TD4 - Variables aléatoires continues et loi normale

Objectif L'objectif de ce TD est de manipuler des **variables aléatoires continues**, prenant des valeurs réelles. Ce type de variable sert à modéliser des mesures comme la taille, la température, la durée, la vitesse, etc. En particulier, nous étudierons la **loi normale**, qui décrit de nombreux phénomènes naturels et représente une notion centrale en statistiques.

Rappel : La variable aléatoire discrète D représentant un dé à six faces non biaisé suit une loi **uniforme discrète** :
- Le nombre total de valeurs possibles S_D est **fini** : il y en a 6.
- Le dé est non biaisé, les valeurs de D sont toutes **équiprobables** : $P(D=1) = \ldots = P(D=6)$.
- La probabilité de chaque face vaut 1 sur le nombre total de possibilités, soit $P(D=i) = 1/6$ pour tout $1 \leq i \leq 6$.
- Plus généralement, $P(X=i) = 1/n$ pour une variable uniforme discrète X ayant $|S_X| = n$ valeurs possibles.

Exercice 4.1: Variables aléatoires continues

Nous définissions une variable aléatoire T représentant la température de l'eau à l'état liquide, pouvant varier entre 0 et 100 degrés Celcius (°C). Le thermomètre utilisé est très précis : une température peut être une **valeur réelle** quelconque.

1. Combien y a t-il de valeurs réelles différentes dans l'intervalle $[0, 100]$?
2. Par conséquent, combien de valeurs différentes peut prendre T ? Cet ensemble est-il fini ou infini ?
3. Si les températures sont équiprobables, leurs probabilités devraient valoir 1 sur le nombre total de possibilités (comme pour D). Dans ce cas, combien vaudrait $P(T = 3,592)$, la probabilité que la température soit égale à exactement 3,592°C ?
4. Pus généralement, quelle serait la valeur de $P(T = t)$ pour n'importe quelle valeur de $t \in [0, 100]$?

Ce résultat contre-intuitif montre que, pour les variables aléatoires continues, on ne peut pas faire comme pour les variables discrètes : la probabilité qu'une variable continue soit *égale* à une valeur précise n'a pas de sens, et **vaut toujours zéro** ! Cela ne nous empêche pas de calculer des probabilités pour des intervalles, p.ex. $P(5 \leq T \leq 10)$. Mais comment faire ?

Exercice 4.2: Loi uniforme continue (température Celsius)

Une variable aléatoire **uniforme continue** prend des valeurs dans l'intervalle $[\alpha, \beta]$, entre la **borne inférieure** α et la **borne supérieure** β. Si deux sous-intervalles de $[\alpha, \beta]$ sont de même taille, alors ils ont la même probabilité.

1. Si T représente la température de l'eau à l'état liquide, quelles sont les valeurs des bornes α et β ?
2. Quelle est la probabilité que la température de l'eau à l'état liquide soit entre les bornes α et β, $P(\alpha \leq T \leq \beta)$?
3. Quelle est la probabilité que la température de l'eau à l'état liquide soit entre 0 et 50°C, $P(0 \leq T \leq 50)$.
4. Quelle est la probabilité que la température de l'eau à l'état liquide <u>ne</u> soit <u>pas</u> entre 70°C et 85°C, $P(T \notin [70, 85])$.

La probabilité qu'une variable **uniforme continue** prenne des valeurs dans l'intervalle $[a, b]$ est égale à la *proportion* que cet intervalle représente par rapport à l'intervalle total de valeurs possibles $[\alpha, \beta]$.

Exercice 4.3: Loi uniforme continue (variable quelconque)

On représenté la loi uniforme continue par un **rectangle** de largeur $\beta - \alpha$ et hauteur h. Les probabilités sont données par l'**aire** sous ce rectangle. La probabilité d'un intervalle $[a, b]$ est égale à l'aire de la "tranche" du rectangle, soit $(b - a) \times h$.

1. Dessiner le rectangle pour la variable T (température de l'eau). Exprimez l'aire du rectangle en fonction de h, α et β.
2. La somme des probabilités pour toutes les valeurs possibles vaut 1 pour une variable aléatoire discrète. Dans le cas continu, c'est <u>l'aire du rectangle</u> qui doit valoir 1. Trouvez la valeur de h telle que l'aire vaut 1.
3. Représenter dans ce rectangle les tranches correspondant aux questions 3 et 4 de l'exercice 4.2.
4. La variable aléatoire F désigne la température de l'eau à l'état liquide en <u>degrés Fahrenheit</u> (°F), allant de 32°F à 212°F. Dessinez le rectangle de la loi de F en fonction des bornes α, β, et de la <u>hauteur h</u> (calculée comme pour T).
5. L'espérance d'une loi uniforme est la valeur découpant le rectangle en deux moitiés de même aire. Calculez $E[T]$ et $E[F]$.
6. Généralisons à une variable uniforme continue X quelconque. Trouvez des formules qui expriment h, $P(a \leq X \leq b)$ et $E[X]$ en fonction de a, b, α et β. Dessinez le rectangle $[\alpha, \beta] \times h$, la tranche $[a, b]$, et la ligne verticale $E[X]$ pile au milieu.

Formellement, on définit une fonction appelée **fonction de densité** de probabilité, notée $f(x)$. Pour la loi uniforme continue, $f(x) = \begin{cases} h & \text{si } \alpha \leq x \leq \beta \\ 0 & \text{sinon} \end{cases}$. $P(a \leq X \leq b)$ correspond à l'aire sous $f(x)$ entre a et b, c'est-à-dire, un rectangle. Notez que ce n'est pas la valeur de $f(x)$ qui permet de calculer des probabilités, mais **l'aire sous la courbe** $f(x)$ dans un intervalle.

Exercice 4.4: Fonction de densité, fonction de répartition et espérance

Une fonction de densité de probabilité $f(x)$ doit respecter deux contraintes : (1) la fonction doit être positive, soit $f(x) \geq 0 \ \forall x$; et (2) l'aire sous $f(x)$ doit valoir 1, sinon on ne peut pas la voir comme une probabilité.

1. Comment calculer l'aire sous la courbe quand $f(x)$ n'est pas une constante (rectangle), mais une courbe quelconque ?
2. Exprimez la contrainte (2) formellement avec une jolie formule mathématique, selon votre réponse à la question 1.
3. On définit $f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{sinon} \end{cases}$. Dessinez le graphique de $f(x)$, puis démontrez que $f(x)$ respecte la contrainte (2).
4. Calculez $P(0 \leq X \leq \frac{2}{3})$ pour la fonction de densité $f(x)$ de la question 3 et indiquez l'aire sous la courbe correspondante.
5. Généralisez ce résultat : exprimez la probabilité $P(X \in [a, b])$ pour l'intervalle $[a, b]$ en fonction de l'aire sous $f(x)$.

Comme pour les variables discrètes, on définit la **fonction de répartition** $F_X(a) = P(X \leq a)$. Mais au lieu d'additionner les probabilités des valeurs $\leq a$, on calcule l'aire sous la courbe $f(x)$ à gauche de a, c'est-à-dire $F_X(a) = \int_{-\infty}^{a} f(x)\, dx$.

6. Trouvez la fonction de répartition $F_X(a)$ pour la fonction de densité $f(x)$ de la question 3, puis calculez $F_X(1/2)$.

De même, pour calculer l'espérance d'une variable continue, l'intégrale "remplace" la somme : $E[X] = \int_{-\infty}^{+\infty} x f(x)\, dx$

7. Calculez l'espérance pour la fonction de densité $f(x)$ de la question 3.

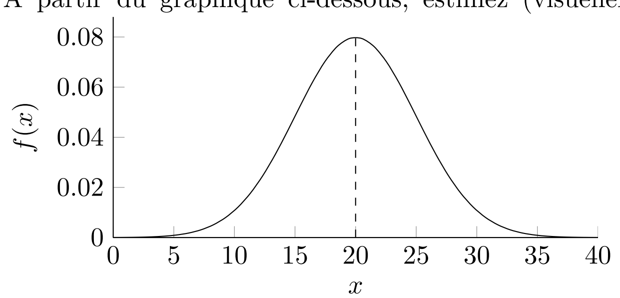
Exercice 4.5: Loi "binomiale continue"

La loi uniforme continue est une généralisation de la loi uniforme discrète. Une autre loi discrète que nous avons étudié est la **loi binomiale**, qui donne la probabilité d'obtenir k succès en faisant n essais indépendants avec probabilité de succès p.

1. À l'aide d'un simulateur en ligne, observez la loi binomiale en variant n de 2 à 100, avec une valeur $p = 1/2$ fixe : https://www.stat.berkeley.edu/~stark/Java/Html/BinHist.htm
2. Dessinez une courbe continue qui épouserait la forme de la loi binomiale quand $n \to +\infty$.

Cette courbe en forme de cloche est la fonction de densité $f(x)$ d'une variable X suivant une **loi normale** (aussi appelée Gaussienne). Sa formule est $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, définie sur l'intervalle infini $\mathbb{R}$. Étonnement, l'aire sous cette courbe est finie, et vaut 1 ! Comme la loi binomiale, la loi normale a deux paramètres : μ et σ^2, et on note $X \sim N(\mu, \sigma^2)$.

3. Il se trouve que pour $X \sim N(\mu, \sigma^2)$, le paramètre μ correspond à l'espérance $E[X]$ et σ^2 correspond à $Var(X)$. Représentez μ sur votre dessin comme une ligne verticale qui divise la cloche en 2 moitiés symétriques (valeur centrale).
4. Ajoutez au dessin des lignes verticales $\mu - i\sigma$ et $\mu + i\sigma$ avec $i = 1, 2, 3$. Pour connaître leurs emplacements approximatifs, sachez que $P(\mu - 1\sigma \leq X \leq \mu + 1\sigma) = 0,683$, $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0,955$, et $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0,997$.
5. À partir du graphique ci-dessous, estimez (visuellement) les paramètres μ et σ^2 de cette la courbe normale $f(x)$.



6. Dessinez la courbe de $f(x)$ pour une variable suivant une loi normale avec $\mu = 7$ et $\sigma^2 = 4$ (et donc $\sigma = 2$).
7. Calculez $E[Y] = np$ et $Var(Y) = np(1-p)$ pour la variable Y suivant une loi binomiale avec $n = 100$, $p = \frac{1}{2}$ (cf. TD 3).
8. Quand n est grand, calculer des probabilités avec la loi binomiale est fastidieux. Par exemple, si $Y \sim B(100, \frac{1}{2})$, le calcul de $P(Y \leq 55)$ est une somme de 56 termes, de 0 à 55 ! Pour n suffisamment grand, la normale donne une bonne approximation. Si on suppose $P(Y \leq 55) \approx P(X \leq 55)$, quels sont les paramètres μ et σ^2 de la variable normale X ?

Exercice 4.6: Loi normale centrée réduite

Nous savons à quoi ressemble la fonction de densité $f(x)$ de la loi normale selon ses paramètres μ et σ^2. Cependant, cela ne nous dit pas comment calculer des probabilités. Calculer des intégrales avec cette fonction paraît compliqué, non ? Bonne nouvelle : on pourra s'en passer, car on utilisera une astuce pour calculer des probabilités avec la loi normale. [1]

1. Vous aurez besoin de résultats du TD3, exercices 3.4 et 3.6.

1. $X \sim N(175, 49)$ représente la taille d'un.e adulte en France. Dessinez sa courbe normale symétrique autour de $E[X]$.
2. On propose la transformation $Z = a(X - b)$. Exprimez l'espérance $E[Z]$ en fonction de $E[X]$ et des constantes a et b.
3. Si on suppose $a = 1$ et $b = \mu$, combien vaut l'espérance $E[Z]$ de la nouvelle variable Z ?
4. Dessinez la courbe normale de la taille des adultes en France modifiée $Z = X - \mu$, avec $\mu = 175$ et $\sigma^2 = 49$.

La transformation $Z = X - \mu$ translate la courbe de X horizontalement, on dit que Z est **centrée** autour de zéro.

5. On s'intéresse désormais à la variance $Var(X)$ pour $X \sim N(\mu, \sigma^2)$. Exprimez $Var(Z)$ en fonction de a, b et σ^2.
6. Si on suppose $a = \frac{1}{\sigma}$ et $b = \mu$, combien vaut la variance $Var(Z)$?
7. Dessinez la courbe normale de la taille des adultes en France modifiée $Z = \frac{X-\mu}{\sigma}$, avec $\mu = 175$ et $\sigma^2 = 49$.

On dit que $Z = \frac{X-\mu}{\sigma}$ suit une loi normale centrée **réduite**, car sa variabilité (largeur) est limitée par une constante.

8. Par conséquent, si X suit une loi normale quelconque $N(175, 49)$, quelle loi suit la variable centrée réduite Z ?

On peut convertir n'importe quelle variable normale X en la variable **normale centrée réduite** Z, dont les paramètres $\mu = 0$ et $\sigma^2 = 1$ sont des constantes. Les probabilités pour Z sont pré calculées dans un tableau. Voyons comment s'en servir.

Exercice 4.7: Loi normale : calcul de probabilités

Le tableau en page 4 donne des valeurs pré-calculées de la fonction de répartition de Z, appelée $\Phi(a)$. Il s'agit des probabilités $P(Z \leq a)$ pour 400 valeurs positives de a échantillonnées entre 0 (tout en haut à gauche) et 3,99 (tout en bas à droite). Les indices de ligne et de colonne indiquent le seuil a, par exemple ligne 1.6 colonne 0.05 contient $\Phi(1,6 + 0,05) = P(Z \leq 1,65) = 0,9505$.

1. À partir du tableau, obtenez la probabilité $P(Z \leq 1)$. Dessinez la courbe et indiquez l'aire correspondant à $P(Z \leq 1)$.
2. À partir du tableau et de la symétrie de la courbe autour de zéro, déduisez les probabilités $P(Z \leq -1)$ et $P(Z > 1)$.

On cherche désormais à calculer des probabilités sur X, la taille des adultes en France. Cependant, on n'a pas de table pré-calculée pour $X \sim N(175, 49)$. Il faut donc **centrer et réduire** X pour connaître les intervalles correspondants de Z.

3. On veut connaître la probabilité qu'un adulte mesure 189cm ou moins, soit $P(X \leq 189)$. Appliquez la transformation $\frac{X-\mu}{\sigma}$ des deux côtés de l'inégalité : X devient Z et le seuil 189 devient une nouvelle valeur seuil a.
4. À partir de la valeur a trouvée ci-dessus, obtenez depuis le tableau $\Phi(a)$ la probabilité $P(Z \leq a) = P(X \leq 189)$.
5. Utilisez le même raisonnement pour obtenir la probabilité $P(X \leq 157,5)$ qu'un adulte soit très petit.
6. Utilisez le même raisonnement pour obtenir la probabilité $P(X > 196)$ qu'un adulte soit très grand.

Exercice 4.8: Loi normale : intervalles

Nous savons désormais calculer des probabilités que Z soit inférieur ou supérieur à un seuil a. Mais comment calculer la probabilité $P(b \leq Z \leq a)$ pour un intervalle $[a, b]$ quelconque ?

1. Sous la courbe normale de Z, coloriez l'aire correspondant à $P(Z \leq a)$ et $P(Z \leq b)$ pour un intervalle $[a, b]$ où $a < b$.
2. Quelle opération permet d'obtenir l'aire <u>entre</u> a et b, c'est-à-dire $P(a \leq Z \leq b)$, à partir des aires $P(Z \leq a)$ et $P(Z \leq b)$?
3. À partir de ce résultat, calculez la probabilité que la taille d'un adulte se situe entre 168cm et 189cm.
4. **(Bonus)** On souhaite approximer $P(Y = 55)$ avec $Y \sim B(100, \frac{1}{2})$ via une variable normale $X \sim N(50, 25)$. Cependant, $P(X = 55)$ vaut zéro, car X est continue. L'approximation $P(Y = k) \approx P(k - 0,5 \leq X \leq k + 0,5)$ est appelée <u>correction de continuité</u>. Approximez $P(Y = 55)$ par la normale X. Est-elle proche de la vraie probabilité ?

En pratique, on utilise souvent des logiciels plutôt que le tableau $\Phi(Z)$ pour calculer les probabilités d'une variable qui suit une loi normale. Néanmoins, il est intéressant de connaître la technique pour centrer et réduire une variable, car elle est souvent employée pour normaliser (ou "standardiser") des données réelles.

Rappels calcul de primitives et intégrales

1. $\int f(x)\,dx = F(x) + C$ tel que $F'(x) = f(x)$
2. $\int f(x) \pm g(x)\,dx = \int f(x)\,dx \pm \int g(x)\,dx$
3. $\int kf(x)\,dx = k \int f(x)\,dx$
4. $\int_a^b f(x)\,dx = [F(x)]_a^b = F(b) - F(a)$
5. $\int dx = x + C$
6. $\int k\,dx = kx + C$
7. $\int x^n\,dx = \dfrac{x^{n+1}}{n+1} + C, n \neq -1$
8. $\int x^{-1}\,dx = \int \dfrac{1}{x}\,dx = \ln|x| + C, x \neq 0$

Table de la fonction de répartition $\Phi(a) = P\{Z \leq a\}$ de la loi normale centrée réduite $N(0,1)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Ligne : première décimale, colonne : deuxième décimale, p. ex. : $\Phi(1,67) = 0,9525$ (ligne 1.6, colonne 0.07)