Tilastollisen päättelyn kurssi - 1. välikoe (14.12.2007, klo 13-15 A111)

1. Toistokokeessa suoritetaan n riippumatonta toistoa ja yksittäisen kokeen (toiston) onnistumistodennäköisyys on tuntematon θ. Satunnaismuuttuja Y_i saa arvon 1, jos i:s koe onnistuu ja arvon 0, jos se epäonnistuu, joten $P\{Y_i = 1\} = \theta$ ja $P\{Y_i = 0\} = 1 - \theta$. Muodosta tätä toistokoetta kuvaavan tilastollisen mallin lauseke. Johda sen jälkeen huolellisesti perustellen parametrin θ suurimman uskottavuuden estimaatti $\hat{\theta}$, kun havainnot ovat $y_1, ..., y_n$. *(5 p.)*

2. Tarkastellaan tilastollista mallia $f_{\mathbf{Y}}(\mathbf{y}; \theta)$, jonka parametri θ on reaalinen.
a) Miten määritellään mallin pistemääräfunktio?
b) Miten määritellään Fisherin informaatio $i(\theta)$?
c) Millaiset tulokset ovat voimassa pistemääräfunktion momenteille, kun malli on säännöllinen? Erityisesti, mikä yhteys on tällöin pistemääräfunktion ja Fisherin informaation välillä? (Perusteluja ei vaadita.) *(5 p.)*

3. Oletetaan, että havaintoja vastaavat satunnaismuuttujat ovat $Y_1, ..., Y_n \sim P(\mu) \perp\!\!\!\perp$. Perustele, että mille tahansa parametrin μ harhattomalle estimaattorille T on voimassa

$$var(T) \geq \mu/n.$$

Mitä nimitystä estimaattorista T käytetään, kun tässä pätee yhtäsuuruus? *(5 p.)*

4. Erään elektronisen komponentin kestoaika noudattaa eksponenttijakaumaa odotusarvona μ. Tutkittaessa n komponenttia toisistaan riippumatta saatiin niiden kestoajoiksi $y_1, ..., y_n$ aikayksikköä. Johda huolellisesti perustellen suurimman uskottavuuden estimaatit odotusarvolle μ ja todennäköisyydelle, että umpimähkään valitun komponentin kestoikä on vähintään y_0 aikayksikköä. *(5 p.)*

Muistin tueksi

Satunnaismuuttujan $Y \sim P(\mu)$ pistetodennäköisyysfunktio on $f(y; \mu) = \mu^y e^{-\mu}/y!$ ($y = 0, 1, ...,$ $\mu > 0$). Lisäksi $E(Y) = \mu$ ja $var(Y) = \mu$.

Jos satunnaismuuttuja Y noudattaa eksponenttijakaumaa parametrina λ (eli $Y \sim Exp(\lambda)$), niin sen tiheysfunktio on $f(y; \mu) = \lambda e^{-\lambda y}$ ja kertymäfunktio on $F(y; \mu) = 1 - e^{-\lambda y}$, $y > 0$. Lisäksi $E(Y) = 1/\lambda$ ja $\mathrm{var}(Y) = 1/\lambda^2$.

Huom.: Kurssin hyväksytyyn suoritukseen vaaditaan, että molemmista kurssikokeista saa vähintään 7 pistettä.

TILASTOLLINEN PÄÄTTELY, 10 OP (aineopinnot). Kirjallisuus: Pekka Nieminen ja Pentti Saikkonen: Tilastollisen päättelyn kurssi. Luennoi: yliopistonlehtori Pekka Pere.

1. välikoe 19.12.2008

Kolme ensimmäistä kysymystä ovat kuuden pisteen arvoisia. Kahdessa ensimmäisessä kysymyksessä oleta, että mallin parametri on yksiulotteinen. Palauta kysymyspaperi (kiitos!).

1. Määrittele ja selitä huolellisesti
a) uskottavuusfunktio, ja miten se eroaa pistetodennäköisyys- tai tiheysfunktiosta.
b) uskottavuusyhtälö ja suurimman uskottavuuden (SU) estimaattori.
c) odotettu informaatio. Esitä kaksi kaavaa, joiden avulla se voidaan määritellä. Tulkitse kaavat intuitiivisesti (kuten luennoilla tehtiin).

2. Merkitään parametrin θ estimaattoria T:llä. Määrittele ja selitä huolellisesti
a) estimaattorin T harhattomuus.
b) estimaattorin T keskineliövirhe. Miten se riippuu estimaattorin varianssista? Johda tämä tulos.
c) SU-estimaatti ja sen asymptoottiset ominaisuudet (riittävien säännöllisyysehtojen vallitessa).

3. Olkoot $Y_1, \ldots, Y_n \sim \mathsf{N}(\mu_0, \sigma^2)$ $\perp\!\!\!\perp$ (samoin normaalijakautuneita ja riippumattomia). Jakauman odotusarvo (μ_0) on tunnettu, mutta sen varianssi (σ^2) on tuntematon.
a) Mikä on havaintojen (n kappaletta) yhteistiheysfunktio? Esitä kaava, ja perustele huolellisesti.
b) Johda SU-estimaattori $\hat{\sigma}^2$ varianssille σ^2. (Toisen derivaatan merkkiä ei tarvitse tarkistaa.)
c) Johda Fisher-informaatio $i(\sigma^2)$ varianssille σ^2. (Vihje: Laske $i(\sigma^2)$ kaavalla, joka on monisteen mukaan ylipäänsä helpommin laskettavissa. Jos aikaa jää, tarkista tuloksesi laskemalla $i(\sigma^2)$ huomattavasti työläämmällä toisella kaavalla.)
d) Mikä on SU-estimaattorin $\hat{\sigma}^2$ approksimaativinen jakauma suurilla havaintomäärillä? Vaihtoehtoisesti vastaa, mikä on sopivasti normeeratun SU-estimaattorin $\hat{\sigma}^2$ asymptoottinen jakauma ja mikä on sopiva normeeraus. Tulkitse esittämäsi kaava intuitiivisesti.

4. Joululahja oppilailleni! Saat ylimääräisen pisteen, jos osaat vastata oikein ja perustellusti tähän kysymykseen.

Nordean mainosesitteessä Säästämisen uutiset (marraskuu 2008) selitetään korrelaatio näin:

Korrelaatiolla tarkoitetaan kahden eri tekijän kehitystä toisiinsa nähden. Korrelaatio(kerroin) vaihtelee lukujen +1 ja −1 välillä. Kun korrelaatio on yksi, kahden osakemarkkinan tuotot kehittyvät samalla tavoin, eli tuotot nousevat ja laskevat yhtä paljon samaan aikaan. Kun korrelaatio on miinus yksi, tuotot kehittyvät toistensa peilikuvina, eli toisen markkinan tuotto nousee saman verran kuin toisen laskee. Kun korrelaatio on nolla, tuottojen välillä ei voida havaita mitään vastaavuutta.

Onko korrelaation selitys hyvä? Perustele.

Seuraavat laskuharjoitukset tulevat piakkoin kurssisivulle. Uusintakuulustelu on yleistentissä 22.1. Siihen tulee ilmoittautua tavalliseen tapaan. Huom! Mainitkaa että tulette tenttimään vain kurssin ensimmäistä puolikasta!

Oikein Hyvää ja Rauhaisaa Joulua kaikille kurssilaisilleni!

Aputuloksia:

- Normaalijakautuneen satunnaismuuttujan $Y \sim \mathsf{N}(\mu, \sigma^2)$ tiheysfunktio on $(2\pi\sigma^2)^{-1/2}\exp[-(y_i-\mu)^2/2\sigma^2]$ (ilmeisin merkinnöin).
- Normaalijakaumaa $\mathsf{N}(\mu, \sigma^2)$ noudattavan satunnaismuuttujan Y neljäs keskusmomentti $\mathsf{E}(Y-\mu)^4$ on $3\sigma^4$.
- $(\sum_{i=1}^n y_i)^2 = (\sum_{i=1}^n \sum_{j=i}^n y_i y_j)^2 = \sum_{i=1}^n y_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^n y_i y_j$.
- $\sum_{i=1}^{n-1}\sum_{j=i+1}^n 1 = (n-1)+(n-2)+\cdots+1 = n(n-1)/2$.

Kolmea viimeistä tulosta et välttämättä tarvitse kokeessa.

TILASTOLLINEN PÄÄTTELY, 10 OP (aineopinnot). Kirjallisuus: Pekka Nieminen ja Pentti Saikkonen: Tilastollisen päättelyn kurssi.

1. välikokeen uusintakuulustelu yleistentissä 22.1.2009

Kysymykset ovat kuuden pisteen arvoisia. Kaikkien kysymysten kohdalla voit olettaa (luentomonisteessa kerrottujen) säännöllisyysehtojen pätevän. Palauta kysymyspaperi (kiitos!).

1. Olkoon mallin parametri $\theta \in \Omega$ yksiulotteinen. Määrittele ja selitä huolellisesti

a) SU-estimaattori $\hat{\theta}$ ja sen asymptoottiset ($n \longrightarrow \infty$) ominaisuudet (mm. jakauma).

b) parametrin θ estimaattorin (T) harha. Onko θ:n SU-estimaattori aina harhaton? Perustele huolellisesti.

c) SU-estimaattorien invarianssiominaisuus eli $g(\theta)$:n (monisteen merkinnöin) estimointi SU-menetelmällä.

d) havaittu ja odotettu eli Fisher-informaatio. Miten ne eroavat tulkinnallisesti? Miten odotettu informaatio liittyy SU-estimaattorin asymptoottisiin ominaisuuksiin?

2. Olkoon T parametrin $\theta \in \Omega$ (yksiulotteinen) harhaton estimaattori.

a) Esitä, ja selitä sanoin, informaatioepäyhtälö. Miten se liittyy SU-estimaattorin ominaisuuksiin?

b) Määrittele, ja selitä sanoin, estimaattorin T tehokkuus. Millainen estimaattori on täystehokas? Onko täystehokas estimaattori aina olemassa? Saat lisäpisteen, jos osaat vastata perustellusti oikein tähän kysymykseen: Milloin ainakin SU-estimaattori on täystehokas?

3. Olkoot $Y_1, \ldots, Y_n \sim \mathsf{N}(\mu_0, \sigma^2)$ $\perp\!\!\!\perp$ (samoin normaalijakautuneita ja riippumattomia). Jakauman odotusarvo (μ_0) on tunnettu, mutta sen keskihajonta ($\sigma > 0$) on tuntematon.

a) Mikä on havaintojen (n kappaletta) yhteistiheysfunktio? Esitä kaava, ja perustele huolellisesti.

b) Johda SU-estimaattori $\hat{\sigma}$:lle eli keskihajonnalle σ. Huom! Tehtävässä kysytään σ:n SU-estimaattoria eikä σ^2:n SU-estimaattoria. Tarkista, että olet löytänyt uskottavuusfunktion (paikallisen) maksimin.

c) Johda odotettu eli Fisher-informaatio $i(\sigma)$ keskihajonnalle σ. (Vihje: Laske $i(\sigma)$ kaavalla, joka on monisteen mukaan ylipäänsä helpommin laskettavissa. Jos aikaa jää, tarkista tuloksesi laskemalla $i(\sigma)$ huomattavasti työläämmällä toisella kaavalla.)

d) Mikä on SU-estimaattorin $\hat{\sigma}$ approksimatiivinen jakauma suurilla havaintomäärillä? Vaihtoehtoisesti vastaa, mikä on sopivasti normeeratun SU-estimaattorin $\hat{\sigma}$ asymptoottinen jakauma ja mikä on sopiva normeeraus. Tulkitse esittämäsi kaava intuitiivisesti.

Aputuloksia:

- Normaalijakautuneen satunnaismuuttujan $Y \sim \mathsf{N}(\mu, \sigma^2)$ tiheysfunktio on $(2\pi\sigma^2)^{-1/2} \exp[-(y_i - \mu)^2/2\sigma^2]$ (ilmeisin merkinnöin).
- Normaalijakaumaa $\mathsf{N}(\mu, \sigma^2)$ noudattavan satunnaismuuttujan Y neljäs keskusmomentti $\mathsf{E}(Y - \mu)^4$ on $3\sigma^4$.
- $(\sum_{i=1}^{n} y_i)^2 = (\sum_{i=1}^{n} \sum_{j=i}^{n} y_i y_j)^2 = \sum_{i=1}^{n} y_i^2 + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} y_i y_j$.
- $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 1 = (n-1) + (n-2) + \cdots + 1 = n(n-1)/2$.

Kolmea viimeistä tulosta et välttämättä tarvitse kokeessa.