Estimation

Frédérique Leblanc

n observations d'un indicateur d'intérêt X (dosage d'un produit, indicateur physiologique, mesure d'un poids,...)
Données : $x_1, ..., x_n$
Modèle : les données sont un tirage d'un n-échantillon $X_1, ..., X_n$ de X de loi dépendant d'un ou deux paramètres.
dans ce cours : deux modèles seront utilisés

- Modèle Normal $\mathcal{N}(\mu, \sigma^2)$ avec μ et (ou) σ inconnus
- Modèle de Bernoulli $\mathcal{B}(p)$ avec p inconnu

Le problème Proposer un ou plusieurs estimateurs du ou des paramètres inconnus du modèle en établir les propriétés pour choisir le meilleur estimateur.

Estimateur : Un estimateur de θ est une fonction connue de $X_1, ..., X_n$ qu'on notera T_n.
Ex : $T_n = \bar{X}_n$; $T'_n = min(X_i)$, $T''_n = X_1$,... estimateurs de $\mu = E(X)$
Qualités attendues d'un estimateur :

- **Sans biais** : $E(T_n) = \theta$ (retrouve en moyenne θ)
- **De variabilité décroissante** : de plus en plus précis si n augmente
$$V(T_n) \to 0 \text{ si } n \to \infty$$
- **Convergent** : sans biais et de variance tendant vers zéro avec n
- **Vocabulaire** : T_n calculé avec les données $x_1, ..., x_2$ produit une estimation de θ et est noté $\hat{\theta}$

- **Moyenne** $\mu = E(X)$: $\bar{X}_n$ est un estimateur sans biais et convergent de μ. On notera l'estimation sans biais obtenue $\hat{\mu} = \bar{x}_n$.
- **Variance** $\sigma^2 = V(X)$:

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \text{et} \quad S_n'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

  sont deux estimateurs de variance en $O(1/n)$ tels que

$$E(S_n^2) = \frac{n-1}{n}\sigma^2 \ \text{ et } \ E(S_n'^2) = \sigma^2$$

  On notera $\hat{\sigma^2} = s'^2$ pour l'estimation sans biais de σ^2.
- **Proportion** p : $F_n = \bar{X}_n$ esb convergent pour p et l'estimation de p est $\hat{p} = \bar{x}_n = f_n$

Définition : On dira que $[T_1; T_2]$ est un intervalle de niveau de confiance $1-\alpha$ pour θ si T_1 et T_2 ne dépendent que des X_i et si

$$P(\theta \in I) = 1 - \alpha$$

Outils de proba : la loi de l'estimateur de θ
Loi des Estimateurs usuels : Pour X de loi $\mathcal{N}(\mu, \sigma^2)$

- $\bar{X}_n$ suit une loi normale $\mathcal{N}(\mu, \sigma^2/n)$
- $nS_n^2/\sigma^2 = (n-1)S_n'^2/\sigma^2$ suit une loi du Chi2 à $n-1$ degrés de liberté : $\mathcal{X}^2_{n-1}$
- Pour X de loi $\mathcal{B}(p)$; F_n suit approx. une $\mathcal{N}(p, p(1-p)/n)$

Modèle normal

- **Moyenne** μ :
  si σ^2 **connue :**

$$I(\mu, \alpha, \sigma^2) = \left[\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right]$$

  si σ^2 **inconnue :**

$$I(\mu, \alpha) = \left[\bar{X} - \frac{S'}{\sqrt{n}} t_{n-1,1-\alpha/2}, \bar{X} + \frac{S'}{\sqrt{n}} t_{n-1,1-\alpha/2}\right]$$

- **Variance** σ^2 **pour** μ **inconnu** :

$$J(\sigma^2, \alpha) = \left[\frac{nS^2}{z_{n-1,1-\alpha/2}}, \frac{nS^2}{z_{n-1,\alpha/2}}\right]$$

Modèle de Bernoulli pour p : intervalle de niveau approximatif $1-\alpha$

$$I(p,\alpha) = \left[F_n - \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} u_{1-\alpha/2}, F_n + \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} u_{1-\alpha/2} \right]$$

à condition que $np > 10$ et $n(1-p) > 10$.