Probabilités et Statistiques :
Exemples et Applications

Table des matières

2 Lois de probabilité discrètes

L'ensemble du texte est un recueil de texte de :

- **Aide-mémoire Statistique et probabilités pour l'ingénieur, Dunod, 2014.**

Les lois discrètes sont utilisées pour modéliser les résultats des jeux de hasard, les sondages d'opinion, les phénomènes biologiques, les processus aléatoires (files d'attente, évolution de l'état de matériels) ... Les plus utilisées sont la loi uniforme, la loi binomiale et les lois dérivées, la loi de Poisson.

<u>Exemple :</u>

Soit X la variable aléatoire prenant trois valeurs 0, 1, 2 avec les probabilités :

$$\Pr(X = 0) = 1/2 \qquad \Pr(X = 1) = 1/3 \qquad \Pr(X = 2) = 1/6$$

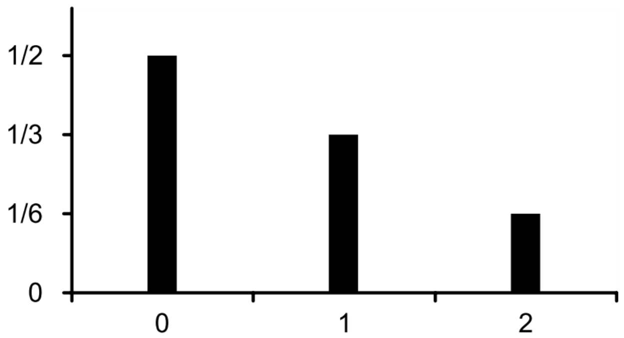
$(1/2 + 1/3 + 1/6 = 1)$

Espérance mathématique :

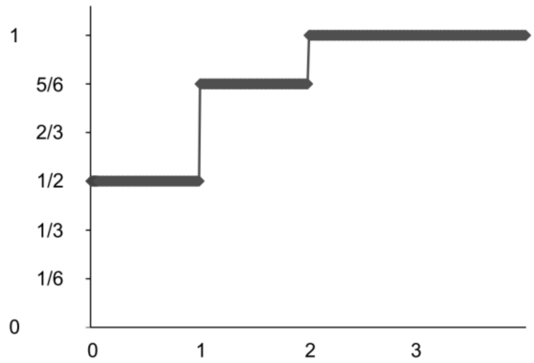
$$E(X) = 0 \times 1/2 + 1 \times 1/3 + 2 \times 1/6 = 2/3$$

Variance :

$$\mathrm{Var}(X) = (\sigma_x)^2 = E(X^2) - [E(X)]^2 = 1 \times 1/3 + 4 \times 1/6 - (2/3)^2 = 5/9$$



Histogramme de la loi donnée dans l'exemple



Fonction de répartition de la loi donnée dans l'exemple

2.1 Loi uniforme

2.1.1 Domaines d'application

La loi de probabilité uniforme intervient dans de nombreux domaines comme les jeux de pile ou face ou les jeux de dés (avec une pièce ou un dé parfaitement équilibré(e)), les jeux de cartes, les loteries, les sondages...

2.2 Loi binomiale ou loi des tirages avec remise

2.2.1 Domaines d'application

- La loi binomiale décrit des phénomènes ne pouvant prendre que deux états s'excluant mutuellement, succès ou échec dans un jeu, bonne pièce ou pièce défectueuse dans une fabrication, lot acceptable ou lot refusé, défaillance ou fonctionnement d'un matériel...
- Elle est utilisée dans le domaine technique pour déterminer la probabilité de défaillance à la sollicitation de matériels, en contrôle qualité, mais elle ne peut s'appliquer rigoureusement que si les expériences sont non exhaustives, c'est la loi du tirage avec remise.
- Les événements considérés doivent être indépendants et la probabilité de réalisation d'un événement doit être constante.

2.2.2 Exemples

Exemple

On veut réaliser une étude clinique sur des malades se présentant à une consultation hospitalière. Pour cette étude, seuls les malades répondant à un ensemble de critères C sont retenus. Des statistiques antérieures ont montré que 20 % des consultants présentent les critères C.

10 malades viennent consulter le premier jour.

Soit X la variable aléatoire « nombre de malades retenus » c'est-à-dire répondant à l'ensemble des critères C. La variable X suit *la loi binomiale* $B(10\,;0{,}20)$.

La probabilité qu'aucun malade ne soit recruté ce jour est égale à :

$$\Pr(X = 0) = C_{10}^{0}\,(0{,}20)^{0}\,(0{,}80)^{10} = 0{,}107$$

La probabilité pour qu'il y ait au moins un malade recruté est égale à :

$$\Pr(X \geqslant 1) = 1 - \Pr(X = 0) = 1 - 0{,}107 = 0{,}803$$

EXAMPLE | Digital Channel

The chance that a bit transmitted through a digital transmission channel is received in error is 0.1. Also, assume that the transmission trials are independent. Let X = the number of bits in error in the next four bits transmitted. Determine $P(X = 2)$.

Let the letter E denote a bit in error, and let the letter O denote that the bit is okay, that is, received without error. We can represent the outcomes of this experiment as a list of four letters that indicate the bits that are in error and those that are okay. For example, the outcome $OEOE$ indicates that the second and fourth bits are in error and the other two bits are okay. The corresponding values for x are

Outcome	x	Outcome	x
$OOOO$	0	$EOOO$	1
$OOOE$	1	$EOOE$	2
$OOEO$	1	$EOEO$	2
$OOEE$	2	$EOEE$	3
$OEOO$	1	$EEOO$	2
$OEOE$	2	$EEOE$	3
$OEEO$	2	$EEEO$	3
$OEEE$	3	$EEEE$	4

The event that $X = 2$ consists of the six outcomes:

$$\{EEOO, EOEO, EOOE, OEEO, OEOE, OOEE\}$$

Using the assumption that the trials are independent, the probability of $\{EEOO\}$ is

$$P(EEOO) = P(E)P(E)P(O)P(O) = (0.1)^2(0.9)^2 = 0.0081$$

Also, any one of the six mutually exclusive outcomes for which $X = 2$ has the same probability of occurring. Therefore,

$$P(X = 2) = 6(0.0081) = 0.0486$$

In general, $P(X = x)$ = (number of outcomes that result in x errors) $\times (0.1)^x(0.9)^{4-x}$.

To complete a general probability formula, only an expression for the number of outcomes that contain x errors is needed. An outcome that contains x errors can be constructed by partitioning the four trials (letters) in the outcome into two groups. One group is of size x and contains the errors, and the other group is of size $n - x$ and consists of the trials that are okay. The number of ways of partitioning four trials into two groups, one of which is of size x, is $\binom{4}{x} = \frac{4!}{x!(4-x)!}$. Therefore, in this example,

$$P(X = x) = \binom{4}{x}(0.1)^x(0.9)^{4-x}$$

Exemple 4.3

L'installation de fabrication représentée schématiquement à la figure 4.5 produit chaque jour des milliers d'unités.

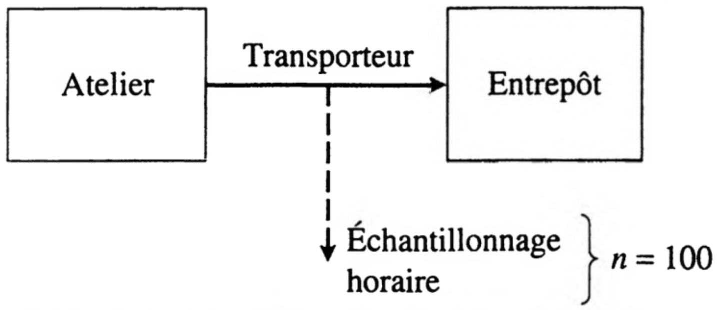


Figure 4.5 **Un échantillonnage pour la vérification d'attributs**

Le pourcentage moyen d'unités défectueuses se situe à 1 % et ne varie pas au fil du temps. Une fois l'heure, on prélève au hasard 100 unités du transporteur de manutention. Une inspectrice en examine et en mesure plusieurs caractéristiques et qualifie chaque unité de parfaite ou de défectueuse. Si l'on envisage ce processus d'échantillonnage comme $n = 100$ épreuves de Bernoulli où $p = 0{,}01$, le nombre total X d'unités défectueuses de l'échantillon obéit à la loi binomiale

$$p(x) = \begin{cases} \binom{100}{x}(0{,}01)^x(0{,}99)^{100-x} & \text{si } x = 0, 1, 2, \ldots, 100 \\ 0 & \text{sinon.} \end{cases}$$

L'inspectrice doit faire cesser la production lorsqu'un échantillon renferme plus de 2 unités défectueuses. Or, $P(X > 2) = 1 - P(X \leq 2)$, et on peut calculer

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \sum_{x=0}^{2} \binom{100}{x}(0{,}01)^x(0{,}99)^{100-x} \\ &= (0{,}99)^{100} + 100(0{,}01)^1(0{,}99)^{99} + 4950(0{,}01)^2(0{,}99)^{98} \\ &\approx 0{,}921. \end{aligned}$$

Il y a donc une probabilité d'environ $1 - 0{,}921 = 0{,}079$ que l'inspectrice fasse cesser la production. Le nombre moyen d'unités défectueuses trouvées dans les échantillons de taille 100 correspond à $E(X) = np = 100(0{,}01) = 1$, avec un écart-type de $\sqrt{V(X)} = \sqrt{np(1-p)} = \sqrt{0{,}99} \approx 0{,}995$.

2.3 Loi géométrique

2.3.1 Domaines d'application

La loi géométrique est la loi de la variable Y « loi du nombre d'essais nécessaires » pour qu'un événement de probabilité p apparaisse pour la première fois, les hypothèses étant les mêmes que pour la loi binomiale, en particulier, la probabilité p reste constante au cours des essais

2.3.2 Exemples

Exemple

Un certain matériel a une probabilité $p = 0{,}02$ constante de défaillance à chaque mise en service. On procède à l'expérience suivante, l'appareil est mis en marche, arrêté, remis en marche, arrêté, jusqu'à ce qu'il tombe en panne. Le nombre d'essais nécessaires pour obtenir la panne est une variable aléatoire suivant la loi géométrique de paramètre p. La probabilité que ce matériel tombe en panne (pour la première fois) au dixième essai est égale à :

$$\Pr(Y = 10) = (0{,}02)(1 - 0{,}02)^9 = 0{,}0167$$

Exemple 4.5

Soit X le nombre de tentatives nécessaires pour obtenir un six en lançant un dé équilibré. On a déjà effectué 5 tentatives sans obtenir le résultat voulu. La probabilité qu'il faille plus de 2 autres tentatives pour y arriver (donc plus de 7 tentatives au total) correspond à la probabilité que le nombre d'essais nécessaires soit supérieur à 2 dès le début de l'expérience.

$$\begin{aligned} P(X>7|X>5) &= \frac{P(X>7)}{P(X>5)} \\ &= \frac{1-P(X\le 7)}{1-P(X\le 5)} \\ &= \frac{1-\sum_{x=1}^{7}\left(\frac{5}{6}\right)^{x-1}\left(\frac{1}{6}\right)}{1-\sum_{x=1}^{5}\left(\frac{5}{6}\right)^{x-1}\left(\frac{1}{6}\right)} \\ &= \frac{25}{36} \end{aligned}$$

$$\begin{aligned} P(X>2) &= 1 - P(X\le 2) \\ &= 1-\sum_{x=1}^{2}(1-p)^{x-1}p \\ &= 1-\left(\frac{1}{6}+\frac{5}{6}\cdot\frac{1}{6}\right) \\ &= \frac{25}{36} \end{aligned}$$

2.4 Loi hypergéométrique

Exemple 5.5

Dans une assemblée de 30 personnes, il y a 20 hommes et 10 femmes. On tire un échantillon de 15 personnes (tirage sans remise).

Soit X la variable aléatoire « nombre d'hommes » dans cet échantillon.

Valeurs extrêmes de cette variable :

– 5 (la valeur 0 ne peut pas être obtenue car il n'y a pas 15 femmes dans l'assemblée),

– 15 (dans l'échantillon, il n'y aura pas de femmes).

La probabilité d'avoir 10 hommes dans un échantillon de taille 15 est égale à :

$$\Pr(X = 10) = \frac{C_{20}^{10}\, C_{10}^{5}}{C_{30}^{15}} = 0{,}30$$

Exemple 4.7

Un client reçoit un lot de 25 articles d'un fournisseur parmi lesquels 3 sont non conformes. Le client tire au hasard et sans remise un échantillon de 5 articles du lot et les inspecte. Soit X le nombre d'articles non conformes qu'il observe dans l'échantillon.

1. **Donnez le nom ainsi que les paramètres de la loi de probabilité de X.**
2. **Calculez la probabilité que le client observe au moins un article non conforme.**
3. **Donnez la moyenne et la variance de X.**

Réponses :

1. La variable X est distribuée selon une loi hypergéométrique de paramètres $N = 25$, $D = 3$ et $n = 5$. Sa fonction de masse est

$$p(x) = \begin{cases} \dfrac{\binom{3}{x}\binom{22}{5-x}}{\binom{25}{5}} & \text{si } x = 0, 1, 2, 3 \\[2ex] 0 & \text{sinon.} \end{cases}$$

2. $P(X \geq 1) = 1 - P(X = 0) = 1 - \dfrac{\binom{3}{0}\binom{22}{5}}{\binom{25}{5}} = 1 - 0{,}496 = 0{,}504.$

3. $E(X) = 5 \cdot \dfrac{3}{25} = 0{,}60$, et $V(X) = 5 \cdot \left(\dfrac{3}{25}\right) \cdot \left(1 - \dfrac{3}{25}\right) \cdot \left(\dfrac{25-5}{25-1}\right) = 0{,}44.$

Exemple 4.8

On reçoit périodiquement des lots de 100 axes de pompe qui doivent faire l'objet d'un contrôle de réception. Voici le plan d'échantillonnage utilisé pour ce contrôle. On prélève du lot, sans remise, un échantillon aléatoire de 10 axes. Si cet échantillon renferme au plus 1 axe défectueux, on accepte le lot. Supposons qu'un lot reçu a une proportion d'axes défectueux de p'. Quelle est la probabilité qu'on l'accepte ?

$$P(\text{accepter un lot}) = P(X \leq 1) = \frac{\sum_{x=0}^{1} \binom{100p'}{x}\binom{100[1-p']}{10-x}}{\binom{100}{10}}$$

$$= \frac{\binom{100p'}{0}\binom{100[1-p']}{10} + \binom{100p'}{1}\binom{100[1-p']}{9}}{\binom{100}{10}}.$$

La probabilité qu'on accepte le lot est manifestement fonction de la qualité de ce lot. Par exemple, si $p' = 0{,}05$, alors

$$P(\text{accepter un lot}) = \frac{\binom{5}{0}\binom{95}{10} + \binom{5}{1}\binom{95}{9}}{\binom{100}{10}} \approx 0{,}923.$$

Si $p' = 0{,}03$, alors la probabilité d'accepter un lot grimpe à 0,974.

EXAMPLE 3.24 | Parts from Suppliers

A batch of parts contains 100 from a local supplier of circuit boards and 200 from a supplier in the next state. If four parts are selected randomly and without replacement, what is the probability they are all from the local supplier?

Let X equal the number of parts in the sample from the local supplier. Then X has a hypergeometric distribution and the requested probability is $P(X = 4)$. Consequently,

$$P(X = 4) = \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}} = 0.0119$$

What is the probability that two or more parts in the sample are from the local supplier?

$$P(X \geq 2) = \frac{\binom{100}{2}\binom{200}{2}}{\binom{300}{4}} + \frac{\binom{100}{3}\binom{200}{1}}{\binom{300}{4}} + \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}$$

$$= 0.298 + 0.098 + 0.0119 = 0.407$$

What is the probability that at least one part in the sample is from the local supplier?

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{100}{0}\binom{200}{4}}{\binom{300}{4}} = 0.804$$

Practical Interpretation: Sampling without replacement is frequently used for inspection and the hypergeometric distribution simplifies the calculations.

EXAMPLE 3.25 | Mean and Variance

In Example 3.24, the sample size is four. The random variable X is the number of parts in the sample from the local supplier. Then, $p = 100/300 = 1/3$. Therefore,

$$E(X) = 4\left(\frac{100}{300}\right) = 1.33$$

and

$$V(X) = 4\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\left(\frac{300-4}{299}\right) = 0.88$$

2.5 Loi multinomiale

2.5.1 Définitions

La loi multinomiale est une généralisation de la loi binomiale.

Une population P est composée d'individus appartenant à k types différents, dans des proportions $p_1, p_2 \dots p_k$ telles que $\sum_{i=1}^{k} p_i = 1$.

On tire un échantillon de n individus, de façon équiprobable et indépendante et on s'intéresse à la composition de l'échantillon.

Soit X_i la variable aléatoire représentant le nombre d'individus de type i dans l'échantillon. Par définition, le vecteur $\underline{X} = (X_1, \dots, X_k)$ est un vecteur aléatoire suivant une *loi multinomiale de paramètres* $(n\ ; p_1, \dots, p_k)$, notée $M(n\ ; p_1, \dots, p_k)$.

$$\begin{aligned} \Pr[X = (x_1, \dots, x_k)] &= \mathrm{C}_n^{x_1}\,\mathrm{C}_{n-x_1}^{x_2} \dots \mathrm{C}_{n-(x_1+\dots+x_{k-1})}^{x_k}\, p_1^{x_1} \dots p_k^{x_k} \\ &= \frac{n!}{x_1! \dots x_k!}\, p_1^{x_1} \dots p_k^{x_k} \end{aligned}$$

$$\sum_{i=1}^{k} p_i = 1 \qquad \sum_{i=1}^{k} x_i = n$$

Le coefficient $\frac{n!}{x_1! \dots x_k!}$ est le nombre de partitions d'un échantillon de taille n en sous-populations d'effectifs x_i (voir annexe 1).

Le nombre d'individus du type i dans l'échantillon suit donc la loi binomiale $B(n\ ; p_i)$. D'où :

$$E(X_i) = np_i \qquad \mathrm{Var}(X_i) = np_i(1 - p_i)$$

Les variables X_i et X_k ne sont pas indépendantes.

Le couple (X_i, X_k) suit une loi multinomiale de dimension 3. En effet, un élément tiré est :

– soit du type i (probabilité p_i),
– soit du type k (probabilité p_k),
– soit de n'importe quel autre type (probabilité $1 - p_i - p_k$).

En partant de ces propriétés, on démontre que :

$$E\left(X_i X_k\right) = n\left(n-1\right)\, p_i\, p_k$$

$$\operatorname{Cov}\left(X_i,\ X_k\right) = E\left(X_i X_k\right) - E\left(X_i\right)\, E\left(X_k\right) = -n\, p_i\, p_k$$

Les variables X_i et X_k ne peuvent donc pas être indépendantes.

2.5.2 Domaines d'application

Soit X une variable aléatoire continue de densité $f(x)$. On suppose que l'espace D_X des valeurs prises par cette variable est partagé en k classes distinctes C_i d'extrémités e_{i-1} et e_i, par exemple tranches d'âges, de revenus, d'impôts...

On considère un échantillon $(X_1, \ldots, X_n)$ de n observations de cette variable et on cherche le nombre de points N_i de l'échantillon dans la classe C_i.

2.5.3 Exemples

Exemple

Un produit d'éclairage de l'entreprise M peut présenter des défectuosités regroupées en trois catégories : défectuosité critique, défectuosité majeure, défectuosité mineure.

Un contrôle final est effectué une semaine après la sortie du produit pour vérifier si certaines défectuosités se seraient développées au cours de cette période. Le résultat du contrôle est le suivant :

- 80 % du produit ne présente aucune défectuosité (ensemble E_1),
- 10 % du produit présente des défectuosités mineures (ensemble E_2),
- 6 % du produit présente des défectuosités majeures (ensemble E_3),
- 4 % du produit présente des défectuosités critiques (ensemble E_4).

Un échantillon de taille $n = 20$ est prélevé au hasard dans un grand lot et vérifié selon les critères précédents.

Soit X_i le nombre d'unités appartenant au sous-ensemble E_i dans l'échantillon contrôlé. L'ensemble (X_1, X_2, X_3, X_4) suit une loi multinomiale qui a pour paramètres les pourcentages donnés par le contrôle. D'où la probabilité :

$$\Pr(X_1 = x_1,\ X_2 = x_2,\ X_3 = x_3,\ X_4 = x_4)$$
$$= C_{20}^{x_1} C_{20-x_1}^{x_2} C_{20-(x_1+x_2)}^{x_3} C_{20-(x_1+x_2+x_3)}^{x_4} (0{,}80)^{x_1} (0{,}10)^{x_2} (0{,}06)^{x_3} (0{,}04)^{x_4}$$
$$= \frac{20!}{x_1!\,x_2!\,x_3!\,x_4!} (0{,}80)^{x_1} (0{,}10)^{x_2} (0{,}06)^{x_3} (0{,}04)^{x_4}$$

Espérance mathématique des variables X_i :

$$E(X_1) = 0{,}80 \times 20 = 16 \qquad E(X_2) = 0{,}10 \times 20 = 2$$
$$E(X_3) = 0{,}06 \times 20 = 1{,}2 \qquad E(X_2) = 0{,}04 \times 20 = 0{,}8$$

On peut calculer différentes probabilités :

$\Pr(X_1 = 10,\ X_2 = 6,\ X_3 = 3,\ X_4 = 1) = 0{,}0001439$

$\Pr(X_1 = 20) = 0{,}0115292$

$\Pr(X_1 = 15,\ X_2 = 5) = 0{,}0054549$

2.6 Loi de poisson

2.6.1 Domaines d'application

La loi de Poisson est la loi discrète d'une variable aléatoire représentant un nombre d'événements. Elle est utilisée pour décrire :

- la réalisation d'événements peu probables, dans une succession d'épreuves très nombreuses, au moins 50,
- le nombre d'accidents dans un atelier, le nombre de défauts sur un appareil,

Elle a des applications dans le domaine des files d'attente. Elle est la loi limite de la loi binomiale quand n tend vers l'infini et p tend vers zéro, le produit *np* restant fini.

2.6.2 Exemples

Exemple

Selon les données recueillies depuis plusieurs années, le nombre de pannes hebdomadaires du système informatique d'une entreprise suit une loi de Poisson de paramètre $\lambda = 0{,}05$.

Soit X la variable aléatoire « nombre de pannes hebdomadaires » :

$$\Pr(X = k) = \frac{e^{-0{,}05}\,(0{,}05)^{k}}{k!}$$

La probabilité que le système tombe en panne une fois au cours d'une semaine quelconque ($k = 1$) est égale à 0,04756.

La probabilité qu'il fonctionne sans panne ($k = 0$) est égale à 0,95122.

On considère une année (50 semaines) de fonctionnement de ce système. Le nombre de pannes Y obéit à une loi de Poisson de paramètre $\mu = 0{,}05 \times 50 = 2{,}5$.

$$\Pr(Y = k) = \frac{e^{-2{,}5}\,(2{,}5)^{k}}{k!}$$

La probabilité d'observer 2 pannes au cours de l'année ($k = 2$) est égale à 0,2565 et la probabilité d'en observer 4 est égale à 0,1336.

THÉORÈME 4.1

Si $X_1, X_2, \ldots, X_k$ **sont des variables aléatoires** indépendantes distribuées selon une loi de Poisson de paramètre c_i, **où** $i = 1, 2, \ldots, k$, **et si** $Y = X_1 + X_2 + \cdots + X_k$, alors Y obéit à une loi de Poisson de paramètre

$$c = c_1 + c_2 + \cdots + c_k.$$

Exemple 4.9

Un détaillant a établi que la demande d'un certain appareil électroménager durant une semaine obéit à une loi de Poisson de paramètre $c = 15$. Il veut déterminer la quantité k qu'il doit avoir en stock en début de semaine pour que la probabilité de remplir toutes les commandes reçues durant la période soit d'au moins 0,95. Le détaillant souhaite éviter toute rupture de stock et tout réapprovisionnement en cours de période. Soit X le nombre de commandes reçues. Le détaillant veut connaître la valeur de k telle que

$$F(k) = P(X \leq k) \geq 0{,}95,$$

de sorte que

$$\sum_{x=0}^{k} \frac{e^{-15} \cdot 15^x}{x!} \geq 0{,}95.$$

On peut trouver la solution en consultant une table de la fonction de répartition de la loi de Poisson ou en utilisant un logiciel. Dans le cas présent, $k = 22$, donc le détaillant doit avoir 22 appareils en stock pour répondre à la demande avec une probabilité de 95 %.

EXAMPLE 3.27 | Calculations for Wire Flaws

For the case of the thin copper wire, suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per millimeter.

Determine the probability of 10 flaws in 5 millimeters of wire. Let X denote the number of flaws in 5 millimeters of wire. Then, X has a Poisson distribution with

$$\lambda T = 2.3 \text{ flaws/mm} \times 5 \text{ mm} = 11.5 \text{ flaws}$$

Therefore,

$$P(X = 10) = e^{-11.5} \frac{11.5^{10}}{10!} = 0.113$$

Determine the probability of at least one flaw in 2 millimeters of wire. Let X denote the number of flaws in 2 millimeters of wire. Then X has a Poisson distribution with

$$\lambda T = 2.3 \text{ flaws/mm} \times 2 \text{ mm} = 4.6 \text{ flaws}$$

Therefore,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-4.6} = 0.9899$$

Practical Interpretation: Given the assumptions for a Poisson process and a value for λ, probabilities can be calculated for intervals of arbitrary length. Such calculations are widely used to set product specifications, control processes, and plan resources.

3 Lois de probabilité continues

L'ensemble du texte est un recueil de texte de :

- **Aide-mémoire Statistique et probabilités pour l'ingénieur, Dunod, 2014.**

3.1 Loi exponentielle

3.1.1 Domaines d'application

– La distribution exponentielle est associée aux processus de Poisson. Un tel processus génère des événements dont les temps d'occurrence sont indépendants et distribués suivant une loi exponentielle (chapitre 9).

– La loi exponentielle est utilisée en *fiabilité* (chapitre 18), le paramètre λ représente le taux de défaillance alors que son inverse $\theta = 1/\lambda$ est le temps moyen de bon fonctionnement MTBF (*Mean Time Between Failure*). Avec le paramètre θ, la densité de probabilité s'écrit :

$$f(x) = \frac{1}{\theta}\ e^{-\frac{x}{\theta}}$$

et les moments sont égaux à :

$$E(X) = \theta$$

$$\mathrm{Var}(X) = \theta^2$$

– La loi exponentielle s'applique bien aux matériels électroniques, c'est-à-dire aux matériels fonctionnant pratiquement sans usure, aux matériels subissant des défaillances brutales ou à des systèmes complexes dont les composants ont des lois de fiabilité différentes. Elle permet de décrire la période de fonctionnement durant laquelle le taux de défaillance est constant ou presque constant.

3.1.2 Exemples

Exemple

On suppose que le temps, en heures, nécessaire pour réparer une machine est une variable aléatoire suivant une loi exponentielle de paramètre $\lambda = 0{,}5$.
La densité de probabilité est $f(t) = 0{,}5\,e^{-0{,}5\,t}$ et la fonction de répartition $F(t) = 1 - e^{-0{,}5\,t}$.
La probabilité pour que le temps de réparation dépasse 2 heures est :

$$\Pr(T > 2) = 1 - \Pr(T < 2) = 1 - F(2) = e^{-1} = 0{,}368$$

Sachant que la réparation a déjà dépassé 9 heures, quelle est la probabilité qu'elle prenne au moins 10 heures ?
La loi exponentielle étant une loi sans « mémoire », on obtient :

$$\Pr(T > 10/T > 9) = \Pr(T > 10 - 9 = 1) = e^{-0{,}5} = 0{,}606$$

Exemple 5.4

Soit un composant électronique dont on peut représenter la durée de vie utile (X) par une fonction de densité exponentielle indiquant un taux de défaillance de 10^{-5} à l'heure (c'est-à-dire $\lambda = 10^{-5}$), d'où une durée moyenne de fonctionnement avant défaillance, $E(X)$, de $1/\lambda = 10^5$ heures. On veut déterminer la proportion des composants de ce type qui feront défaut avant d'avoir atteint leur durée de vie moyenne ou prévue, soit

$$P\left(X \leq \frac{1}{\lambda}\right) = \int_0^{1/\lambda} \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_0^{1/\lambda} = 1 - e^{-1}$$
$$= 0{,}632.$$

Le résultat sera le même pour toute valeur de λ supérieure à zéro. Dans cet exemple, 63,2 % des composants fonctionneront moins de 10^5 heures avant de connaître une défaillance (*voir la figure 5.6*).

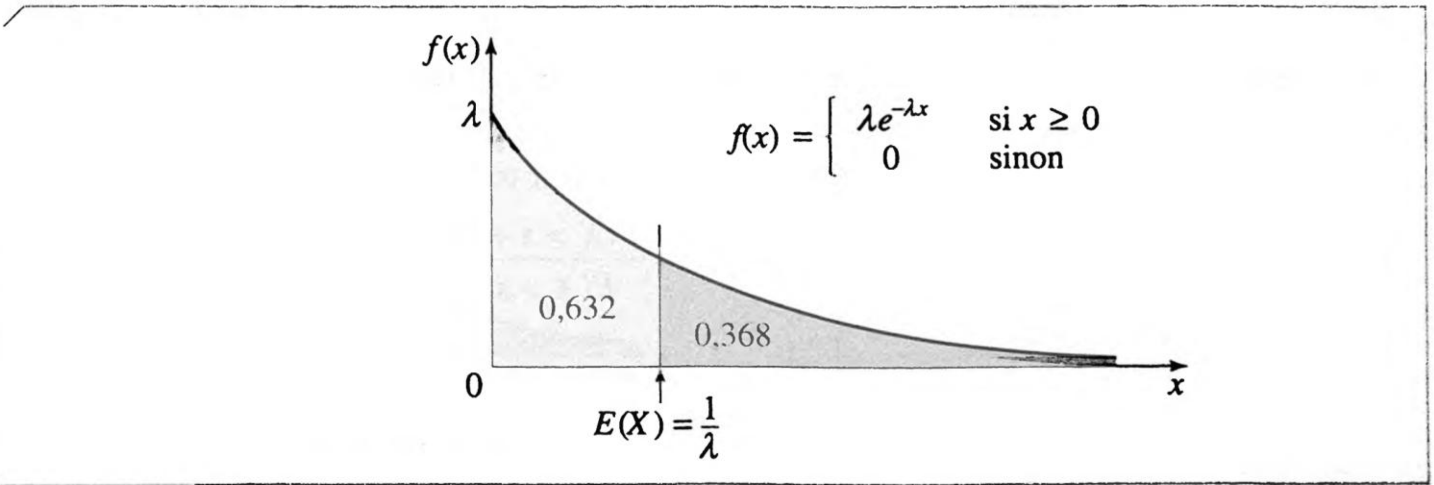


Figure 5.6 **La moyenne d'une variable de loi exponentielle**

Exemple 5.5

On hésite entre deux procédés pour la fabrication d'un composant. Le coût de revient à l'unité sera de 100 dollars si l'on choisit le procédé A et de $k \cdot 100$ dollars (où $k > 1$) si l'on choisit le procédé B. Les deux procédés génèrent des composants dont la durée de fonctionnement avant défaillance obéit à une loi exponentielle, mais le procédé A présente un taux de défaillance de 1/200 à l'heure et le procédé B, un taux de défaillance de 1/300 à l'heure. Les composants fabriqués selon le procédé A ont ainsi une durée de vie moyenne de 200 heures et ceux qui sont fabriqués selon le procédé B (plus chers), une durée de vie moyenne de 300 heures.

Or, une clause de garantie oblige le fabricant à verser une indemnité de 50 dollars pour tout composant qui fonctionne moins de 400 heures avant de connaître une défaillance. Si X représente la durée de fonctionnement avant défaillance de chaque composant, on peut noter les coûts comme suit :

$$C_A = \begin{cases} 100 & \text{si } X \geq 400 \\ 150 & \text{si } X < 400 \end{cases}$$

et

$$C_B = \begin{cases} k \cdot 100 & \text{si } X \geq 400 \\ k \cdot 100 + 50 & \text{si } X < 400. \end{cases}$$

Pour quelle valeur du facteur multiplicatif k le procédé B devient-il avantageux à long terme ?

Les coûts moyens s'établissent ainsi à

$$\begin{aligned} E(C_A) &= 150 \int_0^{400} \frac{1}{200} e^{-x/200}\, dx + 100 \int_{400}^{\infty} \frac{1}{200} e^{-x/200}\, dx \\ &= 150 \left[-e^{-x/200} \Big|_0^{400} \right] + 100 \left[-e^{-x/200} \Big|_{400}^{\infty} \right] \\ &= 100 + 50(1 - e^{-2}) \end{aligned}$$

et à

$$\begin{aligned} E(C_B) &= (k \cdot 100 + 50) \int_0^{400} \frac{1}{300} e^{-x/300}\, dx + k \cdot 100 \int_{400}^{\infty} \frac{1}{300} e^{-x/300}\, dx \\ &= k \cdot 100 + 50(1 - e^{-4/3}). \end{aligned}$$

On peut s'attendre à ce que le fabricant opte pour le procédé B si $E(C_B) < E(C_A)$. Cela survient lorsque $k < 1 - \frac{50}{100} (e^{-2} - e^{-4/3}) = 1{,}0641$. Il faudrait donc que le coût de revient de B soit inférieur à 106,41 \$.

Exemple 5.6

Considérons un détecteur de mouvement activé en moyenne toutes les 6 minutes par un passant. Le temps écoulé entre deux passants suit donc une loi exponentielle de paramètre $\lambda = 1/6$, dont la densité est illustrée à la figure 5.7.

La probabilité que le prochain passant à générer un signal arrive dans plus de 4 minutes correspond à la surface A de la figure 5.7a), et se calcule ainsi :

$$P(X > 4) = 1 - F(4) = e^{-(1/6)(4)} = 0{,}513.$$

Supposons que personne n'est passé depuis 6 minutes. Quelle est la probabilité qu'aucun passant n'active le détecteur d'ici les 4 prochaines minutes ? Autrement dit, quelle est la probabilité qu'un intervalle de 10 minutes s'écoule sans passant, sachant que rien n'a été détecté dans les 6 premières minutes ?

$P(X > 10 | X > 6)$ correspond au rapport des surfaces $C/(B + C)$ de la figure 5.7b), et se calcule ainsi :

$$P(X > 10 \mid X > 6) = \frac{P(X > 10)}{P(X > 6)} = \frac{e^{-(1/6)(10)}}{e^{-(1/6)(6)}} = e^{-(1/6)(4)} = 0{,}513.$$

Le fait que personne ne soit passé dans les 6 premières minutes ne change donc rien à la probabilité que quelqu'un passe dans les 4 prochaines minutes. Voilà pourquoi on parle d'absence de mémoire pour la loi exponentielle.

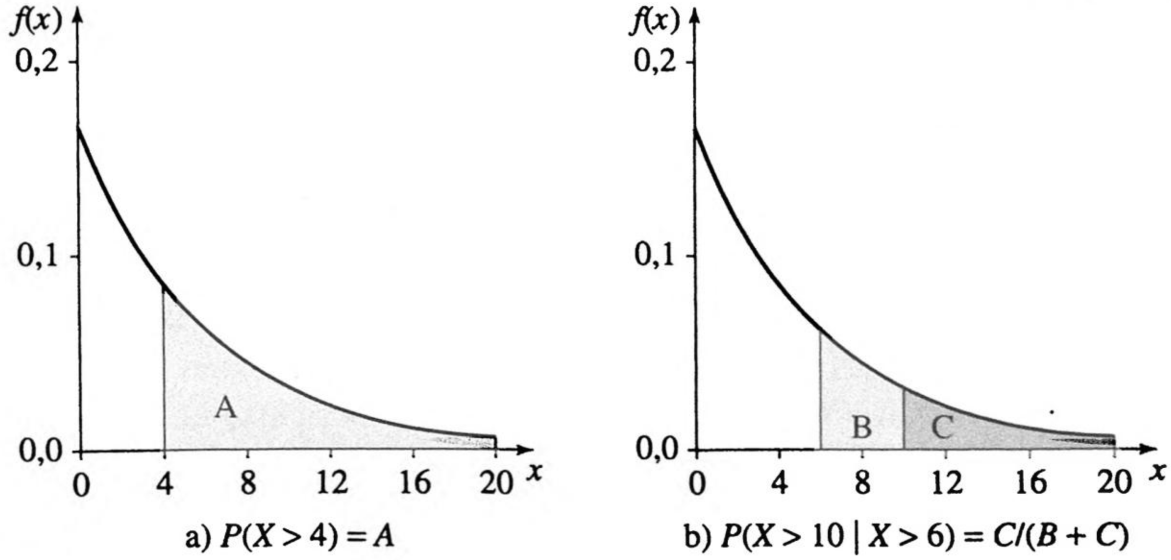


Figure 5.7 **La densité de la loi exponentielle de paramètre $\lambda = 1/6$**

3.2 Loi normale

Exemple Utilisation de la table de la loi normale

Soit X une variable suivant la loi normale $N(3\,;2)$, donc de moyenne 3 et d'écart-type 2. On veut calculer les probabilités suivantes : $\Pr(X < 4)$, $\Pr(X < -1)$, $\Pr(X > 1)$ ou les nombres a_i tels que $\Pr(X < a_1) = 0{,}75$, $\Pr(X > a_2) = 0{,}85$. On utilise la variable centrée réduite U associée à la variable X :

$$U = \frac{X-3}{2} \quad \text{et} \quad X = 2\,U + 3$$

$$\Pr(X < 4) = \Pr(2\,U + 3 < 4) = \Pr(U < 0{,}50) = 0{,}6915$$

$$\Pr(X < -1) = \Pr(2\,U + 3 < -1) = \Pr(U < -2) = 0{,}0228$$

$$\Pr(X > 1) = \Pr(2\,U + 3 > 1) = \Pr(U > -1) = 0{,}8413$$

$$\Pr(X < a_1) = \Pr(2\,U + 3 < a_1) = \Pr\left(U < \frac{a_1-3}{2}\right) = 0{,}75$$

$$\Pr(U < 0{,}6745) = 0{,}75 \quad \text{D'où :} \; a_1 = 4{,}35$$

$$\Pr(X > a_2) = \Pr(2\,U + 3 > a_2) = \Pr\left(U > \frac{a_2-3}{2}\right) = 0{,}85$$

$$\Pr(U < -1{,}0364) = 0{,}15 \quad \Pr(U > -1{,}0364) = 0{,}85$$

$$\text{D'où :} \; a_2 = -1{,}0364 \times 2 + 3 = 0{,}9272$$

EXAMPLE 4.10 | Normal Distribution Calculations

The following calculations are shown pictorially in Figure 4.13. In practice, a probability is often rounded to one or two significant digits.

1. $P(Z > 1.26) = 1 - P(Z \leq 1.26) = 1 - 0.89616 = 0.10384$.
2. $P(Z < -0.86) = 0.19490$.
3. $P(Z > -1.37) = P(Z < 1.37) = 0.91465$.
4. $P(-1.25 < Z < 0.37)$. This probability can be found from the difference of two areas, $P(Z < 0.37) - P(Z < -1.25)$. Now,

$$P(Z < 0.37) = 0.64431$$

and

$$P(Z < -1.25) = 0.10565$$

Therefore,

$$P(-1.25 < Z < 0.37) = 0.64431 - 0.10565 = 0.53866$$

5. $P(Z \leq -4.6)$ cannot be found exactly from Appendix Table III. However, the last entry in the table can be used to find that $P(Z \leq -3.99) = 0.00003$. Because $P(Z \leq -4.6) < P(Z \leq -3.99)$, $P(Z \leq -4.6)$ is nearly zero.
6. Find the value z such that $P(Z > z) = 0.05$. This probability expression can be written as $P(Z \leq z) = 0.95$. Now Table III is used in reverse. We search through the probabilities to find the value that corresponds to 0.95. The solution is illustrated in Figure 4.13. We do not find 0.95 exactly; the nearest value is 0.95053, corresponding to $z = 1.65$.
7. Find the value of z such that $P(-z < Z < z) = 0.99$. Because of the symmetry of the normal distribution, if the area of the shaded region in Figure 4.13(7) is to equal 0.99, the area in each tail of the distribution must equal 0.005. Therefore, the value for z corresponds to a probability of 0.995 in Table III. The nearest probability in Table III is 0.99506 when $z = 2.58$.

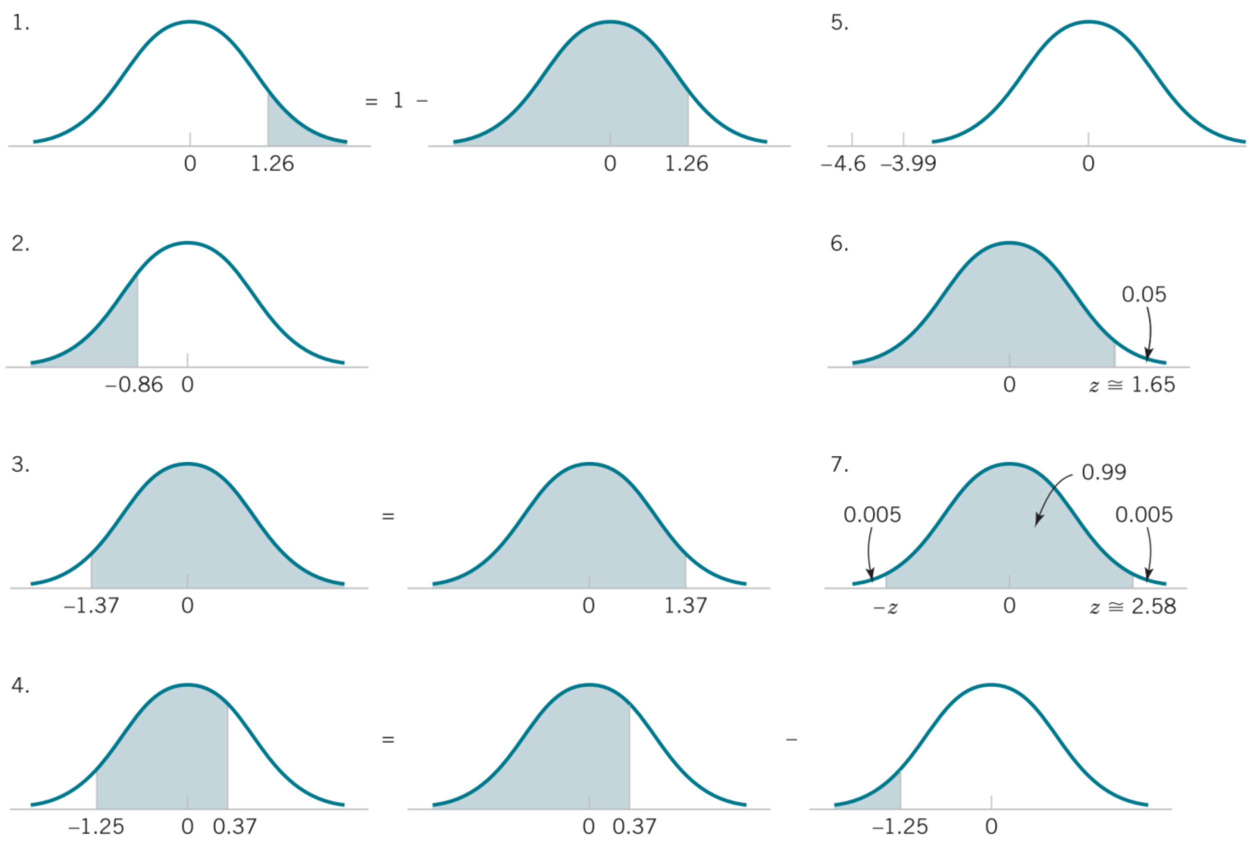


FIGURE 4.13

Graphical displays for standard normal distributions.

EXAMPLE 4.11 | Normally Distributed Current

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 $(\text{milliamperes})^2$. What is the probability that a measurement exceeds 13 milliamperes?

Let X denote the current in milliamperes. The requested probability can be represented as $P(X > 13)$. Let $Z = (X - 10)/2$. The relationship between the several values of X and the transformed values of Z are shown in Figure 4.14. We note that $X > 13$ corresponds to $Z > 1.5$. Therefore, from Appendix Table III,

$$P(X > 13) = P(Z > 1.5) = 1 - P(Z \le 1.5)$$
$$= 1 - 0.93319 = 0.06681$$

Rather than using Figure 4.14, the probability can be found from the inequality $X > 13$. That is,

$$P(X > 13) = P\left(\frac{X-10}{2} > \frac{13-10}{2}\right)$$
$$= P(Z > 1.5) = 0.06681$$

Practical Interpretation: Probabilities for any normal random variable can be computed with a simple transform to a standard normal random variable.

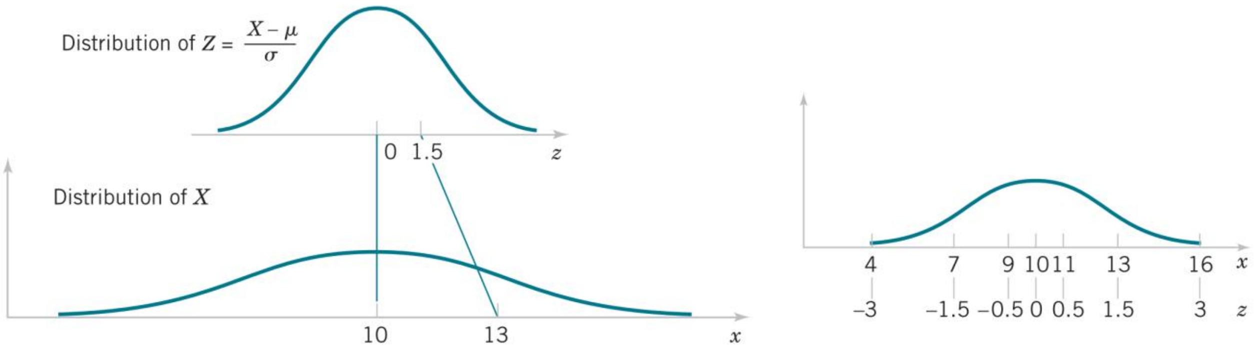


FIGURE 4.14

Standardizing a normal random variable.

EXAMPLE 4.12 | Normally Distributed Current

Continuing Example 4.11, what is the probability that a current measurement is between 9 and 11 milliamperes? From Figure 4.14, or by proceeding algebraically, we have

$$\begin{aligned} P(9 < X < 11) &= P\left(\frac{9-10}{2} < \frac{X-10}{2} < \frac{11-10}{2}\right) \\ &= P(-0.5 < Z < 0.5) \\ &= P(Z < 0.5) - P(Z < -0.5) \\ &= 0.69146 - 0.30854 = 0.38292 \end{aligned}$$

Determine the value for which the probability that a current measurement is less than this value is 0.98. The requested value is shown graphically in Figure 4.15. We need the value of x such that $P(X < x) = 0.98$. By standardizing, this probability expression can be written as

$$\begin{aligned} P(X < x) &= P\left(\frac{X-10}{2} < \frac{x-10}{2}\right) \\ &= P\left(Z < \frac{x-10}{2}\right) = 0.98 \end{aligned}$$

Appendix Table III is used to find the z-value such that $P(Z < z) = 0.98$. The nearest probability from Table III results in

$$P(Z < 2.06) = 0.980301$$

Therefore, $(x - 10)/2 = 2.06$, and the standardizing transformation is used in reverse to solve for x. The result is

$$x = 2(2.06) + 10 = 14.1 \text{ mA}$$

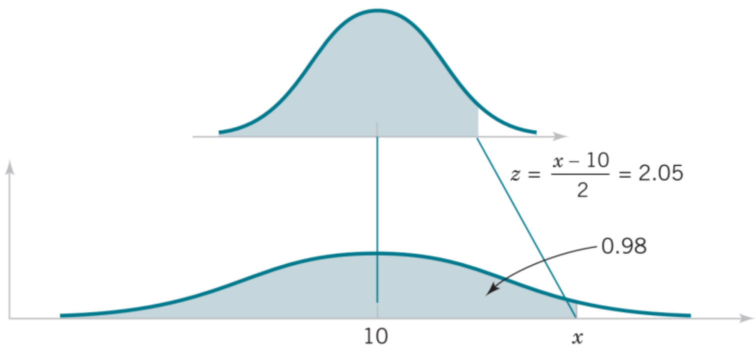


FIGURE 4.15

Determining the value of x to meet a specified probability.

3.3 Limit approximations

3.3.1 Normal Approximation to the Binomial and Poisson Distributions

EXAMPLE 4.13

Assume that in a digital communication channel, the number of bits received in error can be modeled by a binomial random variable, and assume that the probability that a bit is received in error is 1×10^{-5}. If 16 million bits are transmitted, what is the probability that 150 or fewer errors occur?

Let the random variable X denote the number of errors. Then X is a binomial random variable and

$$P(X \le 150) = \sum_{x=0}^{150} \binom{16{,}000{,}000}{x} (10^{-5})^x (1 - 10^{-5})^{16{,}000{,}000-x}$$

Practical Interpretation: Clearly, this probability is difficult to compute. Fortunately, the normal distribution can be used to provide an excellent approximation in this example.

Normal Approximation to the Binomial Distribution

If X is a binomial random variable with parameters n and p,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \qquad (4.12)$$

is approximately a standard normal random variable. To approximate a binomial probability with a normal distribution, a **continuity correction** is applied as follows:

$$P(X \le x) = P(X \le x + 0.5) \approx P\left(Z \le \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

and

$$P(x \le X) = P(x - 0.5 \le X) \approx P\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}} \le Z\right)$$

The approximation is good for $np > 5$ and $n(1 - p) > 5$.

EXAMPLE 4.14

The digital communication problem in Example 4.13 is solved as follows:

$$P(X \le 150) = P(X \le 150.5)$$
$$= P\left(\frac{X - 160}{\sqrt{160(1 - 10^{-5})}} \le \frac{150.5 - 160}{\sqrt{160(1 - 10^{-5})}}\right)$$
$$\approx P(Z \le -0.75) = 0.227$$

Because $np = (16 \times 10^6)(1 \times 10^{-5}) = 160$ and $n(1 - p)$ is much larger, the approximation is expected to work well in this case.

Practical Interpretation: Binomial probabilities that are difficult to compute exactly can be approximated with easy-to-compute probabilities based on the normal distribution.

EXAMPLE 4.15 | Normal Approximation to Binomial

Again consider the transmission of bits in Example 4.14. To judge how well the normal approximation works, assume that only $n = 50$ bits are to be transmitted and that the probability of an error is $p = 0.1$. The exact probability that two or fewer errors occur is

$$P(X \le 2) = \binom{50}{0} 0.9^{50} + \binom{50}{1} 0.1(0.9^{49}) + \binom{50}{2} 0.1^2(0.9^{48}) = 0.112$$

Based on the normal approximation,

$$P(X \le 2) = P\left(\frac{X-5}{\sqrt{50(0.1)(0.9)}} \le \frac{2.5-5}{\sqrt{50(0.1)(0.9)}}\right) \approx P(Z < -1.18) = 0.119$$

We can also approximate $P(X = 5)$ as

$$P(5 \le X \le 5) = P(4.5 \le X \le 5.5) \approx P\left(\frac{4.5-5}{2.12} \le Z \le \frac{5.5-5}{2.12}\right) = P(-0.24 \le Z \le 0.24) = 0.19$$

and this compares well with the exact answer of 0.1849.

Practical Interpretation: Even for a sample as small as 50 bits, the normal approximation is reasonable, when $p = 0.1$.

Normal Approximation to the Poisson Distribution

If X is a Poisson random variable with $E(X) = \lambda$ and $V(X) = \lambda$,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \tag{4.13}$$

is approximately a standard normal random variable. The same continuity correction used for the binomial distribution can also be applied. The approximation is good for

$$\lambda > 5$$

EXAMPLE 4.16 | Normal Approximation to Poisson

Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the probability that 950 or fewer particles are found?

This probability can be expressed exactly as

$$P(X \le 950) = \sum_{x=0}^{950} \frac{e^{-1000} 1000^x}{x!}$$

The computational difficulty is clear. The probability can be approximated as

$$P(X \le 950) = P(X \le 950.5) \approx P\left(Z \le \frac{950.5 - 1000}{\sqrt{1000}}\right) = P(Z \le -1.57) = 0.058$$

Practical Interpretation: Poisson probabilities that are difficult to compute exactly can be approximated with easy-to-compute probabilities based on the normal distribution.

3.3.2 Exponential Distribution

Exponential Distribution

The random variable X that equals the distance between successive events from a Poisson process with mean number of events $\lambda > 0$ per unit interval is an **exponential random variable** with parameter λ. The probability density function of X is

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \le x < \infty \tag{4.14}$$

Mean and Variance

If the random variable X has an exponential distribution with parameter λ,

$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2} \tag{4.15}$$

EXAMPLE | Computer Usage

In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour. What is the probability that there are no log-ons in an interval of 6 minutes?

Let X denote the time in hours from the start of the interval until the first log-on. Then X has an exponential distribution with $\lambda = 25$ log-ons per hour. We are interested in the probability that X exceeds 6 minutes. Because λ is given in log-ons per hour, we express all time units in hours. That is, 6 minutes = 0.1 hour. Therefore,

$$P(X > 0.1) = \int_{0.1}^{\infty} 25e^{-25x}dx = e^{-25(0.1)} = 0.082$$

The cumulative distribution function can also be used to obtain the same result as follows:

$$P(X > 0.1) = 1 - F(0.1) = e^{-25(0.1)}$$

An identical answer is obtained by expressing the mean number of log-ons as 0.417 log-ons per minute and computing the probability that the time until the next log-on exceeds 6 minutes. Try it.

What is the probability that the time until the next log-on is between 2 and 3 minutes? Upon converting all units to hours,

$$P(0.033 < X < 0.05) = \int_{0.033}^{0.05} 25e^{-25x}dx$$
$$= -e^{-25x}\Big|_{0.033}^{0.05} = 0.152$$

An alternative solution is

$$P(0.033 < X < 0.05) = F(0.05) - F(0.033) = 0.152$$

Determine the interval of time such that the probability that no log-on occurs in the interval is 0.90. The question asks for the length of time x such that $P(X > x) = 0.90$. Now,

$$P(X > x) = e^{-25x} = 0.90$$

Take the (natural) log of both sides to obtain $-25x = \ln(0.90) = -0.1054$. Therefore,

$$x = 0.00421 \text{ hour} = 0.25 \text{ minute}$$

Furthermore, the mean time until the next log-on is

$$\mu = 1/25 = 0.04 \text{ hour} = 2.4 \text{ minutes}$$

The standard deviation of the time until the next log-on is

$$\sigma = 1/25 \text{ hours} = 2.4 \text{ minutes}$$

Practical Interpretation: Organizations make wide use of probabilities for exponential random variables to evaluate resources and staffing levels to meet customer service needs.

4 Estimation

4.1 Recettes

7.1 Recettes

Une usine produit deux objets A et B. La quantité d'objets A produite en une journée suit une loi normale de moyenne 2 000 et d'écart-type 500. La quantité d'objet B produite en une journée suit une loi normale de moyenne 1 000 et d'écart-type 300. Un objet A est vendu 30 euros et un objet B est vendu 50 euros. On suppose que tous les objets produits sont vendus.

a) Donner la loi de R représentant la recette réalisée en une journée. Donner sa moyenne et son écart-type.

b) Les dépenses liées à la production engendrent un coût fixe de 80 000 euros quelle que soit la production. Déterminer la probabilité que l'usine perde de l'argent en une journée.

c) On observe les recettes de cette usine sur un mois de 30 jours. Donner un intervalle de confiance à 5 % de la recette moyenne sur ces 30 jours.

d) Expliquer comment évoluerait cet intervalle de confiance si le nombre de jours d'observation augmentait.

Solution

a) Notons X le nombre d'objets A vendus et Y le nombre d'objets B vendus. La recette est $R = 30X + 50Y$.

On a :

$$\begin{aligned} E(R) &= E(30X + 50Y) = 30E(X) + 50E(Y) \\ &= 30 \times 2\,000 + 50 \times 1\,000 = 110\,000 \text{ euros} \end{aligned}$$

$$\begin{aligned} V(R) &= V(30X + 50Y) = 30^2 V(X) + 50^2 V(Y) \\ &= 30^2 \times 500^2 + 50^2 \times 300^2 = 450\,000\,000 \end{aligned}$$

Ainsi l'écart-type de R est $\sigma_R = \sqrt{450\,000\,000} = 21\,213{,}2$ euros.

R suit donc une loi normale de moyenne 110 000 et d'écart-type 21 213,2.

b) On cherche la probabilité :

$$\begin{aligned} P(R < 80\,000) &= P\left(U < \frac{80\,000 - 110\,000}{21213{,}2}\right) = P(U < -1{,}41) \\ &= P(U > 1{,}41) = 1 - P(U < 1{,}41) = 1 - 0{,}92073 \\ &= 7{,}927\ \% \end{aligned}$$

c) L'intervalle de confiance est :

$$\left[110\,000 - 1{,}96\frac{21\,213{,}2}{\sqrt{30}}\ ;\ \ 110\,000 + 1{,}96\frac{21\,213{,}2}{\sqrt{30}}\right]$$
$$= [102\,408{,}95\ ;\ \ 117\,591{,}05]$$

d) Si le nombre de jours d'observations augmente, la valeur de n augmente et par conséquent la largeur de l'intervalle diminue. Si on dispose de plus d'observations, l'estimation de la moyenne sera plus précise.

4.2 Tueur en liberté

7.2 Tueur en liberté

Un tueur est recherché par la police, il a été signalé 23 fois dans un village à 6km du lieu du crime, 47 fois dans un supermarché à 25 km du lieu du crime et 3 fois dans une autre région à 150 km du lieu du crime.

À quelle distance du lieu du crime la police doit-elle effectuer ses recherches pour avoir 95 % de chances de trouver le tueur ?

Solution :

Il s'agit de faire un intervalle de confiance à 95 % pour la distance moyenne à laquelle se trouve le tueur du lieu du crime. On a $t = 1{,}96$, la moyenne est $m = \frac{23\times6+47\times25+3\times150}{23+47+3} = 24{,}15$ km, le coefficient

$$s = \sqrt{\frac{23+47+3}{23+47+3-1}\left(\frac{23\times6^2+47\times25^2+3\times150^2}{23+47+3} - 24{,}15^2\right)} = 27{,}67\text{km et}$$

$n = 23 + 47 + 3 = 73$

Ainsi l'intervalle de confiance est :

$$\left[24{,}15 - 1{,}96\frac{27{,}67}{\sqrt{73}}\ ;\ 24{,}15 - 1{,}96\frac{27{,}67}{\sqrt{73}}\right] = [17{,}8\ ;\ 30{,}5]$$

Les recherches vont pouvoir se concentrer dans un rayon d'au plus 30,5 km et d'au moins 17,8 km du lieu du crime.

4.3 Référendum

7.3 Référendum

On effectue un sondage sur 100 personnes avant un référendum, 51 personnes affirment qu'elles vont voter oui.

a) Donner un intervalle de confiance au risque 5 % de la proportion de personnes qui vont voter oui.

b) Peut-on affirmer que le oui va l'emporter ?

c) Pour quelle taille de population sondée, peut-on affirmer à 95 % que le oui va l'emporter sachant que la proportion de « votant oui » demeure constante dans l'échantillon ?

Solution :

a) On a $t = 1{,}96$, $n = 100$ et $p = 0{,}51$ ainsi l'intervalle de confiance est :

$$\left[0{,}51 - 1{,}96\sqrt{\frac{0{,}51(1-0{,}51)}{100}}\ ;\ 0{,}51 + 1{,}96\sqrt{\frac{0{,}51(1-0{,}51)}{100}}\right]$$

$= [41{,}2\ \%\ ;\ 60{,}8\ \%]$

b) On peut affirmer au risque 5 % que le résultat du référendum sera entre 41,2 % et 60,8 %, or pour l'emporter le oui doit atteindre 50 % ou plus. On ne peut donc rien affirmer quant au résultat du vote.

c) Pour être assuré au risque 5 % que le oui l'emporte, il faudrait que la borne inférieure de l'intervalle de confiance soit supérieure à 50 %. On suppose que le pourcentage de « votants oui » demeure constant à 51 %.

Ainsi $0{,}51 - 1{,}96\sqrt{\dfrac{(0{,}51(1-0{,}51)}{n}} > 0{,}5$

Soit $0{,}51 - 0{,}5 > 1{,}96\sqrt{\dfrac{(0{,}51(1-0{,}51)}{n}}$

D'où $\dfrac{0{,}01}{1{,}96} > \sqrt{\dfrac{(0{,}51(1-0{,}51)}{n}}$

Soit $\left(\dfrac{0{,}01}{1{,}96}\right)^2 > \dfrac{(0{,}51(1-0{,}51)}{n}$

Ainsi $n > \dfrac{0{,}51(1-0{,}51)}{\left(\dfrac{0{,}01}{1{,}96}\right)^2}$

Soit $n > 9\,600{,}16$. Il est donc nécessaire d'interroger au moins 9 601 personnes pour s'assurer au risque 5 % du résultat du vote.

4.4 Contrôlé de fabrication

7.4 Contrôle de fabrication

On effectue un contrôle de fabrication sur des pièces dont une proportion p est défectueuse. On contrôle un lot de 400 pièces et on trouve 30 pièces défectueuses. Donner un intervalle de confiance pour l'estimation de p au niveau 95 %, puis 99 %.

Solution :

Il s'agit d'une application directe des intervalles de confiance pour une proportion. La proportion dans l'échantillon est de $\frac{30}{400} = 0{,}075$

– Au niveau 95 %, on a $t = 1{,}96$ ainsi :

$$\left[p - t\sqrt{\frac{p(1-p)}{n}}\ ;\ p + t\sqrt{\frac{p(1-p)}{n}}\right]$$

$$= \left[0{,}075 - 1{,}96\sqrt{\frac{0{,}075(1-0{,}075)}{400}}\ ;\ 0{,}075 + 1{,}96\sqrt{\frac{0{,}075(1-0{,}075)}{400}}\right]$$

$$= [0{,}049\ ;\ 0{,}101]$$

– Au niveau 99 %, on a $t = 2{,}5758$ ainsi :

$$\left[p - t\sqrt{\frac{p(1-p)}{n}}\ ;\ p + t\sqrt{\frac{p(1-p)}{n}}\right]$$

$$= \left[0{,}075 - 2{,}5758\sqrt{\frac{0{,}075(1-0{,}075)}{400}}\ ;\ 0{,}075 + 2{,}5758\sqrt{\frac{0{,}075(1-0{,}075)}{400}}\right]$$

$$= [0{,}041\ ;\ 0{,}108]$$

5 Introduction à l'apprentissage automatique

Références :

- https://eduscol.education.fr/sti/sites/eduscol.education.fr.sti/files/ressources/pedagogiques/14512/14512-introduction-lapprentissage-automatique-ensps.pdf

Le domaine de l'intelligence artificielle a pour objectif de parvenir à simuler l'intelligence humaine et en particulier l'apprentissage de nombreuses tâches. Deux méthodes sont alors possibles pour apprendre :

- L'apprentissage par cœur consiste à mémoriser explicitement tous les exemples possibles afin de pouvoir les restituer ;
- L'apprentissage par généralisation a pour objectif d'extraire des règles implicites à partir d'une quantité d'exemples afin de les réappliquer à de nouvelles situations jamais rencontrées. L'apprentissage par cœur est relativement aisé pour une machine à condition de disposer des exemples. En revanche, l'apprentissage par généralisation est difficile car il demande d'extraire des règles qui ne sont pas explicitement mentionnées dans les exemples. Ce défi constitue le cœur l'apprentissage automatique.

Comme expliqué précédemment, l'apprentissage automatique est un champ de l'IA porté sur l'analyse statistiques de données d'apprentissage. Historiquement, cette branche est définie comme le développement de machines capables d'apprendre sans avoir été explicitement programmées à apprendre une tâche.

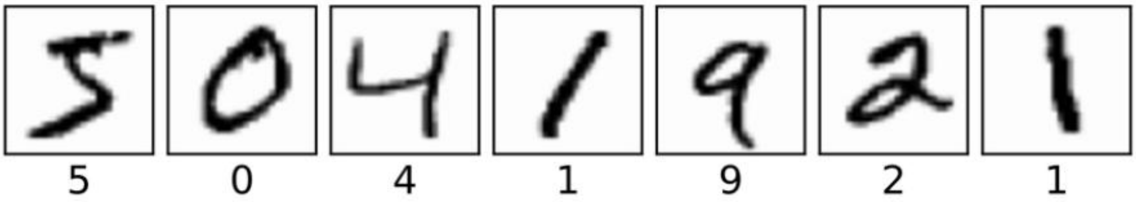
5.1 Exemple 1

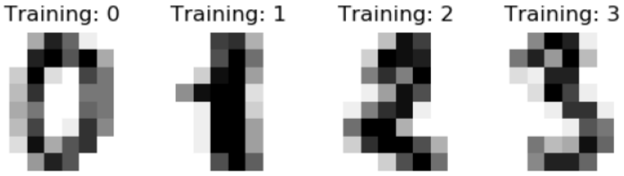
Figure : Exemple de chiffres manuscrits

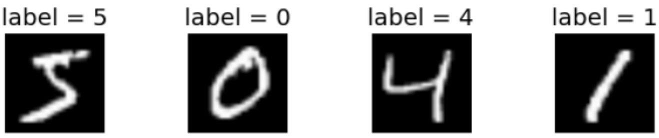
Extrait de la base de données MNIST, couramment utilisée en apprentissage automatique

Etapes d'apprentissage automatique (supervisé) :

1. **Classification manuelle** des objets de la base d'apprentissage :





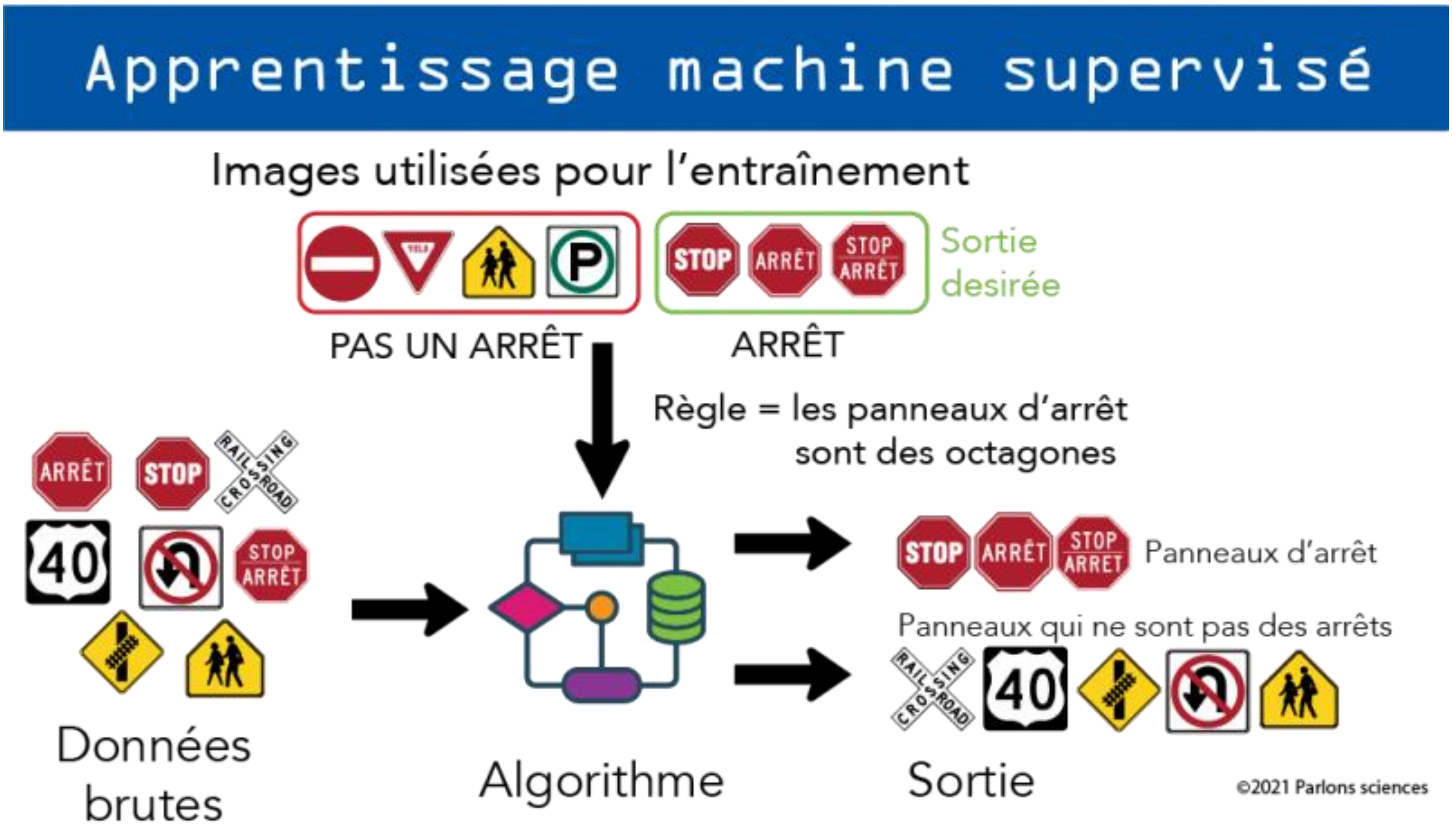


…………………………

…………………………

2. Application d'un **algorithme d'apprentissage** des classes et leurs objets
3. Application grandeur réelle pour **reconnaitre de nouveaux objets** :

5.2 Exemple 2


Apprentissage machine supervisé
Images utilisées pour l'entraînement
STOP
ARRÊT
STOP ARRÊT
Sortie desirée
PAS UN ARRÊT
ARRÊT
Règle = les panneaux d'arrêt sont des octagones
RAILROAD CROSSING
40
Données brutes
Algorithme
Panneaux d'arrêt
Panneaux qui ne sont pas des arrêts
Sortie
©2021 Parlons sciences