Séances 5 et 6 Probabilités et statistiques Image SCIA EPITA

Plan

- Comparaison d'estimateurs
- 2 Intervalle de confiance pour la moyenne *m*
 - Cas $1:\sigma^2$ connu (variance connue)
 - Cas $2:\sigma^2$ inconnu
- $oldsymbol{3}$ Intervalle de confiance pour la variance σ^2
 - Cas 1: m connu
 - Cas 2: m inconnu

But : déterminer les estimateurs sans biais dont la variance sera la plus faible.

Théorème

Sous les hypothèses du cours (cf chapitre 2), soit T un estimateur sans biais de $g(\theta)$ de variance finie.

On a:

- $g(\theta)$ est dérivable.
- pour tout θ ,

$$V(T) \geq \frac{g'^2(\theta)}{I_n(\theta)}$$

La borne $\frac{g'^2(\theta)}{I_n(\theta)}$ est appelée borne de Fréchet ou borne de Cramer-Rao.

Définition

Un estimateur sans biais T de $g(\theta)$ est dit efficace lorsque la borne de Fréchet est atteinte i.e.

$$V(T) = \frac{g'^2(\theta)}{I_n(\theta)}$$

- Deux types d'estimation :
- estimation ponctuelle,
- estimation par intervalle.

- Deux résultats probabilistes :
- loi forte des grands nombres,
- théorème central limite.

Point de départ

- (X_1, \ldots, X_n) échantillon i.i.d. de taille n,
- (x_1,\ldots,x_n) réalisations de cet échantillon,
- $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ estimation ponctuelle de la moyenne (espérance) m,
- $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i \bar{x}_n)^2$ estimation (ponctuelle) de la variance σ^2 .



- En général, la loi de probabilité associée à une expérience aléatoire n'est pas connue.
- ② Dans le modèle de Bernoulli, i.e. avec un échantillon i.i.d. de loi $\mathcal{B}(p)$.
- Intervalle de confiance au niveau de confiance 0,95 :

$$\left[f-\frac{1}{\sqrt{n}};f+\frac{1}{\sqrt{n}}\right]$$

Cet intervalle fournit un encadrement de la valeur réelle de p.

Théorème

Pour 95% des échantillons, la proportion p appartient à cet intervalle sous les conditions suivantes :

- **1** $n \ge 30$,
- **2** $nf \ge 5$,
- **③** n(1-f) ≥ 5.

- Niveaux de confiance fréquents : 95% et 99%,
- Risques associés : 5% et 1%.

- Deux cas :
- n quelconque : variables aléatoires normales,
- n grand et utilisation du TCL.

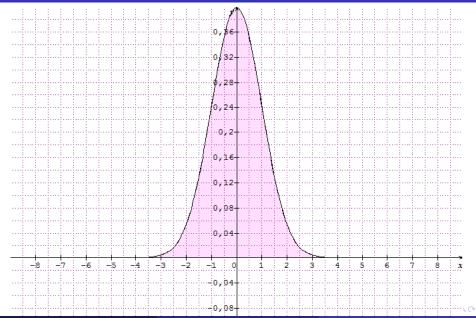
Théorème central limite

- Soit (X_i) une suite de v.a. i.i.d. telle que $E(X_1^2) < +\infty$.
- ② Notons $m := E(X_i)$ et $\sigma^2 = V(X_i)$.
- 3 $\frac{\sqrt{n}(\bar{X}_n m)}{\sigma}$ converge en loi vers une loi normale centrée réduite.

Loi normale centrée réduite

- $\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = 0, 5,$
- **3** $\mathbb{P}(-1,96 \le X \le 1,96) \approx 0,95$ et $\mathbb{P}(-2,58 \le X \le 2,58) \approx 0,99$.

Loi normale centrée réduite



Cas gaussien

- Si X₁ suit une loi normale,
- ② Pour tout entier $n \ge 1$, $\frac{\sqrt{n}(\bar{X}_n m)}{\sigma}$ suit une loi normale centrée réduite,

$$m \in \left[\bar{X}_n - 1,96\frac{\sigma}{\sqrt{n}}; \bar{X}_n + 1,96\frac{\sigma}{\sqrt{n}}\right]$$

au niveau de confiance 0,95.

A partir d'observations (x_1, \ldots, x_n) de moyenne \bar{x} , l'intervalle de confiance au niveau 0,95 pour la moyenne m est :

$$\left[\bar{x}_n - 1,96\frac{\sigma}{\sqrt{n}}; \bar{x}_n + 1,96\frac{\sigma}{\sqrt{n}}\right]$$

Cas général

- X₁ non nécessairement gaussienne,
- $E(X_1^2) < +\infty$.

Grâce au TCL,

$$m \in \left[\bar{X}_n - 1,96\frac{\sigma}{\sqrt{n}}; \bar{X}_n + 1,96\frac{\sigma}{\sqrt{n}}\right]$$

au niveau de confiance 0,95.

A partir d'observations (x_1, \ldots, x_n) de moyenne \bar{x} , l'intervalle de confiance asymptotique au niveau 0,95 pour la moyenne m est :

$$\left[\bar{x}_n - 1,96\frac{\sigma}{n}; \bar{x}_n + 1,96\frac{\sigma}{n}\right]$$

Exercice : Proposer un intervalle de confiance asymptotique au niveau 0,90 pour la moyenne *m* d'une variable aléatoire.

Tableau N [1]

Aire sous la courbe normale à gauche de z, c'est à dire $P[Z \leq z]$, ou $Z \sim N(0;1)$.



| | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| .60 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | .9292 | .9306 | .9319 |
| | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |

F.L. 2006 C Tableau construit avec SAS, Metapost et ConTgX

Plus généralement, l'intervalle de confiance (asymptotique) au niveau $1-\alpha$ pour la moyenne m est :

$$\left[\bar{x}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{x}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

où $z_{1-rac{lpha}{2}}$ désigne le fractile d'ordre $1-rac{lpha}{2}$ de la loi $\mathcal{N}(0,1).$

Cas particulier du modèle de Bernoulli

- intervalle de confiance pour la proportion de personnes porteuses d'un caractère dans une population donnée,
- variance inconnue toutefois l'approximation par la loi normale est possible d'après le théorème suivant :

Théorème (Théorème de Moivre-Laplace)

Soit X_n une variable aléatoire suivant une loi $\mathcal{B}(n,p)$.

Posons q := 1 - p.

Pour tout réel x,

$$\lim_{n \to +\infty} \mathbb{P}\left(\frac{X_n - np}{\sqrt{npq}} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = F(x)$$

où F désigne la fonction de répartition de la loi normale centrée réduite.

Intervalle de confiance pour la proportion p

L'intervalle de confiance asymptotique au niveau $1-\alpha$ pour la proportion p est :

$$\left[\hat{\rho}-z_{1-\frac{\alpha}{2}}\frac{\sqrt{\hat{\rho}(1-\hat{\rho})}}{\sqrt{n}};\hat{\rho}+z_{1-\frac{\alpha}{2}}\frac{\sqrt{\hat{\rho}(1-\hat{\rho})}}{\sqrt{n}}\right]$$

où $z_{1-\frac{\alpha}{2}}$ désigne le fractile d'ordre $1-\frac{\alpha}{2}$.

Exercice

François prélève 300 serpents dans une forêt et constate que 70 d'entre eux sont venimeux.

Déterminer un intervalle de confiance asymptotique pour la proportion de serpents venimeux dans cette forêt au niveau de confiance 0,95.

Rappelons les conditions d'application du résultat :

- **1** n > 30,
- $n\hat{p} \geq 5,$
- $n(1-\hat{p}) \geq 5.$

- En remarquant que $\sqrt{\hat{\rho}(1-\hat{\rho})} \leq \frac{1}{2}$, on a :
- 2 $1,96\sqrt{\hat{p}(1-\hat{p})} < 1$ d'où l'intervalle de l'introduction :

Loi du Khi-deux

Soient (X_1, \ldots, X_n) la donnée de n variables aléatoires indépendantes normales centrées réduites.

La variable aléatoire $U_n := \sum_{i=1}^n X_i^2$ suit une loi du Khi-deux à n degrés de

liberté notée $\chi^2(n)$.

Propriétés

- **1** densité : $f_{U_n}(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}e^{-\frac{x}{2}}x^{\frac{n}{2}-1}$ pour $x \ge 0$,
- $E(U_n) = n,$
- $V(U_n) = 2n,$
- $\ \, \textbf{ 1} \ \, \text{fonction caractéristique} : \phi_{U_n}(t) = \frac{1}{(1-2it)^{\frac{n}{2}}}.$

Théorème

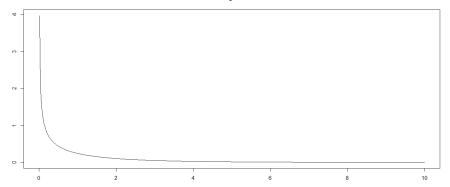
Soient X et Y deux variables aléatoires indépendantes suivant respectivement des lois $\chi^2(m)$ et $\chi^2(n)$ alors la variable aléatoire X+Y suit une loi $\chi^2(m+n)$.

Proposition

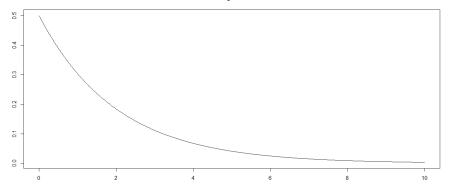
La loi $\chi^2(n)$ est égale à la loi $\Gamma\left(\frac{n}{2},\frac{1}{2}\right)$.

(cf Compléments sur la loi Gamma)

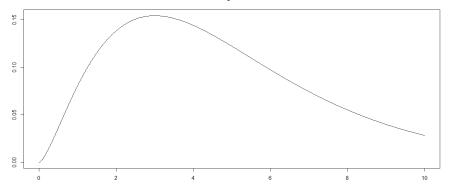
Chi2 à 1 degré de liberté



Chi2 à 2 degrés de liberté



Chi2 à 5 degrés de liberté



Loi de Student

Soient X et Y deux variables aléatoires indépendantes suivant respectivement des lois $\mathcal{N}(0,1)$ et $\chi^2(n)$.

La variable $T_n = \frac{X}{\sqrt{\frac{Y}{n}}}$ suit une loi de Student à n degrés de liberté. Elle

est notée \mathcal{T}_n

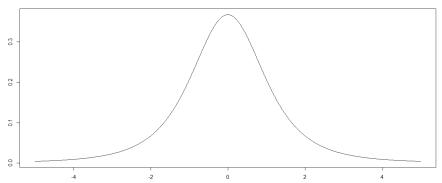


Propriétés

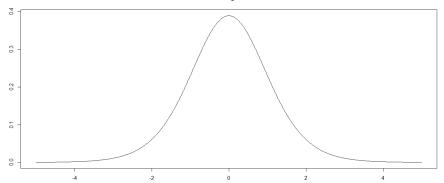
- $E(T_n) = 0$ (symétrie comme la loi normale centrée réduite),
- ② $V(T_n) = \frac{n}{n-2}$ pour n > 2.



Student à 3 degrés de liberté



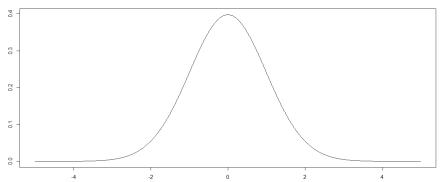
Student à 10 degrés de liberté



Théorème

 T_n converge en loi vers $\mathcal{N}(0,1)$ lorsque n tend vers $+\infty$.

Student à 100 degrés de liberté



Rappel

Théorème

$$S_n^2:=rac{1}{n-1}\sum_{i=1}^n (X_i-ar{X}_n)^2$$
 est un estimateur sans biais de σ^2 .



Cas gaussien

- \bullet X_1 suit une loi normale.
- $T_n := \frac{\sqrt{n}(\bar{X}_n m)}{\sqrt{S_n^2}}$ suit une loi de Student à n-1 degrés de liberté.

L'intervalle de confiance au niveau $1-\alpha$ pour la moyenne m est :

$$\left[\bar{x}_n-t_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}};\bar{x}_n+t_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}}\right]$$

où $t_{1-\frac{\alpha}{2}}$ désigne le fractile d'ordre $1-\frac{\alpha}{2}$ de la loi de Student à n-1 degrés de liberté.

Tableau T1 [1/2]

Tableau de t^* tel qu'une variable de Student à dl degrés de liberté ait probabilité p d'être supérieure à t^*



| | | | | | | P T> | t^*] = p | | | | | |
|----|-------|-------|-------|-------|-------|-------|-------------|-------|-------|--------|-------|--------|
| dl | 0.25 | 0.2 | 0.15 | 0.1 | 0.05 | 0.025 | 0.02 | 0.01 | 0.005 | 0.0025 | 0.001 | 0.0005 |
| 1 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.71 | 15.89 | 31.82 | 63.66 | 127.3 | 318.3 | 636.6 |
| 2 | .8165 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 4.849 | 6.965 | 9.925 | 14.09 | 22.33 | 31.60 |
| 3 | .7649 | .9785 | 1.250 | 1.638 | 2.353 | 3.182 | 3.482 | 4.541 | 5.841 | 7.453 | 10.21 | 12.92 |
| 4 | .7407 | .9410 | 1.190 | 1.533 | 2.132 | 2.776 | 2.999 | 3.747 | 4.604 | 5.598 | 7.173 | 8.610 |
| 5 | .7267 | .9195 | 1.156 | 1.476 | 2.015 | 2.571 | 2.757 | 3.365 | 4.032 | 4.773 | 5.893 | 6.869 |
| 6 | .7176 | .9057 | 1.134 | 1.440 | 1.943 | 2.447 | 2.612 | 3.143 | 3.707 | 4.317 | 5.208 | 5.959 |
| 7 | .7111 | .8960 | 1.119 | 1.415 | 1.895 | 2.365 | 2.517 | 2.998 | 3.499 | 4.029 | 4.785 | 5.408 |
| 8 | .7064 | .8889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.449 | 2.896 | 3.355 | 3.833 | 4.501 | 5.041 |
| 9 | .7027 | .8834 | 1.100 | 1.383 | 1.833 | 2.262 | 2.398 | 2.821 | 3.250 | 3.690 | 4.297 | 4.781 |
| 10 | .6998 | .8791 | 1.093 | 1.372 | 1.812 | 2.228 | 2.359 | 2.764 | 3.169 | 3.581 | 4.144 | 4.587 |
| 11 | .6974 | .8755 | 1.088 | 1.363 | 1.796 | 2.201 | 2.328 | 2.718 | 3.106 | 3.497 | 4.025 | 4.437 |
| 12 | .6955 | .8726 | 1.083 | 1.356 | 1.782 | 2.179 | 2.303 | 2.681 | 3.055 | 3.428 | 3.930 | 4.318 |
| 13 | .6938 | .8702 | 1.079 | 1.350 | 1.771 | 2.160 | 2.282 | 2.650 | 3.012 | 3.372 | 3.852 | 4.221 |
| 14 | .6924 | .8681 | 1.076 | 1.345 | 1.761 | 2.145 | 2.264 | 2.624 | 2.977 | 3.326 | 3.787 | 4.140 |
| 15 | .6912 | .8662 | 1.074 | 1.341 | 1.753 | 2.131 | 2.249 | 2.602 | 2.947 | 3.286 | 3.733 | 4.073 |
| 16 | .6901 | .8647 | 1.071 | 1.337 | 1.746 | 2.120 | 2.235 | 2.583 | 2.921 | 3.252 | 3.686 | 4.015 |
| 17 | .6892 | .8633 | 1.069 | 1.333 | 1.740 | 2.110 | 2.224 | 2.567 | 2.898 | 3.222 | 3.646 | 3.965 |
| 18 | .6884 | .8620 | 1.067 | 1.330 | 1.734 | 2.101 | 2.214 | 2.552 | 2.878 | 3.197 | 3.610 | 3.922 |
| | 0.25 | 0.2 | 0.15 | 0.1 | 0.05 | 0.025 | 0.02 | 0.01 | 0.005 | 0.0025 | 0.001 | 0.0005 |

F.L. 2006 (C) Tableau construit avec SAS, Metapost et ConTgX

Exercice

Déterminer l'intervalle de confiance au niveau 0,9 pour la moyenne m d'une variable aléatoire gaussienne en supposant $n=15, \bar{x}=30$ et $s^2=25$.

• X_1 suit une loi normale.

•
$$S_n^{2\star} := \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$$

• $\frac{nS_n^{2\star}}{\sigma^2}$ suit une loi $\chi^2(n)$.

$$\bullet \ \mathbb{P}\left(\chi_{\frac{\alpha}{2}}^2 \le \frac{nS_n^{2\star}}{\sigma^2} \le \chi_{1-\frac{\alpha}{2}}^2\right) = 1 - \alpha$$

- où $\chi^2_{\frac{\alpha}{2}}$ désigne le fractile d'ordre $\frac{\alpha}{2}$ de la loi $\chi^2(n)$.
- La loi $\chi^2(n)$ n'est pas symétrique!

L'intervalle de confiance au niveau $1-\alpha$ pour la variance σ^2 est :

$$\left[n\frac{s_n^{2\star}}{\chi_{1-\frac{\alpha}{2}}^2}; n\frac{s_n^{2\star}}{\chi_{\frac{\alpha}{2}}^2}\right]$$

- X_1 suit une loi normale.
- \bar{X}_n est un estimateur sans biais de m.

•
$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$
,

•
$$\frac{(n-1)S_n^2}{\sigma^2}$$
 suit une loi $\chi^2(n-1)$.

$$\bullet \ \mathbb{P}\left(\chi_{\frac{\alpha}{2}}^2 \le \frac{nS_n^2}{\sigma^2} \le \chi_{1-\frac{\alpha}{2}}^2\right) = 1 - \alpha$$

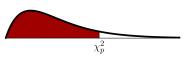
• où $\chi^2_{\frac{\alpha}{2}}$ désigne le fractile d'ordre $\frac{\alpha}{2}$ de la loi $\chi^2(n-1)$.

L'intervalle de confiance au niveau 1-lpha pour la variance σ^2 est :

$$\left[(n-1)\frac{s_n^2}{\chi_{1-\frac{\alpha}{2}}^2};(n-1)\frac{s_n^2}{\chi_{\frac{\alpha}{2}}^2}\right]$$

Tableau C [1/2]

Percentiles de la distribution du $\chi^2.$ Valeurs de χ^2_P correspondant à P



| dl | $\chi^{2}_{0.005}$ | $\chi^{2}_{0.01}$ | $\chi^{2}_{0.025}$ | $\chi^{2}_{0.05}$ | $\chi^{2}_{0.1}$ | $\chi^{2}_{0.9}$ | $\chi^{2}_{0.95}$ | $\chi^{2}_{0.975}$ | $\chi^{2}_{0.99}$ | $\chi^2_{0.995}$ |
|----|--------------------|-------------------|--------------------|-------------------|------------------|------------------|-------------------|--------------------|-------------------|--------------------|
| 1 | .0000 | .0002 | .0010 | .0039 | .0158 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 |
| 2 | .0100 | .0201 | .0506 | .1026 | .2107 | 4.605 | 5.991 | 7.378 | 9.210 | 10.60 |
| 3 | .0717 | .1148 | .2158 | .3518 | .5844 | 6.251 | 7.815 | 9.348 | 11.34 | 12.84 |
| 4 | .2070 | .2971 | .4844 | .7107 | 1.064 | 7.779 | 9.488 | 11.14 | 13.28 | 14.86 |
| 5 | .4117 | .5543 | .8312 | 1.145 | 1.610 | 9.236 | 11.07 | 12.83 | 15.09 | 16.75 |
| 6 | .6757 | .8721 | 1.237 | 1.635 | 2.204 | 10.64 | 12.59 | 14.45 | 16.81 | 18.55 |
| 7 | .9893 | 1.239 | 1.690 | 2.167 | 2.833 | 12.02 | 14.07 | 16.01 | 18.48 | 20.28 |
| 8 | 1.344 | 1.646 | 2.180 | 2.733 | 3.490 | 13.36 | 15.51 | 17.53 | 20.09 | 21.95 |
| 9 | 1.735 | 2.088 | 2.700 | 3.325 | 4.168 | 14.68 | 16.92 | 19.02 | 21.67 | 23.59 |
| 10 | 2.156 | 2.558 | 3.247 | 3.940 | 4.865 | 15.99 | 18.31 | 20.48 | 23.21 | 25.19 |
| 11 | 2.603 | 3.053 | 3.816 | 4.575 | 5.578 | 17.28 | 19.68 | 21.92 | 24.72 | 26.76 |
| 12 | 3.074 | 3.571 | 4.404 | 5.226 | 6.304 | 18.55 | 21.03 | 23.34 | 26.22 | 28.30 |
| 13 | 3.565 | 4.107 | 5.009 | 5.892 | 7.042 | 19.81 | 22.36 | 24.74 | 27.69 | 29.82 |
| 14 | 4.075 | 4.660 | 5.629 | 6.571 | 7.790 | 21.06 | 23.68 | 26.12 | 29.14 | 31.32 |
| 15 | 4.601 | 5.229 | 6.262 | 7.261 | 8.547 | 22.31 | 25.00 | 27.49 | 30.58 | 32.80 |
| 16 | 5.142 | 5.812 | 6.908 | 7.962 | 9.312 | 23.54 | 26.30 | 28.85 | 32.00 | 34.27 |
| 17 | 5.697 | 6.408 | 7.564 | 8.672 | 10.09 | 24.77 | 27.59 | 30.19 | 33.41 | 35.72 |
| 18 | 6.265 | 7.015 | 8.231 | 9.390 | 10.86 | 25.99 | 28.87 | 31.53 | 34.81 | 37.16 |
| dl | $\chi^{2}_{0.005}$ | $\chi^{2}_{0.01}$ | $\chi^{2}_{0.025}$ | $\chi^{2}_{0.05}$ | $\chi^{2}_{0.1}$ | $\chi^{2}_{0.9}$ | $\chi^{2}_{0.95}$ | $\chi^{2}_{0.975}$ | $\chi^{2}_{0.99}$ | $\chi^{2}_{0.995}$ |

F.L. 2006 (C) Tableau construit avec SAS, Metapost et ConTgX

Exercice

Déterminer l'intervalle de confiance au niveau 0,95 pour la variance σ^2 d'une variable aléatoire gaussienne en supposant n=17 et $s_n^2=40$.