

STK1100 våren 2023

Estimeringsmetoder

Svarer til avsnitt 7.2 i læreboka

Matematisk institutt
Universitetet i Oslo

Bakgrunn

- Vi antar at vi har observasjoner x_1, \dots, x_n av de stokastiske variablene X_1, \dots, X_n .

Bakgrunn

- Vi antar at vi har observasjoner x_1, \dots, x_n av de stokastiske variablene X_1, \dots, X_n .
- X_1, \dots, X_n har simultanpunktsannsynlighet/-sannsynlighetstetthet $f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$, der $\theta_1, \dots, \theta_m$ er ukjente parametere.

Bakgrunn

- Vi antar at vi har observasjoner x_1, \dots, x_n av de stokastiske variablene X_1, \dots, X_n .
- X_1, \dots, X_n har simultanpunktsannsynlighet/-sannsynlighetstetthet $f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$, der $\theta_1, \dots, \theta_m$ er ukjente parametere.
- Vi ønsker å estimere disse ved hjelp av estimatorene $\hat{\theta}_1 = \hat{\theta}_1(X_1, \dots, X_n), \dots, \hat{\theta}_m = \hat{\theta}_m(X_1, \dots, X_n)$.

Bakgrunn

- Vi antar at vi har observasjoner x_1, \dots, x_n av de stokastiske variablene X_1, \dots, X_n .
- X_1, \dots, X_n har simultanpunktsannsynlighet/-sannsynlighetstetthet $f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$, der $\theta_1, \dots, \theta_m$ er ukjente parametere.
- Vi ønsker å estimere disse ved hjelp av estimatorene $\hat{\theta}_1 = \hat{\theta}_1(X_1, \dots, X_n), \dots, \hat{\theta}_m = \hat{\theta}_m(X_1, \dots, X_n)$.
- For noen fordelinger har vi “naturlige” estimatorer, slik som $\hat{p} = \frac{Y}{n}$ for p i en binomisk fordeling.

Bakgrunn

- Vi antar at vi har observasjoner x_1, \dots, x_n av de stokastiske variablene X_1, \dots, X_n .
- X_1, \dots, X_n har simultanpunktsannsynlighet/-sannsynlighetstetthet $f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$, der $\theta_1, \dots, \theta_m$ er ukjente parametere.
- Vi ønsker å estimere disse ved hjelp av estimatorene $\hat{\theta}_1 = \hat{\theta}_1(X_1, \dots, X_n), \dots, \hat{\theta}_m = \hat{\theta}_m(X_1, \dots, X_n)$.
- For noen fordelinger har vi “naturlige” estimatorer, slik som $\hat{p} = \frac{Y}{n}$ for p i en binomisk fordeling.
- Dersom vi ikke har noen slik estimator, trenger vi **estimeringsmetoder**. Vi skal ta for oss to slike: momentmetoden og maksimum likelihood.

Momentmetoden

- Vi antar at X_1, \dots, X_n er uavhengige og identisk fordelt med punktsannsynlighet/sannsynlighetstetthet $f(x; \theta_1, \dots, \theta_m)$.

Momentmetoden

- Vi antar at X_1, \dots, X_n er uavhengige og identisk fordelt med punktsannsynlighet/sannsynlighetstetthet $f(x; \theta_1, \dots, \theta_m)$.
- Det k -te **teoretiske momentet** til X_i , for $i = 1, \dots, n$, er da gitt ved

$$E(X_i^k) = \begin{cases} \sum_x x^k f(x; \theta_1, \dots, \theta_m) & \text{hvis } X_i \text{ er diskret} \\ \int x^k f(x; \theta_1, \dots, \theta_m) dx & \text{hvis } X_i \text{ er kontinuerlig.} \end{cases}$$

Momentmetoden

- Vi antar at X_1, \dots, X_n er uavhengige og identisk fordelt med punktsannsynlighet/sannsynlighetstetthet $f(x; \theta_1, \dots, \theta_m)$.
- Det k -te **teoretiske momentet** til X_i , for $i = 1, \dots, n$, er da gitt ved

$$E(X_i^k) = \begin{cases} \sum_x x^k f(x; \theta_1, \dots, \theta_m) & \text{hvis } X_i \text{ er diskret} \\ \int x^k f(x; \theta_1, \dots, \theta_m) dx & \text{hvis } X_i \text{ er kontinuertlig.} \end{cases}$$

- Det k -te **empiriske momentet** er gitt ved

$$\frac{1}{n} \sum_{i=1}^n X_i^k.$$

Momentmetoden

- Vi antar at X_1, \dots, X_n er uavhengige og identisk fordelt med punktsannsynlighet/sannsynlighetstetthet $f(x; \theta_1, \dots, \theta_m)$.
- Det k -te **teoretiske momentet** til X_i , for $i = 1, \dots, n$, er da gitt ved

$$E(X_i^k) = \begin{cases} \sum_x x^k f(x; \theta_1, \dots, \theta_m) & \text{hvis } X_i \text{ er diskret} \\ \int x^k f(x; \theta_1, \dots, \theta_m) dx & \text{hvis } X_i \text{ er kontinuerlig.} \end{cases}$$

- Det k -te **empiriske momentet** er gitt ved

$$\frac{1}{n} \sum_{i=1}^n X_i^k.$$

- **Momentestimatorene** $\hat{\theta}_1, \dots, \hat{\theta}_m$ for $\theta_1, \dots, \theta_m$ er da løsningene til likningene

$$\frac{1}{n} \sum_{i=1}^n X_i^k = E(X_j^k), \quad k = 1, \dots, m.$$

Momentmetoden (forts.)

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Bernoulli}(p)$.

Momentmetoden (forts.)

Eksempel

$$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Bernoulli}(p).$$

Eksempel

$$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Gamma}(\alpha, \beta).$$

Momentmetoden (forts.)

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Bernoulli}(p)$.

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Gamma}(\alpha, \beta)$.

Eksempel

Momentestimer for nedbørsdataene.

Maksimum likelihood-estimering

Eksempel

$X_1, \dots, X_n \stackrel{\text{uif}}{\sim} \text{Bernoulli}(p)$.

Maksimum likelihood-estimering

Eksempel

$X_1, \dots, X_n \stackrel{\text{uif}}{\sim} \text{Bernoulli}(p)$.

- La $f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$ være simultantettheten/-punktsannsynligheten til X_1, \dots, X_n .

Maksimum likelihood-estimering

Eksempel

$X_1, \dots, X_n \stackrel{\text{uif}}{\sim} \text{Bernoulli}(p)$.

- La $f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$ være simultantettheten/-punktsannsynligheten til X_1, \dots, X_n .
- Når vi ser på denne som en funksjon av $\theta_1, \dots, \theta_m$ for gitte data x_1, \dots, x_n , kalles den **likelihood-funksjonen**.

Maksimum likelihood-estimering

Eksempel

$X_1, \dots, X_n \stackrel{\text{uif}}{\sim} \text{Bernoulli}(p)$.

- La $f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$ være simultantettheten/-punktsannsynligheten til X_1, \dots, X_n .
- Når vi ser på denne som en funksjon av $\theta_1, \dots, \theta_m$ for gitte data x_1, \dots, x_n , kalles den **likelihood-funksjonen**.
- **Makimum likelihood-estimatene** (sannsynlighetsmaksimeringsestimatene) $\hat{\theta}_1, \dots, \hat{\theta}_m$ er de verdiene av $\theta_1, \dots, \theta_m$ som maksimerer likelihood-funksjonen.

Maksimum likelihood-estimering

Eksempel

$X_1, \dots, X_n \stackrel{\text{uif}}{\sim} \text{Bernoulli}(p)$.

- La $f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$ være simultantettheten/-punktsannsynligheten til X_1, \dots, X_n .
- Når vi ser på denne som en funksjon av $\theta_1, \dots, \theta_m$ for gitte data x_1, \dots, x_n , kalles den **likelihood-funksjonen**.
- **Makimum likelihood-estimatene** (sannsynlighetsmaksimeringsestimatene) $\hat{\theta}_1, \dots, \hat{\theta}_m$ er de verdiene av $\theta_1, \dots, \theta_m$ som maksimerer likelihood-funksjonen.
- Tilsvarende maksimum likelihood-estimatorer (MLE) får en ved å erstatte de observerte dataene med de stokastiske variablene X_1, \dots, X_n .

Maksimum likelihood-estimering (forts.)

Eksempel

$X_1, \dots, X_n \stackrel{\text{uif}}{\sim} \text{Ekspoenensiell}(\lambda)$ og vi vil finne MLE for λ .

Maksimum likelihood-estimering (forts.)

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Eksponensiell}(\lambda)$ og vi vil finne MLE for λ .

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} N(\mu, \sigma^2)$ og vi vil finne MLE for μ og σ^2 .

Maksimum likelihood-estimering (forts.)

Eksempel

$X_1, \dots, X_n \stackrel{\text{uif}}{\sim} \text{Eksponensiell}(\lambda)$ og vi vil finne MLE for λ .

Eksempel

$X_1, \dots, X_n \stackrel{\text{uif}}{\sim} N(\mu, \sigma^2)$ og vi vil finne MLE for μ og σ^2 .

Eksempel

$X_1, \dots, X_n \stackrel{\text{uif}}{\sim} \text{Gamma}(\alpha, \beta)$ og vi vil finne MLE for α og β .

Maksimum likelihood-estimering (forts.)

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Ekspoenensiell}(\lambda)$ og vi vil finne MLE for λ .

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} N(\mu, \sigma^2)$ og vi vil finne MLE for μ og σ^2 .

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Gamma}(\alpha, \beta)$ og vi vil finne MLE for α og β .

- Ofte er det ikke mulig å finne et analytisk uttrykk for maksimum likelihood-estimatorene.

Maksimum likelihood-estimering (forts.)

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Eksponensiell}(\lambda)$ og vi vil finne MLE for λ .

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} N(\mu, \sigma^2)$ og vi vil finne MLE for μ og σ^2 .

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Gamma}(\alpha, \beta)$ og vi vil finne MLE for α og β .

- Ofte er det ikke mulig å finne et analytisk uttrykk for maksimum likelihood-estimatorene.
- Da må en i stedet maksimere likelihood-funksjonen numerisk. I Python finnes en funksjon `stats.gamma.fit` som gjør dette. Tilsvarende for andre fordelinger.

Maksimum likelihood-estimering (forts.)

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Eksponensiell}(\lambda)$ og vi vil finne MLE for λ .

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} N(\mu, \sigma^2)$ og vi vil finne MLE for μ og σ^2 .

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Gamma}(\alpha, \beta)$ og vi vil finne MLE for α og β .

- Ofte er det ikke mulig å finne et analytisk uttrykk for maksimum likelihood-estimatorene.
- Da må en i stedet maksimere likelihood-funksjonen numerisk. I Python finnes en funksjon `stats.gamma.fit` som gjør dette. Tilsvarende for andre fordelinger.

Eksempel

Maksimum likelihood-estimerer for nedbørsdataene.

Viktige egenskaper ved MLE

- MLE er under veldig generelle betingelser konsistent, hvilket betyr at $\hat{\theta} \xrightarrow{P} \theta$.

Viktige egenskaper ved MLE

- MLE er under veldig generelle betingelser konsistent, hvilket betyr at $\hat{\theta} \xrightarrow{P} \theta$.
- Med andre ord vil $\hat{\theta}$ være nær θ i verdi bare n er stor nok.

Viktige egenskaper ved MLE

- MLE er under veldig generelle betingelser konsistent, hvilket betyr at $\hat{\theta} \xrightarrow{P} \theta$.
- Med andre ord vil $\hat{\theta}$ være nær θ i verdi bare n er stor nok.
- En annen viktig egenskap ved MLE er at dersom $\phi = \phi(\theta_1, \dots, \theta_m)$ er en en-entydig transformasjon av parameterne og $\hat{\theta}_1, \dots, \hat{\theta}_m$ er MLE for $\theta_1, \dots, \theta_m$, så er $\hat{\phi} = \phi(\hat{\theta}_1, \dots, \hat{\theta}_m)$ MLE for ϕ . Kalles INVARIANS-prinsippet.

Viktige egenskaper ved MLE

- MLE er under veldig generelle betingelser konsistent, hvilket betyr at $\hat{\theta} \xrightarrow{P} \theta$.
- Med andre ord vil $\hat{\theta}$ være nær θ i verdi bare n er stor nok.
- En annen viktig egenskap ved MLE er at dersom $\phi = \phi(\theta_1, \dots, \theta_m)$ er en en-entydig transformasjon av parameterne og $\hat{\theta}_1, \dots, \hat{\theta}_m$ er MLE for $\theta_1, \dots, \theta_m$, så er $\hat{\phi} = \phi(\hat{\theta}_1, \dots, \hat{\theta}_m)$ MLE for ϕ . Kalles INVARIANS-prinsippet.

Eksempel

$X_1, \dots, X_n \stackrel{\text{uif}}{\sim} \text{Eksponensiell}(\lambda)$ og vi vil finne MLE for $\phi = \frac{1}{\lambda}$.

Viktige egenskaper ved MLE

- MLE er under veldig generelle betingelser konsistent, hvilket betyr at $\hat{\theta} \xrightarrow{P} \theta$.
- Med andre ord vil $\hat{\theta}$ være nær θ i verdi bare n er stor nok.
- En annen viktig egenskap ved MLE er at dersom $\phi = \phi(\theta_1, \dots, \theta_m)$ er en en-entydig transformasjon av parameterne og $\hat{\theta}_1, \dots, \hat{\theta}_m$ er MLE for $\theta_1, \dots, \theta_m$, så er $\hat{\phi} = \phi(\hat{\theta}_1, \dots, \hat{\theta}_m)$ MLE for ϕ . Kalles INVARIANS-prinsippet.

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} \text{Eksponensiell}(\lambda)$ og vi vil finne MLE for $\phi = \frac{1}{\lambda}$.

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} N(\mu, \sigma^2)$ og vi vil finne MLE for σ .

MLE og den deriverte av likelihooden

- Så langt har vi bare sett på tilfeller der de partiellderiverte til (log-)likelihood-funksjonen er lik 0 i maksimum likelihood-estimatene.

MLE og den deriverte av likelihooden

- Så langt har vi bare sett på tilfeller der de partiellderiverte til (log-)likelihood-funksjonen er lik 0 i maksimum likelihood-estimatene.
- Det er imidlertid ikke alltid tilfellet.

MLE og den deriverte av likelihooden

- Så langt har vi bare sett på tilfeller der de partiellderiverte til (log-)likelihood-funksjonen er lik 0 i maksimum likelihood-estimatene.
- Det er imidlertid ikke alltid tilfellet.

Eksempel

$X_1, \dots, X_n \stackrel{uif}{\sim} U[0, \theta]$ og vi vil finne MLE for θ .