Cours Statistique Bayésienne

Dhafer Malouche

3ième année ESSAI, 2015-2016

Introduction au calcul Bayésien

(Séance 1, 3/9/2016)

Théorème de Bayes

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soient X et Y deux évènements de $\mathcal{A}$. Alors

$$\mathbb{P}\left(X \mid Y\right) = \frac{\mathbb{P}(X \cap Y)}{\mathbb{P}(Y)}$$

Exercice

On considère un test pour une infection. On suppose que la sensitivité du test est égale à 99% (la probabilité qu'un infecté soit détécté positif) et que la spécificité du test est égale à 97% (la probabilité qu'une personne saine soit détéctée négative). Sachant que la proportion des personnes infectées dans une population est égale à 1%.

- Question 1 Quelle est la propbabilité qu'une personne ayant eu un résultat positif au test soit détectée infectée.
- Question 2 Cette personne passe ce test une deuxième fois et il est encore positif. Quelle est la la propbabilité qu'elle soit infectée.

Solution avec `R`

Question 1

- On simule une population avec un taux d'infection égal à 1%.

```
> N=10000 # Taille de la population
> p=0.01 # Probabilité d'infection
> population=rbinom(N,1,prob = p)
> N_infected=sum((population==1));N_infected ## Nombre d'infectés
[1] 97
> infected=which(population==1)
```

- On simule le test

```
> sensitivity=.99
> specificity=.97
> test=rbinom(N,1,(population==1)*sensitivity+(population==0)*(1-specificity))
> N_test=sum((test==1));N_test ## Nombre de positifs.
[1] 380
```

- La matrice de confusion `test x population`

```
> confusion=xtabs(~population+test);confusion
          test
population    0    1
         0 9620  283
         1    0   97
```

Donc la probabilité qu'une personne ayant eu un résultat positif au test soit détectée infectée se calcule de la façon suivante

```
> confusion[2,2]/(confusion[1,2]+confusion[2,2])
[1] 0.2552632
```

Question 2

- Simulant une deuxième fois le test

```
> test2=rbinom(N,1,(population==1)*sensitivity+(population==0)*(1-specificity))
```

- Consideérons le croisement entre les trois variables `population` , `test` et `test2`

```
> confusion2=xtabs(~population+test2+test);confusion2
, , test = 0

          test2
population    0    1
         0 9329  291
         1    0    0

, , test = 1

          test2
population    0    1
         0  271   12
         1    0   97
```

- Donc la probabilité qu'une personne infectée sachant qu'elle a réagit positivement deux fois au test est

```
> confusion2[2,2,2]/(confusion2[2,2,2]+confusion2[2,1,2])
[1] 1
```

Solution théorique

On note par M l'événement qu'une personne soit infecetée et par P qu'elle soit détectée positive. Alors

$$\text{sensitivité} = \mathbb{P}(\mathrm{P} \mid \mathrm{M}) = \frac{\mathbb{P}(\mathrm{P} \cap \mathrm{M})}{\mathbb{P}(\mathrm{M})} = .99$$

et

$$\text{spécificité} = \mathbb{P}(\overline{\mathrm{P}} \mid \overline{\mathrm{M}}) = \frac{\mathbb{P}(\mathrm{P} \cap \mathrm{M})}{\mathbb{P}(\overline{\mathrm{M}})} = .97$$

Donc la la probabilité qu'une personne ayant eu un résultat positif au test soit détectée infectée se calcule

$$\mathbb{P}(\mathrm{M} \mid \mathrm{P}) = \frac{\mathbb{P}(\mathrm{P} \cap \mathrm{M})}{\mathbb{P}(\mathrm{P})} = \frac{\mathbb{P}(\mathrm{P} \mid \mathrm{M})\mathbb{P}(\mathrm{M})}{\mathbb{P}(\mathrm{P} \mid \overline{\mathrm{M}})\mathbb{P}(\overline{\mathrm{M}}) + \mathbb{P}(\mathrm{P} \mid \mathrm{M})\mathbb{P}(\mathrm{M})} = \frac{.99 \times .01}{(1 - .97) \times (1 - .01) + .99 \times .01} = 0.25$$

Le modèle Bayésien et théorie de la décision

(Séance 2, 10/9/2015)

Notions de la théorie de décision.

Soit $\mathcal{M}$ un modèle paramétrique : Soit $\underline{\mathrm{X}} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ un n−échantillon de loi $\mathbb{P}$ dépendant d'un parameètre$^?$ $\in \Theta \subseteq \mathbb{R}^p$.

- Rappelons un estimateur de θ est une fonction du n−échantillon $\underline{X}$:

$$\hat{\theta}(\underline{X}) \in \mathcal{E} \supseteq \Theta$$

- Comment évaluer la qualité d'un estimateur $\hat{\theta}$?
- Fonction de Perte :

$$L\ :\ \Theta \times \mathcal{E} \longrightarrow \mathbb{R}^{+}$$

telle que $L(\theta, \theta) = 0$ pour tout $\theta \in \Theta$. Donc l'erreur de l'estimation de $\hat{\theta}$ est mesurée par $L(\theta, \hat{\theta})$.

- Fonction de Risque

$$\theta \longmapsto R(\theta, \hat{\theta}) = \mathbb{E}\left[L(\theta, \hat{\theta})\right]$$

Exemples : $\Theta = \mathcal{E} = \mathbb{R}$

Perte quadratique

$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$ et le risque associée est le risque quadratique

$$R(\theta, \hat{\theta}) = \mathbb{E}\left[(\theta - \hat{\theta})^2\right]$$

Perte absolue

$L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$ et le risque associée est le risque quadratique

$$R(\theta, \hat{\theta}) = \mathbb{E}\left[|\theta - \hat{\theta}|\right]$$

Intuitivement plus le risque de l'estimateur est faible plus l'estimateur $\hat{\theta}$ est considéré comme performant.

Exercice

Soit $\underline{X} = (X_1, \ldots, X_n)$ telle que pour tout $i = 1, \ldots, n$, $X_i \sim \mathcal{N}(\theta, 1)$. On considère $\overline{X}$ comme estimateur de θ. Calculer le risque quadratique de $\overline{X}$.

Le cas bayésien.

Risque Bayésien

Dans le cas bayésien on suppose qu'on dispose d'une certaine information a priori sur le paramètre θ qu'on la traduit avec une distribution de probabilité θ noteée $\pi(\theta)$.

Notre objectif est de mettre aà jour cette information une fois qu'on a observé l'échantillon des données.

Si on suppose que $\underline{X} = (X_1, \ldots, X_n)$ est issu d'une loi de probabilité de densiteé $f(x \mid \theta)$. Alors la fonction vraisemblance

$$L(\theta, \underline{x}) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

On calcule alors ce qu'on appelle la loi a posteriori de θ :

$$\hat{\pi}(\theta \mid \underline{x}) \propto L(\theta, \underline{x})\, \pi(\theta)$$

Exemples

Calculer la loi a posteriori dans les cas suivants

1. f est la loi Bernoulli(θ) et $\theta \sim \mathrm{Beta}(a, b)$.
2. f est la loi $\mathcal{N}(\theta, 1)$ et $\theta \sim \mathcal{N}(0, 1)$.

- Risque Bayésien On consideère une fonction $L(\cdot;\cdot)$ et modèle statistique paramètrique $\mathcal{M} = \{\mathcal{X}, \mathcal{A}, \mathbb{P}_\theta;\ \theta \in \Theta\}$. On consideère une loi a priori π sur le paramètre θ. Le risque bayésien d'une estimateur $\hat{\theta}(\underline{x})$

$$\mathcal{R}_\pi\left(\hat{\theta}(\underline{x})\right) = \int_\Theta R\left(\theta, \hat{\theta}(\underline{x})\right) \pi(d\theta)$$

Estimateur Bayésien

- L'estimateur $\hat{\theta}^B(\underline{X})$ est appelé estimateur de Bayes associé à la fonction de perte L et à l'a priori π c'est l'estimateur qui minimise le risque bayésien

$$\mathcal{R}_\pi\left(\hat{\theta}^B(\underline{X})\right) = \min_{\hat{\theta}(\underline{x}) \in \mathcal{E}} \mathcal{R}_\pi\left(\hat{\theta}(\underline{x})\right)$$

- Remarque : L'expression de l'estimateur bayeésien dépend du choix de la loi a priori π et de la fonction perte L.

- Théorème : Supposons que
  1. Il existe un estimateur $\hat{\theta}_0(\underline{X})$ tel que $\mathcal{R}_\pi\left(\hat{\theta}_0(\underline{X})\right) < \infty$.
  2. Pour presque tout $x \in \mathcal{X}$ il existe

$$\hat{\theta}(\underline{x}) = \mathrm{argmin}_{y \in \mathcal{E}} \mathbb{E}\left[L(\theta, y) \mid X = x\right]$$

  3. La fonction $x \longrightarrow \hat{\theta}(\underline{x})$ est mesurable.

Alors

$$\hat{\theta}(\underline{X}) = \hat{\theta}^B(\underline{X})$$

- Corollaire Considérons le cas $\Theta = \mathcal{E} = \mathbb{R}$
  1. Si $L(\theta, \theta') = (\theta - \theta')^2$ est la fonction perte quadratique, alors l'estimateur bayésien est donné par la moyenne a posteriori

$$\hat{\theta}^B(\underline{X}) = \mathbb{E}(\theta \mid X)$$

De plus $\mathcal{R}\left(\hat{\theta}^B(\underline{X})\right) = \mathbb{E}\left[\mathbb{V}(\theta \mid X)\right]$

2. Si $L(\theta, \theta') = |\theta - \theta'|$, alors l'estimateur bayésien est donné par la médiane a posteriori

$$\hat{\theta}^B(\underline{X}) = \text{Médiane}_{\theta|X}$$

- Théorème Si un estmateur bayésien est sans biais alors il a un risque bayésien nul.

Exercice

Soit $X = (X_1, \dots, X_n)$ un eéchantillon i.i.d de loi normale de moyenne θ inconnue et de variance σ^2. Consideérons la loi a priori $\pi = \mathcal{N}(0, \sigma_0^2)$ 1. Calculer la loi a posteriori sur θ. 2. En deéduire l'expression de l'estimateur bayésien associé aà la fonction perte quadratique. 3. Calculer le risque bayésien de l'estimateur bayésien. 4. Reépondre aux questions 1,2,3 avec la fonction perte $L(\theta, \theta') = (\theta - \theta')^2 \exp(\theta^2)$. On considérera le cas où $\sigma^2 = 1$.

Le choix de la loi *a priori*

(Séance 3, 17/9/2016)

Pour construire une loi a priori on doit distinguer 3 cas possibles :

1. On dispose d'une information a priori partielle, donnée par l'expérience.

2. On dispose d'une quantité limité : lois conjuguées.

3. On dispose d'une quantité limité ou même aucune information sur θ : la loi a priori de Jeffrey.

Lois *a priori* conjuguées.

Une loi a priori conjuguée sur θ pour un modèle statistique donné si la loi a posteriori sur θ appartient à la même famille de la loi a priori.

Exercice.

Calculer dans les cas suivants la loi a priori conjuguée sur θ et calculer la loi a posteriori :

- $\mathcal{N}(\theta, \sigma^2)$ où σ^2 est connue.
- Poisson(θ)
- $\Gamma(\alpha, \theta)$ où α est connue.
- Binomiale(N, θ)
- $\mathcal{N}(\mu, \theta^{-1})$ où μ est connue.

Lois *a priori* de Jeffrey.

On suppose que le n−échantillon $X_1, \dots, X_n$ de densité $f(\bullet \mid \theta)$. La loi de Jefferey se construit aà partir seulement la connaissance de la densiteé $f(\bullet \mid \theta)$. Cette famille de lois est basée sur le calcul de l'information de Fisher :

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial \log f(X\,|\theta)}{\partial \theta}\right)^2\right]$$

Dans le cas de modèles réguliers.

$$I(\theta) = \mathbb{E}\left[\frac{\partial^2 \log f(X\,|\theta)}{\partial \theta^2}\right] = \mathbb{E}\left[\frac{\partial^2 \log l_n(\theta)}{\partial \theta^2}\right]$$

où l_n est le log-vraisemblance. Ces lois sont souvent des lois a priori non-informatives.

Définition la loi a priori de Jeffrey qu'on note $^ c'est la loi qui admet la densité suivante

- Si $\Theta \subseteq \mathbb{R}$

$$q(\theta) \propto I^{-1/2}(\theta)$$

- Si $\Theta \subseteq \mathbb{R}^d$

$$q(\theta) \propto |I(\theta)|^{-1/2}$$

Exercice

Reprendre les modèles statistiques de l'exercice précedent et calculer les lois a priori de Jeffrey correspondantes.

Exercice

On considère $\underline{x} = (x_1, \dots, x_n)$ un échantillon i.i.d d'une variable X de loi de densité

$$f(x\,|\alpha, \beta) \propto \exp\left(-\alpha x^2 + \beta x\right)$$

définie sur tout $\mathbb{R}$.

1. Montrer que la loi de X est une loi $\mathcal{N}\left(\frac{\beta}{2\alpha}, \frac{1}{2\alpha}\right)$
2. En déduire la constante de normalisation de la densité f.
3. Ecrire la fonction vraisemblance $L\left(\alpha, \beta \,|\, \underline{x}\right)$.
4. Supposons que α est connue.
   a. En déduire une loi a priori conjuguée pour β
   b. Calculer la loi a posteriori de β
   c. Calculer l'estimateur de Bayes de β en considèrant la fonction de perte quadratique.
   d. Reprendre les questions a,b et c. en considérant la loi a priori de Jeffrey.
5. On suppose maintenant que β est connue
   a. En déduire une loi a priori conjuguée pour α
   b. Calculer la loi a posteriori de α

Exercice

On considère le code suivant

```
>  p = seq(0.05, 0.95, by = 0.1)
>  prior = c(1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0)
>  prior = prior/sum(prior)
>  Likelyhood=function(p,s,f) p^s*(1-p)^f
>  Lp=sapply(p,function(x) Likelyhood(x,11,16))
>  post=Lp*prior
>  post=post/sum(post)
>  round(post,4)
 [1] 0.0000 0.0023 0.1291 0.4768 0.3338 0.0559 0.0021 0.0000 0.0000 0.0000
```

1. Quel est le modèle Bayésien considéré dans ce code?
2. Calculer l'estimateur bayèsien du param`etre considèré dans ce code.
3. Donner une loi a priori conjuguée du mode`ele Bayésien.

Modèle Bernoulli

- $y_1, \ldots, y_n$ un n−échantillon de loi Bernoulli $\mathcal{B}(p)$.
- La loi a priori conjuguée

$$g(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

- De moyenne

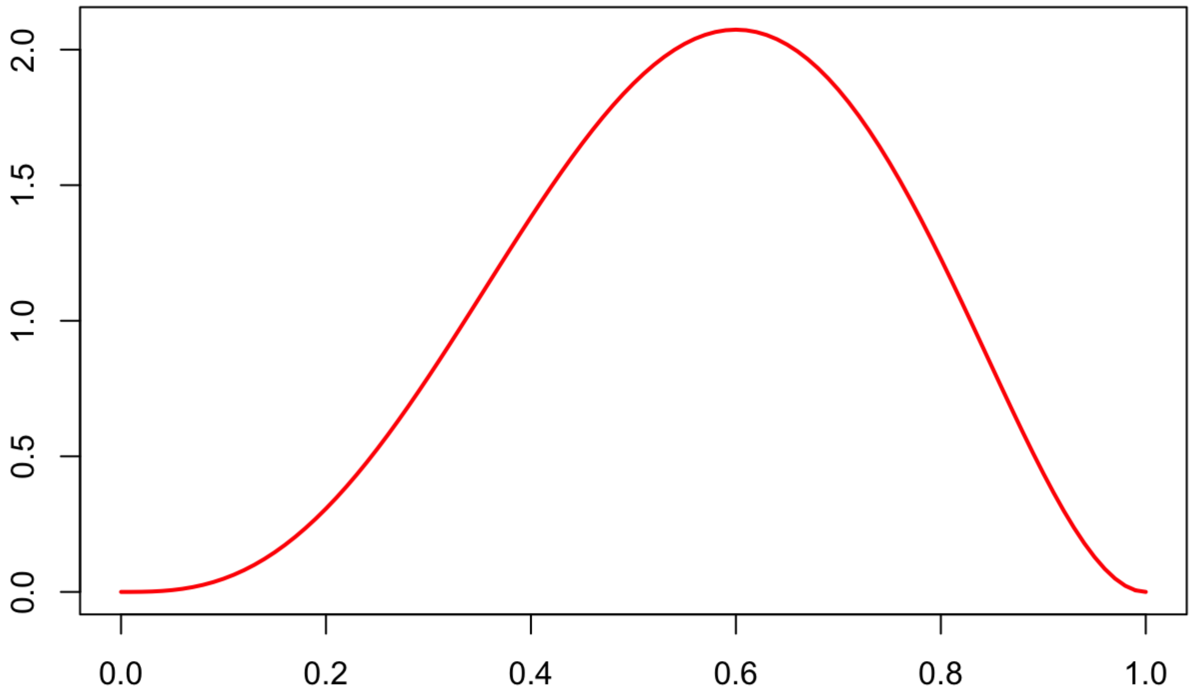
$$E(p) = \frac{\alpha}{\alpha+\beta}$$

- De Variance

$$Var(p) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- De densité

```
> curve(dbeta(x,4,3),from = 0,to = 1,col="red",lwd=2,xlab = "",ylab="")
```



- Calcul d'une loi a priori : On croit que la médiane et 90ième percentile de p sont respectivement égales à 0,3 et 0,5.

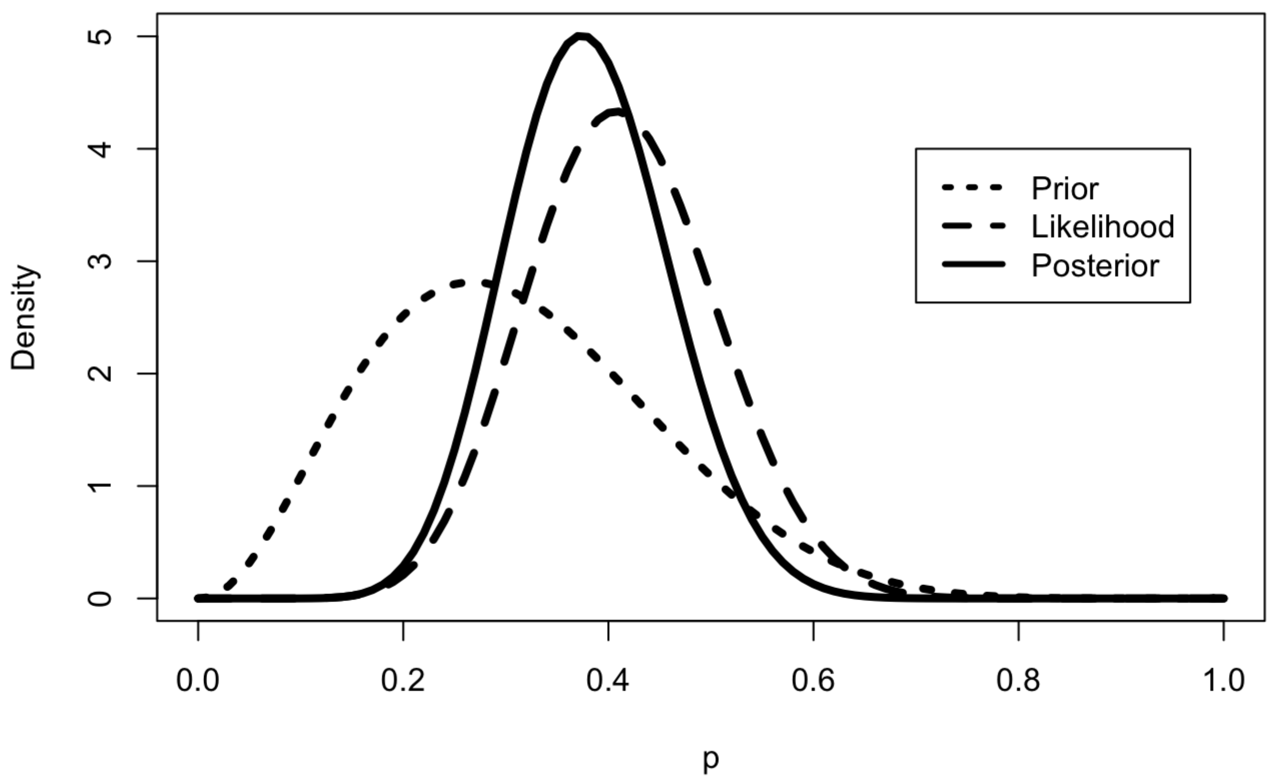
```
>  library(LearnBayes)
> quantile2=list(p=.9,x=.5)
> quantile1=list(p=.5,x=.3)
> beta.select(quantile1,quantile2)
[1] 3.26 7.19
```

- La loi a posteriori

$$g(p \mid \text{ données }) \propto p^{s+ -1}(1-p)^{N-s+ -1}$$

- Représentation des lois a priori, a posteriori et vraisemblance.

```
>  a  =  beta.select(quantile1,quantile2)[1]
>  b  =  beta.select(quantile1,quantile2)[2]
> s  =  11
> f  =  16
>  curve(dbeta(x,a+s,b+f),  from=0,  to=1,
+  xlab="p",ylab="Density",lty=1,lwd=4)
>   curve(dbeta(x,s+1,f+1),add=TRUE,lty=2,lwd=4)
>   curve(dbeta(x,a,b),add=TRUE,lty=3,lwd=4)
>   legend(.7,4,c("Prior","Likelihood","Posterior"),
+          lty=c(3,2,1),lwd=c(3,3,3))
```



- Calcul de la probabilité P(p $\geq$ .5 | données).

```
> 1 - pbeta(0.5,a+s,b+f)
[1] 0.0690226
```

- Estimer une valeur future $\widetilde{y}$ (le nombre de fois de la réalisations d'événement de probabilité p quand l'expérience s'est répétée n fois). On calcule la densité prédictive de $\widetilde{y}$:

$$f(\widetilde{y}) = \int f(\widetilde{y} \mid p) g(p) dp.$$

- Deux type densités prédictives : a priori si g est la loi a priori ou a posteriori si g est la loi a posteriori

Exemple

On lance une pièce de monnaie n fois, on observe y fois piles, on considère une loi a priori sur p. Supposons qu'on veut jeter m fois la pièce et on veut prédire le nombre $\tilde{y}$ d'apparitions de la face Pile.

- Solution fréquentiste : estimer p par $\hat{p} = y/n$, et prédire $\tilde{y}$ par $\tilde{y} = m\hat{p} = my/n$?
- Solution bayésienne : calculer la densité prédictive a posteriori de $\tilde{y}$.

$$
\begin{aligned}
g(\tilde{y} \mid y) &= \int g(\tilde{y},\ p \mid y)dp \\
&= \int g(\tilde{y} \mid\ p)\, g(p \mid y)dp
\end{aligned}
$$

La loi Beta-Binomiale

On a déjà vu que si $p \sim \mathrm{Beta}(a, b)$ alors $p \mid y \sim \mathrm{Beta}(y + a, n - y + b)$. La loi densité prédictive a posteriori s'écrit

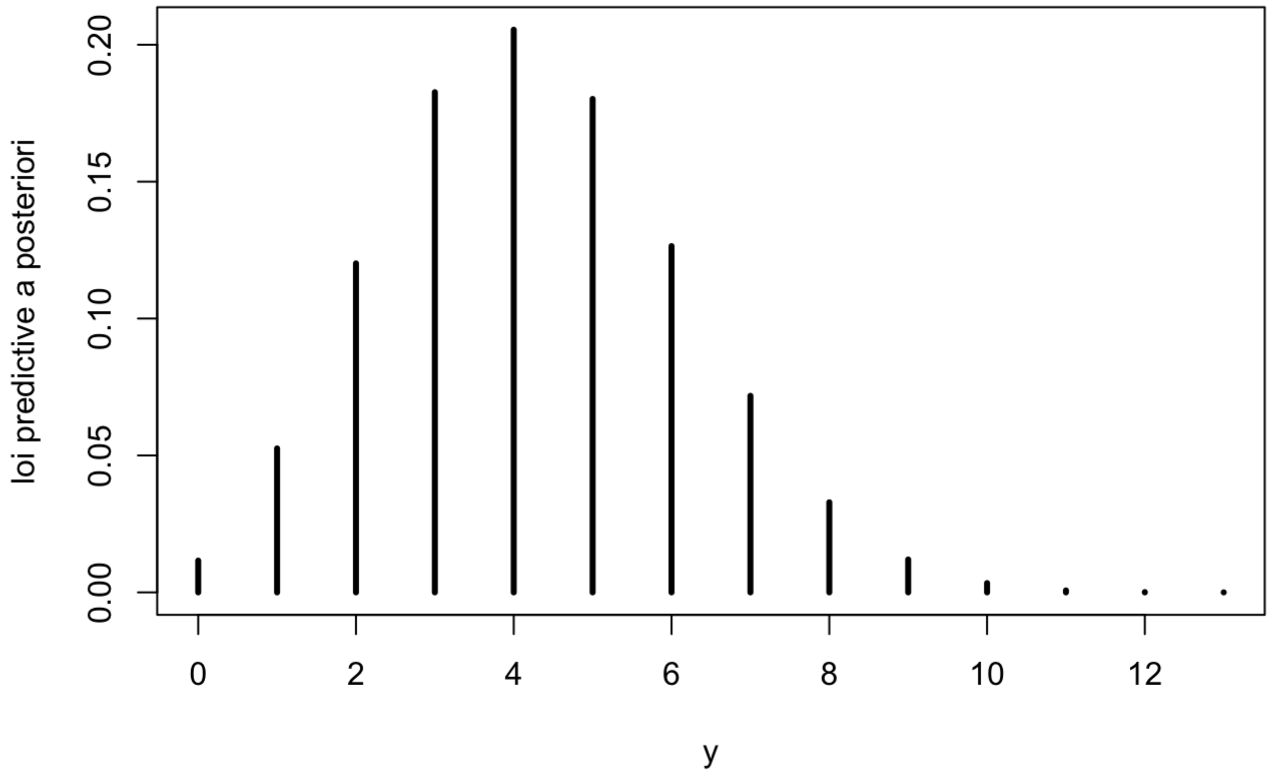
$$
\begin{aligned}
g(\tilde{y} \mid y) &= \frac{1}{B(y + a, n - y + b)} \int_0^1 C_m^{\tilde{y}} p^{\tilde{y}} \\
&\quad (1-p)^{m-\tilde{y}}\, p^{y+a}\, (1-p)^{n-y+b}\, dp \\
&= C_m^{\tilde{y}} \frac{B(\tilde{y} + y + a, m - \tilde{y} + n - \tilde{y} + b)}{B(y + a, n - y + b)}
\end{aligned}
$$

On dit que $\tilde{y}$ suit une loi Beta-Binomiale de paramètres $m, y + a, n - y + b$.

Application

On observe pour $n = 40$, $y = 13$ et on veut prédire les probabilités d'observer le nombre de Piles après $m = 13$

```
> library(emdbook)
>  y=13
>  n=40
>  ### a priori beta 2,4
>  a=2
>  b=4
>  m=13
>  yy=0:m
>  ## Calcul de la loi predictive a posteriori.
>  predy=sapply(yy,function(x) dbetabinom(x,shape1 = y+a,shape2 = n-y+b,size = m))
>  names(predy)=yy
>  plot(yy,predy,xlab="y",ylab="loi predictive a posteriori",
+  type="h",lwd=3)
```



Exemple : combien de fois on a jeté un dé

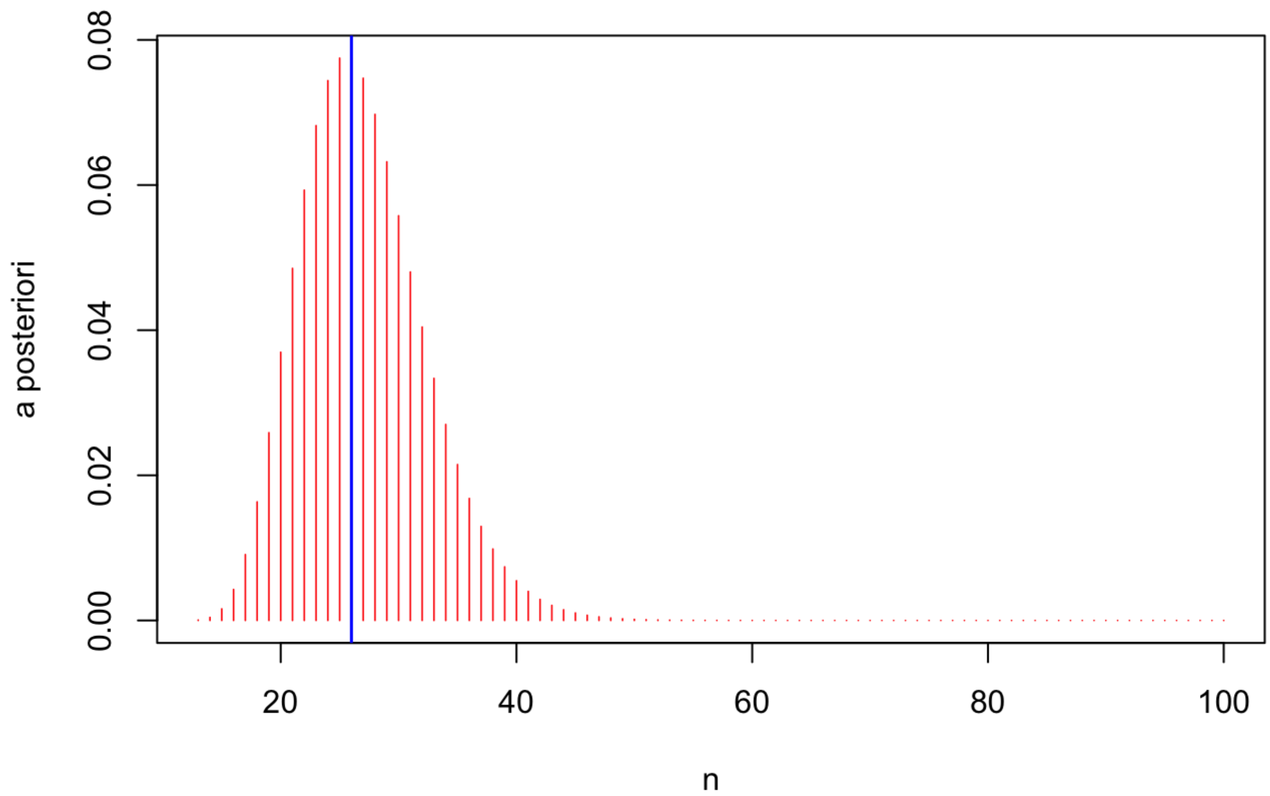
- Supposons qu'on a jeté plusieurs fois une pi`{e}ce de monnaie équilibrée et on vous dit qu'on a obtenu ``face" 13 fois.
- Question : Pourrions-nous estimer le nombre de jets de la pi`{e}ce ?
- Solution : On considère un a priori non-informative sur n : $g(n) \propto 1$ et la vraisemblance est alors

$$L(n) \propto C_n^{13} \left(\frac{1}{2}\right)^n.$$

  Ainsi la loi a posteriori s'exprime

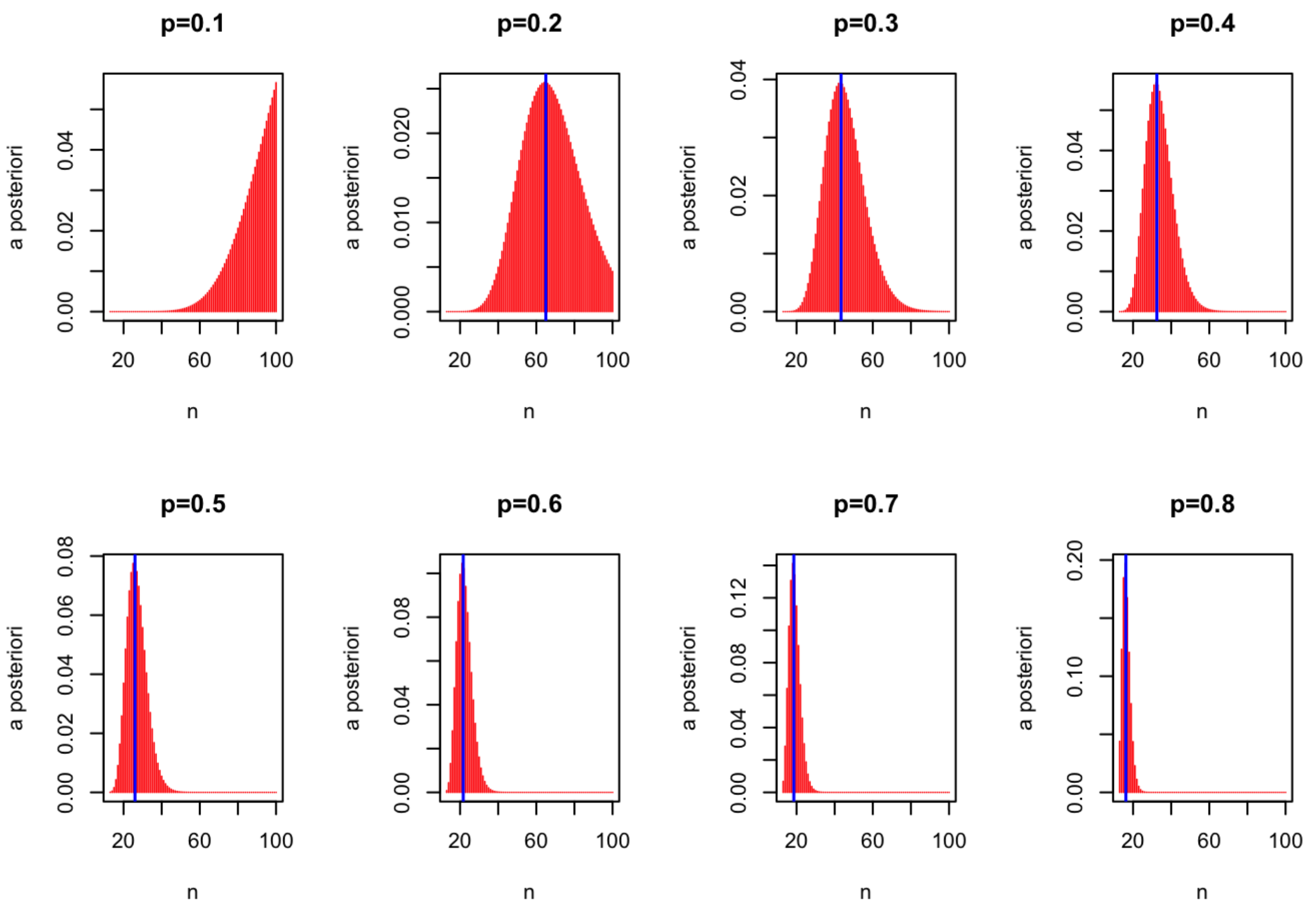
$$g(n \mid x) \propto L(n)g(n) = C_n^{13} \left(\frac{1}{2}\right)^n.$$

```
>  n=13:100
>  posteriori=function(n,p) choose(n,13)*p^13*(1-p)^(n-13)
>  ### le cas p=.5
>  postn=sapply( n,function(n) posteriori(n,.5))
>  postn=postn/sum(postn)
>  #### Graphique et comparer avec  2*13=26.
>  plot(n,postn,type="h",lwd=.8,col="red",xlab="n",
+ ylab="a posteriori")
>  abline(v=26,col="blue",lwd=1.5)
```



- Comparons n avec x/p pour diff'erentes valeurs de p.

```
>  p=seq(0.1,0.8,by=0.1)
>  par(mfrow=c(2,4))
>  for(i in 1:8){
+  postn=c()
+  postn=sapply( n,function(n) posteriori(n,p[i]))
+  postn=postn/sum(postn)
+  plot(n,postn,type="h",lwd=.8,col="red",xlab="n",
+  ylab="a posteriori", main=paste("p=",p[i],sep=""))
+  abline(v=13/p[i],col="blue",lwd=1.5)
+  }
```


p=0.1
p=0.2
p=0.3
p=0.4
p=0.5
p=0.6
p=0.7
p=0.8
a posteriori
n
20
60
100