Tilastollisen päättelyn kurssi - 2. välikoe (27.2.2008, klo 13-15 A111)

1. Havaintoja vastaavat satunnaismuuttujat ovat $Y_1, ..., Y_n \sim P(\mu)$ $\underline{\|}$. Etsi parametrille μ yksiulotteinen tyhjentävä tunnusluku. (*4 p.*)

2. a) Tilastollinen malli on $f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta})$. Testataan parametria $\boldsymbol{\theta}$ koskevaa nollahypoteesia $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ testisuuretta $t = t(\mathbf{y})$ käyttäen. Määrittele huolellisesti tähän liittyvä p-arvo eli havaittu merkitsevyystaso.
b) Selosta lyhyesti p-arvoihin liittyvä valintakorjauksen käsite: missä tilanteessa valintakorjauksen suorittaminen on tarpeen ja miten se voidaan käytännössä tehdä. (*4 p.*)

3. Toistokokeessa (onnistumistodennäköisyys θ) suoritetaan 8 toistoa ja lasketaan onnistumisten lukumäärä k. Halutaan testata hypoteesia $H_0 : \theta = 0.3$ vaihtoehtoa $H_1 : \theta > 0.3$ vastaan. Testisuureena on k.
a) Millaiset k:n arvot (pienet vai suuret) todistavat H_0:aa vastaan ja H_1:n puolesta?
b) Päätetään toimia merkitsevyystasolla 0.01. Ilmoita vastaava kriittinen alue (eli ne testisuureen arvot, jotka johtavat H_0:n hylkäämiseen ja H_1:n hyväksymiseen). Perustele tarkasti.
c) Kuinka suuri on hyväksymisvirheen todennäköisyys pisteessä $\theta = 0.5$. (*6 p.*)

Käytä hyväksesi oheista taulukkoa, jossa on lueteltu $Bin(8, p)$-jakauman pistetodennäköisyydet eräillä onnistumistodennäköisyyden p arvoilla

4. Havaintoja vastaavat satunnaismuuttujat ovat $Y_1, ..., Y_n \sim Exp(\lambda)$. Parametrin λ suurimman uskottavuuden estimaattori on tunnetusti $\hat{\lambda} = 1/\overline{Y}$, jossa $\overline{Y} = (Y_1 + \cdots + Y_n)/n$.
a) Laske Fisherin informaatio $i(\lambda)$.
b) Mitä normaalijakaumaa $\hat{\lambda}$ approksimatiivisesti noudattaa, kun n on suuri?
c) Oletetaan, että havaintoja on $n = 25$ ja niiden keskiarvoksi saadaan $\bar{y} = 5$. Muodosta Waldin testiin eli yo normaaliapproksimaatioon perustuva approksimatiivinen 95 %:n luottamusväli λ:lle. (*6 p.*)

Muistin tueksi

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

Satunnaismuuttujan $Y \sim Exp(\lambda)$ tiheysfunktio on $f(y; \lambda) = \lambda e^{-\lambda y}$, $y > 0, \lambda > 0$. Lisäksi $E(Y) = 1/\lambda$ ja $\mathrm{var}(Y) = 1/\lambda^2$.

Jos Φ on $N(0,1)$-jakauman kertymäfunktio, niin $\Phi(1.645) \approx 0.95$, $\Phi(1.96) \approx 0.975$, $\Phi(2.336) \approx 0.01$

Ole hyvä ja vastaa kurssikyselyyn: http://mathstat.helsinki.fi/kurssit/kysely/

Table of Binomial Probabilities (*Continued*)

n	k	$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$
8	0	.4305	.1678	.0576	.0168	.0039
	1	.3826	.3355	.1977	.0896	.0312
	2	.1488	.2936	.2965	.2090	.1094
	3	.0331	.1468	.2541	.2787	.2188
	4	.0046	.0459	.1361	.2322	.2734
	5	.0004	.0092	.0467	.1239	.2188
	6	.0000	.0011	.0100	.0413	.1094
	7	.0000	.0001	.0012	.0079	.0312
	8	.0000	.0000	.0001	.0007	.0039

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

2. välikoe 5.3.2009

Kolme ensimmäistä kysymystä ovat kuuden pisteen arvoisia. Korkeimman arvosanan voi saada vastaamalla pelkästään niihin. Palauta kysymyspaperi. Kiitos!

1. Olkoon tilastollisen mallin (mahdollisesti moniulotteinen) parametri $\boldsymbol{\theta} \in \Omega$. Määrittele ja selitä huolellisesti kaavoin ja sanoin
a) tyhjentävä tunnusluku.
b) faktorointikriteeri ja miksi se on hyödyllinen.
c) yksinkertainen ja yhdistetty hypoteesi.
d) Neyman–Pearson apulause.

2. Olkoon parametri $\theta \in \Omega$ yksiulotteinen. Määrittele ja selitä huolellisesti kaavoin ja sanoin
a) testin havaittu merkitsevyystaso (p).
b) uskottavuusosamäärä-, Waldin ja Raon testisuureet. Selitä niiden intuitiivinen geometrinen tulkinta (kuten luennoilla tehtiin) ja käyttö (jakauma, johon verrataan ja perustelu sille).
c) luottamusväli (sen frekventistinen tulkinta).
d) perustelu paljon käytetylle luottamusvälille $\hat{\theta} \pm 2 \times s.e.(\hat{\theta})$ (monisteen merkinnöin).

3. Tutkitaan eksponenttijakautuneita riippumattomia satunnaismuuttujia $Y_1, \ldots, Y_n \sim \mathsf{Exp}(1/\mu)$. Testataan nollahypoteesia $H_0 : \mu = \mu_0$, jossa $\mu_0 > 0$.
a) Muodosta uskottavuusosamäärä-, Waldin ja Raon testisuureet.
b) Olkoot otoskoko $n = 50$ ja $\bar{y} = 800$. Testaa mainituilla testisuureilla, onko syytä hylätä nollahypoteesia $H_0 : \mu = 1000$ merkitsevyystasolla $0,05$ (kaksisuuntainen testi).

4. Saat lisäpisteen, jos osaat selittää millaisten tilastotieteellisten argumenttien perusteella Fisher ei uskonut epidemiologisia tutkimuksia tupakoinnin ja syövän yhteydestä.

P.S. Katso uusimman National Geographic -lehden 2/2009 (Suomi) sivulta 34. millä aivan toisella tieteenalalla Fisher on myös kuuluisa! (Samalla sivulla mainitaan tärkeän instrumenttimuuttujaestimointitekniikan innovoinnissa mukana ollut Sewall Wright niinikään aivan toisen tieteenalan harjoittajana.)

Aputuloksia:

- Eksponenttijakautuneen ($\mathsf{Exp}(1/\mu)$) satunnaismuuttujan Y tiheysfunktio on $(1/\mu)\exp(-y/\mu)$ (monisteen s. 3).
- Riippumattomien eksponenttijakautuneiden ($\mathsf{Exp}(1/\mu)$) havaintojen tilanteessa SU-estimaatttori μ:lle on $\bar{y}$, jossa $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$ (monisteen s. 16).
- Oheiset jakaumataulukot.