

*Hibridna umjetna inteligencija,
bejsovska statistika
i kauzalno zaključivanje*

Zlatko Sirotić, univ. spec. inf.
ISTRA TECH d.o.o., Pula

Savjetovanje nastavnika elektrotehnike i
računarstva Istarske županije
Pula, 30.06.2025.

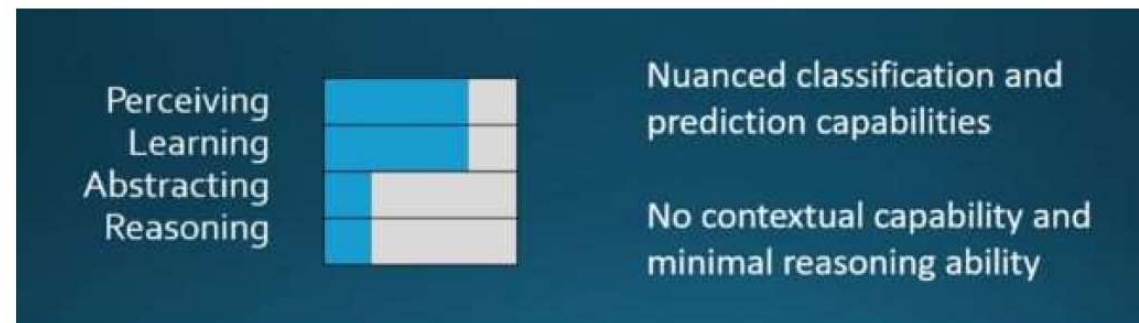
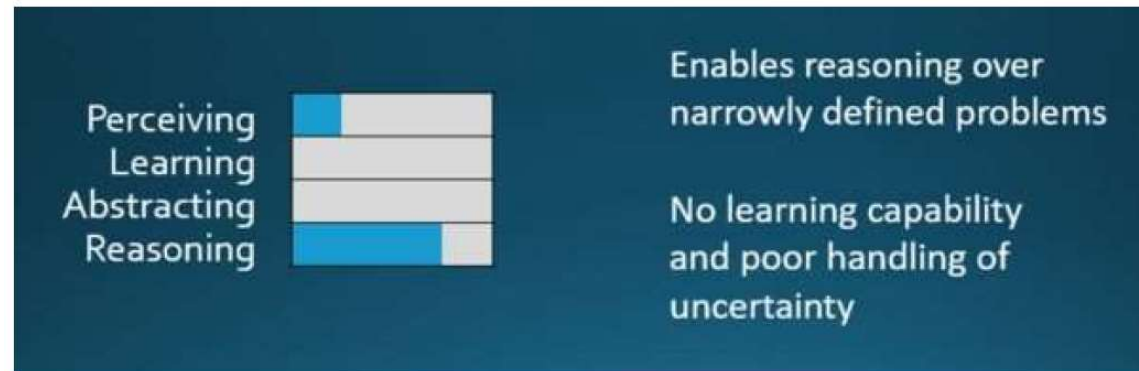
Umjetna inteligencija (UI, engl. AI)

- Umjetna inteligencija (UI) se standardno dijeli na **simboličku** (često se naziva logička UI ili GOF AI) i **subsimboličku** (često se naziva statistička UI).
- Danas su u centru pažnje (duboke) neuronske mreže, kao najvažniji dio subsimboličke UI, no u određenim aplikacijama (npr. DeepMind-ovi AlphaGeometry, AlphaFold, AlphaEvolve) izvrsne rezultate postiže spoj neuronskih mreža i simboličke UI, tzv. **neuro-simbolička UI**.
- **Kauzalno zaključivanje** (causal inference) smatra se važnim za postizanje naprednije UI, npr. tog mišljenja su "očevi" dubokih neuronskih mreža Yoshua Bengio i Yann LeCun (dobitnici Turingove nagrade 2018. godine).
- U kauzalno zaključivanje ovdje uključujemo i **kauzalno učenje** (causal learning), kod kojega se gradi kauzalni model, i **kauzalno rezoniranje** (causal reasoning), gdje se kauzalni model koristi zajedno s podacima za kauzalno rezoniranje.

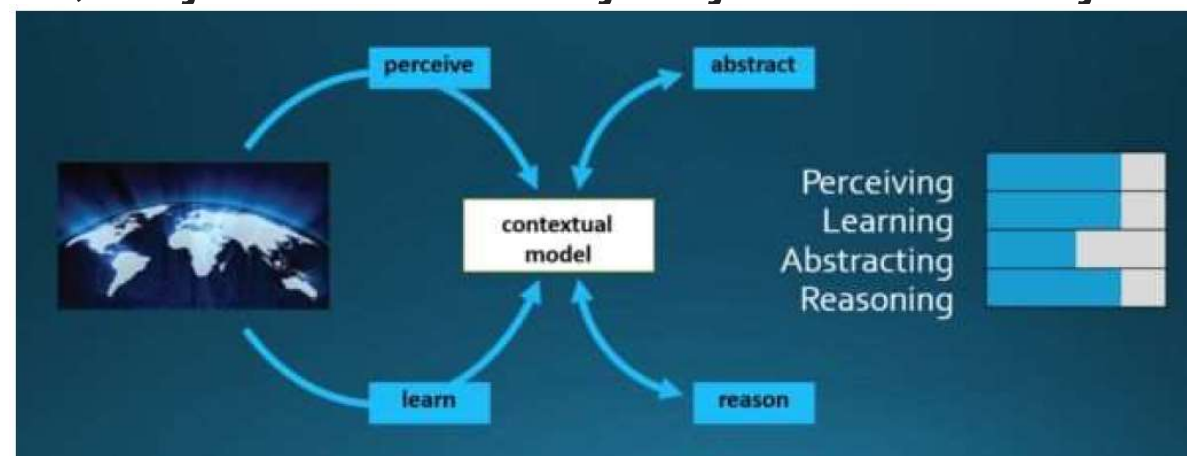
Umjetna inteligencija – kratka povijest

- Pojam 'umjetna inteligencija' prvi se put koristio 1956. na konferenciji u Dartmouthu, ali prvi rad na području UI su još 1943. napisali Walter Pitts i Warren McCulloch "A Logical Calculus of the ideas Imminent in Nervous Activity", prvi matematički model neuronske mreže.
- Dakle, neuronske mreže su prethodile simboličkoj UI. Međutim, neuronske mreže, kao i šira statistička UI, bile su do prije desetak godina inferiornije od simboličke UI.
- Može se reći da postoje tri vala (engl. wave) umjetne inteligencije. U prvom valu je prevladavala simbolička UI, koja je vrlo dobra u rasuđivanju (Reasoning), ali ima dosta slabe sposobnosti opažanja (Perceiving), a vrlo je loša kod učenja (Learning) i stvaranja apstrakcija (Abstracting).
- Drugi val, korištenje statističke UI (i to naročito dubokih neuronskih mreža), doveo je do velikih poboljšanja kod percepcije i učenja, ali je jako pala sposobnost rasuđivanja.

Umjetna inteligencija



- Radi se na trećem valu, spoju simboličke i statističke UI u hibridnu UI, koja bi imala najbolje strane obje.



Logika, matematika, statistika - neka pitanja

- Koje su granice matematičke logike?
- Koje su granice (aksiomatizirane) matematičke teorije?
- Postoji li slučajnost, ili je to samo privid?
- Je li simbolička UI bolja od subsimboličke UI,
ili obrnuto,
ili je bolje da koristimo obje?

Logika sudova (Propositional Logic)

□ **Operatori (veznici) logike sudova (račun/algebra sudova):**

negacija (negation) $\neg$

disjunkcija (disjunction) $\vee$

konjunkcija (conjunction) $\wedge$

kondicional (conditional) $\rightarrow$

bikondicional (biconditional) $\leftrightarrow$

eksluzivno ili (exclusive or) $\oplus$

nili (nor) $\downarrow$ ni (nand) $\uparrow$

□ Primjer formule logike sudova:

$$(p \rightarrow q) \leftrightarrow (\neg q \rightarrow \neg p)$$

□ Napomena: mala slova (npr. p, q) označavaju **atomarne (atomic) sudove** ili atome.

□ Jedna interpretacija (u prirodnom jeziku):

ako **pada kiša** (p), tada **trava je mokra** (q)

ekvivalentno je

ako **trava nije mokra** ($\neg q$), tada **ne pada kiša** ($\neg p$).

Logika sudova (Propositional Logic)

- Logika sudova (za razliku od složenije logike prvog reda) ima **proceduru odlučivanja** (engl. decision procedure), tj. algoritam koji u konačnom vremenu vraća odgovor da li je formula valjana (ili da li je zadovoljiva).
- Jedna procedura odlučivanja je izrada **tablice istinitosti (semantička tablica)**.
- Sljedeća tablica istinitosti pokazuje da je (prije prikazana) formula logike sudova $(p \rightarrow q) \leftrightarrow (\neg q \rightarrow \neg p)$ valjana:

p	q	$(p \rightarrow q)$	$\leftrightarrow$	$(\neg q \rightarrow \neg p)$
T	T	T	T	T
T	F	F	T	T
F	T	T	T	F
F	F	T	T	T

Logika prvog reda (First-Order Logic)

- Logika prvog reda (**logika predikata**, ili **račun predikata**) je nadskup logike sudova, i dodaje joj sljedeće:
 - **predikate**: n-arni predikat se interpretira kao n-arna relacija nad određenom domenom
 - **varijable i konstante** (iz određene domene), koje čine argumente predikata
 - **univerzalni kvantifikator** (universal quantifier)
 $\forall$ (čita se "za svaki")
 - **egzistencijalni kvantifikator** (existential quantifier)
 $\exists$ (čita se "postoji").
- **LPR s funkcijama** dodaje još **funkcije**, tako da argumenti u predikatima mogu biti i funkcije.
- Kod LPR, **kvantifikatori se primjenjuju samo na varijable**, a ne npr. na predikate (to je logika drugog reda).

Logika prvog reda (First-Order Logic)

- Primjer jednostavne formule (ovaj primjer nema funkcije):
 $\forall x \, p(a, x)$ čita se "**za svaki x vrijedi $p(a, x)$** "
(x je varijabla, p je dvomjesni predikat, a je konstanta).
- Jedna interpretacija, kod koje je formula istinita:
 $I1 = (N, \{\leq\}, \{0\})$ - domena je N , tj. skup prirodnih brojeva;
predikat p se zamjenjuje relacijom $\leq$, konstanta a s nulom,
pa za svaki x element od N **vrijedi** da $0 \leq x$.
- Jedna interpretacija, kod koje formula nije istinita:
 $I2 = (Z, \{\leq\}, \{0\})$ - domena je Z , tj. skup cijelih brojeva
(negativni cijeli brojevi, nula i pozitivni cijeli brojevi);
predikat p se zamjenjuje relacijom $\leq$, konstanta a s nulom,
pa **ne vrijedi** za svaki x element od Z da $0 \leq x$
(ne vrijedi za negativne brojeve).
- **LPR nema proceduru odlučivanja, ali je poluodlučiva**
(procedura odlučivanja može ne završiti).

Kako se gradi matematička teorija (aksiomatski sustav)

- Treba imati određeni broj **nedefiniranih termina**.
- Na temelju njih, prema utvrđenim **pravilima definiranja**, grade se **definirani termini**.
- Treba imati određeni broj **aksioma**.
- Na temelju njih, prema utvrđenim **pravilima zaključivanja**, izvode se **teoremi**.

Obavezne ili (vrlo) poželjne osobine aksiomatskog sustava

- Mora biti **konzistentan**, tj. u njemu se ne može dokazati i neka tvrdnja A i njena negacija.
Nekonzistentan aksiomatski sustav je bezvrijedan.
- Vrlo je poželjno da je **(sintaksno) potpun**, tj. da se unutar njega može dokazati ili opovrći bilo koja tvrdnja izrečena u terminima tog sustava. **Nepotpuni sustav nije bezvrijedan.**
- Vrlo poželjno je da je **odlučiv**, tj. da ima opću metodu kojom se može utvrditi (ne)valjanost svake tvrdnje izrečene u terminima tog sustava.
- Većinom je poželjno da je **neredundantan**, što znači da su aksiomi međusobno nezavisni.
- Često se smatra da je **finitni aksiomatski sustav** metamatematički elegantniji od sustava s (prebrojivo) beskonačnim brojem aksioma. **Peanova aritmetika nije finitna, jer ima aksiom indukcije** (koji je, zapravo, **shema aksioma** - prebrojivo beskonačan skup aksioma).

Neeuklidske geometrije - 5. Euklidov aksiom

- Najvažniji pokretač formalizacije aksiomatskih sustava bilo je uvođenje **neeuklidskih geometrija**, koje su nastale negacijom **5. Euklidovog aksioma o paralelama**.
- Aksiom kaže (ne u izvornoj Euklidovoj interpretaciji) da se kroz točku koja se nalazi izvan pravca (u istoj ravnini) može povući točno jedan paralelan pravac.
Dugo se smatralo da 5. aksiom zavisi od drugih.
- U 19. stoljeću stvoreni su aksiomatski sustavi (**Gauss, Lobačevski, Bolay, Riemann**) u kojima je 5. aksiom zamijenjen aksiomom da se kroz točku ne može povući niti jedan paralelan pravac (**eliptičke geometrije**), ili da se može povući beskonačno puno pravaca (**hiperboličke geometrije**).
- U tim se geometrijama više ne može "prirodno dokazivati". Ali, matematičari su tada ustvrdili da se geometrija može sagraditi na temelju aritmetike, i rekli:
ako je aritmetika konzistentna, onda je i geometrija.

Kurt Gödel i nepotpunost aritmetike

- 1929. je (sa 23 godine!) **dokazao (semantičku) potpunost predikatnog računa 1. reda**. To nije lako, za razliku od daleko lakšeg problema dokazivanja potpunosti (ili konzistentnosti) računa sudova.
- **1931. godine objavljen je njegov rad kojim je srušio nade Hilbertovog programa.**
- Prvo, **dokazao je (sintaksnu) nepotpunost aksiomatskog sustava aritmetike** (ili bilo kog sustava sličnog njemu). Pritom se koristio "**prevođenjem**" **metamatematičkih tvrdnji u aritmetičke tvrdnje**.
- To znači da se ne može napraviti konačan ili prebrojivo beskonačan skup (konzistentnih) aksioma iz kojih će se moći dokazati sve aritmetičke istine – **formalizacija ima granice!**

Usporedba osobina logike sudova, logike 1. reda i aritmetike

Logički sustav / matematička teorija	Pouzdana / konzistentna	Potpuna	Odlučiva
Logika sudova	DA (pouzdana)	DA (semantički) (Post, 1921.)	DA
Logika prvog reda	DA (pouzdana)	DA (semantički) (Gödel, 1929.)	NE Churchov teorem (Church, Turing, 1936.) = rješenje za Hilbertov Entscheidungs- problem (1928.)
Aritmetika	DA (konzistentna) (Gentzen, 1936., nefinitnim metodama); NE može se dokazati unutar nje same (Gödel, 1931.)	NE (sintaktički) (Gödel, 1931.)	NE Churchov teorem

Digresija – četiri znanstvena šoka u 20. stoljeću

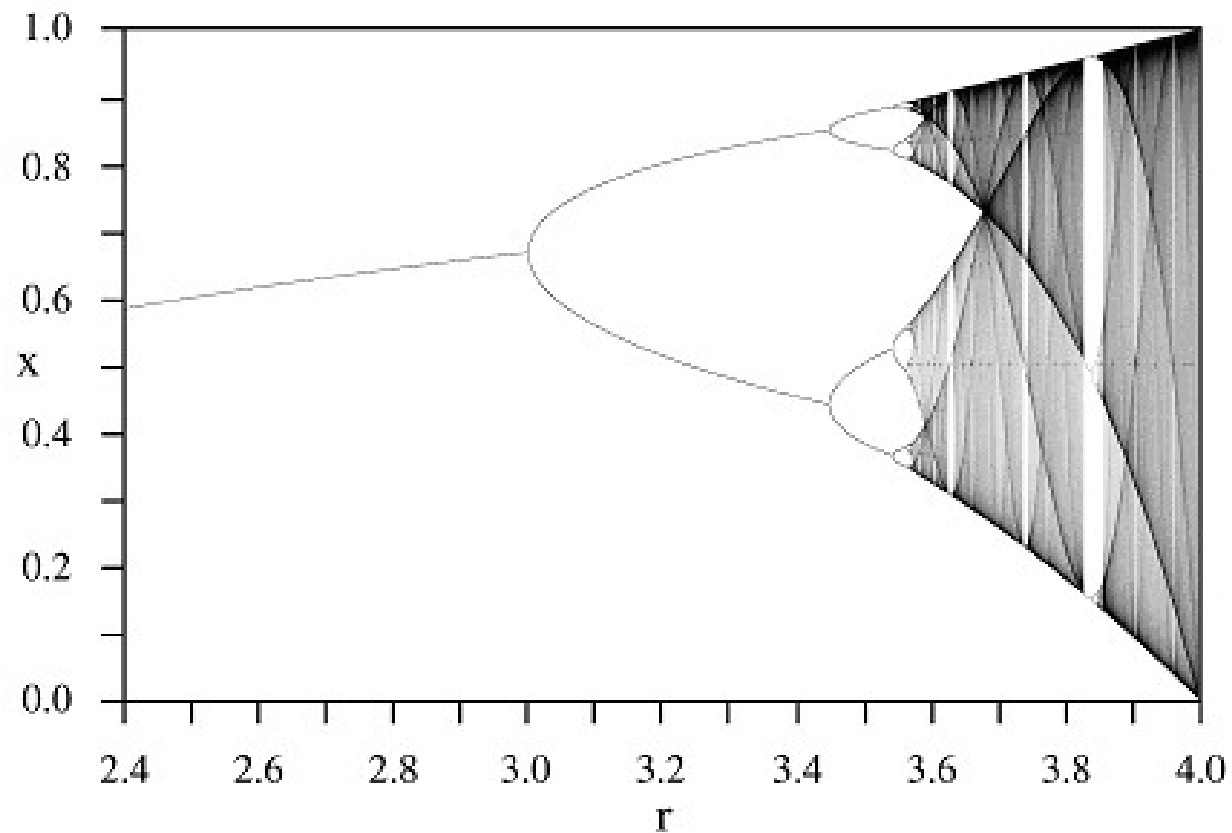
- ❑ **Specijalna i opća teorija relativnosti.**
- ❑ **Kvantna mehanika.** Čak niti sam Einstein nije prihvatio standardnu interpretaciju kvantne mehanike (njegova izreka: "**Bog se ne kocka!**").
Opća teorija relativnosti i kvantna mehanika su **međusobno kontradiktorne** - to znači da najviše jedna (od njih dvije) može biti istinita.
- ❑ **Gödelovi teoremi** o nepotpunosti i nemogućnosti finitnog dokazivanja konzistentnosti aritmetike.
- ❑ **Teorija determinističkog kaosa** - nemjerljivo mala različita početna stanja dovode do značajno različitih ishoda **nakon nekog vremena** (za razliku od stohastičkih sustava).

Deterministički kaos

- **Deterministički kaos je (relativno) nova paradigma za matematičko modeliranje svijeta.** Prije su postojale dvije paradigme za matematičko modeliranje svijeta.
- Prva, starija, bila je rješavanje svega pomoću **matematičke analize**, tj. pomoću diferencijalnih jednažbi. Druga se paradigma pojavila krajem 19.stoljeća, a to je **statistička analiza**, temeljena na teoriji vjerojatnosti.
- Kako bi se mogla opisati razlika između slučajnih procesa i determinističkog kaosa:
 - neki sustav je slučajan (stohastički) ako nemjerljivo mala različita početna stanja (a ponekad i ista početna stanja) **odmah** dovode do značajno različitih ishoda
 - neki sustav je deterministički kaotičan ako nemjerljivo mala različita početna stanja ne dovode odmah do značajno različitih ishoda, nego **tek nakon nekog vremena**.

Bifurkacijski dijagram kod logističkog preslikavanja

$$X_{n+1} = r * X_n * (1 - X_n) \text{ (r je neki broj između 0 i 4)}$$



Vjerojatnost kao proširena logika (probabilističko zaključivanje) – time se bavio još George Boole!

□ Iz knjige Introduction to Artificial Intelligence (Ertel W.)

□ "Klasični" modus ponens (pravilo zaključivanja):

$$(A, A \rightarrow B) \vdash B$$

□ "Generalizirani" modus ponens:

$$(P(A) = \alpha, P(B|A) = \beta) \vdash P(B) = ?$$

$$P(B) = P(A, B) + P(\neg A, B) = P(B|A) P(A) + P(B|\neg A) P(\neg A)$$

$$p_1 = P(A, B), p_2 = P(A, \neg B), p_3 = P(\neg A, B), p_4 = P(\neg A, \neg B)$$

$$p_1 = \alpha \beta$$

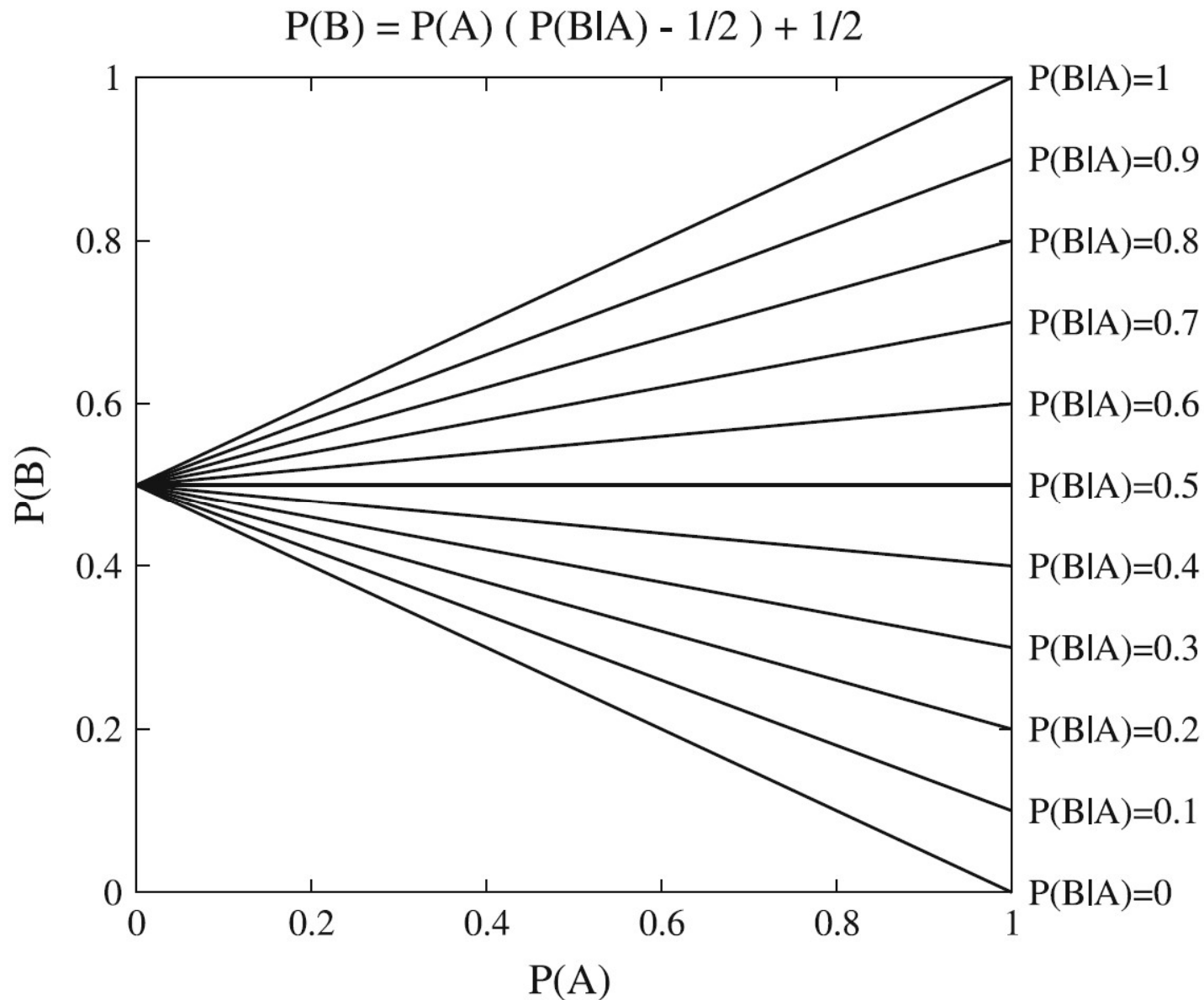
$$p_1 + p_2 = \alpha \quad \Rightarrow p_2 = \alpha (1 - \beta)$$

$$p_1 + p_2 + p_3 + p_4 = 1 \quad \Rightarrow \mathbf{p_3 + p_4 = 1 - \alpha}$$

□ Na formulu za entropiju $H(p) = -p_3 \ln p_3 - p_4 \ln p_4$ primjenjujemo metodu Lagrangeovih multiplikatora.

Vjerojatnost kao proširena logika

□ Entropija je maksimalna za: $p_3 = p_4 = (1 - \alpha) / 2$

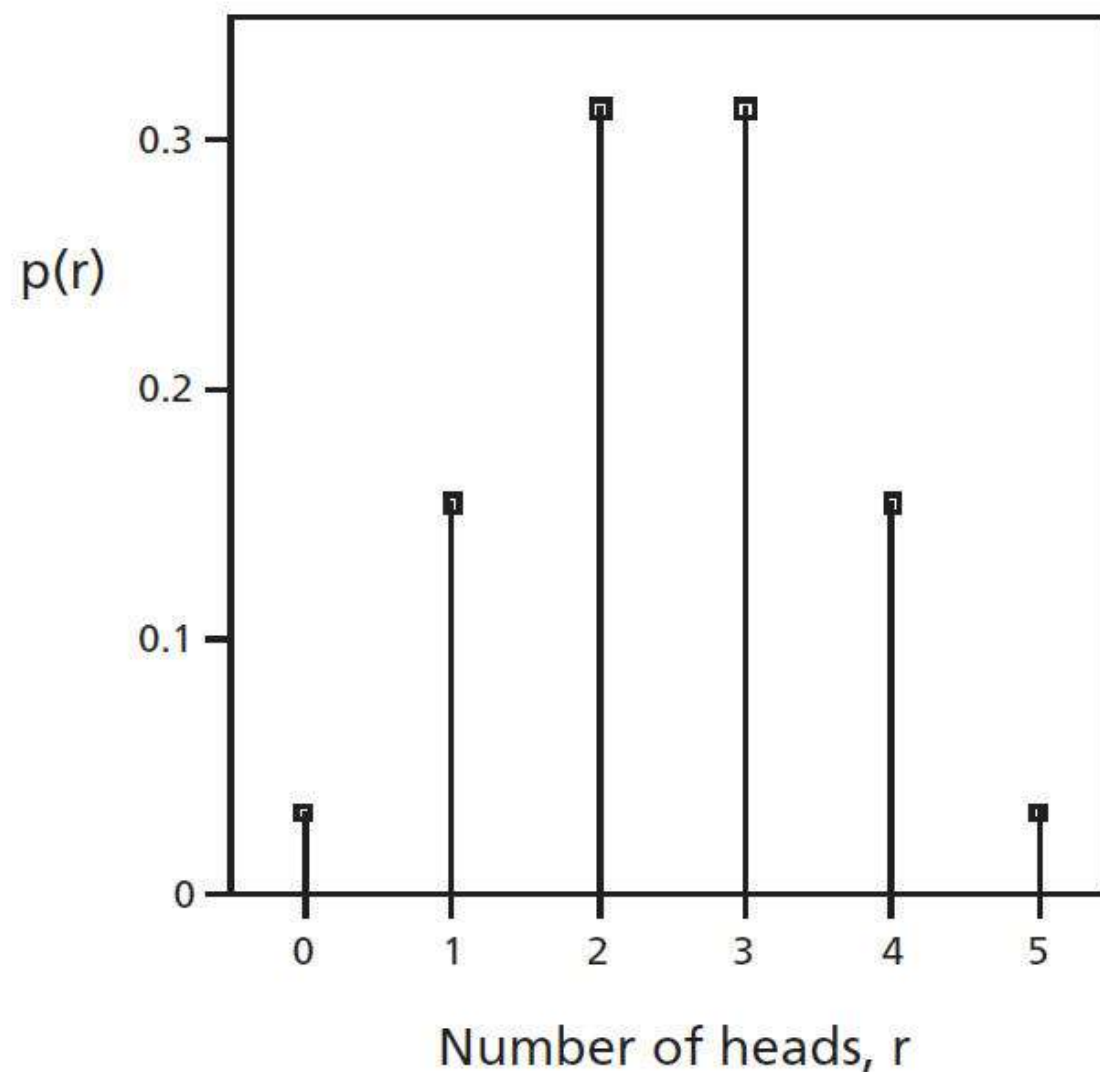


Matematička statistika – podsjetnik;

Diskretne slučajne varijable

- Postoje diskretne (diskontinuirane) i kontinuirane slučajne varijable.
- **Diskretna slučajna varijabla** takva je varijabla x koja
 - prima niz vrijednosti $x_1, x_2, \dots$
 - ali svaku od njih s određenom vjerojatnošću $p(x_1), p(x_2), \dots$
 - pri čemu vjerojatnosti $p(x_i)$ zadovoljavaju jednakost
$$\sum p(x_i) = 1$$
- Zakon $p(x)$ po kojem svakoj vrijednosti x_i pripada vjerojatnost $p(x_i)$ zovemo **funkcijom gustoće vjerojatnosti diskretne slučajne varijable** x (najčešći engl. termin je Probability Mass Function - PMF).

Primjer: Funkcija gustoće vjerojatnosti varijable r (broj glava kod bacanja pet novčića)

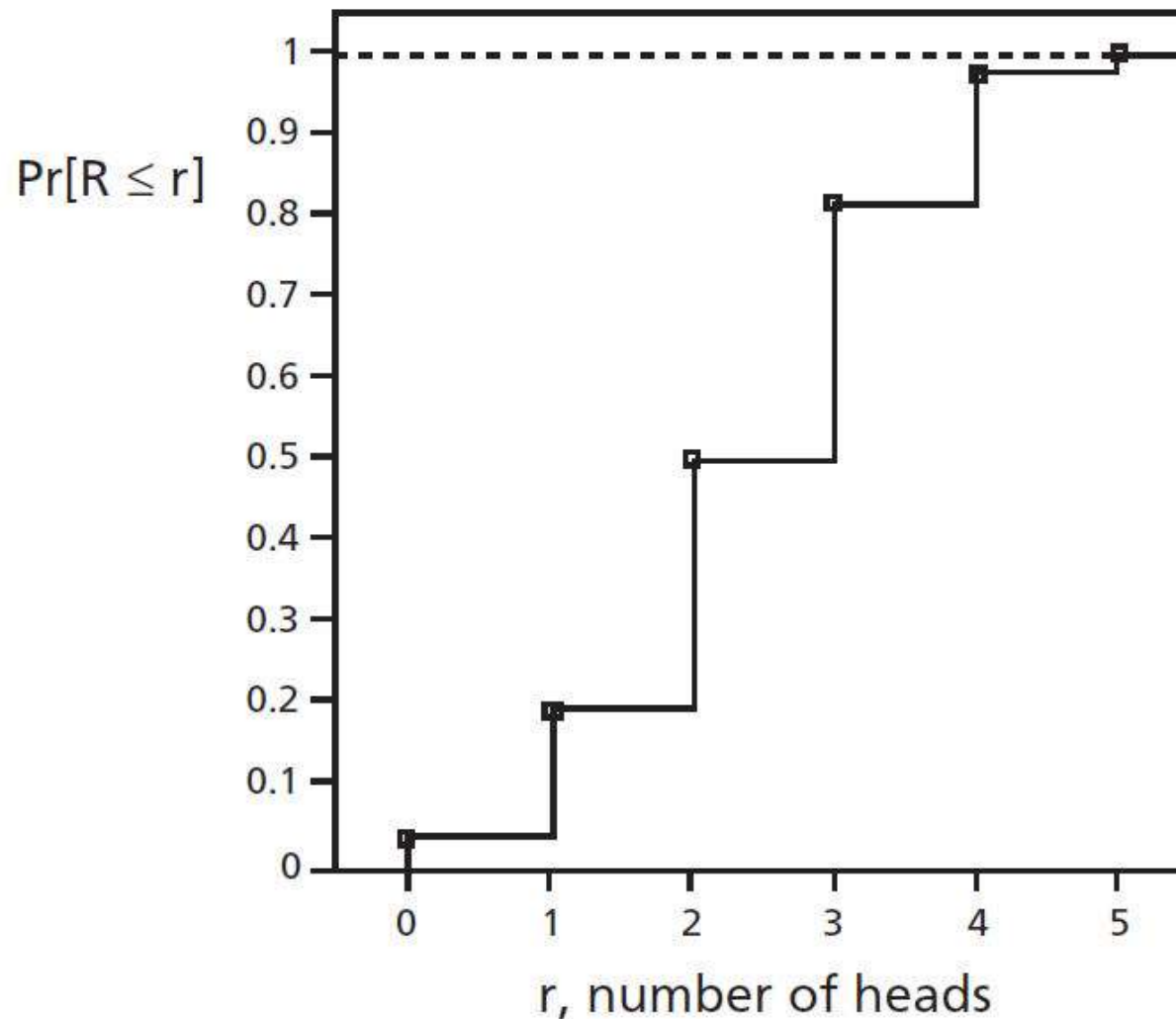


Funkcija distribucije vjerojatnosti slučajne varijable

- Osim funkcije vjerojatnosti, kod diskretnih slučajnih varijabli važna je **funkcija distribucije (ili razdiobe) vjerojatnosti slučajne varijable** (engl. Cumulative Distribution Functions - CDF).
- Ona pokazuje kolika je vjerojatnost da slučajna varijabla x poprimi bilo koju vrijednost $\leq x_0$:

$$F(x_0) = \sum_{x_i \leq x_0} p(x_i) \quad \text{tj.} \quad F(x_0) = P\{x \leq x_0\}$$

Primjer: Funkcija distribucije varijable r (broj glava kod bacanja pet novčića)

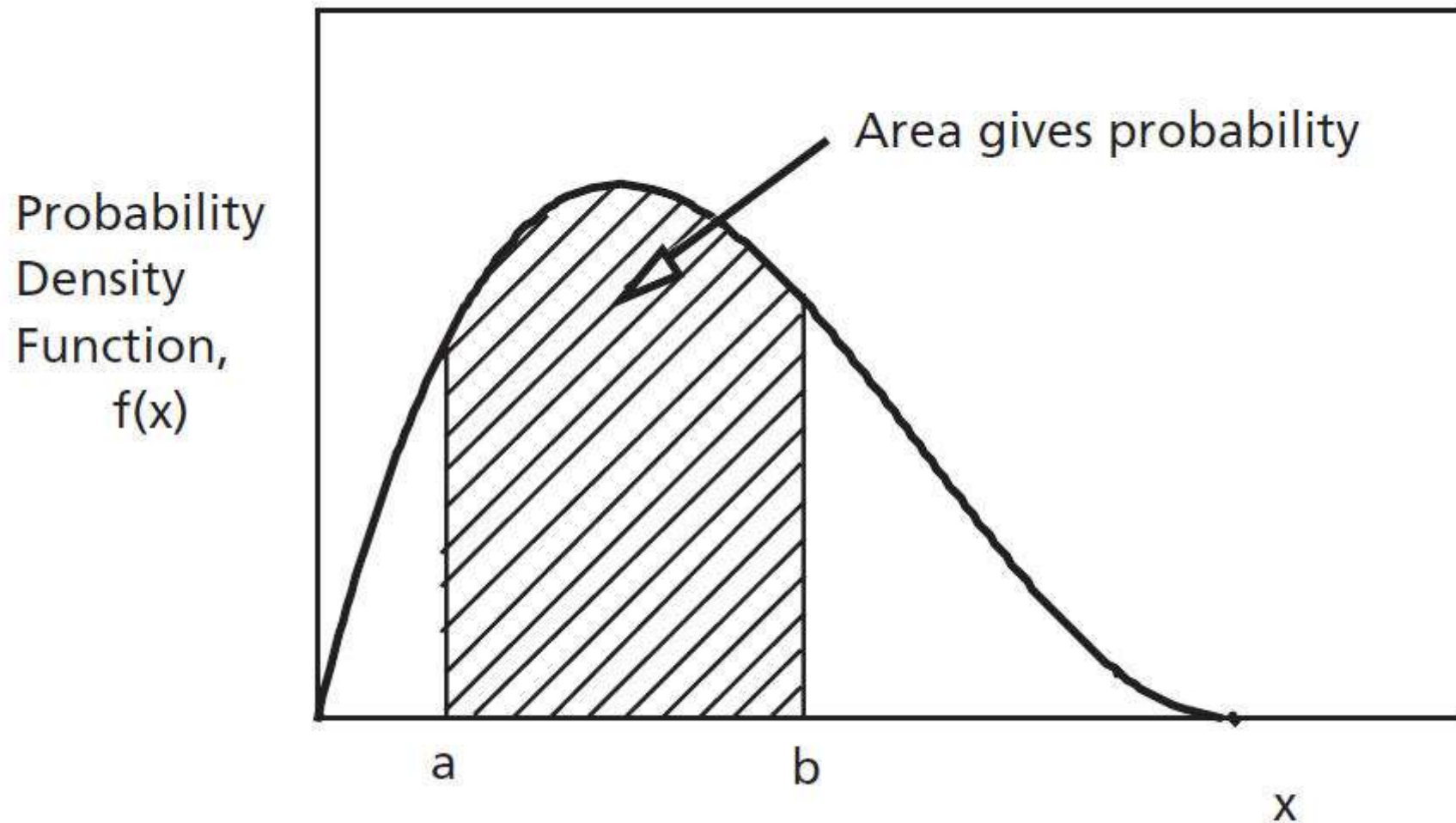


Kontinuirane slučajne varijable

□ **Funkcija gustoće vjerojatnosti kontinuirane slučajne varijable** x (engl. Probability Density Function - PDF) je takva funkcija $f(x)$ koja ima svojstva:

1. $f(x) \geq 0$ za svaki x iz domene funkcije $[a, b]$
(a može biti $-\infty$, b može biti $+\infty$)
2. $\int_a^b f(x) dx = 1$
(površina ispod funkcije unutar domene $[a, b]$ je 1)
3. $\int_{x_1}^{x_2} f(x) dx = P\{x_1 \leq x \leq x_2\}$
(površina ispod funkcije unutar domene $[x_1, x_2]$ jednaka je vjerojatnosti da varijabla poprimi vrijednost iz te domene).

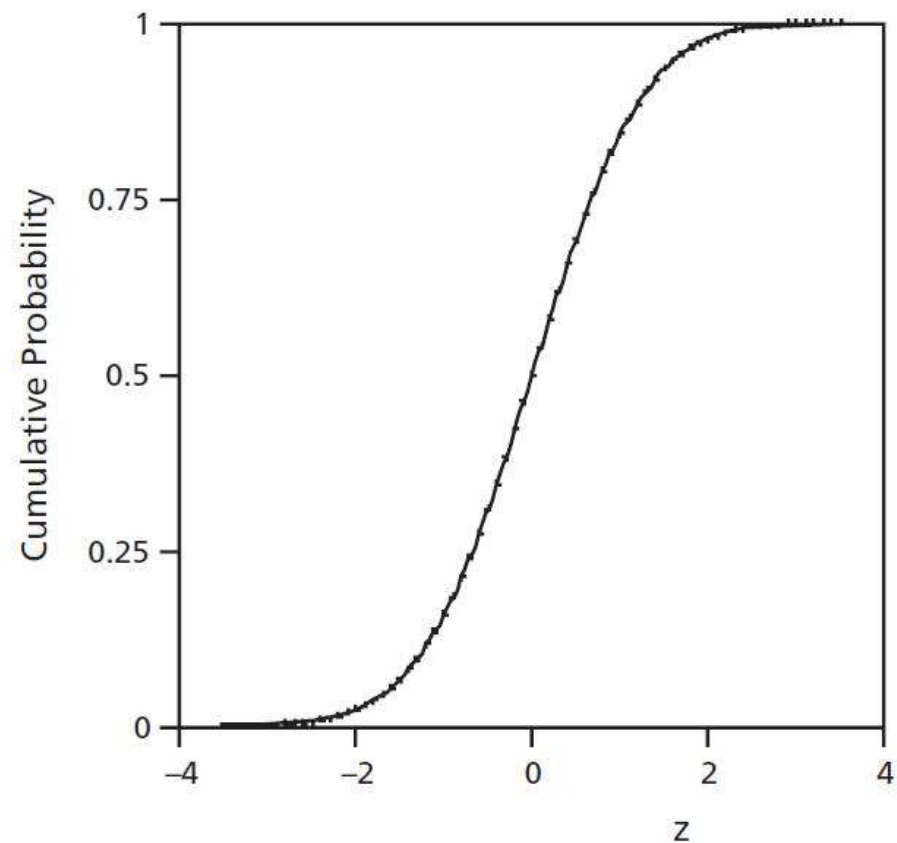
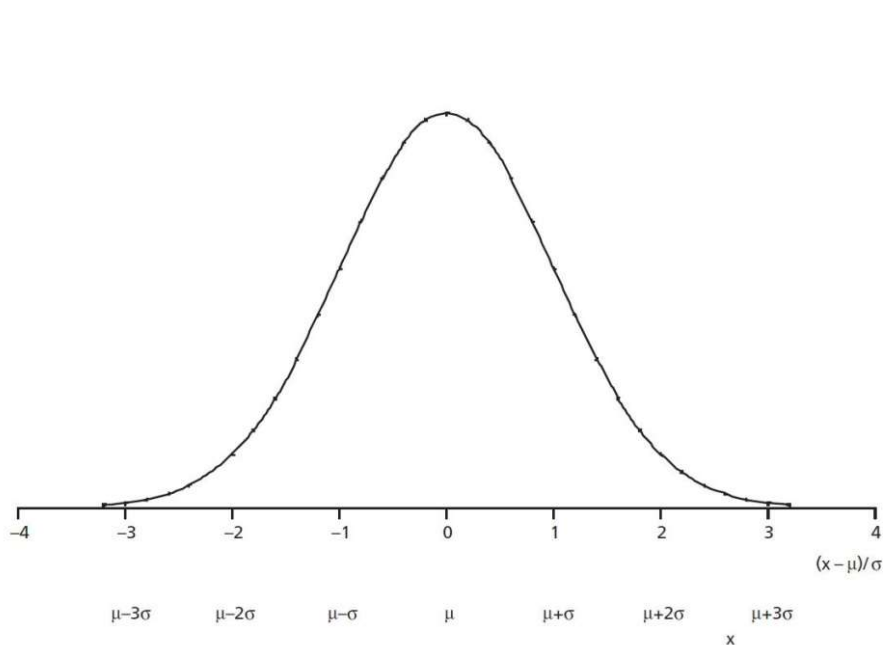
Vjerojatnost kod kontinuirane slučajne varijable



Normalna razdioba (Gaussova razdioba)

- Jedna od najpoznatijih funkcija gustoće vjerojatnosti kontinuirane varijable je tzv. **normalna razdioba** (poznata i kao **Gaussova razdioba**).
- Važna je po tome što mnoge druge razdiobe (diskretne i kontinuirane) graniče prema njoj ako neki parametri rastu u beskonačnost.
- Prema njoj graniči i razdioba aritmetičkih sredina uzoraka, bez obzira na razdiobu osnovnog skupa, ako broj elemenata teži ka beskonačnosti (centralni granični teorem o razdiobi aritmetičkih sredina velikih uzoraka).
- Posebno postoji tzv. **jedinična (ili standardna) normalna razdioba**, kod koje je matematičko očekivanje = 0, a standardna devijacija = 1.

Funkcija gustoće vjerojatnosti i funkcija distribucije vjerojatnosti kod jedinične normalne razdiobe



Testiranje statističkih hipoteza – mogući ishodi

Hipoteza H_0	Istinita	Neistinita
Odbacuje se	Greška 1.vrste (vjerojatnost je α)	Pravilan zaključak
Prihvaća se	Pravilan zaključak	Greška 2.vrste (vjerojatnost je β)

- Manja greška α rezultira većom greškom β (i obrnuto), ali **ne vrijedi** $\alpha + \beta = 1$.
- Uobičajeno se zadaje (mali) α (5% ili 1%), dok β ne znamo unaprijed. Jakost testa $p = 1 - \beta$.

Kod frekventističkog pristupa hipoteze nemaju vjerojatnosti, tj. mi znamo npr. $P(B|H_0)$, a ne $P(H_0|B)$

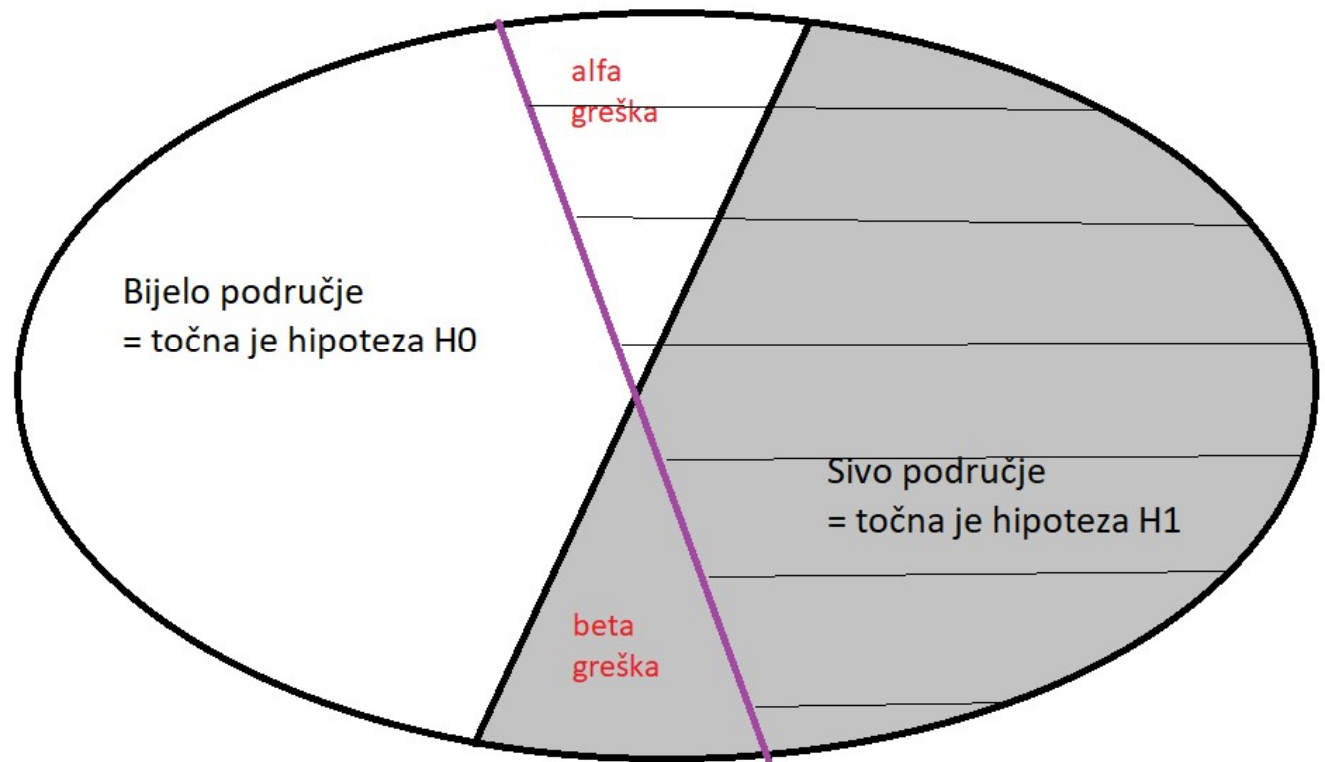
$$P(A|H_0) = 1 - \alpha$$
$$P(B|H_0) = \alpha$$

$$P(A|H_1) = \beta$$
$$P(B|H_1) = 1 - \beta$$

$$(\alpha + \beta \neq 1)$$

$$P(A) + P(B) = 1$$
$$P(H_0) + P(H_1) = 1$$

B područje
= prihvaća se hipoteza H1
(šrafirano)



A područje
= prihvaća se hipoteza H0

Višedimenzijska slučajna varijabla (slučajni vektor) – npr. ovdje dvodimenzijska

- Funkcija gustoće vjerojatnosti: $f(x, y)$
- Marginalne funkcije gustoće: $f(x) = \int f(x, y) dy$
- Uvjetne funkcije gustoće: $f(y|x) = f(x, y) / f(x)$
- Matematička očekivanja: $E(x) = \int x f(x) dx$
- Varijance: $V(x) = \int x^{**2} f(x) dx - E(x)^{**2}$
- Standardne devijacije: $\sigma(x) = \sqrt{V(x)}$
- **Uvjetna očekivanja:** $E(y|x) = \int y f(y|x) dy$
(u regresijskoj analizi zovu se **funkcije regresije**; nemamo "ravnopravne" slučajne varijable, nego zavisnu(e) i nezavisne)
- **Kovarijanca:** $Cov(x, y) = \int \int x y f(x, y) dx dy - E(x) E(y)$
- **Koeficijent korelacije:** $\rho = Cov(x, y) / (\sigma(x) \sigma(y))$

Ako su slučajne varijable nezavisne, kovarijanca je nula.
Obrnuto ne vrijedi: kod nekih teoretskih razdioba kovarijanca je nula i kada su slučajne varijable zavisne.

Bayesov teorem

(Thomas Bayes, 1701-1761)

□ Iz uvjetnih vjerojatnosti:

$$P(B|A) = P(A \text{ presjek } B) / P(A)$$

$$P(A|B) = P(A \text{ presjek } B) / P(B)$$

dobijemo jednu verziju Bayesovog teorema:

$$P(B|A) = P(B) P(A|B) / P(A)$$

$$P(A|B) = P(A) P(B|A) / P(B)$$

□ Verzija Bayesovog teorema s totalnim vjerojatnostima, gdje:

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1$$

□ Specijalno za $n = 2$, gdje je $P(A) + P(\neg A) = 1$

$$P(A|B) = P(A) * P(B|A) / P(B)$$

$$P(B) = P(A) * P(B|A) + P(\neg A) * P(B|\neg A)$$

pa dobijemo

$$P(A|B) = \frac{P(A) * P(B|A)}{(P(A) * P(B|A) + P(\neg A) * P(B|\neg A))}$$

Primjer korištenja Bayesovog teorema

A = sportaš uzima doping

$\neg A$ = sportaš ne uzima doping

B = test tvrdi da uzima doping

$\neg B$ = test tvrdi da ne uzima doping

$P(B|A)$ = vjerojatnost da će test reći

da uzima doping onaj koji uzima (True Positive, TP)

$P(B|\neg A)$ = vjerojatnost da će test reći

da uzima doping onaj koji ne uzima (False Positive, FP)

Imamo ove podatke:

$P(A) = 0,1$ i $p(\neg A) = 0,9$

$P(B|A) = 0,9$ i $p(B|\neg A) = 0,1$

$P(A|B)$ = vjerojatnost da uzima doping ako test tvrdi da uzima

$P(A|B) = P(A) P(B|A) / (P(A) P(B|A) + P(\neg A) P(B|\neg A))$

$P(A|B) = 0,1 * 0,9 / (0,1 * 0,9 + 0,9 * 0,1) = 0,5$

Bejsovski (Bayesian) pristup statistici

- Dugo je prevladavao frekventistički (frekvencionistički) pristup statistici (Fisher, Egon Pearson, Neyman).
- 90-tih godina prošlog stoljeća poraslo je interesiranje za bejsovski pristup, kad su računala postala dovoljno snažna.
- Bejsovski pristup bi se vjerojatno trebao zvati laplasovski pristup (Pierre-Simon Laplace, 1749.-1827.).
- Što kaže prof. Zvonimir Šikić na <https://ideje.hr/kako-je-manje-ispravna-metoda-zamijenila-puno-ispravniju-laplace-vs-fisher/>

Statističari s kraja 19. i s početka 20. stoljeća počeli su inzistirati da se **vjerojatnosti moraju shvaćati isključivo kao granične frekvencije u tzv. slučajnim eksperimentima.**

Rezultat je bila Fisherova metoda, iako nema razloga da se ne upotrebljava generalna, precizna i matematička Laplaceova, tj. bejesovska metoda.

Bejsovski (Bayesian) pristup statistici

- Bejsovski pristup poznaje **apriornu gustoću vjerojatnosti**, koja zajedno s **funkcijom vjerodostojnosti** (dobijenom na temelju uzorka) daje **posteriornu gustoću vjerojatnosti**.
- Dalje, za razliku od frekventističkog pristupa, kod bejsovskog pristupa:
 - **parametri populacije** (npr. matematičko očekivanje, varijanca ...) **su slučajne varijable**, koje imaju svoju razdiobu, a njihovi parametri se nazivaju hiperparametri
 - frekventistički pristup poznaje **pouzdana interval** (confidence interval), a bejsovski **vjerodostojni interval** (credible interval), kod kojeg zaista možemo govoriti o vjerojatnosti da se parametar nalazi unutar tog intervala
 - hipoteze imaju vjerojatnost.

Odnos između pristupa statistici i pristupa kauzalnom zaključivanju

- Većina onih koji podržavaju kauzalno zaključivanje zastupa bejsovski pristup statistici, iako ima i onih koji zastupaju frekventistički pristup.
- Ako netko zastupa bejsovski pristup, ne znači da automatski zastupa i kauzalno zaključivanje.
- Ipak, može se reći da postoji značajna korelacija između onih koji zastupaju bejsovski pristup statistici i onih koji zastupaju kauzalno zaključivanje.
- Napomena: Prevladava mišljenje da bi termin **korelacija** trebalo koristiti samo kod linearnih veza između varijabli, a u općenitom slučaju bi trebalo koristiti termin **asocijacija**.

Kauzalno zaključivanje = kauzalno učenje + kauzalno rezoniranje

□ Iz knjige Elements of Causal Inference (Peters J. i ostali)

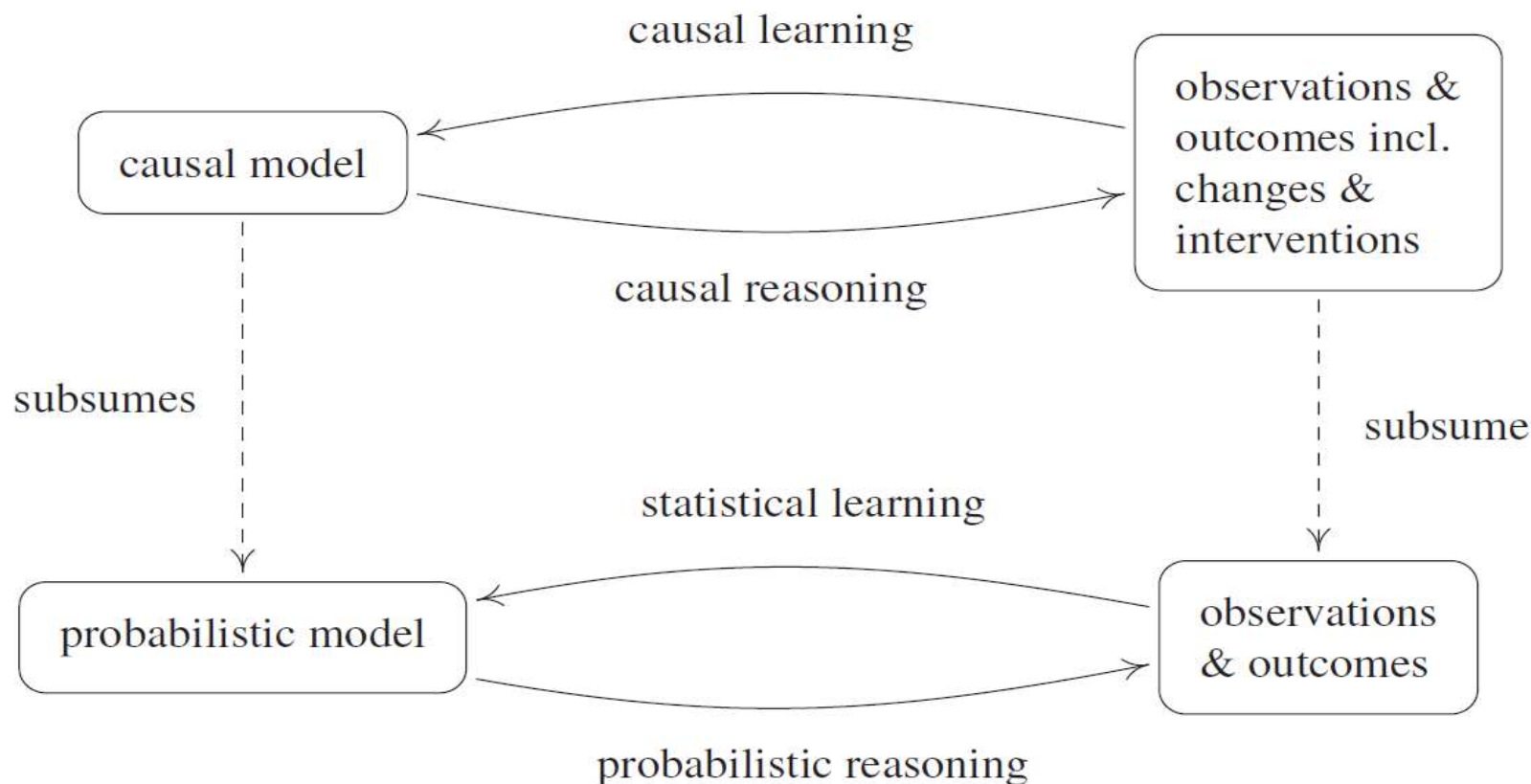


Figure 1.1: Terminology used by the present book for various **probabilistic inference** problems (bottom) and **causal inference** problems (top); see Section 1.3. Note that we use the term “inference” to include both learning and reasoning.

Knjiga "The Book of Why"

- U toj knjizi **Judea Pearl**, računalni znanstvenik i statističar, dobitnik Turingove nagrade 2011. za doprinos u području umjetne inteligencije, pokazuje (za širu publiku, a ne samo za matematičare) kako je razumijevanje uzročnosti važno za znanost, onda i za umjetnu inteligenciju.
- **"Correlation is not causation."** (korelacija nije uzročna veza)
 - ova mantra, koju znanstvenici izgovaraju više od jednog stoljeća, dovela je do praktične zabrane kauzalnog govora. Kauzalna revolucija, koju su potaknuli Judea Pearl, njegovi kolege, te ostali znanstvenici uspostavila je **proučavanje uzroka i posljedice na čvrstoj znanstvenoj osnovi.**
- Njihov rad objašnjava kako možemo znati jednostavne stvari, i kako odgovoriti na teška pitanja, poput toga je li lijek izliječio bolest. Omogućuje nam da znamo ne samo uzrokuje li jedna stvar drugu: omogućuje nam istraživanje svijeta koji jest i svjetova koji su mogli biti.

Pearl – ljestve uzročnosti

□ Protučinjenično zaključivanje

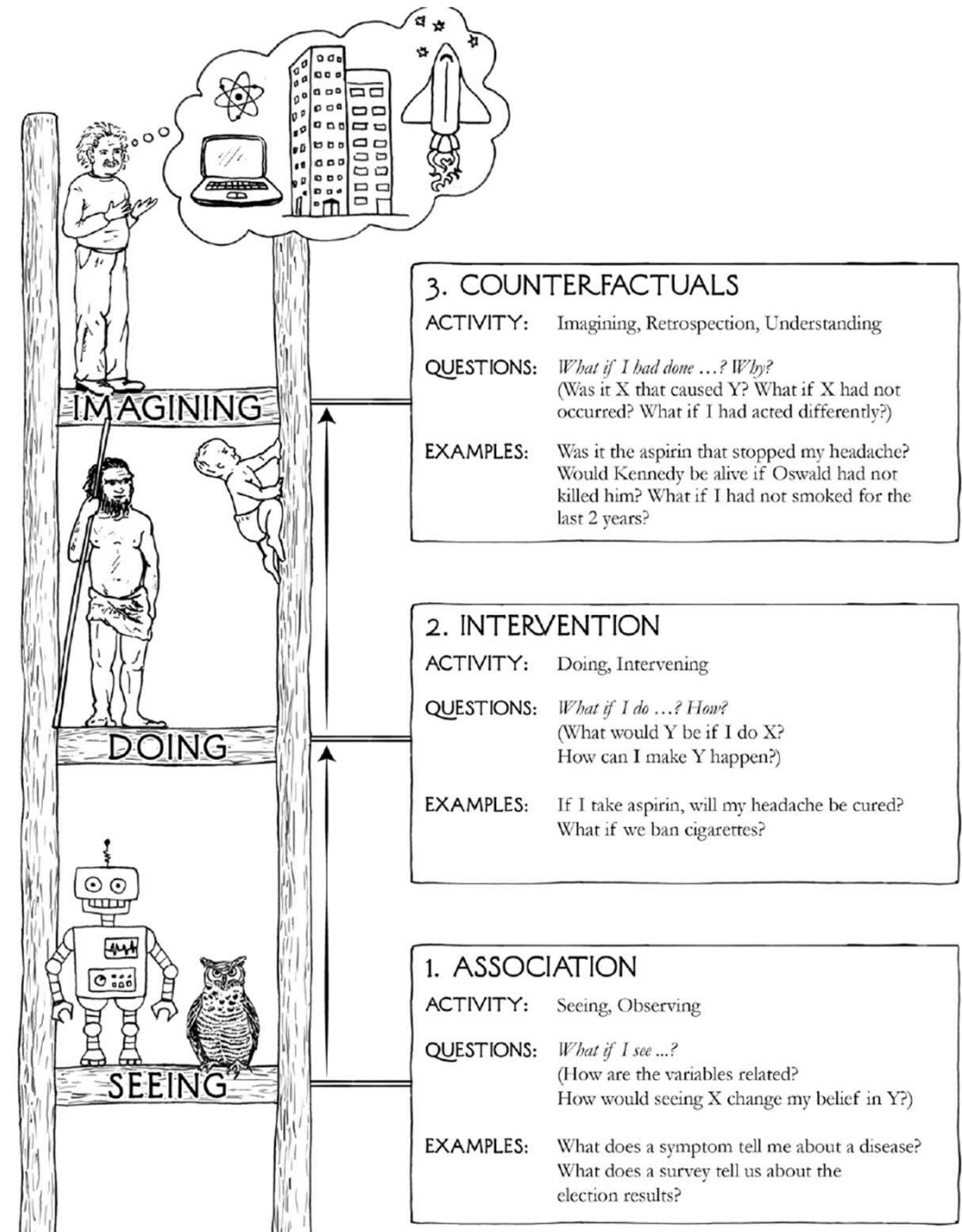
- aktivnost: imaginacija
- Što bi bilo da sam umjesto ... napravio ...?

□ Intervencija

- aktivnost: intervencija
- Što ako napravim ...?

□ Asocijacija

- aktivnost: opservacija
- Što ako vidim ...?



Pearl – Ijestve uzročnosti

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y x)$	Seeing	What is? How would seeing X change my belief in Y ?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y do(x), z)$	Doing	What if? What if I do X ?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x x', y')$	Imagining, Retrospection	Why? Was it X that caused Y ? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Eksperimenti (Randomized Control Trials, RCT) nasuprot opservacijskim studijama

- Eksperimenti (naročito dvostruko slijepi, gdje npr. pacijenti ne znaju primaju li lijek ili placebo, ali to ne znaju niti liječnici koji im daju lijek) predstavljaju **zlatni standard u statističkim testovima**.
- No često se moramo zadovoljiti s podacima koje imamo, tj. **raditi opservacijske studije, bez intervencija**. Razlozi mogu biti:
 - **etički**; npr. nije etički da teškim bolesnicima ne damo lijek za koji vjerujemo (iako nismo sigurni) da je učinkovit
 - **nemogućnost**; npr. ne možemo promijeniti DNK osobe pri rođenju, kako bismo mjerili učinak određenog gena
 - **financijski**; testiranje je skupo, a postojeće podatke možemo analizirati relativno jeftino

...

Sewall Wright, američki genetičar (1889-1988)

– otac kauzalnih dijagrama (path diagrams)

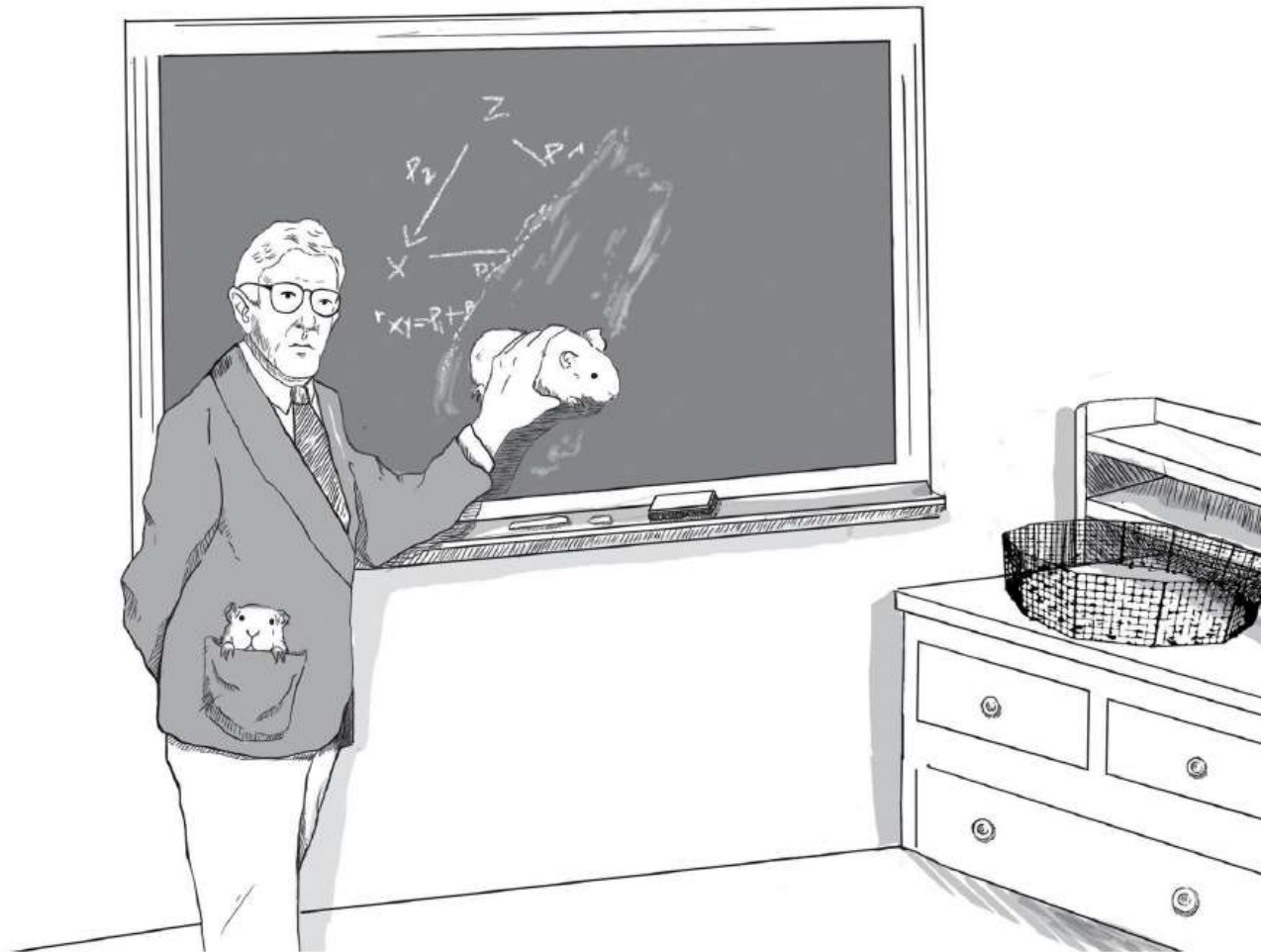


FIGURE 2.6. Sewall Wright was the first person to develop a mathematical method for answering causal questions from data, known as path diagrams. His love of mathematics surrendered only to his passion for guinea pigs. (Source: Drawing by Dakota Harr.)

Berkson's paradox

- Godine 1946. Joseph Berkson, biostatističar na klinici Mayo, ukazao je na neobičnost opservacijskih studija provedenih u bolničkom okruženju: čak i ako dvije bolesti nemaju nikakve veze jedna s drugom u općoj populaciji, može se činiti da su povezane među pacijentima u bolnici.

	General Population			Hospitalized in Last Six Months		
<i>Respiratory disease?</i> ↓	<i>Bone disease?</i> ↓			<i>Bone disease?</i> ↓		
	Yes	No	% Yes	Yes	No	% Yes
<i>Yes</i>	17	207	7.6	5	15	25.0
<i>No (control)</i>	184	2,376	7.2	18	219	7.6

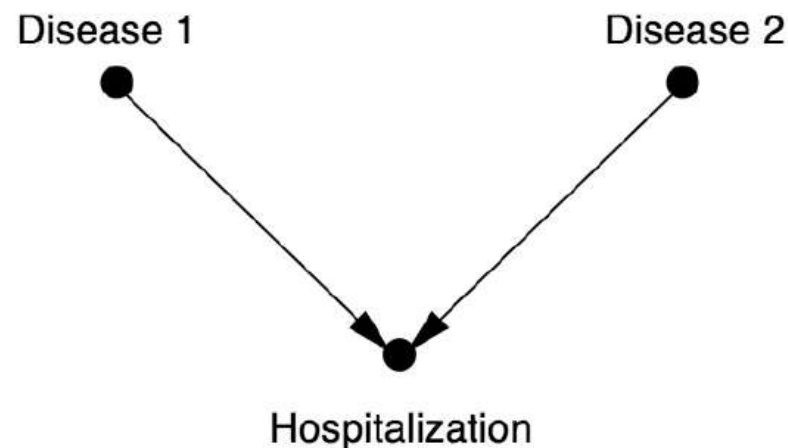
Berkson's paradox

– humoristička verzija

- U filmskoj industriji, glumci/glumice koji su postali filmske zvijezde moraju imati barem jednu od ove dvije osobine:
 - jako su zgodni
 - jako su dobri glumci/glumice.
- Statistički podaci pokazuju sljedeće:
Ako je filmska zvijezda zgodna,
manja je vjerojatnost da je dobar glumac/glumica.

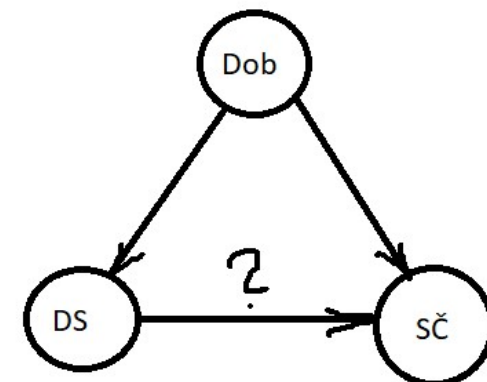
Berkson's paradox

- Razrješenje Berksonovog paradoksa za slučaj (ne)vezanih bolesti dobije se gledajući donji kauzalni graf. Obje bolesti utječu na hospitalizaciju, kao što pokazuje graf.
- To je jedan od tri osnovna grafa s tri čvora – srednji čvor zove se **mjesto sudara** ili **sudarač** ili **kolider (collider)**.
- Vidi se da su bolesti međusobno nezavisne. No **ako kontroliramo kolider** (gledamo određenu vrijednost), **odblockirali smo put** između dva nezavisna čvora – oni su tada postali korelirani! **Kolider ne smijemo kontrolirati!**



Statističko testiranje u osnovnoj školi

- U osnovnoj školi mjerili su učenicima duljinu stopala i testirali sposobnost čitanja. Statistički podaci pokazali su da učenici s duljim stopalima bolje čitaju. Dakle, našli su značajnu korelaciju između duljine stopala i sposobnosti čitanja.
- Naravno, korelacija nije znači uzročnost (kako je govorio i Fisher). Ali, uzročnost ipak postoji, a to je u ovom slučaju dob.
- Graf odnosa između tri čvora, Dob, DS (duljina stopala) i SČ (sposobnost čitanja) pokazuje da je Dob **zajednički uzrok** ili **zbunjivač (confounder)**.
- Da bismo vidjeli postoji li zaista veza između DS i SČ (označeno sa ?), **trebamo blokirati put Dob → DS tj. kontrolirati confounder!**
U ovom slučaju, trebamo podatke gledati zasebno za svaku dob (ili školski razred).



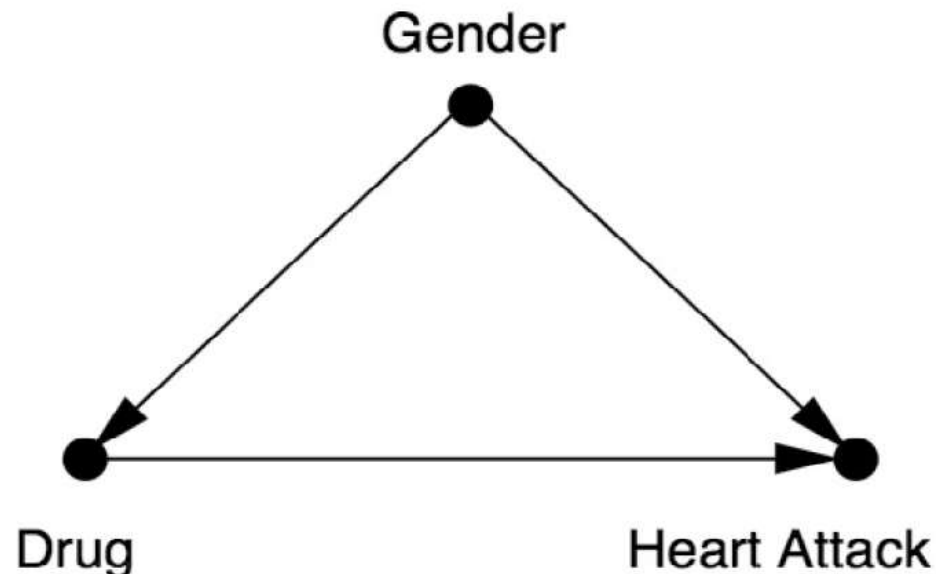
Simpson's paradox

- Paradoks je prvi otkrio statističar Edward Simpson 1951.
- Donja tablica (s fiktivnim podacima) pokazuje da je lijek loš za žene ($3/40$ srčanih udara kod žena koje su uzele lijek u odnosu na $1/20 = 2/40$ kod žena koje nisu uzele lijek) i loš za muškarce ($8/20 = 16/40$ u odnosu na $12/40$).
- No ukupno je lijek dobar ($11/60 < 13/60$).
- **Znači li to da postoji lijek BBG - loš za žene i za muškarce, dobar ako ne znamo spol ?!**

	Control Group (No Drug)		Treatment Group (Took Drug)	
	<i>Heart attack</i>	<i>No heart attack</i>	<i>Heart attack</i>	<i>No heart attack</i>
Female	1	19	3	37
Male	12	28	8	12
Total	13	47	11	49

Simpson's paradox

- Razrješenje Simpsonovog paradoksa za slučaj čudnovatog lijeka BBG dobije se tako da se kontrolira confounder, a to je spol (Gender), jer su muškarci manje skloni da uzimaju propisane lijekove i više su skloni srčanim udarima.
- Nakon kontroliranja confoundera, ostaje samo direktna zavisnost između lijeka (Drug) i srčanog udara (Heart Attack).
- Dakle, lijek je loš za žene i za muškarce, ali je loš i ukupno. (naravno, to je i intuitivno logično). Lijek BBG ne postoji!



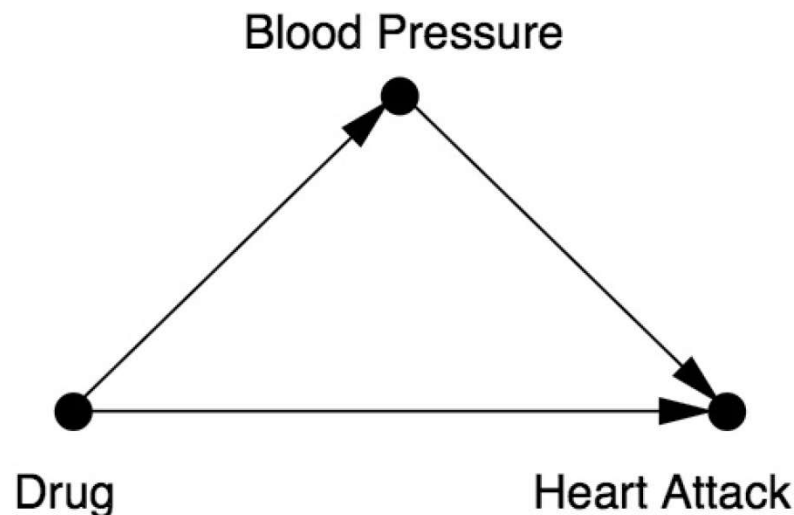
Simpson's paradox - drugi primjer

- Pogledajmo sada drugu tablicu, koja pokazuje jednake brojeve kao prethodna tablica, ali je primjer drugačiji. Tablica pokazuje da je lijek loš i kod pacijenata koji su imali nizak krvni i kod pacijenata koji su imali visok krvni tlak, ali je ukupno dobar.
- Možemo li zaključiti kao u prvom primjeru - da je, zapravo, lijek uvijek loš?

	Control Group (No Drug)		Treatment Group (Took Drug)	
	<i>Heart attack</i>	<i>No heart attack</i>	<i>Heart attack</i>	<i>No heart attack</i>
Low blood pressure	1	19	3	37
High blood pressure	12	28	8	12
Total	13	47	11	49

Simpson's paradox – drugi primjer

- Ovdje je odgovor suprotan: lijek je dobar i za one s niskim tlakom i za one s visokim tlakom, pa onda i ukupno.
- Za razliku od prvog primjera, gdje je spol bio confounder, ovdje tlak nije confounder, nego **posrednik** ili **medijator** (mediator). Lijek utječe i na tlak (snižava ga) i na srčane udare (smanjuje ih), pa lijek djeluje na smanjenje srčanih udara i direktno i indirektno (preko medijatora).
- **Ovdje bi bilo pogrešno kontrolirati medijator**, jer bi se izgubila indirektna veza. Medijator kontroliramo kad hoćemo mjeriti direktnu vezu.



Kontrolirati ili ne kontrolirati, pitanje je sad

□ Kolider ne smijemo kontrolirati

$$X \rightarrow K \leftarrow Y$$

$$X \perp\!\!\!\perp Y \mid \emptyset$$

(X je nezavisno od Y ako ne kontroliramo)

□ Confounder trebamo kontrolirati

$$X \leftarrow C \rightarrow Y$$

$$X \perp\!\!\!\perp Y \mid C$$

(X je nezavisno od Y ako C kontroliramo)

□ Medijator uglavnom ne kontroliramo

$$X \rightarrow M \rightarrow Y$$

$$X \perp\!\!\!\perp Y \mid M$$

(X je nezavisno od Y ako M kontroliramo)

osim kad želimo mjeriti direktnu vezu (između X i Y),
tj. eliminirati indirektnu vezu (preko M).

Što (ne) možemo saznati iz podataka

- Ova tri osnovna usmjerena aciklička grafa (Directed Acyclic Graph, DAG), lanac u oba smjera i račvanje, **spadaju u istu klasu ekvivalencije** – **nemoguće ih je razlikovati samo na temelju podataka:**

$$X \rightarrow M \rightarrow Y$$

$$X \leftarrow M \leftarrow Y$$

$$X \leftarrow C \rightarrow Y$$

- **Četvrti graf spada u drugu klasu:**

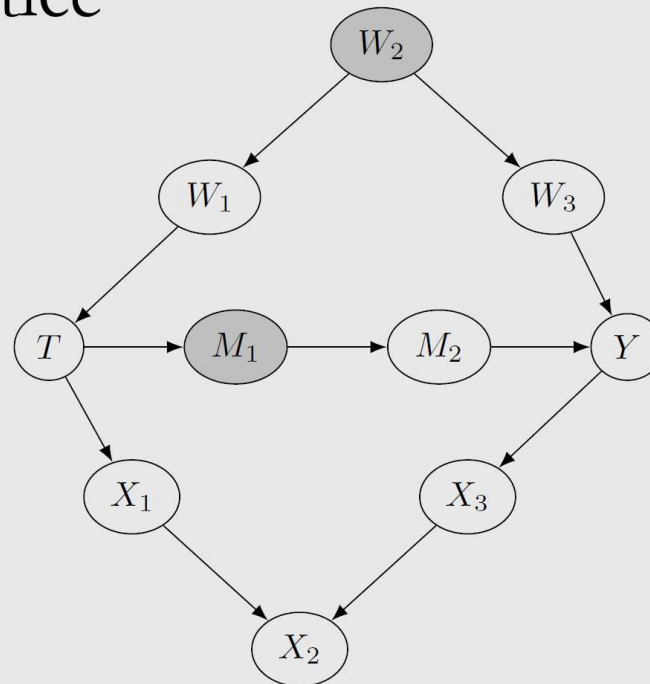
$$X \rightarrow K \leftarrow Y$$

- **Na temelju podataka možemo saznati je li riječ o prvoj ili drugoj klasi.**

d-separation

- Dva skupa čvorova X i Y su d-separirani (d-separated) skupom čvorova Z ako su **svi putovi** između (bilo kojeg čvora u) X i (bilo kojeg čvora u) Y **blokirani s Z** .
- d-separirani čvorovi (G kao graf) su i statistički nezavisni (P):
$$X \perp\!\!\!\perp_G Y \mid Z \Rightarrow X \perp\!\!\!\perp_P Y \mid Z \quad (\text{obrnuto ne vrijedi})$$
- U donjem primjeru su čvorovi T i Y d-separirani pomoću skupa čvorova $\{W_2, M_1\}$ (moguće su i druge kombinacije):

d-separation practice



do-operator i do-calculus

- **do-operator označava intervenciju, a ne opservaciju:**
 $P(Y \mid \text{do}(X)) \neq P(Y \mid X)$ - kauzalna formula $\neq$ statistička
- Npr. ako je X boja kose, a Y brzina trčanja, intervencija $\text{do}(X = \text{crna})$, kojom ofarbamo sijedu kosu u crnu, neće utjecati da se Y poboljša, za razliku od opservacije $P(Y = \text{brzo} \mid X = \text{crna})$ u odnosu na $P(Y = \text{brzo} \mid X = \text{sijeda})$.
- **Ponekad možemo zamijeniti kauzalne formule sa statističkim formulama, tj. ne moramo raditi intervencije.**
do-calculus omogućava nalaženje tih statističkih formula, ako je to moguće (ako nije, znamo da nije).
- do-calculus temelji se na tri pravila (nisu aksiomi, jer se mogu izvesti iz teorije vjerojatnosti) i ima ove osobine:
 - pouzdan
 - potpun: ako se do-operator može pretvoriti u statistički izraz, do-calculus to može
 - postoji algoritam koji to radi u polinomijalnom vremenu.

Literatura (dio)

- Bolstad W.M., Curran J. M. (2017):
Introduction to Bayesian statistics, John Wiley & Sons
- Ertel W. (2017): Introduction to Artificial Intelligence, Springer
- J. Launchbury (2017), A DARPA Perspective on AI,
<https://www.darpa.mil/about-us/darpa-perspective-on-ai>
- Neal B. (2020): Introduction to Causal Inference, predavanja
<https://www.bradyneal.com/causal-inference-course>
- Peters J., Janzing D., Schoelkopf B. (2017):
Elements of Causal Inference, The MIT Press
- Pearl J., Mackenzie D. (2018): The Book of Why –
The New Science of Cause and Effect, Basic Books
- Šikić Z. (2023): Što je vjerojatnost, 9 članaka,
<https://ideje.hr/category/rasprave/>
- Tetrad – softver za kauzalno zaključivanje,
<https://www.cmu.edu/dietrich/philosophy/tetrad/>
- JASP – softver za bejsovsku i frekventističku statistiku,
<https://jasp-stats.org/>

Na CASE 2020 prikazali smo "Mrav i med na prizmi i valjku" (verzija 1)

<http://www.istrattech.hr/mrav-i-med-na-pravilnim-prizmama-i-valjku-verzija-4/>
(napomena: slika prikazuje realizaciju verzije 3)

