

Insinöörimatematiikka: Todennäköisyyslaskenta

Mika Hirvensalo
mikhirve@utu.fi

Matematiikan ja tilastotieteen laitos
Turun yliopisto

2026

Satunnaismuuttujien yhteisjakauma

Yhteisjakauma

Olkoon $p(x, y)$ satunnaismuuttujien X ja Y yhteinen jakauma ja $f(X, Y)$ jokin satunnaismuuttujien X ja Y funktio. Tällöin

$$\mathbb{E}(f(X, Y)) = \sum_x \sum_y f(x, y)p(x, y).$$

Satunnaismuuttujien yhteisjakauma

Yhteisjakauma

Olkoon $p(x, y)$ satunnaismuuttujien X ja Y yhteinen jakauma ja $f(X, Y)$ jokin satunnaismuuttujien X ja Y funktio. Tällöin

$$\mathbb{E}(f(X, Y)) = \sum_x \sum_y f(x, y)p(x, y).$$

Riippumattomuus

Satunnaismuuttujat X ja Y ovat riippumattomat, jos $p(x, y) = p_x(x)p_y(y)$ ja lisäksi samoin jakautuneet, jos $p_x = p_y$.

Satunnaismuuttujien yhteisjakauma

Yhteisjakauma

Olkoon $p(x, y)$ satunnaismuuttujien X ja Y yhteinen jakauma ja $f(X, Y)$ jokin satunnaismuuttujien X ja Y funktio. Tällöin

$$\mathbb{E}(f(X, Y)) = \sum_x \sum_y f(x, y)p(x, y).$$

Riippumattomuus

Satunnaismuuttujat X ja Y ovat riippumattomat, jos $p(x, y) = p_x(x)p_y(y)$ ja lisäksi samoin jakautuneet, jos $p_x = p_y$. Jos satunnaismuuttujat ovat riippumattomat, on

- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$

Satunnaismuuttujien yhteisjakauma

Yhteisjakauma

Olkoon $p(x, y)$ satunnaismuuttujien X ja Y yhteinen jakauma ja $f(X, Y)$ jokin satunnaismuuttujien X ja Y funktio. Tällöin

$$\mathbb{E}(f(X, Y)) = \sum_x \sum_y f(x, y)p(x, y).$$

Riippumattomuus

Satunnaismuuttujat X ja Y ovat riippumattomat, jos $p(x, y) = p_x(x)p_y(y)$ ja lisäksi samoin jakautuneet, jos $p_x = p_y$. Jos satunnaismuuttujat ovat riippumattomat, on

- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$
- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

Satunnaismuuttujien yhteisjakauma

Yhteisjakauma

Olkoon $p(x, y)$ satunnaismuuttujien X ja Y yhteinen jakauma ja $f(X, Y)$ jokin satunnaismuuttujien X ja Y funktio. Tällöin

$$\mathbb{E}(f(X, Y)) = \sum_x \sum_y f(x, y)p(x, y).$$

Riippumattomuus

Satunnaismuuttujat X ja Y ovat riippumattomat, jos $p(x, y) = p_x(x)p_y(y)$ ja lisäksi samoin jakautuneet, jos $p_x = p_y$. Jos satunnaismuuttujat ovat riippumattomat, on

- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$
- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
- $\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y).$

Otoskeskiarvo, sen varianssi

Jos $X_1, \dots, X_n$ ovat samoin jakautuneita $((\mu, \sigma^2))$, riippumattomia satunnaismuuttujia ja

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$

näiden aritmeettinen keskiarvo (ns. otoskeskiarvo), on

- $\mathbb{E}(\bar{X}_n) = n \cdot \frac{1}{n}\mathbb{E}(X_1) = \mu.$
- $\text{Var}(\frac{1}{n}(X_1 + \dots + X_n)) = n \cdot \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}.$

Markovin epäyhtälö

Jos $\mathbb{P}(X < 0) = 0$ ja $a > 0$, on

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}.$$

Markovin epäyhtälö

Jos $\mathbb{P}(X < 0) = 0$ ja $a > 0$, on

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}.$$

Chebyschevin epäyhtälö

$$\mathbb{P}(|X - \mathbb{E}(X)| > a) \leq \frac{\text{Var}(X)}{a^2}$$

Lause (Suurten lukujen laki)

Jos $X_1, \dots, X_n$ ovat samoin jakautuneita $((\mu, \sigma^2))$, riippumattomia satunnaismuuttujia ja

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$

näiden aritmeettinen keskiarvo (ns. otoskeskiarvo), on jokaista $\epsilon > 0$ kohti

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

Lause (Suurten lukujen laki)

Jos $X_1, \dots, X_n$ ovat samoin jakautuneita $((\mu, \sigma^2))$, riippumattomia satunnaismuuttujia ja

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$

näiden aritmeettinen keskiarvo (ns. otoskeskiarvo), on jokaista $\epsilon > 0$ kohti

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Esimerkki

Heitetään noppaa n kertaa ja olkoon X_i heitolla i saatu silmäluku.

Esimerkki

Heitetään noppaa n kertaa ja olkoon X_i heitolla i saatu silmäluku. Merkitään

$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n).$$

Kuinka suuri pitää heittojen määrän n olla, että varmasti

$$\mathbb{P}(|\overline{X}_n - \mu| > 0,1) \leq 0,01?$$

Keskeinen raja-arvolause

Olkoot X_i ja $\bar{X}_n$ kuten edellisessä lauseessa. Suurilla n :n arvoilla $\bar{X}_n$ jakautuu likimain normaalijakauman $N(\mu, \frac{\sigma^2}{n})$ mukaisesti.

Keskeinen raja-arvolause

Olkoot X_i ja $\bar{X}_n$ kuten edellisessä lauseessa. Suurilla n :n arvoilla $\bar{X}_n$ jakautuu likimain normaalijakauman $N(\mu, \frac{\sigma^2}{n})$ mukaisesti.

Toisin sanoen,

$$\mathbb{P}(\bar{X}_n \leq x) \approx \Phi\left(\frac{x - \mu}{\sigma/\sqrt{n}}\right)$$

kun n on suuri.

Satunnaismuuttujien summat

Keskeinen raja-arvolause

Olkoot X_i ja $\bar{X}_n$ kuten edellisessä lauseessa. Suurilla n :n arvoilla $\bar{X}_n$ jakautuu likimain normaalijakauman $N(\mu, \frac{\sigma^2}{n})$ mukaisesti.

Toisin sanoen,

$$\mathbb{P}(\bar{X}_n \leq x) \approx \Phi\left(\frac{x - \mu}{\sigma/\sqrt{n}}\right)$$

kun n on suuri.

Huomautus

Tietyin ehdoin (Lindbergin ehto) väittämä pätee, vaikka satunnaismuuttujilla ei olisikaan samanlainen jakauma. Tulos voidaan muotoilla myös summan $S_n = X_1 + X_2 + \dots + X_n$ avulla seuraavasti:

$$\mathbb{P}(S_n \leq x) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

Esimerkki

Alueelle suunnitellaan asuntoja tuhannelle perheelle. Perheissä on 0, 1, 2 tai 3 lasta todennäköisyyksillä 0.35, 0.45, 0.15 ja 0.05. Mikä on todennäköisyys sille, että 1000 koulupaikkaa ei riitä alueen lapsille?

Esimerkki

Alueelle suunnitellaan asuntoja tuhannelle perheelle. Perheissä on 0, 1, 2 tai 3 lasta todennäköisyyksillä 0.35, 0.45, 0.15 ja 0.05. Mikä on todennäköisyys sille, että 1000 koulupaikkaa ei riitä alueen lapsille?
Olkoon X_i lapsiluku perheessä i .

Esimerkki

Alueelle suunnitellaan asuntoja tuhannelle perheelle. Perheissä on 0, 1, 2 tai 3 lasta todennäköisyyksillä 0.35, 0.45, 0.15 ja 0.05. Mikä on todennäköisyys sille, että 1000 koulupaikkaa ei riitä alueen lapsille?

Olkoon X_i lapsiluku perheessä i . Tällöin $\mathbb{E}(X_i) = 0.9$ (laske) ja $\text{Var}(X_i) = 0.69$ (laske).

Esimerkki

Alueelle suunnitellaan asuntoja tuhannelle perheelle. Perheissä on 0, 1, 2 tai 3 lasta todennäköisyyksillä 0.35, 0.45, 0.15 ja 0.05. Mikä on todennäköisyys sille, että 1000 koulupaikkaa ei riitä alueen lapsille?

Olkoon X_i lapsiluku perheessä i . Tällöin $\mathbb{E}(X_i) = 0.9$ (laske) ja $\text{Var}(X_i) = 0.69$ (laske). Jos merkitään $L = X_1 + \dots + X_{1000}$, on L (kokonaislapsiluku) satunnaismuuttuja, jonka odotusarvo on $0.9 \cdot 1000 = 900$ ja varianssi 690.

Esimerkki

Alueelle suunnitellaan asuntoja tuhannelle perheelle. Perheissä on 0, 1, 2 tai 3 lasta todennäköisyyksillä 0.35, 0.45, 0.15 ja 0.05. Mikä on todennäköisyys sille, että 1000 koulupaikkaa ei riitä alueen lapsille?

Olkoon X_i lapsiluku perheessä i . Tällöin $\mathbb{E}(X_i) = 0.9$ (laske) ja $\text{Var}(X_i) = 0.69$ (laske). Jos merkitään $L = X_1 + \dots + X_{1000}$, on L (kokonaislapsiluku) satunnaismuuttuja, jonka odotusarvo on $0.9 \cdot 1000 = 900$ ja varianssi 690. Tällöin $\sigma = \sqrt{690} \approx 26.3$.

Esimerkki

Alueelle suunnitellaan asuntoja tuhannelle perheelle. Perheissä on 0, 1, 2 tai 3 lasta todennäköisyyksillä 0.35, 0.45, 0.15 ja 0.05. Mikä on todennäköisyys sille, että 1000 koulupaikkaa ei riitä alueen lapsille?

Olkoon X_i lapsiluku perheessä i . Tällöin $\mathbb{E}(X_i) = 0.9$ (laske) ja $\text{Var}(X_i) = 0.69$ (laske). Jos merkitään $L = X_1 + \dots + X_{1000}$, on L (kokonaislapsiluku) satunnaismuuttuja, jonka odotusarvo on $0.9 \cdot 1000 = 900$ ja varianssi 690. Tällöin $\sigma = \sqrt{690} \approx 26.3$. Näin ollen

$$\begin{aligned}\mathbb{P}(L > 1000) &= 1 - \mathbb{P}(L \leq 1000) \approx 1 - \Phi\left(\frac{1000 - 900}{\sqrt{690}}\right) \\ &= 0,0000703499 \dots\end{aligned}$$

Esimerkki

Eräässä kaupungissa huomattiin että 344 kaikista 747:stä kuolemantapauksesta oli sattunut korkeintaan 3 kk syntymäpäivän jälkeen.

Esimerkki

Eräässä kaupungissa huomattiin että 344 kaikista 747:stä kuolemantapauksesta oli sattunut korkeintaan 3 kk syntymäpäivän jälkeen. Mikä on tällaisen todennäköisyys jos kuolemantapaukset jakautuvat tasaisesti?

Esimerkki

Eräässä kaupungissa huomattiin että 344 kaikista 747:stä kuolemantapauksesta oli sattunut korkeintaan 3 kk syntymäpäivän jälkeen. Mikä on tällaisen todennäköisyys jos kuolemantapaukset jakautuvat tasaisesti?

Jos kuolemat jakautuisivat tasaisesti, pitäisi näitä olla 3 kk syntymäpäivän jälkeen noin $747 \cdot \frac{1}{4} = 186,75$.

Esimerkki

Eräässä kaupungissa huomattiin että 344 kaikista 747:stä kuolemantapauksesta oli sattunut korkeintaan 3 kk syntymäpäivän jälkeen. Mikä on tällaisen todennäköisyys jos kuolemantapaukset jakautuvat tasaisesti?

Jos kuolemat jakautuisivat tasaisesti, pitäisi näitä olla 3 kk syntymäpäivän jälkeen noin $747 \cdot \frac{1}{4} = 186,75$. 344 on kuitenkin noin 1,8-kertainen tähän nähden.

Esimerkki

Eräässä kaupungissa huomattiin että 344 kaikista 747:stä kuolemantapauksesta oli sattunut korkeintaan 3 kk syntymäpäivän jälkeen. Mikä on tällaisen todennäköisyys jos kuolemantapaukset jakautuvat tasaisesti?

Jos kuolemat jakautuisivat tasaisesti, pitäisi näitä olla 3 kk syntymäpäivän jälkeen noin $747 \cdot \frac{1}{4} = 186,75$. 344 on kuitenkin noin 1,8-kertainen tähän nähden.

Määritetään todennäköisyys $\mathbb{P}(X \geq 344)$ sillä otaksumalla, että $p = 0,25$ on todennäköisyys kuolla minä hyvänsä 3 kuukauden ajanjaksona.

Binomijakauma

Binomijakauman perustella voidaan laskea

$$\mathbb{P}(X \geq 344) = \sum_{k=344}^{747} \binom{747}{k} 0,25^k \cdot 0,75^{747-k}.$$

Binomijakauma

Binomijakauman perustella voidaan laskea

$$\mathbb{P}(X \geq 344) = \sum_{k=344}^{747} \binom{747}{k} 0,25^k \cdot 0,75^{747-k}.$$

Suorassa laskussa voi tulla paljon pyöristysvirheitä.

Binomijakauma

Binomijakauman perustella voidaan laskea

$$\mathbb{P}(X \geq 344) = \sum_{k=344}^{747} \binom{747}{k} 0,25^k \cdot 0,75^{747-k}.$$

Suorassa laskussa voi tulla paljon pyöristysvirheitä. Varianssi
 $np(1 - p) = 747 \cdot 0.25 \cdot 0.75 = 140,06,$

Binomijakauma

Binomijakauman perustella voidaan laskea

$$\mathbb{P}(X \geq 344) = \sum_{k=344}^{747} \binom{747}{k} 0,25^k \cdot 0,75^{747-k}.$$

Suorassa laskussa voi tulla paljon pyöristysvirheitä. Varianssi $np(1-p) = 747 \cdot 0.25 \cdot 0.75 = 140,06$, joten keskeinen raja-arvause antaa

$$\begin{aligned}\mathbb{P}(X \geq 344) &= 1 - \mathbb{P}(0 \leq X \leq 343) \\ &\approx 1 - \left(\Phi\left(\frac{343,5 - 186,75}{\sqrt{140,06}}\right) - \Phi\left(\frac{-0,5 - 186,75}{\sqrt{140,06}}\right) \right) \\ &= 1,09491 \cdot 10^{-56} = 0,00000000 \dots 000109491 \dots\end{aligned}$$

Binomijakauma

Binomijakauman perustella voidaan laskea

$$\mathbb{P}(X \geq 344) = \sum_{k=344}^{747} \binom{747}{k} 0,25^k \cdot 0,75^{747-k}.$$

Suorassa laskussa voi tulla paljon pyöristysvirheitä. Varianssi $np(1-p) = 747 \cdot 0.25 \cdot 0.75 = 140,06$, joten keskeinen raja-arvolause antaa

$$\begin{aligned}\mathbb{P}(X \geq 344) &= 1 - \mathbb{P}(0 \leq X \leq 343) \\ &\approx 1 - \left(\Phi\left(\frac{343,5 - 186,75}{\sqrt{140,06}}\right) - \Phi\left(\frac{-0,5 - 186,75}{\sqrt{140,06}}\right) \right) \\ &= 1,09491 \cdot 10^{-56} = 0,00000000 \dots 000109491 \dots\end{aligned}$$

Vrt. Suoraan binomijakauman perusteella laskettu arvo on $1,19478 \cdot 10^{-35}$.

Otoskeskiarvo, sen varianssi

Jos $X_1, \dots, X_n$ ovat samoin jakautuneita $((\mu, \sigma^2))$, riippumattomia satunnaismuuttujia ja $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$ näiden aritmeettinen keskiarvo (ns. otoskeskiarvo) sekä summa $S_n = X_1 + \dots + X_n$, on

- $\mathbb{E}(\bar{X}_n) = n \cdot \frac{1}{n} \mathbb{E}(X_1) = \mu.$
- $\text{Var}(\frac{1}{n}(X_1 + \dots + X_n)) = n \cdot \frac{1}{n^2} \sigma^2 = \frac{\sigma^2}{n}.$

Otoskeskiarvo, sen varianssi

Jos $X_1, \dots, X_n$ ovat samoin jakautuneita $((\mu, \sigma^2))$, riippumattomia satunnaismuuttujia ja $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$ näiden aritmeettinen keskiarvo (ns. otoskeskiarvo) sekä summa $S_n = X_1 + \dots + X_n$, on

- $\mathbb{E}(\bar{X}_n) = n \cdot \frac{1}{n} \mathbb{E}(X_1) = \mu.$
- $\text{Var}(\frac{1}{n}(X_1 + \dots + X_n)) = n \cdot \frac{1}{n^2} \sigma^2 = \frac{\sigma^2}{n}.$
- $\mathbb{E}(S_n) = n \cdot \mathbb{E}(X_1) = n\mu.$
- $\text{Var}(S_n) = n\sigma^2$

Keskeinen raja-arvolause

Suurilla n :n arvoilla $\bar{X}_n$ jakautuu likimain normaalijakauman $N(\mu, \frac{\sigma^2}{n})$ mukaisesti.

Keskeinen raja-arvolause

Suurilla n :n arvoilla $\bar{X}_n$ jakautuu likimain normaalijakauman $N(\mu, \frac{\sigma^2}{n})$ mukaisesti. Toisin sanoen,

$$\mathbb{P}(\bar{X}_n \leq x) \approx \Phi\left(\frac{x - \mu}{\sigma/\sqrt{n}}\right)$$

kun n on suuri.

Keskeinen raja-arvolause

Suurilla n :n arvoilla $\bar{X}_n$ jakautuu likimain normaalijakauman $N(\mu, \frac{\sigma^2}{n})$ mukaisesti. Toisin sanoen,

$$\mathbb{P}(\bar{X}_n \leq x) \approx \Phi\left(\frac{x - \mu}{\sigma/\sqrt{n}}\right)$$

kun n on suuri.

Toinen formulointi

$$\mathbb{P}(S_n \leq x) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

Satunnaismuuttujien summat

Keskeinen raja-arvolause

Suurilla n :n arvoilla $\bar{X}_n$ jakautuu likimain normaalijakauman $N(\mu, \frac{\sigma^2}{n})$ mukaisesti. Toisin sanoen,

$$\mathbb{P}(\bar{X}_n \leq x) \approx \Phi\left(\frac{x - \mu}{\sigma/\sqrt{n}}\right)$$

kun n on suuri.

Toinen formulointi

$$\mathbb{P}(S_n \leq x) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

Huomautus

Jatkuvan jakauman diskretisointi

Lähtökohdat

- Havainnot $x_1, \dots, x_n$ jostakin ilmiöstä.

Lähtökohdat

- Havainnot $x_1, \dots, x_n$ jostakin ilmiöstä.
- Tyypillisesti ei tiedetä miten havainnoitava ilmiö on jakautunut.

Lähtökohdat

- Havainnot $x_1, \dots, x_n$ jostakin ilmiöstä.
- Tyypillisesti ei tiedetä miten havainnoitava ilmiö on jakautunut.
- i :s havainto x_i on satunnaismuuttujan X_i arvo.

Lähtökohdat

- Havainnot $x_1, \dots, x_n$ jostakin ilmiöstä.
- Tyypillisesti ei tiedetä miten havainnoitava ilmiö on jakautunut.
- i :s havainto x_i on satunnaismuuttujan X_i arvo.
- Satunnaismuuttujat X_i riippumattomat.

Lähtökohdat

- Havainnot $x_1, \dots, x_n$ jostakin ilmiöstä.
- Tyypillisesti ei tiedetä miten havainnoitava ilmiö on jakautunut.
- i :s havainto x_i on satunnaismuuttujan X_i arvo.
- Satunnaismuuttujat X_i riippumattomat.
- Satunnaisotos: $X_1, X_2, \dots, X_n$.

Lähtökohdat

- Havainnot $x_1, \dots, x_n$ jostakin ilmiöstä.
- Tyypillisesti ei tiedetä miten havainnoitava ilmiö on jakautunut.
- i :s havainto x_i on satunnaismuuttujan X_i arvo.
- Satunnaismuuttujat X_i riippumattomat.
- Satunnaisotos: $X_1, X_2, \dots, X_n$.
- Pyrkimys: Hyvät arviot ainakin odotusarvolle $\mathbb{E}(X_i) = \mu$ ja varianssille $\text{Var}(X_i) = \sigma^2$.

Otoskeskiarvo

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

missä $\mathbb{E}(X_i) = \mu$ ja $\text{Var}(X_i) = \sigma^2$.

Otoskeskiarvo

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

missä $\mathbb{E}(X_i) = \mu$ ja $\text{Var}(X_i) = \sigma^2$.

- $\mathbb{E}(\bar{X}_n) = \mu$

Otoskeskiarvo

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

missä $\mathbb{E}(X_i) = \mu$ ja $\text{Var}(X_i) = \sigma^2$.

- $\mathbb{E}(\bar{X}_n) = \mu$
- $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$.

Otoskeskiarvo

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

missä $\mathbb{E}(X_i) = \mu$ ja $\text{Var}(X_i) = \sigma^2$.

- $\mathbb{E}(\bar{X}_n) = \mu$
- $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$.

Otosvarianssi

$$\text{Var}(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad ?$$

Odotusarvo

Otoskeskiarvo

$$\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n).$$

Odotusarvo

Otoskeskiarvo

$$\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n).$$

Tällöin

$$\mathbb{E}(\overline{X}_n) = \frac{1}{n} \cdot (\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)) = \mu.$$

Odotusarvo

Otoskeskiarvo

$$\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n).$$

Tällöin

$$\mathbb{E}(\overline{X}_n) = \frac{1}{n} \cdot (\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)) = \mu.$$

Tätä ominaisuutta sanotaan *harhattomuudeksi*

Tarkentuvuus

$$\text{Var}(\overline{X}_n) = \text{Var}\left(\frac{1}{n}X_1\right) + \dots + \text{Var}\left(\frac{1}{n}X_n\right)$$

Tarkentuvuus

$$\begin{aligned}\text{Var}(\overline{X_n}) &= \text{Var}\left(\frac{1}{n}X_1\right) + \dots + \text{Var}\left(\frac{1}{n}X_n\right) \\ &= \frac{1}{n^2} \text{Var}(X_1) + \dots + \frac{1}{n^2} \text{Var}(X_n)\end{aligned}$$

Tarkentuvuus

$$\begin{aligned}\text{Var}(\overline{X_n}) &= \text{Var}\left(\frac{1}{n}X_1\right) + \dots + \text{Var}\left(\frac{1}{n}X_n\right) \\ &= \frac{1}{n^2} \text{Var}(X_1) + \dots + \frac{1}{n^2} \text{Var}(X_n) \\ &= n \cdot \frac{1}{n^2} \sigma^2 = \frac{\sigma^2}{n}\end{aligned}$$

Tarkentuvuus

$$\begin{aligned}\text{Var}(\overline{X}_n) &= \text{Var}\left(\frac{1}{n}X_1\right) + \dots + \text{Var}\left(\frac{1}{n}X_n\right) \\ &= \frac{1}{n^2} \text{Var}(X_1) + \dots + \frac{1}{n^2} \text{Var}(X_n) \\ &= n \cdot \frac{1}{n^2} \sigma^2 = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0.\end{aligned}$$

Tarkentuvuus

$$\begin{aligned}\text{Var}(\overline{X}_n) &= \text{Var}\left(\frac{1}{n}X_1\right) + \dots + \text{Var}\left(\frac{1}{n}X_n\right) \\ &= \frac{1}{n^2} \text{Var}(X_1) + \dots + \frac{1}{n^2} \text{Var}(X_n) \\ &= n \cdot \frac{1}{n^2} \sigma^2 = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0.\end{aligned}$$

Tällöin sanotaan, että $\overline{X}_n$ on *tarkentuva* estimaattori.

Varianssi

Koska

$$\sigma^2 = \text{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = \mathbb{E}(X_i^2) - \mu^2,$$

Varianssi

Koska

$$\sigma^2 = \text{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = \mathbb{E}(X_i^2) - \mu^2,$$

on

$$\mathbb{E}(X_i^2) = \sigma^2 + \mu^2$$

Varianssi

Koska

$$\sigma^2 = \text{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = \mathbb{E}(X_i^2) - \mu^2,$$

on

$$\mathbb{E}(X_i^2) = \sigma^2 + \mu^2$$

Samoin

$$\frac{\sigma^2}{n} = \text{Var}(\overline{X}_n) = \mathbb{E}(\overline{X}_n^2) - \mathbb{E}(\overline{X}_n)^2 = \mathbb{E}(\overline{X}_n^2) - \mu^2,$$

Varianssi

Koska

$$\sigma^2 = \text{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = \mathbb{E}(X_i^2) - \mu^2,$$

on

$$\mathbb{E}(X_i^2) = \sigma^2 + \mu^2$$

Samoin

$$\frac{\sigma^2}{n} = \text{Var}(\overline{X}_n) = \mathbb{E}(\overline{X}_n^2) - \mathbb{E}(\overline{X}_n)^2 = \mathbb{E}(\overline{X}_n^2) - \mu^2,$$

josta

$$\mathbb{E}(\overline{X}_n^2) = \frac{\sigma^2}{n} + \mu^2.$$

Varianssi

$$\mathbb{E}\left(\sum_{k=1}^n (X_k - \bar{X}_n)^2\right) = (n-1)\sigma^2.$$

Varianssi

$$\mathbb{E}\left(\sum_{k=1}^n (X_k - \bar{X}_n)^2\right) = (n-1)\sigma^2.$$

Tämän vuoksi määritellään *otosvarianssi*

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2,$$

Varianssi

$$\mathbb{E}\left(\sum_{k=1}^n (X_k - \bar{X}_n)^2\right) = (n-1)\sigma^2.$$

Tämän vuoksi määritellään *otosvarianssi*

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2,$$

jolloin

$$\mathbb{E}(S_n^2) = \sigma^2.$$

Varianssi

$$\mathbb{E}\left(\sum_{k=1}^n (X_k - \bar{X}_n)^2\right) = (n-1)\sigma^2.$$

Tämän vuoksi määritellään *otosvarianssi*

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2,$$

jolloin

$$\mathbb{E}(S_n^2) = \sigma^2.$$

S_n^2 on siis harhaton.

Varianssi

$$\mathbb{E}\left(\sum_{k=1}^n (X_k - \bar{X}_n)^2\right) = (n-1)\sigma^2.$$

Tämän vuoksi määritellään *otosvarianssi*

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2,$$

jolloin

$$\mathbb{E}(S_n^2) = \sigma^2.$$

S_n^2 on siis harhaton. Myös tarkentuva: $\text{Var}(S_n^2) = \frac{\mu_4}{n} + \frac{3-n}{n(n-1)}\sigma^4$.

Lause

Jos satunnaismuuttujat $X_1, \dots, X_n$ ovat jakautuneet $N(\mu, \sigma^2)$ mukaisesti, on

$$T = \frac{\overline{X}_n - \mu}{S_n / \sqrt{n}}$$

jakautunut jakauman $t(n-1)$ (Student's t -distribution) mukaisesti.

Lause

Jos satunnaismuuttujat $X_1, \dots, X_n$ ovat jakautuneet $N(\mu, \sigma^2)$ mukaisesti, on

$$T = \frac{\overline{X}_n - \mu}{S_n / \sqrt{n}}$$

jakautunut jakauman $t(n-1)$ (Student's t -distribution) mukaisesti.

Seuraus

Olkoon $\alpha \in (0, 1)$ ja a sellainen, että $\mathbb{P}(-a < T < a) = 1 - \alpha$.
Tällöin

$$\mathbb{P}\left(\overline{X}_n - a \frac{S_n}{\sqrt{n}} \leq \mu \leq \overline{X}_n + a \frac{S_n}{\sqrt{n}}\right) = 1 - \alpha.$$

Jos esim. $\alpha = 0,05$, on kyseessä 95% *luottamusväli*.

Lause

Jos satunnaismuuttujat $X_1, \dots, X_n$ ovat jakautuneet $N(\mu, \sigma^2)$ mukaisesti, on

$$T = \frac{\overline{X}_n - \mu}{S_n / \sqrt{n}}$$

jakautunut jakauman $t(n-1)$ (Student's t -distribution) mukaisesti.

Seuraus

Olkoon $\alpha \in (0, 1)$ ja a sellainen, että $\mathbb{P}(-a < T < a) = 1 - \alpha$.
Tällöin

$$\mathbb{P}\left(\overline{X}_n - a \frac{S_n}{\sqrt{n}} \leq \mu \leq \overline{X}_n + a \frac{S_n}{\sqrt{n}}\right) = 1 - \alpha.$$

Jos esim. $\alpha = 0,05$, on kyseessä 95% *luottamusväli*. Huom.
Tyypillisesti merkitään $a = t_{n-1; \alpha/2}$.

Lähtökohta

- Havainnot $X_1, \dots, X_n$.

Lähtökohta

- Havainnot $X_1, \dots, X_n$.
- Oletus: $X_i \sim N(\mu, \sigma^2)$.

Lähtökohta

- Havainnot $X_1, \dots, X_n$.
- Oletus: $X_i \sim N(\mu, \sigma^2)$.
- $\overline{X}_n$ ja S_n havainnoista.

Lähtökohta

- Havainnot $X_1, \dots, X_n$.
- Oletus: $X_i \sim N(\mu, \sigma^2)$.
- $\overline{X}_n$ ja S_n havainnoista.
- Onko $\mu = \mu_0$ (nollahypoteesi) vai $\mu \neq \mu_0$ (vastahypoteesi)?

Lähtökohta

- Havainnot $X_1, \dots, X_n$.
- Oletus: $X_i \sim N(\mu, \sigma^2)$.
- $\overline{X}_n$ ja S_n havainnoista.
- Onko $\mu = \mu_0$ (nollahypoteesi) vai $\mu \neq \mu_0$ (vastahypoteesi)?
- Varmuutta ei voi yleensä saavuttaa: p -arvo

Lähtökohta

- Havainnot $X_1, \dots, X_n$.
- Oletus: $X_i \sim N(\mu, \sigma^2)$.
- $\overline{X}_n$ ja S_n havainnoista.
- Onko $\mu = \mu_0$ (nollahypoteesi) vai $\mu \neq \mu_0$ (vastahypoteesi)?
- Varmuutta ei voi yleensä saavuttaa: p -arvo
- p -arvo: Todennäköisyys sille, että nollahypoteesi pitää paikkansa.

Lause

Jos satunnaismuuttujat $X_1, \dots, X_n$ ovat jakautuneet $N(\mu, \sigma^2)$ mukaisesti, on

$$W = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

jakautunut jakauman $\chi^2(n-1)$ mukaisesti.

Määritelmä

Lähtökohtana on diskreetti satunnaismuuttuja X , joka voi saada n arvoa $x_1, \dots, x_n$ todennäköisyyksillä $p_1, \dots, p_n$.

Määritelmä

Lähtökohtana on diskreetti satunnaismuuttuja X , joka voi saada n arvoa $x_1, \dots, x_n$ todennäköisyyksillä $p_1, \dots, p_n$.

Jakauman $p_1, \dots, p_n$ *Shannon-entropia* on

$$H(p_1, \dots, p_n) = -K(p_1 \ln p_1 + \dots + p_n \ln p_n),$$

missä $K > 0$ on vakio.

Määritelmä

Lähtökohtana on diskreetti satunnaismuuttuja X , joka voi saada n arvoa $x_1, \dots, x_n$ todennäköisyyksillä $p_1, \dots, p_n$.

Jakauman $p_1, \dots, p_n$ *Shannon-entropia* on

$$H(p_1, \dots, p_n) = -K(p_1 \ln p_1 + \dots + p_n \ln p_n),$$

missä $K > 0$ on vakio. Jos $K = \frac{1}{\ln 2}$, sanotaan entropiaa binääriseksi.

Määritelmä

Lähtökohtana on diskreetti satunnaismuuttuja X , joka voi saada n arvoa $x_1, \dots, x_n$ todennäköisyyksillä $p_1, \dots, p_n$.

Jakauman $p_1, \dots, p_n$ *Shannon-entropia* on

$$H(p_1, \dots, p_n) = -K(p_1 \ln p_1 + \dots + p_n \ln p_n),$$

missä $K > 0$ on vakio. Jos $K = \frac{1}{\ln 2}$, sanotaan entropiaa binääriseksi. Tällöin $\frac{\ln p_i}{\ln 2} = \log_2 p_i$.

Määritelmä

Lähtökohtana on diskreetti satunnaismuuttuja X , joka voi saada n arvoa $x_1, \dots, x_n$ todennäköisyyksillä $p_1, \dots, p_n$.

Jakauman $p_1, \dots, p_n$ *Shannon-entropia* on

$$H(p_1, \dots, p_n) = -K(p_1 \ln p_1 + \dots + p_n \ln p_n),$$

missä $K > 0$ on vakio. Jos $K = \frac{1}{\ln 2}$, sanotaan entropiaa binääriseksi. Tällöin $\frac{\ln p_i}{\ln 2} = \log_2 p_i$. Määritellään lisäksi $0 \cdot \ln 0 = 0$.

Määritelmä

Lähtökohtana on diskreetti satunnaismuuttuja X , joka voi saada n arvoa $x_1, \dots, x_n$ todennäköisyyksillä $p_1, \dots, p_n$.

Jakauman $p_1, \dots, p_n$ *Shannon-entropia* on

$$H(p_1, \dots, p_n) = -K(p_1 \ln p_1 + \dots + p_n \ln p_n),$$

missä $K > 0$ on vakio. Jos $K = \frac{1}{\ln 2}$, sanotaan entropiaa binääriseksi. Tällöin $\frac{\ln p_i}{\ln 2} = \log_2 p_i$. Määritellään lisäksi $0 \cdot \ln 0 = 0$. Jos diskreetillä satunnaismuuttujalla X on jakauma $p_1, \dots, p_n$, merkitään ylläolevaa entropiaa myös $H(X)$:llä.

Esimerkki

Jos kaikki arvot x_i ilmaantuvat samalla todennäköisyydellä $\frac{1}{n}$, on

$$H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = -\left(\frac{1}{n} \log_2 \frac{1}{n} + \dots + \frac{1}{n} \log_2 \frac{1}{n}\right) = \log_2 n.$$

Esimerkki

Jos kaikki arvot x_i ilmaantuvat samalla todennäköisyydellä $\frac{1}{n}$, on

$$H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = -\left(\frac{1}{n} \log_2 \frac{1}{n} + \dots + \frac{1}{n} \log_2 \frac{1}{n}\right) = \log_2 n.$$

Tämä merkitsee sitä, että satunnaismuuttujan X esittämän viestin välittämiseen tarvitaan keskimäärin $\log_2 n$ bittiä.

Esimerkki

Jos kaikki arvot x_i ilmaantuvat samalla todennäköisyydellä $\frac{1}{n}$, on

$$H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = -\left(\frac{1}{n} \log_2 \frac{1}{n} + \dots + \frac{1}{n} \log_2 \frac{1}{n}\right) = \log_2 n.$$

Tämä merkitsee sitä, että satunnaismuuttujan X esittämän viestin välittämiseen tarvitaan keskimäärin $\log_2 n$ bittiä.

Esimerkki

Jos satunnaismuuttuja X saa arvon x_1 todennäköisyydellä 1 ja muut arvot $x_2, \dots, x_n$ todennäköisyydellä 0, on

$$H(X) = -(1 \cdot \log_2 1 + 0 \cdot \log_2 0 + \dots + 0 \cdot \log_2 0) = 0$$

Esimerkki

Satunnaismuuttuja X saa kaksi arvoa x_0 ja x_1 , molemmat todennäköisyydellä $\frac{1}{2}$.

Esimerkki

Satunnaismuuttuja X saa kaksi arvoa x_0 ja x_1 , molemmat todennäköisyydellä $\frac{1}{2}$. Tällöin

$$H(X) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right) = 1.$$

Esimerkki

Satunnaismuuttuja X saa kaksi arvoa x_0 ja x_1 , molemmat todennäköisyydellä $\frac{1}{2}$. Tällöin

$$H(X) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right) = 1.$$

Esimerkki

Satunnaismuuttuja X saa arvon x_0 todennäköisyydellä $\frac{1}{4}$ ja arvon x_1 todennäköisyydellä $\frac{3}{4}$.

Esimerkki

Satunnaismuuttuja X saa kaksi arvoa x_0 ja x_1 , molemmat todennäköisyydellä $\frac{1}{2}$. Tällöin

$$H(X) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right) = 1.$$

Esimerkki

Satunnaismuuttuja X saa arvon x_0 todennäköisyydellä $\frac{1}{4}$ ja arvon x_1 todennäköisyydellä $\frac{3}{4}$. Tällöin

$$H(X) = -\left(\frac{1}{4} \log_2 \frac{1}{4} + \frac{3}{4} \log_2 \frac{3}{4}\right) = 0,811278 \dots$$

Lause

Oletetaan että reaaliarvoinen funktio $H(p_1, \dots, p_n)$ toteuttaa seuraavat ehdot:

- $H(p_1, \dots, p_n)$ on symmetrinen ja jatkuva
- $H(\frac{1}{n}, \dots, \frac{1}{n})$ on ei-negatiivinen, aidosti kasvava
- Todennäköisyysjakaumille $(p_1, \dots, p_n)$ ja $(q_1, \dots, q_m)$ on voimassa

$$\begin{aligned} & H(p_1, \dots, p_n) + p_n H(q_1, \dots, q_m) \\ = & H(p_1, \dots, p_{n-1}, p_n q_1, \dots, p_n q_m). \end{aligned}$$

Lause

Oletetaan että reaaliarvoinen funktio $H(p_1, \dots, p_n)$ toteuttaa seuraavat ehdot:

- $H(p_1, \dots, p_n)$ on symmetrinen ja jatkuva
- $H(\frac{1}{n}, \dots, \frac{1}{n})$ on ei-negatiivinen, aidosti kasvava
- Todennäköisyysjakaumille $(p_1, \dots, p_n)$ ja $(q_1, \dots, q_m)$ on voimassa

$$\begin{aligned} & H(p_1, \dots, p_n) + p_n H(q_1, \dots, q_m) \\ &= H(p_1, \dots, p_{n-1}, p_n q_1, \dots, p_n q_m). \end{aligned}$$

Tällöin

$$H(p_1, \dots, p_n) = -K(p_1 \ln p_1 + \dots + p_n \ln p_n),$$

missä $K > 0$ on vakio.

Yhdistetty entropia

$$H(X, Y) = - \sum p(x, y) \log_2 p(x, y)$$

Yhdistetty entropia

$$H(X, Y) = - \sum p(x, y) \log_2 p(x, y)$$

Ehdollinen entropia

$$H(X \mid y_j) = - \sum p(x \mid y_j) \log_2 p(x \mid y_j)$$

Yhdistetty entropia

$$H(X, Y) = - \sum p(x, y) \log_2 p(x, y)$$

Ehdollinen entropia

$$H(X \mid y_j) = - \sum p(x \mid y_j) \log_2 p(x \mid y_j)$$

$$H(X \mid Y) = \sum p(y_j) H(X \mid y_j)$$

Yhdistetty entropia

$$H(X, Y) = - \sum p(x, y) \log_2 p(x, y)$$

Ehdollinen entropia

$$H(X \mid y_j) = - \sum p(x \mid y_j) \log_2 p(x \mid y_j)$$

$$H(X \mid Y) = \sum p(y_j) H(X \mid y_j)$$

Informaatio

$$I(X \mid Y) = H(X) - H(X \mid Y)$$

Määritelmä

Mikäli n -pituisen jono satunnaismuuttujan X arvoista muodostuvia viestejä voidaan koodata $m = \lfloor rn \rfloor$ -pituisilla bittijonoilla, sanotaan, että koodaus onnistuu tahdilla (rate) r .

Määritelmä

Mikäli n -pituinen jono satunnaismuuttujan X arvoista muodostuvia viestejä voidaan koodata $m = \lfloor rn \rfloor$ -pituisilla bittijonoilla, sanotaan, että koodaus onnistuu tahdilla (rate) r .

Shannonin lause

Jos $r > H(X)$, on mahdollista koodata X :n arvot binääriseen aakkostoon tahdilla r siten että dekodeausvirheen todennäköisyys lähenee nollaa.

Määritelmä

Mikäli n -pituisen jono satunnaismuuttujan X arvoista muodostuvia viestejä voidaan koodata $m = \lfloor rn \rfloor$ -pituisilla bittijonoilla, sanotaan, että koodaus onnistuu tahdilla (rate) r .

Shannonin lause

Jos $r > H(X)$, on mahdollista koodata X :n arvot binääriseen aakkostoon tahdilla r siten että dekodausvirheen todennäköisyys lähenee nollaa.

Jos $r < H(X)$, edellämainittu koodaus ei ole mahdollista, vaan dekodausvirheen todennäköisyys lähenee ykköstä.

Määritelmä

Binäärisen symmetrisen kanavan *kapasiteetti* on

$$C(p) = 1 - H_2(p),$$

missä p on kanavan virhetodennäköisyys.

Määritelmä

Binäärisen symmetrisen kanavan *kapasiteetti* on

$$C(p) = 1 - H_2(p),$$

missä p on kanavan virhetodennäköisyys.

Koodin C_n *informaatiosuhde* on

$$R(C_n) = \frac{\log_2 |C_n|}{n}.$$

Määritelmä

Binäärisen symmetrisen kanavan *kapasiteetti* on

$$C(p) = 1 - H_2(p),$$

missä p on kanavan virhetodennäköisyys.

Koodin C_n *informaatiosuhde* on

$$R(C_n) = \frac{\log_2 |C_n|}{n}.$$

Shannonin lause 2

Jos $p < \frac{1}{2}$ ja $R < C(p)$ ja $\epsilon > 0$, niin on sellainen rajaluku $N = N(p, R, \epsilon)$, että aina kun $n \geq N$, niin on olemassa n -pituinen binäärikoodi C_n jolle $R(C_n) \geq R$ ja $\mathbb{P}_{er}(C_n) < \epsilon$.