

Statisztikai tanulás az idegrendszerben, 2015.

Valószínűségi modellek

Bányai Mihály

banyai.mihaly@wigner.mta.hu

<http://golab.wigner.mta.hu/people/mihaly-banyai/>

- Hogyan kezeljük formálisan a bizonytalan tudást?
- A valószínűségi kalkulus alapelemei
- Hogyan építsünk és kezeljünk modelleket?
- Egy eloszlás információtartalma

He deals the cards to find the answer
The sacred geometry of chance
The hidden law of a probable outcome
The numbers lead a dance
Sting - Shape Of My Heart

To understand God's thoughts we
must study statistics, for these are
the measure of His purpose.
Florence Nightingale

- **Hogyan kezeljük formálisan a bizonytalan tudást?**
- A valószínűségi kalkulus alapelemei
- Hogyan építsünk és kezeljünk modelleket?
- Egy eloszlás információtartalma

Tudás reprezentációja

- Az idegrendszer a környezetéről gyűjtött információkat mentális modellbe tömöríti
- A mentális modell leírásához formális tudásreprezentációra van szükség
- A reprezentációnak alkalmasnak kell lennie következtetésre és a környezet generatív modelljeként való használatra (elképzeltek helyzetek)
- Szimbolikus rendszerek (klasszikus MI)
 - rugalmas szintaxis
 - egymásba ágyazható állítások
- Numerikus / statisztikai rendszerek (machine learning, data science)
 - általánosítási képesség
 - hatékony algoritmusok



Logikai rendszerek

- Formális nyelvek
 - szintaxis: formai szabályok állítások létrehozására
 - szemantika (igazságérték-szemantika): igaz-e egy állítás?
- például: elsőrendű logika vagy prédikátumkalkulus
 - a világ objektumainak megfelelő értékű változók, amelyekhez igazságértéket prédikátumfüggvényekkel rendelhetünk
 - jól formalizálhatók vele természetes nyelvi mondatok:
“Aki nem ismeri Jaynest, meg fog bukni.”
 $\forall x: \text{Diák}(x) \wedge \neg \text{Ismer}(x, \text{JAYNES}) \rightarrow \neg \text{Nagyobb}(\text{vizsgajegy}(x), 1)$

A tudás reprezentációja logikában

- Axiómák (tudásbázis) - állítások halmaza, amiket szemantikailag igaznak tekintünk
- Egy állítás tétel (=igaz), ha logikai következménye az axiómáknak, és hamis, ha ellentmondást képez velük
 - azaz helyes következtetési szabály alkalmazásával előállítható belőlük
 - $T = \{\text{Minden görög tógát hord. Szókratész görög.}\}$, $A = \{\text{Szókratész tógát hord.}\}$
- ha egyik sem bizonyítható, nem mondhatunk semmit a logikai rendszeren belül
 - mit mondhatunk intuitíve arról, hogy $A = \{\text{Kovács Akhilleusz tógát hord.}\}$?
 - Hogyan befolyásolja ezt a tudásbázis bináris tartalma?

Kiterjesztés bizonytalanságra

- kvantifikáljuk, hogy mennyire tartjuk hihetőnek, hogy egy A állítás tétel egy T tudásbázis mellett, ha nem tudjuk se bizonyítani, se cáfolni: $H(A | T)$
- Keynes, 1921, Cox, 1961, Jaynes, 2003, Dupré & Tipler, 2009
- Kívánságlista
 - valós számmal ábrázoljuk a hihetőséget
 - konzisztencia: ugyanabból a rendelkezésre álló információból mindig ugyanaz a hihetőség jön ki, akárhogy számolunk
 - a változások iránya stimmeljen: ha $H(A | T')$ nő, akkor $H(A \wedge B | T')$ is nőjön, $H(\neg A | T')$ pedig csökkenjen
- Ha egy hihetőségi mérték megfelel ezeknek az elvárásoknak, akkor izomorf a feltételes valószínűségi mértékkel, $H(A | T) = \Pr(A | T)$, amit az összeg- és a szorzatszabály definiál
 - Igazából a fenti feltételek technikailag kiegészítésre szorulnak, ez csak az intuíció

Összegszabály

- Úgy döntünk, hogy 0 legyen az ellentmondás, 1 a tétel valószínűsége
- Komplementer állítások valószínűsége egyre összegződik

$$Pr(A \mid T) + Pr(\neg A \mid T) = 1$$

- Például változók numerikus értékéről szóló állításokra ($A: a=1, T = \{\}$)

$$Pr(a = 1) = \sum_i Pr(a = 1 \mid b = i)$$

Szoratszabály

$$Pr(A, B \mid T) = Pr(A \mid B, T)Pr(B \mid T)$$

- Feltételes valószínűség definíciója

$$Pr(A \mid B) = \frac{Pr(B, A)}{Pr(B)}$$

- Bayes-tétel

$$Pr(A \mid B) = \frac{Pr(B \mid A)Pr(A)}{Pr(B)}$$

Kolmogorov-axiómák

- Kell egy eseményalgebra
- kell egy mérték, ami az eseményalgebra elemein értelmezhető, ez lesz a valószínűség
- a valószínűségi mérték pozitív, egyre összegződik a teljes eseménytéren és egymást kölcsönösen kizáró események uniójának valószínűsége additív
- feltételes valószínűség definícióját fel kell venni axiómának
- Ekvivalens a logikai bevezetéssel

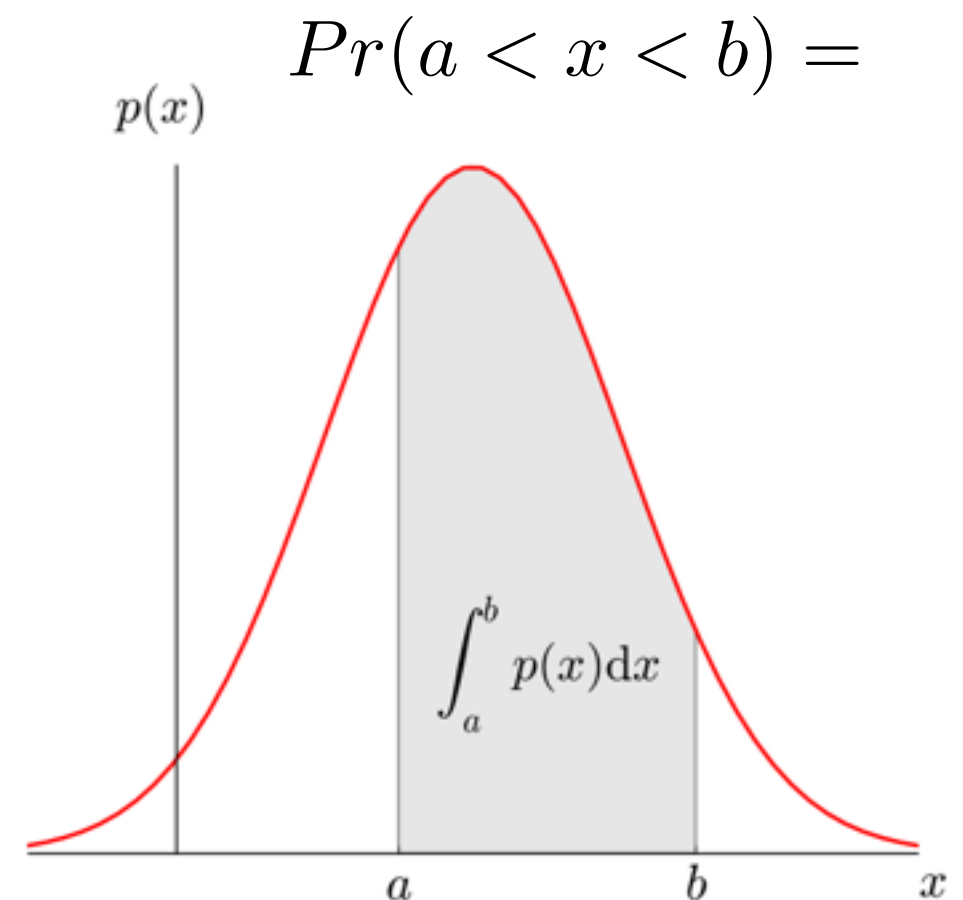
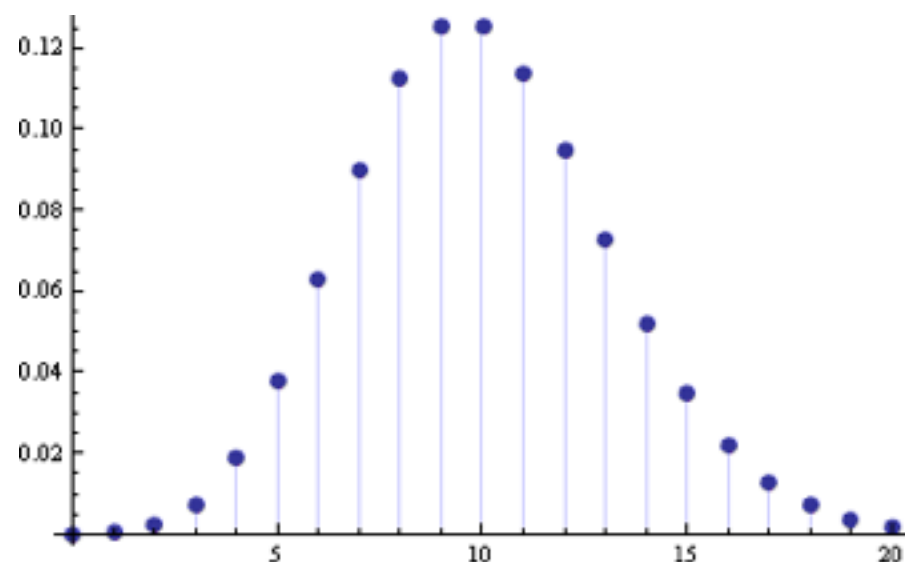
Valószínűség értelmezései

- Gyakoriság-alapú (frekventista)
 - a valószínűség csak ismételt kísérletekben értelmezhető, az adott kimenetel összes próbálkozáson belüli relatív frekvenciájaként
- Információ-alapú (Bayesian, Laplacian)
 - a valószínűség nem a vizsgált jelenség tulajdonsága, hanem arra jellemző, hogy egy megfigyelő milyen információkkal rendelkezik róla
- Szubjektív (de Finetti)
 - a valószínűség egyéni hiedelmeket reprezentál
- Logikai (objektív)

- Hogyan kezeljük formálisan a bizonytalan tudást?
- **A valószínűségi kalkulus alapelemei**
- Hogyan építsünk és kezeljünk modelleket?
- Egy eloszlás információtartalma

Diszkrét és folytonos valószínűségi változók

- Foglaljuk össze minden lehetséges érték valószínűségét egy függvényben
- Diszkrét értékkészletű változók
 - egyszerűen megadjuk az összes értéket
 - valószínűségi tömegfüggvény (pr. mass function)
- Folytonos értékkészletű változók
 - minden konkrét érték valószínűsége nulla
 - intervallumoknak van pozitív valószínűsége
 - sűrűségfüggvény (pr. density function)

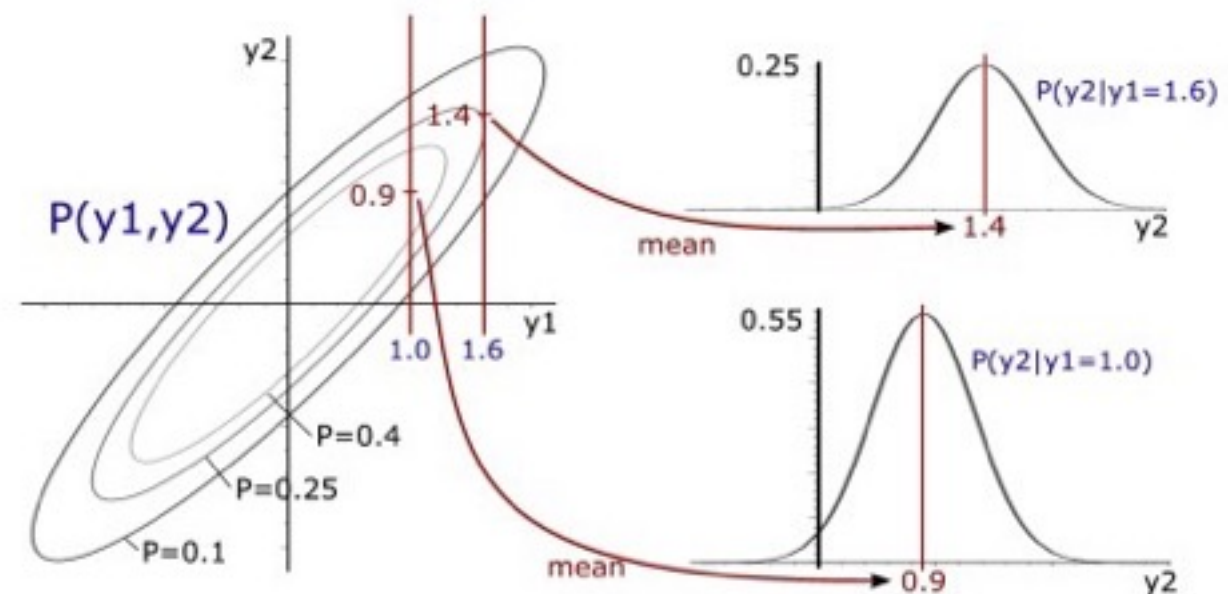


Együttes eloszlások

- valváltozó függvényének várható értéke: integrál a változó eloszlása szerint
- marginalizáció: $f(.)=1$ várható értéke a változók egy része szerint
- kondicionálás: a változók egy részének fixáljuk az értékét
- megszabadulás a feltételben lévő változótól: a feltételes eloszlás várható értéke a kérdéses változó eloszlása szerint

$$E(f(x)) = \int_{-\infty}^{\infty} f(x)p(x)dx$$

$$p(x) = \int_{-\infty}^{\infty} p(x, y)dy$$



$$p(x | z) = \int_{-\infty}^{\infty} p(x, y | z)dy$$

$$p(x | z) = \int_{-\infty}^{\infty} p(x | y, z)p(y | z)dy$$

Mindent p-vel jelölünk

- Az események valószínűségét én itt most Pr -rel
- Az összes különböző eloszlásfüggvény és sűrűségfüggvényt
 - $p(x)$, $p(y)$, $p(x,y)$, $p(x|y)$, $p(y|x)$, ...
 - igazából ezek teljesen különböző függvények, sokszor más az értelmezési tartományuk is
- Nagyon elvetemült emberek még mindenféle paramétereket is
- P-érték ...

Bayes-tétel sűrűségfüggvényekre

likelihood $\swarrow$ prior $\swarrow$

posterior $\longrightarrow$ $p(x \mid y, z) = \frac{p(y \mid x, z)p(x \mid z)}{p(y \mid z)}$

evidence vagy marginal likelihood $\nearrow$

$p(y \mid z) = \int_{-\infty}^{\infty} p(y \mid x, z)p(x \mid z)dx$

The diagram illustrates the components of Bayes' theorem. At the top, 'likelihood' and 'prior' are labeled with arrows pointing to the numerator of the posterior probability formula. The posterior probability formula is shown as $p(x \mid y, z) = \frac{p(y \mid x, z)p(x \mid z)}{p(y \mid z)}$. Below this, 'evidence vagy marginal likelihood' is labeled with an arrow pointing to the denominator of the formula. At the bottom, the integral formula for the evidence is shown: $p(y \mid z) = \int_{-\infty}^{\infty} p(y \mid x, z)p(x \mid z)dx$. An arrow points from this integral formula up to the denominator of the posterior probability formula.

Függetlenség

$$x \perp y \equiv p(x, y) = p(x)p(y)$$

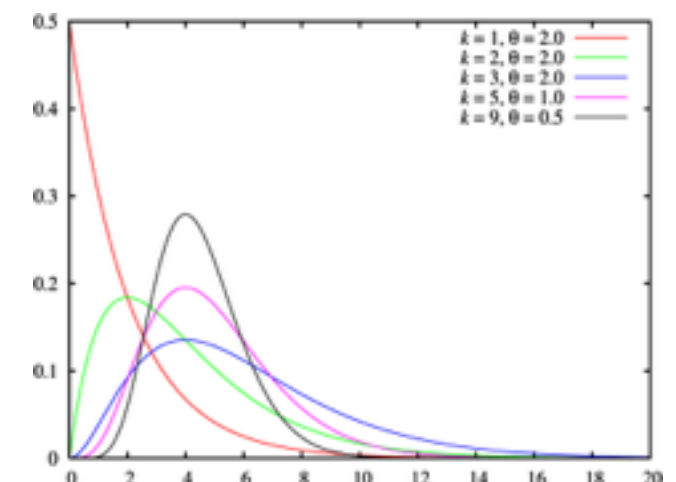
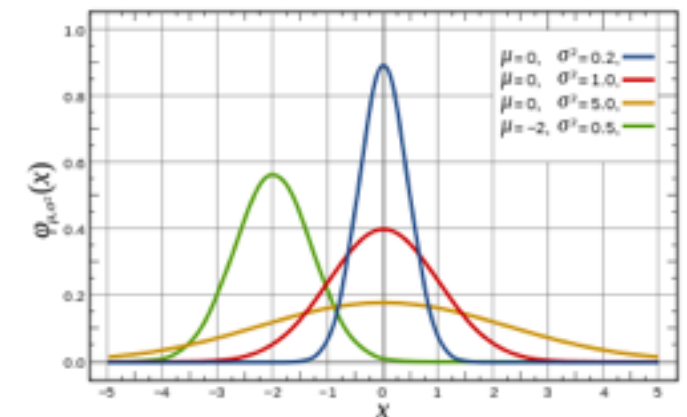
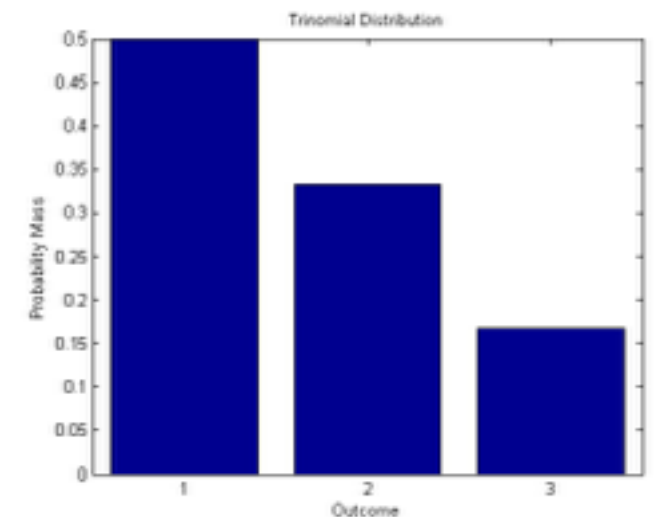
- feltételes függetlenség

$$x \perp y \mid z \equiv p(x, y \mid z) = p(x \mid z)p(y \mid z)$$

- nem implikálja egymást a marginális és feltételes függetlenség
- mikor ekvivalens a függetlenség azzal, hogy két valváltozó korrelációja nulla?

Parametrikus eloszláscsaládok

- Diszkrét értékű
 - Bernoulli: pénzfeldobás - érme aszimmetriája - $p(x) = \text{Ber}(x;\beta)$
 - multinomiális: kockadobás - cinkelés - $\text{Mult}(x;\beta_1 \dots \beta_{k-1})$
- Folytonos értékű
 - egyenletes - minimum, maximum - $U(x;\alpha, \beta)$
 - Gauss (normál): szimmetrikus - átlag, szórás (kovarianciamátrix) - $N(x;\mu, C)$
 - Gamma: pozitív - alak, skála - $\text{Gam}(x;k, \theta)$



- Hogyan kezeljük formálisan a bizonytalan tudást?
- A valószínűségi kalkulus alapelemei
- **Hogyan építsünk és kezeljünk modelleket?**
- Egy eloszlás információtartalma

Valószínűségi modell

- Együttes eloszlás faktorizációja

- n - teszt nehézsége
- q - diák IQ-ja
- f - felvételi pontszám
- s - teszten elért pontszám
- j - kapott jegy

$$p(n, q, f, s, j) = p(n \mid q, f, s, j)p(q \mid f, s, j)p(f \mid s, j)p(s \mid j)p(j)$$

$$p(n, q, f, s, j) = p(j \mid n, q, f, s)p(s \mid n, q, f)p(f \mid n, q)p(n \mid q)p(q)$$

- Függetlenség a változók között

$$n \perp \{q, f\}$$

$$s \perp f \mid q$$

$$j \perp \{n, q, f\} \mid s$$



$$p(n, q, f, s, j) = p(n \mid s, j)p(q \mid f, s)p(f \mid s)p(s \mid j)p(j)$$

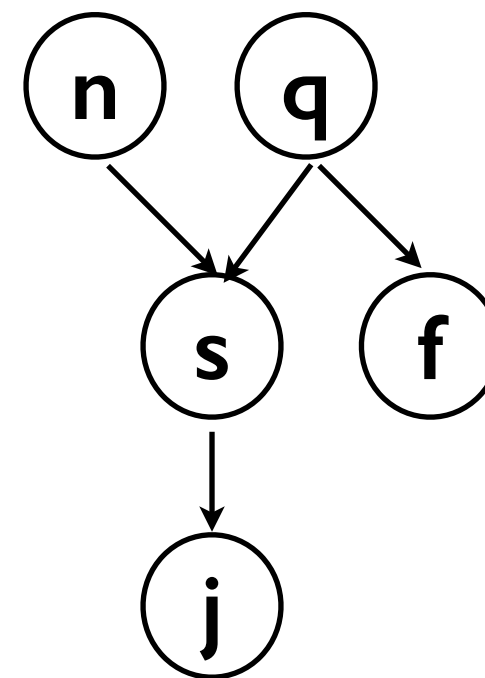
$$p(n, q, f, s, j) = p(j \mid s)p(s \mid n, q)p(f \mid q)p(n)p(q)$$

Jelölés gráfokkal

$$p(n, q, f, s, j) = p(j \mid s)p(s \mid n, q)p(f \mid q)p(n)p(q)$$

- n - teszt nehézsége
- q - diák IQ-ja
- f - felvételi pontszám
- s - teszten elért pontszám
- j - kapott jegy

- Directed Acyclic Graph
- A faktorizációban szereplő feltételes és marginális eloszlásokat kell megadni a modell teljes leírásához
- Kauzalitás nincs



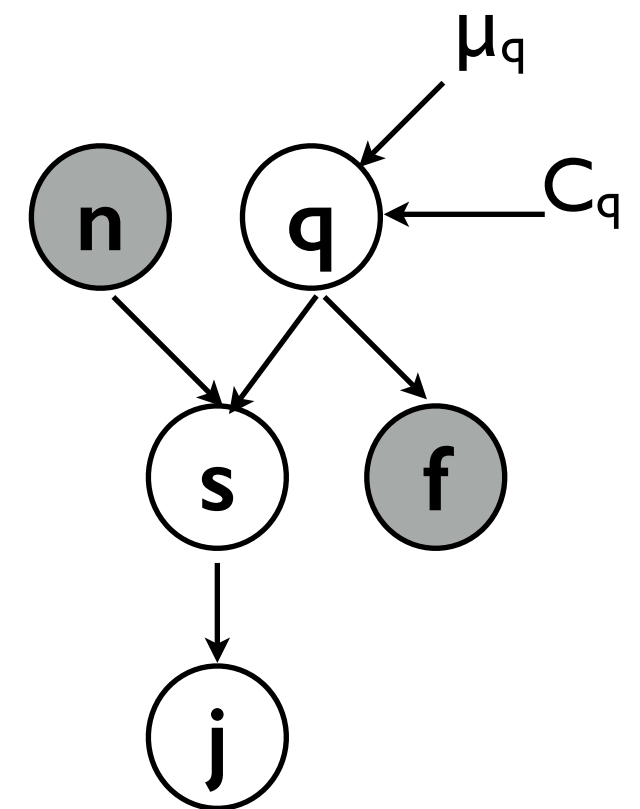
$$n \perp \{q, f\}$$

$$s \perp f \mid q$$

$$j \perp \{n, q, f\} \mid s$$

Megfigyelt és rejtett változók

- bizonyos változókról rendelkezünk adatokkal
 - jelölés: sötét kör
- a többi csak a jelenség struktúrájáról alkotott feltételezéseinket reprezentálja
 - látens, rejtett változók: üres kör
- a prediktív eloszlás: a megfigyelt változók marginális eloszlása
- eloszláscsaládok determinisztikus paraméterei
 - ugyanolyanok, mint a látens változók, csak nincs eloszlásuk, hanem egyetlen értékük
 - eloszlások feltételében jelenhetnek csak meg
 - kör nélküli betű

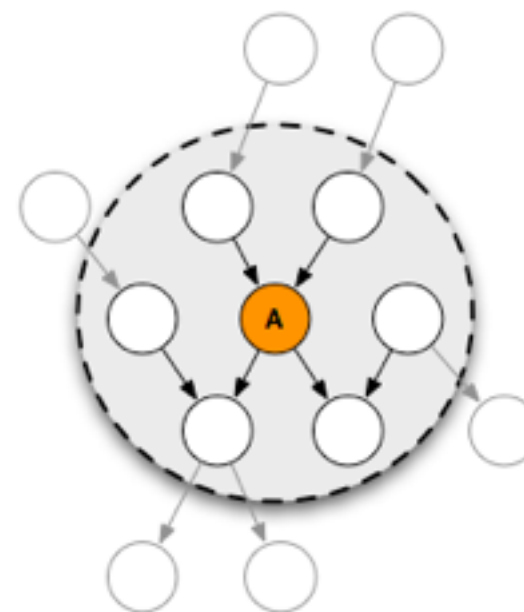
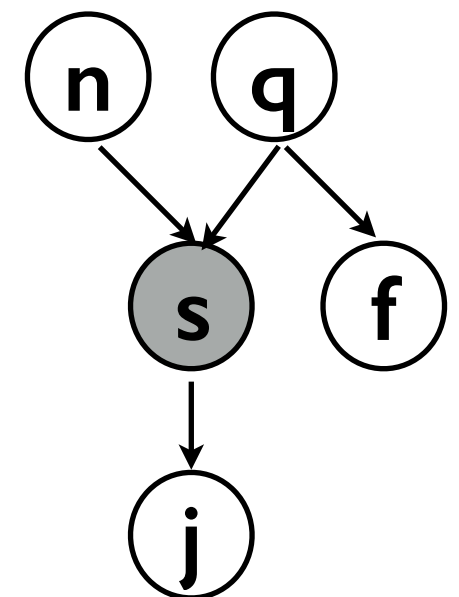


$$p(q) = \mathcal{N}(q; \mu_q, C_q)$$

Függetlenség grafikus modellben

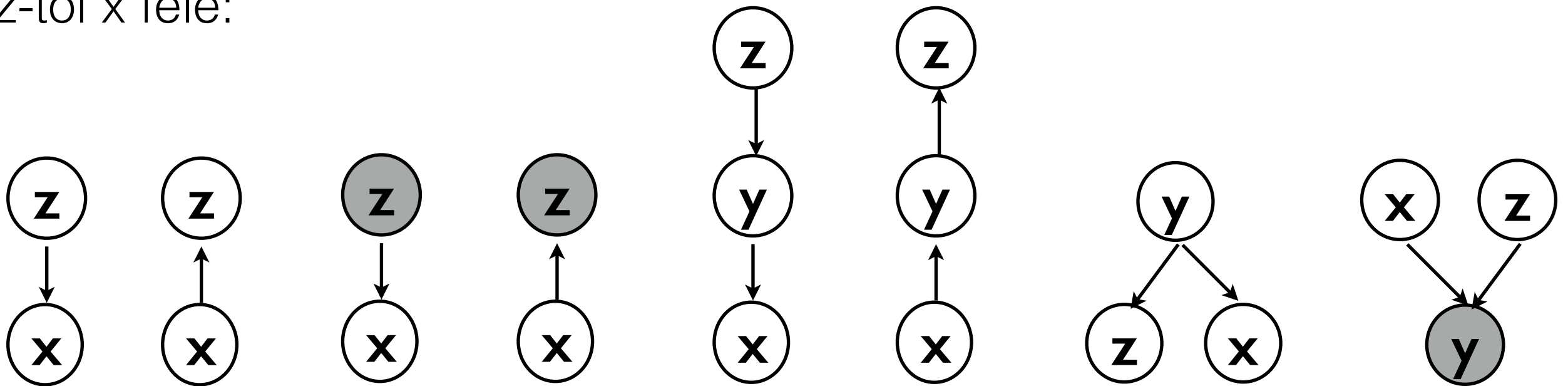
- Egy változó megfigyelése megváltoztathatja más változók függetlenségét
 - mivel a feltételes függetlenség nem ugyanaz, mint a marginális
 - Explaining away: egy konkrét pontszám mellett, ha tudjuk, hogy a teszt nagyon nehéz volt, megnő a valószínűsége a magas IQ-nak
- Markov-takaró: azon változók halmaza, amelyekre kondicionálva az összes többiől független a csúcs
 - minden megfigyeltségi állapot mellett
 - szülők, gyerekek, házastársak

- n - teszt nehézsége
- q - diák IQ-ja
- f - felvételi pontszám
- s - teszten elért pontszám
- j - kapott jegy

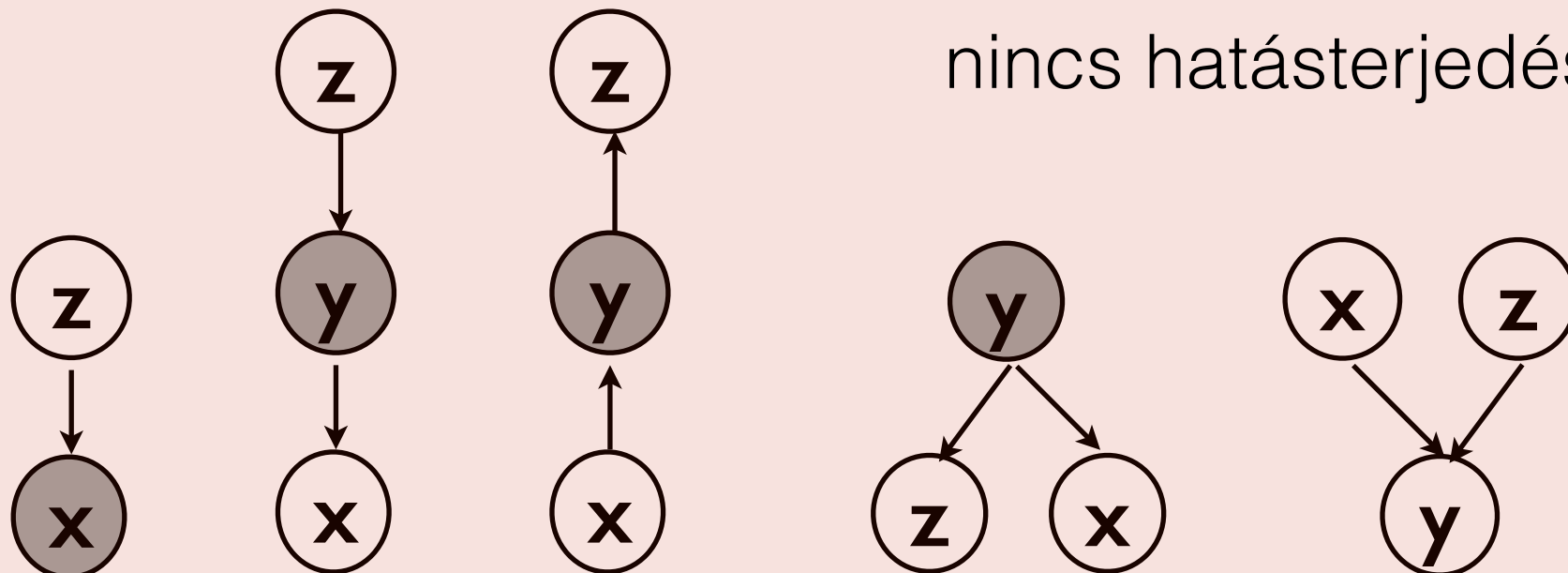


Valószínűségi hatás terjedése

z-től x felé:

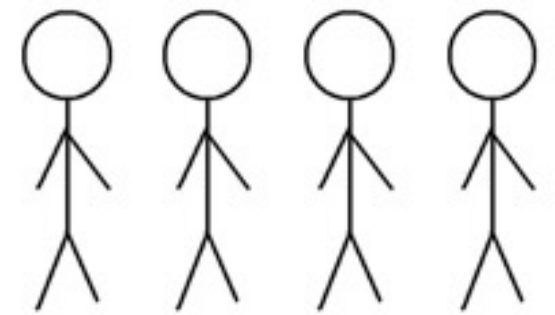


nincs hatásterjedés

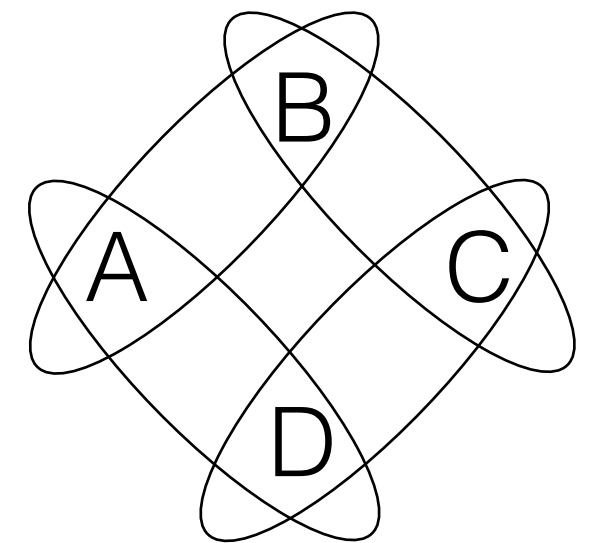


Minden eloszláshoz van gráf?

- Nem, tanulópárok problémája pl
- Ahol körkörös szimmetria van, ott a DAG felírásánál mindig ráerőltetünk egy irányítást, így nem lesz tökéletes a leképezés, elvesznek függetlenségi viszonyok
- Léteznek más, nem-DAG grafikus nevezéktanok



A B C D

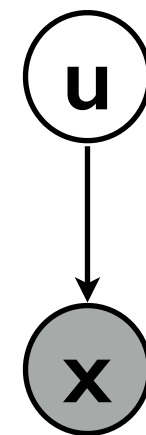
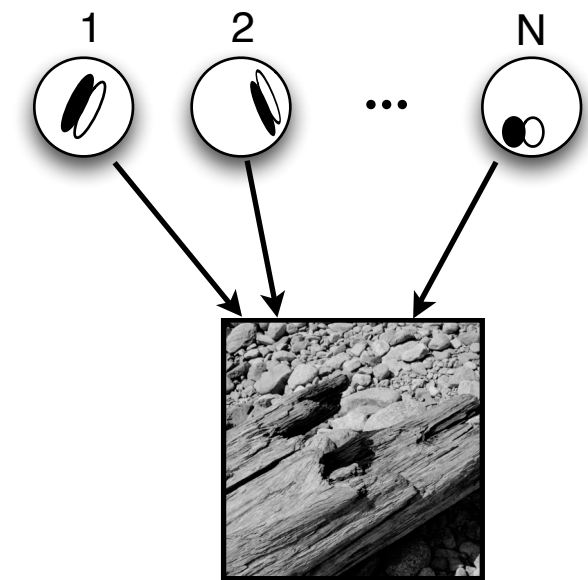


$A \perp C \mid \{B, D\}$

...

Valószínűségi modellek az idegrendszerben

- vizuális rendszer működésének megértéséhez használhatók generatív képmodellek
- a kép pixelei egyszerű grafikai elemeket reprezentáló feature-vektorok lineáris kombinációja, plusz megfigyelési zaj
- a feature-ökoefficiensei a rejtett változók, a pixelek a megfigyeltek
- percepció mint következtetés a modell rejtett változóira
- a rejtett változók prediktálhatják neuronok elektromos aktivitását
- vizuális szabályosságok tanulása a modell építése

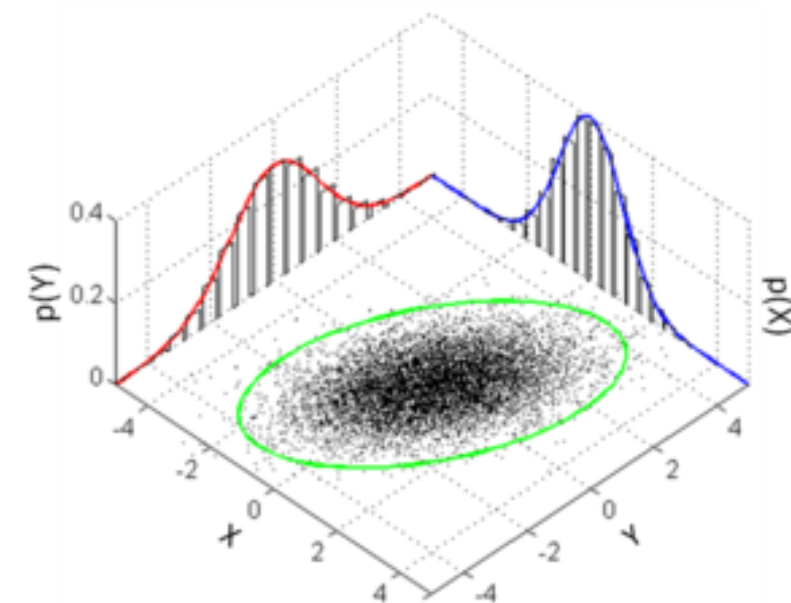
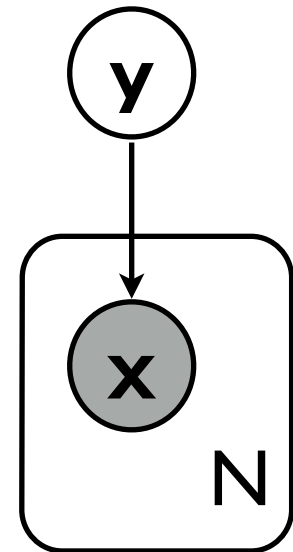


$$p(x|u) = \mathcal{N}(x; f_1 u_1 + \dots + f_N u_N, \sigma I)$$

$$p(u) = ?$$

Szintetikus adatok generálása

- Plate: ha egy változóból N darab i.i.d. példány van
- véletlenszámgenerálás
 - feltételezzük, hogy van egyenletes eloszlású pseudorandom számokat előállító algoritmusunk
 - az egyenletes eloszlású véletlenszámokat transzformáljuk a kvánt eloszlásra
- ancestral sampling
 - először a szülő nélküli változók prior (marginális) eloszlásából veszünk mintát
 - a fán lefelé lépkedve a feltételes eloszlásokból tudunk mintát venni
 - minden változónak a marginális eloszlásából lesznek mintáink, így a prediktív eloszlásból is
- az álom funkciója, Crick & Mitchinson, 1983

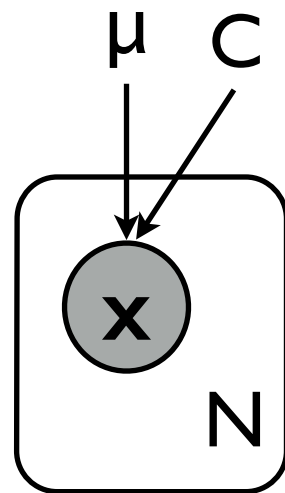
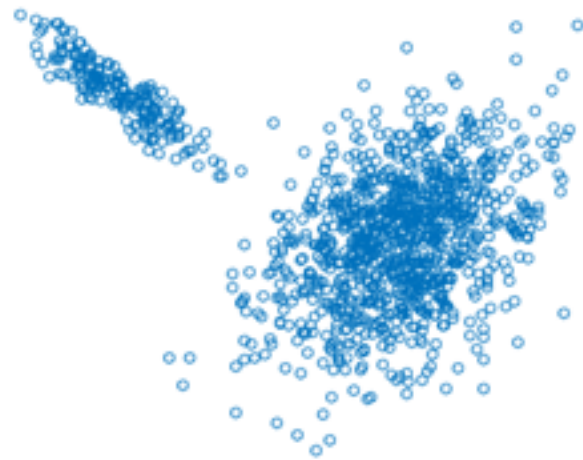


Modellek építése

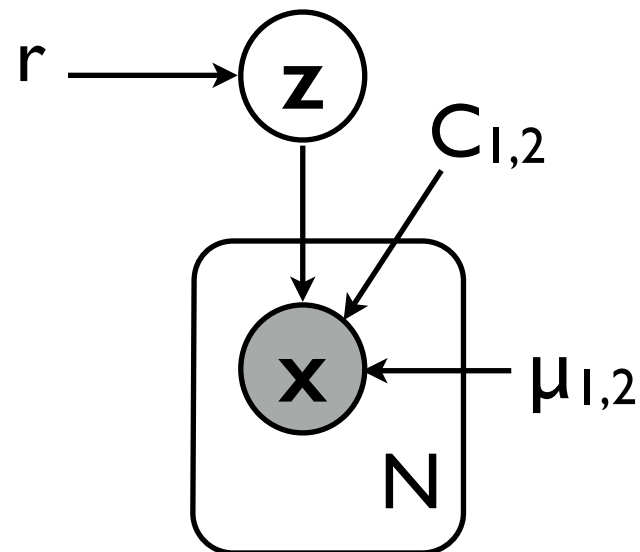
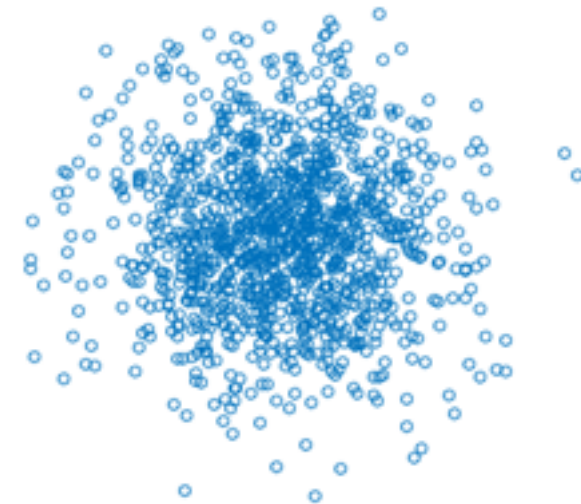
Megfigyelés

Modell

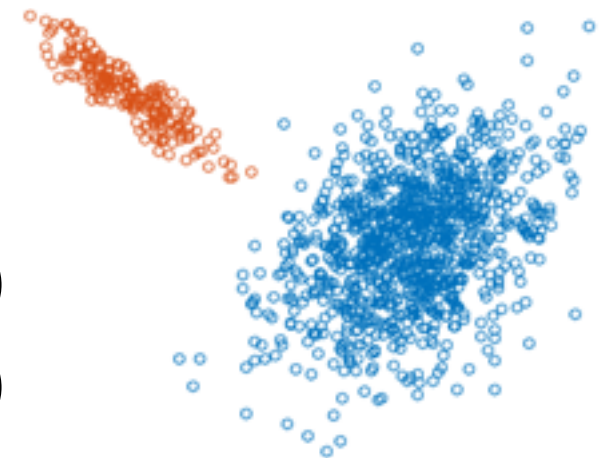
Szintetikus adat



$$p(x) = \mathcal{N}(x; \mu, C)$$



$$p(x | z) = \mathcal{N}(x; \mu_z, C_z)$$
$$p(z) = Mult(z; r)$$



$z=1$, $z=2$

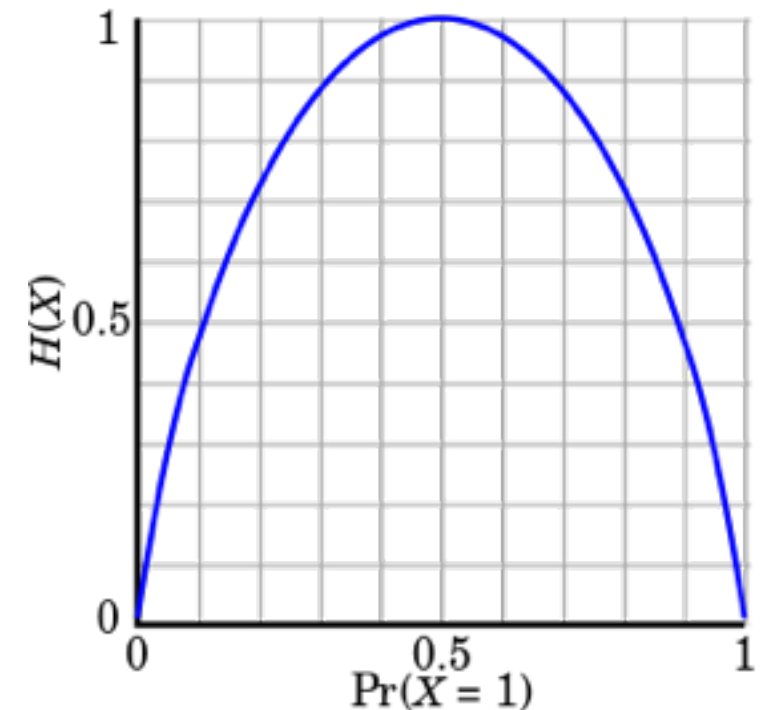
- Hogyan kezeljük formálisan a bizonytalan tudást?
- A valószínűségi kalkulus alapelemei
- Hogyan építsünk és kezeljünk modelleket?
- **Egy eloszlás információtartalma**

Eloszlások információelméleti leírása

- Mennyire meglepő egy adott megfigyelés?

$$I(x) = -\log p(x)$$

- mértékegység a bit vagy a nat a logaritmus bázisától függően



- Mennyi információt tartalmaz egy eloszlás az adott mennyiségről?

$$H(x) = \sum_i p(x_i) I(x_i)$$

- Magas entrópia - keveset tudunk arról, hogy mik a jellemző értékek
- Alacsony entrópia - vannak jellemző régiók az értékkészletben

Maximális entrópiájú eloszlások

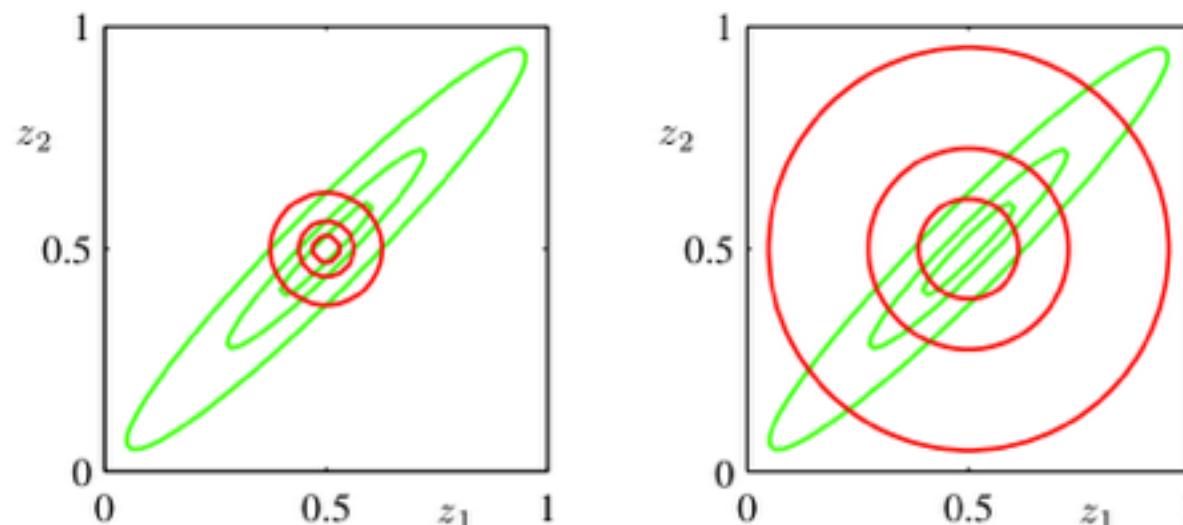
- A cél, hogy minimális pluszfeltételezést tegyünk a modellünkbe a matematikai formalizmus kedvéért
 - a valószínűség logikai értelmezéséből adódik, hogy pontosan azokat az információkat szeretnénk reprezentálni az eloszlásunkkal, amik a tudásbázisunkban vannak, és lehetőleg semmi mást
- Különböző ismert tulajdonságok mellett különböző formájú eloszlások maximalizálják az entrópiát
 - ha csak azt a véges intervallumot ismerjük, ahol pozitívak a valószínűségek
 - egyenletes
 - ha ismerjük az átlagot és a szórást, és a valós számokon van értelmezve az eloszlás
 - Gauss

Sűrűségfüggvények távolsága

- Szeretnénk összehasonlítani a modelljeinket a megfigyelésekkel és egymással is
 - a kérdés, hogy egy modell prediktív eloszlása mennyire különbözik a tapasztalati vagy a másik modell által leírt prediktív eloszlástól
 - definiáljuk távolságmértéket eloszlások között azon keresztül, hogy ha $p(x)$ eloszlást $p'(x)$ segítségével közelítjük, mennyi információt veszítünk el
- Kullback-Leibler divergencia

$$D_{KL}(p(x) || p'(x)) = \sum_i p(x_i) \ln \frac{p(x_i)}{p'(x_i)}$$

- Szimmetrizálással metrikát is képezhetünk: Jensen-Shannon metrika



Ajánlott források

- C. M. Bishop: Pattern Recognition and Machine learning, fejezetek: 2,8
- D. Koller: Probabilistic Graphical Models, online kurzus
- E. T. Jaynes: Probability Theory, the Logic of Science

Házi feladat

- Készíts generatív valószínűségi modellt, ami autógyártók éves bevételeinek a predikciójára használható
 - válaszd ki a fontos változókat
 - a változók közötti függetlenségi viszonyok alapján rajzolj grafikus modellt
 - válassz diszkrét vagy folytonos eloszlásokat a szükséges marginálisok és kondicionálisok formájául
 - gondolkodj el rajta, hogy mik azok a feltételezések, amiket beleépítettél a modellbe, de sejtetően nem egyeznek a valósággal