Tõenäosusteooria ja statistika II

Loeng 10

Suurima tõepära meetod

Märt Möls

Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Kui teame mudelit andmetele (näiteks seda, millisest jaotuste perest on pärit vaatlused) kuid ei tea teatud arvu tundmatute parameetrite väärtuseid…

… siis suurima tõepära meetodi korral valitakse tundmatute parameetrite väärtused selliselt, et meie poolt nähtud valimi (andmete) saamine oleks kõige tõenäolisem.

Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Andmed (sõltumatud mündivisked):

KULL, KULL, KIRI

Mis on kulli saamise tõenäosus, P(KULL)=?

Arvame tõenäosuse P(KULL) olevat sellise, mille puhul oleks vaadeldud andmete nägemise tõenäosus kõige suurem…

Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Andmed (sõltumatud mündivisked):

KULL, KULL, KIRI

Mis on kulli saamise tõenäosus, P(KULL)=?

Kui

P(KULL) = 0,1

siis vaadeldud andmete nägemise tõenäosus (likelihood) oleks

$$L = 0{,}1 \cdot 0{,}1 \cdot 0{,}9 = 0{,}009$$

Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Andmed (sõltumatud mündivisked):

KULL, KULL, KIRI

Mis on kulli saamise tõenäosus, P(KULL)=?

Kui

P(KULL) = 0,9

siis vaadeldud andmete nägemise tõenäosus (likelihood) oleks

$L = 0{,}9 \cdot 0{,}9 \cdot 0{,}1 = 0{,}081$

Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Andmed: KULL, KULL, KIRI

Mis on kulli saamise tõenäosus, P(KULL)=?

Kui

P(KULL) = 2/3

siis vaadeldud andmete nägemise tõenäosus (likelihood) oleks

$L = 2/3 \cdot 2/3 \cdot 1/3 = 4/27 = 0{,}148$

Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Andmed: KULL, KULL, KIRI

Mis on kulli saamise tõenäosus, P(KULL)=?

Kui P(KULL) = p

siis $L = p \cdot p \cdot (1-p) = p^2 - p^3$

$L' = 2p - 3p^2$

maksimum: $2p-3p^2 = 0$

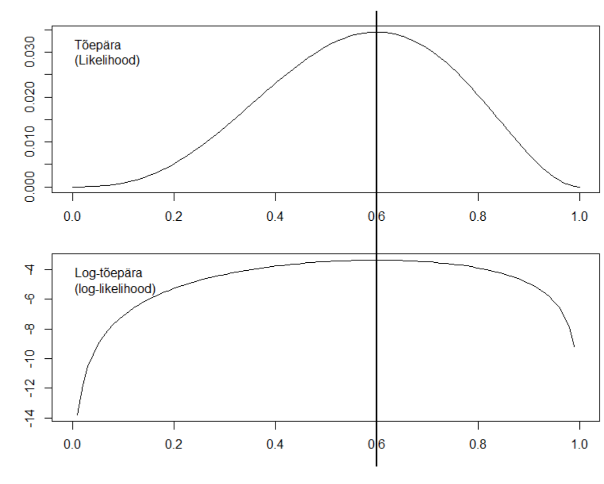
$2p = 3p^2$

$2 = 3p \rightarrow \hat{p} = 2/3$

Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Logaritm on monotoonne funktsioon – kui x kasvab, siis ka log(x) kasvab ja vastupidi.

Seega asuvad tõepärafunktsiooni ja log-tõepära maksimumid samas kohas – ja tõepära maksimiseerimise asemel võime ka maksimiseerida log-tõepära.



Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Vaatlused (sõltumatud): $X_1, \ldots, X_n \sim N(\mu; \sigma^2)$

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2}\right) \cdot \ldots \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_n-\mu)^2}{2\sigma^2}\right)$$

$$L(\mu;\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2}\right) \cdot \ldots \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_n-\mu)^2}{2\sigma^2}\right)$$

$$l(\mu;\sigma^2) = -n/2 \log(2\pi\sigma^2) - 1/(2\sigma^2) \sum_{i=1}^{n} (x_i-\mu)^2$$

Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Vaatlused (sõltumatud): $X_1, \ldots, X_n \sim N(\mu; \sigma^2)$

$$l(\mu;\sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2$$

Maksimum

$$\begin{cases} \frac{\partial l(\mu;\sigma^2)}{\partial \mu} = 0 \\ \frac{\partial l(\mu;\sigma^2)}{\partial \sigma^2} = 0 \end{cases}$$

Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Vaatlused (sõltumatud): $X_1, \ldots, X_n \sim N(\mu; \sigma^2)$

$$l(\mu;\sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2$$

$$\frac{\partial l}{\partial \mu} = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} 2(x_i-\mu) = -\frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - n\mu\right)$$

$$\frac{\partial l}{\partial \mu} = 0$$

$$\sum_{i=1}^{n} x_i = n\hat{\mu} \qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Vaatlused (sõltumatud): $X_1, \ldots, X_n \sim N(\mu; \sigma^2)$

$$l(\mu;\sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{2\pi\sigma^2} \cdot 2\pi + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i-\mu)^2 \qquad \frac{\partial l}{\partial \sigma^2} = 0$$

$$n\hat{\sigma}^2 = \sum_{i=1}^{n} (x_i-\hat{\mu})^2$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i-\hat{\mu})^2$$

Suurima tõepära hinnang ei sõltu parametriseeringust

Mõnikord võime tõepära vaadata kui parameetri λ
või kui parameetri Θ=g(λ) funktsiooni:

$$L(\theta) = L(g(\lambda)) = L^*(\lambda)$$

Näide:
Normaaljaotus

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Kas parameetriteks on: μ; σ
või on normaaljaotuse parameetriteks: μ; $\sigma^2 = g(\sigma)$

Suurima tõepära hinnang ei sõltu parametriseeringust

Mõnikord võime tõepära vaadata kui parameetri λ
või kui parameetri Θ=g(λ) funktsiooni.

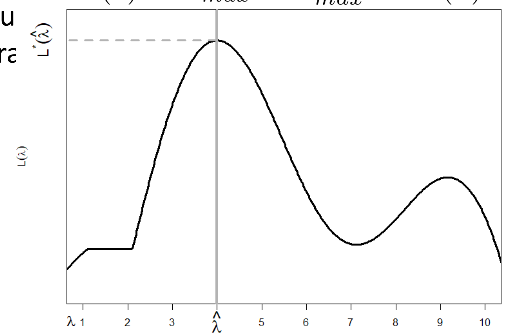
$$L(\theta) = L(g(\lambda)) = L^*(\lambda)$$

$$L(\hat{\theta}) = L_{max} = L^*_{max} = L^*(\hat{\lambda})$$

Mõlema parameetriseeringu korral leitud suurima tõepära hinnangud on omavahel seotud:

$$\hat{\theta} = g(\hat{\lambda}).$$

Võin leida suurima tõepära hinnangu σ-le, $\hat{\sigma}$, ja arvutada suurima tõepära hinnangu σ^2-le valemiga $\hat{\sigma}^2 = (\hat{\sigma})^2$.



Suurima tõepära meetod (Maximum Likelihood, ML, EM, …)

Üldjuhul (sõltumatud vaatlused)

$$L(\theta_1; \ldots; \theta_k) = g(x_1) \cdot \ldots \cdot g(x_2)$$

kus

$$g(x_i) = P(X_i = x_i),$$

kui jaotusfunktsioon pole pidev kohal x_i

$$g(x_i) = f_{X_i}(x_i)$$

kui jaotusfunktsioon on pidev kohal x_i

Näide. $X_i \sim N(\mu;\sigma=5)$. Soovime hinnata μ väärtust.

$$\bar{x} = 8{,}25 \qquad \hat{\mu} = 8{,}25?$$

6.87	10.00	5.82	10.00	10.00	5.90	10.00	10.00	10.00	8.47	10.00	10.00
6.89	0.00	10.00	9.78	9.92	10.00	10.00	10.00	10.00	10.00	10.00	0.05
10.00	9.72	9.22	2.65	7.61	10.00	10.00	9.49	10.00	9.73	3.11	7.93
8.03	9.70	10.00	10.00	9.18	8.73	10.00	10.00	6.56	6.46	10.00	10.00
9.44	10.00	10.00	6.94	10.00	4.35	10.00	10.00	8.16	4.78	10.00	9.32
10.00	9.80	10.00	10.00	6.28	10.00	0.98	10.00	10.00	10.00	10.00	6.45
10.00	5.33	3.73	10.00	7.78	10.00	10.00	7.05	7.16	9.32	10.00	2.38
10.00	10.00	10.00	8.48	10.00	10.00	7.29	10.00	10.00	10.00	10.00	10.00
3.62	7.13	3.88	7.63	6.90	10.00	5.45	10.00	6.73	10.00	10.00	10.00
10.00	10.00	6.82	7.69	10.00	6.75	8.96	8.04	8.40	8.60	10.00	9.11
7.47	10.00	8.93	9.10	9.50	10.00	9.63	9.81	6.59	8.38	10.00	7.06
10.00	2.41	10.00	2.32	8.50	7.36	6.74	9.72	0.43	10.00	1.68	7.68
4.42	6.25	10.00	10.00	3.57	1.80	10.00	9.91	8.41	5.35	2.56	4.62
10.00	6.89	3.08	10.00	10.00	8.81	10.00	10.00	6.90	10.00	8.72	2.88
9.28	10.00	10.00	10.00	10.00	9.61	8.33	9.83	10.00	10.00	10.00	10.00
3.84	10.00	10.00	2.66	10.00	9.21	10.00	6.17	7.85	5.37	9.11	10.00
6.34	10.00	3.96	4.76	10.00	4.92	10.00	8.09				

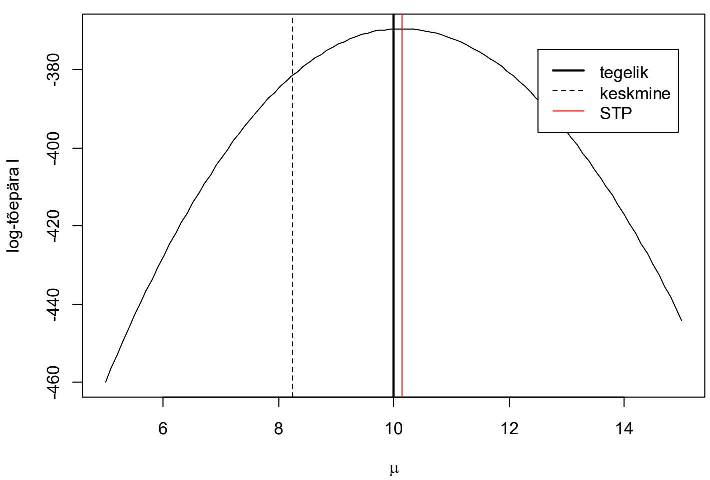
Mõõteriist…

Vaatluse panus valimi tõepärafunktsiooni

$$\frac{1}{\sqrt{2\pi\cdot 25}}\exp\left(-\frac{(x_i-\mu)^2}{2\cdot 25}\right)$$ kui $0 < x_i < 10$

$$\phi\left(\frac{0-\mu}{5}\right)$$ kui $x_i=0$

$$1-\phi\left(\frac{10-\mu}{5}\right)$$ kui $x_i=10$



Suurima tõepära hinnang on mõjus

$$\hat{\theta}_n = \arg\max_{\theta} L_n(\theta) = \arg\max_{\theta} \frac{1}{n}\sum_{i=1}^{n}\log f(X_i|\theta)$$

$$L_n(\theta) = \frac{1}{n} l(\theta)$$

defineerime

$$L(\theta) = \mathrm{E}\log f(X|\theta) = \int_{-\infty}^{\infty} \log f(x|\theta) f(x|\theta_{tegelik})\, dx$$

suurte arvude seaduse järgi (kui defineeritud keskväärtus eksisteerib):

$$L_n(\theta) \to L(\theta)$$

Vaja veel näidata, et

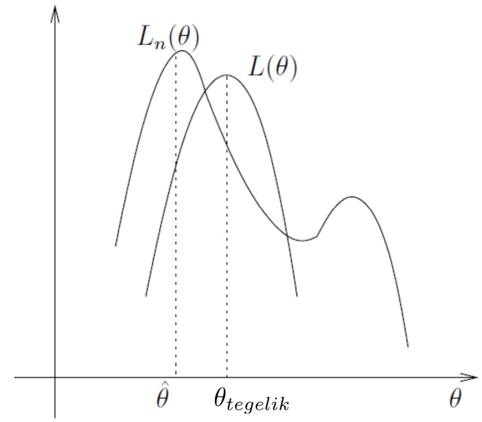
$$L(\theta) \le L(\theta_{tegelik})$$

iga võimaliku parameetri väärtuse Θ korral.

$$
\begin{aligned}
L(\theta) - L(\theta_{tegelik}) &= \mathrm{E}\left(\log f(X|\theta) - \log f(X|\theta_{tegelik})\right) \\
&= \mathrm{E}\left(\log \frac{f(X|\theta)}{f(X|\theta_{tegelik})}\right)
\end{aligned}
$$

kuna $\log(x) \leq x - 1$

$$
\begin{aligned}
\mathrm{E}\left(\log \frac{f(X|\theta)}{f(X|\theta_{tegelik})}\right) &\leq \mathrm{E}\left(\frac{f(X|\theta)}{f(X|\theta_{tegelik})} - 1\right) \\
&= \int \left(\frac{f(X|\theta)}{f(X|\theta_{tegelik})} - 1\right) f(X|\theta_{tegelik})\, dx \\
&= \int \left(f(X|\theta) - f(X|\theta_{tegelik})\right)\, dx \\
&= 0
\end{aligned}
$$



Funktsioon $L_n(\theta)$ saavutab maksimumi kohas $\hat{\theta}$;
Funktsioon $L(\theta)$ saavutab maksimumi kohas $\theta_{tegelik}$;
$\forall\theta\ \ L_n(\theta) \to L(\theta)$

Suurima tõepära hinnang

On mõjus

On asümptootiliselt normaaljaotusega (teatud eeldustel),

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0; \sigma^2_{ST})$$

On asümptootiliselt nihketa (teatud eeldustel)

On asümptootiliselt efektiivne (teatud eeldustel)

Kui soovime hinnata parameetri Θ funktsiooni $g(\Theta)$ väärtust, siis pole vahet kas leiame STH parameetrile Θ ja arvutame seejärel kasutades leitud hinnangut hinnangu $g(\widehat{\Theta})$ või hindame otse suurust $g(\Theta)$, $\widehat{g(\Theta)}$, tulemus on sama. Näiteks võime hinnata σ^2 väärtust ja arvutada standardhälbe hinnangu valemiga $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$. Või võime otse hinnata suurima tõepära meetodil suurust $\hat{\sigma}$. Saame tulemuseks ikka sama hinnangu.