

# Cours

## **Maîtrise statistique des procédés**

**Walter TINSSON**

Université de Pau

Année 2022 / 2023



## Introduction générale

Ce cours a pour objectif de présenter une introduction aux méthodes relevant de la MSP (pour Maîtrise Statistique des Procédés) ou, en langue anglaise, SPC (pour Statistical Process Control). Une des définitions possibles de la MSP est donnée ci-dessous :

« la maîtrise statistique des procédés est l'ensemble des méthodes et des actions permettant d'évaluer dans le temps de façon statistique les performances d'un processus de production et de décider de le régler, si nécessaire, pour maintenir les caractéristiques des produits stables et conformes aux spécifications retenues ».

*Joe ≈ demande général de la qualité des produits*

En pratique une des préoccupations principales du service qualité est que les procédés industriels utilisés restent **sous contrôle** (c'est-à-dire ne se dérèglent pas au fil du temps). Concrètement, sur une chaîne de production, le statisticien doit être capable de proposer des techniques permettant de savoir de manière fiable (à partir d'échantillons prélevés à intervalles réguliers) si :

- la production semble être sous contrôle, elle peut **continuer**, *ajuster*
- la production semble être hors-contrôle, il faut la **stopper** et effectuer des réglages. *desco oforé*

L'une ou l'autre de ces décisions va être associée à un **risque d'erreur** que l'on souhaite, bien entendu, le plus faible possible. Nous traduirons souvent l'hypothèse "le procédé est sous contrôle" par le fait que les observations suivent une loi de probabilité connue, par exemple de type **normale** (avec des paramètres connus ou à estimer). Si l'on ne peut pas raisonnablement supposer cette hypothèse vraie le procédé sera alors jugé hors contrôle.

Ce cours va aborder principalement les trois grands thèmes suivants :

- 1) outils mathématiques pour la MSP,
- 2) aptitude d'un processus,
- 3) cartes de contrôle.

*Par exemple, dans l'industrie automobile, régir vite  
est-ce qu'il faut s'inquiéter? => corriger la production*



# Chapitre 1

## Outils mathématiques pour la MSP

### 1.1 Outils probabilistes

Nous allons toujours dans la suite observer un phénomène aléatoire à partir d'un échantillon  $x_1, \dots, x_n$  de  $n$  valeurs mesurées le plus souvent sur une chaîne de production (masse d'un sac de ciment, nombre de défauts d'un produit fini, etc...). De manière classique on va supposer que chacun des  $x_i$  est une réalisation de la variable aléatoire réelle  $X_i$  telle que (hypothèses iid) :

$$\begin{cases} \text{les } X_i \text{ sont toutes de même loi,} \\ \text{les } X_i \text{ sont indépendantes.} \end{cases}$$

Chacune des variables aléatoires  $X_i$  suit donc la même loi qu'une variable aléatoire de référence  $X$  qui modélise le phénomène sous contrôle. La plupart des lois de référence utilisées en MSP sont des lois classiques telles que :

- la loi normale (mesure d'une caractéristique numérique),
- la loi binomiale (nombre de produits non-conformes),
- la loi de Poisson (nombre de défauts d'une grande production).

Rappelons le résultat primordial suivant, vrai pour toute loi normale (règle des  $3\sigma$ ) :

#### Proposition 01

Si  $X \sim N(\mu, \sigma)$  alors :

$$P[\mu - 3\sigma < X < \mu + 3\sigma] = 0.9973$$

Pour chaque loi de probabilité utilisée il sera fréquemment nécessaire de considérer :

- 1) l'espérance de  $X$ , notée  $E(X)$ , mesurant la « valeur théorique moyenne » de  $X$ .
- 2) l'écart-type de  $X$ , noté  $\sigma_X = \sqrt{\text{Var}(X)}$ , mesurant la « dispersion théorique » de  $X$  autour de  $E(X)$ . *desviación estándar*

Remarquons enfin que les observations  $x_i$  doivent parfois être ordonnées (ainsi que, en parallèle, les variables aléatoires réelles  $X_i$  appelées alors statistiques d'ordre). On notera :

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)} \text{ et } X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

La valeur  $r = x_{(n)} - x_{(1)}$  est appelée étendue de l'échantillon.

*longueur*

## 1.2 Outils statistiques

### 1.2.1 Histogramme

Considérons un phénomène aléatoire observé via un échantillon de valeurs  $x_1, \dots, x_n$ . Un **histogramme** est alors une représentation graphique ayant pour but d'approcher la densité de probabilité à l'aide d'une fonction **constante par morceaux** (sur chacune des classes considérées). Les questions suivantes se posent alors légitimement lors de la construction de tout histogramme :

Comment déterminer le nombre de classes ?

Plusieurs choix classiques existent et sont motivés par la recherche d'un ajustement théorique le plus fidèle à une loi normale. Nous retiendrons ici celui qui consiste à prendre l'entier  $k$  le plus proche de  $\sqrt{n}$  (sans aller au delà de 20 classes).

Comment déterminer l'amplitude des classes ?

On considère classiquement des classes de même amplitude  $h = r/k$ .

Comment construire les classes ?

La première classe est construite, évidemment, à partir du minimum observé  $x_{(1)}$ . La convention la plus courante est que chacune des classes est un intervalle ouvert à gauche et fermé à droite (sauf pour la première classe fermée obligatoirement à gauche afin de pouvoir y inclure la valeur minimale observée).

Comment construire les barres ?

La convention mathématique est que la surface de chacune des barres est égale à fréquence observée (ainsi la somme des surfaces vaut un) par analogie avec la propriété de toute densité de probabilité). Plus généralement on utilise aussi en pratique des surfaces proportionnelles à la fréquence observée. Il en découle l'utilisation de hauteurs de barres proportionnelles à la fréquence observée (vu qu'elles ont toutes même largeur), un choix courant étant que ces hauteurs soient simplement égales aux effectifs observés.

### 1.2.2 Boîte de dispersion

La **boîte de dispersion** (boxplot en anglais) est un outil graphique permettant de représenter diverses caractéristiques de position de l'échantillon étudié. La méthode de construction le plus souvent utilisée par les statisticiens est présentée ci-dessous :

- 1) la **boîte** est un rectangle limité par les valeurs  $Q_1$  et  $Q_3$  (premier et troisième quartile de la série statistique utilisée),
- 2) la médiane de la série statistique est représentée par une barre à l'intérieur de la boîte (ainsi que, parfois, la moyenne par une croix),
- 3) les « **moustaches** » ont pour longueur  $1.5Eg = 1.5(Q_3 - Q_1)$  si cela est possible (on parle alors de moustache complète). Une moustache peut être moins longue si elle est limité par le maximum ou par le minimum de la série statistique étudiée (on parle alors de moustache tronquée).
- 4) toute valeur à l'extérieur des moustaches est représentée par un point, on dit qu'il s'agit d'une valeur **atypique**.

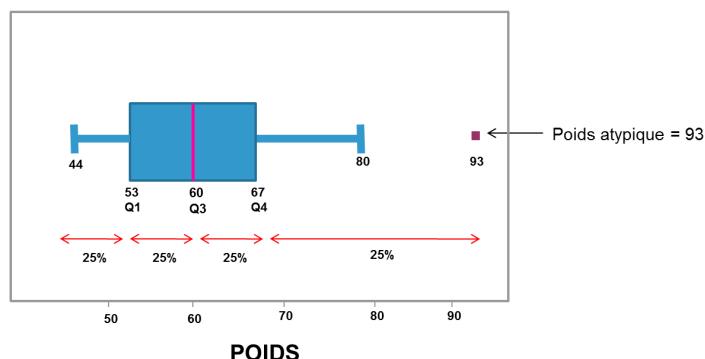


Figure 01. Boîte de dispersion.

### 1.2.3 Estimateurs classiques

Considérons une variable aléatoire  $X$  telle que  $E(X) = \mu$  et  $Var(X) = \sigma^2$ . Les estimateurs sans biais et convergents classiques pour ces deux caractéristiques sont rappelés ci-dessous.

1) pour l'espérance mathématique :

$$\bar{X} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

2) pour la variance :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \hat{\sigma}^2$$

Attention au fait que généralement  $S$  n'est pas un estimateur sans biais de  $\sigma$  !

Remarquons que dans le cas particulier primordial où les observations sont indépendantes et de même loi normale alors la loi de probabilité de  $S^2$  est donnée par :

#### Proposition 02

Si les variables aléatoires réelles  $X_i$  sont indépendantes avec  $\forall i = 1, \dots, n$ ,  $X_i \sim N(\mu, \sigma)$  alors :

$$\frac{(n-1)}{\sigma^2} S^2 \sim \chi_{n-1}^2$$

### 1.2.4 Ajustement d'une loi : test du khi-deux

$\chi^2$  chi-cuadra

Le test d'ajustement du khi-deux est souvent utilisé dans le domaine de la MSP. L'objectif de ce test est de déterminer si les observations semblent être ou non des réalisations d'une loi théorique postulée. L'idée générale de ce test consiste à évaluer une « distance » entre la distribution de probabilité observée et la distribution de probabilité théorique. On rejette alors l'hypothèse que ces deux distributions sont identiques lorsque cette distance est « grande ». Plus précisément la méthodologie est la suivante :

1) Répartissons tout d'abord les observations en un ensemble de  $k$  classes et désignons par  $n_i^{obs}$  l'effectif observé dans la classe  $i$  et par  $n_i^{the}$  l'effectif théorique prédit dans cette même classe à l'aide de la loi théorique postulée. La « distance » du khi-deux est alors donnée par la relation suivante :

$$\chi^2 = \sum_{i=1}^k \frac{(n_i^{obs} - n_i^{the})^2}{n_i^{the}}$$

2) On montre que lorsque  $n$  tend vers  $+\infty$  et que la loi des observations est bien la loi théorique postulée alors la statistique de test  $\chi^2$  est distribuée selon une loi du **khi-deux** à  $(k-s-1)$  ddl où  $s$  désigne le nombre de paramètres inconnus de la loi théorique postulée. On va donc rejeter l'hypothèse  $H_0$  : « la distribution observée est issue de la loi théorique postulée » lorsque la statistique de test va prendre de « grandes » valeurs donc :

on rejette  $H_0$  au niveau  $\alpha$  si  $\chi^2 > \chi_{1-\alpha, k-s-1}^2$

où  $\chi_{\alpha, n}^2$  désigne le fractile d'ordre  $\alpha$  de la loi du  $\chi^2$  à  $n$  degrés de liberté (c'est à dire toute valeur telle que  $P[X < \chi_{\alpha, n}^2] = \alpha$  avec  $X$  de loi du  $\chi^2$  à  $n$  degrés de liberté).

**Remarque.** Attention au fait que ces résultats théoriques sont utilisables en pratique lorsque l'échantillon est, évidemment, d'assez grande taille mais aussi lorsque les effectifs théoriques de chacune des classes ne sont pas trop faibles (on demande souvent qu'ils soient strictement supérieurs à 5). Si ce n'est pas le cas regrouper alors des classes contiguës pour atteindre cet objectif.

mailloune résultats et  
avec  $\chi^2$  1 petite échantillon

### 1.2.5 Ajustement d'une loi normale : test de Shapiro-Wilk

Ce test d'hypothèse, spécifique à la loi normale, est particulièrement puissant dans le cadre d'échantillons de petite taille ( $n \leq 50$ ). Ceci le rend très populaire en MSP où les échantillons utilisés (prélevés par le service qualité) sont couramment « petits ».

D'un point de vue théorique la statistique de ce test est construite par comparaison entre la série des quantiles générés à partir d'une loi normale et la série des quantiles empiriques obtenus à partir des données. Sa mise en oeuvre pratique est réalisée de la manière suivante :

1) ordonner les observations en  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  et déterminer la valeur moyenne observée  $\bar{x}$ ,

2) calculer la valeur :

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

3) déterminer toutes les **étendues partielles** de la forme :

$$d_1 = x_{(n)} - x_{(1)}, d_2 = x_{(n-1)} - x_{(2)}, \dots, d_{(i)} = x_{(n+1-i)} - x_{(i)} > 0$$

Il existe donc  $p$  étendues partielles non nulles avec  $n = 2p$  (parité) ou  $n = 2p + 1$  (imparité).

4) calculer la **statistique de test** :

$$w_{obs} = \frac{\left( \sum_{i=1}^p a_i d_i \right)^2}{z_n} = \frac{\left( \sum_{i=1}^p a_i d_i \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

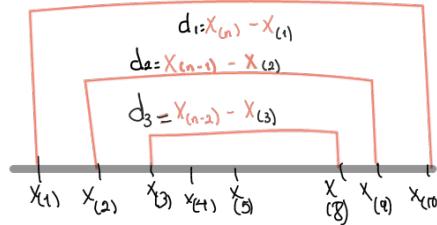
où les coefficients  $a_i$  sont tabulés.

5) plus la valeur calculée  $w_{obs}$  est grande plus la compatibilité avec une loi normale est crédible. Plus précisément, et avec un coefficient de sécurité  $\alpha$  fixé, la règle de décision est donnée par :

on rejette l'hypothèse de normalité lorsque  $w_{obs} < w_\alpha$

avec  $w_\alpha$  valeur critique tabulée.

**Remarque.** Ce test d'ajustement est souvent effectué en parallèle à une représentation graphique d'un diagramme quantile-quantile (QQ-plot en anglais) qui procède du même type de recherche (comparaison entre quantile théoriques postulés et quantiles réels observés).



**Maîtrise Statistique des Procédés**

Master MSiD 2ème année  
Travaux dirigés - Série 1

**Exercice 1** → *Quelques résultats classiques.*

On considère une variable aléatoire  $X$  de loi normale  $N(\mu, \sigma)$ .

- 1** Déterminer un intervalle  $I$  symétrique par rapport à  $\mu$  tel que  $P[X \in I] = 0.95$ .
- 2** Calculer les probabilités de la forme  $P[\mu - k\sigma < X < \mu + k\sigma]$  pour l'entier  $k$  prenant les valeurs 1, 2 et 3.
- 3** Que peut-on dire de la probabilité associée à  $k = 3$  lorsque la normalité n'est pas assurée ?

**Exercice 2** → *Loi de Poisson.*

On considère une variable aléatoire  $X$  de loi de Poisson  $\mathcal{P}(5)$ . Evaluer numériquement la probabilité  $P[\mu - 3\sigma < X < \mu + 3\sigma]$ .

**Exercice 3** → *Loi normale et qualité.*

Une étude a permis d'établir que l'épaisseur du papier produit par une entreprise suit une loi normale  $N(\mu, \sigma)$  de paramètres : es pesos

$$\mu = 0.0197 \text{ mm et } \sigma = 0.0014 \text{ mm.}$$

Le papier livré aux clients suit actuellement les spécifications suivantes concernant son épaisseur :  $0.02 \pm 0.005$  mm (le papier non-conforme est éliminé à l'aide de trieuses automatiques puis recyclé). Les clients se plaignent cependant de la mauvaise qualité du papier produit qui les oblige à régler souvent leurs machines. Ils souhaiteraient avoir du papier ayant comme limites de tolérance :

$$0.02 \pm 0.0015 \text{ mm.}$$

Les actions requises afin de respecter ces nouvelles tolérances nécessitent des aménagements et des investissements. Dans l'immédiat il est donc décidé de ne pas améliorer le processus industriel et de donner de nouvelles tolérances aux trieuses automatiques de feuilles. Les questions suivantes sont alors posées au service du contrôle de qualité.

- 1** Evaluer la proportion de rebus si les nouvelles spécifications sont utilisées.
- 2** Si cette proportion de rebus dépasse 2% l'entreprise va alors demander une dérogation à ses clients afin d'augmenter temporairement la qualité de la production tout en limitant la proportion de rebus à 2%. Proposer alors de nouvelles limites de tolérance  $T_i$  et  $T_s$  respectant ces contraintes (rechercher des limites symétriques).

**Exercice 4** → Boîtes de dispersion.

Une entreprise produit des batteries pour caméscope. Afin d'évaluer la longévité de celles-ci un échantillon de  $n = 30$  batteries est prélevé au sein de la production et leur durée de vie (exprimée en mn) est mesurée :

181	185	181	180	175	192	189	182	176	178
183	179	180	188	174	177	178	154	181	182
184	174	188	186	184	195	190	183	180	197

- 1** Représenter ces données sous forme de boîte de dispersion. Commenter.

On considère ici une variable aléatoire  $X$  de loi normale  $N(50, 2)$ .

- 2** On considère maintenant une variable aléatoire  $X$  de loi normale. Donner les caractéristiques de la boîte de dispersion théorique associée à cette loi. Quelle est la probabilité que  $X$  prenne des valeurs en dehors des moustaches ?

**Exercice 5** → Test d'ajustement du  $\chi^2$ . *Ainsi devrait*

Une industrie produit des sacs de ciment dont la masse moyenne devrait théoriquement être de 50kg. On souhaite alors connaître la distribution de la masse des différents sacs produits et notamment vérifier si elle est normale ou non afin de pouvoir par la suite effectuer divers traitements statistiques. Pour cela on relève un échantillon de 40 sacs d'une production journalière et les différentes masses sont données ci-dessous (exprimées en kg avec une précision de 10 grammes) :

49.73	49.07	50.54	49.01	49.58	49.81	48.94	50.06	49.52	50.05
49.23	49.92	50.08	50.01	50.41	51.46	49.36	51.37	50.28	50.53
49.56	50.77	51.11	50.43	49.40	49.63	51.80	49.29	49.69	50.07
50.51	50.99	50.70	51.03	48.37	50.19	50.75	49.65	50.23	50.62

- 1** Evaluer les caractéristiques de position et de dispersion de cet échantillon (moyenne et écart-type).

- 2** Représenter soigneusement cette distribution sous forme d'histogramme.

- 3** Effectuer un test du  $\chi^2$  pour juger si la distribution des masses suit une loi normale ou non.

- 4** Représenter la carte aux valeurs individuelles pour ces données. Quel est son intérêt ? Qu'apporte-t-elle de plus que l'histogramme ?

**Exercice 6** → Approximation de lois.

Dans son projet de reconquête des voyageurs la SNCF d'Ile de France affirme que 90% des trains en circulation sur l'ensemble de son réseau arrivent à l'heure.

- 1** Considérons un francilien prenant quotidiennement 3 de ces trains. Quelle est la probabilité qu'aucun des trains empruntés durant une journée n'arrive en retard ?

- 2** Toujours avec les mêmes hypothèses évaluer la probabilité qu'au moins 2 trains soient en retard.

Conciendo

3 Sachant que 5000 trains circulent chaque jour quelle est alors la probabilité qu'au moins 4550 trains arrivent à l'heure ?  $\rightarrow$  puisque  $n \rightarrow \infty$  on tourne la Binomial à la normal

4 Quelle est la probabilité qu'exactement 4550 trains arrivent à l'heure ? Puisque nous avons  $N \sim N(\mu, \sigma)$

Il est continue  $\Rightarrow P[Y = 4550] = 0$ , on va alors utiliser l'approximation :

### Exercice 7 → Nombre de défauts d'une production.

On s'intéresse ici au nombre de pièces défectueuses dans une production d'écrous. Les écrous fabriqués sont conditionnés par boîtes de 200 unités. Le tableau ci-dessous contient alors le nombre d'écrous défectueux observés pour 30 boîtes prélevées au hasard dans la production :

4	3	4	5	5	3	9	2	12	7
5	1	6	1	5	7	7	5	8	3
3	6	4	6	7	5	4	6	5	4

1 Quel est le nombre moyen d'écrous défectueux par boîte ? Quelle est la dispersion du nombre d'écrous défectueux par boîte ?

2 Justifier qu'il est possible en théorie d'approcher le nombre d'écrous défectueux par boîte à l'aide d'une loi de probabilité usuelle. Poisson

3 Vérifier alors ce qui a été énoncé à la question 2 à l'aide d'un raisonnement statistique portant sur l'échantillon observé.

4 Un nouveau client souhaite réaliser régulièrement des commandes par lot de 100 boîtes. Pour contrôler si le lot livré lui convient ou non la procédure de son service de la qualité est la suivante :

choisir 3 boîtes parmi les 100 livrées,  
déterminer le nombre d'écrous défectueux,  
refuser la commande si ce nombre dépasse 10.

Déterminer alors la probabilité qu'une commande ne soit pas refusée.

# Exercices

## Ex 1

1) On cherche  $t > 0$  telle que  $P[\mu - t < X < \mu + t] = 0.95$

On sait que  $X^* = \frac{X - \mu}{\sigma} \sim N(0, 1)$

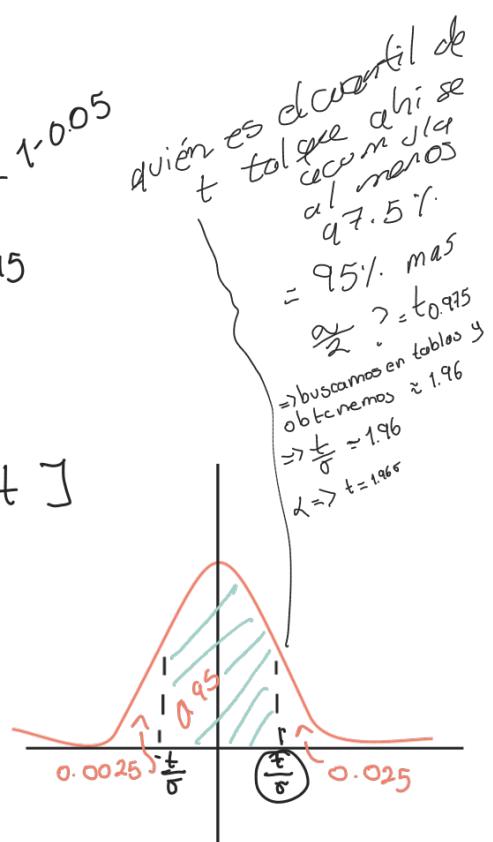
Donc  $P[\mu - t < X < \mu + t] = P[-t < X - \mu < t]$

$$= P\left[-\frac{t}{\sigma} < X^* < \frac{t}{\sigma}\right] = 0.95$$

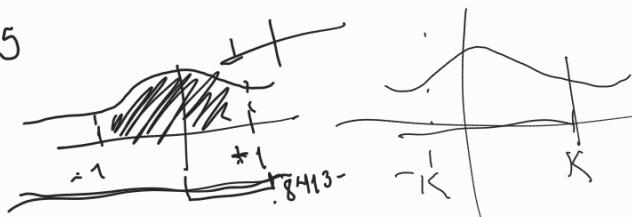
$$\frac{t}{\sigma} \approx 1.96$$

$$1 - \alpha + \frac{\alpha}{2}$$

$$\Rightarrow \text{il faut que } \frac{t}{\sigma} = 1.96$$



Donc  $P[\mu - 1.96\sigma < X < \mu + 1.96\sigma] = 0.95$



2) Calcul de  $P[\mu - k\sigma < x < \mu + k\sigma]$ ,  $k \in \mathbb{N}$

$$x^* = \frac{x-\mu}{\sigma} \sim N(0,1)$$

On a  $P_k = P[\mu - k\sigma < x < \mu + k\sigma] = P[-k\sigma < x - \mu < k\sigma] = P[-k < x^* < k]$

par symétrie de la loi:  $N(0,1)$  on a:

$$P[x^* < -k] = P[x^* > k] = 1 - P[x^* < k]$$

Pour les calculs numériques on a, alors:

$$P_k = 1 - 2P[x^* < k] = 2(0.8413) - 1 = 1.6826 \cdot 10^{-1}$$

$$K=1 \quad \bullet \quad P[\mu - \sigma < x < \mu + \sigma] \approx 0,6826 = 2P[x^* < 1] - 1 = 2(0.8413) - 1 = 0,6826$$

$$K=2 \quad \bullet \quad P[\mu - 2\sigma < x < \mu + 2\sigma] \approx 0,9546 = 2P[x^* < 2] - 1 = 2(0.9772) - 1 = 0,9546$$

$$K=3 \quad \bullet \quad P[\mu - 3\sigma < x < \mu + 3\sigma] \approx 0,9973 = 2P[x^* < 3] - 1 = 2(0.9987) - 1 = 0,9973 \quad \leftarrow \text{Règle des } 3\sigma$$

Démarche 65 sion cherche la littérature

Conclusion

$$p_k = 2P[x^* < k] - 1$$

3) D'après l'inégalité de Bienaymé-Tchebychev on sait que pour toute variable  $x$

on a:  $\forall \varepsilon > 0, P[|x - E(x)| > \varepsilon] \leq \frac{\text{Var}(x)}{\varepsilon^2}$ ,

On ajuste l'intervalle pour  $\varepsilon = 3\sigma$  on a donc:  $P[|x - \mu| > 3\sigma] \leq \frac{\sigma^2}{9\sigma^2} = \frac{1}{9} \approx 0,111$ .

En conclusion la probabilité de sortir de l'intervalle  $[\mu - 3\sigma, \mu + 3\sigma]$  ne peut pas dépasser 0,111

## Exercice 2

Réglé des 3σ avec  $X \sim P(5)$

On a donc ici  $\mu = 5$  et  $\sigma = \sqrt{5} \approx 2,236$

Calculons  $P[\mu - 3\sigma < x < \mu + 3\sigma] = P[5 - 3\sqrt{5} < x < 5 + 3\sqrt{5}] \approx 0.9971$

$= P[X \leq 11] \approx 0,9946$  Résultat très proche au cas de la loi normale

### Exercice 3)

1) Notons  $x$  l'épaisseur d'une feuille. La probabilité qu'elle soit non conforme vaut:

$$P = P[x < 0,0185] + P[x > 0,0215] \text{ avec } x \sim N(0,0197, 0,0014)$$

Donc

$$\begin{aligned} P &= P\left[\frac{x-\mu}{\sigma} < 0,0185\right] + P\left[\frac{x-\mu}{\sigma} < 0,0215\right] \text{ nous dénotons } x^* = \frac{x-\mu}{\sigma} \\ P &= P[x^* < -0,857] + P[x^* > 1,2857] \\ &= P[x^* > 0,857] + [1 - P[x^* < 1,2857]] \\ &= 1 - P[x^* < 0,857] + 1 - P[x^* < 1,2857] = 2 - P[x^* < 0,857] - P[x^* < 1,2857] \end{aligned}$$

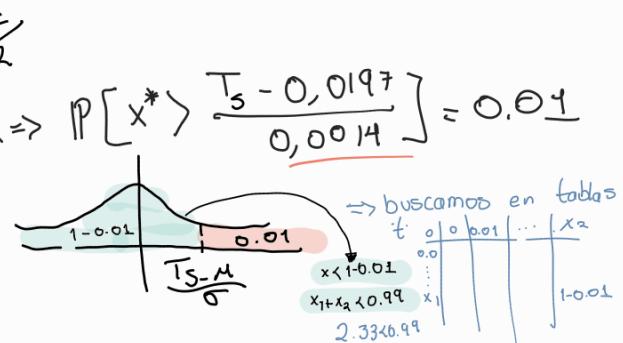
$$= 2 \cdot [0.8023 + 0.899] \approx 0.2934$$

Résultat mauvais parce que près qu'on a été dépeché

2) Déterminons  $T_5$  on a:  $P[x > T_5] = 0,01 \Leftrightarrow P[x^* > \frac{T_5 - 0,0197}{0,0014}] = 0,01$

mais on sait que  $P[x^* > 2,33] = 0,01$  i.e.

Conclusion:  $\frac{T_5 - 0,0197}{0,0014} = 2,33 \Leftrightarrow T_5 = 0,0230$



De même on trouve  $T_i = 0,0170$   $P[x^* < -\frac{x-\mu}{\sigma}] = 0,01 \Rightarrow \frac{T_i - 0,0197}{0,0014} = -2,33$

On peut proposer comme réglage  $0,02 \text{ mm} \pm 0,003 \text{ mm}$   
(amélioration par rapport au premier réglage)



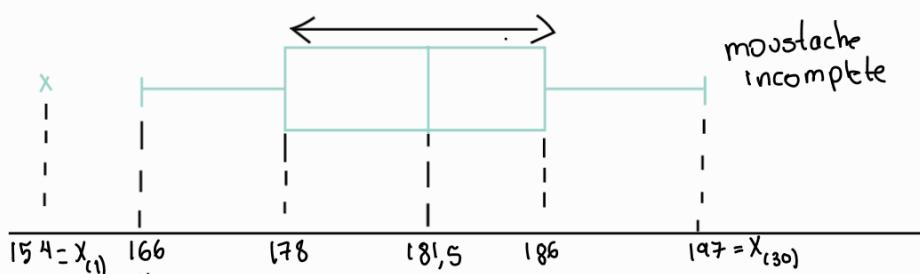
### Exercice 4)

1) Trouve la médiane,  $M = \begin{cases} X_{(\frac{n+1}{2})} & \text{si } n \text{ est impair} \\ \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2} & \text{si } n \text{ est pair} \end{cases}$

$$\Rightarrow M = \frac{X_{15} + X_{16}}{2} = \frac{181 + 182}{2} = 181,5$$

Pour  $Q_1$  il suffit de prendre la médiane des 15 premières valeurs, donc  $Q_1 = X_{(8)} = 178$

Pour  $Q_3$ , on prends la médiane des 15 plus grandes valeurs soit  $Q_3 = X_{(23)} = 186$   
 $Eg = Q_3 - Q_1 = 8 \rightarrow$  longueur d'une moustache complète  $12 = 1,5 Eg$



$$178 - 1,5 Eg$$

$$178 - 12$$

2) Soit  $X \sim N(\mu, \sigma)$ , boîte de dispersion théorique

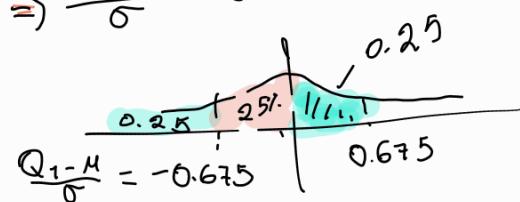
La médiane de  $X$  est égale à:  $M = \mu$  (par symétrie)

Détermination de  $Q_3$ :  $P[X < Q_3] = 0,75$

$$\text{On a: } P[X < Q_3] = P[X^* < \frac{Q_3 - \mu}{\sigma}] = 0,75 \quad Q_3$$

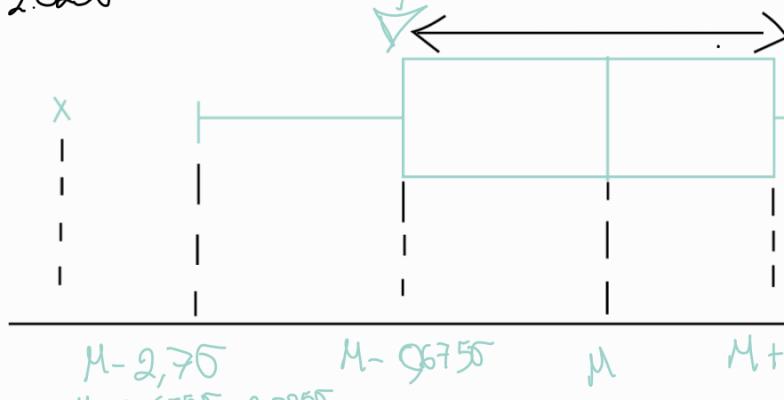
On a:  $X^* \sim N(0,1)$  on sait que  $P[X^* < 0,675] = 0,75 \Rightarrow \frac{Q_3 - \mu}{\sigma} = 0,675$

$$\rightarrow Q_3 = \mu + 0,675\sigma, \text{ par symétrie} \quad Q_1 = \mu - 0,675\sigma$$



$$Eg = 1.5(Q_3 - Q_1)$$

$$= 1.5(\mu + 0,675\sigma - \mu + 0,675\sigma) \\ = 1.5(1.35)\sigma = 2.025\sigma$$



$$\mu + 2.75 = \mu + 0.675 + 2.025$$

longueur des moustaches

$$1.5 eg = 2.025\sigma$$

Probabilité de sortir de moustaches

$$P = P[X < \mu - 2.75] + P[X > \mu + 2.75] \\ = 2 P[X > \mu + 2.75] \text{ par symétrie} \\ = 2 P[X^* > 2.7] = 2[1 - P[X^* \leq 2.7]] = 2[1 - 99.65] = 2[3.5 \times 10^{-3}] = 0.007$$

Probabilité très faible

Exercice 5) 1a. On obtient les estimateurs de  $\mu$  et  $\sigma^2$

$$S^2 = \frac{1}{39} \left( \sum_{i=1}^{40} (x_i - \bar{x})^2 \right) \approx 0,58 \Rightarrow S = 0,76 \quad \bar{x} = \frac{1}{40} \left( \sum_{i=1}^{40} x_i \right) = 50,10$$

Rappel

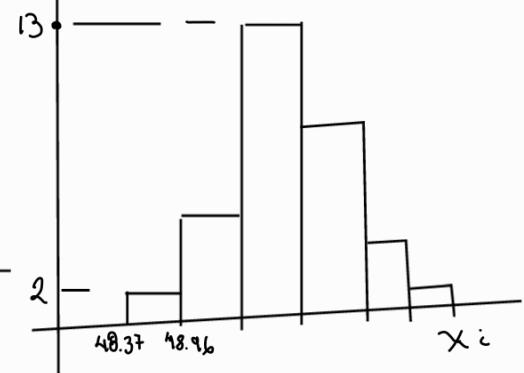
- $\bar{x}$  est un ESB de  $\mu = E(x)$
- $S^2$  est un ESB de  $\sigma^2 = \text{Var}(X)$

Le meilleur valeur pour la quantité de cas à faire dans un histogram c'est l'entier le plus proche de  $\sqrt{n}$

D'après le rappel on va utiliser  $k=6$  classes (entier le plus proche de  $\sqrt{n}, \sqrt{40}$ )  
 $\Rightarrow$  amplitude de chaque classe:  $h = \frac{x_{\max} - x_{\min}}{k} = \frac{51,89 - 48,37}{6} \approx 0,59$

Intervalle	Freq.
48,37 - 50,96	1
50,96 - 52,55	1
52,55 - 54,14	1
54,14 - 55,73	1
55,73 - 57,32	1
57,32 - 58,91	1

Classe	Effectif	Fréquence	Eff. théorique	Théorique
[48,37 ; 48,96]	2	0,005	2.222	0.0555
(48,96 ; 49,54]	7	0,18	6,51	0.1628
:	:	:		
[51,30 ; 51,89]	3	0,08	1,92	0.0480



Si  $X \sim N(50,1, 0.76)$  on calcule

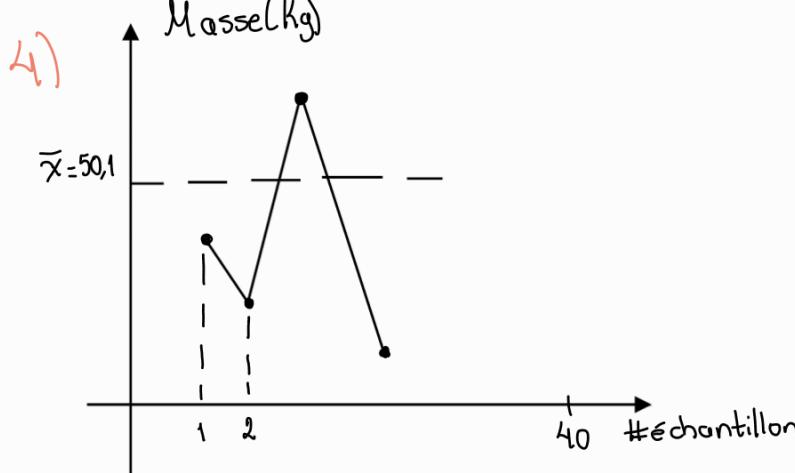
$$P[48.37, 48.96] \approx 0.0555$$

3. Statistique de test  $\chi^2 \approx 1,14$

On compare 1,14 au fractile  $\chi_{0,95,3}^2 \approx 7.81$

$\Rightarrow$  Il est impossible de rejeter l'hypothèse de normalité à  $\alpha = 5\%$ .

4) Massel(kg)



Il ajoute une perspective au cours de temps

6) 1. Désignons par  $x$  le nombre de trains à l'heure. On va donc supposer que  $X \sim B(3, 0.9)$  Bernoulli:  $X \sim B(n, p)$ ,  $P[x=k] = C_n^k p^k (1-p)^{n-k}$   $C_n^k = \frac{n!}{k!(n-k)!}$

$$\text{Ici: } P[x=3] = C_3^3 (0.9)^3 (0.1)^0 = (0.9)^3 \approx 0.729$$

$$2. \text{ Ici: } P[x \leq 1] = P[x=0] + P[x=1] = 0.28$$

3. D'un point de vue théorique le nombre de trains à l'heure est une réalisation de la VAR  $Y \sim B(5000, 0.9)$  et on cherche  $P[Y \geq 4550]$

La grande valeur de  $n$  peut poser des problèmes de calcul numérique

Rappel: d'après le théorème central-limite si  $(X_n)_{n \in \mathbb{N}}$  est une suite de V.AR. indépendantes et de même loi alors:

$$\sqrt{n} \left( \frac{X_n - \mu}{\sigma} \right) = \left( \frac{\sum_{i=1}^n X_i - n\mu}{\sigma \sqrt{n}} \right) \xrightarrow{D} N(0, 1)$$

$$\text{avec } \sum_{i=1}^n X_i = \sum_{i=1}^n x_i$$

variables aléatoires

On sait que si  $Y_n \sim B(n, p)$  alors  $Y_n = \sum_{i=1}^n X_i$  avec les  $X_i$  V.AR indépen-

antes de loi de Bernoulli de paramètre  $p$

Conclusion  $\left( \frac{Y_n - np}{\sqrt{np(1-p)}} \right) \xrightarrow{D} N(0, 1)$

$$\left( \frac{Y - 4500}{\sqrt{np(1-p)}} \right) \xrightarrow{\sim} N(0,1)$$

Puisque ici  $n$  est grand (5,000) on va donc supposer que  $Y^* = \left( \frac{Y - 4500}{\sqrt{450}} \right)$  suit une loi  $N(0,1)$

$$\text{Conclusion: } P[Y \geq 4550] = P[Y - 4500 \geq 50] = P[Y^* \geq 2,357] \approx 0,0091$$

A) Si on approche  $Y$  par une loi normale on a donc  $P[Y=4550]=0$ .  
On va alors utiliser l'approximation suivante pour la probabilité cherchée:

$$P[4549,5 \leq Y \leq 4550,5] = P[2,333 \leq Y^* \leq 2,384] \approx 0,0012$$

Par autre côté, si on le fait avec la Binomiale on obtient:  
 $P[Y=4550] = 0,0011237$

Ex 1)  $\bar{x} = \frac{1}{30} \left( \sum_{i=1}^{30} x_i \right) = 5,07 \quad s^2 = \frac{1}{29} \left( \sum_{i=1}^{29} (x_i - \bar{x})^2 \right) \approx 5,37$

2. Comme l'exercice en 6, on va utiliser la loi Binomial  
D'un point de vue théorique un écrou est défectueux avec une probabilité  $p$  ou pas  
En désignant par  $X$  le nombre d'écrous défectueux par boîte on a donc  
 $X \sim \mathcal{B}(200, p)$  on n'a pas la probabilité  $p$  alors il faut l'estimer\*

Rappel: On montre que si  $X_n \sim \mathcal{B}(n, p_n)$  avec  $(p_n)_{n \in \mathbb{N}}$  suite de réels tels que  $\lim_{n \rightarrow \infty} np_n = \lambda$   
alors  $X_n \xrightarrow{\text{poisson}} X \sim \mathcal{P}(\lambda)$

\* Ici on a bien une loi normale avec  $n$  "grand" ( $n=200$ ) et  $p$  "petit"

( $p \approx \frac{5}{200} \approx 0,025$ )  $\Rightarrow$  on peut approcher le nombre d'écrous défectueux par boîte  
par  $X \sim \mathcal{P}(5)$

3. Test du  $\chi^2$  pour  $H_0$ : "les données sont des réalisations de la loi  $\mathcal{P}(5)$ "

Valeur	Effet	Freq	Freq th	Eff. th	$P[X=0] = e^{-5}$	
0	0	0	0,00674	0,20		
1	2	0,07	0,0369	1,01		
:	:	:	:	:		
12	1	0,03	0,00343	0,1		

Regrouper les classes pour avoir des effectifs  $\geq 5$

$\chi^2 \approx 0,96$  et  $\chi^2_{0,95,3} \approx 7,81$  Impossible de rejeter  $H_0$

4. On sait que le nombre d'écrous défectueux par boîte suit une loi  $\mathcal{P}(5)$   
(approximation) En notant  $Y$  la VAR modélisant le nombre d'écrous défectueux  
sur 3 boîtes on a donc  $Y \sim \mathcal{P}(15)$  pour l'indépendance de loi Poisson

$$10 - 15 / \left( \sum_{k=0}^{10} 15^k \right) \approx 0,12$$

$$\text{Probabilité cherchée: } p = P[Y \leq 10] = \sum_{k=0}^{\infty} P[Y=k] = e^{-\lambda} \left( \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right)$$



# Chapitre 2

## Aptitude d'un processus

**Objectif Général :** rendre compte à l'aide de différents indices (réels) de la capacité ou non d'un processus de fabrication à respecter les tolérances liées aux objectifs de qualité fixés.

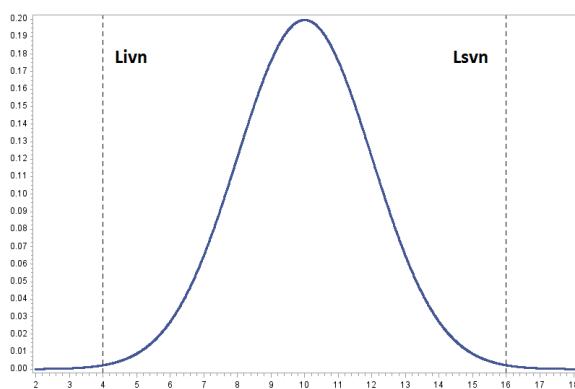
## 2.1 Indice $C_p$ d'aptitude d'un processus

### 2.1.1 Définitions

De manière classique on associe à tout processus de fabrication ses **limites inférieure et supérieure de variabilité naturelle** (notées *L<sub>vn</sub>* et *L<sub>sxn</sub>*) par :

⎩ 0.135% des observations sont supérieures à  $L_{svn}$ ,  
 ⎩ 0.135% des observations sont inférieures à  $L_{ivn}$ .

En d'autres termes 99.73% de la production se situe dans l'intervalle  $[Lvn, Lsvn]$  appelé aussi couramment **intervalle de dispersion** du processus. L'amplitude  $Lsvn - Lvn$  de cet intervalle est la **tolérance naturelle** du processus (encore appelée aptitude du processus). Nous avons que cette tolérance naturelle est égale à  $6\sigma$  dans le cas d'un processus de loi normale.



**Figure 02.** Limites de variabilité naturelles.

Désignons enfin par  $[T_i, T_s]$  la **zone de tolérance** que doit vérifier le processus pour être jugé « sous contrôle » (dans une production industrielle cette zone est fixée par le client, elle correspond aux spécifications souhaitées).

La construction de l'indice  $C_p$  (la lettre  $C$  provenant du terme anglais « capability ») est alors basée sur une simple comparaison de la tolérance naturelle du processus avec l'amplitude de la zone de tolérance imposée via la relation suivante :

### Définition 03

Etant donné un processus, de valeur moyenne  $\mu$  et d'écart-type  $\sigma$ , devant respecter la zone de tolérance  $[T_i, T_s]$  on définit son indice d'aptitude  $C_p$  par :

$$C_p = \frac{T_s - T_i}{6\sigma}$$

Remarque. L'indice  $C_p$  est aussi appelé indice d'aptitude potentielle ou indice de « capabilité » potentielle.

#### 2.1.2 Propriétés de remarques

Les propriétés suivantes sont immédiates :

- 1) On a toujours  $C_p > 0$  et plus la valeur de  $C_p$  est grande meilleur est le processus étudié.
- 2) La valeur théorique pivot est  $C_p = 1$  (le processus a alors une « largeur » - selon la tolérance naturelle - qui coïncide parfaitement avec l'amplitude de la zone de tolérance imposée). Lorsque  $C_p > 1$  alors le processus est qualifié de apte pour cet indice. Dans la pratique industrielle il est courant d'imposer des normes plus contraignantes et de déclarer un processus apte pour l'indice  $C_p$  si et seulement si :

$$C_p > 1.33 \text{ (encore écrit sous la forme } C_p > 133\%).$$

- 3) L'indice  $C_p$  est couramment évalué à l'aide de la formule de la définition 03 sans se préoccuper de la loi de probabilité suivie par le processus étudié. Prenons garde au fait que cette formule n'est (en théorie) licite que sous hypothèse de normalité. Si ce n'est pas le cas alors elle doit être remplacée par :

$$C_p = \frac{T_s - T_i}{L_{svn} - L_{vn}}$$

L'évaluation numérique de cette quantité est alors subordonnée à la détermination des fractiles d'ordre 0.99865 et 0.00135 de la loi réelle suivie par le processus (mais cette opération n'est pas forcément aisée en pratique...).

### Utilisation pratique

La grande qualité de l'indice  $C_p$  est sa simplicité et sa facilité de calcul. A contrario son principal défaut est de ne tenir aucun compte du centrage ou non du processus (c'est-à-dire de sa valeur moyenne). Il en découle (voir la figure 03) que des processus parfaitement centrés ou totalement décentrés peuvent être associés à une même valeur de  $C_p$ .

L'indice  $C_p$  est donc particulièrement utile pour détecter des processus non-aptes (c'est-à-dire pour des valeurs  $C_p < 1$ ).

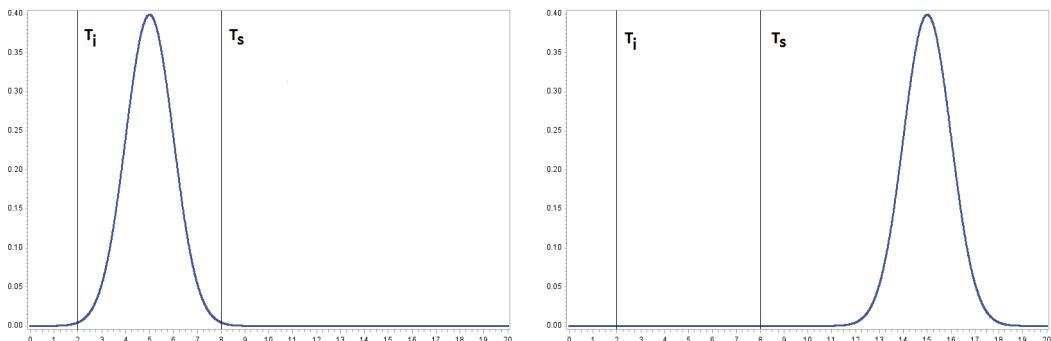


Figure 03. Exemple de deux processus à  $C_p > 1$  égaux.

### 2.1.3 Intervalle de confiance pour $C_p$

Le calcul de l'indice  $C_p$  est lié à l'écart-type  $\sigma$  du processus qui est généralement inconnu. Il est alors nécessaire d'utiliser l'estimateur classique  $S^2$  en remplacement de la variance  $\sigma^2$ . Sachant que  $(n - 1) S^2 / \sigma^2$  suit une loi du  $\chi^2$  (proposition 02) il vient donc :

$$P \left[ \chi_{\alpha/2, n-1}^2 \leq \frac{(n-1)}{\sigma^2} S^2 \leq \chi_{1-\alpha/2, n-1}^2 \right] = 1 - \alpha$$

$$\iff P \left[ \sqrt{\frac{\chi_{\alpha/2, n-1}^2}{n-1}} \leq \frac{S}{\sigma} \leq \sqrt{\frac{\chi_{1-\alpha/2, n-1}^2}{n-1}} \right] = 1 - \alpha$$

Sachant que :

$$C_p = \frac{T_s - T_i}{6\sigma} = \left( \frac{T_s - T_i}{6S} \right) \frac{S}{\sigma}$$

on en déduit donc que :

$$P \left[ \left( \frac{T_s - T_i}{6S} \right) \sqrt{\frac{\chi_{\alpha/2, n-1}^2}{n-1}} \leq C_p \leq \left( \frac{T_s - T_i}{6S} \right) \sqrt{\frac{\chi_{1-\alpha/2, n-1}^2}{n-1}} \right] = 1 - \alpha$$

D'où le résultat suivant :

#### Proposition 04

Sous hypothèse de normalité un **intervalle de confiance** pour l'indice  $C_p$  est donné, au coefficient de sécurité  $\alpha$ , par :

$$\left[ \hat{C}_p \sqrt{\frac{\chi_{\alpha/2, n-1}^2}{n-1}}, \hat{C}_p \sqrt{\frac{\chi_{1-\alpha/2, n-1}^2}{n-1}} \right]$$

avec  $\hat{C}_p = (T_s - T_i) / 6S$  estimateur de l'indice  $C_p$  et  $\chi_{\alpha, n}^2$  fractile d'ordre  $\alpha$  de la loi du  $\chi^2$  à  $n$  degrés de liberté.

## 2.2 Indice $C_{pk}$ d'aptitude d'un processus

### 2.2.1 Définitions

L'idée est d'enrichir l'indice  $C_p$  en tenant compte aussi de la valeur théorique moyenne  $\mu = E(X)$  du processus étudié. On définit alors au préalable les indices d'aptitude unilatéraux par :

$$C_{pI} = \frac{\mu - T_i}{3\sigma} \text{ et } C_{pS} = \frac{T_s - \mu}{3\sigma}.$$

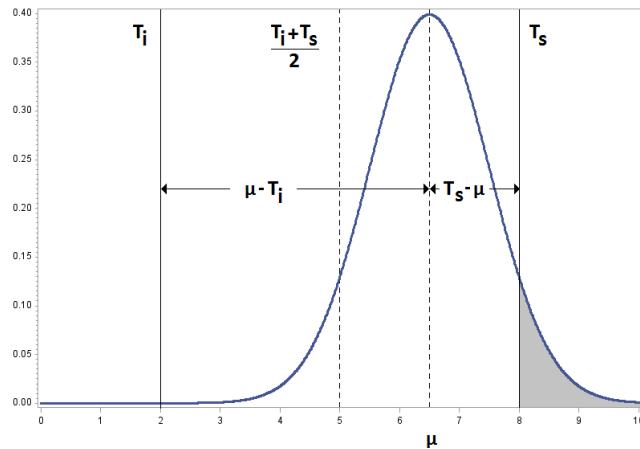
L'indice  $C_{pk}$  est alors défini comme étant la plus petite de ces deux valeurs quantifiant le décentrage à droite ou à gauche :

#### Définition 05

Etant donné un processus, de valeur moyenne  $\mu$  et d'écart-type  $\sigma$ , devant respecter la zone de tolérance  $[T_i, T_s]$  on définit son **indice d'aptitude**  $C_{pk}$  par :

$$C_{pk} = \min \left( \frac{T_s - \mu}{3\sigma}, \frac{\mu - T_i}{3\sigma} \right)$$

**Remarque.** L'indice  $C_{pk}$  est aussi appelé **indice d'aptitude effective** ou indice de « **capabilité** » **effective**.



**Figure 04.** Indice d'aptitude  $C_{pk}$ .

### 2.2.2 Propriétés et remarques

Les propriétés suivantes sont immédiates :

- 1) On a toujours  $C_{pk} > 0$  (car la valeur moyenne  $\mu$  ne peut raisonnablement pas être en dehors de l'intervalle  $[T_i, T_s]$ ) et plus la valeur de  $C_{pk}$  est grande **meilleur** est le processus étudié.
- 2) La valeur théorique pivot est  $C_{pk} = 1$ . Lorsque  $C_{pk} > 1$  alors le processus est qualifié de **apte** pour cet indice. Dans la pratique industrielle il est courant d'imposer des normes plus contraignantes et de déclarer un processus apte pour l'indice  $C_{pk}$  si et seulement si :

$$C_{pk} > 1.33 \text{ (encore écrit sous la forme } C_{pk} > 133\%).$$

3) Comparons maintenant les indices  $C_p$  et  $C_{pk}$ . Pour un processus parfaitement centré (c'est-à-dire tel que  $\mu = (T_i + T_s) / 2$ ) il vient :

$$\frac{T_s - \mu}{3\sigma} = \frac{\mu - T_i}{3\sigma} = \frac{T_i + T_s - 2T_i}{6\sigma} = \frac{T_s - T_i}{6\sigma} \Rightarrow C_{pk} = C_p.$$

On démontre sans difficulté la réciproque donc :

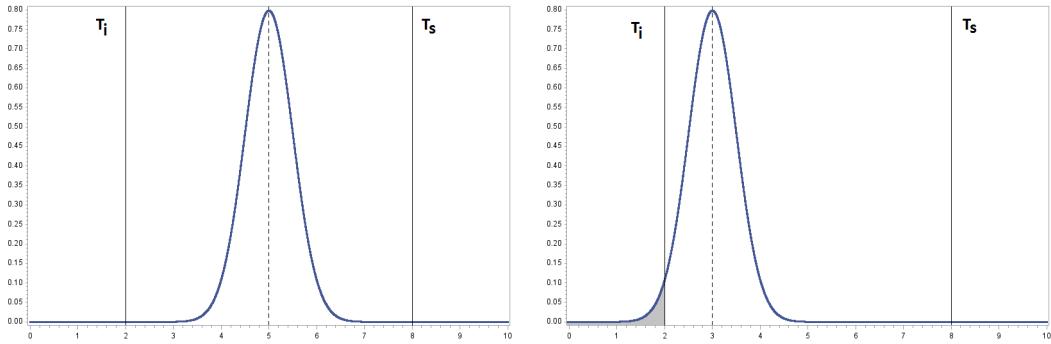
$$C_{pk} = C_p \text{ si et seulement si le processus étudié est centré.}$$

De même, il est évident que  $C_{pk} < C_p$  si et seulement si le processus étudié est décentré.

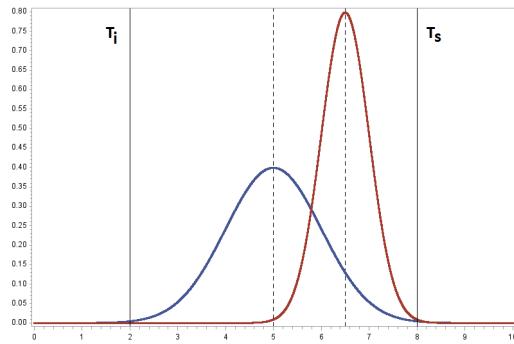
#### Utilisation pratique

La grande qualité de l'indice  $C_{pk}$  est d'enrichir les informations fournies par l'indice précédent  $C_p$ . Le calcul simultané de ces deux indices est alors particulièrement utile (voir la figure 05). Deux aspects négatifs sont cependant à signaler :

- 1) la loi de probabilité de  $C_{pk}$  est souvent complexe à déterminer,
- 2) deux processus ayant des comportements très différents peuvent être associés à une même valeur de l'indice  $C_{pk}$  (voir la figure 06 ... l'idée étant que diminuer la valeur du décalage - donc du numérateur - et diminuer de manière identique la dispersion - donc le dénominateur - conduit à une valeur identique du  $C_{pk}$ ).



**Figure 05.** Exemple de comparaison de deux processus de production  
 $C_p > 1$  et  $C_{pk} > 1$  (à gauche),  $C_p > 1$  et  $C_{pk} < 1$  (à droite).



**Figure 06.** Exemple de deux processus à  $C_{pk}$  égaux.

## 2.3 Indice $C_{pm}$ d'aptitude d'un processus

### 2.3.1 Définitions

Au lieu de se focaliser uniquement sur les limites de tolérance  $T_i$  et  $T_s$  (comme pour les deux indices précédents) il est parfois plus intéressant d'utiliser un indice faisant intervenir la **valeur cible**  $\tau$  souhaitée (c'est-à-dire la valeur optimale attendue pour la production). Cette valeur cible est souvent située au milieu de la zone de tolérance  $[T_i, T_s]$  mais ceci n'a rien d'obligatoire.

Il est alors naturel de faire intervenir l'**erreur quadratique moyenne** du processus, c'est-à-dire la quantité suivante (puisque  $\sigma^2 = \text{Var}(X) = E(X^2) - \mu^2$ ) :

$$\begin{aligned}\theta^2 &= E[(X - \tau)^2] = E(X^2 - 2\tau X + \tau^2) = E(X^2) - 2E(\tau X) + E(\tau^2) \\ &= E(X^2) - 2\tau\mu + \tau^2 \\ &= (\sigma^2 + \mu^2) - 2\tau\mu + \tau^2 \\ \theta^2 &= \sigma^2 + (\mu - \tau)^2\end{aligned}$$

L'indice  $C_{pm}$  est alors défini de la manière suivante :

$$\Rightarrow \theta = \sqrt{\sigma^2 + (\mu - \tau)^2}$$

### Définition 06

Etant donné un processus, de valeur moyenne  $\mu$  et d'écart-type  $\sigma$ , on définit son **indice d'aptitude**  $C_{pm}$  par :

$$C_{pm} = \frac{T_s - T_i}{6\theta}$$

Il en découle donc que :

$$C_{pm} = \left( \frac{T_s - T_i}{6\sigma} \right) \frac{\sigma}{\theta} = \frac{C_p}{\sqrt{1 + \frac{(\mu - \tau)^2}{\sigma^2}}}$$

### 2.3.2 Propriétés et remarques

Les propriétés suivantes sont immédiates :

- 1) On a toujours  $C_{pm} > 0$  et plus la valeur de  $C_{pm}$  est grande **mieux** est le processus étudié.
- 2) La valeur théorique pivot est  $C_{pm} = 1$ . Lorsque  $C_{pm} > 1$  alors le processus est qualifié de **apte** pour cet indice. Dans la pratique industrielle il est courant d'imposer des normes plus contraignantes et de déclarer un processus apte pour l'indice  $C_{pm}$  si et seulement si :

$$C_{pm} > 1.33 \text{ (encore écrit sous la forme } C_{pm} > 133\%).$$

- 3) On a  $C_{pm} = C_p$  si et seulement si  $\mu = \tau$  (c'est-à-dire si et seulement si le processus est parfaitement centré sur la cible visée). Dans tous les autres cas il est clair que  $C_{pm} < C_p$ .

### Utilisation pratique

L'indice  $C_{pm}$  présente l'avantage d'être directement lié à la valeur cible  $\tau$  mais en contrepartie son interprétation en terme de zone de tolérance est complexe (par exemple le lien entre valeur de  $C_{pm}$  et proportion attendue de produits hors de la zone de tolérance n'est pas du tout évidente).

Il en découle que cet indice  $C_{pm}$  ne doit pas, en général, être calculé seul mais **simultanément** aux deux autres indices  $C_p$  et  $C_{pk}$  étudiés précédemment. Ceci est illustré par l'exemple ci-dessous.

**Exemple.** Supposons qu'une production suive une loi normale de type  $N(16, 2/3)$  et que la zone de tolérance imposée soit  $[10, 18]$ . Admettons aussi que la valeur cible soit au milieu de la zone de tolérance. Il vient alors (voir la figure 07 pour une représentation graphique) :

- 1)  $C_p = 8/6\sigma = 2$ ,  $\left( \frac{T_s - \tau}{6\sigma} \right) = \frac{8}{6\sigma} = \frac{8}{6 \cdot \frac{2}{3}} = \frac{8}{4} = 2$
- 2)  $C_{pk} = \min(6/3\sigma, 2/3\sigma) = 1$ ,  $C_{pk} = \min \left( \frac{\frac{18-16}{2/3}}{3\sigma}, \frac{\frac{16-10}{2/3}}{3\sigma} \right) = \min \left\{ \frac{18-16}{3 \cdot \frac{2}{3}}, \frac{16-10}{3 \cdot \frac{2}{3}} \right\} = \min \left\{ \frac{2}{3}, \frac{8}{3} \right\} = \min \{ 1, \frac{8}{3} \} = \min \{ 1, 2.67 \}$
- 3)  $C_{pm} = 2/\sqrt{1 + 4/\sigma^2} = 2/\sqrt{10} \approx 0.63$ .  $C_{pm} = \frac{2}{\sqrt{1 + \frac{(16-14)^2}{\frac{4}{9}}}} = \frac{2}{\sqrt{1 + \frac{4 \cdot 16}{4}}} = \frac{2}{\sqrt{1 + 16}} = \frac{2}{\sqrt{17}} = \frac{2}{\sqrt{10}}$

On en déduit alors que ce processus est associé à une **bonne valeur** de l'indice  $C_p$  (i.e. sa tolérance naturelle est compatible avec la zone de tolérance imposée), une **valeur limite** de l'indice  $C_{pk}$  (i.e. il commence à être dangereusement décentré par la droite) et une **mauvaise valeur** de l'indice  $C_{pm}$  (i.e. ce processus est trop éloigné en général de la valeur cible fixée - le milieu de la zone de tolérance ici -).

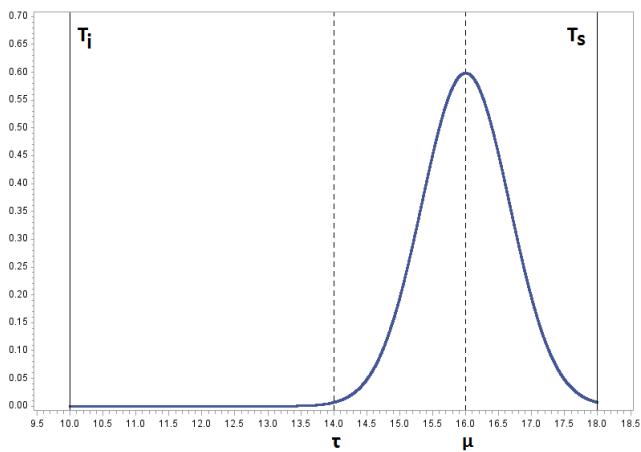


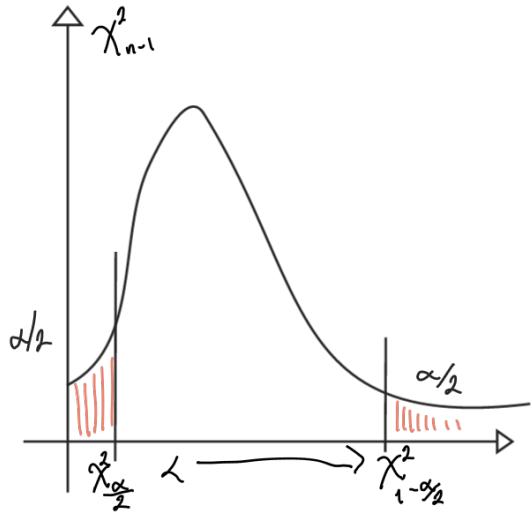
Figure 07. Exemple.

Problème: Intervalle de confiance pour  $C_p$   
Comment calculer en pratique

$$C_p = \frac{T_s - T_i}{6\sigma}$$

On estime  $\sigma$  par la valeur  $s \Rightarrow \hat{C}_p = \frac{\bar{T}_s - \bar{T}_i}{6s}$

Or on sait que  $\frac{n-1}{\sigma^2} s^2 \sim \chi^2_{n-1}$  donc  $P\left[\frac{\chi^2}{\alpha/2, n-1} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{1-\alpha/2, n-1}\right] = 1-\alpha$



Quand  $C_{pk} = C_p$ ? Si le processus est centré sur  $[T_i, T_s]$  alors  $M = \frac{T_i + T_s}{2}$

$$C_{pk} = \min\left(\frac{T_s - M}{3\sigma}, \frac{M - T_i}{3\sigma}\right) = \frac{T_s - M}{3\sigma} = \frac{T_s - \frac{T_i + T_s}{2}}{3\sigma} = \frac{T_s - T_i}{6\sigma} = C_p$$

$$X \sim N(16, \frac{2}{3}) \quad \text{avec} \quad T_i = 10 \quad T_s = 18 \quad \Rightarrow \bar{\tau} = 14$$

$$\cdot C_p = \frac{8}{14} = 2$$

$$\cdot C_{pk} = \min\left(\frac{T_s - M}{3\sigma}, \frac{M - T_i}{3\sigma}\right) = \frac{T_s - M}{3\sigma} = \frac{2}{3} = \boxed{1}$$

$$\cdot C_{pk_m} = \frac{2}{\sqrt{1 + 4\left(\frac{2}{3}\right)^2}} = \frac{2}{\sqrt{10}} \approx 0.63$$

**Maîtrise Statistique des Procédés**

Master MSiD 2ème année

Travaux dirigés - Série 2

**Exercice 1** *Indice d'aptitude.*

On s'intéresse ici au diamètre extérieur de pistons utilisés dans le système de freinage de véhicules. Un échantillon de 15 pistons a été prélevé dans la production et les mesures effectuées (en mm) sont données ci-dessous

33.9974	34.0161	33.9982	34.0119	33.9999
34.0092	34.0014	34.0083	34.0027	34.0060
34.0041	34.0056	34.0042	34.0052	34.0044

Un client souhaiterait lancer une production industrielle avec les limites de tolérance  $T_l = 33.99$  et  $T_S = 34.01$ .

**1** Déterminer l'indice  $\hat{C}_p$  estimé à l'aide de l'échantillon. Que peut-on en conclure ?

Un autre client (moins exigeant) est prêt à travailler avec les limites de tolérance  $T_l = 33.975$  et  $T_S = 34.025$ .

**2** Déterminer l'indice  $\hat{C}_p$  estimé à l'aide de l'échantillon.

**3** On suppose ici que la i-ème observation  $x_i$  est une réalisation de la v.a.r.  $X_i$  de loi normale (avec les  $X_i$  indépendantes) et on note :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Démontrer que  $(n-1)S^2/\sigma^2$  suit une loi du  $\chi^2$  à  $(n-1)$  degrés de liberté.

*A partir de cette*

**4** En déduire un intervalle de confiance pour  $C_p$  (avec un coefficient de sécurité de 5%). Que peut-on en conclure ?

**Exercice 2** *Indice d'aptitude.*

Considérons l'installation d'une chaîne industrielle de fabrication de limonade. Le produit fini doit se présenter sous forme d'une bouteille contenant (théoriquement) un litre de limonade. Le service qualité de l'entreprise estime que cet objectif doit être atteint avec une précision de  $\pm 10\text{ ml}$  pour que le produit soit jugé acceptable. Après étude des divers processus industriels envisagés sur cette chaîne de production il apparaît que les quantités de limonade réellement produites vont suivre une loi normale  $N(x, 2)$ . On aimerait alors savoir de quelle "marge de manoeuvre" on dispose ici pour le réglage de la valeur moyenne  $x$ .

Conseils

[1] Quelles informations sont données par le calcul de l'indice  $C_p$  ?

[2] Quelles informations sont données par le calcul de l'indice  $C_{pk}$ ? Quel réglage optimal du paramètre  $x$  peut-on proposer?

[3] Quelles informations sont données par le calcul de l'indice  $C_{pm}$  (considérer la valeur cible  $\tau = 1000$ )?

## Exercices

### Exercice 1 { 1) Calcul des estimateurs classiques de $\mu$ et $\sigma^2$

$$\bar{x}, s^2$$

On a  $\bar{x} = 34,0050$  et  $s^2 = \frac{1}{n-1} \left( \sum_{i=1}^{15} (x_i - \bar{x})^2 \right) \approx 2 \cdot 10^{-5}$

Conclusion  $C_p = \frac{\bar{T}_S - \bar{T}_U}{6\sigma} \approx \hat{C}_p = \frac{\bar{T}_S - \bar{T}_U}{6s} = \frac{0,002}{6 \cdot 2 \cdot 10^{-5}} \approx 0,67$

⇒ Les exigences du client sont impossible à satisfaire,  $C_p > 1$  *no pasa*

2)  $\bar{T}_U = 33,975 / \bar{T}_S = 34,025$

On obtient ici  $\hat{C}_p = \frac{\bar{T}_S - \bar{T}_U}{6s} \approx 1,67$  (Valeur de la question 1.28)

Donc si les processus est bien centré il devrait ici satisfaire le client.

3) Montrer que  $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$   
Pour  $i=1, \dots, n$  notons par la suite  $*U_i = \frac{(x_i - \mu)}{\sigma}$   
On sait que les  $U_i$  sont des v.ar. indépendantes de même loi  $N(0,1)$   
on a  $\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i - \mu + \mu - \bar{x})^2 = \sum_{i=1}^n (x_i - \mu)^2 + 2 \sum_{i=1}^n (x_i - \mu)(\mu - \bar{x}) = n(\bar{x} - \mu)^2 + 2(\mu - \bar{x}) \sum_{i=1}^n (x_i - \mu)$

Conclusion  $\underbrace{\sum_{i=1}^n (x_i - \bar{x})^2}_{(n-1)S^2} = \sum_{i=1}^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2$   
 $* \sigma^2 (\sum_{i=1}^n U_i^2)$

On a alors  $\sum_{i=1}^n U_i^2 = (n-1) \frac{\sigma^2}{\sigma^2} + n \left( \frac{\bar{x} - \mu}{\sigma} \right)^2$

Notons  $V = \begin{pmatrix} V_1 \\ \vdots \\ V_n \end{pmatrix} \sim N(0, I_n)$  et posons  $V = AU$  avec  $A$  matrice orthogonale de première ligne  $(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$

On sait que  $V = AU \sim N_n(0, A I_n^T A) = N_n(0, I_n)$

On en déduit alors que:

$$\textcircled{1} V_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n (U_i) = \frac{1}{\sigma \sqrt{n}} \left( \sum_{i=1}^n (x_i - \mu) \right) = \frac{\sqrt{n}}{\sigma} (\bar{x} - \mu)$$

$$\textcircled{2} \sum_{i=1}^n V_i^2 = {}^T W = {}^T U^T A A U = {}^T U U = \sum_{i=1}^n U_i^2$$

Conclusion

$$\Rightarrow \sum_{i=1}^n V_i^2 = \frac{(n-1)\sigma^2}{\sigma^2} + V_1^2 \text{ mais } \sum_{i=1}^n V_i^2 = \sum_{i=1}^n V_i^2$$

$$\Rightarrow \sum_{i=1}^n V_i^2 = \frac{(n-1)\sigma^2}{\sigma^2} + V_1^2 \Rightarrow \frac{(n-1)\sigma^2}{\sigma^2} = \left( \sum_{i=2}^n V_i^2 \right) \sim \chi^2(n-1)$$

$C_p$

4) Intervalle de confiance pour  $C_p$   
D'après le cours on sait que l'intervalle de confiance au niveau pour  $C_p$  s'écrit

$$\left[ \hat{C}_p \sqrt{\frac{\chi^2_{\alpha/2, n-1}}{n-1}}, \hat{C}_p \sqrt{\frac{\chi^2_{1-\alpha/2, n-1}}{n-1}} \right]$$

avec ici  $n-1=14$  et  $\alpha=0,005$  et  $\chi^2_{0,025, 14} \approx 5.63$  et  $\chi^2_{0.975, 14} \approx 26.12$

En conclusion il y a 95% de chance que la vraie valeur de  $C_p$

soit dans l'intervalle  $[1.057, 2.277]$

Donc ce processus est opté (pour l'indice  $C_p$ ) au niveau 5%.

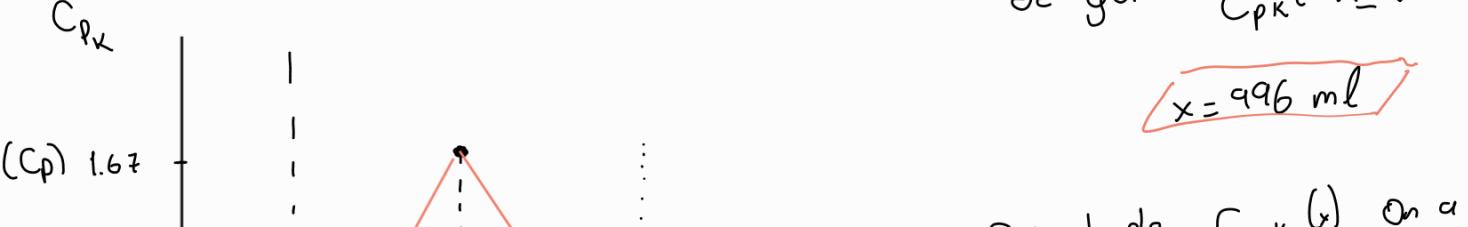
## Exercice 2

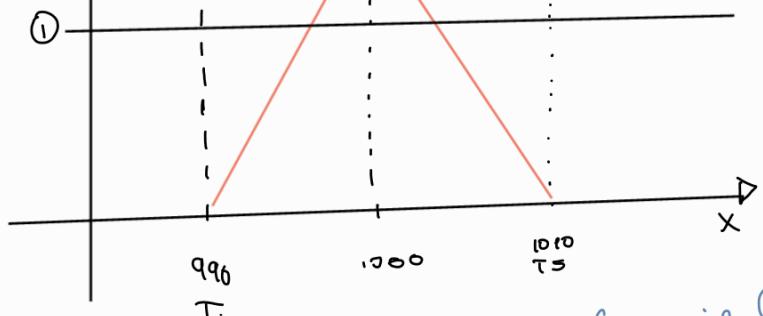
$$1) \text{ On a ici } C_p = \frac{T_s - T_i}{6\sigma} = \frac{1010 - 990}{12} = \frac{20}{12} \approx 1.67$$

Donc si le processus est bien centré il va parfaitement "entier" dans la zone de tolérance  $[T_i, T_s]$

2) Valeur minimale de  $x$  permettant de garder  $C_{pk}(+) \geq 1$ ?

$$x = 996 \text{ ml}$$





$$\text{Calcul de } C_{pk}$$

$$C_{pk}(x) = \min \left( \frac{T_U - M}{3\sigma}, \frac{M - T_L}{3\sigma} \right)$$

$$= \min \left( \frac{1010 - x}{6}, \frac{x - 990}{6} \right)$$

Quelle stratégie utiliser? i.e.  $C_{pk} = C_p$  i.e.  $C_{pk} = 1$   
 $x = 1,000$  ou  $x = 996?$

Raisonnons sur 1 000 000 de litres

① Choix "naturel"  $x = 1000$  (exactement sur la cible)  
 Le choix devrait permettre en moyenne de remplir 1 000 000 de bouteilles

Probabilité de rejet d'une bouteille trop remplie  $P[x > 1010]$  avec  $x \sim N(1000, 2)$

On a  $P[x > 1010] = P[x^* > 5] \approx 2.867 \times 10^{-7}$

Nombre moyen de bouteilles refusées sur le million produites  $ppm = 10^6 (2.867 \times 10^{-7}) \approx 0.58$   
 Il devrait n'y avoir aucune refus ici

② Choix alternatif de  $x = 996$   
 En moyenne on devrait ici remplir 1 004 016 bouteilles. Cherchons la probabilité de refus d'une bouteille pas assez remplie (trop remplie quasiment nul):

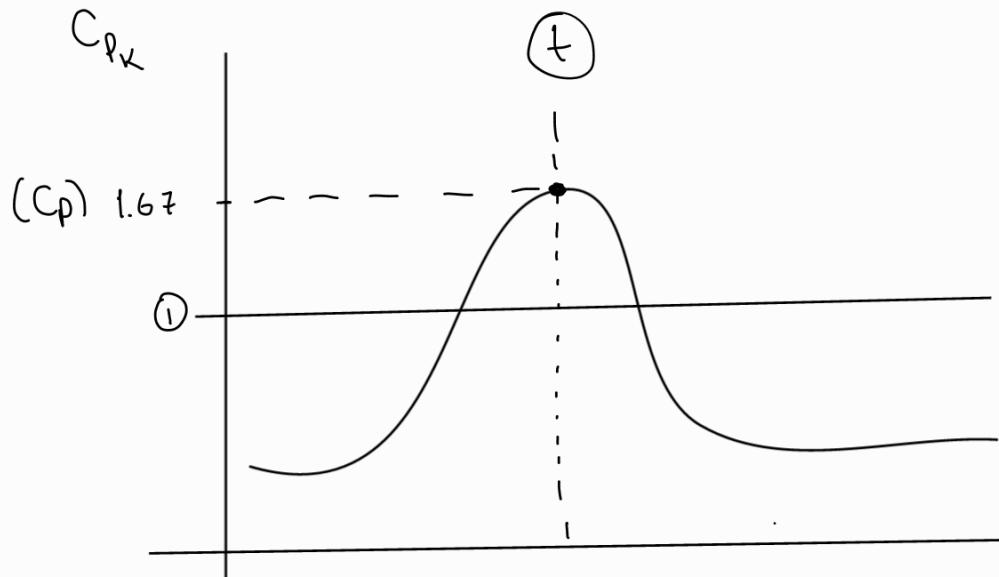
$P[x < 990]$  avec  $x \sim (996, 2)$

Alors  $P[x < 990] = P[x^* < -3] \approx 0.00135$

En conclusion sur la totalité des bouteilles produites les nombre moyen de bouteilles pas assez remplies est  
 $ppm = 1004016 \times 0.00135 \approx 1355$

Ce réglage conduit donc à la vente de plus de bouteilles que le choix  $x = 1,000$   
 Le plus rentable en termes de  $C_{pk}$  est  $x = 996$

③ Recherche  $C_{pm} = C_{pm}(x)$ ,  $C_{pm} = \frac{C_p}{\sqrt{1 + \frac{(M-T)^2}{\sigma^2}}} = \frac{1.67}{\sqrt{1 + \frac{(x-1000)^2}{4}}}$   
 La approximation avec le cible



Plus petit x solution de  $C_{pm}(x) = 1$ :  
 $x \approx 997,3 \text{ ml}$

Dans l'optique de produits plus de bouteilles tout en gardant des indices C "connectés" on peut donc prendre  $x = 997,3 \text{ ml}$

$$\left( C_p > 1, C_{PK} > 1, C_{PM} > 1 \right)$$

## Chapitre 3

# Cartes de contrôle aux mesures

### 3.1 Introduction

Une carte de contrôle est un **outil graphique** permettant de visualiser l'évolution dans le temps d'une caractéristique de qualité mesurable (masse, diamètre, longueur, etc ...) tout en étant capable de détecter à tout moment si le processus est **sous contrôle ou non**.

D'après la norme internationale ISO un processus est dit **sous contrôle** (ou encore stable) lorsque « les écarts entre les résultats observés sur un échantillon peuvent être attribués à un ensemble de causes aléatoires qui ne paraît pas se modifier dans le temps » (en d'autres termes si et seulement si toutes les observations  $x_i$  sont des réalisations d'une même loi  $X$ ).

Graphiquement une carte de contrôle a souvent une structure identique à celle que la figure 08 présentée ci-dessous.

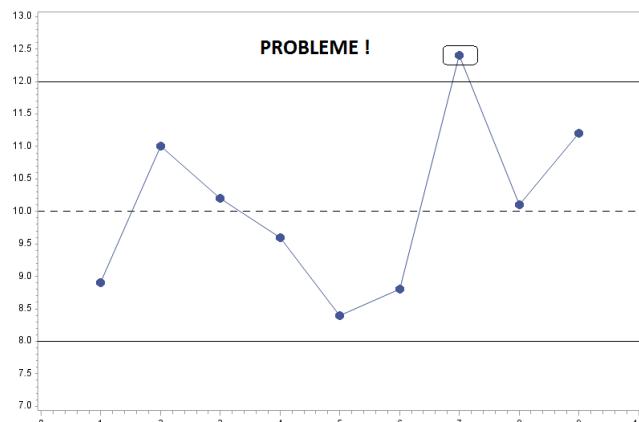


Figure 08. Exemple de carte de contrôle.

Les caractéristiques de qualité étudiées sont représentées sous forme de points (associés à l'échantillon testé au fil du temps). Une référence « centrale » (ligne centrale représentée classiquement en pointillés) correspond à la position optimale théorique pour le processus sous contrôle. Deux **limites de contrôle** (supérieure et inférieure) constituent la spécificité de toute carte : lorsque la situation est sous contrôle quasiment toute la production doit se retrouver entre ces deux limites. A contrario sortir de ces limites va être identifié comme une situation atypique.

Sauf cas particulier toute carte de contrôle est basée sur l'observation périodique d'échantillons constitués par  $n > 1$  observations, appelés **sous-groupes rationnels** (observer plus d'une valeur à la fois permet d'avoir des observations concernant la dispersion du processus). Ces échantillons, recueillis à intervalles de temps régulier par le service qualité, ont tout intérêt (pour des raisons de coût et de temps)

à être de petite taille. Remarquons enfin que l'on distingue en général les deux types suivants de cartes de contrôles (présentées ici dans le cadre d'un processus de loi normale) :

1) les cartes de **phase I** (ou d'étude initiale) lorsque le processus sous contrôle est de loi  $N(\mu, \sigma)$  avec  $\mu$  et  $\sigma$  inconnus,

2) les cartes de **phase II** (ou aux valeurs standard) lorsque le processus sous contrôle est de loi  $N(\mu_0, \sigma_0)$  avec  $\mu_0$  et  $\sigma_0$  donnés.

Concrètement la phase I correspond à un processus industriel nouveau dont on ne connaît pas le comportement (les paramètres des lois devront être estimés) alors que la phase II correspond à l'industrialisation d'un produit bien connu (les valeurs des paramètres des lois sont alors données par diverses normes).

## 3.2 Carte de contrôle de type $(\bar{X}, R)$ *Carte de contrôle moyen étendue*

Considérons cette première carte de contrôle, très connue et très utilisée, encore appelée carte de contrôle moyenne-étendue de Shewhart (en référence à W.A. Shewhart - chercheur à la Western Electric - inventeur des cartes de contrôle dans les années 30). Nous allons considérer dans toute cette section le cas le plus complexe d'une carte de phase I (en cas de phase II sauter simplement l'étape d'estimation des paramètres de la loi).

L'idée de ce type de cartes, adaptées à un processus suivant sous contrôle une loi normale  $N(\mu, \sigma)$  est :

- 1) de contrôler la position du processus à l'aide des diverses moyennes des sous-groupes rationnels,
- 2) de contrôler la dispersion du processus à l'aide des diverses étendues des sous-groupes rationnels.

### 3.2.1 Analyse des observations d'un sous-groupe rationnel

De manière générale supposons que l'on dispose de  $k$  sous-groupes rationnels tels que chacun d'eux est constitué par  $n$  observations (voir les travaux dirigés pour une situation où la taille des sous-groupes rationnels n'est pas constante au fil du temps). Désignons les observations du sous-groupe rationnel  $i$  ( $1 \leq i \leq k$ ) par  $x_{i1}, \dots, x_{in}$  (elles sont donc des réalisations des variables aléatoires réelles  $X_{i1}, \dots, X_{in}$ ).

**Carte  $\bar{X}$**  Pour ce type de carte l'idée est de s'intéresser aux moyennes de chacun des sous-groupes rationnels, c'est-à-dire aux réalisations des variables aléatoires réelles : La stabilité à rapport de la moyenne

$$\forall i = 1, \dots, k, \quad \bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}$$

Si chacune des variables aléatoires réelles  $X_{ij}$  suit une loi normale  $N(\mu, \sigma)$  il est alors immédiat que (d'après l'hypothèse d'indépendance des ces variables aléatoires)

$$\bar{X}_i \rightsquigarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

**Carte  $R$**  Pour ce type de carte l'idée est de s'intéresser aux étendues de chacun des sous-groupes rationnels, c'est-à-dire aux réalisations des variables aléatoires réelles :

avec  $K$  sous-groupes rationnels  $\forall i = 1, \dots, k, \quad R_i = X_{i(n)} - X_{i(1)}$

Lorsque les  $X_{ij}$  sont des variables aléatoires indépendantes de même loi normale alors  $R_i$  suit une loi de probabilité complexe mais que l'on peut néanmoins déterminer explicitement par le biais de sa fonction de répartition. On montre alors que si  $X_{ij}$  suit une loi normale  $N(\mu, \sigma)$  et si chaque sous-groupe rationnel est de taille  $n$  alors :

$$E(R_i) = d_2(n)\sigma \text{ et } \sigma(R_i) = d_3(n)\sigma$$

où les coefficients  $d_2$  et  $d_3$  sont tabulés.

#### Avantages et inconvénients de l'utilisation de l'étendue

L'étendue n'est un indicateur de dispersion couramment utilisé en statistique « générale ». Ses avantages et inconvénients sont résumés ci-dessous :

- ✓ grande facilité de calcul,
- ✓ le calcul de  $R_i$  ne dépend pas de celui de  $\bar{X}_i$ . Ceci est particulièrement intéressant lorsque la moyenne du processus n'est plus sous contrôle.
- ✗ l'étendue est clairement moins riche en informations qu'un indicateur tel que l'écart-type puisqu'elle n'utilise que les deux termes extrêmaux. Cette perte d'information est cependant peu importante lorsque  $n$  est faible (ce qui est le cas en MSP où les sous-groupes rationnels sont en général de petite taille).

#### 3.2.2 Estimation de $\mu$ et $\sigma$

Dans le cas d'une carte de contrôle de phase I il est nécessaire d'estimer les quantités  $\mu$  et  $\sigma$ . L'idée est alors d'utiliser un maximum d'information, c'est-à-dire la totalité des  $nk$  observations.

**Estimation de  $\mu$**  Nous avons déjà vu que, pour chaque sous-groupe rationnel,  $\bar{X}_i$  est un estimateur sans biais de  $\mu$ . Ces variables aléatoires ayant toutes une même dispersion on sait qu'alors un estimateur sans biais plus efficace que chacun des  $\bar{X}_i$  est obtenu à partir de la population totale via la relation suivante :

$$\hat{\mu} = \bar{\bar{X}} = \frac{1}{k} \sum_{i=1}^k \bar{X}_i$$

**Estimation de  $\sigma$**  Nous avons déjà vu de même que, pour chaque sous-groupe rationnel,  $R_i/d_2(n)$  est un estimateur sans biais de  $\sigma$ . Ces variables aléatoires ayant toutes une même dispersion on sait qu'alors un estimateur sans biais plus efficace que chacun des  $R_i/d_2(n)$  est obtenu à partir de la population totale via la relation suivante :

$$\hat{\sigma} = \frac{\bar{R}}{d_2(n)} = \frac{1}{kd_2(n)} \sum_{i=1}^k R_i \Rightarrow \text{para los límites}$$

$$3 + \frac{\hat{\sigma}}{\sqrt{n}} = 3 + \frac{\bar{R}}{d_2(n)\sqrt{n}}$$

#### 3.2.3 Construction des cartes de contrôle

Appliquons alors tous les résultats vus précédemment afin de constituer des cartes de contrôle selon le procédé proposé par Shewhart c'est-à-dire en positionnant les limites de contrôle à l'aide de la règle des  $3\sigma$ .

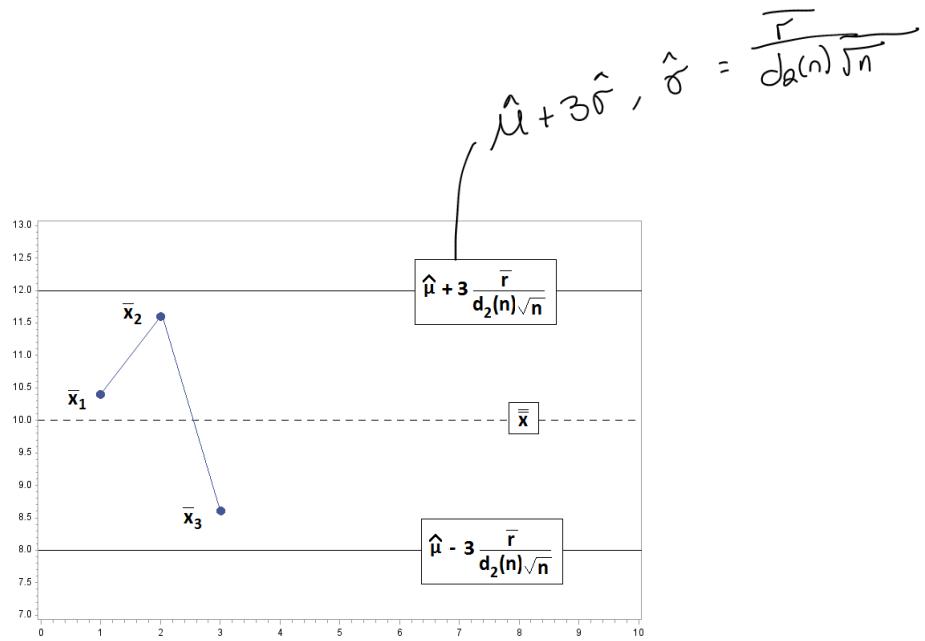
**Carte  $\bar{X}$**  Le principe de construction est le suivant :

1) positionnement de la ligne centrale à la valeur théorique moyenne suivante :

$$E(\bar{X}_i) = \mu \simeq \hat{\mu} = \bar{x}$$

2) positionnement des limites de contrôle selon la règle des  $3\sigma$  à partir de :

$$\sigma(\bar{X}_i) = \frac{\sigma}{\sqrt{n}} \simeq \frac{\hat{\sigma}}{\sqrt{n}} = \frac{\bar{r}}{d_2(n)\sqrt{n}}$$



**Figure 09.** Carte de contrôle  $\bar{X}$ .

**Remarque.** Dans le cas d'une carte de contrôle de phase II la partie « estimation » est inutile et les limite inférieure de contrôle, ligne centrale et limite supérieure de contrôle sont respectivement données par (avec  $\mu_0$  et  $\sigma_0$  caractéristiques connues pour le processus étudié) :

$$\begin{array}{c} \mu_0 + 3 \frac{\sigma_0}{\sqrt{n}} / \mu / \mu_0 + 3 \frac{\sigma_0}{\sqrt{n}} \\ \text{limite inférieure} / \text{ligne} \text{ centrale} / \text{limite supérieure} \end{array}$$

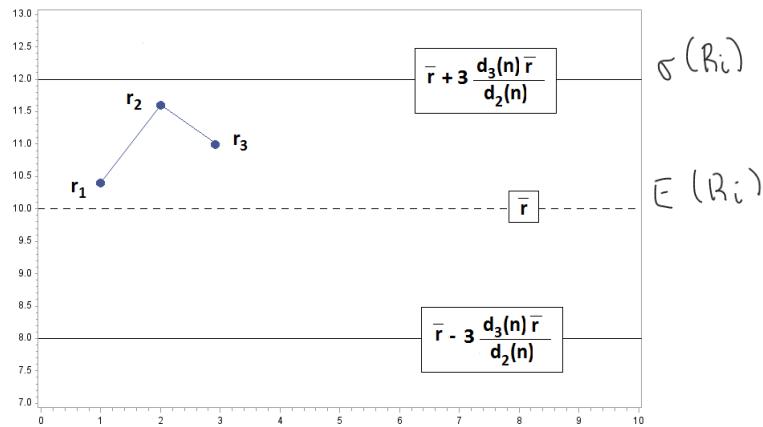
[Carte R] Le principe de construction est le suivant :

- 1) positionnement de la ligne centrale à la valeur théorique moyenne suivante :

$$E(R_i) = d_2(n) \sigma \simeq d_2(n) \hat{\sigma} = \bar{r}$$

- 2) positionnement des limites de contrôle selon la règle des  $3\sigma$  à partir de :

$$\sigma(R_i) = d_3(n) \sigma \simeq d_3(n) \hat{\sigma} = \frac{d_3(n)}{d_2(n)} \bar{r}$$



**Figure 10.** Carte de contrôle  $R$ .

**Remarque.** Dans le cas d'une carte de contrôle de phase II la partie « estimation » est inutile et les limite inférieure de contrôle, ligne centrale et limite supérieure de contrôle sont respectivement données par (avec  $\mu_0$  et  $\sigma_0$  caractéristiques connues pour le processus étudié) :

$$d_2(n) \sigma_0 - 3d_3(n) \sigma_0 / d_2(n) \sigma_0 / d_2(n) \sigma_0 + 3d_3(n) \sigma_0$$

*Observaciones de variables que siguen una normal, pero las observaciones no lo hacen necesariamente*

**Remarque.** Attention au fait que, contrairement à la carte de contrôle précédente, les quantités représentées sont ici des réalisations des variables aléatoires réelles  $R_i$  qui, même sous hypothèse de normalité des observations, ne suivent pas naturellement une loi normale ! Il peut en découler des problèmes concernant les limites de contrôle ainsi calculées (par exemple obtenir LIC négative, ce qui n'a pas de sens).

### 3.3 Carte de contrôle de type $(\bar{X}, S)$

Si l'on estime que l'étendue n'est pas un indicateur de dispersions assez précis (ce qui est notamment le cas pour des sous-groupes rationnels assez grands) on peut alors souhaiter remplacer l'étendue par l'écart-type empirique  $S$ . Nous détaillons ici la construction de la carte de contrôle liée à cette quantité et toujours selon la technique proposée par Shewhart (la carte  $\bar{X}$  ayant pour sa part déjà été construite à la section 3.2).

#### → Analyse des observations d'un sous-groupe rationnel

Considérons toujours  $k$  sous-groupes rationnels de taille  $n$ . La carte  $S$  consiste en la représentation des diverses valeurs  $s_i$  ( $1 \leq i \leq k$ ) qui sont des réalisations des variables aléatoires  $S_i$  telles que :

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$$

D'après la proposition 02 lorsque toutes les variables aléatoires réelles sont indépendantes et de loi normale  $N(\mu, \sigma)$  alors  $(n-1)S_i^2/\sigma^2$  suit une loi du khi-deux de type  $\chi_{n-1}^2$ . La variable aléatoire  $S_i^2$  est donc aussi un estimateur sans biais classique de la variance  $\sigma^2$ . Attention au fait qu'en général  $S_i$  n'est pas un estimateur sans biais de l'écart-type  $\sigma$  ! Afin de lever ce type de difficulté un coefficient de correction, noté  $c_4(n)$  et tabulé, a été proposé selon la définition suivante :

$$\frac{S_i}{c_4(n)} \text{ est un estimateur sans biais de } \sigma$$

(en d'autres termes  $E(S_i) = c_4(n)\sigma$ ).

#### → Estimation de $\sigma$

Dans le cadre d'une carte de contrôle de phase I on dispose alors des quantités  $S_1/c_4(n), \dots, S_k/c_4(n)$  qui sont des estimateurs sans biais de  $\sigma$ . Ces variables aléatoires ayant toutes une même dispersion on sait qu'alors un estimateur sans biais plus efficace que chacun des  $S_i/c_4(n)$  est obtenu à partir de la population totale via la relation suivante :

$$\hat{\sigma} = \frac{1}{kc_4(n)} \sum_{i=1}^k S_i = \frac{\bar{S}}{c_4(n)}$$

#### → Construction de la carte de contrôle

Le principe de construction est le suivant :

1) positionnement de la ligne centrale à la valeur théorique moyenne suivante :

$$E(S_i) = c_4(n)\sigma \simeq c_4(n)\hat{\sigma} = \bar{s}$$

2) positionnement des limites de contrôle selon la règle des  $3\sigma$  à partir de  $\sigma(S_i)$  qui est, à priori, inconnu. Remarquons cependant que :

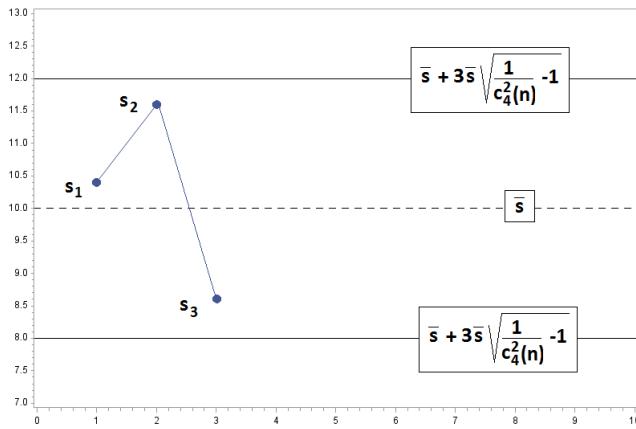
$$Var(S_i) = E(S_i^2) - [E(S_i)]^2 = E(S_i^2) - (c_4(n)\sigma)^2$$

On sait de plus que :

$$\frac{(n-1)}{\sigma^2} S_i^2 \rightsquigarrow \chi^2(n-1) \text{ donc } E\left[\frac{(n-1)}{\sigma^2} S_i^2\right] = n-1 \xrightarrow{\sigma^2 = \bar{s}^2} E(S_i^2) = \sigma^2$$

Il en découle que  $E(S_i^2) = \sigma^2$  donc  $Var(S_i) = \sigma^2(1 - c_4^2(n))$  et :

$$\sigma(S_i) = \sigma\sqrt{1 - c_4^2(n)} \simeq \hat{\sigma}\sqrt{1 - c_4^2(n)} = \bar{s}\sqrt{\frac{1}{c_4^2(n)} - 1}$$



**Figure 11.** Carte de contrôle  $S$ .

**Remarque.** Dans le cas d'une carte de contrôle de phase II la partie « estimation » est inutile et les limite inférieure de contrôle, ligne centrale et limite supérieure de contrôle sont respectivement données par (avec  $\mu_0$  et  $\sigma_0$  caractéristiques connues pour le processus étudié) :

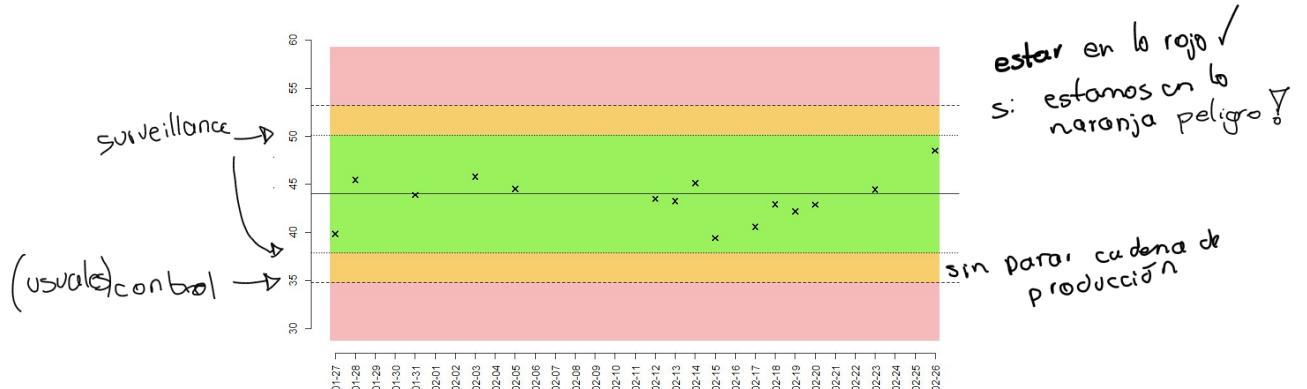
$$c_4(n)\sigma_0 - 3\sigma_0\sqrt{1 - c_4^2(n)} \quad / \quad c_4(n)\sigma_0 \quad / \quad c_4(n)\sigma_0 + 3\sigma_0\sqrt{1 - c_4^2(n)}$$

**Remarque.** Attention au fait que, une nouvelle fois, les quantités représentées dans cette carte de contrôle sont des réalisations de variables aléatoires réelles qui ne suivent pas naturellement une loi normale ! Il peut en découler des problèmes concernant les limites de contrôle ainsi calculées (par exemple obtenir LIC négative, ce qui n'a pas de sens).

## 3.4 Autres types de constructions

### 3.4.1 Limites de surveillance

Une convention consiste aussi à placer sur les cartes de contrôle de Shewhart deux lignes supplémentaires selon la « règle des  $2\sigma$  ». Ces deux valeurs sont appelées **limites inférieure et supérieure de surveillance**. Sous hypothèse de normalité la probabilité qu'un point de la carte se retrouve entre les limites de surveillance et les limites de contrôle est de l'ordre de 5%. Ceci est suffisamment faible pour envisager dans ce cas de procéder à un réglage des machines (sans cependant stopper la chaîne de production comme pour une situation hors contrôle) ou tout au moins à un examen plus poussé du processus en cours.



**Figure 12.** Carte de contrôle avec limites de surveillance.

Qué pasa cuando el proceso no es modelado por una approximación normal ?

### 3.4.2 Cartes à limites probabilistes

Les cartes de contrôle construites jusqu'à présent (dites « de Shewhart ») reposent sur la règle des  $3\sigma$ . Nous savons qu'une telle règle peut s'avérer imprécise dans tous les cas où le processus étudié ne peut être correctement approché par une loi normale. Une alternative consiste à utiliser alors la définition suivante :

#### Définition 07

Une carte de contrôle est dite à **limites de contrôle probabilistes** si et seulement si, pour une valeur  $0 < \alpha < 1$  donnée, chaque point de la carte a une probabilité égale à  $(1 - \alpha)$  de se trouver entre ces limites lorsque le processus est sous contrôle.

**Remarque.** Lorsque les observations sont de loi normale la carte de contrôle de type  $\bar{X}$  est une carte de contrôle à limites probabilistes associée à la valeur  $\alpha = 0.0027$  (application « exacte » de la règle des  $3\sigma$ ).

Exemple de construction d'une carte de contrôle à limites probabilistes

Considérons une carte de contrôle de type  $S$  dans le cas de la phase II et supposons que des limites de contrôles associées au risque  $\alpha = 1\%$  soient recherchées. Il faut donc déterminer ici des limites de contrôle  $LIC$  et  $LSC$  telles que, lorsque le processus est sous contrôle :

$$P[LIC < S < LSC] = 1 - \alpha$$

Déterminons-les de manière symétrique, c'est-à-dire telles que :

$$P[S < LIC] = P[S > LSC] = \alpha/2 = 0.005$$

Effectuons uniquement la construction de  $LSC$  (la méthode est identique pour  $LIC$ ). Il faut donc que, lorsque le processus est sous contrôle :

$$P[S > LSC] = P[S^2 > LSC^2] = P\left[\frac{(n-1)S^2}{\sigma_0^2} > \frac{(n-1)LSC^2}{\sigma_0^2}\right] = 0.005$$

Or nous savons que, lorsque le processus est sous contrôle,  $(n-1)S_i^2/\sigma_0^2$  suit une loi du khi-deux de type  $\chi_{n-1}^2$ . En désignant toujours par  $\chi_{\alpha,n}^2$  le fractile d'ordre  $\alpha$  de la loi du  $\chi^2$  à  $n$  degrés de liberté la limite de contrôle  $LSC$  doit donc vérifier :

$$\frac{(n-1)LSC^2}{\sigma_0^2} = \chi_{0.995,n-1}^2 \iff LSC = \sigma_0 \sqrt{\frac{\chi_{0.995,n-1}^2}{n-1}}$$

**Remarque.** Lorsque la carte de contrôle est de phase II il convient alors de remplacer  $\sigma_0$  par une estimation  $\hat{\sigma}$ .

## 3.5 Efficacité d'une carte de contrôle

### 3.5.1 Carte de contrôle et tests d'hypothèses

Remarquons que toute carte de contrôle n'est autre, d'un point de vue théorique, qu'une **suite de tests d'hypothèse**. En effet, pour chacun des sous-groupes rationnels, on teste alors l'hypothèse  $H_0$  : "le processus est sous-contrôle" (c'est-à-dire que la quantité mesurée semble bien suivre la loi de probabilité  $P_0$  postulée) contre  $\bar{H}_0$  : "le processus est déréglé". La règle de décision est donnée par :

on rejette  $H_0$  lorsque le point de contrôle est hors des limites de contrôle.

Comme pour tout test d'hypothèse les erreurs classiques suivantes sont à considérer :

1) l'erreur de **première espèce**, de probabilité  $\alpha$ , consistant à rejeter  $H_0$  alors qu'en réalité elle est vraie (concrètement ici cela correspond à une situation où l'on pense que le processus est déréglé alors qu'il n'en est rien), arrêter la chaîne de production alors que tous va bien

Test	Sous Contrôle	Hors contrôle
Réalité		
Sous contrôle	$1 - \alpha$	$\alpha$

23

que este fuera de control puede ser por  $\bar{r}$ , por  $M$  por variaciones

Hors contrôle	$\beta$	$1 - \beta$
---------------	---------	-------------

2) l'erreur de **seconde espèce**, de probabilité  $\beta$ , consistant à ne pas rejeter  $H_0$  alors qu'en réalité elle est fausse (concrètement ici cela correspond à une situation où l'on pense que le processus est sous contrôle alors qu'il n'en est rien).

La valeur de  $\alpha$  est en général fixée par l'utilisateur lors de la construction de la carte de contrôle. La valeur de  $\beta$ , difficile à déterminer dans le cas général, est souvent évaluée lorsque les paramètres du processus sous contrôle subissent un « déraglage type » (par exemple, dans le cas de la loi normale de référence  $N(\mu_0, \sigma_0)$  on peut transformer  $\mu_0$  en  $\mu = \mu_0 + \Delta$  ou encore  $\sigma_0$  en  $\sigma = \Delta\sigma_0$ , etc ...).

### 3.5.2 Période opérationnelle moyenne (POM) *Qualité des cartes de contrôle*

Pour toute carte de contrôle utilisée la question de la qualité de la carte se pose naturellement. Il en va de même si l'on souhaite comparer entre elles plusieurs cartes de contrôle adaptées à une même situation. On propose alors classiquement de quantifier la qualité d'une carte via la notion suivante :

#### Définition 08

Etant donnée une carte de contrôle sa **période opérationnelle moyenne** (encore notée POM) est égale au **nombre moyen** de **sous-groupes rationnels** successifs nécessaires afin de détecter **un dérèglement** du processus.

Déterminons la valeur théorique de la POM. Pour chaque sous-groupe rationnel la **probabilité de détecter le dérèglement** est égale à  $(1 - \beta)$ . Désignons par  $Y$  la variable aléatoire réelle égale au **nombre de sous-groupes rationnels nécessaires pour détecter le dérèglement**. Il est bien connu que  $Y$  suit alors une loi géométrique de paramètre  $(1 - \beta)$  et donc :

$$E(Y) = \frac{1}{1 - \beta}$$

D'où la conclusion suivante :

#### Proposition 09

Pour toute carte de contrôle sa **période opérationnelle moyenne** est donnée par :

$$POM = \frac{1}{1 - \beta}$$

avec  $\beta$  probabilité d'erreur de seconde espèce associée à la carte de contrôle considérée.

**Remarque.** Une nouvelle fois la détermination pratique de cette valeur est complexe à obtenir. Il est possible de le faire uniquement par rapport à un « dérèglement type » postulé à priori.

**Maîtrise Statistique des Procédés**

Master MSID 2ème année

Travaux dirigés - Série 3

**Exercice 1** Carte de contrôle de type  $(\bar{X}, R)$ .

On étudie ici le fonctionnement d'une pompe hydraulique et on mesure 5 fois par jour la pression à son embouchure (en  $kg$  par  $cm^2$ ). Les relevés sur 6 jours sont donnés ci-dessous et forment donc des sous-groupes de 5 observations :

1	2	3	4	5	6
344.4	355.9	351.6	350.4	345.9	339.5
341.5	345.4	361.9	346.1	335.1	333.8
360.4	372.3	348.4	351.9	343.8	336.2
358.8	341.2	361.9	364.5	342.5	336.4
349.6	345.8	343.4	348.0	334.0	353.9
$\bar{x}_i$	350.94	352.12	353.44	352.18	340.26
					340

On suppose ici que les observations relevées sont de loi normale.

- 1** Construire la carte de contrôle de Shewhart d'étude initiale (phase I) pour la moyenne et l'étendue. Détailler l'obtention des limites de contrôle.
- 2** Préciser quels sont les avantages et les inconvénients de l'étendue pour mesurer la dispersion.
- 3** Peut-on dire que pour l'instant la pression de la pompe est sous contrôle ?

**Exercice 2** Carte de contrôle de type  $(\bar{X}, S)$ .

On veut étudier ici la stabilité du processus d'usinage de rotors fabriqués au rythme de 1000 par semaine. Dans un premier temps 20 échantillons de taille  $n = 6$  sont prélevés avec une fréquence de prélèvement d'un échantillon toutes les deux heures. Pour chaque échantillon le technicien calcule la moyenne ainsi que la variance des diamètres mesurés (en  $mm$ ). Ces résultats figurent dans le tableau ci-dessous :

Ech.	01	02	03	04	05	06	07	08
Moy.	135.1	134.8	135.2	135.4	135.0	134.9	135.1	134.8
Var.	0.05	0.17	0.12	0.07	0.18	0.11	0.09	0.09
Ech.	09	10	11	12	13	14	15	16
Moy.	135.0	134.9	135.1	135.3	134.7	135.3	135.4	135.6
Var.	0.15	0.14	0.23	0.17	0.24	0.24	0.20	0.25
Ech.	17	18	19	20				
Moy.	135.9	135.5	135.8	135.2				
Var.	0.18	0.26	0.20	0.15				

$$\hat{\mu} = \frac{1}{K} \sum_{i=1}^K \bar{x}_i = \frac{\bar{\bar{x}}}{C_4(n)}$$

**1** Construire la carte de contrôle de Shewhart d'étude initiale (phase I) pour la moyenne et l'écart-type. Détailler l'obtention des limites de contrôle.

**2** Peut-on dire que le processus de production est sous contrôle ?

**Exercice 3** *Efficacité d'une carte de contrôle.*

Considérons ici un procédé industriel qui, lorsqu'il est sous contrôle, suit une loi  $N(\mu_0, \sigma_0)$ . On suppose que la caractéristique de position étudiée est la moyenne et qu'une carte de contrôle de Shewhart aux valeurs standard (phase II) de type  $\bar{X}$  est utilisée avec des échantillons de taille  $n$ . Considérons alors un décentrage se produisant à un moment donné, c'est-à-dire un déréglage affectant la moyenne tel que la nouvelle valeur de la moyenne du procédé devient :

$$\mu = \mu_0 + \Delta. \quad \begin{matrix} \text{amplitude du} \\ \text{déréglage} \end{matrix}$$

On dira que  $\Delta$  est l'amplitude du déréglage.

**1** Déterminer alors la probabilité  $\beta$  de ne pas détecter le déréglage avec un seul échantillon.

**2** Quelle est alors la POM (Période Opérationnelle Moyenne) de la carte de contrôle ?

**3** Soit un processus qui sous contrôle vérifie :  $\mu_0 = 10$  et  $\sigma_0 = 1$ . Quelle est la probabilité de détecter un décentrage de  $\Delta = 0.5$  pour des échantillons de taille  $n = 4$  puis  $n = 6$  (déterminer les POM correspondantes).

Considérons maintenant la dispersion de ce procédé industriel et supposons qu'elle est contrôlée à l'aide d'une carte de Shewhart aux valeurs standard (phase II) de type  $S$  toujours utilisée avec des échantillons de taille  $n$ . Supposons qu'un déréglage survienne en affectant cette fois l'écart-type qui passe de la valeur  $\sigma_0$  à la valeur  $\sigma > \sigma_0$ .

**4** Déterminer alors la probabilité  $\beta$  de ne pas détecter le déréglage avec un seul échantillon.

**5** Soit un processus qui sous contrôle vérifie :  $\mu_0 = 10$  et  $\sigma_0 = 1$ . Quelle est la probabilité de détecter un déréglage tel que de  $\sigma = 1.1$  pour des échantillons de taille  $n = 4$  puis  $n = 10$  (déterminer les POM correspondantes).

**Exercice 4** *Carte de contrôle aux limites probabilistes.*

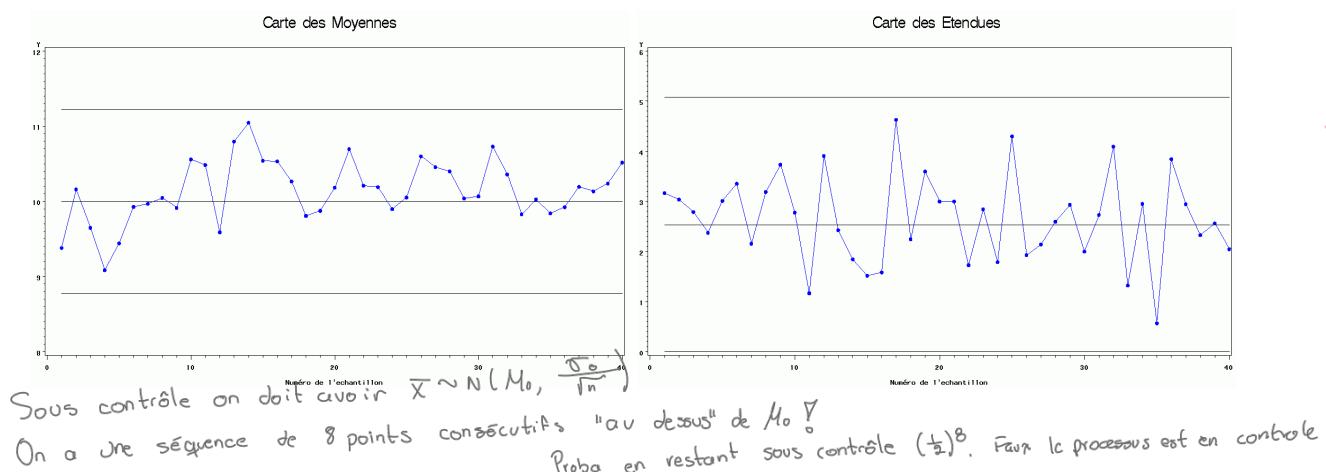
Une entreprise fabriquant des fibres optiques s'intéresse à l'indice de réfraction qui, pour être sous contrôle, devrait suivre une loi  $N(\mu, \sigma)$  avec  $\mu = 96.5$  et  $\sigma = 1.2$ . On considère alors 12 sous-groupes rationnels correspondant à des échantillons issus de 12 jours de production. Les relevés des 6 premiers jours sont obtenus à partir d'échantillons de taille  $n = 4$ . Afin d'obtenir des résultats plus fiables le service de qualité a prélevé pour les 6 jours suivants des échantillons de taille  $n = 6$ .

Ech.	1	2	3	4	5	6
n	4	4	4	4	4	4
Moy.	95.73	95.47	96.65	97.46	96.87	96.25
Ec-type	1.15	1.25	1.59	0.82	1.59	1.46
Ech.	7	8	9	10	11	12
n	6	6	6	6	6	6
Moy.	96.52	96.08	97.23	96.54	96.63	96.46
Ec-type	1.33	1.10	1.48	0.80	1.42	0.77

- 1** Déterminer les estimateurs sans biais classiques de  $\mu$  et  $\sigma^2$ . Est-il possible d'obtenir ici un estimateur sans biais de  $\sigma$ ? Il y aura 2 parce que on a différents tamaños de muestra
- 2** Construire la carte de contrôle aux valeurs standard (phase II) pour la moyenne et l'écart-type avec un risque d'erreur de première espèce fixé à 0.2%.
- 3** Peut-on dire que le processus de production est sous contrôle ?

### Exercice 5 Examen de cartes de contrôle.

Voici deux cartes de contrôle de type  $(\bar{X}, R)$  de phase II obtenues après une journée de production dans une entreprise. Peut-on dire que durant cette journée le procédé industriel est resté sous contrôle ? Est-il plausible de dire que le procédé ne semble pas avoir subi le moindre dérangement ?



### Exercice 6 Problème inverse.

Le fonctionnement d'un procédé est surveillé en temps réel à l'aide de deux cartes de contrôle de Shewhart : l'une pour la moyenne et l'autre pour la variabilité en utilisant l'écart-type. Les caractéristiques des deux cartes sont données ci-dessous :

$$\begin{cases} \text{Carte } \bar{X} : LC = 700, LSC = 713 \text{ et } LIC = 687. \\ \text{Carte } S : LC = 7.979, LSC = 18.08 \text{ et } LIC = 0. \end{cases}$$

- 1** Déterminer alors les valeurs standard  $\mu_0$  et  $\sigma_0$  utilisées pour construire ces cartes. Peut-on savoir quelle est la taille de chacun des échantillons prélevés ? D'après le valeur de la ligne centrale de la carte  $\bar{X}$  on a :  $\mu_0 = 700$

Les spécifications de la production sont de rester dans l'intervalle  $705 \pm 15$ .

**[2]** Déterminer la proportion de produits non-conformes puis la probabilité d'erreur de première espèce pour la moyenne.

La valeur moyenne de la caractéristique surveillée est en fait de 693 alors que son écart-type est de 12.

**[3]** Calculer la probabilité que cette différence de moyenne soit mise en évidence par l'un des échantillons prélevé. A partir de combien de prélèvements peut-on espérer détecter cette différence de moyenne ?

### Coefficients de correction

$n$	$d_2(n)$	$d_3(n)$	$c_4(n)$
2	1.128	0.853	0.798
3	1.693	0.888	0.886
4	2.059	0.880	0.921
5	2.326	0.864	0.940
6	2.534	0.848	0.951
7	2.704	0.833	0.959
8	2.847	0.820	0.965
9	2.970	0.808	0.969
10	3.078	0.797	0.973
11	3.173	0.787	0.975
12	3.258	0.778	0.978
13	3.336	0.770	0.979
14	3.407	0.763	0.981
15	3.472	0.756	0.982

## Exercice 1

$$\overline{x}_i \in \{ 350.94, 352.12, 353.44, 352.18, 340.26, 340 \} \quad i \in \{1, \dots, 6\}$$

$$d_2(5) = 2.326$$

$$\bar{x} = 348.1 = \hat{\mu}$$

$$\begin{matrix} R \\ \approx 19.8 \end{matrix}$$

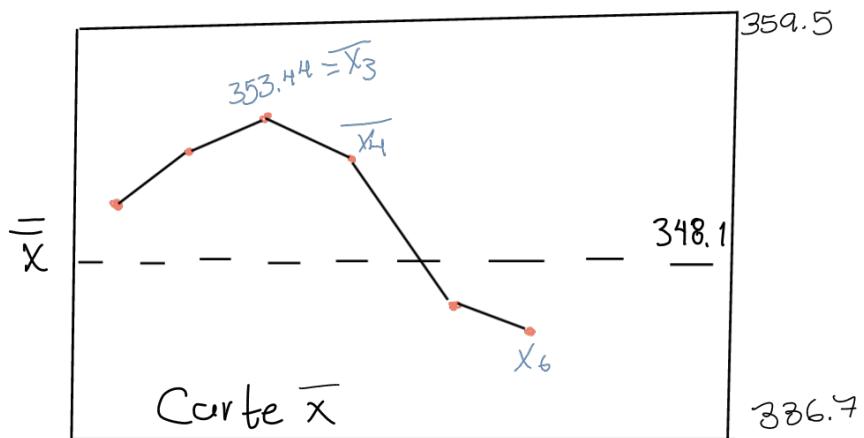
$$\text{Étendues}$$

$$R_1 = X_{(n)} - X_{(1)} = 360.4 - 341.5$$

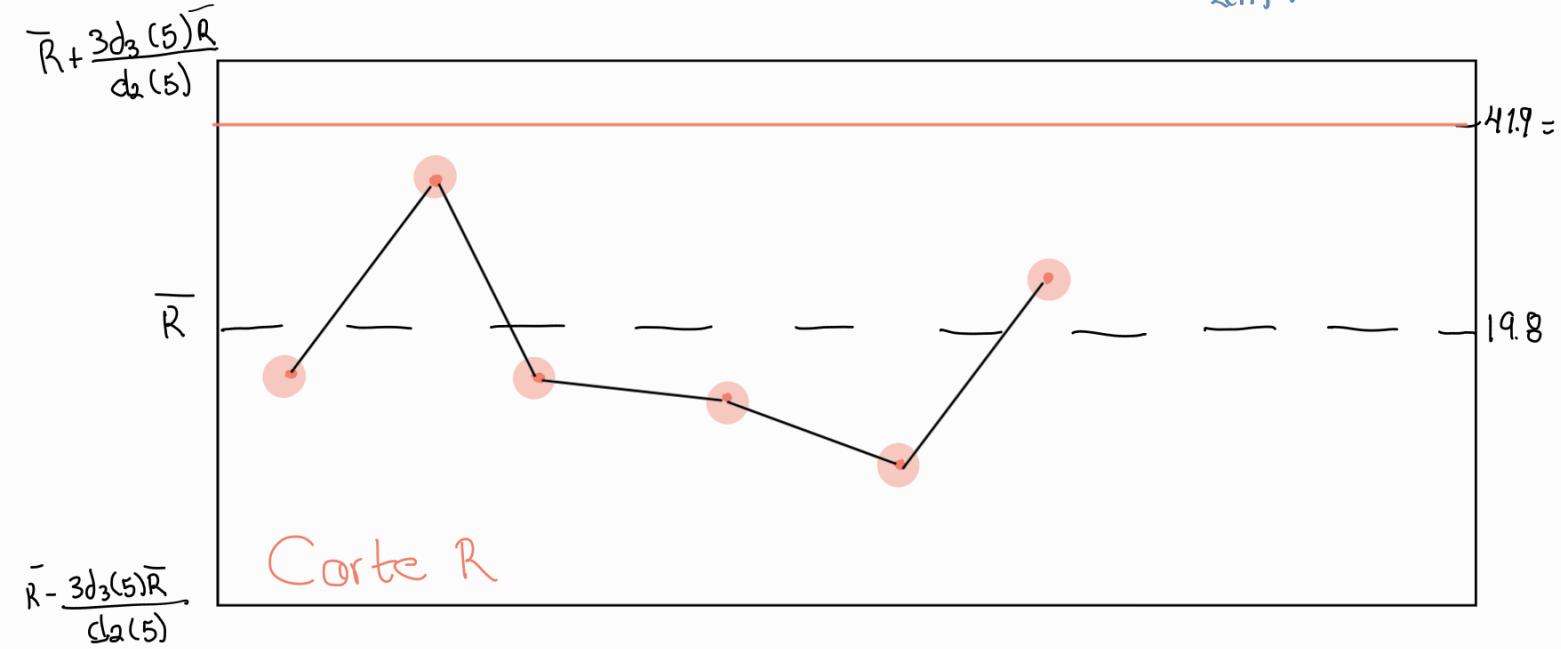
$$R_2 = 18.9 \quad R_3 = 31.1 \quad R_4 = 18.5 \quad R_5 = 18.4 \quad R_6 = 11.9 \quad R_7 = 20.1$$

$$\liminf = \bar{x} - \frac{3R}{d_2(5)\sqrt{5}} \quad \limsup = \bar{x} + \frac{3R}{d_2(5)\sqrt{5}} \quad \text{ou} \quad \bar{x} \pm \frac{3R}{d_2(5)\sqrt{5}} = 359.5, 336.7$$

$$\begin{aligned} \hat{\mu} &= \sum_{i=1}^6 x_i \\ \hat{\mu} &= 348.1 \\ \hat{\sigma} &= \frac{R}{d_2(5)} \approx 8.52 \end{aligned}$$



Pour la carte R  $M \pm 3\sigma = M \pm 3\left(\frac{S_{d_2(n)}}{d_2(n)}\right)\sigma = (LSC, LUC)$



$LIC \approx -2,3 < 0 \Rightarrow$  sans intérêt

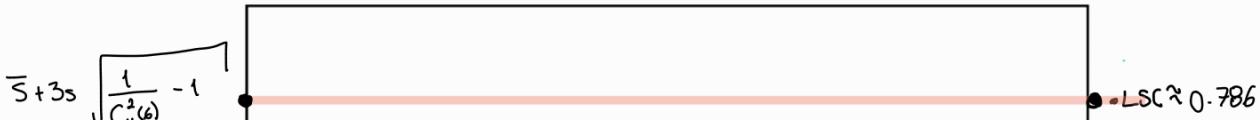
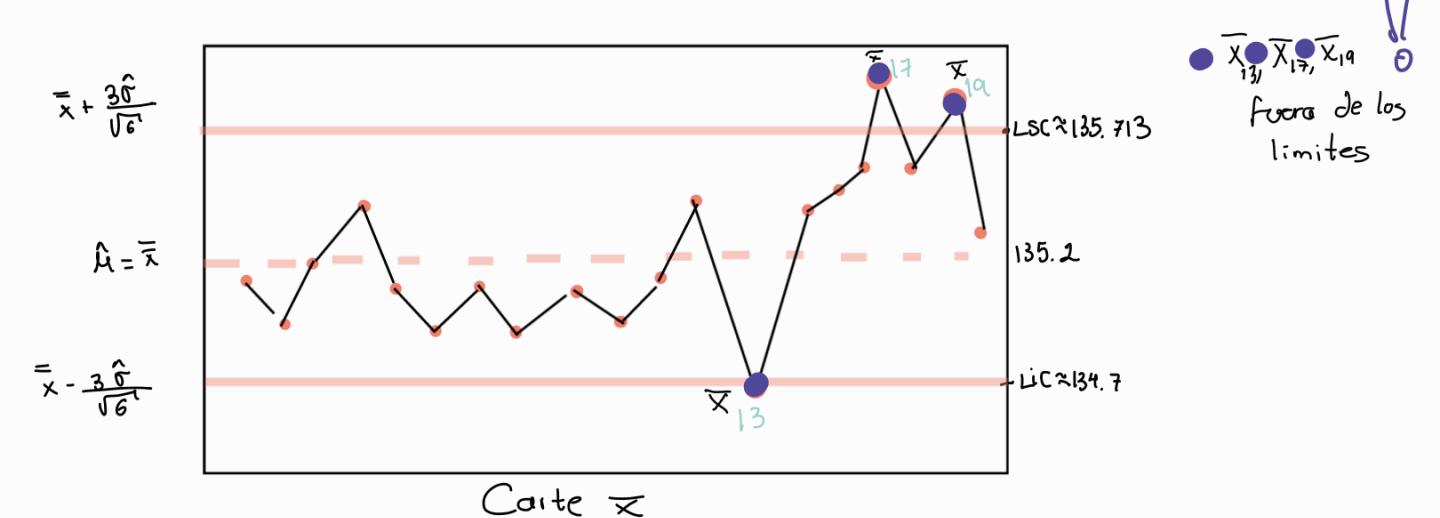
Une fois que des données nous donnent les cartes, si nous avons une résultat assez bien  $\Rightarrow$  nous continuons avec la phase 2 en fixant les valeurs qu'en phase 1 nous avions estimées

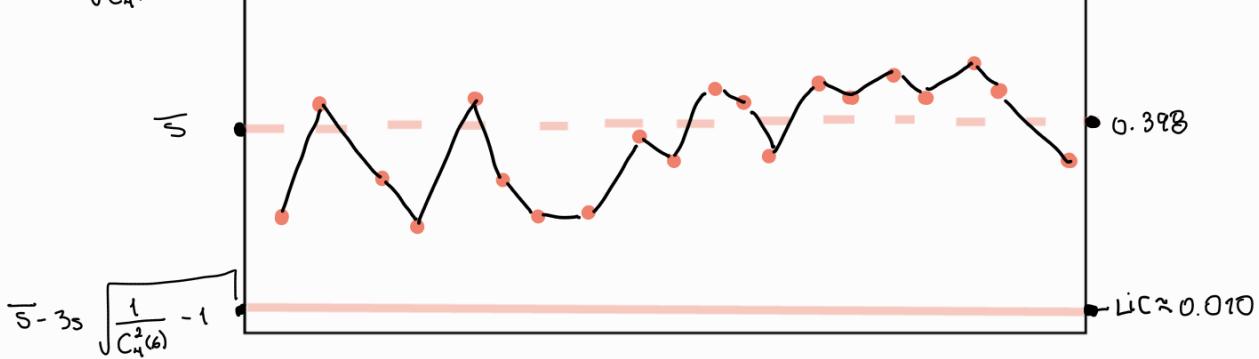
## Exercice 2

Estimateurs On a  $\hat{\mu} = \bar{x} = \frac{1}{20} \sum_{i=1}^{20} \bar{x}_i = 135.2$   $\hat{\sigma} = \sqrt{\frac{s}{C_4(6)}} = \frac{0.398}{0.951} \approx 0.418$

Carte  $\bar{x}$ : On a ici des SGR de taille  $n=6$  donc, chaque  $\bar{x}_i = (i=1, \dots, 20)$  est une réalisation sous-contrôle de la VAR,  $\bar{x}_i \sim N(\mu, \frac{\sigma^2}{6})$

Ici,  $M \approx \hat{\mu} = 135.2$  et  $\frac{\sigma}{\sqrt{6}} \approx \frac{\hat{\sigma}}{\sqrt{6}} \approx 0.171$





Carte S

### Exercice 3

1) On cherche  $\beta = P\{LiC < \bar{X} \text{ LSC}\}$

Avec  $\bar{X}$  calculé par rapport à  $\mu = M_0 + \Delta$  et  $LiC = M_0 - \frac{3\sigma_0}{\sqrt{n}}$  et  $LSC = M_0 + \frac{3\sigma_0}{\sqrt{n}}$ . On a déréglage  $X_i \sim N(\mu, \sigma_0)$  donc  $\bar{X} \sim N(\mu, \frac{\sigma_0^2}{n})$  avec  $M = M_0 + \Delta$

$$\text{On a } \beta = P\left[\frac{\mu_0 - \frac{3\sigma_0}{\sqrt{n}} - \mu}{\frac{\sigma_0}{\sqrt{n}}} \leq \frac{\bar{X} - \mu}{\frac{\sigma_0}{\sqrt{n}}} \leq \frac{\mu_0 + \frac{3\sigma_0}{\sqrt{n}} - \mu}{\frac{\sigma_0}{\sqrt{n}}}\right] \text{ Si } Y \sim N(0, 1) \Rightarrow \beta = P\left[-3 - \frac{4\sqrt{n}}{\sigma_0} \leq Y \leq 3 - \frac{4\sqrt{n}}{\sigma_0}\right]$$

$$\Rightarrow \beta = F\left(3 - \frac{4\sqrt{n}}{\sigma_0}\right) - F\left(-3 - \frac{4\sqrt{n}}{\sigma_0}\right) \text{ avec } F \text{ la fonction de répartition de la loi } N(0, 1)$$

2)  $POM = \frac{1}{1-\beta}$  (voir le cours) signe une ley geométrica de posiciones totales que su medio es?  $\uparrow$   $\uparrow$  cuántas veces para impedir error?  $\uparrow$   $\uparrow$   $POM \approx 44$

$$3) \begin{cases} M_0 = 10 \\ \sigma_0 = 1 \end{cases} \quad \begin{matrix} \text{M} = 10.5 \text{ grande} \\ \text{grande} \end{matrix} \quad \Delta = 0.5 \quad \text{Il vient} \quad \begin{cases} n=4 : \beta = F(2) - F(4) \approx F(2) \approx 0.9772 \\ n=6 : \beta \approx F(1.77) - F(4.22) \approx F(1.77) \approx 0.9616 \\ POM \approx 26 \end{cases} \quad \frac{1}{1-\beta}$$

4) Problème modèle ajustée on connaît  $\sigma^2$

Problème  $\sigma_0 \rightarrow \sigma > \sigma_0$ , On cherche  $\beta = P\{LiC < S^2 \text{ LSC}\}$  avec

$$\begin{cases} LiC = (C_4 - 3\sqrt{1-C_4^2})\sigma \\ LSC = (C_4 + 3\sqrt{1-C_4^2})\sigma \end{cases}$$

On connaît la loi de  $(n-1)\frac{S^2}{\sigma^2}$  qui est une loi

du  $\chi^2$  à  $(n-1)$  ddl. Donc:

$$\beta = P\{LiC^2 < S^2 < LSC^2\} = P\left[\frac{(n-1)LiC^2}{\sigma^2} < \frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)LSC^2}{\sigma^2}\right] \text{ avec } F_{n-1} \text{ fonction de répartition de la loi } \chi^2(n-1)$$

$$5) M_0 = 10 / \sigma_0 = 1 \rightarrow \sigma = 1.1$$

$$C_4(4) = 0.921 \Rightarrow C_4^2(4) = 0.8482$$

Taille de sondage:

$$\begin{aligned} ①_{n=4} \quad & \left\{ LSC = C_4(4) + 3\sqrt{1-C_4^2(4)} \right\} \approx 2.09 \quad \text{Donc } \beta \approx F_3(10.827) \approx 0.987 \\ & LiC = (C_4(4) - 3\sqrt{1-C_4^2(4)})\sigma \leftarrow \text{inutilisable} \end{aligned}$$

venfie

$$\begin{aligned} LSC &= \left( C_4(6) + 3 \sqrt{1 - C_4^2(6)} \right) \approx 1.665 \\ LIC &= \left( C_4(6) - 3 \sqrt{1 - C_4^2(6)} \right) \approx 0.281 \\ * C_4(10) &= \Rightarrow C_4^2(10) = \end{aligned}$$

Donc  $B = F_q(1.665) - F_q(0.281) \approx 0.985$   
 $POM = \frac{1}{1-B} = 6^q$

## Exercice 4)

1) Estimateur sans biais de  $M$ :  $\hat{M} = \bar{x} = \frac{\sum_{k=1}^{12} n_k \bar{x}_k}{\sum_{k=1}^{12} n_k}$  Promedio ponderado

$$\Rightarrow \hat{M} = 96.51$$

• Estimateur sans biais de  $\sigma^2$ :

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^{12} n_k s_k^2}{\sum_{k=1}^{12} n_k} \approx 1.558$$

avec  $n_k$  taille du sgn  $k$   
 $(n_k=4$  pour ceux du début  
 $n_k=6$  pour ceux de la fin)

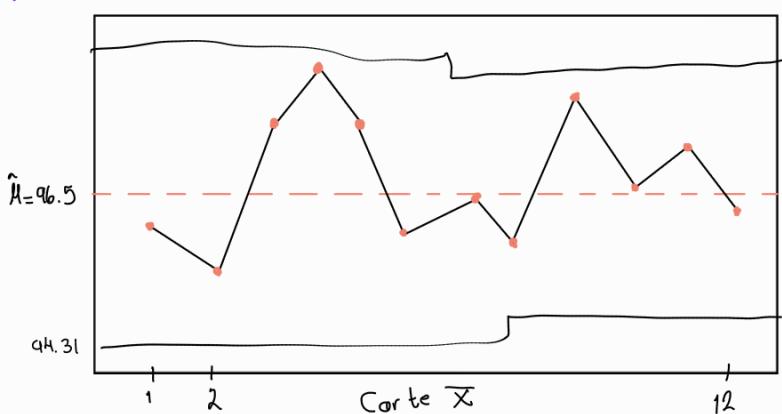
• On peut avoir un estimateur sans biais du  $\sigma$ ?

Attention: Comme estimateur sans biais de  $\sigma$  on peut prendre (cours)  $s/c_4$   
 On a donc ici deux estimateurs sans biais de  $\sigma$ , un avec  $C_4(4)$  et autre avec  $C_4(6)$ , mais  $C_4 = C_4(6)$

$$\hat{\sigma}_{deb} = \frac{\bar{s}_{deb}}{C_4(4)} = \frac{1.310}{0.921} \approx 1.422$$

$$\hat{\sigma}_{fin} = \frac{\bar{s}_{fin}}{C_4(6)} = \frac{1.15}{0.951} \approx 1.209$$

## 2) Carte X



Sous-contôle

$$x_{ij} \sim N(96.5, 1.2)$$

Donc  $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$  sous-contôle on veut que  
 $P[LIC < \bar{x} < LSC] = 0.998$ , de manière symétrique  
 on cherche donc LIC et LSC de manière à

$$P[\bar{x} > LSC] = 0.001 \text{ et } P[\bar{x} < LIC] = 0.001$$

$$P[\bar{x} > LSC] = 0.001 \Leftrightarrow P\left[\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{LSC - \mu}{\sigma/\sqrt{n}}\right] = 0.001$$

Si  $y \sim N(0, 1)$  on a  $P[y > 3.09] \approx 0.001$

$$\text{Conclusion: } \frac{LSC - \mu}{\sigma/\sqrt{n}} = 3.09 \Leftrightarrow LSC = \mu + 3.09 \frac{\sigma}{\sqrt{n}} \quad (\text{de même } LIC = \mu - 3.09 \frac{\sigma}{\sqrt{n}})$$

$$LSC = \begin{cases} 98.7 \text{ au début} \\ 98.01 \text{ à la fin} \end{cases}$$

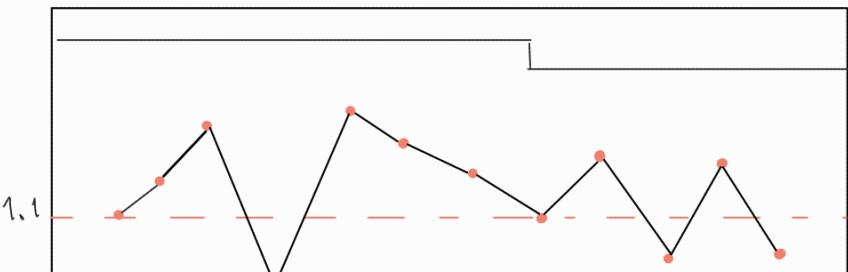
$$LIC = \begin{cases} 94.31 \text{ au début} \\ 94.99 \text{ à la fin} \end{cases}$$

## 2) Carte S

Ligne centrale!

$$C_4(4)\sigma = 1.105 \text{ au début}$$

$$C_4(6)\sigma = 1.145 \text{ à la fin}$$



Recherche de LSC (limites probabilistes sous forme symétrique):

$$P[S > LSC] = 0.001 \text{ avec } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Carte S

$$P[S > LSC] = 0.001 \Leftrightarrow P[S^2 > LSC^2] = 0.001$$

$$\Leftrightarrow P\left[\frac{(n-1)S^2}{\sigma^2} > (n-1) \frac{LSC^2}{\sigma^2}\right] = 0.001$$

S.  $\chi^2_{0.99, n}$  est le fractile de la loi du  $\chi^2$  à  $(n-1)$  ddl alors

$$(n-1) \frac{LSC^2}{\sigma^2} = \chi^2_{0.99, n-1} \Leftrightarrow LSC = \sigma \sqrt{\frac{\chi^2_{0.99, n-1}}{n-1}} \rightarrow LSC = \begin{cases} 2.79 \text{ pour } n=4 \\ 2.43 \text{ pour } n=6 \end{cases}$$

$$\text{de même } LC = \sigma \sqrt{\frac{\chi^2_{0.001, n-1}}{n-1}} \rightarrow LC = \begin{cases} 0.11 \text{ point } n=4 \\ 0.25 \text{ point } n=6 \end{cases}$$

## Exercice 6

1) D'après la valeur de la ligne centrale de la carte  $\bar{x}$  on a

$$\mu_0 = 700$$

A partir de la carte S il vient:

$$\begin{aligned} LC &\rightarrow C_4(n)\sigma_0 = 7.979 \Rightarrow \sigma_0 = \frac{7.979}{C_4(n)} \\ LSC &\rightarrow \begin{cases} C_4(n)\sigma_0 + 3\sigma_0 \sqrt{1-C_4^2(n)} = 18.08 \\ 23.937 \sqrt{1-C_4^2(n)} = 10.101 C_4(n) \end{cases} \end{aligned}$$

$$\begin{aligned} \text{et } 7.979 + 3\left(\frac{7.979}{C_4(n)}\right)\sqrt{1-C_4^2(n)} &= 18.08 \\ \Rightarrow 23.937 \sqrt{1-C_4^2(n)} &= 10.101 C_4(n) \quad \text{et } \sigma_0 = \frac{7.979}{0.94} \approx 8.663 \\ \Rightarrow 572.98 (1-C_4^2(n)) &= 102.03 C_4^2(n) \\ \Rightarrow 675 C_4^2(n) &= 572.98 \quad \Rightarrow C_4(n) \approx 0.92 \text{ donc } n=4 \end{aligned}$$

2) On cherche:  $p = P[\bar{x} < 690] + P[\bar{x} > 720]$  avec  $X \sim N(\mu_0, \sigma_0) = N(700, 8.663)$

Centrer de réduire (classique)  $p \approx 0.1355$

Erreur de première espèce:  $\alpha \approx 0.0026$  (règle des 3σ) puis cada una sigue una ley normal  $\Rightarrow$  la suma tmbn

3) Sous contrôle  $\mu_0 = 700$  et  $\sigma_0 = 8.663$

▷ Dérglage  $\mu_0 \rightarrow \mu = 693$  et  $\sigma_0 \rightarrow \sigma = 12$

La proba de que la gráfica detecte un problema, i.e. que encuentre fuera de sus límites

Probabilité de détecter le dérglage avec un sgn:

$$p = P[\bar{x} < 713] + P[\bar{x} > 687] \quad (\text{il s'agit de } \alpha - \beta)$$

$$p = P[\bar{x} < 713] + P[\bar{x} > 687]$$

$$\text{Ici: } X \sim N(693, 12) \text{ donc, (puisque } n=4) \quad \bar{X} \sim N(693, 6)$$

Donc (centrage et réduction)

$$p = P[\bar{x}^* > 3.33] + P[\bar{x}^* < -1] = 2 \cdot P[\bar{x}^* < 3.33] - P[\bar{x}^* < 1] \approx 0.159$$

$$\text{Donc } \beta \approx 0.841$$

¿Cuántos prélévements necesito para detectar esto? cuando me de cuenta se paran operaciones

¿Cuánto es el tiempo de producción?

Et donc, pour un  $\alpha = 0.05$ , le déréglage devrait être décelé en moyenne 6.28 échantillons.

$$PO M = \frac{1}{1-\beta} \approx 6.28$$

# Chapitre 4

## Cartes de contrôle aux attributs

**Objectif Général** : construire des cartes de contrôle pour des données qualitatives (de type produit accepté ou refusé) ou discrètes (de type nombre de défauts d'une production).

### 4.1 Cartes de contrôle de type $np$

On considère ici les cartes de contrôle utilisées pour le suivi du nombre de pièces non-conformes d'une population.

La situation étudiée ici est celle où une production est évaluée, pour chaque sous-groupe rationnel, à partir du nombre de produits non-conformes (défectueux) observés. Désignons par  $p$  la probabilité de fabrication d'un produit non-conforme et par  $n$  le nombre d'observations de chacun des sous-groupes rationnels. L'hypothèse que tous les sous-groupes rationnels soient de même taille est obligatoire ici afin que les comparaisons des nombres de produits défectueux aient un sens.

#### — Analyse des observations d'un sous-groupe rationnel

La carte  $np$  consiste en la représentation des diverses valeurs  $c_i$  ( $1 \leq i \leq k$ ) donnant le nombre de pièces non-conformes observées dans le sous-groupe rationnel  $i$ . Il est clair que, sous contrôle, ces quantités sont des réalisations des variables aléatoires  $C_i$  de loi binomiale  $B(n, p)$ . On en déduit alors que (toujours sous l'hypothèse que le processus est sous contrôle) :

$$E(C_i) = np \text{ et } Var(C_i) = np(1 - p)$$

#### — Estimation de $p$

Dans le cadre d'une carte de contrôle de phase I la probabilité  $p$  est inconnue. Cette probabilité de fabrication d'un produit non-conforme peut alors classiquement être estimée par la proportion de produits non-conformes sur la totalité des observations soit :

$$\hat{p} = \frac{\sum_{i=1}^k C_i}{nk}$$

Remarquons que  $\hat{p}$  est un estimateur sans biais de  $p$  mais il s'agit aussi de l'estimateur du maximum de vraisemblance de cette même quantité.

#### — Construction de la carte de contrôle

Le principe de construction le plus courant est le suivant :

- 1) positionnement de la ligne centrale à la valeur théorique moyenne suivante :

$$E(C_i) = np \simeq n\hat{p} = \frac{\sum_{i=1}^k c_i}{nk} = \bar{c}$$

2) positionnement des limites de contrôle selon la règle des  $3\sigma$  à partir de :  $\sigma^2 = np(1-p)$

$$\Rightarrow \sigma(C_i) = \sqrt{np(1-p)} \simeq \sqrt{n\hat{p}(1-\hat{p})} \text{ avec } \hat{p} = \left( \sum_{i=1}^k c_i \right) / nk$$

Remarquons que l'utilisation de la règle des  $3\sigma$  n'est pas absurde d'un point de vue théorique puisque le théorème central-limite assure qu'une loi binomiale  $B(n, p)$  peut être approchée par une loi normale  $N(np, \sqrt{np(1-p)})$  (ceci étant vrai, bien évidemment, si la taille  $n$  des sous-groupes rationnels est « assez grande » ...).

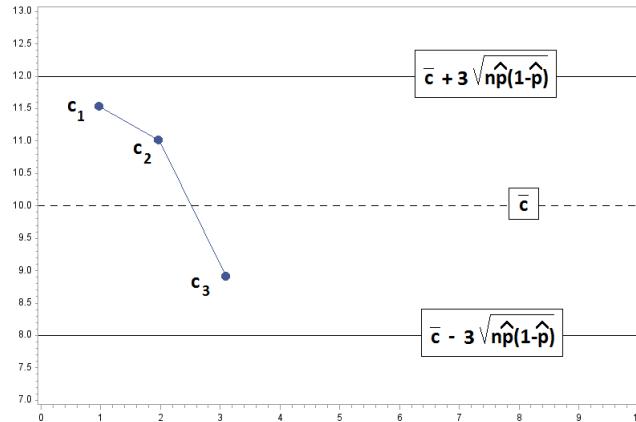


Figure 13. Carte de contrôle  $np$ .

**Remarque.** Dans le cas d'une carte de contrôle de phase II la partie « estimation » est inutile et les limite inférieure de contrôle, ligne centrale et limite supérieure de contrôle sont respectivement données par (avec  $p_0$  valeur de référence connue pour la probabilité de fabrication d'un produit non-conforme) :

$$np_0 - 3\sqrt{np_0(1-p_0)} \quad / \quad np_0 \quad / \quad np_0 + 3\sqrt{np_0(1-p_0)}$$

**Remarque.** Lorsque les sous-groupes rationnels ne sont pas de taille suffisante pour réaliser une approximation correcte via la loi normale il est toujours possible de construire ces cartes avec des limites de contrôle probabilistes issues de la loi binomiale.

## 4.2 Cartes de contrôle de type $p$

On considère ici les cartes de contrôle utilisées pour le suivi de la proportion de pièces non-conformes d'une population.

La situation étudiée ici est identique à celle de la carte  $np$  sauf que cette fois on suppose les sous-groupes rationnels de tailles éventuellement distinctes  $n_1, n_2, \dots, n_k$ . Comparer alors les différentes proportions a bien un sens. Désignons toujours par  $p$  la probabilité de fabrication d'un produit non-conforme

### → Analyse des observations d'un sous-groupe rationnel

La carte  $p$  consiste en la représentation des diverses valeurs  $p_i$  ( $1 \leq i \leq k$ ) donnant la proportion de pièces non-conformes observées dans le sous-groupe rationnel  $i$ . Nous avons déjà vu que, sous contrôle, le nombre  $c_i$  de pièces non-conformes du sous-groupe rationnel  $i$  est une réalisation de la variable aléatoire  $C_i$  de loi binomiale  $B(n_i, p)$ . On en déduit alors (toujours sous l'hypothèse que le processus est sous contrôle) que la proportion observée  $p_i = c_i/n_i$  est une réalisation de la variable aléatoire réelle  $P_i = C_i/n_i$  et donc :

$$E(P_i) = \frac{E(C_i)}{n_i} = p \text{ et } Var(P_i) = \frac{Var(C_i)}{n_i^2} = \frac{p(1-p)}{n_i}$$

### → Estimation de $p$

Dans le cadre d'une carte de contrôle de phase I la probabilité  $p$  est inconnue. Cette probabilité de fabrication d'un produit non-conforme peut alors classiquement être estimée par la proportion de produits non-conformes sur la totalité des observations soit :

$$\hat{p} = \frac{\sum_{i=1}^k C_i}{\sum_{i=1}^k n_i}$$

On vérifie encore que  $\hat{p}$  est un estimateur sans biais de  $p$  et il s'agit aussi de l'estimateur du maximum de vraisemblance de cette même quantité.

### → Construction de la carte de contrôle

Le principe de construction le plus courant est le suivant :

- 1) positionnement de la ligne centrale à la valeur théorique moyenne suivante :

$$E(P_i) = p \simeq \hat{p} = \frac{\sum_{i=1}^k c_i}{\sum_{i=1}^k n_i}$$

- 2) positionnement des limites de contrôle selon la règle des  $3\sigma$  à partir de :

$$\sigma(P_i) = \sqrt{\frac{p(1-p)}{n_i}} \simeq \sqrt{\frac{\hat{p}(1-\hat{p})}{n_i}} \text{ avec } \hat{p} = \left( \sum_{i=1}^k c_i \right) / \left( \sum_{i=1}^k n_i \right)$$

**Remarque.** Dans le cas d'une carte de contrôle de phase II la partie « estimation » est inutile et les limite inférieure de contrôle, ligne centrale et limite supérieure de contrôle sont respectivement données par (avec  $p_0$  valeur de référence connue pour la probabilité de fabrication d'un produit non-conforme) :

$$p_0 - 3\sqrt{\frac{p_0(1-p_0)}{n_i}} \quad / \quad p_0 \quad / \quad p_0 + 3\sqrt{\frac{p_0(1-p_0)}{n_i}}$$

**Remarque.** L'obtention des limites de contrôle est ici plus complexe puisqu'elles varient en fonction de la taille du sous-groupe rationnel considéré. Remarquons que si ceci est jugé trop contraignant alors les deux alternatives suivantes sont couramment utilisées en pratique : *restrictive*

1) représenter uniquement les limites de contrôle « larges » et « étroites » obtenues avec respectivement le plus petit et le plus grand échantillon,

2) représenter seulement deux limites de contrôle construites à partir de la taille moyenne  $\bar{n}$  des sous-groupes rationnels. En pratique cette technique est souvent utilisée lorsque les tailles des sous-groupes rationnels varient « peu » dans le sens suivant :

$$\forall i = 1, \dots, k, \quad 0.75\bar{n} \leq n_i \leq 1.25\bar{n}$$

On parle alors de taille de sous-groupes rationnels pseudo-constante.

Cómo es la proporción, cambia con el tamaño d'enchantillon que se tenga en cada sgp

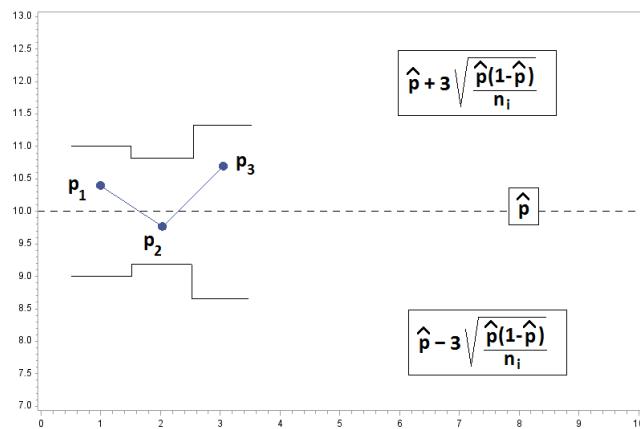


Figure 14. Carte de contrôle  $p$ .

### 4.3 Cartes de contrôle de type $c$

On considère ici les cartes de contrôle utilisées pour le suivi du **nombre de non-conformité** (**défauts**) d'une population.

L'idée est ici d'évaluer la qualité d'une production à partir, pour chaque sous-groupe rationnel étudié, du nombre total de défauts observés (par exemple le nombre de soudures défectueuses d'un lot de circuits imprimés, etc...). Evidemment les défaut sont supposés ici de « même importance » (donc cumulables) et les sous-groupes rationnels de même taille. *hay más de un defecto por posible errors*

#### → Analyse des observations d'un sous-groupe rationnel

La carte  $c$  consiste en la représentation des diverses valeurs  $c_i$  ( $1 \leq i \leq k$ ) donnant **le nombre total de défauts observés** dans le sous-groupe rationnel  $i$ . D'un point de vue théorique chaque  $c_i$  est alors une réalisation de la variable aléatoire réelle  $C_i$  de loi binomiale  $B(n, p)$  avec  $n$  très grand (car un très grand nombre de défauts peuvent survenir) et  $p$  très petit (car normalement chaque défaut a une très faible probabilité d'apparition). C'est pourquoi il est alors classique de supposer (lorsque tous les sous-groupes rationnels sont de même taille) que toutes les variables aléatoires  $C_i$  suivent une **loi de Poisson de type  $\mathcal{P}(\lambda)$** . On en déduit alors que, lorsque le processus est sous contrôle, il vient :

$$E(C_i) = \lambda \text{ et } Var(C_i) = \lambda$$

#### → Estimation de $\lambda$

Dans le cadre d'une carte de contrôle de phase I la quantité  $\lambda$  est inconnue. Cette quantité étant égale à l'espérance de la loi de Poisson considérée elle peut donc être classiquement estimée via la moyenne empirique soit :

$$\hat{\lambda} = \bar{C} = \frac{1}{k} \sum_{i=1}^k C_i$$

#### → Construction de la carte de contrôle

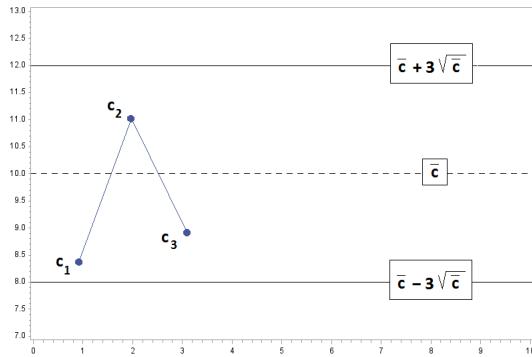
Le principe de construction le plus courant est le suivant :

- 1) positionnement de la ligne centrale à la valeur théorique moyenne suivante :

$$E(C_i) = \lambda \simeq \hat{\lambda} = \bar{c}$$

- 2) positionnement des limites de contrôle selon la règle des  $3\sigma$  à partir de :

$$\sigma(C_i) = \sqrt{\lambda} \simeq \sqrt{\hat{\lambda}} = \sqrt{\bar{c}}$$



**Figure 15.** Carte de contrôle  $c$ .

**Remarque.** Dans le cas d'une carte de contrôle de phase II la partie « estimation » est inutile et les limite inférieure de contrôle, ligne centrale et limite supérieure de contrôle sont respectivement données par (avec  $\lambda_0$  valeur de référence pour la loi de Poisson suivie sous contrôle) :

$$\lambda_0 - 3\sqrt{\lambda_0} \quad / \quad \lambda_0 \quad / \quad \lambda_0 + 3\sqrt{\lambda_0}$$

**Remarque.** En pratique ce type de carte de contrôle est souvent jugée peu satisfaisante lorsque  $\lambda < 5$  (la limite de contrôle inférieure est alors négative). Ceci se justifie théoriquement par le fait que pour de petites valeurs de  $\lambda$  la loi de Poisson ne peut être approchée par une loi normale (elle est alors fortement asymétrique) et donc la règle des  $3\sigma$  ne s'applique pas. Il est alors recommandé dans ce cas d'utiliser des limites de contrôle probabilistes.

**Maîtrise Statistique des Procédés**

Master MSiD 2ème année  
Travaux dirigés - Série 4

**Exercice 1** *Carte de contrôle aux attributs.*

Une entreprise de matériel électronique expérimente une nouvelle chaîne de production pour un nouveau type de jauge à essence. Un certain nombre de ces jauge sont extraites chaque jour au hasard dans la production et testées en laboratoire afin de déterminer si elles vérifient ou non les normes imposées par le constructeur automobile. Voici alors les résultats relatifs aux dix derniers jours (avec  $j$  repérant le jour,  $n_j$  la taille de l'échantillon testé le jour  $j$  et enfin  $c_j$  le nombre de jauge jugées non-conformes) :

$j$	1	2	3	4	5	6	7	8	9	10	$= 1000$
$n_j$	80	110	90	75	130	120	70	125	105	95	
$c_j$	4	7	5	8	6	6	4	5	8	9	

- 1** Expliquer quel type de carte de contrôle (de phase I) va être adapté à l'exploitation de ces données. Déterminer alors la forme générale des limites de contrôle ainsi que du ou des divers paramètres à estimer ici.
- 2** Construire alors la carte de contrôle proposée à la question précédente.
- 3** Le procédé est-il sous-contrôle ? Dans l'affirmative à partir de quel nombre de jauge défectueuses par lot de 100 doit-on donner l'alerte ? *Si, pues ningún valor rebasa el límite superior*

**Exercice 2** *Carte de contrôle aux attributs.*

Un constructeur automobile s'intéresse à la qualité de la peinture des capots d'un modèle de sa gamme. Pour cela divers capots sont choisis au sein de la production journalière et pour chacun d'eux le nombre de bulles d'air présentes à leur surface est mesurée. Ces données sont alors résumées dans le tableau ci-dessous,  $i$  désignant le numéro du capot testé et  $n_i$  le nombre de bulles d'air trouvées :

$i$	01	02	03	04	05	06	07	08	09	10
$n_i$	8	9	5	6	10	11	6	7	7	8
$i$	11	12	13	14	15	16	17	18	19	20
$n_i$	6	8	6	10	6	4	6	7	1	8

- 1** Justifier que le suivi de la qualité peut alors être effectué à l'aide d'une carte de contrôle de type  $c$ . Déterminer la forme générale des limites de contrôle usuelles. Justifier que ces limites

posent des problèmes de mise en oeuvre pratique lorsque le paramètre de la loi de Poisson (donné ou estimé à partir de l'échantillon) est faible.

**2** Construire alors la carte de contrôle de phase II pour ces observations sachant que pour être sous contrôle le nombre de bulles d'air par capot doit suivre une loi de Poisson de moyenne égale à 9. On utilisera ici des limites probabilistes associées à un risque d'erreur de première espèce ne dépassant pas 0.5%.

### **3** Le procédé est-il sous-contrôle ?

# Exercices

## Exercice 1)

① les signes étant de tailles différentes, on va ici considérer

Pour chaque signe  $i$  ( $1 \leq i \leq 10$ ) on va représenter la proportion  $p_i = \frac{c_i}{n_i}$  avec  $c_i$

Pour chaque signe ( $i = 1, \dots, n$ )  
 nombre de produits défectueux  
 Chaque  $p_i$  est donc une réalisation de  $p_i = \frac{c_i}{n_i}$  avec  $C_i \sim B(n_i, p)$   
 où  $p$  est la probabilité qu'un produit soit défectueux

Comme  $E(C_i) = n_i p$  et  $\text{Var}(C_i) = n_i p(1-p)$  on en déduit que  
 $E(P_i) = p$  et  $\text{Var}(P_i) = \frac{p(1-p)}{n_i}$

Donc par application de la règle des 3σ on va construire une carte de contrôle telle que ligne centrale  $p$  / limites de contrôle:  $p \pm 3\sqrt{\frac{p(1-p)}{n_i}}$

1º Calculamos  $\hat{p}_i$ ,  $\hat{p}_i = \frac{\text{fallo}}{\text{total}}$

i	1	2	3	4	5	6	7	8	9	10
p <sub>i</sub>	0.05	0.064	0.056	0.107	0.046	0.05	0.057	0.040	0.076	0.095

$$2^{\circ} \text{ Calculamos } \hat{p}$$

Donde  $\hat{p} = \frac{\sum_{i=1}^{10} C_i}{\sum_{i=1}^{10} n_i} = \frac{62}{1000} = 0.062$

3º Obtenemos los límites de control

$$\hat{p} + \frac{3}{n_i} \sqrt{\hat{p}(1-\hat{p})}$$

4<sup>e</sup> Gráficosmos (3)

Le procédé semble bien être sous contrôle ici. Pour un échantillon de taille  $n_i = 100$  on va se retrouver hors contrôle si

$C_i \geq 14$

## exercice 2

① Le nombre de bulles d'air peut être modélisé par une loi  $\mathcal{B}(n, p)$  avec  $n$  grand et  $p$  petit soit par  $C_i \sim \text{Poisson}(\lambda)$

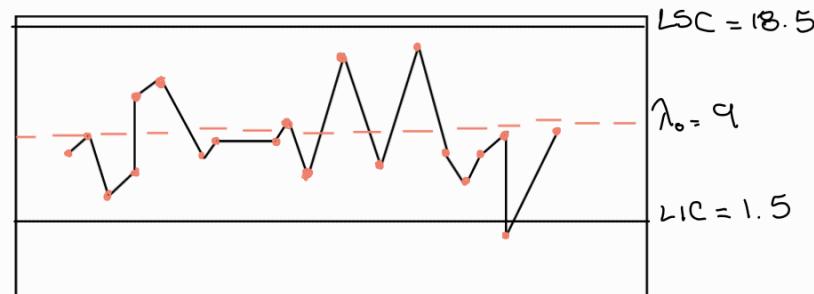
On a donc  $E(C_i) = \lambda$  et  $\text{Var}(C_i) = \lambda$

Donc, d'après la règle de 3σ on peut construire une carte de contrôle telle que ligne centrale  $\lambda$  limites de contrôle  $\lambda \pm 3\sqrt{\lambda}$

Remarquons que LiC est sonst intérêt lorsqu'elle est négative soit

$$\lambda - 3\sqrt{\lambda} < 0 \Rightarrow 0 < \lambda < 9$$

Si  $\lambda$  es pequeño no podemos decir que se acerca a una normal  $\Rightarrow$  habrá problemas



$$0.5\% \\ 0.005$$

Construction des limites de contrôles probabilistes (et symétriques) avec  $\alpha$  (maximal) de 0.5%.

• Pour LiC

Considérons, sous contrôle,  $C \sim P(\lambda)$  donc  $P[C=k] = e^{-\lambda} \frac{\lambda^k}{k!}, k \in \mathbb{N}$

Alors  $P[C=0] \approx 0.0001$        $P[C=1] \approx 0.0011$        $P[C=2] \approx 0.0050$

probabilité du valeur le plus petit       $P[C < 1.5] \approx 0.0012$  que queda por debajo de 0.5%  
 $0.0012 < 0.005 \Rightarrow LiC = 1.5$

• Pour LSC

On trouve  $P[C > 18] \approx 0.0024$

Le pioude n'est pas sous contrôle. El elemento  $1^{\text{a}}$  ver que pasa  
0.025 & se x oror?

Pour LSC

On cherche  $P[C > LSC] \leq 0.005$

On écrit  $\text{ppois}(19, \lambda=9, \text{lower.tail} = \text{FALSE}) = 0.0010$

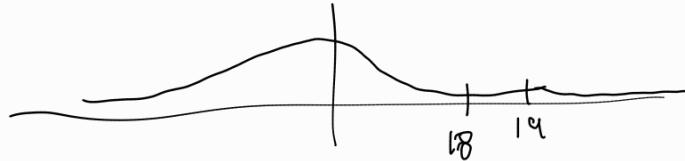
$$P[C > 19]$$

↑  
lower.tail

Análogamente podemos escribir

$$1 - \text{ppois}(19, \lambda=9, \text{lower.tail} = \text{TRUE})$$

sin embargo nos interesa estar lo más cerca posible de 19



Veamos que pasa con 18 y 17

$$\text{ppois}(17, \lambda=9, \text{lower.tail} = \text{FALSE}) = 0.005319 = P[C > 17]$$

ya es más del 0.5%.

$\hookrightarrow$  veamos 18

$$\Rightarrow \text{ppois}(18, \lambda=9, \text{lower.tail} = \text{F}) = 0.0024 = P[C > 18]$$

$$P[C > 18] < 0.005$$

$$\Rightarrow LSC = 18.5$$

# Chapitre 5

## Cartes de contrôle pour petits dérégagements

### 5.1 Introduction

Les cartes de contrôle utilisées jusqu'à présent ont en commun de considérer les sous-groupes rationnels de manière indépendante les uns des autres. Il en résulte qu'elles sont en général peu efficaces afin de détecter de petits déréglements. L'idée est alors de **considérer** non pas seulement le **sous-groupe rationnel  $i$**  mais aussi **une partie du processus passé**. Ceci permettra, toujours dans le cas de petits déréglements, de « **cumuler** » ces **petites quantités** qui seront ainsi détectées de manière beaucoup plus efficace.

Toutes les cartes présentées ici vont se focaliser sur le contrôle de la moyenne du processus (mais le principe est applicable en théorie à tout autre paramètre). Elles seront de plus construites ici en phase II, par soucis de simplification, à partir d'un processus issu sous contrôle d'une loi normale  $N(\mu_0, \sigma_0)$  (pour des situations relevant de la phase I il suffira de remplacer les valeurs de référence sous contrôle par des estimateurs classiques).

### 5.2 Cartes de contrôle MA

Il s'agit des cartes de contrôle de type « Moving Average » (donc cartes de contrôle moyennes mobiles). Considérons par la suite la situation classique où  $k$  sous-groupes rationnels de même taille  $n$  sont observés.

#### 5.2.1 Définitions

Nous savons qu'une carte de type  $\bar{X}$  est basée sur la représentation des diverses moyennes par sous-groupe rationnel  $\bar{x}_1, \dots, \bar{x}_k$ . L'idée des cartes de contrôle de type  $MA$  est de simplement remplacer ces moyennes par des moyennes mobiles d'ordre  $h$ . Donc :

##### Définition 10

Une **carte de contrôle de type MA**, construite avec des moyennes mobiles d'ordre  $h$ , consiste en la représentation des valeurs  $m_1, \dots, m_k$  réalisations des variables aléatoires  $M_1, \dots, M_k$  définies par :

$$M_i = \frac{\bar{X}_i + \bar{X}_{i-1} + \dots + \bar{X}_1}{i} \quad \text{si } i < h$$

$$M_i = \frac{\bar{X}_i + \bar{X}_{i-1} + \dots + \bar{X}_{i-h+1}}{h} \quad \text{si } i \geq h$$

\* Dans le cas où  $h=1$  on a  $\bar{X}$  régulier

moyennes mobiles incomplètes

moyennes mobiles complètes



On parlera de moyennes mobiles incomplètes pour les  $(h - 1)$  premiers relevés et de moyennes mobiles complètes ensuite.

**Remarque.** Il est clair que dans la situation où  $h = 1$  on retombe alors sur la carte classique de type  $\bar{X}$ . Lorsque  $h \geq k$  alors la carte de contrôle utilisée tient compte de l'intégralité du passé quel que soit le sous-groupe rationnel considéré. *Cuando  $h \geq k \Leftrightarrow$  se tendría en cuenta la historia para todos los  $\bar{x}_j$*

### 5.2.2 Constructions

On utilise ici la méthode classique de construction avec la ligne centrale positionnée à la valeur moyenne théorique et les limites de contrôle découlant de la règle des  $3\sigma$ . Il en découle que :

- 1) positionnement de la ligne centrale à la valeur théorique  $E(M_i)$  avec donc (valeur identique pour le cas des moyennes mobiles incomplètes) :

$$E(M_i) = E\left(\frac{\bar{X}_i + \bar{X}_{i-1} + \dots + \bar{X}_{i-h}}{h}\right) = \frac{1}{h} \left( \sum_{j=0}^{h-1} E(\bar{X}_{i-j}) \right) = \frac{h\mu_0}{h} = \mu_0$$

puisque si le processus est sous contrôle chaque observation est une réalisation de la loi normale  $N(\mu_0, \sigma_0)$  et donc  $E(\bar{X}_{i-j}) = \mu_0$ .

- 2) positionnement des limites de contrôle selon la règle des  $3\sigma$ . Il faut alors distinguer le cas des moyennes mobiles complètes ou incomplètes puisqu'il vient (d'après l'indépendance des observations) :

$$\text{si } M_i = \frac{\bar{X}_i + \bar{X}_{i-1} + \dots + \bar{X}_1}{i} \text{ alors } Var(M_i) = \frac{1}{i^2} \left( \sum_{j=0}^{i-1} Var(\bar{X}_{i-j}) \right) = \frac{i\sigma_0^2}{i^2 n} = \frac{\sigma_0^2}{ni}$$

$$\text{si } M_i = \frac{\bar{X}_i + \bar{X}_{i-1} + \dots + \bar{X}_{i-h}}{h} \text{ alors } Var(M_i) = \frac{1}{h^2} \left( \sum_{j=0}^{h-1} Var(\bar{X}_{i-j}) \right) = \frac{h\sigma_0^2}{h^2 n} = \frac{\sigma_0^2}{nh}$$

puisque si le processus est sous contrôle chaque observation est une réalisation de la loi normale  $N(\mu_0, \sigma_0)$  et donc  $Var(\bar{X}_{i-j}) = \sigma_0^2/n$ .

D'où la proposition suivante :

#### Proposition 11

Une **carte de contrôle de type MA**, utilisée avec des moyennes mobiles d'ordre  $h$ , est déterminée par :

- 1) la ligne centrale à la valeur  $\mu_0$ ,
- 2) les limites de contrôle de valeurs (pour le sous-groupe rationnel  $i = 1, \dots, k$ ) :

$$LIC = \mu_0 - 3 \frac{\sigma_0}{\sqrt{ni}} \quad \text{et} \quad LSC = \mu_0 + 3 \frac{\sigma_0}{\sqrt{ni}} \quad \text{si } i < h$$

$$LIC = \mu_0 - 3 \frac{\sigma_0}{\sqrt{hi}} \quad \text{et} \quad LSC = \mu_0 + 3 \frac{\sigma_0}{\sqrt{nh}} \quad \text{si } i \geq h$$

Ceci conduit graphiquement à des limites de contrôle « en créneaux » comme sur la carte suivante (construite ici avec des moyennes mobiles d'ordre  $h = 3$ ).

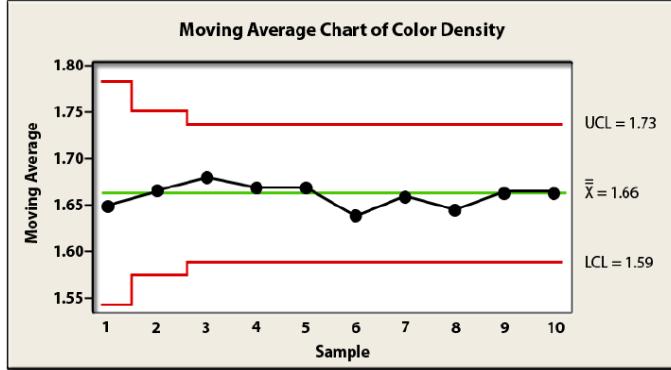


Figure 16. Carte de contrôle MA.

### 5.3 Cartes de contrôle EMWA

Il s'agit des cartes de contrôle de type « Exponentially Moving Weighted Average » (donc cartes de contrôle moyennes mobiles avec pondération exponentielle). Considérons par la suite la situation classique où  $k$  sous-groupes rationnels de même taille  $n$  sont observés.

L'idée de ces cartes de contrôle est de tenir compte de tout le passé du processus mais via des pondérations d'autant plus faibles que l'observation est « ancienne ».

ou fut et à m

#### 5.3.1 Définitions

Nous savons qu'une carte de type  $\bar{X}$  est basée sur la représentation des diverses moyennes par sous-groupe rationnel  $\bar{x}_1, \dots, \bar{x}_k$ . Nous allons maintenant tenir compte de l'intégralité du « passé » du processus en représentant les valeurs  $z_1, \dots, z_k$  réalisations des variables aléatoires  $Z_1, \dots, Z_k$  définies par la relation de récurrence suivante :

exponentially moving weighted average

##### Définition 12

Une **carte de contrôle de type EMWA** consiste en la représentation des valeurs  $z_1, \dots, z_k$  réalisations des variables aléatoires  $Z_1, \dots, Z_k$  définies par la relation de récurrence :

$$\forall i = 1, \dots, k, Z_i = \omega \bar{X}_i + (1 - \omega) Z_{i-1} \quad \text{où} \quad Z_0 = \mu_0$$

et  $0 < \omega \leq 1$  pondération donnant un poids de  $(1 - \omega)$  aux informations passées et un poids de  $\omega$  aux informations issues du sous-groupe rationnel  $i$ .

**Remarque.** En pratique il est courant de prendre de « petites » valeurs pour le coefficient  $\omega$ , par exemple  $\omega = 0.05, 0.10, 0.15$ . Il est clair que dans la situation où  $\omega = 1$  on retombe alors sur la carte classique de type  $\bar{X}$ .  
Lorsque  $\omega \approx 0.8$

Remarquons que  $Z_i$  est une statistique qui tient compte de la totalité du passé du processus puisqu'il vient explicitement :

$$\begin{aligned} \forall i = 1, \dots, k, \quad Z_i &= \omega \bar{X}_i + (1 - \omega) Z_{i-1} \\ Z_{i-1} &= \omega (\bar{X}_i + (1 - \omega) \bar{X}_{i-1}) + (1 - \omega)^2 Z_{i-2} \\ Z_{i-2} &= \omega (\bar{X}_i + (1 - \omega) \bar{X}_{i-1} + (1 - \omega)^2 \bar{X}_{i-2}) + (1 - \omega)^3 Z_{i-3} \dots \end{aligned}$$

D'où la relation suivante découlant de cette écriture « en cascade » :

$$\boxed{\forall i = 1, \dots, k, Z_i = \omega \sum_{j=0}^{i-1} (1 - \omega)^j \bar{X}_{i-j} + (1 - \omega)^i Z_0}$$

Il sera plus fiable de descendre le + possible en el posado

On constate alors que  $Z_i$  tient compte de toutes les valeurs  $\bar{X}_i, \bar{X}_{i-1}, \dots, \bar{X}_1$ , pondérées respectivement par  $\omega, \omega(1-\omega), \dots, \omega(1-\omega)^{i-1}$  (les poids forment donc une suite géométrique de raison  $1-\omega$ ).

**Remarque.** Dans le cas d'une carte de contrôle de phase I la valeur de référence  $\mu_0$  est inconnue et doit être remplacée par un estimateur de l'espérance. On pose classiquement alors :

$$Z_0 = \bar{\bar{X}}$$

### 5.3.2 Constructions

On utilise ici la méthode classique de construction avec la ligne centrale positionnée à la valeur moyenne théorique et les limites de contrôle découlant de la règle des  $3\sigma$ . Il en découle que :

1) positionnement de la ligne centrale à la valeur théorique  $E(Z_i)$  avec donc :

$$E(Z_i) = \omega \sum_{j=0}^{i-1} (1-\omega)^j E(\bar{X}_{i-j}) + (1-\omega)^i \mu_0$$

Si le processus est sous contrôle chaque observation est une réalisation de la loi normale  $N(\mu_0, \sigma_0)$  et donc  $E(\bar{X}_{i-j}) = \mu_0$ . Il en découle que :

$$E(Z_i) = \mu_0 \left[ \omega \sum_{j=0}^{i-1} (1-\omega)^j + (1-\omega)^i \right]$$

Or :

$$\sum_{j=0}^{i-1} (1-\omega)^j = \frac{1 - (1-\omega)^i}{w} \quad \text{donc} \quad E(Z_i) = \mu_0$$

2) positionnement des limites de contrôle selon la règle des  $3\sigma$  à partir de (d'après l'indépendance des observations) :

$$Var(Z_i) = \omega^2 Var \left[ \sum_{j=0}^{i-1} (1-\omega)^j \bar{X}_{i-j} \right] = \omega^2 \sum_{j=0}^{i-1} (1-\omega)^{2j} Var(\bar{X}_{i-j})$$

Si le processus est sous contrôle chaque observation est une réalisation de la loi normale  $N(\mu_0, \sigma_0)$  et donc  $Var(\bar{X}_{i-j}) = \sigma_0^2/n$ . Il en découle que :

$$Var(Z_i) = \frac{\omega^2 \sigma_0^2}{n} \sum_{j=0}^{i-1} (1-\omega)^{2j} = \frac{\omega^2 \sigma_0^2}{n} \left[ \frac{1 - (1-\omega)^{2i}}{1 - (1-\omega)^2} \right] = \frac{\sigma_0^2}{n} \left( \frac{\omega}{2-\omega} \right) \left[ 1 - (1-\omega)^{2i} \right]$$

En pratique on représente souvent sur la carte les limites de contrôle sous forme de fonctions. D'où la proposition suivante :

#### Proposition 13

Pour une valeur du coefficient  $0 < \omega \leq 1$  donné une **carte de contrôle de type EWMA** est déterminée par :

- 1) la ligne centrale à la valeur  $\mu_0$ ,
- 2) les limites de contrôle représentées par les fonctions :

$$t \mapsto \mu_0 \pm 3\sigma_0 \sqrt{\frac{\omega}{n(2-\omega)} \left[ 1 - (1-\omega)^{2t} \right]}$$

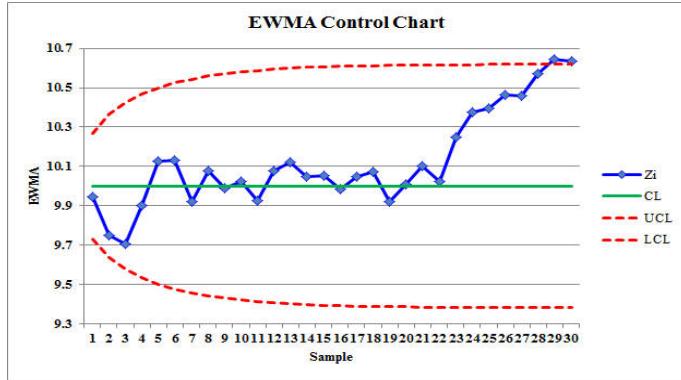


Figure 17. Carte de contrôle EWMA.

**Remarque.** Dans le cas d'une carte de contrôle de phase I les valeurs de référence  $\mu_0$  et  $\sigma_0$  sont alors inconnues et doivent être remplacées par des estimateurs.

**Remarque.** Asymptotiquement les limites de contrôle tendent vers :

$$\lim_{t \rightarrow +\infty} LC(t) = \mu_0 \pm 3\sigma_0 \sqrt{\frac{\omega}{n(2-\omega)}}$$

Ce comportement asymptotique est en général atteint très rapidement (car  $(1-\omega)^{2t}$  converge rapidement vers zéro).

En pratique ce type de carte de contrôle s'avère être plus performante qu'une carte classique aux moyennes de Shewhart afin de détecter une dérive lente. Par contre elle est moins performante afin de détecter une dérive rapide ! Elle ne doit donc pas se substituer à une carte classique de Shewhart, il est conseillé d'utiliser les deux cartes en parallèle.

## 5.4 Cartes de contrôle CUSUM

Il s'agit des cartes de contrôle de type « Cumulative Sum » (donc cartes de contrôle à sommes cumulées). Considérons par la suite la situation classique où  $k$  sous-groupes rationnels de même taille  $n$  sont observés.

### 5.4.1 Définitions

L'idée des cartes de contrôle de type CUSUM est de tout simplement cumuler toutes les moyennes par sous-groupes rationnels obtenues (et non pas réaliser des moyennes ou autres pondérations comme précédemment). Il est cependant classique avec ce type de cartes de contrôle de centrer voire centrer et réduire les observations au préalable. Considérons alors ici la définition suivante :

#### Définition 14

Une **carte de contrôle de type CUSUM** consiste en la représentation des valeurs  $c_1, \dots, c_k$  réalisations des variables aléatoires  $C_1, \dots, C_k$  définies par la relation :

$$\forall i = 1, \dots, k, C_i = \sum_{j=1}^i \sqrt{n} \left( \frac{\bar{X}_j - \mu_0}{\sigma_0} \right)$$

### 5.4.2 Constructions

En première approche on pourrait être tenté ici de construire des limites de contrôle toujours selon la même technique proposée par Shewhart, c'est-à-dire en les positionnant à  $\pm 3\sigma_i$  de la valeur moyenne où  $\sigma_i$  désigne l'écart-type de chacune des variables aléatoires  $C_i$ . Cette méthode va cependant s'avérer très inefficace afin de détecter rapidement un déréglage. Ceci vient du fait que les quantités  $C_i$  considérées ici tiennent compte de tout le passé du processus (avec chaque fois le même « poids ») et donc lors d'un déréglage les premiers relevés vont passer inaperçus à cause du poids trop important du passé correspondant au processus sous contrôle.

On privilégie alors aux limites de contrôle une stratégie axée sur la **détection de rupture** dans le processus étudié. Considérons les deux hypothèses suivantes :

$$H_0 : \forall i = 1, \dots, k, \forall j = 1, \dots, n, X_{ij} \sim N(\mu_0, \sigma_0), \text{ no hay problema}$$

$$H_1 : \exists r < k \text{ tel que } \forall i = 1, \dots, r-1, \forall j = 1, \dots, n, X_{ij} \sim N(\mu_0, \sigma_0) \text{ et}$$

$$\forall i = r, \dots, k, \forall j = 1, \dots, n, X_{ij} \sim N(\mu_0 + \delta, \sigma_0) \text{ avec } \delta > 0.$$

En d'autres termes l'hypothèse  $H_0$  correspond à une situation où le processus est toujours sous contrôle alors que  $H_1 = \bar{H}_0$  correspond à un décentrage de la moyenne d'une amplitude égale à  $\delta$  intervenant à partir du sous-groupe rationnel  $r$ .

**Remarque.** L'hypothèse  $H_1$  présentée ici correspond à un décentrage positif d'amplitude  $\delta$ . L'hypothèse d'un décentrage négatif peut aussi être testée en prenant une valeur de  $\delta$  telle que  $\delta < 0$ . On peut aussi considérer une hypothèse bilatérale en postulant que le processus est hors contrôle si les observations suivent une loi  $N(\mu_0 + \delta, \sigma_0)$  ou bien une loi  $N(\mu_0 - \delta, \sigma_0)$ .

Ecrivons maintenant les hypothèses  $H_0$  et  $H_1$  avec les quantités utilisées sur ce type de cartes de contrôle c'est-à-dire les variables aléatoire  $C_i$ . Le résultat suivant est immédiat (en remarquant pour la dernière relation qu'elle s'écrit comme la somme de  $i - (r-1)$  éléments, d'où l'écart-type donné) :

$$H_0 : \forall i = 1, \dots, k, C_i \sim N(0, \sqrt{i}),$$

$$H_1 : \exists r < k \text{ tel que } \forall i = 1, \dots, r-1, C_i \sim N(0, \sqrt{i}) \text{ et}$$

$$\forall i = r, \dots, k, (C_i - C_{r-1}) \sim N\left(\sqrt{n} \frac{\delta}{\sigma_0}, \sqrt{i - (r-1)}\right) \text{ avec } \delta > 0.$$

D'un point de vue géométrique ces formulations traduisent simplement le fait que sous l'hypothèse  $H_0$  (processus sous contrôle) les valeurs  $c_i$  doivent être en moyenne réparties autour d'une **droite de pente nulle** (l'axe des abscisses en l'occurrence) alors que sous l'hypothèse  $H_1$  (processus déréglé) les valeurs  $c_i$  doivent être en moyenne réparties autour d'une **droite de pente égale à  $\sqrt{n}(\delta/\sigma_0)$** . Cette technique va permettre de détecter assez facilement des petites erreurs systématiques qui, en s'accumulant, vont induire une pente non-nulle pour la droite associée à l'hypothèse  $H_1$ .

Concrètement il faut maintenant être capable de réaliser un choix entre les hypothèses  $H_0$  et  $H_1$ . Voici la règle de décision couramment utilisée :

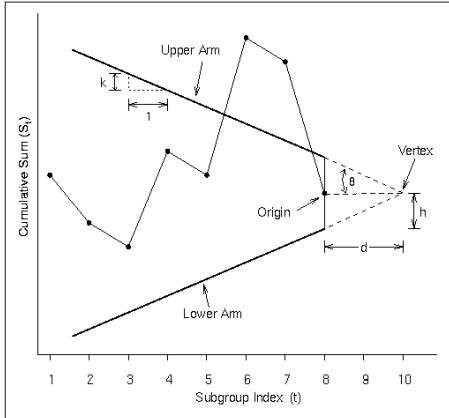
#### Règle de décision

Etant donné un seuil  $h > 0$  fixé arbitrairement (le seuil jouant un rôle similaire à celui du seuil de signification  $\alpha$  d'un test d'hypothèse) le choix entre l'hypothèse  $H_0$  et l'hypothèse  $H_1$  peut être fait pour le sous-groupe rationnel  $i$  de la manière suivante. On rejette l'hypothèse  $H_0$  si et seulement si il existe au moins un entier  $j \leq i$  tel que :

$$c_j \geq (i-j)k + h + c_i \text{ avec } k = \sqrt{n}(\delta/2\sigma_0)$$

**Interprétation géométrique.** Afin d'effectuer un choix entre l'hypothèse  $H_0$  (points répartis autour d'une droite de pente nulle) et  $H_1$  (points répartis autour d'une droite de pente  $\sqrt{n}(\delta/\sigma_0)$ ) on utilise une droite de pente « moyenne » (en coupant la poire en deux) égale donc à  $k$ . On construit ensuite la droite de pente  $k$  passant par le point de coordonnées  $(i, c_i - h)$ . La règle de décision proposée ci-dessus (effectuée au niveau du sous-groupe rationnel  $i$ ) traduit simplement le fait qu'au moins un des sous-groupes rationnels prélevé avant  $i$  est représenté par un point en dessous de la droite construite.

Si l'hypothèse  $H_1$  correspond cette fois à la loi  $N(\mu_0 - \delta, \sigma_0)$  elle conduit alors géométriquement à construire la droite de pente  $-k$  passant par le point de coordonnées  $(i, c_i + h)$ . On en déduit alors une représentation graphique classique pour le test bilatéral via un **masque en V** (voir la figure 18).



**Figure 18.** *Masque en V.*

**Remarque.** Le masque en V a été entièrement défini ici par la donnée de la distance  $h$  et de la pente  $k$ . De manière équivalente il est possible aussi de le définir à l'aide de la distance  $d$  et de l'angle  $\theta$  (voir la figure 18).

### Proposition 15

Considérons un procédé qui, sous contrôle, suit une loi  $N(\mu_0, \sigma_0)$ . Afin de détecter un décentrage de la moyenne d'une amplitude égale à  $\pm\delta$  une carte de type **CUSUM** est construite au niveau du sous-groupe rationnel  $i$  (pour  $i = 1, \dots, k$ ) à l'aide d'un **masque en V** ayant pour seuil une valeur  $h > 0$  et pour pentes associées aux deux bras les valeurs  $\pm k$  avec  $k = \sqrt{n}(\delta/2\sigma_0)$ . Le procédé sera jugé **hors contrôle** si au moins un des sous-groupes rationnels prélevé avant  $i$  est représenté par un point au dessus du bras supérieur ou en dessous du bras inférieur.

**Remarque.** L'utilisateur est donc libre de choisir toute valeur pour le seuil  $h$ . Des études menées sur ce sujet conseillent de, par exemple, poser :

$$h = -\frac{\sigma_0}{\sqrt{n}} \left( \frac{\ln(\alpha/2)}{\delta} \right)$$

dans le but d'obtenir ainsi une règle de décision ayant une erreur de première espèce proche de la valeur  $\alpha$  souhaitée.

## 5.5 Résumé : exemple numérique

Considérons les 10 valeurs suivantes correspondant aux moyennes relevées sur  $k = 10$  sous-groupes rationnels de taille  $n = 4$ .

10.14	10.78	09.24	09.79	10.10	10.88	11.20	11.30	10.53	10.66
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Ces données (simulées) correspondent aux moyennes relevées lorsque le procédé est sous contrôle pour les cinq premières (chaque observations étant alors une réalisation de la loi normale  $N(10, 1)$ ) et hors contrôle avec un léger déréglage de la moyenne pour les cinq dernières (chaque observations étant alors une réalisation de la loi normale  $N(11, 1)$ ).

**Carte de contrôle  $\bar{X}$**  La carte est ici construite selon le principe de Shewhart avec donc :

$$LIC = 8.5 \quad / \quad LC = 10 \quad / \quad LSC = 11.5$$

Il vient donc :

$LIC$	08.50	08.50	08.50	08.50	08.50	08.50	08.50	08.50	08.50	08.50
$\bar{x}_i$	10.14	10.78	09.24	09.79	10.10	10.88	11.20	11.30	10.53	10.66
$LSC$	11.50	11.50	11.50	11.50	11.50	11.50	11.50	11.50	11.50	11.50

Cette carte de contrôle ne détecte à priori aucun problème. Ceci provient du fait que le processus étudié a pour moyenne sous-contrôle  $\mu_0 = 10$  et pour moyenne une fois déréglé  $\mu = 11 = \mu_0 + \Delta$  avec  $\Delta = \sigma_0$ . Ce déréglage égal à « seulement » un écart type est insuffisant pour que cette carte construite à partir de la règle « des  $3\sigma$  » détecte efficacement les problèmes (concrètement sous contrôle quasiment toutes les moyennes par sous-groupe rationnel vont être dans l'intervalle  $[8.5, 11.5]$  alors qu'en situation de déréglage elles vont être dans l'intervalle  $[9.5, 12.5]$  ...). On peut aussi quantifier ceci à l'aide de la POM de cette carte. D'après les formules établies en TD la probabilité d'erreur de seconde espèce est égale à :

$$\beta = F\left(3 - \frac{\Delta\sqrt{n}}{\sigma_0}\right) - F\left(-3 - \frac{\Delta\sqrt{n}}{\sigma_0}\right)$$

avec  $F$  fonction de répartition de la loi normale centrée et réduite. Il vient alors ici :

$$\beta = F(1) - F(-5) \simeq F(1) \simeq 0.8413$$

On en déduit que :

$$POM = \frac{1}{1 - \beta} \simeq 6.3$$

Il faut donc s'attendre dans ce cas à avoir besoin, en moyenne, de 6 ou 7 sous-groupes rationnels collectés après le déréglage afin que la carte  $\bar{X}$  détecte le problème.

**Remarque.** *Cette carte de contrôle va devenir d'autant plus efficace que l'on augmente la taille  $n$  des sous-groupes rationnels considérés. Ceci est cependant en général impossible à réaliser pour des raisons de coût.*

**Carte de contrôle MA** Construisons ici une carte aux moyennes mobiles d'ordre  $h = 5$  (l'utilisateur pouvant bien entendu prendre d'autres valeurs de  $h$ ). D'après la proposition 11 il vient alors :

$LIC$	08.50	08.94	09.13	09.25	09.33	09.33	09.33	09.33	09.33	09.33
$m_i$	10.14	10.46	10.05	09.99	10.01	10.16	10.24	10.65	<u>10.80</u>	<u>10.91</u>
$LSC$	11.50	11.06	10.87	10.75	10.67	10.67	10.67	10.67	10.67	10.67

Cette carte de contrôle se montre donc plus efficace que la précédente puisque le problème est ici détecté à partir du sous-groupe rationnel 9 (soit après 4 relevés effectués en situation de déréglage).

**Carte de contrôle EWMA** Construisons maintenant une carte  $EWMA$  avec (par exemple) la pondération classique  $\omega = 0.1$ . D'après la définition 12 les valeurs à représenter ici sont donc obtenues itérativement via la formule :

$$\forall i = 1, \dots, 10, \quad Z_i = 0.1\bar{X}_i + 0.9Z_{i-1} \quad \text{avec} \quad Z_0 = \mu_0 = 10$$

D'après la proposition 13 les limites de contrôle sont données par les fonctions :

$$t \mapsto 10 \pm 3\sqrt{\frac{1}{76}(1 - 0.9^{2t})}$$

Il vient alors :

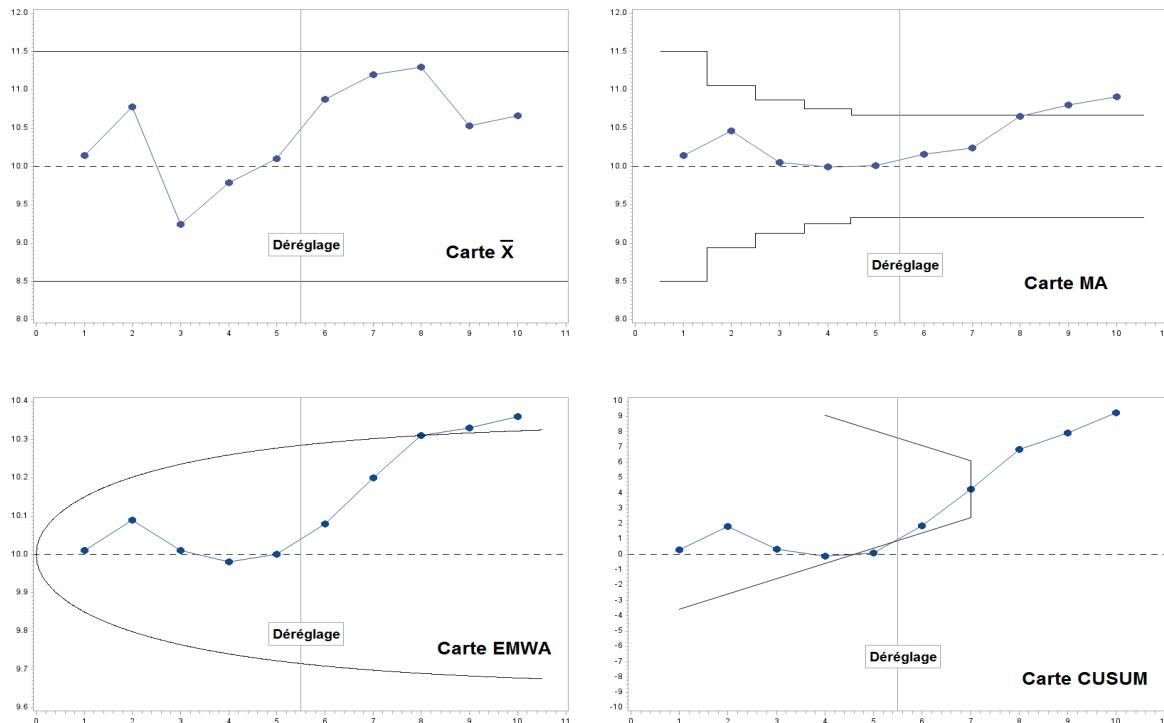
$LIC$	09.85	09.80	09.76	09.74	09.72	09.70	09.69	09.69	09.68	09.68
$z_i$	10.01	10.09	10.01	09.98	10.00	10.08	10.20	10.31	<u>10.33</u>	<u>10.36</u>
$LSC$	10.15	10.20	10.24	10.26	10.28	10.30	10.31	10.31	10.32	10.32

Une nouvelle fois cette carte de contrôle se montre efficace puisque le problème est détecté à partir du sous-groupe rationnel 9 (soit après 4 relevés effectués en situation de déréglage).

**Carte de contrôle CUSUM** Les différentes valeurs relevées donnent sous forme cumulée (après centrage et réduction à partir des valeurs sous contrôle  $\mu_0 = 10$  et  $\sigma_0 = 1$ ) :

$c_i$	0.28	1.84	0.32	-0.10	<u>0.10</u>	1.86	4.26	6.86	7.92	9.24
-------	------	------	------	-------	-------------	------	------	------	------	------

On construit ensuite les différents « masques en V » en chacun des sous-groupes rationnels. Ils sont configurés ici avec  $\delta = 1$  (détection d'un décentrage d'une unité, voir la proposition 15) et  $h \simeq 1.84$  (pour une erreur de première espèce de l'ordre de  $\alpha = 0.05$ , voir la remarque faisant suite à la proposition 15). On constate alors la très bonne réactivité de cette carte de contrôle puisqu'elle détecte le déréglage au niveau du sous-groupe rationnel 5.



**Figure 19.** Diverses cartes de contrôle.

**Maîtrise Statistique des Procédés**  
Master MSID 2ème année  
Travaux dirigés - Série 5

**Exercice 1** *Cartes de contrôle pour petits dérégagements*

On s'intéresse à une production de sacs de ciment. Si la production est sous contrôle alors la masse des sacs doit suivre une loi normale de type  $N(40, 0.5)$ . On extrait alors toutes les heures 4 sacs de la production et on mesure leur masse (en kg). Voici les résultats obtenus après 6 heures de fonctionnement de la chaîne de production :

	1	2	3	4	5	6
1	40.2	40.3	39.8	40.2	39.6	40.6
2	39.8	40.1	40.3	40.8	40.4	40.8
3	39.6	40.4	40.1	40.3	40.8	41.0
4	40.4	39.6	40.6	40.3	40.8	40.4
$\bar{X} =$	40.1	40.1	40.2	40.3	40.4	40.7

- 1** Construire la carte de contrôle classique de type  $\bar{X}$ . Commenter les résultats obtenus.
- 2** Construire la carte de contrôle de type  $MA$  d'ordre 4. Comparer les résultats avec ceux de la première question. Quelle conclusion peut-on tirer ?
- 3** Construire maintenant une carte EWMA avec la valeur classique  $\omega = 0.15$ . Commenter les résultats obtenus.
- 4** Que devient ce type de carte de contrôle pour les valeurs extrêmes  $\omega = 0$  et  $\omega = 1$  ?

**Exercice 2** *Carte de contrôle multivariée.*

Une entreprise veut lancer la production d'une nouvelle colle et pour ceci teste sa nouvelle chaîne de production durant 5 jours. La qualité de la colle produite est très dépendante des deux variables numériques que sont le pH de la colle ainsi que sa viscosité. Durant chaque demi-journée de production 3 échantillons de colle sont prélevés et ces deux caractéristiques sont mesurées. Les résultats sont résumés dans le tableau ci-dessous :

	pH	vis.												
J1	7.7	10.0	J2	8.3	10.8	J3	7.9	10.3	J4	7.7	10.0	J5	8.2	10.2
	8.5	11.0		7.5	09.8		8.1	10.5		8.3	10.8		7.7	08.6
	7.5	09.8		7.6	09.9		8.1	10.5		8.2	10.6		7.9	09.4

- 1** Réaliser une analyse statistique simple des deux variables (sans tenir compte de l'effet temporel lié aux sous-groupes rationnels).

2] Construire séparément des cartes de contrôle de type  $\bar{X}$  pour ces deux variables. Que peut-on en conclure ?

3] Construire maintenant une carte de contrôle multivariée de Hotelling. La production réalisée durant ces 5 jours peut-elle être supposée sous contrôle ?

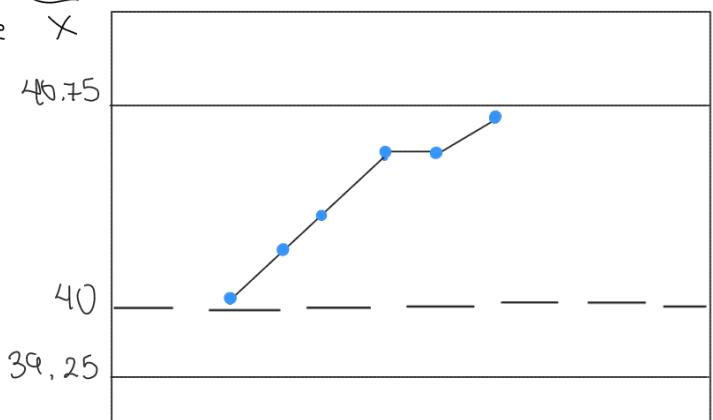
# Exercices

## Exercice 1)

1) On a

$$\begin{array}{c} \bar{x}_1 \\ 40 \\ \bar{x}_2 \\ 40.1 \\ \bar{x}_3 \\ 40.2 \\ \bar{x}_4 \\ 40.4 \\ \bar{x}_5 \\ 40.4 \\ \bar{x}_6 \\ 40.7 \end{array}$$

Carte  $\bar{X}$



$$LSC = 40 + 3 \frac{\sigma_0}{\sqrt{m}}$$

$\Rightarrow$  On a une procédé sous-contrôle

$$LSC = 39.25$$

$$M_1 = \bar{x}_1 = 40$$

$$M_2 = \frac{\bar{x}_1 + \bar{x}_2}{2} = \frac{40 + 40.1}{2} = 40.05$$

2) Construction de la carte MA (avec  $h=4$ )

Valeurs représentées ( $m_i, i=1, \dots, 6$ ):  $40 / 40.05 / 40.1 / 40.18 / 40.28 / 40.43$

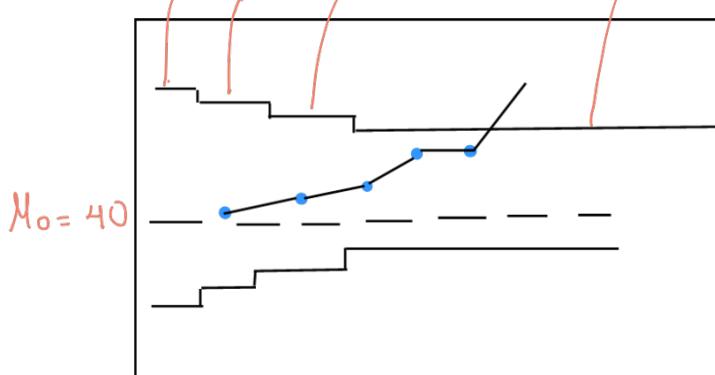
moyennes mobiles incomplètes

Calcul pour les 3 premières est  $\neq$  de celui qui sera utilisé à partir du  $m_4$

pour  $m_2$

$$LSC = 40 + \frac{3 \sigma_0}{\sqrt{8}}$$

$$40.15 \quad 40.53 \quad 40.43$$



Processus hors-contrôle!

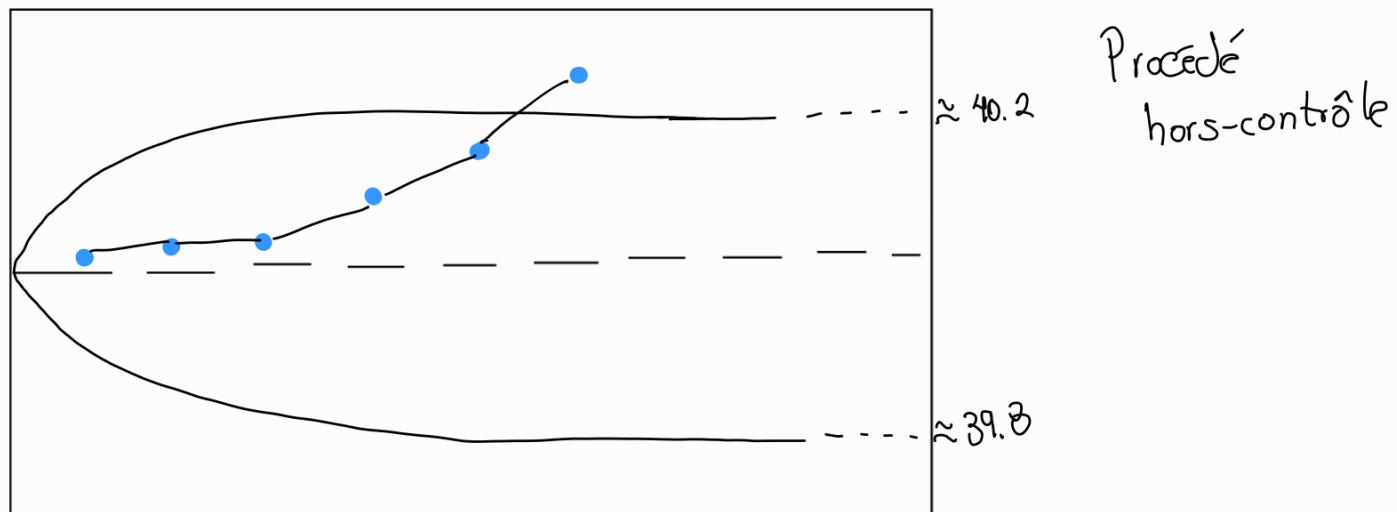
### 3) Carte EWMA avec $\omega = 0.15$

Donc  $Z_i = 0.15\bar{x}_i + 0.85Z_{i-1}$  où  $Z_0 = M_0 = 40$   
donc les  $Z_i$  valent (pour  $i=1, \dots, 6$ )  $40 / 40.02 / 40.04 / 40.11 / 40.14 / 40.23$

limites de contrôle

$$t \rightarrow 40 \pm 1.5 \sqrt{\frac{0.15}{4(1.85)} [1 - 0.85^{2t}]} \Rightarrow \text{on représente la fonction}$$

pour  $t$  grand  $\approx \sqrt{\frac{0.15}{4(1.85)}} \approx 0.14$



Procédé hors-contrôle

### 4) Cartes EWMA avec valeurs extrêmes pour $\omega$

$$\boxed{\omega = 1} \rightarrow Z_i = \bar{x}_i \rightarrow \text{carte } \bar{X}$$

$$\boxed{\omega = 0} \rightarrow Z_i = Z_{i-1} = Z_{i-2} = \dots = Z_0 = M_0 \text{ absurdes}$$

Nous avons des valeurs constantes sur la carte

$\omega$  grande no habrá gran diferencia con la carta  $\bar{X}$

cada que retrocederemos en  $i$ , hacemos pequeña la información que nos pasa i.e. así no se toma en cuenta el pasado

### Exercice 2)

1) On remarque:  $R \approx 0.738 = \cos(\phi_h, v_{ist})$

A priori deux variables sont corrélées (positivement) et donc pas indépendantes. Nuage des points

- Faire une carte de contrôle pour le pH et une autre

2) Si on considère pour la viscosité elles ne détectent aucun problème.  
 (utiliser les étendues)

On a:	moy pH	moy Vis	Var pH	Var Vis	Corrélation
Sgn 1	7.9	10.27	0.28	0.41	0.34
Sgn 2	7.8	10.17	0.19	0.3	0.24
Sgn 3	8.03	10.43	0.01	0.01	0.01
Sgn 4	8.07	10.47	0.1	0.17	0.13
Sgn 5	7.93	9.4	0.06	0.64	0.20

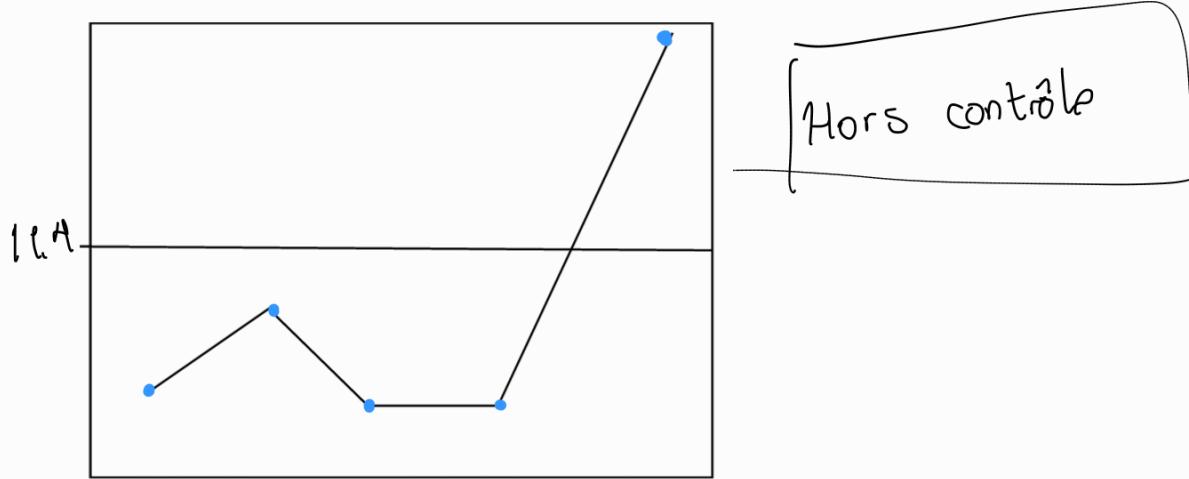
Vemos que la dispersion es grande para sgn 5

Estimateurs  $\hat{\mu}_1 = \frac{18.4}{5} = 7.95$ ,  $\hat{\mu}_2 = 10.15$ ,  $\hat{\sigma}_1^2 = 0.13$ ,  $\hat{\sigma}_2^2 = 0.31$ ,  $\hat{\sigma}_{12}^2 = 0.19$

D'où les  $t_i^2$  ( $i=1,2,\dots,5$ ):  $2.4/4 / 12 / 1.8 / 35.7$

Limite de contrôle

$$LSC = \left( \frac{24}{9} \right) F_{1-\alpha, 2, 9} \approx 11.4 \text{ pour } \alpha = 5\%$$



# Chapitre 6

## Cartes de contrôle multivariées

### 6.1 Introduction

Nous avons jusqu'à présent considéré uniquement des situations où **une seule** caractéristique de contrôle était mesurée. Il existe en pratique une multitude de situations où **plusieurs** caractéristiques de contrôle simultanées présentent un intérêt. A titre d'exemple lors de la fabrication d'un roulement à billes les diamètre intérieur et extérieur sont mesurés. La pièce usinée va bien fonctionner lorsque ces deux valeurs sont dans une « **région de contrôle** » donnée, la pièce va être défectueuse sinon.

Une première approche consiste simplement à construire deux cartes de contrôle séparément (une pour le diamètre intérieur et l'autre pour le diamètre extérieur) et considérer que le processus est sous contrôle si et seulement si il est sous contrôle pour les deux cartes. Géométriquement ceci revient donc à construire une région de contrôle rectangulaire comme présenté à la figure 20.

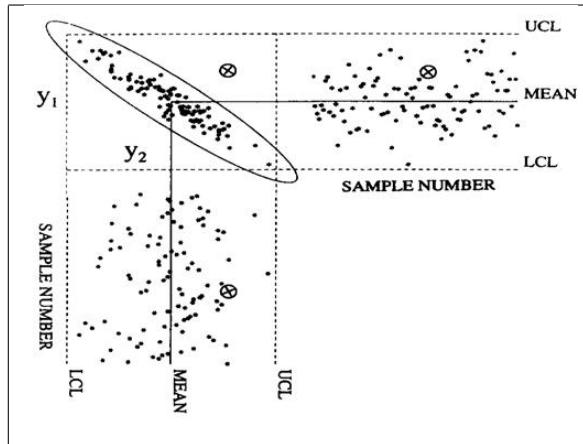


Figure 20. *Région de contrôle.*

Cette technique présente comme problème principal le fait que l'on suppose alors que les deux caractéristiques de contrôle sont **indépendantes**, ce qui bien entendu n'est pas forcément toujours le cas. Remarquons aussi que sous hypothèse de normalité pour les deux cartes de contrôle ce type de technique ne permet plus d'utiliser la « règle des  $3\sigma$  » sous sa forme usuelle. En effet la probabilité d'être entre les limites de contrôles alors que le procédé est bien sous contrôle est de 0.9973 pour chacune des deux cartes donc la probabilité de se trouver dans la zone de contrôle rectangulaire alors que le procédé est bien sous contrôle est maintenant égale à  $(0.9973)^2 = 0.9946$ . Ces différences sont importantes en pratique car l'erreur de première espèce pour une carte de contrôle de Shewhart est donc de 0.0027 (soit de l'ordre de une chance sur 370) alors qu'ici, pour l'utilisation de deux variables, elle devient égale à 0.0054 (soit de l'ordre de une chance sur 185). Cette tendance va, bien entendu, s'aggraver avec le nombre de caractéristiques de contrôle considérées.

Une autre approche consiste alors à ne plus considérer séparément ces deux caractéristiques de contrôle mais, au contraire, à s'intéresser à leur **distribution conjointe**. Ceci va permettre, le cas échéant, d'introduire une corrélation entre les variables étudiées. Les « zones de contrôles » seront ainsi plus fines et plus adaptées à la nature du phénomène étudié. La figure 20 montre une zone de contrôle sous forme d'ellipse qui semble mieux traduire la réalité que la zone de contrôle rectangulaire présentée précédemment (dans la mesure où une des observations - représentée par une croix dans un cercle - semble vraiment atypique par rapport aux autres et se retrouve hors contrôle uniquement par rapport à la région de contrôle ellipsoïdale).

Dans la suite nous présenterons souvent des cartes de contrôle pour la surveillance simultanée de **deux** caractéristiques de contrôle afin de pouvoir traduire géométriquement les résultats obtenus. Les méthodes proposées pourront cependant être étendues sans problème à un plus grand nombre de variables, ceci permettra de généraliser les cartes de contrôle de Shewart de la moyenne.

## 6.2 Rappels et compléments

### 6.2.1 Loi gaussienne multidimensionnelle

Rappelons ici quelques notions élémentaires de probabilité relatives au cas multivarié. Pour tout **vecteur aléatoire**  $Y \in \mathbb{R}^n$  désignons par  $\mu$  son espérance mathématique et par  $V$  sa matrice de covariances (que l'on supposera ici toujours exister). On étend alors au cas vectoriel la loi normale en disant que le vecteur aléatoire  $Y$  va suivre une **loi gaussienne** de type  $N(\mu, V)$  si et seulement si il admet une densité de probabilité de la forme (dès lors que sa matrice des covariances  $V$  est inversible) :

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(V)}} \exp \left[ -\frac{1}{2} {}^t(x - \mu) V^{-1} (x - \mu) \right].$$

Limitons-nous à deux caractéristiques de contrôle. Dans ce cas particulier on considère donc  $Y \in \mathbb{R}^2$  admettant une espérance mathématique et une matrice de covariances de la forme :

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ et } V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Notons que l'on a alors (avec  $\sigma_{12}$  la covariance entre les deux variables étudiées) :

$$\det(V) = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2 \quad \text{et} \quad V^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}.$$

Il en résulte que :

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2}} \exp \left[ -\frac{1}{2(\sigma_1^2 \sigma_2^2 - \sigma_{12}^2)} \left( \sigma_2^2 (x_1 - \mu_1)^2 + \sigma_1^2 (x_2 - \mu_2)^2 - 2\sigma_{12}(x_1 - \mu_1)(x_2 - \mu_2) \right) \right]$$

### 6.2.2 Carte de contrôle multivariée

Le cas d'une carte de contrôle multivariée est celui, de manière très générale, où  $p$  caractéristiques de contrôle numériques sont relevées via  $k$  sous-groupes rationnels de taille  $n$ . Désignons les observations du sous-groupe rationnel  $i$  ( $1 \leq i \leq k$ ), relatives à la variable de contrôle  $l$  ( $1 \leq l \leq p$ ), par  $x_{i1}^{[l]}, \dots, x_{in}^{[l]}$  (elles sont donc des réalisations des variables aléatoires réelles  $X_{i1}^{[l]}, \dots, X_{in}^{[l]}$ ). Notons alors  $X_{ij}$  le vecteur aléatoire modélisant l'ensemble des  $p$  relevés effectués lors de la mesure  $j$  relative au sous-groupe rationnel  $i$  et  $\bar{X}_i$  le vecteur modélisant l'ensemble des valeurs moyennes pour ce même sous-groupe rationnel, donc :

$$X_{ij} = \begin{pmatrix} X_{ij}^{[1]} \\ \vdots \\ X_{ij}^{[p]} \end{pmatrix} \text{ et } \bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij} = \begin{pmatrix} \bar{X}_i^{[1]} \\ \vdots \\ \bar{X}_i^{[p]} \end{pmatrix}.$$

Nous traduirons par la suite la situation sous contrôle en supposant que toutes les variables aléatoires vectorielles  $X_{ij}$  suivent une même loi gaussienne de type  $N(\mu, V)$ .

## 6.3 Cartes de contrôle de Hotelling de phase II

### 6.3.1 Définitions

Commençons par détailler le cas le plus simple. La phase II correspond dans le cas multidimensionnel à une situation où, sous contrôle, chaque relevé  $X_{ij}$  est une réalisation de la loi gaussienne  $N(\mu_0, V_0)$  avec le vecteur des espérances  $\mu_0$  et la matrice des covariances  $V_0$  supposés connus. Afin de pouvoir se ramener, quelle que soit la valeur de  $p$ , à des valeurs numériques à une seule dimension Hotelling a proposé la définition suivante :

#### Définition 16

Une **carte de contrôle de Hotelling** (encore appelée carte de contrôle du  $T^2$ ), dans le cas de la phase II, consiste en la représentation des valeurs numériques  $t_1^2, \dots, t_k^2$  réalisations des variables aléatoires  $T_1^2, \dots, T_k^2$  définies par la relation :

$$\forall i = 1, \dots, k, T_i^2 = n^t (\bar{X}_i - \mu_0) V_0^{-1} (\bar{X}_i - \mu_0)$$

**Remarque.** Les valeurs  $T_i^2$  proposées ici généralisent au cas multidimensionnel la technique classique de « centrage et réduction ». En effet dans le cas où  $p = 1$  et si chaque variable aléatoire réelle  $X_{ij}$  a pour espérance  $\mu_0$  et variance  $\sigma_0^2$  la forme centrée et réduite de  $\bar{X}_i$  est alors donnée par :

$$\bar{X}_i^* = \frac{\bar{X}_i - \mu_0}{\sigma_0 / \sqrt{n}} \quad \text{donc} \quad T_i^2 = n \left( \frac{\bar{X}_i - \mu_0}{\sigma_0} \right)^2 = (\bar{X}_i^*)^2.$$

Dans le cas particulier (très courant en pratique) où  $p = 2$  caractéristiques de contrôle sont mesurées il vient alors (reprendre les résultats et les notations du paragraphe précédent) :

$$T_i^2 = \frac{n}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \begin{pmatrix} \bar{X}_i^{[1]} - \mu_1 & \bar{X}_i^{[2]} - \mu_2 \end{pmatrix} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix} \begin{pmatrix} \bar{X}_i^{[1]} - \mu_1 \\ \bar{X}_i^{[2]} - \mu_2 \end{pmatrix}$$

donc  $T_i^2$  est finalement donné par l'expression développée suivante :

$$T_i^2 = \frac{n}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \left( \sigma_2^2 (\bar{X}_i^{[1]} - \mu_1)^2 + \sigma_1^2 (\bar{X}_i^{[2]} - \mu_2)^2 - 2\sigma_{12} (\bar{X}_i^{[1]} - \mu_1)(\bar{X}_i^{[2]} - \mu_2) \right)$$

### 6.3.2 Constructions

Remarquons que l'indépendance des vecteurs  $X_{ij}$  nous assure que :

$$\forall i = 1, \dots, k, \bar{X}_i = \left( \frac{1}{n} \sum_{j=1}^n X_{ij} \right) \rightsquigarrow N \left( \mu_0, \frac{1}{n} V_0 \right).$$

Utilisons maintenant le résultat général suivant faisant le lien entre loi gaussienne et loi du khi-deux :

#### Lemme 17

Soit  $Y$  un vecteur aléatoire de  $\mathbb{R}^n$  tel que  $Y$  suit une loi gaussienne de type  $N(\mu_0, V_0)$  avec  $V$  matrice des covariances inversible. Alors :

$${}^t (Y - \mu_0) V_0^{-1} (Y - \mu_0) \rightsquigarrow \chi_n^2$$

Ceci permet de proposer la technique de construction suivante :

### Proposition 18

Pour une **carte de contrôle de Hotelling** (encore appelée carte de contrôle du  $T^2$ ) de phase II la limite de contrôle supérieure probabiliste est placée à la valeur suivante pour une erreur de première espèce égale à  $\alpha$  :

$$LSC = \chi_{1-\alpha,p}^2$$

avec  $p$  nombre de caractéristiques de contrôle mesurées et  $\chi_{\alpha,n}^2$  fractile d'ordre  $\alpha$  de la loi du khi-deux à  $n$  degrés de liberté.

**Remarque.** Avec ce type de carte de contrôle il est classique de se limiter à cette seule limite de contrôle. On ne représente ni la limite inférieure de contrôle (peu intéressante en pratique) ni la « ligne moyenne » (pas forcément facile à interpréter concrètement dans le cas multivarié).

Dans le cas particulier (très courant en pratique) où  $p = 2$  caractéristiques de contrôle sont mesurées la région de contrôle est alors définie par (reprendre les résultats des paragraphes précédents) :

$$Cte \left( \sigma_2^2 \left( \bar{X}_i^{[1]} - \mu_1 \right)^2 + \sigma_1^2 \left( \bar{X}_i^{[2]} - \mu_2 \right)^2 - 2\sigma_{12} \left( \bar{X}_i^{[1]} - \mu_1 \right) \left( \bar{X}_i^{[2]} - \mu_2 \right) \right) \leq \chi_{1-\alpha,2}^2$$

avec  $Cte = n (\sigma_1^2 \sigma_2^2 - \sigma_{12}^2)^{-1} = n \sigma_1^2 \sigma_2^2 (1 - r^2) \geq 0$  où  $r$  est le coefficient de corrélation linéaire entre les deux variables étudiées. Géométriquement la région de confiance est alors une **ellipse** centrée en  $(\mu_1, \mu_2)$ .

La figure 21 donne un exemple concrèt d'une carte de type  $T^2$  (avec ici un sous-groupe rationnel très clairement hors contrôle). Notons qu'un des problèmes liés à ce type de carte de contrôle est la **difficulté d'interprétation** d'une situation hors contrôle. En effet l'utilisation des valeurs  $T_i^2$  a permis de réduire la dimension du problème étudié mais en contrepartie lorsqu'une des valeurs  $T_i^2$  est « trop grande » il n'apparaît pas clairement quelle en est la cause. Il conviendra alors de se pencher dans ce cas de figure sur chacune des  $p$  variables étudiées afin d'essayer de déterminer laquelle (voire lesquelles) est (sont) hors contrôle. Une des stratégies proposée dans ce but consiste à évaluer les diverses quantités numériques suivantes :

$$\forall i = 1, \dots, p, d_i = T^2 - T_{(i)}^2$$

avec  $T^2$  valeur calculée de la statistique de Hotelling au point hors-contrôle considéré et  $T_{(i)}^2$  valeur calculée en ce même point en supprimant la  $i$ -ème variable de contrôle. En d'autres termes les  $d_i$  permettent d'évaluer les « contributions » de chacunes des variables de contrôle par rapport à la valeur de  $T^2$ .

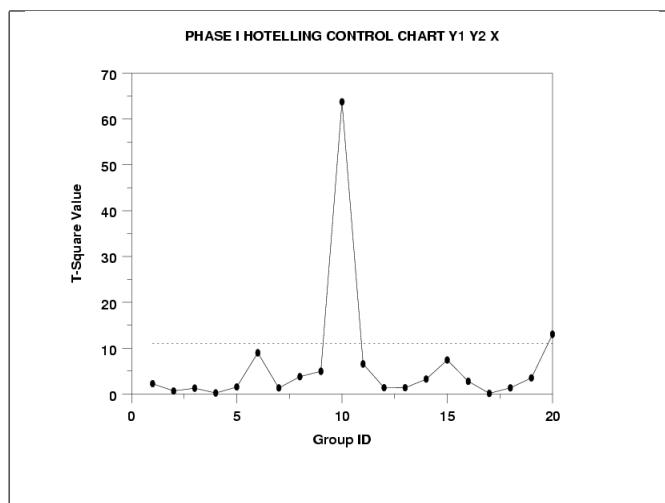


Figure 21. Exemple de carte de type  $T^2$ .

## 6.4 Cartes de contrôle de Hotelling de phase I

### 6.4.1 Définitions

Considérons ici le cas où le vecteur  $X_{ij}$  correspondant au  $j$ -ème ( $1 \leq j \leq n$ ) relevé du sous groupe rationnel  $i$  ( $1 \leq i \leq k$ ) suit une loi gaussienne de type  $N(\mu, V)$  avec le vecteur des espérances  $\mu$  et la matrice des covariances  $V$  tous deux **inconnus** (il existe des situations intermédiaires où seulement l'une de ces deux quantités est inconnue). Attention au fait qu'il ne suffit pas ici de modifier simplement les limites de contrôle obtenues en phase II puisque ces limites sont probabilistes et la loi de probabilité de la quantité précédente vue à la section 6.3 (soit  $T = n^t (\bar{X}_i - \mu) V^{-1} (\bar{X}_i - \mu)$ ) est inconnue maintenant ! Tout comme dans le cas classique à  $p = 1$  caractéristique de contrôle on **estime** alors tout d'abord  $\mu$  et  $V$  au sein du sous-groupe rationnel  $i$  par :

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij} \quad \text{et} \quad V_i = \frac{1}{n-1} \sum_{j=1}^n (X_{ij} - \bar{X}_i)^t (X_{ij} - \bar{X}_i)$$

De manière gloable il en découle alors les estimateurs suivants obtenus pour la totalité des  $k$  sous-groupes rationnels. Le vecteur  $\mu$  est estimé par :

$$\hat{\mu} = \bar{\bar{X}} = \frac{1}{k} \sum_{i=1}^k \bar{X}_i = \frac{1}{nk} \sum_{i=1}^k \sum_{j=1}^n X_{ij}.$$

La matrice  $V$  est estimée par :

$$\hat{V} = \frac{1}{k} \sum_{i=1}^k V_i = \frac{1}{k(n-1)} \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)^t (X_{ij} - \bar{X}_i)$$

Il en découle la définition suivante :

#### Définition 19

Une **carte de contrôle de Hotelling** (encore appelée carte de contrôle du  $T^2$ ), dans le cas de la phase I, consiste en la représentation des valeurs numériques  $t_1^2, \dots, t_k^2$  réalisations des variables aléatoires  $T_1^2, \dots, T_k^2$  définies par la relation :

$$\forall i = 1, \dots, k, \quad T_i^2 = n^t (\bar{X}_i - \hat{\mu}) \hat{V}^{-1} (\bar{X}_i - \hat{\mu})$$

Dans le cas particulier (très courant en pratique) où  $p = 2$  caractéristiques de contrôle sont mesurées il vient alors (reprendre les résultats et les notations du paragraphe précédent) :

$$T_i^2 = \frac{n}{\hat{\sigma}_1^2 \hat{\sigma}_2^2 - \hat{\sigma}_{12}^2} \begin{pmatrix} \bar{X}_i^{[1]} - \hat{\mu}_1 & \bar{X}_i^{[2]} - \hat{\mu}_2 \end{pmatrix} \begin{pmatrix} \hat{\sigma}_2^2 & -\hat{\sigma}_{12} \\ -\hat{\sigma}_{12} & \hat{\sigma}_1^2 \end{pmatrix} \begin{pmatrix} \bar{X}_i^{[1]} - \hat{\mu}_1 \\ \bar{X}_i^{[2]} - \hat{\mu}_2 \end{pmatrix}$$

donc  $T_i^2$  est finalement donné par l'expression développée suivante :

$$T_i^2 = \frac{n}{\hat{\sigma}_1^2 \hat{\sigma}_2^2 - \hat{\sigma}_{12}^2} \left( \hat{\sigma}_2^2 (\bar{X}_i^{[1]} - \hat{\mu}_1)^2 + \hat{\sigma}_1^2 (\bar{X}_i^{[2]} - \hat{\mu}_2)^2 - 2\hat{\sigma}_{12} (\bar{X}_i^{[1]} - \hat{\mu}_1) (\bar{X}_i^{[2]} - \hat{\mu}_2) \right)$$

où  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$  et  $\hat{\sigma}_{12}$  sont les estimateurs respectifs de la variance des caractéristiques de contrôle 1 et 2 ainsi que de leur covariance, ces estimateurs étant obtenus à partir de la moyenne des  $k$  estimateurs construits sur chacun des sous-groupes rationnels.

### 6.4.2 Constructions

On montre que lorsque le processus étudié est sous contrôle, avec donc tous les vecteurs aléatoires  $X_{ij}$  indépendants et de même loi gaussienne, alors la quantité  $T_i^2$  présentée précédemment peut être reliée à la loi de Fisher via la relation suivante :

$$\left( \frac{k(n-1) + 1 - p}{(k+1)(n-1)p} \right) T_i^2 \rightsquigarrow \mathcal{F}_{p,k(n-1)+1-p}$$

On en déduit alors le résultat suivant :

**Proposition 20**

Pour une **carte de contrôle de Hotelling** (encore appelée carte de contrôle du  $T^2$ ) de phase I la limite de contrôle supérieure probabiliste est placée à la valeur suivante pour une erreur de première espèce égale à  $\alpha$  :

$$LSC = \left( \frac{(k+1)(n-1)p}{k(n-1) + 1 - p} \right) f_{1-\alpha,p,k(n-1)+1-p}$$

avec  $p$  nombre de caractéristiques de contrôle mesurées,  $k$  nombre de sous-groupes rationnels utilisés,  $n$  taille commune de ces sous-groupes rationnels et  $f_{\alpha,n,m}$  fractile d'ordre  $\alpha$  de la loi de Fisher à  $n$  et  $m$  degrés de liberté.

## 6.5 Généralisations

La méthodologie présentée précédemment afin de, par exemple, détecter de petits déréglages peut tout à fait être étendue au cas multivarié. Il est alors possible de construire des cartes de contrôle multivariées de type EMWA ainsi que des cartes contrôle multivarées de type CUSUM (on parle alors de carte MCUSUM, le premier M désignant Multivariate).