Meeldetuletuseks

Loeng 10

1) Suurima tõepära hinnang on mõjus. Olgu meie vaatlused pärit jaotusest $f(\cdot|\theta)$. Siis STP hinnang $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ koondub tõenäosuse järgi parameetri tegelikuks väärtuseks. Formaalselt

$$X_1, \ldots, X_n \sim f(\cdot|\theta),$$
$$\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} = h(X_1, \ldots, X_n),$$
$$\forall \epsilon > 0, \quad \mathbf{P}(|\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \theta| > \epsilon) \xrightarrow{n} 0.$$

Intuitiivselt tähendab mõjusus seda, et vaatluste arvu kasvades saame STP hinnanguga parameetril tõelisele väärtusele kuitahes lähedale.

2) Suurima tõepära hinnang on asümptootiliselt normaaljaotusega:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_{STP}).$$

NB! Tulemused (1) ja (2) kehtivad teatud tehnilistel eeldustel, mille sõnastamine nõuab mõningaid teadmisi funktsionaalanalüüsist. Selle kursuse raames võib vaikimisi eeldada, et STP hinnangute korral on need eeldused rahuldatud ja omadused kehtivad.

Loeng 11

Olgu g diferentseeruv ja kehtigu $g'(\theta) = 0$.

1) Olgu $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ mõjus hinnang parameetrile θ keskväärtusega $E\hat{\boldsymbol{\theta}}_{\boldsymbol{n}} = \theta$ ja dispersiooniga $D\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$. Funktsiooni $g(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})$ ligikaudne keskväärtus ja dispersioon avalduvad järgmiselt:

$$E[g(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})] \approx g(\theta),$$

$$D[g(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})] \approx g'(\theta)^2 D(\hat{\boldsymbol{\theta}}).$$

2) Olgu hinnang $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ asümptootiliselt jaotusega $\mathcal{N}(0, \sigma_h^2)$. Siis ka hinnangu funktsioon $g(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})$ on asümptootiliselt normaalne, s.t

$$\sqrt{n}\big(g(\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}) - g(\theta)\big) \xrightarrow{d} \mathcal{N}(0, \sigma_h^2 \cdot g'(\theta)^2).$$

Ülesanne 1

Mõningates maades armastatakse juhuslikkuse kirjeldamisel kasutada tõenäosuse asemel šanssi (*odds*). Näiteks räägitakse, et hobuse šansid võita on 1:1 ehk 1 (s.t võidu tõenäosus on 0,5). Kui sündmuse toimumise tõenäosus on p, siis on selle sündmuse toimumise šanss on $\frac{p}{1-p}$.

Kui nuputatakse välja uusi ravimeetodeid siis võrreldakse sageli ka šansse. Näiteks võib küsida, kas patsienti uuel moel opereerides on operatsiooni õnnestumise šansid paremad kui tavapärase operatsiooni korral. Võrdlus tehakse sageli šansside suhet (*odds ratio, OR*) kasutades. Kui näiteks $p1$ tähistab tõenäosust, et uus raviviis töötab (viib soovitud tulemuseni) ja $p2$ tähistab traditsioonilise ravi õnnestumise tõenäosust, siis uue ravi šansid olla edukad on $OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$ korda suuremad võrreldes vana raviga.

Tegelikus elus muidugi ei tea me raviviiside õnnestumiste tõenäosuseid ja seega tuleb need tõenäosused ja šansside suhe hinnata. Kus on tegemist aga hinnangutega tekib vajadus kirjeldada hinnangute täpsust: leida näiteks 95%-usaldusintervall šansside suhtele. Šansside suhte hinnangu standardviga tahame ehk ka teada.

Tänaseks ülesandeks ongi tuletada arvutusvalem usaldusintervalli leidmiseks ja tuleks ka leida šansside suhte hinnang ja 95%-usaldusintervall järgmise uuringu jaoks:

	õnnestunud operatsioone	**patsiente kokku**
kohene operatsioon	174	297
op seisundi stabiliseerumise järel	178	286

Vihjed:

a) Kasutades delta meetodit leia suuruse $\ln(\widehat{\textbf{šanss}})$ ligikaudne jaotus.

b) Milline võiks olla suuruse $\ln(\widehat{\mathbf{OR}})$ ligikaudne jaotus?

c) Kasutades suuruse $\ln(\widehat{\mathbf{OR}})$ ligikaudset jaotust tuleta 95%-usaldusintervall suurusele $\ln(\mathbf{OR})$.

d) Kasutades juhuslikule suurusele $\ln(\mathbf{OR})$ leitud usaldusintervalli, leia usaldusintervall suurusele $\mathbf{OR}$.

e) Võrdle saadud tulemust internetis esitatud valemitega (näiteks tee Google otsing märksõnadele: *confidence interval for odds ratio*). Kas väljapakutud arvutusvalem on sama või teistsugune kui sinu poolt saadud?

f) Rakenda saadud valemit näiteuuringu andmetel. Millise usaldusintervallini jõuad?

g) Proovi, kas $\ln(\widehat{\mathbf{OR}})$ ligikaudne jaotus tuleb sama mitmemõõtmelist delta meetodit kasutades?

Ülesanne 2

Teame, et tudengite keskmine pikkus on 1,71 m (± 0.003) ja kaal 63kg ($\pm 0{,}5$). Sulgudes antud hinnangute standardvead (keskmised pikkused ja keskmised kaalud on hinnatud erinevaid valimeid kasutades). Kui täpselt me teame nn tüüptudengi kehamassiindeksit (tudeng, kelle kaal on kaalu keskväärtus ja pikkus pikkuse keskväärtus)?