Hagnýting Markov líkana

Helgi Tómasson, helgito@hi.is
Hagfræðideild HÍ
Málstofa Hagfræðideildar HÍ
Reykjavík, 19. nóvember, 2010.

Skipulag fyrirlestrar

Hvað er Markov líkan?
Lauslegt upprifjun á helstu eiginleikum
Ýmsar hagnýtingar
Einfalt sýnidæmi tekjudreifingu
Aðeins flóknara tekjudreifingardæmi
Notkun í survival analysis
Hermunardæmi um lífslíkur á Íslandi í dag
Hjónabandslíkan
Lokaorð

Um Markov líkön

- Hagnýting líkindafræði er að búa til reglur um óvissuna. Vísindin um óvissuna eru „exact" vísindi.
- Á 19. öld voru vísindamenn að leita að „Diffurjöfnunni" sem átti að lýsa hreyfilögmáli í tíma (dynamics) fyrirbæra í náttúrunni.
- Markov hugmyndin gengur út á að tengja líkindafræði og hreyfilögmál í tíma.
- Deterministískt hreyfilögmál getur t.d. verið á forminu.

$$X_{t+1} = g(X_t, X_{t-1}, \ldots, t)$$

- Ef hægt er að draga saman allar fortíða upplýsingar í eina stærð,

$$X_{t+1} = g(X_t, t), \text{ eða time-homogeneous } X_{t+1} = g(X_t).$$

Þá er sagt að lýsa megi ástandinu (state) með (X_t, t).

- Vil búa til líkindafræðilega útgáfu.
- Hugsum okkur einfalt dæmi. X_t er hending (random breyta) sem bara tekur gildi 0 eða 1.

$$P(X_{t+1}|X_t, X_{t-1}, \ldots) = P(X_{t+1}|X_t),$$

þ.e. líkindadreifing X_{t+1} gefin öll fortíð er bara háð X_t.

- Segjum að:

$$P(X_{t+1} = 1|X_t = 1) = p \text{ og } P(X_{t+1} = 0|X_t = 1) = 1 - p,$$
$$P(X_{t+1} = 1|X_t = 0) = 1 - q \text{ og } P(X_{t+1} = 0|X_t = 0) = q.$$

- Ástandið á tíma t er því aðeins gildið á X_t. T.d. gæti $X_t = 0$ þýtt lág laun og $X_t = 1$ há laun.

- Eiginleikum hreyfilögmálsins (milli launaflokka) má þá lýsa með fylkinu (transition-matrix=tilfærslufylki):

$$T = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}$$

- Í þessu tilfelli er til jafnvægisdreifing. Þ.e. líkindadreifing sem fylkið T heldur fastri. Einföld algebra gefur að í jafnvægi þá er:

$$P(X_t = 1) = \frac{1-q}{(1-q)+(1-p)} \qquad \text{og}$$
$$P(X_t = 0) = \frac{1-p}{(1-q)+(1-p)}.$$

- Ef t.d. $p = 1$ þá verða allir að lokum með há laun. Ef jafnframt allir byrja með lág laun og q er nálgægt 1 þá verður aðlögun að jafnvægi hæg. Hér ætla ég að hugsa mér L=lágnmarkslaun, Laun=$L(1 + X_t)$. Þ.e, laun byrji í 1.

- Reikna má væntanlegt gildi:

$$E(X_{t+1}|X_t = 1) = p, \quad E(X_{t+1}|X_t = 0) = 1 - q.$$

- Því gildir að:

$$E(X_{t+1}|X_t) = pX_t + (1-q)(1-X_t) =$$
$$1 - q + (p + q - 1)X_t = \mu + \phi X_t.$$

- Þetta er stationary AR(1) ferli. Jafnvægisdreifing er Bernoulli:

$$E(X_t) = \frac{\mu}{1-\phi} = \frac{1-q}{2-p-q} = \pi_H$$
$$V(X_t) = \pi_H(1-\pi_H).$$

- Ef AR(1) Bernoulli líkanið:

$$X_{t+1} = E(X_{t+1}|X_t) + \varepsilon_{t+1} = \mu + \phi X_t + \varepsilon_{t+1}$$

er umskrifað:

$$\begin{aligned} & X_{t+1} - X_t = \mu + \phi X_t - X_t + \varepsilon_{t+1} = \\ & \Delta X_{t+1} = \mu + (\phi - 1)X_t + \varepsilon_{t+1} = \\ & \kappa(\pi_H - X_t) + \varepsilon_{t+1}. \end{aligned}$$

- Hér er $\kappa = \mu/(\phi - 1)$ aðlögunarhraði (klifurhraði) að langtímameðaltalinu $\pi_H = \mu/(1 - \phi) = \frac{1-q}{2-p+q}$.

- Hreyfilögmálinu milli lágra (1) og hárra (2) launa má því lýsa með:

  $$\Delta X_{t+1} = \kappa(\pi_H - X_t) + \varepsilon_{t+1}.$$

  Ef dt tímaeiningar eru á milli mælinga þá er breyting á tímaeiningu:

  $$\Delta X_{t+dt}/dt = \kappa(\pi_H - X_t) + \varepsilon_{t+dt}$$

  Hliðstæða í samfelldum tíma er stochastic diffurjafna (SDE):

  $$dX(t) = \kappa(\beta - X(t))dt + \text{"noise"}$$

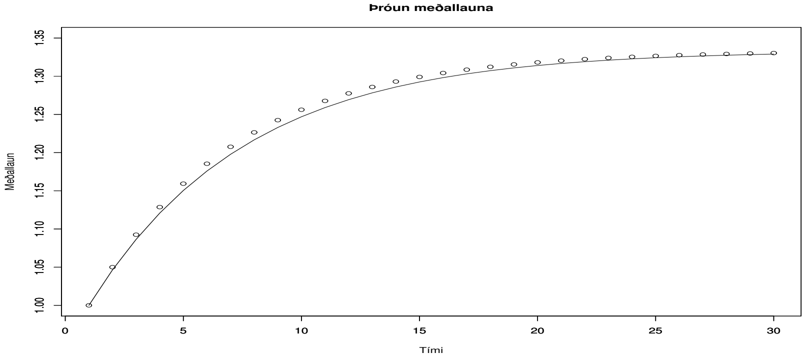
  Þar sem væntanlegt gildi á "noise" er 0. Hliðstæð deterministísk diffurjafna er

  $$dX = \kappa(\beta - X(t)).$$

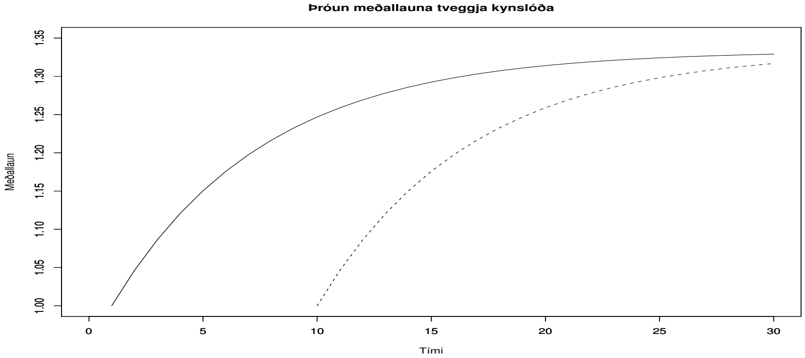
  Ef $\kappa > 0$, $\beta = \pi_H$, $X(0) = 0$. Þá er lausn diffurjöfnunnar:

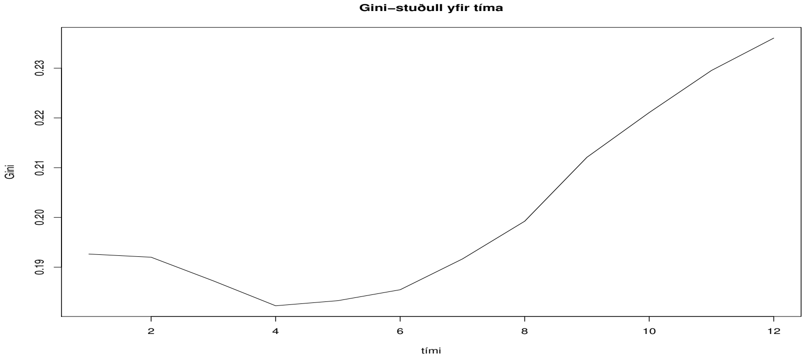
  $$X(t) = \pi_H(1 - exp(-\kappa t)).$$

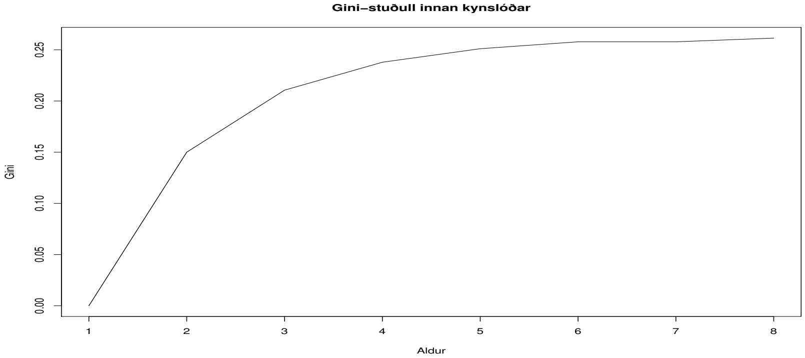
  Launin á tíma t eru t.d. 1+X(t)



Mynd: Þróun meðallauna. Hér er $\kappa = 0.15$ sem þýðir að helmingunar tími fráviks frá langtíma jafnvægi er c.a 4.6 tímabil.


Þróun meðallauna tveggja kynslóða
Meðallaun
1.00
1.05
1.10
1.15
1.20
1.25
1.30
1.35
0
5
10
15
20
25
30
Tími


Gini−stuðull yfir tíma
Gini
0.19
0.20
0.21
0.22
0.23
2
4
6
8
10
12
tími


Gini−stuðull innan kynslóðar
Gini
0.00
0.05
0.10
0.15
0.20
0.25
1
2
3
4
5
6
7
8
Aldur

- Champernowne (1953) gerði fræga grein byggða á þessari hugmynd þar sem hann lýsir raunhæfu hreyfilögmáli og sem leiðir af sér að tekjudreifing sé Pareto-dreifð. Margir höfundar höfðu áður undrast hvað tekjudreifingar hefðu líkt form.
- Whittle (1992) rekur í kennslubók (bls. 185) dæmi sem hann og Hermann Wold hönnuðu í grein (Wold & Whittle, 1957). Dæmið gengur út á að einstaklingar hafi einungis fasta prósentu í fjármagnstekjur og að líftíminn sé exponential dreifður (konstant lífshætta). Allir eignast m afkomendur, hver verður tekjudreifinging? Svar: Pareto dreifing.
- Í næsta dæmi á eftir er aukaóvissu bætt við, þ.e. fjöldi afkomenda má vera breytilegur. Þá fæst dreifing með Pareto-tail (heavy-tail).

- Hvað eru heavy-tail og mortality?
- Heavy-tail þýðir að í dreififallið $F(x)$ hefur þann eiginleika að $(1 - F(x)) \propto x^{-\alpha}$ fyrir stór x. Öfgakennd gildi möguleg, sbr. Cauchy og Pareto dreifingar.
- Mortality-rate (eða bara rate) þýðir:

$$\lambda(x) = \frac{f(x)}{1 - F(x)}.$$

- Fyrir rússneska-rúllettu er mortality-rate fasti. Þ.e. ekkert þjálfu og ekkert slit.
- Fyrir exponential-dreifingu gildir að:

$$\lambda(x) = \frac{\lambda exp(-\lambda x)}{(1 - (1 - exp(-\lambda x))}$$

Nokkrar útgáfur af AR(1) líkönum

$$dX(t) = \kappa(\beta - X(t))dt + \sigma dW$$

Ornstein-Uhlenbeck, normaldreift (W=Wiener/Brownian)

$$dX(t) = \kappa(\beta - X(t))dt + \sigma\sqrt{X}dW$$

Square-root process, CIR, jafnvægisdreifing gamma

$$dX(t) = \kappa(\beta - X(t))dt + \sigma X dW$$

CEV, jafnvægisdreifing inverse-gamma, heavy-tail

$$dX(t) = \kappa(\beta - X(t))dt + \sigma dS$$

AR(1) drifið áfram með stable-dreifingu, heavy-tail

Nánari skoðun á hazard-fallinu

- Hazardfallið lýsir yfirvofandi (lífs)hættu.
- Tengsl dreififalls líftíma T, $P(T \leq t) = F(t)$, þéttifalls líftímans, $f(t)$ og hazard fallsins eru:

$$F(t) = 1 - exp^{-\int_0^t \lambda(s) ds}, \quad \lambda(t) = f(t)/(1 - F(t)).$$

- Árið 1825 sagði Gompertz að hazard-fall fyrir líftíma manna (sumir segja allra lífvera) sé:

$$\lambda(t) = \underbrace{\lambda_0}_{A} + \underbrace{\alpha_1 exp(\alpha_2 t)}_{B}$$

- Liður A er áhætta óháð aldri og liður B er slitáhætta.

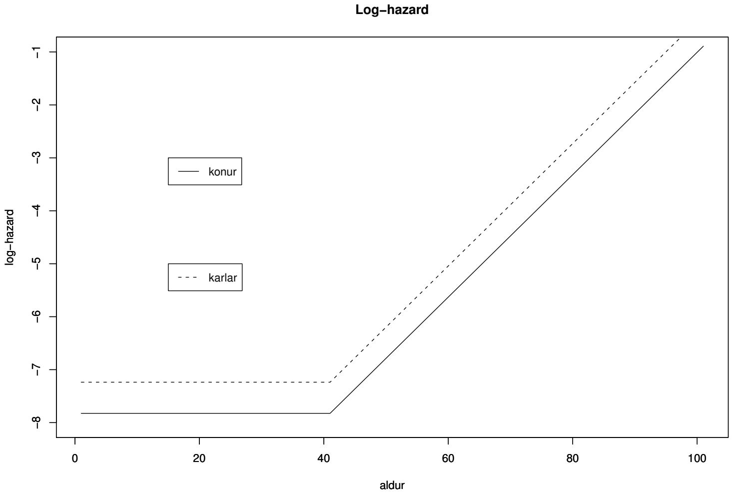
Einfalt hermunardæmi um líftíma

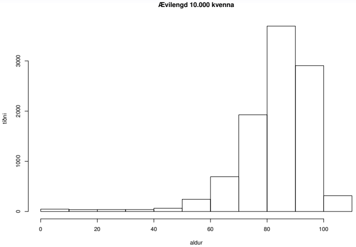
- Veit að væntanlega ævilengd er rúmlega 80 ár fyrir konur og tæplega 80 ár fyrir karlmenn.
- Veit að lífshætta karla er um það bil tvöföld lífshætta kvenna út ævina.
- Geri ráð fyrir fastri lífshættu fram yfir 40 ára aldur en síðan að lífshættan tvöfaldist á ákveðu árabili.
- Vel eftirfarandi fyrir konur (female-mortality):

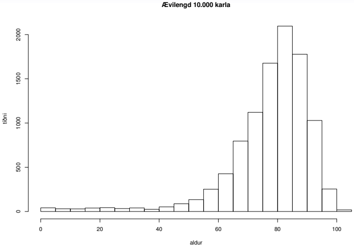
$$\lambda_{f,m}(t) = \begin{cases} \lambda_0 = 0.0004 & t < t0 = 40 \\ \lambda_0 2^{(t-t_0)/6} & t > t0 \end{cases}$$

- Fyrir karla, $\lambda_{m,m}(t) = 1.8\lambda_{f,m}(t)$.

Hazard-fall


Log−hazard
log−hazard
aldur
konur
karlar
−1
−2
−3
−4
−5
−6
−7
−8
0
20
40
60
80
100

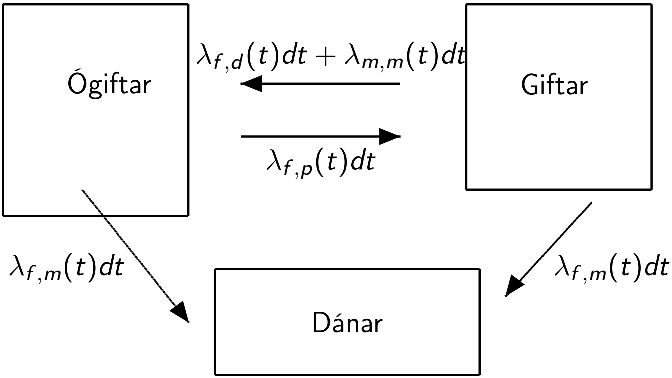




- Meðaldánaraldur kvenna 83 ár
- 2% kvenna deyr fyrir 48 ára aldur.

- Meðaldánaraldur karla er 78 ár.
- 4% karla deyr fyrir 48 ára aldur.

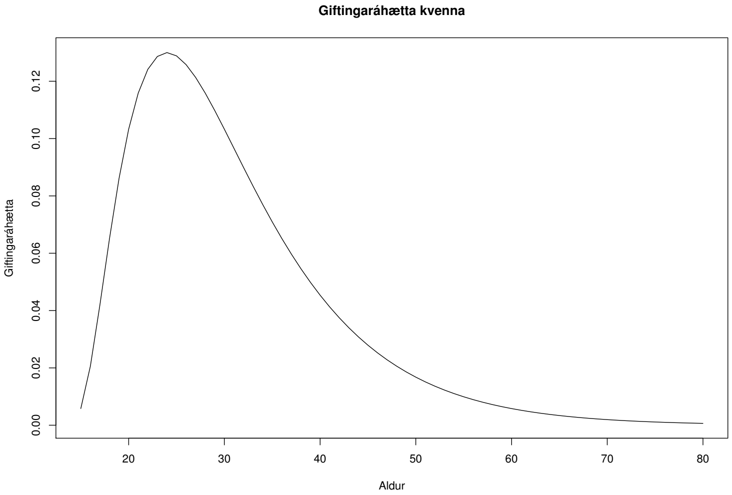
Hættur sem steðja að konum



Hvernig á að gera líkan fyrir hjónabandsáhættuna?

- Vandamál að það eru tvö kyn. Hvernig hegða kynin sér? Flestir leysa hlutina með one-sex nálgun.
- Lausn mín er að taka eitthver hazard fall fyrir konur. Síðan fyrir hverja konu para hana við „nálægan mann" með einhverri formúlu.
- Ég tók hazard-falla úr 40 ára tryggingafærðiglósum frá Kaupmannahöfn. (Beregningsgrundlag G-82)

$$\lambda_{f,p}(t) = 0.13 \times 10^{\frac{(t-24)^2}{20(t-12)}}$$

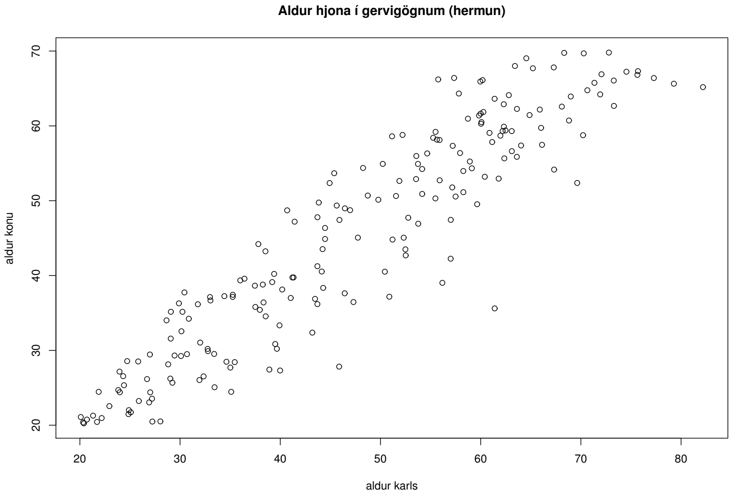

Giftingaráhætta kvenna
Giftingaráhætta
0.00
0.02
0.04
0.06
0.08
0.10
0.12
20
30
40
50
60
70
80
Aldur

Hvað nálægur maður?

- Í þessari tilraun var nálægur=nálægur í aldri.
- Ef gefið var að kona á aldri t_f gifti sig þá var aldur mannsins:

$$T_m = \alpha_0 + \alpha_1 t_f + \varepsilon.$$

- Gengið var út frá að staðalfrávik ε væri fall af t_f og að dreifingin væri tiltekin gammadreifing (til að fá mátulegt skewness). Giskað út frá þekktum gögnum.


Aldur hjona í gervigögnum (hermun)
70
60
50
40
30
20
aldur konu
20
30
40
50
60
70
80
aldur karls

- Í hjónabandi er karlinn oftar eldri.
- Hjónabandið má líta á sem efnahagsbandalag. Hjóna skipta með s sér verkum þannig að tekjur hámarkist. Þá er skynsamlegt að láta þann aðila sem er lengra kominn á aldurskúrfinni (með hærra kaup) vinna meira.
- Hugsalegt er að ógiftir einstaklingar þrói tekjur sínar með:

$$
\begin{aligned}
L(t) &= (1 + X(t)) \\
dX(t) &= \kappa(\beta - X(t))dt + \sigma X dW
\end{aligned}
$$

- Þegar hjónabandið hefst er líklegt að eldri einstaklingurinn sé með hærra X. Hjónin geta e.t.v. ákveðið að skipta þannig með sér verkum að β verði hærri og e.t.v einnig að klifurstuðullinn κ verði stærri. Lágt σ hátt κ gefa möguleika á mjög háum launm.

- Ef gefnir eru n aldursflokkar þá eru á hverju tíma er ástand mannfjöldans á eftirfarandi formi:

$$\boldsymbol{M}(t) = [M_1(t), \dots, M_n(t)] \quad \text{einhleypir karlar}$$
$$\boldsymbol{F}(t) = [F_1(t), \dots, F_n(t)] \quad \text{einhleypar konur}$$
$$\boldsymbol{MF}(t) = \begin{bmatrix} MF_{1,1}(t) & MF_{1,2} & \cdots & MF_{1,n}(t) \\ MF_{2,1}(t) & \cdots & \cdots & MF_{2,n}(t) \\ \vdots & \vdots & & \vdots \\ MF_{n,1}(t) & \cdots & \cdots & MF_{n,n} \end{bmatrix}$$

$MF_{i,j}(t)$=fjöldi hjóna á tíma t þar sem karl er i ára og kona er j ára.

- Eyjólfur Sigurðsson (2010) metur laun eftir hjónabandsstétt.

	Giftir	Ógiftir
Karlar	2.6	1.65
Konur	1.75	1.55

Tafla 1: Laun kynja eftir hjónabandsstétt.

Með því að láta κ og β vera mismunandi milli hjóna er hægt að fá svona meðaltöl út á ýmsa vegu.

Tengsl hjónabands og mortalitets

-

Tengsl hjónabands og mortalitets

- Til eru gögn (ca. 60-100 punktar) frá Ensku biskupakirkjunni og Enska landlæknisembættinu sem sýna fylgni milli mortalitets og markaðshlutdeildar kirkunnar í brúðkaupum ($p<0.05$).

Tengsl hjónabands og mortalitets

- Til eru gögn (ca. 60-100 punktar) frá Ensku biskupakirkjunni og Enska landlæknisembættinu sem sýna fylgni milli mortalitets og markaðshlutdeildar kirkunnar í brúðkaupum ($p<0.05$).
- Þetta eru vandaðar stofnanir með góð gögn

Tengsl hjónabands og mortalitets

- Til eru gögn (ca. 60-100 punktar) frá Ensku biskupakirkjunni og Enska landlæknisembættinu sem sýna fylgni milli mortalitets og markaðshlutdeildar kirkunnar í brúðkaupum ($p<0.05$).
- Þetta eru vandaðar stofnanir með góð gögn
- Útkoman frægt nonsense því ekki var notað tímaraðalíkan

Ekki slá saman misleitum hópum

Ekki slá saman misleitum hópum

	Há laun	Lág laun
Karlar	18	12
Konur	7	3

Tafla 2: Launadreifing í fyrirtæki A. 70% kvenna með há laun, 60% karla með há laun.

Ekki slá saman misleitum hópum

	Há laun	Lág laun
Karlar	18	12
Konur	7	3

Tafla 2: Launadreifing í fyrirtæki A. 70% kvenna með há laun, 60% karla með há laun.

	Há laun	Lág laun
Karlar	2	8
Konur	9	21

Tafla 3: Launadreifing í fyrirtæki B. 30% kvenna með há laun, 20% karla með há laun.

Ekki slá saman misleitum hópum

	Há laun	Lág laun
Karlar	18	12
Konur	7	3

Tafla 2: Launadreifing í fyrirtæki A. 70% kvenna með há laun, 60% karla með há laun.

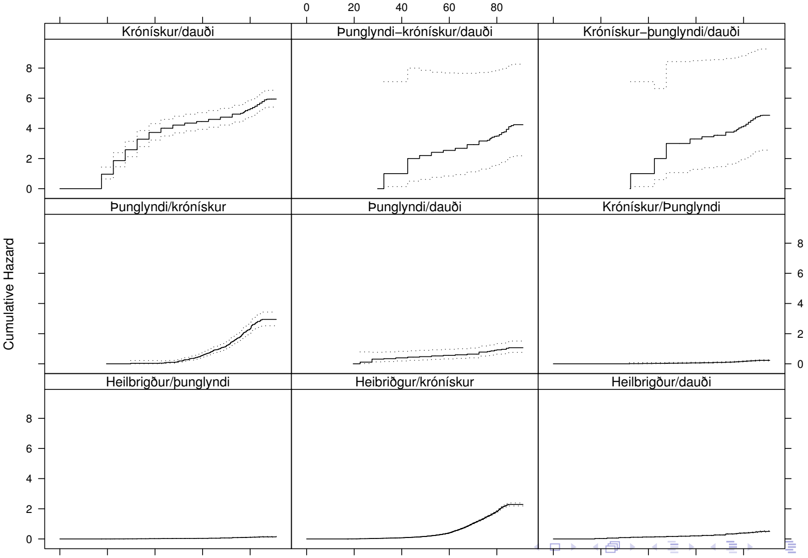
	Há laun	Lág laun
Karlar	2	8
Konur	9	21

Tafla 3: Launadreifing í fyrirtæki B. 30% kvenna með há laun, 20% karla með há laun.

	Há laun	Lág laun
Karlar	20	20
Konur	16	24

Tafla 4: Launadreifing í fyrirtækjum A+B. 40% kvenna með há laun, 50% karla með há laun.

Áhættur milli margra flokka



$$\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} - & h_{01} & h_{02} & 0 & 0 & h_{05} \\ 0 & - & 0 & h_{13} & 0 & h_{15} \\ 0 & 0 & - & 0 & h_{24} & h_{25} \\ 0 & 0 & 0 & - & 0 & h_{35} \\ 0 & 0 & 0 & 0 & - & h_{45} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

0 = heilbrigður

1 = þunglyndi

2 = krónsískur sjúkdómur

3 = fyrst 1 og síðan króniskur

4 = fyrst 2 og síðan þunglyndi

5 = dauði

Lokaorð

- Markov líkön eru stór flokkur líkana.
- Við dýnamísk vandamál á að nota tímaraðalíkön.
- Í bayesískri tölfræði gagnast Markov-keðjur við hermanir á flóknum „a posteriori" dreifingum. Ég ætlaði að sýna „line"-dæmið (regression með 5 punktum) en ákvað að ekki væri pláss.

Champernowne, D. (1953). A model of income distribution. *Economic Journal*, *63*, 318–351.

Whittle, P. (1992). *Probability via expectation* (Third ed.). Springer-Verlag New York Inc,.

Wold, H. & Whittle, P. (1957). A model explaining the pareto distribution of wealth. *Econometrica*, *25(4)*, 591–595.