

Intro to Bayesian Thinking

With Dogs and a Serial Killer

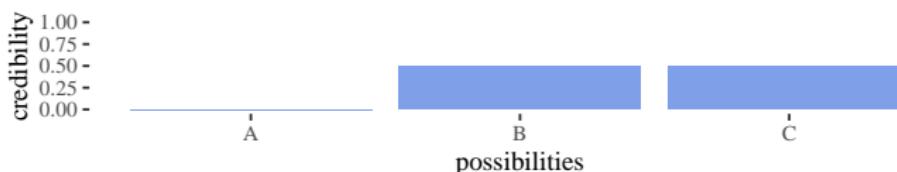
Nikodem Lewandowski



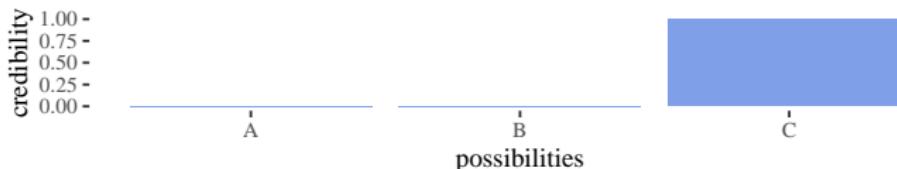
Sherlock's naivete



Posterior 1



Posterior 2



A rather unhelpful piece of advice

“...when you have eliminated the impossible, whatever remains, however, improbable, must be the truth.”

Trouble in paradise

- Data have only probabilistic relations to hypotheses
Many people may have similar footprints
- Measurements only probabilistically narrow down the range
We mathematically can describe the footprints up to some level of precision
- Association does not directly translate into causation
There may be various confounding factors explaining why people who received a given drug have lower blood pressure
- There often is a natural variation
The weight of a newborn baby may vary naturally due to genetics and environmental factors, rather than a specific cause

We love dogs

- Think about a scenario where there is a room with 4 dogs
- Only two possible breeds are: huskies and doges
- You can't see the room, the door opens, and one dog walks in and returns to the room
- We count the observations of breeds
- The task: let's assess how many dogs are inside (Bayesian style!)
- Let's formulate possible hypotheses and asses how likely they are!



Bayesian Counting

Five Doge hypotheses

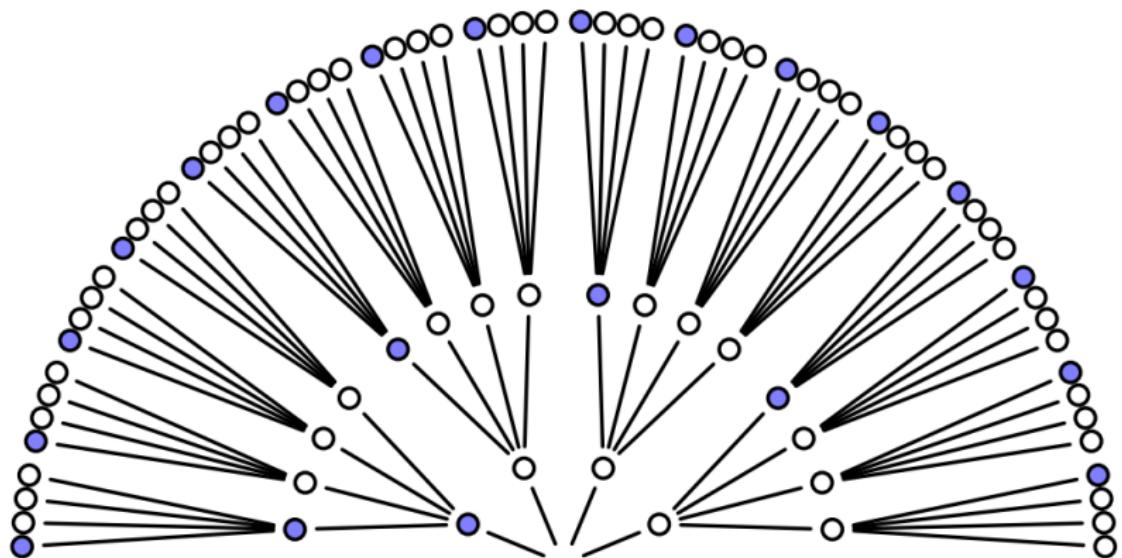


The evidence

Our first evidence: husky, doge, husky

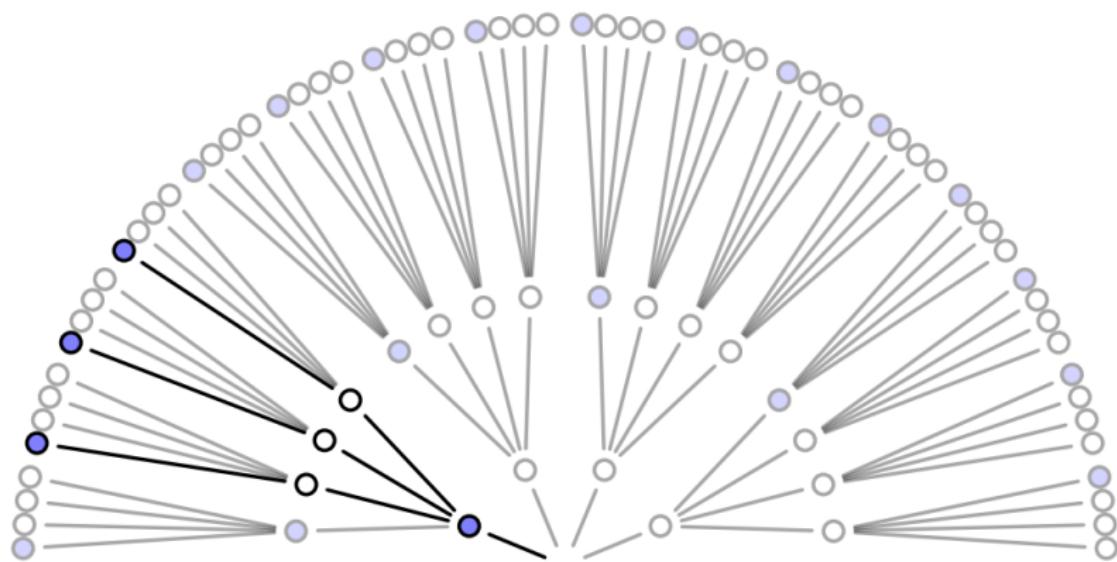


Discovering the possibilities



- color points - huskies

Discovering the possibilities (highlited observed paths)



- color points - huskies

Ways dogs could be (likelihoods)

Ways to observe (h,d,h)

					h	d	h	(h,d,h)
four -					0	4	0	0
three -					1	3	1	3
two -					2	2	2	8
one -					3	1	3	9
zero -					4	0	4	0

Updating with new observations

Ways to observe (h,d,h,h)

	h	d	h	(h,d,h)	h	(h,d,h,h)				
four -					0	4	0	0	0	0
three -					1	3	1	3	1	3
two -					2	2	2	8	2	16
one -					3	1	3	9	3	27
zero -					4	0	4	0	4	0

Now with probabilities

$$\text{Plausibility of } p \text{ after } D_{\text{new}} = \frac{\text{Ways } p \text{ can produce } D_{\text{new}} \times \text{Prior Plausibility } p}{\text{sum of products}}$$

	p	ways0	ways0pr	ways1	ways1pr
d,d,d,d	0.00	0	0.00	0	0.0000000
d,d,d,h	0.25	3	0.15	3	0.0652174
d,d,h,h	0.50	8	0.40	16	0.3478261
d,h,h,h	0.75	9	0.45	27	0.5869565
h,h,h,h	1.00	0	0.00	0	0.0000000

More dogs, Bayesian style!

Probability Mass Function (PMF) for a Binomial Distribution:

$$P(D = d, H = h|\theta) = \frac{(d+h)!}{d!h!} \theta^d (1-\theta)^h$$

Product Rule of Probability:

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

Joint Probability:

$$P(d, h, \theta) = P(d, h|\theta)P(\theta)$$

$$P(d, h, \theta) = P(\theta|d, h)P(d, h)$$

$$P(\theta|d, h)P(d, h) = P(d, h|\theta)P(\theta)$$

Bayes' Theorem:

$$\underbrace{P(\theta|d, h)}_{\text{Posterior}} = \frac{\overbrace{P(d, h|\theta)}^{\text{Likelihood}} \overbrace{P(\theta)}^{\text{Prior}}}{\underbrace{P(d, h)}_{\substack{\text{(Average)} \\ \text{Data}}}}$$

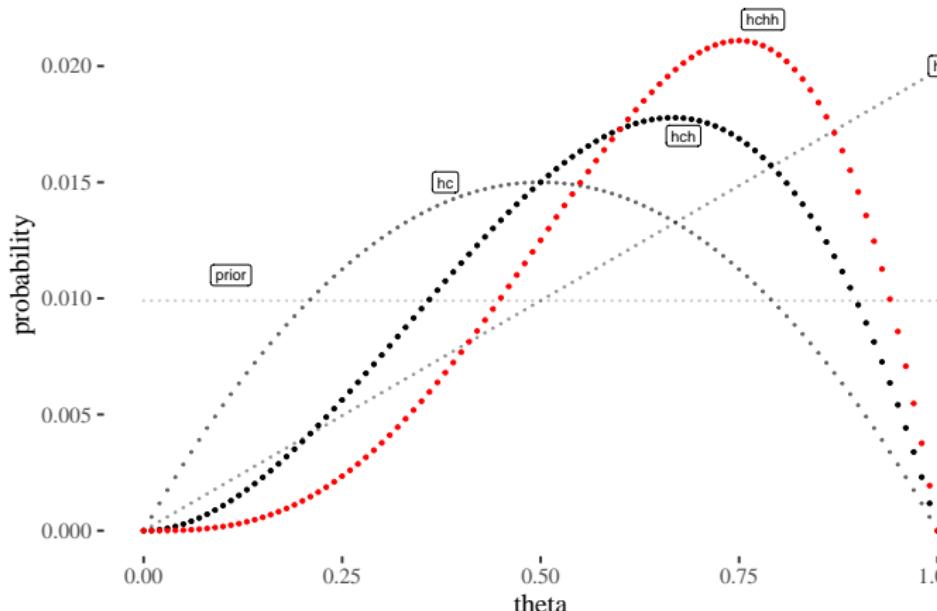


The underlying mechanism

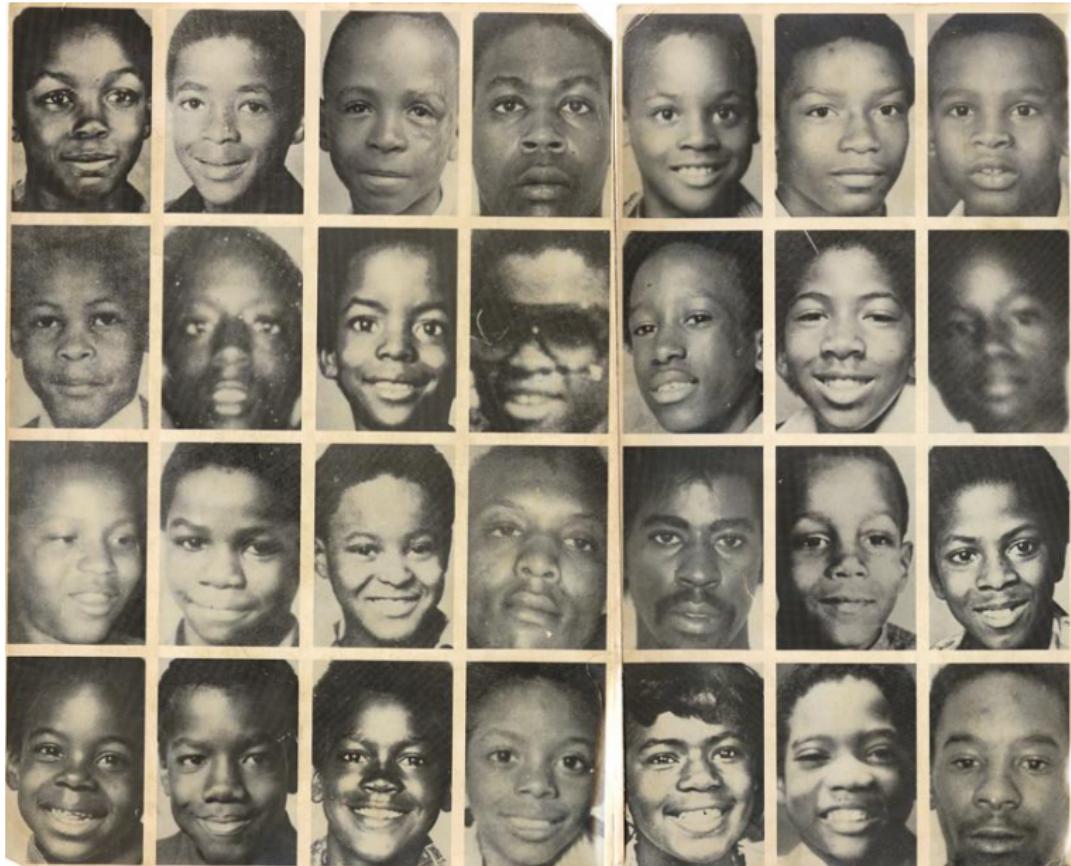
plausibility(hypothesis $n|\text{data}$) \propto

ways hypothesis n can produce data \times prior plausibility of hypothesis n

Proportion learning from flat prior



Wayne Williams case



University
of Gdańsk

Two items of evidence

- Dog hair evidence, random match probability (RMP) is about 0.0256.
- Human hair evidence, RMP is about 0.0253

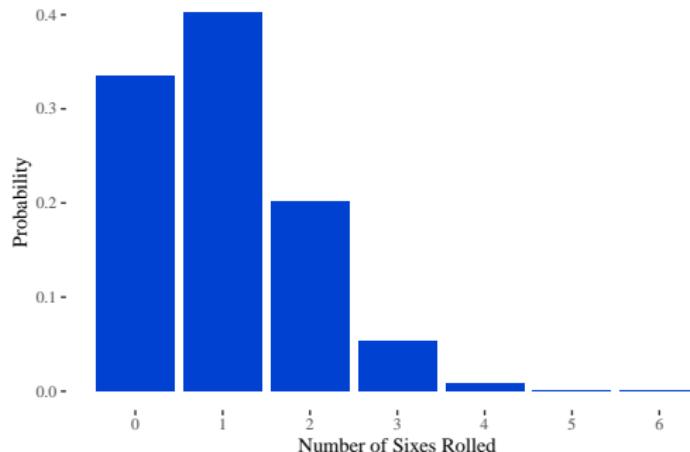
Questions that come to mind?



Binomial Dist. to the Rescue!

```
number_of_rolls <- 6 # Number of dice rolls  
success_probability <- 1/6 # Probability of rolling a 6  
successes <- 0:number_of_rolls # values from 0 to 6  
  
probabilities <- dbinom(x = successes, size = number_of_rolls,  
                           prob = success_probability)
```

Binomial Distribution for Rolling a 6 (6 Rolls)



- $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
- k successes in n independent trials with probability p

The fur testimony (grid approximation)

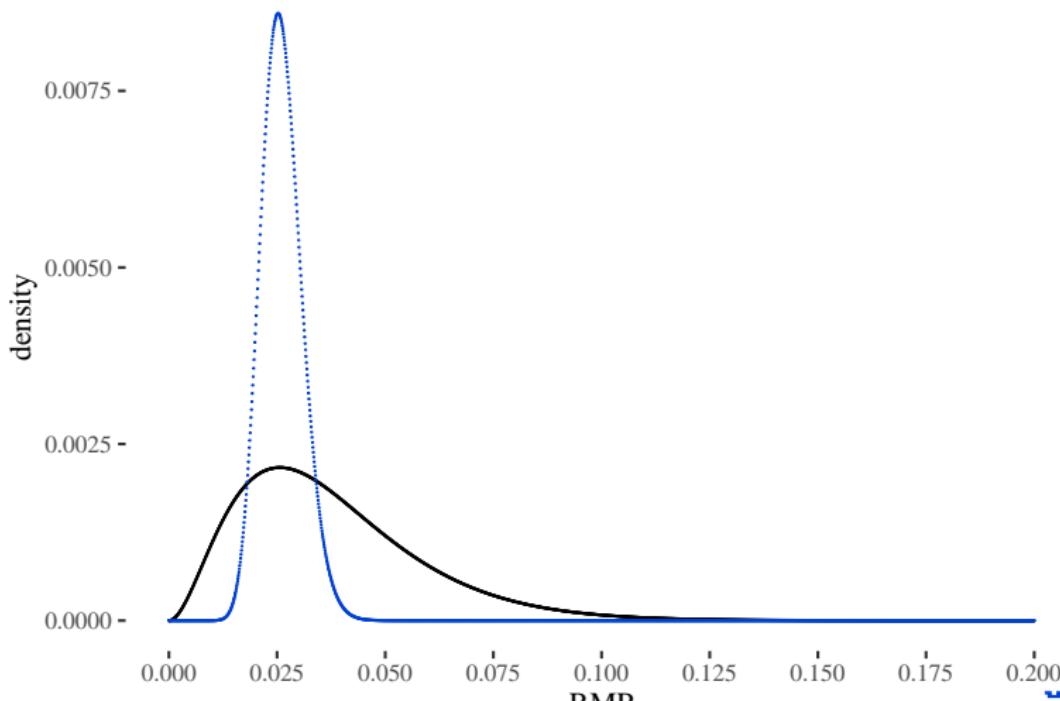
```
theta <- seq(0,1, length.out = 10001)
prior <- rep(1/10001, 10001)

likelihoodDog <- dbinom(2,78, theta)
likelihoodHuman <- dbinom(29,1148, theta)

posteriorDogUnst <- likelihoodDog * prior
posteriorHumanUnst <- likelihoodHuman * prior

posteriorDog <- posteriorDogUnst/sum(posteriorDogUnst)
posteriorHuman <- posteriorHumanUnst/sum(posteriorHumanUnst)
```

The fur testimony (grid approximation)



Steps of Bayesian data analysis

- ① Identify the data, variables, predictors
- ② Define a descriptive and appropriate model
- ③ Specify a prior distribution (over parameters)
- ④ Use Bayesian inference to reallocate credibility in light of the training data
- ⑤ Test whether the posterior predictions are reasonable as compared to validation data

Let's get back to the Dogs

Binomial Distribution:

$$H \sim \text{Binomial}(N, \theta)$$

$$\theta \sim \text{Uniform}(0, 1)$$

We use **variables** here:

- $N = h + d$, number of observations
- target variable H , the proportion of Huskies
- θ that represents our prior belief about the distribution of huskies (set of hypotheses)

Build your first model!

```
dogsModel <- quap(  
  alist(  
    h ~ dbinom( h + d , theta) ,  
    theta ~ dunif(0,1)  
  ) ,  
  data=list(h=50,d=13))
```

- `quap`: is an engine of our model (quadratic approximation)
- `h`: a target that we want to model with a binomial distribution
- `theta`: a prior
- in `data` we represent our observations, 50 huskies and 13 dogs

Build your first model!

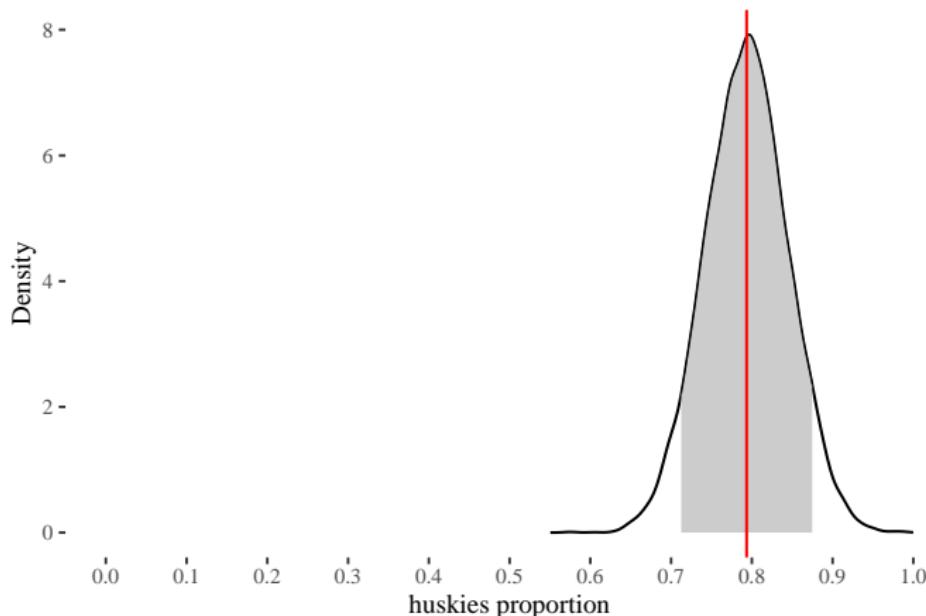
```
precis(dogsModel) # summary table
```

```
mean          sd      5.5%     94.5%
```

```
theta 0.7936496 0.05098465 0.7121663 0.8751329
```

```
vis_post_samples_dogs # importing visualization
```

Summary of 1k Samples



What we did?

- We created a model that generates probabilities for hypotheses regarding the proportion of huskies
- A model consists of variables (prior, likelihood, posterior), and an **engine** (to perform calculations)
- We represent variables with distributions rather than single point estimates
- Prior: a belief about the likelihood of hypotheses before any observations (represented by a uniform prior)
- Likelihood: a set of observations given the hypotheses
- Posterior: the final distribution indicating the likelihood of our hypotheses

Engines

- The function of the motor is to update our prior based on the given data
- Manual counting is not only inefficient but sometimes practically impossible

Possible methods:

- Grid approximation: This method involves limiting the number of hypotheses to a finite set of values
- QUAP: It operates on the premise that the posterior distribution will be nearly Gaussian or ‘normal’. Gaussian approximation is called “quadratic approximation” because the logarithm of a Gaussian distribution forms a parabola
- MCMC methods (Markov Chain Monte Carlo): Instead of attempting to compute or approximate the posterior distribution directly, MCMC techniques merely draw samples from the posterior. We will use the *Ulam* function in R



STANISŁAW MARCIN ULAM

UR. 13 IV 1909 R., ZM. 13 V 1984 R.

$$f: V \rightarrow W$$



$$\frac{(+1)}{\sqrt{2^2}} \times \text{MATEMATYK}$$

- 1** STUDIOWAŁ MATEMATYKĘ I FIZYKĘ NA POLITECHNICE LWOWSKIEJ, GDE UZYSKAŁ PIERWSZY DOKTORAT NA WYDZIALE OGÓLNYM.
- 2** W LATACH 1943–47 UCZESTNICZYŁ W PRACACH ZESPOŁU W OSRODKU BADAŃ JĄDROWYCH W LOS ALAMOS W RAMACH TZW. PROJEKTU „MANHATTAN”
- 3** W LATACH 1955–67 BYŁ DORADCĄ NAUKOWYM OŚRODKA LOS ALAMOS
- 4** JEST WSPÓŁTWÓRCĄ AMERYKAŃSKIEJ BOMBY TERMOJĄDROWEJ
- 5** W PROJEKCIE BOMBY WODOROWEJ ZMODYFIKOWAŁ JEJ KSZTAŁT, PROPONUJĄC WŁASNE ORYGINALNE ROZWIĄZANIE ZNANE JAKO KONFIGURACJA TELLERA-ULAMA
- 6** MA DOKONANIA W MATEMATYCE I FIZYCE MATEMATYCZNEJ, W DZIEDZINACH TOPOLOGII, TEORII MNOGOŚCI, TEORII MIARY I PROCESÓW GALAŻKOWYCH
- 7** BYŁ JEDNYM Z PIERWSZYCH NAUKOWCÓW W PEŁNI WYKORZYSTUJĄCYCH W SWOICH PROJEKTACH KOMPUTERY
- 8** JEST TWÓRCĄ METODY MONTE CARLO, STOSOWANEJ DO MODELOWANIA MATEMATYCZNEGO PROCESÓW ZBITY ZŁOŻONYCH, ABY MOŻNA BYŁO PRZEWIDZIEĆ ICH WYNIKI ZA POMOCĄ PODEJŚCIA ANALITYCZNEGO



17	86	93	35	32	12
18	27	16	11	14	13
19	28	17	12	15	14
20	29	18	13	16	15
21	30	19	14	17	16
22	31	20	15	18	17
23	32	21	16	19	18
24	33	22	17	20	19
25	34	23	18	21	20
26	35	24	19	22	21
27	36	25	20	23	22
28	37	26	21	24	23
29	38	27	22	25	24
30	39	28	23	26	25
31	40	29	24	27	26
32	41	30	25	28	27
33	42	31	26	29	28
34	43	32	27	30	29
35	44	33	28	31	30
36	45	34	29	32	31
37	46	35	30	33	32
38	47	36	31	34	33
39	48	37	32	35	34
40	49	38	33	36	35
41	50	39	34	37	36
42	51	40	35	38	37
43	52	41	36	39	38
44	53	42	37	40	39
45	54	43	38	41	40
46	55	44	39	42	39
47	56	45	40	43	40
48	57	46	41	44	40
49	58	47	42	45	40
50	59	48	43	46	40
51	60	49	44	47	40
52	61	50	45	48	40
53	62	51	46	49	40
54	63	52	47	50	40
55	64	53	48	51	40
56	65	54	49	52	40
57	66	55	50	53	40
58	67	56	51	54	40
59	68	57	52	55	40
60	69	58	53	56	40
61	70	59	54	57	40
62	71	60	55	58	40
63	72	61	56	59	40
64	73	62	57	60	40
65	74	63	58	61	40
66	75	64	59	62	40
67	76	65	60	63	40
68	77	66	61	64	40
69	78	67	62	65	40
70	79	68	63	66	40
71	80	69	64	67	40
72	81	70	65	68	40
73	82	71	66	69	40
74	83	72	67	70	40
75	84	73	68	71	40
76	85	74	69	72	40
77	86	75	70	73	40
78	87	76	71	74	40
79	88	77	72	75	40
80	89	78	73	76	40
81	90	79	74	77	40
82	91	80	75	78	40
83	92	81	76	79	40
84	93	82	77	80	40
85	94	83	78	81	40
86	95	84	79	82	40
87	96	85	80	83	40
88	97	86	81	84	40
89	98	87	82	85	40
90	99	88	83	86	40
91	100	89	84	87	40
92	101	90	85	88	40
93	102	91	86	89	40
94	103	92	87	90	40
95	104	93	88	91	40
96	105	94	89	92	40
97	106	95	90	93	40
98	107	96	91	94	40
99	108	97	92	95	40
100	109	98	93	96	40



• S. ULAM WIRÓŁ UDOWODNIŁ MATEMATYCZNIE, 1930 R., WYR. INNEGO PROJEKTU
• S. ULAM TAKŻE JEST AUTOREM FORMUŁY, JAKIEJ ANGELIUSZ SZMIDT
JAKI HODZI SMIEĆ CZAŚNIE, DRUGIEJ WERSJI MATEMATYCZNO-
FIZYCZNEJ, WYKONANEJ W 1946 R. W WARSZAWIE, W KIERUNKU
PRZYGOTOWANIA DO WYKONANIA WYKŁADÓW MATEMATYCZNO-
FIZYCZNYCH W KOLEGIUM NAUKOWYM W WARSZAWIE.
OPRAWIONE MATEMATYCZNE MAGDALENA RUTKOWSKA, KOMISJA WERTEROWA:
MAREK EKSTROMSKI, ROMAŁEWICZ, M. M. GŁAŻKO



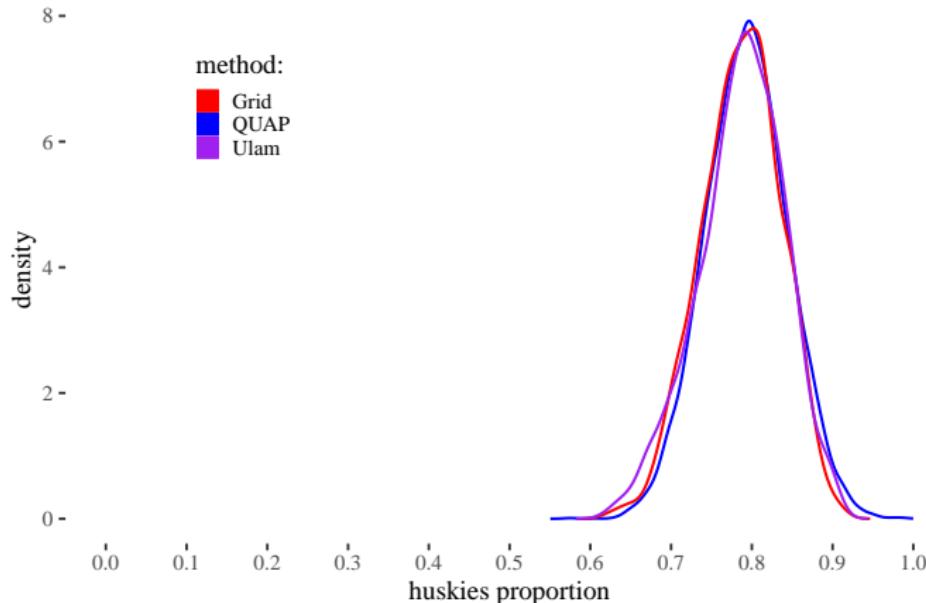
INSTYTUT
PAMIĘCI
NARODOWEJ



University
of Gdańsk

3 engines comparison

1k samples from 3 Models



Model	Mean	SD	HPDI_Lower	HPDI_Upper
1 QUAP	0.7934879	0.05103144	0.7143529	0.8777590
2 Grid	0.7861011	0.04944935	0.7037037	0.8598599
3 Ulam	0.7864292	0.05319343	0.6942720	0.8632100

Vampires village

The Setup

- In a population of 100,000 people, 100 of them are vampires
- Of the 100 who are vampires, 95 of them will test positive for vampirism
- Of the 99,900 mortals, 999 of them will test positive for vampirism

A randomly chosen person is tested positive

Vampires village

The Setup

- In a population of 100,000 people, 100 of them are vampires
- Of the 100 who are vampires, 95 of them will test positive for vampirism
- Of the 99,900 mortals, 999 of them will test positive for vampirism

A randomly chosen person is tested positive

So what?

- number of people who test positive: $95 + 999 = 1094$

$$P(\text{vampire}|\text{positive}) = \frac{95}{1094} \approx 0.087$$

Vampires village

$$P(\text{vampire}|\text{positive}) = \frac{P(\text{positive}|\text{vampire})P(\text{vampire})}{P(\text{positive})}$$

where $P(\text{positive})$ is the average probability of a positive test result, that is,

$$P(\text{positive}) = P(\text{positive}|\text{vampire})P(\text{vampire}) + P(\text{positive}|\text{mortal})(1 - P(\text{vampire}))$$

```
Pr_Positive_Vampire <- 0.95
Pr_Positive_Mortal <- 0.01
Pr_Vampire <- 0.001
Pr_Positive <- Pr_Positive_Vampire * Pr_Vampire +
Pr_Positive_Mortal * ( 1 - Pr_Vampire )
( Pr_Vampire_Positive <- Pr_Positive_Vampire*Pr_Vampire / Pr_Positive )
```

[1] 0.08683729



Signal detection and why data can't save us

Simplified structure of the goal of science

- some binary state is hidden
- we observe imperfect hints
- we use Bayes to learn

Simplified assumptions

- sensitivity is .95
- false positive rate is .05
- base rate: most hypotheses are false, with $\text{pr} = .01$

A simplified observation

The posterior is only .16.

End slide

What to read?

- 2 first chapters of 'Statistical Rethinking' by Richard McElreath
- Also available as video-lectures: [here](#)