Méthode des percentiles

- $\widehat{\theta}$ estimateur de θ (par exemple EMV)

Méthode des percentiles

- $\widehat{\theta}$ estimateur de θ (par exemple EMV)
- B échantillons bootstrap ($B \geq 1000$)
  $\widehat{\theta}_1^*, \ldots, \widehat{\theta}_B^*$

Méthode des percentiles

- $\widehat{\theta}$ estimateur de θ (par exemple EMV)
- B échantillons bootstrap ($B \geq 1000$)
  $\widehat{\theta}_1^*, \ldots, \widehat{\theta}_B^*$
- Réplications ordonnées : $\widehat{\theta}_{(1)}^* \leq \cdots \leq \widehat{\theta}_{(B)}^*$

Méthode des percentiles

- $\widehat{\theta}$ estimateur de θ (par exemple EMV)
- B échantillons bootstrap ($B \geq 1000$)

  $\widehat{\theta}_1^*, \ldots, \widehat{\theta}_B^*$
- Réplications ordonnées : $\widehat{\theta}_{(1)}^* \leq \cdots \leq \widehat{\theta}_{(B)}^*$
- IC des percentiles de $100 \cdot (1 - 2\alpha)\%$: $[\widehat{\theta}_B^{*(\alpha)}, \widehat{\theta}_B^{*(1-\alpha)}]$

  Les bornes sont les réplications de $\widehat{\theta}^*$ de rangs k et $B + 1 - k$, où $k = \lfloor (B+1)\alpha \rfloor$ (partie entière).

Méthode des percentiles

- $\widehat{\theta}$ estimateur de θ (par exemple EMV)
- B échantillons bootstrap ($B \geq 1000$)

  $\widehat{\theta}^*_1, \ldots, \widehat{\theta}^*_B$
- Réplications ordonnées : $\widehat{\theta}^*_{(1)} \leq \cdots \leq \widehat{\theta}^*_{(B)}$
- IC des percentiles de $100 \cdot (1 - 2\alpha)\%$: $[\widehat{\theta}^{*(\alpha)}_B, \widehat{\theta}^{*(1-\alpha)}_B]$

  Les bornes sont les réplications de $\widehat{\theta}^*$ de rangs k et $B + 1 - k$, où $k = \lfloor (B+1)\alpha \rfloor$ (partie entière).
- n grand, $\widehat{\theta}$ est souvent approx. normal (IC standard proche IC des percentiles). Lorsque n petit, la loi de $\widehat{\theta}$ peut être très asymétrique. (Exemple : r) S'il existe une transformation stabilisante et symétrisante ϕ, la méthode des percentiles est applicable.

Méthode des percentiles (suite)

- Données `mouse.t` : groupe traité de 7 souris

  > mouse.t   #durées de vie en jours après une chirurgie

  [1] 94 197 16 38 99 141 23

  $\overline{x} = 86.86$, $\sqrt{v_{\text{Boot}}(\overline{x})} = 25.23549$

Méthode des percentiles (suite)

- Données `mouse.t` : groupe traité de 7 souris

  > mouse.t   #durées de vie en jours après une chirurgie

  [1] 94 197 16 38 99 141 23

  $\overline{x} = 86.86$, $\sqrt{v_{\text{Boot}}(\overline{x})} = 25.23549$

- 1000 réplications bootstrap (non paramétriques) de $\overline{x}$

  > theta.etoile = bootstrap(mouse.t,1000,mean)[[1]]

Méthode des percentiles (suite)

- Données `mouse.t` : groupe traité de 7 souris

  > mouse.t #durées de vie en jours après une chirurgie

  [1] 94 197 16 38 99 141 23

  $\overline{x} = 86.86$, $\sqrt{v_{\text{Boot}}(\overline{x})} = 25.23549$
- 1000 réplications bootstrap (non paramétriques) de $\overline{x}$

  > theta.etoile = bootstrap(mouse.t,1000,mean)[[1]]
- IC standard de 90%

$$[86.86 \pm 1.645 \cdot 25.24] = [45.34, 128.37]$$

Méthode des percentiles (suite)

- Données `mouse.t` : groupe traité de 7 souris

  > mouse.t #durées de vie en jours après une chirurgie

  [1] 94 197 16 38 99 141 23

  $\overline{x} = 86.86$, $\sqrt{v_{\text{Boot}}(\overline{x})} = 25.23549$

- 1000 réplications bootstrap (non paramétriques) de $\overline{x}$

  > theta.etoile = bootstrap(mouse.t,1000,mean)[[1]]

- IC standard de 90%

$$[86.86 \pm 1.645 \cdot 25.24] = [45.34, 128.37]$$

- 5^e et 95^e percentiles des 1000 réplications $\overline{x}^*$

  > quantile(theta.etoile,c(.05,.95))

  5% 95%

  48.0000 129.6214 #IC de 90% des percentiles

Méthode des percentiles (suite)

- $x_1, \ldots, x_{10}$ tirés de $N(0,1)$, $\theta = e^{\mu} = 1$ à estimer par IC

  $\widehat{\theta} = e^{\overline{x}} \quad \Longrightarrow$ il existe une transformation ϕ stabilisante et symétrisante $\Longrightarrow$ la méthode des percentiles s'applique

Méthode des percentiles (suite)

- $x_1, \ldots, x_{10}$ tirés de $N(0, 1)$, $\theta = e^{\mu} = 1$ à estimer par IC
  $\widehat{\theta} = e^{\overline{x}} \quad \Longrightarrow$ il existe une transformation ϕ stabilisante et symétrisante $\Longrightarrow$ la méthode des percentiles s'applique
- 1000 réplications bootstrap (non paramétriques) de $\widehat{\theta}$
  > x = rnorm(10) ; theta = function(x) exp(mean(x))
  > y = bootstrap(x,1000,theta)[[1]]

Calcul des intervalles de confiance

- quantiles 0.025 et 0.975 #IC de 95%

  > quantile(y,c(.025,.975))

  2.5% 97.5%

  0.4377564 2.2845846

Calcul des intervalles de confiance

- quantiles 0.025 et 0.975 #IC de 95%

  > quantile(y,c(.025,.975))

  2.5% 97.5%

  0.4377564 2.2845846

- IC standard de 95%

  > bootstrap(x,1000,theta,func=sd)[[2]]

  [1] 0.5058786 #estimation écart type de $e^{\overline{x}}$

  $[e^{\overline{x}} \pm 1.96 \cdot 0.51] = [-0.04, 1.96]$

Calcul des intervalles de confiance

- quantiles 0.025 et 0.975 #IC de 95%

  > quantile(y,c(.025,.975))

  2.5% 97.5%

  0.4377564 2.2845846

- IC standard de 95%

  > bootstrap(x,1000,theta,func=sd)[[2]]

  [1] 0.5058786 #estimation écart type de $e^{\overline{x}}$

  $[e^{\overline{x}} \pm 1.96 \cdot 0.51] = [-0.04, 1.96]$

- On veut estimer $\theta = e^{\mu} > 0$, et l'intervalle standard contient des valeurs négatives ! Cela ne peut jamais se produire avec la méthode des percentiles.

Méthode BC_a

- Basée sur l'hypothèse d'existence de ϕ cr. telle que

$$\frac{\phi(\widehat{\theta}) - \phi(\theta)}{1 + a\phi(\theta)} + z_0 \sim N(0,1),$$

a constante dite d'accélération, z_0 : constante de correction du biais

Méthode BC_a

- Basée sur l'hypothèse d'existence de ϕ cr. telle que

$$\frac{\phi(\widehat{\theta}) - \phi(\theta)}{1 + a\phi(\theta)} + z_0 \sim N(0,1),$$

  a constante dite d'accélération, z_0 : constante de correction du biais

- $1 + a\phi(\theta)$ joue le rôle de $\sqrt{\text{Var}(\phi(\widehat{\theta}))}$

Méthode BC_a

- Basée sur l'hypothèse d'existence de ϕ cr. telle que

$$\frac{\phi(\widehat{\theta}) - \phi(\theta)}{1 + a\phi(\theta)} + z_0 \sim N(0,1),$$

a constante dite d'accélération, z_0 : constante de correction du biais

- $1 + a\phi(\theta)$ joue le rôle de $\sqrt{\text{Var}(\phi(\widehat{\theta}))}$
- IC BC_a : $[\widehat{\theta}^{*(\alpha_1)}, \widehat{\theta}^{*(\alpha_2)}]$

Méthode BC_a

- Basée sur l'hypothèse d'existence de ϕ cr. telle que

$$\frac{\phi(\widehat{\theta}) - \phi(\theta)}{1 + a\phi(\theta)} + z_0 \sim N(0,1),$$

a constante dite d'accélération, z_0 : constante de correction du biais

- $1 + a\phi(\theta)$ joue le rôle de $\sqrt{\text{Var}(\phi(\widehat{\theta}))}$
- IC BC_a : $[\widehat{\theta}^{*(\alpha_1)}, \widehat{\theta}^{*(\alpha_2)}]$
- $\alpha_1 = \Phi\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z_\alpha}{1 - \widehat{a}(\widehat{z}_0 + z_\alpha)}\right)$, $\alpha_2 = \Phi\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z_{1-\alpha}}{1 - \widehat{a}(\widehat{z}_0 + z_{1-\alpha})}\right)$

Méthode BC_a

- Basée sur l'hypothèse d'existence de ϕ cr. telle que

$$\frac{\phi(\widehat{\theta})-\phi(\theta)}{1+a\phi(\theta)}+z_0\sim N(0,1),$$

a constante dite d'accélération, z_0 : constante de correction du biais

- $1+a\phi(\theta)$ joue le rôle de $\sqrt{\text{Var}(\phi(\widehat{\theta}))}$
- IC BC_a : $[\widehat{\theta}^{*(\alpha_1)},\widehat{\theta}^{*(\alpha_2)}]$
- $\alpha_1=\Phi\left(\widehat{z}_0+\frac{\widehat{z}_0+z_\alpha}{1-\widehat{a}(\widehat{z}_0+z_\alpha)}\right)$, $\alpha_2=\Phi\left(\widehat{z}_0+\frac{\widehat{z}_0+z_{1-\alpha}}{1-\widehat{a}(\widehat{z}_0+z_{1-\alpha})}\right)$
- $\widehat{z}_0=\Phi^{-1}\left(\frac{\#\{\widehat{\theta}_b^*<\widehat{\theta}\}}{B}\right)$, $\quad\widehat{a}=\frac{\sum_1^n(\widehat{\theta}_{(\cdot)}-\widehat{\theta}_{(i)})^3}{6(\sum_1^n(\widehat{\theta}_{(\cdot)}-\widehat{\theta}_{(i)})^2)^{3/2}}$

Méthode BC_a (données)

- Tableau `spatial` dans le package `bootstrap`. Variables A et B : tests de perception spatiale chez $n = 26$ enfants handicapés

Méthode BC_a (données)

- Tableau `spatial` dans le package `bootstrap`. Variables A et B : tests de perception spatiale chez $n = 26$ enfants handicapés
- > x = spatial[,1] #variable A

Méthode BC_a (données)

- Tableau `spatial` dans le package `bootstrap`. Variables A et B : tests de perception spatiale chez $n = 26$ enfants handicapés
- > x = spatial[,1] #variable A
- La méthode BC_a sera appliquée à un estimateur biaisé :

$$\widehat{\theta} = \frac{\sum_i (X_i - \overline{X})^2}{26} \quad \text{(estimateur biaisé de } \theta = \sigma^2\text{)}$$

> va = function(x) var(x)*(length(x)-1)/length(x)

Méthode BC_a (données)

- Tableau `spatial` dans le package `bootstrap`. Variables A et B : tests de perception spatiale chez $n = 26$ enfants handicapés
- > x = spatial[,1] #variable A
- La méthode BC_a sera appliquée à un estimateur biaisé :

$$\widehat{\theta} = \frac{\sum_i (X_i - \overline{X})^2}{26} \quad \text{(estimateur biaisé de } \theta = \sigma^2\text{)}$$

> va = function(x) var(x)*(length(x)-1)/length(x)

- 2000 réplications bootstrap (non paramétriques) de $\widehat{\theta}$

> y = bootstrap(x,2000,va)[[1]]

Méthode BC_a (données)

- Tableau `spatial` dans le package `bootstrap`. Variables A et B : tests de perception spatiale chez $n = 26$ enfants handicapés
- > x = spatial[,1] #variable A
- La méthode BC_a sera appliquée à un estimateur biaisé :

$$\widehat{\theta} = \frac{\sum_i (X_i - \overline{X})^2}{26} \quad \text{(estimateur biaisé de } \theta = \sigma^2\text{)}$$

> va = function(x) var(x)*(length(x)-1)/length(x)

- 2000 réplications bootstrap (non paramétriques) de $\widehat{\theta}$
  > y = bootstrap(x,2000,va)[[1]]
- > va(x) #$\widehat{\theta} =$ 171.5340

Méthode BC_a (fonction `bcanon`)

- Fonction `bcanon` du package `bootstrap`

  ```
  > bcanon(x, nboot, theta, ...,
  alpha=c(0.025, 0.05, 0.1, 0.16, 0.84,
  0.9, 0.95, 0.975))
  ```

Méthode BC_a (fonction `bcanon`)

- Fonction `bcanon` du package `bootstrap`

  `> bcanon(x, nboot, theta, ..., alpha=c(0.025, 0.05, 0.1, 0.16, 0.84, 0.9, 0.95, 0.975))`

- `x` : vecteur des données (cas univarié)

  `nboot` : nombre d'échantillons bootstrap

  `theta` : estimateur ponctuel

  `...` : autres arguments possibles de `theta`

  `alpha` : quantiles spécifiant les niveaux de confiance

Méthode BC_a (fonction `bcanon`)

- Fonction `bcanon` du package `bootstrap`

  ```
  > bcanon(x, nboot, theta, ...,
  alpha=c(0.025, 0.05, 0.1, 0.16, 0.84,
  0.9, 0.95, 0.975))
  ```

- `x` : vecteur des données (cas univarié)

  `nboot` : nombre d'échantillons bootstrap

  `theta` : estimateur ponctuel

  `...` : autres arguments possibles de `theta`

  `alpha` : quantiles spécifiant les niveaux de confiance

- Résultat : liste à 5 composantes dont les 3 premières :

  `confpoint` : bornes des IC

  `z0` : correction pour le biais estimée

  `acc` : accélération estimée

Méthode BC_a appliquée

- Application à l'objet x (valeurs de A)

  > bcanon(x,1000,va,alpha=c(.05,.95))   #IC de 90%

  $confpoints

  alpha bca point

  [1,] 0.05 115.9822

  [2,] 0.95 266.4630

  $z0 # correction du biais estimée

  [1] 0.1967796

  $acc #constante d'accélération estimée

  [1] 0.06124012

Méthode BC_a appliquée

- Application à l'objet x (valeurs de A)

  > bcanon(x,1000,va,alpha=c(.05,.95))   #IC de 90%

  $confpoints

  alpha bca point

  [1,] 0.05 115.9822

  [2,] 0.95 266.4630

  $z0 # correction du biais estimée

  [1] 0.1967796

  $acc #constante d'accélération estimée

  [1] 0.06124012

- L'IC BC_a de 90% pour σ^2 est donc

$$[115.98, 266.46].$$

Intervalles non paramétriques

- 4 IC bootstrap non paramétriques de 90% pour σ^2 `#va(x)=171.53` ($\widehat{\theta}$ estimateur ponctuel biaisé)

Méthode	0.05	0.95	longueur	forme
standard	101.59	241.48	139.89	1.00
boostrap-t	117.69	297.37	179.68	2.34
percentile	95.01	233.90	138.89	0.81
BC_a	115.98	266.46	150.48	1.71

Intervalles non paramétriques

- 4 IC bootstrap non paramétriques de 90% pour σ^2
  #va(x)=171.53 ($\widehat{\theta}$ estimateur ponctuel biaisé)

Méthode	0.05	0.95	longueur	forme
standard	101.59	241.48	139.89	1.00
boostrap-t	117.69	297.37	179.68	2.34
percentile	95.01	233.90	138.89	0.81
BC_a	115.98	266.46	150.48	1.71

- sd(y) #$\sqrt{v_{\text{Boot}}^{(2000)}(\widehat{\theta})} =$ 42.52 (écart type, méth. standard)

Intervalles non paramétriques

- 4 IC bootstrap non paramétriques de 90% pour σ^2
`#va(x)=171.53` ($\widehat{\theta}$ estimateur ponctuel biaisé)

Méthode	0.05	0.95	longueur	forme
standard	101.59	241.48	139.89	1.00
boostrap-t	117.69	297.37	179.68	2.34
percentile	95.01	233.90	138.89	0.81
BC_a	115.98	266.46	150.48	1.71

- sd(y) #$\sqrt{v^{(2000)}_{\text{Boot}}(\widehat{\theta})}$ = 42.52 (écart type, méth. standard)
- forme $= (BS - \widehat{\theta})/(\widehat{\theta} - BI) \quad \widehat{\theta} = \sum_i (X_i - \overline{X})^2/n$

Intervalles non paramétriques

- 4 IC bootstrap non paramétriques de 90% pour σ^2
  #va(x)=171.53 ($\widehat{\theta}$ estimateur ponctuel biaisé)

Méthode	0.05	0.95	longueur	forme
standard	101.59	241.48	139.89	1.00
boostrap-t	117.69	297.37	179.68	2.34
percentile	95.01	233.90	138.89	0.81
BC_a	115.98	266.46	150.48	1.71

- sd(y) #$\sqrt{v^{(2000)}_{\text{Boot}}(\widehat{\theta})} = 42.52$ (écart type, méth. standard)
- forme $= (BS - \widehat{\theta})/(\widehat{\theta} - BI) \quad \widehat{\theta} = \sum_i (X_i - \overline{X})^2/n$
- Efron : la théorie indique qu'en bootstrap non paramétrique la méthode BC_a est supérieure.