Sovellettu todennäköisyyslaskenta B

Antti Rasila

3. lokakuuta 2007

1. Jatkuvia jakaumia (jatkoa)
   - Kertausta: tiheysfunktio, odotusarvo
   - Eulerin Gamma-funktio
   - Gamma-jakauma

2. Keskeinen raja-arvolause
   - Binomijakauman normaalijakauma-approksimaatio
   - Poisson-jakauman normaaliapproksimaatio

3. Normaalijakaumasta johdettuja jakaumia
   - χ^2-jakauma
   - Fisherin F-jakauma
   - Studentin t-jakauma

Esimerkki (Laininen) 1/2

- Oletetaan, että lentokentälle saapuvan matkustajan odotusaika T (minuuttia) noudattaa jakaumaa, jonka tiheysfunktio on muotoa

$$f_T(t) = \begin{cases} a + bt, & \text{kun } 30 \le t \le 90, \\ 0 & \text{muuten,} \end{cases}$$

  missä $a > 0$ ja $b < 0$ ovat vakioita ja $a + 90b = 0$.

- Vakoiden arvot saadaa ehdosta

$$\int_{30}^{90} (a + bt)\, dt = 1.$$

- Ratkaistaan $a = 1/20$ ja $b = -1/1800$, joten tiheysfunktio

$$f_T(t) = \begin{cases} 1/20 - t/1800, & \text{kun } 30 \le t \le 90, \\ 0 & \text{muuten.} \end{cases}$$

Esimerkki (Laininen) 2/2

- Integroimalla saadaan kertymäfunktio

$$F_T(t) = \begin{cases} 0, & \text{kun } t < 30, \\ t/20 - t^2/3600 - 5/4, & \text{kun } 30 \le t \le 90, \\ 1, & \text{kun } t > 90. \end{cases}$$

- Todennäköisyys sille, että odotusaika on vähintään 40 mutta enintään 60 minuuttia on

$$Pr(40 \le T \le 60) = F_T(60) - F_T(40) = 0.444$$

- Matkustaja joutuu odottamaan keskimäärin

$$\mathrm{E}(T) = \int_{-\infty}^{\infty} x f_T(x)\, dt = \int_{30}^{90} t(1/20) - t/1800)\, dt = 50$$

minuuttia.

Eulerin Gamma-funktio

- Eulerin Gamma-funktio määritellään

$$\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z}\, dz, \qquad \alpha > 0.$$

- Gamma-funktio on kertomafunktion laajennus reaaliakselille (oikeastaan kompleksitasoon) lukuunottamatta nollaa ja negatiivisia kokonaislukuja.
- Erityisesti, jos n on positiivinen kokonaisluku, niin $\Gamma(n) = (n-1)!$.
- Gamma-funktiolle pätee myös $\Gamma(z+1) = z\Gamma(z)$.

Gamma-jakauma

- Merkitään $X \sim \mathrm{Gamma}(\alpha, \beta)$.
- Gammajakaumaa käytetään luotettavuustekniikassa komponentin eliniän jakautumismallina. Muita sovelluksia ovat mm. jonomallit ja sademäärän arviointi (esim. kuukaudessa).
- Tiheysfunktio

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \; x > 0, \; \alpha, \beta > 0.$$

- Odotusarvo: $\mathrm{E}(X) = \alpha\beta$, varianssi $\mathrm{D}^2(X) = \alpha\beta^2$.
- Erikoistapauksia: Eksponenttijakauma, kun $\alpha = 1$, kun $\beta = 2$, saadaan $\chi^2 -$ *jakauma* (esitellään myöhemmin).

Keskeinen raja-arvolause

- Olkoon $X_i, i = 1, 2, 3, \dots, n$ riippumattomia, samoin jakautuneita satunnaismuuttujia, joiden odotusarvo ja varianssi ovat $\mathrm{E}(X_i) = \mu$ ja $\mathrm{D}^2(X_i) = \sigma^2$.
- Tällöin satunnaismuuttujien $X_i, i = 1, 2, 3, \dots, n$ summan $Y_n = \sum_{i=1}^{n} X_i$ odotusarvo ja varianssi ovat $\mathrm{E}(Y_n) = n\mu$ ja $\mathrm{D}^2(Y_n) = n\sigma^2$.
- Standardoidaan Y_n:

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}}$$

Lause (Keskeinen raja-arvolause)

Satunnaismuuttujan Z_n jakauma lähestyy standardoitua normaalijakaumaa $\mathrm{N}(0, 1)$, kun $n \to +\infty$.

Keskeinen raja-arvolause, huomautuksia

- Keskeisen raja-arvolauseen mukaan

$$\lim_{n\to+\infty} \Pr\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le z\right) = \Phi(z)$$

- Käytännössä siis n:n suurille, mutta äärellisille arvoille:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim_a \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

- Erityisesti tämä pätee binomi- ja Poisson-jakaumille.

Esimerkki (Niemi)

- Kirjassa on 300 sivua ja keskimäärin yksi painovirhe viittä sivua kohden. Millä todennäköisyydella kirjassa on yli 50 painovirhettä?
- Merkitään X_k:lla sivulla k olevien painovirheiden lukumäärää, ja

$$X = \sum_{k=1}^{300} X_k.$$

- Koska $X_k \sim \mathrm{Poisson}(0.2)$ kaikilla k, saadaan keskeisen raja-arvolauseen nojalla $X \sim_a N(60, 60)$, koska

$$\mathrm{E}(X_k) = D^2(X_k) = 0.2.$$

- Siis

$$Pr(X \geq 50) = Pr\Big(\frac{X - 60}{\sqrt{60}} \geq -1.291\Big) \approx 90.2\%.$$

Esimerkki (Laininen) 1/2

- Asuntoalueelle suunnitellaan rakennettavaksi 1000 huoneistoa. Kuinka monta parkkipaikkaa on varattava, jotta ne 90% varmuudella riittäisivät kaikille asukkaille?
- Oletetaan, että kutakin huoneistoa kohden asukkaista 30% ei omista autoa lainkaan, 60% omistaa yhden auton ja 10% kaksi autoa.
- Huoneistoa kohden omistettujen autojen lukumäärä X on diskreetti satunnaismuuttuja.
- Saadaan odotusarvo

$$\mathrm{E}(X) = 0 \cdot 0.30 + 1 \cdot 0.60 + 2 \cdot 0.10 = 0.8,$$

ja vastaavasti varianssi $\mathrm{D}^2(X) = 0.36$.

- Alueelle tulevien autojen kokonaismäärä Z on

$$Z = X_1 + \ldots + X_{1000},$$

missä satunnaismuuttujat X_k ovat riippumattomia ja jakautuneita samoin kuin X.

Esimerkki (Laininen) 2/2

- Saadaan

$$\mathrm{E}(Z) = 1000 \cdot 0.8, \qquad \mathrm{D}^2(Z) = 1000 \cdot 0.36 = 360.$$

- Keskeisen raja-arvolauseen nojalla

$$Z \sim_a N(800, 360).$$

- On siis löydettää pienin n, jolla

$$0.9 \leq Pr(Z \leq n) = \Phi\big((n - 800)/\sqrt{360}\big).$$

- Taulukosta saadaan $\Phi(1.282) = 0.90$, joten asetetaan $(n - 800)/\sqrt{360} = 1.282$, eli ratkaisu on $n = 825$.

Binomijakauman normaalijakauma-approksimaatio

- Olkoon nyt $X \sim \mathrm{Bin}(n, p)$.
- Standardoidaan X kuten Y_n edellä:

$$Z = \frac{X - np}{\sqrt{npq}} \sim_a \mathrm{N}(0, 1),$$

  missä $q = 1 - p$.
- Tällöin voidaan suurille n:n arvoille käyttää approksimaatiota:

$$\Pr(a < X \leq b) \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right)$$

- Käytännössä parempi approksimaatio saadaan, jos tehdään jatkuvuuskorjaus:

$$\Pr(a < X \leq b) \approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{npq}}\right)$$

Esimerkki (Laininen)

- Heitetään virheetöntä noppaa 1002 kertaa. Kuinka suurella todennäköisyysdellä saadaan vähintään 150 mutta enintään 190 kertaa silmäluku 6?
- Olkoon X saatujen kuutosten lukumäärä. Nyt $X \sim \mathrm{Bin}(1002, 1/6)$.
- Saadaan odotusarvo $\mathrm{E}(X) = 1002 \cdot (1/6)$ ja ja varianssi $\mathrm{D}^2(X) = 1002 \cdot (1/6) \cdot (5/6) = 139.167$.
- Normaalijakauma-approksimaatiosta saadaan

$$
\begin{aligned}
Pr(150 \leq X \leq 190) &= \Phi\left(\frac{190.5 - 167}{\sqrt{139.167}}\right) - \Phi\left(\frac{149.5 - 167}{\sqrt{139.167}}\right) \\
&= \Phi(1.99) - \Phi(-1.48) \\
&= 0.977 - (1 - 0.931) = 0.908.
\end{aligned}
$$

Poisson-jakauman normaaliapproksimaatio

- Olkoon nyt $X \sim \mathrm{Poisson}(\lambda)$.
- Standardoidaan uusi X:

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \sim_a \mathrm{N}(0, 1)$$

- Tällöin voidaan suurille λ:n arvoille käyttää approksimaatiota:

$$\Pr(a < X \leq b) \approx \Phi\left(\frac{b - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - \lambda}{\sqrt{\lambda}}\right)$$

- Jälleen parempi approksimaatio saadaan jatkuvuuskorjauksen avulla:

$$\Pr(a < X \leq b) \approx \Phi\left(\frac{b + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right)$$

Normaalijakaumasta johdettuja jakaumia

- Monet tilastotieteessä keskeiset jakaumat määritellään normaalijakauman avulla.
- Esimerkkejä tällaisista jakaumista ovat χ^2, t ja F-jakaumat.
- Näillä jakaumilla on tärkeä rooli otosjakaumien teoriassa, estimoinnissa ja tilastollisessa testauksessa.

χ^2-jakauma

- Olkoot $Z_i \sim \mathrm{N}(0,1), i = 1, 2, \dots, n$ riippumattomia satunnaismuuttujia.
- $X = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$ eli X noudattaa khin neliön jakaumaa vapausasteilla n.
- Jakauma kuvaa siis normaalijakautuneiden satunnaismuuttujien neliöden summan jakaumaa.
- Tiheysfunktio:

$$f_X(x) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2}, \quad x > 0, \quad n > 0.$$

- Odotusarvo:

$$\mathrm{E}(X) = n$$

- Varianssi:

$$\mathrm{D}^2(X) = 2n$$

Fisherin F-jakauma

- Olkoot $X_i \sim \mathrm{N}(0,1), i = 1,2,\ldots,n$ ja $Y_i \sim \mathrm{N}(0,1), i = 1,2,\ldots,m$ riippumattomia satunnaismuuttujia.
- Olkoot $X = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$ ja $Y = \sum_{i=1}^{m} Y_i^2 \sim \chi^2(m)$ sekä

$$F = \frac{\frac{1}{m}Y}{\frac{1}{n}X}$$

  Tällöin $F \sim \mathrm{F}(m,n)$ eli F noudattaa Fisherin F-jakaumaa vapausasteilla m ja n.
- Jos $F \sim \mathrm{F}(m,n)$ niin silloin

$$\frac{1}{F} \sim \mathrm{F}(n,m)$$

- Odotusarvo:

$$\mathrm{E}(X) = \frac{n}{n-2}, n > 2$$

- Varianssi:

$$\mathrm{D}^2(X) = \frac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}, n > 4$$

Studentin t-jakauma

- Olkoot $X_i \sim \mathrm{N}(0,1), i = 1,2,\dots,n$ ja $Y \sim \mathrm{N}(0,1)$ riippumattomia satunnaismuuttujia.
- Olkoon $X = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$ sekä

$$T = \frac{Y}{\sqrt{\frac{1}{n}X}}$$

  Tällöin $T \sim \mathrm{t}(n)$
- Odotusarvo:

$$\mathrm{E}(T) = 0, n > 1$$

- t-jakauma lähestyy $\mathrm{N}(0,1)$-jakaumaa vapausasteiden kasvaessa.