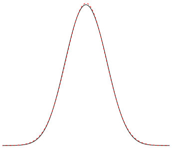
Kubikkrotasymptotikk

Jonas Moss

December 15, 2015



Presentasjon

- Hva er kubikkrotasympotikk?
  - $n^{\frac{1}{3}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$.
  - Vanlig rate: $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$. Kubikkrotasymptotikk er ineffisient. (Variansen er $\frac{\sigma^2}{n^{\frac{2}{3}}}$ istedenfor $\frac{\sigma^2}{n}$.)
- Outline av presentasjonen:
  - M-estimering
  - Eksempler og egenskaper ved kubikkrotasymptotikk
  - Irregulære histogrammer

Presentasjon

- Hva er kubikkrotasymptotikk?
  - $n^{\frac{1}{3}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$.
  - Vanlig rate: $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$. Kubikkrotasymptotikk er ineffisient. (Variansen er $\frac{\sigma^2}{n^{\frac{2}{3}}}$ istedenfor $\frac{\sigma^2}{n}$.)
- Outline av presentasjonen:
  - M-estimering
  - Eksempler og egenskaper ved kubikkrotasymptotikk
  - Irregulære histogrammer

Presentasjon

- Hva er kubikkrotasympotikk?
  - $n^{\frac{1}{3}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$.
  - Vanlig rate: $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$. Kubikkrotasymptotikk er ineffisient. (Variansen er $\frac{\sigma^2}{n^{\frac{2}{3}}}$ istedenfor $\frac{\sigma^2}{n}$.)
- Outline av presentasjonen:
  - M-estimering
  - Eksempler og egenskaper ved kubikkrotasymptotikk
  - Irregulære histogrammer

Presentasjon

- Hva er kubikkrotasympotikk?
    - $n^{\frac{1}{3}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$.
    - Vanlig rate: $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$. Kubikkrotasymptotikk er ineffisient. (Variansen er $\frac{\sigma^2}{n^{\frac{2}{3}}}$ istedenfor $\frac{\sigma^2}{n}$.)
- Outline av presentasjonen:
    - M-estimering
    - Eksempler og egenskaper ved kubikkrotasymptotikk
    - Irregulære histogrammer

Presentasjon

- Hva er kubikkrotasympotikk?
  - $n^{\frac{1}{3}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$.
  - Vanlig rate: $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$. Kubikkrotasymptotikk er ineffisient. (Variansen er $\frac{\sigma^2}{n^{\frac{2}{3}}}$ istedenfor $\frac{\sigma^2}{n}$.)
- Outline av presentasjonen:
  - M-estimering
  - Eksempler og egenskaper ved kubikkrotasymptotikk
  - Irregulære histogrammer

Presentasjon

- Hva er kubikkrotasympotikk?
  - $n^{\frac{1}{3}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$.
  - Vanlig rate: $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$. Kubikkrotasymptotikk er ineffisient. (Variansen er $\frac{\sigma^2}{n^{\frac{2}{3}}}$ istedenfor $\frac{\sigma^2}{n}$.)
- Outline av presentasjonen:
  - M-estimering
  - Eksempler og egenskaper ved kubikkrotasymptotikk
  - Irregulære histogrammer

Presentasjon

- Hva er kubikkrotasympotikk?
  - $n^{\frac{1}{3}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$.
  - Vanlig rate: $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) \xrightarrow{d} Z$. Kubikkrotasymptotikk er ineffisient. (Variansen er $\frac{\sigma^2}{n^{\frac{2}{3}}}$ istedenfor $\frac{\sigma^2}{n}$.)
- Outline av presentasjonen:
  - M-estimering
  - Eksempler og egenskaper ved kubikkrotasymptotikk
  - Irregulære histogrammer

Notasjon

La $m : \Omega \times \Theta \to \mathbb{R}$.

- Kalles m_θ, "objektfunksjon"
- Eksempel: $m_\theta(X_i) = \log f_\theta(X_i)$

Funksjonalnotasjon.

- $Pm_\theta = \int m_\theta(x) dP(x)$
- Når $X_i \sim P$ er iid stokastiske variabler, $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$
- Følgelig: $P_n m_\theta = \frac{1}{n} \sum_{i=1}^n m_\theta(X_i)$

Notasjon

La $m : \Omega \times \Theta \to \mathbb{R}$.

- Kalles m_θ, "objektfunksjon"
- Eksempel: $m_\theta(X_i) = \log f_\theta(X_i)$

Funksjonalnotasjon.

- $Pm_\theta = \int m_\theta(x) dP(x)$
- Når $X_i \sim P$ er iid stokastiske variabler, $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$
- Følgelig: $P_n m_\theta = \frac{1}{n}\sum_{i=1}^n m_\theta(X_i)$

Notasjon

La $m : \Omega \times \Theta \to \mathbb{R}$.

- Kalles m_θ, "objektfunksjon"
- Eksempel: $m_\theta(X_i) = \log f_\theta(X_i)$

Funksjonalnotasjon.

- $Pm_\theta = \int m_\theta(x) dP(x)$
- Når $X_i \sim P$ er iid stokastiske variabler, $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$
- Følgelig: $P_n m_\theta = \frac{1}{n}\sum_{i=1}^n m_\theta(X_i)$

Notasjon

La $m : \Omega \times \Theta \to \mathbb{R}$.

- Kalles m_θ, "objektfunksjon"
- Eksempel: $m_\theta(X_i) = \log f_\theta(X_i)$

Funksjonalnotasjon.

- $Pm_\theta = \int m_\theta(x)dP(x)$
- Når $X_i \sim P$ er iid stokastiske variabler, $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$
- Følgelig: $P_n m_\theta = \frac{1}{n}\sum_{i=1}^n m_\theta(X_i)$

Notasjon

La $m : \Omega \times \Theta \to \mathbb{R}$.

- Kalles m_θ, "objektfunksjon"
- Eksempel: $m_\theta(X_i) = \log f_\theta(X_i)$

Funksjonalnotasjon.

- $Pm_\theta = \int m_\theta(x) dP(x)$
- Når $X_i \sim P$ er iid stokastiske variabler, $P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$
- Følgelig: $P_n m_\theta = \frac{1}{n}\sum_{i=1}^{n} m_\theta(X_i)$

Definisjon av M-estimatorer

M-estimatorer har formen $\widehat{\theta} = \text{argmax}_\theta P_n m_\theta$.

- $m_\theta(x) = \log f_\theta(x)$ gir maximum likelihood
- Kjent fra robust statistikk. (Huber, 1964). Veldig stor klasse av estimatorer.
- $m_\theta(x) = -|x-\theta|$ (median), $m_\theta(x) = -(x-\theta)^2$ (snitt)

Vil $\widehat{\theta} \xrightarrow{p} \text{argmax}_\theta Pm_\theta = \theta_0$? Trenger kontinuitet av argmaxfunksjonalen.

- $||P_n m_\theta - Pm_\theta||_K \xrightarrow{p} 0$ for alle kompakte $K \subseteq \Theta$. En *Glivenko-Cantelli* betingelse. (STL: $P_n m_\theta \xrightarrow{p} Pm_\theta$ punktvis i θ.)

Definisjon av M-estimatorer

M-estimatorer har formen $\widehat{\theta} = \mathrm{argmax}_{\theta} P_n m_{\theta}$.

- $m_{\theta}(x) = \log f_{\theta}(x)$ gir maximum likelihood
- Kjent fra robust statistikk. (Huber, 1964). Veldig stor klasse av estimatorer.
- $m_{\theta}(x) = -|x - \theta|$ (median), $m_{\theta}(x) = -(x - \theta)^2$ (snitt)

Vil $\widehat{\theta} \xrightarrow{p} \mathrm{argmax}_{\theta} P m_{\theta} = \theta_0$? Trenger kontinuitet av argmaxfunksjonalen.

- $||P_n m_{\theta} - P m_{\theta}||_K \xrightarrow{p} 0$ for alle kompakte $K \subseteq \Theta$. En *Glivenko-Cantelli* betingelse. (STL: $P_n m_{\theta} \xrightarrow{p} P m_{\theta}$ punktvis i θ.)

Definisjon av M-estimatorer

M-estimatorer har formen $\widehat{\theta} = \operatorname{argmax}_\theta P_n m_\theta$.

- $m_\theta(x) = \log f_\theta(x)$ gir maximum likelihood
- Kjent fra robust statistikk. (Huber, 1964). Veldig stor klasse av estimatorer.
- $m_\theta(x) = -|x-\theta|$ (median), $m_\theta(x) = -(x-\theta)^2$ (snitt)

Vil $\widehat{\theta} \xrightarrow{p} \operatorname{argmax}_\theta P m_\theta = \theta_0$? Trenger kontinuitet av argmaxfunksjonalen.

- $||P_n m_\theta - P m_\theta||_K \xrightarrow{p} 0$ for alle kompakte $K \subseteq \Theta$. En *Glivenko-Cantelli* betingelse. (STL: $P_n m_\theta \xrightarrow{p} P m_\theta$ punktvis i θ.)

Definisjon av M-estimatorer

M-estimatorer har formen $\widehat{\theta} = \mathrm{argmax}_\theta P_n m_\theta$.

- $m_\theta(x) = \log f_\theta(x)$ gir maximum likelihood
- Kjent fra robust statistikk. (Huber, 1964). Veldig stor klasse av estimatorer.
- $m_\theta(x) = -|x - \theta|$ (median), $m_\theta(x) = -(x-\theta)^2$ (snitt)

Vil $\widehat{\theta} \xrightarrow{p} \mathrm{argmax}_\theta P m_\theta = \theta_0$? Trenger kontinuitet av argmaxfunksjonalen.

- $||P_n m_\theta - P m_\theta||_K \xrightarrow{p} 0$ for alle kompakte $K \subseteq \Theta$. En *Glivenko-Cantelli* betingelse. (STL: $P_n m_\theta \xrightarrow{p} P m_\theta$ punktvis i θ.)

Definisjon av M-estimatorer

M-estimatorer har formen $\widehat{\theta} = \mathrm{argmax}_\theta P_n m_\theta$.

- $m_\theta(x) = \log f_\theta(x)$ gir maximum likelihood
- Kjent fra robust statistikk. (Huber, 1964). Veldig stor klasse av estimatorer.
- $m_\theta(x) = -|x - \theta|$ (median), $m_\theta(x) = -(x - \theta)^2$ (snitt)

Vil $\widehat{\theta} \xrightarrow{p} \mathrm{argmax}_\theta P m_\theta = \theta_0$? Trenger kontinuitet av argmaxfunksjonalen.

- $||P_n m_\theta - P m_\theta||_K \xrightarrow{p} 0$ for alle kompakte $K \subseteq \Theta$. En *Glivenko-Cantelli* betingelse. (STL: $P_n m_\theta \xrightarrow{p} P m_\theta$ punktvis i θ.)

Definisjon av M-estimatorer

M-estimatorer har formen $\widehat{\theta} = \mathrm{argmax}_\theta P_n m_\theta$.

- $m_\theta(x) = \log f_\theta(x)$ gir maximum likelihood
- Kjent fra robust statistikk. (Huber, 1964). Veldig stor klasse av estimatorer.
- $m_\theta(x) = -|x - \theta|$ (median), $m_\theta(x) = -(x - \theta)^2$ (snitt)

Vil $\widehat{\theta} \xrightarrow{p} \mathrm{argmax}_\theta P m_\theta = \theta_0$? Trenger kontinuitet av argmaxfunksjonalen.

- $||P_n m_\theta - P m_\theta||_K \xrightarrow{p} 0$ for alle kompakte $K \subseteq \Theta$. En *Glivenko-Cantelli* betingelse. (STL: $P_n m_\theta \xrightarrow{p} P m_\theta$ punktvis i θ.)

Konvergensrate

Rateteoremet er et generelt verktøy for å finne grensefordelinger, fra "Weak convergence" van der Vaart and Wellner (1996, teorem, 3.2.10).

Theorem

Anta

$$\lim_{\delta \searrow 0} \frac{P(m_{\theta_0+\delta g} m_{\theta_0+\delta h})}{\delta} = K(g, h), \tag{1}$$

for en Gaussisk prosess G med kovarianskjerne K. Sett $V = \frac{d^2}{d\theta^2} Pm_\theta \mid_{\theta=\theta_0}$. Da vil

$$n^{\frac{1}{3}}(\widehat{\theta} - \theta_0) \xrightarrow{d} argmax_h \left[G(h) + \frac{1}{2} h^T V h \right].$$

Om δ^2 istedenfor δ i nevneren i (1) får vi $\sqrt{n}$-konvergens.

Konvergensrate

Rateteoremet er et generelt verktøy for å finne grensefordelinger, fra "Weak convergence" van der Vaart and Wellner (1996, teorem, 3.2.10).

Theorem

Anta

$$\lim_{\delta \searrow 0} \frac{P(m_{\theta_0+\delta g} m_{\theta_0+\delta h})}{\delta} = K(g,h), \tag{1}$$

for en Gaussisk prosess G med kovarianskjerne K. Sett $V = \frac{d^2}{d\theta^2} Pm_\theta \mid_{\theta=\theta_0}$. Da vil

$$n^{\frac{1}{3}}(\widehat{\theta} - \theta_0) \xrightarrow{d} argmax_h \left[G(h) + \frac{1}{2} h^T V h \right].$$

Om δ^2 istedenfor δ i nevneren i (1) får vi $\sqrt{n}$-konvergens.

Eksempler

Estimatorer med kubikkrotasymptotikk:

- Chernoffs (1964) estimator for topppunktet : $\text{argmax}_\theta P_n[\theta - 1, \theta + 1]$
- (Rousseeuw, 1984) Lineær regresjon. Minste median av kvadrater: $\text{argmin}_\beta \text{median}(y_i - X_i^T \beta)^2$.
  - Veldig robust: *Breakdown point* på 0.5.
- Histogrammer og histogramregresjon (Banerjee and McKeague, 2007).
- Manskis (1975) maximum score estimator. $X_i \sim f$ og $Y_i = 1_{[X_i^T\beta + \varepsilon_i \geq 0]}$. $\widehat{\theta} = \text{argmax}_\beta \sum_{i=1}^n (Y_i - \frac{1}{2}) 1_{X_i^T\beta \geq 0}$ (Grensefordeling: Kim and Pollard (1990))

Eksempler

Estimatorer med kubikkrotasymptotikk:

- Chernoffs (1964) estimator for topppunktet : $\text{argmax}_\theta P_n[\theta - 1, \theta + 1]$
- (Rousseeuw, 1984) Lineær regresjon. Minste median av kvadrater: $\text{argmin}_\beta \text{median}(y_i - X_i^T \beta)^2$.
  - Veldig robust: *Breakdown point* på 0.5.
- Histogrammer og histogramregresjon (Banerjee and McKeague, 2007).
- Manskis (1975) maximum score estimator. $X_i \sim f$ og $Y_i = 1_{[X_i^T \beta + \varepsilon_i \geq 0]}$. $\widehat{\theta} = \text{argmax}_\beta \sum_{i=1}^n (Y_i - \frac{1}{2}) 1_{X_i^T \beta \geq 0}$ (Grensefordeling: Kim and Pollard (1990))

Eksempler

Estimatorer med kubikkrotasymptotikk:

- Chernoffs (1964) estimator for topppunktet : $\text{argmax}_\theta P_n[\theta - 1, \theta + 1]$
- (Rousseeuw, 1984) Lineær regresjon. Minste median av kvadrater: $\text{argmin}_\beta \text{median}(y_i - X_i^T \beta)^2$.
  - Veldig robust: *Breakdown point* på 0.5.
- Histogrammer og histogramregresjon (Banerjee and McKeague, 2007).
- Manskis (1975) maximum score estimator. $X_i \sim f$ og $Y_i = 1_{[X_i^T \beta + \varepsilon_i \geq 0]}$. $\widehat{\theta} = \text{argmax}_\beta \sum_{i=1}^n (Y_i - \frac{1}{2}) 1_{X_i^T \beta \geq 0}$ (Grensefordeling: Kim and Pollard (1990))

Eksempler

Estimatorer med kubikkrotasymptotikk:

- Chernoffs (1964) estimator for topppunktet : $\text{argmax}_\theta P_n[\theta - 1, \theta + 1]$
- (Rousseeuw, 1984) Lineær regresjon. Minste median av kvadrater: $\text{argmin}_\beta \text{median}(y_i - X_i^T \beta)^2$.
    - Veldig robust: *Breakdown point* på 0.5.
- Histogrammer og histogramregresjon (Banerjee and McKeague, 2007).
- Manskis (1975) maximum score estimator. $X_i \sim f$ og $Y_i = 1_{[X_i^T \beta + \varepsilon_i \geq 0]}$. $\widehat{\theta} = \text{argmax}_\beta \sum_{i=1}^n (Y_i - \frac{1}{2}) 1_{X_i^T \beta \geq 0}$ (Grensefordeling: Kim and Pollard (1990))

Eksempler

Estimatorer med kubikkrotasymptotikk:

- Chernoffs (1964) estimator for topppunktet : $\text{argmax}_\theta P_n[\theta - 1, \theta + 1]$
- (Rousseeuw, 1984) Lineær regresjon. Minste median av kvadrater: $\text{argmin}_\beta \text{median}(y_i - X_i^T \beta)^2$.
    - Veldig robust: *Breakdown point* på 0.5.
- Histogrammer og histogramregresjon (Banerjee and McKeague, 2007).
- Manskis (1975) maximum score estimator. $X_i \sim f$ og $Y_i = 1_{[X_i^T \beta + \varepsilon_i \geq 0]}$. $\widehat{\theta} = \text{argmax}_\beta \sum_{i=1}^n (Y_i - \frac{1}{2}) 1_{X_i^T \beta \geq 0}$ (Grensefordeling: Kim and Pollard (1990))

Kvadratrøtter

Typiske egenskaper ved $\sqrt{n}$-konsistente M-estimatorer ($\sqrt{n}(\widehat{\theta} - \theta) \to Z$).

- Både $P_n m_\theta$ og $P m_\theta$ er to ganger kontinuerlig deriverbar i θ. "Lokalt kvadratisk".
- Grensefordelingen er normal, $\sqrt{n}(\widehat{\theta} - \theta) \to N(0, J^{-1}KJ^{-1})$.
- Kan beregnes via Newton-Raphson etc. Noen ganger analytisk.
- Bootstrap funger fint, grensefordelinga kan enkelt beregnes og konvergensen er rask.
- Enkelt å identifisere grensefordelinga via "klassisk" teori.

Kvadratrøtter

Typiske egenskaper ved $\sqrt{n}$-konsistente M-estimatorer ($\sqrt{n}(\widehat{\theta} - \theta) \to Z$).

- Både $P_n m_\theta$ og $P m_\theta$ er to ganger kontinuerlig deriverbar i θ. "Lokalt kvadratisk".
- Grensefordelingen er normal, $\sqrt{n}(\widehat{\theta} - \theta) \to N(0, J^{-1}KJ^{-1})$.
- Kan beregnes via Newton-Raphson etc. Noen ganger analytisk.
- Bootstrap funger fint, grensefordelinga kan enkelt beregnes og konvergensen er rask.
- Enkelt å identifisere grensefordelinga via "klassisk" teori.

Kvadratrøtter

Typiske egenskaper ved $\sqrt{n}$-konsistente M-estimatorer ($\sqrt{n}(\widehat{\theta} - \theta) \to Z$).

- Både $P_n m_\theta$ og $P m_\theta$ er to ganger kontinuerlig deriverbar i θ. "Lokalt kvadratisk".
- Grensefordelingen er normal, $\sqrt{n}(\widehat{\theta} - \theta) \to N(0, J^{-1} K J^{-1})$.
- Kan beregnes via Newton-Raphson etc. Noen ganger analytisk.
- Bootstrap funger fint, grensefordelinga kan enkelt beregnes og konvergensen er rask.
- Enkelt å identifisere grensefordelinga via "klassisk" teori.

Kvadratrøtter

Typiske egenskaper ved $\sqrt{n}$-konsistente M-estimatorer ($\sqrt{n}(\widehat{\theta}-\theta) \to Z$).

- Både $P_n m_\theta$ og Pm_θ er to ganger kontinuerlig deriverbar i θ. "Lokalt kvadratisk".
- Grensefordelingen er normal, $\sqrt{n}(\widehat{\theta}-\theta) \to N(0, J^{-1}KJ^{-1})$.
- Kan beregnes via Newton-Raphson etc. Noen ganger analytisk.
- Bootstrap funger fint, grensefordelinga kan enkelt beregnes og konvergensen er rask.
- Enkelt å identifisere grensefordelinga via "klassisk" teori.

Kvadratrøtter

Typiske egenskaper ved $\sqrt{n}$-konsistente M-estimatorer ($\sqrt{n}(\widehat{\theta}-\theta) \to Z$).

- Både $P_n m_\theta$ og $P m_\theta$ er to ganger kontinuerlig deriverbar i θ. "Lokalt kvadratisk".
- Grensefordelingen er normal, $\sqrt{n}(\widehat{\theta}-\theta) \to N(0, J^{-1}KJ^{-1})$.
- Kan beregnes via Newton-Raphson etc. Noen ganger analytisk.
- Bootstrap funger fint, grensefordelinga kan enkelt beregnes og konvergensen er rask.
- Enkelt å identifisere grensefordelinga via "klassisk" teori.

Kvadratrøtter

Typiske egenskaper ved $\sqrt{n}$-konsistente M-estimatorer ($\sqrt{n}(\widehat{\theta} - \theta) \to Z$).

- Både $P_n m_\theta$ og $P m_\theta$ er to ganger kontinuerlig deriverbar i θ. "Lokalt kvadratisk".
- Grensefordelingen er normal, $\sqrt{n}(\widehat{\theta} - \theta) \to N(0, J^{-1} K J^{-1})$.
- Kan beregnes via Newton-Raphson etc. Noen ganger analytisk.
- Bootstrap funger fint, grensefordelinga kan enkelt beregnes og konvergensen er rask.
- Enkelt å identifisere grensefordelinga via "klassisk" teori.

Kubikkrøtter

Egenskaper ved kubikkrotasymptikk.

- $P_n m_\theta$ er ikke deriverbar, men $P m_\theta$ er to ganger kontinuerlig deriverbar.
  - m_θ er diskontinuerlig i θ, mens $P m_\theta$ ikke er det. "Integration smoothens business."
- Grenseprosessen er *ikke* normal. $\operatorname{argmax}_s \left[\frac{1}{2} s^T V s + \sum_{i=1}^{k} \alpha_i W_i(s)\right]$ istedenfor.
- Må bruke kombinatorisk optimering. (NP-hard).
- Inkonsistent bootstrap, problematisk grensefordeling med treg konvergens.
- En jobb å finne grensefordelinga. Klassisk teori fungerer ikke.

Kubikkrøtter

Egenskaper ved kubikkrotasymptikk.

- $P_n m_\theta$ er ikke deriverbar, men Pm_θ er to ganger kontinuerlig deriverbar.
    - m_θ er diskontinuerlig i θ, mens Pm_θ ikke er det. "Integration smoothens business."
- Grenseprosessen er *ikke* normal. $\operatorname{argmax}_s \left[\frac{1}{2}s^T Vs + \sum_{i=1}^k \alpha_i W_i(s)\right]$ istedenfor.
- Må bruke kombinatorisk optimering. (NP-hard).
- Inkonsistent bootstrap, problematisk grensefordeling med treg konvergens.
- En jobb å finne grensefordelinga. Klassisk teori fungerer ikke.

Kubikkrøtter

Egenskaper ved kubikkrotasymptikk.

- $P_n m_\theta$ er ikke deriverbar, men Pm_θ er to ganger kontinuerlig deriverbar.
  - m_θ er diskontinuerlig i θ, mens Pm_θ ikke er det. "Integration smoothens business."
- Grenseprosessen er *ikke* normal. $\text{argmax}_s \left[\frac{1}{2} s^T V s + \sum_{i=1}^k \alpha_i W_i(s)\right]$ istedenfor.
- Må bruke kombinatorisk optimering. (NP-hard).
- Inkonsistent bootstrap, problematisk grensefordeling med treg konvergens.
- En jobb å finne grensefordelinga. Klassisk teori fungerer ikke.

Kubikkrøtter

Egenskaper ved kubikkrotasymptikk.

- $P_n m_\theta$ er ikke deriverbar, men $P m_\theta$ er to ganger kontinuerlig deriverbar.
  - m_θ er diskontinuerlig i θ, mens $P m_\theta$ ikke er det. "Integration smoothens business."
- Grenseprosessen er *ikke* normal. $\operatorname{argmax}_s \left[\frac{1}{2} s^T V s + \sum_{i=1}^{k} \alpha_i W_i(s)\right]$ istedenfor.
- Må bruke kombinatorisk optimering. (NP-hard).
- Inkonsistent bootstrap, problematisk grensefordeling med treg konvergens.
- En jobb å finne grensefordelinga. Klassisk teori fungerer ikke.

Kubikkrøtter

Egenskaper ved kubikkrotasymptikk.

- $P_n m_\theta$ er ikke deriverbar, men Pm_θ er to ganger kontinuerlig deriverbar.
  - m_θ er diskontinuerlig i θ, mens Pm_θ ikke er det. "Integration smoothens business."
- Grenseprosessen er *ikke* normal. $\operatorname{argmax}_s \left[\frac{1}{2}s^T V s + \sum_{i=1}^{k} \alpha_i W_i(s)\right]$ istedenfor.
- Må bruke kombinatorisk optimering. (NP-hard).
- Inkonsistent bootstrap, problematisk grensefordeling med treg konvergens.
- En jobb å finne grensefordelinga. Klassisk teori fungerer ikke.

Kubikkrøtter

Egenskaper ved kubikkrotasymptikk.

- $P_n m_\theta$ er ikke deriverbar, men Pm_θ er to ganger kontinuerlig deriverbar.
  - m_θ er diskontinuerlig i θ, mens Pm_θ ikke er det. "Integration smoothens business."
- Grenseprosessen er *ikke* normal. $\text{argmax}_s \left[\frac{1}{2}s^T Vs + \sum_{i=1}^k \alpha_i W_i(s)\right]$ istedenfor.
- Må bruke kombinatorisk optimering. (NP-hard).
- Inkonsistent bootstrap, problematisk grensefordeling med treg konvergens.
- En jobb å finne grensefordelinga. Klassisk teori fungerer ikke.

Histogrammodellen

Definition (Histogrammer)

Histogramtettheter på $[0,1]$ er $h \ll \lambda$ på formen

$$h(x) = \sum_{i=1}^{k} \frac{w_i}{a_i - a_{i-1}} 1_{[a_{i-1}, a_i)}(x),$$

hvor $\sum w_i = 1$, $w_i > 0$, $0 = a_0 < a_1 < \ldots < a_k = 1$. $(a = (a_0, ..., a_k)$, $w = (w_1, ..., w_k))$

- Udefinert i høyre endepunkt.
- Definer $\mathscr{H}_k = \{h \mid h \text{ er et histogram med } k \text{ blokker}\}$
- $\mathscr{H}_k$ er en "diskontinuerlig" *modell* for tettheter på $[0,1]$ (i sup-norm).
- Elementer i $\mathscr{H}_k$ kalles h_k

Histogrammodellen

Definition (Histogrammer)

Histogramtettheter på $[0,1]$ er $h \ll \lambda$ på formen

$$h(x) = \sum_{i=1}^{k} \frac{w_i}{a_i - a_{i-1}} 1_{[a_{i-1}, a_i)}(x),$$

hvor $\sum w_i = 1$, $w_i > 0$, $0 = a_0 < a_1 < \ldots < a_k = 1$. $(a = (a_0, \ldots, a_k)$, $w = (w_1, \ldots, w_k))$

- Udefinert i høyre endepunkt.
- Definer $\mathscr{H}_k = \{h \mid h \text{ er et histogram med } k \text{ blokker}\}$
- $\mathscr{H}_k$ er en "diskontinuerlig" *modell* for tettheter på $[0,1]$ (i sup-norm).
- Elementer i $\mathscr{H}_k$ kalles h_k

Histogrammodellen

Definition (Histogrammer)

Histogramtettheter på $[0,1]$ er $h \ll \lambda$ på formen

$$h(x) = \sum_{i=1}^{k} \frac{w_i}{a_i - a_{i-1}} 1_{[a_{i-1}, a_i)}(x),$$

hvor $\sum w_i = 1$, $w_i > 0$, $0 = a_0 < a_1 < \ldots < a_k = 1$. $(a = (a_0, ..., a_k)$, $w = (w_1, ..., w_k))$

- Udefinert i høyre endepunkt.
- Definer $\mathscr{H}_k = \{h \mid h \text{ er et histogram med } k \text{ blokker}\}$
- $\mathscr{H}_k$ er en "diskontinuerlig" *modell* for tettheter på $[0,1]$ (i sup-norm).
- Elementer i $\mathscr{H}_k$ kalles h_k

Histogrammodellen

Definition (Histogrammer)

Histogramtettheter på $[0,1]$ er $h \ll \lambda$ på formen

$$h(x) = \sum_{i=1}^{k} \frac{w_i}{a_i - a_{i-1}} 1_{[a_{i-1}, a_i)}(x),$$

hvor $\sum w_i = 1$, $w_i > 0$, $0 = a_0 < a_1 < \ldots < a_k = 1$. $(a = (a_0, ..., a_k)$, $w = (w_1, ..., w_k))$

- Udefinert i høyre endepunkt.
- Definer $\mathscr{H}_k = \{h \mid h \text{ er et histogram med } k \text{ blokker}\}$
- $\mathscr{H}_k$ er en "diskontinuerlig" *modell* for tettheter på $[0,1]$ (i sup-norm).
- Elementer i $\mathscr{H}_k$ kalles h_k

Histogrammodellen

Definition (Histogrammer)

Histogramtettheter på $[0,1]$ er $h \ll \lambda$ på formen

$$h(x) = \sum_{i=1}^{k} \frac{w_i}{a_i - a_{i-1}} 1_{[a_{i-1},a_i)}(x),$$

hvor $\sum w_i = 1$, $w_i > 0$, $0 = a_0 < a_1 < \ldots < a_k = 1$. $(a = (a_0, ..., a_k)$, $w = (w_1, ..., w_k))$

- Udefinert i høyre endepunkt.
- Definer $\mathscr{H}_k = \{h \mid h \text{ er et histogram med } k \text{ blokker}\}$
- $\mathscr{H}_k$ er en "diskontinuerlig" *modell* for tettheter på $[0,1]$ (i sup-norm).
- Elementer i $\mathscr{H}_k$ kalles h_k

Statistiske divergenser

- **Mål:** Approksimer en sannsynlighet P med tetthet f på $[0,1]$ med et histogram h_k.
- **Metode:** Velg h_k ved minimering av *statistiske divergenser.*
- **Eksempler:** Kullback-Leibler (d_{KL}), L_p, Kolmogorovavstanden, BHHJ, *etc.*

Identifiserer h_k med (w, a). Da er

$$\operatorname{argmin}_{h_k \in \mathscr{H}_k} d_{KL}(f, h_k) = \operatorname{argmin}_{(a,w)} P m_{(a,w)} = (a^0, w^0),$$

hvor $m_{(a,w)} = \sum_{i=1}^{k} \log \frac{w_i}{a_i - a_{i-1}} 1_{[a_{i-1}, a_i)}$. Lett å vise at $w_i^0 = P[a_{i-1}^0, a_i^0)$.

Statistiske divergenser

- **Mål:** Approksimer en sannsynlighet P med tetthet f på $[0,1]$ med et histogram h_k.
- **Metode:** Velg h_k ved minimering av *statistiske divergenser*.
- **Eksempler:** Kullback-Leibler (d_{KL}), L_p, Kolmogorovavstanden, BHHJ, *etc.*

Identifiserer h_k med (w,a). Da er

$$\text{argmin}_{h_k \in \mathscr{H}_k} d_{KL}(f, h_k) = \text{argmin}_{(a,w)} Pm_{(a,w)} = (a^0, w^0),$$

hvor $m_{(a,w)} = \sum_{i=1}^{k} \log \frac{w_i}{a_i - a_{i-1}} 1_{[a_{i-1}, a_i)}$. Lett å vise at $w_i^0 = P[a_{i-1}^0, a_i^0)$.

Statistiske divergenser

- **Mål:** Approksimer en sannsynlighet P med tetthet f på $[0,1]$ med et histogram h_k.
- **Metode:** Velg h_k ved minimering av *statistiske divergenser*.
- **Eksempler:** Kullback-Leibler (d_{KL}), L_p, Kolmogorovavstanden, BHHJ, *etc.*

Identifiserer h_k med (w,a). Da er

$$\operatorname{argmin}_{h_k \in \mathscr{H}_k} d_{KL}(f, h_k) = \operatorname{argmin}_{(a,w)} P m_{(a,w)} = (a^0, w^0),$$

hvor $m_{(a,w)} = \sum_{i=1}^{k} \log \frac{w_i}{a_i - a_{i-1}} 1_{[a_{i-1}, a_i)}$. Lett å vise at $w_i^0 = P[a_{i-1}^0, a_i^0)$.

Statistiske divergenser

- **Mål:** Approksimer en sannsynlighet P med tetthet f på $[0,1]$ med et histogram h_k.
- **Metode:** Velg h_k ved minimering av *statistiske divergenser*.
- **Eksempler:** Kullback-Leibler (d_{KL}), L_p, Kolmogorovavstanden, BHHJ, *etc.*

Identifiserer h_k med (w,a). Da er

$$\text{argmin}_{h_k \in \mathscr{H}_k} d_{KL}(f, h_k) = \text{argmin}_{(a,w)} Pm_{(a,w)} = (a^0, w^0),$$

hvor $m_{(a,w)} = \sum_{i=1}^{k} \log \frac{w_i}{a_i - a_{i-1}} 1_{[a_{i-1}, a_i)}$. Lett å vise at $w_i^0 = P[a_{i-1}^0, a_i^0)$.

Statistiske divergenser

- Vil approksimere (a^0, w^0) med $\text{argmin}_{(a,w)} P_n m_{(a,w)}$.

Kan vises at

$$\text{argmin}_{(a,w)} P_n m_{(a,w)} \xrightarrow{p} (a^0, w^0),$$

via enten *Vapnik-Chervonenkisteori* eller *bracketing entropy*.
Også mulig å se på delmengder av $\mathscr{H}_k$:

- Om a er konstant: Regulære histogrammer med ($\sqrt{n}$-konvergens!)
- Om w er konstant: Irregulære "kvantilhistogrammer" ($n^{\frac{1}{3}}$-konvergens)

Statistiske divergenser

- Vil approksimere (a^0, w^0) med $\text{argmin}_{(a,w)} P_n m_{(a,w)}$.

Kan vises at

$$\text{argmin}_{(a,w)} P_n m_{(a,w)} \xrightarrow{p} (a^0, w^0),$$

via enten *Vapnik-Chervonenkisteori* eller *bracketing entropy*.

Også mulig å se på delmengder av $\mathscr{H}_k$:

- Om a er konstant: Regulære histogrammer med ($\sqrt{n}$-konvergens!)
- Om w er konstant: Irregulære "kvantilhistogrammer" ($n^{\frac{1}{3}}$-konvergens)

Statistiske divergenser

- Vil approksimere (a^0, w^0) med $\text{argmin}_{(a,w)} P_n m_{(a,w)}$.

Kan vises at

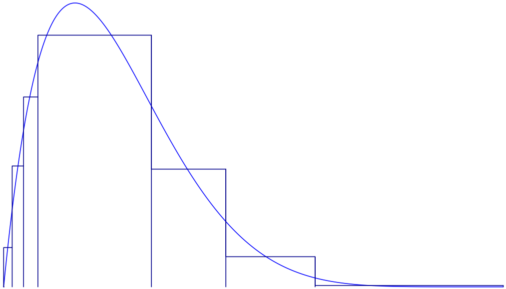
$$\text{argmin}_{(a,w)} P_n m_{(a,w)} \xrightarrow{p} (a^0, w^0),$$

via enten *Vapnik-Chervonenkisteori* eller *bracketing entropy*.
Også mulig å se på delmengder av $\mathscr{H}_k$:

- Om a er konstant: Regulære histogrammer med ($\sqrt{n}$-konvergens!)
- Om w er konstant: Irregulære "kvantilhistogrammer" ($n^{\frac{1}{3}}$-konvergens)

Et histogram

Kullback-Leibler approksimering av Beta$(2,7)$ med $k = 7$ blokker.



Grensefordeling

Antakelser om f er veldig viktige. Vi skiller mellom to tilfeller:

- f er en glatt tetthet: modellen $\mathscr{H}_k$ inneholder ikke f for noen k. Gir kubikkrotasymptotikk.
- f er en stykkvis konstant tetthet: $f \in \mathscr{H}_k$ for en k. Gir n-asymptotikkk på brekkpunktene (for den k-en!)

Theorem

La $k > 1$ være antall blokker, f være glatt, og $V = \frac{d^2}{d(a,w)^2} Pm_{(a,w)} \big|_{(a,w)=(a^0,w^0)}$. La W_i være uavhengige tosidige Wienerprosesser som starter i 0. Da vil $n^{\frac{1}{3}}(\widehat{(a,w)} - (a^0, w^0)) \xrightarrow{d} \text{argmax}_h \left[\frac{1}{2} h^T V h + G(h)\right]$, hvor

$$G(h) = \sum_{i=1}^{k-1} f(a_i^0)^{\frac{1}{2}} \left| \log \frac{w_{i+1}^0}{a_{i+1}^0 - a_i^0} - \log \frac{w_i^0}{a_i^0 - a_{i-1}^0} \right| W_i(h_i).$$

Grensefordeling

Antakelser om f er veldig viktige. Vi skiller mellom to tilfeller:

- f er en glatt tetthet: modellen $\mathscr{H}_k$ inneholder ikke f for noen k. Gir kubikkrotasymptotikk.
- f er en stykkvis konstant tetthet: $f \in \mathscr{H}_k$ for en k. Gir n-asymptotikkk på brekkpunktene (for den k-en!)

Theorem

La $k > 1$ være antall blokker, f være glatt, og $V = \frac{d^2}{d(a,w)^2} Pm_{(a,w)} \big|_{(a,w)=(a^0,w^0)}$. La W_i være uavhengige tosidige Wienerprosesser som starter i 0. Da vil $n^{\frac{1}{3}}(\widehat{(a,w)} - (a^0, w^0)) \xrightarrow{d} \text{argmax}_h \left[\frac{1}{2} h^T V h + G(h)\right]$, hvor

$$G(h) = \sum_{i=1}^{k-1} f(a_i^0)^{\frac{1}{2}} \left| \log \frac{w_{i+1}^0}{a_{i+1}^0 - a_i^0} - \log \frac{w_i^0}{a_i^0 - a_{i-1}^0} \right| W_i(h_i).$$

Grensefordeling

Antakelser om f er veldig viktige. Vi skiller mellom to tilfeller:

- f er en glatt tetthet: modellen $\mathscr{H}_k$ inneholder ikke f for noen k. Gir kubikkrotasymptotikk.
- f er en stykkvis konstant tetthet: $f \in \mathscr{H}_k$ for en k. Gir n-asymptotikkk på brekkpunktene (for den k-en!)

Theorem

La $k > 1$ være antall blokker, f være glatt, og $V = \frac{d^2}{d(a,w)^2} Pm_{(a,w)} \big|_{(a,w)=(a^0,w^0)}$. La W_i være uavhengige tosidige Wienerprosesser som starter i 0. Da vil $n^{\frac{1}{3}}(\widehat{(a,w)} - (a^0,w^0)) \xrightarrow{d} \text{argmax}_h \left[\frac{1}{2} h^T V h + G(h)\right]$, hvor

$$G(h) = \sum_{i=1}^{k-1} f(a_i^0)^{\frac{1}{2}} \left| \log \frac{w_{i+1}^0}{a_{i+1}^0 - a_i^0} - \log \frac{w_i^0}{a_i^0 - a_{i-1}^0} \right| W_i(h_i).$$

Grensefordeling

Antakelser om f er veldig viktige. Vi skiller mellom to tilfeller:

- f er en glatt tetthet: modellen $\mathscr{H}_k$ inneholder ikke f for noen k. Gir kubikkrotasymptotikk.
- f er en stykkvis konstant tetthet: $f \in \mathscr{H}_k$ for en k. Gir n-asymptotikkk på brekkpunktene (for den k-en!)

Theorem

La $k > 1$ være antall blokker, f være glatt, og $V = \frac{d^2}{d(a,w)^2} Pm_{(a,w)} \mid_{(a,w)=(a^0,w^0)}$. La W_i være uavhengige tosidige Wienerprosesser som starter i 0. Da vil $n^{\frac{1}{3}}(\widehat{(a,w)} - (a^0, w^0)) \xrightarrow{d} argmax_h \left[\frac{1}{2} h^T Vh + G(h)\right]$, hvor

$$G(h) = \sum_{i=1}^{k-1} f(a_i^0)^{\frac{1}{2}} \left| \log \frac{w_{i+1}^0}{a_{i+1}^0 - a_i^0} - \log \frac{w_i^0}{a_i^0 - a_{i-1}^0} \right| W_i(h_i).$$

Kommentar om grensefordelinga

- Når $k = 2$ blir $G(h) = cW(h)$ for én Wienerprosess. Grensefordelinga kan beskrives via Chernoffs fordeling, som er godt studert. (Dukker ofte opp i kubikkrotasymptotikk.)
- Fordelinga for $k > 2$ er horribel å jobbe med teoretisk, og vanskelig å simulere fra.
- En kan bruke subsampling for å lage konfidensintervaller osv. Kanskje også "smoothed bootstrap".
- Lignende grensefordeling i "histogramregresjon".

Kommentar om grensefordelinga

- Når $k = 2$ blir $G(h) = cW(h)$ for én Wienerprosess. Grensefordelinga kan beskrives via Chernoffs fordeling, som er godt studert. (Dukker ofte opp i kubikkrotasymptotikk.)
- Fordelinga for $k > 2$ er horribel å jobbe med teoretisk, og vanskelig å simulere fra.
- En kan bruke subsampling for å lage konfidensintervaller osv. Kanskje også "smoothed bootstrap".
- Lignende grensefordeling i "histogramregresjon".

Kommentar om grensefordelinga

- Når $k = 2$ blir $G(h) = cW(h)$ for én Wienerprosess. Grensefordelinga kan beskrives via Chernoffs fordeling, som er godt studert. (Dukker ofte opp i kubikkrotasymptotikk.)
- Fordelinga for $k > 2$ er horribel å jobbe med teoretisk, og vanskelig å simulere fra.
- En kan bruke subsampling for å lage konfidensintervaller osv. Kanskje også “smoothed bootstrap”.
- Lignende grensefordeling i “histogramregresjon”.

Kommentar om grensefordelinga

- Når $k = 2$ blir $G(h) = cW(h)$ for én Wienerprosess. Grensefordelinga kan beskrives via Chernoffs fordeling, som er godt studert. (Dukker ofte opp i kubikkrotasymptotikk.)
- Fordelinga for $k > 2$ er horribel å jobbe med teoretisk, og vanskelig å simulere fra.
- En kan bruke subsampling for å lage konfidensintervaller osv. Kanskje også "smoothed bootstrap".
- Lignende grensefordeling i "histogramregresjon".

Modellvalg (CIC)

En variant av AIC for histogrammer, for å velge antall blokker.

- AIC: $2nP_n \log f_{\widehat{\theta}} - 2p$. Oppstår fordi vi ønsker å minimere $d_{KL}(f_\theta, f)$. Ekvivalent med å minimere $P \log f_{\widehat{\theta}}$, som kan estimeres forventningsrett med $P_n \log f_{\widehat{\theta}} - n^{-1}p$, hvor $n^{-1}p$ er biasen. ($n^{-1}\mathrm{Tr}(J^{-1}K)$ mer generelt.)

- Riktig bias for histogrammer: $n^{-\frac{2}{3}} \sum_{i=1}^{k-1} f(a_i)^{\frac{1}{2}} \left| \log\left(\frac{w_{i+1}}{a_{i+1}-a_i}\right) - \log\left(\frac{w_i}{a_i - a_{i-1}}\right) \right| E(W_i(h_i))$, hvor $\mathbf{h}$ maksimerer $\frac{1}{2}\mathbf{h}^T V \mathbf{h} + G(\mathbf{h})$

- Veldig vanskelig å regne ut. Men *subsampling* med blokkstørrelse $b = 0.5$ funker.

- Tentativt: Bedre ytelse (via Hellinger avtand) enn andre metoder. For beregningsintensiv til å bruke i praksis.

Modellvalg (CIC)

En variant av AIC for histogrammer, for å velge antall blokker.

- AIC: $2nP_n \log f_{\widehat{\theta}} - 2p$. Oppstår fordi vi ønsker å minimere $d_{KL}(f_\theta, f)$. Ekvivalent med å minimere $P \log f_{\widehat{\theta}}$, som kan estimeres forventningsrett med $P_n \log f_{\widehat{\theta}} - n^{-1}p$, hvor $n^{-1}p$ er biasen. ($n^{-1}\mathrm{Tr}(J^{-1}K)$ mer generelt.)
- Riktig bias for histogrammer: $n^{-\frac{2}{3}} \sum_{i=1}^{k-1} f(a_i)^{\frac{1}{2}} \left| \log\left(\frac{w_{i+1}}{a_{i+1}-a_i}\right) - \log\left(\frac{w_i}{a_i - a_{i-1}}\right) \right| E(W_i(h_i))$, hvor $\mathbf{h}$ maksimerer $\frac{1}{2}\mathbf{h}^T V \mathbf{h} + G(\mathbf{h})$
- Veldig vanskelig å regne ut. Men *subsampling* med blokkstørrelse $b = 0.5$ funker.
- Tentativt: Bedre ytelse (via Hellinger avtand) enn andre metoder. For beregningsintensiv til å bruke i praksis.

Modellvalg (CIC)

En variant av AIC for histogrammer, for å velge antall blokker.

- AIC: $2nP_n \log f_{\widehat{\theta}} - 2p$. Oppstår fordi vi ønsker å minimere $d_{KL}(f_\theta, f)$. Ekvivalent med å minimere $P \log f_{\widehat{\theta}}$, som kan estimeres forventningsrett med $P_n \log f_{\widehat{\theta}} - n^{-1}p$, hvor $n^{-1}p$ er biasen. ($n^{-1}\mathrm{Tr}(J^{-1}K)$ mer generelt.)
- Riktig bias for histogrammer: $n^{-\frac{2}{3}} \sum_{i=1}^{k-1} f(a_i)^{\frac{1}{2}} \left| \log\left(\frac{w_{i+1}}{a_{i+1}-a_i}\right) - \log\left(\frac{w_i}{a_i - a_{i-1}}\right) \right| E(W_i(h_i))$, hvor $\mathbf{h}$ maksimerer $\frac{1}{2}\mathbf{h}^T V \mathbf{h} + G(\mathbf{h})$
- Veldig vanskelig å regne ut. Men *subsampling* med blokkstørrelse $b = 0.5$ funker.
- Tentativt: Bedre ytelse (via Hellinger avtand) enn andre metoder. For beregningsintensiv til å bruke i praksis.

Modellvalg (CIC)

En variant av AIC for histogrammer, for å velge antall blokker.

- AIC: $2nP_n \log f_{\widehat{\theta}} - 2p$. Oppstår fordi vi ønsker å minimere $d_{KL}(f_\theta, f)$. Ekvivalent med å minimere $P \log f_{\widehat{\theta}}$, som kan estimeres forventningsrett med $P_n \log f_{\widehat{\theta}} - n^{-1}p$, hvor $n^{-1}p$ er biasen. ($n^{-1}\mathrm{Tr}(J^{-1}K)$ mer generelt.)

- Riktig bias for histogrammer: $n^{-\frac{2}{3}} \sum_{i=1}^{k-1} f(a_i)^{\frac{1}{2}} \left| \log\left(\frac{w_{i+1}}{a_{i+1}-a_i}\right) - \log\left(\frac{w_i}{a_i - a_{i-1}}\right)\right| E(W_i(h_i))$, hvor $\mathbf{h}$ maksimerer $\frac{1}{2}\mathbf{h}^T V \mathbf{h} + G(\mathbf{h})$

- Veldig vanskelig å regne ut. Men *subsampling* med blokkstørrelse $b = 0.5$ funker.

- Tentativt: Bedre ytelse (via Hellinger avtand) enn andre metoder. For beregningsintensiv til å bruke i praksis.

Modellvalg (CIC)

En variant av AIC for histogrammer, for å velge antall blokker.

- AIC: $2nP_n \log f_{\widehat{\theta}} - 2p$. Oppstår fordi vi ønsker å minimere $d_{KL}(f_\theta, f)$. Ekvivalent med å minimere $P \log f_{\widehat{\theta}}$, som kan estimeres forventningsrett med $P_n \log f_{\widehat{\theta}} - n^{-1}p$, hvor $n^{-1}p$ er biasen. ($n^{-1}\mathrm{Tr}(J^{-1}K)$ mer generelt.)

- Riktig bias for histogrammer: $n^{-\frac{2}{3}} \sum_{i=1}^{k-1} f(a_i)^{\frac{1}{2}} \left| \log\left(\frac{w_{i+1}}{a_{i+1}-a_i}\right) - \log\left(\frac{w_i}{a_i - a_{i-1}}\right) \right| E(W_i(h_i))$, hvor $\mathbf{h}$ maksimerer $\frac{1}{2}\mathbf{h}^T V \mathbf{h} + G(\mathbf{h})$

- Veldig vanskelig å regne ut. Men *subsampling* med blokkstørrelse $b = 0.5$ funker.

- Tentativt: Bedre ytelse (via Hellinger avtand) enn andre metoder. For beregningsintensiv til å bruke i praksis.

Konsistens

Er histogrammene konsistente?

- (L_1) Finnes en k_n s.a. $\int |\widehat{h}_{k_n}(x) - f(x)| dx \to 0$ i sannsynlighet eller med sannsynlighet 1?
- Korollar av resultatene i Lugosi and Nobel (1996): Ja, forutsatt at sannsynligheten på hver blokk går mot 0 og $k^{-1}n \to \infty$ etter som $n \to \infty$, $k \to \infty$.
- Uvisst om det er sant for alle tettheter såfremt $k^{-1}n \to \infty$, men virker rimelig.

Konsistens

Er histogrammene konsistente?

- (L_1) Finnes en k_n s.a. $\int |\widehat{h}_{k_n}(x) - f(x)| dx \to 0$ i sannsynlighet eller med sannsynlighet 1?

- Korollar av resultatene i Lugosi and Nobel (1996): Ja, forutsatt at sannsynligheten på hver blokk går mot 0 og $k^{-1}n \to \infty$ etter som $n \to \infty$, $k \to \infty$.

- Uvisst om det er sant for alle tettheter såfremt $k^{-1}n \to \infty$, men virker rimelig.

Konsistens

Er histogrammene konsistente?

- (L_1) Finnes en k_n s.a. $\int |\widehat{h}_{k_n}(x) - f(x)| dx \to 0$ i sannsynlighet eller med sannsynlighet 1?
- Korollar av resultatene i Lugosi and Nobel (1996): Ja, forutsatt at sannsynligheten på hver blokk går mot 0 og $k^{-1}n \to \infty$ etter som $n \to \infty$, $k \to \infty$.
- Uvisst om det er sant for alle tettheter såfremt $k^{-1}n \to \infty$, men virker rimelig.

Konsistens

Er histogrammene konsistente?

- (L_1) Finnes en k_n s.a. $\int |\widehat{h}_{k_n}(x) - f(x)| dx \to 0$ i sannsynlighet eller med sannsynlighet 1?
- Korollar av resultatene i Lugosi and Nobel (1996): Ja, forutsatt at sannsynligheten på hver blokk går mot 0 og $k^{-1}n \to \infty$ etter som $n \to \infty$, $k \to \infty$.
- Uvisst om det er sant for alle tettheter såfremt $k^{-1}n \to \infty$, men virker rimelig.

Oppsummering

- Teorien om kubikkrotasymptotikk bygger på en komplisert, generell teori om M-estimering. Kubikkrotasympotikk er ille på flere måter enn lav effisiens.
- Bruddpunktene i irregulære histogrammer konvergerer med kubikkrotrate når den underliggende tettheten er tilstrekkelig glatt. Histogrammene er konsistente under svake ekstrabetingelser. CIC, en variant av AIC, er god til å velge antall blokker, men vanskelig å beregne.

Oppsummering

- Teorien om kubikkrotasymptotikk bygger på en komplisert, generell teori om M-estimering. Kubikkrotasympotikk er ille på flere måter enn lav effisiens.
- Bruddpunktene i irregulære histogrammer konvergerer med kubikkrotrate når den underliggende tettheten er tilstrekkelig glatt. Histogrammene er konsistente under svake ekstrabetingelser. CIC, en variant av AIC, er god til å velge antall blokker, men vanskelig å beregne.

Banerjee, M. and McKeague, I. W. (2007), 'Confidence sets for split points in decision trees', *The Annals of Statistics* **35**(2), 543–574.

Chernoff, H. (1964), 'Estimation of the mode', *Annals of the Institute of Statistical Mathematics* **16**(1), 31–41.

Kim, J. and Pollard, D. (1990), 'Cube root asymptotics', *The Annals of Statistics* pp. 191–219.

Lugosi, G. and Nobel, A. (1996), 'Consistency of data-driven histogram methods for density estimation and classification', *The Annals of Statistics* **24**(2), 687–706.

Manski, C. F. (1975), 'Maximum score estimation of the stochastic utility model of choice', *Journal of econometrics* **3**(3), 205–228.

Rousseeuw, P. J. (1984), 'Least median of squares regression', *Journal of the American statistical association* **79**(388), 871–880.

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence*, Springer.