OS13 : Tests d'estimation et loi extrêmes des lois de Cauchy et géométrique

Adrien WARTELLE
TRAN Quoc Nhat Han

4 octobre 2018

Sommaire

1 Loi géométrique

La loi géomètrique est une loi discrète de distribution utilisé dans le cadre d'un enchainement (non fini) d'épreuves de Bernouilli. Soit X une varaible aléatoire suivant cette loi, on a :

$$P(X = k) = (1 - p)^(k - 1)p \ k \in \mathbb{N}_+$$

La variable X est le numéro de la première épreuve où l'on obtient un succès qui a une probabilité p. Ainsi, p est le seul paramètre de loi (à estimer). La fonction de répartition $F(k) = P(X \leq k) \ k \in \mathbb{N}_+$ est :

$$F(k) = 1 - (1 - p)^k$$

Démonstration. On peut remarquer que :

$$P(X = k) = (1 - p)^{k-1}(1 - (1 - p)) = (1 - p)^{k-1} - (1 - p)^k$$

Donc :

$$F(k) = \sum_{l=1}^{k} P(X = l) = \sum_{l=1}^{k}((1 - p)^{l-1} - (1 - p)^l)$$

$$F(k) = \sum_{l=0}^{k-1}(1 - p)^l - \sum_{l=1}^{k}(1 - p)^l$$

Soit $q = 1 - p$, on a :

$$F(k) = \sum_{l=0}^{k-1} q^l - \sum_{l=1}^{k} q^l$$

$$F(k) = \frac{1 - p^k}{1 - q} - (\frac{1 - q^{k+1}}{1 - q} - 1)$$

$$F(k) = \frac{1 - q^k - 1 + q^{k+1} + 1 - q}{1 - q}$$

$$F(k) = \frac{1 - q + q^{k+1} - q^k}{1 - q}$$

$$F(k) = \frac{1 - q - q^k(1 - q)}{1 - q}$$

$$F(k) = \frac{1 - q - q^k(1 - q)}{1 - q}$$

$$F(k) = 1 - q^k$$

On retrouve bien :

$$F(k) = 1 - (1 - p)^k$$

$\square$

1.1 Génération de n variables aléatoires

Afin de voir à quoi ressemble la distribution et de pouvoir tester l'estimateur que nous allons calculer par la suite, on génère des échantillons de 50, 500 et 5000 variables. On utilise ainsi le code Matlab (Octave) ce dessous :

```
p=0.1; %probability to win
pop1 = geomGen(50,p);
pop2 = geomGen(500,p);
pop3 = geomGen(5000,p);
k=1:max(pop3)+1;
distrib = (1-p).^k.*p;
subplot(221);
hist(pop1,k,1);
xlabel('k');
ylabel('Freq(k)');
title('Histogramme de 50 variables');
subplot(222);
hist(pop2,k,1);
xlabel('k');
ylabel('Freq(k)');
title('Histogramme de 500 variables');
subplot(223);
hist(pop3,k,1);
xlabel('k');
ylabel('Freq(k)');
title('Histogramme de 5000 variables');
subplot(224);
bar(k,distrib);
xlabel('k');
ylabel('P(X=k)');
title('Distribution théorique géomètrique');
```

On obtient ainsi la figure 1.

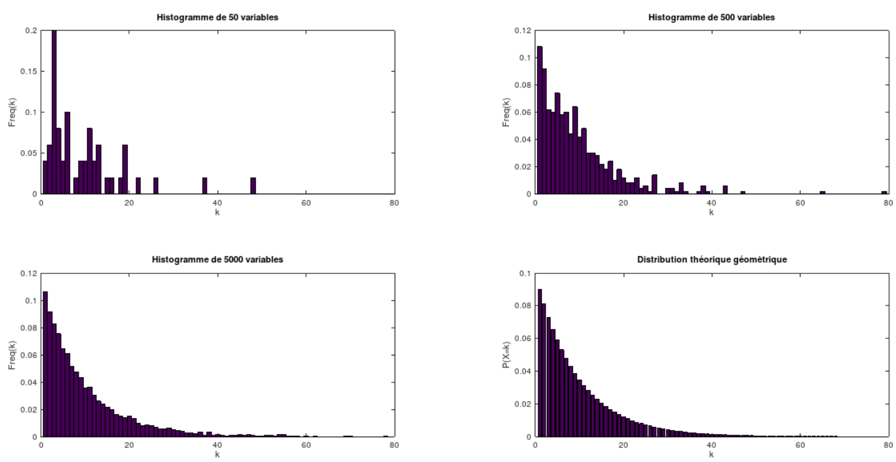

FIGURE 1 – Histogrammes et distribution de loi géométrique

On observe bien que les histogrammes se rapprochent de la distribution théorique quand la taille de l'échantillon est grande (plusieurs milliers).

Pour générer les populations , on utilise la fonction "geomGen" qui simule n expériences où, pour chacune d'entre elle, on effectue un enchainement d'épreuve de Bernouilli en générant une variable de loi uniforme sur $[0;1]$. Si la valeur obtenu est supérieure à p, on arrête la boucle et la valeur générée est égale au nombre d'épreuves, sinon on continue jusqu'à obtenir cette condition. Le code Matlab (Octave) utilisé est :

```
function[X]=geomGen(n,p)
X=ones(1,n);
for (i=1:n)
  trial=rand(1,1);
  k=1;
  while (trial > p)
    trial=rand(1,1);
    k++;
  endwhile
  X(i)=k;
endfor
endfunction
```

1.2 Estimateurs du maximum de vraisemblance

L'estimateur $\hat{p}$ du maximum de vraisemblance est la valeur qui maximise la loi de vraisemblance L(p) : $\hat{p} = argmax(L(p))$. On calcule L(p) dans un premier temps :

$$L(p) = L(X_1(p), X_2(p), ..., X_n(p)) = \prod_{i=1}^{n} (1-p)^{X_i-1} p$$

avec $X_i,\ i \in \{1, .., n\}$ les variables d'échantillon.

$$L(p) = p^n (1-p)^{(\sum\limits_{i=1}^{n} X_i)-n}$$

On peut effectuer un passage au logarithme pour trouver un maximum car il s'agit d'une fonction strictement croissante (et défini sur $]0;1]$). L'argument du maximum du logarithme de $L(p)$ et du maximum de $L(p)$ sont les mêmes.

$$\log L(p) = n \log p + ((\sum_{i=1}^{n} X_i) - n) \log(1-p)$$

$$\frac{d \log L(p)}{dp} = \frac{n}{p} - \frac{(\sum\limits_{i=1}^{n} X_i) - n}{1-p}$$

$$\text{On a } \frac{d\log L(\hat{p})}{dp} = 0$$

$$\text{Donc } \frac{n - n\hat{p} - ((\sum_{i=1}^{n} X_i) - n)\hat{p}}{\hat{p}(1-\hat{p})} = 0$$

$$\text{Ce qui implique (dans le cas différent de 0 et 1) } n - (\sum_{i=1}^{n} X_i)\hat{p} = 0$$

Finalement :

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i}$$

De plus on peut vérifier qu'il s'agit bien d'un maximum (unique et donc global) en regardant si la fonction $L(p)$ est concave :

$$\frac{d^2 \log L(p)}{dp^2} = -\frac{n}{p^2} - \frac{\sum_{i=1}^{n} X_i) - n}{1-p}$$

Sachant que $\sum_{i=1}^{n} X_i > n$ et que $p \in [0;1]$, on a :

$$\frac{d^2 \log L(p)}{dp^2} < 0 \; \forall p \in [0;1]$$

La fonction $L(p)$ est donc concave et l'estimateur trouvé $\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i}$ maximise bien la loi de vraisemblance.

Si la résolution analytique pour trouver l'estimateur avait été impossible, l'utilisation d'un algorithme d'optimisation (comme une descente de gradient) ou encore l'utilisation d'une fonction d'intégration aurait été possible.

1.3 Test de l'estimateur

Afin de tester l'estimateur du maximum de vraisemblance pour différent taille d'échantillon, on utilise le code suivant :

```
close all;
p=1/10;
n=10.^(1:0.02:4);
delta_est=zeros(1,length(n));
for i=1:length(n)
  Y=geomGen(n(i),p);
  p_est = n(i)/sum(Y);
  delta_est(i) = abs(p_est - p);
```

```
  i++;
endfor
plot(log10(n),delta_est);
title('Evolution du biais de p_{est}');
xlabel('log10(n)');
ylabel('|p_{est} - p|');
```

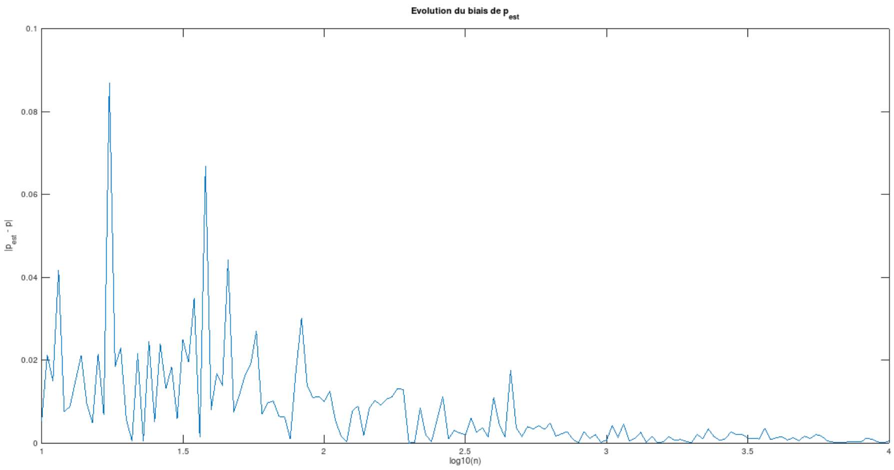

FIGURE 2 – Evolution du biais de l'estimateur

On voit que
On effectue la ...

```
close all;
p=1/10;
n=10.^(1:0.2:4);
delta_esp_est=zeros(1,length(n));
for i=1:length(n)
  esp_p_est=0;
  for k=1:100
    Y=geomGen(n(i),p);
    esp_p_est += n(i)/sum(Y);
  endfor
  esp_p_est/=k;
  delta_esp_est(i) = abs(esp_p_est - p);
  i++;
endfor
plot(log10(n),delta_esp_est);
title('Evolution du biais moyen de p_{est}');
xlabel('log10(n)');
ylabel('|p_{est} - p|');
```

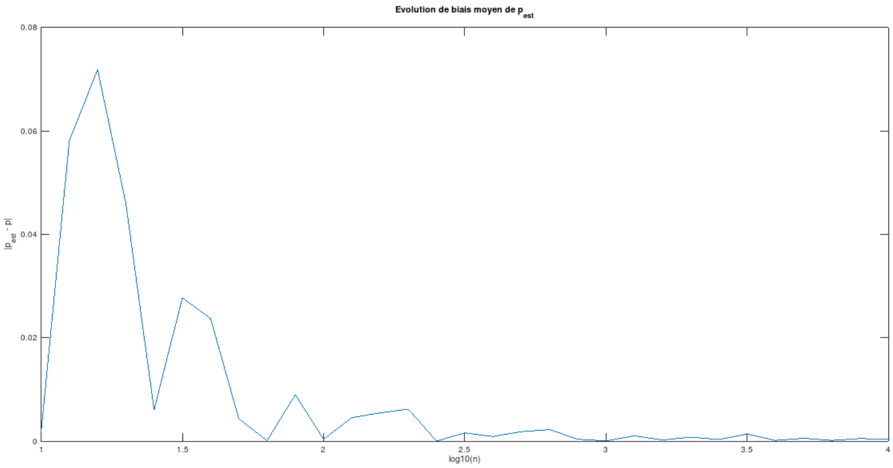

FIGURE 3 – Evolution du biais moyen de l'estimateur

1.4 Loi extrême

2 Loi de Cauchy