d) Types d'histogramme.

→ **Histogramme à classe de m̂ largeur :**

- On a $\forall j \in [\![1, m]\!]$ $h_j = cte$ ⇒ $h_j = \frac{a_K}{K}$ (K : nombre des class)

⟹ On trace les ≠ts classes

classes $[a_{j-1}, a_j[$	...	.
effectif n_j		
freq : n_j/n		
hauteur : $\frac{n_j}{n\, h_j}$		

- À partir de ce tableau on peut dessiner facilmt l'histogramme.

→ **Histogramme à classes de même effectif :**

- Histogramme tq : $\forall j \in [\![1, n]\!]$ $n_j = cte$ ⟹ $n_j = n/K$ (K : nombre de classes)
- les classes de h_j ≠ts

classe $]a_{j-1}, a_j]$			
effectif n_j	n/K	...	...
freq : n_j/n			
hauteur : $\frac{n_j}{n\, h_j}$			

- À partir de ce tableau on peut dessiner facilmt l'histogramme

Rq : à partir des listes $(x_i)_i$ des observations on peut trouver les (h_j)

Rq

* Les histogrammes à classe de même effectif décrivent plus finement la distribution que les histogrammes à classe de m̂ larger.

5] Fonction de répartition empirique :

a) Déf.

* La fonction de répartition empirique (FdRE) F_n associée à un échantillon $x_1, \longrightarrow, x_n$ est la fonction définie par :

- $\forall x \in \mathbb{R}$, $F_n(x)$ = pourcentage d'observations inférieures à x

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{x_i \leq x\}} = \begin{cases} 0 & \text{si } x < x_1^* \\ \frac{i}{n} & \text{si } x_i^* \leq x \leq x_{i+1}^* \\ 1 & \text{si } x \geq x_n^* \end{cases}$$

Re

- $F_n(x)$ est une estimation de $F(x)$
- On peut montrer que cette estimation est d'excellente qualité
- déterminer le modèle probabiliste à partir de la FdRE n'est pas tout à fait correcte mais on utilise plutôt graphe de probabilités

b] Graphe de probabilités :

* Déf : est un nuage de pts tracé à partir de la fonction de répartition empirique tq : les pts doivent être approximativement alignés si les observations proviennent d'une loi de probabilité bien précise

* Méthode :

* Obj : On souhaite savoir si les observations sont issues de loi probabilité dépendant d'un paramètre θ inconnu, dont la fonction de répartition est F.

* Principe : chercher une relation affine du type $h[F(x)] = \alpha(\theta)\, g(x) + \beta(\theta)$

* Résultat : si la vraie fonction de répartition est F, les pts :
$\left(g(x_i^*),\ h(F_n(x_i)) = h\left(\frac{i}{n}\right)\right)$ seront approximativement alignés

$\Longrightarrow$ la pente et l'ordonnée à l'origine fournit une approximation de $\underline{\underline{\theta}}$

Re. * Le nuage $\left(g(x_i^*),\ h\left(\frac{i}{n}\right)\right)$, $i \in \{1, \ldots, n\}$ est appelé graphe de probabilités pour la loi de fonction de R F

Exemple 1 : graphe de probabilité d'un $X \sim \mathcal{E}(\lambda)$:

$\forall x > 0, \quad F(x) = \mathbb{P}(X \leq x) = 1 - e^{-\lambda x}$

$\forall x > 0 \quad \ln(1 - F(x)) = -\lambda x$

Ainsi : le graphe de probabilité d'une loi $\mathcal{E}(\lambda)$ est le nuage des pts :

$$\left(x_i^*, \ln\left(1 - \frac{i}{n}\right)\right) \quad i \in \{1, \dots, n-1\}$$

Rq

✓ Estimation de λ

Exemple 2 : graphe de probabilité de $X \sim \mathcal{N}(m, \sigma)$:

- $X \sim \mathcal{N}(m, \sigma)$ Alors $U = \frac{X-m}{\sigma} \sim \mathcal{N}(0,1)$

$\forall x \in \mathbb{R} \quad F(x) = \mathbb{P}(X \leq x) = \phi\left(\frac{x-m}{\sigma}\right)$ ou $\phi = F_U$

Ainsi $\quad \phi^{-1}(F(x)) = \frac{1}{\sigma} x - \frac{m}{\sigma}$

* **Ainsi** : le graphe de probabilité d'une $\mathcal{N}(m, \sigma)$ est le nuage de pts :

$$\left(x_i^*, \phi^{-1}\left(\frac{i}{n}\right)\right) \quad i \in \{1, \dots, n-1\}.$$

6] Indicateurs statistiques :

a) Def

- Indicateurs statistiques : indicateurs numériques permettant de caractériser au mieux des données quantitatives
- Deux familles d'indicateurs :
  * Indicateurs de localisation
  * Indicateurs de dispersion

b) Indicateur de localisation :

* **obj** : donner un ordre de grandeur général des observations ; un nombre unique qui résume au mieux les données.
* **La moyenne empirique :** $\quad \overline{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$

Rq D'ap LFGN $\overline{X}_n \longrightarrow E(X)$

* les valeurs extrêmes :

- $x_1^* = \min_i (x_i)$ et $x_n^* = \max (x_i)$

Alors $(x_1^* + x_n^*) / 2$ est un indicateur de localisation

Rq :
Inconvénient de la moyenne et des valeurs extrêmes : Sensibilité aux valeurs aberrantes.

* La médiane empirique : $\tilde{x}_n$ ou $\tilde{q}_{n, 1/2}$

- Un réel qui partage l'echantillon ordonné en deux parties de même effectif

$\Longrightarrow$ la moitié des observations sont inférieures à $\tilde{x}_n$ et l'autre moitié lui sont supérieurs

- $$\tilde{x}_n = \tilde{q}_{n, 1/2} = \begin{cases} \frac{1}{2}\left(x^*_{\frac{n}{2}} + x^*_{\frac{n}{2}+1} \right) & \text{si } n \text{ est paire} \\ x^*_{\frac{n}{2}+1} & \text{si } n \text{ im} \end{cases}$$

Rq
La médiane empirique est insensible aux valeurs aberrantes

c) Indicateurs de dispersion :

*) obj

* But : mesurer la variabilité des observations

*) Variance Empirique :

- $$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}_n^2$$

- S_n^2 : mesure l'ecart quadratique moyenne de l'echantillon à sa moyenne

Rq
En R : var(x) donne $S_n'^2 = \frac{n}{n-1} S_n^2$ au lieu de S_n^2

*) Ecart-type empirique : $S_n = \sqrt{S_n^2}$, Interêt : il s'exprime dans le m̂ unité que les données

*) Coefficient de variation empirique : $\boxed{CV_n = \frac{S_n}{\bar{x}_n}}$

Rq - si $CV_n > 0{,}15 \Longrightarrow$ l'echantillon possède une variabilité significative

- si $CV_n < 0{,}15 \Rightarrow$ la moyenne empirique seule un bon résumé de tout l'echantillon

d) quantiles empiriques :

* Les quantiles empiriques sont des valeurs qui partagent l'échantillon ordonné en un certain nombre de parties de même effectif.

Ex 2 parties ⟹ médiane $\tilde{x}_n$

4 parties ⟹ quartiles : $\tilde{q}_{n,1/4}$, $\tilde{q}_{n,2/4}$, $\tilde{q}_{n,\frac{3}{4}}$

10 parties ⟹ déciles : $\tilde{q}_{n,1/10}$, ——— , $\tilde{q}_{n,9/10}$

* quantiles empiriques de l'échantillon $x_1, \ldots, x_n$:

$$\forall p \in]0,1[\quad \tilde{q}_{n,p} = \begin{cases} \frac{1}{2}\left(x^*_{np} + x^*_{np+1}\right) & \text{si } np \text{ est entier} \\ x^*_{\lfloor np \rfloor + 1} & \text{sinon} \end{cases}$$

Rq : toutes les quantités définies dans ce chapitre sont ~~...~~ -ĉ aussi définies par des VA.

i) FdRE : $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$

ii) $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

iii) $S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$

iv) $$\hat{Q}_{n,p} = \begin{cases} \frac{1}{2}\left(X^*_{np} + X^*_{np+1}\right) & \text{si } np \text{ est entier} \\ X^*_{\lfloor np \rfloor + 1} & \text{si non} \end{cases}$$