

## TP du CC2–QQ-plot (graphiques quantile-quantile)

Soit  $x_1, \dots, x_n$  une série statistique. On peut chercher à savoir si la distribution des données est gaussienne ou Poisson etc... Notons  $F_0$  la fonction de répartition de cette loi de probabilité d'intérêt. Le QQ-plot est un outil graphique permettant de visualiser rapidement l'adéquation de la distribution d'une série numérique à une distribution de référence. Dans ce graphe, on reporte sur l'axe des ordonnées les fractiles correspondant à la distribution observée et sur l'axe des abscisses ceux correspondant à la distribution théorique.

### 2.1 En pratique

*Tableau des quantiles et nuage de points*

- Dans le cas d'une variable quantitative dont les valeurs sont regroupées par modalités :

Soient  $m_1, \dots, m_J$  les modalités de la série  $x_1, \dots, x_n$  que l'on appellera quantile observés ( $q_i = m_i$ ). On remplit le tableau des fréquences cumulées. Pour chaque fréquence cumulée, on calcule le quantile théorique i.e.  $q_j^*$  tel que

$$F_0(q_j^*) = F_j \quad \Leftrightarrow \quad q_j^* = F_0^{-1}(F_j)$$

En général ce calcul se fait par l'utilisation des tables statistiques ou par un logiciel.

*On reporte dans un graphique le nuage de points  $(q_j, q_j^*)_{j=1 \dots J}$ .*

Modalités ordonnées, quantiles observés	$q_1 = m_1$	$\dots$	$q_J = m_J$
Fréquences cumulées	$F_1$	$\dots$	$F_J$
Quantiles théoriques	$q_1^* = F_0^{-1}(F_1)$	$\dots$	$q_J^* = F_0^{-1}(F_J)$

**Note :** Les quantiles théoriques d'une loi normale  $q_j^* = F_0^{-1}(F_j)$  sont accessibles sur R via `qnorm( $F_j, m, \sigma$ )`.

### 2.2 Interprétation

- Si les points sont alignés sur la diagonale du carré de côté 1 (1ère bissectrice), alors la loi théorique proposée (de fonction de répartition  $F_0$ ) est adaptée aux observations.
- Si les points sont alignés sur une droite parallèle à la diagonale du carré de côté 1 on soupçonnera une erreur sur les paramètres de position de la loi théorique.
- Si les points sont alignés sur une droite passant par l'origine mais inclinée par rapport à la diagonale du carré de côté 1 on soupçonnera une erreur sur les paramètres de dispersion de la loi théorique.
- Si les points sont alignés sur une droite ne passant pas par l'origine et inclinée par rapport à la diagonale du carré de côté 1 on soupçonnera une erreur sur les paramètres de dispersion et de position de la loi théorique.
- Si les points ne sont pas alignés sur une droite la loi théorique n'est pas adaptée aux observations.

#### Remarque

*Si l'on dispose des données individuelles d'une variable aléatoire continue, les modalités sont toutes les valeurs prises par la série (ordonnées) et les fréquences cumulées sont du type  $\frac{i}{n}$ .*

#### Remarque

*Les QQ-plot peuvent servir aussi à comparer les distributions de deux séries.*

## 2.3 Exercice

On considère la série de relevés de longueurs d'une série manufacturée de 500 pièces au dixième de millimètre.

$x_1$	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10	10.1	10.2	10.3	10.4	10.5	10.6	10.7	10.8	10.9
$n_i$	1	4	9	28	23	37	51	63	65	60	48	38	32	23	12	3	1	2

- Calculer  $m$  et  $\sigma$  les moyenne et écart-type de la série  $x$ .
- On souhaite savoir si cette série est proche d'une loi Gaussienne de moyenne  $m$  et écart-type  $\sigma$ . Tracer le QQ-plot et conclure.
- Vous enverrez par e-mail à l'adresse `julien.grepat@univ-grenoble-alpes.fr` un fichier contenant votre code R, vos interprétations et le graphe R.