Probabilités

Vincent Arel-Bundock

1 Concepts

2 Probabilités

3 Théorème de Bayes

4 Distributions

Concepts

Concepts

Événement:

> *Un* résultat possible d'un processus physique ou social.

Concepts

Événement:

> *Un* résultat possible d'un processus physique ou social.

Espace échantillonal:

> L'*ensemble* de tous les événements que ce processus peut produire.

Concepts

Événement:

> *Un* résultat possible d'un processus physique ou social.

Espace échantillonal:

> L'*ensemble* de tous les événements que ce processus peut produire.

Variable aléatoire:

> Représentation algébrique de l'espace échantillonnal.

Exemple

Processus:

> Lancer d'une pièce de monnaie

Espace échantillonal:

> Un ensemble de 2 événements: {Pile, Face}

Variable aléatoire:

> $$X = \begin{cases} 1 \text{ si Pile} \\ 0 \text{ si Face} \end{cases}$$

Types de Variables

Continue

- N'importe quelle valeur sur un intervalle

Binaire ou dichotomique

- Deux événements possibles

Dénombrement

- Nombres entiers non-négatifs

Types de Variables

Continue

- N'importe quelle valeur sur un intervalle

Binaire ou dichotomique

- Deux événements possibles

Dénombrement

- Nombres entiers non-négatifs

Ordinale

- Catégories ordonnées.

Nominale

- Catégories distinctes, mais non ordonnées

Probabilités

Probabilités

Probabilité d'un événement:

> Un chiffre entre 0 et 1 qui correspond au risque d'observer cet événement.

Loi de distribution d'une variable X:

> L'ensemble des probabilités associées à toutes les valeurs possibles de X.

Probabilités: Exemple

Si X représente le lancer d'une pièce de monnaie, alors

$$P(Pile) = 0,5$$

$$P(Face) = 0,5$$

et la loi de distribution de X peut s'exprimer ainsi:

$$P(X) = \begin{cases} P(X = \text{Pile}) = \frac{1}{2} \\ P(X = \text{Face}) = \frac{1}{2} \end{cases}$$

Probabilité d'une variable

Règle 1

La somme des probabilités de tous les événements mutuellement exclusifs est égale à 1.

$$P(Pile) + P(Face) = 1$$

Probabilité d'une variable

Règle 1

La somme des probabilités de tous les événements mutuellement exclusifs est égale à 1.

$$P(Pile) + P(Face) = 1$$

Règle 2: Complément

$$P(X \neq a) = 1 - P(X = a)$$

Probabilité d'une variable

Règle 1

La somme des probabilités de tous les événements mutuellement exclusifs est égale à 1.

$$P(Pile) + P(Face) = 1$$

Règle 2: Complément

$$P(X \neq a) = 1 - P(X = a)$$

Règle 3: Addition

Si a et b sont mutuellement exlusifs, alors la probabilité que X soit égale à a *ou* à b est égale à:

$$P(X = a) + P(X = b)$$

Probabilité de plusieurs variables (conjointe)

$P(Y, X)$ représente la probabilité conjointe de deux variables.

Par exemple,

$$X \in \{\text{Fruits, Chocolat}\}$$

$$Y \in \{\text{Tarte, Gâteau}\}$$

L'expression $P(Y, X)$ se lit:

> Probabilité de Y *et* X.

"et" = "intersection"

	Fruits	Chocolat
Tarte	1	2
Gâteau	3	4

Probabilité à plusieurs variables (conjointe)

Probabilités d'une variable:

$$P(Y = \text{Tarte}) = \frac{3}{10}$$

$$P(X = \text{Fruits}) = \frac{4}{10}$$

Probabilité conjointe: "et", "intersection"

$$P(Y = \text{Tarte}, X = \text{Fruits}) = \frac{1}{10}$$

	Fruits	Chocolat
Tarte	1	2
Gâteau	3	4

Probabilité à plusieurs variables (conjointe)

ET vs. OU

Lorsque nous voulons calculer la probabilité d'un événement *ou* d'un autre événement, il faut examiner la probabilité conjointe du "complément":

Probabilité de piger une tarte *ou* un dessert aux fruits:

$$1 - P(X \neq \text{Tarte}, Y \neq \text{Fruits})$$

$$1 - 4/10 = 6/10$$

	Fruits	Chocolat
Tarte	1	2
Gâteau	3	4

Probabilité conditionnelle

La probabilité conditionnelle s'écrit:

$$P(Y|X)$$

et se dit:

> "probabilité de Y étant donné X."

Probabilité conditionnelle

La probabilité conditionnelle s'écrit:

$$P(Y|X)$$

et se dit:

> "probabilité de Y étant donné X."

Intuition: Calculer la probabilité conditionnelle équivaut à

- contrôler,
- tenir constant,
- fixer

la variable de conditionnement.

Probabilité conditionnelle

Définition:

$$P(Y|X) = \frac{P(Y,X)}{P(X)},$$

	Fruits	Chocolat
Tarte	1	2
Gâteau	3	4

Calcul:

$$P(Y = \text{Tarte}|X = \text{Fruits}) = \frac{P(Y = \text{Tarte}, X = \text{Fruits})}{P(X = \text{Fruits})} = \frac{1/10}{4/10} = \frac{1}{4}$$

Si les seuls desserts disponibles sont garnis de fruits, la probabilité de piger une tarte au hasard est de 1/4.

Indépendance

Indépendance

Lorsque $Y \perp X$, nous pouvons faire deux constats:

Constat 1: La probabilité conjointe est égale au produit des probabilités individuelles

$$P(Y, X) = P(Y)P(X)$$

Constat 2: La probabilité conditionnelle est égale à la probabilité

$$P(Y|X) = P(Y)$$

$$P(X|Y) = P(X)$$

Indépendance

Constat 1: La probabilité conjointe est égale au produit des probabilités individuelles:

$$P(Y, X) = P(Y)P(X)$$

Indépendance

Constat 1: La probabilité conjointe est égale au produit des probabilités individuelles:

$$P(Y, X) = P(Y)P(X)$$

Exemple

- Lancer 3 dés indépendants à 6 faces (X, Y, Z).
- Quelle est la probabilité d'obtenir le chiffre 5 trois fois?

$$\begin{aligned} P(X = 5, Y = 5, Z = 5) &= P(X = 5)P(Y = 5)P(Z = 5) \\ &= \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216} \end{aligned}$$

Indépendance

Constat 2: La probabilité conditionnelle est égale à la probabilité

$$P(Y|X) = P(Y)$$

$$P(X|Y) = P(X)$$

Indépendance

Constat 2: La probabilité conditionnelle est égale à la probabilité

$$P(Y|X) = P(Y)$$

$$P(X|Y) = P(X)$$

Exemple

Manger un sandwich à Montréal $\perp$ Température à New Delhi

$$P(\text{Sandwich}|\text{T=40°C}) = P(\text{Sandwich}|\text{T=10°C}) = P(\text{Sandwich})$$

Qu'il fasse 10°C ou 40°C à New Delhi, la probabilité que je mange un sandwich à Montréal demeure la même.

Indépendance

Une nouvelle diplômée de l'UdeM postule pour 10 postes en relations publiques. La probabilité d'être choisie est de 5% pour chaque poste, et les résultats sont indépendants d'un poste à l'autre.

La probabilité conjointe d'obtenir *tous* les postes (poste 1 ET poste 2 ET...), est de:

$$0,05^{10} = 0.0000000000009$$

La probabilité conjointe d'obtenir *un* poste (poste 1 OU poste 2 OU...) est obtenue en considérant le complément:

$$1 - (1 - 0,05)^{10} = 0,40$$

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}$$

Théorème de Bayes

Une des plus importantes idées... tout court:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Permet de comparer 2 probabilités conditionnelles.

Révérend Thomas Bayes (1701-1761)

Théorème de Bayes: Exemple 1

Les tests de dépistage pour la Covid-19 ne sont pas parfaits. Parfois, ils font des erreurs: un individu malade peut recevoir un test négatif et un individu en santé peut recevoir un test positif. Comparez 2 quantités:

Probabilité d'obtenir un test positif si un individu est affecté par la Covid-19:

$$P(+|\text{Malade})$$

Probabilité d'être affecté par la Covid-19 si un individu reçoit un test positif:

$$P(\text{Malade}|+)$$

Théorème de Bayes: Exemple 1

Imaginez que le gouvernement canadien décide de tester tous les citoyens du pays, et que:

- 20% de la population est malade (80% en santé):
  - $P(\text{Malade}) = 0,2$
- 90% des personnes malades reçoivent un test positif (correctement):
  - $P(+|\text{Malade}) = 0,9$
- 10% des personnes en santé reçoivent un test positif (par erreur):
  - $P(+|\text{En Santé}) = 0,1$

Dans ce contexte, la probabilité d'obtenir un test positif est:

$$P(+) = P(\text{Malade})P(+|\text{Malade}) + P(\text{En Santé})P(+|\text{En Santé})$$

$$P(+) = 0,2 \cdot 0,9 + 0,8 \cdot 0,1 = 0,26$$

Théorème de Bayes: Exemple 1

$$P(\text{Malade}) = 0,2$$

$$P(+|\text{Malade}) = 0,9$$

$$P(+) = 0,26$$

$$P(\text{Malade}|+) = \frac{P(+|\text{Malade})P(\text{Malade})}{P(+)}$$

$$P(\text{Malade}|+) = \frac{0,9 \cdot 0,2}{0,26} = 0,69$$

Si un individu reçoit un test positif, la probabilité qu'il soit réellement malade est de 69% et ce, même si le test détecte 90% des cas de Covid-19!

Théorème de Bayes: Exemple 2

Un journal publie la manchette suivante :

> "75% des personnes condamnées pour voie de fait ont joué à des jeux vidéos."

Formellement, le journal affirme:

$$P(\text{Jeux vidéos}|\text{Violence}) = 0.75$$

Ce qui nous intéresse plus, c'est la probabilité conditionnelle inverse:

$$P(\text{Violence}|\text{Jeux vidéos})$$

Théorème de Bayes: Exemple 2

Ces deux probabilités sont liées par le théorème de Bayes :

$$P(\text{Jeux vidéos}|\text{Violence}) = \frac{P(\text{Violence}|\text{Jeux vidéos}) \cdot P(\text{Jeux vidéos})}{P(\text{Violence})}$$

3 phénomènes peuvent expliquer la statistique publiée par le journal :

1. $P(\textbf{Violence})$ est élevée parmi les joueurs ;
2. $P(\textbf{Jeux vidéos})$ est élevée ;
3. $P(\textbf{Condamnation pour voie de fait})$ est faible.

Pour pouvoir tirer des conclusions quant à $P(\textbf{Violence}|\textbf{Jeux vidéos})$, il faut connaître les éléments 2 et 3.

Théorème de Bayes: Exemple 3

RESEARCH ARTICLE

Officer characteristics and racial disparities in fatal officer-involved shootings

David J. Johnson, Trevor Tress, Nicole Burkel, Carley Taylor, and Joseph Cesario

PNAS August 6, 2019 116 (32) 15877-15882; first published July 22, 2019
https://doi.org/10.1073/pnas.1903856116

Étude publiée en 2019 dans les *Proceedings of the National Academy of Sciences*.

Barr Says There Is No Systemic Racisn Policing

The attorney general's remarks, which mirrored those of other administration officials, came as the president was scheduled to meet with law enforcement officials at the White House.

Citée fréquemment par des politiciens et bureaucrates.

Théorème de Bayes: Exemple 3

La question scientifique qui nous intéresse:

> Est-ce que les noirs ont une plus grandes de recevoir des coups de feu par la police que les blancs?

$$P(\text{Violence policière}|\text{Couleur})$$

Problème! Les chercheurs estiment:

$$P(\text{Couleur}|\text{Violence policière})$$

Grâce au Théorème de Bayes, nous savons que:

$$P(\text{Couleur}|\text{Violence}) = \frac{P(\text{Violence}|\text{Couleur})P(\text{Couleur})}{P(\text{Violence})}$$

Théorème de Bayes: Exemple 3

Comme l'expliquent deux critiques (Knox & Mummolo 2020), il est impossible de répondre à la question de recherche sans tenir compte du Théorème de Bayes!

L'article est retiré par la revue (à la demande des auteurs).

> This article has a Correction, but has also been retracted and


Distributions

Distributions

Lorsque le nombre d'événements dans l'espace échantillonal est petit, il est facile de décrire une variable:

$$X = \begin{cases} 1 \text{ si Pile} \\ 0 \text{ si Face} \end{cases}$$

Problèmes:

- Complexité
- Événements non-dénombrables

Solution:

- Distributions

Distribution

Définition:

> Une fonction mathématique qui décrit la probabilité qu'un processus physique ou social produise certains événements.

Bénéfices:

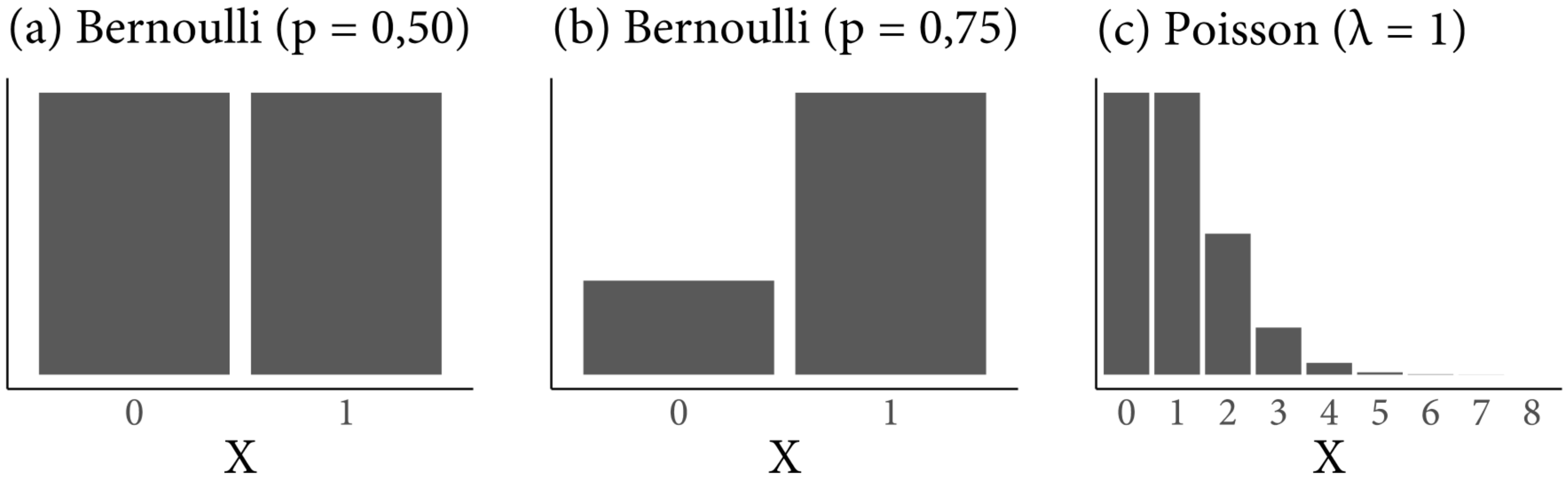
- Lorsqu'un processus se conforme à une distribution donnée, nous sommes en mesure de quantifier la probabilité d'observer un résultat plutôt qu'un autre.

Types de distributions:

- Discrètes
- Continues

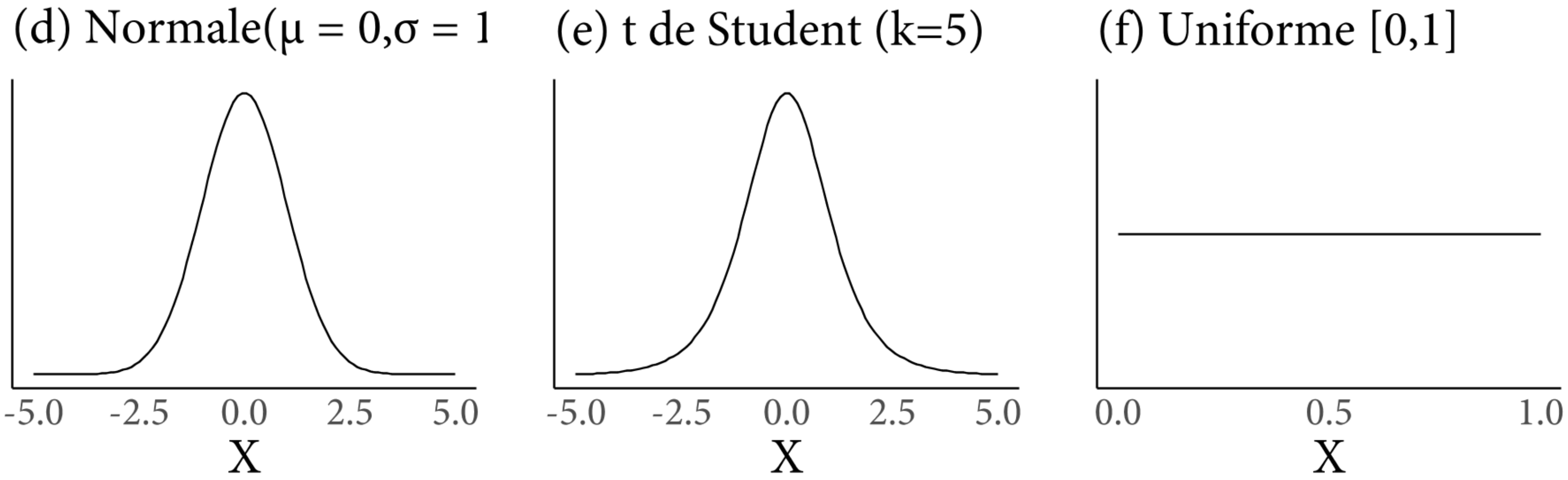
Distributions discrètes

Variables à valeurs dénombrables :



- Bernoulli (ou binomiale): deux valeurs possibles
- Poisson: variables de dénombrement, nombres entiers positifs

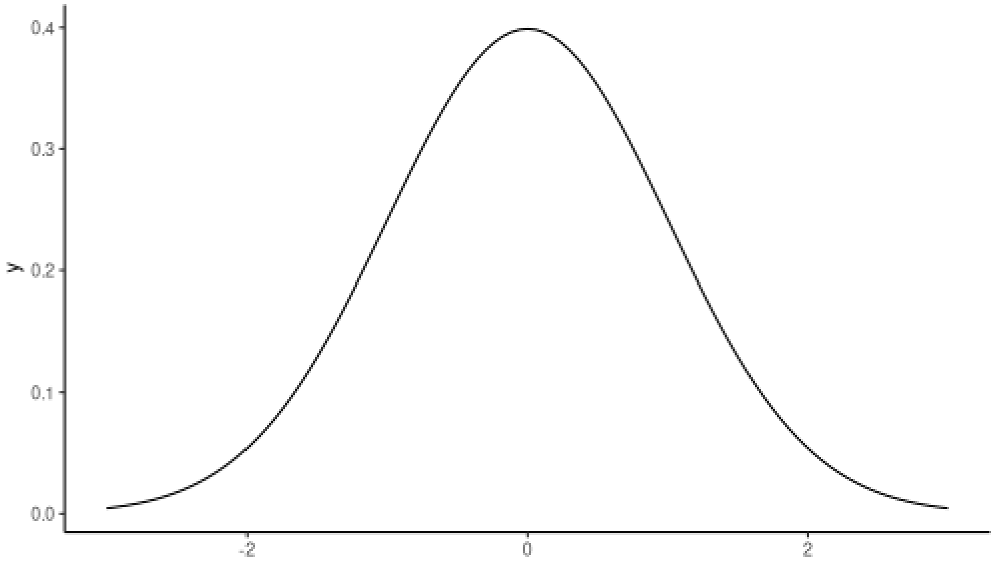
Distributions continues



Distribution normale

3 charactéristiques de la courbe normale :

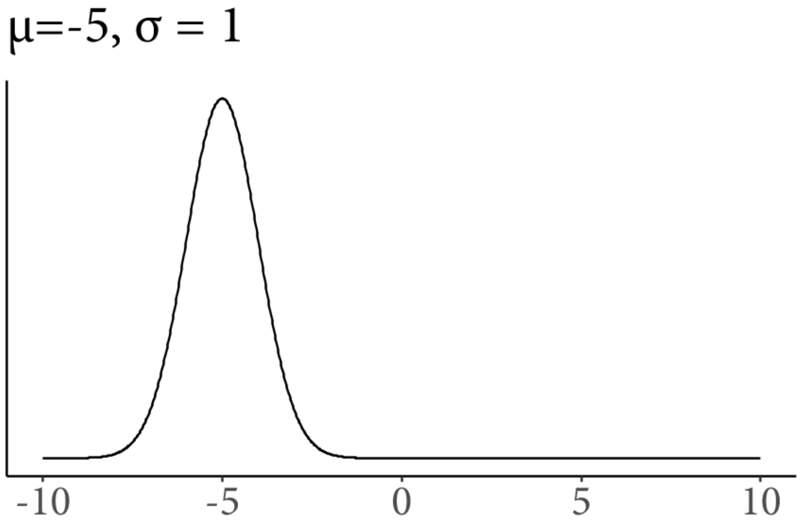
1. Symétrie
2. Pic
3. Forme: moyenne (μ) et écart type (σ)

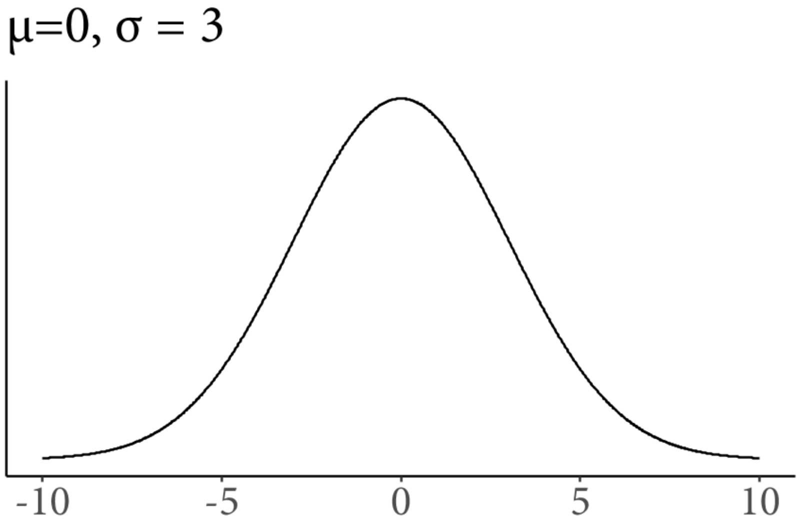


Distribution normale

La distribution normale a deux paramètres:

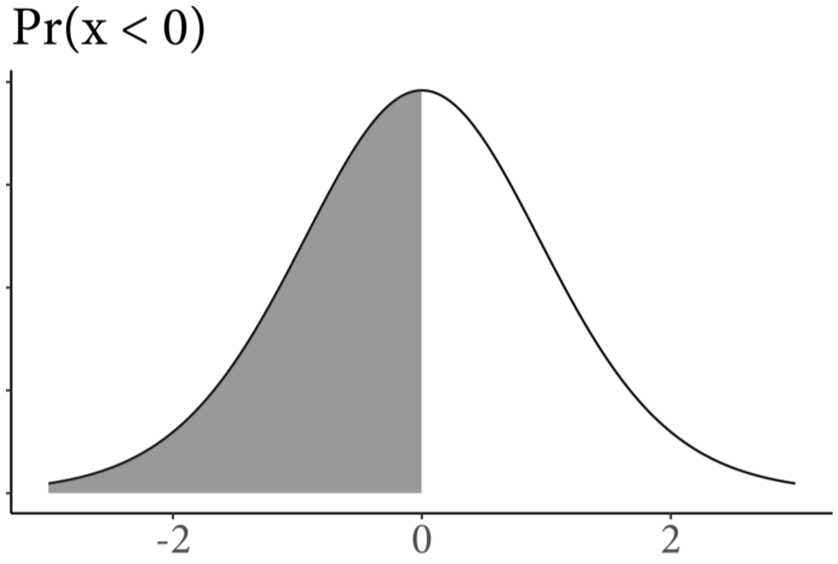
1. Moyenne
2. Variance

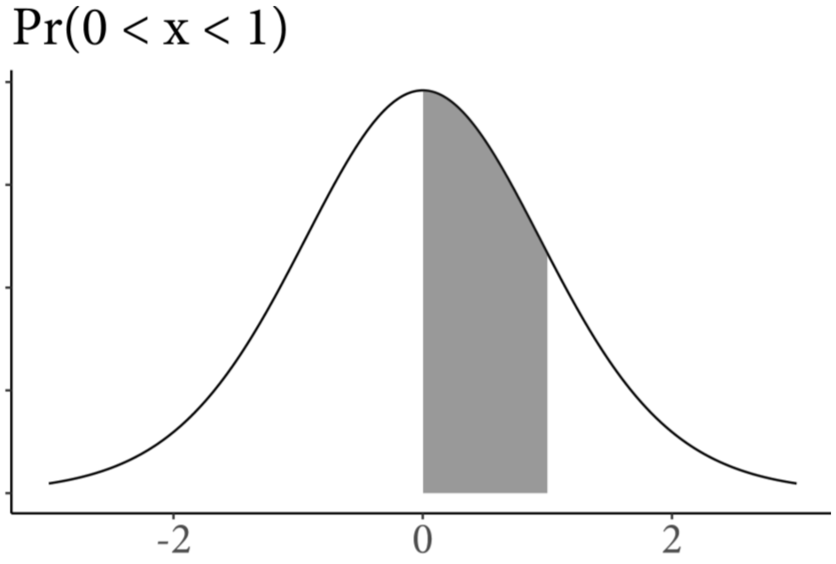




Aire sous la courbe de distribution

L'aire sous la courbe de distribution est toujours égale à 1.





Aire sous la courbe de distribution

La fonction `pnorm()` dans `R` calcule l'aire sous la courbe d'une distribution normale à gauche de x.
$Pr(x < 0)$:

```
pnorm(0)
```

```
## [1] 0.5
```

$Pr(0 < x < 1)$:

```
pnorm(1) - pnorm(0)
```

```
## [1] 0.3413447
```

Merci!