

IFT-3655, Modèles Stochastiques

Introduction et rappels sur les probabilités

Prof. Pierre L'Ecuyer

DIRO, Université de Montréal

Ces “diapos” sont surtout un support pour les présentations en classe. Elles ne contiennent évidemment pas toutes les explications détaillées (ce n'est pas un livre). Pour cela il est recommandé de bien écouter le cours et d'étudier le livre de Sheldon Ross. Le livre contient beaucoup d'exemples additionnels et de nombreux exercices.

Outils

Lien sur la page web du cours. Voir le plan de cours.

Lien IFT3655 sur Studium

Livre de base: Introduction to Probability Models, Sheldon M. Ross, 11-ième édition (2014), ou 12-ième édition (2019), ou 13-ième édition (2023).

Modèles stochastiques

Définition. Un **modèle stochastique** (ou probabiliste) est un modèle mathématique (une abstraction) pouvant représenter un système comportant de l'incertitude.

L'incertitude peut être due à des mesures imprécises, ou simplement à un manque de connaissances du système, ou au fait que le système est trop complexe pour que l'on puisse prédire exactement son comportement.

Modèles stochastiques

Définition. Un **modèle stochastique** (ou probabiliste) est un modèle mathématique (une abstraction) pouvant représenter un système comportant de l'incertitude.

L'incertitude peut être due à des mesures imprécises, ou simplement à un manque de connaissances du système, ou au fait que le système est trop complexe pour que l'on puisse prédire exactement son comportement.

On modélise alors le comportement par des lois de probabilité, qui sont des abstractions mathématiques avec lesquelles il est beaucoup plus facile de travailler.

On a des variables aléatoires (dans \mathbb{R}), des vecteurs aléatoires (dans \mathbb{R}^d), des processus stochastiques, des fonctions aléatoires, etc.

Nous allons commencer par reviser ces notions de base, qui sont essentiels pour l'étude des modèles stochastiques.

Exemple: lancer un dé

Quand on lance un dé à 6 faces, on peut adopter le **modèle abstrait** que chaque face a exactement la probabilité $1/6$, même si cela n'est **pas vraiment exact**, car pour un vrai dé les probabilités peuvent varier légèrement.

Exemple: lancer un dé

Quand on lance un dé à 6 faces, on peut adopter le **modèle abstrait** que chaque face a exactement la probabilité $1/6$, même si cela n'est **pas vraiment exact**, car pour un vrai dé les probabilités peuvent varier légèrement.

On pourrait imaginer qu'un robot soit capable de lancer le dé en mettant juste assez de force de rotation, vitesse, etc., pour avoir la valeur voulue quasiment à tout coup. Dans ce cas, le modèle ne tient plus si on sait ce que le robot va faire, mais si on ne sait pas et si le robot varie suffisamment ses choix, il se peut qu'un observateur non averti ne puisse pas prédire le résultat et que le modèle soit quand même approprié. On pourra alors dire que le robot **simule** des dés équitables.

Exemple: lancer un dé

Quand on lance un dé à 6 faces, on peut adopter le **modèle abstrait** que chaque face a exactement la probabilité $1/6$, même si cela n'est **pas vraiment exact**, car pour un vrai dé les probabilités peuvent varier légèrement.

On pourrait imaginer qu'un robot soit capable de lancer le dé en mettant juste assez de force de rotation, vitesse, etc., pour avoir la valeur voulue quasiment à tout coup. Dans ce cas, le modèle ne tient plus si on sait ce que le robot va faire, mais si on ne sait pas et si le robot varie suffisamment ses choix, il se peut qu'un observateur non averti ne puisse pas prédire le résultat et que le modèle soit quand même approprié. On pourra alors dire que le robot **simule** des dés équitables.

Il se pourrait aussi qu'un robot puisse **prédire** le résultat quasiment à tout coup pendant que le dé est encore en l'air, juste après qu'on l'a lancé, sur la base de la vitesse de rotation, etc. Ce robot pourrait alors mieux prédire que le modèle abstrait simplifié. Mieux prédire les choses est l'un des grands objectifs en intelligence artificielle (IA), en statistique, et dans de nombreux autres domaines. Pour cela, on utilise aussi des modèles probabilistes.

Exemple: modèles de files d'attente

Beaucoup d'applications: centres d'appels, urgence d'un hôpital, comptoir de glaces, cafés, restaurants, boulangerie, etc. Comment estimer la distribution des temps d'attente?

Supposons qu'il arrive λ clients par minute en moyenne, qu'il y a s serveurs, et que chaque serveur peut servir en moyenne μ clients par minute (la durée moyenne d'un service est de $1/\mu$ minutes). La demande totale est λ/μ minutes de service par minute. Il faut $s \geq \lambda/\mu$ pour pouvoir couvrir la demande. Ex.: si $\lambda = 1$ et $1/\mu = 2.9$, il faut $s \geq 3$.

Exemple: modèles de files d'attente

Baucoup d'applications: centres d'appels, urgence d'un hôpital, comptoir de glaces, cafés, restaurants, boulangerie, etc. Comment estimer la distribution des temps d'attente?

Supposons qu'il arrive λ clients par minute en moyenne, qu'il y a s serveurs, et que chaque serveur peut servir en moyenne μ clients par minute (la durée moyenne d'un service est de $1/\mu$ minutes). La demande totale est λ/μ minutes de service par minute. Il faut $s \geq \lambda/\mu$ pour pouvoir couvrir la demande. Ex.: si $\lambda = 1$ et $1/\mu = 2.9$, il faut $s \geq 3$.

Dans un modèle déterministe qui suppose qu'il arrive un client exactement à toutes les minutes et que chaque durée de service est exactement $1/\mu$, si $s > \lambda/\mu$ il n'y aura jamais d'attente.

Exemple: modèles de files d'attente

Baucoup d'applications: centres d'appels, urgence d'un hôpital, comptoir de glaces, cafés, restaurants, boulangerie, etc. Comment estimer la distribution des temps d'attente?

Supposons qu'il arrive λ clients par minute en moyenne, qu'il y a s serveurs, et que chaque serveur peut servir en moyenne μ clients par minute (la durée moyenne d'un service est de $1/\mu$ minutes). La demande totale est λ/μ minutes de service par minute. Il faut $s \geq \lambda/\mu$ pour pouvoir couvrir la demande. Ex.: si $\lambda = 1$ et $1/\mu = 2.9$, il faut $s \geq 3$.

Dans un modèle déterministe qui suppose qu'il arrive un client exactement à toutes les minutes et que chaque durée de service est exactement $1/\mu$, si $s > \lambda/\mu$ il n'y aura jamais d'attente. Mais pas réaliste!

Exemple: modèles de files d'attente

Beaucoup d'applications: centres d'appels, urgence d'un hôpital, comptoir de glaces, cafés, restaurants, boulangerie, etc. Comment estimer la distribution des temps d'attente?

Supposons qu'il arrive λ clients par minute en moyenne, qu'il y a s serveurs, et que chaque serveur peut servir en moyenne μ clients par minute (la durée moyenne d'un service est de $1/\mu$ minutes). La demande totale est λ/μ minutes de service par minute. Il faut $s \geq \lambda/\mu$ pour pouvoir couvrir la demande. Ex.: si $\lambda = 1$ et $1/\mu = 2.9$, il faut $s \geq 3$.

Dans un modèle déterministe qui suppose qu'il arrive un client exactement à toutes les minutes et que chaque durée de service est exactement $1/\mu$, si $s > \lambda/\mu$ il n'y aura jamais d'attente. Mais pas réaliste!

Pour un modèle stochastique qui donne une formule simple, supposons que les clients arrivent complètement au hasard et indépendamment les uns des autres (processus de Poisson) et que les durées de service suivent une loi sans mémoire (loi exponentielle).

Exemple: modèles de files d'attente

Beaucoup d'applications: centres d'appels, urgence d'un hôpital, comptoir de glaces, cafés, restaurants, boulangerie, etc. Comment estimer la distribution des temps d'attente?

Supposons qu'il arrive λ clients par minute en moyenne, qu'il y a s serveurs, et que chaque serveur peut servir en moyenne μ clients par minute (la durée moyenne d'un service est de $1/\mu$ minutes). La demande totale est λ/μ minutes de service par minute. Il faut $s \geq \lambda/\mu$ pour pouvoir couvrir la demande. Ex.: si $\lambda = 1$ et $1/\mu = 2.9$, il faut $s \geq 3$.

Dans un modèle déterministe qui suppose qu'il arrive un client exactement à toutes les minutes et que chaque durée de service est exactement $1/\mu$, si $s > \lambda/\mu$ il n'y aura jamais d'attente. Mais pas réaliste!

Pour un modèle stochastique qui donne une formule simple, supposons que les clients arrivent complètement au hasard et indépendamment les uns des autres (processus de Poisson) et que les durées de service suivent une loi sans mémoire (loi exponentielle).

Dans ce cas, si $\lambda/\mu < s = 1$ (un seul serveur), l'attente moyenne par client à long terme est $w = \lambda/[\mu(\mu - \lambda)]$. Nous verrons plus tard comment trouver ce genre de formule.

Pour $\lambda = 1$, si $1/\mu = 0.9$ on a $w = 8.1$. Si $1/\mu = 0.99$ on a $w = 98.01$.

On trouve aussi des formules (plus compliquées) pour le cas général où $s \geq 1$, pour w , pour la proportion des clients qui attendent plus de s secondes, etc.

Aussi pour des modèles où les clients en attente peuvent abandonner, où la file d'attente a une longueur limitée (ou même nulle), où on a un réseau de plusieurs files d'attente, etc.

Mot clé: "[Erlang formulas](#)".

Dans la vraie vie, λ et s peuvent varier dans le temps, etc., et la réalité peut être trop compliquée pour trouver des formules. Solution: [simulation stochastique](#).

On peut construire un modèle aussi réaliste et compliqué que l'on veut et le simuler sur ordinateur pour apprendre le comportement du système qui nous intéresse.

Autres exemples

Centre d'appels 911 de Montréal: Les appels (différents types) arrivent "au hasard". On doit décider du nombre d'agents de chaque type qui répondent aux appels à chaque heure du jour, chaque jour de la semaine. Aussi: combien d'ambulances, de voitures de police, où les placer, etc.

Autres exemples

Centre d'appels 911 de Montréal: Les appels (différents types) arrivent "au hasard". On doit décider du nombre d'agents de chaque type qui répondent aux appels à chaque heure du jour, chaque jour de la semaine. Aussi: combien d'ambulances, de voitures de police, où les placer, etc.

Magasin de détail: Combien d'employés sur le plancher du magasin pour chaque période de 15 min de la semaine? Quoi mettre dans les étagères et quoi avoir en inventaire? Les arrivées des clients, leurs demandes, et l'impact du nombre d'employés sur les ventes sont incertains.

Autres exemples

Centre d'appels 911 de Montréal: Les appels (différents types) arrivent "au hasard". On doit décider du nombre d'agents de chaque type qui répondent aux appels à chaque heure du jour, chaque jour de la semaine. Aussi: combien d'ambulances, de voitures de police, où les placer, etc.

Magasin de détail: Combien d'employés sur le plancher du magasin pour chaque période de 15 min de la semaine? Quoi mettre dans les étagères et quoi avoir en inventaire? Les arrivées des clients, leurs demandes, et l'impact du nombre d'employés sur les ventes sont incertains.

Autres: Réseaux de communication, fiabilité, entretien préventif, taxis, véhicules de livraison, transport de marchandises, transport aérien, hôtels, restaurants, boulangeries, médecine, etc.

Autres exemples

Centre d'appels 911 de Montréal: Les appels (différents types) arrivent "au hasard". On doit décider du nombre d'agents de chaque type qui répondent aux appels à chaque heure du jour, chaque jour de la semaine. Aussi: combien d'ambulances, de voitures de police, où les placer, etc.

Magasin de détail: Combien d'employés sur le plancher du magasin pour chaque période de 15 min de la semaine? Quoi mettre dans les étagères et quoi avoir en inventaire? Les arrivées des clients, leurs demandes, et l'impact du nombre d'employés sur les ventes sont incertains.

Autres: Réseaux de communication, fiabilité, entretien préventif, taxis, véhicules de livraison, transport de marchandises, transport aérien, hôtels, restaurants, boulangeries, médecine, etc.

Quel est le meilleur modèle? Plus simple? Plus réaliste?

Pas de réponse universelle. Cela dépend de ce que l'on veut faire. Discuter.

En général, on veut un modèle le plus simple possible, mais suffisamment réaliste.

Brève introduction à la théorie des probabilités

On va (re)voir comment définir et calculer des probabilités, ainsi que certains concepts et outils de base.

Notion fondamentale: Espace de probabilité

Définition. Un **espace de probabilité** est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$ où Ω est l'**espace échantillonnal** (un ensemble de **réalisations** possibles), \mathcal{F} est une famille d'**événements** (des sous-ensembles mesurables de Ω) qui forme une σ -algèbre, et \mathbb{P} est une **mesure de probabilité** qui associe à chaque événement $E \in \mathcal{F}$ un nombre réel $\mathbb{P}(E) \in [0, 1]$.

Conditions pour que \mathcal{F} soit une σ -algèbre: (1) si $E \in \mathcal{F}$, alors $E^c \stackrel{\text{def}}{=} \Omega \setminus E \in \mathcal{F}$;

(2) si $E_1, E_2, \dots \in \mathcal{F}$, alors $\cup_{n=1}^{\infty} E_n \in \mathcal{F}$ et $\cap_{n=1}^{\infty} E_n \in \mathcal{F}$.

Conditions pour que \mathbb{P} soit une **mesure de probabilité**: (1) $\mathbb{P}(\Omega) = 1$;

(2) si $E_1, E_2, \dots \in \mathcal{F}$ avec $E_i \cap E_j = \emptyset$ pour tout $i \neq j$, alors $\mathbb{P}(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mathbb{P}(E_n)$.

Notion fondamentale: Espace de probabilité

Définition. Un **espace de probabilité** est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$ où Ω est l'**espace échantillonnal** (un ensemble de **réalisations** possibles), \mathcal{F} est une famille d'**événements** (des sous-ensembles mesurables de Ω) qui forme une σ -algèbre, et \mathbb{P} est une **mesure de probabilité** qui associe à chaque événement $E \in \mathcal{F}$ un nombre réel $\mathbb{P}(E) \in [0, 1]$.

Conditions pour que \mathcal{F} soit une σ -algèbre: (1) si $E \in \mathcal{F}$, alors $E^c \stackrel{\text{def}}{=} \Omega \setminus E \in \mathcal{F}$;
 (2) si $E_1, E_2, \dots \in \mathcal{F}$, alors $\bigcup_{n=1}^{\infty} E_n \in \mathcal{F}$ et $\bigcap_{n=1}^{\infty} E_n \in \mathcal{F}$.

Conditions pour que \mathbb{P} soit une **mesure de probabilité**: (1) $\mathbb{P}(\Omega) = 1$;
 (2) si $E_1, E_2, \dots \in \mathcal{F}$ avec $E_i \cap E_j = \emptyset$ pour tout $i \neq j$, alors $\mathbb{P}(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mathbb{P}(E_n)$.

Quand Ω est fini ou dénombrable, on peut prendre \mathcal{F} comme la famille de tous les sous-ensembles de Ω . Mais quand Ω est non dénombrable, c'est plus compliqué.

Quand $\Omega = \mathbb{R}$ ou $\Omega = [a, b]$ (un intervalle), on prend habituellement \mathcal{F} comme la famille \mathcal{B} des **sous-ensembles de Borel**, qui sont les sous-ensembles obtenus par l'union et l'intersection dénombrable d'intervalles. On ne va pas considérer ces questions rigoureusement dans ce cours. Sauf indication contraire, on va supposer que les événements considérés sont dans \mathcal{F} .

Exemples

Exemple: Pile ou face.

Si on lance 1 pièce, on peut poser $\Omega = \{P, F\}$, $\mathcal{F} = \{\emptyset, \{P\}, \{F\}, \{P, F\}\}$.

Si la pièce est équitale (ou équilibrée), on a $\mathbb{P}(\{P\}) = \mathbb{P}(\{F\}) = 1/2$.

Si la pièce est biaisée, on peut avoir par exemple $\mathbb{P}(\{P\}) = 2/3$ et $\mathbb{P}(\{F\}) = 1/3$.

Dans tous les cas, on a $\mathbb{P}(\{\emptyset\}) = 0$ et $\mathbb{P}(\{P, F\}) = 1$.

Exemples

Exemple: Pile ou face.

Si on lance 1 pièce, on peut poser $\Omega = \{P, F\}$, $\mathcal{F} = \{\emptyset, \{P\}, \{F\}, \{P, F\}\}$.

Si la pièce est équitale (ou équilibrée), on a $\mathbb{P}(\{P\}) = \mathbb{P}(\{F\}) = 1/2$.

Si la pièce est biaisée, on peut avoir par exemple $\mathbb{P}(\{P\}) = 2/3$ et $\mathbb{P}(\{F\}) = 1/3$.

Dans tous les cas, on a $\mathbb{P}(\{\emptyset\}) = 0$ et $\mathbb{P}(\{P, F\}) = 1$.

Si on lance 2 pièces, on peut poser $\Omega = \{(P, P), (P, F), (F, P), (F, F)\}$, \mathcal{F} = la famille des 2^4 sous-ensembles de Ω , et donner la probabilité $1/4$ à chacune des 4 réalisations possibles (les 4 éléments de Ω).

On aura alors par exemple $\mathbb{P}(\{(P, P), (F, F)\}) = 1/2$, etc.

En général, si Ω contient m réalisations possibles, il y a 2^m sous-ensembles dans \mathcal{F} .

On lance deux dés à 6 faces. Une réalisation a la forme (i, j) où i est la valeur sur le premier dé et j est la valeur sur le second dé. L'espace Ω contient 36 réalisations possibles, chacune ayant une probabilité $1/36$.

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, \dots, (5, 6), (6, 6)\}.$$

Ici, \mathcal{F} contient tous les sous-ensembles de Ω (il y en a 2^{36}).

On lance deux dés à 6 faces. Une réalisation a la forme (i, j) où i est la valeur sur le premier dé et j est la valeur sur le second dé. L'espace Ω contient 36 réalisations possibles, chacune ayant une probabilité $1/36$.

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, \dots, (5, 6), (6, 6)\}.$$

Ici, \mathcal{F} contient tous les sous-ensembles de Ω (il y en a 2^{36}).

Par exemple, l'événement "la somme des valeurs sur les deux dés est 4" est $E_1 = \{(1, 3), (2, 2), (3, 1)\} \in \mathcal{F}$ et on a $\mathbb{P}(E_1) = 3/36 = 1/12$. Voir page suivante.

La probabilité de l'événement complémentaire est $\mathbb{P}(E_1^c) = 1 - \mathbb{P}(E_1) = 11/12$.

On lance deux dés à 6 faces. Une réalisation a la forme (i, j) où i est la valeur sur le premier dé et j est la valeur sur le second dé. L'espace Ω contient 36 réalisations possibles, chacune ayant une probabilité $1/36$.

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, \dots, (5, 6), (6, 6)\}.$$

Ici, \mathcal{F} contient tous les sous-ensembles de Ω (il y en a 2^{36}).

Par exemple, l'événement "la somme des valeurs sur les deux dés est 4" est $E_1 = \{(1, 3), (2, 2), (3, 1)\} \in \mathcal{F}$ et on a $\mathbb{P}(E_1) = 3/36 = 1/12$. Voir page suivante.

La probabilité de l'événement complémentaire est $\mathbb{P}(E_1^c) = 1 - \mathbb{P}(E_1) = 11/12$.

Si $E_2 = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$, alors on a

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(\{(2, 2)\}) = 1/36 \text{ et}$$

$$\mathbb{P}(E_1 \cup E_2) = 8/36 = 2/9 = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2).$$

36 paires:

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

36 sommes:

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Formules d'inclusion-exclusion pour les probabilités

Si E_1, E_2, E_3 sont des événements quelconques, on a

$$\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2),$$

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) - \mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_1 \cap E_3) - \mathbb{P}(E_2 \cap E_3) + \mathbb{P}(E_1 \cap E_2 \cap E_3),$$

et plus généralement:

Pour des événements E_1, \dots, E_n , on a toujours

$$\begin{aligned} \mathbb{P}(E_1 \cup E_2 \cup \dots \cup E_n) &= \sum_{m=1}^n (-1)^{m+1} \sum_{1 \leq j_1 < \dots < j_m \leq n} \mathbb{P}(E_{j_1} \cap E_{j_2} \cap \dots \cap E_{j_m}) \\ &= \sum_{j=1}^n \mathbb{P}(E_j) - \sum_{1 \leq j_1 < j_2 \leq n} \mathbb{P}(E_{j_1} \cap E_{j_2}) + \sum_{1 \leq j_1 < j_2 < j_3 \leq n} \mathbb{P}(E_{j_1} \cap E_{j_2} \cap E_{j_3}) \\ &\quad - \dots + (-1)^{n+1} \mathbb{P}(E_1 \cap E_2 \cap \dots \cap E_n). \end{aligned}$$

Probabilités conditionnelles

Si on lance deux dés, sachant que l'événement $E_1 = \{(1, 3), (2, 2), (3, 1)\}$ (la somme est 4) se produit, quelle est la probabilité que l'un des 2 dés donne un 3?

Probabilités conditionnelles

Si on lance deux dés, sachant que l'événement $E_1 = \{(1, 3), (2, 2), (3, 1)\}$ (la somme est 4) se produit, quelle est la probabilité que l'un des 2 dés donne un 3?

On pose $E_2 = \{(1, 3), (3, 1)\}$ et on cherche $\mathbb{P}[E_2 | E_1]$.

C'est $\mathbb{P}[\{(1, 3), (3, 1)\} | \{(1, 3), (2, 2), (3, 1)\}] = 2/3$.

Probabilités conditionnelles

Si on lance deux dés, sachant que l'événement $E_1 = \{(1, 3), (2, 2), (3, 1)\}$ (la somme est 4) se produit, quelle est la probabilité que l'un des 2 dés donne un 3?

On pose $E_2 = \{(1, 3), (3, 1)\}$ et on cherche $\mathbb{P}[E_2 | E_1]$.

C'est $\mathbb{P}[\{(1, 3), (3, 1)\} | \{(1, 3), (2, 2), (3, 1)\}] = 2/3$.

En général, pour deux événements E_1 et E_2 ,

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2 | E_1)$$

ou de manière équivalente

$$\mathbb{P}(E_2 | E_1) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_1)}.$$

Événements indépendants

Deux événements E_1 et E_2 sont dits **indépendants** lorsque

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2).$$

Cela tient si et seulement si $\mathbb{P}(E_2 \mid E_1) = \mathbb{P}(E_2)$.

Événements indépendants

Deux événements E_1 et E_2 sont dits **indépendants** lorsque

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2).$$

Cela tient si et seulement si $\mathbb{P}(E_2 \mid E_1) = \mathbb{P}(E_2)$.

Exemple: Dans l'exemple des 2 dés, si $E_1 =$ "la somme des deux dés est 6" et $E_2 =$ "le premier dé donne un 4", alors $\mathbb{P}(E_1) = 5/36$, $\mathbb{P}(E_2) = 1/6$, et $\mathbb{P}(E_1 \cap E_2) = 1/36 \neq \mathbb{P}(E_1)\mathbb{P}(E_2)$. Ces deux événements ne sont donc **pas indépendants**. La dépendance est liée au fait que si le premier dé donne 6, alors on ne peut pas avoir E_1 . Ainsi, si on sait si E_2 s'est produit ou pas, cela change la probabilité de E_1 .

Événements indépendants

Deux événements E_1 et E_2 sont dits **indépendants** lorsque

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2).$$

Cela tient si et seulement si $\mathbb{P}(E_2 | E_1) = \mathbb{P}(E_2)$.

Exemple: Dans l'exemple des 2 dés, si $E_1 =$ "la somme des deux dés est 6" et $E_2 =$ "le premier dé donne un 4", alors $\mathbb{P}(E_1) = 5/36$, $\mathbb{P}(E_2) = 1/6$, et $\mathbb{P}(E_1 \cap E_2) = 1/36 \neq \mathbb{P}(E_1)\mathbb{P}(E_2)$. Ces deux événements ne sont donc **pas indépendants**. La dépendance est liée au fait que si le premier dé donne 6, alors on ne peut pas avoir E_1 . Ainsi, si on sait si E_2 s'est produit ou pas, cela change la probabilité de E_1 .

Par contre, si $E_3 =$ "la somme des deux dés est 7", on a $\mathbb{P}(E_3) = 1/6$ et $\mathbb{P}(E_3 \cap E_2) = 1/36 = \mathbb{P}(E_3)\mathbb{P}(E_2)$, donc les deux événements sont **indépendants!** On peut vérifier aussi que $\mathbb{P}(E_3 | E_2) = \mathbb{P}(E_3)$, ce qui implique l'indépendance.

Événements indépendants

Des événements E_1, \dots, E_n sont dits (mutuellement) indépendants si pour tout sous-ensemble de $m \leq n$ indices j_1, j_2, \dots, j_m ,

$$\mathbb{P}(E_{j_1} \cap E_{j_2} \cap \dots \cap E_{j_m}) = \mathbb{P}(E_{j_1})\mathbb{P}(E_{j_2}) \cdots \mathbb{P}(E_{j_m}).$$

Ces événements sont dits indépendants par paires si pour toute paire (i, j) , on a

$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_i)\mathbb{P}(E_j).$$

Cette notion est plus faible que l'indépendance mutuelle.

Événements indépendants

Des événements E_1, \dots, E_n sont dits (mutuellement) indépendants si pour tout sous-ensemble de $m \leq n$ indices j_1, j_2, \dots, j_m ,

$$\mathbb{P}(E_{j_1} \cap E_{j_2} \cap \dots \cap E_{j_m}) = \mathbb{P}(E_{j_1})\mathbb{P}(E_{j_2}) \cdots \mathbb{P}(E_{j_m}).$$

Ces événements sont dits indépendants par paires si pour toute paire (i, j) , on a

$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_i)\mathbb{P}(E_j).$$

Cette notion est plus faible que l'indépendance mutuelle.

Supposons que l'on effectue n tirages et chaque tirage donne un succès ou un échec. Si les événements $E_j =$ "le tirage j donne un succès" sont indépendants, on dit qu'on a des tirages (ou essais) indépendants.

Formule de Bayes

Si les événements F_1, \dots, F_n forment une **partition** de Ω , i.e., $F_j \cap F_i = \emptyset$ pour tout $i \neq j$ et $F_1 \cup \dots \cup F_n = \Omega$, et si $E \in \mathcal{F}$ est un autre événement, alors

$$E = \bigcup_{i=1}^n (E \cap F_i) \quad \text{et} \quad \mathbb{P}(E) = \sum_{i=1}^n \mathbb{P}(E \cap F_i) = \sum_{i=1}^n \mathbb{P}(E | F_i) \mathbb{P}(F_i),$$

parce que les $E \cap F_i$ sont disjoints.

Ce type de **décomposition** peut souvent faciliter le calcul de $\mathbb{P}(E)$.

Supposons maintenant que l'on sait que E s'est produit, et on veut calculer la probabilité conditionnelle de l'un des F_j . On a la **formule de Bayes**:

$$\mathbb{P}(F_j | E) = \frac{\mathbb{P}(E \cap F_j)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E | F_j) \mathbb{P}(F_j)}{\sum_{i=1}^n \mathbb{P}(E | F_i) \mathbb{P}(F_i)}.$$

Cette formule est **fondamentale** en statistique et en apprentissage automatique.

Le E représente habituellement ce que l'on voit et les F_j ce que l'on ne voit pas.

Exemples

Un test médical pour une maladie donne un résultat positif avec probabilité 0.95 quand la personne a la maladie et 0.01 quand elle ne l'a pas. Si 0.5% des gens ont la maladie, quelle est la probabilité qu'une personne ayant un résultat de test positif ait la maladie?

Exemples

Un test médical pour une maladie donne un résultat positif avec probabilité 0.95 quand la personne a la maladie et 0.01 quand elle ne l'a pas. Si 0.5% des gens ont la maladie, quelle est la probabilité qu'une personne ayant un résultat de test positif ait la maladie?

On considère les événements P = "le test est positif" et M = "elle a la maladie".
La formule de Bayes donne

$$\mathbb{P}[M | P] = \frac{\mathbb{P}(M \cap P)}{\mathbb{P}(P)}$$

Exemples

Un test médical pour une maladie donne un résultat positif avec probabilité 0.95 quand la personne a la maladie et 0.01 quand elle ne l'a pas. Si 0.5% des gens ont la maladie, quelle est la probabilité qu'une personne ayant un résultat de test positif ait la maladie?

On considère les événements P = "le test est positif" et M = "elle a la maladie".
La formule de Bayes donne

$$\begin{aligned}\mathbb{P}[M | P] &= \frac{\mathbb{P}(M \cap P)}{\mathbb{P}(P)} = \frac{\mathbb{P}(P | M)\mathbb{P}(M)}{\mathbb{P}(P | M)\mathbb{P}(M) + \mathbb{P}(P | M^c)\mathbb{P}(M^c)} \\ &= \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times 0.995} = \frac{95}{294} \approx 0.323.\end{aligned}$$

La décomposition judicieuse au dénominateur permet un calcul facile!

Certains pourraient penser que la probabilité d'avoir la maladie quand le test est positif est de 99% car la probabilité d'un faux positif est de 1%, mais ce n'est pas du tout le cas!

Exemples

Un document que vous cherchez se trouve quelque part dans l'un de vos n dossiers. Soit F_i l'événement "il est dans dossier i ", pour $i = 1, \dots, n$. Ces événements sont exclusifs; ils forment une partition de $\Omega =$ "il est dans l'un des n dossiers". Soit $p_i = \mathbb{P}(F_i)$ la probabilité à priori qu'il soit dans le dossier i .

Exemples

Un document que vous cherchez se trouve quelque part dans l'un de vos n dossiers. Soit F_i l'événement "il est dans dossier i ", pour $i = 1, \dots, n$. Ces événements sont exclusifs; ils forment une partition de $\Omega =$ "il est dans l'un des n dossiers". Soit $p_i = \mathbb{P}(F_i)$ la probabilité à priori qu'il soit dans le dossier i .

Soit α_j la probabilité de trouver le document dans le dossier j par une recherche rapide s'il y est, et soit E l'événement qu'avec une **recherche rapide dans le dossier j , on ne l'y trouve pas**. Conditionnellement à E , quelle est la probabilité qu'il s'y trouve quand même? C'est

$$\mathbb{P}(F_j | E) = \frac{\mathbb{P}(E | F_j)\mathbb{P}(F_j)}{\sum_{i=1}^n \mathbb{P}(E | F_i)\mathbb{P}(F_i)} = \frac{(1 - \alpha_j)p_j}{(1 - \alpha_j)p_j + (1 - p_j)} = \frac{(1 - \alpha_j)p_j}{1 - \alpha_j p_j}.$$

Cette probabilité sera grande si p_j est proche de 1.

Espace échantillonnal infini et continu

Dans les exemples précédents, l'espace échantillonnal Ω était fini.

Mais souvent, il est infini et non dénombrable.

Les deux possibilités suivantes pour Ω (et \mathbb{P} , mais je ne définis pas \mathcal{F}) permettent de couvrir essentiellement toutes les applications. Un élément $\omega \in \Omega$ représente toute la source d'aléa dans le modèle. Tout ce qui est aléatoire peut être vu comme une fonction de ω .

Espace échantillonnal infini et continu

Dans les exemples précédents, l'espace échantillonnal Ω était fini.

Mais souvent, il est infini et non dénombrable.

Les deux possibilités suivantes pour Ω (et \mathbb{P} , mais je ne définis pas \mathcal{F}) permettent de couvrir essentiellement toutes les applications. Un élément $\omega \in \Omega$ représente toute la source d'aléa dans le modèle. Tout ce qui est aléatoire peut être vu comme une fonction de ω .

Modèle 1: Bits aléatoires indépendants. Chaque $\omega \in \Omega$ est une suite infinie de bits. Chaque bit est 1 avec probabilité 1/2 et 0 avec probabilité 1/2, indépendamment des autres.

011110100110110101001101100101000111?...

Espace échantillonnal infini et continu

Dans les exemples précédents, l'espace échantillonnal Ω était fini.

Mais souvent, il est infini et non dénombrable.

Les deux possibilités suivantes pour Ω (et \mathbb{P} , mais je ne définis pas \mathcal{F}) permettent de couvrir essentiellement toutes les applications. Un élément $\omega \in \Omega$ représente toute la source d'aléa dans le modèle. Tout ce qui est aléatoire peut être vu comme une fonction de ω .

Modèle 1: Bits aléatoires indépendants. Chaque $\omega \in \Omega$ est une suite infinie de bits. Chaque bit est 1 avec probabilité $1/2$ et 0 avec probabilité $1/2$, indépendamment des autres.

01111?100110?1?101001101100101000111...

Exemple: on 8 possibilités pour les 3 bits ? ? ?:

000, 001, 010, 011, 100, 101, 110, 111

On veut une probabilité de $1/8$ pour chacune, peu importe les autres bits.

Espace échantillonnal infini et continu

Dans les exemples précédents, l'espace échantillonnal Ω était fini.

Mais souvent, il est infini et non dénombrable.

Les deux possibilités suivantes pour Ω (et \mathbb{P} , mais je ne définis pas \mathcal{F}) permettent de couvrir essentiellement toutes les applications. Un élément $\omega \in \Omega$ représente toute la source d'aléa dans le modèle. Tout ce qui est aléatoire peut être vu comme une fonction de ω .

Modèle 1: Bits aléatoires indépendants. Chaque $\omega \in \Omega$ est une suite infinie de bits. Chaque bit est 1 avec probabilité $1/2$ et 0 avec probabilité $1/2$, indépendamment des autres.

01111?100110?1?101001101100101000111...

Exemple: on 8 possibilités pour les 3 bits ???:

000, 001, 010, 011, 100, 101, 110, 111

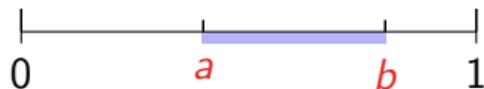
On veut une probabilité de $1/8$ pour chacune, peu importe les autres bits.

Pour s bits cachés, on veut une probabilité de $1/2^s$ pour chacune des 2^s possibilités.

Modèle 2: Nombres aléatoires uniformes sur $(0, 1)$ et indépendants.

On a une suite (finie ou infinie) de nombres réels U_0, U_1, U_2, \dots telle que:

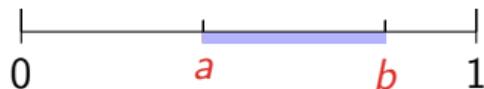
Uniformité: $\mathbb{P}[a \leq U_j \leq b] = b - a$ pour tout j et tout $0 \leq a < b \leq 1$;



Modèle 2: Nombres aléatoires uniformes sur $(0, 1)$ et indépendants.

On a une suite (finie ou infinie) de nombres réels U_0, U_1, U_2, \dots telle que:

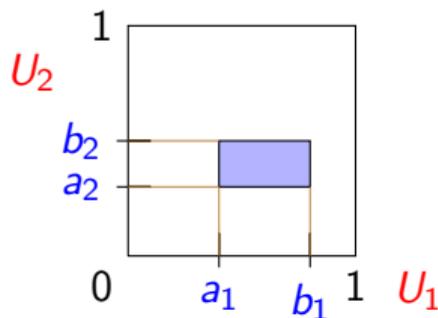
Uniformité: $\mathbb{P}[a \leq U_j \leq b] = b - a$ pour tout j et tout $0 \leq a < b \leq 1$;



Indépendance: Pour (U_1, \dots, U_s) en s dimensions, on veut

$$\mathbb{P}[a_j \leq U_j \leq b_j \text{ pour } j = 1, \dots, s] = (b_1 - a_1) \cdots (b_s - a_s).$$

Exemple pour $s = 2$:



Variables aléatoires

Habituellement, il n'est pas très important de savoir exactement comment sont construits Ω , \mathcal{F} et \mathbb{P} . On va définir des **variables aléatoires** sur l'espace de probabilité, et on va surtout s'intéresser aux **lois de probabilité** de ces variables aléatoires.

Une variable aléatoire associe à chaque $\omega \in \Omega$ un nombre réel $X = X(\omega)$.

Variables aléatoires

Habituellement, il n'est pas très important de savoir exactement comment sont construits Ω , \mathcal{F} et \mathbb{P} . On va définir des **variables aléatoires** sur l'espace de probabilité, et on va surtout s'intéresser aux **lois de probabilité** de ces variables aléatoires.

Une variable aléatoire associe à chaque $\omega \in \Omega$ un nombre réel $X = X(\omega)$.

Exemple: Dans le **modèle 2**, si $X_i = 1 + \lfloor 6U_i \rfloor$, alors $X_i \in \{1, \dots, 6\}$ peut représenter le résultat du tirage d'un dé équilibré. On obtient ainsi une suite de tirages d'un dé.

Variables aléatoires

Habituellement, il n'est pas très important de savoir exactement comment sont construits Ω , \mathcal{F} et \mathbb{P} . On va définir des **variables aléatoires** sur l'espace de probabilité, et on va surtout s'intéresser aux **lois de probabilité** de ces variables aléatoires.

Une variable aléatoire associe à chaque $\omega \in \Omega$ un nombre réel $X = X(\omega)$.

Exemple: Dans le **modèle 2**, si $X_i = 1 + \lfloor 6U_i \rfloor$, alors $X_i \in \{1, \dots, 6\}$ peut représenter le résultat du tirage d'un dé équilibré. On obtient ainsi une suite de tirages d'un dé.

Exemple: Dans le **modèle 1**, pour simuler un dé équilibré, on peut tirer trois bits et on utilise la correspondance suivante, où **R** veut dire "recommencer" (méthode de rejet):

000	001	010	011	100	101	110	111
1	2	3	4	5	6	R	R

Variables aléatoires

Habituellement, il n'est pas très important de savoir exactement comment sont construits Ω , \mathcal{F} et \mathbb{P} . On va définir des **variables aléatoires** sur l'espace de probabilité, et on va surtout s'intéresser aux **lois de probabilité** de ces variables aléatoires.

Une variable aléatoire associe à chaque $\omega \in \Omega$ un nombre réel $X = X(\omega)$.

Exemple: Dans le **modèle 2**, si $X_i = 1 + \lfloor 6U_i \rfloor$, alors $X_i \in \{1, \dots, 6\}$ peut représenter le résultat du tirage d'un dé équilibré. On obtient ainsi une suite de tirages d'un dé.

Exemple: Dans le **modèle 1**, pour simuler un dé équilibré, on peut tirer trois bits et on utilise la correspondance suivante, où **R** veut dire "recommencer" (méthode de rejet):

000	001	010	011	100	101	110	111
1	2	3	4	5	6	R	R

À chaque ω , correspond un résultat X du tirage d'un dé.

Et puisque ω est une suite infinie de bits, on peut continuer à tirer des dés et obtenir une suite illimitée (infinie) de tirages indépendants d'un dé, X_1, X_2, \dots .

Dans les deux cas, les X_i sont des **variables aléatoires**.

Variables aléatoires

Définition. Une **variable aléatoire (v.a.)** est une fonction $X : \Omega \rightarrow \mathbb{R}$ mesurable; i.e., telle que pour tout $x \in \mathbb{R}$, $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$.

La mesurabilité implique que l'on peut définir la **fonction de répartition** (“cumulative distribution function (cdf)”) de X :

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega : X(\omega) \leq x\}), \quad \text{pour } x \in \mathbb{R},$$

et la **fonction de survie (ou de fiabilité)**:

$$\bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x).$$

- Exemples:**
1. On lance 2 dés et soit X la somme des valeurs sur les 2 dés.
 2. On tire à pile ou face n fois et soit X le nombre de “pile”.
 3. Un athlète va lancer son javelot aux jeux olympiques et soit X la distance qu'il va réaliser.
 4. Soient X_1, X_2, X_3, \dots les temps d'attente des patients à l'urgence d'un hôpital.

Lois discrètes et continues

Loi discrète. X prend ses valeurs dans un ensemble dénombrable. On a $\mathbb{P}(X = x_i) = p(x_i) = p_i$ pour $i = 0, 1, 2, \dots$, et

$$F(x) = \sum_{i=0}^{\infty} p_i \mathbb{I}[x_i \leq x] = \sum_{\{i | x_i \leq x\}} p_i.$$

Lois discrètes et continues

Loi discrète. X prend ses valeurs dans un ensemble dénombrable. On a $\mathbb{P}(X = x_i) = p(x_i) = p_i$ pour $i = 0, 1, 2, \dots$, et

$$F(x) = \sum_{i=0}^{\infty} p_i \mathbb{I}[x_i \leq x] = \sum_{\{i|x_i \leq x\}} p_i.$$

Loi continue. La fonction de répartition de X est continue et s'écrit comme

$$F(x) = \int_{-\infty}^x f(y) dy$$

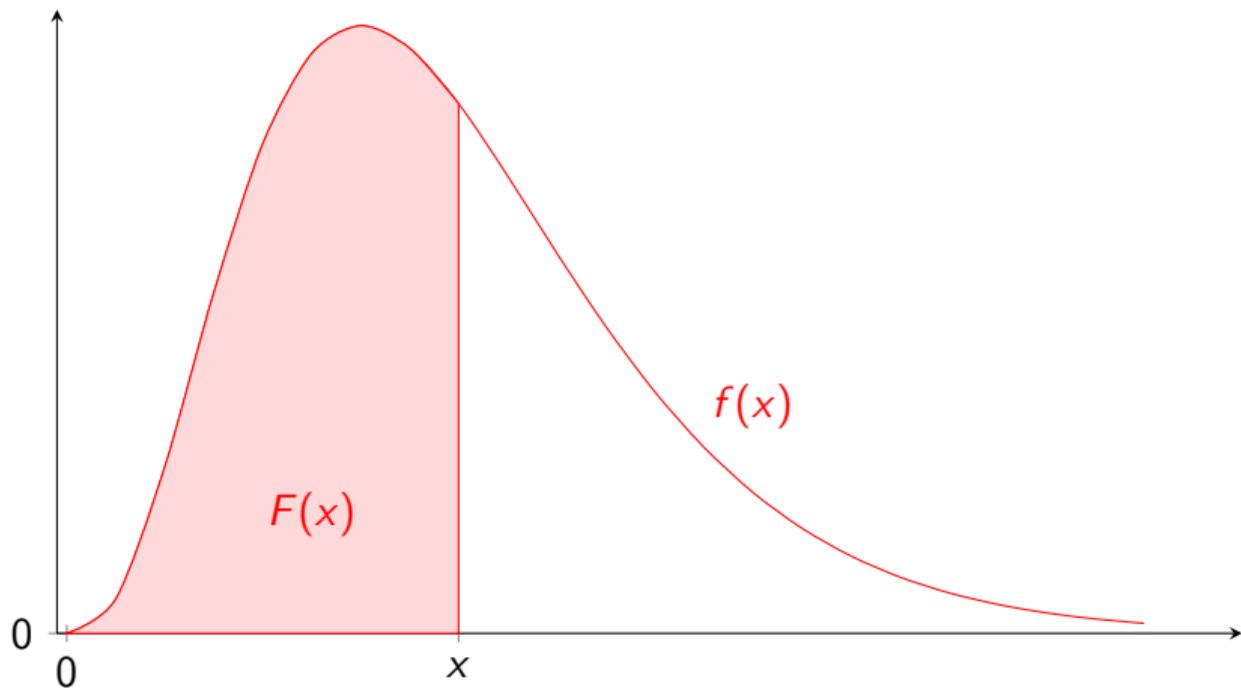
pour une fonction $f : \mathbb{R} \rightarrow [0, \infty)$ appelée la **fonction de densité** de X .

On a $f(x) = F'(x)$ quand la dérivée existe, $\int_{-\infty}^{\infty} f(y) dy = F(\infty) = 1$,

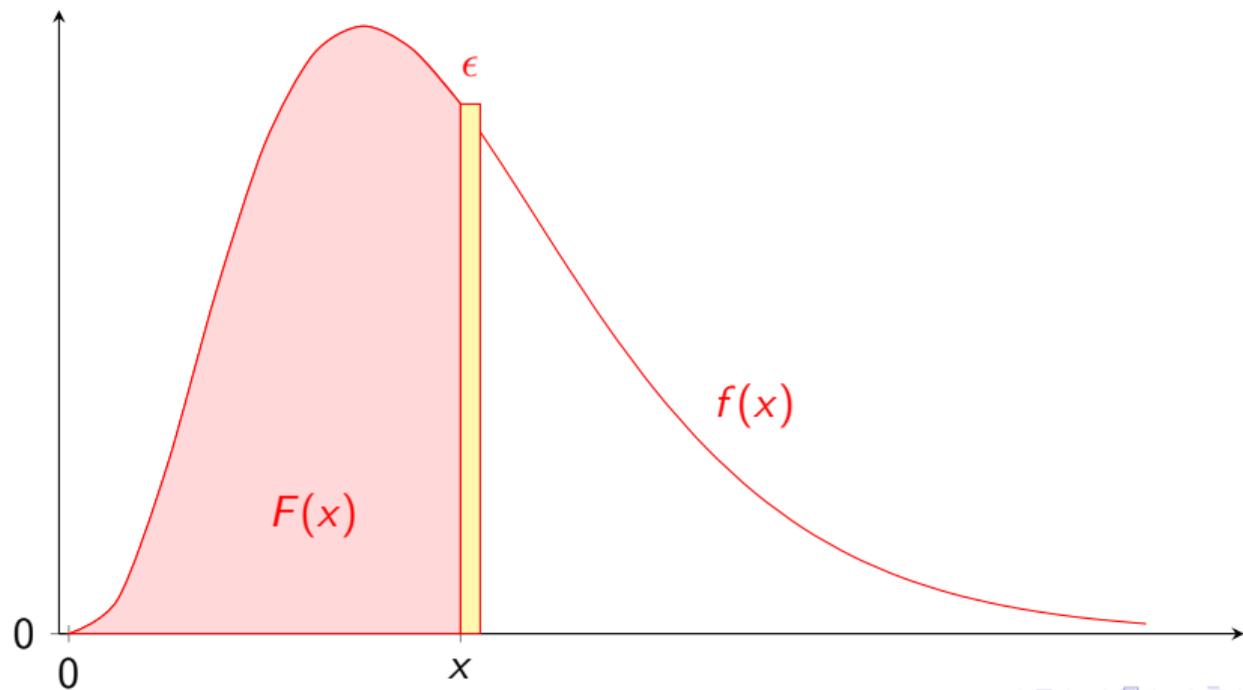
$\mathbb{P}(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(y) dy$, et $\mathbb{P}(X = x) = \int_x^x f(y) dy = 0$ pour tout $x \in \mathbb{R}$.

Note: Parfois, on a une combinaison des deux types de loi (discrète et continue).

Fonction de densité f et cdf F . On a $F(x) = \int_{-\infty}^x f(y)dy$ (surface en rose).



La surface à droite de x est $1 - F(x)$ et la surface jaune est $f(x)\epsilon$.



Taux de panne

Quand X n'est jamais négatif, on peut interpréter X comme la **durée de vie** d'un système. Utilisé en fiabilité, entretien préventif, et assurance-vie, par exemple.

Si X est continue, le **taux de panne** (ou taux de mortalité) est défini par

$$r(x) = \frac{f(x)}{1 - F(x)}.$$

Pour $\epsilon > 0$ très petit, on a

$$\mathbb{P}[X < x + \epsilon \mid X > x] = \frac{\mathbb{P}[x < X < x + \epsilon]}{\mathbb{P}[X > x]} \approx \frac{f(x)\epsilon}{1 - F(x)} = r(x)\epsilon.$$

Autrement dit, si X représente l'âge de la première panne d'un système qui est rendu à l'âge x et qui n'est pas encore tombé en panne, alors $r(x)\epsilon$ représente (approximativement) la probabilité d'une panne dans les prochains ϵ unités de temps.

Espérance mathématique

Si X est une variable aléatoire et $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction (mesurable), on définit

$$\mathbb{E}[g(X)] \stackrel{\text{def}}{=} \begin{cases} \sum_{i=0}^{\infty} g(x_i) p_i & \text{si } X \text{ est discrète;} \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{si } X \text{ est continue.} \end{cases}$$

En particulier, $\mathbb{E}[X^r]$ est le **moment d'ordre r** de X ,

$\mu = \mathbb{E}[X]$ est l'espérance de X (la **moyenne**),

$\sigma^2 = \text{Var}[X] = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mu^2$ est la **variance**, et σ est l'**écart-type**.

Le **coefficient de variation** (CV) est σ/μ .

Espérance mathématique

Si X est une variable aléatoire et $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction (mesurable), on définit

$$\mathbb{E}[g(X)] \stackrel{\text{def}}{=} \begin{cases} \sum_{i=0}^{\infty} g(x_i) p_i & \text{si } X \text{ est discrète;} \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{si } X \text{ est continue.} \end{cases}$$

En particulier, $\mathbb{E}[X^r]$ est le **moment d'ordre r** de X ,

$\mu = \mathbb{E}[X]$ est l'espérance de X (la **moyenne**),

$\sigma^2 = \text{Var}[X] = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mu^2$ est la **variance**, et σ est l'**écart-type**.

Le **coefficient de variation** (CV) est σ/μ .

En général **on n'a pas** $\mathbb{E}[g(X)] = g(\mathbb{E}(X))$, sauf si g est linéaire. Exemple: Si $\mu = 0$, $\sigma^2 > 0$ et $g(x) = x^2$, alors $g(\mathbb{E}(X)) = g(0) = 0$ mais $\mathbb{E}(g(X)) = \mathbb{E}(X^2) > 0$.

Si $Y = g(X) = aX + b$, alors $\mathbb{E}[Y] = \mathbb{E}[aX + b] = a\mathbb{E}[X] + b$

et $\text{Var}[Y] = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[(aX + b - a\mu - b)^2] = a^2\mathbb{E}[(X - \mu)^2] = a^2\text{Var}[X]$.

Autre expression utile pour l'espérance de X

Proposition. Si $\mathbb{P}(X \geq 0) = 1$, alors on a (les deux intégrales sont toujours égales):

$$\mathbb{E}[X] = \int_0^{\infty} \bar{F}(x) dx = \int_0^{\infty} \mathbb{P}[X > x] dx.$$

Preuve. Puisque $\int_0^{\infty} \mathbb{I}[X > x] dx = \int_0^X dx = X$, on a

$$\mathbb{E}[X] = \mathbb{E} \left[\int_0^{\infty} \mathbb{I}[X > x] dx \right] = \int_0^{\infty} \mathbb{E}[\mathbb{I}[X > x]] dx = \int_0^{\infty} \mathbb{P}(X > x) dx. \quad \square$$

Autre expression utile pour l'espérance de X

Proposition. Si $\mathbb{P}(X \geq 0) = 1$, alors on a (les deux intégrales sont toujours égales):

$$\mathbb{E}[X] = \int_0^{\infty} \bar{F}(x) dx = \int_0^{\infty} \mathbb{P}[X > x] dx.$$

Preuve. Puisque $\int_0^{\infty} \mathbb{I}[X > x] dx = \int_0^X dx = X$, on a

$$\mathbb{E}[X] = \mathbb{E} \left[\int_0^{\infty} \mathbb{I}[X > x] dx \right] = \int_0^{\infty} \mathbb{E}[\mathbb{I}[X > x]] dx = \int_0^{\infty} \mathbb{P}(X > x) dx. \quad \square$$

Plus généralement, si X peut aussi prendre des valeurs négatives, on a

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-] = \int_0^{\infty} \bar{F}(x) dx - \int_{-\infty}^0 F(x) dx.$$

Quelques lois de probabilité discrètes:

Loi binomiale(n, p)

Soit X le nombre de succès dans une **expérience de Bernoulli**, où on fait n tirages (ou essais) indépendants et la probabilité de succès est p pour chaque essai. On a

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad \text{pour } i = 0, \dots, n,$$

Quelques lois de probabilité discrètes:

Loi binomiale(n, p)

Soit X le nombre de succès dans une **expérience de Bernoulli**, où on fait n tirages (ou essais) indépendants et la probabilité de succès est p pour chaque essai. On a

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad \text{pour } i = 0, \dots, n,$$

D'où vient cette formule? Si on fixe la position des i succès, ce qui correspond à choisir une chaîne de n bits contenant i uns et $n - i$ zéros, la probabilité d'avoir cette chaîne particulière est $p^i (1 - p)^{n-i}$. Mais le nombre de ces chaînes qui donnent $X = i$ est le nombre de façons de choisir les positions des i bits qui sont à 1, et ce nombre est $\binom{n}{i}$.

Quelques lois de probabilité discrètes:

Loi binomiale(n, p)

Soit X le nombre de succès dans une **expérience de Bernoulli**, où on fait n tirages (ou essais) indépendants et la probabilité de succès est p pour chaque essai. On a

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad \text{pour } i = 0, \dots, n,$$

D'où vient cette formule? Si on fixe la position des i succès, ce qui correspond à choisir une chaîne de n bits contenant i uns et $n - i$ zéros, la probabilité d'avoir cette chaîne particulière est $p^i (1 - p)^{n-i}$. Mais le nombre de ces chaînes qui donnent $X = i$ est le nombre de façons de choisir les positions des i bits qui sont à 1, et ce nombre est $\binom{n}{i}$.

On note $X \sim \text{Binomiale}(n, p)$. On a $\mathbb{E}[X] = np$ et $\text{Var}[X] = np(1 - p)$.

Quelques lois de probabilité discrètes:

Loi binomiale(n, p)

Soit X le nombre de succès dans une **expérience de Bernoulli**, où on fait n tirages (ou essais) indépendants et la probabilité de succès est p pour chaque essai. On a

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad \text{pour } i = 0, \dots, n,$$

D'où vient cette formule? Si on fixe la position des i succès, ce qui correspond à choisir une chaîne de n bits contenant i uns et $n - i$ zéros, la probabilité d'avoir cette chaîne particulière est $p^i (1 - p)^{n-i}$. Mais le nombre de ces chaînes qui donnent $X = i$ est le nombre de façons de choisir les positions des i bits qui sont à 1, et ce nombre est $\binom{n}{i}$.

On note $X \sim \text{Binomiale}(n, p)$. On a $\mathbb{E}[X] = np$ et $\text{Var}[X] = np(1 - p)$.

Quand $n \rightarrow \infty$ avec $np = \lambda$ fixé, on a $\mathbb{P}(X = i) \rightarrow \lambda^i e^{-\lambda} / i!$.

Quand $n \rightarrow \infty$ pour $0 < p < 1$ fixé, $(X - np) / \sqrt{np(1 - p)} \Rightarrow \text{Normale}(0, 1)$.

Récurrances pour calculer ces probabilités quand n est grand:

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}.$$

On a $\mathbb{P}(X = 0) = (1-p)^n$ et

$$\mathbb{P}(X = i) = \mathbb{P}(X = i-1) \frac{(n-i)}{i} \frac{p}{(1-p)}.$$

Ou encore à reculons: $\mathbb{P}(X = n) = p^n$ et

$$\mathbb{P}(X = i) = \mathbb{P}(X = i+1) \frac{(i+1)}{(n-i)} \frac{(1-p)}{p}.$$

Exemple: moteurs d'avions.

On a des avions à 2 et 4 moteurs pour lesquels chaque moteur va fonctionner correctement lors d'un vol avec probabilité p , indépendamment des autres moteurs. Donc si l'avion a n moteurs, le nombre X de moteurs fonctionnels suit la loi binomiale(n, p).

Est-ce que la probabilité $p_{\text{ok}}(n)$ qu'au moins la moitié des moteurs fonctionnent est plus grande avec $n = 2$ moteurs ou avec $n = 4$ moteurs?

En fait, cela dépend de p . Voyons voir.

Exemple: moteurs d'avions.

On a des avions à 2 et 4 moteurs pour lesquels chaque moteur va fonctionner correctement lors d'un vol avec probabilité p , indépendamment des autres moteurs. Donc si l'avion a n moteurs, le nombre X de moteurs fonctionnels suit la loi binomiale (n, p) .

Est-ce que la probabilité $p_{\text{ok}}(n)$ qu'au moins la moitié des moteurs fonctionnent est plus grande avec $n = 2$ moteurs ou avec $n = 4$ moteurs?

En fait, cela dépend de p . Voyons voir. On a

$$p_{\text{ok}}(4) = \mathbb{P}[X \geq 2 \mid n = 4] = \sum_{i=2}^4 \binom{4}{i} p^i (1-p)^{4-i} = 6p^2(1-p)^2 + 4p^3(1-p) + p^4,$$

$$p_{\text{ok}}(2) = \mathbb{P}[X \geq 1 \mid n = 2] = \sum_{i=1}^2 \binom{2}{i} p^i (1-p)^{2-i} = 2p(1-p) + p^2.$$

Exemple: moteurs d'avions.

On a des avions à 2 et 4 moteurs pour lesquels chaque moteur va fonctionner correctement lors d'un vol avec probabilité p , indépendamment des autres moteurs. Donc si l'avion a n moteurs, le nombre X de moteurs fonctionnels suit la loi binomiale (n, p) .

Est-ce que la probabilité $p_{\text{ok}}(n)$ qu'au moins la moitié des moteurs fonctionnent est plus grande avec $n = 2$ moteurs ou avec $n = 4$ moteurs?

En fait, cela dépend de p . Voyons voir. On a

$$p_{\text{ok}}(4) = \mathbb{P}[X \geq 2 \mid n = 4] = \sum_{i=2}^4 \binom{4}{i} p^i (1-p)^{4-i} = 6p^2(1-p)^2 + 4p^3(1-p) + p^4,$$

$$p_{\text{ok}}(2) = \mathbb{P}[X \geq 1 \mid n = 2] = \sum_{i=1}^2 \binom{2}{i} p^i (1-p)^{2-i} = 2p(1-p) + p^2.$$

L'avion à 4 moteurs est supérieur ssi $6p^2(1-p)^2 + 4p^3(1-p) + p^4 > 2p(1-p) + p^2$,
ssi $(1-p)^2(3p-2) > 0$, ssi $p > 2/3$.

Si les moteurs ne sont pas fiables, mieux vaut prendre seulement 2 moteurs!

Loi géométrique(p)

Encore des essais indépendants avec probabilité de succès p .

Soit X le nombre d'échecs **avant** le premier succès. On a

$$\mathbb{P}(X = i) = p(1 - p)^i \quad \text{pour } i = 0, 1, \dots$$

Attention: Ross (2014) compte le succès, i.e., il compte $X + 1$, donc sa définition est différente. Les deux définitions sont populaires dans la littérature.

Cette loi est **sans mémoire**: pour $i \geq 0$,

$$\mathbb{P}[X = y + i \mid X \geq y] = \frac{\mathbb{P}[X = y + i]}{\mathbb{P}[X \geq y]} = \frac{p(1 - p)^{y+i}}{(1 - p)^y} = \mathbb{P}[X = i].$$

On a $\mathbb{E}[X] = (1 - p)/p$ et $\text{Var}[X] = (1 - p)/p^2$.

Loi binomiale négative (n, p)

Soient $n \in (0, \infty)$ et $0 < p < 1$. On pose

$$\mathbb{P}(X = i) = \frac{\Gamma(n+i)}{\Gamma(n)\Gamma(i+1)} p^n (1-p)^i \quad \text{pour } i = 0, 1, \dots$$

où

$$\Gamma(y) = \int_0^{\infty} x^{y-1} e^{-x} dx \quad \text{pour } y > 0.$$

On a $\mathbb{E}[X] = n(1-p)/p$ et $\text{Var}[X] = n(1-p)/p^2$.

Note: si y est un entier positif, $\Gamma(y) = (y-1)!$.

Pour n entier: X est le nombre d'échecs avant d'avoir n succès. On a alors

$$\frac{\Gamma(n+i)}{\Gamma(n)\Gamma(i+1)} = \frac{(n+i-1)!}{(n-1)!i!} = \binom{n+i-1}{i}.$$

Pour $n = 1$, c'est la loi géométrique.

Loi de Poisson(λ)

$$\mathbb{P}(X = i) = \frac{\lambda^i e^{-\lambda}}{i!} \quad \text{pour } i = 0, 1, 2, \dots$$

On note $X \sim \text{Poisson}(\lambda)$. On a $\mathbb{E}[X] = \text{Var}[X] = \lambda$.

Loi de Poisson(λ)

$$\mathbb{P}(X = i) = \frac{\lambda^i e^{-\lambda}}{i!} \quad \text{pour } i = 0, 1, 2, \dots$$

On note $X \sim \text{Poisson}(\lambda)$. On a $\mathbb{E}[X] = \text{Var}[X] = \lambda$.

Qu'est-ce qui motive la loi de Poisson? Si $X \sim \text{Binomiale}(n, p)$ et $(n, np) \rightarrow (\infty, \lambda)$, alors pour tout i fixé, $\mathbb{P}(X = i) \rightarrow \lambda^i e^{-\lambda} / i!$. En effet, quand $n \rightarrow \infty$:

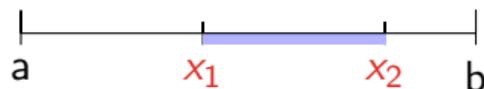
$$\begin{aligned} \mathbb{P}(X = i) &= \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \rightarrow \frac{n!}{(n-i)!i!} (\lambda/n)^i (1-\lambda/n)^{n-i} \\ &\rightarrow \frac{n!}{(n-i)!n^i} \frac{\lambda^i (1-\lambda/n)^n}{i! (1-\lambda/n)^i} \rightarrow \frac{\lambda^i}{i!} (1-\lambda/n)^n \rightarrow \frac{\lambda^i}{i!} e^{-\lambda}. \end{aligned}$$

Exemples: nombre d'erreurs typographiques dans un livre, nombres d'accidents sur une route, nombre d'appels aux pompiers demain, nombres de voitures vendues par un détaillant dans le mois, nombre de personnes achetant des actions de Bombardier demain, etc.

Quelques lois continues: la loi uniforme

$X \sim \text{Uniforme}(a, b)$ a la densité

$$f(x) = 1/(b - a) \quad \text{pour } a < x < b.$$



$$\mathbb{P}[x_1 \leq X \leq x_2] = (x_2 - x_1)/(b - a).$$

On a $\mathbb{E}[X] = (a + b)/2$ et $\text{Var}[X] = (b - a)^2/12$.

La loi exponentielle

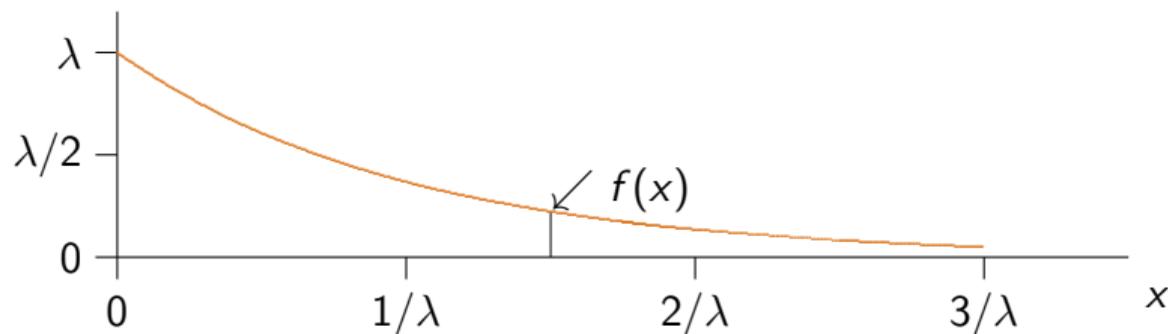
X suit la loi exponentielle de taux $\lambda > 0$, noté $X \sim \text{Expon}(\lambda)$, si

$$f(x) = \lambda e^{-\lambda x} \quad \text{pour } x > 0,$$

$$F(x) = 1 - e^{-\lambda x},$$

$$r(x) = \lambda.$$

$\mathbb{E}[X] = 1/\lambda$ et $\text{Var}[X] = 1/\lambda^2$. $\text{CV} = 1$.



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

$$\mathbb{P}[X > t + x \mid X > t] = \frac{\mathbb{P}[X > t + x]}{\mathbb{P}[X > t]}$$

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

$$\mathbb{P}[X > t + x \mid X > t] = \frac{\mathbb{P}[X > t + x]}{\mathbb{P}[X > t]} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}}$$

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

$$\mathbb{P}[X > t + x \mid X > t] = \frac{\mathbb{P}[X > t + x]}{\mathbb{P}[X > t]} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x} = \mathbb{P}[X > x].$$

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

$$\mathbb{P}[X > t + x \mid X > t] = \frac{\mathbb{P}[X > t + x]}{\mathbb{P}[X > t]} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x} = \mathbb{P}[X > x].$$

Cette propriété simplifie beaucoup l'analyse mathématique et l'explique la grande popularité de la loi exponentielle. On va l'exploiter en particulier dans la définition des processus de Poisson et des chaînes de Markov en temps continu.

La loi gamma

$X \sim \text{Gamma}(\alpha, \lambda)$ si

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad \text{pour } x > 0,$$

où

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad \text{pour } \alpha > 0$$

On a $\mathbb{E}[X] = \alpha/\lambda$ et $\text{Var}[X] = \alpha/\lambda^2$. Souvent utilisé pour modéliser des **durées d'activités**.

La loi gamma

$X \sim \text{Gamma}(\alpha, \lambda)$ si

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad \text{pour } x > 0,$$

où

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad \text{pour } \alpha > 0$$

On a $\mathbb{E}[X] = \alpha/\lambda$ et $\text{Var}[X] = \alpha/\lambda^2$. Souvent utilisé pour modéliser des **durées d'activités**.

Pour $\alpha = 1$: c'est la loi exponentielle.

La loi gamma

$X \sim \text{Gamma}(\alpha, \lambda)$ si

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad \text{pour } x > 0,$$

où

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad \text{pour } \alpha > 0$$

On a $\mathbb{E}[X] = \alpha/\lambda$ et $\text{Var}[X] = \alpha/\lambda^2$. Souvent utilisé pour modéliser des **durées d'activités**.

Pour $\alpha = 1$: c'est la loi exponentielle.

Pour 2α entier et $\lambda = 1/2$: c'est la loi du **chi-deux**.

La loi gamma

$X \sim \text{Gamma}(\alpha, \lambda)$ si

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad \text{pour } x > 0,$$

où

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad \text{pour } \alpha > 0$$

On a $\mathbb{E}[X] = \alpha/\lambda$ et $\text{Var}[X] = \alpha/\lambda^2$. Souvent utilisé pour modéliser des **durées d'activités**.

Pour $\alpha = 1$: c'est la loi exponentielle.

Pour 2α entier et $\lambda = 1/2$: c'est la loi du **chi-deux**.

Pour $\alpha = k$ entier, $X \sim \text{Gamma}(\alpha, \lambda)$ équivaut à $X \sim \text{Erlang}(k, \lambda)$,

i.e., $X = X_1 + \dots + X_k$ est une somme de k exponentielles(λ) indépendantes.

La loi gamma

$X \sim \text{Gamma}(\alpha, \lambda)$ si

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad \text{pour } x > 0,$$

où

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad \text{pour } \alpha > 0$$

On a $\mathbb{E}[X] = \alpha/\lambda$ et $\text{Var}[X] = \alpha/\lambda^2$. Souvent utilisé pour modéliser des **durées d'activités**.

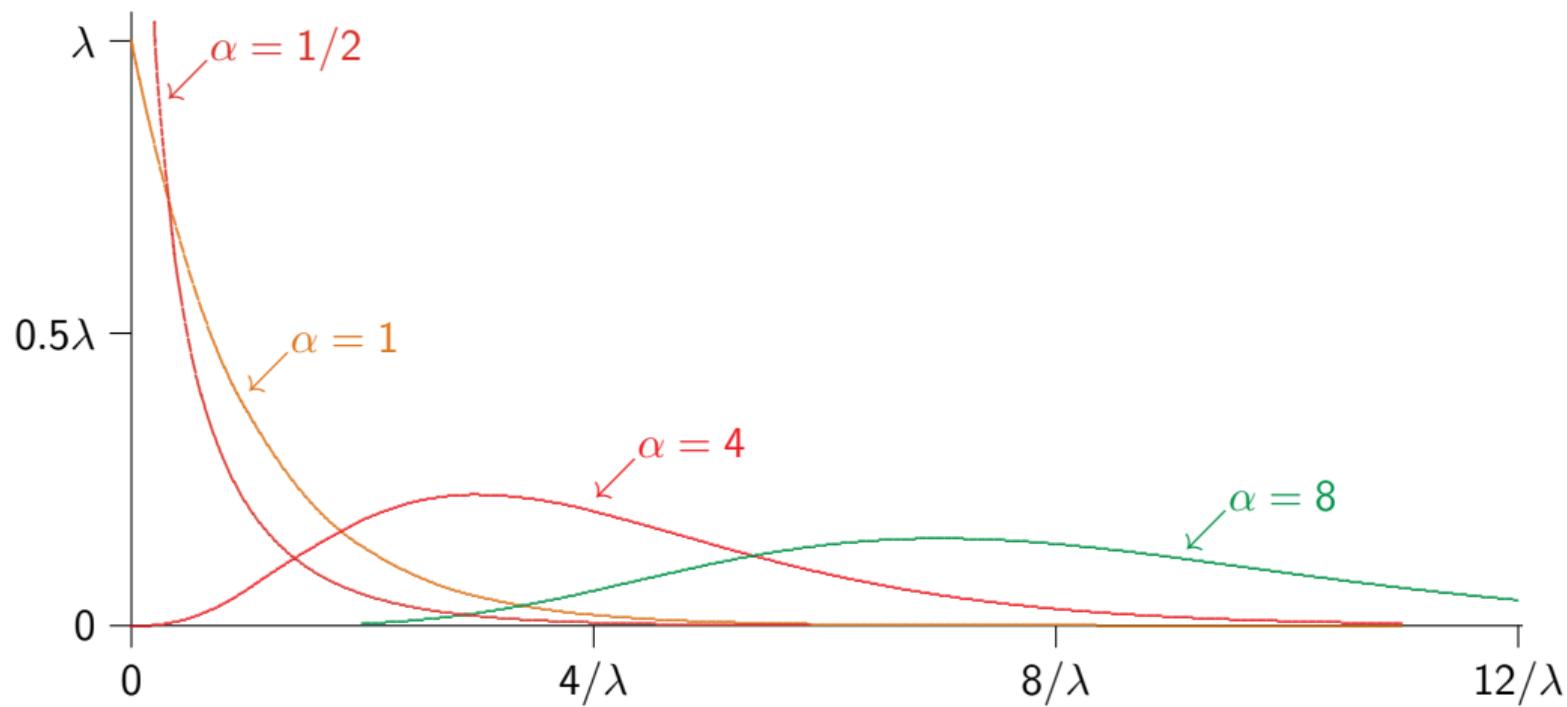
Pour $\alpha = 1$: c'est la loi exponentielle.

Pour 2α entier et $\lambda = 1/2$: c'est la loi du **chi-deux**.

Pour $\alpha = k$ entier, $X \sim \text{Gamma}(\alpha, \lambda)$ équivaut à $X \sim \text{Erlang}(k, \lambda)$,
i.e., $X = X_1 + \dots + X_k$ est une somme de k exponentielles(λ) indépendantes.

Si $X \sim \text{Erlang}(k, \lambda)$ et $Y \sim \text{Poisson}(\lambda x)$, alors $\mathbb{P}[X \leq x] = \mathbb{P}[Y \geq k]$.

La loi gamma



La loi beta

C'est une loi continue sur l'intervalle $(0, 1)$. On a $X \sim \text{Beta}(\alpha, \beta)$ si

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad \text{pour } 0 < x < 1.$$

On a $\mathbb{E}[X] = \alpha/(\alpha + \beta)$.

On peut ajouter des paramètres d'échelle et de localisation: $Y = a + (b - a)X$.
Donne une densité très flexible sur un intervalle borné $[a, b]$.

La loi beta

C'est une loi continue sur l'intervalle $(0, 1)$. On a $X \sim \text{Beta}(\alpha, \beta)$ si

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad \text{pour } 0 < x < 1.$$

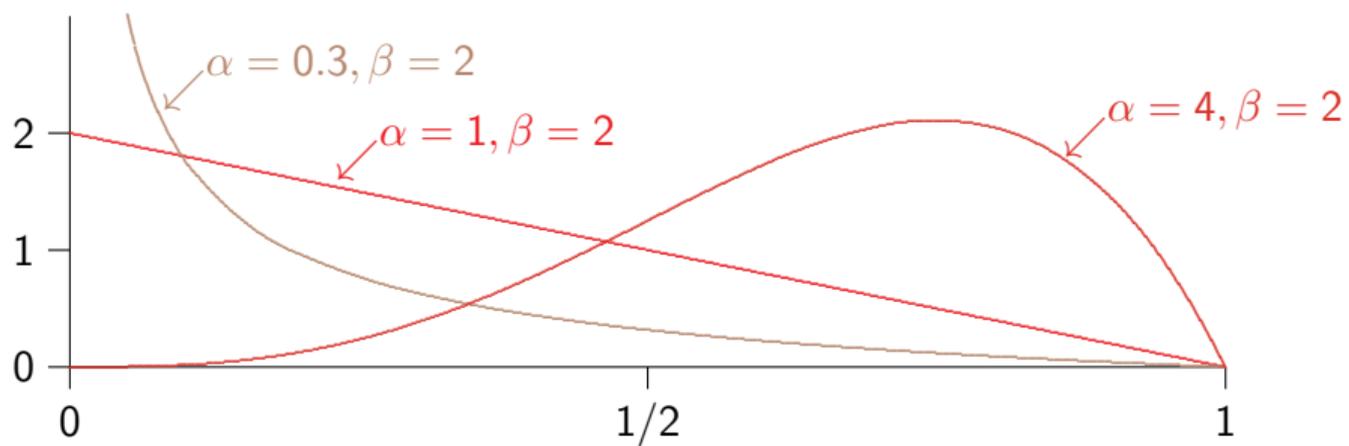
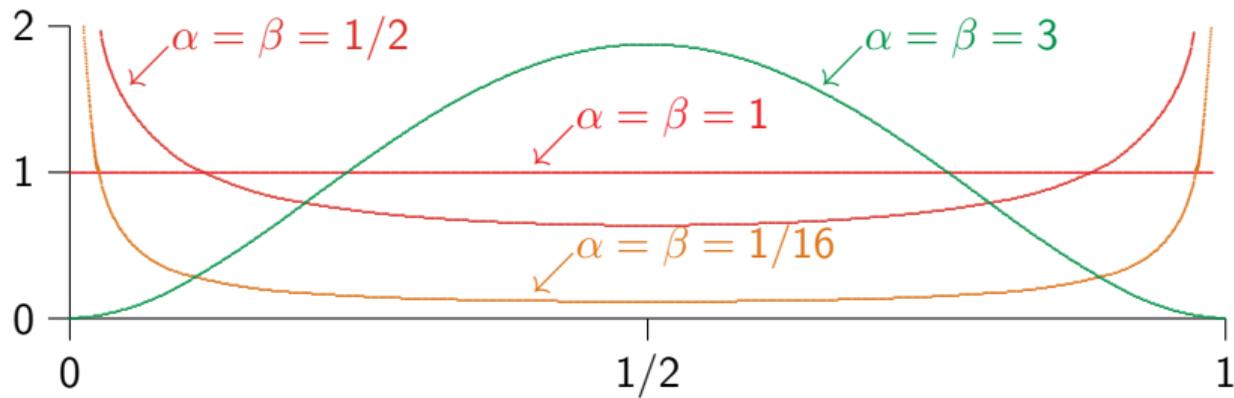
On a $\mathbb{E}[X] = \alpha/(\alpha + \beta)$.

On peut ajouter des paramètres d'échelle et de localisation: $Y = a + (b - a)X$.
Donne une densité très flexible sur un intervalle borné $[a, b]$.

Si α et β sont des entiers positifs, on a

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{(\alpha + \beta - 1)!}{(\alpha - 1)!(\beta - 1)!} = \binom{\alpha + \beta - 2}{\alpha - 1} (\alpha + \beta - 1).$$

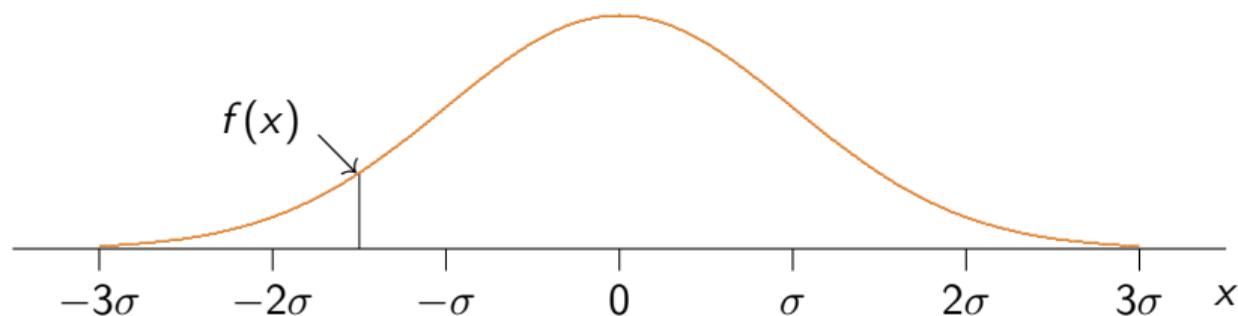
Pour $\alpha = \beta = 1$, on a la **loi uniforme** sur $(0, 1)$.



La loi normale

$X \sim \text{Normale}(\mu, \sigma^2)$ a la densité

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\} \quad \text{pour } x \in \mathbb{R}.$$



Si $X \sim \text{Normale}(\mu, \sigma^2)$, alors $Y = aX + b \sim \text{Normale}(a\mu + b, a^2\sigma^2)$.

En particulier, $Y = (X - \mu)/\sigma \sim \text{Normale}(0, 1)$.

Si X_1, \dots, X_k sont indépendantes, $X_i \sim \text{Normale}(\mu_i, \sigma_i^2)$, et $Y = X_1 + \dots + X_k$, alors $Y \sim \text{Normale}(\mu, \sigma^2)$ où $\mu = \mu_1 + \dots + \mu_k$ et $\sigma^2 = \sigma_1^2 + \dots + \sigma_k^2$.

Fonction génératrice

La **fonction génératrice des moments** d'une variable aléatoire X est définie par

$$\phi(t) = M_X(t) = \mathbb{E} \left[e^{tX} \right] \quad \text{pour tout } t \in \mathbb{R},$$

lorsque cette espérance existe. Quand $M_X(t)$ existe dans un petit voisinage de $t = 0$, elle contient dans ce voisinage suffisamment d'information pour déterminer toute la distribution de X de manière unique!

On obtient le moment d'ordre j de X en évaluant la j -ième dérivée de $M_X(t)$ à $t = 0$:

$$\mathbb{E}[X^j] = \left. \frac{d^j M_X(t)}{dt^j} \right|_{t=0}.$$

Tous ces moments existent ssi $M_X(t)$ existe dans un intervalle ouvert contenant 0.

$M_X(-t)$ est la transformée de Laplace de la fonction de densité (ou de masse) de X .

Exemple d'utilisation de la fonction génératrice

Si $X \sim \text{Poisson}(\lambda)$, alors

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{i=0}^{\infty} e^{ti} \frac{e^{-\lambda} \lambda^i}{i!} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}.$$

En dérivant, on obtient

$$M'_X(t) = \lambda e^t \exp[\lambda(e^t - 1)],$$

$$M''_X(t) = (\lambda e^t)^2 \exp[\lambda(e^t - 1)] + \lambda e^t \exp[\lambda(e^t - 1)],$$

ce qui donne

$$\mathbb{E}[X] = M'_X(0) = \lambda,$$

$$\mathbb{E}[X^2] = M''_X(0) = \lambda^2 + \lambda,$$

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \lambda.$$

Vecteurs aléatoires, loi conjointe, et indépendance

Si des v.a. X_1, \dots, X_d sont définies sur un même espace de probabilité, la **fonction de répartition (conjointe)** (“joint cdf”) $F : \mathbb{R}^d \rightarrow [0, 1]$ du **vecteur aléatoire** (transposé) $\mathbf{X} = (X_1, \dots, X_d)^t$ est définie par

$$F(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$$

pour tout $\mathbf{x} = (x_1, \dots, x_d)^t \in \mathbb{R}^d$.

La cdf **marginale** de X_j est F_j , définie par $F_j(x_j) = F(\infty, \dots, \infty, x_j, \infty, \dots, \infty) = \mathbb{P}[X_j \leq x_j]$.

Vecteurs aléatoires, loi conjointe, et indépendance

Si des v.a. X_1, \dots, X_d sont définies sur un même espace de probabilité, la **fonction de répartition (conjointe)** ("joint cdf") $F : \mathbb{R}^d \rightarrow [0, 1]$ du **vecteur aléatoire** (transposé) $\mathbf{X} = (X_1, \dots, X_d)^t$ est définie par

$$F(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$$

pour tout $\mathbf{x} = (x_1, \dots, x_d)^t \in \mathbb{R}^d$.

La cdf **marginale** de X_j est F_j , définie par $F_j(x_j) = F(\infty, \dots, \infty, x_j, \infty, \dots, \infty) = \mathbb{P}[X_j \leq x_j]$.

On dit que les v.a. X_1, \dots, X_d sont **mutuellement indépendantes** lorsque les événements $\{X_j \leq x_j\}$ sont indépendants, i.e., si $F(x_1, \dots, x_d) = F_1(x_1) \cdots F_d(x_d)$, pour tout $(x_1, \dots, x_d) \in \mathbb{R}^d$.

Elles sont dites **indépendantes par paires** si $F(x_i, x_j) = F_i(x_i)F_j(x_j)$ pour toute paire $i \neq j$ et pour tout $(x_i, x_j) \in \mathbb{R}^2$.

Loi discrète multivariée

Si X_1, \dots, X_d sont des v.a. **discrètes**, leur **fonction de masse de probabilité conjointes** p est définie par

$$p(x_1, \dots, x_d) = \mathbb{P}(X_1 = x_1, \dots, X_d = x_d)$$

pour toutes les valeurs x_1, \dots, x_d que ces variables peuvent prendre. La fonction de masse **marginale** de X_j est

$$p_j(x_j) = \mathbb{P}(X_j = x_j).$$

Les variables X_1, \dots, X_d sont mutuellement **indépendantes** ssi

$$p(x_1, \dots, x_d) = p_1(x_1) \cdots p_d(x_d)$$

pour toutes valeurs x_1, \dots, x_d .

Loi continue multivariée

Si X_1, \dots, X_d sont continues, leur fonction de densité conjointe f satisfait

$$F(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(y_1, \dots, y_d) dy_1 \cdots dy_d$$

et la densité marginale de X_j , notée f_j , satisfait

$$F_j(x) = \int_{-\infty}^x f_j(y) dy.$$

Les variables X_1, \dots, X_d sont (mutuellement) indépendantes ssi

$$f(x_1, \dots, x_d) = f_1(x_1) \cdots f_d(x_d)$$

pour tout $(x_1, \dots, x_d) \in \mathbb{R}^d$.

Si $Y = g(X_1, \dots, X_d)$ où $g : \mathbb{R}^d \rightarrow \mathbb{R}$ est mesurable, alors Y est une variable aléatoire. Dans le cas où X_1, \dots, X_d sont **continues**, on a

$$\mathbb{E}[Y] = \mathbb{E}[g(X_1, \dots, X_d)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_d) f(x_1, \dots, x_d) dx_1 \cdots dx_d.$$

Dans le cas **discret**, les intégrales sont remplacées par des sommes.

Si $Y = g(X_1, \dots, X_d)$ où $g : \mathbb{R}^d \rightarrow \mathbb{R}$ est mesurable, alors Y est une variable aléatoire. Dans le cas où X_1, \dots, X_d sont **continues**, on a

$$\mathbb{E}[Y] = \mathbb{E}[g(X_1, \dots, X_d)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_d) f(x_1, \dots, x_d) dx_1 \cdots dx_d.$$

Dans le cas **discret**, les intégrales sont remplacées par des sommes.

Si X_1, \dots, X_d sont mutuellement indépendantes, alors pour toutes fonctions mesurables g_1, \dots, g_d , où $g_j : \mathbb{R} \rightarrow \mathbb{R}$, $g_1(X_1), \dots, g_d(X_d)$ le sont aussi, et par conséquent on a

$$\mathbb{E}[g_1(X_1) \cdots g_d(X_d)] = \mathbb{E}[g_1(X_1)] \cdots \mathbb{E}[g_d(X_d)].$$

Trouver la loi d'une somme avec la fonction génératrice

Si X_1, \dots, X_d sont des v.a. indépendantes et $Y = X_1 + \dots + X_d$, alors

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(X_1 + \dots + X_d)}] = \mathbb{E}[e^{tX_1}] \dots \mathbb{E}[e^{tX_d}] = \prod_{j=1}^d M_{X_j}(t).$$

Trouver la loi d'une somme avec la fonction génératrice

Si X_1, \dots, X_d sont des v.a. indépendantes et $Y = X_1 + \dots + X_d$, alors

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(X_1 + \dots + X_d)}] = \mathbb{E}[e^{tX_1}] \dots \mathbb{E}[e^{tX_d}] = \prod_{j=1}^d M_{X_j}(t).$$

Exemple. Supposons que X_1, \dots, X_d sont indépendantes avec $X_j \sim \text{Poisson}(\lambda_j)$, $Y = X_1 + \dots + X_d$, et $\lambda = \lambda_1 + \dots + \lambda_d$, alors

$$M_Y(t) = \prod_{j=1}^d M_{X_j}(t) = \prod_{j=1}^d e^{\lambda_j(e^t - 1)} = e^{\lambda(e^t - 1)}.$$

Donc $Y \sim \text{Poisson}(\lambda)$ (exactement).

Exemple: approximation par la loi de Poisson

Soient X_1, \dots, X_n indépendantes avec $X_j \sim \text{Binomiale}(1, p_j)$, où chaque p_j est très petit et n est grand. On peut interpréter $\{X_j = 1\}$ comme un événement rare.

$Y = X_1 + \dots + X_n$ est le nombre d'événements rares qui se produisent.

Example: approximation par la loi de Poisson

Soient X_1, \dots, X_n indépendantes avec $X_j \sim \text{Binomiale}(1, p_j)$, où chaque p_j est très petit et n est grand. On peut interpréter $\{X_j = 1\}$ comme un événement rare.

$Y = X_1 + \dots + X_n$ est le nombre d'événements rares qui se produisent.

La fonction génératrice de X_j est

$$M_{X_j}(t) = \mathbb{E}[e^{tX_j}] = p_j e^t + 1 - p_j = 1 + p_j(e^t - 1) \approx \exp[p_j(e^t - 1)].$$

quand t est proche de 0, car $1 + x \approx e^x$ quand x est proche de 0.

Example: approximation par la loi de Poisson

Soient X_1, \dots, X_n indépendantes avec $X_j \sim \text{Binomiale}(1, p_j)$, où chaque p_j est très petit et n est grand. On peut interpréter $\{X_j = 1\}$ comme un événement rare.

$Y = X_1 + \dots + X_n$ est le nombre d'événements rares qui se produisent.

La fonction génératrice de X_j est

$$M_{X_j}(t) = \mathbb{E}[e^{tX_j}] = p_j e^t + 1 - p_j = 1 + p_j(e^t - 1) \approx \exp[p_j(e^t - 1)].$$

quand t est proche de 0, car $1 + x \approx e^x$ quand x est proche de 0. Il s'ensuit que

$$M_Y(t) \approx \prod_{j=1}^n \exp[p_j(e^t - 1)] = \exp \left[(e^t - 1) \sum_{j=1}^n p_j \right],$$

qui est la fonction génératrice d'une v.a. de Poisson de paramètre $\lambda = \sum_{j=1}^n p_j$.

Donc Y , le nombre total d'événements rares, suit à peu près la loi de Poisson(λ).

Ceci tient même si les p_j ne sont pas tous égaux. Il suffit qu'ils soient petits.

Example: approximation par la loi de Poisson

Soient X_1, \dots, X_n indépendantes avec $X_j \sim \text{Binomiale}(1, p_j)$, où chaque p_j est très petit et n est grand. On peut interpréter $\{X_j = 1\}$ comme un événement rare.

$Y = X_1 + \dots + X_n$ est le nombre d'événements rares qui se produisent.

La fonction génératrice de X_j est

$$M_{X_j}(t) = \mathbb{E}[e^{tX_j}] = p_j e^t + 1 - p_j = 1 + p_j(e^t - 1) \approx \exp[p_j(e^t - 1)].$$

quand t est proche de 0, car $1 + x \approx e^x$ quand x est proche de 0. Il s'ensuit que

$$M_Y(t) \approx \prod_{j=1}^n \exp[p_j(e^t - 1)] = \exp \left[(e^t - 1) \sum_{j=1}^n p_j \right],$$

qui est la fonction génératrice d'une v.a. de Poisson de paramètre $\lambda = \sum_{j=1}^n p_j$.

Donc Y , le nombre total d'événements rares, suit à peu près la loi de Poisson(λ).

Ceci tient même si les p_j ne sont pas tous égaux. Il suffit qu'ils soient petits.

L'approximation tient aussi quand les X_j sont (un peu) dépendants (souvent le cas en pratique).

Il suffit en gros que la **dépendance soit faible**. Les preuves sont plus compliquées...

Covariance

La **covariance** entre deux v.a. X et Y est définie par

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

Si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$. Mais l'inverse n'est pas toujours vrai.

Covariance

La **covariance** entre deux v.a. X et Y est définie par

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

Si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$. Mais l'inverse n'est pas toujours vrai.

Exemple. Si A et B sont deux événements et $X = \mathbb{I}[A]$ et $Y = \mathbb{I}[B]$, alors $\text{Cov}(X, Y) = \mathbb{E}[\mathbb{I}[A] \cdot \mathbb{I}[B]] - \mathbb{E}[\mathbb{I}[A]] \cdot \mathbb{E}[\mathbb{I}[B]] = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$, et donc

$$\text{Cov}(X, Y) > 0 \iff \mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} > \mathbb{P}(A).$$

Une covariance positive indique que la probabilité de A est plus grande quand B se produit.

Propriétés de la covariance

Proposition. Si X, Y, Z sont des variables aléatoires et c est une constante, alors

1. $\text{Var}(X) = \text{Cov}(X, X)$;
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
3. $\text{Cov}(cX, Y) = c \text{Cov}(X, Y)$;
4. $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$.

La propriété 4 se généralise facilement (par induction) à :

$$\text{Cov} \left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j).$$

Propriétés de la covariance

Proposition. Si X, Y, Z sont des variables aléatoires et c est une constante, alors

1. $\text{Var}(X) = \text{Cov}(X, X)$;
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
3. $\text{Cov}(cX, Y) = c \text{Cov}(X, Y)$;
4. $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$.

La propriété 4 se généralise facilement (par induction) à :

$$\text{Cov} \left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j).$$

On en déduit en particulier

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \text{Cov} \left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j \right) = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^n \text{Cov}(X_i, X_j).$$

Quand les X_i sont indépendantes, la dernière somme s'annule.

Échantillon i.i.d

Soient X_1, \dots, X_n des variables aléatoires **indépendantes et identiquement distribuées (i.i.d.)** (même loi de probabilité) de moyenne $\mathbb{E}[X] = \mu$ et variance $\text{Var}[X] = \sigma^2 < \infty$.
Les réalisations de ces variables constituent un **échantillon i.i.d.**

Échantillon i.i.d

Soient X_1, \dots, X_n des variables aléatoires **indépendantes et identiquement distribuées (i.i.d.)** (même loi de probabilité) de moyenne $\mathbb{E}[X] = \mu$ et variance $\text{Var}[X] = \sigma^2 < \infty$.

Les réalisations de ces variables constituent un **échantillon i.i.d.**

La **moyenne échantillonnale**: $\bar{X} = \bar{X}_n = (X_1 + \dots + X_n)/n$.

La **variance échantillonnale**: $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

On a $\mathbb{E}[\bar{X}] = \mu$, $\text{Var}[\bar{X}] = \sigma^2/n$, $\text{Cov}(\bar{X}, X_i - \bar{X}) = 0$, $\mathbb{E}[S_n^2] = \sigma^2$.

Échantillon i.i.d

Soient X_1, \dots, X_n des variables aléatoires **indépendantes et identiquement distribuées (i.i.d.)** (même loi de probabilité) de moyenne $\mathbb{E}[X] = \mu$ et variance $\text{Var}[X] = \sigma^2 < \infty$.

Les réalisations de ces variables constituent un **échantillon i.i.d.**

La **moyenne échantillonnale**: $\bar{X} = \bar{X}_n = (X_1 + \dots + X_n)/n$.

La **variance échantillonnale**: $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

On a $\mathbb{E}[\bar{X}] = \mu$, $\text{Var}[\bar{X}] = \sigma^2/n$, $\text{Cov}(\bar{X}, X_i - \bar{X}) = 0$, $\mathbb{E}[S_n^2] = \sigma^2$.

Les v.a. i.i.d. X_1, \dots, X_n dont les valeurs sont triées en ordre croissant s'appellent les **statistiques d'ordre** de l'échantillon. On les dénote $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$.

Types de convergence de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$

Soit $\{Y_n, n \geq 1\}$ une suite de variables aléatoires dont on veut étudier la convergence vers une v.a. Y quand $n \rightarrow \infty$. Il y a plusieurs façons de définir la convergence. Exemples:

Convergence forte: $Y_n \rightarrow Y$ presque sûrement, ou avec probabilité 1, noté $Y_n \xrightarrow{\text{w.p.1}} Y$, si

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} Y_n = Y \right] \stackrel{\text{def}}{=} \mathbb{P} \left[\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega) \right\} \right] = 1.$$

Types de convergence de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$

Soit $\{Y_n, n \geq 1\}$ une suite de variables aléatoires dont on veut étudier la convergence vers une v.a. Y quand $n \rightarrow \infty$. Il y a plusieurs façons de définir la convergence. Exemples:

Convergence forte: $Y_n \rightarrow Y$ presque sûrement, ou avec probabilité 1, noté $Y_n \xrightarrow{\text{w.p.1}} Y$, si

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} Y_n = Y \right] \stackrel{\text{def}}{=} \mathbb{P} \left[\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega) \right\} \right] = 1.$$

Convergence faible: $Y_n \rightarrow Y$ en probabilité, noté $Y_n \xrightarrow{\mathbb{P}} Y$, si pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} [|Y_n - Y| > \epsilon] = 0.$$

Types de convergence de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$

Soit $\{Y_n, n \geq 1\}$ une suite de variables aléatoires dont on veut étudier la convergence vers une v.a. Y quand $n \rightarrow \infty$. Il y a plusieurs façons de définir la convergence. Exemples:

Convergence forte: $Y_n \rightarrow Y$ presque sûrement, ou avec probabilité 1, noté $Y_n \xrightarrow{\text{w.p.1}} Y$, si

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} Y_n = Y \right] \stackrel{\text{def}}{=} \mathbb{P} \left[\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega) \right\} \right] = 1.$$

Convergence faible: $Y_n \rightarrow Y$ en probabilité, noté $Y_n \xrightarrow{\text{P}} Y$, si pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} [|Y_n - Y| > \epsilon] = 0.$$

Convergence en loi: $Y_n \rightarrow Y$ en loi (ou en distribution), noté $Y_n \xrightarrow{\text{D}} Y$ ou $Y_n \Rightarrow Y$, si pour tout $x > 0$ où F est continue, on a $\lim_{n \rightarrow \infty} F_n(x) = F(x)$.

Types de convergence de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$

Soit $\{Y_n, n \geq 1\}$ une suite de variables aléatoires dont on veut étudier la convergence vers une v.a. Y quand $n \rightarrow \infty$. Il y a plusieurs façons de définir la convergence. Exemples:

Convergence forte: $Y_n \rightarrow Y$ presque sûrement, ou avec probabilité 1, noté $Y_n \xrightarrow{\text{w.p.1}} Y$, si

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} Y_n = Y \right] \stackrel{\text{def}}{=} \mathbb{P} \left[\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega) \right\} \right] = 1.$$

Convergence faible: $Y_n \rightarrow Y$ en probabilité, noté $Y_n \xrightarrow{\text{P}} Y$, si pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} [|Y_n - Y| > \epsilon] = 0.$$

Convergence en loi: $Y_n \rightarrow Y$ en loi (ou en distribution), noté $Y_n \xrightarrow{\text{D}} Y$ ou $Y_n \Rightarrow Y$, si pour tout $x > 0$ où F est continue, on a $\lim_{n \rightarrow \infty} F_n(x) = F(x)$.

Convergence en norme L^p : Pour $p \geq 1$, $Y_n \rightarrow Y$ en norme L^p , noté $Y_n \xrightarrow{L^p} Y$, si $\lim_{n \rightarrow \infty} \mathbb{E}[|Y_n - Y|^p] = 0$. Quand $p = 2$, on dit que $Y_n \rightarrow Y$ en moyenne quadratique.

Types de convergence de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$

Soit $\{Y_n, n \geq 1\}$ une suite de variables aléatoires dont on veut étudier la convergence vers une v.a. Y quand $n \rightarrow \infty$. Il y a plusieurs façons de définir la convergence. Exemples:

Convergence forte: $Y_n \rightarrow Y$ presque sûrement, ou avec probabilité 1, noté $Y_n \xrightarrow{\text{w.p.1}} Y$, si

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} Y_n = Y \right] \stackrel{\text{def}}{=} \mathbb{P} \left[\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega) \right\} \right] = 1.$$

Convergence faible: $Y_n \rightarrow Y$ en probabilité, noté $Y_n \xrightarrow{\text{P}} Y$, si pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} [|Y_n - Y| > \epsilon] = 0.$$

Convergence en loi: $Y_n \rightarrow Y$ en loi (ou en distribution), noté $Y_n \xrightarrow{\text{D}} Y$ ou $Y_n \Rightarrow Y$, si pour tout $x > 0$ où F est continue, on a $\lim_{n \rightarrow \infty} F_n(x) = F(x)$.

Convergence en norme L^p : Pour $p \geq 1$, $Y_n \rightarrow Y$ en norme L^p , noté $Y_n \xrightarrow{L^p} Y$, si $\lim_{n \rightarrow \infty} \mathbb{E}[|Y_n - Y|^p] = 0$. Quand $p = 2$, on dit que $Y_n \rightarrow Y$ en moyenne quadratique.

On a que $\{Y_n \xrightarrow{\text{w.p.1}} Y\} \Rightarrow \{Y_n \xrightarrow{\text{P}} Y\} \Rightarrow \{Y_n \xrightarrow{\text{D}} Y\}$ et aussi $\{Y_n \xrightarrow{L^p} Y\} \Rightarrow \{Y_n \xrightarrow{\text{P}} Y\}$.
Mais les implications inverses ne tiennent pas.

Exemples

Soit $\{Y_n, n \geq 1\}$ une suite de v.a.'s indépendantes telles que $\mathbb{P}[Y_n = 1] = 1/n$ et $\mathbb{P}[Y_n = 0] = 1 - 1/n$.

Ici $Y_n \xrightarrow{L^p} 0$ pour tout $p \geq 1$, car $\mathbb{E}[|Y_n|^p] = \mathbb{P}[Y_n = 1] = 1/n \rightarrow 0$,
mais on n'a pas $Y_n \xrightarrow{\text{w.p.1}} 0$.

Supposons maintenant que $\mathbb{P}[Y_n = n] = 1/n$ et $\mathbb{P}[Y_n = 0] = 1 - 1/n$.

Ici on a $Y_n \xrightarrow{P} 0$, car $\mathbb{P}[|Y_n| > \epsilon] = \mathbb{P}[Y_n = n] = 1/n \rightarrow 0$,
mais on n'a pas $Y_n \xrightarrow{\text{w.p.1}} 0$.

Quelques inégalités et théorèmes limites importants

Inégalité de Markov. Si X est une v.a. qui prend ses valeurs dans $[0, \infty)$, alors pour tout $c > 0$, $\mathbb{P}(X \geq c) \leq \mathbb{E}[X]/c$.

Quelques inégalités et théorèmes limites importants

Inégalité de Markov. Si X est une v.a. qui prend ses valeurs dans $[0, \infty)$, alors pour tout $c > 0$, $\mathbb{P}(X \geq c) \leq \mathbb{E}[X]/c$. Preuve: $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq x)dx \geq \int_0^c \mathbb{P}(X \geq c)dx = c\mathbb{P}(X \geq c)$.

Quelques inégalités et théorèmes limites importants

Inégalité de Markov. Si X est une v.a. qui prend ses valeurs dans $[0, \infty)$, alors pour tout $c > 0$, $\mathbb{P}(X \geq c) \leq \mathbb{E}[X]/c$. Preuve: $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq x)dx \geq \int_0^c \mathbb{P}(X \geq c)dx = c\mathbb{P}(X \geq c)$.

Inégalité de Tchebychev. Si X est une v.a. de moyenne μ et de variance σ^2 , alors pour toute constante $c > 0$, $\mathbb{P}(|X - \mu| \geq c) \leq \sigma^2/c^2$.

Quelques inégalités et théorèmes limites importants

Inégalité de Markov. Si X est une v.a. qui prend ses valeurs dans $[0, \infty)$, alors pour tout $c > 0$, $\mathbb{P}(X \geq c) \leq \mathbb{E}[X]/c$. Preuve: $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq x)dx \geq \int_0^c \mathbb{P}(X \geq c)dx = c\mathbb{P}(X \geq c)$.

Inégalité de Tchebychev. Si X est une v.a. de moyenne μ et de variance σ^2 , alors pour toute constante $c > 0$, $\mathbb{P}(|X - \mu| \geq c) \leq \sigma^2/c^2$.

Preuve: Appliquer Markov à $Y = (X - \mu)^2$. $\mathbb{P}(|X - \mu| \geq c) = \mathbb{P}(Y \geq c^2) \leq \mathbb{E}[Y]/c^2 = \sigma^2/c^2$.

Quelques inégalités et théorèmes limites importants

Inégalité de Markov. Si X est une v.a. qui prend ses valeurs dans $[0, \infty)$, alors pour tout $c > 0$, $\mathbb{P}(X \geq c) \leq \mathbb{E}[X]/c$. Preuve: $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq x)dx \geq \int_0^c \mathbb{P}(X \geq c)dx = c\mathbb{P}(X \geq c)$.

Inégalité de Tchebychev. Si X est une v.a. de moyenne μ et de variance σ^2 , alors pour toute constante $c > 0$, $\mathbb{P}(|X - \mu| \geq c) \leq \sigma^2/c^2$.

Preuve: Appliquer Markov à $Y = (X - \mu)^2$. $\mathbb{P}(|X - \mu| \geq c) = \mathbb{P}(Y \geq c^2) \leq \mathbb{E}[Y]/c^2 = \sigma^2/c^2$.

Dans ce qui suit, X_1, \dots, X_n sont des v.a. i.i.d. de moyenne μ et de variance $\sigma^2 < \infty$, et \bar{X}_n est leur moyenne.

Loi forte des grands nombres. $\bar{X}_n \xrightarrow{\text{w.p.1}} \mu$ quand $n \rightarrow \infty$.

Loi faible des grands nombres. $\bar{X}_n \xrightarrow{\text{P}} \mu$ quand $n \rightarrow \infty$. Preuve: utiliser Tchebychev.

Quelques inégalités et théorèmes limites importants

Inégalité de Markov. Si X est une v.a. qui prend ses valeurs dans $[0, \infty)$, alors pour tout $c > 0$, $\mathbb{P}(X \geq c) \leq \mathbb{E}[X]/c$. Preuve: $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq x)dx \geq \int_0^c \mathbb{P}(X \geq c)dx = c\mathbb{P}(X \geq c)$.

Inégalité de Tchebychev. Si X est une v.a. de moyenne μ et de variance σ^2 , alors pour toute constante $c > 0$, $\mathbb{P}(|X - \mu| \geq c) \leq \sigma^2/c^2$.

Preuve: Appliquer Markov à $Y = (X - \mu)^2$. $\mathbb{P}(|X - \mu| \geq c) = \mathbb{P}(Y \geq c^2) \leq \mathbb{E}[Y]/c^2 = \sigma^2/c^2$.

Dans ce qui suit, X_1, \dots, X_n sont des v.a. i.i.d. de moyenne μ et de variance $\sigma^2 < \infty$, et \bar{X}_n est leur moyenne.

Loi forte des grands nombres. $\bar{X}_n \xrightarrow{\text{w.p.1}} \mu$ quand $n \rightarrow \infty$.

Loi faible des grands nombres. $\bar{X}_n \xrightarrow{\text{P}} \mu$ quand $n \rightarrow \infty$. Preuve: utiliser Tchebychev.

Théorème central limite (TCL). Preuve via les fonction génératrices.

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \Rightarrow \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \Rightarrow Z \sim \text{Normale}(0, 1) \quad \text{quand } n \rightarrow \infty.$$

Théorème central limite (TCL).

Il y a de nombreuses autres versions du TCL: des X_i de lois différentes, dépendance, TCLs multivariés, TCL fonctionnels, etc.

En gros, $X \approx$ normale si X est une somme de nombreux petits effets indépendants ou pour lesquels la dépendance s'estompe quand $n \rightarrow \infty$.

Théorème central limite (TCL).

Il y a de nombreuses autres versions du TCL: des X_i de lois différentes, dépendance, TCLs multivariés, TCL fonctionnels, etc.

En gros, $X \approx$ normale si X est une somme de nombreux petits effets indépendants ou pour lesquels la dépendance s'estompe quand $n \rightarrow \infty$.

Exemple: Théorème de Berry-Essen. Soient X_1, X_2, \dots indep., $\mathbb{E}[X_i] = \mu_i$ et $\text{Var}[X_i] = \sigma_i^2 < \infty$. Posons $s_n^2 = \sigma_1^2 + \dots + \sigma_n^2$,

$$Y_n = \frac{(X_1 - \mu_1) + \dots + (X_n - \mu_n)}{s_n},$$

et $F_n(x) = \mathbb{P}[Y_n \leq x]$. Alors, $\mathbb{E}[Y_n] = 0$, $\text{Var}[Y_n] = 1$, et

$$\sup_{n \geq 1, x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \kappa \frac{\mathbb{E}(|X_1 - \mu_1|^3) + \dots + \mathbb{E}(|X_n - \mu_n|^3)}{s_n^3}$$

où $\kappa = 3$ si les X_i sont i.i.d. et $\kappa = 6$ sinon. La borne dépend de l'asymétrie.

Processus stochastiques

Un **processus stochastique** est une famille $Y = \{Y_t, t \in T\}$ de variables aléatoires définies sur un même espace de probabilité.

Processus stochastiques

Un **processus stochastique** est une famille $Y = \{Y_t, t \in T\}$ de variables aléatoires définies sur un même espace de probabilité.

L'indice t est souvent interprété comme le **temps**.

Le processus est en **temps continu** si T est continu (e.g., $T = [0, \infty)$),
et en **temps discret** si T est discret (e.g., $T = \{0, 1, 2, \dots\}$).

Processus stochastiques

Un **processus stochastique** est une famille $Y = \{Y_t, t \in T\}$ de variables aléatoires définies sur un même espace de probabilité.

L'indice t est souvent interprété comme le **temps**.

Le processus est en **temps continu** si T est continu (e.g., $T = [0, \infty)$), et en **temps discret** si T est discret (e.g., $T = \{0, 1, 2, \dots\}$).

Lorsque t est continu, on note souvent Y_t par $Y(t)$.

Processus stochastiques

Un **processus stochastique** est une famille $Y = \{Y_t, t \in T\}$ de variables aléatoires définies sur un même espace de probabilité.

L'indice t est souvent interprété comme le **temps**.

Le processus est en **temps continu** si T est continu (e.g., $T = [0, \infty)$), et en **temps discret** si T est discret (e.g., $T = \{0, 1, 2, \dots\}$).

Lorsque t est continu, on note souvent Y_t par $Y(t)$.

L'**espace d'états** du processus est l'ensemble dans lequel Y_t ou $Y(t)$ prend ses valeurs. C'est souvent \mathbb{R}^d ou un sous-ensemble de \mathbb{R}^d , mais ce peut être aussi un espace plus général.

Processus markoviens

Un processus stochastique est **markovien** si, conditionnellement à sa valeur présente au temps t , son évolution future est indépendante de son passé.

Processus markoviens

Un processus stochastique est **markovien** si, conditionnellement à sa valeur présente au temps t , son évolution future est indépendante de son passé.

C'est-à-dire: pour toute variable aléatoire X fonction de $\{Y_s, s > t\}$, la loi de X conditionnelle à $\{Y_s, s \leq t\}$, est la même que celle conditionnelle à Y_t .

Processus markoviens

Un processus stochastique est **markovien** si, conditionnellement à sa valeur présente au temps t , son évolution future est indépendante de son passé.

C'est-à-dire: pour toute variable aléatoire X fonction de $\{Y_s, s > t\}$, la loi de X conditionnelle à $\{Y_s, s \leq t\}$, est la même que celle conditionnelle à Y_t .

On peut toujours rendre un processus markovien en ajoutant suffisamment d'information dans l'état Y_t .

Processus markoviens

Un processus stochastique est **markovien** si, conditionnellement à sa valeur présente au temps t , son évolution future est indépendante de son passé.

C'est-à-dire: pour toute variable aléatoire X fonction de $\{Y_s, s > t\}$, la loi de X conditionnelle à $\{Y_s, s \leq t\}$, est la même que celle conditionnelle à Y_t .

On peut toujours rendre un processus markovien en ajoutant suffisamment d'information dans l'état Y_t .

Chaîne de Markov temps discret: lorsque $T = \{0, 1, \dots\}$.

Exemple: marche aléatoire sur une chaîne circulaire

Un processus en temps discret a $\{0, 1, \dots, m\}$ comme espace d'états et ces états sont placés sur un cercle. Soit X_n l'état du processus à l'étape n . L'état initial est $X_0 = 0$. Si $X_n = i$, alors X_{n+1} sera l'un des deux états voisins, $(i - 1) \bmod (m + 1)$ ou $(i + 1) \bmod (m + 1)$, chacun avec probabilité $1/2$.

Ce processus évolue jusqu'à ce que tous les états aient été visités au moins une fois. Quelle est la probabilité que le dernier état visité soit l'état i , pour chaque $i \neq 0$?

Exemple: marche aléatoire sur une chaîne circulaire

Un processus en temps discret a $\{0, 1, \dots, m\}$ comme espace d'états et ces états sont placés sur un cercle. Soit X_n l'état du processus à l'étape n . L'état initial est $X_0 = 0$. Si $X_n = i$, alors X_{n+1} sera l'un des deux états voisins, $(i - 1) \bmod (m + 1)$ ou $(i + 1) \bmod (m + 1)$, chacun avec probabilité $1/2$.

Ce processus évolue jusqu'à ce que tous les états aient été visités au moins une fois. Quelle est la probabilité que le dernier état visité soit l'état i , pour chaque $i \neq 0$?

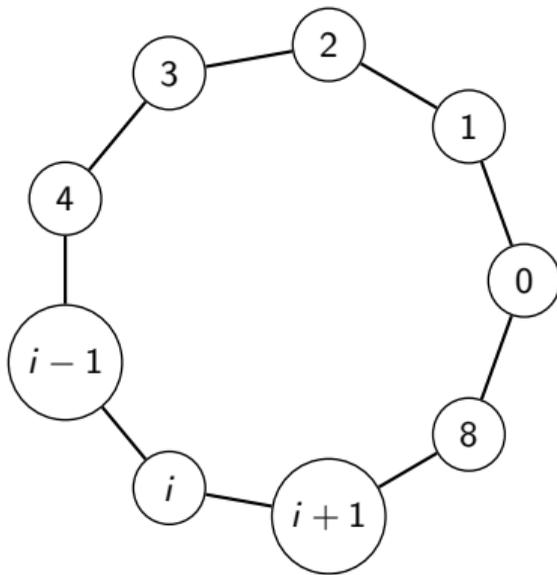
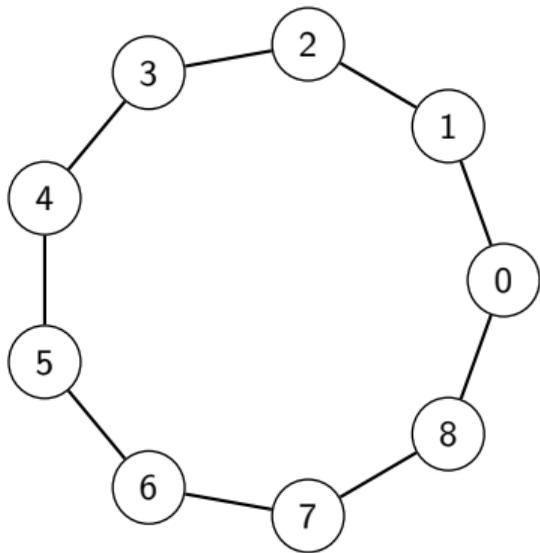
On sait que cette probabilité est nulle pour $i = 0$.

Exemple: marche aléatoire sur une chaîne circulaire

Un processus en temps discret a $\{0, 1, \dots, m\}$ comme espace d'états et ces états sont placés sur un cercle. Soit X_n l'état du processus à l'étape n . L'état initial est $X_0 = 0$. Si $X_n = i$, alors X_{n+1} sera l'un des deux états voisins, $(i - 1) \bmod (m + 1)$ ou $(i + 1) \bmod (m + 1)$, chacun avec probabilité $1/2$.

Ce processus évolue jusqu'à ce que tous les états aient été visités au moins une fois. Quelle est la probabilité que le dernier état visité soit l'état i , pour chaque $i \neq 0$?

On sait que cette probabilité est nulle pour $i = 0$. Pour $i \neq 0$, considérons l'ensemble $V(i)$ constitué des deux voisins de l'état i , qui sont $i - 1$ et $(i + 1) \bmod (m + 1)$. Avec probabilité 1, il y aura un instant (numéro d'étape) où $V(i)$ est visité pour la première fois. À partir de cet instant, la probabilité que i soit visité le dernier est la probabilité que le processus fasse le tour de tous les autres états avant de visiter i . Cette probabilité est la même pour tout $i \neq 0$. Cela implique que la probabilité que i soit le dernier visité est $1/m$ pour tout $i \neq 0$.



Autre interprétation:

Un joueur dispose de 1 dollar et à chaque étape d'un jeu, il gagne un dollar avec probabilité $p = 1/2$ et perd 1 dollar avec probabilité $1 - p = 1/2$. Quelle est la probabilité que sa fortune atteigne m dollars avant de descendre à zéro?

Autre interprétation:

Un joueur dispose de 1 dollar et à chaque étape d'un jeu, il gagne un dollar avec probabilité $p = 1/2$ et perd 1 dollar avec probabilité $1 - p = 1/2$. Quelle est la probabilité que sa fortune **atteigne m dollars avant de descendre à zéro?**

Dans l'exemple précédent, c'est la probabilité que si $X_0 = 1$, l'état 0 soit le dernier état visité. Cette probabilité est $1/m$.

Pour $m = 2$, en particulier, c'est simplement la probabilité de gagner 1 dollar au premier coup.

Autre interprétation:

Un joueur dispose de 1 dollar et à chaque étape d'un jeu, il gagne un dollar avec probabilité $p = 1/2$ et perd 1 dollar avec probabilité $1 - p = 1/2$. Quelle est la probabilité que sa fortune **atteigne m dollars avant de descendre à zéro?**

Dans l'exemple précédent, c'est la probabilité que si $X_0 = 1$, l'état 0 soit le dernier état visité. Cette probabilité est $1/m$.

Pour $m = 2$, en particulier, c'est simplement la probabilité de gagner 1 dollar au premier coup.

Pour une valeur de p plus générale, on peut résoudre le problème par induction sur m .
À faire en exercice.