bayesian inference

Duke UNIVERSITY

Dr. Mine Çetinkaya-Rundel
Duke University

set-up

win: ≥4

6-sided

12-sided

Image sources:

6-sided: http://commons.wikimedia.org/wiki/File:Sixsided_Dice_inJapan.jpg

12-sided: http://commons.wikimedia.org/wiki/File:12-sided_die.jpg

probabilities

What is the probability of rolling ≥4 with a 6-sided die?

What is the probability of rolling ≥4 with a 12-sided die?

probabilities

What is the probability of rolling ≥4 with a 6-sided die?

$S = \{1,2,3,\boxed{4,5,6}\}$

$P(\geq 4) = 3/6 = 1/2 = 0.5$

What is the probability of rolling ≥4 with a 12-sided die?

$S = \{1,2,3,\boxed{4,5,6,7,8,9,10,11,12}\}$

$P(\geq 4) = 9/12 = 3/4 = 0.75$

“good die”

Say you're playing a game where the goal is to roll ≥ 4. If you could get your pick, which die would you prefer to play this game with?

(a) 

6-sided

(b) 

12-sided

“good die”

Say you're playing a game where the goal is to roll ≥ 4. If you could get your pick, which die would you prefer to play this game with?

good die

(a) 

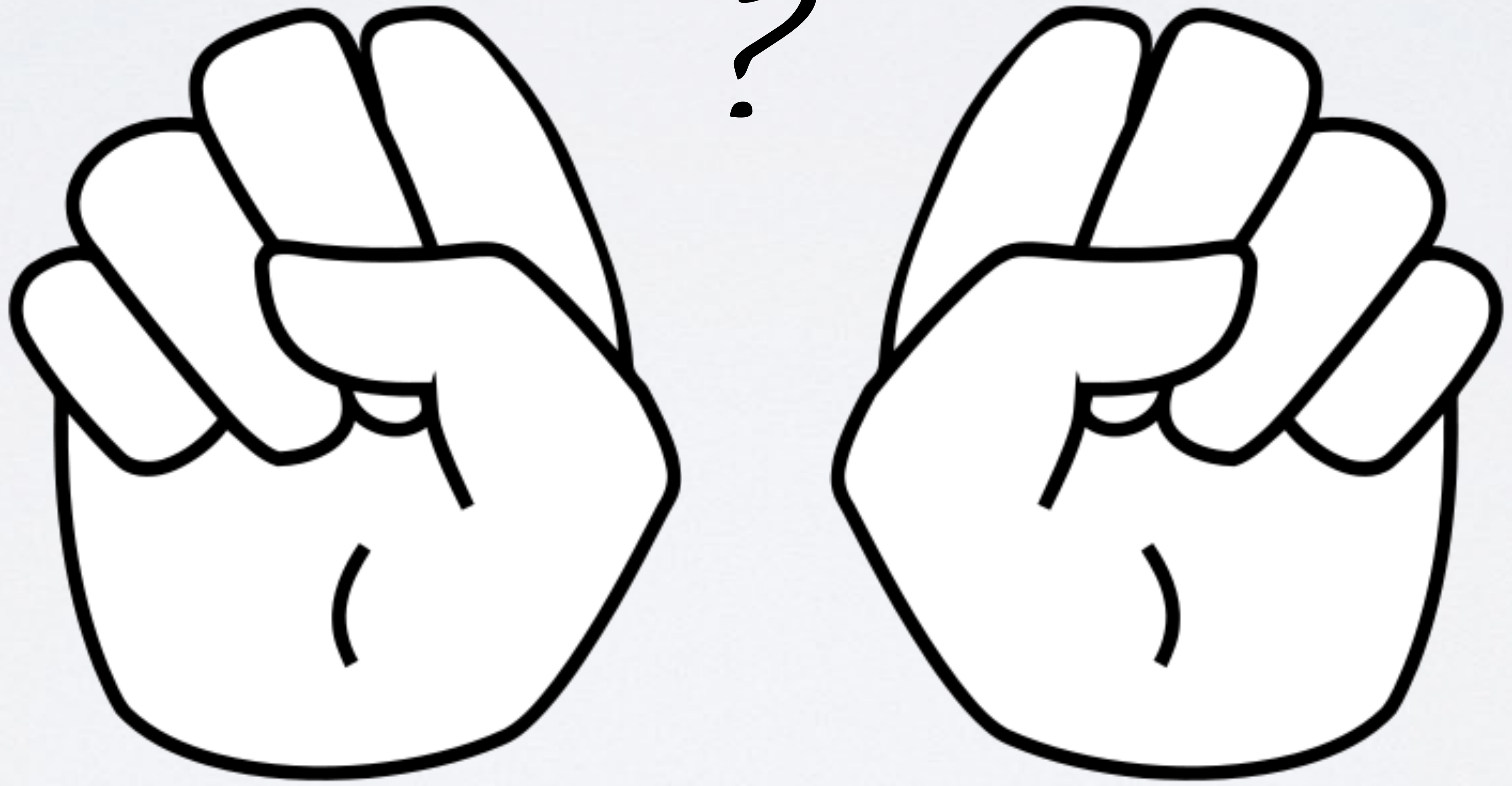
(b) 

$P(\geq 4) = 0.5$

$P(\geq 4) = 0.75$

rules

LEFT ? RIGHT



Hands: http://commons.wikimedia.org/wiki/File:Hand_0_Alternativ.svg

hypotheses and decisions

		Truth	
		Right good, Left bad	Right bad, Left good
Decision	pick Right	**You win the game!**	**You lose :(**
	pick Left	**You lose :(**	**You win the game!**

before you collect data

Before we collect any data, you have no idea if I am holding the good die (12-sided) on the right hand or the left hand. Then, what are the probabilities associated with the following hypotheses?

H_1: good die on the Right (bad die on the Left)

H_2: good die on the Left (bad die on the Right)

	P(H_1: good die on the Right)	P(H_2: good die on the Left)
(a)	0.33	0.67
(b)	0.5	0.5
(c)	0	1
(d)	0.25	0.75

before you collect data

Before we collect any data, you have no idea if I am holding the good die (12-sided) on the right hand or the left hand. Then, what are the probabilities associated with the following hypotheses?

H_1: good die on the Right (bad die on the Left)

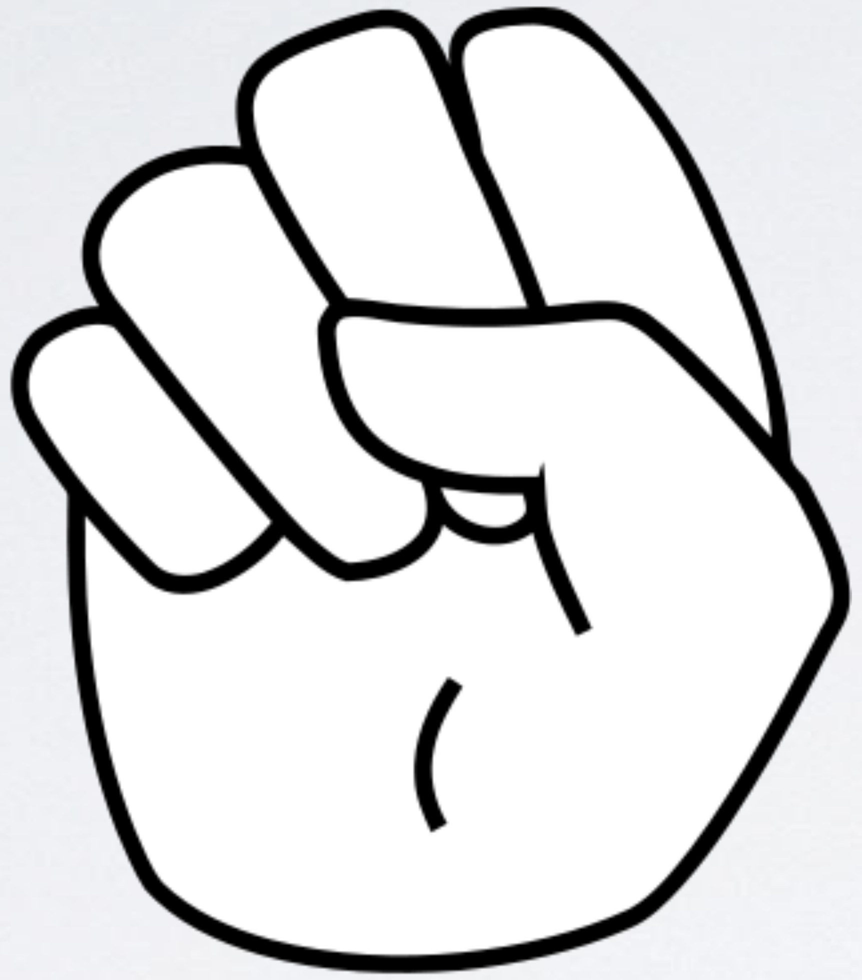
H_2: good die on the Left (bad die on the Right)

	P(H_1: good die on the Right)	P(H_2: good die on the Left)
(a)	0.33	0.67
(b)	0.5	0.5
(c)	0	1
(d)	0.25	0.75

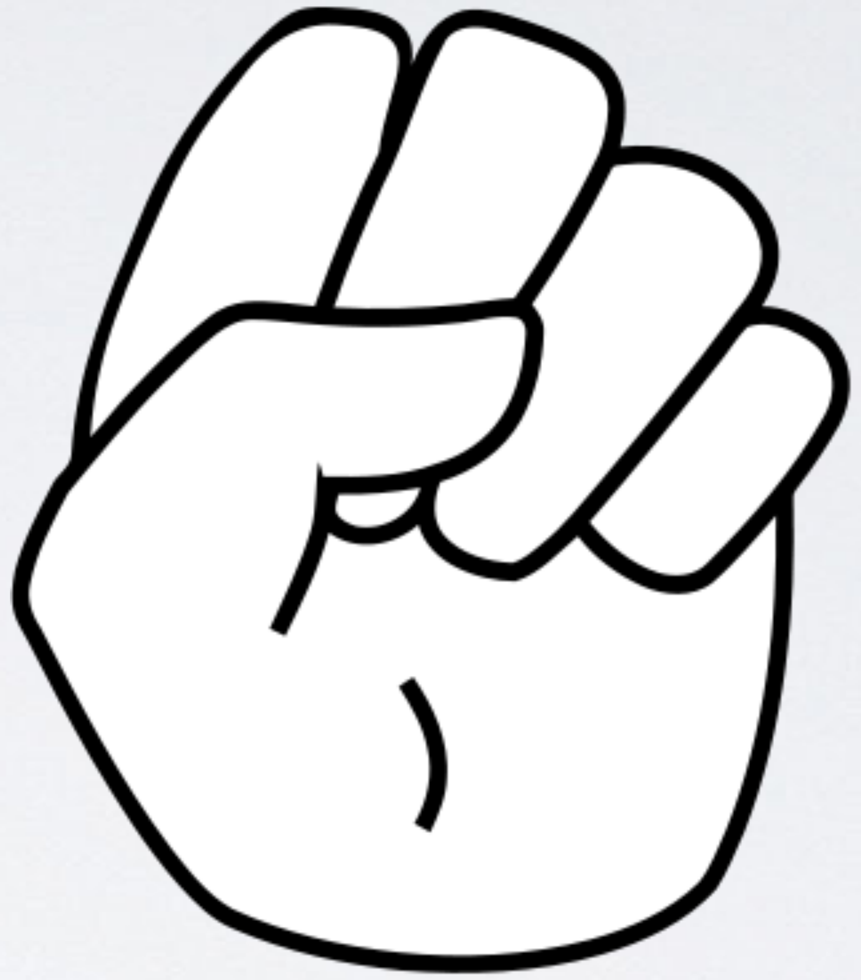
→ prior

data collection: round #1

LEFT



RIGHT



≥4

after you see the data

You chose the right hand, and you won (rolled a number ≥4). Having observed this data point how, if at all, do the probabilities you assign to the same set of hypotheses change?

H_1: good die on the Right (bad die on the Left)

H_2: good die on the Left (bad die on the Right)

	P(H_1: good die on the Right)	P(H_2: good die on the Left)
(a)	0.5	0.5
(b)	more than 0.5	less than 0.5
(c)	less than 0.5	more than 0.5

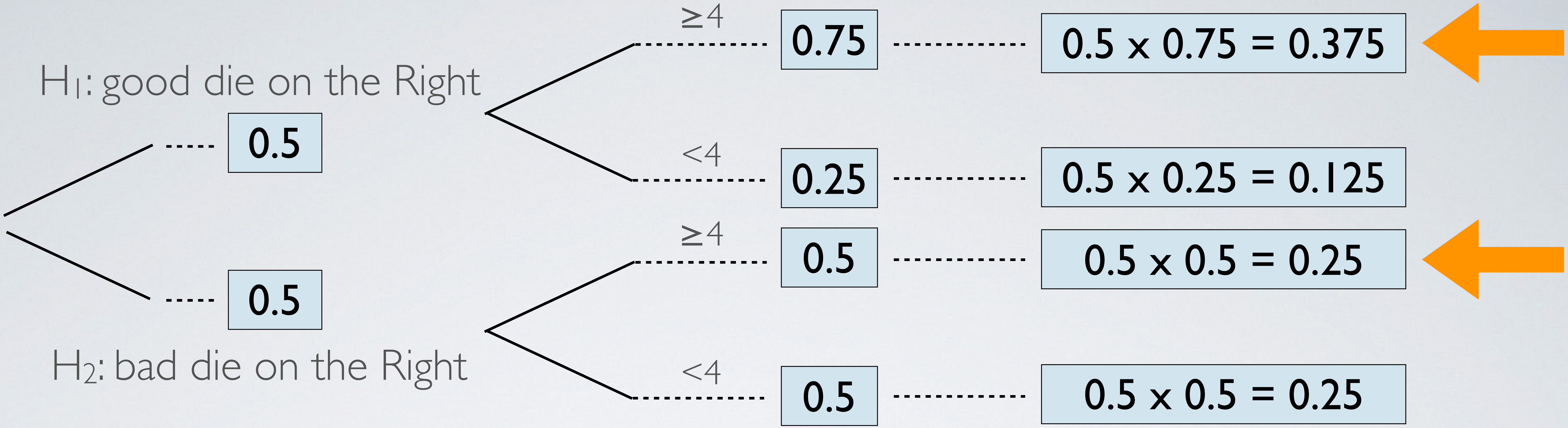
after you see the data

You chose the right hand, and you won (rolled a number ≥4). Having observed this data point how, if at all, do the probabilities you assign to the same set of hypotheses change?

H_1: good die on the Right (bad die on the Left)

H_2: good die on the Left (bad die on the Right)

	P(H_1: good die on the Right)	P(H_2: good die on the Left)
(a)	0.5	0.5
(b)	more than 0.5	less than 0.5
(c)	less than 0.5	more than 0.5



$P(H_1\text{: good die on the Right} \mid \text{you rolled} \geq 4 \text{ with the die on the Right}) =$

$$= \frac{P(\text{good Right} \,\&\, \geq 4 \text{ Right})}{P(\geq 4 \text{ Right})} = \frac{0.375}{0.375 + 0.25} = 0.6$$

posterior

- The probability we just calculated is also called the posterior probability.

  P(H_1: good die on the Right | you rolled $\geq$4 with the die on the Right)
- Posterior probability is generally defined as P(hypothesis | data).
- It tells us the probability of a hypothesis we set forth, given the data we just observed.
- It depends on both the prior probability we set and the observed data.
- This is different than what we calculated at the end of the randomization test on gender discrimination – the probability of observed or more extreme data given the null hypothesis being true, i.e. P(data | hypothesis), also called a p-value.

updating the prior

- In the Bayesian approach, we evaluate claims iteratively as we collect more data.
- In the next iteration (roll) we get to take advantage of what we learned from the data.
- In other words, we update our prior with our posterior probability from the previous iteration.

updated:

$P(H_1$: good die on the Right)	$P(H_2$: good die on the Left)
0.6	0.4

recap

- Take advantage of prior information, like a previously published study or a physical model.
- Naturally integrate data as you collect it, and update your priors.
- Avoid the counter-intuitive definition of a p-value:

  P(observed or more extreme outcome | H0 is true)
- Instead base decisions on the posterior probability:

  P(hypothesis is true | observed data)
- A good prior helps, a bad prior hurts, but the prior matters less the more data you have.
- More advanced Bayesian techniques offer flexibility not present in Frequentist models.