Exercice 1.  

Mean vs Median

Let $X$ be a real random variable of mean $\mathbb{E}(X) = \mu$ and variance $\text{Var}(X) = \sigma^2$, both finite. Let $m \in \mathbb{R}$ be the median value of $X$, defined by $P(X < m) \leq 1/2$ and $P(X > m) \leq 1/2$.

Distance between the mean and the median

1. Given the box plot of $X$ drawn below, where can you position the mean?

![Box plot of X](image)

2. Prove the Cantelli inequality (a unilateral Bienaymé-Tchebychev bound): for all real $a > 0$,

$$P(X - \mu \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

3. Deduce the gap between the mean and the median is bounded by the standard deviation: $|m - \mu| \leq \sigma$.

Exercice 2.  

Yule-Simpson paradox

A selective school hires students through different exams: cooking, ping-pong, belote. Each candidate can take only one of those exams. Over the last session, there were as many male candidates as female candidates. For each exam, the success rate of boys is better than the success rate of the girls. Can you infer that the success rate of boys over all the three exams is better than the one of girls? If YES, prove it. If NO, provide a counterexample.

Exercice 3.  

Ponctual estimation

Let $(x_1, \ldots, x_n)$ be a sample of values in $E$. We wish to model this sample as a realization of an iid sequence of random variables following an unknown law which is characterized by some parameters. An estimator $\theta_n$ of a parameter $\theta \in \mathbb{R}$ of the law is a function from $E^n$ into $\mathbb{R}$, to work with samples of any size. By extension, we will denote a bit differently $\hat{\theta}_n$ the random variable $\theta_n(x_1, \ldots, x_n)$ where $X_1, \ldots, X_n$ is an iid sequence of random variables following the searched law.

For this iid model of sample with a law of density/mass $f_\theta(x)$ where $\theta$ is a parameter, the likelihood function gives the weight associated with the occurrence of $(x_1, \ldots, x_n)$, that is $L_\theta(x_1, \ldots, x_n) \defeq f_\theta(x_1) \cdots f_\theta(x_n)$ ($L$ like Likelihood). An maximum likelihood estimator is an estimator $\theta_n$ such that for all $(x_1, \ldots, x_n)$, $\theta = \theta_n(x_1, \ldots, x_n)$ maximizes $L_\theta(x_1, \ldots, x_n)$.

A family of estimators $(\theta_n)_{n \in \mathbb{N}^*}$ of $\theta$ is said unbiased if $\forall n \in \mathbb{N}^*$, $\mathbb{E}(\hat{\theta}_n) = \theta$. It is said asymptotically unbiased if $\lim_{n \to +\infty} \mathbb{E}(\hat{\theta}_n) = \theta$. It is said weakly converging if the random variables $\hat{\theta}_n$ tend to $\theta$ in probabilities when $n \to +\infty$.

1. Suggest an unbiased estimator of the mean viewed as a parameter. Under which condition(s) about the underlying law, your estimator is converging?

2. Suggest an unbiased estimator of the variance viewed as a parameter. Under which condition(s) about the underlying law, your estimator is converging?

3. Same question if we assume that the mean $\mu$ of the underlying law is known.

4. Find a maximum likelihood estimator $\lambda_n$ of the parameter $\lambda$ of a Poisson law $\mathcal{P}(\lambda)$.  

1
5. Let $E = \mathbb{R}^+$ be the space of values for a sample $(x_1, \ldots, x_n)$. We are looking for an underlying law uniform over $[0, \theta]$ which correspond to this sample. For the next estimators, argue whether they are unbiased, asymptotically unbiased, converging:

- $\hat{\theta}_1$ arithmetical mean of $x_i$, multiplied by 2;
- $\hat{\theta}_2$ maximum of $x_i$;

**Exercice 4.**

"L’important c’est de participer" (from J.-F. Delmas)

At the London 2012 Olympic Games, France got 34 medals including 11 gold medals, over a total number of 962 medals including 302 gold medals. The earth population is roughly $6.10^9$ people with 60 millions in France. Can we say the elite sportsmen are uniformly distributed inside the earth population?