Líkindafræðileg undirstaða

Tölfræði frá grunni

Anna Helga Jónsdóttir og Sigrún Helga Lund

Háskóli Íslands

Helstu atriði:

1. Slembistærðir
2. Stærðfræðilegir eiginleikar slembistærða
3. Strjálar líkindadreifingar - hluti 1
4. Strjálar líkindadreifingar - hluti 2
5. Samfelldar líkindadreifingar - hluti 1
6. Samfelldar líkindadreifingar - hluti 2
7. Samfelldar líkindadreifingar - hluti 3

Yfirlit

Slembistærð

Slembistærð

Slembistærð (random variable) lýsir útkomu breytu áður en hún er mæld.

Ritháttur slembistærða

Við táknum slembistærð með **stórum** staf, oft X.

Við táknum gildi sem slembistærð **hefur tekið** með **litlum** staf, oft x.

Ávallt er notaður sami bókstafur fyrir slembistærðina og gildið sem hún tekur.

Strjálar og samfelldar slembistærðir

Strjálar slembistærðir

Strjálar slembistærðir (discrete random variables) lýsa strjálum breytum. Þær geta eingöngu tekið endanlega mörg gildi á sérhverju takmörkuðu bili.

Samfelldar slembistærðir

Samfelldar slembistærðir (continuous random variables) lýsa samfelldum breytum. Þær geta tekið hvaða gildi sem er á einhverju bili.

Líkindadreifingar slembistærða eru strjálar ef slembistærðirnar eru strjálar en annars samfelldar.

Ritháttur fyrir líkindi slembistærðai

Ritháttur fyrir líkindi slembistærða

$P(X \leq a)$: Táknar líkur þess að útkoma slembistærðarinnar X verði **minni eða jöfn** gildinu a.

$P(X \geq a)$: Táknar líkur þess að útkoma slembistærðarinnar X verði **stærri eða jöfn** gildinu a.

$P(a \leq X \leq b)$: Táknar líkur þess að útkoma slembistærðarinnar X verði **á milli** a og b, bæði gildin meðtalin.

$P(X = a)$: Táknar líkur þess að útkoma slembistærðarinnar X verði **nákvæmlega** gildið a.

Líkindadreifing slembistærða

Líkindadreifing slembistærða

Líkindadreifing (probability distribution) slembistærðar er regla sem segir okkur hvaða gildi slembistærðin getur tekið og ennfremur:

- $P(X = a)$ fyrir öll gildi a sem hún getur tekið ef líkindadreifingin er **strjál**.
- $P(a \leq X \leq b)$ fyrir öll gildi a og b ef líkindadreifingin er **samfelld**.

Líkindadreifing slembistærðar gefur okkur allar þær upplýsingar sem hægt er að fá um hana!

Gerð líkindadreifinga

Gerð líkindadreifinga

Slembni margra þeirra breyta sem við skoðum er svipuð í eðli sínu.

Þá haga slembistærðirnar sem þær lýsa sér svipað.

Þær hafa þar af leiðandi svipaða líkindadreifingu.

Við segjum þá að líkindadreifingar slembistærðanna séu af sömu gerð.

Stiki

Stiki

Sérhverri gerð líkindadreifingar er lýst með tölum sem kallast **stikar** (parameters) líkindadreifingarinnar.

Mismunandi stikar lýsa mismunandi líkindadreifingum.

Yfirleitt eru stikarnir bara einn eða tveir.

Vitum við af hvaða gerð líkindadreifing slembistærðar er, gefa gildin á stikum hennar okkur allar þær upplýsingar sem hægt er að fá um slembistærðina.

Stutt samantekt

- Hægt að tala um líkur þess að slembistærðir taki tiltekin gildi.
- Þeim líkum er lýst með líkindadreifingu slembistærðanna, sem gefa okkur allar mögulegar upplýsingar um þær.
- Margar slembistærðir hafa líkindadreifingar af ákveðnum þekktum gerðum.
- Hverri gerð líkindadreifingar er lýst með tölum sem kallast stikar.
- Til hverrar gerðar tilheyra mismunandi stikar og þeir eru yfirleitt einn eða tveir.
- Ef við vitum af hvaða gerð líkindadreifing er, þá gefa gildi stika hennar okkur allar þær upplýsingar sem hægt er að fá um líkinadreifinguna.

Yfirlit

Óháðar slembistærðir

Óháðar slembistærðir

Við segjum að tvær slembistærðir séu **óháðar** (independent) ef útkoma annarrar slembistærðarinnar hefur engin áhrif á hver útkoma hinnar slembistærðarinnar verður.

Háðar slembistærðir

Við segjum að tvær slembistærðir séu **háðar** (dependent) ef þær eru ekki óháðar, það er ef útkoma annarrar slembistærðarinnar veldur því að einhverjar útkomur hinnar slembistærðarinnar verði líklegri eða ólíklegri en ella.

Óháðar og einsdreifðar slembistærðir

Við segjum að slembistærðir $X_1, \ldots, X_n$ séu **óháðar** (indipendent) ef hver þeirra er óháð öllum hinum og **einsdreifðar** (identically distributed) ef þær hafa allar sömu líkindadreifinguna.

Væntigildi og dreifni slembistærða

Væntigildi slembistærða

Væntigildi slembistærðar (Expected value) er **raunverulegt** meðaltal slembistærðarinnar. Það er ýmist táknað með μ eða $E[X]$. Það er einnig kallað **meðaltal þýðis** (population mean) þegar við á.

Dreifni slembistærða, $Var[X]$

Alveg eins slembistærðirnar okkar hafa raunverulegt meðaltal, þá hafa þær einnig **raunverulega dreifni**. Hana táknum við ýmist með σ^2, eða $Var[X]$. Hún er einnig kölluð **dreifni þýðis** (population variance) þegar við á.

Lögmál mikils fjölda

Lögmál mikils fjölda

Eftir því sem fjöldi mælinga á slembistærð X eykst, þá stefnir meðaltal mælinganna, táknað $\bar{x}$, nær **væntigildi** slembistærðarinnar, táknað μ eða $E[X]$.

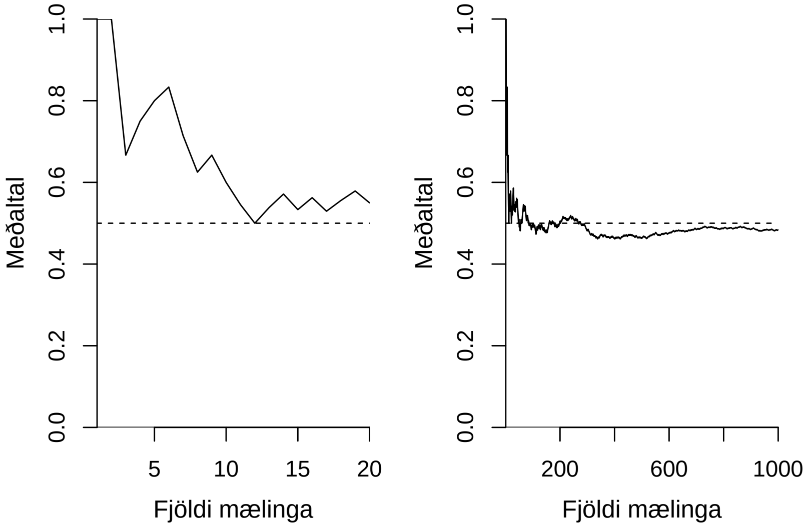
Lögmál mikils fjölda - krónukast

Látum X vera slembistærðina sem tekur gildið 0 ef upp kemur þorskur þegar krónu er kastað, en 1 ef landvættirnir koma upp. Krónunni var kastað 20 sinnum og upp komu þessar útkomur:

$$1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0$$

Fjöldi:	1	2	3	4	5	6	7	8	9	10
Meðaltal:	1.00	1.00	0.67	0.75	0.80	0.83	0.71	0.62	0.67	0.60
Fjöldi:	11	12	13	14	15	16	17	18	19	20
Meðaltal:	0.55	0.50	0.54	0.57	0.53	0.56	0.53	0.56	0.58	0.55

Lögmál mikils fjölda - krónukast



Línuleg umbreyting á samfelldum slembistærðum

Oft eru mælingar á samfelldum breytum ekki á þeim kvarða sem við hefðum viljað. Oftar en ekki dugar **línuleg umbreyting** til að koma gögnunum á kvarða sem við skiljum.

Línuleg umbreyting

Línuleg umbreyting slembistærðarinnar X með samlagningarstuðulinn a og margföldunarstuðulinn b er slembistærðin $a + bX$.

Við margföldum sem sagt sérhvert gildi með tölunni b og leggjum sérhverja útkomu við töluna a.

Línuleg umbreyting á samfelldum slembistærðum

Væntigildi og dreifni eftir línulega umbreytingu

Ef X er slembistærð og a og b eru gefnar tölur, þá eru væntigildi og dreifni línulegu umbreytingarinnar $a + bX$ stærðirnar:

$$E[a + bX] = a + b \cdot E[X]$$

og

$$Var[a + bX] = b^2 Var[X]$$

Yfirlit

Líkindadreifingar slembistærða

- Margar slembistærðir fylgja líkindadreifingum af ákveðnum þekktum gerðum.
- Líkindadreifingar slembistærða eru strjálar ef slembistærðirnar eru strjálar en annars samfelldar.
- Skoðum þær tvær strjálu líkindadreifingar sem mest eru notaðar:
  - **Tvíkostadreifingin** (binomial distribution)
  - **Poisson dreifingin** (Poisson distribution).
- Munum sjá hvernig nota má þessar tvær líkindadreifingar til að lýsa fjöldamörgum slembnum fyrirbærum.

Massafall

Massafall (mass function)

Strjálum dreifingum er lýst með **massafalli** (mass function) og notum við það til að reikna líkur einstakra útkoma strjálla slembistærða. Við táknum massafallið með $f(x)$ og það má skrifa sem

$$f(x) = P(X = x).$$

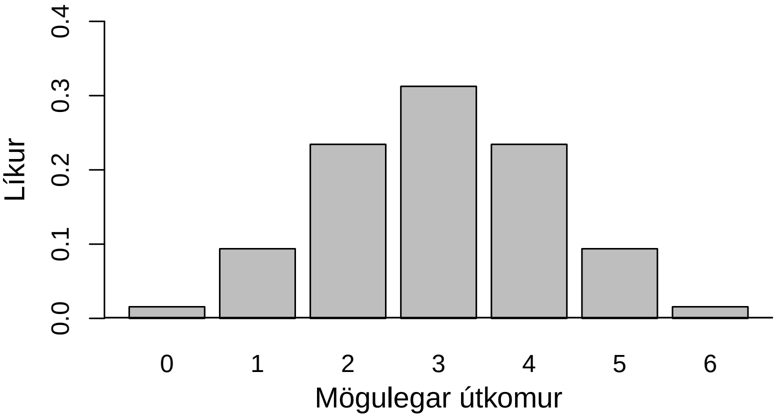
þar sem X er strjál slembistærð. Útkomumengi X táknum við með Ω og inniheldur það allar mögulegar útkomur X.
Um massafallið gildir að

$$f(x) \geq 0$$
$$\sum_{\text{yfir öll x}} f(x) = 1.$$

Við notum stöplarit til að lýsa massafallinu myndrænt.

Stöplarit massafalls

X er fjöldi þorska þegar krónu er kastað sex sinnum.



Reiknireglur fyrir strjálar slembistærðir

Reiknireglur fyrir strjálar slembistærðir

Þegar reikna á líkur fyrir strjálu slembistærðina X má oft auðvelda útreikninga með því að snúa líkunum við

$$
\begin{aligned}
P(X \leq k) &= 1 - P(X > k) \\
P(X < k) &= 1 - P(X \geq k) \\
P(X \geq k) &= 1 - P(X < k) \\
P(X > k) &= 1 - P(X \leq k)
\end{aligned}
$$

þar sem k getur verið hvaða tala sem er í útkomumengi X.

Bernoulli tilraun

Bernoulli tilraun (Bernoulli trial)

Sérhver tilraun í safni endurtekinna tilrauna flokkast sem **Bernoulli tilraun** ef eftirfarandi gildir:

1. Hver tilraun hefur aðeins tvær mögulegar útkomur. Það er venja að kalla þessar útkomur **jákvæða útkomu** (success) og **neikvæða útkomu** (failure).
2. Líkurnar á jákvæðri útkomu eru þær sömu í hverri tilraun fyrir sig. Líkurnar á neikvæðri útkomu eru þar af leiðandi þær sömu í hverri tilraun fyrir sig þar sem líkurnar á neikvæðri útkomu eru ávallt 1 mínus líkurnar á jákvæðri útkomu.
3. Útkoma í einni tilraun hefur ekki áhrif á útkomu í annarri tilraun, þ.e.a.s útkomurnar eru óháðar (independent).

Tvíkostadreifingin

- Við höfum oft áhuga á því að reikna hversu oft við sjáum jákvæða útkomu meðal safns Bernoulli tilrauna.
- Við gætum til dæmis viljað reikna líkurnar á því að fá tvær sexur (sem væru þá jákvæða útkoman) þegar teningi er kastað fimm sinnum.
- Lítum þá á heildarfjölda jákvæðra útkoma sem slembistærðina X.
- Hún hefur þekkta líkindadreifingu, sem kallast **tvíkostadreifingin** og er henni lýst með stikunum n, sem er fjöldi Bernoulli tilrauna sem framkvæmdar eru, og p sem er líkurnar á því að hver og ein Bernoulli tilraun heppnist.

Tvíkostadreifingin

Tvíkostadreifingin (binomial distribution)

Látum slembistærðina X tákna fjölda jákvæðra tilrauna úr n Bernoulli tilraunum. X fylgir þá tvíkostadreifingu með stikana n og p, skrifað $X \sim B(n, p)$, þar sem n er fjöldi tilrauna og p eru líkurnar á jákvæðri útkomu í hverri tilraun fyrir sig. Útkomumengi X er $\Omega = \{0,1,2,...n\}$. Ef k er eitthvert þessara gilda má reikna líkurnar á að slembistærðin X taki gildi k með massafalli tvíkostadreifingarinnar:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \qquad k = 0, 1, 2, ...n$$

$\binom{n}{k}$ er tvíliðustuðullinn. Hann er jafn fjölda möguleika á að fá k jákvæðar útkomur í n tilraunum reiknaður sem

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Þar sem $k! = k \cdot (k-1) \cdot (k-2) \cdot ... \cdot (1)$. Sér í lagi er $0! = 1$.

Tvíkostadreifingin

- Við höfum nú séð að reikna má líkurnar á að slembistærðin X taki eitthvert ákveðið gildi k.
- Auk þess að reikna $P(X = k)$ höfum við oft áhuga á að reikna
  - $P(a \leq X \leq b)$ eða $P(a < X < b)$
  - $P(X \leq k)$ eða $P(X < k)$
  - $P(X \geq k)$ eða $P(X > k)$
- Við getum reiknað allar þessar líkur með að nota jöfnuna fyrir massafall tvíkostadreifingarinnar ásamt því að nota reglurnar á glæru 23.

Væntigildi og dreifni tvíkostadreifingarinnar

Væntigildi og dreifni tvíkostadreifingar

Ef X fylgir tvíkostadreifingu, $X \sim B(n, p)$ þá gildir

$$\mathsf{E}[X] = np$$
$$\mathsf{Var}[X] = np(1-p)$$

Dæmi - stjálar líkindadreifingar

Dæmi

a) Hvert er gildi tvíliðustuðulsins $\binom{8}{2}$?

b) Á hversu marga vegu er hægt að fá tvo þorska þegar krónu er kastað átta sinnum?

Dæmi - stjálar líkindadreifingar

Dæmi

a) Hvert er gildi tvíliðustuðulsins $\binom{8}{2}$?

b) Á hversu marga vegu er hægt að fá tvo þorska þegar krónu er kastað átta sinnum?

Svar:

a) 28

b) 28

Dæmi - stjálar líkindadreifingar

Dæmi

Siggi kastar krónu 9 sinnum. Táknum fjölda þorska með X, $X \sim B(9, 0.5)$.

a) Hverjar eru líkurnar á því að Sigga fái á milli 2 og 5 þorska?

b) Hverjar eru líkurnar á því að Sigga fái 1 eða færri þorsk?

c) Hverjar eru líkurnar á því að Sigga fái 8 eða fleiri þorska?

d) Hverjar eru líkurnar á því að Sigga fái fleiri en 2 þorska?

Dæmi - stjálar líkindadreifingar

Dæmi

Siggi kastar krónu 9 sinnum. Táknum fjölda þorska með X, $X \sim B(9, 0.5)$.

a) Hverjar eru líkurnar á því að Sigga fái á milli 2 og 5 þorska?

b) Hverjar eru líkurnar á því að Sigga fái 1 eða færri þorsk?

c) Hverjar eru líkurnar á því að Sigga fái 8 eða fleiri þorska?

d) Hverjar eru líkurnar á því að Sigga fái fleiri en 2 þorska?

Svar:

a) 0.73

b) 0.02

c) 0.02

d) 0.91

Dæmi - stjálar líkindadreifingar

Dæmi

Hlutfall kvenkyns eðla af ákveðinni gerð sem fæðast á hitabeltiseyju langt langt í burtu er 0.48. Hverjar eru líkurnar á að í sex (einbura) fæðingum fæðist nákvæmlega þrjár kvenkyns eðlur?

Dæmi - stjálar líkindadreifingar

Dæmi

Hlutfall kvenkyns eðla af ákveðinni gerð sem fæðast á hitabeltiseyju langt langt í burtu er 0.48. Hverjar eru líkurnar á að í sex (einbura) fæðingum fæðist nákvæmlega þrjár kvenkyns eðlur?

Svar: 0.31

Yfirlit

Poisson dreifingin

- Poisson dreifingin er oft notuð til að lýsa fjölda slembinna fyrirbæra sem eiga sér stað **á ákveðinni einingu** en mögulegar útkomur hafa engin efri mörk.
- Einingarnar geta verið **tímabil** (time interval), **svæði** (spatial interval), eða einhver **hlutur** (physical object).
- Sem dæmi má nefna fjölda símtala til nemendaskrár **á** mínútu, fjölda hreindýra **á** ferkílómetra og fjölda innsláttarvillna **á** blaðsíðu.

Poisson dreifingin

Poisson dreifingin (Poisson distribution)

Poisson dreifingin hefur einn stika sem við köllum λ. Ef X fylgir Poisson dreifingu með stikanum λ má finna líkurnar á að slembistærðin X taki eitthvert gildi k, $k = 0, 1, 2, ...$ með massafalli Poisson dreifingarinnar:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Við skrifum að $X \sim \text{Pois}(\lambda)$. Útkomumengi X er $\Omega = \{0,1,2,...\}$. Stikinn λ er væntigildi slembistærðarinnar X, þ.e. raunverulegt meðaltal hennar. Hann lýsir því hvað við væntum að margar jákvæðar útkomur eigi sér stað að meðaltali á hverri einingu.

Poisson dreifingin

- Við höfum nú séð að reikna má líkurnar á að slembistærðin X sem fylgir Poisson dreifingunni taki eitthvert ákveðið gildi k með massafalli Poisson dreifingarinnar.
- Við höfum oft áhuga á að reikna aðrar líkur:
  - $P(a \leq X \leq b)$ eða $P(a < X < b)$
  - $P(X \leq k)$ eða $P(X < k)$
  - $P(X \geq k)$ eða $P(X > k)$.

Við getum reiknað allar þessar líkur með massafalli Poisson dreifingarinnar ásamt því að nota reglurnar glæru 23.

Poisson dreifingin

- Þegar reikna á líkur á að slembistærð sem fylgir Poisson dreifingu taki eitthvert gildi fáum við λ oft uppgefið sem fjölda á annarri einingu en þeirri sem við viljum vinna með.
- Við gætum til dæmis vitað fjölda bilana í fólksbíl á mánuði, en við viljum finna fjölda bilana á ári. Þá væri gefna einingin mánuður en einingin sem við viljum vinna með er ár. Þá þarf að laga λ að nýrri einingu.

Væntigildi og dreifni Poisson dreifingar

Væntigildi og dreifni Poisson dreifingar

Ef X fylgir Poisson dreifingu, $X \sim \text{Pois}(\lambda)$ þá gildir

$$\begin{aligned} \mathsf{E}[X] &= \lambda \\ \mathsf{Var}[X] &= \lambda. \end{aligned}$$

Dæmi - stjálar líkindadreifingar

Dæmi

Skiptiborðið í banka fær að meðaltali 2.4 símtöl á mínútu.

Reiknið líkurnar á að skiptiborðið fái:

a) þrjú símtöl á einni mínútu.

b) að minnsta kosti tvö símtöl á einni mínútu.

c) sex símtöl á tveimur mínútum.

d) sjö símtöl á þremur mínútum.

Dæmi - stjálar líkindadreifingar

Dæmi

Svör:

a) 0.21

b) 0.69

c) 0.14

d) 0.15

Dæmi

Dæmi

Látum X tákna slembistærð sem fylgir Poisson dreifingu með $\lambda = 6$. Finnið væntigildi og dreifni X.

Dæmi

Dæmi

Látum X tákna slembistærð sem fylgir Poisson dreifingu með $\lambda = 6$. Finnið væntigildi og dreifni X.

Svar:
$\mathsf{E}[X] = \lambda = 6$
$\mathsf{VAR}[X] = \lambda = 6$

Yfirlit

Líkindadreifingar slembistærða

- Margar slembistærðir fylgja líkindadreifingum af ákveðnum þekktum gerðum.
- Líkindadreifingar slembistærða eru samfelldar ef slembistærðirnar eru samfelldar en annars strjálar.
- Skoðum þá samfelldu líkindadreifingu sem mest er notuð:
  - **Normaldreifingin** (normal distribution)
- Skoðum þrjár samfelldar dreifingar sem við munum nota þegar kemur að ályktunartölfræði
  - t-dreifing
  - χ^2-dreifing
  - F-dreifing

Líkur samfelldra slembistærða

Um samfelldar slembistærðir gildir að

$$P(X = x) = 0.$$

Jafnan segir okkur að líkurnar á því að samfelld slembistærð taki eitthvert eitt ákveðið gildi eru alltaf núll, sama hvert gildið er. Því gildir að

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

þegar X er samfelld. Munið að þetta gildir almennt ekki um strjálar slembistærðir!

Dreififall

Dreififall (Distribution function)

Með **dreififalli** reiknum við líkurnar á að samfelld slembistærð X taki gildi sem er minna en viðmiðunargildið x. Við táknum dreififallið með $F(x)$ og það má skrifa sem

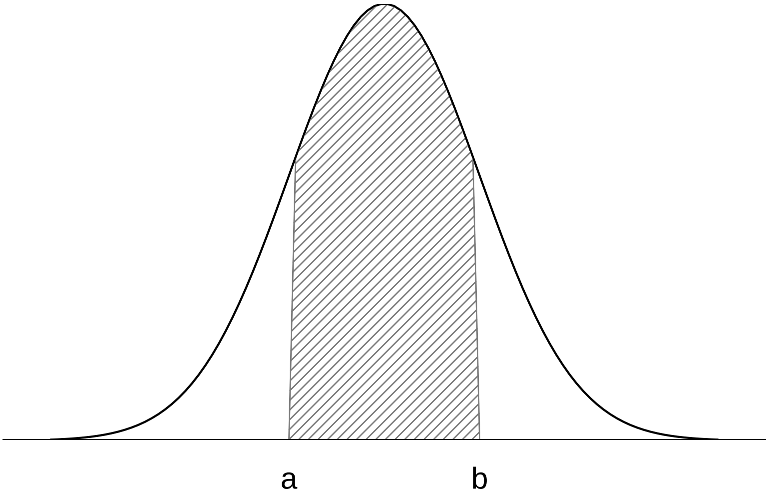
$$F(x) = P(X < x)$$

Þéttifall og þéttiferill

Þéttifall og þéttiferill (Density function and density curve)

Þéttifall (density function) er táknað með $f(x)$ og kallast graf þess **þéttiferill** (density curve). Flatarmálið undir þéttiferlinum milli tveggja stærða a og b er jafnt $P(a < X < b)$, líkunum á því að slembistærðin taki gildi á milli a og b.

Þéttiferill



Líkur samfelldra slembistærða

Reiknireglur fyrir samfelldar slembistærðir

Eftirfarandi reglur sem gilda um samfelldar slembistærðir koma oft að góðum notum.

$$P(X > a) = 1 - P(X < a)$$

$$P(a < X < b) = P(X < b) - P(X < a).$$

Normaldreifing

- Normaldreifingin mest notaða dreifingin innan tölfræðinnar.
- Margs konar fyrirbærum má lýsa með normaldreifingu svo sem hæð, blóðþrýstingi, villur í mælitækjum, ...
- Við kynnumst einnig mikilvægi dreifingarinnar þegar farið verður í höfuðsetningu tölfræðinnar síðar.
- Þéttiferill normaldreifingarinnar er bjöllulaga og hefur dreifingin tvo stika sem ráða lögun hennar.

Normaldreifingin

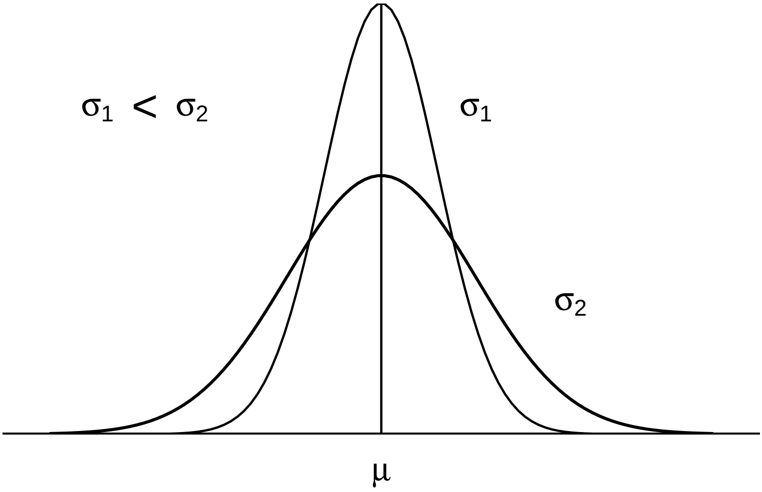
Normaldreifingin (normal distribution)

Þéttifall normaldreifingarinnar er oft táknað með $\phi(x)$ og má skrifa það sem

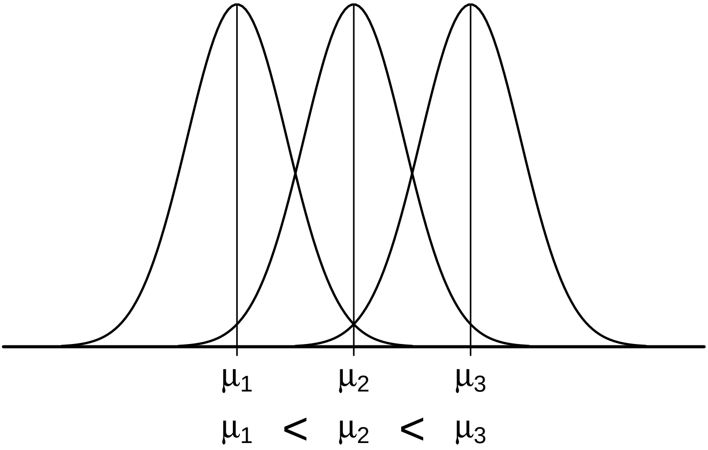
$$f(x) = \phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Fallið er stikað með tveimur **stikum** (parameters), μ og σ. μ táknar meðaltal normaldreifingarinnar og ræður það staðsetningu (location) hennar. σ^2 táknar dreifni dreifingarinnar og ræður hún dreifð (spread) hennar. Ef slembistærðin X fylgir normaldreifingu með meðaltal μ og dreifni σ^2 skrifum við að $X \sim N(\mu, \sigma^2)$. Dreififall normaldreifingarinnar er táknað með $\Phi(x)$.

Tvær normaldreifingar með sama meðaltal en ólíka dreifni



Þrjár normaldreifingar með sömu dreifni en ólík meðaltöl



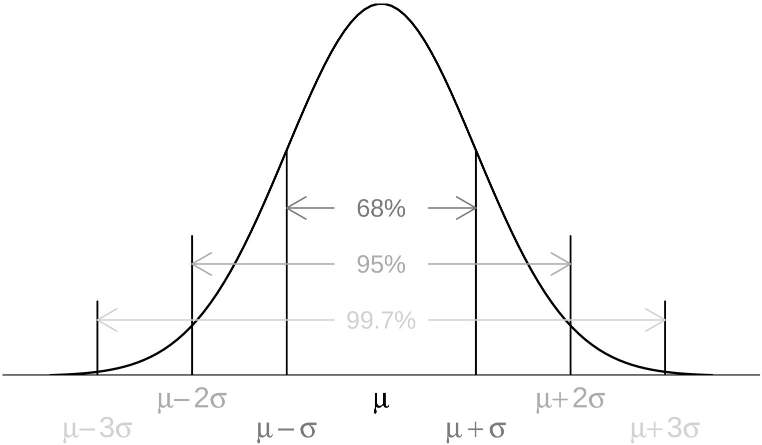
68-95-99.7% reglan

68-95-99.7% reglan

Fyrir sérhverja normaldreifingu með meðaltal μ og staðalfrávik σ gildir að

- u.þ.b 68% mælinga munu liggja innan við eitt staðalfrávik frá meðaltalinu
- u.þ.b 95% mælinga munu liggja innan við tvö staðalfrávik frá meðaltalinu
- u.þ.b 99.7% mælinga munu liggja innan við þrjú staðalfrávik frá meðaltalinu

68-95-99.7% reglan



Yfirlit

Stöðluð normaldreifing

Normaldreifing með meðaltal $\mu = 0$ og staðalfrávik $\sigma = 1$ er kölluð staðlaða normaldreifingin.

Staðlaða normaldreifingin (Standardized normal distribution)

Ef slembistærðin X fylgir normaldreifingu með meðaltal μ, staðalfrávik σ og dreifni σ^2, skrifað

$$X \sim N(\mu, \sigma^2)$$

þá fylgir

$$Z = \frac{X - \mu}{\sigma}$$

normaldreifingu með meðaltal $\mu = 0$ og staðalfrávik $\sigma = 1$, skrifað

$$Z \sim N(0, 1).$$

Samband X og Z

Samband X og Z

Ef slembistærðin X fylgir normaldreifingu með meðaltal μ og dreifni σ^2, $X \in N(\mu, \sigma^2)$, og slembistærðin Z fylgir stöðluðu normaldreifingunni, $Z \in N(0, 1)$, þá gildir að

$$P(X \leq x) = P(Z \leq z)$$

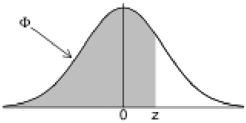
þar sem $z = \frac{x-\mu}{\sigma}$.

Líkur normaldreifðra slembistærða

- Líkur normaldreifðra slembistærða má reikna sem flatarmálið undir þéttiferlinum.
- Ef finna á líkurnar á að normaldreifð slembistærð liggi á bilinu frá a til b þarf að heilda þéttifallið milli a og b. Þetta gerum við ekki í höndunum en notum til þess töflur.
- Töflur fyrir stöðluðu normaldreifinguna má yfirleitt finna í kennslubókum um tölfræði.

Tafla stöðluðu normaldreifingarinnar

Normaldreifing - neikvæð z-gildi



Taflan gefur gildi á Φ, það er líkurnar á að Z taki gildi sem er minna en z, þar sem Z fylgir normaldreifingu með meðaltal 0 og staðalfrávik 1.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
-3.50	0.0002	**-3.15**	0.0008	**-2.80**	0.0026	**-2.45**	0.0071
-3.49	0.0002	**-3.14**	0.0008	**-2.79**	0.0026	**-2.44**	0.0073
-3.48	0.0003	**-3.13**	0.0009	**-2.78**	0.0027	**-2.43**	0.0075
-3.47	0.0003	**-3.12**	0.0009	**-2.77**	0.0028	**-2.42**	0.0078
-3.46	0.0003	**-3.11**	0.0009	**-2.76**	0.0029	**-2.41**	0.0080
-3.45	0.0003	**-3.10**	0.0010	**-2.75**	0.0030	**-2.40**	0.0082
-3.44	0.0003	**-3.09**	0.0010	**-2.74**	0.0031	**-2.39**	0.0084
-3.43	0.0003	**-3.08**	0.0010	**-2.73**	0.0032	**-2.38**	0.0087
-3.42	0.0003	**-3.07**	0.0011	**-2.72**	0.0033	**-2.37**	0.0089
-3.41	0.0003	**-3.06**	0.0011	**-2.71**	0.0034	**-2.36**	0.0091

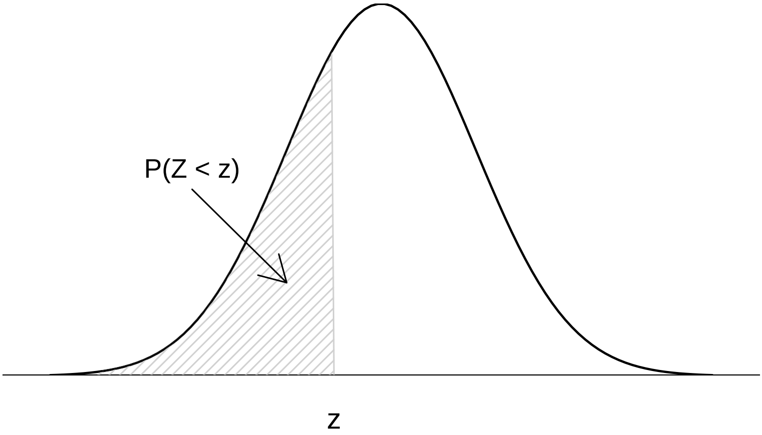
Líkur normaldreifðra slembistærða

Áður en við getum notað töflurnar þarf að koma normaldreifingunni á staðlað form. Taflan gefur eftirfarandi líkur:

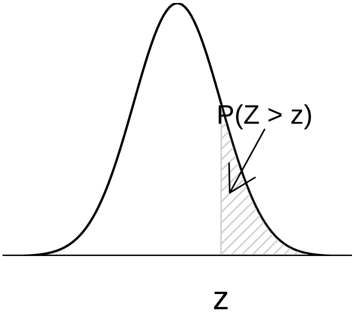
$$\Phi(z) = P(Z < z),$$

það er að segja taflan gefur okkur líkurnar á að slembistærðin Z sem fylgir stöðluðu normaldreifingunni taki gildi minna en töluna z, kallað z-gildið. Þetta má hugsa sem svo að taflan horfi til vinstri.

Líkur normaldreifðra slembistærða



$P(Z > z)$ og $P(a < Z < b)$ þar sem $Z \sim N(0, 1)$



Notkun töflu stöðluðu normaldreifingarinnar

Notkun töflu stöðluðu normaldreifingarinnar

Hægt er að nota töfluna á tvo vegu:

1. Ef við viljum finna líkurnar sem svara til ákveðins viðmiðunargildis:
   Ef gildið er fengið úr staðlaðri normaldreifingu er það gildi sjálft z-gildið. Ef gildið er ekki fengið úr staðlaðri normaldreifingu finnum við staðlaða z-gildið með

   $$z = \frac{x - \mu}{\sigma}$$

   Við finnum z-gildið í töflunni (feitletrað) og líkurnar eru $\Phi(z)$ gildið því á hægri hlið.

2. Ef við viljum finna hvaða viðmiðunargildi svarar til ákveðinna líkinda:
   Við finnum líkurnar, eða þær líkur sem þeim eru næstar, meðal $\Phi(z)$-gildanna í töflunni og z-gildið stendur (feitletrað) því á vinstri hlið.
   Ef viðmiðunargildið er ekki fengið úr staðlaðri normaldreifingu þurfum við að **varpa** z-gildinu aftur í upphaflegu dreifinguna, svo tilsvarandi gildi verður

   $$x = \mu + z\sigma$$

Rithátturinn z_a

Rithátturinn z_a

Með z_a táknum við það z-gildi sem er þannig að slembistærð sem fylgir stöðluðu normaldreifingunni hefur líkurnar a að taka gildi sem er **minna** en z_a. Þetta má skrifa sem:

$$a = P(Z < z_a).$$

þar sem Z fylgir stöðluðu normaldreifingunni. z_a er því a-ta prósentumark stöðluðu normaldreifingarinnar.

Normaldreifingarrit

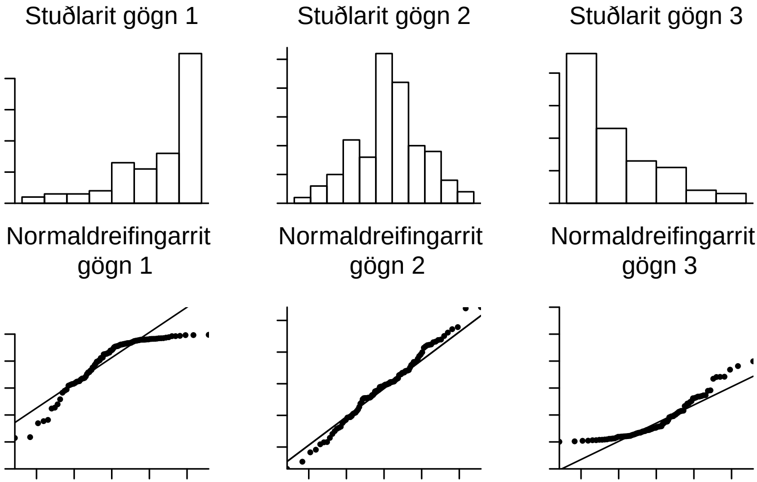
- Margar tölfræðilegar aðferðir eru háðar því að gögnin sem þeim er beitt á fylgi normaldreifingu.
- Áður en aðferðirnar eru notaðar þarf því að ganga úr skugga um að svo sé raunin.
- Ýmsar aðferðir finnast til þess og er algengt að nota svonefnt normaldreifingarrit.
- Normaldreifingarrit er myndræn aðferð til að kanna hvort gögn fylgi normaldreifingu eða ekki.
- Normaldreifingarrit eru sjaldan gerð í höndunum heldur er notast við tölfræðiforrit en mikilvægt er að kunna að túlka þau.

Normaldreifingarrit

Normaldreifingarrit (normal probability plot)

Ef punktarnir á normaldreifingarritinu liggja nálægt beinu línunni sem sýnd er á ritinu og endapunktarnir báðum megin sveigjast ekki afgerandi upp eða niður þá er ásættanlegt að gera ráð fyrir að gögnin fylgi normaldreifingu.

Normaldreifingarrit



Dæmi - samfelldar líkindadreifingar

Dæmi

Í USA þurfa nemendur að þreyta staðlað próf, svokallað SAT próf, áður en þeir fara í menntaskóla. Gera má ráð fyrir því að einkunnir nemenda á prófinu séu u.þ.b normaldreifðar með meðaltal 1026 og staðalfrávik 209. Köllum nú einkunnirnar X, $X \sim N(1026, 209^2)$.

a) Reiknið líkurnar á að einkunn nemanda sem þreytir prófið og valinn er af handahófi sé lægri en 720, þ.e.a.s $P(X \leq 720)$.

b) Reiknið líkurnar á að einkunn nemanda sem þreytir prófið og valinn er af handahófi sé hærri en 820, þ.e.a.s $P(X \geq 820)$.

c) Reiknið líkurnar á að einkunn nemanda sem þreytir prófið og valinn er af handahófi sé á bilinu 720 og 820, þ.e.a.s $P(720 \leq X \leq 820)$.

d) Hvaða einkunn þarf nemandi að ná í SAT prófinu ætli hann sér að vera í hópi 10% efstu nemendanna?

Dæmi - samfelldar líkindadreifingar

Dæmi

Svar:

a) 0.07

b) 0.84

c) 0.09

d) 1293.84

Yfirlit

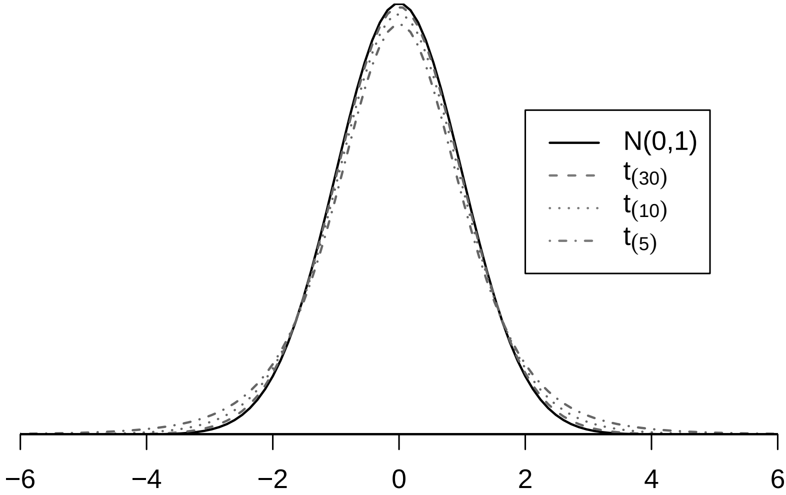
t-dreifingin

- t-dreifingin, eða Student's t, er samfelld líkindadreifing sem minnir á normaldreifinguna.
- Hún er bjöllulaga og samhverf um meðaltal dreifingarinnar sem er 0. Við þurfum á t-dreifingunni að halda síðar meir þegar við tölum um ályktunartölfræði.
- t-dreifingin hefur einn stika, sem kallast frígráður. Við notum k til að tákna fjölda frígráða. t-dreifingu með k frígráðum táknum við með $t_{(k)}$.

t-dreifingin

- t-töflu má finna í flestum kennslubókum um tölfræði.
- Við flettum upp í töflunni eftir fjölda frígráða. Gildin sem við fáum út úr töflunni táknum við með $t_{a,(k)}$.
- Um $t_{a,(k)}$ gildir að slembistærð sem fylgir t-dreifingu með k frígráður hefur líkurnar a að taka gildi sem er **minna** en $t_{a,(k)}$.
- Dálkur er valinn eftir a-gildinu en lína eftir fjölda frígráða.
- Eftir því sem fjöldi frígráða eykst því meira líkist t-dreifingin stöðluðu normaldreifingunni.

Nokkrar t-dreifingar



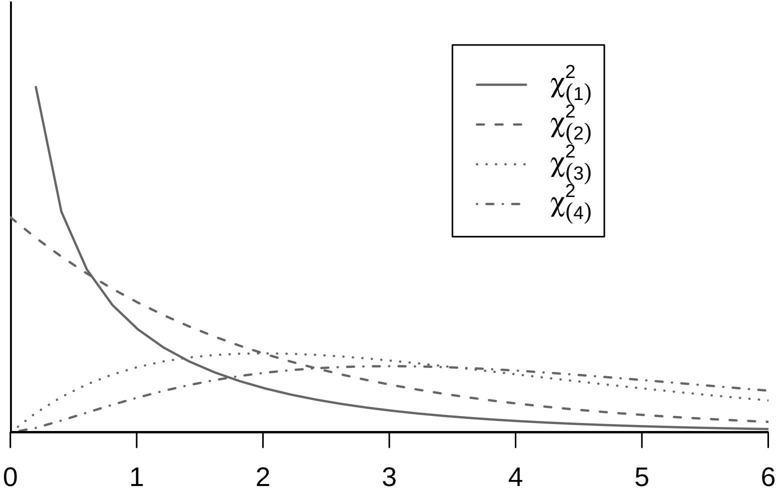
χ^2-dreifing

- χ^2-dreifingin, lesist kí-kvaðrat dreifingin, er samfelld líkindadreifing og er hún mikið notuð í ályktunartölfræði.
- Hún er ekki samhverf eins og normaldreifingin.
- χ^2-dreifingin hefur einn stika, fjölda frígráða, sem við köllum k.
- χ^2-dreifingu með k frígráðum táknum við með $\chi^2_{(k)}$.
- Meðaltal χ^2-dreifingar er jafnt fjölda frígráða hennar.

χ^2-dreifing

- χ^2-töflu má finna í flestum kennslubókum um tölfræði.
- Við flettum upp í töflunni eftir fjölda frígráða. Gildin sem við fáum út úr töflunni táknum við með $\chi^2_{a,(k)}$.
- Um $\chi^2_{a,(k)}$ gildir að slembistærð sem fylgir χ^2-dreifingu með k frígráður hefur líkurnar a á að taka gildi sem er **minna** en $\chi^2_{a,(k)}$.
- Við veljum dálk eftir a-gildinu en línu eftir fjölda frígráða.

Nokkrar χ^2-dreifingar



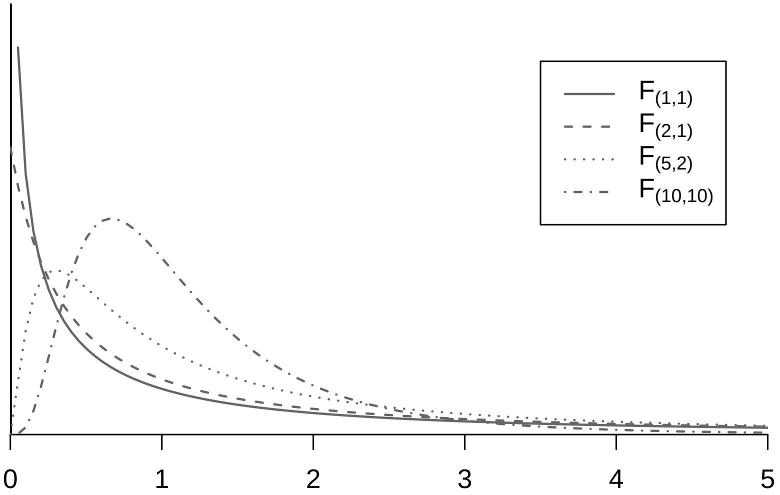
F-dreifingin

- F-dreifingin er samfelld líkindadreifing sem við munum nota þegar kemur að ályktunartölfræði.
- Líkt og χ^2-dreifingin er hún ekki samhverf.
- F-dreifingin hefur tvo stika sem við köllum frígráður og táknum með v_1 og v_2.
- F-dreifingu með v_1 og v_2 fjölda frígráða táknum við með $F_{(v_1, v_2)}$.

F-dreifingin

- F töflur má finna í flestum kennslubókum um tölfræði og eru þær oft fjórar og þarf að passa vel að nota þá réttu hverju sinni.
- Töflurnar fjórar eru fyrir fjögur mismunandi a-gildi, $a = 0.90$, $a = 0.95$, $a = 0.975$ og $a = 0.99$.
- Dálkarnir í töflunum tákna mismunandi gildi á v_1 og línurnar mismunandi gildi á v_2.
- Gildin sem við fáum út úr töflunni táknum við svo með $F_{a,(v_1,v_2)}$.
- Um $F_{a,(v_1,v_2)}$ gildir að slembistærð sem fylgir F-dreifingu með v_1 og v_2 frígráðum hefur líkurnar a á að taka gildi sem er **minna** en $F_{a,(v_1,v_2)}$.

Nokkrar F-dreifingar



Dæmi - z_a, t_a, χ^2_a, F_a ritháturinn

Dæmi

a) Finnið $z_{0.95}$.

b) Finnið $t_{0.975,(17)}$.

c) Finnið $\chi^2_{0.025,(1)}$.

d) Finnið $F_{0.95,(7,12)}$.

Dæmi - z_a, t_a, χ^2_a, F_a ritháturinn

Dæmi

a) Finnið $z_{0.95}$.
b) Finnið $t_{0.975,(17)}$.
c) Finnið $\chi^2_{0.025,(1)}$.
d) Finnið $F_{0.95,(7,12)}$.

a) 1.645
b) 2.11
c) 0.001
d) 2.913