

# Líkindi og líkindalíkön

## Fyrirlestur í Aðferðafræði II

© 1998, 2000–2003 Guðmundur Arnkelsson

All rights reserved. Copying or distribution prohibited without explicit permission. Students in Methodology II at the University of Iceland may print a copy for their own private use.



# Líkindafræði

Líkindafræði segir til um hve oft hlutfallslega atburður verður.

Gert er ráð fyrir að eitthvert ferli liggi til grundvallar atburðinum.

Ferlið er óbreytt; það er forsenda þess að tilgreina ákveðin líkindi.

Allir atburðir eru skilgreindir fyrirfram

Sá atburður sem verður er hluti af mengi allra atburða (*sample space*).

Atburður getur verið einstakur eða samsettur; tveir eða fleiri atburðir geta orðið, ýmist allir í einu eða hver fyrir sig.

Atburður í líkindafræði byggist á tilviljun

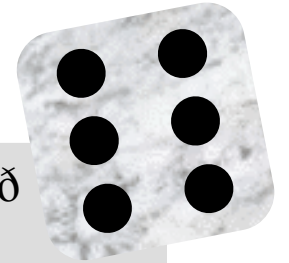
Við vitum ekki hvort atburðurinn verður eða verður ekki, aðeins líkindi hans.

Teningskast er líkindaatburður

Við vitum ekki hvaða tala kemur næst. Ef við fáum sexu, segir það okkur ekkert um hvað komi næst.

Ef við vitum hvaða ferli liggur að baki niðurstöðunni og ef ferlið er óbreytt, getum við ákvarðað *líkindi* atburða.

# Fræðilíkur



Fræðilíkur eru þau líkindi sem við búumst við fyrirfram

Við getum lagt mat á fræðilíkur ef við þekkjum það ferli sem liggur að baki atburðunum.

Oftast höfum við of litlar upplýsingar til að meta líkindi. Hver eru t.d. líkindi þess að bíllinn minn bili í næstu viku?

Ef allir atburðir eru jafnlíklegir er þó auðvelt að meta fræðilíkur; þær eru þá hlutfallsleg tíðni miðað við aðra atburði.

Við vitum því að líkindin er  $\frac{1}{6}$  á því að fá hverja tölu á teningi.

Fræðilíkindi miðast alltaf við einhverjar forsendur

Fræðilíkur tenings eiga aðeins við „réttu“ teninga; ef eitthvað er átt við teninginn verða fræðilíkurnar aðrar.

Stundum miðast fræðilíkurnar við það að ákveðin aðferð sé notuð til að fá atburðinn. Líkindi þess að draga spaða úr venjulegum spilastokki eru 25%. Þetta er þó aðeins rétt ef spilin eru vel stokkuð.

Mikilvægt viðfangsefni líkindafræði er að ákvarða fræðilíkindi og þær forsendur sem þau byggja á.

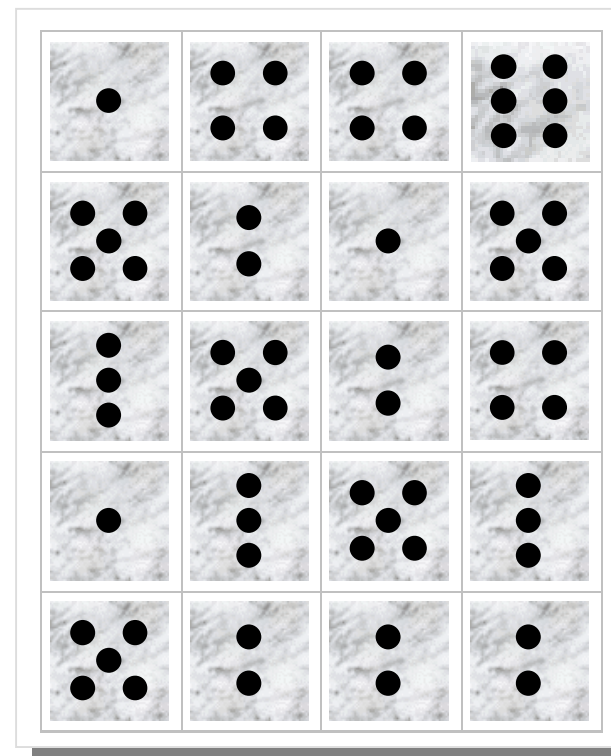
# Raunlíkur

Raunlíkur samsvara hlutfallslegri tíðni til lengri tíma litið

Við getum metið líkindi atburðar með því að endurtaka aftur og aftur ferlið sem orsakar hann, t.d. kastað teningi aftur og aftur.

Þar sem tilviljun ræður hvaða atburður verður, er vandasamt að finna réttar raunlíkur. Við getum gert líkindatilraunir (*random experiments*); til hliðar eru 20 teningsköst og þrjú þeirra eru fernur eða 15%—þetta eru raunlíkur þess að fá fernu.

Í sumum slíkum tilraunum verða engar fernur en í öðrum verða þær fleiri. Til lengri tíma litið fáum við þó 16,7% fernur, þ.e. að meðaltali 3,3 fernur í hverri 20 teningskasta tilraun. Raunlíkur nálgast þannig fræðilíkur eftir því sem teningsköstum fjölgar.



Niðurstöður tuttugu teningskasta

# Raunlíkur sem hlutfallsleg tíðni

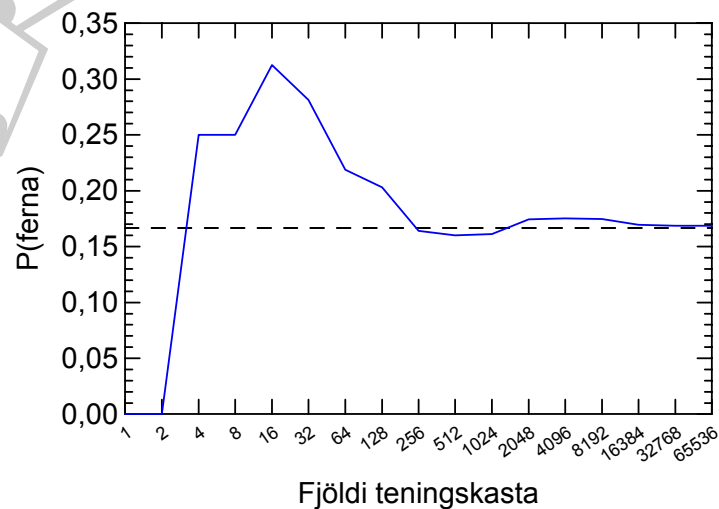
Það getur þurft mjög mörg skipti fyrir raunlíkur til að nálgast fræðilíkur

Gerum ráð fyrir að við séum með „réttan“ tening. Fræðilíkindi þess að fá fjóra eru því 16,7%. Prófum nú að kasta teningnum aftur og aftur og sjáum hvaða raunlíkindi fást.

Taflan sýnir hvernig raunlíkur breytast þegar teningsköstum fjölga. Brotalínan sýnir fræðilíkindin 0,167.

Myndin sýnir miklar sveiflur í raunlíkum þar til köstin eru komin yfir 200. Það er síðan ekki fyrr en komið er upp í mörg þúsund köst sem raunlíkur verða jafnar fræðilíkum.

Við getum því þurft miklar upplýsingar ef staðfesta á að teningurinn er með „réttur“.



Fræðilíkurnar eru heimurinn eins og við höldum að hann sé—líkan af heiminum—en raunlíkur heimurinn eins og við þekkjum hann—upplýsingar um heiminn.

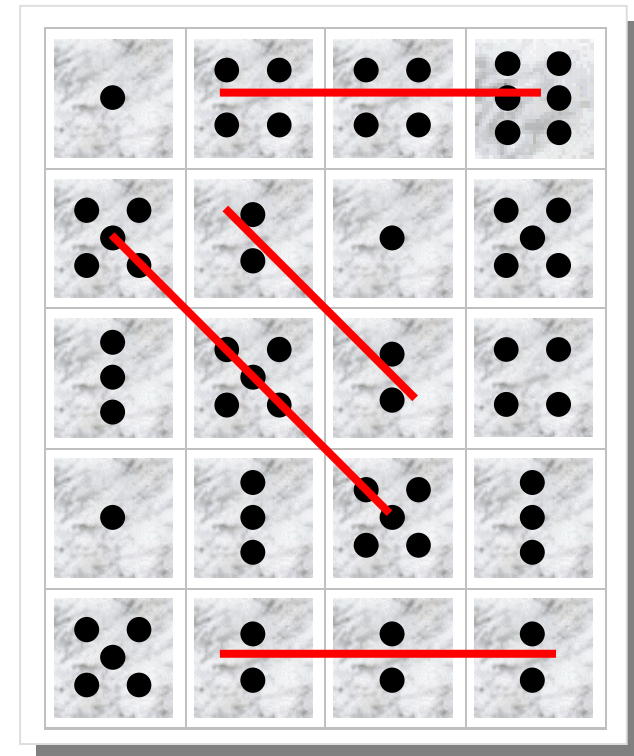
# Eiginleikar tilviljunar

Það er erfitt að bera kennsl á tilviljun

Ef tilviljunarraðir eru skoðaðar sjást fljótt mynstur í þeim. Myndin til hliðar sýnir fern skýr mynstur þrátt fyrir að tilviljun hafi ráðið því hvernig teningarnir lentu.

Fólk hafnar oft tilviljun sem skýringu því niðurstaðan virðist of regluleg. Þess vegna er oft leitað skýringa þegar engra skýringa er þörf.

Þegar reynt er að búa til tilviljunarraðir „viljandi“ verða þær of óreglulegar. Með því að forðast öll mynstur og endurtekningar hverfa þau mynstur sem verða fyrir tilviljun og niðurstaðan því ekki lengur tilviljunarröð.



Mynstrin verða til fyrir tilviljun

# Eiginleikar líkinda

- Líkindi eru á bilinu 0,0–1,0
  - 0,0 samsvarar engum og 1,0 fullkomnum líkindum
- Summa líkinda allra möguleika er 1,0
  - $P(\text{kona}) + P(\text{karl}) = 1,0$
- Ef atburðum er skipt í ósamrýmanleg mengi, er summa líkinda fyrir öll mengin jafnt og 1,0
  - $P(\text{bókasafnsfr.}) + P(\text{félagsfr.}) + \dots + P(\text{mannfr.}) = 1,0$ ;  
þar sem aðeins er hægt að hafa eina aðalgrein

# Samlagningarreglan

$$P(\text{Atburður A eða B}) = P(\text{Atburður A}) + P(\text{Atburður B})$$

$$P(2 \text{ eða } 4) = P(2) + P(4) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

- Líkindi tveggja eða fleiri ósamrýmanlegra (*mutually exclusive*) atburða
  - Líkur þess að „2“ eða „4“ komi upp á teningi
  - Líkur þess að nemandi sé úr mannfræði eða félagsfræði
- Ósamrýmanlegir atburðir geta ekki orðið báðir í einu



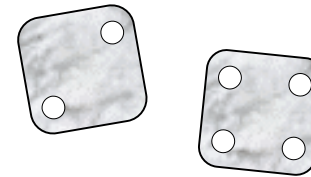
# Almenna samlagningarreglan

$$P(\text{Atburður A eða B}) = P(\text{Atburður A}) + P(\text{Atburður B}) - P(\text{A og B})$$

- Notuð þegar atburðir eru ekki ósamrýmanlegir
  - Hverjar eru líkur þess að vera í mannfræði *eða* vera kona
  - Um 6% nema eru í mannfræði og um 60% (?) eru konur
  - Með samlagningu fáum við 66% sem stenst ekki því sumir mannfræðinemar eru konur og því ekki ósamrýmanlegir atburðir
  - Við þyrftum því að draga frá líkindi þess að vera í mannfræði og kona til að reikna réttar líkur

# Að fá 2 í fyrsta eða 4 í seinna kasti

- Ekki ósamrýmanlegir atburðir



- $P(2 \text{ í fyrsta}) = 1/6$
- $P(4 \text{ í öðru}) = 1/6$
- $P(\text{fyrst } 2 \text{ eða svo } 4) \neq 2/6$

- Líkurnar  $P(2 \text{ og } 4)$  eru tvítaldar, sbr. mynd

- Réttar líkur eru  $1/6 + 1/6 - 1/36 = 11/36$

		Fyrri kast						
		1	2	3	4	5	6	Samtals
S e i n n a	1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
Samtals		1/6	1/6	1/6	1/6	1/6	1/6	1/1

# Kennari í félagsvísindadeild eða prófessor

Háskóladeild	Starfsheiti			Samtals
	Lektor	Dósent	Prófessor	
Guðfræðideild	0,0	0,3	2,2	2,5
Læknadeild	3,1	13,2	7,5	23,9
Lagadeild	0,0	0,6	2,5	3,1
Viðsk. & hagfræðideild	1,6	2,8	2,5	6,9
Heimspékideild	4,4	6,9	7,5	18,9
Tannlæknadeild	1,9	1,3	1,3	4,4
Verkfræðideild	0,9	1,6	5,0	7,5
Raunvísindadeild	1,9	9,1	11,6	22,6
Félagsvísindadeild	2,2	3,8	4,1	10,1
Samtals	16,0	39,6	44,3	100,0

Við viljum ekki telja þennan reit tvisvar  
 $P(\text{fél. eða prófessor}) = 0,443 + 0,101 - 0,041 = 0,503$

# Margföldunarreglan

- Líkindi *óháðra* samrýmanlegra atburða
  - Líkur þess að fá fyrst 2 og síðan 4 í tveimur teningsköstum
  - Teningsköstin eru *óháð*; hið fyrra hefur ekki áhrif á það seinna
  - Því má margfalda líkindin saman;  
 $P(2 \text{ og síðan } 4) = 1/6 \cdot 1/6 = 1/36 \approx 2,8\%$

$$P(\text{Atburður A og B}) = P(\text{Atburður A}) \cdot P(\text{Atburður B})$$

# Að fá 2 og síðan 4

Fjöldi kasta	Tíðni „2“ og síðan „4“	%
1 0	1	10,0
1 0 0	5	5,0
1 0 0 0	3 2	3,2
1 0 0 0 0	2 5 5	2,6
1 0 0 0 0 0	2 8 7 4	2,9

- Rauntíðnin nálgast fræðitíðnina eftir því sem fjöldi kasta eykst

# Háðir og óháðir atburðir

- Margföldunarreglan á aðeins við óháða atburði
  - Tenings- og krónuköst eru innbyrðis óháð
- Fjöldamargir atburðir eru *háðir* en ekki óháðir
  - Spil dregin úr stökk *án skila*
  - Val í úrtak án hreinnar tilviljunaraðferðar
- Óháð val er mikilvæg forsenda margs konar ályktana í aðferða- og tölfræði

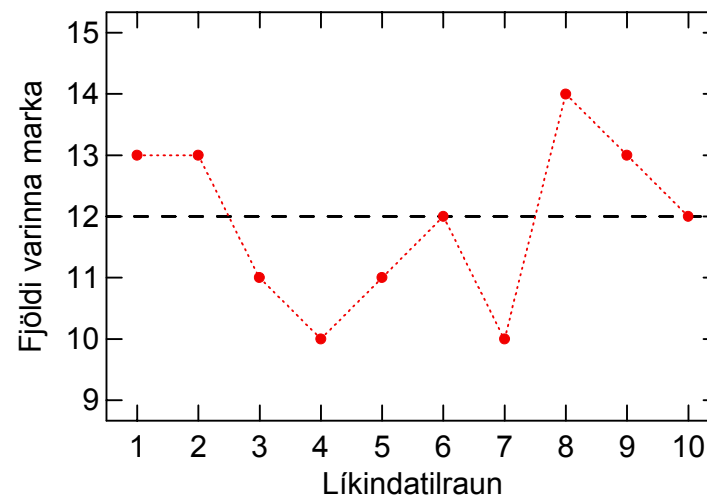
# Líkindabreytur

Líkindabreyta fær gildi úr líkindatíraun

Niðurstöður ýmissa fyrirbæra hafa eiginleika líkindabreyta (*random variable*; niðurstaðan ræðst af tilviljun og má lýsa með líkindum.

Gerum t.d. ráð fyrir því að markmaður í handbolta verji að jafnaði 40% af öllum skotum að marki. Við vitum ekki hver markskotanna verða varin, né heldur hve mörg hann muni verja í röð heldur aðeins að til langs tíma litið mun hann verja 40% allra skota.

Ef við gerum ráð fyrir 30 markskotum í leik og að varin skot séu innbyrðis óháð, fást niðurstöðurnar hér til hliðar.



Varin skot breytast tilviljunarbundið frá einni tíraun til annarrar. Við getum ekki spáð fyrir um varin skot en vitum að þau sveiflast í kringum 12, þ.e. 40% markskota.

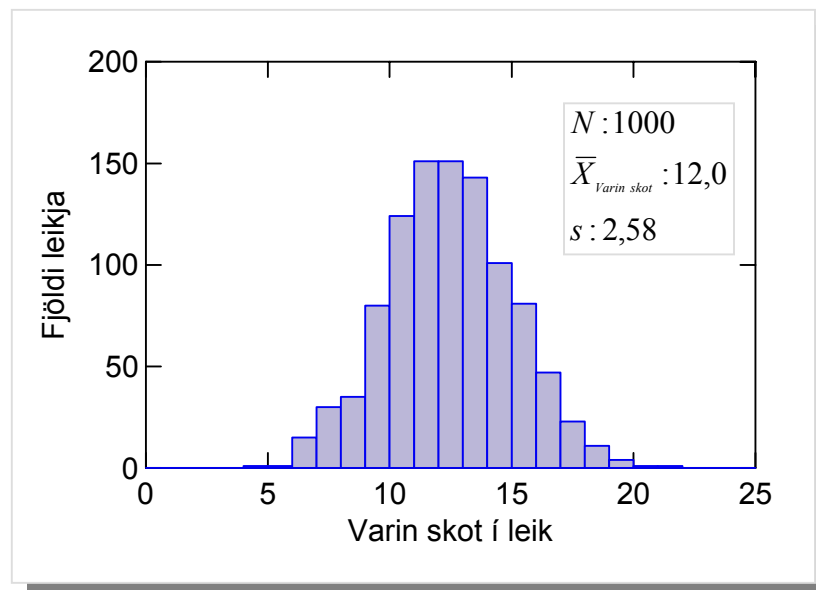
# Líkindabreyta sem líkan af ferli

Ef niðurstaða er líkindabreyta getum við lýst niðurstöðunni með líkindalíkani

Við vitum ekki hve mörg skot markmaðurinn ver í næsta leik, en við vitum hvernig varin skot ættu að dreifast ef líkanið er rétt.

Hér til hliðar er líkan af fjölda varinna skota miðað við að það sé líkindabreyta, hlutfall varinna skota sé 0,4 og varin skot séu innbyrðis óháðir atburðir.

Í því felst m.a. að markmaðurinn á hvorki góðan né slæman leik, þ.e.ferlið sem liggur til grundvallar frammistöðunni er alltaf hið sama; það breytist ekki milli leikja, fer ekki eftir því hve góð vörnin er, o.s.frv.



Niðurstaðan segir til um líklegan fjölda varinna skora í hverjum leik. Ef líkanið er rétt, ættu varin skot í leik að dreifast eins og myndin sýnir.



# Markvarsla á Evrópumeistaramóti

Er markvarslan í samræmi við líkanið?

Í leiknum við Slóvena á Evrópumeistaramóti árið 2002 varði Guðmundur Hrafnkelsson 19 skot af 44 eða 43%. Þetta var besta frammi- staða hans á mótinu. Á móti Dönum varði hann aðeins 8 skot af 36 eða 22%. Samtals varði hann 86 af 261 skoti eða að jafnaði 33% á mótinu öllu.

Efri myndin sýnir dreifingu varinna skota samkvæmt líkindalíkani miðað við 44 skot (leikurinn við Slóvena) og neðri myndin miðað við 36 skot (leikurinn við Dani).

Í ljós kemur að það er ekkert óvenjulegt við þessa leiki, miðað við 33% markvörslu að jafnaði getur markvarslan farið niður í 22% og upp í 43% fyrir tilviljun eina.

