Ei-parametriset ja robustit menetelmät, kevät 2015

Harjoitus 2

1. a) Johda kaava mediaanin asymptoottiselle suhteelliselle tehokkuudelle keskiarvoon verrattuna, kun havainnot ovat t-jakaumasta ν vapausasteella. Laske saadun kaavan avulla asymptoottisia suhteellisia tehokkuuksia eri vapausasteilla. Millä vapausasteiden arvoilla mediaani on tehokkaampi kuin keskiarvo.

   b) (Vapaaehtoinen tehtävä). Johda kaava HL-estimaatin asymptoottiselle suhteelliselle tehokkuudelle keskiarvoon verrattuna, kun havainnot ovat t-jakaumasta ν vapausasteella. Laske saadun kaavan avulla asymptoottisia suhteellisia tehokkuuksia eri vapausasteilla. Millä vapausasteiden arvoilla HL-estimaatti on tehokkaampi kuin keskiarvo.

   Avuksi: t_ν-jakauman tiheysfunktio on

   $$f_\nu(x) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\nu\pi}\Gamma(\nu/2)} \frac{1}{(1+x^2/\nu)^{(\nu+1)/2}}, \quad x \in \mathbb{R}, \quad \nu \geq 3.$$

   Jakauman varianssi on $\nu/(\nu-2)$. Rajalla $\nu \to \infty$ löytyy normaalijakauma. Funktio $\Gamma(x)$ on gamma-funktio, jolla on mm. seuraavat ominaisuudet:

   - $\Gamma(x+1) = x\Gamma(x)$.
   - $\Gamma(n+1) = n!$, kun n on ei-negatiivinen kokonaisluku.
   - $\Gamma(1/2) = \sqrt{\pi}$.

   R-funktioita ja vakioita

   Gamma-funktion arvo kohdassa x eli $\Gamma(x)$

   ```
   > gamma(x)
   ```

   Potenssiin korotus 6^3

   ```
   > 6^3
   ```

   Vakio π

   ```
   > pi
   ```

   Neliöjuuri luvusta x eli $\sqrt{x}$

   ```
   > sqrt(x)
   ```

2. Tarkastellaan niin sanottua **Tukeyn mallia** (Gross error model) $T(\epsilon, \tau)$, jolloin kertymäfunktio on

$$F(x) = (1-\epsilon)\Phi(x) + \epsilon\Phi\left(\frac{x}{\tau}\right), \ x \in \mathbb{R},$$

jossa $\Phi(x)$ on $N(0,1)$-jakauman kertymäfunktio. Etsi mediaanin suhteellinen tehokkuus keskiarvoon verrattuna, kun havainnot ovat peräisin Tukeyn mallista, kun $\tau = 2, 4, 10$ ja $\epsilon = 0.01,\ 0.05,\ 0.10$.

Avuksi: Odotusarvon ja varianssin laskemiseen voi käyttää seuraavaa tulosta. Olkoon $u \sim Bin(1, \epsilon)$ ja $z \sim N(0,1)$ riippumattomia satunnaismuuttujia. Silloin

$$x = (1-u)z + u\tau z \sim T(\epsilon, \tau).$$

3. Olkoon $x_1, ..., x_n$, $n = 2k+1$, satunnaisotos jatkuvasta jakaumasta, jonka kertymäfunktio on $F(x)$ ja tiheysfunktio on $f(x)$. Voidaan osoittaa, että mediaanin

$$M \ = \ Med\{x_1, ..., x_n\}$$

kertymäfunktio ja tiheysfunktio ovat

$$F_M(m) = \sum_{i=k+1}^{2k+1} \left[\binom{n}{i}(F(m))^i(1-F(m))^{n-i}\right]$$

ja

$$f_M(m) = \frac{n!}{k!k!} f(m)(F(m))^k(1-F(m))^k.$$

Vertaa graafisesti (samassa kuvassa) otoskeskiarvon ja mediaanin tiheysfunktioita, kun $x_1, ..., x_{2k+1}$ on satunnaisotos $N(0,1)$-jakaumasta (otoskeskiarvon jakaumahan on tunnetusti $N(\mu, \sigma^2/n)$).

R-komentoja

Potenssiin korotus 6^3

```
> k<-3
```

```
> 6^k
```

Kertoma $n!$

```
> factorial(n)
```

Vektori $x = (-2.00, -1.95, \ldots, 1.95, 2.00)$

```
> x <- seq(-2,2,0.05)
```

$N(3, 5^2)$-jakauman tiheysfunktion arvot kohdissa x[1],x[2],...

```
> dnorm(x, mean = 3, sd = 5)
```

$N(3, 5^2)$-jakauman kertymäfunktion arvot kohdissa x[1],x[2],...

```
> pnorm(x, mean = 3, sd = 5)
```

Tehdään hajontakuvio, jossa on pisteet (x[1],y[1)), (x[2],y[2]),... ja yhdistetään ne viivoilla (type=”l”).

```
> plot(x,y,type="l")
```

Lisätään kuvaan pisteet (x[1],y[1)), (x[2],y[2]),... ja yhdistetään ne katkoviivoilla (type=”l”,lty=2).

```
> points(x,y,type="l",lty=2)
```