Statistiques - TP2 noté

A. Enfroy, T. Ngo, T. Vo

7 février 2025

Simulation et convergence

Simulation dans R

Ex1. Si $X \sim \mathcal{N}(\mu = 20, \sigma^2 = 60)$, quelle est la probabilité qu'on observe une valeur supérieure à 35? Utiliser la fonction `pnorm(n=, mean=, sd = )` dans `R` pour évaluer cette probabilité théorique.

```
?rnorm
# list of available distributions
?Distributions
```

$p = \mathbb{P}(X > 35) = ?$

```
p =
print(c("La probabilité d'observer une valeur supérieure à 35 est", p))
```

Ex2. Pour la suite, nous voulons simuler un échantillon de taille $n = 20$ d'un loi de X. Complétez les codes suivants pour générer cet échantillon.

```
n = 20
mu = 20 # moyenne
sigma = sqrt(60) # l'écart type
x =
```

Obtenez la probabilité empirique $\hat{p}$ que la valeur dépasse 35. Créez un histogramme de votre échantillon et donnez des commentaires sur la forme de l'histogramme obtenu. Superposer la vraie densité.

```
p_chap =
hist(x, probability = TRUE, main = paste("Emp.prob=", p_chap), ylim = c(0, 0.1),
    xlim = c(min(x), max(x)))
## evaluer la densite de la loi normale en grille xx
xx = seq(-10,50,by=1)
yy = dnorm(xx, mean=20, sd=sqrt(60))
## add density plot
lines(xx, yy, col="red")
abline(v=c(20,35), col=c("red","gray"), lty=5)
```

Ex3. Répétez cette opération 10 fois et enregistrez les données dans une matrice ou un data frame (disons, `datmat`). Calculez la probabilité empirique pour chaque ensemble de données. Sont-elles proches de la valeur théorique ? Y a-t-il des différences entre les histogrammes que vous obtenez à chaque fois ? Utilisez la même limite sur les axes pour faciliter la comparaison.

```
m <- 10
datmat <- matrix(c(0), nrow = n, ncol = m)
for(j in 1:m){
  datmat[,j] = rnorm(n, mean = mu, sd = sigma)
```

```
}
```

```
## plots en multi panel
ncol(datmat) #
par(mfrow=c(2,2)) # 2 x 2 panel
for (j in 1:4)
{
  hist(datmat[,j], xlim= c(-10,50), ylim = c(0,0.1), probability = TRUE,
      main=paste("histogram of X_",j ), xlab=paste("X_",j))
}
```

Ex4. Augmentez la taille de votre échantillon à 50 et répétez votre expérience. Que remarquez-vous?

```
n<-50
m <- 10
datmat <-
for(j in 1:m){
  datmat[,j] =
}
```

```
## plots in multi panel
ncol(datmat) #
par(mfrow=c(2,2)) # 2 x 2 panel
for (j in 1:4)
{
  hist(datmat[,j], xlim= c(-10,50), ylim = c(0,0.1), probability = TRUE,
      main=paste("histogram of X_",j ), xlab=paste("X_",j))
}
```

Estimation - Méthode de Monte Carlo

Dans la suite (des exercices 5-8), nous voulons estimer les parametres d'un modèle statistique des lois uniformes $U \sim \mathcal{U}([a,b]), (a,b) \in \mathbb{R} \times \mathbb{R}$.

Ex5. Simuler un échantillon $(U_1, \dots, U_n)$ de taille $n = 10$ d'une loi $\mathcal{U}([-1,1])$ et enregister le maximum $M_n = \max_{1 \leq i \leq n} U_i$ et le minimum $N_n = \min_{1 \leq i \leq n} U_i$ de cet échantillon.

Ex6. Répéter les 2 étapes ci-dessus $m = 50$ fois, en écrivant le maximum et le minimum de l'échantillon à chaque fois. Donner les commentaires sur la variabilité des maxima obtenus des m échantillons.

Ex7. Maintenant, répéter $m = 100$ fois, faire un histogramme et une `boxplot` des M_n et des N_n.

Quelle est la loi de M_n, N_n? (Indication: voir TD1) Superposer la densité théorique sur l'histogramme. Comparer la moyenne de M_n (*rep.* N_n) avec les bornes de la loi de U_i. Que remarquez-vous ?

Ex8 Augmenter la taille de votre échantillon à $n = 50$ et répéter l'expérience. Comparer la moyenne de M_n (*rep.* N_n) avec les bornes de la loi de U_i. Que remarquez-vous ?

Théorème Central Limite

Ex9. Simuler $B = 500$ échantillons i.i.d de loi commune Bernoulli(p) (avec votre choix de paramètres) de taille $m = 20, 100, 200$ et calculer les moyennes et variances empiriques $\bar{X}_{m,i}$ et $S_{m,i}, i = 1, \dots, B$. Tracer l'histogramme des moyennes empiriques.

Ex10. A l'aide d'une renormalisation adéquate (a_m, b_m), montrer que $U_{m,i} = \frac{\bar{X}_{m,i} - a_m}{b_m}$ a une loi que vous pouvez approchez. Comparez histogramme de les moyennes empiriques normalisées, $U_{m,i}$, et *distribution théorique approchée.* Quelle est l'influence de la taille de l'échantillon m sur la qualité de cette approximation?