IV. Etude statistique

IV.1 Ecart Type , Variance et Inégalité de Tchébychev

Nous avons calculé une longueur moyenne de file sur un temps d'étude que nous avons pris assez grand (50000) mais néanmoins limité . Il serait bon d'avoir à notre disposition un grand nombre de valeurs , même si on doit alors diminuer le temps d'étude afin d'avoir un temps de calcul raisonnable, afin d'en faire la moyenne . La variable statistique sera alors la longueur moyenne de file , obtenue pour chaque exécution . En faisant un nombre d'exécutions suffisamment elevés , nous pourrons alors avoir une bonne estimation de la longueur moyenne de file réelle.

$$L_o = \frac{1}{N} \sum_{i=0}^{N} L_i \sim E(L) \quad \textbf{si N assez grand}$$

Où L_o est la moyenne arithmétique des longueurs moyennes L_i obtenuent pour une exécution . **N** est le nombre d'exécutions.

Le calcul de la variance et de l'écart type de la variable statistique nous permettra alors à partir d'une mesure isolée de savoir si elle s'écarte beaucoup de la moyenne .

$$V_o = \frac{1}{N} \sum_{i=0}^{N} (Li-L_o)^2 \sim V(L) \quad \textbf{si N assez grand}$$

$$\sigma = \sqrt{V_o} \quad \textbf{, l'écart type.}$$

L'inégalité de Tchébychev s'écrit :

$$P[\ |X-E(X)| > \alpha] < \beta$$

ce qui nous donne une majoration **ß** de la probabilité que l'écart entre une valeur isolée **X** et la moyenne réelle **E(X)** soit supérieur à α .

/////////////
.... $E(X) + \alpha$

.... **E(X)**

.... $E(X) - \alpha$
/////////////

Plus la variance est grande et plus la probabilité que la variable aléatoire sorte de l'intervalle est grande .

De façon plus précise , on a :

$$(1) \qquad P[\ |X-E(X)| > k\sigma] < \frac{1}{k^2}$$

On utilise alors l'inégalité comme suit :

- Si on prend ($1/k^2$) = 0.25 , donc k = 2 , on peut alors dire que la probabilité pour que la valeur isolée observée s'éloigne de k fois l'écart type soit 0.25 .

On a de plus:

$$\alpha = k\sigma \;\; \text{-->} \;\; k = \frac{\alpha}{\sigma} \;\; \text{-->} \;\; \frac{1}{k^2} = \frac{\sigma^2}{\alpha^2} = \frac{V}{\alpha^2}$$

On appelle intervalle de confiance la quantité 2α et la quantité σ^2/α^2 est la confiance le plus souvent exprimé en pourcentage .

On s'intéresse le plus souvent à :

$$(2) \qquad P[\ |X-E(X)| < k\sigma] > 1 - \frac{1}{k^2}$$

Où $1 - \frac{1}{k^2}$ est le degré de confiance .

IV.2 Résultats

Les résultats suivants ont été obtenus pour un temps d'étude de 3000 unités . Seule la longueur moyenne de file est reportée. Le nombre d'exécutions est de 100.

L_MOYEN = 3.8806	**L_MOYEN = 4.1780**	**L_MOYEN = 4.2580**
L_MOYEN = 3.5010	L_MOYEN = 5.1572	**L_MOYEN = 3.6854**
L_MOYEN = 4.1164	**L_MOYEN = 3.5293**	L_MOYEN = 3.1350
L_MOYEN = 3.8652	**L_MOYEN = 4.1987**	L_MOYEN = 6.1263
L_MOYEN = 4.3853	**L_MOYEN = 3.7328**	L_MOYEN = 5.1021
L_MOYEN = 2.6432	**L_MOYEN = 4.6243**	**L_MOYEN = 3.5088**
L_MOYEN = 3.2327	**L_MOYEN = 3.2598**	**L_MOYEN = 3.2347**
L_MOYEN = 4.1007	**L_MOYEN = 3.7777**	**L_MOYEN = 3.3147**
L_MOYEN = 3.4463	**L_MOYEN = 3.7447**	**L_MOYEN = 3.5958**
L_MOYEN = 4.0328	**L_MOYEN = 3.9285**	**L_MOYEN = 3.7179**
L_MOYEN = 4.9354	**L_MOYEN = 3.7252**	L_MOYEN = 3.0862
L_MOYEN = 4.2803	**L_MOYEN = 4.2653**	**L_MOYEN = 4.5994**
L_MOYEN = 3.4135	**L_MOYEN = 3.6450**	L_MOYEN = 4.8910
L_MOYEN = 3.2415	**L_MOYEN = 3.2409**	**L_MOYEN = 4.2954**
L_MOYEN = 4.2662	**L_MOYEN = 3.5899**	**L_MOYEN = 3.4592**
L_MOYEN = 3.4488	**L_MOYEN = 3.9402**	**L_MOYEN = 3.2448**
L_MOYEN = 3.4642	**L_MOYEN = 4.1283**	L_MOYEN = 5.5712
L_MOYEN = 4.0865	**L_MOYEN = 3.5491**	L_MOYEN = 4.9631
L_MOYEN = 3.8471	**L_MOYEN = 3.7788**	**L_MOYEN = 3.9733**
L_MOYEN = 4.2420	**L_MOYEN = 3.8167**	**L_MOYEN = 3.6533**
L_MOYEN = 4.4444	**L_MOYEN = 3.4928**	L_MOYEN = 5.2900
L_MOYEN = 6.0149	**L_MOYEN = 4.0955**	L_MOYEN = 4.6686
L_MOYEN = 5.1891	**L_MOYEN = 3.6611**	**L_MOYEN = 3.5949**
L_MOYEN = 4.1340	**L_MOYEN = 3.7561**	**L_MOYEN = 4.1490**
L_MOYEN = 3.8061	**L_MOYEN = 4.2215**	L_MOYEN = 2.7603
L_MOYEN = 3.5678	**L_MOYEN = 4.4370**	**L_MOYEN = 3.4026**
L_MOYEN = 5.3534	**L_MOYEN = 3.9491**	**L_MOYEN = 3.6836**
L_MOYEN = 3.5809	**L_MOYEN = 3.8339**	**L_MOYEN = 3.7797**
L_MOYEN = 4.0801	**L_MOYEN = 3.9568**	**L_MOYEN = 3.4260**
L_MOYEN = 4.0020	**L_MOYEN = 3.5623**	**L_MOYEN = 3.3371**
L_MOYEN = 3.3598	**L_MOYEN = 4.3387**	**L_MOYEN = 3.9088**
L_MOYEN = 4.1635	**L_MOYEN = 3.9108**	**L_MOYEN = 3.6769**
L_MOYEN = 3.3024	L_MOYEN = 2.9404	**L_MOYEN = 4.6396**
L_MOYEN = 3.8099		

$L_O = 3.9494 \quad V_O = 0.4095 \quad \text{-->} \; \sigma = 0.6399$

IV.3 Utilisation

Utilisons l'inégalité (2) et regardons pour $\mathbf{1 - \frac{1}{k^2} = 0.2}$

On en tire alors $\mathbf{1 / k^2 = 0.8}$ d'où $\mathbf{k = 1.1180}$ **.** On en déduit alors la valeur de $\alpha = \mathbf{k\sigma = 0.7154}$ **.** Donc les bornes pour notre étude sont :

$$\mathbf{Lo-k\sigma < Li < Lo+k\sigma} \qquad \mathbf{(E)}$$

On va regarder combien de valeur Li sont dans cette intervalle , en sachant que l'inégalité doit nous donner proportion supérieur à 20 % .

Les **Li** qui sont dans l'intervalle ont été marqués en gras , on constate alors que leur nombre dépasse largement les 20.

Si on utilise l'inégalité (1) , sous les mêmes conditions , c'est à dire $\mathbf{k = 1.1180}$ et $\mathbf{k\sigma = 0.7154}$, on doit trouver que le nombre des **Li** ne vérifiant par (E) soit inférieur à $\mathbf{1 / k^2}$, soit 80 % .

On voit aisément que ces **Li** , en caractère normal , sont bien inférieur à 80 .

L'intervalle de confiance , $\mathbf{2\alpha}$, est alors de **1.4308** .