

3. Variables aléatoires discrètes

et leurs distributions

Sam Gyetvay

ECO 2273 – Économetrie I

September 19, 2025

Introduction

La dernière classe, nous avons appris comment calculer la probabilité de différents événements

- ▶ La probabilité qu'une pièce tombe sur pile est $1/2$
- ▶ La probabilité d'obtenir un "snake eyes" en lançant des dés est $1/36$
- ▶ La probabilité d'obtenir une "royal flush" est $\frac{4}{2598960}$

Connaître les probabilités est utile, mais pour des espaces d'échantillonnage larges, il n'est pas pratique ou faisable de montrer la probabilité de chaque événement possible

À la place, nous définissons une **variable aléatoire** qui prend des valeurs numériques, puis étudions sa **distribution**

Nous pouvons résumer la distribution d'une variable aléatoire avec quelques nombres tels que son **espérance** (moyenne), sa **variance** et sa **skewness**

Les variables aléatoires sont ce qui relie la théorie des probabilités aux statistiques descriptives que nous avons apprises lors de la premier cours

Variables Aléatoires

Une **variable aléatoire** X associe une valeur numérique X à chaque résultat d'une expérience

Exemples:

- ▶ Lancer une pièce de monnaie : $X = 1$ si face, $X = 0$ si pile
- ▶ Lancer un dé : X est égal au nombre affiché sur le dé
- ▶ Cartes : X est égal à la somme des valeurs des cartes

Mathématiquement, X est une **fonction** dont le **domaine** est l'espace d'échantillonnage S et dont l'**image** est l'ensemble des nombres réels \mathbb{R}

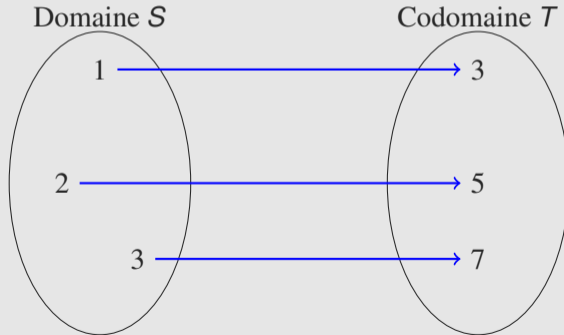
$$X : S \rightarrow \mathbb{R}$$

Revue : fonctions, domaine et image

Une **fonction** $f(\cdot)$ est une règle qui associe à chaque valeur x exactement une valeur $f(x)$

Domaine : l'ensemble de toutes les valeurs x possibles

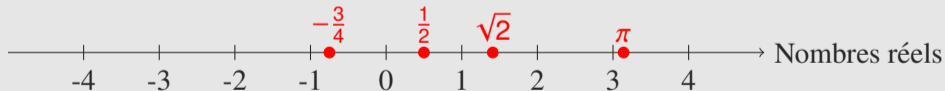
Image : l'ensemble de toutes les valeurs $f(x)$ possibles



Révision : nombres réels

Nombres réels \mathbb{R} incluent :

- ▶ Tous les nombres positifs : $1, 2, 3, \dots$
- ▶ Tous les nombres négatifs : $-1, -2, -3, \dots$
- ▶ Zéro : 0
- ▶ Fractions : $\frac{1}{2}, -\frac{3}{4}$
- ▶ Nombres irrationnels : $\pi, \sqrt{2}$



Variables aléatoires discrètes vs continues

Variables aléatoires discrètes prennent des valeurs dans un ensemble **fini**

Variables aléatoires continues: prennent des valeurs dans un ensemble **infini**
(typiquement \mathbb{R} ou une partie de celui-ci comme $[0, +\infty)$)

Exemples:

- ▶ Discret : nombre de faces obtenues en 10 lancers de pièce ($X \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$)
- ▶ Continu : le temps mis par un coureur pour terminer une course ($X \in [0, +\infty)$)

Pourquoi le temps est-il continu ?

Variables aléatoires discrètes vs continues

Variables aléatoires discrètes prennent des valeurs dans un ensemble **fini**

Variables aléatoires continues: prennent des valeurs dans un ensemble **infini**
(typiquement \mathbb{R} ou une partie de celui-ci comme $[0, +\infty)$)

Exemples:

- ▶ Discret : nombre de faces obtenues en 10 lancers de pièce ($X \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$)
- ▶ Continu : le temps mis par un coureur pour terminer une course ($X \in [0, +\infty)$)

Pourquoi le temps est-il continu ?

- ▶ Si nous mesurons le temps en années, heures, secondes ou minutes, il est techniquement discret !

Variables aléatoires discrètes vs continues

Variables aléatoires discrètes prennent des valeurs dans un ensemble **fini**

Variables aléatoires continues: prennent des valeurs dans un ensemble **infini**
(typiquement \mathbb{R} ou une partie de celui-ci comme $[0, +\infty)$)

Exemples:

- ▶ Discret : nombre de faces obtenues en 10 lancers de pièce ($X \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$)
- ▶ Continu : le temps mis par un coureur pour terminer une course ($X \in [0, +\infty)$)

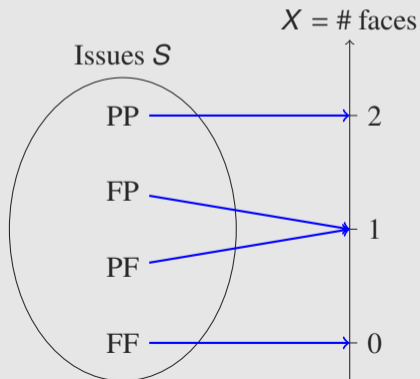
Pourquoi le temps est-il continu ?

- ▶ Si nous mesurons le temps en années, heures, secondes ou minutes, il est techniquement discret !
- ▶ Mais nous pouvons mesurer le temps en unités arbitrairement petites (par exemple, en millisecondes)
- ▶ Sam a terminé la course en 98,14367464234578346784289 secondes
- ▶ Lorsqu'on peut calculer la valeur d'une variable de manière très précise, nous la considérons généralement comme continue. Cela dépend du contexte

Variables aléatoires discrètes

Aujourd'hui, nous nous concentrerons sur les variables aléatoires discrètes. Nous aborderons les variables continues la semaine prochaine.

Exemple : X = nombre de faces obtenues en lançant deux pièces



$$X(s) = \begin{cases} 2 & \text{si } s = PP \\ 1 & \text{si } s = FP \text{ ou } PF \\ 0 & \text{si } s = FF \end{cases}$$

Fonction de masse de probabilité

Une **fonction de masse de probabilité** $p(x)$ attribue des probabilités aux valeurs possibles d'une variable aléatoire discrète X :

$$p(x) = P(X = x)$$

Exemple : si X = nombre de faces lors du lancer de deux pièces, alors

$$p(x) = \begin{cases} P(X = 0) = \frac{1}{4} & (\text{PP}) \\ P(X = 1) = \frac{2}{4} & (\text{FP, PF}) \\ P(X = 2) = \frac{1}{4} & (\text{FF}) \end{cases}$$

Remarque : chaque valeur de X correspond à un **événement** !

- ▶ Rappel : un **événement** est un résultat ou un ensemble de résultats
- ▶ $X = 0 \iff \{PP\}$, $X = 1 \iff \{FP, PF\}$, $X = 2 \iff \{FF\}$

La fonction de masse de probabilité est également appelée la **distribution** de X .

Propriétés d'une fonction de masse de probabilité

Non-négativité : Pour chaque valeur possible x que X peut prendre, la probabilité n'est jamais négative :

$$p(x) \geq 0$$

Pourquoi cela a-t-il du sens ?

- ▶ Rappel : $p(x) = 0$ signifie que x n'arrivera jamais
- ▶ Rappel : $p(x) = 1$ signifie que x arrivera toujours
- ▶ Ne peut être moins que "jamais", ou plus que "toujours"

Somme à un : Les probabilités sur toutes les issues possibles doivent sommer à 1 :

$$\sum_x p(x) = 1$$

Exemple : Lancer une pièce deux fois, X = nombre de faces :

$$p(0) + p(1) + p(2) = \frac{1}{4} + \frac{2}{4} + \frac{1}{4} = 1.$$

Distribution discrète uniforme

Une variable aléatoire discrète X prenant n valeurs a une **distribution uniforme** si chaque valeur est également probable :

$$p(x) = \frac{1}{n} \quad \text{pour chaque valeur possible de } X.$$

Exemple 1 : Lancer un dé et laisser $X =$ le nombre. Alors

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$$

Exemple 2 : Lancer une pièce deux fois et laisser $X = 1$ si (H, H) ou (T, T) et $X = 0$ si (H, T) ou (T, H) . Alors

$$p(0) = p(1) = \frac{1}{2}$$

Exemple 3 : Choisir une carte et laisser $X = 1$ si trèfle, 2 si cœur, 3 si pique, 4 si carreau. Alors

$$p(1) = p(2) = p(3) = p(4) = \frac{1}{4}$$

Quiz : est-ce uniforme ?

Question 1 : X = somme des nombres obtenus en lançant deux dés. Est-ce que $p(x)$ est uniforme ?

Quiz : est-ce uniforme ?

Question 1 : X = somme des nombres obtenus en lançant deux dés. Est-ce que $p(x)$ est uniforme ?

- ▶ Non ! Pourquoi pas ?

Quiz : est-ce uniforme ?

Question 1 : X = somme des nombres obtenus en lançant deux dés. Est-ce que $p(x)$ est uniforme ?

- ▶ Non ! Pourquoi pas ?
- ▶ Il n'y a qu'une seule façon d'obtenir un 2 : “snake eyes” $X = 2 \iff \{(1, 1)\}$
- ▶ Mais il y a six façons d'obtenir un sept :

$$X = 7 \iff \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

- ▶ Donc $p(1) = 1/36$, mais $p(7) = 6/36$!

Quiz : est-ce uniforme ?

Question 1 : X = somme des nombres obtenus en lançant deux dés. Est-ce que $p(x)$ est uniforme ?

- ▶ Non ! Pourquoi pas ?
- ▶ Il n'y a qu'une seule façon d'obtenir un 2 : “snake eyes” $X = 2 \iff \{(1, 1)\}$
- ▶ Mais il y a six façons d'obtenir un sept :

$$X = 7 \iff \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

- ▶ Donc $p(1) = 1/36$, mais $p(7) = 6/36$!

Question 2 : X = nombre de faces obtenues en lançant deux pièces. Est-ce que $p(Y)$ est uniforme ?

Quiz : est-ce uniforme ?

Question 1 : X = somme des nombres obtenus en lançant deux dés. Est-ce que $p(x)$ est uniforme ?

- ▶ Non ! Pourquoi pas ?
- ▶ Il n'y a qu'une seule façon d'obtenir un 2 : “snake eyes” $X = 2 \iff \{(1, 1)\}$
- ▶ Mais il y a six façons d'obtenir un sept :

$$X = 7 \iff \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

- ▶ Donc $p(1) = 1/36$, mais $p(7) = 6/36$!

Question 2 : X = nombre de faces obtenues en lançant deux pièces. Est-ce que $p(Y)$ est uniforme ?

- ▶ Non ! Pourquoi pas ?
- ▶ $p(0) = 1/4, p(1) = 2/4, p(2) = 1/4$

Valeur attendue

La **valeur attendue** d'une variable aléatoire X est

$$E[X] = \sum_x x \cdot p(x)$$

Cela vous semble-t-il familier ?

Valeur attendue

La **valeur attendue** d'une variable aléatoire X est

$$E[X] = \sum_x x \cdot p(x)$$

Cela vous semble-t-il familier ?

Rappelez-vous les définitions de la **moyenne** et de la moyenne pondérée du Cours 1 :

$$\bar{x} = \sum_{i=1}^n \frac{1}{n} x_i$$

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

La moyenne est la valeur attendue lorsque $p(x_i) = \frac{1}{n}$

La moyenne pondérée est lorsque $p(x_i) = \frac{w_i}{\sum_i w_i}$

Interprétation de la valeur attendue I

La **valeur attendue** peut être interprétée comme la valeur que vous vous attendriez à obtenir en jouant à un jeu où vous gagnez la valeur x chaque fois que $X = x$.

Exemple : Vous gagnez 1 \$ à chaque fois que vous obtenez face et perdez 1 \$ à chaque fois que vous obtenez pile. Alors $X = +1$ si F et $X = -1$ si P , et $p(-1) = p(1) = \frac{1}{2}$, donc

$$E[X] = p(-1) \cdot -1 + p(1) \cdot 1 = \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 1 = 0.$$

Exemple : Vous gagnez 2 \$ si le dé affiche 6, sinon vous perdez 0,50 \$. Soit Y le montant que vous gagnez. Alors $Y = 2$ si 6, $Y = -0,5$, $p(2) = \frac{1}{6}$, $p(-0,5) = \frac{5}{6}$. Alors

$$E[Y] = p(2) \cdot 2 + p(-0,5) \cdot (-0,5) = \frac{1}{6} \cdot 2 + \frac{5}{6} \cdot (-0,5) = -\frac{1}{6}$$

Quel jeu préféreriez-vous jouer ?

Interprétation de la valeur attendue II

Une autre façon d'interpréter la valeur attendue est d'imaginer que vous répétez le jeu un nombre infini de fois

Lorsque vous jouez au jeu de pile ou face un petit nombre de fois, le nombre de fois où cela tombera sur pile ne sera pas exactement égal au nombre de fois où cela tombera sur face

Mais si vous jouiez au jeu un très grand nombre de fois (par exemple, 1000000 lancers de pièce), ces nombres seraient presque exactement les mêmes (par exemple 499999 contre 500001)

Même si vous ne lancerez jamais une pièce autant de fois, il est utile de penser aux valeurs attendues *comme si* vous l'aviez fait

La **variance** d'une variable aléatoire X mesure à quel point ses valeurs sont éloignées de leur valeur attendue $E[X]$

$$\text{Var}(X) = E[(X - E[X])^2]$$

- ▶ $X - E[X]$ est la différence entre X et sa valeur attendue
- ▶ $(X - E[X])^2$ est le carré de la différence. Les valeurs du carré sont non négatives, et elles sont plus grandes lorsque $X - E[X]$ est un grand nombre négatif ou positif
- ▶ $E(X - E[X])^2 = \sum_x p(x)(x - E[X])^2$ est la distance carrée attendue entre X et sa valeur attendue

Calcul de la variance

Il est souvent plus facile de calculer la variance en utilisant cette formule :

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

- ▶ $E[X^2]$ est appelé le **deuxième moment** de X . C'est la valeur attendue de X^2 :

$$E[X^2] = \sum_x x^2 p(x)$$

- ▶ $(E[X])^2$ est le carré de la **valeur attendue** :

$$E[X] = \sum_x x p(x), \quad (E[X])^2 = \left(\sum_x x p(x) \right)^2$$

Calcul de la variance : exemples

Exemple 1 : Soit X le nombre sur un dé **équitable**. Probabilités : $p(x) = 1/6$ pour $x = 1, 2, \dots, 6$.

$$E[X] = 3.5, \quad E[X^2] = \frac{91}{6} \approx 15.17$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{91}{6} - 3.5^2 \approx 2.92$$

Exemple 2 : Soit X le nombre sur un dé **truqué** (angl. **loaded**) biaisé vers 3 et 4

$$p(1) = p(2) = 0.1, \quad p(3) = p(4) = 0.25, \quad p(5) = p(6) = 0.15$$

$$E[X] = 3.55, \quad E[X^2] = 15.75$$

$$\text{Var}(X) = 15.75 - 3.55^2 \approx 2.23$$

Distributions conjointes

Pour deux variables aléatoires X et Y , une **fonction de probabilité conjointe** ou une **distribution conjointe** $p(x, y)$ donne la probabilité que X prenne une valeur spécifique et que Y prenne une valeur spécifique simultanément :

$$p(x, y) = P(X = x, Y = y)$$

Exemple : Lancer deux dés

- ▶ X = valeur sur le premier dé, Y = valeur sur le deuxième dé.
- ▶ Les valeurs possibles pour (X, Y) sont tous les couples (x, y) avec $x, y \in \{1, 2, 3, 4, 5, 6\}$.

$$p(x, y) = \frac{1}{36} \quad \text{pour tout } x, y$$

Exemple : Lancer un dé

- ▶ X = valeur sur le dé, $Y = 1$ si la valeur est supérieure à 3
- ▶ Les valeurs possibles pour (X, Y) sont $(1, 0), (2, 0), (3, 0), (4, 1), (5, 1), (6, 1)$

$$p(1, 0) = p(2, 1) = p(3, 0) = p(4, 1) = p(5, 1) = p(6, 1) = \frac{1}{6}$$

Propriétés des distributions conjointes

Pour une fonction de probabilité conjointe $p(x, y)$ de variables aléatoires discrètes X et Y , les propriétés suivantes sont toujours vérifiées :

Non-négativité :

$$p(x, y) \geq 0 \quad \text{pour tout } x, y$$

Somme à un :

$$\sum_x \sum_y p(x, y) = 1$$

Distributions marginales : Nous pouvons retrouver la distribution de X (en ignorant Y) en faisant la somme sur toutes les valeurs de Y :

$$p(x) = \sum_y p(x, y)$$

Dans ce contexte, nous appelons $p(x)$ la **distribution marginale** de X

Covariance et corrélation

Tout comme la moyenne et la variance décrivent la distribution d'une variable aléatoire X , la **covariance** et la **corrélation** résument la distribution conjointe de deux variables aléatoires X et Y

La **covariance** est

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

La **corrélation** est

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

La corrélation est toujours comprise entre -1 et 1

- ▶ Corrélation = 0 : aucune relation (linéaire) entre X et Y
- ▶ Corrélation = -1 : relation linéaire négative parfaite entre X et Y
- ▶ Corrélation = +1 : relation linéaire positive parfaite entre X et Y

Calcul de la covariance

Parfois, il est plus facile de calculer la covariance en utilisant cette formule:

$$\text{Cov}(X, Y) = \frac{1}{2} \left(\text{Var}(X + Y) - \text{Var}(X) - \text{Var}(Y) \right)$$

- ▶ $\text{Var}(X + Y)$ est la variance de la somme. C'est la valeur attendue de $(X + Y)^2$ moins le carré de sa moyenne:

$$\text{Var}(X + Y) = E[(X + Y)^2] - (E[X + Y])^2$$

- ▶ $\text{Var}(X)$ et $\text{Var}(Y)$ sont les variances habituelles de X et Y :

$$\text{Var}(X) = E[X^2] - (E[X])^2, \quad \text{Var}(Y) = E[Y^2] - (E[Y])^2$$

Exemple : calcul de la covariance et de la corrélation

Exemple 1 : Lancer deux dés, X = valeur sur le premier dé, Y = valeur sur le deuxième dé

$$E[X] = E[Y] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3,5, \quad E[XY] = \sum_{x=1}^6 \sum_{y=1}^6 xy \frac{1}{36} = 12,25$$

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 12,25 - 3,5 * 3,5 = 0$$

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = 0$$

Exemple 2 : Lancer un dé, X = valeur, $Y = 1$ si $X > 3$, 0 sinon

$$E[X] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3,5, \quad E[Y] = \frac{0 + 0 + 0 + 1 + 1 + 1}{6} = 0,5$$

$$E[XY] = 2,5, \quad \text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 2,5 - 3,5 * 0,5 = 0,75$$

$$\text{Var}(X) \approx 2,9 \quad \text{Var}(Y) = 0,25$$

$$\rho_{X,Y} = \frac{0,75}{\sqrt{2,9 * 0,25}} \approx 0,88$$

Distributions discrètes classiques

Il existe certaines distributions discrètes célèbres et qui portent leur propre nom.

Ces distributions apparaissent souvent en économie et dans les jeux de hasard.

Binomiale : nombre d'événements se produisant lors de n essais indépendants avec probabilité de succès p

- ▶ Exemple : si l'on lance une pièce 10 fois, X = nombre de fois qu'elle tombe sur face

Poisson : nombre d'événements indépendants se produisant dans un intervalle fixe

Hypergéométrique : nombre de succès dans un échantillon de taille n tiré sans remise d'une population de taille N avec K succès

Variables aléatoires binomiales

Les binomiales apparaissent chaque fois que l'on répète un processus ayant un **résultat binaire** (succès ou échec) plusieurs fois.

À chaque répétition, la probabilité de succès reste la même et le résultat d'un essai n'influence pas les suivants.

Une variable aléatoire binomiale est le nombre de succès dans n essais, lorsque la probabilité de succès est p .

- ▶ Lancer une pièce n fois, X = nombre de faces
- ▶ Lancer deux dés n fois, X = nombre de "double six"
- ▶ Tirer une carte n fois, X = nombre d'As

Si X est une variable binomiale provenant de n essais avec probabilité de succès p , on écrit $X \sim \text{Bin}(n, p)$

Distribution binomiale

Si $X \sim \text{Bin}(n, p)$, alors sa fonction de masse de probabilité est

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

Moyenne : $E[X] = np$ Variance : $\text{Var}(X) = np(1-p)$

$E[X]$ augmente avec n et p . Pourquoi ?

Distribution binomiale

Si $X \sim \text{Bin}(n, p)$, alors sa fonction de masse de probabilité est

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

Moyenne : $E[X] = np$ Variance : $\text{Var}(X) = np(1-p)$

$E[X]$ augmente avec n et p . Pourquoi ?

- ▶ Si l'on répète le processus plus de fois, on observera plus de succès
- ▶ Si le succès est plus probable, on observera plus de succès

$\text{Var}(X)$ augmente avec n et est maximale lorsque $p = 1/2$. Pourquoi ?

Distribution binomiale

Si $X \sim \text{Bin}(n, p)$, alors sa fonction de masse de probabilité est

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

Moyenne : $E[X] = np$ Variance : $\text{Var}(X) = np(1-p)$

$E[X]$ augmente avec n et p . Pourquoi ?

- ▶ Si l'on répète le processus plus de fois, on observera plus de succès
- ▶ Si le succès est plus probable, on observera plus de succès

$\text{Var}(X)$ augmente avec n et est maximale lorsque $p = 1/2$. Pourquoi ?

- ▶ Plus d'essais \rightarrow plus de chances d'observer différents résultats
- ▶ $p = 1/2$ signifie que succès et échec sont également probables : le résultat est le plus imprévisible

Exemples : Binomiale

Exemple 1 : Lancer de pièce

Lancer une pièce $n = 5$ fois, $X =$ nombre de faces

- ▶ $n = 5, p = 0.5$
- ▶ $E[X] = np = 5 \cdot 0.5 = 2.5$
- ▶ $Var(X) = np(1 - p) = 5 \cdot 0.5 \cdot 0.5 = 1.25$

Exemple 2 : Dés

Lancer deux dés $n = 12$ fois, $X =$ nombre de “double six”

- ▶ $p = 1/36, n = 12$
- ▶ $E[X] = np = 12 \cdot \frac{1}{36} = \frac{12}{36} = \frac{1}{3} \approx 0.33$
- ▶ $Var(X) = np(1 - p) = \frac{1}{3} \cdot \frac{35}{36} = \frac{35}{108} \approx 0.32$

Exemple 3 : Cartes

Tirer une carte $n = 5$ fois avec remise, $X =$ nombre d'As

- ▶ $p = 4/52 = 1/13, n = 5$
- ▶ $E[X] = np = 5 \cdot \frac{1}{13} = \frac{5}{13} \approx 0.38$
- ▶ $Var(X) = np(1 - p) = \frac{5}{13} \cdot \frac{12}{13} = \frac{60}{169} \approx 0.36$

Variables aléatoires de Poisson

Les variables de Poisson apparaissent lorsque l'on compte combien de fois un événement se produit dans un **intervalle de temps fixe**.

Chaque événement se produit indépendamment des autres, avec un taux moyen constant λ par intervalle.

Une variable de Poisson est le nombre d'événements observés dans un intervalle

- ▶ Nombre de voitures passant à un carrefour en 1 minute
- ▶ Nombre d'appels téléphoniques reçus par un centre d'appels en 1 heure

Si X est une variable de Poisson avec taux λ , on écrit $X \sim \text{Poisson}(\lambda)$

Distribution de Poisson

Si $X \sim \text{Poisson}(\lambda)$, alors sa fonction de masse de probabilité est

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

Moyenne : $E[X] = \lambda$ Variance : $\text{Var}(X) = \lambda$

Interprétation :

- ▶ $E[X] = \lambda$ signifie que le nombre moyen d'événements par intervalle est égal au taux
- ▶ $\text{Var}(X) = \lambda$ signifie que l'incertitude est du même ordre que la moyenne

Variables aléatoires hypergéométriques

Les hypergéométriques apparaissent lorsque l'on tire des objets **sans remise** d'un nombre fini d'objets

On a N objets, dont K sont des “succès” et $N - K$ des “échecs”. On tire n objets au hasard simultanément, sans les remettre.

Une variable aléatoire hypergéométrique est le nombre de succès parmi les n tirages.

- ▶ Un jeu de 52 cartes, 4 sont des As. Tirer $n = 5$ cartes sans remise, $X =$ nombre d'As
- ▶ Une urne contient 20 boules, 6 rouges et 14 bleues. Tirer $n = 4$, $X =$ nombre de boules rouges
- ▶ Une classe de 30 étudiants, 18 filles et 12 garçons. Tirer $n = 3$ au hasard, $X =$ nombre de filles

Si X est une variable hypergéométrique, on écrit $X \sim \text{Hypergeo}(N, K, n)$

Distribution hypergéométrique

Si $X \sim \text{Hypergeo}(N, K, n)$, alors sa fonction de masse de probabilité est

$$p(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \dots, n$$

Moyenne : $E[X] = n \frac{K}{N}$ Variance : $\text{Var}(X) = n \frac{K}{N} \left(1 - \frac{K}{N}\right) \frac{N-n}{N-1}$

$E[X]$ augmente avec n et K/N . Pourquoi ?

Distribution hypergéométrique

Si $X \sim \text{Hypergeo}(N, K, n)$, alors sa fonction de masse de probabilité est

$$p(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \dots, n$$

Moyenne : $E[X] = n \frac{K}{N}$ Variance : $\text{Var}(X) = n \frac{K}{N} \left(1 - \frac{K}{N}\right) \frac{N-n}{N-1}$

$E[X]$ augmente avec n et K/N . Pourquoi ?

- ▶ Plus d'objets tirés \rightarrow plus de succès attendus
- ▶ Si la population a une plus grande proportion de succès, on en tire davantage
- ▶ K/N joue le rôle de p dans la distribution binomiale

$\text{Var}(X)$ augmente avec n jusqu'à un certain point, puis diminue. Pourquoi ?

- ▶ Plus de tirages \rightarrow plus de chances de variation
- ▶ Mais si $n = N$, il n'y a aucune variance : on obtient toujours K succès
- ▶ Plus imprévisible lorsque $K/N = 1/2$ (succès et échecs également probables)

Exemples : Hypergéométrique

Exemple 1 : Tirer $n = 5$ cartes sans remise, $X =$ nombre d'as

- ▶ $N = 52, K = 4, n = 5$
- ▶ $E[X] = n \cdot K/N = 5 \cdot 4/52 = 20/52 \approx 0.38$
- ▶ $Var(X) = n \cdot \frac{K}{N} \cdot \left(1 - \frac{K}{N}\right) \cdot \frac{N-n}{N-1}$
- ▶ $= 5 \cdot \frac{4}{52} \cdot \frac{48}{52} \cdot \frac{47}{51} \approx 0.34$

Exemple 2 : Une urne contient 20 boules, 6 rouges et 14 bleues. Tirer $n = 4$, $X =$ nombre de boules rouges

- ▶ $N = 20, K = 6, n = 4$
- ▶ $E[X] = 4 \cdot 6/20 = 24/20 = 1.2$
- ▶ $Var(X) = 4 \cdot \frac{6}{20} \cdot \frac{14}{20} \cdot \frac{16}{19} \approx 0.71$

Exemple 3 : Une classe a 30 étudiants, 18 femmes et 12 hommes. Choisir $n = 3$ au hasard, $X =$ nombre de femmes

- ▶ $N = 30, K = 18, n = 3$
- ▶ $E[X] = 3 \cdot 18/30 = 54/30 = 1.8$
- ▶ $Var(X) = 3 \cdot \frac{18}{30} \cdot \frac{12}{30} \cdot \frac{27}{29} \approx 0.67$

Exemple : Compter les poissons avec une hypergéométrie

Jacques Cousteau (biologiste marin) veut estimer le nombre de poissons dans un lac

Étape 1 : attraper K poissons, les peindre en rouge, puis les remettre dans le lac

Étape 2 : attendre un peu pour que les poissons se mélangent, puis attraper n poissons

Soit X = nombre de poissons rouges (marqués) dans le second échantillon

Alors $X \sim \text{Hyper}(N, K, n)$, où N = nombre total de poissons dans le lac

Donc $E[X] = n \cdot K/N$

Si l'on connaît K et n , et en supposant $X = E[X]$, on peut estimer le nombre total de poissons :

$$N = n \cdot \frac{K}{X}$$



Prochaine leçon

Je sais que cette leçon était ennuyeuse. Désolé...

Mais, souvenez-vous que nous construisons une base pour appliquer ces concepts aux données économiques

De nombreuses variables économiques peuvent être vues comme suivant ces distributions

- ▶ Un travailleur envoie n candidatures, chacune avec probabilité p de mener à un entretien (Binomiale)
- ▶ Un chômeur reçoit des offres d'emploi au taux λ (Poisson)
- ▶ Une banque examine au hasard n prêts parmi N au total, dont K sont en défaut (Hypergéométrique)

La prochaine séance, nous étudierons les variables aléatoires continues et leurs distributions (Normale, Chi-carré, Student- t , etc.). Ce sera peut-être plus intéressant !