Sovellettu todennäköisyyslaskenta B

Antti Rasila

28. syyskuuta 2007

Negatiivinen binomijakauma

- $X \sim \mathrm{Negbin}(r, p)$
- Negatiivinen binomijakauma on geometrisen jakauman yleistys.
- Toistetaan koetta niin monta kertaa, kunnes tapahtuma A on sattunut r kertaa.
- Geometrinen jakauma on siis negatiivinen binomijakauma parametrilla $r = 1$.
- Pistetodennäköisyysfunktio:

$$Pr(X = k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r, \qquad k = r, r+1, r+2, \ldots$$

- Odotusarvo ja varianssi:

$$\mathrm{E}(X) = \frac{r}{p}, \qquad \mathrm{D}^2(X) = \frac{r(1-p)}{p^2}.$$

Poisson-jakauma

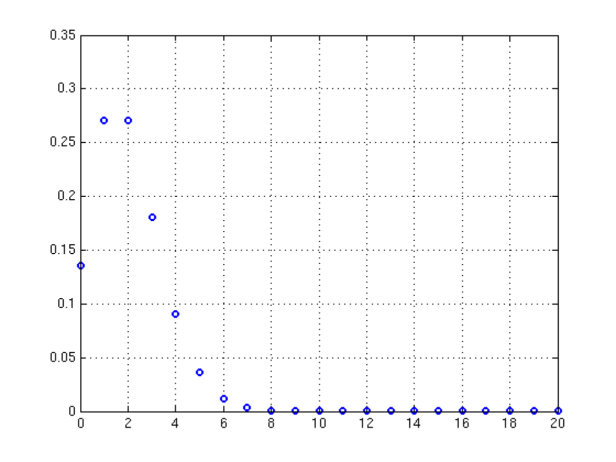
- $X \sim \mathrm{Poisson}(\lambda)$
- Kun tapahtuma A esiintyy keskimäärin λ kertaa aikayksikössä, on A:n esiintymiskertojen lukumäärä aikayksikössä Poisson-jakautunut parametrina λ. Esimerkiksi jonoon saapuvien ruokailijoiden määrä puolen tunnin aikana.
- Pistetodennäköisyysfunktio:

$$f(x) = Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots .$$

- Odotusarvo ja varianssi:

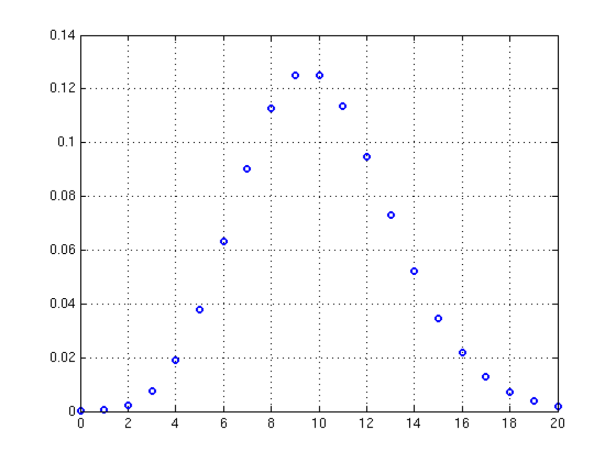
$$\mathrm{E}(X) = \mathrm{D}^2(X) = \lambda.$$

Poisson-jakauma



Poisson(2)

Poisson-jakauma



$\text{Poisson}(10)$

Esimerkki (Laininen) 1/2

- ENIAC tietokoneen (1946) rakentamisessa oli käytetty n. 18 000 radioputkea. Oletetaan, että yksi radioputki rikkoutui keskimäärin joka kymmenes tunti.
- Olkoon satunnaismuuttuja X kahdessakymmenessä tunnissa rikkoutuneiden radioputkien määrä. Kuinka X jakautuu?
- Vikaantumisten lukumäärän mallintamisessa voidaan käyttää Poissonin jakaumaa.
- Kymmenessä tunnissa palaneiden putkien lukumäärän malliksi sopii $\mathrm{Poisson}(\lambda)$, missä $\lambda = 1$ on rikkoutuneiden putkien määrän odotusarvo kymmenessä tunnissa.
- Kahdessakymmenessä tunnissa palaa keskimäärin kaksi putkea, joten X:n jakauma on $\mathrm{Poisson}(\lambda)$, missä $\lambda = \mathrm{E}(X) = 2$.

Esimerkki (Laininen) 1/2

- Jos kone menee toimintakyvyttömäksi kolmen tai useamman putken rikkouduttua, kuinka todennäköistä on, että 20 tuntia kestävä laskutoimitus saadaan laskettua loppuun?
- Laskutoimitus saadaan lasketuksi loppuun, jos enintaan 2 putkea on palanut. Siis todennäköisyys on

$$Pr(X \leq 2) = \sum_{x=0}^{2} \frac{2^x}{x!} e^{-2} \approx 0.677.$$

- Numeerinen arvo saadaan taulukosta tai tietokoneella.

Esimerkki (Milton-Arnold)

- Terveellä ihmisellä on 6000 valkosolua kuutiomillimetrissä verta. Valkosolujen puuttumisen havaitsemiseksi tehdään mittaus verikokeesta, jossa on 0.001 kuutiomillimetriä verta. Mikä on todennäköisyys sille, että terveen ihmisen tapauksessa kokeessa löytyy enintään 2 valkosolua?
- Valkosolujen esiintymiskertojen määrää voidaan tarkastella Poisson-jakautuneena satunnaismuuttujana. Tässä "aikayksikkö" on pisara verta.
- Koska yhdessä kuutiomillimetrissä esiintyy keskimäärin 6000 valkosolua, saadaan parametriksi $\lambda = 6000 \cdot 0.001 = 6$.
- Todennäköisyys $Pr(X \leq 2)$ on siis

$$\sum_{x=0}^{2} f(x) = \sum_{0}^{2} \frac{e^{-6}6^{x}}{x!} \approx 0.062$$

- Numeerinen arvo saadaan jälleen taulukosta tai tietokoneella.

Diskreettien jakaumien yhteenveto

Jakauma	Toistot	Otantamenetelmä	$n(\text{suotuisat})$
Bernoulli	1 (tiedossa etukäteen)		x (0 tai 1)
Binomi	n (tiedossa etukäteen)	yksi kerrallaan takaisinpanolla	x
Geom.	1 - ∞ (ei tiedossa etukäteen)	yksi kerrallaan	1 (tiedossa etukäteen)
Neg.Bin.	r - ∞ (ei tiedossa etukäteen)	yksi kerrallaan	r (tiedossa etukäteen)
Hyp.Geom.	n (tiedossa etukäteen)	kertaotos ilman takaisin panoa	x

Jatkuva tasainen jakauma

- $X \sim \mathrm{Uniform}(a, b)$ tai $X \sim \mathrm{Tas}(a, b)$
- Jakauma kuvaa tietyllä välillä tasaisesti jakautunutta jatkuvaa suuretta. Esimerkiksi onnenpyörän viisarin kulma lähtötilanteen suhteen.
- Tiheysfunktio:
$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$
- Odotusarvo:
$$\mathrm{E}(X) = \frac{a+b}{2}$$
- Varianssi:
$$\mathrm{D}^2(X) = \frac{(b-a)^2}{12}$$

Eksponenttijakauma

- $X \sim \mathrm{Exp}(\lambda)$
- Tapahtuma A tapahtuu keskimäärin λ kertaa tunnissa. Hetki jolloin tapahtuma A tapahtuu seuraavan kerran on eksponenttijakautunut parametrina λ. Esimerkiksi hetki, jolloin seuraava asiakas saapuu ruokalan jonoon.
- Tiheysfunktio:
$$f(x) = \lambda e^{-\lambda x}$$
- Odotusarvo:
$$\mathrm{E}(X) = \frac{1}{\lambda}$$
- Varianssi:
$$\mathrm{D}^2(X) = \frac{1}{\lambda^2}$$

Normaalijakauma

- Normaalijakauma esiintyy monien ilmiöiden yhteydessä luonnossa ja tekniikassa. Esimerkiksi suomalaisten jalan kokoa ja fysikaalisen mittauksen mittausvirhettä voidaan pitää normaalijakautuneina.
- Tiheysfunktio:

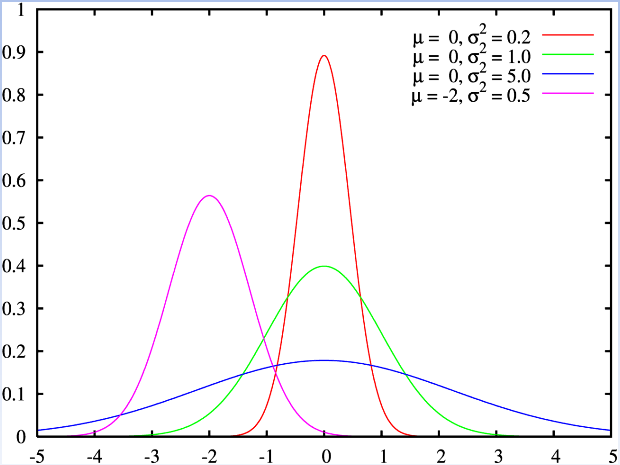
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Odotusarvo ja varianssi:

$$\mathrm{E}(X) = \mu \qquad \mathrm{D}^2(X) = \sigma^2$$

- Saadaanko normaalijakaumalle kertymäfunktio?

Normaalijakauman tiheysfunktio



Normaalijakauman standardointi 1/2

- Olkoon $X \sim \mathrm{N}(\mu, \sigma^2)$. Tällöin *standardoitu muuttuja*

$$Z = \frac{X - \mu}{\sigma} \sim \mathrm{N}(0,1)$$

- Operaatiota kutsutaan *standardoinniksi* ja jakaumaa $\mathrm{N}(0,1)$ *standardoiduksi normaalijakaumaksi*.
- Z:n kertymäfunktio $\Phi(z)$ saadaan integraalina

$$\Phi(z) = Pr(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}u^2}\, du.$$

- Kertymäfunktion arvoja voidaan laskea tietokoneella ja niitä löytyy myös taulukkokirjoista.

Normaalijakauman standardointi 2/2

- Jakauman symmetrisyyden johdosta pätee

$$\Phi(z) = 1 - \Phi(-z).$$

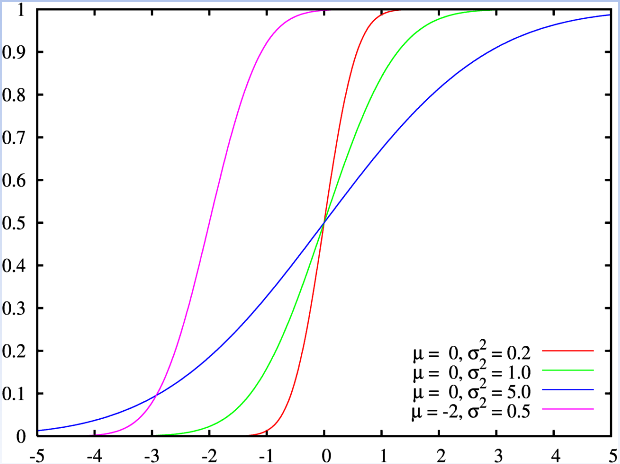
- Jos $X \sim N(\mu, \sigma^2)$, ja $a < b$, niin $Pr(a \leq X \leq b)$ saadaan standardoidun normaalijakauman avulla seuraavasti:

$$\begin{aligned} Pr(a \leq X \leq b) &= Pr(a - \mu \leq X - \mu \leq b - \mu) \\ &= Pr\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \end{aligned}$$

  eli

$$Pr(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Normaalijakauman kertymäfunktio (approksimaatio tietokoneella)



Esimerkki (68-95-99.7 -sääntö)

- Soveltamalla kaavaa nähdään, että

$$Pr(\mu - \sigma \leq X \leq \mu + \sigma) = Pr(-1 \leq Z \leq 1) = 0.6827,$$

$$Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = Pr(-2 \leq Z \leq 2) = 0.9545,$$

$$Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = Pr(-3 \leq Z \leq 3) = 0.9973.$$

- Siis normaalijakautuneen suureen X arvoista noin 68% on välillä $\mu \pm \sigma$, välillä $\mu \pm 2\sigma$ noin 95% ja noin 99.7% välillä $\mu \pm 3\sigma$.

Esimerkki (Milton-Arnold) 1/2

- Merkitään X:ällä säteilyannosta, joka ihmisen on saatava kuollakseen altistukseen.
- Oletetaan, että X on normaalijakautunut parametreilla $\mu = 500$ röntgeniä ja $\sigma = 150$ röntgeniä.
- Kuinka paljon säteilyä on tarvitaan, jotta vain 5% altistuneista jäisi henkiin?
- Tämän selvittämiseksi on löydettävä piste x_0, jolle

$$Pr(X \geq x_0) = 0.05.$$

- Standardoimalla saadaan

$$Pr(X \geq x_0) = Pr\left(\frac{X - 500}{150} \geq \frac{x_0 - 500}{150}\right) = Pr\left(Z \geq \frac{x_0 - 500}{150}\right) = 0.05.$$

Esimerkki (Milton-Arnold) 2/2

- Siis $(x_0 - 500)/150$ on piste, jossa standardoidun normaalijakauman tiheysfunktion alle jäävästä pinta-alasta 95% on pisteen vasemmalla ja 5 % oikealla puolella.
- Katsotaan taulukosta, mikä on z, kun $\Phi(z) = 0.95$.
- Numeeriseksi arvoksi saadaan noin 1.645 (valitaan arvo kahden taulukosta löytyvän arvon puolivälistä).
- Ratkaistaan lopuksi yhtälö

$$\frac{x_0 - 500}{150} = 1.645.$$

- Saadaan

$$x_0 = 150 \cdot 1.645 + 500 = 746.75 \text{ röntgeniä.}$$