 0.4472136
```

```
> abs(T)<qt(1-alpha/2,df=n)
```

```
[1] TRUE
```

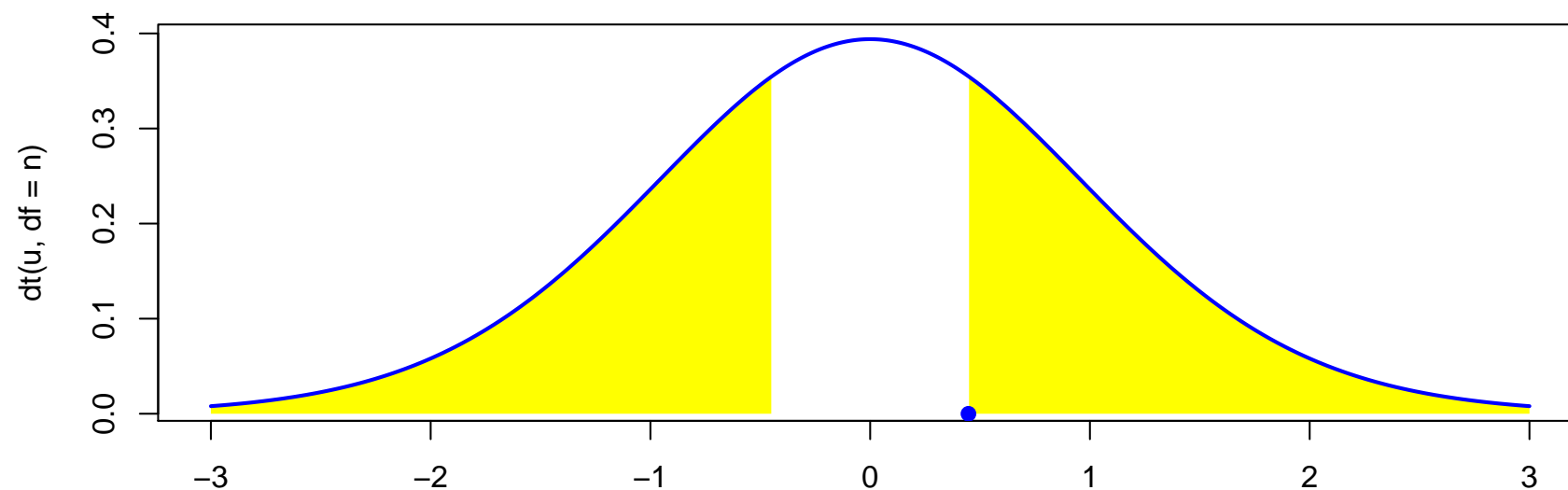


On est ici dans la région d'acceptation du test.

On peut aussi calculer la  $p$ -value,  $\mathbb{P}(|T| > |t_{obs}|)$ ,

```
> 2*(1-pt(abs(T),df=n))
```

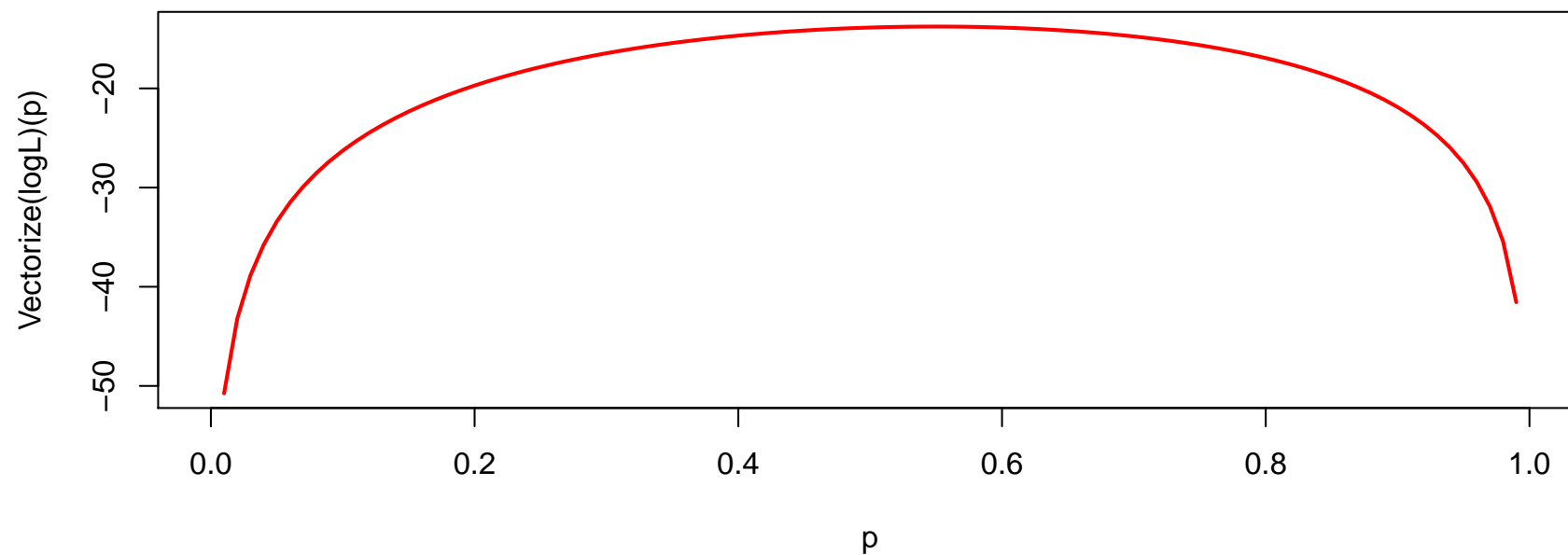
```
[1] 0.6595265
```



## Mise en oeuvre pratique

Pour l'estimateur de maximum de vraisemblance, on peut faire les calculs, ou le faire numériquement. On commence par tracer la fonction  $\theta \mapsto \log \mathcal{L}(\theta)$

```
> p=seq(0,1,by=.01)
> logL=function(p){sum(log(dbinom(X,size=1,prob=p)))}
> plot(p,Vectorize(logL)(p),type="l",col="red",lwd=2)
```



On peut trouver numériquement le maximum de  $\log \mathcal{L}$ ,

```
> neglogL=function(p){-sum(log(dbinom(X,size=1,prob=p)))}  
> pml=optim(fn=neglogL,par=p0,method="BFGS")  
> pml  
$par  
[1] 0.5499996
```

```
$value
```

```
[1] 13.76278
```

i.e. on retrouve ici  $\hat{\pi} = \bar{y}$ .

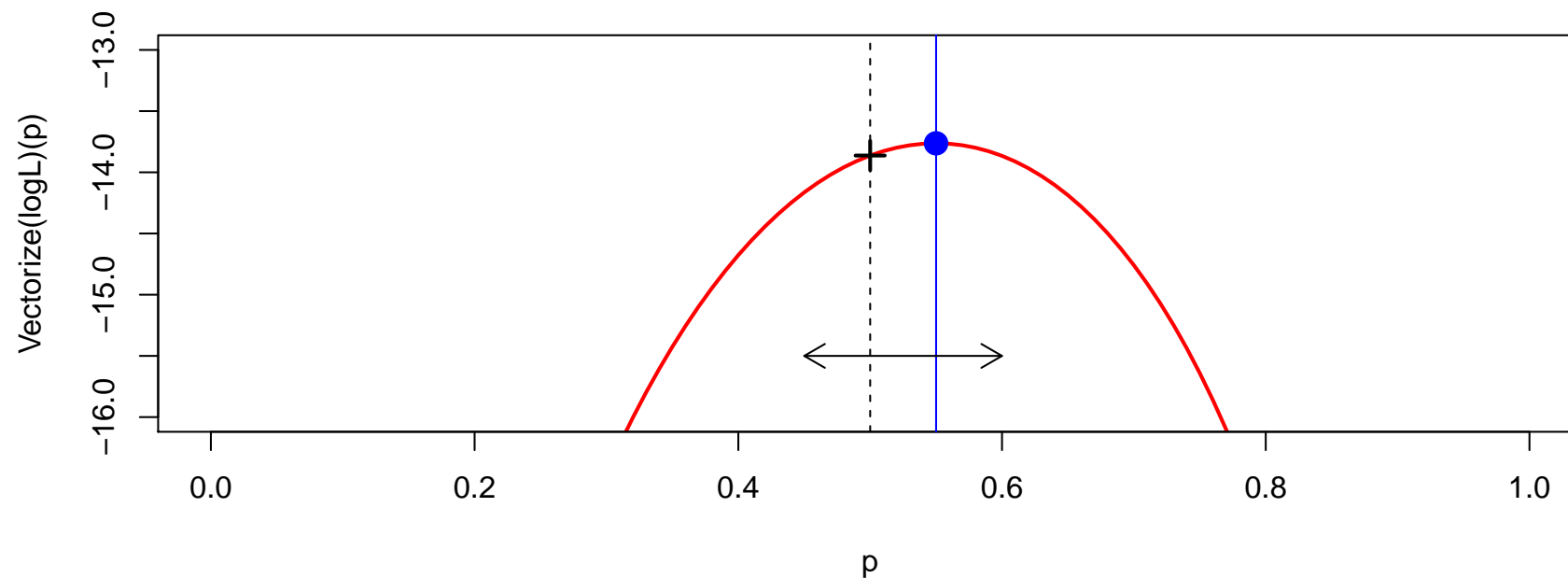
Maintenant, on peut tester  $H_0 : \pi = \pi_\star = 50\%$  versus  $H_1 : \pi \neq 50\%$ . Pour le test de Wald, on peut commencer par calculer  $nI(\theta_\star)$  ou ,

```
> nx=sum(X==1)
> f = expression(nx*log(p)+(n-nx)*log(1-p))
> Df = D(f, "p")
> Df2 = D(Df, "p")
> p=p0=0.5
> (IF=-eval(Df2))
[1] 80
```

On peut d'ailleurs comparer cette valeur avec la valeur théorique, car

$$I(\pi)^{-1} = \pi(1 - \pi)$$

```
> 1/(p0*(1-p0)/n)
[1] 80
```



La statistique du test de Wald est ici

```
> pml=optim(fn=neglogL,par=p0,method="BFGS")$par  
> (T=(pml-p0)^2*IF)  
[1] 0.199997
```

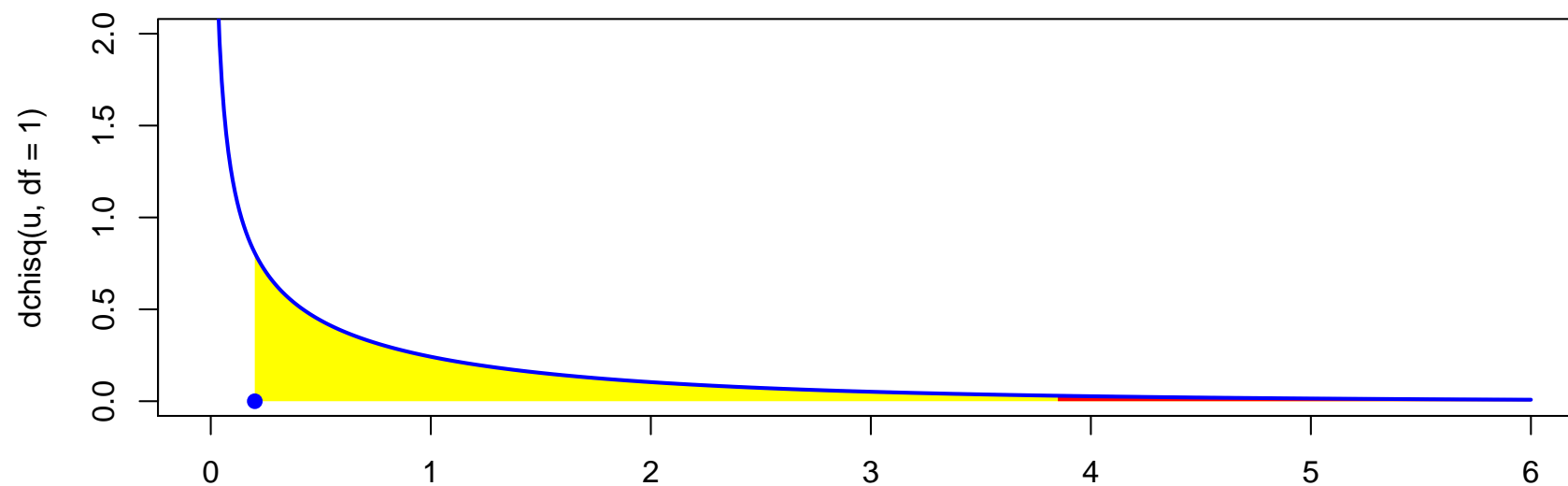
que l'on va chercher à comparer avec le quantile de la loi du  $\chi^2$ ,

```
> T<qchisq(1-alpha,df=1)
[1] TRUE
```

i.e. on est dans la région d'acceptation du test. De manière duale, on peut calculer la  $p$ -value du test,

```
> 1-pchisq(T,df=1)
[1] 0.6547233
```

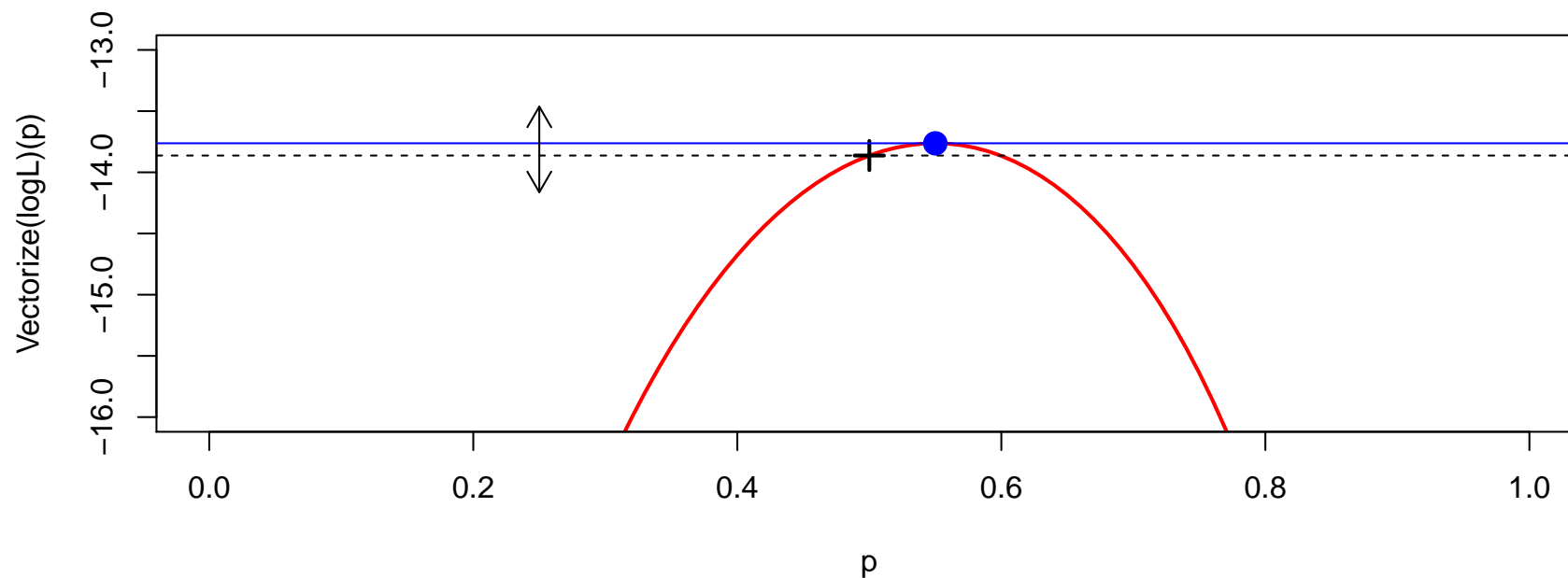
Donc on va accepter  $H_0$ .



Pour le test du rapport de vraisemblance,  $T$  est ici

```
> (T=2*(logL(pml)-logL(p0)))  
[1] 0.2003347
```





Là encore, on est dans la région d'acceptation,

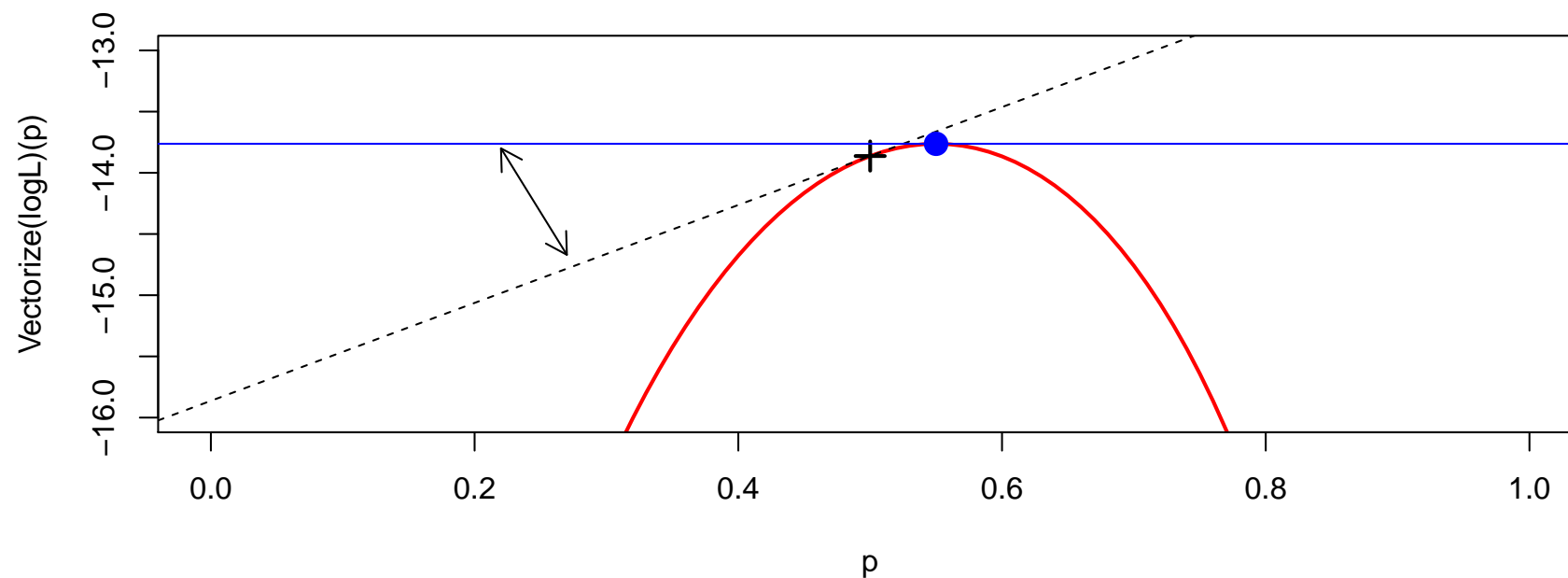
```
> T<qchisq(1-alpha,df=1)
[1] TRUE
```

Enfin, on peut faire le test du score. Le score se calcule facilement,

```
> nx=sum(X==1)
> f = expression(nx*log(p)+(n-nx)*log(1-p))
> Df = D(f, "p")
> p=p0
> score=eval(Df)
```

La statistique de test est alors

```
> (T=score^2/IF)
[1] 0.2
```



Là encore, on est dans la région d'acceptation,

```
> T<qchisq(1-alpha,df=1)
[1] TRUE
```