Tilastollinen päättely I   Harj. 4   vko 6 2012

(4.1.) Otos $X_1, ..., X_{18}$ jak. $f(x) = 1 - \frac{x}{2}$, $0 \le x \le 2$

a) $\mu = \int_0^2 x(1-\frac{x}{2})dx = \Big/_0^2 (\frac{x^2}{2} - \frac{x^3}{6}) = 2 - \frac{4}{3} = \frac{2}{3}$

$\sigma^2 = \int_0^2 x^2(1-\frac{x}{2})dx - (\frac{2}{3})^2 = \Big/_0^2 (\frac{x^3}{3} - \frac{x^4}{8}) - (\frac{2}{3})^2 = (\frac{8}{3} - \frac{16}{8}) - \frac{4}{9} = \frac{2}{9}$

b) $P(\frac{2}{3} \le \bar{X} \le \frac{5}{6}) = P(\frac{\frac{2}{3}-\frac{2}{3}}{\sqrt{2/9/18}} \le Z \le \frac{\frac{5}{6}-\frac{2}{3}}{\sqrt{2/9/18}}) = P(0 \le Z \le 1.5)$

$= \Phi(1.5) - \Phi(0) \approx 0.4332$ R:llä: pnorm(1.5) - pnorm(0) = 0,4332

tässä siis $Z = \frac{\bar{x} - \frac{2}{3}}{\sqrt{2/9/18}} \overset{\text{likim.}}{\sim} N(0,1)$

(4.2.) X:n kertymäfunktio $F(x) = x - \frac{x^2}{4}$

Silloin $Y = F(X) \sim Tas(0,1)$

F on monotoninen ja sillä on käänteiskuvaus F^{-1}

Nyt $X = F^{-1}(y)$ noudattaa jakaumaa F

$F(x) = x - \frac{x^2}{4}$, $0 \le x < 2$

$F(x) = y = x - \frac{x^2}{4}$ ratkaistaan x

$-\frac{1}{4}x^2 + x - y = 0$

$x = \frac{-1 \pm \sqrt{1-y}}{-\frac{1}{2}} = 2 \pm 2\sqrt{1-y}$

koska $0 \le x < 2$

$x = 2 - \sqrt{4-4y}$

Generoidaan Y:n arvot Tas(0,1):stä ja ratk. x   R:llä:

```
x <- 2 - sqrt(4 - 4*runif(1000))
hist(x)
```

(4.3.) Tšebyševin epäyhtälön nojalla (9.6.1) $x \sim Ber(\pi)$, $E(x) = \pi$, $Var(x) = \pi(1-\pi)$

$P(|\hat{\pi} - \pi| \ge 0.03) \le \frac{\pi(1-\pi)}{0.03^2 \cdot 3000}$

$P(|\hat{\pi} - \pi| < 0.03) \ge 1 - \frac{\pi(1-\pi)}{0.03^2 \cdot 3000}$

koska $\pi(1-\pi) \le \frac{1}{2} \cdot \frac{1}{2}$ $0 \le \pi \le 1$

niin $-\pi(1-\pi) \ge -\frac{1}{4}$ ja

$1 - \frac{\pi(1-\pi)}{0.03^2 \cdot 3000} \ge 1 - \frac{1}{4 \cdot 0.03^2 \cdot 3000} = 0.907$

siis

$P(|\hat{\pi} - \pi| < 0.03) \ge 0.907 > 0.90$

(4.4.) a) $\hat{Var}(\hat{\pi}) = \frac{\hat{\pi}(1-\hat{\pi})}{n} = \frac{\hat{\pi}(1-\hat{\pi})}{3000}$ $\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{3000}}$

$$P(|\hat{\pi}-\pi| < 0.03) = P\left(\frac{|\hat{\pi}-\pi|}{\hat{\sigma}_{\hat{\pi}}} < \frac{0.03}{\hat{\sigma}_{\hat{\pi}}}\right) \approx \Phi\left(\frac{0.03}{\hat{\sigma}_{\hat{\pi}}}\right) - \Phi\left(-\frac{0.03}{\hat{\sigma}_{\hat{\pi}}}\right)$$

b) Generointi R:llä

```
sum(as.numeric(abs(rbinom(1000,3000,0.2)/3000 - 0.2) > 0.03))
```

(4.5.) Olkoon $S_n = X_1 + \cdots + X_n$, missä $X_i = 1$, jos haastateltava i kannattaa A:ta, muutoin $X_i = 0$. Ehdokkaan A kannattajien lkm otoksessa on siis S_n, joka noudattaa likimain binomijakaumaa $\mathrm{Bin}(n, p)$, koska populaatio suuri. (Kun otanta on palauttamatta, S_n:n tarkka jakauma on hypergeometrinen jakauma. Kun n on pieni suhteessa populaation kokoon, binomijakauma on hypergeometrisen hyvä likiarvo. Hypergeometrista tässä ei voi käyttää myös siksi, että emme tiedä populaation kokoa.) Keskeisen rajaväittämän nojalla $(S_n - np)/\sqrt{npq}$, $q = 1 - p$, noudattaa likimain normaalijakaumaa $\mathrm{N}(0, 1)$. Silloin

$$\begin{aligned} P(|\frac{S_n}{n} - p| < \epsilon) &= P(|\frac{S_n - np}{\sqrt{npq}}| < \frac{n\epsilon}{\sqrt{npq}}) \\ &\approx \Phi(\frac{n\epsilon}{\sqrt{npq}}) - \Phi(-\frac{n\epsilon}{\sqrt{npq}}) \\ &= 2\Phi(\frac{n\epsilon}{\sqrt{npq}}) - 1 \geq 0.95 \end{aligned}$$

(a) Silloin

$$\frac{\sqrt{n}\epsilon}{\sqrt{pq}} \geq 1.96 \iff n \geq \frac{1.96^2 pq}{\epsilon^2}.$$

Koska p on tuntematon, valitsemalla

$$n \geq \frac{1.96^2 0.5^2}{\epsilon^2} \geq \frac{1.96^2 pq}{\epsilon^2}$$

varmistetaan, että vaadittu epäyhtälö toteutuu.

(b)

$$n \geq \frac{1.96^2 0.5^2}{0.02^2} = 2401.$$

(c) $\frac{\sqrt{n}\varepsilon}{\sqrt{pq}} \geq 1.96 \iff \varepsilon \geq \frac{1.96\sqrt{pq}}{\sqrt{n}}$ $n \leq 1000$

$\varepsilon \geq \frac{1.96\sqrt{pq}}{10\sqrt{10}}$, koska $\sqrt{pq} \leq 0.5$ niin ainakin tarkkuuteen

$\varepsilon = \frac{1.96 \cdot 0.5}{10\sqrt{10}}$ voidaan päästä.

(4.6.) $X \sim \mathrm{Bin}(100, 0.1)$ $E(X) = 100 \cdot 0.1 = 10$ $Var(X) = 100 \cdot 0.1 \cdot 0.9 = 9$

a) $P(12 \leq X \leq 14)$ = pbinom(14, 100, 0.1) - pbinom(11, 100, 0.1) = 0.2244

b) $X \overset{\text{likim.}}{\sim} \mathrm{Poi}(10)$ $[\lambda = np = 100 \cdot 0.1]$

$P(12 \leq X \leq 14)$ = ppois(14,10) - ppois(11,10) = 0.2198

c) $P(12 \leq X \leq 14) \overset{\text{jatkuvuuskorj.}}{=} P(11{,}5 < X < 14.5) \approx P\left(\frac{11.5-10}{3} < Z < \frac{14.5-10}{3}\right)$

$\approx \Phi(1.5) - \Phi(0.5) = 0.2417$

4.7. $Y_i \sim Exp(100)$ Kestoaika $X = \max(Y_1, Y_2)$

X:n tiheysfunktio eli maksimin tiheysfunktio Luku 9.5.1

$f_{(n)}(y) = n[F(y)]^{n-1} f(y)$, missä $n=2$

$$f(y) = \begin{cases} \frac{1}{100} e^{-\frac{y}{100}}, & y>0 \\ 0, & \text{muualla} \end{cases} \qquad F(y) = 1 - e^{-\frac{y}{100}}$$

Nyt siis

$$f_X(y) = \begin{cases} 2 \cdot [1 - e^{-\frac{y}{100}}]^{2-1} \cdot \frac{1}{100} e^{-\frac{y}{100}}, & \text{kun } y>0 \\ 0, & \text{muualla} \end{cases}$$

$$f_X(y) = \begin{cases} \frac{1}{50} [e^{-\frac{y}{100}} - e^{-\frac{y}{50}}], & \text{kun } y>0 \\ 0, & \text{muualla} \end{cases}$$

4.8. Olk. $\bar{Y}$ otoskeskiarvo, kun $n=100$

KRV

$$P(\bar{Y} \leq 57.9) = P\left(\frac{\bar{Y} - 60}{\sqrt{64}/\sqrt{100}} \leq \frac{57.9 - 60}{\sqrt{64}/\sqrt{100}}\right)$$

$$\approx \Phi\left(\frac{57.9-60}{8/10}\right) = \Phi(-2.625) = 0.0043$$

Koska todennäköisyys sille, että pistemäärä otoksesta korkeintaan 57.9 on pieni syntyy epäilys, että koulun taso on keskitasoa alempi.