

Estimation et Intervalles de Confiance

On a un estimateur X pour μ .

Mesure de **qualité**: $\text{Eff}[X] = 1/[C(X)\text{MSE}(X)]$.

Estimation et Intervalles de Confiance

On a un estimateur X pour μ .

Mesure de **qualité**: $\text{Eff}[X] = 1/[C(X)\text{MSE}(X)]$.

Toute mesure de qualité est imparfaite ou incomplète.

Estimation et Intervalles de Confiance

On a un estimateur X pour μ .

Mesure de **qualité**: $\text{Eff}[X] = 1/[C(X)\text{MSE}(X)]$.

Toute mesure de qualité est imparfaite ou incomplète.

$\text{Eff}[X]$ suppose que le coût de l'erreur est symétrique et proportionnel à son carré.

Estimation et Intervalles de Confiance

On a un estimateur X pour μ .

Mesure de **qualité**: $\text{Eff}[X] = 1/[C(X)\text{MSE}(X)]$.

Toute mesure de qualité est imparfaite ou incomplète.

$\text{Eff}[X]$ suppose que le coût de l'erreur est symétrique et proportionnel à son carré.

Autre aspect important que $\text{Eff}[X]$ ne mesure pas: la disponibilité d'une bonne façon d'évaluer erreur d'estimation.

Estimation et Intervalles de Confiance

On a un estimateur X pour μ .

Mesure de qualité: $\text{Eff}[X] = 1/[C(X)\text{MSE}(X)]$.

Toute mesure de qualité est imparfaite ou incomplète.

$\text{Eff}[X]$ suppose que le coût de l'erreur est symétrique et proportionnel à son carré.

Autre aspect important que $\text{Eff}[X]$ ne mesure pas: la disponibilité d'une bonne façon d'évaluer erreur d'estimation.

Par exemple, si on estime cette erreur par la variance de X , il nous faut un bon estimateur de cette variance.

Estimation et Intervalles de Confiance

On a un estimateur X pour μ .

Mesure de **qualité**: $\text{Eff}[X] = 1/[C(X)\text{MSE}(X)]$.

Toute mesure de qualité est imparfaite ou incomplète.

$\text{Eff}[X]$ suppose que le coût de l'erreur est symétrique et proportionnel à son carré.

Autre aspect important que $\text{Eff}[X]$ ne mesure pas: la disponibilité d'une bonne façon d'**évaluer erreur d'estimation**.

Par exemple, si on estime cette erreur par la variance de X , il nous faut un bon estimateur de cette variance.

L'évaluation de l'erreur d'estimation est habituellement fournie en donnant un **intervalle de confiance** (IC) pour μ .

Estimation et Intervalles de Confiance

On a un estimateur X pour μ .

Mesure de **qualité**: $\text{Eff}[X] = 1/[C(X)\text{MSE}(X)]$.

Toute mesure de qualité est imparfaite ou incomplète.

$\text{Eff}[X]$ suppose que le coût de l'erreur est symétrique et proportionnel à son carré.

Autre aspect important que $\text{Eff}[X]$ ne mesure pas: la disponibilité d'une bonne façon d'**évaluer erreur d'estimation**.

Par exemple, si on estime cette erreur par la variance de X , il nous faut un bon estimateur de cette variance.

L'évaluation de l'erreur d'estimation est habituellement fournie en donnant un **intervalle de confiance** (IC) pour μ .

$[I_1, I_2]$ est un IC de niveau $1 - \alpha$ (ou à $100(1 - \alpha)\%$) pour μ si

$$\mathbb{P}[I_1 \leq \mu \leq I_2] = 1 - \alpha.$$

Estimation et Intervalles de Confiance

On a un estimateur X pour μ .

Mesure de **qualité**: $\text{Eff}[X] = 1/[C(X)\text{MSE}(X)]$.

Toute mesure de qualité est imparfaite ou incomplète.

$\text{Eff}[X]$ suppose que le coût de l'erreur est symétrique et proportionnel à son carré.

Autre aspect important que $\text{Eff}[X]$ ne mesure pas: la disponibilité d'une bonne façon d'**évaluer erreur d'estimation**.

Par exemple, si on estime cette erreur par la variance de X , il nous faut un bon estimateur de cette variance.

L'évaluation de l'erreur d'estimation est habituellement fournie en donnant un **intervalle de confiance** (IC) pour μ .

$[I_1, I_2]$ est un IC de niveau $1 - \alpha$ (ou à $100(1 - \alpha)\%$) pour μ si

$$\mathbb{P}[I_1 \leq \mu \leq I_2] = 1 - \alpha.$$

Habituellement, on construit un IC pour un niveau **visé** ou **nominal** $1 - \alpha$, mais la vraie **probabilité de couverture** est différente et inconnue.

Estimation et Intervalles de Confiance

On a un estimateur X pour μ .

Mesure de **qualité**: $\text{Eff}[X] = 1/[C(X)\text{MSE}(X)]$.

Toute mesure de qualité est imparfaite ou incomplète.

$\text{Eff}[X]$ suppose que le coût de l'erreur est symétrique et proportionnel à son carré.

Autre aspect important que $\text{Eff}[X]$ ne mesure pas: la disponibilité d'une bonne façon d'**évaluer erreur d'estimation**.

Par exemple, si on estime cette erreur par la variance de X , il nous faut un bon estimateur de cette variance.

L'évaluation de l'erreur d'estimation est habituellement fournie en donnant un **intervalle de confiance** (IC) pour μ .

$[I_1, I_2]$ est un IC de niveau $1 - \alpha$ (ou à $100(1 - \alpha)\%$) pour μ si

$$\mathbb{P}[I_1 \leq \mu \leq I_2] = 1 - \alpha.$$

Habituellement, on construit un IC pour un niveau **visé** ou **nominal** $1 - \alpha$, mais la vraie **probabilité de couverture** est différente et inconnue.

La différence $\mathbb{P}[I_1 \leq \mu \leq I_2] - (1 - \alpha)$ est l'**erreur de couverture**.

Estimation et Intervalles de Confiance

On a un estimateur X pour μ .

Mesure de **qualité**: $\text{Eff}[X] = 1/[C(X)\text{MSE}(X)]$.

Toute mesure de qualité est imparfaite ou incomplète.

$\text{Eff}[X]$ suppose que le coût de l'erreur est symétrique et proportionnel à son carré.

Autre aspect important que $\text{Eff}[X]$ ne mesure pas: la disponibilité d'une bonne façon d'**évaluer erreur d'estimation**.

Par exemple, si on estime cette erreur par la variance de X , il nous faut un bon estimateur de cette variance.

L'évaluation de l'erreur d'estimation est habituellement fournie en donnant un **intervalle de confiance** (IC) pour μ .

$[I_1, I_2]$ est un IC de niveau $1 - \alpha$ (ou à $100(1 - \alpha)\%$) pour μ si

$$\mathbb{P}[I_1 \leq \mu \leq I_2] = 1 - \alpha.$$

Habituellement, on construit un IC pour un niveau **visé** ou **nominal** $1 - \alpha$, mais la vraie **probabilité de couverture** est différente et inconnue.

La différence $\mathbb{P}[I_1 \leq \mu \leq I_2] - (1 - \alpha)$ est l'**erreur de couverture**.

La **largeur** de l'intervalle est $I_2 - I_1$ (une variable aléatoire).

La **largeur** de l'intervalle est $I_2 - I_1$ (une variable aléatoire).

Idéalement, on veut (a) la bonne couverture et (b) $\mathbb{E}[I_2 - I_1]$ et $\text{Var}[I_2 - I_1]$ petits.

La **largeur** de l'intervalle est $I_2 - I_1$ (une variable aléatoire).

Idéalement, on veut (a) la bonne couverture et (b) $\mathbb{E}[I_2 - I_1]$ et $\text{Var}[I_2 - I_1]$ petits.

Je dénote parfois une **suite** d'estimateurs $\{Y_n, n \geq 1\}$ par Y_n (abus de notation).

La **largeur** de l'intervalle est $I_2 - I_1$ (une variable aléatoire).

Idéalement, on veut (a) la bonne couverture et (b) $\mathbb{E}[I_2 - I_1]$ et $\text{Var}[I_2 - I_1]$ petits.

Je dénote parfois une **suite** d'estimateurs $\{Y_n, n \geq 1\}$ par Y_n (abus de notation).
Exemples: \bar{X}_n et S_n^2 .

La **largeur** de l'intervalle est $I_2 - I_1$ (une variable aléatoire).

Idéalement, on veut (a) la bonne couverture et (b) $\mathbb{E}[I_2 - I_1]$ et $\text{Var}[I_2 - I_1]$ petits.

Je dénote parfois une **suite** d'estimateurs $\{Y_n, n \geq 1\}$ par Y_n (abus de notation).
Exemples: \bar{X}_n et S_n^2 .

Lorsque $n \rightarrow \infty$, Y_n est dit **asymptotiquement sans biais** si $\mathbb{E}[Y_n - \mu] \rightarrow 0$,

La **largeur** de l'intervalle est $I_2 - I_1$ (une variable aléatoire).

Idéalement, on veut (a) la bonne couverture et (b) $\mathbb{E}[I_2 - I_1]$ et $\text{Var}[I_2 - I_1]$ petits.

Je dénote parfois une **suite** d'estimateurs $\{Y_n, n \geq 1\}$ par Y_n (abus de notation).
Exemples: \bar{X}_n et S_n^2 .

Lorsque $n \rightarrow \infty$, Y_n est dit **asymptotiquement sans biais** si $\mathbb{E}[Y_n - \mu] \rightarrow 0$,
consistant si $Y_n \rightarrow \mu$ en probabilité, i.e. $\mathbb{P}[|Y_n - \mu| > \epsilon] \rightarrow 0$ pour tout $\epsilon > 0$.

La **largeur** de l'intervalle est $I_2 - I_1$ (une variable aléatoire).

Idéalement, on veut (a) la bonne couverture et (b) $\mathbb{E}[I_2 - I_1]$ et $\text{Var}[I_2 - I_1]$ petits.

Je dénote parfois une **suite** d'estimateurs $\{Y_n, n \geq 1\}$ par Y_n (abus de notation).
Exemples: \bar{X}_n et S_n^2 .

Lorsque $n \rightarrow \infty$, Y_n est dit **asymptotiquement sans biais** si $\mathbb{E}[Y_n - \mu] \rightarrow 0$,
consistant si $Y_n \rightarrow \mu$ en probabilité, i.e. $\mathbb{P}[|Y_n - \mu| > \epsilon] \rightarrow 0$ pour tout $\epsilon > 0$.
et **fortement consistant** si $Y_n \rightarrow \mu$ avec probabilité 1.

La **largeur** de l'intervalle est $I_2 - I_1$ (une variable aléatoire).

Idéalement, on veut (a) la bonne couverture et (b) $\mathbb{E}[I_2 - I_1]$ et $\text{Var}[I_2 - I_1]$ petits.

Je dénote parfois une **suite** d'estimateurs $\{Y_n, n \geq 1\}$ par Y_n (abus de notation).
Exemples: \bar{X}_n et S_n^2 .

Lorsque $n \rightarrow \infty$, Y_n est dit **asymptotiquement sans biais** si $\mathbb{E}[Y_n - \mu] \rightarrow 0$,
consistant si $Y_n \rightarrow \mu$ en probabilité, i.e. $\mathbb{P}[|Y_n - \mu| > \epsilon] \rightarrow 0$ pour tout $\epsilon > 0$.
et **fortement consistant** si $Y_n \rightarrow \mu$ avec probabilité 1.

Exemples: \bar{X}_n et S_n^2 pour $\mu = \mathbb{E}[X_i]$.

La **largeur** de l'intervalle est $I_2 - I_1$ (une variable aléatoire).

Idéalement, on veut (a) la bonne couverture et (b) $\mathbb{E}[I_2 - I_1]$ et $\text{Var}[I_2 - I_1]$ petits.

Je dénote parfois une **suite** d'estimateurs $\{Y_n, n \geq 1\}$ par Y_n (abus de notation).
Exemples: \bar{X}_n et S_n^2 .

Lorsque $n \rightarrow \infty$, Y_n est dit **asymptotiquement sans biais** si $\mathbb{E}[Y_n - \mu] \rightarrow 0$,
consistant si $Y_n \rightarrow \mu$ en probabilité, i.e. $\mathbb{P}[|Y_n - \mu| > \epsilon] \rightarrow 0$ pour tout $\epsilon > 0$.
et **fortement consistant** si $Y_n \rightarrow \mu$ avec probabilité 1.

Exemples: \bar{X}_n et S_n^2 pour $\mu = \mathbb{E}[X_i]$.

Un IC $(I_{n,1}, I_{n,2})$ est **asymptotiquement valide** si son erreur de couverture converge vers 0, i.e., si $\lim_{n \rightarrow \infty} \mathbb{P}[I_{n,1} \leq \mu \leq I_{n,2}] = (1 - \alpha)$.

La **largeur** de l'intervalle est $I_2 - I_1$ (une variable aléatoire).

Idéalement, on veut (a) la bonne couverture et (b) $\mathbb{E}[I_2 - I_1]$ et $\text{Var}[I_2 - I_1]$ petits.

Je dénote parfois une **suite** d'estimateurs $\{Y_n, n \geq 1\}$ par Y_n (abus de notation).
Exemples: \bar{X}_n et S_n^2 .

Lorsque $n \rightarrow \infty$, Y_n est dit **asymptotiquement sans biais** si $\mathbb{E}[Y_n - \mu] \rightarrow 0$,
consistant si $Y_n \rightarrow \mu$ en probabilité, i.e. $\mathbb{P}[|Y_n - \mu| > \epsilon] \rightarrow 0$ pour tout $\epsilon > 0$.
et **fortement consistant** si $Y_n \rightarrow \mu$ avec probabilité 1.

Exemples: \bar{X}_n et S_n^2 pour $\mu = \mathbb{E}[X_i]$.

Un IC $(I_{n,1}, I_{n,2})$ est **asymptotiquement valide** si son erreur de couverture converge vers 0, i.e., si $\lim_{n \rightarrow \infty} \mathbb{P}[I_{n,1} \leq \mu \leq I_{n,2}] = (1 - \alpha)$.

Outils statistiques classiques vs outils statistiques pour la simulation.

Erreur non prise en compte

Les intervalles de confiance considérés ici prennent en compte l'erreur due aux aléas de la simulation.

Erreur non prise en compte

Les intervalles de confiance considérés ici prennent en compte l'erreur due aux aléas de la simulation.

Mais pas l'erreur dans l'estimation des paramètres du modèle.

Erreur non prise en compte

Les intervalles de confiance considérés ici prennent en compte l'erreur due aux aléas de la simulation.

Mais pas l'erreur dans l'estimation des paramètres du modèle.

Expliquer... Lire les notes.

Horizon fini

On observe X_1, \dots, X_n , des copies i.i.d. de X obtenues en faisant n répétitions de la simulation, et on veut estimer $\mu = \mathbb{E}[X]$.

Horizon fini

On observe X_1, \dots, X_n , des copies i.i.d. de X obtenues en faisant n répétitions de la simulation, et on veut estimer $\mu = \mathbb{E}[X]$.

On estime μ par \bar{X}_n et $\sigma^2 = \text{var}[X]$ par S_n^2 .

Horizon fini

On observe X_1, \dots, X_n , des copies i.i.d. de X obtenues en faisant n répétitions de la simulation, et on veut estimer $\mu = \mathbb{E}[X]$.

On estime μ par \bar{X}_n et $\sigma^2 = \text{var}[X]$ par S_n^2 .

Théorème. Si X_1, \dots, X_n sont i.i.d. $N(\mu, \sigma^2)$, alors pour tout n :

(i) \bar{X}_n et S_n^2 sont indépendants;

Horizon fini

On observe X_1, \dots, X_n , des copies i.i.d. de X obtenues en faisant n répétitions de la simulation, et on veut estimer $\mu = \mathbb{E}[X]$.

On estime μ par \bar{X}_n et $\sigma^2 = \text{var}[X]$ par S_n^2 .

Théorème. Si X_1, \dots, X_n sont i.i.d. $N(\mu, \sigma^2)$, alors pour tout n :

- (i) \bar{X}_n et S_n^2 sont indépendants;
- (ii) $(n-1)S_n^2/\sigma^2 \sim \text{chi-deux}(n-1)$;

Horizon fini

On observe X_1, \dots, X_n , des copies i.i.d. de X obtenues en faisant n répétitions de la simulation, et on veut estimer $\mu = \mathbb{E}[X]$.

On estime μ par \bar{X}_n et $\sigma^2 = \text{var}[X]$ par S_n^2 .

Théorème. Si X_1, \dots, X_n sont i.i.d. $N(\mu, \sigma^2)$, alors pour tout n :

- (i) \bar{X}_n et S_n^2 sont indépendants;
- (ii) $(n-1)S_n^2/\sigma^2 \sim \text{chi-deux}(n-1)$;
- (iii) $\sqrt{n}(\bar{X}_n - \mu)/S_n \sim \text{Student-}t(n-1)$.

Horizon fini

On observe X_1, \dots, X_n , des copies i.i.d. de X obtenues en faisant n répétitions de la simulation, et on veut estimer $\mu = \mathbb{E}[X]$.

On estime μ par \bar{X}_n et $\sigma^2 = \text{var}[X]$ par S_n^2 .

Théorème. Si X_1, \dots, X_n sont i.i.d. $N(\mu, \sigma^2)$, alors pour tout n :

- (i) \bar{X}_n et S_n^2 sont **indépendants**;
- (ii) $(n-1)S_n^2/\sigma^2 \sim \text{chi-deux}(n-1)$;
- (iii) $\sqrt{n}(\bar{X}_n - \mu)/S_n \sim \text{Student-}t(n-1)$.

Permet de calculer un IC exact pour μ au niveau $1 - \alpha$:

$$(\bar{X}_n \pm t^* S_n / \sqrt{n}),$$

où $\mathbb{P}[T_{n-1} \leq t^*] = 1 - \alpha/2$ ($t^* = t_{n-1, 1-\alpha/2}$).

Si n est grand: $\text{Student-}t(n-1) \approx N(0, 1)$.

Aussi un IC pour σ^2 : choisir x_1 et x_2 tels que

$$\mathbb{P}[x_1 < \chi_{n-1}^2 < x_2] = 1 - \alpha,$$

et poser

$$[I_1, I_2] = [(n-1)S_n^2/x_2, (n-1)S_n^2/x_1].$$

Aussi un IC pour σ^2 : choisir x_1 et x_2 tels que

$$\mathbb{P}[x_1 < \chi_{n-1}^2 < x_2] = 1 - \alpha,$$

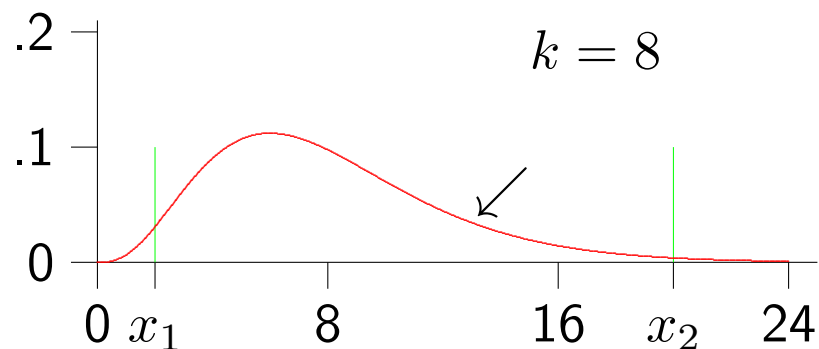
et poser

$$[I_1, I_2] = [(n-1)S_n^2/x_2, (n-1)S_n^2/x_1].$$

On a

$$\begin{aligned} \mathbb{P}[I_1 \leq \sigma^2 \leq I_2] &= \mathbb{P}[(n-1)S_n^2/x_2 \leq \sigma^2 \leq (n-1)S_n^2/x_1] \\ &= \mathbb{P}[x_1 \leq (n-1)S_n^2/\sigma^2 \leq x_2] \\ &= 1 - \alpha. \end{aligned}$$

Mais ceci n'est valide que si les X_i suivent la loi normale.



Bornes $(n-1)/x_2$ et $(n-1)/x_1$ d'un IC sur σ^2/S_n^2 :

	$\alpha = 0.02$		$\alpha = 0.10$	
n	$(n-1)/x_2$	$(n-1)/x_1$	$(n-1)/x_2$	$(n-1)/x_1$
10	0.388	3.518	0.492	2.284
30	0.570	1.939	0.663	1.568
100	0.729	1.413	0.796	1.270
300	0.831	1.216	0.876	1.146
1000	0.902	1.111	0.930	1.077

Bornes $(n-1)/x_2$ et $(n-1)/x_1$ d'un IC sur σ^2/S_n^2 :

	$\alpha = 0.02$		$\alpha = 0.10$	
n	$(n-1)/x_2$	$(n-1)/x_1$	$(n-1)/x_2$	$(n-1)/x_1$
10	0.388	3.518	0.492	2.284
30	0.570	1.939	0.663	1.568
100	0.729	1.413	0.796	1.270
300	0.831	1.216	0.876	1.146
1000	0.902	1.111	0.930	1.077

Par exemple, pour $n = 1000$, un IC à 90% pour σ^2 est donné par

$$[0.930 S_n^2, 1.077 S_n^2].$$

Bornes $(n - 1)/x_2$ et $(n - 1)/x_1$ d'un IC sur σ^2/S_n^2 :

	$\alpha = 0.02$		$\alpha = 0.10$	
n	$(n - 1)/x_2$	$(n - 1)/x_1$	$(n - 1)/x_2$	$(n - 1)/x_1$
10	0.388	3.518	0.492	2.284
30	0.570	1.939	0.663	1.568
100	0.729	1.413	0.796	1.270
300	0.831	1.216	0.876	1.146
1000	0.902	1.111	0.930	1.077

Par exemple, pour $n = 1000$, un IC à 90% pour σ^2 est donné par

$$[0.930 S_n^2, 1.077 S_n^2].$$

Pour $n = 30$, un IC à 90% pour σ^2 est donné par

$$[0.66 S_n^2, 1.56 S_n^2].$$

Si on a deux échantillons **indépendants**, X_1, \dots, X_m i.i.d. normales de variance σ_x^2 et Y_1, \dots, Y_n i.i.d. normales de variance σ_y^2 , on peut calculer un IC sur le **rapport des deux variances**, en utilisant le fait que ⁷

$$F = \frac{S_{x,m}^2 / \sigma_x^2}{S_{y,n}^2 / \sigma_y^2} = \frac{S_{x,m}^2 \sigma_y^2}{S_{y,n}^2 \sigma_x^2} \sim F(m-1, n-1),$$

où $S_{x,m}^2$ et $S_{y,n}^2$ sont les variances échantillonnales.

Si on a deux échantillons **indépendants**, X_1, \dots, X_m i.i.d. normales de variance σ_x^2 et Y_1, \dots, Y_n i.i.d. normales de variance σ_y^2 , on peut calculer un IC sur le **rapport des deux variances**, en utilisant le fait que ⁷

$$F = \frac{S_{x,m}^2 / \sigma_x^2}{S_{y,n}^2 / \sigma_y^2} = \frac{S_{x,m}^2 \sigma_y^2}{S_{y,n}^2 \sigma_x^2} \sim F(m-1, n-1),$$

où $S_{x,m}^2$ et $S_{y,n}^2$ sont les variances échantillonnales.

Si $\mathbb{P}[x_1 < F < x_2] = 1 - \alpha$, l'intervalle est

$$[I_1, I_2] = \left[\frac{1}{x_2} \frac{S_{x,m}^2}{S_{y,n}^2}, \frac{1}{x_1} \frac{S_{x,m}^2}{S_{y,n}^2} \right].$$

Si on a deux échantillons **indépendants**, X_1, \dots, X_m i.i.d. normales de variance σ_x^2 et Y_1, \dots, Y_n i.i.d. normales de variance σ_y^2 , on peut calculer un IC sur le **rapport des deux variances**, en utilisant le fait que ⁷

$$F = \frac{S_{x,m}^2 / \sigma_x^2}{S_{y,n}^2 / \sigma_y^2} = \frac{S_{x,m}^2 \sigma_y^2}{S_{y,n}^2 \sigma_x^2} \sim F(m-1, n-1),$$

où $S_{x,m}^2$ et $S_{y,n}^2$ sont les variances échantillonnales.

Si $\mathbb{P}[x_1 < F < x_2] = 1 - \alpha$, l'intervalle est

$$[I_1, I_2] = \left[\frac{1}{x_2} \frac{S_{x,m}^2}{S_{y,n}^2}, \frac{1}{x_1} \frac{S_{x,m}^2}{S_{y,n}^2} \right].$$

Potentiellement utile lorsqu'on estime le **facteur de réduction de variance** entre deux estimateurs.

Pour un grand n : approx. normale

Si n est grand, \bar{X}_n est approx. normal même si X ne l'est pas, grâce au **TLC**.

Pour un grand n : approx. normale

Si n est grand, \bar{X}_n est approx. normal même si X ne l'est pas, grâce au TLC.

On peut s'y fier pour calculer un IC, sauf si:

- n est trop petit,
- ou α est proche de 0,
- ou les X_i ont une loi très asymétrique
- ou avec des moments supérieurs très élevés.

Pour un grand n : approx. normale

Si n est grand, \bar{X}_n est approx. normal même si X ne l'est pas, grâce au TLC.

On peut s'y fier pour calculer un IC, sauf si:

- n est trop petit,
- ou α est proche de 0,
- ou les X_i ont une loi très asymétrique
- ou avec des moments supérieurs très élevés.

Exemple. $n = 1000$, $X_i \sim \text{Binomiale}(1, p)$, on veut estimer p .

Pour un grand n : approx. normale

Si n est grand, \bar{X}_n est approx. normal même si X ne l'est pas, grâce au TLC.

On peut s'y fier pour calculer un IC, sauf si:

- n est trop petit,
- ou α est proche de 0,
- ou les X_i ont une loi très asymétrique
- ou avec des moments supérieurs très élevés.

Exemple. $n = 1000$, $X_i \sim \text{Binomiale}(1, p)$, on veut estimer p .

Si on a 882 succès, $\bar{X}_n = 0.882$.

Pour un grand n : approx. normale

Si n est grand, \bar{X}_n est approx. normal même si X ne l'est pas, grâce au TLC.

On peut s'y fier pour calculer un IC, sauf si:

- n est trop petit,
- ou α est proche de 0,
- ou les X_i ont une loi très asymétrique
- ou avec des moments supérieurs très élevés.

Exemple. $n = 1000$, $X_i \sim \text{Binomiale}(1, p)$, on veut estimer p .

Si on a 882 succès, $\bar{X}_n = 0.882$.

On a alors $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \bar{X}_n(1 - \bar{X}_n) \approx 0.1042$

et un IC à 95% (approx.) est $(\bar{X}_n \pm 1.96 S_n / \sqrt{n}) \approx (0.862, 0.902)$.

Pour un grand n : approx. normale

Si n est grand, \bar{X}_n est approx. normal même si X ne l'est pas, grâce au TLC.

On peut s'y fier pour calculer un IC, sauf si:

- n est trop petit,
- ou α est proche de 0,
- ou les X_i ont une loi très asymétrique
- ou avec des moments supérieurs très élevés.

Exemple. $n = 1000$, $X_i \sim \text{Binomiale}(1, p)$, on veut estimer p .

Si on a 882 succès, $\bar{X}_n = 0.882$.

On a alors $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \bar{X}_n(1 - \bar{X}_n) \approx 0.1042$

et un IC à 95% (approx.) est $(\bar{X}_n \pm 1.96 S_n / \sqrt{n}) \approx (0.862, 0.902)$.

Mais si $\bar{X}_n = 0.998$, alors on voit que p est trop proche de 1 et l'approx. normale sera très mauvaise.

Pour un grand n : approx. normale

Si n est grand, \bar{X}_n est approx. normal même si X ne l'est pas, grâce au **TLC**.

On peut s'y fier pour calculer un IC, sauf si:

- n est trop petit,
- ou α est proche de 0,
- ou les X_i ont une loi très asymétrique
- ou avec des moments supérieurs très élevés.

Exemple. $n = 1000$, $X_i \sim \text{Binomiale}(1, p)$, on veut estimer p .

Si on a 882 succès, $\bar{X}_n = 0.882$.

On a alors $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \bar{X}_n (1 - \bar{X}_n) \approx 0.1042$

et un IC à 95% (approx.) est $(\bar{X}_n \pm 1.96 S_n / \sqrt{n}) \approx (0.862, 0.902)$.

Mais si $\bar{X}_n = 0.998$, alors on voit que p est trop proche de 1 et l'approx. normale sera très mauvaise.

Dans ce cas, on va plutôt utiliser: $Y = \sum_{i=1}^n (1 - X_i) \approx \text{Poisson}(n(1 - p))$.

IC pour une loi discrète

Y prend ses valeurs dans $\{0, 1, 2, \dots\}$ et suit une loi de paramètre μ , telle que $P_\mu[Y \geq y]$ est croissant en μ .

IC pour une loi discrète

Y prend ses valeurs dans $\{0, 1, 2, \dots\}$ et suit une loi de paramètre μ , telle que $P_\mu[Y \geq y]$ est croissant en μ .
(Le cas décroissant se traite de manière symétrique.)

IC pour une loi discrète

Y prend ses valeurs dans $\{0, 1, 2, \dots\}$ et suit une loi de paramètre μ , telle que $P_\mu[Y \geq y]$ est croissant en μ .

(Le cas décroissant se traite de manière symétrique.)

Exemples: lois binomiale, géométrique, Poisson, ...

IC pour une loi discrète

Y prend ses valeurs dans $\{0, 1, 2, \dots\}$ et suit une loi de paramètre μ , telle que $P_\mu[Y \geq y]$ est croissant en μ .

(Le cas décroissant se traite de manière symétrique.)

Exemples: lois binomiale, géométrique, Poisson, ...

On veut un IC $[I_1, I_2]$ de niveau (approx.) $1 - \alpha$ pour μ .

IC pour une loi discrète

Y prend ses valeurs dans $\{0, 1, 2, \dots\}$ et suit une loi de paramètre μ , telle que $P_\mu[Y \geq y]$ est croissant en μ .

(Le cas décroissant se traite de manière symétrique.)

Exemples: lois binomiale, géométrique, Poisson, ...

On veut un IC $[I_1, I_2]$ de niveau (approx.) $1 - \alpha$ pour μ .

Ou encore pour $\alpha = \alpha_1 + \alpha_2$, on veut $\mathbb{P}[\mu < I_1] \approx \alpha_1$ et $\mathbb{P}[\mu > I_2] \approx \alpha_2$.

IC pour une loi discrète

Y prend ses valeurs dans $\{0, 1, 2, \dots\}$ et suit une loi de paramètre μ , telle que $P_\mu[Y \geq y]$ est croissant en μ .

(Le cas décroissant se traite de manière symétrique.)

Exemples: lois binomiale, géométrique, Poisson, ...

On veut un IC $[I_1, I_2]$ de niveau (approx.) $1 - \alpha$ pour μ .

Ou encore pour $\alpha = \alpha_1 + \alpha_2$, on veut $\mathbb{P}[\mu < I_1] \approx \alpha_1$ et $\mathbb{P}[\mu > I_2] \approx \alpha_2$.

Si on observe $Y = y$, l'intervalle sera $[I_1(y), I_2(y)]$.

IC pour une loi discrète

Y prend ses valeurs dans $\{0, 1, 2, \dots\}$ et suit une loi de paramètre μ , telle que $P_\mu[Y \geq y]$ est croissant en μ .

(Le cas décroissant se traite de manière symétrique.)

Exemples: lois binomiale, géométrique, Poisson, ...

On veut un IC $[I_1, I_2]$ de niveau (approx.) $1 - \alpha$ pour μ .

Ou encore pour $\alpha = \alpha_1 + \alpha_2$, on veut $\mathbb{P}[\mu < I_1] \approx \alpha_1$ et $\mathbb{P}[\mu > I_2] \approx \alpha_2$.

Si on observe $Y = y$, l'intervalle sera $[I_1(y), I_2(y)]$.

Algorithme: Prendre pour $I_1(y)$ et $I_2(y)$ les solutions de

$$\alpha_1 = P_{I_1}[Y \geq y] \quad \text{et} \quad \alpha_2 = P_{I_2}[Y \leq y].$$

IC pour une loi discrète

Y prend ses valeurs dans $\{0, 1, 2, \dots\}$ et suit une loi de paramètre μ , telle que $P_\mu[Y \geq y]$ est croissant en μ .

(Le cas décroissant se traite de manière symétrique.)

Exemples: lois binomiale, géométrique, Poisson, ...

On veut un IC $[I_1, I_2]$ de niveau (approx.) $1 - \alpha$ pour μ .

Ou encore pour $\alpha = \alpha_1 + \alpha_2$, on veut $\mathbb{P}[\mu < I_1] \approx \alpha_1$ et $\mathbb{P}[\mu > I_2] \approx \alpha_2$.

Si on observe $Y = y$, l'intervalle sera $[I_1(y), I_2(y)]$.

Algorithme: Prendre pour $I_1(y)$ et $I_2(y)$ les solutions de

$$\alpha_1 = P_{I_1}[Y \geq y] \quad \text{et} \quad \alpha_2 = P_{I_2}[Y \leq y].$$

Exemple: Si $Y \sim \text{Poisson}(\lambda)$ et on observe $y = 2$, on a $I_1 = \lambda_1$ tel que

$$\alpha_1 = P_{I_1}[Y \geq 2] = 1 - e^{\lambda_1} - \lambda_1 e^{-\lambda_1}$$

et $I_2 = \lambda_2$ tel que

$$\alpha_2 = P_{I_2}[Y \leq 2] = e^{\lambda_2} + \lambda_2 e^{-\lambda_2} + \lambda_2^2 e^{-\lambda_2} / 2.$$

On peut résoudre par recherche binaire, par exemple.

Si on observe $Y = y$, l'intervalle sera $[I_1(y), I_2(y)]$.

Algorithme: Prendre pour $I_1(y)$ et $I_2(y)$ les solutions de

$$\alpha_1 = P_{I_1}[Y \geq y] \quad \text{et} \quad \alpha_2 = P_{I_2}[Y \leq y].$$

Proposition. La probabilité de couverture de cet intervalle est d'au moins $1 - \alpha$.

Si on observe $Y = y$, l'intervalle sera $[I_1(y), I_2(y)]$.

Algorithme: Prendre pour $I_1(y)$ et $I_2(y)$ les solutions de

$$\alpha_1 = P_{I_1}[Y \geq y] \quad \text{et} \quad \alpha_2 = P_{I_2}[Y \leq y].$$

Proposition. La probabilité de couverture de cet intervalle est d'au moins $1 - \alpha$.

Preuve: Soit $y^*(\mu) = \min\{y \in \mathbb{N} : I_1(y) \geq \mu\}$ et $\nu = I_1(y^*(\mu)) \geq \mu$. On a

$$P_\mu[I_1(Y) \geq \mu] \leq P_\nu[I_1(Y) \geq \mu] = P_\nu[Y \geq y^*(\mu)] = P_{I_1(y^*(\mu))}[Y \geq y^*(\mu)] = \alpha_1.$$

On montre de même que $P_\mu[I_2(Y) \leq \mu] \leq \alpha_2$.

Si on observe $Y = y$, l'intervalle sera $[I_1(y), I_2(y)]$.

Algorithme: Prendre pour $I_1(y)$ et $I_2(y)$ les solutions de

$$\alpha_1 = P_{I_1}[Y \geq y] \quad \text{et} \quad \alpha_2 = P_{I_2}[Y \leq y].$$

Proposition. La probabilité de couverture de cet intervalle est d'au moins $1 - \alpha$.

Preuve: Soit $y^*(\mu) = \min\{y \in \mathbb{N} : I_1(y) \geq \mu\}$ et $\nu = I_1(y^*(\mu)) \geq \mu$. On a

$$P_\mu[I_1(Y) \geq \mu] \leq P_\nu[I_1(Y) \geq \mu] = P_\nu[Y \geq y^*(\mu)] = P_{I_1(y^*(\mu))}[Y \geq y^*(\mu)] = \alpha_1.$$

On montre de même que $P_\mu[I_2(Y) \leq \mu] \leq \alpha_2$.

La probabilité de couverture **exacte** dépend de F_μ et est généralement inconnue.

Si on observe $Y = y$, l'intervalle sera $[I_1(y), I_2(y)]$.

Algorithme: Prendre pour $I_1(y)$ et $I_2(y)$ les solutions de

$$\alpha_1 = P_{I_1}[Y \geq y] \quad \text{et} \quad \alpha_2 = P_{I_2}[Y \leq y].$$

Proposition. La probabilité de couverture de cet intervalle est d'au moins $1 - \alpha$.

Preuve: Soit $y^*(\mu) = \min\{y \in \mathbb{N} : I_1(y) \geq \mu\}$ et $\nu = I_1(y^*(\mu)) \geq \mu$. On a

$$P_\mu[I_1(Y) \geq \mu] \leq P_\nu[I_1(Y) \geq \mu] = P_\nu[Y \geq y^*(\mu)] = P_{I_1(y^*(\mu))}[Y \geq y^*(\mu)] = \alpha_1.$$

On montre de même que $P_\mu[I_2(Y) \leq \mu] \leq \alpha_2$.

La probabilité de couverture **exacte** dépend de F_μ et est généralement inconnue.

Exemple. X_1, \dots, X_n i.i.d. Bernouilli(p), $Y = \sum_{i=1}^n X_i \sim \text{Binomiale}(n, p)$.

On veut un IC pour p .

Si on observe $Y = y$, l'intervalle sera $[I_1(y), I_2(y)]$.

Algorithme: Prendre pour $I_1(y)$ et $I_2(y)$ les solutions de

$$\alpha_1 = P_{I_1}[Y \geq y] \quad \text{et} \quad \alpha_2 = P_{I_2}[Y \leq y].$$

Proposition. La probabilité de couverture de cet intervalle est d'au moins $1 - \alpha$.

Preuve: Soit $y^*(\mu) = \min\{y \in \mathbb{N} : I_1(y) \geq \mu\}$ et $\nu = I_1(y^*(\mu)) \geq \mu$. On a

$$P_\mu[I_1(Y) \geq \mu] \leq P_\nu[I_1(Y) \geq \mu] = P_\nu[Y \geq y^*(\mu)] = P_{I_1(y^*(\mu))}[Y \geq y^*(\mu)] = \alpha_1.$$

On montre de même que $P_\mu[I_2(Y) \leq \mu] \leq \alpha_2$.

La probabilité de couverture **exacte** dépend de F_μ et est généralement inconnue.

Exemple. X_1, \dots, X_n i.i.d. Bernouilli(p), $Y = \sum_{i=1}^n X_i \sim \text{Binomiale}(n, p)$.

On veut un IC pour p .

Si n est grand et p est petit: $Y \approx \text{Poisson}(np)$.

Si on observe $Y = y$, l'intervalle sera $[I_1(y), I_2(y)]$.

Algorithme: Prendre pour $I_1(y)$ et $I_2(y)$ les solutions de

$$\alpha_1 = P_{I_1}[Y \geq y] \quad \text{et} \quad \alpha_2 = P_{I_2}[Y \leq y].$$

Proposition. La probabilité de couverture de cet intervalle est d'au moins $1 - \alpha$.

Preuve: Soit $y^*(\mu) = \min\{y \in \mathbb{N} : I_1(y) \geq \mu\}$ et $\nu = I_1(y^*(\mu)) \geq \mu$. On a

$$P_\mu[I_1(Y) \geq \mu] \leq P_\nu[I_1(Y) \geq \mu] = P_\nu[Y \geq y^*(\mu)] = P_{I_1(y^*(\mu))}[Y \geq y^*(\mu)] = \alpha_1.$$

On montre de même que $P_\mu[I_2(Y) \leq \mu] \leq \alpha_2$.

La probabilité de couverture **exacte** dépend de F_μ et est généralement inconnue.

Exemple. X_1, \dots, X_n i.i.d. Bernouilli(p), $Y = \sum_{i=1}^n X_i \sim \text{Binomiale}(n, p)$.

On veut un IC pour p .

Si n est grand et p est petit: $Y \approx \text{Poisson}(np)$.

Si p est proche de 1, on remplace p et X_i par $1 - p$ et $1 - X_i$.

Si on observe $Y = y$, l'intervalle sera $[I_1(y), I_2(y)]$.

Algorithme: Prendre pour $I_1(y)$ et $I_2(y)$ les solutions de

$$\alpha_1 = P_{I_1}[Y \geq y] \quad \text{et} \quad \alpha_2 = P_{I_2}[Y \leq y].$$

Proposition. La probabilité de couverture de cet intervalle est d'au moins $1 - \alpha$.

Preuve: Soit $y^*(\mu) = \min\{y \in \mathbb{N} : I_1(y) \geq \mu\}$ et $\nu = I_1(y^*(\mu)) \geq \mu$. On a

$$P_\mu[I_1(Y) \geq \mu] \leq P_\nu[I_1(Y) \geq \mu] = P_\nu[Y \geq y^*(\mu)] = P_{I_1(y^*(\mu))}[Y \geq y^*(\mu)] = \alpha_1.$$

On montre de même que $P_\mu[I_2(Y) \leq \mu] \leq \alpha_2$.

La probabilité de couverture **exacte** dépend de F_μ et est généralement inconnue.

Exemple. X_1, \dots, X_n i.i.d. Bernouilli(p), $Y = \sum_{i=1}^n X_i \sim \text{Binomiale}(n, p)$.

On veut un IC pour p .

Si n est grand et p est petit: $Y \approx \text{Poisson}(np)$.

Si p est proche de 1, on remplace p et X_i par $1 - p$ et $1 - X_i$.

Si y dépasse quelques dizaines, on peut calculer $\mathbb{P}[Y \leq y]$ via la cdf d'une loi beta ou gamma, si on dispose d'un programme pouvant la calculer:

Si $Y \sim \text{Binomiale}(n, p)$ et $X \sim \text{Beta}(y, n - y + 1)$, alors $\mathbb{P}[Y \geq y] = \mathbb{P}[X \leq p]$.

Si $Y \sim \text{Poisson}(\lambda)$ et $X \sim \text{Gamma}(y, 1)$, alors $\mathbb{P}[Y \geq y] = \mathbb{P}[X \leq \lambda]$.

IC pour la moyenne via Hoeffding

Supposons que $\mathbb{P}[0 \leq Y \leq 1] = 1$ et on veut un IC pour $\nu = \mathbb{E}[Y]$.
On peut généraliser à des bornes quelconques en changeant l'échelle.

IC pour la moyenne via Hoeffding

Supposons que $\mathbb{P}[0 \leq Y \leq 1] = 1$ et on veut un IC pour $\nu = \mathbb{E}[Y]$.
On peut généraliser à des bornes quelconques en changeant l'échelle.

Théorème. (Hoeffding 1963).

Si Y_1, \dots, Y_n est un échantillon de copies i.i.d. de Y , alors

$$\begin{aligned} \mathbb{P}[\bar{Y}_n - \nu \geq \epsilon] &\leq e^{-ng(\nu)\epsilon^2} \leq e^{-2n\epsilon^2} && \text{et} \\ \mathbb{P}[\nu - \bar{Y}_n \geq \epsilon] &\leq e^{-ng(1-\nu)\epsilon^2} \leq e^{-2n\epsilon^2}, \end{aligned}$$

où

$$g(\nu) = \begin{cases} \ln((1-\nu)/\nu)/[1-2\nu] & \text{pour } 0 < \nu < 1/2, \\ 1/[2\nu(1-\nu)] & \text{pour } 1/2 \leq \nu < 1. \end{cases}$$

IC pour la moyenne via Hoeffding

Supposons que $\mathbb{P}[0 \leq Y \leq 1] = 1$ et on veut un IC pour $\nu = \mathbb{E}[Y]$.
On peut généraliser à des bornes quelconques en changeant l'échelle.

Théorème. (Hoeffding 1963).

Si Y_1, \dots, Y_n est un échantillon de copies i.i.d. de Y , alors

$$\begin{aligned} \mathbb{P}[\bar{Y}_n - \nu \geq \epsilon] &\leq e^{-ng(\nu)\epsilon^2} \leq e^{-2n\epsilon^2} & \text{et} \\ \mathbb{P}[\nu - \bar{Y}_n \geq \epsilon] &\leq e^{-ng(1-\nu)\epsilon^2} \leq e^{-2n\epsilon^2}, \end{aligned}$$

où

$$g(\nu) = \begin{cases} \ln((1-\nu)/\nu)/[1-2\nu] & \text{pour } 0 < \nu < 1/2, \\ 1/[2\nu(1-\nu)] & \text{pour } 1/2 \leq \nu < 1. \end{cases}$$

En prenant $\epsilon^2 = -\ln(\alpha/2)/(2n)$, on a

$$\mathbb{P}[|\bar{Y}_n - \nu| \leq \epsilon] \geq 1 - 2e^{-2n\epsilon^2} = 1 - \alpha.$$

et donc $(\bar{Y}_n - \epsilon, \bar{Y}_n + \epsilon)$ est un IC à au moins $100(1 - \alpha)\%$ pour μ .

Important: Ne dépend pas de la loi de probabilité!

Important: Ne dépend pas de la loi de probabilité!

Les bornes impliquant g sont utiles si on sait à l'avance que ν est dans un intervalle ne contenant pas $1/2$. La fonction g est continue, décroissante sur $(0, 1/2)$, croissante sur $(1/2, 1)$, et a un minimum à $g(1/2) = 2$.

Estimation Séquentielle

Pour un IC de niveau $1 - \alpha$, si on fixe n , la largeur $I_2 - I_1$ est aléatoire.

Estimation Séquentielle

Pour un IC de niveau $1 - \alpha$, si on fixe n , la largeur $I_2 - I_1$ est aléatoire.

Si on veut $I_2 - I_1 \leq w$ pour w fixé, la valeur minimale de n requise est une variable aléatoire N . Comment prédire ce N ?

Estimation Séquentielle

Pour un IC de niveau $1 - \alpha$, si on fixe n , la largeur $I_2 - I_1$ est aléatoire.

Si on veut $I_2 - I_1 \leq w$ pour w fixé, la valeur minimale de n requise est une variable aléatoire N . Comment prédire ce N ?

Exemple. Pour $X_i \sim \text{binomiale}(1, p)$, avec $n = 1000$ on a obtenu $\bar{X}_n = 0.882$, $S_n^2 \approx 0.1042$, et l'IC à 95% était $[I_1, I_2] = [0.862, 0.902]$.
La demi-largeur de cet IC est de 0.020.

Estimation Séquentielle

Pour un IC de niveau $1 - \alpha$, si on fixe n , la largeur $I_2 - I_1$ est aléatoire.

Si on veut $I_2 - I_1 \leq w$ pour w fixé, la valeur minimale de n requise est une variable aléatoire N . Comment prédire ce N ?

Exemple. Pour $X_i \sim \text{binomiale}(1, p)$, avec $n = 1000$ on a obtenu $\bar{X}_n = 0.882$, $S_n^2 \approx 0.1042$, et l'IC à 95% était $[I_1, I_2] = [0.862, 0.902]$.

La demi-largeur de cet IC est de 0.020.

Combien de répétitions additionnelles faut-il pour réduire la demi-largeur à 0.005?

Estimation Séquentielle

Pour un IC de niveau $1 - \alpha$, si on fixe n , la largeur $I_2 - I_1$ est aléatoire.

Si on veut $I_2 - I_1 \leq w$ pour w fixé, la valeur minimale de n requise est une variable aléatoire N . Comment prédire ce N ?

Exemple. Pour $X_i \sim \text{binomiale}(1, p)$, avec $n = 1000$ on a obtenu $\bar{X}_n = 0.882$, $S_n^2 \approx 0.1042$, et l'IC à 95% était $[I_1, I_2] = [0.862, 0.902]$.

La demi-largeur de cet IC est de 0.020.

Combien de répétitions additionnelles faut-il pour réduire la demi-largeur à 0.005?

On veut $1.96S_n/\sqrt{n} \leq 0.005$.

Estimation Séquentielle

Pour un IC de niveau $1 - \alpha$, si on fixe n , la largeur $I_2 - I_1$ est aléatoire.

Si on veut $I_2 - I_1 \leq w$ pour w fixé, la valeur minimale de n requise est une variable aléatoire N . Comment prédire ce N ?

Exemple. Pour $X_i \sim \text{binomiale}(1, p)$, avec $n = 1000$ on a obtenu $\bar{X}_n = 0.882$, $S_n^2 \approx 0.1042$, et l'IC à 95% était $[I_1, I_2] = [0.862, 0.902]$.

La demi-largeur de cet IC est de 0.020.

Combien de répétitions additionnelles faut-il pour réduire la demi-largeur à 0.005?

On veut $1.96 S_n / \sqrt{n} \leq 0.005$.

En supposant que S_n ne changera pas significativement, cela donne

$n \geq (1.96 \times S_n / 0.005)^2 \approx 16011.8$.

Estimation Séquentielle

Pour un IC de niveau $1 - \alpha$, si on fixe n , la largeur $I_2 - I_1$ est aléatoire.

Si on veut $I_2 - I_1 \leq w$ pour w fixé, la valeur minimale de n requise est une variable aléatoire N . Comment prédire ce N ?

Exemple. Pour $X_i \sim \text{binomiale}(1, p)$, avec $n = 1000$ on a obtenu $\bar{X}_n = 0.882$, $S_n^2 \approx 0.1042$, et l'IC à 95% était $[I_1, I_2] = [0.862, 0.902]$.

La demi-largeur de cet IC est de 0.020.

Combien de répétitions additionnelles faut-il pour réduire la demi-largeur à 0.005?

On veut $1.96S_n/\sqrt{n} \leq 0.005$.

En supposant que S_n ne changera pas significativement, cela donne

$n \geq (1.96 \times S_n/0.005)^2 \approx 16011.8$.

Recommandation: faire 15012 répétitions additionnelles.

Procédure à deux étapes, pour la loi de Student (ou normale):
Faire n_0 répétitions et calculer $S_{n_0}^2$;

Procédure à **deux étapes**, pour la loi de Student (ou normale):

Faire n_0 répétitions et calculer $S_{n_0}^2$;

La prédiction du n requis est

$$\hat{N}^* = \min \{n \geq n_0 \mid (t_{n-1, 1-\alpha/2}) S_{n_0} / \sqrt{n} \leq w/2\}.$$

Procédure à deux étapes, pour la loi de Student (ou normale):

Faire n_0 répétitions et calculer $S_{n_0}^2$;

La prédiction du n requis est

$$\hat{N}^* = \min \{n \geq n_0 \mid (t_{n-1, 1-\alpha/2}) S_{n_0} / \sqrt{n} \leq w/2\}.$$

On fera $\max(0, \hat{N}^* - n_0)$ répétitions additionnelles.

Procédure à **deux étapes**, pour la loi de Student (ou normale):

Faire n_0 répétitions et calculer $S_{n_0}^2$;

La prédiction du n requis est

$$\hat{N}^* = \min \{n \geq n_0 \mid (t_{n-1, 1-\alpha/2}) S_{n_0} / \sqrt{n} \leq w/2\}.$$

On fera $\max(0, \hat{N}^* - n_0)$ répétitions additionnelles.

Bien sûr, il se peut que ce soit insuffisant, ou trop.

Procédure à **deux étapes**, pour la loi de Student (ou normale):

Faire n_0 répétitions et calculer $S_{n_0}^2$;

La prédiction du n requis est

$$\hat{N}^* = \min \{n \geq n_0 \mid (t_{n-1, 1-\alpha/2}) S_{n_0} / \sqrt{n} \leq w/2\}.$$

On fera $\max(0, \hat{N}^* - n_0)$ répétitions additionnelles.

Bien sûr, il se peut que ce soit insuffisant, ou trop.

Estimation séquentielle:

Après n_0 répét., recalculer S_n^2 et la demi-largeur pour chaque n .

On s'arrête dès que $I_2 - I_1 \leq w$.

Procédure à **deux étapes**, pour la loi de Student (ou normale):

Faire n_0 répétitions et calculer $S_{n_0}^2$;

La prédiction du n requis est

$$\hat{N}^* = \min \{n \geq n_0 \mid (t_{n-1, 1-\alpha/2}) S_{n_0} / \sqrt{n} \leq w/2\}.$$

On fera $\max(0, \hat{N}^* - n_0)$ répétitions additionnelles.

Bien sûr, il se peut que ce soit insuffisant, ou trop.

Estimation séquentielle:

Après n_0 répét., recalculer S_n^2 et la demi-largeur pour chaque n .

On s'arrête dès que $I_2 - I_1 \leq w$.

Cette procédure est biaisée, car on tend à s'arrêter à un N où S_N^2 **sous-estime** la variance.

Procédure à **deux étapes**, pour la loi de Student (ou normale):

Faire n_0 répétitions et calculer $S_{n_0}^2$;

La prédiction du n requis est

$$\hat{N}^* = \min \{n \geq n_0 \mid (t_{n-1, 1-\alpha/2}) S_{n_0} / \sqrt{n} \leq w/2\}.$$

On fera $\max(0, \hat{N}^* - n_0)$ répétitions additionnelles.

Bien sûr, il se peut que ce soit insuffisant, ou trop.

Estimation séquentielle:

Après n_0 répét., recalculer S_n^2 et la demi-largeur pour chaque n .

On s'arrête dès que $I_2 - I_1 \leq w$.

Cette procédure est biaisée, car on tend à s'arrêter à un N où S_N^2 **sous-estime** la variance.

Mais lorsque $w \rightarrow 0$, le bias disparaît,

Procédure à **deux étapes**, pour la loi de Student (ou normale):

Faire n_0 répétitions et calculer $S_{n_0}^2$;

La prédiction du n requis est

$$\hat{N}^* = \min \{n \geq n_0 \mid (t_{n-1, 1-\alpha/2}) S_{n_0} / \sqrt{n} \leq w/2\}.$$

On fera $\max(0, \hat{N}^* - n_0)$ répétitions additionnelles.

Bien sûr, il se peut que ce soit insuffisant, ou trop.

Estimation séquentielle:

Après n_0 répét., recalculer S_n^2 et la demi-largeur pour chaque n .

On s'arrête dès que $I_2 - I_1 \leq w$.

Cette procédure est biaisée, car on tend à s'arrêter à un N où S_N^2 **sous-estime** la variance.

Mais lorsque $w \rightarrow 0$, le bias disparaît, $N/n^* \rightarrow 1$ a.p.1 où n^* est la valeur optimale de N si on connaissait σ^2 ,

Procédure à **deux étapes**, pour la loi de Student (ou normale):

Faire n_0 répétitions et calculer $S_{n_0}^2$;

La prédiction du n requis est

$$\hat{N}^* = \min \{n \geq n_0 \mid (t_{n-1, 1-\alpha/2})S_{n_0}/\sqrt{n} \leq w/2\}.$$

On fera $\max(0, \hat{N}^* - n_0)$ répétitions additionnelles.

Bien sûr, il se peut que ce soit insuffisant, ou trop.

Estimation séquentielle:

Après n_0 répét., recalculer S_n^2 et la demi-largeur pour chaque n .

On s'arrête dès que $I_2 - I_1 \leq w$.

Cette procédure est biaisée, car on tend à s'arrêter à un N où S_N^2 **sous-estime** la variance.

Mais lorsque $w \rightarrow 0$, le bias disparaît, $N/n^* \rightarrow 1$ a.p.1 où n^* est la valeur optimale de N si on connaissait σ^2 , et $\mathbb{P}[|\bar{X}_N - \mu| \leq w/2] \rightarrow 1 - \alpha$.

Régions de confiance pour des vecteurs

On veut estimer un vecteur de d espérances, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$.

Régions de confiance pour des vecteurs

On veut estimer un vecteur de d espérances, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$.
Si on calcule des IC $\mathcal{I}_1, \dots, \mathcal{I}_d$ tels que $\mathbb{P}[\mu_j \in \mathcal{I}_j] = 1 - \alpha$,
alors la probabilité de couverture simultanée est

$$1 - \alpha' = \mathbb{P}[\mu_j \in \mathcal{I}_j \text{ pour tout } j] \neq 1 - \alpha.$$

Régions de confiance pour des vecteurs

On veut estimer un **vecteur** de d espérances, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$.
Si on calcule des IC $\mathcal{I}_1, \dots, \mathcal{I}_d$ tels que $\mathbb{P}[\mu_j \in \mathcal{I}_j] = 1 - \alpha$,
alors la probabilité de couverture **simultanée** est

$$1 - \alpha' = \mathbb{P}[\mu_j \in \mathcal{I}_j \text{ pour tout } j] \neq 1 - \alpha.$$

Quand les \mathcal{I}_j sont dépendants, on ne connaît pas α' .

Régions de confiance pour des vecteurs

On veut estimer un **vecteur** de d espérances, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$.
Si on calcule des IC $\mathcal{I}_1, \dots, \mathcal{I}_d$ tels que $\mathbb{P}[\mu_j \in \mathcal{I}_j] = 1 - \alpha$,
alors la probabilité de couverture **simultanée** est

$$1 - \alpha' = \mathbb{P}[\mu_j \in \mathcal{I}_j \text{ pour tout } j] \neq 1 - \alpha.$$

Quand les \mathcal{I}_j sont dépendants, on ne connaît pas α' .

Quoi faire si on veut une **région de confiance** $\mathcal{I} \subset \mathbb{R}^d$ telle que $\mathbb{P}[\boldsymbol{\mu} \in \mathcal{I}] \geq 1 - \alpha$?

Inégalité de Bonferroni.

Pour chaque j , on définit \mathcal{I}_j tel que $\mathbb{P}[\mu_j \in \mathcal{I}_j] = 1 - \alpha_j$.

Inégalité de Bonferroni.

Pour chaque j , on définit \mathcal{I}_j tel que $\mathbb{P}[\mu_j \in \mathcal{I}_j] = 1 - \alpha_j$.

Soit

$$\mathcal{I} = \{(\mu_1, \dots, \mu_d) \mid \mu_1 \in \mathcal{I}_1, \dots, \mu_d \in \mathcal{I}_d\}$$

(boîte rectangulaire)

Inégalité de Bonferroni.

Pour chaque j , on définit \mathcal{I}_j tel que $\mathbb{P}[\mu_j \in \mathcal{I}_j] = 1 - \alpha_j$.

Soit

$$\mathcal{I} = \{(\mu_1, \dots, \mu_d) \mid \mu_1 \in \mathcal{I}_1, \dots, \mu_d \in \mathcal{I}_d\}$$

(boîte rectangulaire) et $\alpha = \sum_{j=1}^d \alpha_j < 1$.

Inégalité de Bonferroni.

Pour chaque j , on définit \mathcal{I}_j tel que $\mathbb{P}[\mu_j \in \mathcal{I}_j] = 1 - \alpha_j$.
Soit

$$\mathcal{I} = \{(\mu_1, \dots, \mu_d) \mid \mu_1 \in \mathcal{I}_1, \dots, \mu_d \in \mathcal{I}_d\}$$

(boîte rectangulaire) et $\alpha = \sum_{j=1}^d \alpha_j < 1$. Alors

$$\begin{aligned} \mathbb{P}[\boldsymbol{\mu} \in \mathcal{I}] &= \mathbb{P}[\mu_j \in \mathcal{I}_j \text{ pour tout } j] \\ &= 1 - \mathbb{P}[\mu_j \notin \mathcal{I}_j \text{ pour un } j] \\ &\geq 1 - \sum_{j=1}^d \mathbb{P}[\mu_j \notin \mathcal{I}_j] \\ &\geq 1 - \sum_{j=1}^d \alpha_j = 1 - \alpha. \end{aligned}$$

Inégalité de Bonferroni.

Pour chaque j , on définit \mathcal{I}_j tel que $\mathbb{P}[\mu_j \in \mathcal{I}_j] = 1 - \alpha_j$.
Soit

$$\mathcal{I} = \{(\mu_1, \dots, \mu_d) \mid \mu_1 \in \mathcal{I}_1, \dots, \mu_d \in \mathcal{I}_d\}$$

(boîte rectangulaire) et $\alpha = \sum_{j=1}^d \alpha_j < 1$. Alors

$$\begin{aligned} \mathbb{P}[\boldsymbol{\mu} \in \mathcal{I}] &= \mathbb{P}[\mu_j \in \mathcal{I}_j \text{ pour tout } j] \\ &= 1 - \mathbb{P}[\mu_j \notin \mathcal{I}_j \text{ pour un } j] \\ &\geq 1 - \sum_{j=1}^d \mathbb{P}[\mu_j \notin \mathcal{I}_j] \\ &\geq 1 - \sum_{j=1}^d \alpha_j = 1 - \alpha. \end{aligned}$$

Un peu trop conservateur lorsque d est grand.

Ellipsoïde de confiance. Cas de la loi normale.

Soit $\mathbf{X} = (X_1, \dots, X_d)^t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

et $\mathbf{X}_1, \dots, \mathbf{X}_n$ des copies i.i.d. de \mathbf{X} .

On estime $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ par

$$\bar{\mathbf{X}}_n = (\bar{X}_{n1}, \dots, \bar{X}_{nd})^t = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i,$$

$$\hat{\boldsymbol{\Sigma}}_n = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n)^t$$

(moyenne et matrice de covariance empiriques).

Ellipsoïde de confiance. Cas de la loi normale.

Soit $\mathbf{X} = (X_1, \dots, X_d)^t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

et $\mathbf{X}_1, \dots, \mathbf{X}_n$ des copies i.i.d. de \mathbf{X} .

On estime $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ par

$$\bar{\mathbf{X}}_n = (\bar{X}_{n1}, \dots, \bar{X}_{nd})^t = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i,$$

$$\hat{\boldsymbol{\Sigma}}_n = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n)^t$$

(moyenne et matrice de covariance empiriques). Dans ce cas,

$$\frac{n(n-d)}{(n-1)d} (\bar{\mathbf{X}}_n - \boldsymbol{\mu})^t \hat{\boldsymbol{\Sigma}}_n^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \sim F_{d, n-d}$$

Ellipsoïde de confiance. Cas de la loi normale.

Soit $\mathbf{X} = (X_1, \dots, X_d)^t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
 et $\mathbf{X}_1, \dots, \mathbf{X}_n$ des copies i.i.d. de \mathbf{X} .
 On estime $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ par

$$\bar{\mathbf{X}}_n = (\bar{X}_{n1}, \dots, \bar{X}_{nd})^t = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i,$$

$$\hat{\boldsymbol{\Sigma}}_n = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n)^t$$

(moyenne et matrice de covariance empiriques). Dans ce cas,

$$\frac{n(n-d)}{(n-1)d} (\bar{\mathbf{X}}_n - \boldsymbol{\mu})^t \hat{\boldsymbol{\Sigma}}_n^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \sim F_{d, n-d}$$

On cherche f tel que $\mathbb{P}[F_{d, n-d} > f] = \alpha$, et la region où la forme quadratique est $\leq f$ est une **région de confiance** à $(1 - \alpha)\%$. C'est une **ellipsoïde** centrée à $\bar{\mathbf{X}}_n$.

Ellipsoïde de confiance. Cas général.

Quand $n \rightarrow \infty$,

$$n(\bar{\mathbf{X}}_n - \boldsymbol{\mu})^t \hat{\boldsymbol{\Sigma}}_n^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \Rightarrow \chi^2(d).$$

Ellipsoïde de niveau de confiance $1 - \alpha$ pour $\boldsymbol{\mu}$ quand n est grand:

$$\{\boldsymbol{\mu} : n(\bar{\mathbf{X}}_n - \boldsymbol{\mu})^t \hat{\boldsymbol{\Sigma}}_n^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \leq x\},$$

où x est tel que $\mathbb{P}[X \geq x] = \alpha$ si $X \sim \chi^2(d)$.

IC pour fonction d'une ou plusieurs moyennes

On veut un intervalle de confiance pour $g(\boldsymbol{\mu})$ où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ et g est continue mais potentiellement non linéaire.

IC pour fonction d'une ou plusieurs moyennes

On veut un intervalle de confiance pour $g(\boldsymbol{\mu})$ où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ et g est continue mais potentiellement non linéaire.

Cas **déterministe**: si $\mathbf{Y}_n = (Y_{1n}, \dots, Y_{dn}) \rightarrow \boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ et si $g : \mathbb{R}^d \rightarrow \mathbb{R}$ est continue, alors $g(\mathbf{Y}_n) \rightarrow g(\boldsymbol{\mu})$.

Cas **aléatoire**: si $\mathbf{Y}_n \rightarrow \boldsymbol{\mu}$ a.p.1, alors $g(\mathbf{Y}_n) \rightarrow g(\boldsymbol{\mu})$ a.p.1.

IC pour fonction d'une ou plusieurs moyennes

On veut un intervalle de confiance pour $g(\boldsymbol{\mu})$ où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ et g est continue mais potentiellement non linéaire.

Cas **déterministe**: si $\mathbf{Y}_n = (Y_{1n}, \dots, Y_{dn}) \rightarrow \boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ et si $g : \mathbb{R}^d \rightarrow \mathbb{R}$ est continue, alors $g(\mathbf{Y}_n) \rightarrow g(\boldsymbol{\mu})$.

Cas **aléatoire**: si $\mathbf{Y}_n \rightarrow \boldsymbol{\mu}$ a.p.1, alors $g(\mathbf{Y}_n) \rightarrow g(\boldsymbol{\mu})$ a.p.1.

Si \mathbf{Y}_n est une **moyenne de n vecteurs**, on sait que $\sqrt{n}(\mathbf{Y}_n - \boldsymbol{\mu}) \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}_y)$.

IC pour fonction d'une ou plusieurs moyennes

On veut un intervalle de confiance pour $g(\boldsymbol{\mu})$ où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ et g est continue mais potentiellement non linéaire.

Cas **déterministe**: si $\mathbf{Y}_n = (Y_{1n}, \dots, Y_{dn}) \rightarrow \boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ et si $g : \mathbb{R}^d \rightarrow \mathbb{R}$ est continue, alors $g(\mathbf{Y}_n) \rightarrow g(\boldsymbol{\mu})$.

Cas **aléatoire**: si $\mathbf{Y}_n \rightarrow \boldsymbol{\mu}$ a.p.1, alors $g(\mathbf{Y}_n) \rightarrow g(\boldsymbol{\mu})$ a.p.1.

Si \mathbf{Y}_n est une **moyenne de n vecteurs**, on sait que $\sqrt{n}(\mathbf{Y}_n - \boldsymbol{\mu}) \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}_y)$.

Plus généralement, si les \mathbf{Y}_n sont des **vecteurs aléatoires** et si $r(n)(\mathbf{Y}_n - \boldsymbol{\mu}) \Rightarrow \mathbf{Y}$, alors $r(n)(g(\mathbf{Y}_n) - g(\boldsymbol{\mu})) \Rightarrow ?$

IC pour fonction d'une ou plusieurs moyennes

On veut un intervalle de confiance pour $g(\boldsymbol{\mu})$ où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ et g est continue mais potentiellement non linéaire.

Cas **déterministe**: si $\mathbf{Y}_n = (Y_{1n}, \dots, Y_{dn}) \rightarrow \boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ et si $g : \mathbb{R}^d \rightarrow \mathbb{R}$ est continue, alors $g(\mathbf{Y}_n) \rightarrow g(\boldsymbol{\mu})$.

Cas **aléatoire**: si $\mathbf{Y}_n \rightarrow \boldsymbol{\mu}$ a.p.1, alors $g(\mathbf{Y}_n) \rightarrow g(\boldsymbol{\mu})$ a.p.1.

Si \mathbf{Y}_n est une **moyenne de n vecteurs**, on sait que $\sqrt{n}(\mathbf{Y}_n - \boldsymbol{\mu}) \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}_y)$.

Plus généralement, si les \mathbf{Y}_n sont des **vecteurs aléatoires** et si $r(n)(\mathbf{Y}_n - \boldsymbol{\mu}) \Rightarrow \mathbf{Y}$, alors $r(n)(g(\mathbf{Y}_n) - g(\boldsymbol{\mu})) \Rightarrow ?$

Idée: approximer par le premier terme d'une série de Taylor:

$$g(\mathbf{Y}_n) - g(\boldsymbol{\mu}) = (\nabla g(\boldsymbol{\mu}))^\top (\mathbf{Y}_n - \boldsymbol{\mu}) + o(\|\mathbf{Y}_n - \boldsymbol{\mu}\|)$$

quand $n \rightarrow \infty$. Si on multiplie par $r(n)$ des deux cotés, on obtient:

Théorème delta. Soit $g : \mathbb{R}^d \rightarrow \mathbb{R}$ continûment différentiable dans un voisinage de μ , et ∇g son gradient. Si $r(n)(\mathbf{Y}_n - \mu) \Rightarrow \mathbf{Y}$ quand $n \rightarrow \infty$, alors

$$r(n)(g(\mathbf{Y}_n) - g(\mu)) \Rightarrow (\nabla g(\mu))^t \mathbf{Y} \quad \text{quand } n \rightarrow \infty.$$

Théorème delta. Soit $g : \mathbb{R}^d \rightarrow \mathbb{R}$ continûment différentiable dans un voisinage de μ , et ∇g son gradient. Si $r(n)(\mathbf{Y}_n - \mu) \Rightarrow \mathbf{Y}$ quand $n \rightarrow \infty$, alors

$$r(n)(g(\mathbf{Y}_n) - g(\mu)) \Rightarrow (\nabla g(\mu))^t \mathbf{Y} \quad \text{quand } n \rightarrow \infty.$$

Corollaire. Si $\sqrt{n}(\mathbf{Y}_n - \mu) \Rightarrow N(\mathbf{0}, \Sigma_y)$ quand $n \rightarrow \infty$, alors on a le TLC:

$$\frac{\sqrt{n}(g(\mathbf{Y}_n) - g(\mu))}{\sigma_g} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty,$$

où $\sigma_g^2 = (\nabla g(\mu))^t \Sigma_y \nabla g(\mu) = \text{Var}[(\nabla g(\mu))^t \mathbf{Y}]$.

Théorème delta. Soit $g : \mathbb{R}^d \rightarrow \mathbb{R}$ continûment différentiable dans un voisinage de μ , et ∇g son gradient. Si $r(n)(\mathbf{Y}_n - \mu) \Rightarrow \mathbf{Y}$ quand $n \rightarrow \infty$, alors

$$r(n)(g(\mathbf{Y}_n) - g(\mu)) \Rightarrow (\nabla g(\mu))^t \mathbf{Y} \quad \text{quand } n \rightarrow \infty.$$

Corollaire. Si $\sqrt{n}(\mathbf{Y}_n - \mu) \Rightarrow N(\mathbf{0}, \Sigma_y)$ quand $n \rightarrow \infty$, alors on a le TLC:

$$\frac{\sqrt{n}(g(\mathbf{Y}_n) - g(\mu))}{\sigma_g} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty,$$

où $\sigma_g^2 = (\nabla g(\mu))^t \Sigma_y \nabla g(\mu) = \text{Var}[(\nabla g(\mu))^t \mathbf{Y}]$.

Exemples:

- (a) $g(\mu) = \ln \mu$ où $\mu = \mu_1 > 0$;
- (b) $g(\mu_1, \mu_2) = \mu_2 / \mu_1$;
- (c) $\text{Var}[X] = g(\mu_1, \mu_2) = \mu_2 - \mu_1^2$ avec $\mu_k = \mathbb{E}[X^k]$,
- (d) $\text{Cov}[X, Y] = g(\mu_1, \mu_3, \mu_3)$ avec $\mu_1 = \mathbb{E}[X]$, $\mu_2 = \mathbb{E}[Y]$, $\mu_3 = \mathbb{E}[XY]$.

Exemple. (Fonction non linéaire.)

Soit $g(\mu) = \ln \mu$ où $\mu = \mu_1 > 0$;

On a Y_1, \dots, Y_n i.i.d., $\mathbb{E}[Y_i] = \mu$.

On sait calculer un IC à 95% pour μ , disons $[I_1, I_2]$.

Alors $[\ln(I_1), \ln(I_2)]$ est un IC à 95% pour $\ln(\mu)$.

Exemple. (Fonction non linéaire.)

Soit $g(\mu) = \ln \mu$ où $\mu = \mu_1 > 0$;

On a Y_1, \dots, Y_n i.i.d., $\mathbb{E}[Y_i] = \mu$.

On sait calculer un IC à 95% pour μ , disons $[I_1, I_2]$.

Alors $[\ln(I_1), \ln(I_2)]$ est un IC à 95% pour $\ln(\mu)$.

Ce raisonnement tient en général si on remplace \ln par une fonction monotone g .

Exemple. (Fonction non linéaire.)

Soit $g(\mu) = \ln \mu$ où $\mu = \mu_1 > 0$;

On a Y_1, \dots, Y_n i.i.d., $\mathbb{E}[Y_i] = \mu$.

On sait calculer un IC à 95% pour μ , disons $[I_1, I_2]$.

Alors $[\ln(I_1), \ln(I_2)]$ est un IC à 95% pour $\ln(\mu)$.

Ce raisonnement tient en général si on remplace \ln par une fonction monotone g .

Par contre, $\mathbb{E}[\ln(\bar{Y}_n)] < \ln \mathbb{E}[\bar{Y}_n] = \ln(\mu)$ (inégalité de Jensen).

En général, $\mathbb{E}[g(\bar{Y}_n)] \neq g(\mu)$.

Exemple. (Fonction non linéaire.)

Soit $g(\mu) = \ln \mu$ où $\mu = \mu_1 > 0$;

On a Y_1, \dots, Y_n i.i.d., $\mathbb{E}[Y_i] = \mu$.

On sait calculer un IC à 95% pour μ , disons $[I_1, I_2]$.

Alors $[\ln(I_1), \ln(I_2)]$ est un IC à 95% pour $\ln(\mu)$.

Ce raisonnement tient en général si on remplace \ln par une fonction monotone g .

Par contre, $\mathbb{E}[\ln(\bar{Y}_n)] < \ln \mathbb{E}[\bar{Y}_n] = \ln(\mu)$ (inégalité de Jensen).

En général, $\mathbb{E}[g(\bar{Y}_n)] \neq g(\mu)$.

Si on applique le théorème delta pour $g(\mu) = \ln \mu$: $\nabla g(\mu) = \partial \ln \mu / \partial \mu = 1/\mu$.

Si $\sqrt{n}(\bar{Y}_n - \mu)/S_n \Rightarrow N(0, 1)$, alors

$$\sqrt{n}(\ln(\bar{Y}_n) - \ln \mu) \bar{Y}_n / S_n \Rightarrow \sqrt{n}(\ln(\bar{Y}_n) - \ln \mu) / \sigma_g \Rightarrow N(0, 1)$$

où $\sigma_g^2 = \text{Var}[Y_i] / \mu^2 \approx S_n^2 / \bar{Y}_n^2$. L'IC devient:

$$[\ln(\bar{Y}_n) \pm \Phi^{-1}(1 - \alpha/2) S_n / (\bar{Y}_n \sqrt{n})].$$

Il diffère de l'IC précédent. **Exercice:** comparer les deux numériquement.

Terme de second ordre.

En principe, on peut ajouter un terme à la série de Taylor, si g est deux fois continûment différentiable:

$$g(\mathbf{Y}_n) - g(\boldsymbol{\mu}) = (\nabla g(\boldsymbol{\mu}))^t(\mathbf{Y}_n - \boldsymbol{\mu}) + (\mathbf{Y}_n - \boldsymbol{\mu})^t \mathbf{H}(\boldsymbol{\mu})(\mathbf{Y}_n - \boldsymbol{\mu})/2 + o(\|\mathbf{Y}_n - \boldsymbol{\mu}\|^2),$$

où \mathbf{H} est la matrice Hessienne de g à $\boldsymbol{\mu}$. Si on dispose d'un bon estimateur du terme additionnel, on peut le soustraire pour réduire le biais.

Par exemple, si $\mathbf{Y}_n = \bar{\mathbf{X}}_n$, la moyenne de n v.a. i.i.d. $\mathbf{X}_1, \dots, \mathbf{X}_n$, le terme de correction est

$$\frac{(\bar{\mathbf{X}}_n - \boldsymbol{\mu})^t \mathbf{H}(\boldsymbol{\mu})(\bar{\mathbf{X}}_n - \boldsymbol{\mu})}{2} = \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{X}_i - \boldsymbol{\mu})^t \mathbf{H}(\boldsymbol{\mu})(\mathbf{X}_j - \boldsymbol{\mu}).$$

Puisque $\mathbb{E}[(\mathbf{X}_i - \boldsymbol{\mu})^t \mathbf{H}(\boldsymbol{\mu})(\mathbf{X}_j - \boldsymbol{\mu})] = 0$ pour $i \neq j$, on peut estimer le terme de correction par

$$\frac{1}{2n(n-1)} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)^t \mathbf{H}(\bar{\mathbf{X}}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n).$$

Cela donne l'estimateur de $g(\boldsymbol{\mu})$ avec biais réduit:

$$g(\boldsymbol{\mu}) \approx g(\bar{\mathbf{X}}_n) - \frac{1}{2n(n-1)} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)^t \mathbf{H}(\bar{\mathbf{X}}_n) (\mathbf{X}_i - \bar{\mathbf{X}}_n).$$

Moins de biais pour n grand, mais pourrait avoir un plus grand MSE si la variance du terme de correction est grande.

Exige aussi le calcul du Hessien, qui peut être difficile surtout si $d > 1$.

Ne change pas le TLC, car le terme de correction converge en $\mathcal{O}(1/n)$ alors que l'erreur converge en $\mathcal{O}(1/\sqrt{n})$. Donc la correction est potentiellement utile seulement si n n'est pas très grand.

Cela donne l'estimateur de $g(\boldsymbol{\mu})$ avec biais réduit:

$$g(\boldsymbol{\mu}) \approx g(\bar{\mathbf{X}}_n) - \frac{1}{2n(n-1)} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)^t \mathbf{H}(\bar{\mathbf{X}}_n) (\mathbf{X}_i - \bar{\mathbf{X}}_n).$$

Moins de biais pour n grand, mais pourrait avoir un plus grand MSE si la variance du terme de correction est grande.

Exige aussi le calcul du Hessien, qui peut être difficile surtout si $d > 1$.

Ne change pas le TLC, car le terme de correction converge en $\mathcal{O}(1/n)$ alors que l'erreur converge en $\mathcal{O}(1/\sqrt{n})$. Donc la correction est potentiellement utile seulement si n n'est pas très grand.

Exemple: $g(\mu) = \ln \mu$: $g''(\mu) = -1/\mu^2$ et l'estimateur corrigé devient

$$\ln(\bar{Y}_n) + \frac{1}{2n(n-1)(\bar{Y}_n)^2} \sum_{j=1}^n (Y_j - \bar{Y}_n)^2 = \ln(\bar{Y}_n) + \frac{1}{2n(n-1)} \sum_{j=1}^n (Y_j/\bar{Y}_n - 1)^2.$$



Exemple. (Quotient de deux espérances.)

Soient $(X_1, Y_1), \dots, (X_n, Y_n)$ des copies i.i.d. de (X, Y) et supposons que l'on estime $\nu = \mathbb{E}[X]/\mathbb{E}[Y]$ par

$$\hat{\nu}_n = \frac{\bar{X}_n}{\bar{Y}_n} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i}.$$

Exemple. (Quotient de deux espérances.)

Soient $(X_1, Y_1), \dots, (X_n, Y_n)$ des copies i.i.d. de (X, Y) et supposons que l'on estime $\nu = \mathbb{E}[X]/\mathbb{E}[Y]$ par

$$\hat{\nu}_n = \frac{\bar{X}_n}{\bar{Y}_n} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i}.$$

Cet estimateur est **biaisé** mais fortement consistant.

Exemple. (Quotient de deux espérances.)

Soient $(X_1, Y_1), \dots, (X_n, Y_n)$ des copies i.i.d. de (X, Y) et supposons que l'on estime $\nu = \mathbb{E}[X]/\mathbb{E}[Y]$ par

$$\hat{\nu}_n = \frac{\bar{X}_n}{\bar{Y}_n} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i}.$$

Cet estimateur est **biaisé** mais fortement consistant.

Posons $\mu_1 = \mathbb{E}[X]$, $\mu_2 = \mathbb{E}[Y]$, $g(\mu_1, \mu_2) = \mu_1/\mu_2$, $\sigma_1^2 = \text{Var}[X]$, $\sigma_2^2 = \text{Var}[Y]$, et $\sigma_{12} = \text{Cov}[X, Y]$.

Exemple. (Quotient de deux espérances.)

Soient $(X_1, Y_1), \dots, (X_n, Y_n)$ des copies i.i.d. de (X, Y) et supposons que l'on estime $\nu = \mathbb{E}[X]/\mathbb{E}[Y]$ par

$$\hat{\nu}_n = \frac{\bar{X}_n}{\bar{Y}_n} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i}.$$

Cet estimateur est **biaisé** mais fortement consistant.

Posons $\mu_1 = \mathbb{E}[X]$, $\mu_2 = \mathbb{E}[Y]$, $g(\mu_1, \mu_2) = \mu_1/\mu_2$, $\sigma_1^2 = \text{Var}[X]$, $\sigma_2^2 = \text{Var}[Y]$, et $\sigma_{12} = \text{Cov}[X, Y]$.

On suppose que ces quantités sont finies et que $\mu_2 \neq 0$, $\sigma_1^2 > 0$, et $\sigma_2^2 > 0$.

Exemple. (Quotient de deux espérances.)

Soient $(X_1, Y_1), \dots, (X_n, Y_n)$ des copies i.i.d. de (X, Y) et supposons que l'on estime $\nu = \mathbb{E}[X]/\mathbb{E}[Y]$ par

$$\hat{\nu}_n = \frac{\bar{X}_n}{\bar{Y}_n} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i}.$$

Cet estimateur est **biaisé** mais fortement consistant.

Posons $\mu_1 = \mathbb{E}[X]$, $\mu_2 = \mathbb{E}[Y]$, $g(\mu_1, \mu_2) = \mu_1/\mu_2$, $\sigma_1^2 = \text{Var}[X]$, $\sigma_2^2 = \text{Var}[Y]$, et $\sigma_{12} = \text{Cov}[X, Y]$.

On suppose que ces quantités sont finies et que $\mu_2 \neq 0$, $\sigma_1^2 > 0$, et $\sigma_2^2 > 0$.

Le gradient de g est $\nabla g(\mu_1, \mu_2) = (1/\mu_2, -\mu_1/\mu_2^2)^\top$.

Le TLC ordinaire nous dit que

$$\sqrt{n}(\bar{X}_n - \mu_1, \bar{Y}_n - \mu_2)^t \Rightarrow (W_1, W_2)^t \sim N(\mathbf{0}, \Sigma_y)$$

où

$$\Sigma_y = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Le TLC ordinaire nous dit que

$$\sqrt{n}(\bar{X}_n - \mu_1, \bar{Y}_n - \mu_2)^t \Rightarrow (\mathbf{W}_1, \mathbf{W}_2)^t \sim N(\mathbf{0}, \Sigma_y)$$

où

$$\Sigma_y = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Puis, par the théorème delta (ou son corollaire),

$$\sqrt{n}(\hat{\nu}_n - \nu) \Rightarrow (W_1, W_2) \cdot \nabla g(\mu_1, \mu_2) = W_1/\mu_2 - W_2\mu_1/\mu_2^2 \sim N(0, \sigma_g^2)$$

où

$$\begin{aligned} \sigma_g^2 &= (\nabla g(\boldsymbol{\mu}))^t \Sigma_y \nabla g(\boldsymbol{\mu}) \\ &= (1/\mu_2, -\mu_1/\mu_2^2) \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1/\mu_2 \\ -\mu_1/\mu_2^2 \end{pmatrix} \\ &= \sigma_1^2/\mu_2^2 + \sigma_2^2\mu_1^2/\mu_2^4 - 2\sigma_{12}\mu_1/\mu_2^3 \\ &= (\sigma_1^2 + \sigma_2^2\nu^2 - 2\sigma_{12}\nu)/\mu_2^2. \end{aligned}$$

$$\sigma_g^2 = \frac{\sigma_1^2 + \sigma_2^2 \nu^2 - 2\sigma_{12}\nu}{\mu_2^2}.$$

On peut calculer un IC en utilisant ce dernier TLC si on dispose d'un bon estimateur de σ_g^2 . Un candidat évident est:

$$\hat{\sigma}_{g,n}^2 = \frac{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 \hat{\nu}_n^2 - 2\hat{\sigma}_{12}\hat{\nu}_n}{(\bar{Y}_n)^2},$$

où

$$\hat{\sigma}_1^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2,$$

$$\hat{\sigma}_2^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y}_n)^2,$$

$$\hat{\sigma}_{12}^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)(Y_j - \bar{Y}_n).$$

Puisque $\hat{\sigma}_{g,n}^2$ est fortement consistant, on obtient le TLC

$$\frac{\sqrt{n}(\hat{\nu}_n - \nu)}{\hat{\sigma}_{g,n}} \Rightarrow \frac{\sqrt{n}(\hat{\nu}_n - \nu)}{\sigma_g} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty.$$

Puisque $\hat{\sigma}_{g,n}^2$ est fortement consistant, on obtient le TLC

$$\frac{\sqrt{n}(\hat{\nu}_n - \nu)}{\hat{\sigma}_{g,n}} \Rightarrow \frac{\sqrt{n}(\hat{\nu}_n - \nu)}{\sigma_g} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty.$$

L'IC classique pour ν au niveau nominal $1 - \alpha$ est $(\hat{\nu}_n - r, \hat{\nu}_n + r)$ où $r = z_{1-\alpha/2} \hat{\sigma}_{g,n} / \sqrt{n}$.

Puisque $\hat{\sigma}_{g,n}^2$ est fortement consistant, on obtient le TLC

$$\frac{\sqrt{n}(\hat{\nu}_n - \nu)}{\hat{\sigma}_{g,n}} \Rightarrow \frac{\sqrt{n}(\hat{\nu}_n - \nu)}{\sigma_g} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty.$$

L'IC **classique** pour ν au niveau nominal $1 - \alpha$ est $(\hat{\nu}_n - r, \hat{\nu}_n + r)$ où $r = z_{1-\alpha/2} \hat{\sigma}_{g,n} / \sqrt{n}$.

Son erreur de couverture est parfois grande lorsque n n'est pas très grand, ou lorsque la convergence vers $N(0, 1)$ est lente.

Puisque $\hat{\sigma}_{g,n}^2$ est fortement consistant, on obtient le TLC

$$\frac{\sqrt{n}(\hat{\nu}_n - \nu)}{\hat{\sigma}_{g,n}} \Rightarrow \frac{\sqrt{n}(\hat{\nu}_n - \nu)}{\sigma_g} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty.$$

L'IC **classique** pour ν au niveau nominal $1 - \alpha$ est $(\hat{\nu}_n - r, \hat{\nu}_n + r)$ où $r = z_{1-\alpha/2} \hat{\sigma}_{g,n} / \sqrt{n}$.

Son erreur de couverture est parfois grande lorsque n n'est pas très grand, ou lorsque la convergence vers $N(0, 1)$ est lente.

Dans ce cas, on recommande d'utiliser le **bootstrap- t non-paramétrique**, en prenant $\hat{\nu}_n$ et $\hat{\sigma}_{g,n}^2$ comme estimateurs de la moyenne et de la variance.

Pour le cas particulier d'un rapport de deux espérances, la dérivation suivante est plus directe. Les v.a.

$$Z_j = X_j - \mu Y_j,$$

sont i.i.d. de moyenne 0 et variance

$$\sigma_z^2 = \text{Var}[Z_j] = \text{Var}[X_j] + \mu^2 \text{Var}[Y_j] - 2\mu \text{Cov}(X_j, Y_j).$$

Pour le cas particulier d'un rapport de deux espérances, la dérivation suivante est plus directe. Les v.a.

$$Z_j = X_j - \mu Y_j,$$

sont i.i.d. de moyenne 0 et variance

$$\sigma_z^2 = \text{Var}[Z_j] = \text{Var}[X_j] + \mu^2 \text{Var}[Y_j] - 2\mu \text{Cov}(X_j, Y_j).$$

En appliquant le TLC aux Z_j , on obtient

$$\frac{\sqrt{n}\bar{Y}_n(\hat{\mu}_n - \mu)}{\sigma_z} = \frac{\sqrt{n}\bar{Z}_n}{\sigma_z} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty.$$

Pour le cas particulier d'un rapport de deux espérances, la dérivation suivante est plus directe. Les v.a.

$$Z_j = X_j - \mu Y_j,$$

sont i.i.d. de moyenne 0 et variance

$$\sigma_z^2 = \text{Var}[Z_j] = \text{Var}[X_j] + \mu^2 \text{Var}[Y_j] - 2\mu \text{Cov}(X_j, Y_j).$$

En appliquant le TLC aux Z_j , on obtient

$$\frac{\sqrt{n}\bar{Y}_n(\hat{\mu}_n - \mu)}{\sigma_z} = \frac{\sqrt{n}\bar{Z}_n}{\sigma_z} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty.$$

C'est **équivalent**, car $\sigma_z/\bar{Y}_n \xrightarrow{\text{p.s.}} \sigma_z/\mu_2 = \sigma_g$ quand $n \rightarrow \infty$.

Pour le cas particulier d'un rapport de deux espérances, la dérivation suivante est plus directe. Les v.a.

$$Z_j = X_j - \mu Y_j,$$

sont i.i.d. de moyenne 0 et variance

$$\sigma_z^2 = \text{Var}[Z_j] = \text{Var}[X_j] + \mu^2 \text{Var}[Y_j] - 2\mu \text{Cov}(X_j, Y_j).$$

En appliquant le TLC aux Z_j , on obtient

$$\frac{\sqrt{n}\bar{Y}_n(\hat{\mu}_n - \mu)}{\sigma_z} = \frac{\sqrt{n}\bar{Z}_n}{\sigma_z} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty.$$

C'est équivalent, car $\sigma_z/\bar{Y}_n \xrightarrow{\text{p.s.}} \sigma_z/\mu_2 = \sigma_g$ quand $n \rightarrow \infty$.

Remarque importante: on préfère $\text{Cov}(X_j, Y_j) > 0$!

Différence entre deux moyennes

On a n_1 observations i.i.d. X_{11}, \dots, X_{1,n_1} , de moyenne μ_1 ,
et n_2 observations i.i.d. X_{21}, \dots, X_{2,n_2} , de moyenne μ_2 .

Différence entre deux moyennes

On a n_1 observations i.i.d. X_{11}, \dots, X_{1,n_1} , de moyenne μ_1 ,
et n_2 observations i.i.d. X_{21}, \dots, X_{2,n_2} , de moyenne μ_2 .

On veut un IC pour $\mu_1 - \mu_2$.

Différence entre deux moyennes

On a n_1 observations i.i.d. X_{11}, \dots, X_{1,n_1} , de moyenne μ_1 ,
et n_2 observations i.i.d. X_{21}, \dots, X_{2,n_2} , de moyenne μ_2 .

On veut un IC pour $\mu_1 - \mu_2$.

Les deux méthodes suivantes supposent que les X_{ji} suivent la loi normale.
(Pas toujours valide!)

Différence entre deux moyennes

On a n_1 observations i.i.d. X_{11}, \dots, X_{1,n_1} , de moyenne μ_1 ,
et n_2 observations i.i.d. X_{21}, \dots, X_{2,n_2} , de moyenne μ_2 .

On veut un IC pour $\mu_1 - \mu_2$.

Les deux méthodes suivantes supposent que les X_{ji} suivent la loi normale.
(Pas toujours valide!)

Dans la seconde (Welch), les deux échantillons doivent être indépendants mais on peut avoir $n_1 \neq n_2$.

Différence entre deux moyennes

On a n_1 observations i.i.d. X_{11}, \dots, X_{1,n_1} , de moyenne μ_1 ,
et n_2 observations i.i.d. X_{21}, \dots, X_{2,n_2} , de moyenne μ_2 .

On veut un IC pour $\mu_1 - \mu_2$.

Les deux méthodes suivantes supposent que les X_{ji} suivent la loi normale.
(Pas toujours valide!)

Dans la seconde (Welch), les deux échantillons doivent être indépendants mais on peut avoir $n_1 \neq n_2$.

Dans la première, il faut $n_1 = n_2$ mais X_{1i} et X_{2i} peuvent être corrélés.

Observations couplées.

Soit $n_1 = n_2 = n$. Posons $Z_i = X_{1i} - X_{2i}$ pour $1 \leq i \leq n$,

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i, \quad \text{et} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2.$$

Observations couplées.

Soit $n_1 = n_2 = n$. Posons $Z_i = X_{1i} - X_{2i}$ pour $1 \leq i \leq n$,

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i, \quad \text{et} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2.$$

Puisque les Z_i sont i.i.d. normales,

$$\frac{\sqrt{n}[\bar{Z}_n - (\mu_1 - \mu_2)]}{S_n} \sim \text{Student}(n-1).$$

On utilise cela pour calculer l'IC.

Observations couplées.

Soit $n_1 = n_2 = n$. Posons $Z_i = X_{1i} - X_{2i}$ pour $1 \leq i \leq n$,

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i, \quad \text{et} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2.$$

Puisque les Z_i sont i.i.d. normales,

$$\frac{\sqrt{n}[\bar{Z}_n - (\mu_1 - \mu_2)]}{S_n} \sim \text{Student}(n-1).$$

On utilise cela pour calculer l'IC.

Puisque

$$\text{Var}[Z_i] = \text{Var}[X_{1i}] + \text{Var}[X_{2i}] - 2\text{Cov}[X_{1i}, X_{2i}],$$

il est avantageux d'avoir $\text{Cov}[X_{1i}, X_{2i}] > 0$.

Méthode de Welch.

On suppose X_{1i} et X_{2i} indépendants. Soit

$$\bar{X}_{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ki} \quad \text{et} \quad S_{(k)}^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_{ki} - \bar{X}_{(k)})^2,$$

pour $k = 1, 2$. Alors,

$$\frac{\bar{X}_{(1)} - \bar{X}_{(2)} - (\mu_1 - \mu_2)}{[S_{(1)}^2/n_1 + S_{(2)}^2/n_2]^{1/2}} \approx \text{Student}(\hat{\ell})$$

où

$$\hat{\ell} = \frac{[S_{(1)}^2/n_1 + S_{(2)}^2/n_2]^2}{[S_{(1)}^2/n_1]^2/(n_1 - 1) + [S_{(2)}^2/n_2]^2/(n_2 - 1)}.$$

Estimation de quantiles

Si X est de répartition F , le q -quantile de F est

$$\xi_q = F^{-1}(q) = \inf\{x : F(x) \geq q\}.$$

Estimation de quantiles

Si X est de répartition F , le q -quantile de F est

$$\xi_q = F^{-1}(q) = \inf\{x : F(x) \geq q\}.$$

Soit $X_{(1)}, \dots, X_{(n)}$ un échantillon i.i.d. de X , trié, et \hat{F}_n la f.r. empirique.

Estimation de quantiles

Si X est de répartition F , le q -quantile de F est

$$\xi_q = F^{-1}(q) = \inf\{x : F(x) \geq q\}.$$

Soit $X_{(1)}, \dots, X_{(n)}$ un échantillon i.i.d. de X , trié, et \hat{F}_n la f.r. empirique.

Un estimateur simple de ξ_q est le **quantile empirique**

$$\hat{\xi}_{q,n} = \hat{F}_n^{-1}(q) = \inf\{x : \hat{F}_n(x) \geq q\} = X_{(\lceil nq \rceil)}.$$

Estimation de quantiles

Si X est de répartition F , le q -quantile de F est

$$\xi_q = F^{-1}(q) = \inf\{x : F(x) \geq q\}.$$

Soit $X_{(1)}, \dots, X_{(n)}$ un échantillon i.i.d. de X , trié, et \hat{F}_n la f.r. empirique.

Un estimateur simple de ξ_q est le **quantile empirique**

$$\hat{\xi}_{q,n} = \hat{F}_n^{-1}(q) = \inf\{x : \hat{F}_n(x) \geq q\} = X_{(\lceil nq \rceil)}.$$

Théorème.

- (i) Pour chaque q , $\hat{\xi}_{q,n} \xrightarrow{\text{p.s.}} \xi_q$ quand $n \rightarrow \infty$ (fortement consistant).
- (ii) Si X a une densité $f > 0$ et continue dans un voisinage de ξ_q , alors (TLC):

$$\frac{\sqrt{n}(\hat{\xi}_{q,n} - \xi_q)f(\xi_q)}{\sqrt{q(1-q)}} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty.$$

Estimation de quantiles

Si X est de répartition F , le q -quantile de F est

$$\xi_q = F^{-1}(q) = \inf\{x : F(x) \geq q\}.$$

Soit $X_{(1)}, \dots, X_{(n)}$ un échantillon i.i.d. de X , trié, et \hat{F}_n la f.r. empirique.

Un estimateur simple de ξ_q est le **quantile empirique**

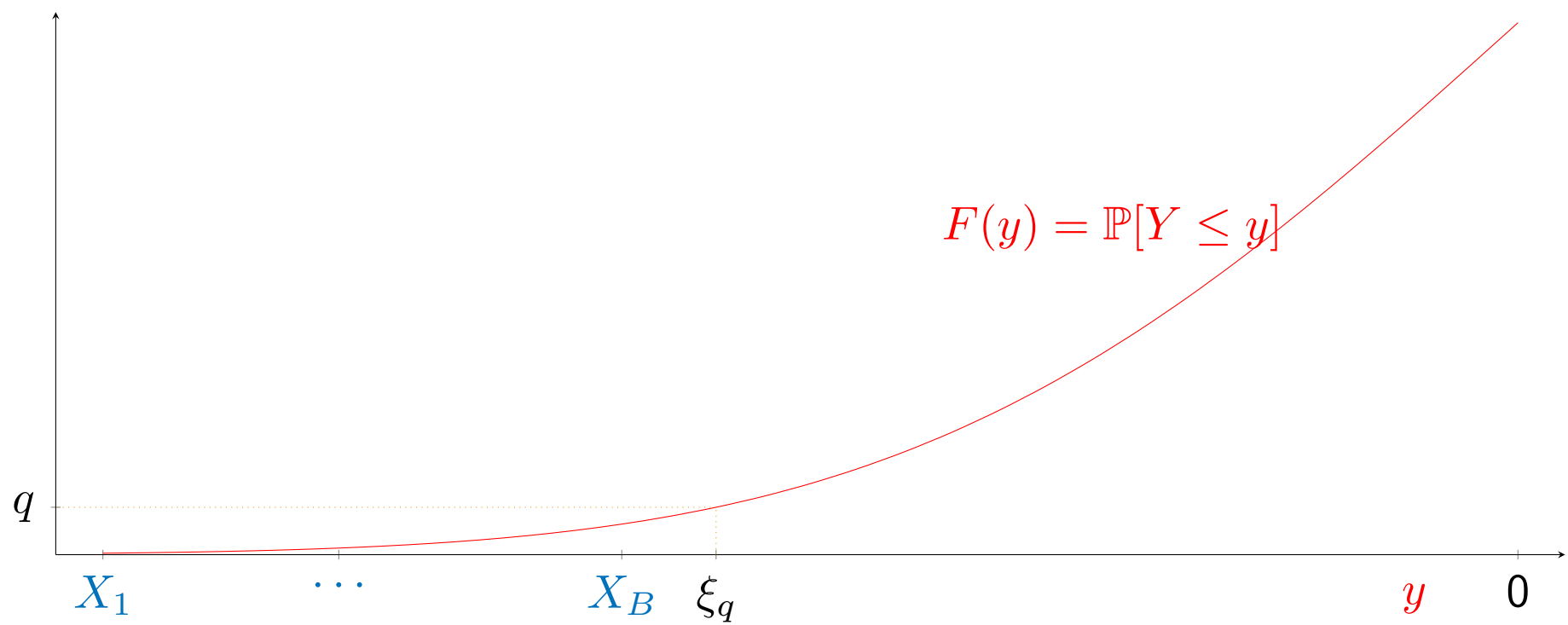
$$\hat{\xi}_{q,n} = \hat{F}_n^{-1}(q) = \inf\{x : \hat{F}_n(x) \geq q\} = X_{(\lceil nq \rceil)}.$$

Théorème.

- (i) Pour chaque q , $\hat{\xi}_{q,n} \xrightarrow{\text{p.s.}} \xi_q$ quand $n \rightarrow \infty$ (fortement consistant).
- (ii) Si X a une densité $f > 0$ et continue dans un voisinage de ξ_q , alors (TLC):

$$\frac{\sqrt{n}(\hat{\xi}_{q,n} - \xi_q)f(\xi_q)}{\sqrt{q(1-q)}} \Rightarrow N(0, 1) \quad \text{quand } n \rightarrow \infty.$$

Cela implique que $\text{Var}[\hat{\xi}_{q,n}] \approx \frac{q(1-q)}{nf^2(\xi_q)}.$



Ce TLC indique: beaucoup de bruit (variance) si $f(\xi_q)$ est petit.
Pour l'utiliser pour IC, il faut estimer $f(\xi_q)$: difficile.

Ce TLC indique: beaucoup de bruit (variance) si $f(\xi_q)$ est petit.
Pour l'utiliser pour IC, il faut estimer $f(\xi_q)$: difficile.

Méthode **non-asymptotique** de calcul d'un IC pour ξ_q :
Supposons que F est continue à ξ_q .

Ce TLC indique: beaucoup de bruit (variance) si $f(\xi_q)$ est petit.
Pour l'utiliser pour IC, il faut estimer $f(\xi_q)$: difficile.

Méthode **non-asymptotique** de calcul d'un IC pour ξ_q :

Supposons que F est continue à ξ_q .

Soit B le nombre d'observations $X_{(i)}$ inférieures à ξ_q .

Puisque $\mathbb{P}[X < \xi_q] = q$, B est binomiale(n, q).

Ce TLC indique: beaucoup de bruit (variance) si $f(\xi_q)$ est petit.
 Pour l'utiliser pour IC, il faut estimer $f(\xi_q)$: difficile.

Méthode **non-asymptotique** de calcul d'un IC pour ξ_q :

Supposons que F est continue à ξ_q .

Soit B le nombre d'observations $X_{(i)}$ inférieures à ξ_q .

Puisque $\mathbb{P}[X < \xi_q] = q$, B est binomiale(n, q).

Si $1 \leq j < k \leq n$, on a

$$\mathbb{P}[X_{(j)} < \xi_q \leq X_{(k)}] = \mathbb{P}[j \leq B < k] = \sum_{i=j}^{k-1} \binom{n}{i} q^i (1-q)^{n-i}.$$

Ce TLC indique: beaucoup de bruit (variance) si $f(\xi_q)$ est petit.
 Pour l'utiliser pour IC, il faut estimer $f(\xi_q)$: difficile.

Méthode **non-asymptotique** de calcul d'un IC pour ξ_q :

Supposons que F est continue à ξ_q .

Soit B le nombre d'observations $X_{(i)}$ inférieures à ξ_q .

Puisque $\mathbb{P}[X < \xi_q] = q$, B est binomiale(n, q).

Si $1 \leq j < k \leq n$, on a

$$\mathbb{P}[X_{(j)} < \xi_q \leq X_{(k)}] = \mathbb{P}[j \leq B < k] = \sum_{i=j}^{k-1} \binom{n}{i} q^i (1-q)^{n-i}.$$

On choisit j et k pour que cette somme soit $\geq 1 - \alpha$.
 (Intervalle unilatéral ou bilatéral.)

Ce TLC indique: beaucoup de bruit (variance) si $f(\xi_q)$ est petit.
 Pour l'utiliser pour IC, il faut estimer $f(\xi_q)$: difficile.

Méthode **non-asymptotique** de calcul d'un IC pour ξ_q :

Supposons que F est continue à ξ_q .

Soit B le nombre d'observations $X_{(i)}$ inférieures à ξ_q .

Puisque $\mathbb{P}[X < \xi_q] = q$, B est binomiale(n, q).

Si $1 \leq j < k \leq n$, on a

$$\mathbb{P}[X_{(j)} < \xi_q \leq X_{(k)}] = \mathbb{P}[j \leq B < k] = \sum_{i=j}^{k-1} \binom{n}{i} q^i (1-q)^{n-i}.$$

On choisit j et k pour que cette somme soit $\geq 1 - \alpha$.

(Intervalle unilatéral ou bilatéral.)

Si n est grand et q pas trop proche de 0 ou 1, on peut approximer la loi binomiale par la loi **normale**: $(B - nq) / \sqrt{nq(1-q)} \approx N(0, 1)$.

Ce TLC indique: beaucoup de bruit (variance) si $f(\xi_q)$ est petit.
 Pour l'utiliser pour IC, il faut estimer $f(\xi_q)$: difficile.

Méthode **non-asymptotique** de calcul d'un IC pour ξ_q :

Supposons que F est continue à ξ_q .

Soit B le nombre d'observations $X_{(i)}$ inférieures à ξ_q .

Puisque $\mathbb{P}[X < \xi_q] = q$, B est binomiale(n, q).

Si $1 \leq j < k \leq n$, on a

$$\mathbb{P}[X_{(j)} < \xi_q \leq X_{(k)}] = \mathbb{P}[j \leq B < k] = \sum_{i=j}^{k-1} \binom{n}{i} q^i (1-q)^{n-i}.$$

On choisit j et k pour que cette somme soit $\geq 1 - \alpha$.

(Intervalle unilatéral ou bilatéral.)

Si n est grand et q pas trop proche de 0 ou 1, on peut approximer la loi binomiale par la loi **normale**: $(B - nq) / \sqrt{nq(1-q)} \approx N(0, 1)$.

On obtient alors $j = \lfloor nq + 1 - \delta \rfloor$ et $k = \lfloor nq + 1 + \delta \rfloor$, où $\delta = \sqrt{nq(1-q)} \Phi^{-1}(1 - \alpha/2)$.

Ce TLC indique: beaucoup de bruit (variance) si $f(\xi_q)$ est petit.
 Pour l'utiliser pour IC, il faut estimer $f(\xi_q)$: difficile.

Méthode **non-asymptotique** de calcul d'un IC pour ξ_q :

Supposons que F est continue à ξ_q .

Soit B le nombre d'observations $X_{(i)}$ inférieures à ξ_q .

Puisque $\mathbb{P}[X < \xi_q] = q$, B est binomiale(n, q).

Si $1 \leq j < k \leq n$, on a

$$\mathbb{P}[X_{(j)} < \xi_q \leq X_{(k)}] = \mathbb{P}[j \leq B < k] = \sum_{i=j}^{k-1} \binom{n}{i} q^i (1-q)^{n-i}.$$

On choisit j et k pour que cette somme soit $\geq 1 - \alpha$.

(Intervalle unilatéral ou bilatéral.)

Si n est grand et q pas trop proche de 0 ou 1, on peut approximer la loi binomiale par la loi **normale**: $(B - nq) / \sqrt{nq(1-q)} \approx N(0, 1)$.

On obtient alors $j = \lfloor nq + 1 - \delta \rfloor$ et $k = \lfloor nq + 1 + \delta \rfloor$, où $\delta = \sqrt{nq(1-q)} \Phi^{-1}(1 - \alpha/2)$.

Pour améliorer l'estimateur, on peut remplacer \hat{F}_n par une fonction plus lisse (e.g., loi quasi-empirique, estimateur de densité par noyau, etc.). Dans tous les cas, l'estimateur est quand même biaisé, mais on peut diminuer le MSE.

Pour améliorer l'estimateur, on peut remplacer \hat{F}_n par une fonction plus lisse (e.g., loi quasi-empirique, estimateur de densité par noyau, etc.).

Dans tous les cas, l'estimateur est quand même biaisé, mais on peut diminuer le MSE.

Exemple: Avramidis et Wilson (1998).

Exemple: la valeur à risque.

Soit L la perte nette de valeur d'un porte-feuille d'actifs pour une période de temps donnée $[0, T]$.

La **valeur à risque (VaR)** (au temps 0) est la valeur de x_p telle que $\mathbb{P}[L > x_p] = p$.
C'est le $(1 - p)$ -quantile de L .

Exemple: la valeur à risque.

Soit L la perte nette de valeur d'un porte-feuille d'actifs pour une période de temps donnée $[0, T]$.

La **valeur à risque (VaR)** (au temps 0) est la valeur de x_p telle que $\mathbb{P}[L > x_p] = p$. C'est le $(1 - p)$ -quantile de L .

Valeurs courantes: $p = 0.01$, $T = 2$ semaines (banques), $T =$ mois ou années (assurance, fonds de pension).

On peut critiquer l'utilisation de la VaR: donne une information très limitée. Par exemple si $x_{0.01} = 10^7$ dollars, que sait-on?

Exemple: la valeur à risque.

Soit L la perte nette de valeur d'un porte-feuille d'actifs pour une période de temps donnée $[0, T]$.

La **valeur à risque (VaR)** (au temps 0) est la valeur de x_p telle que $\mathbb{P}[L > x_p] = p$. C'est le $(1 - p)$ -quantile de L .

Valeurs courantes: $p = 0.01$, $T = 2$ semaines (banques), $T =$ mois ou années (assurance, fonds de pension).

On peut critiquer l'utilisation de la VaR: donne une information très limitée. Par exemple si $x_{0.01} = 10^7$ dollars, que sait-on?

Une mesure complémentaire est la valeur à risque conditionnelle (CVaR):

$$\mathbb{E}[L \mid L > x_p].$$

Exemple: la valeur à risque.

Soit L la perte nette de valeur d'un porte-feuille d'actifs pour une période de temps donnée $[0, T]$.

La **valeur à risque (VaR)** (au temps 0) est la valeur de x_p telle que $\mathbb{P}[L > x_p] = p$. C'est le $(1 - p)$ -quantile de L .

Valeurs courantes: $p = 0.01$, $T = 2$ semaines (banques), $T =$ mois ou années (assurance, fonds de pension).

On peut critiquer l'utilisation de la VaR: donne une information très limitée. Par exemple si $x_{0.01} = 10^7$ dollars, que sait-on?

Une mesure complémentaire est la valeur à risque conditionnelle (CVaR):

$$\mathbb{E}[L \mid L > x_p].$$

Modèles pour estimer la VaR: On doit modéliser l'évolution du prix des actifs (souvent plusieurs milliers, dépendants). Souvent: modèles à facteurs.

On peut remplacer les actifs par des prêts, comptes à payer, etc.

Sauf dans les cas simples, on estime la VaR par simulation.

Quand p est petit: "importance sampling".

CI par rééchantillonnage (“bootstrap”)

Ce sont des techniques de simulation appliquées en statistique.

CI par rééchantillonnage (“bootstrap”)

Ce sont des techniques de simulation appliquées en statistique.

L'idée est d'estimer la distribution (inconnue et quelconque) de l'estimateur en **rééchantillonnant** des échantillons de taille n en tirant **avec remplacement** dans l'échantillon de taille n original. Pour chaque échantillon, on calcule l'estimateur.

CI par rééchantillonnage (“bootstrap”)

Ce sont des techniques de simulation appliquées en statistique.

L'idée est d'estimer la distribution (inconnue et quelconque) de l'estimateur en **rééchantillonnant** des échantillons de taille n en tirant **avec remplacement** dans l'échantillon de taille n original. Pour chaque échantillon, on calcule l'estimateur.

Bootstrap non-paramétrique de base.

On a un échantillon i.i.d. X_1, \dots, X_n d'une loi de f.r. F et un estimateur $Y = Y_n = g(X_1, \dots, X_n)$ d'une valeur réelle inconnue θ .

CI par rééchantillonnage (“bootstrap”)

Ce sont des techniques de simulation appliquées en statistique.

L'idée est d'estimer la distribution (inconnue et quelconque) de l'estimateur en **rééchantillonnant** des échantillons de taille n en tirant **avec remplacement** dans l'échantillon de taille n original. Pour chaque échantillon, on calcule l'estimateur.

Bootstrap non-paramétrique de base.

On a un échantillon i.i.d. X_1, \dots, X_n d'une loi de f.r. F et un estimateur $Y = Y_n = g(X_1, \dots, X_n)$ d'une valeur réelle inconnue θ .

Par exemple, $Y_n = \bar{X}_n$ avec $\theta = \mu$, ou $Y_n = S_n^2$ avec $\theta = \sigma^2$, etc.

CI par rééchantillonnage (“bootstrap”)

Ce sont des techniques de simulation appliquées en statistique.

L'idée est d'estimer la distribution (inconnue et quelconque) de l'estimateur en **rééchantillonnant** des échantillons de taille n en tirant **avec remplacement** dans l'échantillon de taille n original. Pour chaque échantillon, on calcule l'estimateur.

Bootstrap non-paramétrique de base.

On a un échantillon i.i.d. X_1, \dots, X_n d'une loi de f.r. F et un estimateur $Y = Y_n = g(X_1, \dots, X_n)$ d'une valeur réelle inconnue θ .

Par exemple, $Y_n = \bar{X}_n$ avec $\theta = \mu$, ou $Y_n = S_n^2$ avec $\theta = \sigma^2$, etc.

On peut avoir $\mathbb{E}[Y_n] \neq \theta$.

Mais on suppose que g ne dépend pas de l'**ordonnement** des X_i .

CI par rééchantillonnage (“bootstrap”)

Ce sont des techniques de simulation appliquées en statistique.

L'idée est d'estimer la distribution (inconnue et quelconque) de l'estimateur en **rééchantillonnant** des échantillons de taille n en tirant **avec remplacement** dans l'échantillon de taille n original. Pour chaque échantillon, on calcule l'estimateur.

Bootstrap non-paramétrique de base.

On a un échantillon i.i.d. X_1, \dots, X_n d'une loi de f.r. F et un estimateur $Y = Y_n = g(X_1, \dots, X_n)$ d'une valeur réelle inconnue θ .

Par exemple, $Y_n = \bar{X}_n$ avec $\theta = \mu$, ou $Y_n = S_n^2$ avec $\theta = \sigma^2$, etc.

On peut avoir $\mathbb{E}[Y_n] \neq \theta$.

Mais on suppose que g ne dépend pas de l'**ordonnement** des X_i .

Si on connaissait exactement la loi (fct de répartition) de Y_n , on pourrait calculer un IC exact pour θ , comme suit.

Soit $K_n(z) \equiv K_n(F, z) = \mathbb{P}[Y_n - \theta \leq z]$ pour $z \in \mathbb{R}$.

Soit $K_n(z) \equiv K_n(F, z) = \mathbb{P}[Y_n - \theta \leq z]$ pour $z \in \mathbb{R}$.

Un IC **exact** pour θ , au niveau $1 - \alpha_1 - \alpha_2$:

$$(I_1, I_2) = (Y_n - K_n^{-1}(1 - \alpha_1), Y_n - K_n^{-1}(\alpha_2)),$$

où $K_n^{-1}(q)$ est le q -quantile de $K_n(\cdot)$.

Soit $K_n(z) \equiv K_n(F, z) = \mathbb{P}[Y_n - \theta \leq z]$ pour $z \in \mathbb{R}$.

Un IC **exact** pour θ , au niveau $1 - \alpha_1 - \alpha_2$:

$$(I_1, I_2) = (Y_n - K_n^{-1}(1 - \alpha_1), Y_n - K_n^{-1}(\alpha_2)),$$

où $K_n^{-1}(q)$ est le q -quantile de $K_n(\cdot)$. En effet,

$$\mathbb{P}[I_1 > \theta] = \mathbb{P}[Y_n - \theta > K_n^{-1}(1 - \alpha_1)]$$

Soit $K_n(z) \equiv K_n(F, z) = \mathbb{P}[Y_n - \theta \leq z]$ pour $z \in \mathbb{R}$.

Un IC **exact** pour θ , au niveau $1 - \alpha_1 - \alpha_2$:

$$(I_1, I_2) = (Y_n - K_n^{-1}(1 - \alpha_1), Y_n - K_n^{-1}(\alpha_2)),$$

où $K_n^{-1}(q)$ est le q -quantile de $K_n(\cdot)$. En effet,

$$\mathbb{P}[I_1 > \theta] = \mathbb{P}[Y_n - \theta > K_n^{-1}(1 - \alpha_1)] = 1 - K_n(K_n^{-1}(1 - \alpha_1))$$

Soit $K_n(z) \equiv K_n(F, z) = \mathbb{P}[Y_n - \theta \leq z]$ pour $z \in \mathbb{R}$.

Un IC **exact** pour θ , au niveau $1 - \alpha_1 - \alpha_2$:

$$(I_1, I_2) = (Y_n - K_n^{-1}(1 - \alpha_1), Y_n - K_n^{-1}(\alpha_2)),$$

où $K_n^{-1}(q)$ est le q -quantile de $K_n(\cdot)$. En effet,

$$\mathbb{P}[I_1 > \theta] = \mathbb{P}[Y_n - \theta > K_n^{-1}(1 - \alpha_1)] = 1 - K_n(K_n^{-1}(1 - \alpha_1)) = \alpha_1.$$

Soit $K_n(z) \equiv K_n(F, z) = \mathbb{P}[Y_n - \theta \leq z]$ pour $z \in \mathbb{R}$.

Un IC **exact** pour θ , au niveau $1 - \alpha_1 - \alpha_2$:

$$(I_1, I_2) = (Y_n - K_n^{-1}(1 - \alpha_1), Y_n - K_n^{-1}(\alpha_2)),$$

où $K_n^{-1}(q)$ est le q -quantile de $K_n(\cdot)$. En effet,

$$\mathbb{P}[I_1 > \theta] = \mathbb{P}[Y_n - \theta > K_n^{-1}(1 - \alpha_1)] = 1 - K_n(K_n^{-1}(1 - \alpha_1)) = \alpha_1.$$

De même, $\mathbb{P}[I_2 < \theta] = \alpha_2$.

Soit $K_n(z) \equiv K_n(F, z) = \mathbb{P}[Y_n - \theta \leq z]$ pour $z \in \mathbb{R}$.

Un IC **exact** pour θ , au niveau $1 - \alpha_1 - \alpha_2$:

$$(I_1, I_2) = (Y_n - K_n^{-1}(1 - \alpha_1), Y_n - K_n^{-1}(\alpha_2)),$$

où $K_n^{-1}(q)$ est le q -quantile de $K_n(\cdot)$. En effet,

$$\mathbb{P}[I_1 > \theta] = \mathbb{P}[Y_n - \theta > K_n^{-1}(1 - \alpha_1)] = 1 - K_n(K_n^{-1}(1 - \alpha_1)) = \alpha_1.$$

De même, $\mathbb{P}[I_2 < \theta] = \alpha_2$.

Mais habituellement on ne connaît pas $K_n(\cdot)$.

Soit $K_n(z) \equiv K_n(F, z) = \mathbb{P}[Y_n - \theta \leq z]$ pour $z \in \mathbb{R}$.

Un IC **exact** pour θ , au niveau $1 - \alpha_1 - \alpha_2$:

$$(I_1, I_2) = (Y_n - K_n^{-1}(1 - \alpha_1), Y_n - K_n^{-1}(\alpha_2)),$$

où $K_n^{-1}(q)$ est le q -quantile de $K_n(\cdot)$. En effet,

$$\mathbb{P}[I_1 > \theta] = \mathbb{P}[Y_n - \theta > K_n^{-1}(1 - \alpha_1)] = 1 - K_n(K_n^{-1}(1 - \alpha_1)) = \alpha_1.$$

De même, $\mathbb{P}[I_2 < \theta] = \alpha_2$.

Mais habituellement on ne connaît pas $K_n(\cdot)$.

Idée (force brute): on pourrait répéter l'expérience m fois et obtenir m copies i.i.d. de Y_n pour estimer sa distribution?

Si $\mathbb{E}[Y_n] = \theta$, on peut estimer en même temps θ , et estimer ainsi la distribution de $Y_n - \theta$.

Soit $K_n(z) \equiv K_n(F, z) = \mathbb{P}[Y_n - \theta \leq z]$ pour $z \in \mathbb{R}$.

Un IC **exact** pour θ , au niveau $1 - \alpha_1 - \alpha_2$:

$$(I_1, I_2) = (Y_n - K_n^{-1}(1 - \alpha_1), Y_n - K_n^{-1}(\alpha_2)),$$

où $K_n^{-1}(q)$ est le q -quantile de $K_n(\cdot)$. En effet,

$$\mathbb{P}[I_1 > \theta] = \mathbb{P}[Y_n - \theta > K_n^{-1}(1 - \alpha_1)] = 1 - K_n(K_n^{-1}(1 - \alpha_1)) = \alpha_1.$$

De même, $\mathbb{P}[I_2 < \theta] = \alpha_2$.

Mais habituellement on ne connaît pas $K_n(\cdot)$.

Idée (force brute): on pourrait répéter l'expérience m fois et obtenir m copies i.i.d. de Y_n pour estimer sa distribution?

Si $\mathbb{E}[Y_n] = \theta$, on peut estimer en même temps θ , et estimer ainsi la distribution de $Y_n - \theta$.

Mais cela ferait mn simulations!

Trop coûteux. On va seulement “faire semblant” de répéter l'expérience m fois.

Soit $K_n(z) \equiv K_n(F, z) = \mathbb{P}[Y_n - \theta \leq z]$ pour $z \in \mathbb{R}$.

Un IC **exact** pour θ , au niveau $1 - \alpha_1 - \alpha_2$:

$$(I_1, I_2) = (Y_n - K_n^{-1}(1 - \alpha_1), Y_n - K_n^{-1}(\alpha_2)),$$

où $K_n^{-1}(q)$ est le q -quantile de $K_n(\cdot)$. En effet,

$$\mathbb{P}[I_1 > \theta] = \mathbb{P}[Y_n - \theta > K_n^{-1}(1 - \alpha_1)] = 1 - K_n(K_n^{-1}(1 - \alpha_1)) = \alpha_1.$$

De même, $\mathbb{P}[I_2 < \theta] = \alpha_2$.

Mais habituellement on ne connaît pas $K_n(\cdot)$.

Idée (force brute): on pourrait répéter l'expérience m fois et obtenir m copies i.i.d. de Y_n pour estimer sa distribution?

Si $\mathbb{E}[Y_n] = \theta$, on peut estimer en même temps θ , et estimer ainsi la distribution de $Y_n - \theta$.

Mais cela ferait mn simulations!

Trop coûteux. On va seulement “faire semblant” de répéter l'expérience m fois.

Soient x_1, \dots, x_n les réalisations de X_1, \dots, X_n et $y = g(x_1, \dots, x_n)$.

Bootstrap non-paramétrique de base:

On va plutôt répéter m fois l'expérience en tirant les n valeurs de \hat{F}_n au lieu de F (moins coûteux).

Tirer X_1^*, \dots, X_n^* au hasard avec remplacement de $\{x_1, \dots, x_n\}$ et calculer $Y^* = g(X_1^*, \dots, X_n^*)$.

Bootstrap non-paramétrique de base:

On va plutôt répéter m fois l'expérience en tirant les n valeurs de \hat{F}_n au lieu de F (moins coûteux).

Tirer X_1^*, \dots, X_n^* au hasard **avec remplacement** de $\{x_1, \dots, x_n\}$ et calculer $Y^* = g(X_1^*, \dots, X_n^*)$.

Répéter m fois; on obtient Y_1^*, \dots, Y_m^* , copies i.i.d. de Y^* .

Bootstrap non-paramétrique de base:

On va plutôt répéter m fois l'expérience en tirant les n valeurs de \hat{F}_n au lieu de F (moins coûteux).

Tirer X_1^*, \dots, X_n^* au hasard **avec remplacement** de $\{x_1, \dots, x_n\}$ et calculer $Y^* = g(X_1^*, \dots, X_n^*)$.

Répéter m fois; on obtient Y_1^*, \dots, Y_m^* , copies i.i.d. de Y^* .

Soit $\hat{K}_{n,m}$ la f.r. empirique de $Y_1^* - y, \dots, Y_m^* - y$.

Bootstrap non-paramétrique de base:

On va plutôt répéter m fois l'expérience en tirant les n valeurs de \hat{F}_n au lieu de F (moins coûteux).

Tirer X_1^*, \dots, X_n^* au hasard **avec remplacement** de $\{x_1, \dots, x_n\}$ et calculer $Y^* = g(X_1^*, \dots, X_n^*)$.

Répéter m fois; on obtient Y_1^*, \dots, Y_m^* , copies i.i.d. de Y^* .

Soit $\hat{K}_{n,m}$ la f.r. empirique de $Y_1^* - y, \dots, Y_m^* - y$.

Pour $m \rightarrow \infty$, elle converge vers la f.r. de $Y^* - y$, qui est $K_n(\hat{F}_n, \cdot)$.

Bootstrap non-paramétrique de base:

On va plutôt répéter m fois l'expérience en tirant les n valeurs de \hat{F}_n au lieu de F (moins coûteux).

Tirer X_1^*, \dots, X_n^* au hasard **avec remplacement** de $\{x_1, \dots, x_n\}$ et calculer $Y^* = g(X_1^*, \dots, X_n^*)$.

Répéter m fois; on obtient Y_1^*, \dots, Y_m^* , copies i.i.d. de Y^* .

Soit $\hat{K}_{n,m}$ la f.r. empirique de $Y_1^* - y, \dots, Y_m^* - y$.

Pour $m \rightarrow \infty$, elle converge vers la f.r. de $Y^* - y$, qui est $K_n(\hat{F}_n, \cdot)$.

L'**IC** retourné est:

$$(y - \hat{K}_{n,m}^{-1}(1 - \alpha_1), y - \hat{K}_{n,m}^{-1}(\alpha_2)) = (2y - Y_{(\lceil m(1-\alpha_1) \rceil)}^*, 2y - Y_{(\lceil m\alpha_2 \rceil)}^*).$$

Équivaut à remplacer F par \hat{F}_n puis à approximer $K_n(\hat{F}_n, \cdot)$ par $\hat{K}_{n,m}$.

Deux sources d'erreur.

Bootstrap non-paramétrique de base:

On va plutôt répéter m fois l'expérience en tirant les n valeurs de \hat{F}_n au lieu de F (moins coûteux).

Tirer X_1^*, \dots, X_n^* au hasard avec remplacement de $\{x_1, \dots, x_n\}$ et calculer $Y^* = g(X_1^*, \dots, X_n^*)$.

Répéter m fois; on obtient Y_1^*, \dots, Y_m^* , copies i.i.d. de Y^* .

Soit $\hat{K}_{n,m}$ la f.r. empirique de $Y_1^* - y, \dots, Y_m^* - y$.

Pour $m \rightarrow \infty$, elle converge vers la f.r. de $Y^* - y$, qui est $K_n(\hat{F}_n, \cdot)$.

L'IC retourné est:

$$(y - \hat{K}_{n,m}^{-1}(1 - \alpha_1), y - \hat{K}_{n,m}^{-1}(\alpha_2)) = (2y - Y_{(\lceil m(1-\alpha_1) \rceil)}^*, 2y - Y_{(\lceil m\alpha_2 \rceil)}^*).$$

Équivaut à remplacer F par \hat{F}_n puis à approximer $K_n(\hat{F}_n, \cdot)$ par $\hat{K}_{n,m}$.

Deux sources d'erreur.

Permet d'estimer aussi le biais!

Bootstrap non-paramétrique de base:

On va plutôt répéter m fois l'expérience en tirant les n valeurs de \hat{F}_n au lieu de F (moins coûteux).

Tirer X_1^*, \dots, X_n^* au hasard **avec remplacement** de $\{x_1, \dots, x_n\}$ et calculer $Y^* = g(X_1^*, \dots, X_n^*)$.

Répéter m fois; on obtient Y_1^*, \dots, Y_m^* , copies i.i.d. de Y^* .

Soit $\hat{K}_{n,m}$ la f.r. empirique de $Y_1^* - y, \dots, Y_m^* - y$.

Pour $m \rightarrow \infty$, elle converge vers la f.r. de $Y^* - y$, qui est $K_n(\hat{F}_n, \cdot)$.

L'IC retourné est:

$$(y - \hat{K}_{n,m}^{-1}(1 - \alpha_1), y - \hat{K}_{n,m}^{-1}(\alpha_2)) = (2y - Y_{(\lceil m(1-\alpha_1) \rceil)}^*, 2y - Y_{(\lceil m\alpha_2 \rceil)}^*).$$

Équivaut à remplacer F par \hat{F}_n puis à approximer $K_n(\hat{F}_n, \cdot)$ par $\hat{K}_{n,m}$.

Deux sources d'erreur.

Permet d'estimer aussi le biais!

Il y a des cas où on peut calculer (ou approximer) $K_n^{-1}(\hat{F}_n, \cdot)$ directement sans passer par $\hat{K}_{n,m}^{-1}$. Par exemple, IC pour $\mu = \mathbb{E}[X_i]$.

Pas besoin des tirages bootstrap dans ces cas.

Bootstrap-t Non-Parametrique

Souvent meilleur que le bootstrap de base, mais requiert un estimateur de $\text{Var}[Y_n]$, disons $S^2 = h^2(X_1, \dots, X_n)$.

Soit $J_n(\cdot) = J_n(F, \cdot)$ la f.r. de la statistique **studentisée** $(Y_n - \theta)/S$.

Bootstrap-t Non-Parametrique

Souvent meilleur que le bootstrap de base, mais requiert un estimateur de $\text{Var}[Y_n]$, disons $S^2 = h^2(X_1, \dots, X_n)$.

Soit $J_n(\cdot) = J_n(F, \cdot)$ la f.r. de la statistique **studentisée** $(Y_n - \theta)/S$.

Un IC exact de niveau $(1 - \alpha_1 - \alpha_2)$:

$$(I_1, I_2) = (Y - J_n^{-1}(1 - \alpha_1)S, Y - J_n^{-1}(\alpha_2)S).$$

Bootstrap-t Non-Parametrique

Souvent meilleur que le bootstrap de base, mais requiert un estimateur de $\text{Var}[Y_n]$, disons $S^2 = h^2(X_1, \dots, X_n)$.

Soit $J_n(\cdot) = J_n(F, \cdot)$ la f.r. de la statistique **studentisée** $(Y_n - \theta)/S$.

Un IC exact de niveau $(1 - \alpha_1 - \alpha_2)$:

$$(I_1, I_2) = (Y - J_n^{-1}(1 - \alpha_1)S, Y - J_n^{-1}(\alpha_2)S).$$

Algorithme du **bootstrap-t non-paramétrique**:

Pour chacune des m répétitions bootstrap:

Bootstrap-t Non-Parametrique

Souvent meilleur que le bootstrap de base, mais requiert un estimateur de $\text{Var}[Y_n]$, disons $S^2 = h^2(X_1, \dots, X_n)$.

Soit $J_n(\cdot) = J_n(F, \cdot)$ la f.r. de la statistique **studentisée** $(Y_n - \theta)/S$.

Un IC exact de niveau $(1 - \alpha_1 - \alpha_2)$:

$$(I_1, I_2) = (Y - J_n^{-1}(1 - \alpha_1)S, Y - J_n^{-1}(\alpha_2)S).$$

Algorithme du **bootstrap-t non-paramétrique**:

Pour chacune des m répétitions bootstrap:

Générer n observations X_1^*, \dots, X_n^* comme avant;

Bootstrap-t Non-Parametrique

Souvent meilleur que le bootstrap de base, mais requiert un estimateur de $\text{Var}[Y_n]$, disons $S^2 = h^2(X_1, \dots, X_n)$.

Soit $J_n(\cdot) = J_n(F, \cdot)$ la f.r. de la statistique **studentisée** $(Y_n - \theta)/S$.

Un IC exact de niveau $(1 - \alpha_1 - \alpha_2)$:

$$(I_1, I_2) = (Y - J_n^{-1}(1 - \alpha_1)S, Y - J_n^{-1}(\alpha_2)S).$$

Algorithme du **bootstrap-t non-paramétrique**:

Pour chacune des m répétitions bootstrap:

Générer n observations X_1^*, \dots, X_n^* comme avant;

Calculer $Y^* = g(X_1^*, \dots, X_n^*)$, $S^* = h(X_1^*, \dots, X_n^*)$, et $Z^* = (Y^* - y)/S^*$.

Bootstrap-t Non-Parametrique

Souvent meilleur que le bootstrap de base, mais requiert un estimateur de $\text{Var}[Y_n]$, disons $S^2 = h^2(X_1, \dots, X_n)$.

Soit $J_n(\cdot) = J_n(F, \cdot)$ la f.r. de la statistique **studentisée** $(Y_n - \theta)/S$.

Un IC exact de niveau $(1 - \alpha_1 - \alpha_2)$:

$$(I_1, I_2) = (Y - J_n^{-1}(1 - \alpha_1)S, Y - J_n^{-1}(\alpha_2)S).$$

Algorithme du **bootstrap-t non-paramétrique**:

Pour chacune des m répétitions bootstrap:

Générer n observations X_1^*, \dots, X_n^* comme avant;

Calculer $Y^* = g(X_1^*, \dots, X_n^*)$, $S^* = h(X_1^*, \dots, X_n^*)$, et $Z^* = (Y^* - y)/S^*$.

Soient Z_1^*, \dots, Z_m^* les m copies i.i.d. de Z^* et $\hat{J}_{n,m}$ leur f.r. empirique.

Bootstrap-t Non-Parametrique

Souvent meilleur que le bootstrap de base, mais requiert un estimateur de $\text{Var}[Y_n]$, disons $S^2 = h^2(X_1, \dots, X_n)$.

Soit $J_n(\cdot) = J_n(F, \cdot)$ la f.r. de la statistique **studentisée** $(Y_n - \theta)/S$.

Un IC exact de niveau $(1 - \alpha_1 - \alpha_2)$:

$$(I_1, I_2) = (Y - J_n^{-1}(1 - \alpha_1)S, Y - J_n^{-1}(\alpha_2)S).$$

Algorithme du **bootstrap-t non-paramétrique**:

Pour chacune des m répétitions bootstrap:

Générer n observations X_1^*, \dots, X_n^* comme avant;

Calculer $Y^* = g(X_1^*, \dots, X_n^*)$, $S^* = h(X_1^*, \dots, X_n^*)$, et $Z^* = (Y^* - y)/S^*$.

Soient Z_1^*, \dots, Z_m^* les m copies i.i.d. de Z^* et $\hat{J}_{n,m}$ leur f.r. empirique.

Pour calculer l'IC, on remplace $J_n(\cdot)$ par $\hat{J}_{n,m}(\cdot)$:

$$\begin{aligned} (I_1, I_2) &= (y - \hat{J}_{n,m}^{-1}(1 - \alpha_1)S, y - \hat{J}_{n,m}^{-1}(\alpha_2)S) \\ &= (y - Z_{(\lceil m(1-\alpha_1) \rceil)}^* S, y - Z_{(\lceil m\alpha_2 \rceil)}^* S). \end{aligned}$$

Théorie

Supposons que $m = \infty$, $\theta = g(\mathbb{E}[X_i])$ et $Y_n = g(\bar{X}_n)$ où g est une fonction lisse, et $\text{Var}[Y_n] \rightarrow \sigma^2 = h^2(\mathbb{E}[X_i])/n$ quand $n \rightarrow \infty$.

Théorie

Supposons que $m = \infty$, $\theta = g(\mathbb{E}[X_i])$ et $Y_n = g(\bar{X}_n)$ où g est une fonction lisse, et $\text{Var}[Y_n] \rightarrow \sigma^2 = h^2(\mathbb{E}[X_i])/n$ quand $n \rightarrow \infty$.

Sous ces hypothèses et des conditions de régularité sur F , pour les deux types de bootstrap que nous avons vus, Hall (1988) a montré que l'erreur de couverture est

$$n^{-\gamma} p(z_{1-\alpha}) \phi(z_{1-\alpha}) + O(n^{-\gamma-1/2}),$$

où p est un polynôme qui ne dépend que de la sorte d'IC, γ une constante, ϕ la densité normale standard, $z_{1-\alpha}$ son $(1 - \alpha)$ -quantile, $\gamma = 1/2$ pour un IC unilatéral avec le bootstrap de base, $\gamma = 1$ pour les trois autres cas.

Théorie

Supposons que $m = \infty$, $\theta = g(\mathbb{E}[X_i])$ et $Y_n = g(\bar{X}_n)$ où g est une fonction lisse, et $\text{Var}[Y_n] \rightarrow \sigma^2 = h^2(\mathbb{E}[X_i])/n$ quand $n \rightarrow \infty$.

Sous ces hypothèses et des conditions de régularité sur F , pour les deux types de bootstrap que nous avons vus, Hall (1988) a montré que l'erreur de couverture est

$$n^{-\gamma} p(z_{1-\alpha}) \phi(z_{1-\alpha}) + O(n^{-\gamma-1/2}),$$

où p est un polynôme qui ne dépend que de la sorte d'IC, γ une constante, ϕ la densité normale standard, $z_{1-\alpha}$ son $(1 - \alpha)$ -quantile, $\gamma = 1/2$ pour un IC unilatéral avec le bootstrap de base, $\gamma = 1$ pour les trois autres cas.

Empiriquement, le bootstrap- t performe souvent le mieux.

Théorie

Supposons que $m = \infty$, $\theta = g(\mathbb{E}[X_i])$ et $Y_n = g(\bar{X}_n)$ où g est une fonction lisse, et $\text{Var}[Y_n] \rightarrow \sigma^2 = h^2(\mathbb{E}[X_i])/n$ quand $n \rightarrow \infty$.

Sous ces hypothèses et des conditions de régularité sur F , pour les deux types de bootstrap que nous avons vus, Hall (1988) a montré que l'erreur de couverture est

$$n^{-\gamma} p(z_{1-\alpha}) \phi(z_{1-\alpha}) + O(n^{-\gamma-1/2}),$$

où p est un polynôme qui ne dépend que de la sorte d'IC, γ une constante, ϕ la densité normale standard, $z_{1-\alpha}$ son $(1 - \alpha)$ -quantile, $\gamma = 1/2$ pour un IC unilatéral avec le bootstrap de base, $\gamma = 1$ pour les trois autres cas.

Empiriquement, le bootstrap- t performe souvent le mieux.

Le choix de m influence peu l'erreur de couverture, mais un trop petit m donne des IC dont la largeur varie beaucoup.

Choix populaire: $m = 1000$.

Variantes et améliorations.

Il y en a de toutes sortes.

Par exemple, on peut remplacer \hat{F}_n par une meilleure approximation de F , comme une loi quasi-empirique, une fonction obtenue par un estimateur de densité, etc.

Variantes et améliorations.

Il y en a de toutes sortes.

Par exemple, on peut remplacer \hat{F}_n par une meilleure approximation de F , comme une loi quasi-empirique, une fonction obtenue par un estimateur de densité, etc.

Bootstrap paramétrique.

On sait que $F \equiv F_\theta$ et on veut estimer θ .

Variantes et améliorations.

Il y en a de toutes sortes.

Par exemple, on peut remplacer \hat{F}_n par une meilleure approximation de F , comme une loi quasi-empirique, une fonction obtenue par un estimateur de densité, etc.

Bootstrap paramétrique.

On sait que $F \equiv F_\theta$ et on veut estimer θ .

Soit $\hat{\theta}_n = Y = g(X_1, \dots, X_n)$ estimateur de θ .

Variantes et améliorations.

Il y en a de toutes sortes.

Par exemple, on peut remplacer \hat{F}_n par une meilleure approximation de F , comme une loi quasi-empirique, une fonction obtenue par un estimateur de densité, etc.

Bootstrap paramétrique.

On sait que $F \equiv F_\theta$ et on veut estimer θ .

Soit $\hat{\theta}_n = Y = g(X_1, \dots, X_n)$ estimateur de θ .

On génère chaque échantillon bootstrap de $F_{\hat{\theta}_n}$ au lieu de \hat{F}_n , et on l'utilise pour calculer $\theta^* = Y^* = g(X_1^*, \dots, X_n^*)$.

Variantes et améliorations.

Il y en a de toutes sortes.

Par exemple, on peut remplacer \hat{F}_n par une meilleure approximation de F , comme une loi quasi-empirique, une fonction obtenue par un estimateur de densité, etc.

Bootstrap paramétrique.

On sait que $F \equiv F_\theta$ et on veut estimer θ .

Soit $\hat{\theta}_n = Y = g(X_1, \dots, X_n)$ estimateur de θ .

On génère chaque échantillon bootstrap de $F_{\hat{\theta}_n}$ au lieu de \hat{F}_n , et on l'utilise pour calculer $\theta^* = Y^* = g(X_1^*, \dots, X_n^*)$.

On utilise ensuite la f.r. empirique des m copies Y_1^*, \dots, Y_m^* pour approximer la f.r. de $\hat{\theta}_n$ dans le calcul d'un IC.