

## PROPOSITION DE CORRIGÉ DE L'EXAMEN – 10 JUIN 2004

GILLES STOLTZ

L'examen a été noté sur 29 : 7 points pour chacun des exercices et 15 points pour le problème (la note sur 29 est considérée comme une note sur 20, il n'y a pas d'application de règle de proportionnalité). Vous pouvez consulter vos copies à mon bureau<sup>1</sup>, soit le mardi 15 juin (peut-être le mercredi 16 juin), soit à partir du 28 juin.

### 1. CALCULS DE $\pi$

Voici, à grands traits, ce que j'attendais pour réponses aux questions théoriques. Nos variables aléatoires sont les  $g(U_j)$ ,  $j \geq 1$ , et on s'intéresse à la suite de leurs moyennes mobiles. Les garanties théoriques sont données, dans le cas de la variance, par l'inégalité de Chebychev (hypothèse :  $g(U_1)$  admet un moment d'ordre 2), et dans le cas où  $g(U_1)$  est bornée, par le lemme d'Hoeffding. Aucune de ces hypothèses n'est vérifiée pour la fonction proposée par l'élève cosinus, contrairement à celle de l'élève lambda, pour qui les deux garanties s'appliquent. Le lemme de Hoeffding propose notamment un nombre  $N$  majoré par

$$N = \frac{(4-0)^2 \ln(2/\delta)}{2\varepsilon^2} \leq 3 \times 10^5 .$$

Attention, nombre d'entre vous n'ont pas tenu compte du fait que  $g$  prenait ses valeurs dans  $[0,4]$ .

%N calculé par Hoeffding (pour la fonction g)

```
N = 3*10^5;
```

%Approximation de pi par f

```
y = max(abs(2*mean(1./sqrt(1-rand(10,N).^2),2)-pi))
```

%Approximation de pi par g

```
z = max(abs(4*mean(sqrt(1-rand(10,N).^2),2)-pi))
```

J'obtiens par exemple  $y = 0.0206$  et  $z = 0.0023$ . La théorie nous disait que nous devrions avoir davantage confiance en  $g$  que  $f$  et la pratique nous le confirme. Il est donc crucial de s'intéresser aux garanties théoriques offertes par les différentes méthodes de Monte-Carlo!

### 2. APPROXIMATION BINÔMIALE-POISSON

La situation pratique est le nombre de lettres ou d'appels téléphoniques reçus par jour (chacun de nous écrit une lettre ou émet une communication téléphonique selon une loi de Bernoulli, avec un paramètre petit, sinon nous passerions notre temps au téléphone où à écrire des lettres).

Le script est le suivant. Notez que l'on représente la convergence par des diagrammes<sup>2</sup> en bâtons parce que nous avons affaire à des variables discrètes (et ainsi, convergence des bâtons équivaut dans ce cas à convergence des fonctions de répartition, donc à la convergence en loi). Je recentre mes diagrammes sur les classes chargées positivement par l'échantillon (c'est l'objet du travail avec `cl`), sans quoi on ne voit rien qu'un gros pic et plein de blanc...

```
clf; clc; clear;
```

```
pas = [10 25 50 100];
```

```
N = 2*10^4;
```

```
R = zeros(length(pas),1);
```

```
poiss = 5;
```

---

1. Bureau 106, bâtiment 430

2. Commandes `hist` et `bar`, ou `stem`, surtout pas `histo`!

```

for i = 1:length(pas)

    n=pas(i);
    B = sum((rand(N,n) <= poiss/n),2);

    P = exp(-poiss)*(poiss.^(0:pas(i)))./cumprod([1 1:pas(i)]);

    %Proximité: manière graphique
    cl = hist(B,0:pas(i));
    subplot(2,2,i);
    %Pour rendre le graphique plus joli, on s'arrête
    %à la première barre de hauteur nulle
    m = [find(cl == 0) pas(i)+1]; %Pour le cas où cl ne contient pas de 0!
    bar(0:m(1)-1,cl(1:m(1))/N);
    hold on;
    stem(0:m(1)-1,P(1:m(1)),'r');
    title(['Cas n = ' num2str(pas(i))]);

    %Proximité: manière numérique
    R(i) = max(abs(cl/N-P));

end;
disp(R);

```

La différence, pour  $k$  entre 0 et  $n$ , des probabilités (empiriques, calculées sur mes échantillons) que les binômiales ou la Poisson valent  $k$  sont mesurées par  $R$ . Par le lemme de Hoeffding et pour  $N$  assez grand,  $R$  approxime bien empiriquement les valeurs théoriques des différences. J'obtiens pour  $R$  le vecteur suivant :

```

0.0703
0.0212
0.0121
0.0068

```

et pour graphiques :

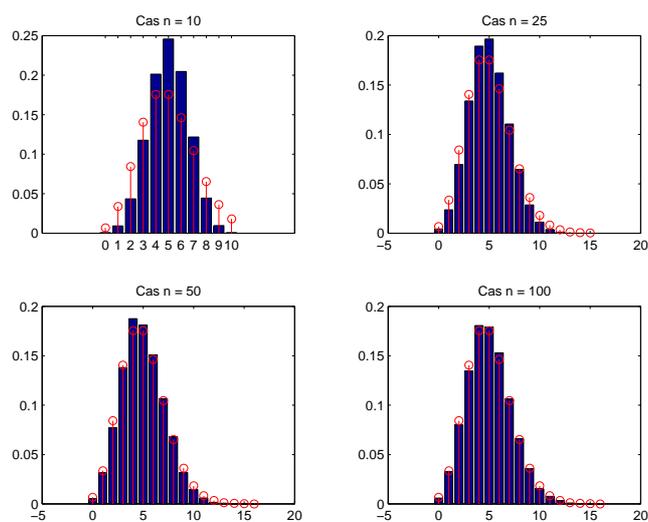


FIG. 1. Approximations binômiale-Poisson à différents rangs  $n$

Les erreurs les plus fréquemment rencontrées sont l'utilisation abusive d'`histo` (auquel cas, généralement, vos graphes n'arrivent pas à illustrer la convergence), de même qu'un appel fautif à `hist` : il faut préciser  $0:k$  classes, où  $k$  est le paramètre de la binômiale  $\mathcal{B}(k, \lambda/k)$  – et non pas un nombre fixe de classes.

### 3. PROCESSUS DE GALTON-WATSON

#### 3.1. Cas sous-critique.

```
%Fonction auxiliaire de simulation des enfants
%p <-- p(1): proba que 0 enfant, p(2): proba que 2 enfants
%Prend un vecteur u, renvoie un vecteur v
function v = enfants(u,p)
v = u;
for i = 1:length(u)
    U = rand(1,u(i));
    v(i) = u(i) - sum(U < p(1)) + sum(U > 1 - p(2));
end;

%Fonction du cas sous-critique
%Ex: n = 100, N = 10000
function [E,R] = SousCrit(n,N)

p = [0.2 0.1];
Z = ones(N,1);

R = N;
E = N;

for i = 1:n
    Z = enfants(Z,p);
    R = [R sum(not(Z==0))];
    E = [E sum(Z)];
end

clf;
subplot(1,2,1);
plot(0:n,1-R/N,'b');
title('Probabilité d''extinction');
subplot(1,2,2);
plot(0:n,E/N,'b',0:n,0.9.^(0:n),'r');
title('Estimation des espérances');
```

$N = 10^4$  par lemme de Hoeffding. Il semble que  $q = 1$  : c'est bien ce que prédit la théorie, il y a extinction presque-sûre dans le cas sous-critique. Attention, beaucoup ont confondu le graphe des probabilités d'extinction et celui de la suite des espérances, traçant le second en lieu et place du premier. Pourtant, l'un montre une fonction dont on savait déjà qu'elle était croissante, et l'autre approxime la fonction (décroissante)  $n \mapsto m^{-n}$ .

#### 3.2. Cas sur-critique.

```
function R = SurCrit(n,N)

p = [0.1 0.2];
m = 1.1;
Z = ones(N,1);
```

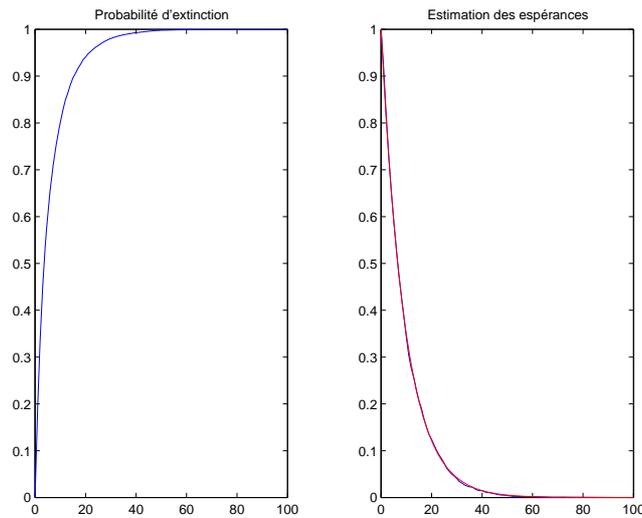


FIG. 2. Probabilités d'extinction et estimation des espérances ( $n = 100$  et  $N = 10000$ )

```

for i = 1:n
    Z = [Z enfants(Z(:,end),p)];
end

R = Z(:,end)/m^n;

disp('');
%Représentation de la loi limite:
%1. Hauteur du Dirac en 0
t = find(not(R==0));
disp(['Probabilité d''extinction : ' num2str(1-length(t)/N)]);
%2. Histogramme des valeurs prises (autres que 0)
%Il faut vraiment traiter le cas 0 (extinction) à part...
histo(R(t),10*N^.25,[],length(t)/N);

disp('Estimation de l''espérance de la loi limite : ')
disp([num2str(mean(R)) ' (à comparer à 1)']);
disp('Estimation de la variance de la loi limite : ')
disp([num2str(var(R)) ' (à comparer à ' num2str(0.29/(1.1^2-1.1)) ')']);
J'obtiens pour résultats numériques :
Probabilité d'extinction : 0.4933
Estimation de l'espérance de la loi limite :
1.0266 (à comparer à 1)
Estimation de la variance de la loi limite :
2.729 (à comparer à 2.6364)
et pour graphique:

```

Notez que la probabilité que  $W = 0$  n'apparaît pas sur la figure. Vu que le point 0 est vraiment chargé par la loi de  $W$  (poids presque  $1/2$ ), il est préférable de ne pas le mettre sur le graphique (ou alors mettre un bâton à part à la main, d'une autre couleur ou forme de ligne), histoire de ne pas écraser les autres classes.

La représentation sous forme d'histogramme serait à justifier mathématiquement : il faudrait plus de renseignement sur les lois en jeu (existence de moments d'ordre 3 par exemple)... mais esthétiquement, pas de doute: les histogrammes sont plus lisibles pour définir une loi que des fonctions de répartition.

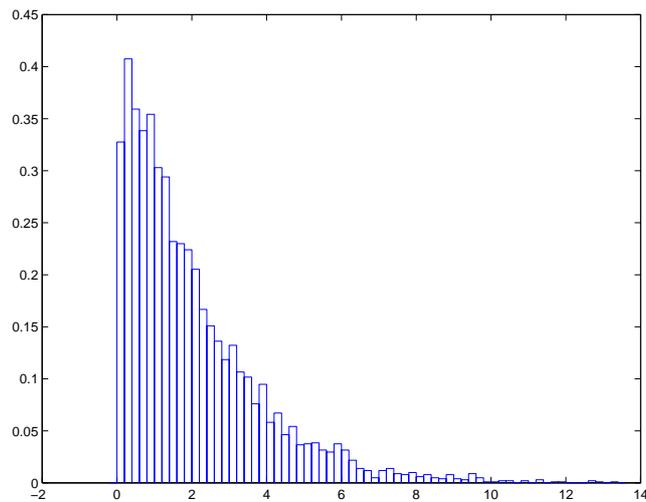


FIG. 3. *Distribution de la loi limite de  $Z_n/m^n$  dans le cas sur-critique (sauf Dirac en 0)*

### 3.3. Cas critique.

```
function Z = Crit(n,N)
```

```
p = [0.15 0.15];
```

```
Z = ones(N,1);
```

```
for i = 1:n
```

```
    Z = [Z enfants(Z(:,end),p)];
```

```
end
```

```
clf
```

```
%Suite des probabilités d'extinction multipliées par n
```

```
figure(1)
```

```
plot((1:n),(1:n).*sum(not(Z(:,2:(n+1)))==0),1)/N);
```

```
%Représentation de la convergence en loi
```

```
%par les fonctions de répartition,
```

```
%empirique et théorique
```

```
t = find(not(Z(:,end))==0);
```

```
D = Z(t,end)/n;
```

```
d = length(D);
```

```
figure(2);
```

```
hold on;
```

```
stairs(sort(D),(1:d)/d);
```

```
fplot('1-exp(-x/.15)',[0 max(D)*1.1],'r');
```

J'obtiens les deux graphiques ci-dessous. Le choix des fonctions de répartition s'impose ici : on n'a pas assez de renseignements pour assurer qu'une suite d'histogrammes empiriques bien choisie convergerait vers l'histogramme d'une loi exponentielle, tandis que la convergence des fonctions de répartition est toujours garantie.

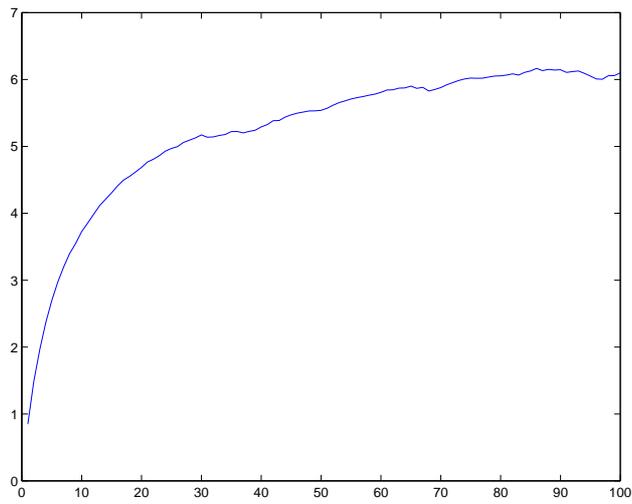


FIG. 4. *Convergence de la suite  $n\mathbb{P}(Z_n > 0)$  vers  $2/\sigma^2$*

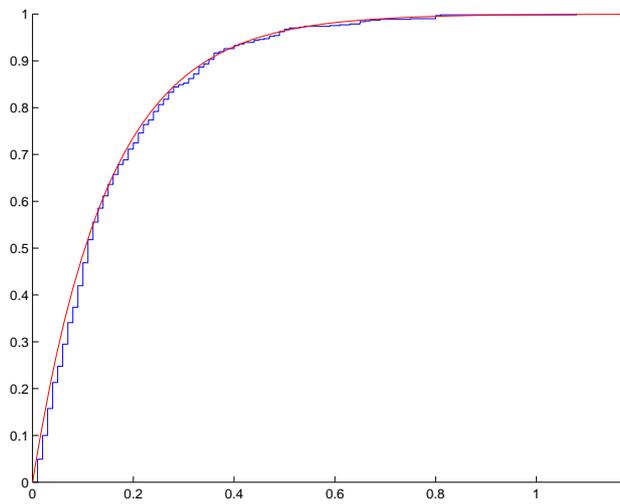


FIG. 5. *Représentation de la convergence en loi par convergence uniforme des fonctions de répartition*