

# **Chapitre 3**

## **Lois continues**

Une variable aléatoire continue  $X$  est une variable dont les valeurs sont un intervalle de nombres réels—le poids d'une pomme, la hauteur d'un arbre, la température d'une soupe. La distribution d'une variable continue ne peut pas s'exprimer par une fonction de probabilité  $p(x) = P(X = x)$  car  $p(x) = 0$  pour tout  $x$ . Elle s'exprime plutôt par une fonction de *densité* et une fonction de *répartition*.

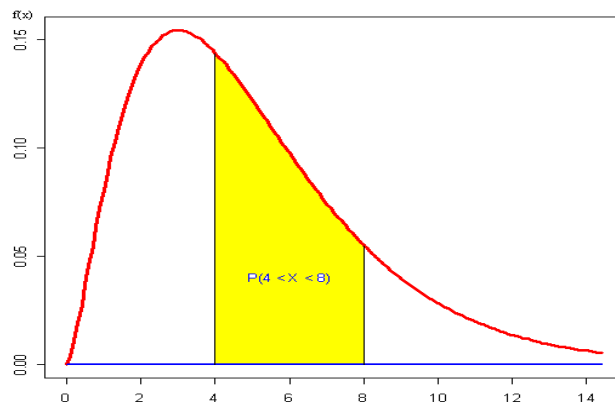
### 3.1 Fonctions densité et de répartition

#### Fonction de densité

La fonction de densité  $f$  d'une variable aléatoire continue  $X$  est une fonction de  $\mathbb{R} \rightarrow \mathbb{R}$  telle que, pour tout  $a$  et tout  $b$  tels que  $a < b$ ,

$$P(X \in (a ; b]) = P(a < X \leq b) = \int_a^b f(t)dt \quad (3.1.1)$$

**Figure 3.1.1**  
Fonction de densité



Une fonction de densité  $f$  doit satisfaire les conditions suivantes :

1.  $f(x) \geq 0, \forall x \in \mathbb{R}$  (sinon, l'intégrale risquerait d'être négative) ;
2.  $\int_{-\infty}^{\infty} f(x)dx = 1$  (la probabilité que  $X$  se trouve entre  $-\infty$  et  $\infty$  doit être égale à 1).

#### Fonction de répartition

La *fonction de répartition*  $F : \mathbb{R} \rightarrow [0 ; 1]$ , est définie, dans le cas continu (comme dans le cas discret), par

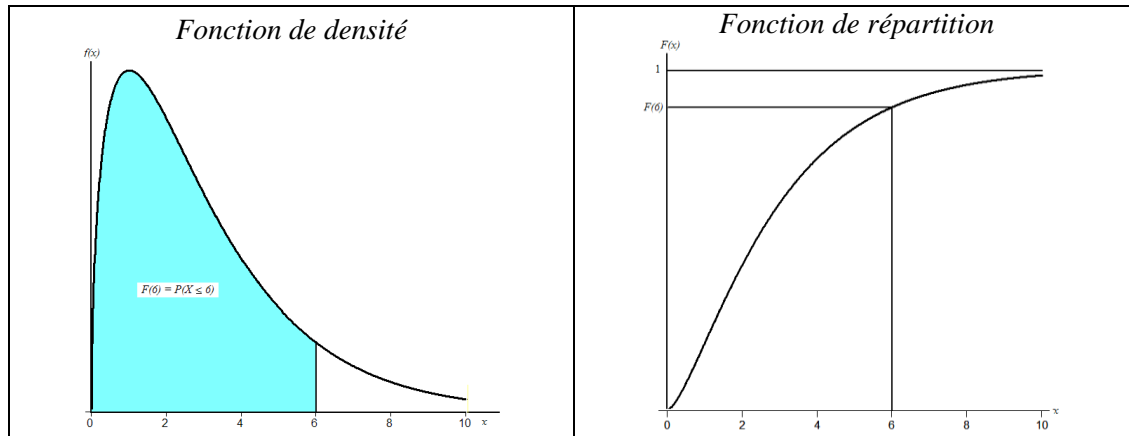
$$F(x) = P(X \leq x), \quad \text{pour tout } x \in \mathbb{R}$$

Par conséquent,

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

Dans le cas continu on obtient cette probabilité par l'intégration de la fonction de densité :

$$F(x) = \int_{-\infty}^x f(t)dt \quad (3.1.2)$$



Du fait que  $F(x) = \int_{-\infty}^x f(t)dt$ , on a que  $f$  est la dérivée de  $F$  :

$$f(x) = \frac{d}{dx} F(x) \quad (3.1.3)$$

Si  $X$  est votre temps d'arrêt au prochain feu rouge, comment trouver la fonction  $f$  qui vous permettra de calculer la probabilité d'attendre, disons, plus de 12 secondes ? Comment trouver la fonction de densité de  $Y$ , le poids de votre prochaine nièce à sa naissance ? Ou la fonction de densité des poids  $X_1, X_2, \dots, X_n$  des pommes cueillies dans un verger ?

La chose n'est pas toujours facile. Parfois, la loi découle de certaines hypothèses concernant le contexte. Dans la plupart des situations on recourra à des lois connues qui, tout comme dans le cas discret, dépendent d'un ou plusieurs paramètres. Certaines caractéristiques du contexte expérimental peuvent guider le choix du modèle. Nous présentons dans ce chapitre la *loi uniforme*, la *loi exponentielle*, la *loi normale*, la *loi Gamma* et la *loi khi-deux*.

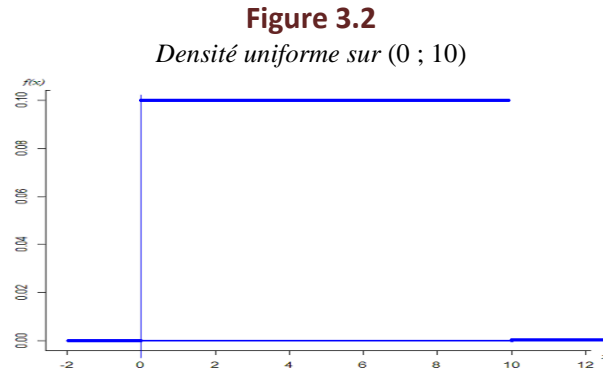
### 3.2 Loi uniforme

Soit  $X$  votre temps d'attente lorsque vous arrivez à une station de métro à une heure tout à fait aléatoire. Sachant qu'un métro passe régulièrement à chaque 10 minutes, quelle est la fonction de densité de  $X$  ? L'ensemble des valeurs possibles de  $X$  est l'intervalle  $(0 ; 10)$ . Aucun point dans cet intervalle n'est plus probable qu'un autre. On exprime ce fait par une fonction de densité  $f$  qui prend la même valeur partout sur  $(0 ; 10)$ . Par ailleurs, pour toute valeur  $x$  à l'extérieur de l'intervalle, on aura  $f(x) = 0$ . Quelle doit être la valeur de  $f$  dans  $(0 ; 10)$  ? Puisque la surface sous la courbe doit être égale à 1, il faudra que  $f(x) = 1/10$  sur  $(0 ; 10)$ . La figure 3.2 présente le graphique de la fonction de densité.

#### Définition Loi uniforme $\mathcal{U}(\alpha ; \beta)$

Une variable aléatoire  $X$  est de loi uniforme sur l'intervalle  $(\alpha ; \beta)$  [ $X \sim \mathcal{U}(\alpha ; \beta)$ ] si sa fonction de densité est donnée par

$$\begin{cases} \frac{1}{\beta - \alpha} & \text{si } \alpha < x \leq \beta \\ 0 & \text{sinon} \end{cases} \quad (3.2.1)$$



**Théorème 3.2.1** Fonction de répartition et propriétés d'une loi  $\mathcal{U}(\alpha ; \beta)$

Soit  $X$  une variable aléatoire de loi  $\mathcal{U}(\alpha ; \beta)$ . Alors

- a) La fonction de répartition de  $X$  est

$$F(x) = \begin{cases} 0 & \text{si } x < \alpha \\ \frac{x - \alpha}{\beta - \alpha} & \text{si } \alpha \leq x \leq \beta \\ 1 & \text{si } x > \beta \end{cases} \quad (3.2.2)$$

- b) L'espérance et la variance sont données par

$$E(X) = \frac{\alpha + \beta}{2} \quad \text{et} \quad \text{Var}(X) = \frac{(\beta - \alpha)^2}{12} \quad (3.2.3)$$

### 3.3 Loi exponentielle

La loi exponentielle sert à modéliser des temps d'attente ou des durées de vie : le temps d'attente entre deux arrivées à un comptoir de service, ou la durée de vie d'une pièce électronique.

**Définition** Loi exponentielle  $\mathcal{E}(\beta)$

Une variable aléatoire  $X$  est distribuée selon une loi exponentielle de moyenne  $\beta$ , notée  $\mathcal{E}(\beta)$ , si elle a pour densité la fonction

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases} \quad (3.3.1)$$

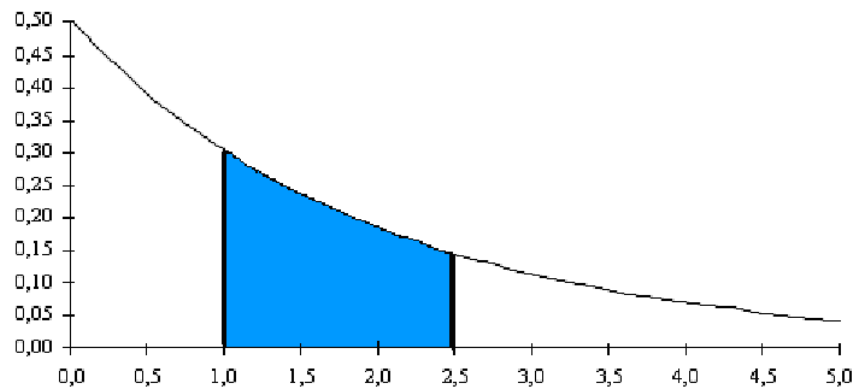
Les probabilités sont données par les surfaces sous la courbe. Considérons la probabilité

$$P(1,0 < X \leq 2,5)$$

lorsque  $X$  est de loi  $\mathcal{E}(2)$ . La surface à considérer est indiquée à la figure ci-dessous.

Figure 3.3.1

$P(1,0 < X \leq 2,5)$  pour une variable de loi  $\mathcal{E}(2)$



La surface entre 1,0 et 2,5 est donnée par  $F(2,5) - F(1,0)$ . Il se trouve que dans le cas d'une variable de loi  $\mathcal{E}(\beta)$ , il existe une expression de la fonction de répartition  $F(x)$  :

**Théorème 3.3.1** Soit  $X \sim \mathcal{E}(\beta)$ . Alors

a) La fonction de répartition de  $X$  est donnée par

$$F(x) = \begin{cases} 1 - e^{-x/\beta} & x > 0 \\ 0 & \text{sinon} \end{cases} \quad (3.3.2)$$

b) L'espérance et la variance de  $X$  sont données par

$$E(X) = \beta ; \text{Var}(X) = \beta^2. \quad (3.3.3)$$

c)  $P(a \leq X \leq b) = e^{-a/\beta} - e^{-b/\beta}$ . (3.3.4)

Le paramètre  $\beta$  est la moyenne de  $X$  et c'est aussi son écart-type.

**Exemple 3.3.1** Loi exponentielle

Soit  $X$  une variable aléatoire de loi  $\mathcal{E}(2)$ . Déterminer  $P(1,0 < X \leq 2,5)$ .

*Solution*  $P(1,0 < X \leq 2,5) = e^{-1,0/2} - e^{-2,5/2} = 0,6065 - 0,2865 = 0,3200$ .

**Exemple 3.3.2** Loi exponentielle

Supposons que la durée de vie d'une pièce électronique est une variable  $X$  de loi exponentielle. Si ces pièces ont une durée moyenne de 18 mois, quelle est la probabilité qu'une pièce donnée dure plus de 2 ans ?

*Solution* Si la durée  $X$  est exprimée en nombre d'années, on a  $\beta = 1,5$  années, puisque  $\beta = E(X) =$  la durée moyenne des pièces. On a alors

$$P(X > 2) = 1 - F(2) = 1 - (1 - e^{-2/\beta}) = e^{-2/1,5} = 0,2636.$$

**Exemple 3.3.3** Absence de mémoire

Dans le dernier exemple, on a montré que la probabilité qu'une pièce neuve dure plus de 2 ans est 0,2636. Nous montrons maintenant que cette probabilité est indépendante de l'âge de la pièce. Plus précisément, calculons la probabilité qu'une pièce produite il y a un an dure plus de 2 ans encore.

*Solution* Soit  $X$  la durée de la pièce. Nous devons calculer la probabilité conditionnelle  $P(B|A)$ , où  $B$  est l'événement « la pièce dure plus de 2 ans encore », soit  $\{X \geq 3\}$ ; et  $A$  est l'événement « la pièce a duré au moins un an », soit  $\{X \geq 1\}$ . Alors  $P(B|A) = P(A \cap B)/P(A) = P(X \geq 3)/P(X \geq 1) = e^{-3/1,5}/e^{-1/1,5} = e^{-2/1,5} = 0,2636$ .

Cette « absence de mémoire » est une des caractéristiques du modèle exponentiel.

### Lien avec la loi de Poisson

Là où il y a une loi de Poisson, il y a une loi exponentielle. Supposons qu'on observe les arrivées à un comptoir de service. On peut noter plusieurs choses, entre autres, le nombre  $Y$  d'arrivées dans un intervalle de temps.  $Y$  est de loi de Poisson. Mais on peut aussi noter le *temps d'attente*  $X$  avant la première arrivée, ou encore le temps d'attente entre deux arrivées successives.  $X$ , une variable continue, suit une loi exponentielle. Voici pourquoi.

Supposons donc que  $Y$ , le nombre d'arrivées dans un intervalle de temps  $(0 ; x)$ , est de loi de Poisson de paramètre  $\lambda = \gamma x$ . Alors

$$P(Y = y) = (\gamma x)^y e^{-\gamma x} / y!$$

Soit  $X$  est la durée de temps entre deux réalisations. La fonction de répartition de  $X$  est  $F_X(x) = P(X \leq x) = 1 - P(X > x)$ . Or  $P(X > x)$  est la probabilité qu'il n'y ait aucune arrivée dans l'intervalle  $(0 ; x)$ . Donc

$$P(X > x) = P[\text{aucune arrivée dans } (0, x)] = P(Y = 0) = (\gamma x)^0 e^{-\gamma x} / 0! = e^{-\gamma x}.$$

Ainsi,  $F_X(x) = 1 - P(X > x) = 1 - e^{-\gamma x}$ .

On reconnaît cette expression comme la fonction de répartition d'une variable de loi exponentielle de paramètre  $\beta = 1/\gamma$ .

### Exemple 3.3.4 Loi exponentielle issue d'une variable de loi de Poisson

Supposons qu'à une certaine intersection, il arrive en moyenne deux accidents en une semaine de 5 jours. En supposant que le nombre d'accidents dans un intervalle de temps suit une loi de Poisson, calculer la probabilité a) que le premier accident de la semaine n'ait pas lieu avant la fin du 3<sup>e</sup> jour ; b) que le premier accident ait lieu au courant de la 4<sup>e</sup> journée ?

*Solution* Le nombre moyen d'accidents en  $t = 5$  jours est  $\lambda = 2$  ( $\gamma$ , le nombre moyen d'accidents par jour, est égal à  $2/5 = 0,4$ ) ; donc le temps d'attente moyen entre deux accidents est  $\beta = t/\lambda = 5/2 = 2,5$  jours. Si  $X$  est la durée de temps qui s'écoule entre le début de la semaine et le premier accident, alors  $X \sim \mathcal{E}(2,5)$ .

a)  $P(X \geq 3) = e^{-3/2,5} = e^{-1,2} = 0,3012$ . On pourrait également résoudre ce problème en invoquant la loi de Poisson. Si  $Y$  est le nombre d'accidents en 3 jours, la probabilité demandée est  $P(Y = 0)$ . Or  $Y \sim \mathcal{P}(\gamma \times 3) = \mathcal{P}(1,2)$  et  $P(Y = 0) = e^{-1,2}$ .

b)  $P(3 < X < 4) = e^{-3/2,5} - e^{-4/2,5} = 0,3012 - 0,2019 = 0,0993$ .

### 3.4 Loi normale

La loi normale est l'une des lois les plus fondamentales de la statistique. Non seulement permet-elle de modéliser des variables observées couramment dans le quotidien et dans les sciences, telles la taille et parfois le poids d'individus, mais elle joue un rôle capital au niveau de l'inférence statistique. Une définition, en termes de la fonction de densité, suit

**Définition :** Loi normale  $N(\mu ; \sigma^2)$

Une variable aléatoire  $X$  est distribuée selon une loi normale de paramètres  $\mu$  et  $\sigma^2$ , désignée par  $X \sim \mathcal{N}(\mu ; \sigma^2)$ , si elle a pour densité la fonction

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)(x-\mu)^2/\sigma^2}, \quad -\infty < x < \infty \quad (3.4.1)$$

Les paramètres  $\mu$  et  $\sigma$  désignent l'espérance et l'écart-type, respectivement. La courbe est symétrique par rapport à sa moyenne  $\mu$ . Elle présente des points d'inflexion de part et d'autre de celle-ci, à une distance d'un écart-type.

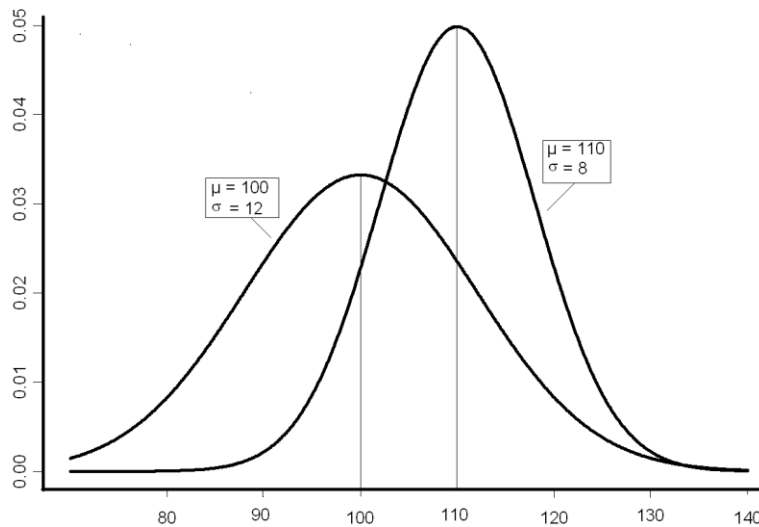
**Théorème 3.4.1** *Espérance et variance d'une  $\mathcal{N}(\mu ; \sigma^2)$*

Si  $X \sim \mathcal{N}(\mu ; \sigma^2)$ , alors

$$E(X) = \mu \text{ et } \text{Var}(X) = \sigma^2. \quad (3.4.2)$$

**Figure 3.4.1**

Deux lois normales :  $\mathcal{N}(100 ; 144)$  et  $\mathcal{N}(110 ; 64)$



### Calculs d'aires

Nous discuterons plus bas la question de savoir dans quelles conditions une variable suit une loi normale. Pour l'instant, considérons une variable  $X$  dont on sait qu'elle est de loi normale, que sa moyenne est  $\mu$  et sa variance  $\sigma^2$ , et supposons qu'on veuille calculer la probabilité  $P(a < X \leq b)$ , où  $a$  et  $b$  sont deux constantes,  $a < b$ . Si  $F$  est la fonction de répartition d'une variable de loi  $\mathcal{N}(\mu ; \sigma^2)$ ,  $P(a < X \leq b) = F(b) - F(a)$ . Or la primitive  $F(x)$  d'une variable de loi normale ne possède pas d'expression explicite. Aussi faut-il recourir à des algorithmes numériques complexes pour l'évaluer. Aujourd'hui, ces calculs ne posent pas de problème car plusieurs logiciels et certaines calculatrices peuvent les effectuer en une fraction de seconde. Malheureusement, ces procédures ne sont pas implantées dans toutes les calculatrices. En attendant qu'elles le soient (ou qu'un ordinateur portable soit à la portée de tous), nous devons nous référer à des tables qui publient les résultats de ces calculs. La plupart des tables publiées, dont une version est présentée en annexe, se concentrent sur une loi normale particulière, la courbe  $\mathcal{N}(0 ; 1)$ , la normale de moyenne 0 et de variance (ou écart-type) 1, appelée **normale centrée-réduite**. On désigne habituellement la fonction de répartition d'une variable  $Z$  normale centrée-réduite par  $\Phi$  :

$$\Phi(z) = P(Z \leq z) \text{ lorsque } Z \sim N(0; 1) \quad (3.4.3)$$

La table en annexe donne les valeurs de  $\Phi(z)$  pour diverses valeurs  $z$ .

**Exemple 3.4.1** Calcul des probabilités pour une normale centrée-réduite.

Soit  $Z \sim N(0; 1)$ . Calculer

- a)  $P(Z > 1,25)$ ;      b)  $P(Z \leq -1)$ ;      c)  $P(1,15 < Z \leq 2,11)$ ;  
 d)  $P(0 < Z \leq 1)$ ;      e)  $P(-2 < Z \leq 1)$ ,

*Solution*

- a)  $P(Z > 1,25) = 0,1056$  ;  
 b)  $P(Z \leq -1) = P(Z \geq 1) = 0,1587$  ;  
 c)  $P(1,15 < Z \leq 2,11) = P(Z > 1,15) - P(Z > 2,11) = 0,1251 - 0,0174 = 0,1077$  ;  
 d)  $P(0 < Z \leq 1) = P(Z > 0) - P(Z > 1) = 0,5 - 0,1587 = 0,3413$  ;  
 e)  $P(-2 < Z \leq 1) = P(Z > -2) - P(Z > 1) = (1 - 0,0228) - 0,1587 = 0,8185$ .

Pour des variables normales de moyenne et variance quelconques, il faut pouvoir « ramener » une normale arbitraire  $N(\mu; \sigma^2)$  à une  $N(0; 1)$ . Le théorème qui nous permet de le faire affirme que toute fonction affine d'une variable normale est normale.

**Théorème 3.4.2** Toute fonction affine d'une variable normale est normale.

Si  $X \sim N(\mu; \sigma^2)$  et  $Y = a + bX$  alors  $Y \sim N(a + b\mu; b^2\sigma^2)$

C'est le cas particulier suivant qui explique pourquoi il est possible de se limiter à une seule table.

**Corollaire** Si  $X : N(\mu; \sigma^2)$ , alors

$$Z = \frac{X - \mu}{\sigma} \sim N(0; 1)$$

C'est la possibilité de transformer une variable  $N(\mu; \sigma^2)$  en une variable centrée-réduite qui nous permet d'utiliser des tables pour calculer des surfaces sous une courbe normale quelconque, car

$$P[a < X \leq b] = P\left[\frac{a - \mu}{\sigma} < Z \leq \frac{b - \mu}{\sigma}\right], \text{ où } Z = \frac{X - \mu}{\sigma} \sim N(0; 1)$$

**Exemple 3.4.2** Calcul des probabilités pour une normale quelconque

Supposons que les montants correspondant à une population de factures sont de moyenne  $\mu = 200$  \$ et d'écart-type  $\sigma = 80$  \$. En supposant que les montants des factures sont de loi normale, déterminer la probabilité qu'une facture tirée au hasard corresponde à un montant compris entre 40 \$ et 280 \$.

*Solution* Si  $X$  est le montant de la facture, nous voulons calculer  $P(40 < X \leq 280)$ . Nous transformons

$$X \rightarrow Z = (X - 200)/80. \text{ Alors } P(40 < X \leq 280) = P\left(\frac{40 - 200}{80} \leq \frac{X - 200}{80} \leq \frac{280 - 200}{80}\right) =$$

$$P(-2 \leq Z \leq 1) = 0,8185.$$

### 3.5 Variables normales indépendantes

Une propriété importante de la loi normale est le fait qu'une somme ou une moyenne de variables aléatoires indépendantes, toutes de loi normale, est normale. Voici l'énoncé formel :



**Propriété Somme de variables normales indépendantes**

Si  $X_1, X_2, \dots, X_n$  sont des variables aléatoires *indépendantes* de loi normale, alors

leur somme  $Y = \sum_{i=1}^n X_i$  et leur moyenne  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$  sont de loi normale.

Nous savons, par ailleurs, que l'espérance de  $Y$  est égale à la somme des espérances et (étant donné l'indépendance des variables  $X_1, X_2, \dots, X_n$ ) la variance de  $Y$  est égale à la somme des variances :

$$E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) \quad \text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) \quad (3.4.4)$$

Par conséquent

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i); \quad \text{Var}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \quad (3.4.5)$$

Nous nous intéresserons particulièrement au cas où les moyennes sont toutes égales et les variances aussi. Nous avons alors le corollaire suivant :

**Propriété Somme de variables normales indépendantes de loi normale identique**

Soit  $X_1, X_2, \dots, X_n$  des variables aléatoires indépendantes de loi normale de moyenne  $\mu$  et de variance  $\sigma^2$ . Alors leur somme et leur moyenne sont de loi normale:

$$Y = \sum_{i=1}^n X_i \sim \mathcal{N}(n\mu; n\sigma^2) \quad \text{et} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

**Exemple 3.5.1** *Distribution de la moyenne d'un échantillon*

On prélève un échantillon de  $n = 15$  factures d'une très grande population de moyenne  $\mu = 300$  \$ et d'écart-type  $\sigma = 60$  \$. On suppose que la population est normale.

- Quelle est la probabilité que la valeur moyenne de l'échantillon soit comprise entre 290 \$ et 310 \$ (en d'autres termes, que la moyenne de l'échantillon se situe à 10 \$ ou moins de la moyenne de la population) ?
- Quelle est la probabilité que la valeur totale des 15 factures soit inférieure à 4549 \$ ?

*Solution*

- Supposer que la population est très grande et qu'elle est normale, c'est dire que les variables aléatoires  $X_1, \dots, X_{15}$  sont indépendantes et de loi normale. Leur moyenne est donc d'espérance  $\mu =$

300 et d'écart-type  $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{15}} = 15,49$ .

Alors  $P(290 \leq \bar{X} \leq 310) = P\left(\frac{290-300}{15,49} \leq Z \leq \frac{310-300}{15,49}\right) = P(-0,65 \leq Z \leq 0,65) = 0,4844$ .

- La valeur totale  $T$  des 15 factures est normale de moyenne  $n\mu = 4500$  et de variance  $n\sigma^2 = 54\,000$ . Alors  $P(T \leq 4549) = P\left(Z \leq \frac{4549-4500}{\sqrt{54000}}\right) = P(Z \leq 0,21) = 0,5835$ .

Les propriétés énoncées ci-dessus (normalité d'une somme et d'une moyenne) découlent d'un théorème plus général que voici :

**Théorème 3.5.1** *Combinaison linéaire de variables normales indépendantes*

Soit  $X_1, \dots, X_n$   $n$  variables aléatoires indépendantes, où  $X_i \sim N(\mu_i; \sigma_i^2)$ ,  $i = 1, \dots, n$ , et  $a_1, \dots, a_n$  sont des constantes. Alors

$$X = \sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i; \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

**Exemple 3.5.2** *Distribution de moyennes, de sommes et de différences*

Considérons deux très grandes populations de mangues de deux variétés différentes. Supposons que leur poids est de loi normale dont les paramètres sont les suivants :

Variété	Moyenne	Écart-type
Julie	$\mu_1 = 352$ g	$\sigma_1 = 40$ g
Amélie	$\mu_2 = 370$ g	$\sigma_2 = 35$ g

Dans les solutions qui suivent, nous supposons que les tirages de mangues dans les populations sont indépendants, ce qui est justifié lorsqu'il s'agit de grandes populations. On désignera par  $Z$  toute variable normale centrée-réduite.

- Quelle est la probabilité que le poids total d'une douzaine de mangues « Julie » soit inférieure à 4357 g ?
- Quelle est la probabilité que le poids *moyen* des mangues dans un sac de 16 mangues « Julie » soit inférieur à 364,5 g ?
- Si on tire deux « Julie » au hasard, quelle est la probabilité que la première pèse plus de 100 g de plus que la deuxième ?
- Quelle est la probabilité que la différence de poids entre deux mangues « Julie » soit supérieure à 120g ?
- Quelle est la probabilité que le poids total de 5 « Julie » soit supérieur au poids total de 4 « Amélie » ?

*Réponses*

- Le poids total de 12 mangues est une somme  $Y = X_1 + \dots + X_{12}$  de 12 variables aléatoires de loi  $N(352; 40^2)$ . Donc  $Y \sim N[12(352); 12(40^2)]$ .  $P(Y < 4357) = P(Z < 0,96) = 0,8315$ .
- Le poids moyen est  $\bar{X} = (X_1 + \dots + X_{16})/16$ .  $\bar{X} \sim N[352; 40^2/16]$  et  $P(\bar{X} \leq 364,5) = P(Z < 1,25) = 0,8944$ .
- Soit  $X_1$  le poids de la première mangue et  $X_2$  le poids de la deuxième. Nous cherchons  $P(Y > 100)$  où  $Y = X_1 - X_2 \sim N[0; 40^2 + 40^2]$ .  $P(Y > 100) = P(Z > 1,77) = 0,0384$ .
- Soit  $X_1$  le poids de la première mangue et  $X_2$  le poids de la deuxième. Nous cherchons  $P(|Y| > 120)$  où  $Y = X_1 - X_2 \sim N[0; 40^2 + 40^2]$ .  $P(|Y| > 120) = 2 P(Y > 120) = 2P(Z > 2,12) = 0,0338$ .
- Soit  $X_1 + \dots + X_5$  le poids total des 5 « Julie »,  $Y_1 + \dots + Y_4$  celui des 4 « Amélie », et  $W = (X_1 + \dots + X_5) - (Y_1 + \dots + Y_4)$ .  $W \sim N[5(352) - 4(370); 5(40^2) + 4(35^2)]$ . La probabilité demandée est  $P(W > 0) = P(Z > -3,26) = 0,9994$ .

**3.6 Théorème limite central**

Le théorème invoqué à l'exemple 3.5.1, selon lequel la moyenne  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  d'un échantillon est de loi normale, a une portée limitée du fait qu'il suppose la normalité des variables  $X_1, X_2, \dots, X_n$ . Or en pratique, cette hypothèse est rarement vérifiée. Ce qu'elle signifie, dans l'exemple 3.5.1, où l'on parle des montants d'une population de factures, c'est que l'histogramme des montants des factures de la population entière a la forme en cloche d'une loi normale. C'est

rarement le cas pour des données financières, dont les distributions sont généralement fortement asymétriques : des petits montants très nombreux, et un nombre décroissant de montants de plus en plus élevés.

Il se trouve, heureusement, que la solution proposée à l'exemple 3.5.1 est néanmoins adéquate, dans le sens que la réponse obtenue ne s'éloigne probablement pas trop de la réponse exacte. Ce qui justifie cette conclusion, c'est un remarquable théorème appelé *théorème limite central*, qui dit à peu près ceci : la variable  $\bar{X}$  est à peu près de loi normale, même si la population n'est pas normale. La seule condition exigée est que  $n$  soit assez grand. Voici l'énoncé :

**Théorème 3.6.1** *Théorème limite central*

Soit  $X_1, X_2, \dots$ , une suite de variables aléatoires indépendantes de même loi, de moyenne  $\mu$  et de variance  $\sigma^2$ . Alors la variable  $Z_{\bar{X}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$  tend en loi vers une loi  $N(0 ; 1)$ .

En pratique, cela signifie que si  $n$  est assez grand, on peut supposer que

$$\bar{X} \sim N(\mu ; \sigma_{\bar{X}}^2) \text{ où } \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad (3.6.1)$$

(ou que  $T = \sum_{i=1}^n X_i \sim N(n\mu ; \sigma_T^2)$ ,  $\sigma_T^2 = n\sigma^2$ ).

On ne peut pas être définitif ni très précis quant à ce qu'on entend par «  $n$  grand ». En pratique on se donne comme limite le nombre 30 : si  $n \geq 30$ , on dit que  $n$  est « grand », ce qui permet de supposer que  $\bar{X}$  est approximativement de loi  $N(\mu ; \sigma^2/n)$  (et donc que  $Z_{\bar{X}} \sim N(0 ; 1)$ ). Mais il est évident que l'approximation ne sera pas toujours bonne pour  $n \geq 30$ , ni nécessairement mauvaise pour  $n < 30$ . D'abord, nous n'avons pas de critère de « qualité » pour une approximation. Est-elle bonne à  $n = 30$  ? Avant de répondre, il faut préciser ce qu'on entend par « bonne ». Cette notion est forcément relative : l'approximation est meilleure avec  $n = 300$  qu'avec  $n = 30$ . Pour une taille donnée, disons  $n = 30$ , la qualité de l'approximation dépend de plusieurs facteurs, le principal étant la normalité ou non-normalité des variables  $X_i$  elles-mêmes. Si les  $X_i$  sont déjà « presque » normales, l'approximation de la distribution de  $\bar{X}$  par une  $N(\mu ; \sigma^2/n)$  est adéquate même lorsque  $n$  est inférieur à 30. Inversement, si les  $X_i$  sont très loin d'être normales (par exemple, exponentielles ou autrement asymétriques), alors même avec  $n$  bien supérieur à 30 la distribution de  $\bar{X}$  n'approchera pas une  $N(\mu ; \sigma^2/n)$ .

**Remarque** *Le théorème limite central est l'une des raisons pour lesquelles la loi normale est utilisée si fréquemment comme modèle : on a souvent affaire à des sommes ou des moyennes. Mais il arrive aussi de supposer une loi normale pour des observations faites directement dans la nature, comme, par exemple, la taille  $X$  d'une personne : Qu'est-ce qui nous fait dire que  $X$  est normale ? Une certaine supposition — dans la mesure où elle semble crédible — permet d'espérer qu'elle le soit, du moins à peu près. C'est la supposition que  $X$  est en fait une somme de variables aléatoires, même si nous ne les avons pas nous-mêmes observées et sommées. Concrètement, c'est dire que la taille  $X$  est l'effet d'une multitude de facteurs, génétiques et environnementaux, des facteurs additifs dont aucun ne prédomine (ceci découle d'une interprétation qu'on peut donner à un certain théorème central limite, plus général que celui que nous avons énoncé ici). C'est le cas des tailles dans une population relativement homogène, comme on l'a souvent observé dans des études empiriques. Une population hétérogène — par exemple, les tailles des membres d'une population composée, en parts égales d'adultes et d'enfants —, ne présenterait pas une loi normale. Dans une moindre mesure, la normalité est altérée si la population est constituée d'hommes et de femmes. Contrairement à la distribution des tailles d'un groupe de femmes (figure 3.6.1-a)), celle d'un groupe mixte est légèrement bimodale (figure 3.6.1-b)).*

En général, la décision d'adopter tel modèle plutôt que tel autre est basée sur un mélange de considérations théoriques et d'observations. Mais le test ultime, c'est l'observation. On compare — graphiquement d'abord, au moyen d'un histogramme — l'échantillon à la fonction de densité du modèle. Par exemple, si la loi considérée comme modèle est normale, on comparera les données de l'échantillon à une loi normale de même moyenne et même variance que l'échantillon. Parfois cette approche graphique suffit pour rejeter d'emblée un modèle. Autrement, on procède à une analyse plus rigoureuse. On n'a jamais vraiment la preuve qu'un modèle est bon, mais tant qu'il n'est pas contredit par l'observation et tant qu'il n'y en a pas de meilleur on le retient et on l'utilise. La figure 3.6.2 illustre l'approche graphique. Il est vrai qu'on a de bonnes raisons de croire a priori que le poids d'un nouveau-né est une variable normale. À l'œil, la distribution observée des poids d'un grand groupe de nouveau-nés (figure 3.6.2-a) semble le confirmer. On la compare à la distribution normale de même moyenne et même variance (figure 3.6.2-b). En tout cas, il est évident que la loi exponentielle est mal ajustée aux données (figure 3.6.2-c). Mais la loi normale est-elle la seule qui s'ajuste bien ? Visiblement non, puisque la loi gamma (figure 3.6.2-d) est elle aussi bien ajustée. Le choix n'est pas évident dans le graphique. Une analyse plus rigoureuse aboutirait à un choix — mais quel que soit le choix proposé par les techniques formelles, il n'est pas exclu qu'on décide quand même d'utiliser la loi normale, d'abord parce qu'elle est plus commode, ensuite parce qu'elle est plus crédible a priori.

### 3.7 Approximation normale de la loi binomiale

Une variable de loi binomiale peut être approchée par une variable de loi normale lorsque  $n$  est grand. Ceci est une application immédiate du théorème limite central, puisque une variable binomiale peut être considérée comme une somme  $X = \sum_i X_i$ , où  $X_i$  est le nombre de succès au  $i^{\text{e}}$  essai. Les variables  $X_i$  prennent donc la valeur 1 ou 0 avec probabilité  $p$  et  $q = 1-p$ , respectivement. Elles sont indépendantes puisque les épreuves sont indépendantes, leur moyenne  $\mu$  et variance  $\sigma^2$  sont données par  $\mu = E(X_i) = p$ ,  $\sigma^2 = \text{Var}(X_i) = pq$ . La variable  $X$  peut donc être approchée par une  $N(np ; npq)$ . En plus d'exiger que  $n$  soit grand il faut que  $p$  ne soit ni trop grand ni trop petit car dans ces cas la loi binomiale est fortement asymétrique. En général, plus  $p$  est proche des extrémités 0 ou de 1, plus il faut que  $n$  soit grand. Une façon de combiner ces conditions est d'exiger que  $np$  et  $nq$  ne soient pas trop petits. La formulation suivante résume ces conditions :

#### **Théorème 3.7.1** Approximation normale de la loi binomiale

Soit  $X \sim \mathcal{B}(n ; p)$  et supposons que  $n \geq 30$ ,  $np > 5$ , et  $nq > 5$ .

Alors, la variable centrée-réduite  $Z = (X - np) / \sqrt{npq}$  suit à peu près une loi  $N(0 ; 1)$  :

$$Z = \frac{X - np}{\sqrt{npq}} \sim N(0 ; 1) \quad (3.7.1)$$

ou encore, que

$$X \sim N(np ; npq). \quad (3.7.2)$$

Figure 3.6.1

Distribution des tailles – échantillon homogène et échantillon hétérogène

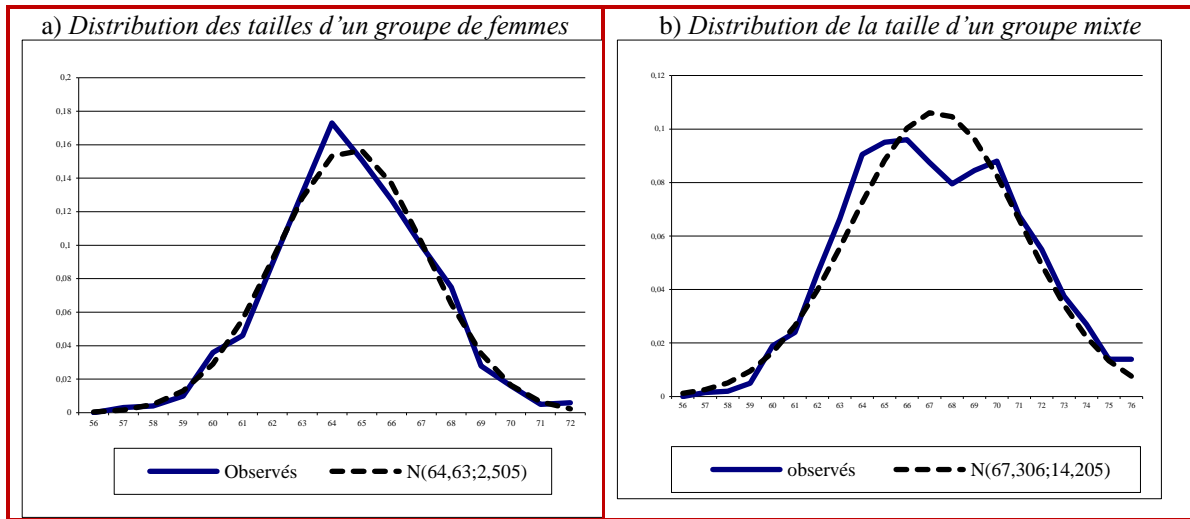
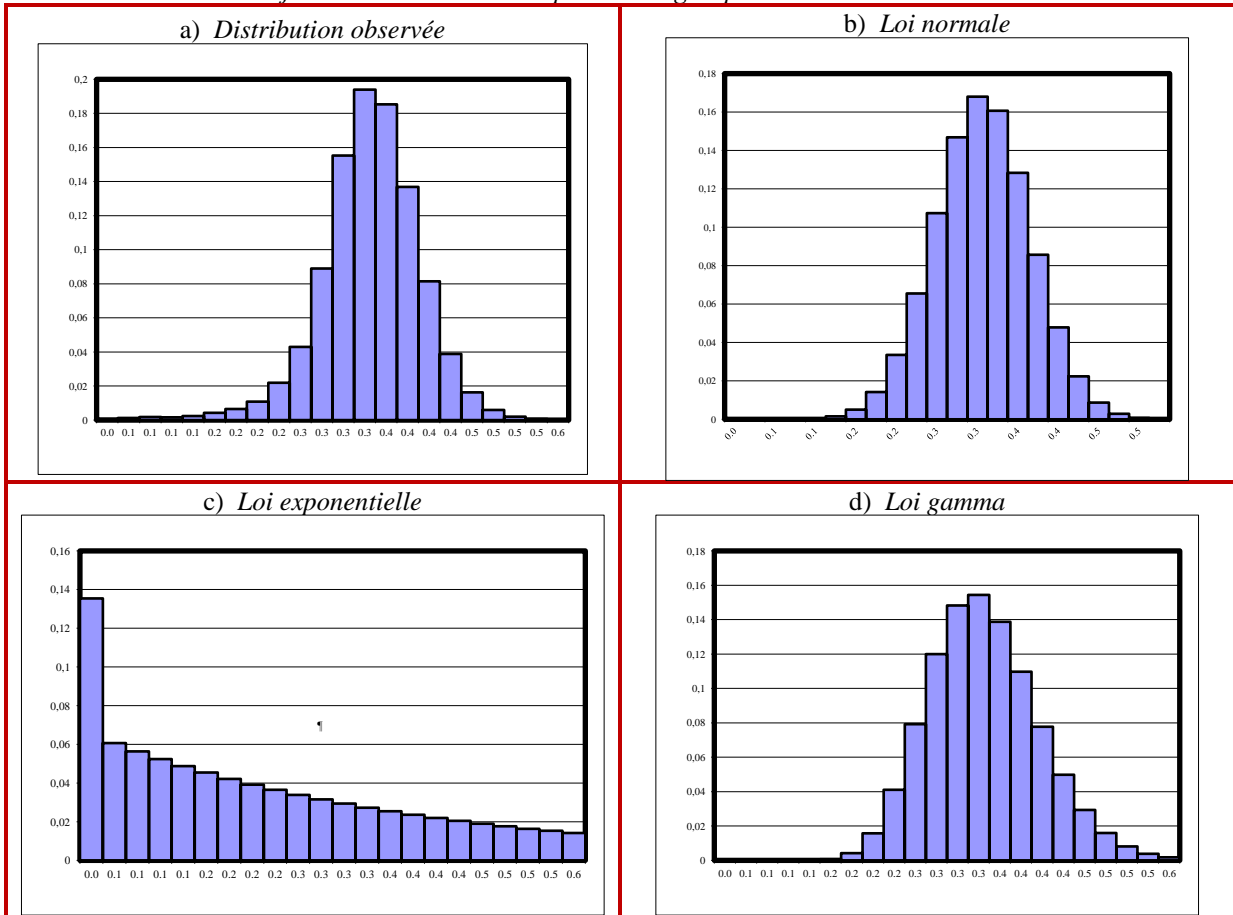


Figure 3.6.2

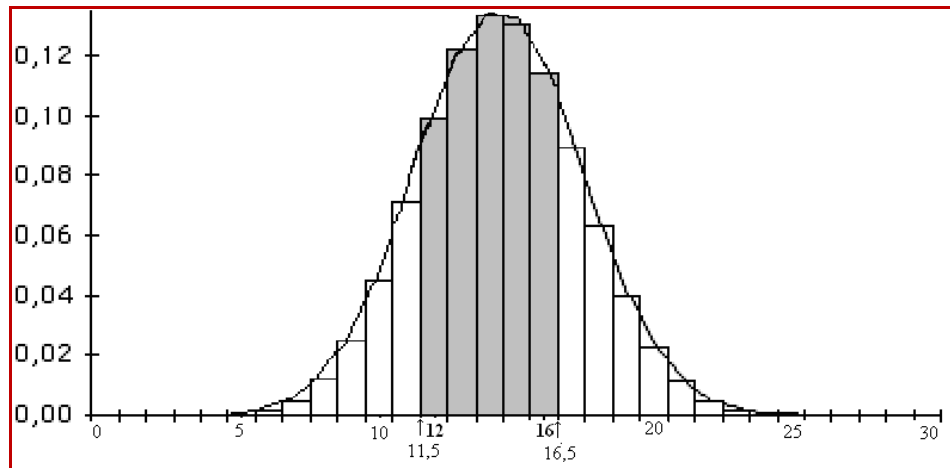
Ajustement à trois lois des poids d'un groupe de nouveau-nés



La probabilité d'une valeur donnée  $x$  est proportionnelle à la *surface* du rectangle de l'histogramme posé au-dessus de  $x$  puisque la base des rectangles est toujours de largeur 1. Cette probabilité peut être approchée par la surface correspondante sous la courbe  $N(np ; npq)$ . Considérons, par exemple, la probabilité  $P(a \leq X \leq b)$  où  $a$  et  $b$  sont des entiers. Elle est représentée par la surface combinée des rectangles dont les bases sont centrées aux points  $a, a+1, \dots, b$ .

**Figure 3.7.1**

Distribution d'une  $N(14,4 ; 8,64)$  et d'une  $\mathcal{B}(36 ; 0,4)$   
 $P(12 \leq X \leq 16)$



Ces rectangles reposent sur l'intervalle  $[a - 1/2 ; b + 1/2]$  et c'est sur cet intervalle que nous devons évaluer l'aire sous la courbe  $N(np ; npq)$  et non sur l'intervalle  $[a ; b]$  ; ce dernier intervalle n'englobe qu'une moitié du rectangle posé sur  $a$  et une moitié du rectangle posé sur  $b$ . L'ajustement aux limites  $a$  et  $b$ , appelé *correction pour la continuité*, est utile lorsque  $n$  est plutôt modeste ; il a peu d'effet lorsque  $n$  est grand.

**Exemple 3.7.1** Approximation normale de la loi binomiale

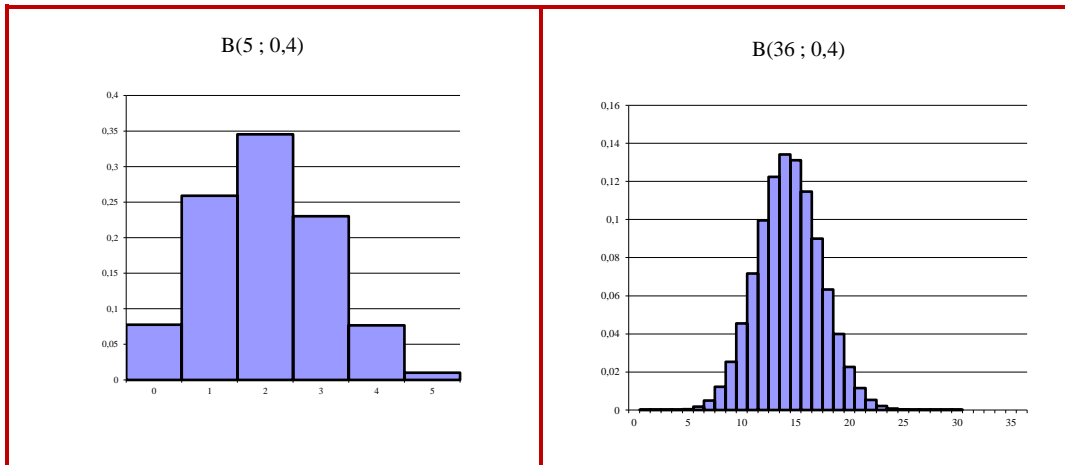
Soit  $X \sim \mathcal{B}(36 ; 0,4)$ . Calculer par la loi normale a)  $P(12 \leq X \leq 16)$  ; b)  $P(X < 13)$  ; c)  $P(X = 10)$ .

*Solution* L'espérance de  $X$  est  $E(X) = np = (36)(0,4) = 14,4$  et sa variance est  $\sigma^2 = npq = 36(0,4)(0,6) = 8,64$ . Nous supposons donc que  $X \sim N(14,4 ; 8,64)$ .

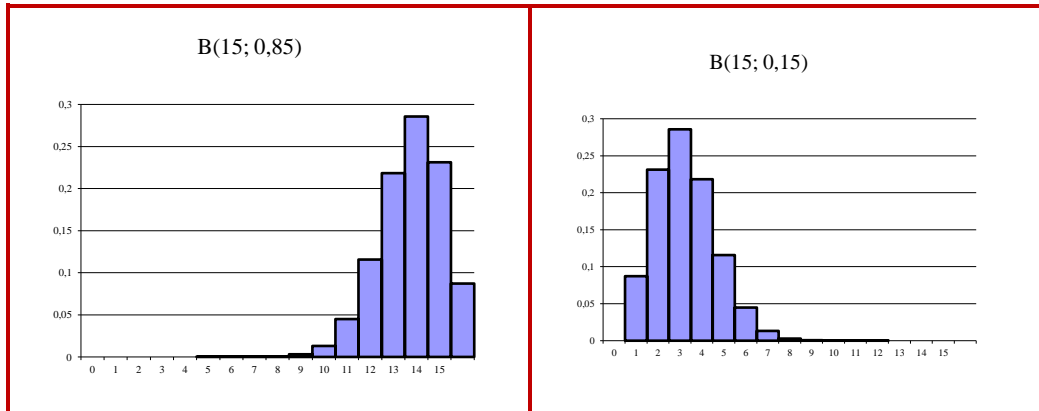
- $P(11,5 \leq X \leq 16,5) = P[(11,5-14,4)/\sqrt{8,64} \leq Z \leq (16,5-14,4)/\sqrt{8,64}] = P(-0,99 \leq Z \leq 0,71) = 0,6000$  ; ce calcul est illustré à la figure 3.6.1.
- $P(X < 12,5) = P(Z < (12,5-14,4)/\sqrt{8,64}) = P(Z < -0,65) = 0,2578$  ;
- $P(9,5 < X < 10,5) = P(-1,67 < Z < -1,33) = 0,0443$ .

Par la loi binomiale, ces probabilités sont, respectivement, 0,602 ; 0,2615 et 0,04550.

**Remarque** Il n'est évidemment pas possible de formuler une consigne catégorique —comme « utiliser l'approximation si  $n \geq 30$  »— concernant l'utilisation de l'approximation normale. Plusieurs facteurs entrent en ligne de compte, dont l'un certainement est ce qu'on entend par « bonne ». Ce qu'on sait, c'est que l'approximation est d'autant meilleure que  $n$  est grand, et d'autant meilleure que  $p$  est proche de  $1/2$ . Ces conditions s'expliquent intuitivement. Lorsque  $n$  est grand, les contours de l'histogramme peuvent mieux être approchés par une courbe, comme le montrent les graphiques suivants, qui comparent  $n = 5$  avec  $n = 36$  (et  $p = 0,4$ ).



La valeur de  $p$  détermine le degré d'asymétrie. Lorsque  $p = 1/2$ , l'histogramme est symétrique et donc facilement approché par une courbe normale. Les graphiques suivants montrent ce qui se passe lorsque  $p$  est très grand ou très petit.



Quelques comparaisons numériques peuvent donner une idée des précisions atteignables. Ces calculs se font aisément avec un logiciel statistique ou avec un tableur comme Excel. Voici, par exemple, une comparaison des calculs exacts et approximatifs de la fonction de répartition  $F(x)$  d'une variable de loi  $\mathcal{B}(20 ; 0,3) : F(x) = P(X \leq x)$ .

$x$	$F(x) = P(X \leq x)$		
	Probabilité exacte	Probabilité approximative	Pourcentage d'erreur
0	0,0008	0,0036	356,23
1	0,0076	0,0141	84,02
2	0,0355	0,0438	23,54
3	0,1071	0,1113	3,89
8	0,8867	0,8887	0,23
9	0,9520	0,9562	0,43
10	0,9829	0,9859	0,31
11	0,9949	0,9964	0,15
12	0,9987	0,9992	0,05
13	0,9997	0,9999	0,01

Ce qu'on voit clairement, c'est que pour les petites valeurs  $x$  les approximations ne sont pas très bonnes, mais qu'elles s'améliorent nettement lorsque  $x$  croît.

### 3.8 Loi Gamma et loi khi-deux

La loi khi-deux, qui joue un rôle important en statistique, est un cas particulier d'une loi plus générale appelée *loi Gamma*, qui elle aussi d'importantes applications. Nous commençons donc par définir la loi Gamma. C'est une loi à deux paramètres positifs,  $\alpha$  et  $\beta$ , dont la fonction de densité est

**Définition** *Loi Gamma*  $\mathcal{G}\mathcal{M}(\alpha ; \beta)$

Une variable  $X$  est dite de loi Gamma de paramètres  $\alpha$  et  $\beta$  si sa fonction de densité est donnée par

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x > 0 \quad (3.8.1)$$

où  $\Gamma(\alpha)$  est la fonction gamma, définie pour  $\alpha > 0$ , par

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy.$$

On désignera la loi gamma de paramètres  $\alpha$  et  $\beta$  par  $\mathcal{G}\mathcal{M}(\alpha ; \beta)$

**Théorème 3.8.1** *Espérance et variance de la loi Gamma*

$$E(X) = \alpha\beta \text{ et } \text{Var}(X) = \alpha\beta^2 \quad (3.8.2)$$

Le prochain théorème ne sera pas démontré.

**Théorème 3.8.2** *Additivité de la loi Gamma*

Soit  $X_1, \dots, X_n$   $n$  variables aléatoires indépendantes,  $X_i \sim \mathcal{G}\mathcal{M}(\alpha_i ; \beta)$ . Alors la somme

$$X = X_1 + \dots + X_n \sim \mathcal{G}\mathcal{M}(\sum \alpha_i ; \beta)$$

où  $v = \sum_{i=1}^n v_i$ , et

$$E(X) = \alpha\beta, \quad \text{Var}(X) = \alpha\beta^2 \quad (3.8.3)$$

**Remarque** La loi exponentielle est une loi  $\mathcal{G}\mathcal{M}(1 ; \beta)$

*La loi  $\chi^2$*

La loi khi-deux est une loi Gamma avec  $\beta = 2$  et  $\alpha = v/2$ , où  $v$  est un entier positif.  $\beta$  étant fixe, le seul paramètre est  $v$ , appelé *nombre de degrés de liberté*.

On écrit  $X \sim \chi_v^2$  pour signifier que  $X$  est de la loi khi-deux à  $v$  degrés de liberté

Sa fonction de densité est

$$f(x) = \frac{1}{\Gamma(v/2)2^{v/2}} x^{v/2-1} e^{-x/2}, \quad x > 0 \quad (3.8.4)$$

L'espérance et la variance de  $Y$  sont :  $E(Y) = v$  et  $\text{Var}(Y) = 2v$ .

*Le rôle de la loi  $\chi^2$*

Nous avons montré qu'une fonction linéaire d'une variable normale est normale. Pour des fonctions non linéaires, il n'existe pas de théorème général; leur loi doit être déterminée au cas par cas. Il y a une fonction non linéaire, cependant, qui, on le verra plus tard, a une importance



particulière : le carré d'une variable de loi normale centrée réduite est de loi khi-deux à 1 degré de liberté.

**Théorème 3.8.3** Le carré d'une  $N(0 ; \sigma^2)$  est une  $\chi_1^2$

$$\text{Si } Z \sim N(0 ; 1), \text{ alors } Z^2 \sim \chi_1^2 .$$

En combinant les théorèmes 3.8.2 et 3.8.3, on a ceci :

**Théorème 3.8.4** Somme de carrés de normales centrées-réduites

Soit  $Z_1, \dots, Z_n$ ,  $n$  variables aléatoires indépendantes, chacune de loi  $N(0 ; 1)$  . Alors

$$X = Z_1^2 + Z_2^2 + \dots + Z_n^2 \sim \chi_n^2 .$$

**Corollaire 1** Soit  $X_1, \dots, X_n$ ,  $n$  variables aléatoires indépendantes,  $X_i \sim N(\mu_i ; \sigma_i^2)$ . Alors

$$Q = \left( \frac{X_1 - \mu_1}{\sigma_1} \right)^2 + \left( \frac{X_2 - \mu_2}{\sigma_2} \right)^2 + \dots + \left( \frac{X_n - \mu_n}{\sigma_n} \right)^2 \sim \chi_n^2$$

**Corollaire 2** Soit  $X_1, \dots, X_n$ ,  $n$  variables aléatoires indépendantes,  $X_i \sim N(\mu ; \sigma^2)$ . Alors

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

### 3.9 Statistiques d'ordre

Soit  $X_1, X_2, \dots, X_n$  un échantillon aléatoire simple d'une population de fonction de densité  $f$  et de fonction de répartition  $F$ . Considérons les variables  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  définies par

- $X_{(1)}$  = la plus petite des  $n$  valeurs  $X_1, X_2, \dots, X_n$  ;
- $X_{(2)}$  = la 2<sup>e</sup> plus petite des  $n$  valeurs  $X_1, X_2, \dots, X_n$  ;
- $\vdots$
- $X_{(n)}$  = La plus grande des  $n$  valeurs  $X_1, X_2, \dots, X_n$

En général.

$$X_{(r)} = \text{La } r^{\text{e}} \text{ plus petite des } n \text{ valeurs } X_1, X_2, \dots, X_n, 1 \leq r \leq n.$$

**Exemple 3.9.1** . Soit  $X_1, X_2, X_3$  et  $X_4$  les tailles de 4 personnes tirées successivement d'une grande population ( $X_1$  est la taille de la première personne tiré,  $X_2$  celle de la deuxième, etc.). Supposons que  $X_1 = 168, X_2 = 179, X_3 = 166$  et  $X_4 = 173$ . Alors  $X_{(1)} = 166 ; X_{(2)} = 168 ; X_{(3)} = 173 ; X_{(4)} = 179$ .

Nous allons déterminer la fonction de densité de  $X_{(r)}$ . Avant d'aborder le cas général, nous traitons les cas particuliers—relativement simples— $X_{(1)}$  et  $X_{(n)}$  : le minimum et le maximum des  $n$  valeurs.

**Exemple 3.9.2** . Soit  $X_1, X_2, \dots, X_n$  un échantillon aléatoire simple d'une population de fonction de densité  $f$  et de fonction de répartition  $F$ . Quelle est la fonction de densité de  $X_{(n)}$  ?

*Solution* Désignons  $X_{(n)}$  par  $Y$ . La fonction de répartition de  $Y$  est  $G(y) = P(Y \leq y)$ .

Nous montrons que  $G(y) = [F(y)]^n$  et donc que la fonction de densité de  $Y$  est  $g(y) = nf(y)[F(y)]^{n-1}$ . L'événement  $\{Y \leq y\}$  est équivalent à l'événement  $\{X_1 \leq y \text{ et } X_2 \leq y, \dots, \text{ et } X_n \leq y\}$ . Donc  $G(y) = P(X_1 \leq y \text{ et } X_2 \leq y, \dots, \text{ et } X_n \leq y) = P(X_1 \leq y) \times P(X_2 \leq y) \times \dots \times P(X_n \leq y) = [F(y)]^n$ . Donc la fonction de densité de  $Y$  est

$$g(y) = \frac{d}{dy} G(y) = \frac{d}{dy} [F(y)]^n = n[F(y)]^{n-1} F'(y) = nf(y)[F(y)]^{n-1}.$$

**Exemple 3.9.3** . Soit  $X_1, X_2, \dots, X_n$  un échantillon aléatoire simple d'une population de fonction de densité  $f$  et de fonction de répartition  $F$ . Quelle est la fonction de densité de  $X_{(1)}$  ?

*Solution* Désignons  $X_{(1)}$  par  $Z$ . La fonction de répartition de  $Z$  est

$$H(z) = P(Z \leq z) = 1 - P(Z > z) = 1 - P(X_1 > z \text{ et } X_2 > z, \dots, \text{ et } X_n > z) \\ = 1 - P(X_1 > z) \times P(X_2 > z) \times \dots \times P(X_n > z) = 1 - [1 - F(z)]^n.$$

Donc la fonction de densité de  $Z$  est  $\frac{d}{dz} H(z) = \frac{d}{dy} (1 - [1 - F(y)]^n) = nf(z)[1 - F(z)]^{n-1}$ .

Ces deux exemples sont des cas particuliers du théorème suivant :

**Théorème 3.9.1** Soit  $X_1, X_2, \dots, X_n$  un échantillon aléatoire simple d'une population de fonction de densité  $f$  et de fonction de répartition  $F$ .

La fonction de densité de  $Y = X_{(r)}$ , pour  $1 \leq r \leq n$  est donnée par

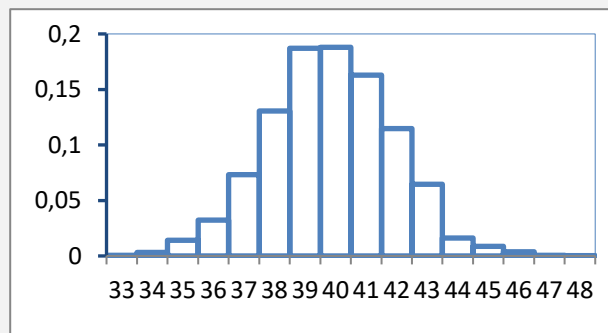
$$g(y) = \frac{n!}{(r-1)!(n-r)!} f(y)[F(y)]^{r-1}[1 - F(y)]^{n-r} \tag{3.9.1}$$

### 3.10 Quelques distributions naturelles

Nous avons justifié (par le théorème limite central) l'emploi de la loi normale pour des moyennes de variables indépendantes— des moyennes qu'on calcule régulièrement dans le cadre d'un sondage. Il suffit donc que nos échantillons ne soient pas trop petits pour n'avoir pas trop d'alarme concernant la normalité ou non de la population. Il reste que la question de savoir quelles variables sont naturellement normales est théoriquement intéressante. Quelles sont les variables, observées dans la nature, dont on peut prévoir qu'elles sont de loi normale ? Nous verrons quelques exemples desquels on pourra tirer quelque leçon. Essentiellement, on peut prévoir qu'une variable est de loi normale si elle est l'effet d'une multitude de facteurs indépendants, et si aucun de ces facteurs ne prédomine. Ainsi donc, on s'attend à ce que les tailles des personnes d'une population homogène soient distribuées normalement parce que les tailles sont déterminées par un grand nombre de facteurs environnementaux et génétiques. Si la population n'est pas homogène — si elle comprend des femmes et des hommes, des adultes et des enfants, des personnes de races différentes — alors on ne peut pas compter sur une distribution normale. Nous verrons, et c'est une conséquence de ces considérations, que les variables observées directement dans la nature — physique ou biologique— ont de fortes chances d'être normales alors que celles issues d'un contexte social ne le sont presque jamais. Voici donc un certain nombre d'exemples.

**Exemple 3.10.1** *Tour de poitrine de 5738 soldats écossais du XIX<sup>e</sup> siècle.*

On le voit ici comme dans beaucoup de mesures anthropométriques, la loi normale semble bien s'ajuster aux données.

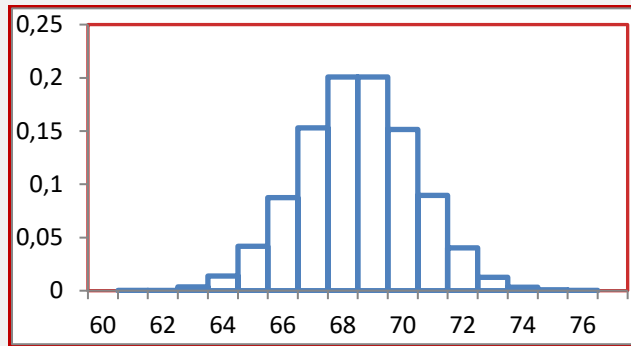


Données de A. Quetelet, dans « Lettres S.A.R. le Duc Régnant de Saxe-Cobourg et Gotha, sur la Théorie des Probabilités, Appliquée aux Sciences Morales et Politiques. » (Bruxelles : M. Hayes, 1846) p. 400. Source : Velleman, P. F. and Hoaglin, D. C. (1981). *Applications, Basics, and Computing of Exploratory Data Analysis*. Belmont. CA : Wadsworth, Inc.

Données disponible dans <http://lib.stat.cmu.edu/DASL/Stories/ChestsizesofMilitiamen.html>

**Exemple 3.10.2** Tailles (en pouces) de 25 000 enfants de Hong Kong.

Bien que, généralement, la distribution des tailles des personnes soit à peu près normale, on ne devrait pas s’y attendre ici, car la population est plutôt hétérogène : elle comprend des enfants des deux sexes et d’âges différents. Malgré cela, la distribution empirique ne s’éloigne pas trop d’une distribution normale.

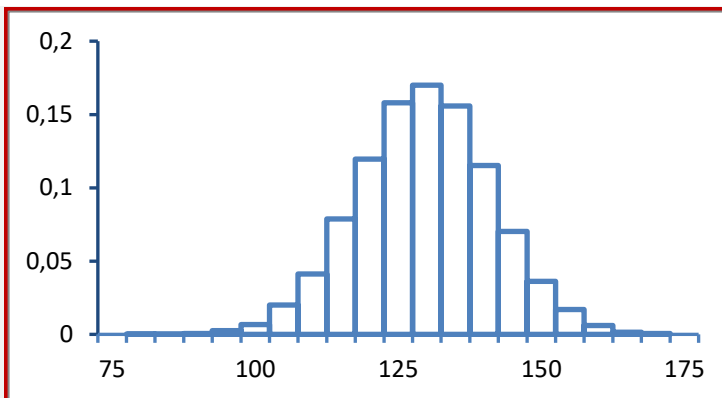


Obtenues lors d’une étude menée en 1993 auprès des maternelles, écoles et centres de santé. Données disponibles dans

[http://wiki.stat.ucla.edu/socr/index.php/SOCR\\_Data\\_Dinov\\_020108\\_HeightsWeights](http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_Dinov_020108_HeightsWeights).

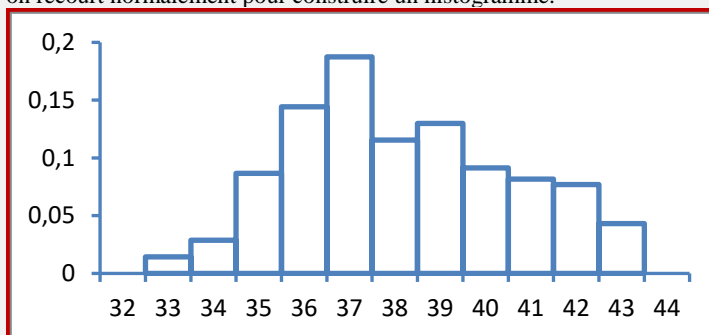
**Exemple 3.10.3** Poids (en livres) des enfants de la population présentée au dernier exemple.

Ici aussi l’hypothèse de normalité est plausible.

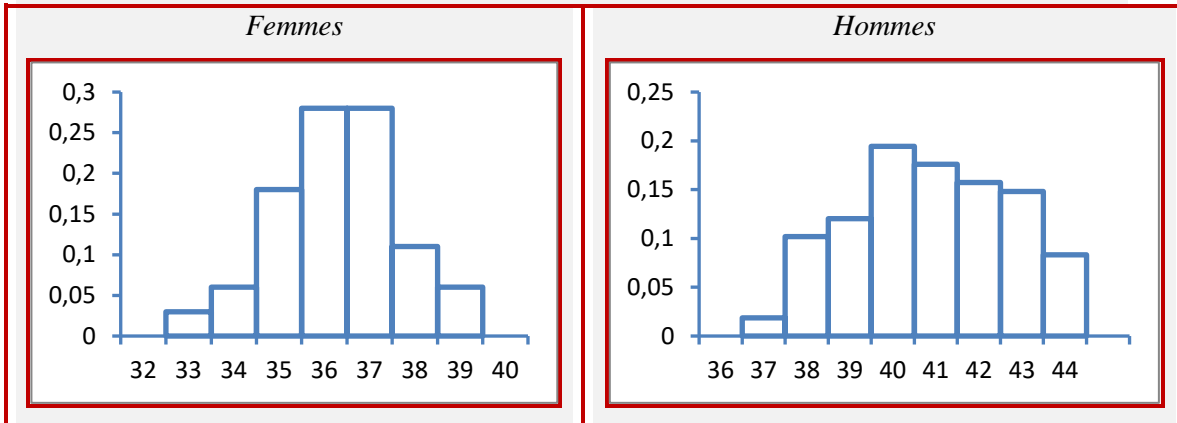


**Exemple 3.10.4** Distribution de la pointure des chaussures au Brésil

La pointure des chaussures est une variable carrément discrète, mais chacune de ses valeurs correspond à un intervalle de longueurs de pied. La longueur du pied est une variable continue, dite « latente », qui, elle, pourrait bien être normale. Dans ce cas, la distribution ci-dessous serait équivalente au regroupement de valeurs auquel on recourt normalement pour construire un histogramme.



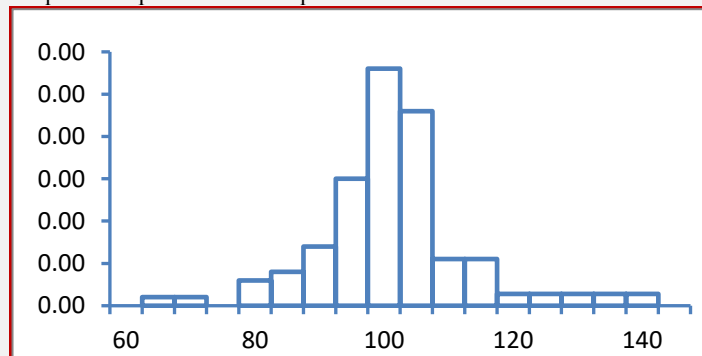
Or la normalité n'est pas évidente ici. Il semblerait même que la distribution soit bimodale (deux sommets). Est-ce dû au fait que l'échantillon comprend des hommes et des femmes ? C'est possible ; donc on considère séparément la distribution pour les femmes et pour les hommes. On constate en effet que les deux distributions sont unimodales, mais elles sont plutôt asymétriques. La distribution pour les hommes, en particulier, surprend un peu par la grande fréquence des hautes pointures.



Étude effectuée auprès de 5312 Brésiliens en 2004. <http://www.zonalatina.com/Zldata377.htm>

**Exemple 3.10.5** Force de rupture d'un échantillon de fils.

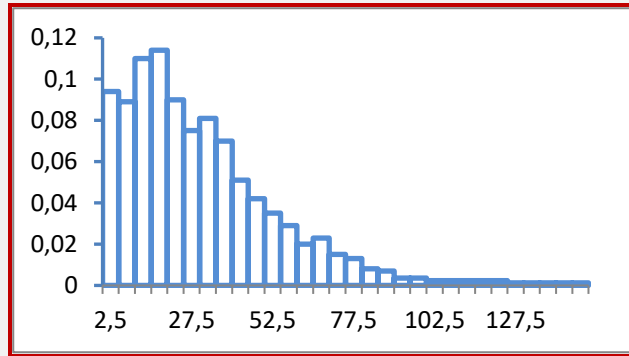
On s'attendrait ici à une distribution normale, car les mesures issues d'un processus de fabrication le sont souvent. Ce qu'on observe plutôt, c'est une distribution relativement asymétrique, due à quelques données extrêmes qui pourraient être le fait d'un événement épisodique. La présomption de normalité est assez ferme pour faire soupçonner que ces cas extrêmes sont symptomatiques d'un problème technique.



[DASL : <http://lib.stat.cmu.edu/DASL/Stories/distribution.html> ; Source : A.J. Duncan, Quality Control and Industrial Statistics, 4th Ed., Richard D. Irwin, 1974, p. 67]

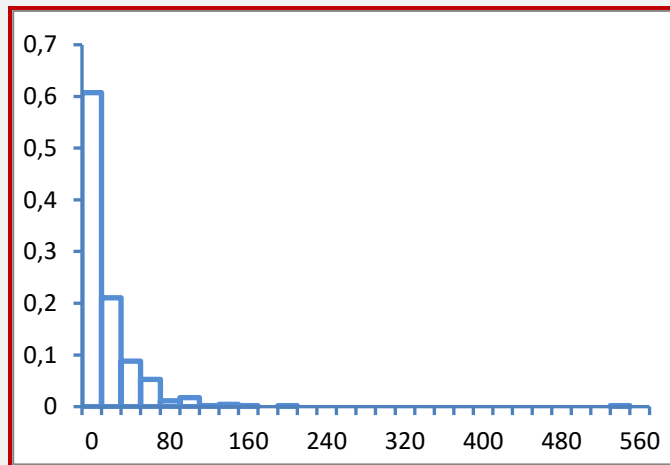
**Exemple 3.10.6** Revenus des Québécois en 2007

C'était prévisible : les choses se gâtent dès qu'on sort du domaine physique ou biologique. La distribution des revenus ci-dessous est typique : forte concentration à gauche, décroissance de la fréquence sur un long intervalle à droite. C'est caractéristique des distributions de revenus.



**Exemple 3.10.7** Distribution des revenus des PME du Québec (en milliers de dollars)

On pouvait prévoir l'asymétrie de cette distribution, bien que le grand nombre de PME avec des revenus très faibles surprend un peu. En particulier quand on voit (dans l'exemple 3.8.7) que la plupart d'entre elles ont 40 employés ou plus.



<http://www.lesaffaires.com/classements/500-quebec-2010>

**3.11 Quelques démonstrations**

*La fonction de densité normale*

On montre que  $f(x) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)(x-\mu)^2/\sigma^2}$  est bien une fonction de densité. Il est évident que  $f(x) > 0 \forall x$ . Il

reste à démontrer que  $\int_{-\infty}^{\infty} f(x)dx = 1$ . Nous commençons par une variable normale centrée-réduite  $Z$  dont

la fonction de densité est  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ . Nous montrons que  $\int_{-\infty}^{\infty} \varphi(x)dx = 1 \Leftrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz$

$= 1 \Leftrightarrow \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-z^2/2} dz = 1 \Leftrightarrow \int_0^{\infty} e^{-z^2/2} dz = \sqrt{\frac{\pi}{2}} \Leftrightarrow \left( \int_0^{\infty} e^{-z^2/2} dz \right)^2 = \frac{\pi}{2}$ . Nous allons donc montrer

que  $\left( \int_0^{\infty} e^{-z^2/2} dz \right)^2 = \frac{\pi}{2}$ .

Or l'intégrale  $L^2 = \left( \int_0^\infty e^{-z^2/2} dz \right)^2$  peut s'exprimer comme une intégrale double :

$$L^2 = \left( \int_0^\infty e^{-z^2/2} dz \right)^2 = \int_0^\infty e^{-u^2/2} du \int_0^\infty e^{-v^2/2} dv = \int_0^\infty \int_0^\infty e^{-(u^2+v^2)/2} dudv$$

En faisant la transformation  $u = r \cos \theta$  et  $v = r \sin \theta$ , nous obtenons

$$L^2 = \int_0^{\pi/2} \int_0^\infty e^{-r^2/2} r dr d\theta$$

$r$  étant le Jacobien de la transformation. Lorsqu'on intègre par rapport à  $\theta$ , on obtient

$$L^2 = \frac{\pi}{2} \int_0^\infty e^{-r^2/2} r dr$$

En faisant la transformation  $s = r^2/2$ , nous obtenons  $L^2 = \frac{\pi}{2} \int_0^\infty e^{-s} ds$ .

On a donc  $L^2 = \frac{\pi}{2}$  car  $\int_0^\infty e^{-s} ds = 1$ .

Dans le cas général, on évalue l'intégrale de la fonction de densité,  $\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^\infty e^{-(x-\mu)^2/2\sigma^2} dx$ , à l'aide de la transformation  $z = \frac{x-\mu}{\sigma}$ . On a alors  $\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^\infty e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-z^2/2} dz = \int_{-\infty}^\infty \varphi(x) dx = 1$ . C'est ce qu'il fallait démontrer.

**Théorème 3.4.2**

À montrer que si  $X \sim N(\mu ; \sigma^2)$  et  $Y = a + bX$ , alors  $Y \sim N(a + b\mu ; b^2\sigma^2)$

Supposons que  $b > 0$ . La fonction de répartition de  $Y$  est  $G(y) = P(Y \leq y) = P(X \leq (y-a)/b) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{(y-a)/b} e^{-(x-\mu)^2/(2\sigma^2)} dx$ . Effectuant la transformation  $u = a + bx$ . On a alors

$$G(y) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{(y-a)/b} e^{-(x-\mu)^2/(2\sigma^2)} dx = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^y e^{-\frac{1}{2\sigma^2}((u-a)/b-\mu)^2} \frac{1}{b} du = \frac{1}{\sqrt{2\pi b\sigma}} \int_{-\infty}^y e^{-\frac{1}{2(b\sigma)^2}[u-(a+b\mu)]^2} du$$

On reconnaît cette expression comme la fonction de répartition d'une variable de loi  $N(a+b\mu ; b^2\sigma^2)$ , ce qui démontre le théorème.

La démonstration du théorème 3.5.1 exige une connaissance des fonctions génératrices des moments et ne sera pas traitée ici.

**La loi khi-deux**

Soit  $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$ ,  $x > 0$ , la fonction de densité d'une variable de loi Gamma de paramètres  $\alpha$

et  $\beta$ .

Pour montrer que

$$\int_0^\infty f(x) dx = 1 \Leftrightarrow \int_0^\infty \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx = \Gamma(\alpha),$$

il suffit d'effectuer la transformation  $x = \beta y$  :

$$\int_0^\infty \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx = \int_0^\infty \frac{1}{\beta^\alpha} (\beta y)^{\alpha-1} e^{-y} \beta dy = \int_0^\infty y^{\alpha-1} e^{-y} dy = \Gamma(\alpha)$$

On écrit  $X \sim \mathcal{G}m(\alpha ; \beta)$  pour signifier que  $X$  suit une loi gamma de paramètre  $\alpha$  et  $\beta$ .

Nous utiliserons les propriétés suivantes de la fonction gamma :

- 1) Lorsque  $\alpha = 1$ ,  $\Gamma(\alpha) = \Gamma(1) = 1$ .
- 2) Si  $\alpha > 1$ , on obtient en intégrant par parties la formule 
$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

Lorsque  $\alpha = n$ , un entier positif, on obtient

3) On peut aussi vérifier que  $\Gamma(n) = (n - 1)!$   
 $\Gamma(1/2) = \sqrt{\pi}$

**Théorème 3.8.2**

Si  $X \sim \mathcal{GM}(\alpha ; \beta)$ , alors  $E(X) = \alpha\beta$  et  $\text{Var}(X) = \alpha\beta^2$ .

Le dernier résultat montre que  $\int_0^\infty x^{\alpha-1}e^{-x/\beta}dx = \beta^\alpha\Gamma(\alpha)$ .

$$\text{Donc } E(X) = \int_0^\infty x \frac{1}{\beta^\alpha\Gamma(\alpha)} x^{\alpha-1}e^{-x/\beta}dx = \frac{1}{\beta^\alpha\Gamma(\alpha)} \int_0^\infty x^{(\alpha+1)-1}e^{-x/\beta}dx = \frac{1}{\beta^\alpha\Gamma(\alpha)} \beta^{(\alpha+1)}\Gamma(\alpha+1) = \alpha\beta.$$

$$E(X^2) = \int_0^\infty x^2 \frac{1}{\beta^\alpha\Gamma(\alpha)} x^{\alpha-1}e^{-x/\beta}dx = \frac{1}{\beta^\alpha\Gamma(\alpha)} \int_0^\infty x^{(\alpha+2)-1}e^{-x/\beta}dx = \frac{1}{\beta^\alpha\Gamma(\alpha)} \beta^{(\alpha+2)}\Gamma(\alpha+2) = \alpha(\alpha+1)\beta^2.$$

De là il vient que  $\text{Var}(X) = E(X^2) - [E(X)]^2 = \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$ .

**Loi khi-deux**

La loi  $\chi_k^2$  est un cas particulier de la loi gamma, avec  $\beta = 2$  et  $\alpha = k/2$ , où  $k$ , le nombre de degrés de liberté, est un entier positif.

**Théorème 3.8.2**

À montrer : si  $Z \sim \mathcal{N}(0 ; 1)$ , alors  $Y = X^2 \sim \chi_1^2$ .

Soit  $Z \sim \mathcal{N}(0 ; 1)$  et  $Y = Z^2$ . La fonction de répartition de  $Y$  est, pour  $y > 0$ ,

$$F_Y(y) = P(Y \leq y) = P(Z^2 \leq y) = P(-\sqrt{y} \leq Z \leq \sqrt{y}) = P(Z \leq \sqrt{y}) - P(Z \leq -\sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}),$$

où  $\Phi$  est la fonction de répartition d'une  $\mathcal{N}(0 ; 1)$ .

La fonction de densité est donc

$$f_Y(y) = \Phi(\sqrt{y}) \frac{d}{dy} \sqrt{y} - \Phi(-\sqrt{y}) \frac{d}{dy} (-\sqrt{y})$$

où

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

$$\text{Donc } f_Y(y) = \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-y/2} y^{-1/2} - \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-y/2} (-y^{-1/2}) = \frac{1}{\sqrt{2\pi}} e^{-y/2} y^{-1/2}.$$

Puisque  $\Gamma(1/2) = \sqrt{\pi}$ , la fonction de densité peut s'écrire comme  $f_Y(y) = \frac{1}{\Gamma(1/2)\sqrt{2}} y^{-1/2} e^{-y/2}$ .

Cette expression est celle de la fonction de densité d'une variable de loi  $\chi_1^2$ .

Plus généralement, une variable  $\chi_n^2$  est une somme de carrés de  $n$  variables aléatoires  $\mathcal{N}(0 ; 1)$  indépendantes. Ceci découle du théorème suivant, corollaire des théorèmes 3.8.1 et 3.8.2,

**Théorème**

Soit  $X_1, \dots, X_n$ ,  $n$  variables aléatoires indépendantes,  $X_i \sim \chi_{\nu_i}^2$ . Alors la somme

$$X = X_1 + \dots + X_n \sim \chi_\nu^2$$

où  $\nu = \sum_{i=1}^n \nu_i$ .

L'énoncé suivant est un corollaire :

**Corollaire**

Soit  $Z_1, \dots, Z_n$ ,  $n$  variables aléatoires indépendantes, chacune de loi  $\mathcal{N}(0; 1)$  et soit  $X = \sum_{i=1}^n Z_i^2$ . Alors  $X \sim \chi_n^2$ .

### Théorème 3.9.1

$g(y) = \lim_{\delta \rightarrow 0} \frac{P(y \leq X_{(r)} \leq y + \delta)}{\delta}$ . Or l'événement  $\{y \leq X_{(r)} \leq y + \delta\}$  signifie qu'une des observations prend une valeur  $t$  dans l'intervalle  $[y; y + \delta]$ ,  $r-1$  autres se situent à gauche de  $t$  et  $n-r$  autres enfin se situent à droite de  $t$ . Le nombre de façons de choisir la variable qui prend la valeur  $t$ , les  $r-1$  variables qui prennent une valeur inférieure à  $t$  et les  $n-r$  variables qui prennent une valeur supérieure à  $t$  est  $\frac{n!}{(r-1)!(n-r)!}$ . La

probabilité  $P(y \leq X_{(r)} \leq y + \delta)$  ne dépend pas du choix. Donc, prenant un choix particulier, nous avons

$$P(y \leq X_{(r)} \leq y + \delta) = \frac{n!}{(r-1)!(n-r)!} P(X_1 \leq X_{(r)}, \dots, X_{r-1} \leq X_{(r)}, y \leq X_r \leq y + \delta, X_{r+1} \geq X_{(r)}, \dots, X_n \geq X_{(r)})$$

$$g(y) = \frac{n!}{(r-1)!(n-r)!} \lim_{\delta \rightarrow 0} \frac{P(X_1 \leq X_{(r)}, \dots, X_{r-1} \leq X_{(r)}, y \leq X_r \leq y + \delta, X_{r+1} \geq X_{(r)}, \dots, X_n \geq X_{(r)})}{\delta}$$

Soit  $B(\delta) = P(X_1 \leq X_{(r)}, \dots, X_{r-1} \leq X_{(r)}, y \leq X_r \leq y + \delta, X_{r+1} \geq X_{(r)}, \dots, X_n \geq X_{(r)})$ .

$$P(X_1 \leq y, \dots, X_{(r-1)} \leq y, y \leq X_r \leq y + \delta, X_{(r+1)} \geq y + \delta, \dots, X_{(n)} \geq y + \delta)$$

$$\leq B(\delta) \leq P(X_1 \leq y + \delta, \dots, X_{(r-1)} \leq y + \delta, y \leq X_r \leq y + \delta, X_{(r+1)} \geq y, \dots, X_{(n)} \geq y)$$

$$[F(y)]^{r-1} P(y \leq X_r \leq y + \delta) [1 - F(y + \delta)]^{n-r} \leq B(\delta) \leq [F(y + \delta)]^{r-1} P(y \leq X_r \leq y + \delta) [1 - F(y)]^{n-r}$$

Donc

$$\frac{n!}{(r-1)!(n-r)!} \lim_{\delta \rightarrow 0} \frac{[F(y)]^{r-1} P(y \leq X_r \leq y + \delta) [1 - F(y + \delta)]^{n-r}}{\delta}$$

$$\leq g(y) \leq \frac{n!}{(r-1)!(n-r)!} \lim_{\delta \rightarrow 0} \frac{[F(y + \delta)]^{r-1} P(y \leq X_r \leq y + \delta) [1 - F(y)]^{n-r}}{\delta}$$

$$\frac{n!}{(r-1)!(n-r)!} [F(y)]^{r-1} \lim_{\delta \rightarrow 0} [1 - F(y + \delta)]^{n-r} \lim_{\delta \rightarrow 0} \frac{P(y \leq X_r \leq y + \delta)}{\delta}$$

$$\leq g(y) \leq \frac{n!}{(r-1)!(n-r)!} [1 - F(y)]^{n-r} \lim_{\delta \rightarrow 0} [F(y + \delta)]^{r-1} \lim_{\delta \rightarrow 0} \frac{P(y \leq X_r \leq y + \delta)}{\delta}$$

$$\frac{n!}{(r-1)!(n-r)!} [F(y)]^{r-1} [1 - F(y)]^{n-r} f(y) \leq g(y) \leq \frac{n!}{(r-1)!(n-r)!} [1 - F(y)]^{n-r} [F(y)]^{r-1} f(y)$$

$$\text{Donc } g(y) = \frac{n!}{(r-1)!(n-r)!} f(y) [F(y)]^{r-1} [1 - F(y)]^{n-r}.$$



**RÉSUMÉ**

1 La fonction de densité d'une variable aléatoire continue  $X$  est une fonction  $f$  telle que pour  $x \in \mathbb{R}$   $F(x) = \int_{-\infty}^x f(t)dt$ , où  $F$  est la fonction de répartition de  $X$ . Donc  $f(x) = F'(x)$  et  $P(a < X \leq b) = \int_a^b f(x)dx$ .

2 Lois continues

Distribution	Modalités de $X$	Fonctions de densité et de répartition	$E(X)$	$Var(X)$	$M_X(t)$
Uniforme $\mathcal{U}(\alpha; \beta)$	$x \in [\alpha; \beta]$	$f(x) = \frac{1}{\beta - \alpha}$ $F(x) = \frac{x - \alpha}{\beta - \alpha}$	$(\alpha + \beta)/2$	$(\beta - \alpha)^2/12$	$\frac{e^{\beta t} - e^{\alpha t}}{t(\beta - \alpha)}$
Loi exponentielle $\mathcal{E}(\beta)$	$x \in [0; \infty)$	$f(x) = \frac{1}{\beta} e^{-x/\beta}$ $F(x) = 1 - e^{-x/\beta}$	$\beta$	$\beta^2$	$(1 - \beta t)^{-1}$
Loi gamma $\mathcal{Gm}(\alpha; \beta)$	$x \in [0; \infty)$	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$	$\alpha\beta$	$\alpha\beta^2$	$(1 - \beta t)^{-\alpha}$
Loi khi-Deux $\chi_v^2$	$x \in [0; \infty)$	$\frac{1}{\Gamma(v/2)2^{v/2}} x^{v/2-1} e^{-x/2}$	$v$	$2v^2$	$(1 - 2t)^{-v/2}$
Normale $\mathcal{N}(\mu; \sigma^2)$	$x \in (-\infty; \infty)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\mu$	$\sigma^2$	$e^{\mu t + \sigma^2 t^2/2}$

4 Une fonction linéaire d'une normale est normale  
Si  $X \sim \mathcal{N}(\mu; \sigma^2)$  et  $Y = a + bX$  alors  $Y \sim \mathcal{N}(a + b\mu; b^2\sigma^2)$ .

5 La cote  $Z$  d'une normale est une normale centrée-réduite :  $Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0; 1)$ .

6 Toute fonction linéaire de normales indépendantes est normale  
Si les variables  $X_i \sim \mathcal{N}(\mu_i; \sigma_i^2)$ ,  $i = 1, \dots, n$ , sont indépendantes et  $a_1, \dots, a_n$  des constantes, alors  $X = \sum_{i=1}^n a_i X_i \sim \mathcal{N}(\sum_{i=1}^n a_i \mu_i; \sum_{i=1}^n a_i^2 \sigma_i^2)$ .

7 Somme et moyenne de normales identiquement distribuées.  
Si  $X_1, X_2, \dots, X_n$  sont indépendantes, toutes de même loi  $\mathcal{N}(\mu; \sigma^2)$ , alors  $T = \sum_{i=1}^n X_i \sim \mathcal{N}(n\mu; n\sigma^2)$  et  $\bar{X} \sim \mathcal{N}(\mu; \sigma^2/n)$ .

8 Théorème limite central.  
Soit  $X_1, X_2, \dots$ , une suite de variables aléatoires indépendantes de même loi, de moyenne  $\mu$  et de variance  $\sigma^2$ . Alors la variable  $Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$  tend en loi vers une normale centrée-réduite lorsque  $n$  tend vers l'infini.

- 9 *Approximation de la loi binomiale par la loi normale*  
Soit  $X \sim \mathcal{N}(n; p)$  et supposons que  $n \geq 30$ ,  $np > 5$ , et  $nq > 5$ . Alors, il est approximativement vrai que  $\frac{X - np}{\sqrt{npq}} \sim \mathcal{N}(0; 1)$ , ou encore, que  $X \sim \mathcal{N}(np; npq)$ .
- 10 *Loi Gamma*, désignée par  $\mathcal{Gm}(\alpha; \beta)$  est une loi à deux paramètres,  $\alpha$  et  $\beta$ . L'espérance d'une variable  $X$  de loi  $\mathcal{Gm}(\alpha; \beta)$  est  $E(X) = \alpha\beta$  et sa variance est  $\text{Var}(X) = \alpha\beta^2$ ,
- 11 Une somme de  $n$  variables aléatoires indépendantes de loi Gamma est une variable de loi Gamma:  $X_i \sim \mathcal{Gm}(\alpha_i; \beta) \Rightarrow \sum_{i=1}^n X_i \sim \mathcal{Gm}(\sum \alpha_i; \beta)$ .
- 12 *La loi exponentielle* La loi exponentielle de paramètre  $\beta$  est une loi  $\mathcal{Gm}(1; \beta)$ . Son espérance est  $\beta$  et sa variance  $\beta^2$ .
- 13 Une somme de  $n$  variables aléatoires indépendantes de loi exponentielle de paramètre  $\beta$  est une variable de loi Gamma:  $\sum_{i=1}^n X_i \sim \mathcal{Gm}(n; \beta)$ .
- 14 *Loi khi-deux*  
La loi khi-deux est une loi gamma avec  $\beta = 2$  et  $\alpha = v/2$ ,  $v$  étant un entier positif appelé *nombre de degrés de liberté*.
- 15 *Espérance et variance d'une khi-deux*  
Si  $X \sim \chi_v^2$ ,  $E(X) = v$  et  $\text{Var}(X) = 2v$ .
- 16 *Le carré d'une normale centrée-réduite est une  $\chi_1^2$* .  
La somme des carrés de  $n$  normales centrées-réduites est une  $\chi_n^2$ .

### Exercices

- 3.1 Le nombre d'arrivées à un comptoir de service suit une loi de Poisson de moyenne deux par heure.
- Quelle est la probabilité qu'aucun client n'arrive durant la première heure ?
  - 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 ?
  - Si l'ouverture est à 9 heures, quelle est la probabilité que la première arrivée ait lieu entre 9 :30 heures et 10 heures ?
- 3.2 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.
- 3.3 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 ?
- 3.4 En moyenne, un moteur d'avion peut fonctionner 2 000 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.
- 3.5 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.
- Quelle est la probabilité qu'il puisse terminer son vol si les 4 moteurs sont nécessaires pour voler ?
  - Quelle est la probabilité qu'il puisse terminer son vol si l'avion peut voler avec un seul moteur ?
- 3.6 Quelle est la probabilité qu'une variable de loi exponentielle prenne une valeur située à plus d'un écart-type de sa moyenne ?
- 3.7 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. À 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.
- 3.8 Sachant que  $X$  suit une loi  $\mathcal{N}(100 ; 256)$ , calculer chacune des probabilités suivantes.
- |                             |                           |                        |
|-----------------------------|---------------------------|------------------------|
| a) $P(100 \leq X \leq 150)$ | b) $P(X \leq 92)$         | c) $P( X-100  \leq 8)$ |
| d) $P(52 < X \leq 148)$     | e) $P(76 \leq X \leq 92)$ | f) $P( X-100  > 8)$    |
- 3.9 Soit  $Z \sim \mathcal{N}(0 ; 1)$ . Déterminer la valeur de  $a$  telle que
- |                        |                            |                          |
|------------------------|----------------------------|--------------------------|
| a) $P(Z > a) = 0,1056$ | b) $P(0 < Z < a) = 0,3554$ | c) $P(Z < a) = 0,7734$   |
| d) $P(Z < a) = 0,0188$ | e) $P( Z  < a) = 0,8836$   | f) $P( Z  > a) = 0,3422$ |
- 3.10 Soit  $X \sim \mathcal{N}(32 ; 81)$ . Déterminer la valeur de  $a$  telle que
- |                        |                             |                             |
|------------------------|-----------------------------|-----------------------------|
| a) $P(X > a) = 0,1056$ | b) $P(32 < X < a) = 0,3554$ | c) $P(X < a) = 0,7734$      |
| d) $P(X < a) = 0,0188$ | e) $P( X-32  < a) = 0,8836$ | f) $P( X-32  > a) = 0,3422$ |
- 3.11 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 ?
- 3.12 Lorsqu'une machine est réglée pour mettre  $\mu$  grammes de petits pois dans des boîtes de conserve, elle n'en met pas exactement  $\mu$  grammes. Le poids réel du contenu varie selon une loi normale de moyenne  $\mu$  et d'écart-type  $\sigma$  grammes. Si  $\sigma = 4$ , à quelle valeur doit-on régler  $\mu$  pour que 1% seulement des boîtes contiennent moins de 300 grammes ?

- 3.13 Soit  $X_1$  et  $X_2$  deux variables aléatoires indépendantes,  $X_1 \sim N(25 ; 16)$  et  $X_2 \sim N(23 ; 9)$ . Calculer
- a)  $P(41 \leq X_1 + X_2 \leq 51,5)$       b)  $P(-1,5 \leq X_1 - X_2 \leq 16)$ ,      c)  $P(X_2 - X_1 > 0)$ .
- 3.14 Supposons que le Q. I. moyen de 10 personnes choisies au hasard d'une population normale d'écart-type  $\sigma = 10$  est  $\bar{X} = 110$ . Calculer  $P(\bar{X} \geq 110)$  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.
- 3.15 Un terrain est découpé en 10 lots identiques. Sans engrais, la production de céréales, en tonnes, pour chaque lot, suit une loi  $N(6 ; 1)$ . En utilisant un certain engrais, la production d'un lot sera de loi  $N(6,3 ; 1)$ . 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 ?
- 3.16 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  $\bar{X}_1 - \bar{X}_2$  ?
- 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 ?
- 3.17 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 ?
- 3.18 D'une population normale de moyenne  $\mu = 200\$$  et d'écart-type  $\sigma = 30\$$ , un statisticien prélève un échantillon de  $n = 25$  unités afin d'estimer la moyenne ( $\mu$ ) de la population par la moyenne échantillonnale  $\bar{X}$ .
- a) Quelle est la probabilité qu'il surestime de plus de 4 \$ la vraie moyenne
- b) Quelle est la probabilité qu'il commette une erreur (de sous-estimation ou de surestimation) de plus de 10 \$ dans son estimation ?
- 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 ?
- 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 ?
- 3.19 On prélève un échantillon de taille  $n = 30$  d'une population d'écart-type  $\sigma = 18\$$  pour estimer la moyenne de la population.
- a) Quelle est la probabilité de surestimer la vraie moyenne de plus de 6 \$ ?
- b) Quelle est la probabilité de commettre une erreur (de sous-estimation ou de surestimation) de plus de 6 \$ ?
- 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 ?
- 3.20 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 %) ?
- 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 de 4 % ?
- 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.
- 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 %.
- 3.21 Un cultivateur sème des graines de haricot par rangs de 100 graines. On suppose que 80 % des graines germeront.
- Quel est le nombre moyen de plants obtenu par rang ?
  - Quelle est la variance du nombre de plants obtenu par rang ?
  - Quelle est la probabilité qu'un certain rang contienne moins de 70 plants ?
  - Le cultivateur a semé 20 000 rangs de haricots. Combien de ces rangs devraient normalement contenir moins de 70 plants ?
  - Quelle est la probabilité que plus de 100 rangs contiennent moins de 70 plants ?
- 3.22 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*)  $X_1, X_2, \dots, X_{16}$ . Supposons que  $X_i : N(\mu ; 144)$ ,  $\mu$  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.
- Quelle doit être la valeur de  $C$  s'il faut que la probabilité de rejeter un lot pour lequel  $\mu = 50$  soit de 5 % ?
  - Avec la valeur de  $C$  trouvée en a), déterminer la probabilité de rejeter un lot pour lequel  $\mu = 51, 55, 60, 70$ .
- 3.23 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 ?
- 3.24 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 ?
- 3.25 Soit  $X \sim \mathcal{B}(14 ; 0,4)$ . Calculer par l'approximation normale :
- $P(X \geq 10)$
  - $P(4 \leq X \leq 8)$
  - $P(X \geq 6)$
  - $P(X > 4)$ .
- 3.26 Soit  $X \sim \mathcal{B}(60 ; 0,2)$ . Calculer par l'approximation normale :
- $P(9 \leq X \leq 15)$
  - $P(X \geq 8)$
  - $P(X > 20)$
  - $P(X \leq 4)$ .
- 3.27 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 échantillonnale 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 % ?
- 3.28 Les homards pêchés sur une certaine côte ont un poids distribué selon une loi  $N(0,5 ; 0,01)$ . 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.
- 3.29 Soit  $X_1, X_2$  et  $X_3$  trois variables aléatoires indépendantes, chacune de loi  $N(\mu ; \sigma^2)$ . Soit  $X = X_1 + 2 X_2 - 2 X_3$ . Quelle est la distribution de  $(X - \mu)^2 / (9\sigma^2)$  ?
- 3.30 Une distributrice à café est réglée de façon à verser  $\mu$  ml de café dans chaque tasse. En fait, la quantité versée est variable. Elle suit une loi normale de moyenne  $\mu$  et d'écart-type  $\sigma = 15$  ml. Supposons qu'on fixe  $\mu = 177$  ml.

- a) Quel pourcentage des tasses contiendra entre 163 ml et 192 ml de café ?
- b) Le café déborde lorsque le montant versé est supérieur à 200 ml. Avec quelle fréquence cet événement se produira-t-il ?
- 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 ?
- d) Dix pour cent des tasses auront une quantité de café inférieure à  $x$  ml. Quelle est la valeur de  $x$  ?
- 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  $d$  ?
- f) Supposons qu'on ne voudrait pas que le café déborde plus de 1 % des fois. À quoi doit-on fixer  $\mu$  ?
- g) Supposons qu'il soit possible de contrôler  $\sigma$  aussi bien que  $\mu$ . Déterminer  $\mu$  et  $\sigma$  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 %.
- 3.31 La propriétaire d'un comptoir de sushi a toujours vendu au moins 2 plats de sushi au courant de la dernière heure d'ouverture, soit de 17h00 à 18h00. Soit  $T$  le moment (nombre d'heures après 17h00) où elle vend le 2<sup>e</sup> sushi, et supposons que la fonction de densité de  $T$  est  $f(t) = 6t(1-t)$ ,  $0 \leq t \leq 1$ . 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.

### Exercices théoriques

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

- a) Soit  $Y = \max\{X_1, \dots, X_n\}$ . Déterminer la fonction de répartition et la fonction de densité de  $Y$ .
- b) Soit  $Z = \min\{X_1, \dots, X_n\}$ . Déterminer la fonction de répartition et la fonction de densité de  $Z$ .

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

$$a) f(x) = \begin{cases} 2xe^{-x^2} & x > 0 \\ 0 & \text{sinon} \end{cases} \quad b) f(x) = \begin{cases} 1+x & -1 < x \leq 0 \\ 1-x & 0 < x < 1 \\ 0 & \text{Ailleurs} \end{cases}$$

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

- a) Vérifier que  $f$  est bien une fonction de densité
- b) Déterminer la fonction de répartition de  $X$
- c) Calculer  $P(|X| > 2)$ ,  $P(|X| \leq 1)$ ,  $P(X \leq -3)$

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

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

Déterminer la fonction de répartition.

3.36 Soit  $X$  une variable aléatoire de densité  $f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{ailleurs} \end{cases}$ . Déterminer la fonction de densité  $g(y)$

pour  $Y = \ln X$ .

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

3.38 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  $0 \leq c \leq 1$ ,  $\alpha > 0$ ,  $\beta > 0$  sont des constantes. Déterminez  $E(T)$  et  $\text{Var}(T)$ .

3.39 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].

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

- Montrez que  $f$  est une fonction de densité, et déterminez la fonction de répartition de  $X$ .
- Déterminez  $E(X)$  et  $\text{Var}(X)$ .

3.41 La probabilité d'atteindre une région quelconque  $R$  sur une cible circulaire de rayon  $1m$  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  $X$ .

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

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

- Montrez qu'il s'agit bien d'une fonction de répartition.
- Déterminez la fonction de densité de  $X$ .

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

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

- Combien d'heures, en moyenne, devez-vous attendre avant qu'elle tombe en panne ?
- Cinq de ces pièces sont employées dans un même appareil, et toutes 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,  $E(U_1)$  et  $\text{Var}(U_1)$ . (Vous supposerez que le fonctionnement de l'appareil dépend uniquement du fonctionnement des 5 pièces. Vous supposerez aussi que les pièces fonctionnent indépendamment l'une de l'autre).
- Soit  $U_2$  la durée de l'appareil si on suppose, contrairement à ce qui a été supposé en b), qu'il suffit qu'une seule des 5 pièces fonctionne pour que l'appareil fonctionne. Déterminez la fonction de densité de  $U_2$ .
- 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 fonction de densité de  $U_3$ .

3.44 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  $y = \sqrt{2x}$  et utiliser le fait que  $\int_0^{\infty} e^{-y^2/2} dy = \sqrt{\pi/2}$ .

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

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

3.46 Soit  $X_1, \dots, X_n$ ,  $n$  variables aléatoires indépendantes de loi exponentielle de paramètre  $\beta$ . Montrer que

$$X = 2 \sum_{i=1}^n \frac{X_i}{\beta} \text{ suit une loi } \chi^2 \text{ à } 2n \text{ degrés de liberté.}$$

3.47 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  $\frac{\sqrt{2X} - \sqrt{2n}}{\sqrt{2}}$  s'approche d'une  $N(0; 1)$ .

**Exercices à faire avec l'ordinateur**

- 3.48 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  $\bar{X}$  le poids moyen d'un échantillon de  $n$  boîtes. Calculez la probabilités que  $\bar{X}$  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.
- 3.49 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  $\bar{X}$  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.
- 3.50 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 ?
- b) Calculez la moyenne des  $\bar{X}$  obtenus. Théoriquement, à quoi cette moyenne devrait-elle être égale ?
- c) Calculez la variance des  $\bar{X}$  obtenus. Théoriquement, à quoi cette moyenne devrait-elle être égale ?
- d) Soit  $Z = (\bar{X} - \mu) / \sigma_{\bar{X}}$ . 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 ;
  - (ii) à l'aide de votre simulation.
- 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  $n = 50$ )

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

- 3.51 Soit  $X_1$  et  $X_2$  deux variables aléatoires indépendantes d'une loi  $\mathcal{U}(0; 1)$  (uniforme sur  $[0; 1]$ ).
  - a) Déterminer la fonction de densité de  $Y_1$ .
  - b) Déterminer  $E(Y_1)$  et  $\text{Var}(Y_1)$
  - c) Déterminer la fonction de densité de  $Y_2$
  - d) Déterminer  $E(Y_2)$  et  $\text{Var}(Y_2)$
- 3.52 Soit  $X_1, X_2, \dots, X_n$   $n$  variables aléatoires indépendantes d'une loi  $\mathcal{U}(0; \theta)$  (uniforme sur  $[0; \theta]$ ).
  - a) Déterminer la fonction de répartition de  $Y_n$
  - b) Déterminer la fonction de densité de  $Y_n$ .
  - c) Déterminer l'espérance et la variance de  $Y_n$ .
- 3.53 Supposons que le temps d'attente au métro est une variables de loi uniforme sur  $[0; 10]$ . Vous prenez le métro 10 fois dans la semaine. Quelle est la probabilité que la plus longue attente soit inférieure à 5 minutes? [Voir le numéro 2].
- 3.54 Soit  $X_1, \dots, X_n$  un échantillon aléatoire simple d'une population exponentielle de paramètre  $\beta$ .
  - a) Montrer que  $Y_1$  est de loi exponentielle de paramètre  $\beta/n$ .
  - b) Montrer que la densité de  $Y_n$  est  $g(y_n) = \frac{n}{\beta} e^{-y_n/\beta} [1 - e^{-y_n/\beta}]^{n-1}$



- c) Déterminer  $E(Y_n)$  lorsque  $n = 2$ .
- d) Montrer que pour un échantillon de taille  $n = 2m+1$  la densité de la médiane  $\tilde{x}$  est

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

3.55 Soit  $X_1, \dots, X_n$  un échantillon aléatoire simple d'une population uniforme sur  $(0 ; 1)$ .

- a) Déterminer les distributions de  $Y_1$  et de  $Y_n$ .
- b) Supposons que  $n$  est impair,  $n = 2m+1$ . Déterminer la distribution de la médiane  $\tilde{x}$ .
- c) Déterminer l'espérance et la variance de  $Y_1$ .
- d) Déterminer la probabilité que dans un échantillon de taille  $n = 4$  d'une population uniforme sur  $(0 ; 1)$  la plus petite valeur soit  $\geq 0,2$ .

3.56 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

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

[ La densité conjointe de deux variables  $(X;Y)$  est définie par  $\lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{P(x < X < x + \Delta x, y < Y < y + \Delta y)}{\Delta x \Delta y}$  ]

- b) Appliquer ce résultat à un échantillon d'une population exponentielle de paramètre  $\beta$ .
- c) Appliquer ce résultat à un échantillon d'une population uniforme sur  $(0 ; 1)$ .

3.57 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  $f(x) = (1/2)e^{-(1/2)(x-4)}$ ,  $x \geq 4$ . Soit  $Y$  votre temps d'attente.

- a) Déterminer la fonction de densité de  $Y$
- b) Déterminer l'espérance de votre temps d'attente.

3.58 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  $Y$
- b) Déterminer l'espérance de votre temps d'attente.