Problématique de l'estimation paramétrique

- **Famille de distributions :**

On dispose d'une famille de lois de variables aléatoires, à un ou deux **paramètres** continus. *(On note X une « variable modèle » de cette loi.)* Par exemple, les familles de lois usuelles :

Lois discrètes à valeurs entières			**Lois continues à densité**		
Loi de probabilités		**Paramètre**	**Loi de probabilités**		**Paramètres**
de Bernoulli	$\mathcal{B}(p)$	$p \in]0\,;1[$	uniforme	$\mathcal{U}\,[a;b]$	$a, b \in \mathbb{R}, \quad a < b$
de Poisson	$\mathcal{P}(\lambda)$	$\lambda > 0$	normale	$\mathcal{N}(\mu, \sigma^2)$	$\mu \in \mathbb{R}, \quad \sigma^2 > 0$
géométrique	$\mathcal{G}(p)$	$p \in]0\,;1[$	exponentielle	$\mathcal{E}(\lambda)$	$\lambda > 0$

On suppose **inconnue à estimer** la vraie valeur *(**valeur théorique**)* du paramètre de la loi.

- **Notion d'échantillon de la loi étudiée**

Un échantillon $X_1, \dots, X_n$ de la loi étudiée, est une famille de variables aléatoires qui sont :

- mutuellement indépendantes,
- de même loi que la « variable modèle » X.

Les valeurs prises par l'échantillon $x_1, \dots, x_n$ s'appellent l'**observation**.

- **Estimateur**

Un **estimateur** est une **statistique** : une variable aléatoire $\varphi(X_1, \dots, X_n)$ définie en termes de l'échantillon, qui ne dépend **pas** du paramètre à estimer, susceptible de donner,

- avec une probabilité élevée
- une valeur proche de la vraie valeur du paramètre à estimer.

- **Biais, risque quadratique**

Notion	*(Interprétation)*	**Formule**
Biais	*(**erreur moyenne**)*	$b_a(A_n) = \mathbb{E}_a[A_n - a] = \mathbb{E}_a[A_n] - a$
Risque quadratique	*(erreur **quadratique** moyenne)*	$r_a(A_n) = \mathbb{E}_a\Big[(A_n - a)^2\Big]$
Décomposition biais-variance	*(méthode de calcul)*	$r_a(A_n) = \Big(b_a(A_n)\Big)^2 + \mathrm{Var}_a(A_n)$

- **Exemple central : l'estimateur de moyenne empirique**

C'est la moyenne arithmétique de l'échantillon, notée $\overline{X}_n = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} X_k$.

- C'est un estimateur **sans biais** de l'espérance de la loi : $\mathbb{E}\Big[\overline{X}_n\Big] = \mathbb{E}\Big[X\Big]$
- Sa **variance** *(et donc aussi son **risque quadratique**)* est : $\mathrm{Var}\Big(\overline{X}_n\Big) = \frac{1}{n}\,\mathrm{Var}\Big(X\Big)$

- **Autres**

On utilise aussi souvent $\min / \max(X_1, \dots, X_n)$: savoir en calculer la fonction de répartition. *(ou d'anti-répartition)*

Connaître les règles de calcul sur l'espérance, et la variance *(combinaison linéaire, indépendance)*