

Examen Final (2h)

Documents autorisés : polycopié de MAP361, notes personnelles de PC et de cours, transparents imprimés, dictionnaire. Tous les appareils électroniques doivent être éteints.

Exercice 1 [Estimation et espérances. (6pts)]

Pour tout réel $\theta > -1$ on définit la densité f_θ par

$$f_\theta(x) = \begin{cases} (1 + \theta)x^\theta & \text{si } 0 < x < 1, \\ 0 & \text{sinon.} \end{cases}$$

1. Soit X une variable aléatoire admettant pour densité f_θ . Démontrer que $^* \log(X)$ est une variable intégrable et calculer $\mathbb{E}[\log(X)]$.

Soit $n \geq 1$. On souhaite estimer le paramètre θ à partir d'un échantillon (X_1, \dots, X_n) de n variables aléatoires i.i.d. de densité f_θ .

2. Vérifier que l'Estimateur du Maximum de Vraisemblance (EMV), que l'on notera $\hat{\theta}_n$, associé à cet échantillon peut s'écrire

$$\hat{\theta}_n = -\frac{n}{\sum_{i=1}^n \log(X_i)} - 1.$$

3. Démontrer que quand $n \rightarrow +\infty$ on a $\hat{\theta}_n \rightarrow \theta$ presque sûrement.
4. En utilisant l'inégalité de Jensen, démontrer que l'EMV est biaisé : pour tout $\theta > -1$ on a $\mathbb{E}[\hat{\theta}_n] > \theta$. (*On peut utiliser sans justification que la variable aléatoire $\hat{\theta}_n$ est intégrable.*)

Solutions :

- 1.

$$\begin{aligned} \mathbb{E}[|\log(X)|] &= (1 + \theta) \int_0^1 |\log(x)| x^\theta dx \\ &= -(1 + \theta) \int_0^1 \log(x) x^\theta dx \\ &= -(1 + \theta) \left(\left[\log(x) \left(\frac{1}{\theta+1} x^{\theta+1} \right) \right]_0^1 - \int_0^1 (1/x) \left(\frac{1}{\theta+1} x^{\theta+1} \right) dx \right) \\ &= -(1 + \theta) \left(0 - \frac{1}{(\theta+1)^2} \right) < +\infty. \end{aligned}$$

Donc $\log(X)$ est intégrable. On écrit ensuite

$$\mathbb{E}[\log(X)] = -\mathbb{E}[|\log(X)|] = -\frac{1}{\theta + 1}.$$

(1pt : intégrable)

2. La vraisemblance s'écrit

$$L_\theta(x_1, \dots, x_n) := (1 + \theta)^n \prod_{i=1}^n x_i^\theta$$

que l'on log-maximise en calculant

$$\frac{\partial}{\partial \theta} \log(L_\theta) = \frac{n}{1 + \theta} + \sum \log(x_i).$$

Finalement,

$$\hat{\theta}_n = \operatorname{argmax}_\theta L_\theta(X_1, \dots, X_n) = -\frac{n}{\sum_{i=1}^n \log(X_i)} - 1.$$

(1pt : calcul)

3. Avec la LFGN et la question 2) on a $\frac{1}{n} \sum_{i=1}^n \log(X_i) \rightarrow -1/(\theta + 1)$. Et donc $\hat{\theta}_n \xrightarrow{n \rightarrow +\infty} \theta$ presque-sûrement.

(1pt : LFGN)

4. La fonction $x \mapsto -1/x$ est convexe sur $(-\infty, 0)$. Puisque $\sum_{i=1}^n \log(X_i) < 0$ l'inégalité de Jensen donne

$$\mathbb{E} \left[-\frac{n}{\sum_{i=1}^n \log(X_i)} \right] \geq -\frac{n}{\mathbb{E} [\sum_{i=1}^n \log(X_i)]} = -\frac{n}{n \times (-1/(\theta + 1))} = \theta + 1.$$

Comme $x \mapsto -1/x$ n'est pas affine et que $\sum_{i=1}^n \log(X_i)$ n'est pas constante on a en fait l'inégalité stricte ci-dessus (voir la Remarque 4.36 dans le Poly de MAP361). Finalement

(1pt : Jensen bien appliqué)

$$\mathbb{E} [\hat{\theta}_n] = -\mathbb{E} \left[\frac{n}{\sum_{i=1}^n \log(X_i)} \right] - 1 > \theta + 1 - 1 = \theta.$$

(1pt : cas d'égalité)

Exercice 2 [Convergence en loi. (2pts)]

Soit $((X_n, Y_n))_{n \geq 1}$ une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^2 telle que pour tout n les v.a. X_n et Y_n sont indépendantes. On suppose que quand $n \rightarrow +\infty$ on a $(X_n) \xrightarrow{\mathcal{L}} X$ et $(Y_n) \xrightarrow{\mathcal{L}} Y$. Démontrer que

$$X_n + Y_n \xrightarrow{\mathcal{L}} X' + Y',$$

où X' et Y' sont indépendantes, X' a même loi que X , et Y' a même loi que Y .

Solutions :

Soit $t \in \mathbb{R}$ fixé, on a pour tout n

(2pts)

$$\begin{aligned} \mathbb{E}[\exp(it(X_n + Y_n))] &= \mathbb{E}[\exp(itX_n)] \times \mathbb{E}[\exp(itY_n)] && \text{(indépendance)} \\ &\rightarrow \mathbb{E}[\exp(itX)] \times \mathbb{E}[\exp(itY)] && \text{(cv en loi)} \\ &= \mathbb{E}[\exp(itX')] \times \mathbb{E}[\exp(itY')] && \text{(trivial)} \\ &= \mathbb{E}[\exp(it(X' + Y'))] && \text{(indépendance)} \end{aligned}$$

Exercice 3 [Processus autorégressif. (3pts)]

Soit a un réel fixé et $(X_n)_{n \geq 0}$ le processus défini par $X_0 = 0$ et, pour tout $n \geq 1$,

$$X_n = aX_{n-1} + \varepsilon_n$$

où $(\varepsilon_n)_{n \geq 1}$ est une suite de variables aléatoires i.i.d. de loi $\mathcal{N}(0, 1)$.

1. Pour tout $n \geq 1$, écrire X_n en fonction de $\varepsilon_1, \dots, \varepsilon_n$. En déduire que, pour tout $n \geq 1$, $X_n \sim \mathcal{N}(\mu_n, s_n^2)$ où μ_n et s_n sont à déterminer.

(1pt)

2. Est-ce que (X_1, \dots, X_n) est un vecteur gaussien ? (Justifier rapidement.)

(0.5pt)

3. Démontrer que le n -uplet (X_1, \dots, X_n) admet une densité sur \mathbb{R}^n qui peut s'écrire :

$$(x_1, \dots, x_n) \mapsto \frac{1}{(\sqrt{2\pi})^n} \times \exp \left(-\frac{1}{2}x_1^2 - \frac{1}{2}(x_2 - ax_1)^2 - \frac{1}{2}(x_3 - ax_2)^2 \cdots - \frac{1}{2}(x_n - ax_{n-1})^2 \right).$$

(On pourra utiliser la méthode de la fonction muette.)

(1.5pt)

Solutions :

1. On a pour tout $n \geq 1$

$$X_n = \varepsilon_1 a^{n-1} + \varepsilon_2 a^{n-2} + \cdots + \varepsilon_n.$$

Ainsi X_n est une somme de gaussiennes indépendantes, c'est donc une gaussienne (Exemple 7.9 du Poly). Clairement on a $\mu_n = 0$ et la variance vaut

$$s_n^2 = \sum_{k=1}^n \text{Var}(\varepsilon_k a^{n-k}) = \sum_{k=1}^n a^{2(n-k)} = \frac{1 - a^{2n}}{1 - a^2}.$$

*. Dans tout cet examen log désigne le logarithme naturel, en base e .

2. Oui car si l'on prend une combinaison linéaire $\sum \alpha_i X_i$ alors elle peut s'écrire comme combinaison linéaire des ε_k et donc c'est une gaussienne. Attention le critère "transformation linéaire de gaussiennes indépendantes \Rightarrow vecteur gaussien" ne semble pas dans le poly.

Remarque : Si des élèves utilisent (en l'admettant) la formule de la densité énoncée à la question suivante j'estime que c'est correct et que ça mérite tous les points. (Après tout ça demande plus de boulot que l'argument direct...).

3. **1ère méthode : Fonction muette.** Soit $\phi : \mathbb{R}^n \mapsto \mathbb{R}$ continue bornée,

$$\begin{aligned} \mathbb{E}[\phi(X_1, \dots, X_n)] &= \mathbb{E} \left[\phi \left(\varepsilon_1, \dots, \sum_{k=1}^n \varepsilon_k a^{n-k}, \dots \right) \right] \\ &= \int_{\mathbb{R}^n} \phi \left(e_1, \dots, \sum_{k=1}^n e_k a^{n-k}, \dots \right) \prod_{k=1}^n \frac{\exp(-e_k^2/2)}{\sqrt{2\pi}} de_1 \dots de_n \\ &= \int_{\mathbb{R}^n} \phi(x_1, \dots, x_n) \prod_{k=1}^n \frac{\exp(-(x_k - ax_{k-1})^2/2)}{\sqrt{2\pi}} dx_1 \dots dx_n. \end{aligned}$$

En effet le changement de variables $\Psi : \{x_i\} \leftrightarrow \{e_i\}$ a les propriétés suivantes :

- Il est bijectif $\mathbb{R}^n \rightarrow \mathbb{R}^n$ (immédiat car c'est une application linéaire clairement inversible) donc le domaine est inchangé
- Le jacobien a pour déterminant 1 :

$$\text{Jac}(\Psi) = \begin{pmatrix} 1 & 0 & \dots & & \\ -a & 1 & 0 & & \\ 0 & -a & 1 & 0 & \\ & & & \ddots & \\ & & & & -a & 1 \end{pmatrix}$$

2ème méthode : Densités conditionnelles. Cette méthode n'est pas 100% certifiée MAP361 car les deux arguments ci-dessous ne peuvent pas réellement être justifiés avec les résultats du cours mais je pense qu'on peut donner les points. À vous de me dire.

Conditionnellement à $\{X_1 = x_1, \dots, X_{k-1} = x_{k-1}\}$ on a que $X_k \sim \mathcal{N}(ax_k, 1)$ donc

$$f_{X_k | X_1=x_1, \dots, X_{k-1}=x_{k-1}}(x_k) = \frac{\exp(-(x_k - ax_{k-1})^2/2)}{\sqrt{2\pi}}.$$

On en déduit le résultat en écrivant

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = f_{X_1}(x_1) \times f_{X_2 | X_1=x_1}(x_2) \times f_{X_n | X_1=x_1, \dots, X_{n-1}=x_{n-1}}(x_n).$$

3ème méthode : Vecteurs gaussiens. Ça ne me paraît pas rentable mais certains pourront vouloir faire comme ça en utilisant la formule de la densité.

Problème. [Un problème d'arrêt optimal. (9pts)]

On considère dans ce problème un jeu aléatoire pour lequel on cherche la stratégie optimale. Le jeu dépend d'un paramètre entier $n \geq 2$ et la règle est la suivante.

- On tire une variable aléatoire continue uniforme sur l'intervalle $[0, n]$;
- on peut alors accepter ce gain et le jeu s'arrête ;
- on peut refuser le gain, payer 1 et tirer à nouveau une variable aléatoire continue uniforme sur $[0, n]$ (indépendante du passé). On peut à nouveau accepter ou payer 1 et relancer, et ce autant de fois que l'on souhaite.

Nous allons chercher à déterminer une stratégie optimale pour maximiser l'espérance de gain. Pour cela on fixe un réel $s \in]0, n[$, appelé *seuil*, la stratégie consiste à accepter le premier tirage supérieur ou égal à s .

Notations. Soit $(Y_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. uniformes sur $[0, n]$ qui vont représenter les tirages successifs. On note

$$\begin{aligned}\tau(n, s) &= \min\{i \geq 1 \text{ tels que } Y_i \geq s\}, \\ X(n, s) &= Y_{\tau(n, s)}.\end{aligned}$$

Ainsi $\tau(n, s) \in \{1, 2, 3, \dots\}$ et $X(n, s) \in [s, n]$. Puisque l'on a dû payer $\tau(n, s) - 1$ fois, le gain est donné par la variable aléatoire

$$G(n, s) := X(n, s) - (\tau(n, s) - 1).$$

Pour illustrer les notations voici le résultat d'une simulation de quelques tirages avec $n = 100$:

i	1	2	3	4	5	6	7	8	9	...
tirage Y_i	67.7	70.1	12.8	44.3	86.3	77.2	4.5	50.1	97.2	...

Pour cette simulation, si le seuil vaut $s = 75$, alors on a

$$\tau(n, 75) = 5, \quad X(n, 75) = Y_5 = 86.3, \quad G(n, 75) = Y_5 - 5 + 1 = 82.3.$$

Le problème se compose de 3 parties. Vous pouvez utiliser les résultats d'une partie précédente même si vous n'avez pas réussi à les démontrer.

Partie 1 : Gain à n fixé

1. On considère le programme python suivant :

```
def SimulationGain(n,s):
    # Entrées : entier n>1 et seuil s avec 0<s<n
    NombreDeLancers=0
    Y=0
    while $$$ :
        NombreDeLancers = NombreDeLancers + 1
        Y=np.random.uniform(0,n) # Tire au sort une uniforme dans l'intervalle [0,n]
    return ***
```

Par quoi faut-il remplacer \$\$\$ et *** pour que cette fonction renvoie une simulation de la variable $G(n, s)$? (Aucune justification n'est demandée.)

2. Démontrer que $\tau(n, s)$ est une variable géométrique de paramètre de succès $(n - s)/n$.
Remarque : Ceci démontre en particulier que $\tau(n, s) < +\infty$ presque sûrement, et ainsi que $X(n, s)$ et $G(n, s)$ sont des variables aléatoires bien définies.
3. Démontrer que pour tout $x \in [s, n]$ et tout $i \in \{1, 2, 3, \dots\}$ on a

$$\mathbb{P}\left(\{X(n, s) \leq x\} \cap \{\tau(n, s) = i\}\right) = \frac{s^{i-1}(x - s)}{n^i}.$$

4. En déduire que $X(n, s)$ est uniforme sur l'intervalle $[s, n]$.
5. Démontrer que les variables aléatoires $X(n, s)$ et $\tau(n, s)$ sont indépendantes.
6. Calculer $\mathbb{E}[G(n, s)]$. Soit maintenant s_n^* le seuil tel que

$$\mathbb{E}[G(n, s_n^*)] = \max_{s \in [0, n]} \mathbb{E}[G(n, s)].$$

Vérifier que pour tout $n \geq 2$ on a $s_n^* = n - \sqrt{2n}$.

Partie 2 : Gain asymptotique

On se place maintenant dans le cas de ce seuil optimal et on va chercher à étudier le comportement asymptotique quand $n \rightarrow +\infty$ du gain, qui est donc défini par :

$$G(n, s_n^*) = X(n, s_n^*) - \tau(n, s_n^*) + 1.$$

7. Démontrer que pour tout $n \geq 2$ la variable aléatoire $X(n, s_n^*)$ a même loi que $n - U\sqrt{n}$, où U est une variable aléatoire uniforme sur l'intervalle $[0, \sqrt{2}]$.
8. Démontrer que quand $n \rightarrow +\infty$

$$\frac{1}{\sqrt{n}}\tau(n, s_n^*) \xrightarrow{\mathcal{L}} E,$$

où E est une variable aléatoire exponentielle de paramètre $\sqrt{2}$ (c'est-à-dire que E a pour densité $x \mapsto \sqrt{2} \exp(-\sqrt{2}x)$ sur \mathbb{R}_+).

9. En utilisant le résultat de l'Exercice 2 prouver que quand $n \rightarrow +\infty$

$$\frac{n - G(n, s_n^*)}{\sqrt{n}} \xrightarrow{\mathcal{L}} U + E,$$

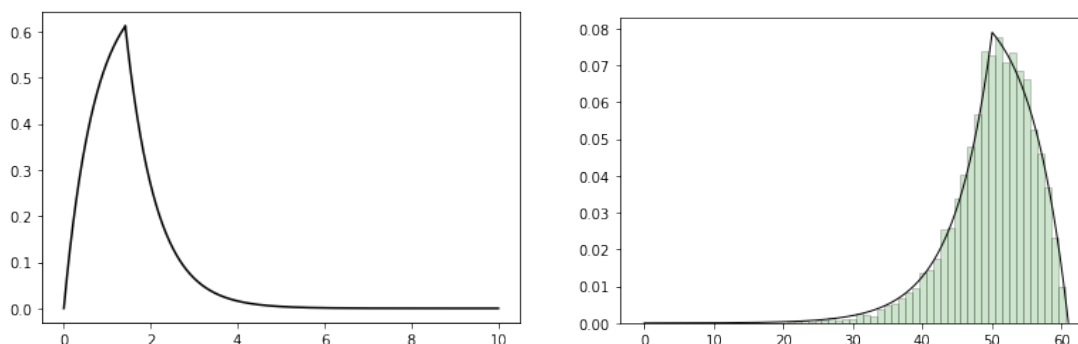
où U est une variable aléatoire uniforme sur $[0, \sqrt{2}]$, E est une variable aléatoire exponentielle de paramètre $\sqrt{2}$, et où U est indépendante de E .

Partie 3 : Étude de la loi limite

10. Justifier en utilisant un résultat du cours que la variable aléatoire $U + E$ définie précédemment admet une densité f et démontrer que f peut s'écrire

$$f(x) = \begin{cases} 0 & \text{si } x < 0, \\ \frac{1}{\sqrt{2}} (1 - \exp(-\sqrt{2}x)) & \text{si } 0 \leq x < \sqrt{2}, \\ \frac{1}{\sqrt{2}} \exp(-\sqrt{2}x) (\exp(2) - 1) & \text{si } \sqrt{2} \leq x. \end{cases}$$

Pour illustrer la convergence démontrée en Question 9 on cherche à comparer des simulations de la variable $G(n, s_n^*)$ avec la densité f . Pour cela on représente sur le même graphique l'histogramme de simulations de $G(n, s_n^*)$ ainsi que la courbe $x \mapsto \frac{1}{b_n} f\left(\frac{a_n - x}{b_n}\right)$, pour des réels $a_n, b_n > 0$ bien choisis.



À gauche : la courbe de f . À droite : un histogramme de 10000 simulations de $G(n, s_n^*)$ pour $n = 60$ et la courbe de $x \mapsto \frac{1}{b_n} f\left(\frac{a_n - x}{b_n}\right)$.

11. Pour tout n , calculer les réels a_n et b_n pour que le graphique illustre bien la convergence en loi démontrée en Question 9.

Solutions :

1. Il faut remplacer \$\$\$ par $Y < s$ et *** par $Y - \text{NombreDeLancers} + 1$.

(1pt)

2. Soit $i \geq 1$, on a

$$\mathbb{P}(\tau(n, s) = i) = \mathbb{P}(Y_1 < s, Y_2 < s, \dots, Y_{i-1} < s, s \leq Y_i) = (s/n)^{i-1}(n-s)/n,$$

qui est bien la probabilité d'une géométrique de probabilité de succès $(n-s)/n$.

3.

$$\begin{aligned} \mathbb{P}\left(X(n, s) \leq x ; \tau(n, s) = i\right) &= \mathbb{P}(Y_1 < s, Y_2 < s, \dots, Y_{i-1} < s, s \leq Y_i \leq x) \\ &= (s/n)^{i-1}(x-s)/n. \end{aligned} \tag{1}$$

(1pt) : les Y_i doivent apparaître

(0.5pt)

4. On calcule la fonction de répartition de $X(n, s)$ en utilisant la formule des probabilités totales : pour $s \leq x \leq n$

$$\mathbb{P}(X(n, s) \leq x) = \sum_{i \geq 1} (s/n)^{i-1}(x-s)/n = \frac{x-s}{n} \frac{1}{1-s/n} = \frac{x-s}{n-s},$$

qui est la fonction de répartition d'une uniforme sur $[s, n]$.

5. On déduit de la formule (1) que la fonction de répartition jointe s'écrit comme un produit. Cf Proposition 4.49 dans le Poly.

(0.5pt)

(0.5pt) : utiliser un critère d'indépendance

6. On a

$$\mathbb{E}[G(n, s)] = \frac{n+s}{2} - \frac{1}{1-s/n} + 1 =: \phi(s).$$

On dérive l'expression ci-dessus par rapport à s .

(1pt)

$$\phi'(s) = \frac{1}{2} - \frac{n}{(n-s)^2}.$$

Si $n \geq 4$ c'est positif puis négatif, et ça s'annule en $s_n^* = n - \sqrt{2n}$.

7. En utilisant ce qui précède on a que $X(n, s_n^*)$ est uniforme sur $[n - \sqrt{2n}, n]$. Calculons pour ϕ continue bornée

$$\begin{aligned} \mathbb{E}[\phi(n - \sqrt{n}U_n)] &= \int \phi(n - \sqrt{n}u) \mathbf{1}_{[0, \sqrt{2}]}(u) \frac{du}{\sqrt{2}} \\ &= \int \phi(x) \mathbf{1}_{[n - \sqrt{2n}, n]}(x) \frac{dx}{\sqrt{2n}} \\ &= \mathbb{E}[\phi(X(n, s_n^*))]. \end{aligned}$$

(0.5pt) : argument vaseux pas accepté!

8. On utilise l'anti-fonction de répartition. Soit $x \geq 0$,

(1pt)

$$\begin{aligned} \mathbb{P}(\tau(n, s_n^*) > x\sqrt{n}) &= \mathbb{P}(\tau(n, s_n^*) \geq \lfloor x\sqrt{n} \rfloor + 1) \\ &= (s_n^*/n)^{\lfloor x\sqrt{n} \rfloor} \\ &= \left(1 - \frac{\sqrt{2}}{\sqrt{n}}\right)^{\lfloor x\sqrt{n} \rfloor} = \left(1 - \frac{\sqrt{2}}{\sqrt{n}}\right)^{x\sqrt{n} + \mathcal{O}(1)} \rightarrow \exp(-\sqrt{2}x), \end{aligned}$$

qui est l'anti-fonction de répartition de l'exponentielle de paramètre $\sqrt{2}$.

9. On applique l'exercice précédent :

(1pt) : 0.5pt si pas de Slutsky

$$\begin{aligned} \frac{n - G(n, s_n^*)}{\sqrt{n}} &= \frac{n - X(n, s_n^*) + \tau(n, s_n^*) - 1}{\sqrt{n}} \\ &\stackrel{(\text{loi})}{=} \frac{n - (n - \sqrt{n}U_n) + \tau(n, s_n^*) - 1}{\sqrt{n}} \quad \text{avec indépendance de } U_n, \tau, \\ &= \underbrace{U_n}_{\rightarrow U} + \underbrace{\frac{\tau(n, s_n^*)}{\sqrt{n}}}_{\rightarrow E} - \underbrace{\frac{1}{\sqrt{n}}}_{\rightarrow 0}. \end{aligned}$$

(On utilise Slutsky pour éliminer le dernier terme.)

10. Pour le fait que $U + E$ a une densité c'est la Proposition 5.35 du cours (produit de convolution) car on a l'hypothèse d'indépendance. (1pt)

Soient maintenant g_U, g_E les densités de U, E à savoir

$$g_U(x) = \frac{1}{\sqrt{2}} \mathbf{1}_{0 \leq x \leq \sqrt{2}}, \quad g_E(x) = \sqrt{2} \exp(-\sqrt{2}x) \mathbf{1}_{0 \leq x}.$$

Alors f est donnée par la convolution

$$\begin{aligned} f(t) &= \int g_U(x) g_E(t-x) dx \\ &= \int \frac{1}{\sqrt{2}} \times \sqrt{2} \times \mathbf{1}_{0 \leq x \leq \sqrt{2}} \exp(-\sqrt{2}(t-x)) \mathbf{1}_{0 \leq t-x} dx \\ &= \exp(-\sqrt{2}t) \int \mathbf{1}_{0 \leq x \leq \min\{\sqrt{2}, t\}} \exp(\sqrt{2}x) dx. \\ &= \exp(-\sqrt{2}t) \left[\frac{1}{\sqrt{2}} \exp(\sqrt{2}x) \right]_0^{\min\{\sqrt{2}, t\}} \\ &= \begin{cases} 0 & \text{si } t < 0, \\ \frac{1}{\sqrt{2}} \exp(-\sqrt{2}t) (\exp(\sqrt{2}t) - 1) = \frac{1}{\sqrt{2}} (1 - \exp(-\sqrt{2}t)) & \text{si } 0 \leq t < \sqrt{2}, \\ \frac{1}{\sqrt{2}} \exp(-\sqrt{2}t) (\exp(2) - 1) & \text{si } \sqrt{2} \leq t \end{cases} \end{aligned}$$

11. À la Question 9 on a démontré que (1pt)

$$G(n, s_n^*) \stackrel{(loi)}{\approx} n + 1 - \sqrt{n} (U + E).$$

Donc si l'on trace sur le même graphique un histogramme de simulations de v.a. $G(n, s_n^*)$ et la densité de $n + 1 - \sqrt{n} (U + E)$ ça devrait coïncider. Or la v.a. $n + 1 - \sqrt{n} (U + E)$ a pour densité

$$x \mapsto \frac{1}{\sqrt{n}} f\left(\frac{n+1-x}{\sqrt{n}}\right),$$

on doit donc prendre $a_n = n + 1$ et $b_n = \sqrt{n}$.

NB : Évidemment on met les points s'il y a $a_n = n$ à la place de $n + 1$...