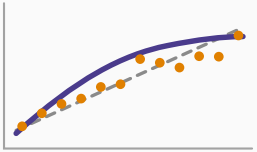
Économétrie avancée

Annexe B : Rappels de probabilité



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Plan du chapitre

Avant de commencer

Pourquoi cette annexe

Pourquoi ce chapitre ?

Tous les résultats des chapitres 1 à 10 — non-biais, convergence, lois t et F — reposent sur un socle probabiliste précis : variables aléatoires, espérance, variance, covariance, et la famille de lois normale/χ^2/t/F. Ce chapitre rassemble ce socle en un seul endroit, comme référence et non comme nouveau cours de probabilité.

Variables aléatoires et leur distribution

Variable aléatoire et fonction de densité

Une **variable aléatoire** prend des valeurs numériques déterminées par le résultat d'une expérience. Une variable **discrète** prend un nombre fini (ou dénombrable) de valeurs, chacune associée à une probabilité $p_j = P(X = x_j)$ (avec $\sum_j p_j = 1$) : c'est la **fonction de densité (pdf)** $f(x)$. Une variable **continue** prend n'importe quelle valeur réelle avec probabilité individuelle nulle ; seules les probabilités sur un intervalle, obtenues par l'aire sous f, ont un sens.

Fonction de répartition (cdf)

$$F(x) = P(X \leq x)$$

est toujours croissante, comprise entre 0 et 1. Elle permet de calculer toute probabilité d'intervalle : $P(X > c) = 1 - F(c)$ et $P(a < X \leq b) = F(b) - F(a)$.

Indépendance

X et Y sont **indépendantes** si et seulement si $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ pour tout (x, y) — la probabilité jointe se factorise. L'indépendance implique l'absence de corrélation, mais l'inverse est faux (voir plus bas).

Distribution conditionnelle et espérance conditionnelle

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

L'**espérance conditionnelle** $E(Y \mid X = x)$ — moyenne de Y pour une valeur donnée de X — est l'outil central de l'économétrie : la fonction de régression $E(y \mid x)$ des chapitres 2 à 10 n'est rien d'autre que cette notion, appliquée à un modèle linéaire en x.

Espérance et variance

Espérance (moyenne de population)

Pour X discrète, $E(X) = \sum_j x_j f(x_j)$ — une moyenne pondérée par les probabilités. Propriétés essentielles : $E(c) = c$ pour toute constante ; $E(aX + b) = aE(X) + b$ (linéarité) ; plus généralement, $E\left(\sum_i a_i X_i\right) = \sum_i a_i E(X_i)$, **sans aucune condition d'indépendance**.

Une mise en garde utile : $E[g(X)] \neq g[E(X)]$

Sauf cas particulier (fonction affine), l'espérance d'une transformation non linéaire de X n'est **pas** la transformation de l'espérance — $E(X^2) \neq [E(X)]^2$ en général. Ce fait rappelle pourquoi $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$ est presque toujours strictement positif.

Variance et écart-type

$$\mathrm{Var}(X) = E[(X-\mu)^2] = E(X^2) - \mu^2, \qquad sd(X) = \sqrt{\mathrm{Var}(X)}.$$

$\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$: ajouter une constante ne change pas la dispersion, la multiplier par a multiplie la variance par a^2 (et l'écart-type par $|a|$) — c'est pourquoi l'écart-type, exprimé dans l'unité d'origine, est souvent plus interprétable que la variance.

Standardisation

$$Z = \frac{X-\mu}{\sigma} \implies E(Z) = 0,\ \mathrm{Var}(Z) = 1.$$

C'est exactement la construction du **coefficient bêta** du chapitre 6 : centrer-réduire une variable pour rendre les comparaisons indépendantes de l'unité de mesure.

Covariance, corrélation, variance d'une somme

Covariance et corrélation

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY)-\mu_X\mu_Y, \qquad \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{sd(X)\,sd(Y)} \in [-1,1].$$

X, Y indépendantes $\Rightarrow \mathrm{Cov}(X,Y) = 0$ (non corrélées) — mais l'**inverse est faux** : X et X^2 peuvent avoir une covariance nulle tout en étant manifestement dépendantes. Covariance et corrélation ne captent que la dépendance **linéaire**.

Un exemple qui piège

Si $E(X) = 0$, alors $Cov(X, X^2) = E(X^3)$, qui s'annule dès que la distribution de X est symétrique — bien que $Y = X^2$ soit une fonction déterministe (donc parfaitement dépendante) de X. La corrélation nulle ne signifie jamais absence de relation, seulement absence de relation **linéaire**.

Variance d'une combinaison linéaire

$$\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y).$$

Si X et Y sont non corrélées, $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ — la variance d'une **différence** est

Propriétés de l'espérance conditionnelle

$E[c(X) \mid X] = c(X)$ (une fonction de X se comporte comme une constante une fois X connu); si $X \perp Y$, $E(Y \mid X) = E(Y)$; et surtout, la **loi des espérances itérées** :

$$E[E(Y \mid X)] = E(Y).$$

Si $E(Y \mid X) = E(Y)$ (l'espérance conditionnelle ne dépend pas de X), alors $\mathrm{Cov}(X, Y) = 0$ – c'est ce résultat qui justifie que $E(u \mid x) = 0$ (hypothèse centrale de tout ce cours depuis le chapitre 2) implique $\mathrm{Cov}(x, u) = 0$, condition nécessaire à l'absence de biais des MCO.

Pourquoi l'espérance conditionnelle est le meilleur prédicteur

Parmi toutes les fonctions $g(X)$ possibles, $E(Y \mid X)$ minimise l'erreur quadratique moyenne de prédiction $E\{[Y - g(X)]^2\}$. C'est la justification probabiliste profonde de la régression : quand on écrit $E(y \mid x) = \beta_0 + \beta_1 x$, on postule que **la meilleure prédiction linéaire** de y est une fonction affine de x.

La famille normale et ses dérivées

Loi normale et loi normale centrée réduite

$X \sim \mathcal{N}(\mu, \sigma^2)$ a pour densité $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x-\mu)^2/2\sigma^2]$, symétrique autour de μ. Standardiser X donne $Z = (X-\mu)/\sigma \sim \mathcal{N}(0,1)$, dont la cdf $\Phi(z)$ est tabulée. Toute combinaison linéaire de variables normales indépendantes est elle-même normale — en particulier, la moyenne de n variables i.i.d. $\mathcal{N}(\mu, \sigma^2)$ suit $\mathcal{N}(\mu, \sigma^2/n)$.

Corrélation nulle = indépendance, un cas spécial à la loi normale

En général, absence de corrélation n'implique pas indépendance (voir plus haut) — sauf pour des variables **conjointement normales**, où les deux notions coïncident. C'est une propriété remarquable, propre à la famille gaussienne, qui simplifie beaucoup l'analyse sous l'hypothèse de normalité (MLR.6, chapitre 4).

Les lois dérivées : χ^2, t, F

Construction des trois lois de l'inférence

Avec $Z, Z_1, \dots, Z_n$ des $\mathcal{N}(0,1)$ indépendantes :

- $\chi^2_n = \sum_{i=1}^{n} Z_i^2$ suit une loi du **chi-deux** à n degrés de liberté ($E = n$, $\mathrm{Var} = 2n$), toujours positive, non symétrique.
- $T = Z / \sqrt{\chi^2_n / n}$ (avec Z et χ^2_n indépendantes) suit une loi de **Student** t_n — plus dispersée que la normale, s'en rapprochant quand $n \to \infty$.
- $F = (\chi^2_{k_1}/k_1)/(\chi^2_{k_2}/k_2)$ suit une loi de **Fisher** F_{k_1,k_2}.

Pourquoi ces trois lois, précisément

Ce n'est pas un hasard si ce sont exactement les trois lois des chapitres 4 et 8 : le test t sur un coefficient MCO combine un numérateur normal ($\hat{\beta}_j$ centré) et un dénominateur qui est la racine d'un χ^2 (l'estimateur de σ^2, via $SSR/\sigma^2 \sim \chi^2_{n-k-1}$ sous normalité) — exactement la construction de T ci-dessus. Le test F pour des restrictions multiples est de même un ratio de deux χ^2 normalisés. Toute la théorie exacte du chapitre 4 est une application directe de ces trois définitions.

Synthèse

Ce qu'il faut retenir

L'essentiel du chapitre 11

- Espérance et variance obéissent à des règles de linéarité simples, valables sans hypothèse d'indépendance pour l'espérance, avec covariance pour la variance.
- Covariance et corrélation ne mesurent que la dépendance linéaire ; l'espérance conditionnelle capture toute la relation, linéaire ou non, et minimise l'erreur quadratique de prédiction.
- $E(u \mid x) = 0 \Rightarrow \mathrm{Cov}(x, u) = 0$ via la loi des espérances itérées : c'est le pont exact entre l'hypothèse fondamentale de la régression et la condition d'orthogonalité utilisée pour dériver les MCO.
- Les lois χ^2, t et F se construisent toutes à partir de normales indépendantes ; ce sont précisément les lois utilisées pour l'inférence exacte en régression (chapitre 4).

Vers le chapitre 12

Ce chapitre a traité les variables aléatoires **individuellement**, comme des objets de population. Le chapitre suivant applique ces mêmes outils à l'**échantillonnage** — comment un estimateur, construit à partir d'un échantillon, hérite lui-même d'une distribution de probabilité, et comment on en tire les propriétés (non-biais, efficacité, convergence) déjà utilisées depuis le chapitre 2.