

Module 4 - Quelques lois de probabilité continues et théorème central limite

STT-1000 Probabilités et statistique

Département de mathématiques et de statistique



UNIVERSITÉ
LAVAL

Loi uniforme

Loi uniforme

Il s'agit de la loi d'une variable aléatoire

- qui peut prendre toutes les valeurs possibles dans un intervalle $[a, b]$;
- dont tous les intervalles de même longueur sont **équiprobables**.

Fonction de densité :

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{sinon.} \end{cases}$$

$$E(X) = \frac{a+b}{2}$$

$$V(X) = \frac{(b-a)^2}{12}$$

Montrez les résultats pour $E(X)$ et $V(X)$.

Loi uniforme

Fonction de répartition :

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(s) \, ds \\ &= \begin{cases} \int_{-\infty}^x 0 \, ds, & x < a \\ \int_{-\infty}^a 0 \, ds + \int_a^x \frac{1}{b-a} \, ds, & a \leq x < b \\ \int_{-\infty}^a 0 \, ds + \int_a^b \frac{1}{b-a} \, ds + \int_b^x 0 \, ds, & x \geq b. \end{cases} \\ &= \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b. \end{cases} \end{aligned}$$

Exemple

La puissance de calcul demandée à un processeur lors de l'exécution d'une tâche peut être n'importe quelle valeur entre 1,5 et 2 GHz, et chaque valeur dans cet intervalle est équiprobable.

Quelle est la probabilité que la puissance demandée excède 1,8 GHz ?

Quelle est l'espérance de la puissance demandée ?

Loi exponentielle

Loi exponentielle

Il s'agit d'une loi liée à la loi de Poisson.

- Soit Y , le nombre d'événements dans un intervalle de temps de longueur t .
Quand on dit que les événements surviennent selon un processus de Poisson, on suppose que $Y \sim \text{Poisson}(\lambda t)$.
- Soit X , le **temps** qui s'écoule entre deux événements.

Alors on peut montrer que X suit une loi **exponentielle de taux λ** , dont la densité est donnée par

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{sinon.} \end{cases}$$

Loi exponentielle

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} f(x) dx \\ &= \int_{-\infty}^0 x 0 dx + \int_0^{\infty} x \lambda e^{-\lambda x} dx \\ \text{(posons } y = \lambda x) \quad &= 0 + \frac{1}{\lambda} \int_{0 \times \lambda}^{\infty \times \lambda} y e^{-y} dy \\ \text{(int. par parties)} \quad &= \frac{1}{\lambda} \left\{ [-y e^{-y}]_0^{\infty} + \int_0^{\infty} e^{-y} dy \right\} \\ &= \frac{1}{\lambda} \end{aligned}$$

Loi exponentielle

Pour la variance :

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_{-\infty}^0 x^2 0 dx + \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx \\ (\text{par parties 2 fois}) \quad &= \frac{2}{\lambda^2} \end{aligned}$$

$$\begin{aligned} \Rightarrow V(X) &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \\ &= \frac{1}{\lambda^2} \end{aligned}$$

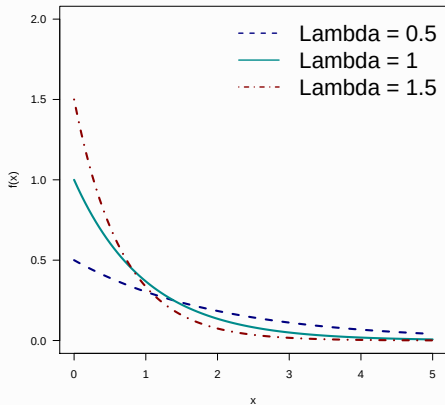
Loi exponentielle

La fonction de répartition :

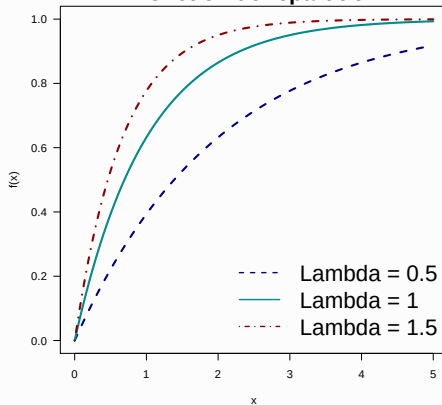
$$\begin{aligned} F(x) &= \int_{-\infty}^x f(s) \, ds \\ &= \begin{cases} \int_{-\infty}^x 0 \, ds, & x < 0 \\ \int_{-\infty}^0 0 \, ds + \int_0^x \lambda e^{-\lambda s} \, ds, & x \geq 0. \end{cases} \\ &= \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases} \end{aligned}$$

Loi exponentielle

Distribution exponentielle – Densité



Distribution exponentielle
Fonction de répartition



Exemple

On suppose que le temps d'attente entre le passage de 2 bateaux sous un pont-levis suit une loi exponentielle de moyenne 3 heures. Quelle est la probabilité que plus de 6 heures s'écoulent entre le passage de 2 bateaux ?

Quelle est la probabilité que 4 bateaux passent en 15 heures ?

Loi exponentielle

La loi exponentielle est **la seule loi continue sans mémoire** :

$$\begin{aligned}P(X > x + s | X > s) &= \frac{P(X > x + s \text{ et } X > s)}{P(X > s)} \\&= \frac{P(X > x + s)}{P(X > s)} = \frac{1 - F_X(x + s)}{1 - F_X(s)} \\&= \frac{1 - \{1 - \exp(-\lambda\{x + s\})\}}{1 - \{1 - \exp(-\lambda s)\}} = \frac{\exp(-\lambda\{x + s\})}{\exp(-\lambda s)} \\&= e^{-\lambda x - \lambda s - (-\lambda s)} = e^{-\lambda x} \\&= 1 - (1 - e^{-\lambda x}) \\&= P(X > x)\end{aligned}$$

Exemple

La loi exponentielle est souvent utilisée pour modéliser la durée de vie des ampoules électriques. On vous dit que la durée de vie d'une certaine ampoule suit une loi exponentielle de moyenne 1000 heures. On vous dit que cette ampoule fonctionne toujours après 500 heures. Quelle est la probabilité qu'elle survivra au moins 1000 autres heures?

Loi Gamma

Loi Gamma

Considérons $Y_1, \dots, Y_n$ des variables aléatoires indépendantes et identiquement distribuées (i.i.d.), telles que $Y_i \sim \exp(\lambda)$, $\forall i = 1, \dots, n$. Alors,

$$X = \sum_{i=1}^n Y_i \sim \text{Gamma}(n, \lambda).$$

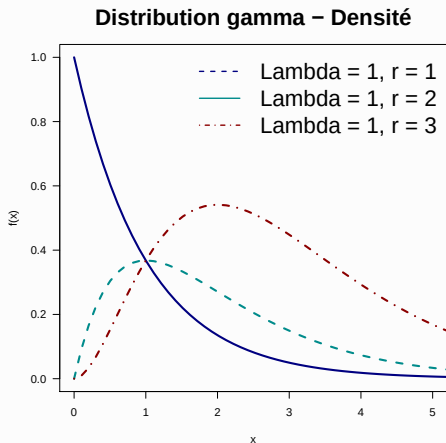
Fonction de densité : Si $X \sim \text{Gamma}(n, \lambda)$, où $n > 0$ et $\lambda > 0$,

$$f_X(x) = \begin{cases} \frac{\lambda^n}{\Gamma(n)} (\lambda x)^{n-1} e^{-\lambda x} & x > 0 \\ 0, & \text{sinon.} \end{cases}$$

$\Gamma(r)$ est une fonction un peu complexe en général, mais pour un entier n on a $\Gamma(n) = (n-1)!$

Fonction de répartition : N'a pas de forme explicite.

Loi Gamma



*notez que le r sur le graphique représente le n dans la fonction de densité présentée plus haut

Loi Gamma

Espérance :

$$E(X) = E(Y_1 + \dots + Y_n) = \sum_{i=1}^n E(Y_i) = \sum_{i=1}^n 1/\lambda = n/\lambda$$

Variance :

$$\begin{aligned} V(X) &= V(Y_1 + \dots + Y_n) \\ &= \sum_{i=1}^n V(Y_i) + 2 \sum_{i < j} \sum_{j} \text{Cov}(Y_i, Y_j) \\ &= \sum_{i=1}^n 1/\lambda^2 + 2 \sum_{i < j} \sum_{j} 0 = \frac{n}{\lambda^2} \end{aligned}$$

Lien avec les processus de Poisson

Si les événements surviennent selon un processus de Poisson de taux λ , alors on a que

- le nombre d'événements dans tout intervalle de longueur t est une v.a. suivant une loi de Poisson de moyenne λt ;
- le temps qui s'écoule entre deux événements (ou avant un premier événement) est une v.a. de loi exponentielle de paramètre λ et donc de moyenne $1/\lambda$;
- le temps qui s'écoule entre $n + 1$ événements (ou avant un n -ème événement) est une v.a. qui suit une loi Gamma(n, λ).

En image

Exemple

Les temps que prend un hacker pour craquer des mots de passe peuvent être vus comme des v.a. indépendantes de loi exponentielle de moyenne 15 minutes. Quelles sont la distribution (loi), la moyenne et la variance du temps qu'il prendra pour craquer les trois mots de passe qui protègent un certain site ?

Exemple (suite)

Si ce hacker a 20 minutes à sa disposition, quelle est la probabilité qu'il craquera 2 mots de passe?

Loi de Weibull

Loi de Weibull

La loi de Weibull permet de modéliser une situation où la probabilité d'événement n'est pas constante dans le temps comme dans le cas de la loi exponentielle. Par exemple, la probabilité de tomber en panne dans un instant donné ne reste pas constante dans le temps, mais augmente avec l'âge.

Le taux de cette augmentation peut être contrôlé par un des paramètres de la loi (paramètre de forme).

Loi de Weibull

Si X suit une loi de Weibull(δ, β, γ) alors

$$f(x) = \begin{cases} \frac{\beta}{\delta} \left(\frac{x-\gamma}{\delta}\right)^{\beta-1} \exp\left\{-\left(\frac{x-\gamma}{\delta}\right)^{\beta}\right\}, & x > \gamma \\ 0, & \text{sinon.} \end{cases}$$

On appelle β “paramètre de forme” et δ “paramètre d’échelle”.

Si $\gamma = 0$ et $\beta = 1$, on retrouve la loi exponentielle de moyenne δ !

Fonction de répartition

En posant $u = \{(s - \gamma)/\delta\}^\beta$ dans $\int_{-\infty}^x f(s) ds$, on trouve

$$F(x) = \begin{cases} 0, & x < \gamma \\ 1 - \exp \left\{ - \left(\frac{x - \gamma}{\delta} \right)^\beta \right\}, & x \geq \gamma. \end{cases}$$

L'espérance et la variance de la loi de Weibull ont des formes un peu compliquées et ne sont pas à l'étude.

Exemple 1

Les panneaux de verre installés sur un édifice ont des durées de vie (en années) indépendantes et toutes de loi de Weibull(20, 2, 0). Ces panneaux sont vendus avec une garantie de 5 ans. Quelle est la probabilité qu'un panneau ne dure pas jusqu'à la fin de sa garantie?

Exemple 1 (suite)

S'il en coûte 1000\$ pour remplacer un panneau qui ne survit pas jusqu'à la fin de la garantie et que l'on vend 500 tels panneaux, quel est le coût total espéré de la garantie ?

Exemple 2

On vous dit que la durée de vie de votre iPod (en années) suit une loi de Weibull avec $\beta = 3$, $\delta = 4$ et $\gamma = 0$.

Quelle est la probabilité que votre iPod survive au moins 3 autres années (i) s'il est neuf?
(ii) s'il est âgé de 2 ans?

Exemple 2 (suite)

Quelle est la durée de vie médiane de tels iPod (c-à-d le 50ème percentile de la distribution, c-à-d le moment tel que 50% des iPod durent moins longtemps et 50% des iPod durent plus longtemps)?

Loi normale

Loi normale

La loi normale est probablement la loi de probabilité la plus utilisée dans les applications de la statistique. Quelques raisons de sa popularité :

- modélise bien la distribution de beaucoup de variables d'intérêt
- approxime la loi de plusieurs variables aléatoires de nature complexe (théorème central limite)
- mène à des résultats analytiques dans plusieurs problèmes d'inférence statistique

Loi normale

La loi normale est définie par son espérance et sa variance.

Si X suit une loi normale d'espérance μ et de variance σ^2 , dénoté $X \sim \mathcal{N}(\mu, \sigma^2)$ alors

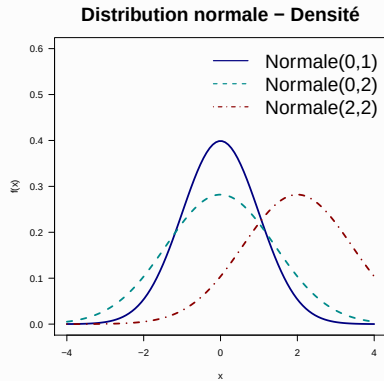
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty.$$

Comme le nom des paramètres l'indique, on a alors

$$E(X) = \mu$$

$$V(X) = \sigma^2$$

Loi normale



Loi normale

- $f(x) \geq 0$ pour tout $x \in \mathbb{R}$
- $\lim_{x \rightarrow -\infty} f(x) = 0$ et $\lim_{x \rightarrow \infty} f(x) = 0$
- $f(\mu - x) = f(\mu + x)$ pour tout $x \in \mathbb{R}$, donc cette densité est **symétrique autour de μ**
- $f(x)$ atteint son unique maximum en $x = \mu$
- $f(x)$ a deux points d'inflexion, un en $\mu - \sigma$ et un en $\mu + \sigma$ (la deuxième dérivée est nulle)

Loi normale

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

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

Donc,

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

Il n'y a pas de solution analytique à cette intégrale. On doit obtenir ces valeurs à l'aide de logiciels, comme R ou d'une table.

Loi normale standard

La loi $\mathcal{N}(0, 1)$ est appelée **loi normale standard** ou encore **loi normale centrée réduite**. Cette loi est très importante, et mérite donc une notation spéciale :

Si $Z \sim \mathcal{N}(0, 1)$ alors on dénote

- $\phi(z)$ la densité de Z

$$\phi(z) = (2\pi)^{-1/2} e^{-z^2/2}$$

- $\Phi(z)$ la fonction de répartition de Z

$\Phi(z)$ n'a toujours pas de forme analytique, mais les livres de statistiques ont une table des valeurs de $\Phi(z)$.

Loi normale standard

Table II Les valeurs de la fonction de répartition de la loi normale centrée réduite*

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Deuxième décimale de z										
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500 00	0,503 99	0,507 98	0,511 97	0,515 95	0,519 94	0,523 92	0,527 90	0,531 88	0,535 86
0,1	0,539 83	0,543 79	0,547 76	0,551,72	0,555 67	0,559 62	0,563 56	0,567 49	0,571 42	0,575 34
0,2	0,579 26	0,583 17	0,587 06	0,590 95	0,594 83	0,598 71	0,602 57	0,606 42	0,610 26	0,614 09
0,3	0,617 91	0,621 72	0,625 51	0,629 30	0,633 07	0,636 83	0,640 58	0,644 31	0,648 03	0,651 73
0,4	0,655 42	0,659 10	0,662 76	0,666 40	0,670 03	0,673 64	0,677 24	0,680 82	0,684 38	0,687 93
0,5	0,691 46	0,694 97	0,698 47	0,701 94	0,705 40	0,708 84	0,712 26	0,715 66	0,719 04	0,722 40
0,6	0,725 75	0,729 07	0,732 37	0,735 65	0,738 91	0,742 15	0,745 37	0,748 57	0,751 75	0,754 90
0,7	0,758 03	0,761 15	0,764 24	0,767 30	0,770 35	0,773 37	0,776 37	0,779 35	0,782 30	0,785 23
0,8	0,788 14	0,791 03	0,793 89	0,796 73	0,799 54	0,802 34	0,805 10	0,807 85	0,810 57	0,813 27
0,9	0,815 94	0,818 59	0,821 21	0,823 81	0,826 39	0,828 94	0,831 47	0,833 97	0,836 46	0,838 91
1,0	0,841 34	0,843 75	0,846 13	0,848 49	0,850 83	0,853 14	0,855 43	0,857 69	0,859 93	0,862 14
1,1	0,864 33	0,866 50	0,868 64	0,870 76	0,872 85	0,874 93	0,876 97	0,879 00	0,881 00	0,882 97
1,2	0,884 93	0,886 86	0,888 77	0,890 65	0,892 51	0,894 35	0,896 16	0,897 96	0,899 73	0,901 47
1,3	0,903 20	0,904 90	0,906 58	0,908 24	0,909 88	0,911 49	0,913 08	0,914 65	0,916 21	0,917 73
1,4	0,919 24	0,920 73	0,922 19	0,923 64	0,925 06	0,926 47	0,927 85	0,929 22	0,930 56	0,931 89
1,5	0,933 19	0,934 48	0,935 74	0,936 99	0,938 22	0,939 43	0,940 62	0,941 79	0,942 95	0,944 08
1,6	0,945 20	0,946 30	0,947 38	0,948 45	0,949 50	0,950 53	0,951 54	0,952 54	0,953 52	0,954 48
1,7	0,955 43	0,956 37	0,957 28	0,958 18	0,959 07	0,959 94	0,960 80	0,961 64	0,962 46	0,963 27
1,8	0,964 07	0,964 85	0,965 62	0,966 37	0,967 11	0,967 84	0,968 56	0,969 26	0,969 95	0,970 62
1,9	0,971 28	0,971 93	0,972 57	0,973 20	0,973 81	0,974 41	0,975 00	0,975 58	0,976 15	0,976 70
2,0	0,977 25	0,977 78	0,978 31	0,978 82	0,979 32	0,979 82	0,980 30	0,980 77	0,981 24	0,981 69
2,1	0,982 14	0,982 57	0,983 00	0,983 41	0,983 82	0,984 22	0,984 61	0,985 00	0,985 37	0,985 74

Loi normale standard

On suppose que $X \sim \mathcal{N}(0, 1)$.

Quelle est la probabilité que X soit plus petit que 1,15 ?

Quelle est la probabilité que X soit plus grand que 1,15 ?

Loi normale standard

Quelle est la probabilité que X soit entre 0,4 et 2,12?

Quelle est la probabilité que X soit plus petit que $-1,4$?

Loi normale standard

Quelle est la probabilité que X soit entre $-1,3$ et $1,98$?

Quelle est la probabilité que $|X|$ soit plus petit que 2 ?

Loi normale standard

On peut aussi lire la table dans l'autre sens, c-à-d trouver les valeurs de x à partir des probabilités :

Quelle est la valeur c telle que $P(X < c) = 78\%$?

Quelle est la valeur c telle que $P(X > c) = 94,3\%$?

Standardisation

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, et qu'on veut calculer des probabilités pour X , on peut utiliser la table pour la loi normale standard grâce au processus de **standardisation** :

$$\begin{aligned} F_X(x) = P(X \leq x) &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \\ &= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned}$$

Pour calculer la valeur de la fonction de répartition en x pour une $\mathcal{N}(\mu, \sigma^2)$, on peut simplement calculer la fonction de répartition de la loi normale standard en $\frac{x-\mu}{\sigma}$, par exemple à l'aide de la table.

Standardisation

Un raisonnement similaire permet de montrer que si

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

alors

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

⇒ Seule une table de la loi normale standard est requise!

Standardisation

Supposons que $X \sim \mathcal{N}(\mu, \sigma^2)$ et que l'on cherche $P(a < X \leq b)$.

Alors si $Z = \frac{(X-\mu)}{\sigma}$, on a que $Z \sim \mathcal{N}(0, 1)$.

Donc

$$\begin{aligned}P(a < X \leq b) &= P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) \\&= P\left(\frac{a - \mu}{\sigma} < Z \leq \frac{b - \mu}{\sigma}\right) \\&= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).\end{aligned}$$

Standardisation

Les problèmes impliquant des calculs de probabilités pour la loi normale se résolvent donc en deux étapes :

- 1- standardiser la v.a. d'intérêt pour obtenir des probabilités impliquant la $\mathcal{N}(0, 1)$;
- 2- aller chercher les valeurs de $\Phi(\cdot)$ dans une table

Comme la densité $\mathcal{N}(0, 1)$ est symétrique autour de 0,

$$\Phi(-x) = 1 - \Phi(x) \text{ pour tout } x \in \mathbb{R}.$$

Les livres donnent donc généralement les valeurs de $\Phi(x)$ pour $x > 0$.

Exemple 1

La masse (en tonnes) des camions passant sur un viaduc peut être vue comme une v.a. normale de moyenne 50 et d'écart-type 10.

(i) Quelle est la probabilité qu'un camion pèse plus de 70 tonnes ?

(ii) Quelle proportion des camions pèse entre 40 et 60 tonnes ?

Exemple 1 (suite)

(iii) Quelle est la masse, disons m_1 , telle que 97,5% des camions ont une masse inférieure à m_1 ?

(iv) Quelle est la masse, disons m_2 , telle que 97,5% des camions ont une masse supérieure à m_2 ?

Loi normale

Une combinaison linéaire de variables aléatoires normales est aussi une variable aléatoire normale, peu importe si les v.a. sont indépendantes ou pas — à condition qu'elles soient conjointement normales (et pas seulement normales chacune de leur côté).

On peut calculer sa moyenne et sa variance avec les formules vues précédemment pour les combinaisons linéaires de variables aléatoires.

Par exemple, si $X_1, X_2, \dots, X_n$ sont des v.a. indépendantes et que $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ pour $i = 1, \dots, n$, alors

$$Y = X_1 + X_2 + \dots + X_n \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

et

$$W = a_1 X_1 + \dots + a_n X_n \sim \mathcal{N}\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

Exemple 2

Le portefeuille d'actions de votre firme est constitué à 20% d'actions de ABC, 30% d'actions de EFG et 50% d'actions de WXY. On vous dit que les rendements des trois titres sont des v.a. normales indépendantes de variance 0,09 et de moyenne 0,12 pour ABC, 0,15 pour EFG et 0,18 pour WXY. Quelle est la probabilité que le rendement global de ce portefeuille excède 0,17?

(Note : Le rendement global du portefeuille est la combinaison linéaire des rendements des actions individuelles dont les poids sont les proportions de chaque action dans le portefeuille.)

Exemple 2 (suite)

Un résultat très (très) utile

Théorème.

Soit $X_1, X_2, \dots, X_n$ des variables aléatoires **indépendantes** de moyenne $\mu < \infty$, de variance $\sigma^2 < \infty$, de loi de probabilité **quelconque**.

Alors si n est assez grand (généralement $n \geq 30$),

$$Y = X_1 + \dots + X_n \approx \mathcal{N}(n\mu, n\sigma^2)$$

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

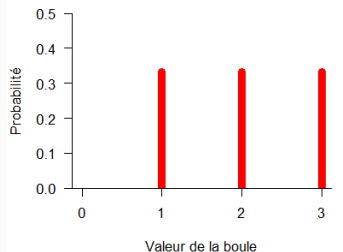
Nous appelons ce théorème, le *théorème central limite* (TCL).

Le seuil $n \geq 30$ est indicatif : la vitesse de convergence dépend en réalité de l'asymétrie de la distribution sous-jacente (une loi très asymétrique peut nécessiter un n plus grand).

1) Vous tirez une boule dans une urne contenant 3 boules numérotées : 1, 2, 3.

Vrai ou Faux ?

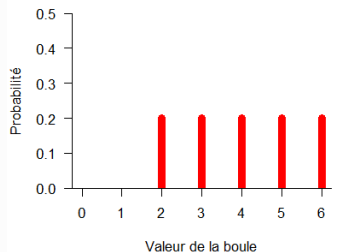
La valeur de la boule tirée a la distribution de probabilité suivante :



2) Vous remettez la boule dans l'urne et tirez une boule à nouveau.

Vrai ou Faux ?

La somme des deux numéros tirés a la distribution de probabilité suivante :



Exemple 3

Supposons que les demandes quotidiennes en électricité de 100 usines peuvent être vues comme des v.a. indépendantes de moyenne 10 kWh et d'écart-type 5 kWh.

Quelle est la probabilité approximative que la demande quotidienne totale en électricité de ces 100 usines excède la capacité de 1100 kWh du réseau qui les alimente ?

Exemple 3 (solution)

Application : approximation de la loi binomiale par la loi normale

Idée générale

Une v.a. binomiale peut être écrite comme une somme de v.a. indépendantes :

On peut donc approximer sa distribution par une loi normale, pourvu que n soit assez grand (par exemple, $np \geq 5$ et $n(1 - p) \geq 5$) :

Exemple

Un étudiant fait une erreur de calcul dans 10% de ses calculs. S'il doit répondre à un examen qui demande 100 calculs différents, quelle est la probabilité qu'il fasse plus de 5 erreurs de calcul ?

Exemple (suite)

Quelle est la probabilité que l'étudiant fasse exactement 5 erreurs ?

Correction pour la continuité

Puisque la loi binomiale est discrète mais que la loi normale est continue, on doit faire une correction pour la continuité.

Pour calculer $P(X = a)$ on calcule $P(a - 0,5 \leq X \leq a + 0,5)$:

De même pour les autres probabilités :

$$P(X \leq a) \approx$$

$$P(X < a) \approx$$

$$P(X \geq a) \approx$$

$$P(X > a) \approx$$

Exemple (suite) avec correction

Un étudiant fait une erreur de calcul dans 10% de ses calculs. S'il doit répondre à un examen qui demande 100 calculs différents, quelle est la probabilité qu'il fasse plus de 5 erreurs de calcul ?

Réponses des exemples

Réponses des exemples (1/3)

- **Exemple (loi uniforme, puissance du processeur) :** $P(X > 1,8) = \frac{2-1,8}{2-1,5} = 0,4$; $E(X) = 1,75$.
- **Exemple (loi exponentielle, bateaux) :** $P(X > 6) = e^{-2} \approx 0,1353$;
 $P(4 \text{ bateaux en } 15 \text{ h}) = P(\text{Poisson}(5) = 4) = e^{-5} 5^4 / 4! \approx 0,1755$.
- **Exemple (loi exponentielle, ampoule) :** par l'absence de mémoire,
 $P(\text{survie} \geq 1000 \text{ h de plus}) = P(X > 1000) = e^{-1} \approx 0,3679$.
- **Exemple (loi Gamma, hacker) :** $X \sim \text{Gamma}(3, 1/15)$; $E(X) = 45 \text{ minutes}$; $V(X) = 675 \text{ minutes}^2$.
- **Exemple (suite) (loi Gamma, hacker en 20 minutes) :** nombre de mots de passe craqués
 $\sim \text{Poisson}(4/3)$; $P(= 2) = e^{-4/3} (4/3)^2 / 2! \approx 0,2343$.

Réponses des exemples (2/3)

- **Exemple 1 (Weibull, panneaux de verre) :** $P(\text{dure} < 5 \text{ ans}) = F(5) = 1 - e^{-0,0625} \approx 0,0606$.
- **Exemple 1 (suite) (coût de la garantie) :** **coût total espéré** $= 500 \times 1000\$ \times 0,0606 \approx 30\,294\$$.
- **Exemple 2 (Weibull, iPod) :** **(i)** $P(X \geq 3 | \text{neuf}) = e^{-(3/4)^3} \approx 0,6558$; **(ii)**
 $P(X \geq 5 | X \geq 2) = e^{-(5/4)^3} / e^{-(2/4)^3} = e^{-1,828125} \approx 0,1607$.
- **Exemple 2 (suite) (médiane du iPod) :** m tel que $(m/4)^3 = \ln 2 \Rightarrow m \approx 3,54 \text{ ans}$.
- **Exemple 1 (loi normale, camions) :** **(i)** $P(X > 70) = 1 - \Phi(2) = 0,0228$; **(ii)**
 $P(40 < X < 60) = \Phi(1) - \Phi(-1) = 0,6826$.

Réponses des exemples (3/3)

- **Exemple 1 (suite) (percentiles des camions) :** (iii) $m_1 = 50 + 1,96 \times 10 = 69,6$; (iv) $m_2 = 50 - 1,96 \times 10 = 30,4$.
- **Exemple 2 (portefeuille) :** $E(W) = 0,159$, $V(W) = 0,0342$; $P(W > 0,17) = P(Z > 0,0595) \approx 0,476$.
- **Exemple 3 (TCL, demande des usines) :** $Y \approx \mathcal{N}(1000, 2500)$; $P(Y > 1100) = P(Z > 2) = 0,0228$.
- **Exemple (approximation binomiale, étudiant, sans correction) :**
 $P(X > 5) \approx P(Z > 1,667) = \Phi(1,667) \approx 0,9522$.
- **Exemple (suite) (exactement 5 erreurs, sans correction) :** $P(X = 5) \approx 0$ (ceci illustre le besoin d'une correction de continuité).
- **Exemple (suite) avec correction :** $P(X > 5) \approx P(Z \geq 1,5) = \Phi(1,5) \approx 0,9332$.

Bibliographie

Ce document est rédigé avec Beamer ainsi qu'une classe .cls fournie par Jérôme Soucy. Les diapositives sont adaptées des diapositives créées par Anne-Sophie Charest pour le cours STT-1000 ainsi que celles créées par Thierry Duchesne et Emmanuelle Reny-Nolin pour le cours STT-1900. Les graphiques proviennent de ces mêmes diapositives.

Pour signaler une erreur dans ce document, veuillez écrire à l'adresse : julien.miron@mat.ulaval.ca.