# STT1000

# Chapitre 3 Lois continues

# Solutions

Lire l'annexe C pour voir comment effectuer les calculs à l'aide d'Excel

## Loi uniforme

3.1 Soit $X$ votre temps d'attente lorsque vous arrivez à une heure aléatoire au quai du métro. Supposons que $X$ est de loi uniforme sur [0 ;10].

a) Déterminer chacune des probabilités suivantes:
i) $P(X \le 3{,}2)$ ii) $P(X > 7)$ iii) $P(2 < X \le 5)$

b) Déterminer la fonction de répartition $F$ de $X$.

c) Utiliser la fonction de répartition pour évaluer chacune des probabilités en a).

d) Utiliser la fonction de répartition pour évaluer les probabilités suivantes:
i) $P(X \le 15)$ ii) $P(X > -1)$ iii) $P(-2 < X \le 5)$ iv) $P(2 < X \le 12)$ v) $P(-2 < X \le 15)$

## Loi exponentielle

3.2 Soit $X$ la durée de vie (en milliers d'heures) d'une composante électronique. Supposons que $X$ suive une loi exponentielle de paramètre $\beta = 3$.

a) Déterminer chacune des probabilités suivantes:
i) $P(X \le 3{,}2)$ ii) $P(X > 7)$ iii) $P(2 < X \le 5)$

a) Déterminer la fonction de répartition $F$ de $X$.

c) Utiliser la fonction de répartition pour évaluer les probabilités suivantes:
i) $P(X \le 15)$ ii) $P(X \le -1)$ iii) $P(-2 < X \le 5)$

3.3 En moyenne, un moteur d'avion peut fonctionner 2000 heures sans panne. En supposant que la durée de fonctionnement suit une loi exponentielle, calculer la probabilité que le moteur ne puisse pas compléter un vol de 10 heures 0,004987

3.4 Supposons que le temps $X$ entre l'inspection d'un certain moteur d'avion et le moment de la première panne de moteur est de loi exponentielle de moyenne 1000 heures. Un avion entreprend un voyage de 25 heures après inspection de ses moteurs.

a) Quelle est la probabilité qu'il puisse terminer son vol si les 4 moteurs sont nécessaires pour voler ?

Le *nombre X* de moteurs qui dureront plus de 25 heures est une variable de loi binomiale de paramètres $n$ = 4 et $p = e^{-25/1000}$.
On cherche la $P(X \ge 4)$, ce qui peut se calculer avec la fonction de probabilité binomiale; On trouve 0,9048374

b) Quelle est la probabilité qu'il puisse terminer son vol si l'avion peut voler avec un seul moteur ?

On cherche ($P(X \ge 1)$, ce qui peut également se calculer avec la fonction de probabilité binomiale.

On trouve $P(X \ge 1) = 0{,}9999996$

3.5 Quelle est la probabilité qu'une variable de loi exponentielle prenne une valeur située à plus d'un écart-type de sa moyenne ?

Soit $X$ une variable de loi exponentielle de paramètre β. β est en même temps la moyenne et l'écart-type de $X$. Donc la probabilité recherchée est $P(X - \beta > \beta) = P(X > 2\beta) = e^{-2\beta/\beta} = e^{-2} = 0{,}1353$.

## Lois Gamma et khi-deux

3.6 La durée $X$ (en milliers d'heures) d'un appareil est une variable de loi gamma de paramètres $\alpha = 2$ et $\beta = 3$. Déterminer chacune des probabilités suivantes:

a) $P(X \le 6)$ b) $P(X > 10)$ c) $P(5 < X \le 20)$

La fonction de densité est $f(x) = \frac{1}{9} x e^{-x/3}$ $(0 < x < \infty)$.

La fonction de répartition est $F(x) = \int_0^x \frac{1}{9} t e^{-t/3} dt = 1 - e^{-x/3}\left[1 + \frac{x}{3} e^{-x/3}\right]$

a) $P(X \le 6) = F(6) = 1 - 3e^{-2} = 0{,}5399$.

b) $P(X > 10) = 1 - F(10) = 1 - [1 - \left[\frac{13}{3} e^{-10/3}\right]$ = 0,154587.

c) $P(5 < X \le 20) = F(20) - F(5) = 0{,}990243 - 0{,}463317 = 0{,}493911$.

3.7 Supposons que la durée de vie (en milliers d'heures) d'une composante électronique suive une loi exponentielle de paramètre $\beta = 1{,}2$. Un appareil fonctionne avec une seule de ces composantes, mais on dispose de 5 pièces de rechange. Quelle est la probabilité que les 6 composantes (celle déjà installée + les 5 de rechange), utilisées successivement, permettent de faire fonctionner l'appareil pour 10 000 heures ou plus.

Soit $X_1$, $X_2$, $X_3$, $X_4$, $X_5$ et $X_6$ les durées des 6 pièces. L'appareil pourra fonctionner pour 10 heures ou plus si et seulement si $X > 10$, où $X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6$. $X$ suit une loi Gamme de paramètres $\alpha = 6$ et $\beta = 1{,}2$. La probabilité $P(X > 10)$ serait pénible à calculer manuellement. Elle peut être calculée par Excel. La commande est `=LOI.GAMMA.N(10;6;1,2;1)`.

3.8 Soit $X_1$, $X_2$ et $X_3$ trois variables aléatoires indépendantes de lois normales de moyennes $\mu_1 = 3$, $\mu_2 = 5$ et $\mu_3 = 1$, respectivement; et d'écart $\sigma = 4$ (commun aux trois variables).

Soit $Q = (X_1 - 3)^2 + (X_2 - 5)^2 + (X_3 - 1)^2$. Déterminer la probabilité $P(Q \le 80)$.

$\frac{Q}{\sigma^2} = \left(\frac{X_1 - 3}{4}\right)^2 + \left(\frac{X_2 - 5}{4}\right)^2 + \left(\frac{X_3 - 1}{4}\right)^2$ est une somme de 4 variables aléatoires indépendantes de loi khi-2 à un degré de liberté. Donc $\frac{Q}{\sigma^2} \sim \chi_3^2$ et $P(Q \le 80) = P\left(\frac{Q}{\sigma^2} \le \frac{80}{16}\right)$ = 0,8282.

Ce calcul peut être effectué par Excel à l'aide de la commande

`=LOI.KHIDEUX.N(80/16;3;1)`.

## Loi normale

3.9 Sachant que $X$ suit une loi $N(100 ; 256)$, calculer chacune des probabilités suivantes.

Les probabilités $P(X \le x)$ sont calculées par Excel à l'aide de la commande

`=LOI.NORMALE(x;mu;sigma;1)`.

a) $P(100 \le X \le 150)$ 0,4991 b) $P(X \le 92)$ 0,3085 c) $P(|X-100| \le 8)$ 0,3829

d) $P(52 < X \le 148)$ 0,9974 e) $P(76 \le X \le 92)$ 0,2417 f) $P(|X-100| > 8)$ 0,6170

3.10 Soit $Z \sim N(0 ; 1)$. Déterminer la valeur de $a$ telle que

a) $P(Z > a) = 0{,}1056$ 1,25 b) $P(0 < Z < a) = 0{,}3554$ 1,0599 c) $P(Z < a) = 0{,}7734$ 0,7501

d) $P(Z < a) = 0{,}0188$ -2,0792 e) $P(|Z| < a) = 0{,}8836$ 1,5701 f) $P(|Z| > a) = 0{,}3422$ 0,9498

3.11 Soit $X \sim N(32 ; 81)$. Déterminer la valeur de $a$ telle que

a) $P(X > a) = 0{,}1056$ 43,25 b) $P(32 < X < a) = 0{,}3554$ 41,54 c) $P(X < a) = 0{,}7734$ 38,75

d) $P(X < a) = 0{,}0188$ 13,29

e) $P(|X-32| < a) = 0{,}8836$ 14,13

f) $P(|X-32| > a) = 0{,}3422$ 8,55

3.12 Supposons que la taille des hommes d'une population est distribuée normalement de moyenne $\mu = 70$ pouces et d'écart-type $\sigma = 10$ pouces. Quelle devrait être la longueur d'un matelas pour qu'il puisse accommoder 99 % des hommes de la population ?

Soit $L$ la longueur du matelas. Si $X$ est la taille d'un homme, il faut que $P(X \le L) = 0{,}99 \Rightarrow L = 93{,}26$ pouces.. Le matelas doit être de longueur supérieure ou égale à 93,26 pouces.

3.13 Lorsqu'une machine est réglée pour mettre μ grammes de petits pois dans des boîtes de conserve, elle n'en met pas exactement μ grammes. Le poids réel du contenu varie selon une loi normale de moyenne μ et d'écart-type σ grammes. Si σ = 4, à quelle valeur doit-on régler μ pour que 1% seulement des boîtes contiennent moins de 300 grammes ?

Si $X$ est le poids du contenu, il faut que $P(X < 300)$ .

$P(X < 300) = 0{,}01 \Rightarrow P(X - \mu < 300 - \mu) = 0{,}01 \Rightarrow 300 - \mu = -9{,}305\,4 \Rightarrow \mu = 309{,}305\,4$.

3.14 Soit $X_1$ et $X_2$ deux variables aléatoires indépendantes, $X_1 \sim \mathcal{N}(25 ; 16)$ et $X_2 \sim \mathcal{N}(23 ; 9)$. Calculer

a) $P(41 \le X_1 + X_2 \le 51{,}5)$

$X_1 + X_2$ est de loi normale de moyenne 48 et de variance 25. $P(41 \le X_1 + X_2 \le 51{,}5) = 0{,}6772$.

b) $P(-1{,}5 \le X_1 - X_2 \le 16)$

$X_1 - X_2$ est de loi normale de moyenne 2 et de variance 25. $P(-1{,}5 \le X_1 + X_2 \le 16) = 0{,}7554$.

c) $P(X_2 - X_1 > 0)$.

$X_2 - X_1$ est de loi normale de moyenne -2 et de variance 25. $P(X_2 - X_1 > 0) = 0{,}3446$.

3.15 Supposons que le Q. I. moyen de 10 personnes choisies au hasard d'une population normale d'écart-type σ = 10 est $\bar{X} = 110$. Calculer $P(\bar{X} \ge 110)$ et dites si la moyenne observée de 110 est trop grande pour qu'on puisse accepter l'hypothèse que la moyenne de la population vaut 100.

$\bar{X}$ est de loi $\mathcal{N}(\mu ; 100/10)$. Si μ = 100, $P(\bar{X} \ge 110) = 0{,}0008$, ce qui signifie qu'il est très improbable d'observer une moyenne aussi élevée sous l'hypothèse que μ = 100. On rejette donc cette hypothèse pour conclure que μ > 100.

3.16 Un terrain est découpé en 10 lots identiques. Sans engrais, la production de céréales, en tonnes, pour chaque lot, suit une loi $\mathcal{N}(6 ; 1)$. En utilisant un certain engrais, la production d'un lot sera de loi $\mathcal{N}(6{,}3 ; 1)$. Parmi les 10 lots, 6 sont semés sans engrais et 4 reçoivent de l'engrais. Quelle est la probabilité que les 6 lots sans engrais produisent, en moyenne, plus de céréales que les 4 lots avec engrais ?

Soit $\bar{X}$ la moyenne des 6 lots semés sans engrais ; et $\bar{Y}$ la moyenne des 4 lots semés avec engrais.

$\bar{X} \sim \mathcal{N}(6 ; 1/6)$ et $\bar{Y} \sim \mathcal{N}(6{,}3 ; 1/4)$. $P(\bar{X} > \bar{Y}) = P(\bar{X} - \bar{Y} > 0)$. Sachant que $\bar{X} - \bar{Y}$ est de loi normale de moyenne -0,3 et de variance 1/6 + 1/4, on trouve $P(\bar{X} - \bar{Y} > 0) = 0{,}3228$.

3.17 Soit $\bar{X}_1$ le Q.I. moyen de 10 personnes choisies au hasard d'une population normale de variance 100 ; soit $\bar{X}_2$ le Q.I. moyen de 15 personnes choisies au hasard d'une population de variance 121.

a) Supposons que les deux populations sont de même moyennes. Quelle est la distribution de $\bar{X}_1 - \bar{X}_2$ ?

$$\bar{X}_1 - \bar{X}_2 \sim \mathcal{N}\left(0 \ ; \ \frac{100}{10} + \frac{121}{15}\right).$$

b) Supposons qu'on observe $\bar{X}_1 - \bar{X}_2 = 20$. Cet écart est-il trop grand pour qu'on trouve acceptable l'hypothèse que les deux populations sont de même moyennes ?

$P\left(\left|\bar{X}_1 - \bar{X}_2\right| \ge 20\right) = 0{,}000002$. Si les deux populations étaient de même moyenne, il serait presque impossible que l'écart entre les deux moyennes échantillonnales soit aussi grand que 20. On conclut que les deux populations ne sont pas de même moyenne.

## Approximation de la loi binomiale

3.18 Soit $X \sim \mathcal{B}(14 ; 0{,}4)$. Calculer par l'approximation normale :

Les probabilités $P(X = x)$ et $P(X \le x)$ pour une variable de loi binomiale de paramètres $n$ et $p$ sont calculées par Excel à l'aide des commandes

`=LOI.BINOMIALE.N(x;n;p;0)` et

`=LOI.BINOMIALE.N(x;n;p;1)`,

respectivement.

a) $P(X \geq 10)$ `0,0167` b) $P(4 \leq X \leq 8)$ `0,8172` c) $P(X \geq 6)$ `0,5218` d) $P(X > 4)$. `0,7258`

3.19 Soit $X \sim \mathcal{B}(60 ; 0{,}2)$. Calculer par l'approximation normale :

a) $P(9 \leq X \leq 15)$ `0,7414` b) $P(X \geq 8)$ `0,9268`

c) $P(X > 20)$ `0,0030` d) $P(X \leq 4)$ `0,0077`.

## Exercices divers

3.20 Supposons que la durée des tubes de télévision est distribuée selon une loi exponentielle de moyenne 8 ans. Un marchand vend 10 appareils munis d'une garantie d'un an. Quelle est la probabilité que 3 appareils ou plus doivent être remplacés à cause d'un tube brûlé ?

Un appareil doit être remplacé si sa durée est de moins d'un an. La probabilité qu'un tube doive être remplacé est donc $p = 1 - e^{-1/8} = 0{,}1175$. Si $X$ est le nombre d'appareils qui doivent être remplacés, alors $X \sim \mathcal{B}(10 ; p)$. $P(X \geq 3) = 1 - P(X \leq 2) = 1- [\binom{10}{0} p^0(1-p)^{10} + \binom{10}{1} p^1(1-p)^9 + \binom{10}{2} p^2(1-p)^8] = 0{,}1034$ (0,0966 avec l'approximation normale).

3.21 Le nombre d'arrivées à un comptoir de service suit une loi de Poisson de moyenne deux par heure.

Les probabilités $P(X = x)$ et $P(X \leq x)$ pour une variable de loi de Poisson de paramètre `lambda` sont calculées par Excel à l'aide des commandes

`=LOI.POISSON.N(x;LAMBDA;0)` et `=LOI.POISSON.N(x;lambda;1)`, respectivement.

a) Quelle est la probabilité qu'aucun client n'arrive durant la première heure ? `0,1353`

b) Quelle est la probabilité qu'aucun client n'arrive durant la première heure sachant qu'aucun client n'est arrivé durant la première demi-heure ? `0,3679`

Soit $X_1$ et $X_2$ le nombre d'arrivées durant la première et la deuxième demi-heure, respectivement. Ces deux variables suivent une loi de Poisson de paramètre $\lambda = 1$, et elles sont indépendantes par les hypothèses d'une loi de Poisson.

$P(X_1 + X_2 = 0 \mid X_1 = 0) = \dfrac{P(X_1 + X_2 = 0 \; et \; X_1 = 0)}{P(X_1 = 0)} = \dfrac{P(X_1 = 0 \; et \; X_2 = 0)}{P(X_1 = 0)}$. Par l'indépendance des variables, cette probabilité est $\dfrac{e^{-1}1^2/0!e^{-1}1^2/0!}{e^{-1}1^2/0!} = e^{-1}$.

c) Si l'ouverture est à 9 heures, quelle est la probabilité que la première arrivée ait lieu entre 9 :30 heures et 10 heures ? `0,2325`

3.22 A un poste de péage, il arrive en moyenne 5 voitures par minute. Quelle est la probabilité que plus de 30 secondes s'écoulent entre deux arrivées ? Écrire la probabilité demandée en fonction d'une variable de Poisson et en fonction d'une variable exponentielle.

Soit $X$ le nombre de voitures qui passent en ½ minute. Alors $X$ suit une loi de Poisson de paramètre $\lambda = 2{,}5$ et $P(X = 0) = e^{-2{,}5} = 0{,}0821$. Soit $Y$ le temps d'attente entre deux arrivées. Alors $Y$ suit une loi exponentielle de paramètre $\beta = 1/5$ et donc $P(Y > ½) = e^{-(1/2)/(1/5)}$.

3.23 On suppose que le poids (en kg) des adultes se distribue avec une moyenne de 64 et un écart-type de 12. Dans un ascenseur, une plaque indique « Capacité maximale : 12 personnes ou 1000 kg ». Si 14 personnes s'entassent dans l'ascenseur, quelle est la probabilité que leur poids total dépasse 1000 kg ?

Le poids total est la somme $X = X_1 + \ldots + X_{14}$ des poids des 12 personnes. $X \sim \mathcal{N}(14(64) ; 14(144))$.

Alors $P(X > 1000) = 0{,}0103$.

3.24 D'une population normale de moyenne μ = 200 $ et d'écart-type σ = 30 $, un statisticien prélève un échantillon de $n$ = 25 unités afin d'estimer la moyenne (μ) de la population par la moyenne échantillonnale $\bar{X}$ .

La variable $\bar{X}$ est de loi normale de moyenne μ = 200 et d'écart-type $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{30}{5} = 6$.

a) Quelle est la probabilité qu'il surestime de plus de 4 $ la vraie moyenne. 0,2525

$P(\bar{X} > 204) = 0{,}2525$.

b) Quelle est la probabilité qu'il commette une erreur (de sous-estimation ou de surestimation) de plus de 10 $ dans son estimation ? 0,0956

$P(|\bar{X} - 200| > 10) = P(\bar{X} > 200 + 10) + P(\bar{X} > 200 - 10) = 0{,}0956$.

c) Quelle est la probabilité qu'il commette une erreur (de sous-estimation ou de surestimation) de plus de 1 % par rapport à la vraie moyenne ? 0,7389

$P\left(\frac{|\bar{X} - 200|}{200} > 0{,}01\right) = P(\bar{X} > 200 + 2) + P(\bar{X} > 200 - 2) = 0{,}73888$.

d) Quelle devrait être la taille de son échantillon pour que la probabilité d'une erreur de plus de 1 % par rapport à la vraie moyenne ne soit pas supérieure 0,05 ? 864

Il faut que $P(|\bar{X} - \mu| > 2) \le 0{,}05$. $\bar{X} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0;1)$. La condition

$P(|\bar{X} - \mu| > 2) \le 0{,}05$ est équivalente à $P\left(\left|\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}\right| > \frac{2\sqrt{n}}{\sigma}\right) \le 0{,}05$ , ou encore, par symétrie,

$P\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} > \frac{2\sqrt{n}}{\sigma}\right) \le 0{,}025$, ce qui entraîne $\frac{2\sqrt{n}}{\sigma} > 1{,}96 \Rightarrow n > \left(\frac{1{,}96\sigma}{2}\right)^2 = \left(\frac{1{,}96(30)}{2}\right)^2 = 864$.

3.25 On prélève un échantillon de taille $n$ = 30 d'une population d'écart-type σ = 18 $ pour estimer la moyenne de la population par la moyenne de l'échantillon. $\bar{X} \sim N(\mu\,;\,10{,}8)$

a) Quelle est la probabilité de surestimer la vraie moyenne de plus de 6 $ ? $P(\bar{X} - \mu > 6) = 0{,}0339$

b) Quelle est la probabilité de commettre une erreur (de sous-estimation ou de surestimation) de plus de 6 $ ? $P(|\bar{X} - \mu| > 6) = 0{,}0679$

c) Quelle doit être la taille de l'échantillon s'il faut que la probabilité d'une erreur de plus de 6 $ ne dépasse pas 0,05 ?

$P(|\bar{X} - \mu| > 6) \le 0{,}05 \Rightarrow P\left(\frac{|\bar{X} - \mu|}{18/\sqrt{n}} \ge \frac{6}{18/\sqrt{n}}\right) \le 0{,}05 \Rightarrow \frac{6}{18/\sqrt{n}} \ge 1{,}96 \Rightarrow n \ge 35$.

*Remarque* : Noter qu'il n'a pas été nécessaire de connaître la moyenne pour répondre à cette question.

3.26 Un cultivateur sème des graines de haricot par rangs de 100 graines. On suppose que 80 % des graines germeront.

a) Quel est le nombre moyen de plants obtenu par rang ?

Le nombre $X$ de plants dans un rang est de loi binomiale de paramètres $n$ = 100 et $p$ = 0,8 $\Rightarrow E(X) = 80$

b) Quelle est la variance du nombre de plants obtenu par rang ?

$\mathrm{Var}(X) = 100(0{,}8)(0{,}2) = 16$

c) Quelle est la probabilité qu'un certain rang contienne moins de 70 plants ?

$X$ se comporte comme une variable $X^*$ de loi normale de moyenne μ = 80 et de variance $\sigma^2 = 16$. On calcule donc $p_n = P(X^* \le 69{,}5) = 0{,}004332448$. Si on effectue le calcul exact avec la loi binomiale, on obtient $p_b = P(X < 70) = 0{,}006059335$.

d) Le cultivateur a semé 20 000 rangs de haricots. Combien de ces rangs devraient normalement contenir moins de 70 plants ?

Le nombre $Y$ de rangs avec moins de 70 plants est de loi binomiale de paramètres $n$ = 20 000 et $p_n$ = 0,004332448 (approximativement) ou $p_b$ = 0,006059335. Donc E($Y$) = 20 000$p_n$ = 86,64897 ou E($Y$) = 20 000$p_b$ = 121,1867. La différence est loin d'être anodine !

e) Quelle est la probabilité que plus de 100 rangs contiennent moins de 70 plants ?

Le nombre $Z$ de rangs avec moins de 70 plants est une variable de loi binomiale de paramètres $n$ = 20 000 et $p_n$ ou $p_b$. On calculera donc P(Z > 100). Les résultats divergent considérablement selon qu'on emploie $p_n$ ou $p_b$ et selon qu'on calcule P(Z > 100) exactement ou approximativement :

| | | P(Z >100) | |
|---|---|---|---|
| | | Calcul exact | Approximation normale |
| $p$ = | $p_b$ | 0,9785 | 0,9761 |
| | $p_n$ | 0,0855 | 0,0838 |

Les très grandes différences viennent du fait que l'erreur d'approximation en c) a été multipliée par 20 000.

3.27 L'épaisseur des pièces de 25 ¢ est une variable de moyenne 1,625 mm et d'écart-type 0,16 mm. Une succursale de banque qui reçoit régulièrement des dépôts sous forme de rouleaux de 40 pièces de 25 ¢ mesure la longueur de chaque rouleau déposé et rejette les rouleaux de moins de 64 mm de longueur. Quelle est la probabilité de rejeter un rouleau de 40 pièces ?

Soit $X$ la longueur d'un rouleau de 40 pièces. $X \sim \mathcal{N}(40(1{,}625)\ ;\ 40(0{,}16)^2)$.

Alors P(Rejeter un rouleau) = P($X$ < 64) = 0,1615.

3.28 Sachant que 20 % de la population étudiante sont gauchers, une administration universitaire meuble son nouvel amphithéâtre de 60 pupitres pour gauchers et 240 pupitres pour droitiers. Quelle est la probabilité que lorsque 250 étudiants entrent dans l'amphithéâtre il n'y ait pas assez de places pour tous les gauchers ?

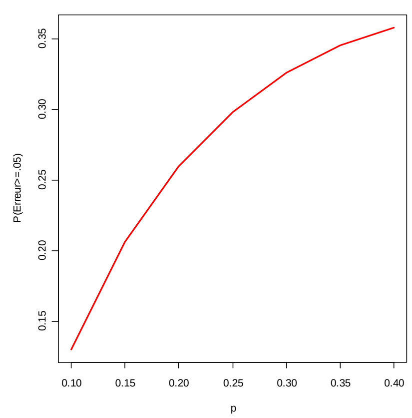
Soit $X$ le nombre de gauchers parmi les 250 étudiants. $X \sim \mathcal{B}(250\ ;\ 0{,}2)$. La probabilité voulue est P($X$ > 60) .
Par l'approximation normale, P($X$ > 60) ≈ 0,0485. La probabilité exacte est 0,0510.

3.29 D'une population d'employés, on compte prélever un échantillon de 100 personnes afin d'estimer le pourcentage $p$ de fumeurs parmi eux. On estimera $p$, bien sûr, par le pourcentage échantillonnal équivalent, soit $X$/100, où $X$ est le nombre de fumeurs dans l'échantillon. Quelle est la probabilité de se tromper d'au moins 5 points de pourcentage si la valeur *réelle* de $p$ est *a*) 10 %, *b*) 20 %, *c*) 30 % ?

Voici les valeurs de $P(|\hat{p}-p|\geq 0{,}05)$ pour quelques valeurs de $p$.

| $p$ | 0,1 | 0,15 | 0,2 | 0,25 | 0,3 | 0,35 | 0,4 |
|---|---|---|---|---|---|---|---|
| Probabilités exactes | 0,1301 | 0,2060 | 0,2599 | 0,2984 | 0,3260 | 0,3454 | 0,3584 |
| Probabilités approximatives avec ajustement | 0,1002 | 0,1655 | 0,2149 | 0,2514 | 0,2781 | 0,2972 | 0,3099 |
| Probabilités approximatives sans ajustement | 0,0956 | 0,1614 | 0,2113 | 0,2482 | 0,2752 | 0,2945 | 0,3074 |

Et en voici un graphique (avec probabilités exactes) :




3.30 Les homards pêchés sur une certaine côte ont un poids distribué selon une loi $\mathcal{N}(0{,}5\ ;\ 0{,}01)$. On les place dans des boîtes contenant 15, 16 ou 17 homards. Les fréquences de ces boîtes sont, respectivement, de 20 %, 50 % et 30 %. Déterminer la probabilité que le contenu d'une boîte choisie au hasard pèse moins de 8 livres.

Soit $X$ le nombre de homards qui se trouvent dans la boîte et $Y$ le poids total des homards dans la boîte.
$P(Y < 8) = P(X = 15) \times P(Y < 8 \mid X = 15) + P(X = 16) \times P(Y < 8 \mid X = 16) + P(X = 17) \times P(Y < 8 \mid X = 17)$
$= \sum_{x=15}^{17} P(X = x)P(Y < 8 \mid X = x)$.
Si $X = x$ alors $Y$ est de loi normale de moyenne $0,5x$ et de variance $0,01x$.

Donc $\sum_{x=15}^{17} P(X = x)P(Y < 8 \mid X = x) = \sum_{x=15}^{17} P(X = x)P\left(Z < \frac{8 - 0,5x}{\sqrt{0,01x}}\right)$.

Les trois probabilités normales $P\left(Z < \frac{8 - 0,5x}{\sqrt{0,01x}}\right)$ sont 0,9016 ; 0,5 ; 0,1126 .

Les produits $P(X = x)P\left(Z < \frac{8 - 0,5x}{\sqrt{0,01x}}\right)$ sont 0,1803 ; 0,25 ; 0,0338.

Leur somme est $\sum_{x=15}^{17} P(X = x)P\left(Z < \frac{8 - 0,5x}{\sqrt{0,01x}}\right) = 0,4641$.

3.31 Une distributrice à café est réglée de façon à verser μ ml de café dans chaque tasse. En fait, la quantité versée est variable. Elle suit une loi normale de moyenne μ et d'écart-type σ = 15 ml. Supposons qu'on fixe μ = 177 ml.

a) Quel pourcentage des tasses contiendra entre 163 ml et 192 ml de café ? 0,6651

b) Le café déborde lorsque le montant versé est supérieur à 200 ml. Avec quelle fréquence cet événement se produira-t-il ? 0,0630

c) Quelle est la probabilité que la quantité versée dans une tasse s'écarte de la moyenne fixée (177 ml) de plus de 5 ml ? 0,7414

d) Dix pour cent des tasses auront une quantité de café inférieure à $x$ ml. Quelle est la valeur de $x$ ? 157,78

e) Dix pour cent des tasses auront une quantité de café s'écartant de la moyenne fixée de plus de $d$ ml. Quelle est la valeur de $d$ ? 24,67

f) Supposons qu'on ne voudrait pas que le café déborde plus de 1 % des fois. À quoi doit-on fixer μ ? 165,10

g) Supposons qu'il est possible de contrôler σ aussi bien que μ. Déterminer μ et σ de telle sorte que la probabilité que la tasse déborde soit de 1%; et la probabilité que la tasse contienne moins de 150 ml soit de 1 %. μ=175 ; σ=10,75.

3.32 La propriétaire d'un comptoir de sushi ferme ses portes à 18 heures. À 17h00, elle se trouve avec 5 sushis non vendus. Entre 17h00 et 18h00 elle a toujours vendu au moins 2 sushis. Soit $T$ le moment (nombre d'heures après 17h00) où elle vend le 2^e^ sushi, et supposons que la fonction de densité de $T$ est $f(t) = 6t(1-t)$, $0 \le t \le 1$. Le nombre de sushis qu'elle vendra après cette heure-là est une variable de loi de Poisson de moyenne 2/heure. Déterminer l'espérance et la variance du nombre de sushis qu'elle vendra entre 17h00 et 18h00.

$E(T) = ½$; $E(T^2) = 3/10$; $Var(T) = 1/20$.
Soit $X$ le nombre de sushis vendus après le deuxième.
$E(X \mid t) = Var(X \mid t) = 2(1-t)$.
Alors $E(X) = E[E(X \mid t] = E[2(1-t)] = 1$;
$Var(X) = Var[E(X \mid t)] + E[Var(X \mid t)] = Var[2(1-t)] + E[2(1-t)] = 4/20 + 1 = 1,2$.
Le nombre de plats de sushi vendu est $Y = 2 + X$. $E(Y) = 3$; $Var(Y) = Var(X) = 1,2$.

## Tests d'hypothèses

3.33 Une standardiste prétend qu'elle reçoit en moyenne 2 appels par minutes. Le patron voudrait démontrer que le nombre d'appels par minute est inférieur à 2. Pour ce faire, il procède de la façon (peu recommandable) suivante. A un moment aléatoire de la journée, il entre chez la standardiste et attend l'arrivée du prochain appel. Il note son temps d'attente : $X = 3$ minutes. Calculer $P(X \geq 3)$ en supposant que le nombre moyen d'appels est réellement 2, et dites si on a l'évidence qu'il faut pour déclarer que le nombre moyen d'appels est inférieur à 2.

Si $X$ est le temps d'attente, $X$ est de loi exponentielle de paramètre $\beta = ½$.
$P(X \geq 3) = 0{,}00248$ sous l'hypothèse que le nombre d'appels à la minute est 2. Nous rejetons cette hypothèse.

3.34 Un acheteur de pièces de manufacture et son fournisseur sont d'accord pour déclarer «acceptable» un lot dont la proportion réelle de pièces défectueuses est de 5% ou moins. Les deux parties s'entendent pour que, à chaque livraison, on effectue une inspection selon le plan suivant : l'acheteur prélève 120 pièces au hasard et rejette le lot si 12 pièces ou plus dans l'échantillon sont défectueuses.

a) Avec le plan d'inspection proposé, quelle est la probabilité que le lot soit rejeté alors qu'il est juste acceptable (c'est-à-dire, si le pourcentage réel de pièces défectueuses est de 5 %) ?

Nous allons calculer approximativement $P(X \geq 12 \mid p = 0{,}05)$ à l'aide de la loi normale.

On calcule $P(X^* \geq 11{,}5)$, où $X^* \sim \mathcal{N}(6\,;\,5{,}7)$.

On trouve $P(X^* \geq 11{,}5) = 0{,}010\,7$. (Probabilité exacte par la loi binomiale : 0,017 3)

b) Avec le plan d'inspection proposé, quelle est la probabilité que le lot soit rejeté alors qu'il est mieux qu'acceptable, c'est-à-dire, si le pourcentage réel de pièces défectueuses est 4 % ?

Nous allons calculer approximativement $P(X \geq 12 \mid p = 0{,}04)$ à l'aide de la loi normale.

c) Supposons qu'on trouve 14 pièces défectueuses parmi les 120 inspectées, et que le fournisseur insiste quand même que le lot est acceptable, attribuant au hasard le fait que l'échantillon contient un grand nombre de pièces défectueuses. Expliquer clairement pourquoi on n'accepte pas cette argumentation.

Supposons que le fournisseur a raison, et que $p = 0{,}05$ : le lot est acceptable. Quelle serait alors la probabilité d'un nombre de pièces défectueuses aussi grand que celui observé ? C'est $P(X \geq 14 \mid p = 0{,}05) = 0{,}000\,8$ [en fait 0,002 77 si on fait le calcul exact avec la loi binomiale]. La probabilité d'avoir autant de pièces défectueuses sous cette hypothèse étant tellement minuscule, nous devons rejeter l'hypothèse que $p = 0{,}05$ et conclure que $p$ est sûrement supérieure à 0,05.

d) Selon le plan proposé, on rejette le lot si le nombre de pièces défectueuses dans l'échantillon est supérieur ou égal à 12. Modifier ce critère de façon que la probabilité de rejeter le lot alors qu'il est juste acceptable (proportion est 5 %) soit inférieure à 0,5 %.

Nous allons rejeter le lot si $X \geq C$, $C$ étant un nombre à déterminer. Il faut que $P(X \geq C \mid p = 0{,}05) < 0{,}005$ $\Rightarrow P(Z \geq (C-np)/\sqrt{120(0{,}05)(0{,}95)}) < 0{,}005 \Rightarrow P(Z \geq (C-6)\big/\sqrt{120(0{,}05)(0{,}95)}) < 0{,}005 \Rightarrow$ $(C-6)\big/\sqrt{120(0{,}05)(0{,}95)} > 2{,}576 \Rightarrow C = 12{,}15$. Puisque $C$ est forcément un entier, on pourrait songer à arrondir cette valeur à 12, mais en fait, on prendra $X \geq 13$ comme critère de rejet. Car si on prenait $X \geq 12$ comme critère, la probabilité de rejet (sous l'hypothèse que $p = 0{,}05$) serait approximativement $P(X^* \geq 11{,}5)$, et cette probabilité est supérieure à 0,005, puisque, comme on vient de conclure, $P(X^* \geq 12{,}15) = 0{,}005$.

3.35 Pour contrôler la qualité de certains sachets pharmaceutiques, on décide de tirer de chaque lot un échantillon de 16 sachets et prendre note des poids (en *mg*) $X_1, X_2, \ldots, X_{16}$. Supposons que $X_i \sim \mathcal{N}(\mu\,;\,144)$, $\mu$ inconnue. Le poids moyen dans la population ne doit pas être supérieur à $\mu = 50$. Aussi, on décide de rejeter le lot si $\bar{X} > C$, où $\bar{X}$ est la moyenne échantillonnale.

a) Quelle doit être la valeur de $C$ s'il faut que la probabilité de rejeter un lot pour lequel $\mu = 50$ soit de 5 % ?

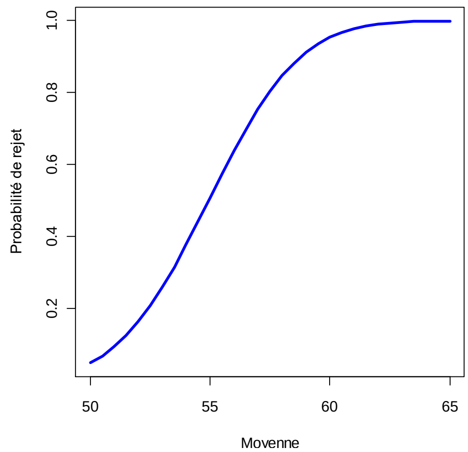
On rejettera le lot si $\bar{X} > 54{,}935$

b) Avec la valeur de $C$ trouvée en a), déterminer la probabilité de rejeter un lot pour lequel $\mu$ = 51, 55, 60, 70.

Voici quelques valeurs de $\mu$ et la probabilité de rejeter le lot étant donné $\mu$

| $\mu$ | 50 | 55 | 60 | 65 | 70 |
|---|---|---|---|---|---|
| P(rejeter\|$\mu$) | 0,05 | 0,5086 | 0,9543 | 0,9996 | 0,99999974 |

Voici un graphique de la probabilité de rejeter le lot, en fonction de la moyenne $\mu$ :




## Exercices théoriques

3.36 Soit $X_1$, $X_2$ et $X_3$ trois variables aléatoires indépendantes, chacune de loi $N(\mu;\sigma^2)$. Soit $X = X_1 + 2X_2 - 2X_3$. Quelle est la distribution de $(X - \mu)^2/(9\sigma^2)$ ?

$X$ est de loi normale de moyenne $\mu_X = (1 + 2 - 2)\mu = \mu$ et de variance $\sigma_X^2 = (1 + 4 + 4)\sigma^2 = 9\sigma^2$. Donc $\frac{X-\mu_X}{\sigma_X} = \frac{X-\mu}{\sqrt{9\sigma^2}} \sim N(0;1)$ . Par conséquent, $\frac{(X-\mu)^2}{9\sigma^2} \sim \chi_1^2$ .

3.37 Soit $X_1, \ldots, X_n$, chacune de fonction de densité $f(x) = \frac{1}{\beta}e^{-x/\beta}$ , $x \geq 0$, $\beta > 0$.

a) Soit $Y = \max\{X_1, \ldots, X_n\}$. Déterminer la fonction de répartition et la fonction de densité de $Y$.

La fonction de répartition de $Y$ est $G(y) = P(Y \leq y) = P(X_1 \leq y ; X_2 \leq y ; \ldots , Y_n \leq y) = \left[1-e^{-y/\beta}\right]^n$. La fonction de densité est donc $g(y) = G'(y) = n\left[1-e^{-y/\beta}\right]^{n-1}\frac{1}{\beta}e^{-y/\beta}$ .

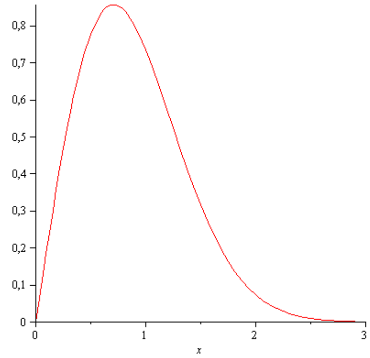
b) Soit $Z = \min\{X_1, \ldots, X_n\}$. Déterminer la fonction de répartition et la fonction de densité de $Z$.

La fonction de répartition de $Y$ est $G(z) = P(Y \leq z) = 1- P(Y > z) = 1-P(X_1>z ; X_2 > z ; \ldots , Y_n > z) = 1- \left[e^{-z/\beta}\right]^n$ $= 1- e^{-nz/\beta}$ . C'est la fonction de répartition d'une variable de loi exponentielle de paramètre $\beta/n$. La fonction de densité est donc $g(z) = G'(z) = \frac{n}{\beta}e^{-nz/\beta}$ .

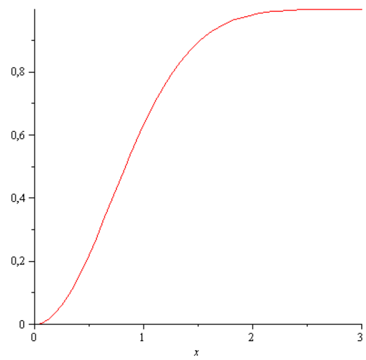
3.38 Pour chacune des fonctions de densité suivantes, déterminer la fonction de répartition et faites-en un graphique.

a) $f(x) = \begin{cases} 2xe^{-x^2} & x > 0 \\ 0 & \text{sinon} \end{cases}$

*Fonction de densité* :

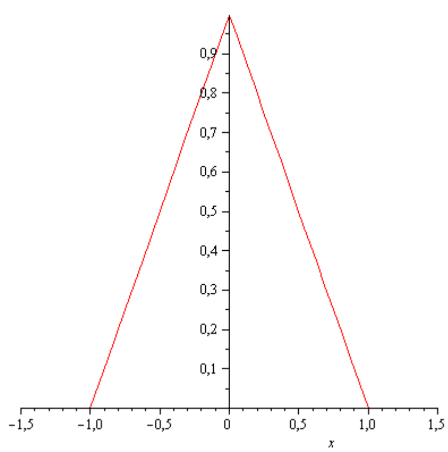



*Fonction de répartition :*

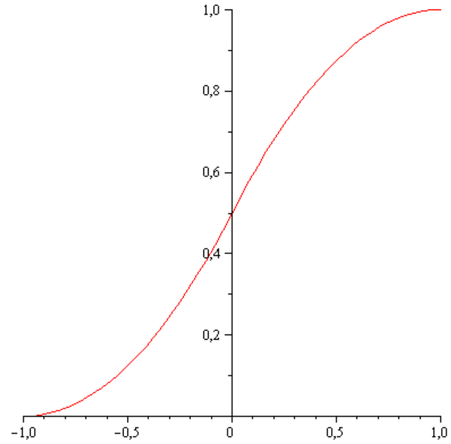



$$b) \quad f(x) = \begin{cases} 1+x & -1 < x \leq 0 \\ 1-x & 0 < x < 1 \\ 0 & \text{Ailleurs} \end{cases}$$

*Fonction de densité :*

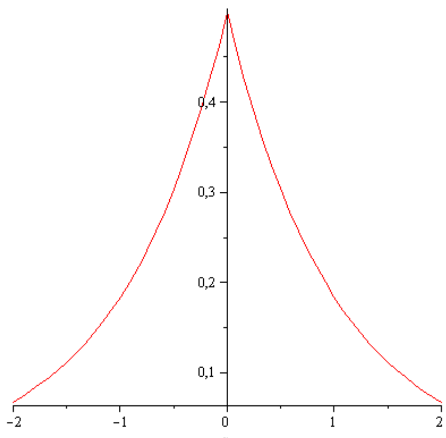


*Fonction de répartition :*

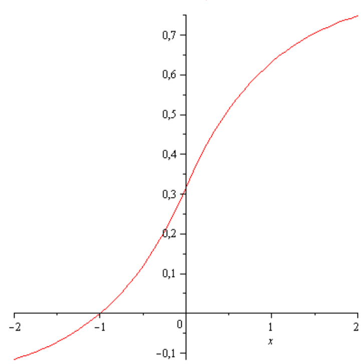



3.39 Soit $X$ une variable aléatoire continue dont la fonction densité est $f(x) = (1/2)e^{-|x|}$, $\forall\ x \in \mathbb{R}$

a) Vérifier que $f$ est bien une fonction de densité




b) Déterminer la fonction de répartition de $X$ : $\begin{cases} \frac{1}{2}e^{x} - \frac{1}{2}e^{-1} & x < 0 \\ 1 - \frac{1}{2}e^{-x} - \frac{1}{2}e^{-1} & x > 0 \end{cases}$

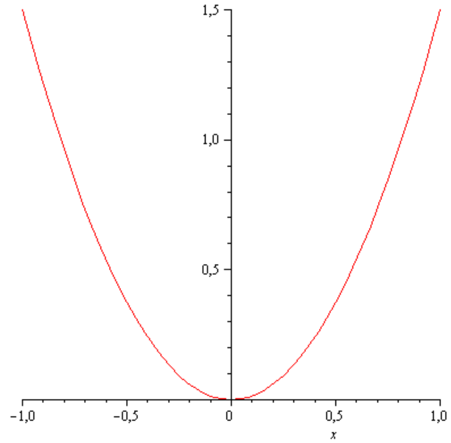



c) Calculer P(|$X$| > 2) $e^{-1}+e^{-2}$ , P(|$X$| ≤ 1) $1-e^{-1}$ , P($X$ ≤ −3) $(1/2)(e^{-3}-e^{-1})$

3.40 Soit $X$ une variable aléatoire continue dont la fonction densité est

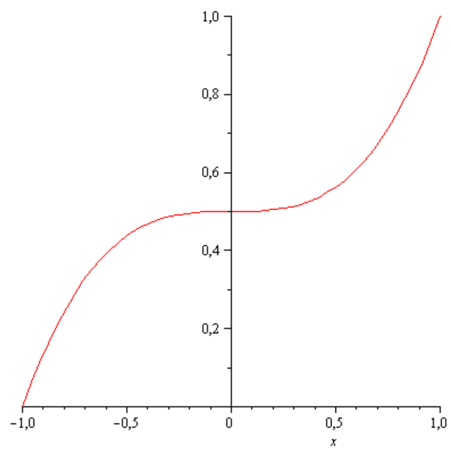
$$f(x) = \begin{cases} \left\{ \frac{3}{2}x^2 \right. & -1 < x < 1 \\ 0 & ailleurs \end{cases}$$

*Voici un dessin de la fonction de densité*




Déterminer la fonction de répartition. $\begin{cases} 0 & x<-1 \\ \frac{1}{2}(x^3+1) & -1\le x\le 1 \\ 1 & x>1 \end{cases}$

*Fonction de répartition*




3.41 Soit $X$ une variable aléatoire de densité $f(x) = \begin{cases} 1 & 0<x<1 \\ 0 & ailleurs \end{cases}$.

Déterminer la fonction de densité $g(y)$ pour $Y = \ln X$.

Soit $F(x)$ la fonction de répartition de $X$ ($F(x) = x$ sur $(0\,;1]$) et $G$ celle de $Y$.

$G(y) = P(Y \le y) = P(\ln X \le y) = P(X \le e^y) = F(e^y) = e^y$ lorsque $0 < e^y \le 1 \Leftrightarrow -\infty < y \le 0$.

La fonction de densité est $g(y) = e^y$ lorsque $-\infty < y \le 0$.

3.42 Soit $f(x) = \begin{cases} e^{-x} & x>0 \\ 0 & ailleurs \end{cases}$. Déterminer la fonction de densité $g(y)$ pour $Y = e^{-X}$.

Soit $F(x)$ la fonction de répartition de $X$ ($F(x) = 1 - e^{-x}$ sur $[0\,;\infty)$) et $G$ celle de $Y$. $G(y) = P(Y \le y) = P(e^{-X} \le y) = P(X \ge -\ln y) = 1-F(-\ln y) = e^{\ln y} = y$ lorsque $0 < -\ln y \Leftrightarrow 0 < y \le 1$. $Y$ est donc uniforme sur $(0\,;1]$.

3.43 La durée $T$ d'un certain type d'appel téléphonique satisfait $P(T > t) = ce^{-\alpha t} + (1-c)e^{-\beta t}$, $t > 0$, où $0 \le c \le 1$, $\alpha > 0$, $\beta > 0$ sont des constantes. Déterminez $E(T)$ et $Var(T)$.

La fonction de densité de $T$ est $f(t) = c\alpha e^{-\alpha t} + (1-c)\beta e^{-\beta t}$ , $t > 0$. E($T$) = $\int_0^\infty \left[c\alpha e^{-\alpha t} + (1-c)\beta e^{-\beta t}\right]dt = \frac{c}{\alpha} + \frac{1-c}{\beta}$.

La variance est E($T^2$)-[E($T$)]$^2$ = $\int_0^\infty \left[ce^{-\alpha t} + (1-c)e^{-\beta t}\right]dt - \left[\frac{c}{\alpha} + \frac{1-c}{\beta}\right]^2 = \left[\frac{2c}{\alpha^2} + \frac{2(1-c)}{\beta^2}\right] - \left[\frac{c}{\alpha} + \frac{1-c}{\beta}\right]^2$ [Pour l'espérance, nous pouvons utiliser le résultat du numéro suivant. E($X$) = $\int_0^\infty \left[ce^{-\alpha t} + (1-c)e^{-\beta t}\right]dt = \frac{c\beta + (1-c)\alpha}{\alpha\beta}$ .]

3.44 Soit $X$ une variable aléatoire continue dont le support est l'ensemble des réels positifs et $F$ est la fonction de répartition. Montrez que E($X$) = $\int_0^\infty [1-F(x)]dx$ . [*Suggestion* : Exprimez cette intégrale comme une intégrale double, puis changez l'ordre d'intégration].

$$\int_0^\infty [1-F(x)]dx = \int_0^\infty \int_x^\infty f(t)dtdx = \int_0^\infty f(t)\int_0^t dxdt = \int_0^\infty tf(t)dt = \mathrm{E}(X)$$

3.45 Soit $X$ une variable aléatoire de fonction de densité $f(x) = xe^{-x}$, $x > 0$.

a) Montrez que $f$ est une fonction de densité, et déterminez la fonction de répartition de $X$.

Il est évident que $f(x) \geq 0$ pour tout $x > 0$. Il faut maintenant montrer que $\int_0^\infty f(t)dt$ =1. On intègre par parties, en posant $u = x$, $dv = e^{-x}dx \Rightarrow du = dx$ et $v = -e^{-x}$ : $\int_0^\infty f(t)dt = -xe^{-x}\Big|_0^\infty + \int_0^\infty e^{-x}dx$ = 0+1 = 1.

On peut montrer le résultat suivant en intégrant par parties :

$\int_0^\infty x^{\alpha-1}e^{-x/\beta}dx = \Gamma(\alpha)\beta^\alpha = (\alpha-1)!\beta^\alpha$ si $\alpha$ est un entier positif.

b) Déterminez E($X$) et Var($X$).

En utilisant le résultat général énoncé en a), on a que E($X$) = $\int_0^\infty x^2e^{-x}dx = \Gamma(3)1^2 = 2$, E($X^2$) = $\int_0^\infty x^3e^{-x}dx = 3! =$ 6 et Var($X$) = 6-4 = 2.

3.46 La probabilité d'atteindre une région quelconque $R$ sur une cible circulaire de rayon 1 m est proportionnelle à l'aire de $R$. Soit $X$ la distance entre le point atteint par un projectile et le centre de la cible. Déterminez la fonction de répartition, la fonction de densité, la variance et l'espérance de $X$.

P($X \leq x$) = $\frac{\pi x^2}{\pi} = x^2$, $0 \leq x \leq 1$. La fonction de densité est $f(x) = F'(x) = 2x$, $0 \leq x \leq 1$.

3.47 Soit $X$ une variable aléatoire de fonction de répartition

$$F(x) = \begin{cases} 1-(1+x)e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Il est évident que $F(x) = 0$ pour $x \leq 0$, que $F(x)$ tend vers 1 lorsque $x \to \infty$. Pour montrer que $F$ est croissante, on détermine la dérivée, $f(x) = F'(x) = xe^{-x} > 0$ sur $(0 ; \infty)$ et $f(x) = 0$ ailleurs.
$f(x) = F'(x) = xe^{-x}$, $x > 0$.

3.48 Considérez une pièce électronique dont la durée de vie (en heures) est une variable de loi exponentielle dont la fonction de densité est :

$$f(t) = \lambda e^{-\lambda t}, \lambda > 0.$$

Le temps d'attente moyen est E($T$) = $1/\lambda$.
Soit $G$ la fonction de répartition de $U_1$. $G(u)$ = P($U_1 \leq u$) = 1-P($U_1 > u$) =1 - P(toutes les pièces ont une durée $>u$) = 1 - $(e^{-\lambda u})^5 = 1-e^{-5\lambda u}$. Il s'agit donc d'une variable de loi exponentielle de paramètre $1/5\lambda$. Son espérance est donc $1/5\lambda$, sa variance $1/25\lambda^2$.
Soit $G$ la fonction de répartition de $U_2$. $G(u)$ = P($U_2 \leq u$) =P(toutes les pièces ont une durée $\leq u$) = $(1-e^{-\lambda u})^5$. La fonction de densité est $5\lambda(1-e^{-\lambda u})^4e^{-\lambda u}$.
Soit $X$ le nombre de pièces dont la durée est supérieure à $u$. Alors $X \sim \mathcal{B}(5 ; p)$, où $p = e^{-\lambda u}$. La fonction de répartition de $U_3$ est

$G(u)$ = 1-P($U_3 > u$) =1- P($X \geq 2$) = 1- $\binom{5}{0}\left(e^{-\lambda u}\right)^0\left(1-e^{-\lambda u}\right)^5 + \binom{5}{1}\left(e^{-\lambda u}\right)^1\left(1-e^{-\lambda u}\right)^4 + \binom{5}{2}\left(e^{-\lambda u}\right)^2\left(1-e^{-\lambda u}\right)^3$.

3.49 Démontrer que $\Gamma(1/2) = \sqrt{\pi}$. *Indice* : Dans l'expression $\Gamma(1/2) = \int_0^\infty x^{-1/2}e^{-x}dx$, faites la transformation $u = \sqrt{2x}$ et utiliser le fait que $\int_0^\infty e^{-u^2/2}dy = \sqrt{\pi/2}$.

$$\int_0^\infty x^{-1/2}e^{-x}dx = \int_0^\infty \left(\frac{u^2}{2}\right)^{-1/2} e^{-u^2/2}du = \frac{1}{2}\int_0^\infty \left(u^2\right)^{-1/2} e^{-u^2/2}udu = \frac{1}{2}\int_0^\infty e^{-u^2/2}du = \frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}}.$$

3.50 Soient $X_1$, …, $X_n$, $n$ variables aléatoires indépendantes de même loi $N(\mu\,;\sigma^2)$. Montrer que

$$\sum_{i=1}^n \frac{(X_i-\mu)^2}{\sigma^2} \sim \chi_n^2.$$

$X_i \sim N(\mu\,;\sigma^2) \Rightarrow Y_i = \frac{X_i-\mu}{\sigma} \sim N(0\,;1) \Rightarrow \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_1^2$ par le théorème 3.8.1, et $\sum_{i=1}^b \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2$ par le théorème 8.8.3.

3.51 Soit $X_1$, …, $X_n$, $n$ variables aléatoires indépendantes de loi exponentielle de paramètre $\beta$. Montrer que $X = 2\sum_{i=1}^n \frac{X_i}{\beta}$ suit une loi $\chi^2$ à $2n$ degrés de liberté.

Il suffira de démontrer que si $X$ est de loi exponentielle de paramètre $\beta$, alors $Y = 2X/\beta$ est de loi $\chi_2^2$.

La fonction de répartition de $X$ est $F(x) = 1-e^{-x/\beta}$ sur $x > 0$.

La fonction de répartition de $Y$ est $G(y) = P(Y \leq y) = P(X \leq \beta y/2) = F(\beta y/2) = 1-e^{-\beta y/2\beta} = 1-e^{-y/2}$, d'où la fonction de densité de $Y$ est $\frac{1}{2}e^{-y/2}$, ce qui est bien la fonction de densité d'une variable de loi $\chi_2^2$. Donc $2X_i/\beta \sim \chi_2^2$ et $\sum_{i=1}^n \frac{2X_i}{\beta} \sim \chi_{2n}^2$ par le théorème 3.8.2.

3.52 Soit $X$ une variable aléatoire qui prend seulement des valeurs positives. Montrer que

$$P(\sqrt{2X} - \sqrt{2n} \leq k) = P[(X-n)/\sqrt{2n} \leq k + k^2/(2\sqrt{2n})]$$

si $n$ est grand. Utiliser ce résultat pour montrer que si $X \sim \chi_n^2$ et $n$ est grand, alors la distribution de $\sqrt{2X} - \sqrt{2n}$ s'approche d'une $N(0\,;1)$.

$$P(\sqrt{2X}-\sqrt{2n} \leq k) = P(\sqrt{2X} \leq k+\sqrt{2n}) = P(2X \leq k^2+2k\sqrt{2n}+2n) = P(2X \leq k^2+2k\sqrt{2n}+2n)$$

$$= P(2X-2n \leq k^2+2k\sqrt{2n}) = P\left(\frac{X-n}{\sqrt{2n}} \leq k + \frac{k^2}{2\sqrt{2n}}\right).$$

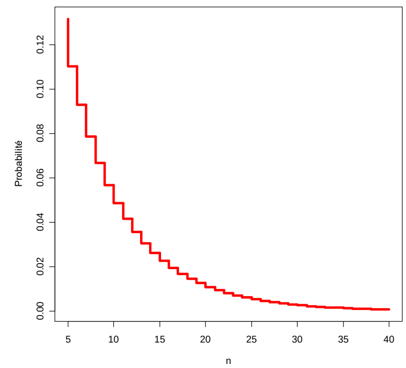
Si $X \sim \chi_n^2$, $X$ est de moyenne $n$ et de variance $2n$, et $X$ peut s'exprimer comme une somme de $n$ variables aléatoires indépendantes de loi $\chi_1^2$. Donc $\frac{X-n}{\sqrt{2n}}$ tend vers une loi normale $N(0\,;1)$ lorsque $n$ tend vers l'infini. Alors

$$P(\sqrt{2X}-\sqrt{2n} \leq k) = P\left(\frac{X-n}{\sqrt{2n}} \leq k + \frac{k^2}{2\sqrt{2n}}\right) \approx P\left(\frac{X-n}{\sqrt{2n}} \leq k\right) \approx \Phi(k)$$ lorsque $n$ est grand.

## Exercices à faire avec l'ordinateur

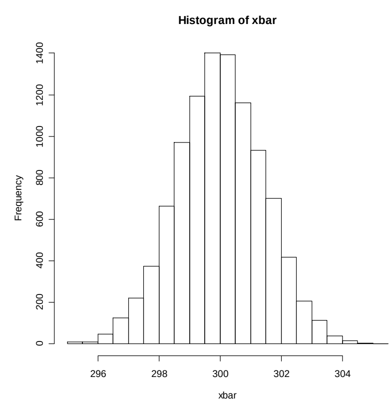
3.53 Le poids du contenu des boîtes de conserves dans une certaine usine est de moyenne $\mu = 300\ g$ et d'écart-type $\sigma = 4\ g$. Soit $\overline{X}$ le poids moyen d'un échantillon de $n$ boîtes. Calculez la probabilité que $\overline{X}$ soit supérieure à 302 pour des valeurs de $n$ allant de 10 à 200. Faites le graphique qui exprime la probabilité calculée en fonction de $n$. Expliquez en vos propres mots pourquoi la fonction est décroissante.

Nous choisissons de considérer de plus petites valeurs de *n*, soit *n* de 5 à 20.




3.54 Le poids du contenu des boîtes de conserves dans une certaine usine est de moyenne μ = 300 *g* et d'écart-type σ = 4 *g*. Soit $\bar{X}$ le poids moyen d'un échantillon de 8 boîtes. Calculez la probabilité que $\bar{X}$ soit supérieure à 302 ; puis estimez cette même probabilité par simulation.

$P(\bar{X} > 302) =$ 0,0786496. J'ai généré 1000 échantillons de taille 8, calculé la moyenne de chacun. La proportion des échantillons dont la moyenne est supérieure à 302 est 0,0795. C'est notre estimation de la probabilité $P(\bar{X} > 302)$ .Voici l'histogramme des moyennes de 10000 échantillons de taille 8 :



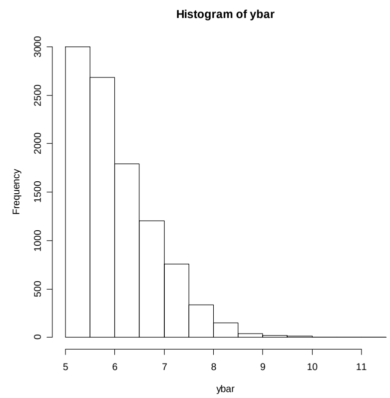

3.55 On note la pointure des chaussures de *n* personnes tirées avec remise d'une population dans laquelle la distribution des pointures est la suivante :

| Pointure | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Fréquence | 0,45 | 0,30 | 0,10 | 0,06 | 0,03 | 0,02 | 0,02 | 0,01 | 0,01 | 1 |

Soit $\bar{X}$ la moyenne d'un échantillon de taille 4.

a) Simulez l'expérience 500 fois (ou plus, si possible). La distribution de $\bar{X}$ a-t-elle l'air normale ?

Loin de là (l'échantillon est beaucoup trop petit) :




b) Calculez la moyenne des $\overline{X}$ obtenus. Théoriquement, à quoi cette moyenne devrait-elle être égale ?

La moyenne est 6,1619. Théoriquement, elle devrait être égale à 3,17.

c) Calculez la variance des $\overline{X}$ obtenus. Théoriquement, à quoi cette moyenne devrait-elle être égale ?

La variance est 0.6604169. Théoriquement elle devrait égaler 0,675275.

d) Soit $Z = (\overline{X} - \mu)/\sigma_{\overline{X}}$ . Calculez $P(|Z| \geq c)$ avec $c$ = 0,84 ; 1,04 ; 1,28 ; 1,645 ; 1,96 ; 2,237 ; 2,576

(i) en utilisant le théorème limite central

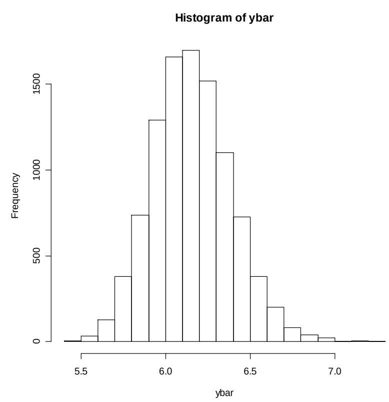
0,401 0,298 0,201 0,100 0,050 0,025 0,010

(ii) à l'aide de votre simulation.

0,3305 0,2771 0,1690 0,0560 0,0355 0,0225 0,0147

e) Répétez les exercices a) à d) avec $n$ = 50, et résumez les différences que vous avez observées entre les deux cas ($n$ = 4 et $n$ = 50).

La moyenne et la variance sont 6,16614 et 0,0535451, respectivement. La variance est théoriquement 0,054022. Les fréquences empiriques sont 0,3788 ; 0,2936 ; 0,1887 ; 0,1007 ; 0,0464 ; 0,0241 ; 0,0110, bien plus proches des fréquences théoriques. La distribution est bien plus proche d'une normale :




## Statistiques d'ordre

Dans ce qui suit, $X_1, \ldots X_n$ est un échantillon aléatoire simple et $Y_1 \leq \ldots \leq Y_n$ sont les statistiques d'ordre.

3.56 Soit $X_1$ et $X_2$ deux variables aléatoires indépendantes d'une loi $\mathfrak{U}(0 ; 1)$ (uniforme sur [0 ; 1]).

a) Déterminer la fonction de densité de $Y_1$ (la plus petite valeur). $2(1-y_1),\ 0 < y_1 < 1$.

b) Déterminer $E(Y_1)$ et $Var(Y_1)$. 1/3 ; 1/18

c) Déterminer la fonction de densité de $Y_2$ (la plus grande valeur). $2y_2,\ 0 < y_2 < 1$

d) Déterminer $E(Y_2)$ et $Var(Y_2)$. 2/3 ; 1/18

e) Déterminer la probabilité que dans un échantillon de taille $n = 4$ d'une population uniforme sur (0 ; 1) la plus petite valeur soit ≥ 0,2.

La plus petite valeur est supérieure à 0,2 si et seulement si toutes les valeurs sont supérieures à 0,2. La probabilité qu'une valeur soit supérieure à 0,2 est 0,8 ; et la probabilité que les quatre valeurs soient supérieures à 0,2 et $(0,8)^4 = 0,4096$.

3.57 Le temps d'attente pour l'arrivée du métro est une variable de loi $\mathfrak{U}(0 ; \theta)$ (uniforme sur $[0 ; \theta]$). Vous prenez le métro $n$ fois. Soit $X_1, X_2, \ldots, X_n$.

a) Déterminer la fonction de répartition de $Y_n$.

$$G(y) = P(Y_n \le y) = [F(y)]^n = \left[\frac{y}{\theta}\right]^n, 0 \le y \le \theta$$

b) Déterminer la fonction de densité de $Y_n$.

$$g(y) = G'(y) = n[F(y)]^{n-1} f(y) = \frac{n}{\theta^n} y^{n-1}, 0 \le y \le \theta$$

c) Déterminer l'espérance et la variance de $Y_n$.

$$E(Y_n) = \frac{n\theta}{n+1}\ ;\ \mathrm{Var}(Y_n) = \frac{n\theta^2}{(n+2)(n+1)^2}$$

$$E(Y_n) = \int_0^\theta y \frac{n}{\theta^n} y^{n-1} dy = \frac{n\theta}{n+1}\ ;\ E(Y_n^2) = \int_0^\theta y^2 \frac{n}{\theta^n} y^{n-1} dy = \frac{n\theta^2}{n+2}\ ;$$

$$\mathrm{Var}(Y_n) = E(Y_n^2) - [E(Y_n)]^2 = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

d) Déterminer la fonction de répartition de $Y_1$.

$$H(z) = P(Y_1 \le z) = 1 - P(Y_1 > z) = 1 - [1 - F(z)]^n = 1 - \left[1 - \frac{z}{\theta}\right]^n, 0 \le z \le \theta.$$

e) Déterminer la fonction de densité de $Y_1$.

$$h(z) = H'(z) = \frac{n}{\theta}\left[1 - \frac{z}{\theta}\right]^{n-1}, 0 \le z \le \theta.$$

f) Déterminer l'espérance et la variance de $Y_1$.

$$E(Y_1) = \int_0^\theta z \frac{n}{\theta}\left(1 - \frac{z}{\theta}\right)^{n-1} dy = \frac{\theta}{n+1}\ ; E(Y_1^2) = \int_0^\theta z^2 \frac{n}{\theta}\left(1 - \frac{z}{\theta}\right)^{n-1} dy = \frac{2\theta^2}{(n+2)(n+1)^2}$$

$$\mathrm{Var}(Y_1) = E(Y_1^2) - [E(Y_1)]^2 = \frac{n\theta^2}{(n+2)(n+1)^2}$$

d) Vous avez pris le métro 20 fois

i) La plus longue attente a été de 7 minutes. Est-ce que cette valeur est compatible avec l'hypothèse que $\theta = 10$? Calculer $P(Y_{20} \le 7)$ sous cette hypothèse et concluez que ce résultat n'est pas compatible avec l'hypothèse. Devriez-vous conclure que $\theta > 10$ ou que $\theta < 10$?

$P(Y_{20} \le 7) = 0,000798$

ii) La plus courte attente a été de 4 minutes. Est-ce que cette valeur est compatible avec l'hypothèse que $\theta = 10$? Calculer $P(Y_1 \ge 4)$ sous cette hypothèse et concluez que ce résultat n'est pas compatible avec l'hypothèse. Devriez-vous conclure que $\theta > 10$ ou que $\theta < 10$?

$P(Y_1 \ge 4) = 0,000037$

3.58 Soit $X_1, \dots, X_n$ un échantillon aléatoire simple d'une population exponentielle de paramètre β.

a) Montrer que $Y_1$ est de loi exponentielle de paramètre $\beta/n$.
La fonction de densité $g$ de $Y_1$ est

$$g(y) = nf(y)[1-F(y)]^{n-1} = n\frac{e^{-y/\beta}}{\beta}[e^{-y/\beta}]^{n-1} = \frac{e^{-\frac{y}{\beta/n}}}{\beta/n}$$, où $f$ est la fonction de densité exponentielle.

b) Montrer que la densité de $Y_n$ est $h(z) = \frac{n}{\beta}e^{-z/\beta}[1-e^{-z/\beta}]^{n-1}$.

La fonction de densité $h$ de $Y_n$ est $h(z) = nf(z)[F(z)]^{n-1} = \frac{n}{\beta}e^{-y/\beta}[1-e^{-y/\beta}]^{n-1}$

c) Déterminer $\mathrm{E}(Y_n)$ lorsque $n = 2$. $E(Y_n) = \int_0^\infty z\frac{2}{\beta}e^{-z/\beta}[1-e^{-z/\beta}]dz = \frac{3\beta}{2}$

d) Montrer que pour un échantillon de taille $n = 2m+1$ la densité de la médiane $\tilde{x}$ est

$$h(\tilde{x}) = \frac{(2m+1)!}{m!m!\beta}e^{-(m+1)\tilde{x}/\beta}[1-e^{-\tilde{x}/\beta}]^m$$

$$g(\tilde{x}) = \lim_{\Delta\tilde{x}\to 0}\frac{\mathrm{P}(\tilde{x} < \tilde{X} < \tilde{x}+\Delta\tilde{x})}{\Delta\tilde{x}}$$

$$= \lim_{\Delta\tilde{x}\to 0}\frac{\mathrm{P}(\text{Une observation entre } \tilde{x} \text{ et } \tilde{x}+\Delta\tilde{x};\ m \text{ observations} < \tilde{x}; m \text{ observations} > \tilde{x}+\Delta\tilde{x})}{\Delta\tilde{x}}$$

$$= \lim_{\Delta\tilde{x}\to 0}\frac{[F(\tilde{x}+\Delta\tilde{x})-F(\tilde{x})]\times[F(\tilde{x})]^m\times[1\text{-}F(\tilde{x}+\Delta\tilde{x})]^m)}{\Delta\tilde{x}}$$

$$= \frac{(2m+1)!}{m!m!}\lim_{\Delta\to 0}\frac{[F(\tilde{x}+\Delta\tilde{x})-F(\tilde{x})]}{\Delta\tilde{x}}\times[F(\tilde{x})]^m\times[1\text{-}F(\tilde{x}+\Delta\tilde{x})]^m)$$

$$= \frac{(2m+1)!}{m!m!}\lim_{\Delta\to 0}\frac{[F(\tilde{x}+\Delta\tilde{x})-F(\tilde{x})]}{\Delta\tilde{x}}\times[F(\tilde{x})]^m\times[1\text{-}F(\tilde{x})]^m) = \frac{(2m+1)!}{m!m!}f(\tilde{x})\times[F(\tilde{x})]^m\times[1\text{-}F(\tilde{x})]^m).$$

Le facteur $\frac{(2m+1)!}{m!m!} = \binom{2m+1}{m}\binom{m+1}{1}\binom{m}{m}$ est le nombre de façons de choisir les $m$ observations inférieures à $\tilde{x}$, celle dans un élément autour de $\tilde{x}$ et finalement les $m$ observations supérieures à $\tilde{x}$.

3.59 Soit $X_1, \dots, X_n$ un échantillon aléatoire simple d'une population uniforme sur (0 ; 1). Supposons que $n$ est impair, $n = 2m+1$. Déterminer la distribution de la médiane $\tilde{x}$.

Selon le numéro précédent, la densité de la médiane est $g(\tilde{x}) = \frac{(2m+1)!}{m!m!}f(\tilde{x})\times[F(\tilde{x})]^m\times[1\text{-}F(\tilde{x})]^m$.

La densité $\mathcal{U}(0\,;1)$ est $f(x) = 1$ et la fonction de répartition $F(x) = x$ sur (0 ; 1).

D'où $g(\tilde{x}) = \frac{(2m+1)!}{m!m!}\tilde{x}^m(1\text{-}\tilde{x})^m$ sur (0 ; 1).

3.60 Soit $X_1, \dots, X_n$ un échantillon aléatoire simple d'une population de densité $f$.

a) Montrer que la densité conjointe de $(Y_1\,;Y_n)$ est donnée par

$$g(y_1, y_n) = n(n-1)f(y_1)f(y_n)\left[F(y_n)-F(y_1)\right]^{n-2} \quad -\infty < y_1 < y_n < \infty$$

[*Suggestion* : Si $G$ est la fonction de répartition conjointe de $(Y_1\,;Y_n)$ et $F$ la fonction de répartition des $X_i$, montrer que $G(y_1\,;y_n) = \left[F(y_n)\right]^n - \left[F(y_n)-F(y_1)\right]^n$, puis déterminer les dérivées secondes $\frac{\partial^2}{\partial y_1\partial y_n}G(y_1;y_n)$ ].

$$\left[\begin{array}{l}\text{La densité conjointe de deux variables } (X\ ;Y) \text{ est définie par} \\ \lim\limits_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\frac{P(x<X<x+\Delta x, y<Y<y+\Delta y)}{\Delta x\Delta y}\end{array}\right]$$

$G(y_1;y_n) = \mathrm{P}(Y_1 \le y_1; Y_n \le y_n) = \mathrm{P}(Y_n \le y_n) - \mathrm{P}(Y_1 > y_1; Y_n \le y_n)$

$\mathrm{P}(Y_n \le y_n)$ = Probabilité que toutes les observations soient inférieures ou égales à $y_1$

$= [F(y_n)]^n$

$\mathrm{P}(Y_1 > y_1; Y_n \le y_n)$

= Probabilité que toutes les observations soient comprises entre $y_1$ et $y_n$

$= [F(y_n) - F(y_1)]^n$.

Donc $G(y_1;y_n) = [F(y_n)]^n - [F(y_n) - F(y_1)]^n$.

$$g(y_1;y_n) = \frac{\partial}{\partial y_n}\left[\frac{\partial}{\partial y_1} G(y_1;y_n)\right] = \frac{\partial}{\partial y_n}\left[-\frac{\partial}{\partial y_1}[F(y_n)-F(y_1)]^n\right] = \frac{\partial}{\partial y_n}\left[-n[F(y_n)-F(y_1)]^{n-1}(-f(y_1)\right]$$

$$= \frac{\partial}{\partial y_n}\left[nf(y_1)[F(y_n)-F(y_1)]^{n-1}\right] = nf(y_1)\frac{\partial}{\partial y_n}[F(y_n)-F(y_1)]^{n-1}$$

$$= nf(y_1)f(y_n)(n-1)[F(y_n)-F(y_1)]^{n-2}$$

b) Appliquer ce résultat à un échantillon d'une population exponentielle de paramètre β.

$$= \frac{n(n-1)}{\beta^2} e^{-(y_1+y_n)/\beta}\left[e^{-y_1/\beta} - e^{-y_2/\beta}\right]^{n-2}, 0 < y_1 \le y_n < \infty$$

c) Appliquer ce résultat à un échantillon d'une population uniforme sur (0 ; 1).

$$= n(n-1)(y_n - y_1)^{n-2}, 0 < y_1 \le y_n < 1$$

3.61 Vous et deux autres personnes attendez l'ouverture d'une banque. Quand la porte s'ouvre, les deux autres se précipitent vers les deux seules caissières disponibles. La durée de service d'un client est une variable aléatoire $X$ dont la fonction de densité est $f(x) = (½)e^{-(½)(x-4)}$, $x \ge 4$. Soit $Y$ votre temps d'attente.

a) Déterminer la fonction de densité de $Y$

$Y$ est la plus petite des deux durées de service, $X_1$ et $X_2$. La fonction de répartition de $X_i$ est $F(x) = 1 - e^{-(x-4)/2}$. La fonction de densité de $Y$ est $g(y) = 2f(y)[1-F(y)] = e^{4-y}$

b) Déterminer l'espérance de votre temps d'attente.

$$\int_4^\infty e^{4-y} dy = 5$$

3.62 Au numéro précédent, supposez que $X$ de densité $f(x) = e^{-(x-\theta)}$, $x > \theta$, et que $n$ personnes se font servir en même temps. Soit $Y$ votre temps d'attente.

a) Déterminer la fonction de densité de $Y$.

La fonction de répartition de $X$ et $F(x) = 1 - e^{-(x-\theta)}$ $x > \theta$. La fonction de densité de Y est $nf(y)[1-F(y)]^{n-1} = ne^{-n(y-\theta)}$.

b) Déterminer l'espérance de votre temps d'attente.

$Y$ est de loi exponentielle de paramètre $1/n$. Son espérance est donc $1/n$.

3.63 La durée d'une composante électronique suit une loi exponentielle de paramètre β = 2 (en milliers d'heures). Cinq de ces pièces sont installées dans un appareil.

a) Supposons que toutes les pièces sont nécessaires au fonctionnement de l'appareil. Soit $U_1$, la durée de l'appareil. Déterminer la fonction de répartition de $U_1$, sa fonction de densité, $\mathrm{E}(U_1)$ et $\mathrm{Var}(U_1)$, et calculer $\mathrm{P}(U_1 \le 3)$. (On suppose que les durées des pièces sont indépendantes et que le fonctionnement de l'appareil dépend uniquement du fonctionnement des 5 pièces).

Soit $F$ la fonction de répartition de la durée d'une pièce, et $p = F(3) = 1\text{-}e^{-3/2}$. Alors $P(U_1 \leq 3) = 1\text{-}(1\text{-}p)^5 =$ 0,9994469.

b) Soit $U_2$ la durée de l'appareil si on suppose, contrairement à ce qui a été supposé en a), qu'il suffit qu'une seule des 5 pièces fonctionne pour que l'appareil fonctionne. Déterminez la probabilité $P(U_2 \leq 3)$.

Soit $F$ la fonction de répartition de la durée d'une pièce, et $p = F(3) = 1\text{-}e^{-3/2} = 0{,}776687$.

Aors $P(U_2 \leq 3) = 1\text{-}p^5 = 0{,}7170294$.

c) Soit $U_3$ la durée de l'appareil lorsqu'on suppose que l'appareil fonctionne si et seulement si au moins deux des pièces fonctionnent. Déterminez la probabilité $P(U_3 \leq 3)$.

$P(U_3 \leq 3) = P(Y \geq 3 \mid Y \sim B(5\ ;\ p)) = 0{,}92277$

d) Soit $U_4$ la durée de l'appareil si les 5 pièces sont installées mais une seule est utilisée à la fois. Quand une pièce cesse de fonctionner, une autre prend la relève. Déterminez la probabilité $P(U_4 \leq 3)$

$U_4$ est une variable de loi gamma de paramètres $\alpha = 5$ et $\beta = 2$. $P(U_4 \leq 3) = 0{,}01858$.