Rappel de Probabilités et Statistiques

Mohamed Ndaoud

1 Définitions importantes

La modélisation statistique comporte une partie essentielle de formalisme mathématique. En particulier, pour se familiariser avec les méthodes fondamentales en statistiques, il est nécessaire de maîtriser un certain nombre d'outil de la théorie des Probabilités. Il n'est pas question ici de faire un cours de Probabilités exhaustif, mais simplement de rassembler un petit nombre de définitions essentielles dans la suite. L'étudiant pourra s'y référer tout au long du cours.

Définition 1 (**Expérience aléatoire).** *Nous appelons* **expérience aléatoire** *une expérience $\mathcal{E}$ qui, reproduite dans des conditions identiques, peut conduire à plusieurs résultats possibles, et dont on ne peut prédire le résultat par avance. L'espace de tous les résultats possibles, appelé* **espace d'états** *associé à l'expérience, sera noté Ω. Un résultat possible de l'expérience est noté ω, $\omega \in \Omega$*

Exemple 1. *Le lancé d'un dé est une expérience aléatoire d'espace d'états $\Omega = \{1, 2, 3, 4, 5, 6\}$.*

Définition 2 (**Événement aléatoire).** *Nous appelons* **événement aléatoire** *(associé à l'expérience $\mathcal{E}$) un sous-ensemble de l'espace d'état Ω dont on peut dire au vu de l'expérience s'il est réalisé ou non. Un événement A est donc une partie de Ω: $A \subset \Omega$.*

Exemple 2. *Pour le lancé d'un dé,*

$$A = \{\textit{la valeur du dé est inférieure ou égale à } 3\} = \{1, 2, 3\} \subset \Omega$$

est un événement aléatoire.

Définition 3 (**Probabilité).** *Une probabilité $\mathbb{P}$ est définie sur l'ensemble $\mathcal{A}$ des événements aléatoires liés à l'expérience et vérifie les propriétés suivantes pour tout A et B dans $\mathcal{A}$:*

- $0 \leq \mathbb{P}(A) \leq 1$
- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ *si* $A \cap B = \emptyset$.

Exemple 3. *Considérons six nombres $p_1, \ldots, p_6$ satisfaisant $0 \leq p_i \leq 1$ pour tout $i \in \{1, \ldots, 6\}$, et $p_1 + \ldots + p_6 = 1$. La loi $\mathbb{P}$ définie par*

- $\{\mathbb{P}(\{i\}) = p_i, i \in \{1, \ldots, 6\}\}$,
- *pour tout* $I \subset \{1, \ldots, 6\}$, $\mathbb{P}(I) = \sum_{i \in I} \mathbb{P}(\{i\})$

est une probabilité.

Définition 4 (**Modèle probabiliste).** *Un modèle probabiliste est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ constitué de l'espace d'états Ω, de l'ensemble des événements aléatoires $\mathcal{A}$ et de la famille des $\mathbb{P}(A)$ pour $A \in \mathcal{A}$.*

Exemple 4. *Pour l'expérience du lancé de dé, le triplet* $(\{1,\ldots,6\},\mathcal{P}(\{1,\ldots,6\}),\mathbb{P})$ *est un modèle probabiliste, avec* $\mathcal{P}(\{1,\ldots,6\})$ *l'ensemble des parties de* $\{1,\ldots,6\}$ *et* $\mathbb{P}$ *la probabilité définie ci-dessus.*

Définition 5 (Tribu)**.** *La classe* $\mathcal{A}$ *est une tribu sur l'espace* Ω *si elle vérifie:*

- $\emptyset \in \mathcal{A}$ *et* $\Omega \in \mathcal{A}$*,*
- $\mathcal{A}$ *est stable par passage au complémentaire :* $A \in \mathcal{A} \Rightarrow \Omega \setminus A \in \mathcal{A}$*,*
- $\mathcal{A}$ *est stable par union et intersection dénombrables : si* $(A_n)_{n\in\mathbb{N}}$ *est une suite d'éléments de* $\mathcal{A}$ *alors* $\cup_{n\in\mathbb{N}}A_n$ *et* $\cap_{n\in\mathbb{N}}A_n$ *sont dans* $\mathcal{A}$*.*

Exemple 5. *Si* Ω *est un ensemble fini alors l'ensemble des parties de* Ω *noté* $\mathcal{P}(\Omega)$ *est une tribu. Si* $\Omega = \mathbb{R}$ *alors l'ensemble des intervalles ouverts de* $\mathbb{R}$ *noté* $\mathcal{B}(\mathbb{R})$ *est une tribu (appelée tribu Borélienne).*

Définition 6 (Espace mesurable)**.** *Un espace mesurable est un couple* $(\Omega,\mathcal{A})$*, où* $\mathcal{A}$ *est une tribu sur* Ω

Exemple 6. *Si* Ω *est un ensemble fini, alors* $(\Omega,\mathcal{P}(\Omega))$ *est un espace mesurable. De même,* $\mathbb{R},\mathcal{B}(\mathbb{R})$ *est un espace mesurable*

Définition 7 (Fonction mesurable)**.** *Soient* $(E,\mathcal{E})$ *et* $(F,\mathcal{F})$ *deux espaces mesurables. Une fonction* $f : E \to F$ *est mesurable si la tribu image réciproque par* f *de la tribu* $\mathcal{F}$ *est incluse dans* $\mathcal{E}$*, c'est-à-dire si*

$$\forall B \in \mathcal{F}, f^{-1}(B) \in \mathcal{E}.$$

Exemple 7. *Une fonction continue de* $\mathbb{R}$ *dans* $\mathbb{R}$ *est mesurable sur* $(\mathbb{R},\mathcal{B}(\mathbb{R}))$*.*

---

2 Loi d'une variable aléatoire

Une variable aléatoire est une grandeur qui dépend du résultat d'une expérience aléatoire. Par exemple, le nombre de 2 obtenus dans un lacés de 3 dés. Formellement, une variable aléatoire est définie comme suit.

Définition 8 (**Variable aléatoire**)**.** *On considère un espace de probabilité* $(\Omega,\mathcal{A},\mathbb{P})$ *et* $(E,\mathcal{E})$ *un espace mesurable. Une variable aléatoire* X *est une application mesurable de* $(\Omega,\mathcal{A})$ *dans* $(E,\mathcal{E})$*,*

$$X : \omega \in \Omega \mapsto X(\omega) \in E.$$

Intuitivement, l'introduction d'une variable aléatoire revient à évaluer le résultat de l'expérience aléatoire non pas par rapport à son résultat direct $\omega \in \Omega$, mais par rapport à la valeur de la fonction $X(\omega)$. L'intérêt fondamental est que le résultat de l'expérience n'est pas forcément accessible, contrairement aux évaluations de $X(\omega)$. La **loi d'une variable aléatoire** est définie comme suit.

Définition 9 (**Loi d'une variable aléatoire**)**.** *Soit* $X : \Omega \mapsto E$ *une variable aléatoire. La loi de* X*, notée* $\mathbb{P}_X$*, est une mesure de probabilité sur* $(E,\mathcal{E})$ *telle que*

$$\mathbb{P}_X[A] = \mathbb{P}[X^{-1}(A)] = \mathbb{P}(X \in A), \quad \forall A \in \mathcal{E}.$$

La distribution de probabilité d'une variable aléatoire est souvent caractérisée par un petit nombre de paramètres, qui ont également une interprétation pratique. Par exemple, il suffit souvent de connaître sa " valeur moyenne ". Celle-ci est donnée par **l'espérance** d'une variable aléatoire (également appelée premier moment).

Exemple 8. *Supposons que* X *suit une loi de Bernoulli de paramètre* $1/3$*.*

- *La loi de* X *est décrite par* $\mathbb{P}[X = 1] = 1/3 = 1 - \mathbb{P}[X = 0]$

2.1 Distribution binomiale

Une vaste classe de problèmes se caractérisent par le fait que l'expérience a exactement deux résultats possibles. Par exemple, lorsqu'on tire à pile ou face, mais également dans certains tests cliniques, et dans des problèmes de classification binaire. Ces problèmes sont appelés des expériences binomiales, ou expériences de Bernoulli. Les caractéristiques d'une expérience binomiale sont les suivantes :

- Il y a un nombre fixe de tirages, noté n
- Les n tirages sont indépendants et répétés sous les mêmes conditions
- Chaque tirage a deux issues possibles : la réussite (1) ou l'échec (0)
- Pour chaque tirage, la probabilité de réussite est la même. Cette probabilité est dénotée p, tandis que la probabilité d'échec est dénotée q. Nous avons $p + q = 1$.

Remarque 1 (Tirage sans remise)**.** *Lorsqu'on fait des tirages sans remise dans une population, les tirages ne sont ni indépendants ni identiquement distribués.*

Formule pour la distribution de probabilité binomiale :

$$\mathbb{P}_X(X = k) = \frac{n!}{k!(n-k)!}p^k q^{n-k} = \binom{n}{r} p^k q^{n-k}$$

- $n =$ nombre de tirages
- $p =$ probabilité de succès de chaque tirage
- $q = 1 - p$ probabilité d'échec de chaque tirage
- $k =$ le nombre de succès parmi les n tirages
- $\binom{n}{k} =$ le coefficient binomial (le nombre de combinaison de taille k de n objets)

La formule de $\mathbb{P}_X$ se décompose en deux parties : le nombre $p^k q^{n-k}$ est la probabilité d'obtenir k succès parmi les n tirages; le coefficient binomial $\binom{n}{r}$ compte le nombre de résultats qui comprennent k succès et $n - k$ échecs.

Exemple 9 (Mots dans un corpus de textes)**.** *La loi binomiale peut être utilisée pour modéliser l'apparition de mots dans un corpus de textes. Supposons qu'on s'intéresse aux occurrences d'un mot fixé, par exemple "statistiques" dans un corpus de n textes. Notons p la fréquence du mot "statistiques" dans le corpus. Alors, sous des hypothèses d'indépendence et d'équidistribution, un modèle pour l'apparition du mot "statistiques" est la loi binomiale de nombre de tirages n et de probabilité de succès p :*

$$\mathbb{P}(\{\text{"statistiques" apparaît } k \text{ fois dans le corpus}\}) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} = \binom{n}{r}p^k(1-p)^{n-k}.$$

2.2 Distribution géométrique

Dans de nombreuses situations, les expériences sont en fait répétées jusqu'à obtenir un "succès". Considérons une expérience dans laquelle nous répétons des tirages binomiaux jusqu'à obtenir le premier succès, puis nous arrêtons. Notons n soit le numéro du tirage pour lequel nous obtenons notre premier succès. Dans ce contexte, n n'est pas un nombre fixe. En fait, n pourrait être n'importe lequel des nombres 1, 2, 3, et ainsi de suite. Quelle est la probabilité que notre premier succès survienne lors du $n$ième essai ? La réponse est donnée par la distribution géométrique. **Formule pour la distribution géométrique:**

$$\mathbb{P}_X(X = n) = p(1-p)^{n-1}$$

- $n =$ nombre de tirages effectués avant l'obtention du premier succès
- $p =$ probabilité de succès à chaque tirage

2.3 Distribution de Poisson

Lorsque, dans la loi binomiale, le nombre de tirages tend vers l'infini tandis que la probabilité de succès tend vers 0, on obtient à la limite la distribution de Poisson. **Formule de la distribution de Poisson:**

$$\mathbb{P}_X(X = k) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}$$

- $\lambda =$ le nombre moyen de succès par unité de temps, volume, etc.
- $k =$ le nombre de succès

Par exemple, soit X une variable aléatoire de Poisson de paramètre 2.

- La loi de X est décrite par :
$$\mathbb{P}_X[X = k] = \frac{2^k}{k!}e^{-2}, \quad k = 0, 1, \ldots$$

Il existe de nombreuses applications de la distribution de Poisson. Par exemple, les "temps d'attente" (d'un bus, des pompiers, etc.) sont souvent modélisés par une distribution de Poisson de la façon suivante. On considère que le temps d'attente peut être subdivisé en de nombreux petits intervalles. L'arrivée effective (du bus, des pompiers, etc.) pendant l'un des très courts intervalles peut être considérée comme un événement peu fréquent (ou rare). La probabilité d'arrivée dans l'intervalle k est donnée par la distribution de Poisson:

$$\mathbb{P}(\{\text{Le bus arrive au temps } k\}) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}.$$

La distribution de Poisson est aussi fréquemment utilisée pour modéliser les données de comptage en écologie. Considérons une expérience de recensement d'espèces sur un site écologique. La probabilité d'observer k individus de cette espèce est donnée par :

$$\mathbb{P}(\{k \text{ individus sont présents sur le site}\}) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}.$$

Approximation de Poisson à la distribution binomiale : Considérons la distribution binomiale. Si le nombre d'essais n est grand alors que la probabilité p de succès est assez faible, nous appelons l'événement (succès) un événement rare. Pour des raisons pratiques, la distribution de Poisson avec $\lambda = np$ sera une très bonne approximation de la distribution binomiale à condition que le nombre d'essais $n \geq 100$ et $\lambda n \leq 10$. Au fur et à mesure que n devient plus grand et p plus petit, l'approximation devient meilleure.

2.4 Distribution normale

La distribution normale est l'une des distributions les plus importantes et les plus fréquemment utilisées pour une variable aléatoire continue. Les caractéristiques importantes de la loi normale de moyenne μ et d'écart-type σ sont les suivantes :

- La courbe de la distribution est en forme de "cloche" et prend sa valeur maximale à la moyenne μ
- La courbe est symmétrique par rapport à l'axe vertical passant par μ
- Les points d'inflexion se situent en $\mu - \sigma$ et $\mu + \sigma$
- Le paramètre σ contrôle l'étalement de la courbe : si l'écart-type σ est grand, la courbe sera plus étalée ; s'il est petit, la courbe sera plus pointue.

Considérons une variable aléatoire X suivant une loi normale de moyenne 0 et d'écart-type 1 (appelée loi normale standard).

- La loi de X est décrite par :
$$\mathbb{P}_X[X \in [a, b]] = \int_{[a,b]} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}dx$$

- La loi de X s'écrit grâce à sa densité:

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

- Le calcul de l'espérance :

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x)\, \mathbb{P}_X(dx) = \int_{\mathbb{R}} \phi(x) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} dx$$

Approximation normale de la distribution binomiale : On sait que la probabilité qu'un nouveau vaccin protège les adultes contre le choléra est de 0,85. Le vaccin est administré à 300 adultes qui doivent se rendre dans une zone où la maladie est prévalente. Quelle est la probabilité que plus de 280 de ces adultes soient protégés contre le choléra par le vaccin ? Cette question entre dans la catégorie d'une expérience binomiale où le nombre d'essais n est égal à 300, la probabilité de succès p est égale à 0,85 et le nombre de succès k est supérieur à 280. Il est possible d'utiliser la formule de la distribution binomiale pour calculer la probabilité que k soit supérieur à 280. Toutefois, cette approche nécessiterait un certain nombre de calculs longs et fastidieux. Une façon plus facile de résoudre ce problème est d'utiliser l'approximation normale de la distribution binomiale. La distribution normale avec $\mu = np$ et $\sigma = \sqrt{npq}$ sera une bonne approximation de la distribution binomiale à condition que $np > 5$ et $nq > 5$. Plus n augmente, meilleure est l'approximation.

2.5 Fonction de répartition

La loi d'une variable aléatoire est un objet "compliqué". Elle peut être :

- Discrète (Bernoulli)
- Continue par rapport à la mesure de Lebesgue (distribution normale)
- Une combinaison des deux

On peut caractériser la loi de X par un objet plus simple : **la fonction de répartition**, qui est une fonction croissante et bornée. Dans un contexte statistique, cette fonction est souvent plus simple à utiliser que la loi de X.

Définition 10 (Fonction de répartition)**.** *La fonction de répartition d'une variable aléatoire X à valeurs réelles, évaluée en x, est la probabilité que la variable X prenne une valeur inférieure ou égale à x :*

$$F(x) = \mathbb{P}[X \leq x], \forall x \in \mathbb{R}\,.$$

La fonction F est croissante, continue à droite (seule la limite à droite est contrainte à être égale à la valeur de la fonction). De plus, $F(-\infty) = 0$ et $F(+\infty) = 1$. La fonction F caractérise la loi $\mathbb{P}_X$:

$$\mathbb{P}_X[(a, b]] = \mathbb{P}[a < X \leq b] = F(b) - F(a).$$

À partir de maintenant, le terme "loi de X" se référera indifféremment à la distribution $\mathbb{P}_X$ ou à la fonction de répartition F.

3 Convergence d'une séquence de variables aléatoires

La convergence des séquences de variables aléatoires vers une variable aléatoire limite est un concept important. Il existe plusieurs notions différentes de convergence pour les variables aléatoires. Elles formalisent l'idée qu'on peut parfois s'attendre à ce qu'une séquence d'événements aléatoires s'installe en un "modèle" lorsque les éléments sont suffisamment éloignés dans la séquence. Par exemple, la séquence finit par prendre une valeur constante ou les valeurs de la séquence continuent à changer mais peuvent être décrites par une distribution de probabilité limite. Dans cette section, nous passons en revue les différents types de convergence pour une variable aléatoire.

3.1 Convergence en distribution

L'idée de la convergence en distribution est la suivante. Pour une séquence de variable aléatoire $(X_n)_{n\geq 0}$, lorsque n grandit, la variable aléatoire X_n est de mieux en mieux modélisé par une certaine distribution de probabilité donnée. La convergence en distribution est la forme de convergence la plus faible (elle est impliquée par tous les autres types de convergence). Elle est très utilisée dans la pratique ; le plus souvent, elle découle de l'application du théorème de la limite centrale.

Définition 11 (Convergence en distribution)**.** *Une séquence (X_n) de variables aléatoires à valeurs réelles converge en distribution (ou en loi) vers une variable aléatoire X si*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

pour tout nombre $x \in \mathbb{R}$ où F est continue. Ici, F_n et F sont les fonctions de répartitions des variables aléatoires X_n et X, respectivement.

La convergence en distribution se note : $X_n \xrightarrow{d} X$.

Remarque 2. *La restriction aux points où F est continue est essentielle. Par exemple, si X_n est distribuée uniformément sur l'intervalle $(0, 1/n)$, alors la séquence (X_n) converge en distribution vers une variable aléatoire dégénérée $X = 0$. Cependant, nous avons $F(0) = 1$, et $F_n(0) = 0$ pour tout n.*

Remarque 3. *Nous avons $\mathbb{E}\, f(X_n) \to \mathbb{E}\, f(X)$ pour toute fonction $f : \mathbb{R} \to \mathbb{R}$ continue bornée.*

Exemple 10 (Lancé de pièces)**.** *Soit X_n la fraction de "face" obtenue après n lancement d'une pièce non biaisée. Alors, X_1 suit une distribution de Bernoulli de valeur moyenne $\mu = 0.5$ et de variance $\sigma^2 = 0.25$. Les variables aléatoires suivantes $X_2, X_3, \ldots$ suivent des distributions binomiales. Lorsque n grandit, la distribution de X_n va graduellement ressembler de plus en plus à une loi normale (approximation normale). Si l'on translate et standardise X_n, $Z_n = \frac{\sqrt{n}}{\sigma}(X_n - \mu)$ converge en distribution vers une loi normale standard (théorème central limite).*

3.2 Convergence en probabilité

L'idée de la convergence en probabilité est la suivante. La probabilité d'un résultat inhabituel devient de plus en plus faible à mesure que la séquence progresse. La convergence en probabilité est très souvent utilisée en statistique. Par exemple, un estimateur est dit consistent si il converge en probabilité vers la quantité que l'on cherche à estimer.

Définition 12 (Convergence en probabilité)**.** *Une séquence X_n de variables aléatoires convergen en probabilité vers la variable aléatoire X si pour tout $\epsilon > 0$*

$$\lim_{n\to 0} \mathbb{P}\left(|X_n - X| > \epsilon\right) = 0.$$

Ainsi, on peut choisir $\epsilon > 0$ et $\delta > 0$. Soit P_n la probabilité que X_n se trouve à l'extérieur de la boule centrée en X de rayon ϵ. Alors, si X_n converge en probabilité vers X, in exist un entier N (qui dépend de δ et ϵ) tel que pour tout $n \geq N$, $P_n < \delta$.

La convergence en distribution se note $X_n \xrightarrow{P} X$.

Remarque 4. *La convergence en probabilité implique la convergence en distribution.*

Remarque 5. *La convergence en distribution implique la convergence en probabilité lorsque la variable aléatoire limite X est une constante.*

Exemple de l'archer : Supposons qu'une personne prenne un arc et commence à tirer des flèches sur une cible. Soit X_n son score au n-ième tir. Au début, il aura de fortes chances de marquer des zéros, mais au fur et à mesure que le temps passe et que son habileté au tir à l'arc augmente, il aura de plus en plus de chances de toucher le mille et de marquer 10 points. Après des années de pratique, la probabilité qu'il touche autre chose que 10 sera de plus en plus faible et convergera vers 0. Ainsi, la séquence X_n converge en probabilité vers $X = 10$. Notez que X_n ne converge cependant pas de façon quasi certaine. Peu importe le professionnalisme de l'archer, il y aura toujours une petite probabilité de faire une erreur. Ainsi, la séquence X_n ne deviendra jamais stationnaire : il y aura toujours des scores non-parfaits, même s'ils sont de moins en moins fréquents.

Vidéo: Vous pouvez regarder le vidéo de Ben Lambert qui explique la différence entre la convergence en probabilité et convergence en distribution: https://www.youtube.com/watch?v=ZKqzA81Nz2Y

3.3 Convergence presque sûre

La convergence presque sûre est le type de convergence qui ressemble le plus à la convergence ponctuelle habituelle.

Définition 13 (Convergence presque sûre)**.** *La séquence (X_n) converge presque sûrement (ou avec probabilité 1) vers X si*

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

Cela signifie que la valeur de X_n approche de la valeur de X, au sens où l'événement où X_n ne converge pas vers X est de probabilité nulle.

La convergence presque sûre se note $X_n \overset{p.s.}{\to} X$.

Remarque 6. *La convergence presque sûre implique la convergence en probabilité (par le lemme de Fatou), et donc aussi la convergence en distribution.*

Exemple 11. *Supposons que l'on commence avec \$1. Chaque jour nous jouons à un jeu où soit on perd, soit on triple sa mise, avec probabilités égales. Soit X_n la variable aléatoire qui représente le montant que l'on possède après n jours. Cette séquence converge presque sûrement vers la variable aléatoire $X = 0$.*

Vidéo: Vous pouvez regarder le vidéo de Mark Huber qui explique la différence entre la convergence en probabilité et convergence presque sure: https://www.youtube.com/watch?v=cRFADVuXCiw