Normalité asymptotique du coefficient de corrélation empirique

À partir d'un n-échantillon $((X_1,Y_1),\ldots,(X_n,Y_n))$ d'une loi bidimensionnelle avec $\mathbb{E}_\theta[X_1]=m_X$, $\mathbb{E}_\theta[Y_1]=m_Y$, $V_\theta(X_1)=\sigma_X^2$, $V_\theta(Y_1)=\sigma_Y^2$ et le coefficient de corrélation $\rho=\rho(X_1,Y_1)$. On suppose l'existence des moments $\mathbb{E}_\theta[(X_1-m_X)^i(Y_1-m_Y)^j]$ pour $i+j\le 4$. L'estimateur naturel de ρ est le coefficient de corrélation empirique :

$$\widehat{\rho}_n=\frac{\frac{1}{n}\sum_i(X_i-\overline{X}_n)(Y_i-\overline{Y}_n)}{\sqrt{S_n^2(X)}\sqrt{S_n^2(Y)}}$$

Cet estimateur est asymptotiquement normal avec pour matrice de covariance asymptotique $J(\theta)\Sigma(\theta)J(\theta)^\top$ où

- $g(\theta_1,\theta_2,\theta_3)=\theta_1/\sqrt{\theta_2\theta_3}$ et
- $\Sigma(\theta)=V_\theta(((X_1-m_X)(Y_1-m_Y),(X_1-m_X)^2,(Y_1-m_Y)^2))$

La propriété est également valable pour l'estimateur n'employant que les estimateurs sans biais de la covariance et des variances.

Cas des EMM

- Rappelons que la consistance (forte) a été établie dans la Pro 16 sous les conditions d'existence et d'unicité de l'EMM $\widehat{\theta}_n=f(\overline{X}_n,\ldots\overline{X^k}_n)$ de $g(\theta)$, et de continuité en $(m_1,\ldots m_k)$ de f.
- Si f est de plus différentiable alors on a la normalité asymptotique

Théorème 17

Sous les conditions de la Pro 16, si f est de plus différentiable en $(m_1,\ldots m_k)$ alors $\widehat{\theta}_n$ est asymptotiquement normal.

Démonstration.

On a déjà mentionné la normalité asymptotique du vecteur $(\overline{X}_n,\ldots\overline{X^k}_n)$. Comme f est différentiable en $(m_1,\ldots m_k)$, la méthode delta (Th 13 ou Cor 7) donne la normalité asymptotique. □

Cas de EMV

- L'existence d'une suite de maxima locaux de la vraisemblance (fortement) consistante a été établie dans le Th 8. Sous des conditions renforcées, on obtient la normalité asymptotique de cette suite.

Théorème 18 (admis voir [Mon97, Th 7, p 109])

1. *Θ est un ouvert de $\mathbb{R}$*
2. *le modèle statistique est homogène*
3. *l'information de Fisher pour une observation $I_1(\theta)>0$ pour tout θ*
4. *la fonction de vraisemblance pour une observation $L_1(\theta,x)$ est 2 fois différentiable en θ et la dérivée partielles seconde $(\partial^2L_1/\partial\theta^2)(\theta,x)$ est continue en θ uniformément en x.*
5. *la condition (CD2) est satisfaite*

Alors toute suite $\widehat{\theta}_n$ de racines de l'équation de vraisemblance convergeant p.s. vers θ vérifie

$$\sqrt{n}(\widehat{\theta}_n-\theta)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,I_1(\theta)^{-1}).$$

Remarque 24

- Sous ces conditions, **l'EMV est asymptotiquement sans biais et asymptotiquement efficace**. Ces propriétés supportent l'idée que pour de grands échantillons, on est presque certain que la méthode du maximum de vraisemblance est la meilleure méthode d'estimation. Ceci explique la place prépondérante occupée par cette méthode dans l'approche « classique » des statistiques.
- Dans le cas des familles exponentielles identifiables, les conditions sont satisfaites.
- Si on s'intéresse à un estimateur de $g(\theta)$ alors

$$\sqrt{n}(g(\widehat{\theta}_n)-g(\theta))\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}\left(0,J(g)I_1(\theta)^{-1}J(g)^\top\right).$$

- Les outils essentiels de la preuve du Th 18 sont la formule de Taylor, la loi forte des grands nombres et le résultat de type Slutsky.

Estimation par région de confiance

On a mesuré la durée de vie en h de 10 ampoules identiques :

$$x_1 = 91.6,\ x_2 = 35.7,\ x_3 = 251.3,\ x_4 = 24.3,\ x_5 = 5.4,$$
$$x_6 = 67.3,\ x_7 = 170.9,\ x_8 = 9.5,\ x_9 = 118.4,\ x_{10} = 57.1$$

- **Modèle statistique :** les 10 données $x_1, \ldots, x_{10}$ sont les réalisations d'un n-échantillon $(X_1, \ldots, X_n)$ d'une loi donnée : dans ce contexte, une loi classique est $\mathrm{Exp}(\theta)$ où λ est un paramètre strictement positif **inconnu**, et on travaille avec un modèle statistique de 10-échantillon d'une loi $\mathbb{P}_\theta = \mathrm{Exp}(\theta)$
- On a proposé plusieurs méthodes pour proposer une valeur au paramètre θ, i.e. une estimation $\widehat{\theta}(x_1, \ldots, x_{10})$ de θ calculée à partir des seules observations $x_1, \ldots, x_{10}$
- Une série de critères qualitatifs associés à **l'estimateur** choisi, i.e. à la fonction du n-échantillon $\widehat{\theta}(X_1, \ldots, X_{10})$, a été proposée (exhaustivité, biais, diverses formes de consistance,. . .) :
- **Bilan :** on dispose du meilleur estimateur possible suivant les précédents critères et

$$\widehat{\theta}(x_1, \ldots, x_{10}) = \frac{1}{\overline{x}_{10}} = \frac{10}{\sum_{i=1}^{10} x_i} = \frac{1}{83.15} \approx 0,012$$

À un jeu de donnée, une valeur du paramètre : estimation dite ponctuelle du paramètre

Une alternative plus séduisante :

déterminer un ensemble (ou région) de valeurs du paramètre θ (sous-ensemble de Θ) « compatibles » avec le jeu d'observation avec un degré (niveau) de confiance $1 - \alpha$ élevé

Dans le cas des ampoules : on se fixe le niveau de confiance $1 - \alpha$ avec $\alpha \in]0, 1[$ petit, et

$$\underline{x} := (x_1, \ldots, x_{10}) \longrightarrow \mathrm{R}_{1-\alpha}(\underline{x}) := [f_1(\underline{x}), f_2(\underline{x})] \subset \Theta$$

tel que si $\underline{X} := (X_1, \ldots, X_{10})$

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta\Big\{\theta \in \mathrm{R}_{1-\alpha}(\underline{X}) = [f_1(\underline{X}), f_2(\underline{X})]\Big\} \geq 1 - \alpha$$

- Devrait s'appliquer à l'estimation de n'importe quelle fonction $g(\theta)$ de θ.
- On verra un point de vue « dual » avec la problématique des tests d'hypothèses

- Avec une convention déjà employée, un modèle statistique $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ désignera le modèle associé à une observation, ou un n-échantillon, ou à un échantillon de « taille infinie » suivant le contexte.
- $\underline{X}_n$ représentera le vecteur aléatoire $(X_1, \ldots, X_n)$ et sa réalisation $\underline{x}_n$.

Définition 25 (Région de confiance)

On appelle région de confiance (« aléatoire ») de niveau $1 - \alpha$ (pour $\alpha \in]0, 1[$) associée à $g(\theta)$, une famille de parties de $g(\Theta)$, notée $\mathbb{R}_{1-\alpha}(g(\theta))$, telle que

1. $\forall \theta \in \Theta$, $\{g(\theta) \in \mathrm{R}_{1-\alpha}(g(\theta))\} := \{\underline{x}_n \mid g(\theta) \in \mathrm{R}_{1-\alpha}(g(\theta))\} \subset \otimes\mathscr{E}$
2. $\forall \theta \in \Theta$,

$$\mathbb{P}_\theta\{g(\theta) \in \mathrm{R}_{1-\alpha}(g(\theta))\} = 1 - \alpha \tag{40}$$

La dépendance en le n-échantillon $\underline{X}_n$ d'une région de confiance n'est plus mise en évidence par la notation.

Remarque 25

1. En général, on définit des régions de confiance associées à un niveau de confiance au moins $1 - \alpha$, c'est à dire que l'égalité dans (40) est remplacée par l'inégalité large $\geq$.
2. Vu le temps restant, on s'intéressera dans le cours qu'au cas unidimensionnel, i.e. d'une fonction $g(\theta) \subset \mathbb{R}$. Alors $\mathbb{R}_{1-\alpha}(g(\theta)) = [f_1(\underline{X}_n), f_2(\underline{X}_n)]$ est appelé un **intervalle de confiance au niveau** $1 - \alpha$, ce que l'on notera également $\mathrm{IC}_{1-\alpha}(g(\theta))$.
3. On verra sur un exemple quel est « le prix » à payer si l'on veut toujours se restreindre au cas unidimensionnel dans un contexte d'estimation conjointe de plusieurs paramètres. Penser au cas de l'estimation du vecteur des coefficients de régression où les estimateurs des paramètres présentent une corrélation.

On peut distinguer deux classes de régions de confiance

1. les régions dites **exactes** car s'appuyant sur des calculs exacts de lois : cela concerne (principalement) le cas fondamental des échantillons gaussiens
2. les régions dites **asymptotiques** car s'appuyant sur des résultats asymptotiques (en la taille de l'échantillon) sur les lois : le cœur de leur construction s'appuiera essentiellement sur des résultats de **normalité asymptotique**
   - ↪ Ce second cas de figure est le plus courant pour des échantillons non-gaussiens

Interprétation pratique d'une région de confiance

L'interprétation suivante d'une $\mathrm{R}_{1-\alpha}(g(\theta))$ peut être justifiée avec la LFGN :

- Supposons connue la vraie valeur du paramètre θ, donc de $g(\theta)$. n est fixé.
- Supposons que l'on obtienne une séquence indépendante de réalisations d'un n-échantillon de la loi $\mathbb{P}_\theta$
  - $x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}$
  - $x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)}$
  - $x_1^{(3)}, x_2^{(3)}, \ldots, x_n^{(3)}$
  - $\ldots$
- Pour chaque jeu de n-observations, on calcule la région de confiance.
- Si cette séquence d'échantillons indépendants est assez longue (i.e. on fait tendre k vers l'infini) alors $(1-\alpha)\%$ des intervalles obtenus contiendront la vraie valeur de $g(\theta)$.

Mais : pour la région calculée à partir d'un échantillon particulier, on ne peut pas savoir si elle « fait partie » des $(1-\alpha)\%$ de régions contenant la vraie valeur de $g(\theta)$ ou des $\alpha\%$ « mauvaises régions »

Construction d'intervalles de confiance de niveau $1-\alpha$: en général les trois étapes

1. Déterminer le meilleur estimateur possible $\widehat{g(\theta)}_n$ du paramètre $g(\theta)$. La loi de probabilité de $\widehat{g(\theta)}_n$ est parfois appelée la loi d'échantillonnage.
2. Détermination d'une seconde fonction $D_n(\theta)$ de l'échantillon, dépendant également du paramètre $g(\theta)$, telle que sa **loi de probabilité** sous $\mathbb{P}_\theta$ soit indépendante de θ (fonction pivot) et on choisit les deux constantes $d_{1,1-\alpha}, d_{2,1-\alpha}$
$$\mathbb{P}_\theta\{d_{1,1-\alpha} \leq D_n(g(\theta)) \leq d_{2,1-\alpha}\} = 1-\alpha$$
3. En déduire l'intervalle de valeurs de $g(\theta)$ de probabilité $1-\alpha$ par une construction de type « inversion des inégalités », comme par exemple
$$\mathbb{P}_\theta\left\{(D_n)^{-1}(d_{1,1-\alpha}) \leq g(\theta) \leq (D_n)^{-1}(d_{2,1-\alpha})\right\} = 1-\alpha$$

Définition 26 (fonction pivot)

Une fonction pivot $D_n(g(\theta))$ pour $g(\theta)$ est une fonction définie sur $E_X^n \times g(\Theta)$ à valeurs dans un espace mesurable telle que

1. *$\forall\theta \in \Theta$, la fonction $\underline{x}_n \mapsto D_n(g(\theta))$ est mesurable*
2. *la loi de $D_n(g(\theta))$ ne dépend pas de $\theta \in \Theta$*

Remarque 26

1. Noter qu'une fonction pivot dépend fonctionnellement de θ. En ce sens, il ne s'agit pas d'une statistique et en particulier, il ne s'agit pas d'une statistique libre même si l'on requiert la même condition sur sa loi.
2. La première condition est juste là pour garantir que la fonction pivot est une variable aléatoire.

Cas des échantillons gaussiens

1. **Estimation par intervalle de confiance de la moyenne** m : notre meilleur estimateur est $\overline{X}_n$
2. **Estimation par intervalle de confiance de la variance** σ^2 : notre meilleur estimateur (si m est inconnu) est

$$\widetilde{S}_n^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X}_n)^2$$

Propriétés fondamentales des n-échantillons gaussiens

1. Soit $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_1, \sigma_1^2)$ telles que $X_1 \perp\!\!\!\perp X_2$ alors

$$a_1 X_1 + a_2 X_2 \sim \mathcal{N}\left(m = a_1 m_1 + a_2 m_2, \sigma^2 = (a_1)^2\sigma_1^2 + (a_2)^2\sigma_2^2\right)$$

2. $\overline{X}_n \perp\!\!\!\perp \widetilde{S}_n^2, S_n^2$ et $(n-1)\widetilde{S}_n^2 = nS_n^2 \sim \chi^2_{n-1}$
3. $\left\|\left(\frac{X_i - m}{\sigma}\right)_{i=1}^n\right\|_2^2 = \sum_{i=1}^n \left(\frac{X_i-m}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - m)^2 \sim \chi^2_n$

Estimation de la moyenne d'un modèle $\mathcal{N}(m, \sigma^2)$ I

1. **Supposons la variance σ^2 connue :**

$$\overline{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \tag{41}$$

et alors on obtient la fonction pivot

$$D_n(m) = \frac{\overline{X}_n - m}{\sigma/\sqrt{n}} = \sqrt{n}\left[\frac{\overline{X}_n - m}{\sigma}\right] \sim \mathcal{N}(0,1)$$

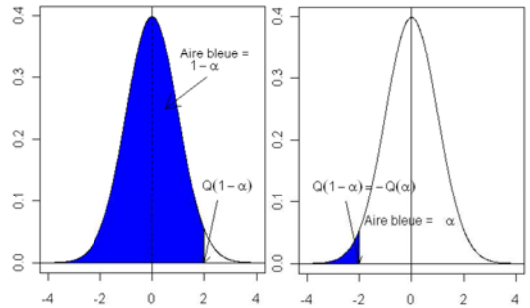


FIGURE 3: Quantiles d'ordre α et $1-\alpha$ pour $\mathcal{N}(0,1)$

Estimation de la moyenne d'un modèle $\mathcal{N}(m, \sigma^2)$ II

d'où les trois choix raisonnables :

$$\mathbb{P}_\theta\left\{z_{\alpha/2} \le D_n(m) \le z_{1-\alpha/2}\right\} = 1-\alpha$$

$$\mathbb{P}_\theta\{D_n(m) \le z_{1-\alpha}\} = 1-\alpha \qquad \mathbb{P}_\theta\{z_\alpha \le D_n(m)\} = 1-\alpha$$

où $z_{\alpha/2}, z_\alpha, z_{1-\alpha/2}, z_{1-\alpha}$ sont les quantiles d'ordre $\alpha/2, \alpha, 1-\alpha/2, 1-\alpha$ d'une $\mathcal{N}(0,1)$

Proposition 21 ($\mathrm{IC}_{1-\alpha}(m)$ d'une loi $\mathcal{N}(m, \sigma^2)$ avec σ^2 connue)

(a) *bilatéral* : $\mathrm{IC}_{1-\alpha}(m) = [\ \overline{X}_n \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\ ]$

(b) *unilatéral supérieur* : $\mathrm{IC}_{1-\alpha}(m) = [\ \overline{X}_n - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, +\infty[$

(c) *unilatéral inférieur* : $\mathrm{IC}_{1-\alpha}(m) =]-\infty, \overline{X}_n + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\ ]$

Estimation de la moyenne d'un modèle $\mathcal{N}(m, \sigma^2)$ III

Un exemple

- Supposons que le rendement (en quintal/ha) d'une parcelle de blé est connu comme suivant une loi normale $\mathcal{N}(m, \sigma^2 = 1)$.
- Sur un échantillon de taille $n = 4$ parcelles de blé, on a observé les rendements suivant en quintal/ha : $\{79, 79, 80, 82\}$: $\overline{X}_4 = 80$
- On choisit un niveau de 0.95, i.e ; $\alpha = 0.05$.
- L'estimation par intervalle (bilatéral) du rendement potentiel moyen d'une parcelle de blé est

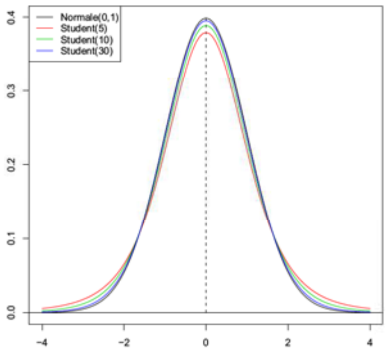
$$\mathrm{IC}_{0.95}(m) = [80 \pm 1.96\frac{1}{\sqrt{4}}] \approx [79, 81]$$

Cas d'une variable X de loi gaussienne $\mathcal{N}(m,\sigma^2)$ IV

2 **La variance σ^2 est inconnue :** on l'estime par $\widetilde{S}_n^2$(ou éventuellement S_n^2) alors

$$\widetilde{D}_n(m) = \sqrt{n}\left[\frac{\overline{X}_n - m}{\sqrt{\widetilde{S}_n^2}}\right] \sim st_{n-1} \qquad D_n(m) = \sqrt{n-1}\left[\frac{\overline{X}_n - m}{\sqrt{S_n^2}}\right] \sim st_{n-1}$$

où st_{n-1} désigne une loi de Student à $(n-1)$ degrés de liberté.



Estimation de la moyenne d'un modèle $\mathcal{N}(m,\sigma^2)$ V

Proposition 22 ($\mathrm{IC}_{1-\alpha}(m)$ d'une loi $\mathcal{N}(m,\sigma^2)$ avec σ^2 inconnue)

Un intervalle de confiance $\mathrm{IC}_{1-\alpha}(m)$ ***au niveau*** $1-\alpha$:

(a) *bilatéral :* $\mathrm{IC}_{1-\alpha}(m) = [\ \overline{X}_n \pm t_{n-1,1-\alpha/2}\frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}\ ]$

(b) *unilatéral supérieur :* $\mathrm{IC}_{1-\alpha}(m) = [\ \overline{X}_n - t_{n-1,1-\alpha}\frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}, +\infty[$

(c) *unilatéral inférieur :* $\mathrm{IC}_{1-\alpha}(m) =]-\infty, \overline{X}_n + t_{n-1,1-\alpha}\frac{\sqrt{\widetilde{S}_n^2}}{\sqrt{n}}\ ]$

où $t_{n-1,1-\alpha/2}, t_{n-1,1-\alpha}$ sont les quantiles d'ordre $\alpha, 1-\alpha/2, 1-\alpha$ d'une loi de Student à $(n-1)$ degrés de liberté.
Dans le cas où on emploie S_n^2 comme estimateur de la variance, on remplace $\widetilde{S}_n^2$ par S_n^2 et $\sqrt{n}$ par $\sqrt{n-1}$ dans les intervalles de confiance ci-dessus.

En pratique, si $n > 30$ alors la loi de Student à n degrés de liberté peut être approchée par une loi normale centrée-réduite (TCL). Dans, ce cas $t_{n-1,\alpha} \approx z_\alpha$ pour tout α, on retrouve les intervalles du cas variance connue.

Estimation de la moyenne d'un modèle $\mathcal{N}(m,\sigma^2)$ VI

Reprise du précédent exemple

- Supposons que le rendement (en quintal/ha) d'une parcelle de blé comme suivant une loi normale $\mathcal{N}(m,\sigma^2)$ où les deux paramètres sont inconnus.
- Sur un échantillon de taille $n=4$ parcelles de blé, on a observé les rendements suivant en quintal/ha : $\{79,79,80,82\}$: $\overline{X}_4 = 80$ et $\widetilde{s}_n^2 = 2$
- On choisit $\alpha = 0.05$: $t_{n-1,1-\alpha/2} = t_{3,0.975} = 3.182$
- L'estimation par intervalle (bilatéral) du rendement potentiel moyen d'une parcelle de blé est

$$\mathrm{IC}_{0.95}(m) = [80 \pm 3.182\frac{\sqrt{2}}{\sqrt{4}}] \approx [77.75, 82.25]$$

L'IC est plus large que dans le cas σ^2 connue et ceci pour deux raisons

- ici $\widetilde{s}_n^2 = 2 > \sigma^2 = 1$
- et on a toujours $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$ pour les faibles valeurs de α. Cette différence est d'autant plus importante que n est petit

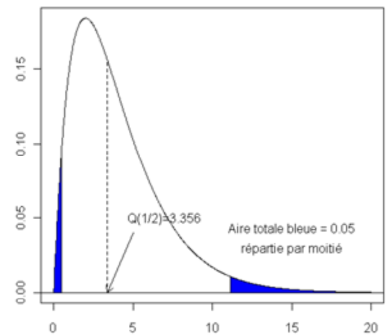
Estimation de la variance d'un modèle $\mathcal{N}(m,\sigma^2)$ I

1 **La moyenne m est connue :** le meilleur estimateur de σ^2 est alors

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - m)^2$$

et

$$D_n(\sigma^2) = \frac{n \times \widehat{\sigma^2}}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - m}{\sigma}\right)^2 \sim \chi_n^2$$



Estimation de la variance d'un modèle $\mathcal{N}(m,\sigma^2)$ II

On a

$$\mathbb{P}_\theta\left\{x_{n,\alpha/2} \le D_n(\sigma^2) = \frac{n \times \widehat{\sigma^2}}{\sigma^2} \le x_{n,1-\alpha/2}\right\} = 1-\alpha$$

Proposition 23 ($\mathrm{IC}_{1-\alpha}(\sigma^2)$ pour une loi $\mathcal{N}(m,\sigma^2)$ avec m connue)

$$\mathrm{IC}_{1-\alpha}(\sigma^2) = \left[\frac{n\,\widehat{\sigma^2}}{x_{n,1-\alpha/2}}, \frac{n\,\widehat{\sigma^2}}{x_{n,\alpha/2}}\right]$$

où $x_{n,\alpha/2}, x_{n,1-\alpha/2}$ sont les quantiles d'ordre $\alpha/2, 1-\alpha/2$ d'une loi du chi-deux à n degrés de liberté.
Une démarche similaire permet de construire les intervalles unilatéraux.

Estimation de la variance d'un modèle $\mathcal{N}(m,\sigma^2)$ III

2. **La moyenne m est inconnue :** on peut utiliser l'un des deux estimateurs de la variance

$$S_n^2 = \frac{1}{n}\sum_i (X_i - \overline{X}_n)^2 \qquad \widetilde{S}_n^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X}_n)^2$$

et

$$D_n(\sigma^2) = \frac{n \times S_n^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{x_i - \overline{X}_n}{\sigma}\right)^2 = \frac{(n-1)\times \widetilde{S}_n^2}{\sigma^2} = \widetilde{D}_n(\sigma^2) \sim \chi^2_{n-1}$$

Proposition 24 ($\mathrm{IC}_{1-\alpha}(\sigma^2)$ d'une loi $\mathcal{N}(m,\sigma^2)$ avec m inconnue)

$$\mathrm{IC}_{1-\alpha}(\sigma^2) = \left[\frac{n\,S_n^2}{x_{n-1,1-\alpha/2}}, \frac{n\,S_n^2}{x_{n-1,\alpha/2}}\right] = \left[\frac{(n-1)\,\widetilde{S}_n^2}{x_{n-1,1-\alpha/2}}, \frac{(n-1)\,\widetilde{S}_n^2}{x_{n-1,\alpha/2}}\right]$$

où $x_{n-1,\alpha/2}, x_{n-1,1-\alpha/2}$ sont les quantiles d'ordre $\alpha/2, 1-\alpha/2$ d'une loi du chi-deux à $(n-1)$ degrés de liberté.

Un exemple

Sur un échantillon de $n=6$ parcelles, on a observé l'écart-type $\tilde{s}_n^2 = 1$ sur le rendement en blé.

- On a donc $(n-1)\tilde{s}_n^2 = 5$.
- Avec un niveau de 0.95, on $x_{5,0.025} \approx 0.831$ et $x_{5,0.975} \approx 12.83$
- l'intervalle de confiance vaut

$$\mathrm{IC}_{0.95}(\sigma^2) = \left[\frac{5}{12.83}, \frac{5}{0.831}\right] \approx [0.39, 6.02]$$

et pour l'écart-type, on peut proposer

$$\mathrm{IC}_{0.95}(\sigma) = \left[\sqrt{\frac{5}{12.83}}, \sqrt{\frac{5}{0.831}}\right] \approx [0.625, 2.45]$$

Notons que cet estimateur de l'écart-type n'est pas le meilleur possible.

Une région de confiance pour (m,σ^2) de $\mathcal{N}(m,\sigma^2)$

1. Utiliser les résultats précédents sur les $\mathrm{IC}_{1-\alpha_0}(m)$ (avec σ^2 inconnu) et $\mathrm{IC}_{1-(\alpha_1+\alpha_2)}(\sigma^2)$ (avec m inconnu) :

$$\mathbb{P}_\theta(A\cap B) = 1 - \mathbb{P}_\theta(\overline{A}\cup\overline{A}) \ge 1 - \mathbb{P}_\theta(\overline{A}) - \mathbb{P}_\theta(\overline{B})$$

d'où pour tout $\theta \in \Theta = \mathbb{R}\times]0,+\infty[$:

$$\mathbb{P}_\theta\left\{-t_{n-1,1-\frac{\alpha_0}{2}} \le D_n(m) \le t_{n-1,1-\frac{\alpha_0}{2}},\ x_{n-1,\alpha_1} \le D_n(\sigma^2) \le x_{n-1,1-\alpha_2}\right\} \ge 1-\alpha_0-(\alpha_1+\alpha_2) = 1-\alpha$$

si $\alpha_0+\alpha_1+\alpha_2 = \alpha$.
Ceci justifie l'utilisation du rectangle suivant comme région de confiance au niveau $1-(\alpha_0+\alpha_1+\alpha_2)$:

$$\mathrm{IC}_{1-(\alpha_0+\alpha_1+\alpha_2)}(m,\sigma^2) = \mathrm{IC}_{1-\alpha_0}(m)\times \mathrm{IC}_{1-(\alpha_1+\alpha_2)}(\sigma^2)$$