



Probabilistično razmišljanje in Programiranje

Erik Štrumbelj

Oris vsebine



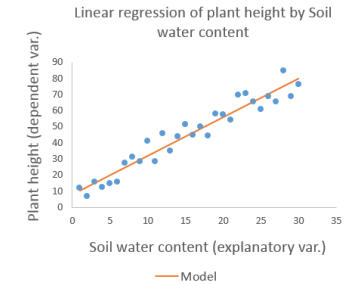
Zakaj naj mi bo mar za probabilistično programiranje?

1. Temelj sodobnega statističnega modeliranja in strojnega učenja.

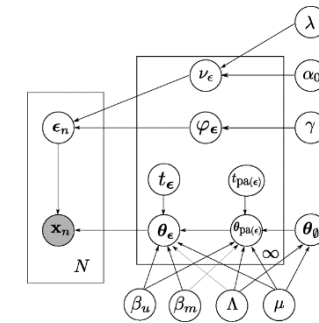
Obvezno orodje za vsakogar, ki se želi resno ukvarjati s kvantitativno analizo podatkov!

2. Zelo koristen način razmišljanja.

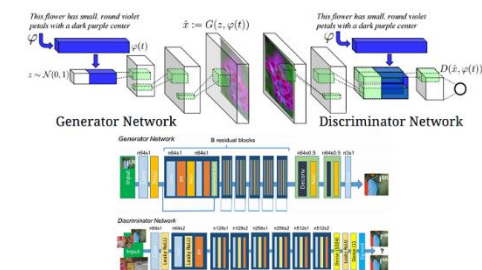
3. Prihodnost “podatkovnega inženirstva”.



uporabna statistika



probabilistični grafični modeli



(generativno) globoko učenje

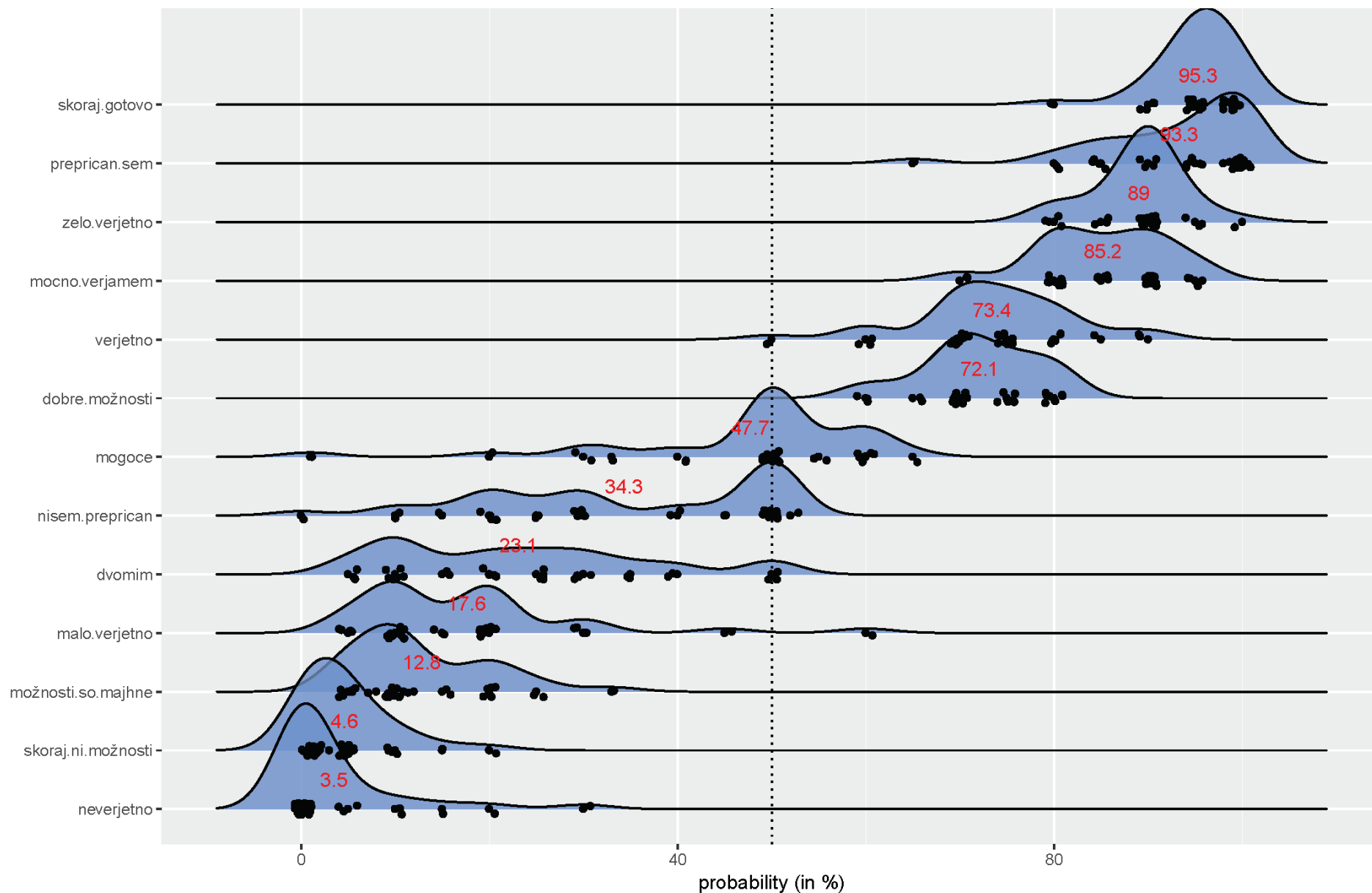
Interaktivni test opreme za delavnico

1

Negotovost in
probabilistično razmišljanje

Q: Ali bo naslednjo sredo
v Ljubljani deževalo?

Izrazi verjetnosti v naravnem jeziku



Q: Kako toplo ($^{\circ}\text{C}$) bo jutri
opoldne v Ljubljani?

Naravni jezik je nekonsistenten, nenatačen in premalo ekspresiven za resno kvantitativno delo!

- **Dobra novica** Primeren jezik so že razvili!
- **Slaba novica** Gre za teorijo verjetnosti – matematiki se ne moremo izogniti.
- **Dobra novica** Ni se nam potrebno naučiti niti vse dodiplomske verjetnosti¹ – potrebujemo le verjetnost kot jezik, računal pa bo računalnik.

¹ Kar pa ne pomeni, da nam ne bo koristilo! Verjetnost je osnova kvantitativne analize podatkov.

Gramatika verjetnosti

Verjetnost P (pogosto Pr) je funkcija, ki dogodkom prireja numerične vrednosti in zadošča tem aksiomom:

$$A1 \quad P(A) \geq 0.$$

$$A2 \quad P(\Omega) = 1.$$

$$A3 \quad P(A_1 \cup A_2 \cup A_3 \cup \dots) = \sum_{i=1}^{\infty} P(A_i),$$

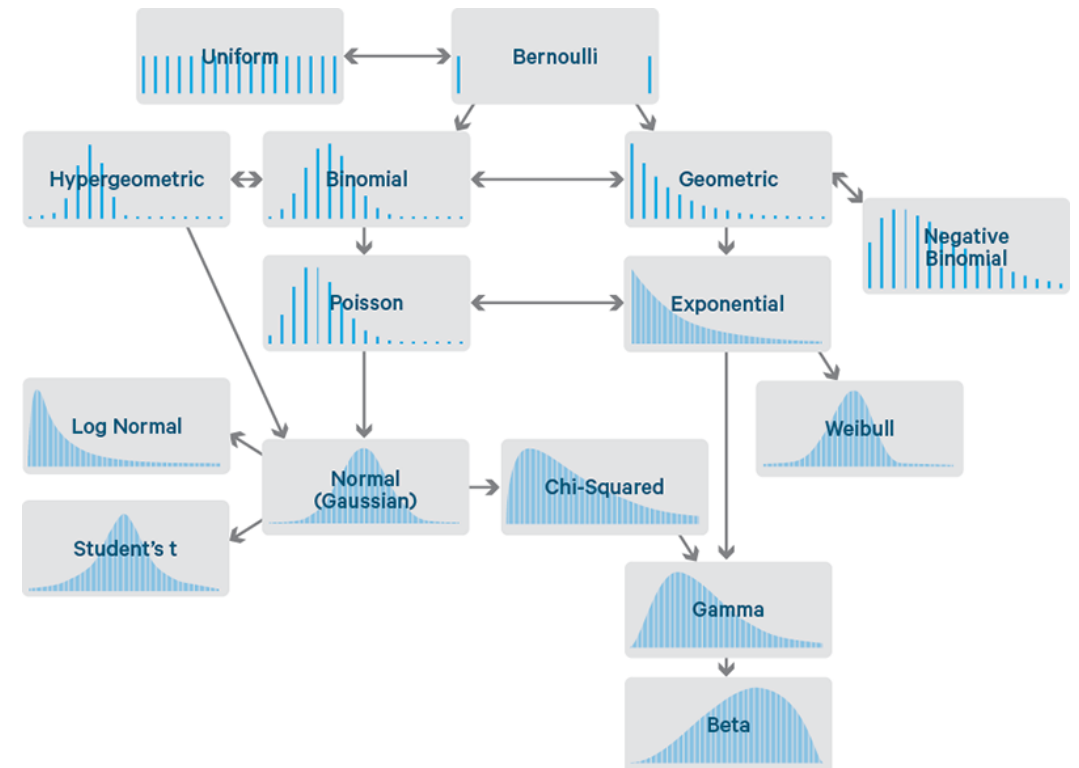
za poljubno sekvenco disjunktih dogodkov.

Definicija **pogojne verjetnosti**:

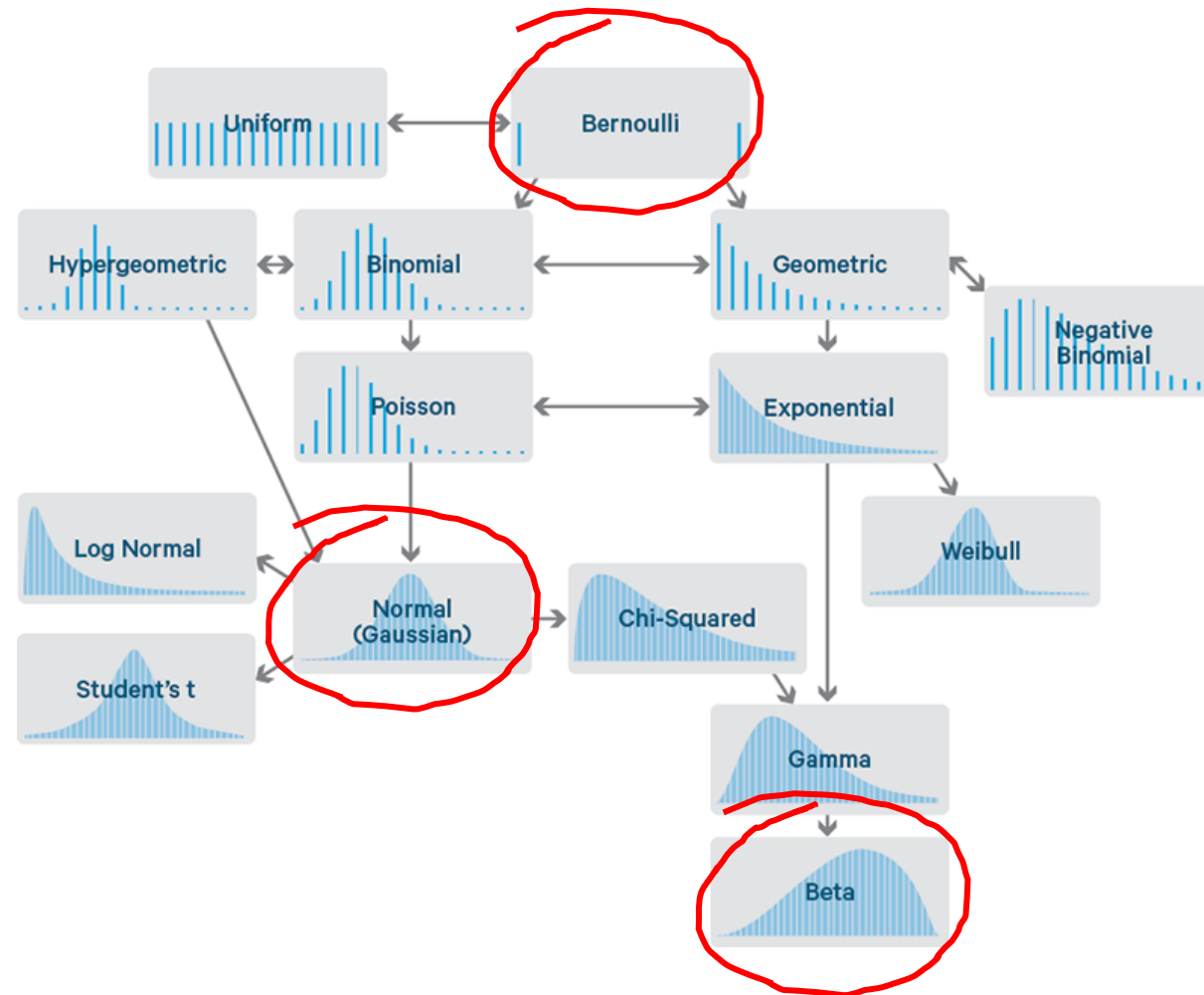
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Porazdelitve

- Porazdelitve so elementarni **izrazi** **probabilističnega razmišljanja** in
- osnovni gradniki statističnih modelov.
- Porazdelitve so v skladu s pravili teorije verjetnosti, zato so **konsistentne** in **natančne** probabilistične izjave.
- Več kot vemo o porazdelitvah, bolj bogato se lahko izražamo.

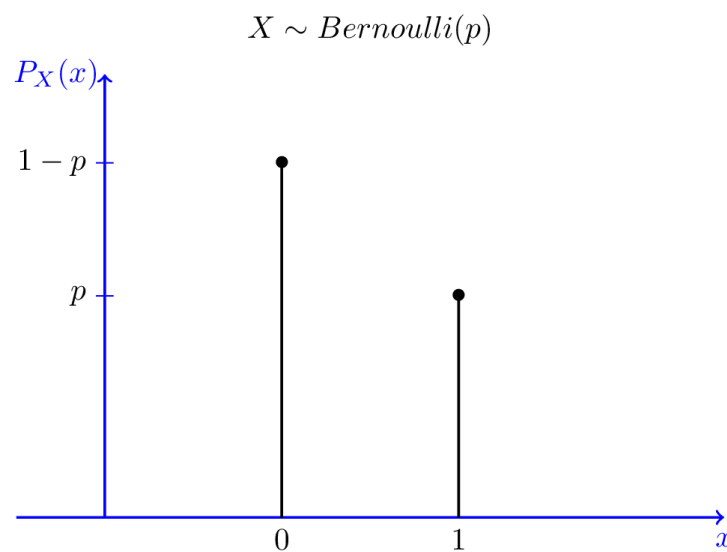


Beseda na dan ...



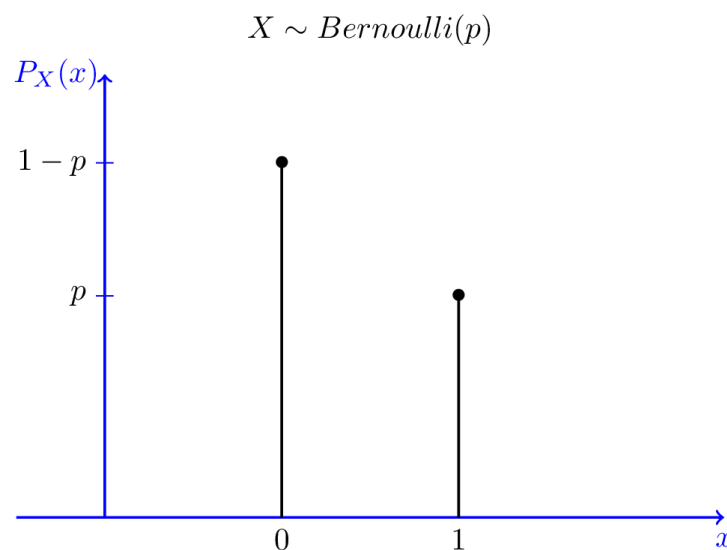
Bernoullijeva porazdelitev

distribution	pmf	mean	variance
Bernoulli(p)	$p^x(1-p)^{1-x}; x = 0, 1; p \in (0, 1)$	p	$p(1-p)$



Bernoullijeva porazdelitev

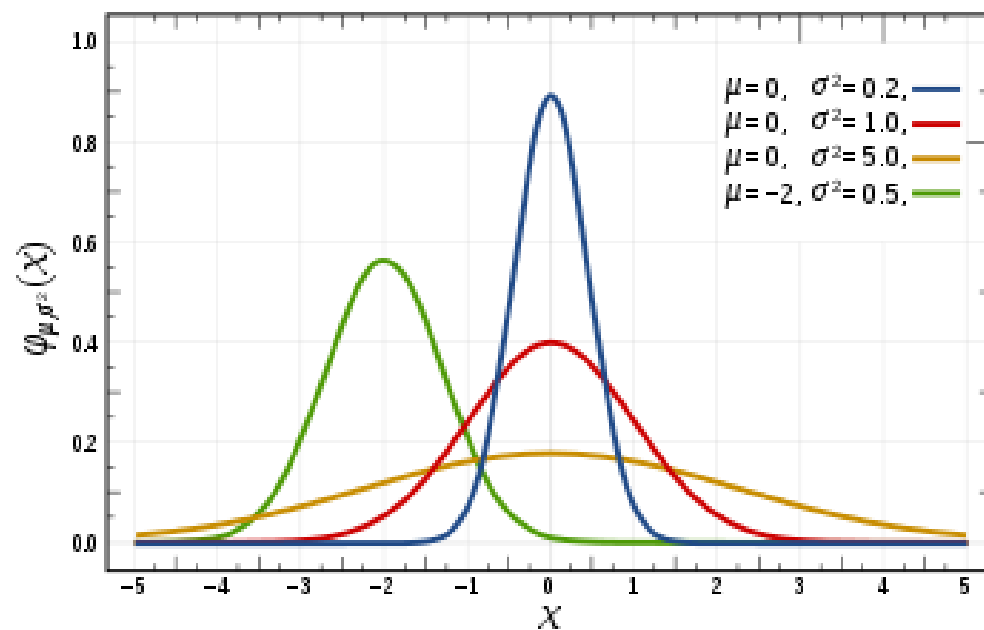
distribution	pmf	mean	variance
Bernoulli(p)	$p^x(1 - p)^{1 - x}; x = 0, 1; p \in (0, 1)$	p	$p(1 - p)$



Q: Ali bo naslednji teden v Ljubljani deževalo?

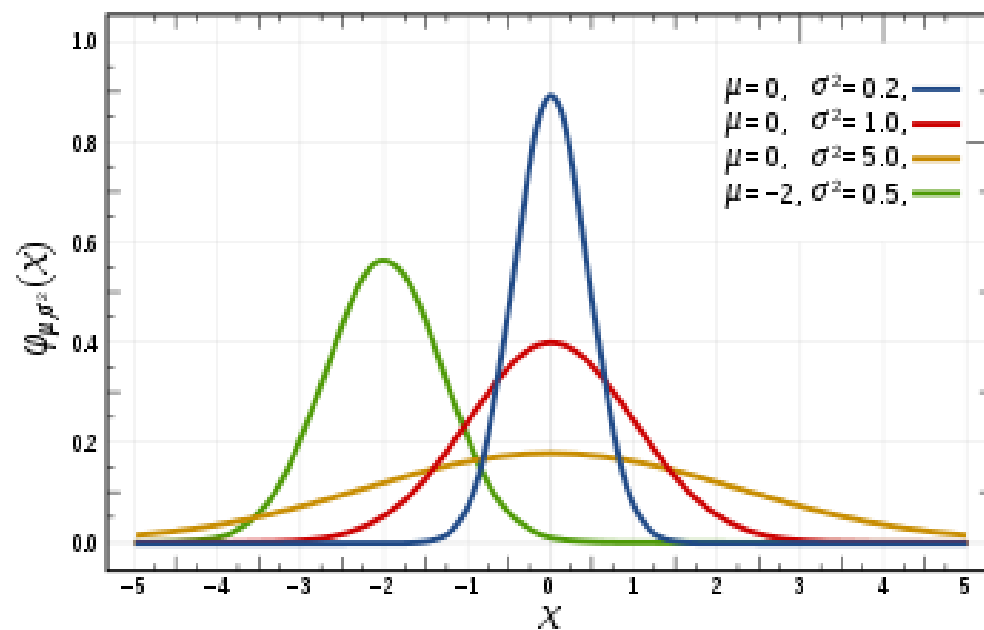
Normalna (Gaussova) porazdelitev

distribution	pdf	mean	variance
Normal(μ, σ^2)	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \sigma > 0$	μ	σ^2



Normalna (Gaussova) porazdelitev

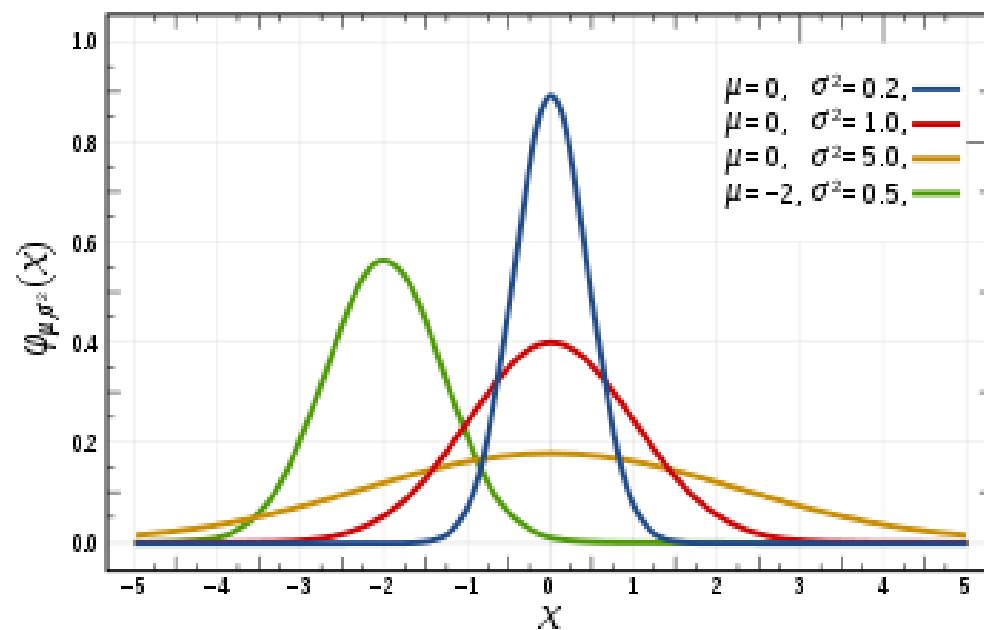
distribution	pdf	mean	variance
Normal(μ, σ^2)	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \sigma > 0$	μ	σ^2



Q: Kako toplo (°C) bo jutri opoldne v Ljubljani?

Normalna (Gaussova) porazdelitev

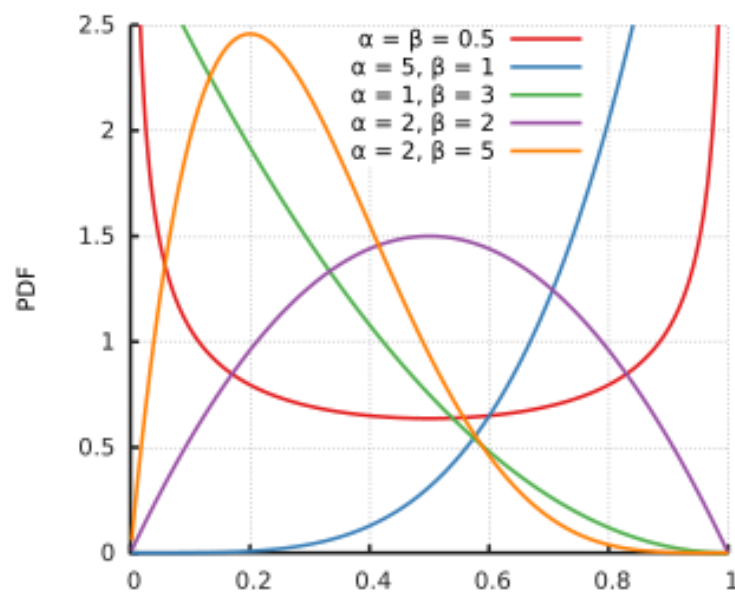
distribution	pdf	mean	variance
Normal(μ, σ^2)	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \sigma > 0$	μ	σ^2



Q: Kako toplo (°C) je bilo na današnji dan pred 50 leti?

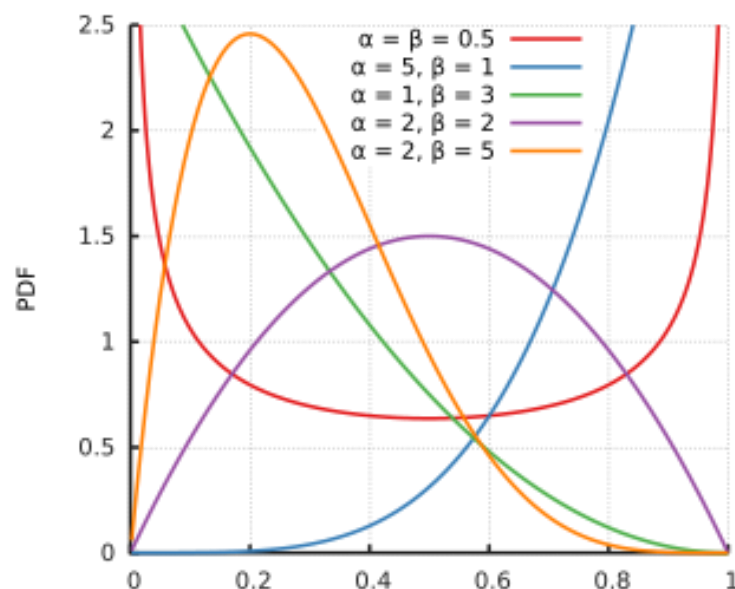
Porazdelitev Beta

distribution	pdf	mean	variance
Beta(α, β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}; x \in (0, 1), \alpha, \beta > 0$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$



Porazdelitev Beta

distribution	pdf	mean	variance
Beta(α, β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}; x \in (0, 1), \alpha, \beta > 0$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$



Q: Kolikšna je verjetnost, da naslednji teden v LJ dežuje?

Preizkus probabilističnega razmišljanja

To so izidi 10 metov (morda nepoštenega) kovanca:

c c g c c g c c c g (?)

Q1: Je enajsti met **c**ifra ali **g**rb?

Q2: Kolikšna je verjetnost p , da na tem kovancu pade grb?

Q3: Je kovanec pošten? Poštenost je npr., da je p med 48% and 52%.

V razmislek ...

Verjetnost je koherenten in natančen jezik za izražanje negotovosti:

- Če ne sledimo zakonom verjetnosti, nas nihče ne bo razumel!
- Sicer pa so probabilistične izjave lahko subjektivne ali navidez popolnoma nesmiselne.
- Precej naravno nam je, da imamo verjetnostno mnenje o stvareh, ki niso naključne. Naključje je samo eden izmed virov negotovosti (in ne preveč pogost).

Uporaba verjetnosti za izražanje negotovosti je bistvo bayesovskega pogleda na statistično sklepanje!

2

Statistično
modeliranje

Model =

**Hipoteza, kako so
nastali naši podatki.**

**Ni modeliranja
brez modela.**

Ne, resno, ni.

Q: Zapišite 1 metodo iz statistike ali strojnega učenja, ki se uporablja za napovedovanje, razpoznavanje vzorcev, gručenje, testiranje hipotez, ipd.

Zaporedje enic in ničel (= podatki):

10010010101101100111111111101

Statistični model (= poskus statistične interpretacije):

Zaporedje je nastalo s 30 neodvisnimi meti kovanca z neznano verjetnostjo enice θ .

Predhodno mnenje o parametrih modela:

Nimam pojma, koliko je θ , zato ne bom izrazil preference do nobene vrednosti θ .

Statistično sklepanje (= učenje):

Pri vseh teh predpostavkah in upoštevajoč zakone verjetnosti, kakšno mora biti moje mnenje o θ , ko vidim podatke?

Zaporedje enic in ničel (= podatki):

100100101011011001111111111101

$$y_1, \dots, y_n \quad y_i \in \{0, 1\}$$

Statistični model (= poskus statistične interpretacije):

Zaporedje je nastalo s 30 neodvisnimi meti kovanca z neznano verjetnostjo enice θ .

$$y_1, y_2, \dots, y_n | \theta \sim_{\text{iid}} \text{Bernoulli}(\theta)$$

Predhodno mnenje o parametrih modela:

Nimam pojma, koliko je θ , zato ne bom izrazil preference do nobene vrednosti θ .

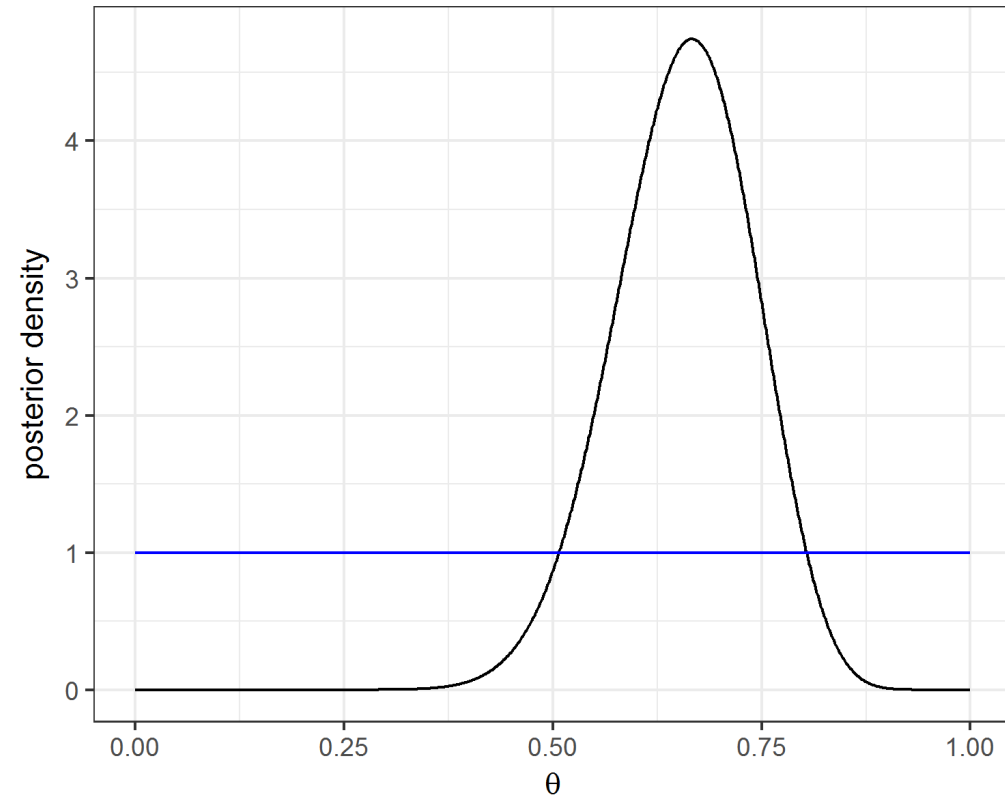
$$\theta \sim \text{Beta}(1, 1)$$

Statistično sklepanje (= učenje):

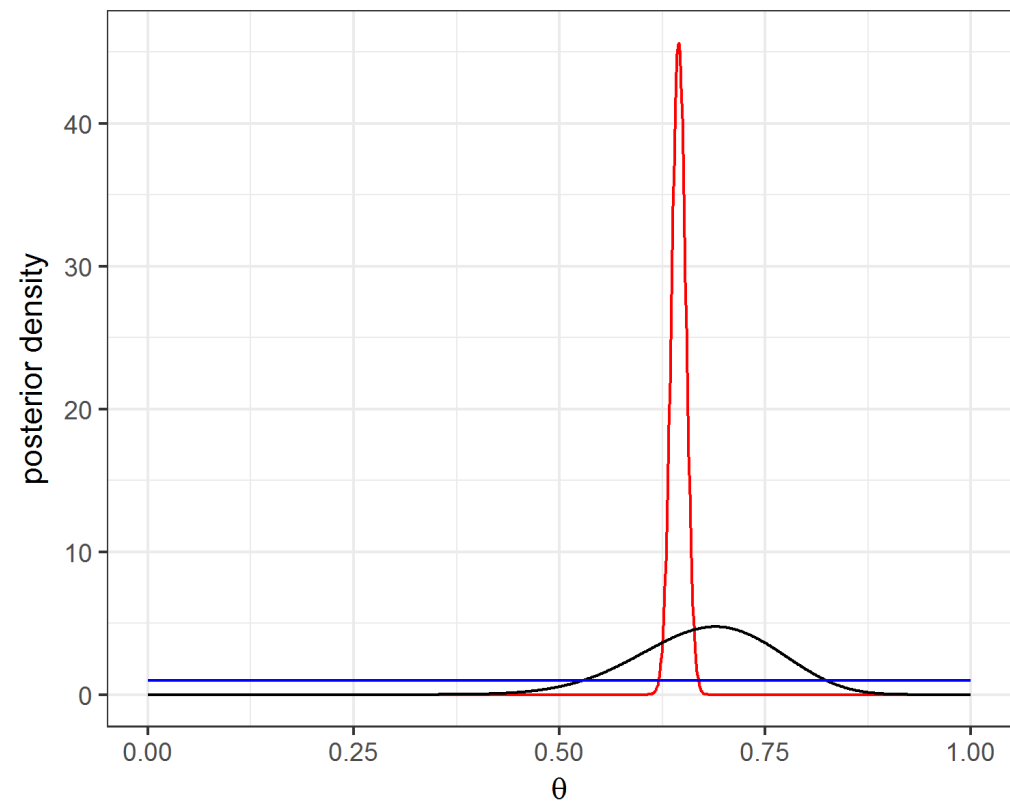
Pri vseh teh predpostavkah in upoštevajoč zakone verjetnosti, kakšno mora biti moje mnenje o θ , ko vidim podatke?

$$p(\theta | y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y | \theta) p(\theta)}{\int p(y | \theta) p(\theta) d\theta}$$

$$\theta | y_1, \dots, y_n \sim \text{Beta}(\sum y_i + 1, n - \sum y_i + 1)$$



Naše mnenje o θ **prej** in **potem**, ko smo videli zaporedje, ki vsebuje 20 enic in 10 ničel.



Naše mnenje o θ **potem**, ko smo videli zaporedje, ki vsebuje 20 enic in 10 ničel.

3

Probabilistično
programiranje

Probabilistični programski jezik (PPL) je programski jezik, ki je zasnovan za opisovanje probabilističnih modelov in računsko sklepanje iz teh modelov.

Vir: Wikipedia

Probabilistični programski jezik nam omogoča, da se osredotočimo na modeliranje in preskočimo matematične in računske probleme pri sklepanju.

Dva primera imperativnega programiranja

```
← → |  |  ☐ Source on Save |  |  |  | 
```

```
1 # Bubble Sort
2 sort <- function(x) {
3   n <- length(x)
4
5   for (k in n:2) {
6     i <- 1
7     while (i < k) {
8       if (x[i] > x[i+1]) {
9         temp <- x[i+1]
10        x[i+1] <- x[i]
11        x[i] <- temp
12      }
13      i <- i + 1
14    }
15  }
16  x
17 }
```

```
← → |  |  ☐ Source on Save |  |  |  | 
```

```
1 # Generate 30 Bernoulli variables
2 bernoulli <- function(p) {
3   x <- c()
4
5   for (i in 1:30) {
6     if (runif(1) > p) {
7       x <- c(x, 0)
8     } else {
9       x <- c(x, 1)
10    }
11  }
12  x
13 }
14
```

Imperativno programiranje

in

Statistično modeliranje

```
← → | | Source on Save | |  
1 # Generate 30 Bernoulli variables  
2 bernoulli <- function(p) {  
3   x <- c()  
4  
5   for (i in 1:30) {  
6     if (runif(1) > p) {  
7       x <- c(x, 0)  
8     } else {  
9       x <- c(x, 1)  
10    }  
11  }  
12  x  
13 }  
14
```

- Podane imamo **vhodne podatke** in **parametre**,
- sprogramiramo algoritem, ki generira zahtevane izhodne podatke.

Sklepanje o relativni frekvenci tega zaporedja

100100101011011001111111111101

$$y_1, \dots, y_n \quad y_i \in \{0, 1\}$$

$$y_1, y_2, \dots, y_n | \theta \sim_{\text{iid}} \text{Bernoulli}(\theta)$$

$$\theta \sim \text{Beta}(1, 1)$$

- Podane imamo **vhodne** in **izhodne podatke**,
- opišemo generator, ki naj bi generiral podatke,
- sklepamo o najbolj verjetnih vrednostih parametrov.

4

MCMC

(Markov Chain Monte Carlo)

5

Programski jezik

Stan

Kaj je Stan?

- Orodje za učinkovito Bayesovo statistično modeliranje:
 - Programski jezik
 - Matematična knjižnica z avtomatskim odvajanjem
 - Algoritmi MCMC
- Najlažje ga uporabljamo preko vmesnikov (RStan, PyStan...).
- Aktivna skupnost, odličen priročnik, konstanten razvoj:
 - <https://mc-stan.org/>
 - <https://discourse.mc-stan.org/>



Stan program je organiziran v bloke

- **data** – blok, v katerem deklariramo vhodne podatke. Vrednosti vhodnih podatkov pripravi uporabnik.
- **parameters** – blok, v katerem deklariramo parametre, ki jih želimo oceniti (kateri parametri našega statističnega modela nas zanimajo).
- **model** – opis statističnega modela.

```
/*  
Primer komentarja, ki obsega  
več vrstic.  
*/  
  
data {  
    // tukaj definiramo vhodne podatke  
}  
  
parameters {  
    // parametri modela, ki jih želimo oceniti  
}  
  
model {  
    // sem spada statistično modeliranje  
}
```

Osnovni tipi spremenljivk

- **int** – celo število
`int n;`
- **real** – realno število
`real r;`
- **seznam** (array) – seznam celih ali realnih števil
`int a[10]; real b[n];`
- **matrika** (matrix) – 2D seznam [vrstice, stolpci]
`int A[10, 10];`
- **vector** – vektor realnih števil (optimiziran seznam)
`vector[n] v;`
- **simplex** – vektor pozitivnih realnih števil, ki se seštevajo v 1
`simplex[n] s;`
- (skoraj) vsem spremenljivkam lahko določimo zgornjo in spodnjo mejo
`real<lower=0> sigma;`
`real<lower=0,upper=1> success_rate;`

Porazdelitve

- **Bernoulli**

y je vektor "uspehov" (1) in "neuspehov" (0)
 θ (theta) predstavlja verjetnost uspeha

$y \sim \text{bernoulli}(\theta);$

- **beta**

y je vektor realnih števil med 0 in 1
 α, β parametra porazdelitve

$y \sim \text{beta}(\alpha, \beta);$

- **normal**

y je vektor realnih števil
 μ, σ sta upanje oziroma varianca

$y \sim \text{normal}(\mu, \sigma);$

- porazdelitve uporabimo tudi za vnašanje predznanja o določenih parametrih modela

$\theta \sim \text{beta}(1, 1);$

6

Praktični
primeri