CentraleSupélec

Statistique et apprentissage

Chargés de cours (ordre alphabétique) :

Julien Bect, Gilles Faÿ, Ziad Kobeissi, Laurent Le Brusquet,
Vincent Lescarret, Arshak Minasyan, Arthur Tenenhaus† & Xujia Zhu

† Coordinateur du cours

Cours 1/9

Introduction et méthodes d'estimation ponctuelle

Objectifs du cours 1

- Illustrer l'intérêt de la statistique
- Définir le cadre mathématique nécessaire à la compréhension du cours
- Présenter les principales méthodes d'estimation ponctuelle

Plan du cours

Plan du cours

Un mot, plusieurs significations...

- Une (ou des) statistique(s) : ensemble d'indicateurs, souvent simples, calculés à partir des données.

  Exemples : moyenne, écart-type, médiane, etc....

- La (ou les) statistique(s) : discipline mathématique qui comporte plusieurs branches, dont
  - la statistique descriptive,
  - l'inférence statistique (partie 1 du cours),
  - la planification d'expériences,
  - l'apprentissage statistique (partie 2 du cours),
  - ...

Remarque : une définition mathématique du mot « statistique » (première acception) sera donnée plus loin.

Un mot, plusieurs significations. . .

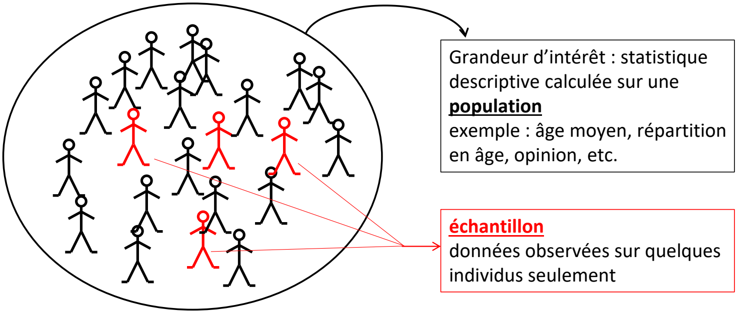
- Une (ou des) statistique(s) : ensemble d'indicateurs, souvent simples, calculés à partir des données.

  Exemples : moyenne, écart-type, médiane, etc. . . .

- La (ou les) statistique(s) : discipline mathématique qui comporte plusieurs branches, dont
  - la statistique descriptive,
  - l'inférence statistique (partie 1 du cours),
  - la planification d'expériences,
  - l'apprentissage statistique (partie 2 du cours),
  - . . .

Remarque : une définition mathématique du mot « statistique » (première acception) sera donnée plus loin.

Exemple historique : le sondage

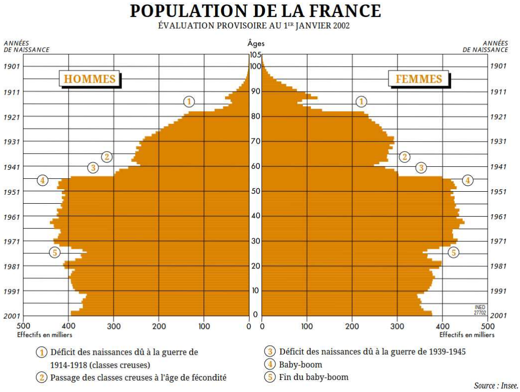


Une statistique descriptive peut être calculée sur :

- toute la population → grandeur d'intérêt
- un échantillon → valeur « approximative » (sens à définir)

Inférer = tirer des conclusions sur une population à partir de données collectées sur un échantillon

Sondage $\neq$ recensement exhaustif



Les statistiques descriptives servent à « explorer » les données

Objectifs : obtenir des résumés numériques (de petite dimension), des visualisations facilement interprétables, etc.

Note : en France, pour les communes de plus de 10000 habitants, le recensement exhaustif a été remplacé depuis 2004 par un échantillonnage aléatoire (mais pas IID) des adresses.

Autre exemple : estimer une proportion

Contexte. Considérons une boite avec B boules blanches et R boules rouges, avec B et R inconnus.

Objectif. Estimer la proportion θ de boules blanches $\left(\theta = \frac{B}{B+R}\right)$.

Données observées. On effectue n tirages avec remise

➠ pour le $i^{\text{ème}}$ tirage, $x_i = 1$ si la boule est blanche, 0 sinon.

Etapes pour estimer θ

1. **modéliser.**
   x_i réalisation d'une VA X_i avec $X_i \overset{\text{iid}}{\sim} \mathrm{Ber}(\theta)$, $0 \leq \theta \leq 1$
2. **inférer.**
   à partir de $\underline{x} = (x_1, \ldots, x_n)$ et du modèle considéré.
   ➠ recours fréquent à des « descripteurs » de l'échantillon $\underline{x}$.
   ➠ On peut voir $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$ comme un candidat naturel.

Liens entre l'inférence statistique et les probabilités

La statistique repose sur la théorie des probabilités :

- probabilités : la loi de probabilité est connue ;
- inférence statistique : seul le modèle de loi est connu et on caractérise les paramètres du modèle à partir d'observations.

Illustration sur l'exemple de la boite :

	Probabilités (B et R connus)	Inférence (B et R inconnus)
exemples de question	• loi du nombre de boules blanches après n tirages ; • loi du nombre de tirages nécessaires pour obtenir une boule blanche	• estimer θ ; • donner un intervalle contenant θ ; • décider si $\theta \leq 0.5$
nature des conclusions	annoncées avec certitude	à n fini, impossible en général de répondre avec certitude

Liens entre l'inférence statistique et les probabilités

La statistique repose sur la théorie des probabilités :

- probabilités : la loi de probabilité est connue ;
- inférence statistique : seul le modèle de loi est connu et on caractérise les paramètres du modèle à partir d'observations.

Illustration sur l'exemple de la boite :

	Probabilités (B et R connus)	Inférence (B et R inconnus)
exemples de question	• loi du nombre de boules blanches après n tirages ; • loi du nombre de tirages nécessaires pour obtenir une boule blanche	• estimer θ ; • donner un intervalle contenant θ ; • décider si $\theta \leq 0.5$
nature des conclusions	annoncées avec certitude	à n fini, impossible en général de répondre avec certitude

Exemples de questions traitées, dans divers domaines

- **Santé :** identifier des bio-marqueurs liés à une maladie à partir de données issues de cohortes.
- **Assurance :** évaluer le risque de ruine d'une compagnie d'assurance.
- **Industrie :** contrôler la qualité d'une chaîne de production à partir de mesures réalisées sur quelques éléments prélevés.
- **Sondage :** prédire le vainqueur d'une élection à partir d'une enquête d'opinion, quantifier la confiance dans la prédiction.
- **Écologie :** estimer la taille d'une population animale à partir d'observations partielles (par ex., capture-marquage-recapture).
- ...

Plan du cours

Des observations aux variables aléatoires

Données (observations)

Soit $\underline{x} \in \underline{\mathcal{X}}$ les données à traiter. Par exemple :

1. une grandeur scalaire collectée sur n objets/individus :
   ➠ $\underline{x} = (x_1, \dots, x_n), \quad x_i \in \mathbb{R}, \quad \underline{\mathcal{X}} = \mathbb{R}^n$;
2. d grandeurs scalaires, de natures potentiellement différentes, collectées sur n objets/individus :
   ➠ $\underline{x} = (x_1, \dots, x_n), \quad x_i \in \mathbb{R}^d, \quad \underline{\mathcal{X}} = \mathbb{R}^{n \times d}$;
3. tout autre ensemble plus complexe de données (séries chronologiques, données symboliques, graphes, etc.).

On modélise les données, a priori, par une variable aléatoire (VA) $\underline{X}$
➠ $\underline{x}$ est considéré comme une réalisation de $\underline{X}$.

Modèle statistique

Espace $(\underline{\mathcal{X}}, \underline{\mathscr{A}})$

Il s'agit de l'espace mesurable dans lequel $\underline{X}$ prend ses valeurs.
La plupart du temps, on considérera :

- $\underline{\mathcal{X}} = \mathbb{R}^n$ avec $\underline{\mathscr{A}} = \mathcal{B}(\mathbb{R}^n)$
- ou, plus généralement, $\underline{\mathcal{X}} = \mathbb{R}^{n\times d}$ avec $\underline{\mathscr{A}} = \mathcal{B}\left(\mathbb{R}^{n\times d}\right)$.

Modélisation statistique

Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace probabilisé permettant de définir :

- la variable aléatoire observée $\underline{X}$,
- toute autre VA (non obs.) dont nous pourrions avoir besoin.

La probabilité $\mathbb{P}$ n'est pas parfaitement connue : on se donne un

- ensemble $\mathscr{P}$ de probabilités sur $(\Omega, \mathscr{F})$ supposé contenir la « vraie » mesure de probabilité.

Modèle statistique

Espace $(\underline{\mathcal{X}}, \underline{\mathscr{A}})$

Il s'agit de l'espace mesurable dans lequel $\underline{X}$ prend ses valeurs.
La plupart du temps, on considérera :

- $\underline{\mathcal{X}} = \mathbb{R}^n$ avec $\underline{\mathscr{A}} = \mathcal{B}(\mathbb{R}^n)$
- ou, plus généralement, $\underline{\mathcal{X}} = \mathbb{R}^{n \times d}$ avec $\underline{\mathscr{A}} = \mathcal{B}\left(\mathbb{R}^{n \times d}\right)$.

Modélisation statistique

Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace probabilisé permettant de définir :

- la variable aléatoire observée $\underline{X}$,
- toute autre VA (non obs.) dont nous pourrions avoir besoin.

La probabilité $\mathbb{P}$ n'est pas parfaitement connue : on se donne un

- ensemble $\mathscr{P}$ de probabilités sur $(\Omega, \mathscr{F})$ supposé contenir la « vraie » mesure de probabilité.

Modèle statistique (suite)

Loi des observations

On note $\mathbb{P}^{\underline{X}}$ la loi de $\underline{X}$ lorsque $\mathbb{P} \in \mathscr{P}$ est la loi sous-jacente.

➠ On a un ensemble $\mathscr{P}^{\underline{X}} = \left\{\mathbb{P}^{\underline{X}}, \mathbb{P} \in \mathscr{P}\right\}$ de lois possibles.

Définition : modèle statistique

Formellement, on appelle modèle statistique le triplet

$$\mathscr{M} = \left(\underline{\mathcal{X}},\ \underline{\mathscr{A}},\ \mathscr{P}^{\underline{X}}\right).$$

Remarques :

- On peut construire différents modèles $(\Omega, \mathscr{F}, \mathscr{P}, \underline{X})$ pour un même $\mathscr{M}$.
- En particulier, lorsqu'on ne s'intéresse qu'à la VA observée $\underline{X}$, on peut se placer sur le modèle *canonique* : $\Omega = \underline{\mathcal{X}}$, $\mathscr{F} = \underline{\mathscr{A}}$, $\mathscr{P} = \mathscr{P}^{\underline{X}}$, $\underline{X} = \mathrm{Id}_{\underline{\mathcal{X}}}$.

Modèle statistique (suite)

Loi des observations

On note $\mathbb{P}^{\underline{X}}$ la loi de $\underline{X}$ lorsque $\mathbb{P} \in \mathscr{P}$ est la loi sous-jacente.

➠ On a un ensemble $\mathscr{P}^{\underline{X}} = \left\{ \mathbb{P}^{\underline{X}}, \mathbb{P} \in \mathscr{P} \right\}$ de lois possibles.

Définition : modèle statistique

Formellement, on appelle modèle statistique le triplet

$$\mathscr{M} = \left(\underline{\mathcal{X}}, \ \underline{\mathscr{A}}, \ \mathscr{P}^{\underline{X}} \right).$$

Remarques :

- On peut construire différents modèles $(\Omega, \mathscr{F}, \mathscr{P}, \underline{X})$ pour un même $\mathscr{M}$.
- En particulier, lorsqu'on ne s'intéresse qu'à la VA observée $\underline{X}$, on peut se placer sur le modèle *canonique* : $\Omega = \underline{\mathcal{X}}$, $\mathscr{F} = \underline{\mathscr{A}}$, $\mathscr{P} = \mathscr{P}^{\underline{X}}$, $\underline{X} = \mathrm{Id}_{\underline{\mathcal{X}}}$.

Inférence statistique

Rappel : les données $\underline{x} \in \underline{\mathcal{X}}$ sont considérées comme une réalisation de $\underline{X} \sim \mathbb{P}^{\underline{X}}$, pour une certaine probabilité $\mathbb{P} \in \mathscr{P}$ (inconnue).

Objectif de l'inférence statistique

Objectif : construire des procédures permettant d'extraire de l'information sur $\mathbb{P}^{\underline{X}}$ à partir

- d'une réalisation de $\underline{X}$,
- de la connaissance de l'ensemble $\mathscr{P}^{\underline{X}}$ des lois possibles.

Important

Comme la vraie probabilité $\mathbb{P}$ est inconnue, on doit concevoir des procédures statistiques « applicables » à **toutes** les proba $\mathbb{P} \in \mathscr{P}$.

Inférence statistique

Rappel : les données $\underline{x} \in \underline{\mathcal{X}}$ sont considérées comme une réalisation de $\underline{X} \sim \mathbb{P}^{\underline{X}}$, pour une certaine probabilité $\mathbb{P} \in \mathscr{P}$ (inconnue).

Objectif de l'inférence statistique

Objectif : construire des procédures permettant d'extraire de l'information sur $\mathbb{P}^{\underline{X}}$ à partir

- d'une réalisation de $\underline{X}$,
- de la connaissance de l'ensemble $\mathscr{P}^{\underline{X}}$ des lois possibles.

Important

Comme la vraie probabilité $\mathbb{P}$ est inconnue, on doit concevoir des procédures statistiques « applicables » à **toutes** les proba $\mathbb{P} \in \mathscr{P}$.

Famille de lois

L'ensemble $\mathscr{P}$ est représenté par une famille paramétrée :

$$\mathscr{P} = \{\mathbb{P}_\theta,\ \theta \in \Theta\}\,.$$

Modèle paramétrique.

Si Θ est de dimension finie, le modèle est dit paramétrique.

- le vecteur de paramètre θ est souvent de petite taille.
- dans la suite, on note p le nombre de paramètres ($\Theta \subset \mathbb{R}^p$).

Exemple. Famille des lois normales sur $\underline{\mathcal{X}} = \mathbb{R}$

$$\mathscr{P}^{\underline{X}} = \left\{\mathcal{N}(\mu,\sigma^2), \quad \mu \in \mathbb{R}, \quad \sigma^2 \in \mathbb{R}_*^+\right\}$$

(Dans cet exemple on suppose que l'on n'a qu'une observation, scalaire.)

Famille de lois

L'ensemble $\mathscr{P}$ est représenté par une famille paramétrée :

$$\mathscr{P} = \{\mathbb{P}_\theta,\ \theta \in \Theta\}\,.$$

Modèle paramétrique.

Si Θ est de dimension finie, le modèle est dit paramétrique.

- le vecteur de paramètre θ est souvent de petite taille.
- dans la suite, on note p le nombre de paramètres ($\Theta \subset \mathbb{R}^p$).

Exemple. Famille des lois normales sur $\underline{\mathcal{X}} = \mathbb{R}$

$$\mathscr{P}^{\underline{X}} = \left\{\mathscr{N}(\mu, \sigma^2), \quad \mu \in \mathbb{R}, \quad \sigma^2 \in \mathbb{R}_*^+\right\}$$

(Dans cet exemple on suppose que l'on n'a qu'une observation, scalaire.)

Modèles d'échantillonnage

n-échantillon

Si $\underline{X} = (X_1, \ldots, X_n)$ est tel que :

- les X_i sont mutuellement indépendantes,
- tous les X_i suivent la même loi P_θ,

alors les X_i sont dites indépendantes et identiquement distribuées (iid) et on dit que $\underline{X}$ est un n-échantillon (iid).

Loi d'un n-échantillon iid.

Considérons le modèle décrivant chacun des X_i individuellement :

- $(\mathcal{X}, \mathscr{A}, \{\mathrm{P}_\theta, \theta \in \Theta\})$

On a alors :

- $(\underline{\mathcal{X}}, \underline{\mathscr{A}}) = (\mathcal{X}^n, \mathscr{A}^{\otimes n})$ (espace produit),
- $\forall \theta \in \Theta,\ \mathbb{P}_\theta^{\underline{X}} = \mathrm{P}_\theta^{\otimes n}$ (loi produit).

Modèles d'échantillonnage

n-échantillon

Si $\underline{X} = (X_1, \dots, X_n)$ est tel que :

- les X_i sont mutuellement indépendantes,
- tous les X_i suivent la même loi P_θ,

alors les X_i sont dites indépendantes et identiquement distribuées (iid) et on dit que $\underline{X}$ est un *n*-échantillon (iid).

Loi d'un *n*-échantillon iid.
Considérons le modèle décrivant chacun des X_i individuellement :

- $(\mathcal{X}, \mathscr{A}, \{\mathrm{P}_\theta, \theta \in \Theta\})$

On a alors :

- $(\underline{\mathcal{X}}, \underline{\mathscr{A}}) = (\mathcal{X}^n, \mathscr{A}^{\otimes n})$ (espace produit),
- $\forall \theta \in \Theta,\ \mathbb{P}_\theta^{\underline{X}} = \mathrm{P}_\theta^{\otimes n}$ (loi produit).

Application « fiabilité composant »

Cette application servira d'illustration dans plusieurs chapitres.

Contexte

- On s'intéresse à la fiabilité d'un composant issu d'une chaîne de production.
- Grandeur d'intérêt (fiabilité) assimilée à la durée de vie du composant.
- Données observées : Prélèvement de $n = 10$ composants pour lesquels on mesure leurs durées de vie : $\underline{x} = (x_1, \ldots, x_n)$.

Modélisation

- On modélise chaque x_i par une VA X_i.
- Les X_i sont supposées iid, à valeurs dans $(\mathcal{X}, \mathscr{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- $(\underline{\mathcal{X}}, \underline{\mathscr{A}}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Application « fiabilité composant »

Cette application servira d'illustration dans plusieurs chapitres.

Contexte

- On s'intéresse à la fiabilité d'un composant issu d'une chaîne de production.
- Grandeur d'intérêt (fiabilité) assimilée à la durée de vie du composant.
- Données observées : Prélèvement de $n = 10$ composants pour lesquels on mesure leurs durées de vie : $\underline{x} = (x_1, \ldots, x_n)$.

Modélisation

- On modélise chaque x_i par une VA X_i.
- Les X_i sont supposées iid, à valeurs dans $(\mathcal{X}, \mathscr{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- $(\underline{\mathcal{X}}, \underline{\mathscr{A}}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Application « fiabilité composant »

Modélisation (suite) : famille de lois

Hypothèse courante* pour la durée de vie d'un composant :

$$X_i \sim \mathcal{E}(\theta), \quad \theta > 0.$$

D'où le modèle statistique :

$$\left(\mathbb{R}^n,\ \mathcal{B}(\mathbb{R}^n),\ \{\mathcal{E}(\theta)^{\otimes n}, \theta > 0\}\right).$$

Rappel. La loi exponentielle $\mathcal{E}(\theta)$ admet la densité :

$$f_\theta(x) = \theta \exp(-\theta x)\, \mathbb{1}_{[0,\infty[}(x).$$

par rapport à la mesure de Lebesgue.

* dans le cas de pannes imprévisibles, sans lien avec l'âge du composant

Application « fiabilité composant »

Modélisation (suite) : famille de lois

Hypothèse courante* pour la durée de vie d'un composant :

$$X_i \sim \mathcal{E}(\theta), \quad \theta > 0.$$

D'où le modèle statistique :

$$\left(\mathbb{R}^n,\ \mathcal{B}(\mathbb{R}^n),\ \{\mathcal{E}(\theta)^{\otimes n}, \theta > 0\}\right).$$

Rappel. La loi exponentielle $\mathcal{E}(\theta)$ admet la densité :

$$f_\theta(x) = \theta \exp(-\theta x)\, \mathbb{1}_{[0,\infty[}(x).$$

par rapport à la mesure de Lebesgue.

* dans le cas de pannes imprévisibles, sans lien avec l'âge du composant

Application « fiabilité composant »

Quelques problèmes statistiques

- estimer θ, ou
- estimer $\eta = \frac{1}{\theta} = \mathbb{E}(X_1)$ (durée de vie moyenne)

  ➠ cours 1 et 2

- fournir des **intervalles de confiance** pour θ et η

  ➠ cours 3

- **tester l'hypothèse** $\eta \leq 10$ afin de déterminer s'il serait pertinent de proposer aux clients une garantie

  ➠ cours 4 sur les tests d'hypothèses

- **estimer** θ en présence d'a priori sur sa valeur (a priori fourni par le constructeur de la chaîne de production)

  ➠ cours 5 sur l'estimation bayésienne

Application « fiabilité composant »

Quelques problèmes statistiques

- **estimer** θ, ou
- **estimer** $\eta = \frac{1}{\theta} = \mathbb{E}(X_1)$ (durée de vie moyenne)

  ➠ cours 1 et 2

- fournir des **intervalles de confiance** pour θ et η

  ➠ cours 3

- **tester l'hypothèse** $\eta \leq 10$ afin de déterminer s'il serait pertinent de proposer aux clients une garantie

  ➠ cours 4 sur les tests d'hypothèses

- **estimer** θ en présence d'a priori sur sa valeur
  (a priori fourni par le constructeur de la chaîne de production)

  ➠ cours 5 sur l'estimation bayésienne

Application « fiabilité composant »

Quelques problèmes statistiques

- **estimer** θ, ou
- **estimer** $\eta = \frac{1}{\theta} = \mathbb{E}(X_1)$ (durée de vie moyenne)

  ➠ cours 1 et 2
- fournir des **intervalles de confiance** pour θ et η

  ➠ cours 3
- **tester l'hypothèse** $\eta \leq 10$ afin de déterminer s'il serait pertinent de proposer aux clients une garantie

  ➠ cours 4 sur les tests d'hypothèses
- **estimer** θ en présence d'a priori sur sa valeur
  (a priori fourni par le constructeur de la chaîne de production)

  ➠ cours 5 sur l'estimation bayésienne

Application « fiabilité composant »

Quelques problèmes statistiques

- **estimer** θ, ou
- **estimer** $\eta = \frac{1}{\theta} = \mathbb{E}(X_1)$ (durée de vie moyenne)

  ➠ cours 1 et 2
- fournir des **intervalles de confiance** pour θ et η

  ➠ cours 3
- **tester l'hypothèse** $\eta \leq 10$ afin de déterminer s'il serait pertinent de proposer aux clients une garantie

  ➠ cours 4 sur les tests d'hypothèses
- **estimer** θ en présence d'a priori sur sa valeur (a priori fourni par le constructeur de la chaîne de production)

  ➠ cours 5 sur l'estimation bayésienne

Application « fiabilité composant » (suite)

Données : un échantillon de taille $n = 10$ [unité arbitraire]

0.5627	16.1121	5.4943	7.9374	1.2658
2.9885	8.6266	43.8877	2.1641	8.9138

Estimer η **: un premier estimateur** (voir cours 2 pour une définition)

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}_\theta(X_1) = \eta \quad \text{(LFGN)}.$$

➠ $\hat{\eta}^{(1)} = \bar{X}$ semble un estimateur « raisonnable » de η.

Application numérique : $\hat{\eta}^{(1)} = 10.1960$

Application « fiabilité composant » (suite)

Données : un échantillon de taille $n = 10$ [unité arbitraire]

0.5627	16.1121	5.4943	7.9374	1.2658
2.9885	8.6266	43.8877	2.1641	8.9138

Estimer η **:** un premier estimateur (voir cours 2 pour une définition)

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[n \to \infty]{\text{p.s.}} \mathbb{E}_\theta(X_1) = \eta \quad \text{(LFGN)}.$$

➠ $\hat{\eta}^{(1)} = \bar{X}$ semble un estimateur « raisonnable » de η.

Application numérique : $\hat{\eta}^{(1)} = 10.1960$

Notations / vocabulaire

Notations. On utilisera souvent les notations :

- $\mathbb{E}_{\theta}(.)$ (espérance),
- $\text{var}_{\theta}(.)$ (variance ou matrice de covariance),
- $f_{\theta}(.)$ (densité), ...

pour rappeler que ces opérateurs ou fonctions dépendent d'une loi $\mathbb{P}_{\theta}$ pour une valeur particulière de θ.

Définition : Statistique

Une statistique est une variable aléatoire (souvent scalaire ou vectorielle) calculée à partir de $\underline{X}$ seulement*.

Exemple : l'estimateur $\hat{\eta}^{(1)} = \bar{X}$ est une statistique.

* Techniquement : s'exprime comme une fonction mesurable de $\underline{X}$.
En particulier, ne dépend ni d'autres VA (non observées) ni de θ.

Notations / vocabulaire

Notations. On utilisera souvent les notations :

- $\mathbb{E}_\theta(.)$ (espérance),
- $\mathrm{var}_\theta(.)$ (variance ou matrice de covariance),
- $f_\theta(.)$ (densité), ...

pour rappeler que ces opérateurs ou fonctions dépendent d'une loi $\mathbb{P}_\theta$ pour une valeur particulière de θ.

Définition : Statistique

Une statistique est une variable aléatoire (souvent scalaire ou vectorielle) calculée à partir de $\underline{X}$ seulement*.

Exemple : l'estimateur $\hat{\eta}^{(1)} = \bar{X}$ est une statistique.

* Techniquement : s'exprime comme une fonction mesurable de $\underline{X}$.
En particulier, ne dépend ni d'autres VA (non observées) ni de θ.

Évaluation numérique des performances de $\hat{\eta}^{(1)}$

En simulation, (presque) tout est permis !

- on choisit une valeur particulière de η (ici $\eta_* = 11,4$), puis
- on simule sur ordinateur des n-échantillons pour un grand nombre m de réalisations (ici $m = 10000$).

Remarques

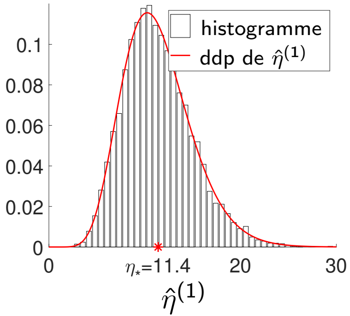
- Les estimations sont, ici, assez peu précises.
- Il serait pertinent de fournir des intervalles de confiance.
- On peut, dans ce cas simple, calculer analytiquement la densité de $\hat{\eta}^{(1)}$.

loi gamma

Évaluation numérique des performances de $\hat{\eta}^{(1)}$

En simulation, (presque) tout est permis !

- on choisit une valeur particulière de η (ici $\eta_* = 11,4$), puis
- on simule sur ordinateur des n-échantillons pour un grand nombre m de réalisations (ici $m = 10000$).



Remarques

- Les estimations sont, ici, assez peu précises.
- Il serait pertinent de fournir des intervalles de confiance.
- On peut, dans ce cas simple, calculer analytiquement la densité de $\hat{\eta}^{(1)}$.

loi gamma

$\hat{\eta}^{(2)}$: un deuxième estimateur

Avec les mêmes considérations que précédemment :

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}_\theta\left(X_1^2\right) = \frac{2}{\theta^2} = 2\eta^2,$$

donc utiliser $\hat{\eta}^{(2)} = \sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2}$ semble également « raisonnable ».

Application numérique. $\hat{\eta}^{(2)} = 11.2228$

Questions

- Comment comparer deux estimateurs ?
- Existe-t-il un estimateur « meilleur » que tous les autres ?
- Comment construire de « bons » estimateurs ?

$\hat{\eta}^{(2)}$: un deuxième estimateur

Avec les mêmes considérations que précédemment :

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}_\theta\left(X_1^2\right) = \frac{2}{\theta^2} = 2\eta^2,$$

donc utiliser $\hat{\eta}^{(2)} = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} X_i^2}$ semble également « raisonnable ».

Application numérique. $\hat{\eta}^{(2)} = 11.2228$

Questions

- Comment comparer deux estimateurs ?
- Existe-t-il un estimateur « meilleur » que tous les autres ?
- Comment construire de « bons » estimateurs ?

Plan du cours

Cadre mathématique

Pour toute la section :

- on considère un modèle statistique

$$\mathscr{M} = \left(\underline{\mathcal{X}}, \underline{\mathscr{A}}, \left\{\mathbb{P}_\theta^{\underline{X}},\, \theta \in \Theta\right\}\right),$$

  le plus souvent paramétrique ($\Theta \subset \mathbb{R}^p$) ;

- lorsque $\underline{X}$ est un n-échantillon iid, on note
  - $\underline{X} = (X_1, \ldots, X_n)$
  - $\underline{\mathcal{X}} = \mathcal{X}^n$, avec $\mathcal{X} = \mathbb{R}$ ou $\mathcal{X} = \mathbb{R}^d$,
  - $\mathbb{P}_\theta^{\underline{X}} = \mathrm{P}_\theta^{\otimes n}$;

- on veut estimer un « paramètre d'intérêt » :
  - soit θ lui-même,
  - soit, plus généralement, $\eta = g(\theta)$.

Cadre mathématique

Pour toute la section :

- on considère un modèle statistique

$$\mathscr{M} = \left(\underline{\mathcal{X}}, \underline{\mathscr{A}}, \left\{\mathbb{P}_\theta^{\underline{X}}, \, \theta \in \Theta\right\}\right),$$

  le plus souvent paramétrique ($\Theta \subset \mathbb{R}^p$) ;

- lorsque $\underline{X}$ est un *n*-échantillon iid, on note
  - $\underline{X} = (X_1, \ldots, X_n)$
  - $\underline{\mathcal{X}} = \mathcal{X}^n$, avec $\mathcal{X} = \mathbb{R}$ ou $\mathcal{X} = \mathbb{R}^d$,
  - $\mathbb{P}_\theta^{\underline{X}} = \mathrm{P}_\theta^{\otimes n}$;

- on veut estimer un « paramètre d'intérêt » :
  - soit θ lui-même,
  - soit, plus généralement, $\eta = g(\theta)$.

Cadre mathématique

Pour toute la section :

- on considère un modèle statistique

$$\mathscr{M} = \left(\underline{\mathcal{X}}, \underline{\mathscr{A}}, \left\{\mathbb{P}_{\theta}^{\underline{X}},\, \theta \in \Theta\right\}\right),$$

  le plus souvent paramétrique ($\Theta \subset \mathbb{R}^p$) ;

- lorsque $\underline{X}$ est un n-échantillon iid, on note
  - $\underline{X} = (X_1, \ldots, X_n)$
  - $\underline{\mathcal{X}} = \mathcal{X}^n$, avec $\mathcal{X} = \mathbb{R}$ ou $\mathcal{X} = \mathbb{R}^d$,
  - $\mathbb{P}_{\theta}^{\underline{X}} = \mathrm{P}_{\theta}^{\otimes n}$;

- on veut estimer un « paramètre d'intérêt » :
  - soit θ lui-même,
  - soit, plus généralement, $\eta = g(\theta)$.

Plan du cours

Méthode de substitution

Supposons que

- l'on ait déjà à notre disposition un estimateur $\hat{\eta}$ de $\eta = g(\theta)$
- et que l'on veuille estimer une autre grandeur d'intérêt η' qui s'écrit : $\eta' = h(\eta)$, avec h une fonction continue.

Méthode de substitution

La méthode de substitution consiste à considérer

$$\hat{\eta}' = h(\hat{\eta}) \text{ comme estimateur de } \eta'.$$

Méthode de substitution

Supposons que

- l'on ait déjà à notre disposition un estimateur $\hat{\eta}$ de $\eta = g(\theta)$
- et que l'on veuille estimer une autre grandeur d'intérêt η' qui s'écrit : $\eta' = h(\eta)$, avec h une fonction continue.

Méthode de substitution

La méthode de substitution consiste à considérer

$$\hat{\eta}' = h\left(\hat{\eta}\right) \text{ comme estimateur de } \eta'.$$

Application « fiabilité composant »

Rappel : $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta), \quad \theta > 0.$

On s'intéresse à la probabilité de tomber en panne avant t_0 :

$$\Rightarrow \quad \eta' = \mathbb{P}_\theta (X_1 \leq t_0) = \int_0^{t_0} \theta \exp(-\theta x) \mathrm{d}x$$
$$= 1 - \exp(-\theta t_0) = 1 - \exp\left(-\frac{t_0}{\eta}\right).$$

En utilisant $\hat{\eta}^{(1)} = \bar{X}$ comme estimateur de $\eta = \frac{1}{\theta}$, on obtient

$$\hat{\eta}' = 1 - \exp\left(-\frac{t_0}{\bar{X}}\right).$$

Application « fiabilité composant »

Rappel : $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta), \quad \theta > 0.$

On s'intéresse à la probabilité de tomber en panne avant t_0 :

$$\Rightarrow \quad \eta' = \mathbb{P}_\theta(X_1 \le t_0) = \int_0^{t_0} \theta \exp(-\theta x)\mathrm{d}x = 1 - \exp(-\theta t_0) = 1 - \exp\left(-\frac{t_0}{\eta}\right).$$

En utilisant $\hat{\eta}^{(1)} = \bar{X}$ comme estimateur de $\eta = \frac{1}{\theta}$, on obtient

$$\hat{\eta}' = 1 - \exp\left(-\frac{t_0}{\bar{X}}\right).$$

Mesure empirique

Soient $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathbb{P}^{X_1}$.

Rappel : la mesure de Dirac en $x \in \mathcal{X}$ est la mesure δ_x définie par

$$\forall A \in \mathscr{A}, \quad \delta_x(A) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{sinon.} \end{cases}$$

Définition : mesure empirique

La mesure empirique est la mesure (aléatoire) définie par :

$$\hat{\mathbb{P}}^{X_1} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

Intérêt : La mesure empirique s'interprète comme un estimateur de $\mathbb{P}^{X_1}$ ➠ permet de construire d'autres estimateurs par substitution.

Mesure empirique

Soient $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathbb{P}^{X_1}$.

Rappel : la mesure de Dirac en $x \in \mathcal{X}$ est la mesure δ_x définie par

$$\forall A \in \mathscr{A}, \quad \delta_x(A) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{sinon.} \end{cases}$$

Définition : mesure empirique

La mesure empirique est la mesure (aléatoire) définie par :

$$\hat{\mathbb{P}}^{X_1} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

Intérêt : La mesure empirique s'interprète comme un estimateur de $\mathbb{P}^{X_1}$ ➠ permet de construire d'autres estimateurs par substitution.

Exemple : estimateur du moment d'ordre k

Supposons $X_1 \in L^k$. Alors

$$m_k = \mathbb{E}\left(X_1^k\right) = \mathscr{G}\left(\mathbb{P}^{X_1}\right)$$

est bien défini, avec $\mathscr{G}(\mu) = \int_{\mathcal{X}} x^k \mu(\mathrm{d}x)$. Par substitution :

$$\hat{m}_k = \mathscr{G}\left(\hat{\mathbb{P}}^{X_1}\right) = \int_{\mathcal{X}} x^k \, \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\mathrm{d}x) = \frac{1}{n} \sum_{i=1}^{n} X_i^k .$$

Autres exemples :

- variance empirique exercice 3
- fonction de répartition empirique complément

Exemple : estimateur du moment d'ordre k

Supposons $X_1 \in L^k$. Alors

$$m_k = \mathbb{E}\left(X_1^k\right) = \mathscr{G}\left(\mathbb{P}^{X_1}\right)$$

est bien défini, avec $\mathscr{G}(\mu) = \int_{\mathcal{X}} x^k \mu(\mathrm{d}x)$. Par substitution :

$$\hat{m}_k = \mathscr{G}\left(\hat{\mathbb{P}}^{X_1}\right) = \int_{\mathcal{X}} x^k \, \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\mathrm{d}x) = \frac{1}{n} \sum_{i=1}^{n} X_i^k .$$

Autres exemples :

- variance empirique — exercice 3
- fonction de répartition empirique — complément

Exemple : estimateur du moment d'ordre k

Supposons $X_1 \in L^k$. Alors

$$m_k = \mathbb{E}\left(X_1^k\right) = \mathscr{G}\left(\mathbb{P}^{X_1}\right)$$

est bien défini, avec $\mathscr{G}(\mu) = \int_{\mathcal{X}} x^k \mu(\mathrm{d}x)$. Par substitution :

$$\hat{m}_k = \mathscr{G}\left(\hat{\mathbb{P}}^{X_1}\right) = \int_{\mathcal{X}} x^k \, \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(\mathrm{d}x) = \frac{1}{n}\sum_{i=1}^{n} X_i^k .$$

Autres exemples :

- variance empirique — exercice 3
- fonction de répartition empirique — complément

Plan du cours

Méthode des moments

Supposons que

- $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, avec $\theta \in \Theta$;
- le modèle est paramétrique : $\Theta \subset \mathbb{R}^p$,
- et on veut estimer θ lui-même.

Considérons la fonction

$$\begin{array}{rccl} h: & \Theta \subset \mathbb{R}^p & \to & h(\Theta) \subset \mathbb{R}^p, \\ & \theta & \mapsto & h(\theta) = \begin{pmatrix} \mathbb{E}_\theta(X_1) \\ \vdots \\ \mathbb{E}_\theta(X_1^p) \end{pmatrix}. \end{array}$$

Remarque : on peut parfois choisir autre chose que les p premiers moments.

Méthode des moments

Supposons que

- $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, avec $\theta \in \Theta$;
- le modèle est paramétrique : $\Theta \subset \mathbb{R}^p$,
- et on veut estimer θ lui-même.

Considérons la fonction

$$\begin{array}{rccl} h : & \Theta \subset \mathbb{R}^p & \to & h(\Theta) \subset \mathbb{R}^p, \\ & \theta & \mapsto & h(\theta) = \begin{pmatrix} \mathbb{E}_\theta\left(X_1\right) \\ \vdots \\ \mathbb{E}_\theta\left(X_1^p\right) \end{pmatrix}. \end{array}$$

Remarque : on peut parfois choisir autre chose que les p premiers moments.

Méthode des moments (suite)

Supposons $h : \Theta \to h(\Theta)$ injective, et donc bijective.

Méthode des moments

La méthode des moments consiste :

- à estimer les p premiers moments $\hat{m}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$, $k \leq p$,
- puis appliquer h^{-1} pour former un estimateur de θ.

D'où l'estimateur des moments : $\hat{\theta} = h^{-1}(\hat{m}_{1:p})$, où

$$\hat{m}_{1:p} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n X_i^p \end{pmatrix}.$$

Remarque : n'est bien défini que si $\hat{m}_{1:p} \in h(\Theta)$ $\mathbb{P}_\theta$-ps, pour tout θ.

Sinon, minimisation d'une distance (méthode des moments généralisée).

Méthode des moments (suite)

Supposons $h : \Theta \to h(\Theta)$ injective, et donc bijective.

Méthode des moments

La méthode des moments consiste :

- à estimer les p premiers moments $\hat{m}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$, $k \leq p$,
- puis appliquer h^{-1} pour former un estimateur de θ.

D'où l'estimateur des moments : $\hat{\theta} = h^{-1}(\hat{m}_{1:p})$, où

$$\hat{m}_{1:p} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n X_i^p \end{pmatrix}.$$

Remarque : n'est bien défini que si $\hat{m}_{1:p} \in h(\Theta)$ $\mathbb{P}_\theta$-ps, pour tout θ.

Sinon, minimisation d'une distance (méthode des moments généralisée).

Méthode des moments (suite)

Supposons $h : \Theta \to h(\Theta)$ injective, et donc bijective.

Méthode des moments

La méthode des moments consiste :

- à estimer les p premiers moments $\hat{m}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$, $k \leq p$,
- puis appliquer h^{-1} pour former un estimateur de θ.

D'où l'estimateur des moments : $\hat{\theta} = h^{-1}(\hat{m}_{1:p})$, où

$$\hat{m}_{1:p} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n X_i^p \end{pmatrix}.$$

Remarque : n'est bien défini que si $\hat{m}_{1:p} \in h(\Theta)$ $\mathbb{P}_\theta$-ps, pour tout θ.

Sinon, minimisation d'une distance (méthode des moments généralisée).

Estimateur des moments : exemples

Application « fiabilité composant »

On a $\mathbb{E}_\theta(X_1) = \theta^{-1}$ (loi exponentielle), donc

$$\theta = \left(\mathbb{E}_\theta(X_1)\right)^{-1} \quad \text{et} \quad \hat{\theta} = \left(\bar{X}\right)^{-1}.$$

Autre exemple : n-échantillon gaussien

✏ TD 1, ex. 1.1

$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, avec $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$.

En considérant les deux premiers moments, on trouve :

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 \end{pmatrix}.$$

Estimateur des moments : exemples

Application « fiabilité composant »

On a $\mathbb{E}_\theta(X_1) = \theta^{-1}$ (loi exponentielle), donc

$$\theta = \left(\mathbb{E}_\theta(X_1)\right)^{-1} \quad \text{et} \quad \hat{\theta} = \left(\bar{X}\right)^{-1}.$$

Autre exemple : n-échantillon gaussien ✏ TD 1, ex. 1.1

$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, avec $\theta = \left(\mu, \sigma^2\right) \in \mathbb{R} \times \mathbb{R}_+^*$.

En considérant les deux premiers moments, on trouve :

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 \end{pmatrix}.$$

Plan du cours

Fonction de vraisemblance

On suppose le modèle dominé : $\mathbb{P}_\theta^X$ admet une densité f_θ par rapport à une mesure ν sur $\underline{\mathcal{X}}$, pour tout $\theta \in \Theta$.

rappel : densité

Définition : vraisemblance

La vraisemblance est la fonction :

$$\begin{array}{rccl} \mathcal{L} : & \Theta \times \underline{\mathcal{X}} & \to & \mathbb{R}_+ \\ & (\theta; \underline{x}) & \mapsto & f_\theta(\underline{x}) \end{array}$$

On appelle log-vraisemblance la fonction $\ln \mathcal{L}$.

Remarque. Si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, alors,

$$\mathcal{L}(\theta; \underline{x}) = \prod_{i=1}^n f_\theta(x_i), \qquad \text{donc} \quad \ln \mathcal{L}(\theta; \underline{x}) = \sum_{i=1}^n \ln f_\theta(x_i).$$

(abus de notation usuel : ici $f_\theta = f_\theta^{X_1}$)

Fonction de vraisemblance

On suppose le modèle dominé : $\mathbb{P}_\theta^X$ admet une densité f_θ par rapport à une mesure ν sur $\underline{\mathcal{X}}$, pour tout $\theta \in \Theta$.

rappel : densité

Définition : vraisemblance

La vraisemblance est la fonction :

$$\begin{array}{rccc} \mathcal{L} : & \Theta \times \underline{\mathcal{X}} & \to & \mathbb{R}_+ \\ & (\theta; \underline{x}) & \mapsto & f_\theta(\underline{x}) \end{array}$$

On appelle log-vraisemblance la fonction $\ln \mathcal{L}$.

Remarque. Si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, alors,

$$\mathcal{L}(\theta; \underline{x}) = \prod_{i=1}^n f_\theta(x_i), \qquad \text{donc} \qquad \ln \mathcal{L}(\theta; \underline{x}) = \sum_{i=1}^n \ln f_\theta(x_i).$$

(abus de notation usuel : ici $f_\theta = f_\theta^{X_1}$)

Fonction de vraisemblance

On suppose le modèle dominé : $\mathbb{P}_\theta^X$ admet une densité f_θ par rapport à une mesure ν sur $\underline{\mathcal{X}}$, pour tout $\theta \in \Theta$. rappel : densité

Définition : vraisemblance

La vraisemblance est la fonction :

$$\begin{array}{cccc} \mathcal{L} : & \Theta \times \underline{\mathcal{X}} & \to & \mathbb{R}_+ \\ & (\theta; \underline{x}) & \mapsto & f_\theta(\underline{x}) \end{array}$$

On appelle log-vraisemblance la fonction $\ln \mathcal{L}$.

Remarque. Si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{P}_\theta$, alors,

$$\mathcal{L}(\theta; \underline{x}) = \prod_{i=1}^n f_\theta(x_i), \qquad \text{donc} \quad \ln \mathcal{L}(\theta; \underline{x}) = \sum_{i=1}^n \ln f_\theta(x_i).$$

(abus de notation usuel : ici $f_\theta = f_\theta^{X_1}$)

Estimateur du maximum de vraisemblance

Définition : EMV

Si $\hat{\theta}$ maximise la fonction $\theta \mapsto \mathcal{L}(\theta; \underline{X})$, alors
$\hat{\theta}$ est un estimateur du maximum de vraisemblance (EMV) de θ.

Remarques :

- Existence et unicité de l'EMV : pas assurées en général.
- De manière équivalente, $\hat{\theta}$ maximise $\theta \mapsto \ln \mathcal{L}(\theta; \underline{X})$.
- Supposons $\Theta \subset \mathbb{R}^p$. Si $\mathcal{L}$ est de classe C^1 en θ sur $\mathrm{int}(\Theta)$, une condition nécessaire pour qu'un point intérieur $\hat{\theta} \in \mathrm{int}(\Theta)$ maximise la vraisemblance est :

$$\left(\nabla_\theta (\ln \mathcal{L})\right)\left(\hat{\theta}; \underline{X}\right) = 0.$$

On l'appelle équation de vraisemblance.

Estimateur du maximum de vraisemblance

Définition : EMV

Si $\hat{\theta}$ maximise la fonction $\theta \mapsto \mathcal{L}(\theta; \underline{X})$, alors
$\hat{\theta}$ est un estimateur du maximum de vraisemblance (EMV) de θ.

Remarques :

- Existence et unicité de l'EMV : pas assurées en général.
- De manière équivalente, $\hat{\theta}$ maximise $\theta \mapsto \ln \mathcal{L}(\theta; \underline{X})$.
- Supposons $\Theta \subset \mathbb{R}^p$. Si $\mathcal{L}$ est de classe C^1 en θ sur $\mathrm{int}(\Theta)$, une condition nécessaire pour qu'un point intérieur $\hat{\theta} \in \mathrm{int}(\Theta)$ maximise la vraisemblance est :

$$\left(\nabla_\theta (\ln \mathcal{L})\right)\left(\hat{\theta}; \underline{X}\right) = 0.$$

On l'appelle équation de vraisemblance.

Estimateur du maximum de vraisemblance

Définition : EMV

Si $\hat{\theta}$ maximise la fonction $\theta \mapsto \mathcal{L}(\theta; \underline{X})$, alors
$\hat{\theta}$ est un estimateur du maximum de vraisemblance (EMV) de θ.

Remarques :

- Existence et unicité de l'EMV : pas assurées en général.
- De manière équivalente, $\hat{\theta}$ maximise $\theta \mapsto \ln \mathcal{L}(\theta; \underline{X})$.
- Supposons $\Theta \subset \mathbb{R}^p$. Si $\mathcal{L}$ est de classe C^1 en θ sur $\mathrm{int}(\Theta)$, une condition nécessaire pour qu'un point intérieur $\hat{\theta} \in \mathrm{int}(\Theta)$ maximise la vraisemblance est :

$$\left(\nabla_\theta \left(\ln \mathcal{L}\right)\right)\left(\hat{\theta}; \underline{X}\right) \;=\; 0.$$

On l'appelle équation de vraisemblance.

Exemple d'EMV : application « fiabilité composant »

Pour $x_1, \ldots, x_n \geq 0$, on a $\mathcal{L}(\theta; \underline{x}) = \prod_{i=1}^{n} \theta \exp(-\theta x_i)$, donc

$$\ln \mathcal{L}(\theta; \underline{x}) = n \ln(\theta) - \theta \sum_{i=1}^{n} x_i.$$

Condition de stationnarité (« équation de vraisemblance »)

$$\frac{\partial(\ln \mathcal{L})}{\partial \theta}(\theta; \underline{x}) = 0 \iff \frac{n}{\theta} - \sum_{i=1}^{n} x_i = 0.$$

✏ Si $\sum_{i=1}^{n} x_i > 0$, unique solution dans $\Theta = \mathbb{R}_+^*$ en $\theta = n \left(\sum_{i=1}^{n} x_i\right)^{-1}$.

✏ Il s'agit bien d'un maximum de la vraisemblance (cf. signe de la dérivée).

➠ Puisque $\sum_{i=1}^{n} X_i > 0$ p.s., un unique EMV existe : $\hat{\theta} = \left(\bar{X}\right)^{-1}$.

Remarque : même estimateur que par la méthode des moments.

Exemple d'EMV : application « fiabilité composant »

Pour $x_1, \ldots, x_n \geq 0$, on a $\mathcal{L}(\theta; \underline{x}) = \prod_{i=1}^{n} \theta \exp(-\theta x_i)$, donc

$$\ln \mathcal{L}(\theta; \underline{x}) = n \ln(\theta) - \theta \sum_{i=1}^{n} x_i.$$

Condition de stationnarité (« équation de vraisemblance »)

$$\frac{\partial(\ln \mathcal{L})}{\partial \theta}(\theta; \underline{x}) = 0 \iff \frac{n}{\theta} - \sum_{i=1}^{n} x_i = 0.$$

✏ Si $\sum_{i=1}^{n} x_i > 0$, unique solution dans $\Theta = \mathbb{R}_+^*$ en $\theta = n \left(\sum_{i=1}^{n} x_i\right)^{-1}$.

✏ Il s'agit bien d'un maximum de la vraisemblance (cf. signe de la dérivée).

➠ Puisque $\sum_{i=1}^{n} X_i > 0$ p.s., un unique EMV existe : $\hat{\theta} = \left(\bar{X}\right)^{-1}$.

Remarque : même estimateur que par la méthode des moments.

Exemple d'EMV : application « fiabilité composant »

Pour $x_1, \ldots, x_n \geq 0$, on a $\mathcal{L}(\theta; \underline{x}) = \prod_{i=1}^n \theta \exp(-\theta x_i)$, donc

$$\ln \mathcal{L}(\theta; \underline{x}) = n \ln(\theta) - \theta \sum_{i=1}^n x_i.$$

Condition de stationnarité (« équation de vraisemblance »)

$$\frac{\partial(\ln \mathcal{L})}{\partial \theta}(\theta; \underline{x}) = 0 \iff \frac{n}{\theta} - \sum_{i=1}^n x_i = 0.$$

✏ Si $\sum_{i=1}^n x_i > 0$, unique solution dans $\Theta = \mathbb{R}_+^*$ en $\theta = n\left(\sum_{i=1}^n x_i\right)^{-1}$.

✏ Il s'agit bien d'un maximum de la vraisemblance (cf. signe de la dérivée).

➠ Puisque $\sum_{i=1}^n X_i > 0$ p.s., un unique EMV existe : $\hat{\theta} = \left(\bar{X}\right)^{-1}$.

Remarque : même estimateur que par la méthode des moments.

Exemple d'EMV : n-échantillon iid gaussien, $\theta = \left(\mu, \sigma^2\right)$

Même démarche que dans l'exemple précédent.

1. On écrit la log-vraisemblance :

$$\ln \mathcal{L}\left(\theta; \underline{x}\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2},$$

2. Les conditions de stationnarité fournissent :

$$\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_i \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 \end{pmatrix}.$$

3. On peut montrer que le maximum est bien atteint en ce point.

✏ TD 1, ex. 1.1

Remarque : même estimateur que par la méthode des moments.

Exemple d'EMV : n-échantillon iid gaussien, $\theta = (\mu, \sigma^2)$

Même démarche que dans l'exemple précédent.

1. On écrit la log-vraisemblance :

$$\ln \mathcal{L}(\theta; \underline{x}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2},$$

2. Les conditions de stationnarité fournissent :

$$\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \frac{1}{n}\sum_{i=1}^n (X_i - \hat{\mu})^2 \end{pmatrix}.$$

3. On peut montrer que le maximum est bien atteint en ce point.

✏ TD 1, ex. 1.1

Remarque : même estimateur que par la méthode des moments.

Exemple d'EMV : n-échantillon iid gaussien, $\theta = (\mu, \sigma^2)$

Même démarche que dans l'exemple précédent.

1. On écrit la log-vraisemblance :

$$\ln \mathcal{L}(\theta; \underline{x}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2},$$

2. Les conditions de stationnarité fournissent :

$$\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_i \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 \end{pmatrix}.$$

3. On peut montrer que le maximum est bien atteint en ce point.

✏ TD 1, ex. 1.1

Remarque : même estimateur que par la méthode des moments.

Conclusion du cours et transition vers la séquence suivante

Nous avons vu et développerons en TD 1 :

- le cadre général et formel de la statistique inférentielle,
- les méthodes classiques de construction d'estimateurs.

Nous aborderons dans la prochaine séquence :

- la quantification de la performance d'un estimateur,
- la comparaison d'estimateurs entre eux,
- l'approche asymptotique ($n \to \infty$).

Conclusion du cours et transition vers la séquence suivante

Nous avons vu et développerons en TD 1 :

- le cadre général et formel de la statistique inférentielle,
- les méthodes classiques de construction d'estimateurs.

Nous aborderons dans la prochaine séquence :

- la quantification de la performance d'un estimateur,
- la comparaison d'estimateurs entre eux,
- l'approche asymptotique ($n \to \infty$).

Plan du cours

Plan du cours

Exercice 1 (modèle de Bernoulli)

corrigé

Soient $X_1, \ldots, X_n$ un n-échantillon d'observations binaires, indépendantes et identiquement distribuées selon la loi de Bernoulli $\mathrm{Ber}(p)$, avec $p \in [0, 1]$.

Questions

1. Préciser un modèle statistique formel $\mathscr{M} = (\underline{\mathcal{X}},\ \underline{\mathscr{A}},\ \mathscr{P}^{\underline{X}})$ correspondant à cette description.
2. Construire un estimateur de p par la méthode des moments.
3. Construire un estimateur de p par maximum de vraisemblance.
4. Calculer l'espérance et la variance de $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Exercice 2 (loi log-normale)

corrigé

Élements de contexte

L'association entre leucémie infantile et exposition aux champs magnétiques très basses fréquences (essentiellement due aux ouvrages et appareils électriques) est statistiquement significative pour une exposition résidentielle, moyennée sur 24 h, dont les niveaux sont supérieurs à $0.4\mu T$.

Source : anses

Hypothèse de modélisation. Pour des logements situés à moins de 50 mètres de lignes HT, l'exposition résidentielle moyennée sur 24h suit une loi log-normale.

loi log-normale

Exercice 2 (loi log-normale)

corrigé

Soient

- $\underline{X} = (X_1, \ldots, X_n)$: n-échantillon de loi log-normale $\mathscr{LN}(\mu, \sigma^2)$, avec $\sigma^2 > 0$ **connu**.
- p_0 : probabilité qu'une VA de loi $\mathscr{LN}(\mu, \sigma^2)$ dépasse le seuil $s_0 = 0.4\mu$T.

Questions

1. Construire un estimateur de μ par maximum de vraisemblance.
2. Utiliser la méthode de substitution pour construire un estimateur de p_0.
3. L'estimateur de p_0 obtenu converge-t'il presque sûrement ? Le cas échéant, vers quelle limite ?

Exercice 3 (variance empirique)

corrigé

Soient $X_1, \dots, X_n$ un n-échantillon d'observations réelles, indépendantes et identiquement distribuées, dont la loi admet un moment d'ordre deux.

On note $\mathbb{M}$ l'ensemble des mesures de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ admettant un moment d'ordre deux.

Questions

1. Montrer que $\operatorname{var}(X_1) = \mathscr{G}\left(\mathbb{P}^{X_1}\right)$, où $\mathscr{G}$ est une fonction définie sur $\mathbb{M}$ que l'on précisera.
2. En déduire par substitution un estimateur de la variance.
3. Étudier la convergence de l'estimateur lorsque $n \to +\infty$.

retour au slide 27

Plan du cours

❶ Modèle statistique $\mathscr{M} = \left(\underline{\mathcal{X}},\ \underline{\mathscr{A}},\ \mathscr{P}^{\underline{X}}\right)$

L'ensemble « naturel » (minimal) pour décrire les valeurs d'une variable binaire est $\mathcal{X} = \{0,1\}$.

➠ $\underline{\mathcal{X}} = \{0,1\}^n$ pour un n-échantillon

Sur un ensemble fini ou dénombre, on utilise (sauf exception) la tribu discrète, c'est-à-dire l'ensemble des parties de $\underline{\mathcal{X}}$.

➠ $\underline{\mathscr{A}} = \mathcal{P}\left(\{0,1\}^n\right) = \mathcal{P}\left(\{0,1\}\right)^{\otimes n}$

La loi d'un n-uplet $(X_1, \ldots, X_n)$ de VA indépendantes est la mesure produit $\mathrm{P}^{X_1} \otimes \cdots \otimes \mathrm{P}^{X_n}$.

➠ $\mathscr{P}^{\underline{X}} = \left\{\mathrm{Ber}(p)^{\otimes n},\ p \in [0,1]\right\}$

Remarque : on aurait pu prendre aussi $\underline{\mathcal{X}} = \mathbb{R}^n$, $\underline{\mathscr{A}} = \mathcal{B}(\mathbb{R}^n)$.

Corrigé de l'exercice 1

retour au sujet

❷ Méthodes des moments

Si $X \sim \mathrm{Ber}(p)$, alors $\mathbb{E}_p(X) = p$.

➠ La méthode des moments, appliquée au moment d'ordre un, donne donc directement l'estimateur $\hat{p}_n = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}_n$.

❸ Maximum de vraisemblance

On écrit la vraisemblance :

$$\begin{aligned}\mathcal{L}(p; \underline{X}) &= \prod_{i=1}^n p^{X_i}(1-p)^{1-X_i} \\ &= p^N(1-p)^{n-N},\end{aligned}$$

où $N = \sum_{i=1}^n X_i$ et $0^0 = 1$,

puis la log-vraisemblance pour $p \in]0,1[$:

$$\begin{aligned}\ell(p;\underline{X}) &= \ln(\mathcal{L}(p;\underline{X}))\\ &= N\ln(p) + (n-N)\ln(1-p).\end{aligned}$$

La log-vraisemblance est dérivable sur $]0,1[$, de dérivée

$$\begin{aligned}\frac{\partial \ell}{\partial p}(p;\underline{X}) &= \frac{N}{p} - \frac{n-N}{1-p}\\ &= \frac{n}{p(1-p)} \cdot \left(\bar{X}_n - p\right).\end{aligned}$$

On a $\frac{\partial \ell}{\partial p}(p;\underline{X}) > 0$ ssi $p < N/n = \bar{X}_n$,
$\frac{\partial \ell}{\partial p}(p;\underline{X}) < 0$ ssi $p > N/n = \bar{X}_n$.

Corrigé de l'exercice 1

➠ retour au sujet

Si $\bar{X}_n = 0$, la log-vraisemblance est strictement décroissante

➠ la vraisemblance est maximale en $p = 0$.

Si $\bar{X}_n = 1$, la log-vraisemblance est strictement croissante

➠ la vraisemblance est maximale en $p = 1$.

Si $0 < \bar{X}_n < 1$, la log-vraisemblance est maximale en $p = \bar{X}_n$.

En résumé : $\hat{p}_n = \bar{X}_n$ est l'unique EMV.

Remarque : la log-vraisemblance peut prendre des valeurs infinies en $p = 0$ et/ou $p = 1$, mais la vraisemblance elle-même est bien définie et continue sur $[0, 1]$.

retour au sujet

❹ Espérance et variance de $\bar{X}$

Rappels

- $\mathbb{E}_p(X_1) = p$ et $\mathrm{var}_p(X_1) = p(1-p)$.
- indépendance $\Rightarrow$ décorrélation $\Rightarrow \mathrm{var}(\sum_i X_i) = \sum_i \mathrm{var}(X_i)$.

En utilisant que les X_i sont identiquement distribuées :

$$\mathbb{E}_p\left(\bar{X}_n\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_p[X_1] = p.$$

En utilisant que les X_i sont IID :

$$\mathrm{var}_p(\bar{X}_n) = \frac{1}{n^2}\mathrm{var}_p\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}_p\left(X_i\right) = \frac{p(1-p)}{n}.$$

❶ Maximum de vraisemblance

On commence par écrire la log-vraisemblance :

$$\ell(\mu; \underline{X}) = -\frac{n}{2}\ln(2\pi\sigma_0^2) - \sum_{i=1}^{n}\ln(X_i) + \frac{(\ln(X_i)-\mu)^2}{2\sigma^2}\mathbb{1}_{(\mathbb{R}_*^+)^n}(\underline{x}).$$

La log-vraisemblable est dérivable, de dérivée (pour $X_1, \ldots, X_n > 0$) :

$$\frac{\partial \ell}{\partial \mu}(\mu; \underline{X}) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(\ln(X_i) - \mu)$$

On vérifie que $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\ln(X_i)$ est l'EMV puisque :

$$\frac{\partial \ell}{\partial \mu}(\mu; \underline{X}) > 0 \quad \text{pour } \mu < \hat{\mu}, \qquad , \frac{\partial \ell}{\partial \mu}(\mu; \underline{X}) < 0 \quad \text{pour } \mu > \hat{\mu}.$$

Corrigé de l'exercice 2

retour au sujet

❷ On exprimer la probabilité de dépasser s_0 en fonction de μ :

$$\begin{aligned} p_0 &= \mathbb{P}(X > s_0) \quad \text{avec } X \sim \mathscr{LN}(\mu, \sigma^2) \\ &= 1 - F_{\mu,\sigma}(s_0) \\ &= 1 - \Phi_0\left(\frac{\ln(s_0) - \mu}{\sigma}\right). \end{aligned}$$

Puis on construit un estimateur de p_0 par substitution, à partir de $\hat{\mu}$:

$$\hat{p}_0 = 1 - \Phi_0\left(\frac{\ln(s_0) - \hat{\mu}}{\sigma}\right).$$

retour au sujet

❸ Soit $Z_i = \ln(X_i)$, $i \geq 1$. Les Z_i sont IID, et admettent un moment d'ordre un égal à μ, puisque $Z_i \sim \mathscr{N}(\mu, \sigma^2)$.

Donc, par la loi forte des grands nombres :

$$\hat{\mu} \;=\; \frac{1}{n}\sum_{i=1}^{n} \ln(X_i) \;\xrightarrow[n\to\infty]{\text{ps}}\; \mathbb{E}(Z_1) = \mu.$$

D'où, par continuité de $h : \mu \mapsto 1 - \Phi_0\left(\frac{\ln(s_0) - \mu}{\sigma}\right)$,

$$\hat{p}_0 \;=\; h\left(\hat{\mu}\right) \;\xrightarrow[n\to\infty]{\text{ps}}\; h(\mu) = p_0.$$

Remarque. La convergence presque-sûre vers le paramètre d'intérêt s'appelle la *consistance forte* (cf. cours 3 à venir).

Corrigé de l'exercice 3

retour au sujet

❶ En utilisant Huygens-König et le théorème de transfert :

$$\mathrm{var}\left(X_1\right) = \mathbb{E}\left(X_1^2\right) - \mathbb{E}\left(X_1\right)^2 = \mathscr{G}\left(\mathbb{P}^{X_1}\right)$$

avec, pour tout $\mu \in \mathbb{M}$,

$$\mathscr{G}(\mu) = \int_{\mathcal{X}} x^2 \mu(\mathrm{d}x) - \left(\int_{\mathcal{X}} x \mu(dx)\right)^2.$$

❷ On applique le principe de substitution, en estimant $\mathbb{P}^{X_1}$ par la mesure empirique :

$$\hat{\mathbb{P}}_n^{X_1} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

Corrigé de l'exercice 3

retour au sujet

On trouve l'estimateur

$$\begin{aligned} S_n^2 &= \int_{\mathcal{X}} x^2 \, \hat{\mathbb{P}}_n^{X_1}(\mathrm{d}x) - \left(\int_{\mathcal{X}} x \, \hat{\mathbb{P}}_n^{X_1}(dx) \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n \left(X_i - \bar{X}\right)^2, \end{aligned}$$

que l'on appelle *variance empirique*.

❸ En appliquant la loi forte des grands nombres aux suites (X_i) et $\left(X_i^2\right)$, qui sont des suites IID de VA admettant un moment d'ordre un, on trouve

$$\bar{X} \xrightarrow{\text{ps}} \mathbb{E}(X_1), \qquad \frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{\text{ps}} \mathbb{E}(X_1^2),$$

Corrigé de l'exercice 3

retour au sujet

et donc

$$S_n^2 \xrightarrow{\text{ps}} \mathrm{var}(X_1).$$

Remarque : en revanche on n'a pas convergence dans L^2 en général, les X_i^2 n'ayant pas nécessairement un moment d'ordre deux (il faudrait pour cela que les X_i aient un moment d'ordre quatre).

Plan du cours

Plan du cours

La famille des lois Gamma

On dit que X suit la loi $\Gamma(p, \lambda)$, de paramètres $p > 0$ et $\lambda > 0$, si elle admet pour densité de probabilité

$$f(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} \exp(-\lambda x) \mathbb{1}_{\mathbb{R}^+}(x).$$

Moments

- moyenne : $\mathbb{E}_\theta(X) = \frac{p}{\lambda}$
- variance : $\mathrm{var}_\theta(X) = \frac{p}{\lambda^2}$

Cas particuliers

- $\mathcal{E}(\lambda) = \Gamma(p = 1, \lambda)$
- $\Gamma(p = \frac{n}{2}, \lambda = \frac{n}{2}) = \chi^2(n)$

Propriétés

- Soit $a > 0$. Si $X \sim \Gamma(p, \lambda)$, alors $aX \sim \Gamma\left(p, \frac{\lambda}{a}\right)$.
- Si X et Y sont indépendantes, avec $X \sim \Gamma(p, \lambda)$ et $Y \sim \Gamma(q, \lambda)$, alors $X + Y \sim \Gamma(p + q, \lambda)$.

retour au slide 21

La loi log-normale $\mathscr{LN}(\mu, \sigma^2)$

Définition

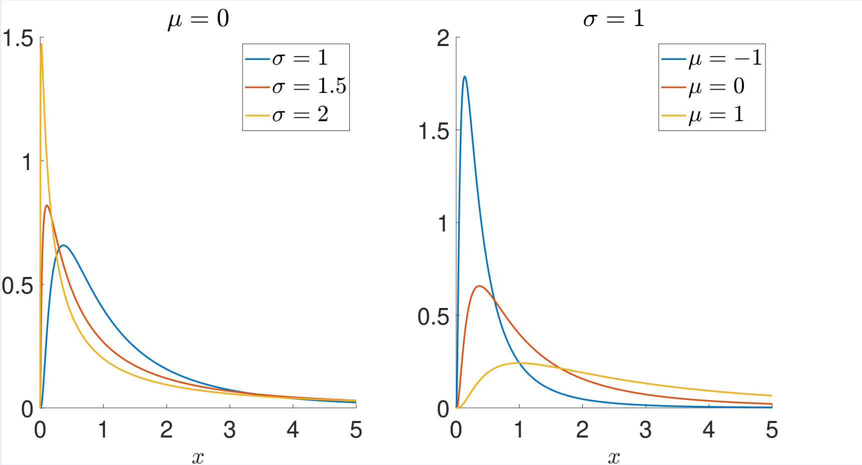
$X \sim \mathscr{LN}(\mu, \sigma^2)$, avec $\mu \in \mathbb{R}$ et $\sigma^2 > 0$, si elle admet la densité

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right) 1_{\mathbb{R}_*^+}(x).$$

Propriétés

- moyenne : $\mathbb{E}_{\mu,\sigma}(X) = \exp\left(\mu + \dfrac{\sigma^2}{2}\right)$
- variance : $\mathrm{var}_{\mu,\sigma}(X) = \left(\exp\left(\sigma^2\right) - 1\right) \exp\left(2\mu + \sigma^2\right)$
- fonction de répartition : $F_{\mu,\sigma} = \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)$, où Φ est la fonction de répartition de la loi $\mathscr{N}(0,1)$.
- $X \sim \mathscr{LN}(\mu, \sigma^2) \quad$ ssi $\quad \ln(X) \sim \mathscr{N}(\mu, \sigma^2)$.

Densité de la loi $\mathscr{LN}(\mu, \sigma^2)$



retour à l'exercice 2

Plan du cours

Rappel : Densité de probabilité par rapport à une mesure

Soit ν une mesure positive sur $(\underline{\mathcal{X}}, \underline{\mathscr{A}})$.

Définition : densité de probabilité

La loi $\mathbb{P}^{\underline{X}}$ d'une VA $\underline{X}$ à valeurs dans $(\underline{\mathcal{X}}, \underline{\mathscr{A}})$ admet une densité par rapport à ν s'il existe $f : \underline{X} \to \mathbb{R}_+$, $\underline{\mathscr{A}}$-mesurable et positive, tq

$$\forall A \in \underline{\mathscr{A}}, \qquad \mathbb{P}\left(\underline{X} \in A\right) = \mathbb{P}^{\underline{X}}(A) = \int_A f(\underline{x})\,\nu(\mathrm{d}\underline{x}).$$

➠ f est la densité de probabilité de la loi $\mathbb{P}^{\underline{X}}$ par rapport à ν.
➠ Elle vérifie $\int f\,\mathrm{d}\nu = 1$.

Dans la plupart des cas, les VA considérées seront :

- soit « continues » : mesure de référence ν = mesure de Lebesgue,
- soit discrètes : mesure de référence ν = mesure de comptage.

➠ retour au slide 31

Complément : la fonction de répartition empirique

Soit $x \in \mathbb{R}$. La fonction de répartition de X_1 en x est

$$F(x) = \mathbb{P}^{X_1}(X_1 \leq x) = \mathscr{G}_x\left(\mathbb{P}^{X_1}\right) \quad \text{avec} \quad \mathscr{G}_x(\mu) = \int_{-\infty}^{x} \mu(\mathrm{d}x).$$

D'ou, par substitution, la fonction de répartition empirique (FRE) :

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}.$$

Pour un n-échantillon $X_1, \ldots, X_n$ IID, de fonction de répartition F, on montre (théorème de Glivenko-Cantelli) que $\hat{F}_n \to F$ uniformément sur $\mathbb{R}$, presque sûrement.

Complément : la fonction de répartition empirique

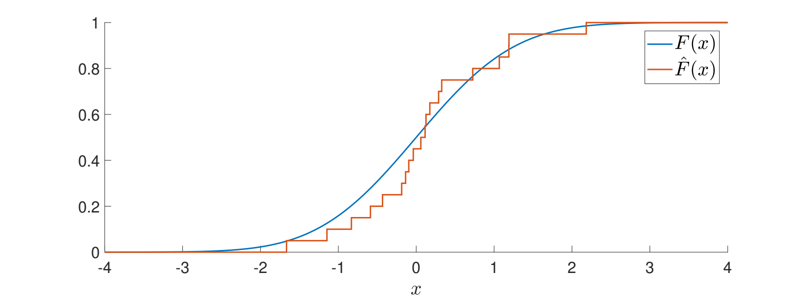


Figure – FRE pour $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ et $n = 20$.

retour au slide 27