

Module 3 - Loi normale et théorème central limite

Département de mathématiques et de statistique



UNIVERSITÉ
LAVAL

Description de la loi normale

Loi importante

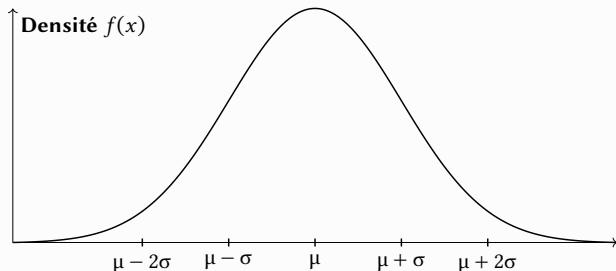
La loi normale est la loi de probabilité la plus utilisée dans les applications de la statistique :

- elle modélise bien la distribution de plusieurs variables;
- elle approxime la loi de variables représentant une somme ou une moyenne de plusieurs mesures (théorème central limite);
- elle mène à des résultats analytiques dans plusieurs problèmes d'inférence statistique

Densité de la loi normale

On dit que X suit une loi normale d'espérance μ et de variance σ^2 , dénotée $X \sim \mathcal{N}(\mu, \sigma^2)$, lorsque sa densité est donnée par

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



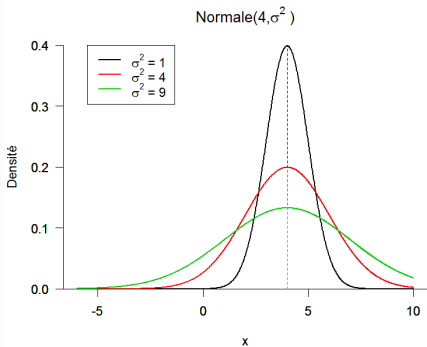
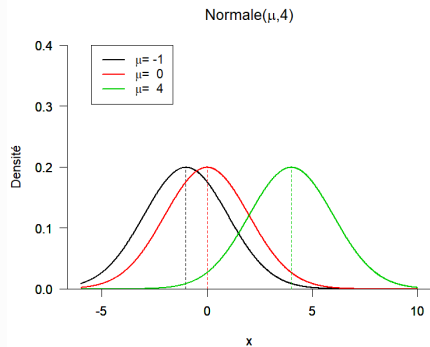
Propriétés les plus importantes de la densité

- $f(x) \geq 0$ pour tout $x \in \mathbb{R}$ et $\int_{-\infty}^{\infty} f(x) dx = 1$
- $\lim_{x \rightarrow -\infty} f(x) = 0$ et $\lim_{x \rightarrow \infty} f(x) = 0$
- $f(\mu - x) = f(\mu + x)$ pour tout $x \in \mathbb{R}$ (symétrique autour de μ)
- $f(x)$ atteint son unique maximum en $x = \mu$
- $f(x)$ a deux points d'inflexion : en $\mu - \sigma$ et en $\mu + \sigma$

Les deux paramètres de la loi $\mathcal{N}(\mu, \sigma^2)$

Pour calculer des probabilités à partir de ce modèle, il faut donner des valeurs à deux paramètres :

- **Espérance** : $E(X) = \mu$.
- **Variance** : $V(X) = E[(X - \mu)^2] = \sigma^2$.



QUIZ!

La longueur X des lattes de bois coupées par une machine est distribuée selon une loi normale de paramètres $\mu = 2400$ mm et $\sigma^2 = 4$ mm². Laquelle des deux probabilités est la plus grande?

1. $P(X < 2396)$ ou $P(X > 2404)$?

2. $P(2396 < X < 2397)$ ou $P(2401 < X < 2402)$?

Loi normale centrée-réduite

- **Loi normale standard (ou loi normale centrée-réduite) :**

Lorsque $\mu = 0$ et $\sigma^2 = 1$: $Z \sim \mathcal{N}(0, 1)$

- **Fonction de répartition de la $\mathcal{N}(0, 1)$:**

$$F(z) = \Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du, \quad -\infty < z < \infty.$$

Il n'y a pas de solution analytique à cette intégrale.

- **Table des valeurs de $\Phi(z)$:**

Disponible dans les livres de probabilités et de statistique, et dans des logiciels tels Excel, Matlab, Maple, R, etc.

- Toutes les variables aléatoires suivant une loi normale $\mathcal{N}(\mu, \sigma^2)$ peuvent se transformer en normales centrées-réduites par standardisation.

Standardisation de $\mathcal{N}(\mu, \sigma^2)$ en $\mathcal{N}(0, 1)$

- **Standardisation** : Pour utiliser la table de $\Phi(z)$, on doit transformer la variable X qui suit une $\mathcal{N}(\mu, \sigma^2)$:

Il faut "centrer" X , i.e. ramener sa moyenne à 0 :

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \Rightarrow \quad X - \mu \sim \mathcal{N}(0, \sigma^2)$$

Il faut "réduire" $X - \mu$, i.e. amener l'écart-type à 1 :

$$X - \mu \sim \mathcal{N}(0, \sigma^2) \quad \Rightarrow \quad \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Standardisation : justification mathématique

- Soit $X \sim \mathcal{N}(\mu, \sigma^2)$. On a que

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

Standardisation : justification mathématique

- Soit $X \sim \mathcal{N}(\mu, \sigma^2)$. On a que

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

- En résumé : 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 $\mathcal{N}(0, 1)$ est requise pour calculer des probabilités impliquant la loi normale, peu importe la valeur des paramètres!

Standardisation

Petit truc qui sera utile plus tard (module 6 et suite) :

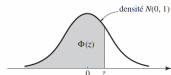
- On sait que si $X \sim \mathcal{N}(\mu, \sigma^2)$, $E(X) = \mu$ et $V(X) = \sigma^2$;
- On peut donc réécrire $X \sim \mathcal{N}(E(X), V(X))$;
- La standardisation est donc $Z = \frac{X - E(X)}{\sqrt{V(X)}}$.

Fonction de répartition de la loi $\mathcal{N}(0, 1)$

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

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

	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
2,2	0,986 10	0,986 45	0,986 79	0,987 13	0,987 45	0,987 78	0,988 09	0,988 40	0,988 70	0,988 99



Exemple 1

Environnement Canada mesure la vitesse du vent (en km/h) à chaque heure à la station météo de l'aéroport Jean-Lesage. Supposons que la vitesse horaire en janvier peut être vue comme une v.a. normale de moyenne 17 et d'écart-type 8.

Exemple 1 (suite)

- (i) Quelle est la probabilité qu'une mesure de la vitesse horaire dépasse 27 km/h ?

Exemple 1 (suite)

(ii) Quelle proportion des mesures sont comprises entre 5 et 25 km/h ?

Exemple 1 (suite)

- (iii) Quelle est la vitesse, disons v_1 , telle que 97,5% des mesures horaires représentent une vitesse inférieure à v_1 ?

QUIZ!

Quelle est la vitesse v_2 telle que 80% des mesures horaires sont supérieures à v_2 ?

Additivité de la loi normale

Loi d'une combinaison linéaire de v.a. normales

Supposons que $X_1, X_2, \dots, X_n$ sont des v.a. **indépendantes** et que $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, etc.

- On s'intéresse à la somme des X_i .

$$X_1 + X_2 + \dots + X_n \sim \mathcal{N}(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2)$$

- On s'intéresse à une combinaison linéaire des X_i .

$$a_1 X_1 + \dots + a_n X_n \sim \mathcal{N}(a_1 \mu_1 + \dots + a_n \mu_n, a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2)$$

Si les X_i sont conjointement normales mais pas indépendantes, leurs combinaisons linéaires suivent quand même des lois normales et avec les moyennes données ci-dessus, mais les **formules de variances doivent être ajustées afin de tenir compte des covariances** (mesures de dépendance).

Exemple 2

Le portefeuille d'actions de votre firme est constitué à

- 20% d'actions de ABC (rendement moyen 0,12)
- 30% d'actions de EFG (rendement moyen 0,15)
- 50% d'actions de WXY (rendement moyen 0,15)

On vous dit que les rendements des trois titres sont des v.a. normales indépendantes de variance 0,09.

Quelle est la probabilité que le rendement global de ce portefeuille excède 0,17 ?

Exemple 2 (solution)

Exemple 3

Le salaire des ingénieurs dans la province A suit une loi $\mathcal{N}(70, 100)$, alors que dans la province B il est distribué selon une loi $\mathcal{N}(80, 64)$.

Si on tire au hasard un ingénieur de chacune des provinces, quelle est la probabilité que le salaire de l'ingénieur de la province B excède celui de l'ingénieur de la province A?

Exemple 3 (solution)

Théorème central limite

exemple

exemple introductif

Un résultat très (très) utile

Théorème central limite

Soit $X_1, X_2, \dots, X_n$ des v.a. indépendantes
d'espérance $\mu < \infty$,
de variance $\sigma^2 < \infty$,
d'une même loi de probabilité (quelconque).

Alors si $n \geq 30$,

$$Y = X_1 + X_2 + \dots + X_n \approx \mathcal{N}(n\mu, n\sigma^2)$$

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Exemple 4

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 4 (solution)

Réponses aux exemples

1. $P(X > 27) = 0,10565$
 $P(5 < X < 25) = 0,77453$
 $v_1 = 32,68 \text{ km/h}$
2. $P(R > 0,17) = 0,44433$
3. $P(X_B > X_A) = 0,78230$
4. $P(Y > 1100) = 0,02275$