

Loi Normale

Greg

24 mai 2007

Énoncé

1. Donnez la fonction de densité caractérisant une variable aléatoire X qui suit une loi normale $N(0; 1)$.
2. Calculez sa moyenne. Sans calcul, donnez la valeur de la variance et de l'écart-type.
3. Soit Φ la fonction de répartition de la loi $N(0; 1)$. Montrez que $\Phi(-u) = 1 - \Phi(u)$.
4. Expliquez comment on calcule $P(X \geq 1)$ et interprétez le résultat sur le graphique de la densité de probabilité de X .
5. Que faire avec une variable aléatoire qui suit une loi normale $N(\mu, \sigma)$?

Démonstration

Point 1

Tout d'abord faisons un rappel sur la notion de fonction de densité. Nous avons vu jusqu'à présent des variables aléatoires discrètes (comptage, ...). Cependant beaucoup de phénomènes sont associés à des variables aléatoires continues (poids d'une personne, température, durée, ...). Prenons l'exemple du poids d'une personne. Quelle est la probabilité de peser 100 kg ? La réponse est 0 ! En effet, étant donné que la variable aléatoire peut prendre *toutes* les valeurs comprises entre 0 et 200 kg (il y a une *infinité* de valeurs dans cet intervalle), la probabilité pour qu'elle en prenne une en particulier est nulle. Nous verrons en détails pourquoi.

Si on ne peut pas calculer la probabilité que la variable aléatoire continue prenne une valeur particulière, on peut calculer la probabilité qu'elle prenne une valeur comprise entre deux bornes. Par exemple, quelle est la probabilité que le poids de la personne soit compris entre 95 et 100 kg ? Pour se faire nous avons besoin d'un nouvel outil : la fonction de densité.

Cette fonction se définit de la manière suivante. Soit X une variable aléatoire continue et $a, b \in \mathbb{R}$. On appelle f la fonction de densité (de probabilité) de X qui satisfait :

1. $\forall x \in \mathbb{R}, \text{ on a } f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$
3. $P(a < X \leq b) = \int_a^b f(x) dx.$

La Figure 1 montre bien que plus on augmente le nombre de mesure (n), plus il y a de tranches dans le graphique qui finalement tend vers une courbe. On remarque aussi que lorsque n tend vers l'infini alors la notion de densité de fréquences relatives se transforme en probabilités. Dans les graphiques (a) et (b), la *surface* de chaque rectangle représente la fréquence relative de la valeur de X correspondante. Par exemple, sur le graphique (a), il y a $0.04 \cdot 5 = 20\%$ des personnes (parmi les 20) qui pèsent entre 60 et 65 kg et 37.5% ($20+17.5$) qui pèsent entre 60 et 70 kg. Si on passe au graphique (b), on remarque qu'il faut additionner l'aire de 3 rectangles pour connaître la proportion de personne qui pèse entre 60 et 70 kg. Plus n augmente et plus l'aire des rectangles tend vers l'aire sous la courbe. Ainsi la probabilité que X soit comprise entre deux valeurs a et b est égale à l'aire sous la courbe entre a et b (une *somme infinie* de rectangle) et donc à l'intégrale :

$$P(a < X \leq b) = \int_a^b f(x) dx$$

Avec cette définition on comprend pourquoi la probabilité qu'une variable aléatoire continue prenne une valeur exacte est nulle. Tout simplement car on aurait : $P(X = a) = \int_a^a f(x) dx = 0$. C'est l'aire d'un rectangle de base nulle et de hauteur $f(a)$, d'où une aire nulle.

La fonction de densité pour une variable aléatoire X qui suit la loi normale centrée réduite vaut :

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

Plus généralement, si X suit une loi normale (mais pas centrée ni réduite), alors la fonction de densité vaut :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

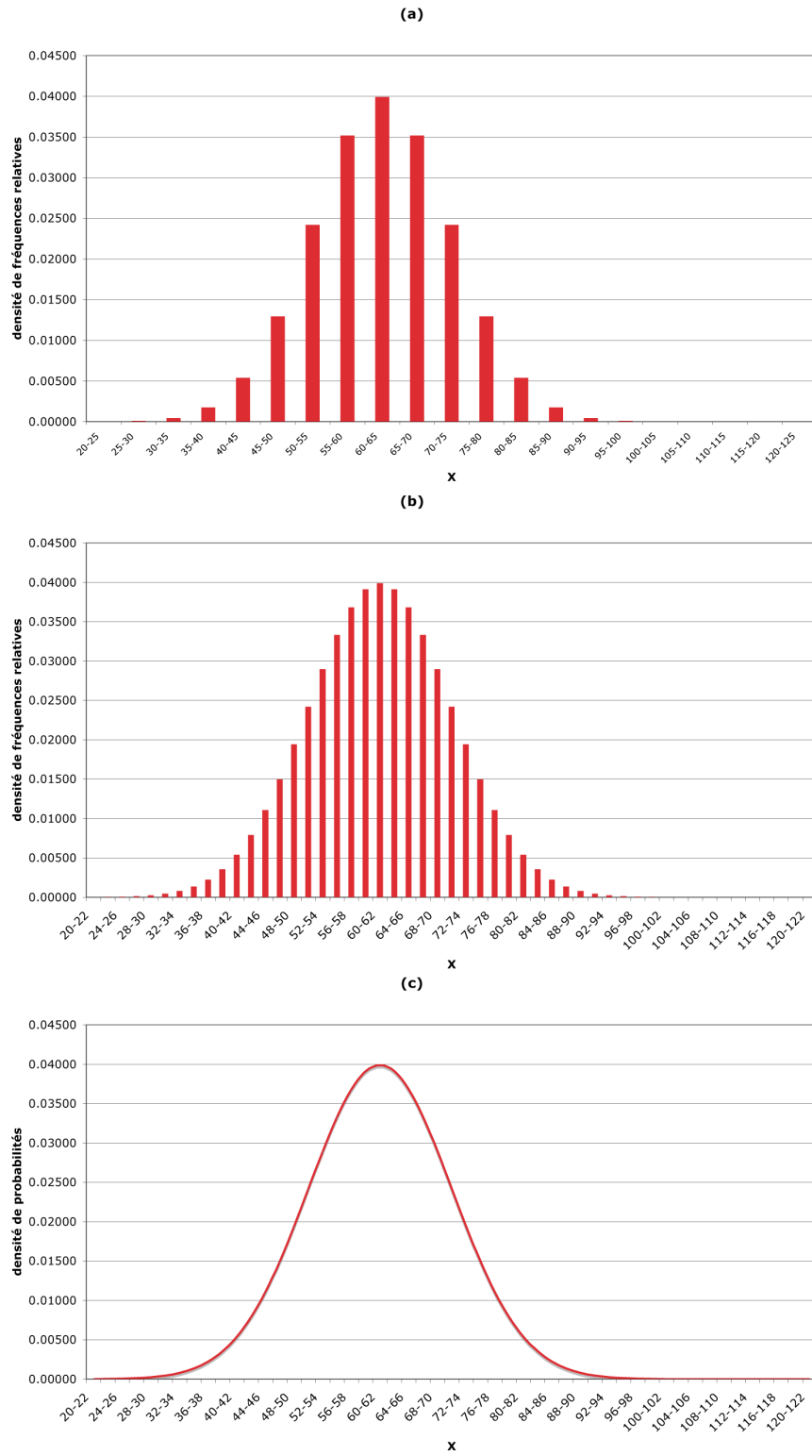


FIG. 1 – (a) – Graphique de densité de fréquences relatives du poids d'un homme avec $n = 20$ (la variable X) peut prendre 20 valeurs. (b) – idem avec $n = 50$. (c) – idem avec $n = \infty$.

Point 2

Dans le cas des variables aléatoires discrètes, nous avons :

$$E(X) = \sum_{i=1}^n x_i \cdot P(X = x_i)$$

Comme nous sommes dans le cas continue, il faut simplement remplacer la somme par une intégrale et on a :

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Dans le cas de la loi normale centrée ($\mu = 0$) et réduite ($\sigma = 1$), il suffit d'y insérer la fonction de densité et on a :

$$E(X) = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx$$

Séparons l'intégrale en deux :

$$\int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx = \int_{-\infty}^0 \frac{x}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx + \int_0^{\infty} \frac{x}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx$$

Comme la fonction est impaire (voir Figure 2), alors l'intégrale est nulle et la moyenne est donc nulle.

Point 3

On sait que :

$$\Phi(x) = \int_{-\infty}^x f(x) dx$$

On sait aussi que l'aire totale sous la courbe $f(x)$ vaut 1. $\Phi(u)$ représente l'aire sous la courbe entre $-\infty$ et u . La courbe $f(x)$ est symétrique ($\Phi(0) = 0.5$), ce qui veut dire que l'aire sous la courbe entre $-\infty$ et $-u$ est la même que celle entre u et ∞ . Ainsi, comme le montre la Figure 3 :

$$\Phi(-u) = 1 - \Phi(u)$$

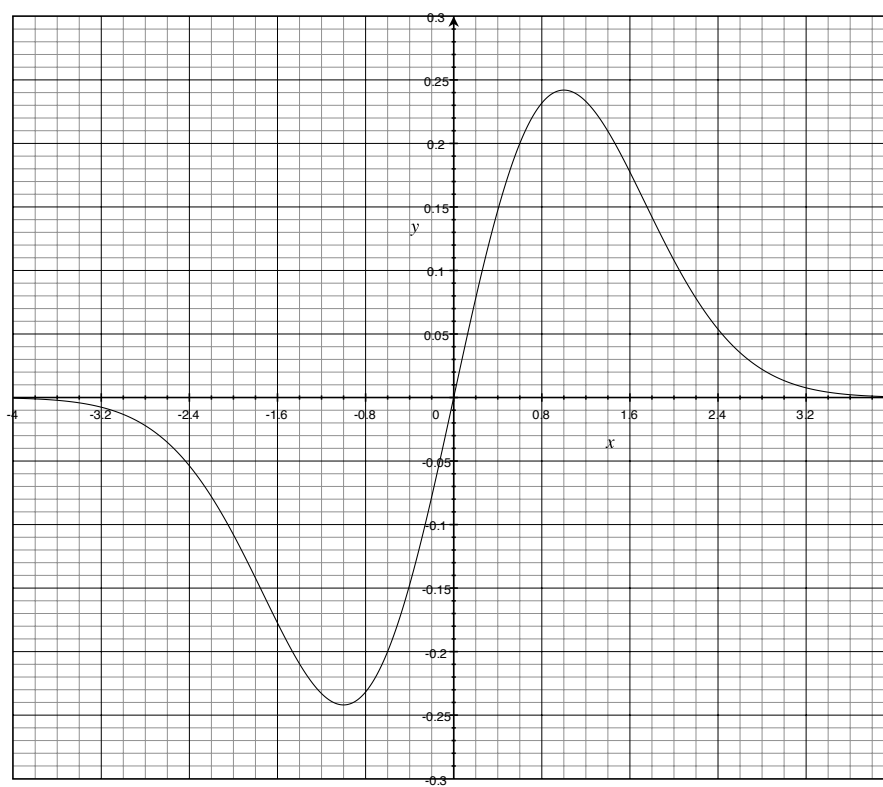


FIG. 2 – Graphique de $f(x) = \frac{x}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$. La fonction est bien impaire.

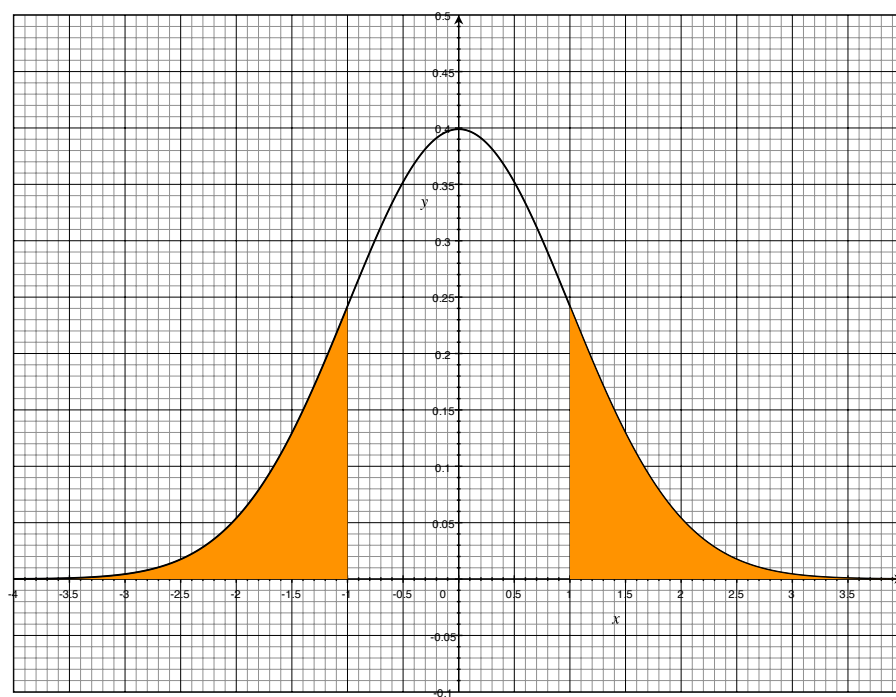


FIG. 3 – En orange, à gauche est représenté l'aire correspondant à $\Phi(-1)$ et à droite en orange l'aire correspondant à $1 - \Phi(1)$.

Point 4

On sait que

$$P(X \leq a) = \Phi(a)$$

La probabilité que X soit inférieure ou égale à a est égale à l'aire sous la courbe entre $-\infty$ et a , qui vaut $\Phi(a)$ comme nous l'avons vu plus haut. La probabilité que X soit supérieure strictement à a est alors l'aire entre a et ∞ . Cette aire se calcule par $1 - \Phi(a)$ comme nous l'avons aussi vu plus haut et donc :

$$P(X > a) = 1 - \Phi(a)$$

La Figure 4 illustre la situation.

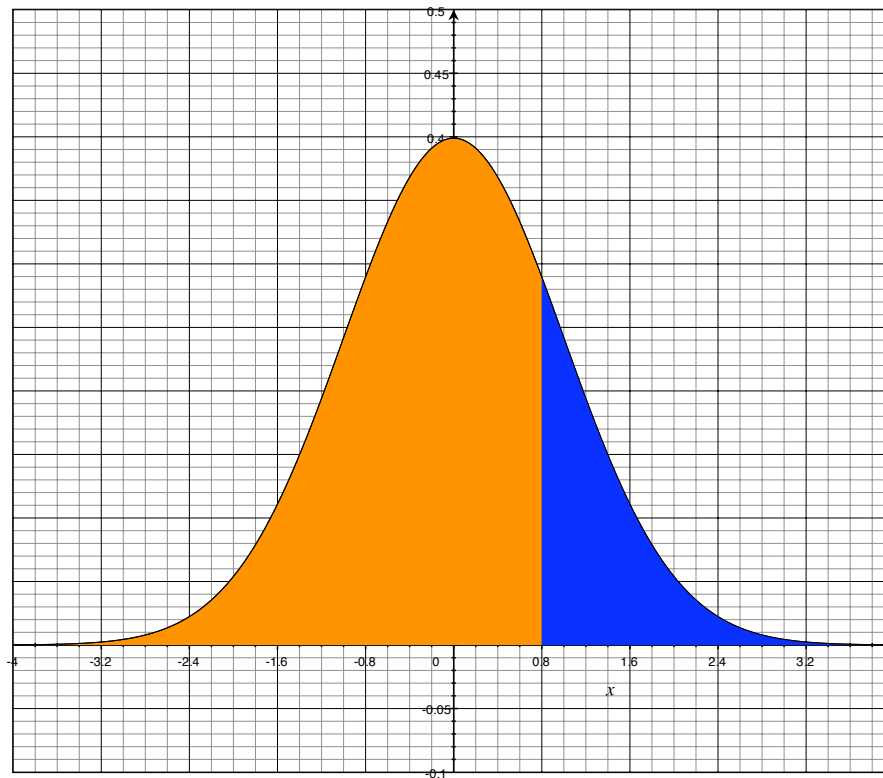


FIG. 4 – Soit $a = 0.8$. L'aire orange vaut $\Phi(a)$. L'aire totale sous la courbe vaut 1. L'aire bleue vaut alors $1 - \Phi(a)$.

Point 5

Supposons que X soit une variable aléatoire qui représente la température à la mi-journée en Suisse au mois de février. X suit une loi normale $N(0; 1)$ (pour simplifier on suppose que la température moyenne (μ) d'une journée de février soit de zéro degré). Quelle est la probabilité pour qu'il fasse plus de 1.5 degré aujourd'hui ?

$P(X > 1.5) = 1 - \Phi(1.5)$. On peut aller lire la valeur de $\Phi(1.5)$ dans la table numérique et on trouve 0.93319. Finalement :

$$P(X > 1.5) = 1 - 0.93319 = 0.06681 = 6\%$$

Point 6

Regardons la Figure 5. On remarque que la courbe rouge est exactement la même que celle de la Figure 3, mais décalée sur la droite (au lieu d'être centrée en 0 elle est centrée en 5). Pour faciliter les calculs, il faudrait pouvoir centrer la courbe rouge sur zéro. On va donc soustraire la valeur de la moyenne (5) à X . En faisant cela on crée une nouvelle variable aléatoire Z qui vaut :

$$Z = X - \mu$$

Cela permet de remener une loi normale non centrée (mais réduite : $\sigma = 1$) à une loi normale centrée réduite. Il reste à voir le cas où la loi est non-centrée en non-réduite. C'est l'exemple de la courbe bleue. Il faut faire comme précédemment : soustraire la moyenne à X pour la centrer, mais il faut aussi l'allonger (la réduire) ! Pour se faire, nous allons encore tout diviser par σ . On a alors une nouvelle variable aléatoire Z qui permet de centrer et de réduire n'importe quelle loi normale $N(\mu; \sigma)$:

$$Z = \frac{X - \mu}{\sigma}$$

Z suit alors une loi normale $N(0; 1)$.

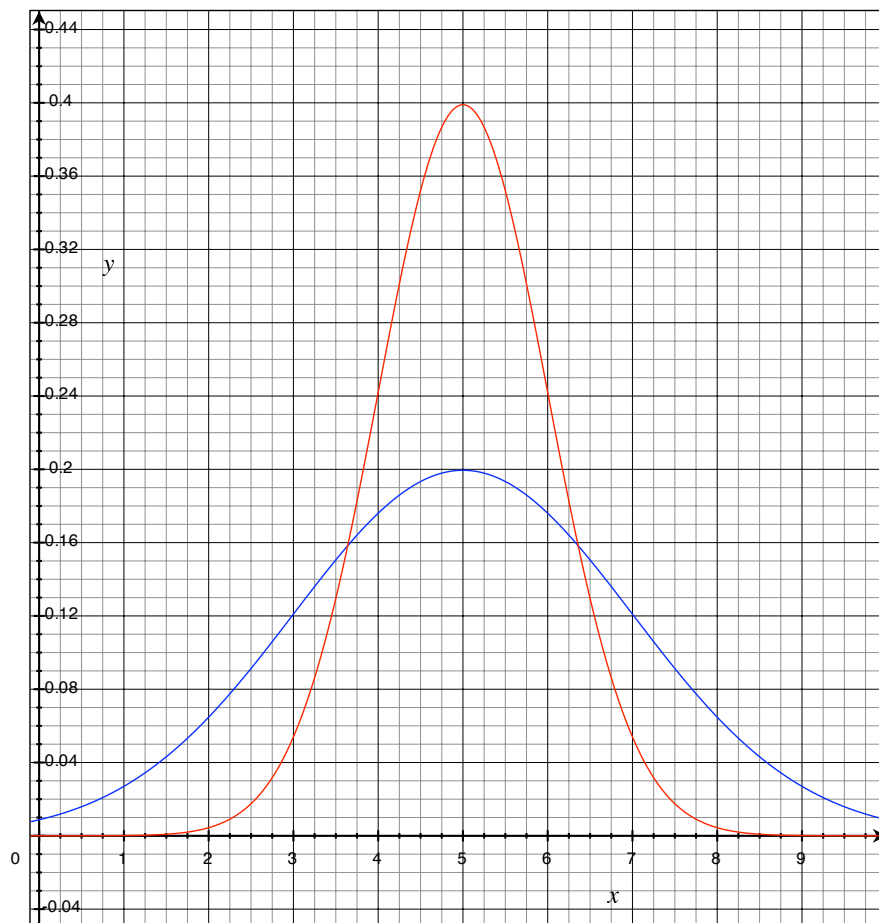


FIG. 5 – La courbe rouge représente la fonction de densité d'une variable aléatoire X qui suit une loi normale $N(5; 1)$. On remarque qu'elle est centrée en 5, sur sa moyenne. La courbe bleue représente la fonction de densité d'une variable aléatoire X qui suit une loi normale $N(5; 2)$. Elle est aussi centrée en 5, mais est plus aplatie. Plus σ augmente, plus la courbe s'aplatit. L'aire sous les deux courbes est identique et vaut 1.