Petits développement autour des estimateurs Statistiques

Il est très rare de trouver une explication claire des différentes constantes utilisées pour les calculs de capabilité. Très souvent on nous renvoie à des tables sans autre explication... Mon propos ici est de donner les grandes lignes qui permettent le calcul de ces constantes.

Bien sûr,il faut quelques pré-requis en mathématique, en particulier être familié avec la loi du χ^2 et la fonction Γ qui n'est rien d'autre qu'une extension de la factorielle aux complexes (Voir ici)

Soit $(X_i)_{1\leq i\leq n}$ n variables aléatoires indépendantes de même loi (cette hypothèse est essentielle on le verra par la suite))

On rappelle que la variable aléatoire égale à la moyenne est :

$$\overline{X} = \frac{X_1 + X_2 + ... + Xn}{n}$$

la variable aléatoire égale à la variance des n valeurs est :

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

Si X est distribuée selon la loi normale $\mathcal{N}(\mu;\sigma)$ alors $\overline{X} \hookrightarrow \mathcal{N}(\mu;\frac{\sigma}{\sqrt{n}})$
En effet : $E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \frac{n\mu}{n} = \mu$ et

$V(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$ (par indépendance) donc :

$V(\overline{X}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

Théorème Central limite : Si n est "grand" alors la loi de $\overline{X}$ se rapproche de la loi $\mathcal{N}(\mu;\frac{\sigma}{\sqrt{n}})$

Soient $(X_i)_{1\leq i\leq n}$ n variables iid suivant toutes la loi $\mathcal{N}(\mu;\sigma)$, alors :

$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \Longleftrightarrow \frac{nS^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2$
par conséquent $\frac{nS^2}{\sigma^2}$ suit la loi du $\chi^2(n-1)$ (Loi du chi2)

On sait que $E\left(\frac{nS^2}{\sigma^2}\right) = n-1$ une estimation ponctuelle de *la variance σ^2* est donc $S^2_{n-1} = \frac{nS^2}{n-1} = \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}$

Une autre approche serait d'écrire : $S^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2$ alors :

$$E(S^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\overline{X}^2) = E(X^2) - E(\overline{X}^2)$$
$$E(S^2) = V(X) + E(X)^2 - \left(V(\overline{X}) + E(\overline{X})^2\right) = V(X) - V(\overline{X})$$
$$E(S^2) = V(X) - \frac{1}{n}V(X) = \frac{n-1}{n}V(X) = \frac{n-1}{n}\sigma^2$$

et donc pour éviter le biais $\frac{n-1}{n}$ on prend pour estimateur ponctuel de la variance $S^2_{n-1} = \frac{nS^2}{n-1}$

Nous avons un estimateur sans biais de la variance mais il ne faut pas penser que sa racine carrée est un estimateur sans biais de l'écart-type...

En effet, posons $K = \frac{nS^2}{\sigma^2} \hookrightarrow \chi^2(n-1)$ de densité :

$$f(t) = \frac{e^{-\frac{t}{2}} t^{\frac{n-1}{2}-1}}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}$$

Alors

$$E(\sqrt{K}) = \int_0^{+\infty} \sqrt{t} f(t)dt = \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \Longleftrightarrow E(\sqrt{n}S) = \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\sigma$$

$$E\left(\sqrt{\frac{n}{n-1}}S\right) = \sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\sigma$$

On pose :

$$c_4(n) = \sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$$

on alors :

$$E\left(\frac{S_{n-1}}{c_4(n)}\right) = \sigma$$

On donc trouver un estimateur ponctuel de σ : $\hat{\sigma} = \frac{S_{n-1}}{c_4(n)}$

Estimateurs ponctuel de l'écart-type σ d'une population à partir de l'étendue R

Soient $(X_i)_{1\leq i\leq n}$ n variables aléatoires indépendantes suivant toutes la loi $\mathcal{N}(\mu;\sigma)$.
Les variables centrées réduites seront notées X_i^* de densité φ et de fonction de répartition Φ
L'étendue R (Range en anglais) est définie par :

$$R = Max_i(X_i) - Min_i(X_i)$$

On a :

$$\frac{R}{\sigma} = \frac{Max_i(X_i) - \mu}{\sigma} - \frac{Min_i(X_i) - \mu}{\sigma}$$

Considérons :
$M_n = \frac{Max_i(X_i)-\mu}{\sigma} = Max_i(X_i^*)$ et $m_n = \frac{Min_i(X_i)-\mu}{\sigma} = Min_i(X_i^*)$

La fonction de répartition de M_n est donnée pour tout réel t par :

$$P(M_n < t) = P\left(\bigcap_i X_i^* < t\right) = \prod_i P\left(X_i^* < t\right) = \left(\Phi(t)\right)^n$$

Ainsi, la fonction de densité de M_n, notée f_{M_n} est obtenue par dérivation :

$$f_{M_n}(t) = n\varphi(t) \times \left(\Phi(t)\right)^{n-1}$$

En outre, $m_n = Min_i(X_i^*) = -Max_i(-X_i^*)$ ainsi :

$\mathbb{E}\left(\frac{R}{\sigma}\right) = 2\mathbb{E}\left(M_n\right)$

On notera $d_2 = 2\mathbb{E}\left(M_n\right)$

et on a :

$$d_2 = 2 \int_{-\infty}^{+\infty} nt\varphi(t) \times (\Phi(t))^{n-1}\, dt$$

soit :

$$\mathbb{E}\left(\frac{R}{\sigma}\right) = d_2$$

Ainsi :

$$\hat{\sigma} = \frac{\bar{R}}{d_2}$$

d_2 et c_4 se calculeront numériquement ou nous utiliserons des tables