conditional probabilities & bayes' rule

Duke University

Dr. Mine Çetinkaya-Rundel

2015 Gallup poll on use of online dating sites:

		Age				
		18-29	30-49 B	50-64	65+	Total
Used online dating site	Yes A	60	86	58	21	225
	No	255	426	450	382	1513
	Total	315	512	508	403	1738

Bayes' rule:

$$P(A \mid B) = \frac{P(A \mathbin{\&} B)}{P(B)}$$

% of 30-49 year olds using online dating sites =

$\underbrace{86}_{A\&B} / \underbrace{512}_{B} \approx 0.17$

P(use online dating site | 30-49 year old)

Source: http://www.pewinternet.org/2016/02/11/15-percent-of-american-adults-have-used-online-dating-sites-or-mobile-dating-apps/

Bayes' rule

Thomas Bayes
(1702 –1761)

Source: http://bayesian.org/bayes

2015 Gallup poll on use of online dating sites:

		Age				
		18-29	30-49	50-64	65+	Total
Used online dating site	Yes	60	86	58	21	225
	No	255	426	450	382	1513
	Total	315	512	508	403	1738

$$P(\text{use online dating site} \mid \text{30-49 year old}) =$$

$$= \frac{P(\text{use online dating site \& 30-49 year old})}{P(\text{30-49 year old})}$$

$$= \frac{86 / 1738}{512 / 1738} = \frac{86}{512} \approx 0.17$$

Source: http://www.pewinternet.org/2016/02/11/15-percent-of-american-adults-have-used-online-dating-sites-or-mobile-dating-apps/

bayes' rule & diagnostic testing

Dr. Mine Çetinkaya-Rundel

early HIV testing in the military

- first screen with ELISA
- if positive, two more rounds of ELISA
- if either positive, two Western blot assays
- only if both positive, determine HIV infection

data

ELISA

- sensitivity (true positive): 93%
- specificity (true negative): 99%

$P(+ \mid HIV) = 0.93$

$P(- \mid no\ HIV) = 0.99$

Western blot

- sensitivity: 99.9%
- specificity: 99.1%

prevalence: 1.48 / 1000

$P(HIV) = 0.00148$

$$P(has\ HIV \mid ELISA\ +) = ?$$

Sources:

- Petricciani (1985). Licensed tests for antibody to human T-lymphotropic virus type III: sensitivity and specificity. Annals of internal medicine, 103(5), 726-729.
- Burke et. al. (1987). Diagnosis of human immunodeficiency virus infection by immunoassay using a molecularly cloned and expressed virus envelope polypeptide: comparison to Western blot on 2707 consecutive serum samples. Annals of internal medicine, 106(5), 671-676.
- Burke et. al. (1987). Human immunodeficiency virus infections among civilian applicants for United States military service, October 1985 to March 1986. New England Journal of Medicine, 317(3), 131-136.

prior probability

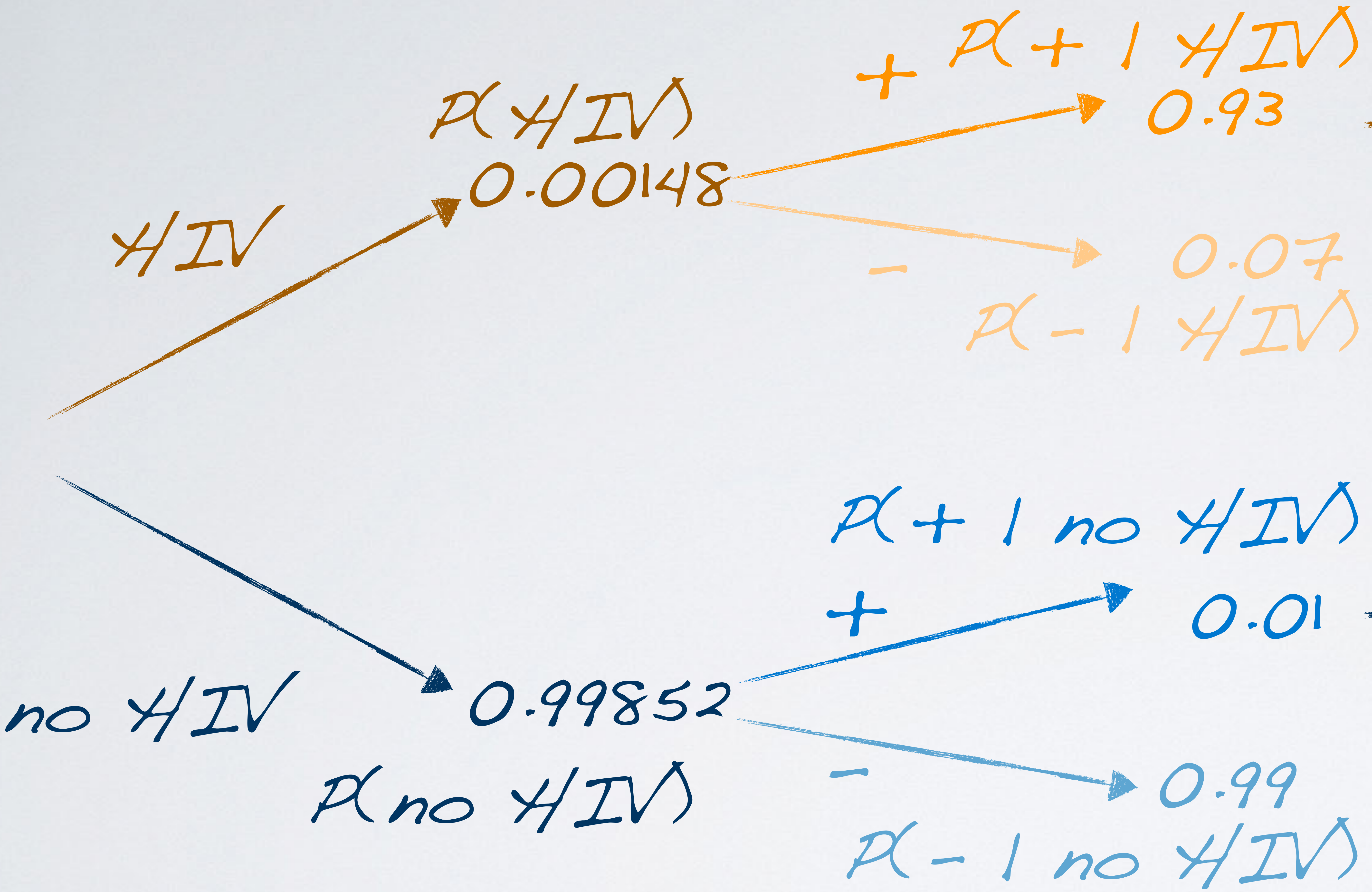
Prior to any testing, what probability should be assigned to a recruit having HIV?

$P(HIV) = 0.00148$

posterior probability

When a recruit goes through HIV screening there are two competing claims: recruit has HIV and recruit doesn't have HIV. If the ELISA yields a positive result, what is the probability this recruit has HIV?

posterior probability



P(HIV and +)

$0.00148 \times 0.93 = 0.0013764$

$$P(HIV \mid +) = \frac{P(HIV \text{ and } +)}{P(+)}$$

$$= \frac{0.0013764}{0.0013764 + 0.0099852} \approx 0.12$$

$0.99852 \times 0.01 = 0.0099852$

P(no HIV and +)

bayesian updating

Dr. Mine Çetinkaya-Rundel

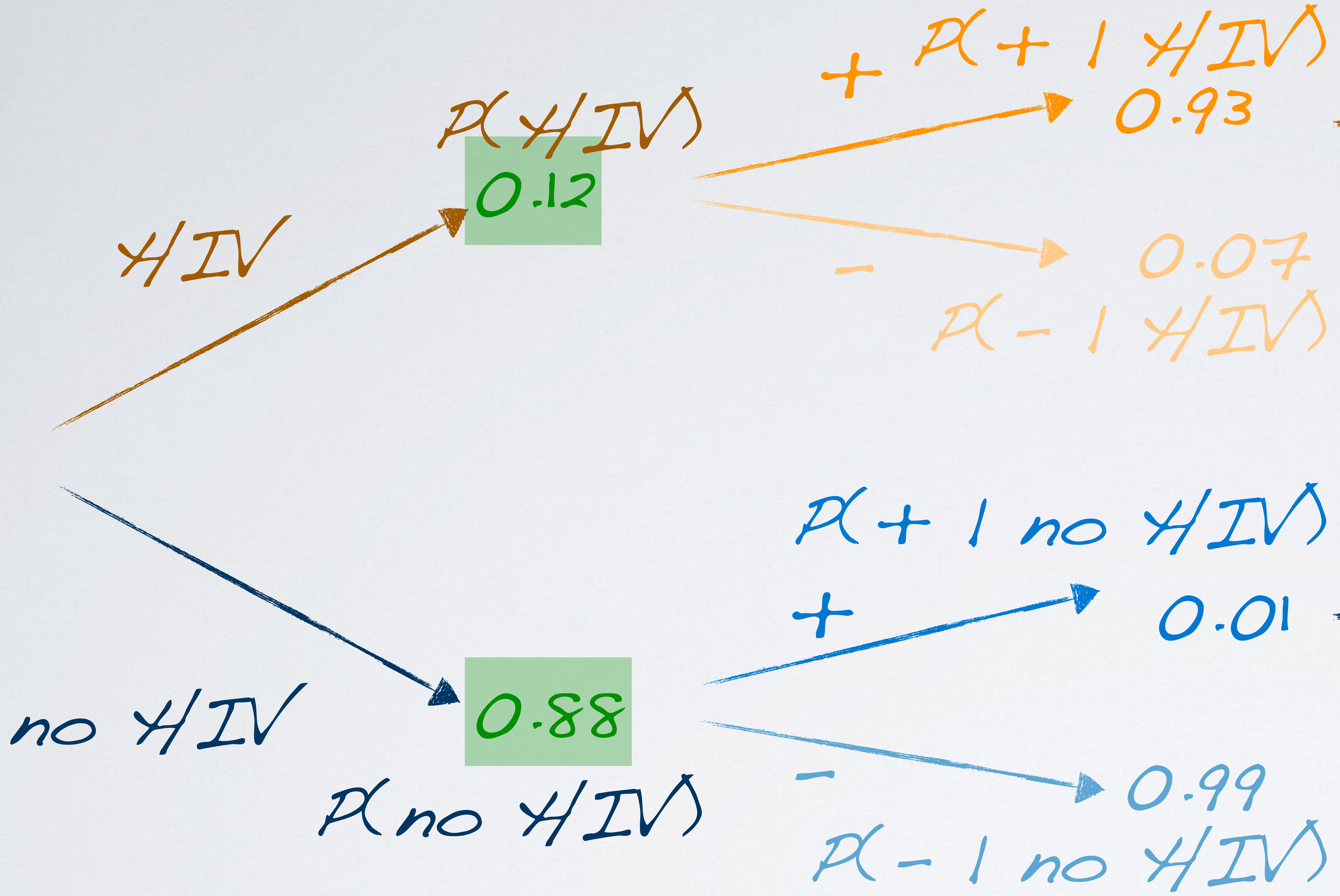
early HIV testing in the military

- first screen with ELISA
- if positive, two more rounds of ELISA
- if either positive, two Western blot assays
- only if both positive, determine HIV infection

retesting

Since a positive outcome on the ELISA doesn't necessarily mean that the recruit actually has HIV, they are retested. What is the probability of having HIV if this second ELISA also yields a positive result?

retesting



P(HIV and +)

$0.12 \times 0.93 = 0.1116$

$$P(HIV \mid +) = \frac{P(HIV \text{ and } +)}{P(+)}$$

$$= \frac{0.1116}{0.1116 + 0.0088} \approx 0.93$$

$0.88 \times 0.01 = 0.0088$

P(no HIV and +)

summary

- individual vs. group diagnostics
- updating only the prior vs. also updating sensitivity and specificity
- Bayesian updating

bayesian & frequentist definitions of probability

Duke UNIVERSITY

Dr. Mine Çetinkaya-Rundel

- the probability of flipping a coin and getting heads is ½
- the probability of rolling snake eyes, that is, two 1s on two dice, is 1/36
- the probability of Apple's stock price going up today is 0.75

frequentist definition of probability

$$P(E) = \lim_{n \to \infty} \frac{n_E}{n}$$

Bayesian definition of probability

- indifferent between winning
  - \$1 if event E occurs, or
  - winning \$1 if you draw a blue chip from a box with 1,000 × p blue chips + 1,000 × (1-p) white chips
- equating the probability of event E, P(E), to the probability of drawing a blue chip from this box, p

$$P(E) = p$$

confidence intervals

Example: Based on a 2015 Pew Research poll on 1,500 adults: "We are 95% confident that 60% to 64% of Americans think the federal government does not do enough for middle class people."

- 95% of random samples of 1,500 adults will produce confidence intervals that contain the true proportion of Americans who think the federal government does not do enough for middle class people
- common misconceptions:
  - there is a 95% chance that this confidence interval includes the true population proportion
  - the true population proportion is in this interval 95% of the time

Source: http://www.pewsocialtrends.org/2016/02/04/most-americans-say-government-doesnt-do-enough-to-help-middle-class/

credible intervals

- allow us to describe the unknown true parameter not as a fixed value but with a probability distribution
- this will let us construct something like a confidence interval, except we can make probabilistic statements about the parameter falling within that range
  - **Example**: "*The posterior distribution yields a 95% credible interval of 60% to 64% for the proportion of Americans who think the federal government does not do enough for middle class people.*"
- these are called credible intervals

Source: http://www.pewsocialtrends.org/2016/02/04/most-americans-say-government-doesnt-do-enough-to-help-middle-class/

inference for a proportion - frequentist approach

Duke UNIVERSITY

Dr. Mine Çetinkaya-Rundel

morning after

- **research question:** Is RU-486 an effective "morning after" contraceptive?
- **participants:** 40 women who came to a health clinic asking for emergency contraception
- **design:** Random assignment to RU-486 or standard therapy (20 in each group)
- **data:**
  - 4 out of 20 in RU-486 (treatment) became pregnant
  - 16 out of 20 in standard therapy (control) pregnant
- **question:** How strongly do these data indicate that the treatment is more effective than the control?

framework

- simplification: one proportion
  - consider the 20 total pregnancies
  - question: How likely is it that 4 pregnancies occur in the treatment group?
- if treatment and control are equally effective + sample sizes for the two groups are the same

$$P(\text{pregnancy comes from treatment group}) = p = 0.5$$

hypotheses

$p =$ probability that a given pregnancy comes from the treatment group

$H_0 : p = 0.5$ - no difference, a pregnancy is equally likely to come from the treatment or control group

$H_A : p < 0.5$ - treatment is more effective, a pregnancy is less likely to come from the treatment group

p-value

- $k = 4$ and $n = 20$ - since there are 20 pregnancies total, and 4 occur in the treatment group
- $p = 0.5$ - assuming H_0 is true
- p−value = $P(k \leq 4)$

```r
sum(dbinom(0:4, size = 20, p = 0.5))
```

```
## [1] 0.005908966
```

inference for a proportion - Bayesian approach

Duke UNIVERSITY

Dr. Mine Çetinkaya-Rundel

framework

- consider the 20 total pregnancies
  - **question**: how likely is it that 4 pregnancies occur in the treatment group?
- if treatment and control are equally effective + sample sizes for the two groups are the same

P(pregnancy comes from treatment group) $= p = 0.5$

hypotheses, i.e. models

- delineate plausible models:
  - assume p could be

10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90%

- consider 9 models, instead of 1 as in the frequentist paradigm
  - $p = 20\%$: given a pregnancy occurs, there is a 2:8 or 1:4 chance that it will occur in the treatment group

specifying the prior

- prior probabilities reflect state of belief prior to the current experiment
- incorporate information learned from all relevant research up to the current point in time, but not incorporate information from the current experiment
- suppose my prior probability for each of the 9 models is as presented below:

Model (p)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Total
Prior	0.06	0.06	0.06	0.06	0.52	0.06	0.06	0.06	0.06	1

  - benefit of treatment is symmetric — equally likely to be better or worse than the standard treatment
  - 52% chance that there is no difference between the treatments

likelihood

- calculate *P(data | model)* for each model considered.
- this probability is called the **likelihood**:

$$P(data \mid model) = P(k = 4 \mid n = 20, p)$$

calculating the likelihood

```r
p <- seq(from = 0.1, to = 0.9, by = 0.1)
prior <- c(rep(0.06, 4), 0.52, rep(0.06, 4))
likelihood <- dbinom(4, size = 20, prob = p)
```

Model (p)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Total
Prior, $P(model)$	0.06	0.06	0.06	0.06	0.52	0.06	0.06	0.06	0.06	1
Likelihood, $P(data \mid model)$	0.0898	0.2182	0.1304	0.035	0.0046	0.0003	0	0	0	

posterior

use Bayes' rule to calculate the posterior probability, i.e. *P(model | data)*

$$P(model \mid data) = \frac{P(model \ \& \ data)}{P(data)}$$

$$= \frac{P(data \mid model) \times P(model)}{P(data)}$$

calculating the posterior

```
numerator <- prior * likelihood
denominator <- sum(numerator)
posterior <- numerator / denominator
sum(posterior)
```

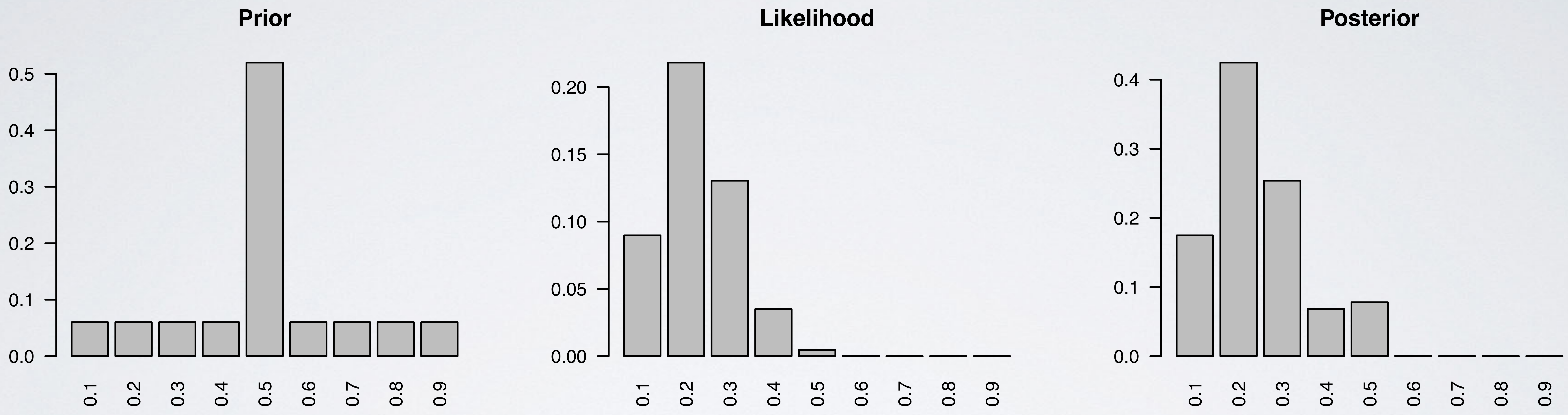
```
## [1] 1
```

Model (p)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Total
Prior, $P(model)$	0.06	0.06	0.06	0.06	0.52	0.06	0.06	0.06	0.06	1
Likelihood, $P(data \mid model)$	0.0898	0.2182	0.1304	0.035	0.0046	0.0003	0	0	0	
$P(data \vert model) \times P(model)$	0.0054	0.0131	0.0078	0.0021	0.0024	0	0	0	0	0.0308
Posterior, $P(model \vert data)$	0.1748	0.4248	0.2539	0.0681	0.0780	0.0005	0	0	0	1

decision making

- posterior probability that $p = 0.2$ is 42.48%
  - this model has the highest posterior probability
- calculation of the posterior incorporated prior information and likelihood of data observed
  - data "at least as extreme as observed" plays no part in the Bayesian paradigm
- note that probability that $p = 0.5$ dropped from 52% in the prior to about 7.8% in the posterior
  - this demonstrates how we update our beliefs based on observed data

prior, likelihood, and posterior, visualized



synthesis

- Bayesian paradigm allows us to make direct probability statements about our models
- we can also calculate the probability that RU-486 (the treatment) is more effective than the control
  - this is the sum of the posteriors of the models where $p < 0.5$

Model (p)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Total
Posterior, $P(model \mid data)$	0.1748	0.4248	0.2539	0.0681	0.0780	0.0005	0	0	0	1

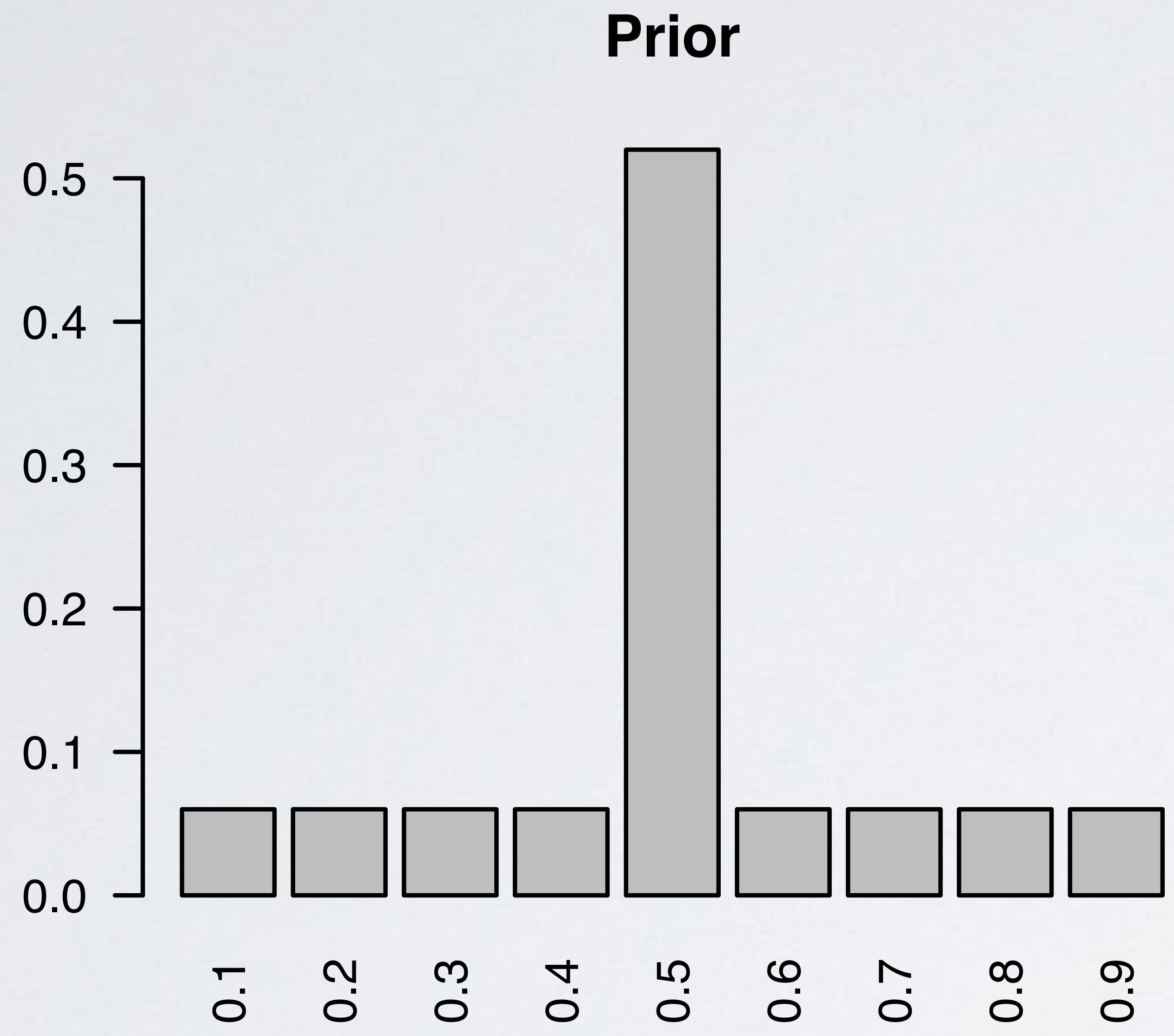
0.9216

effect of sample size on the posterior

Dr. Mine Çetinkaya-Rundel

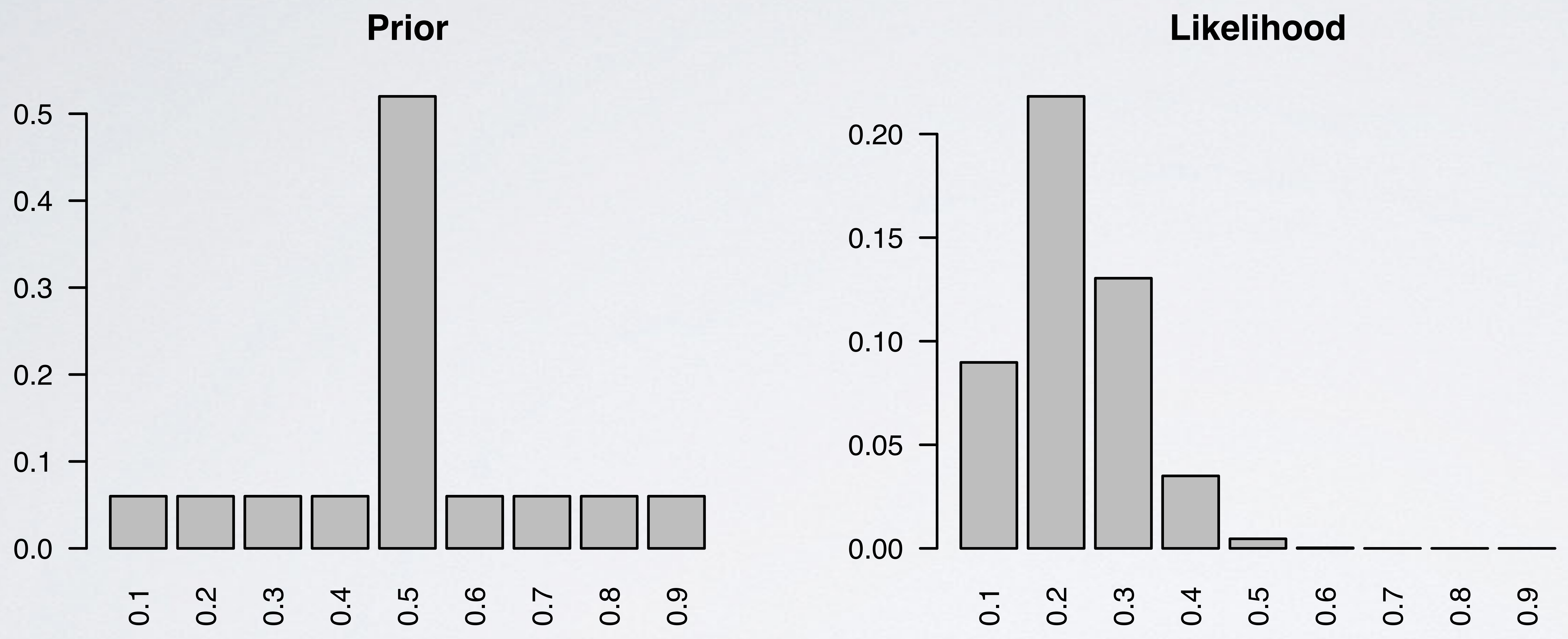
original results

$n = 20, k = 4$



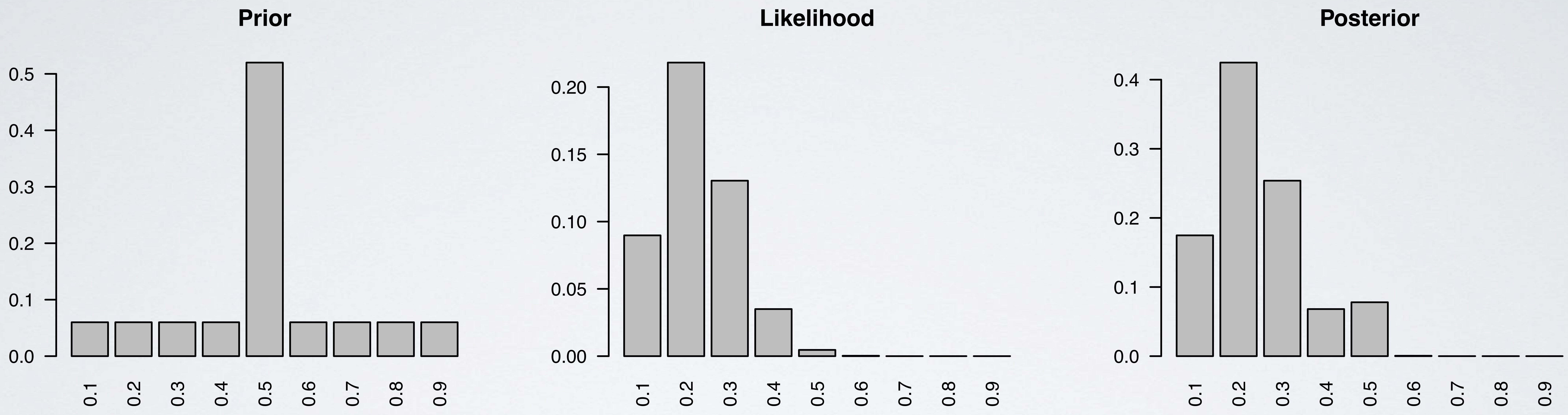
original results

$n = 20, k = 4$



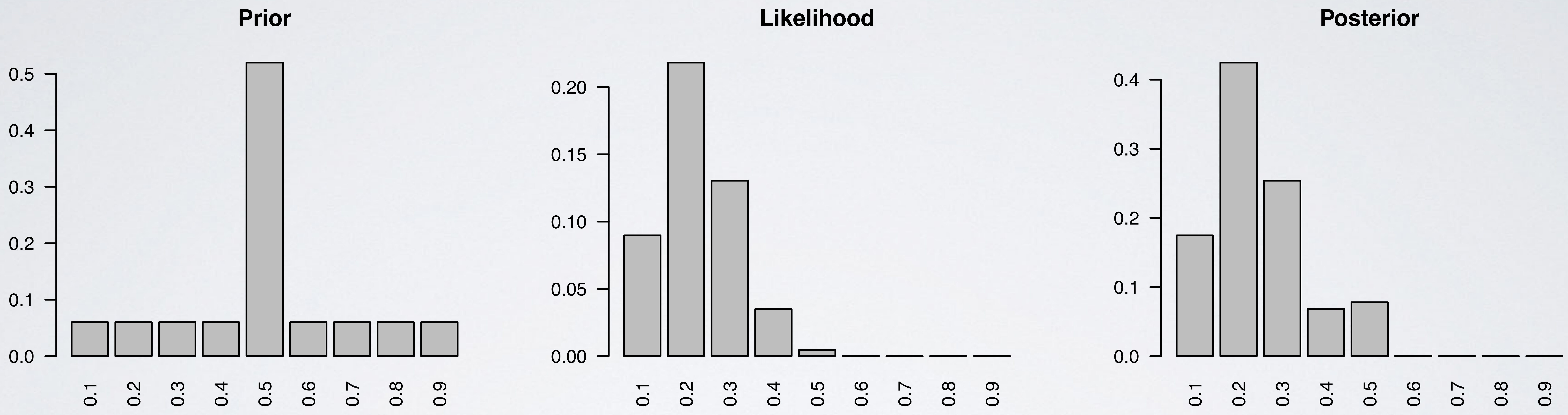
original results

$n = 20,\ k = 4$



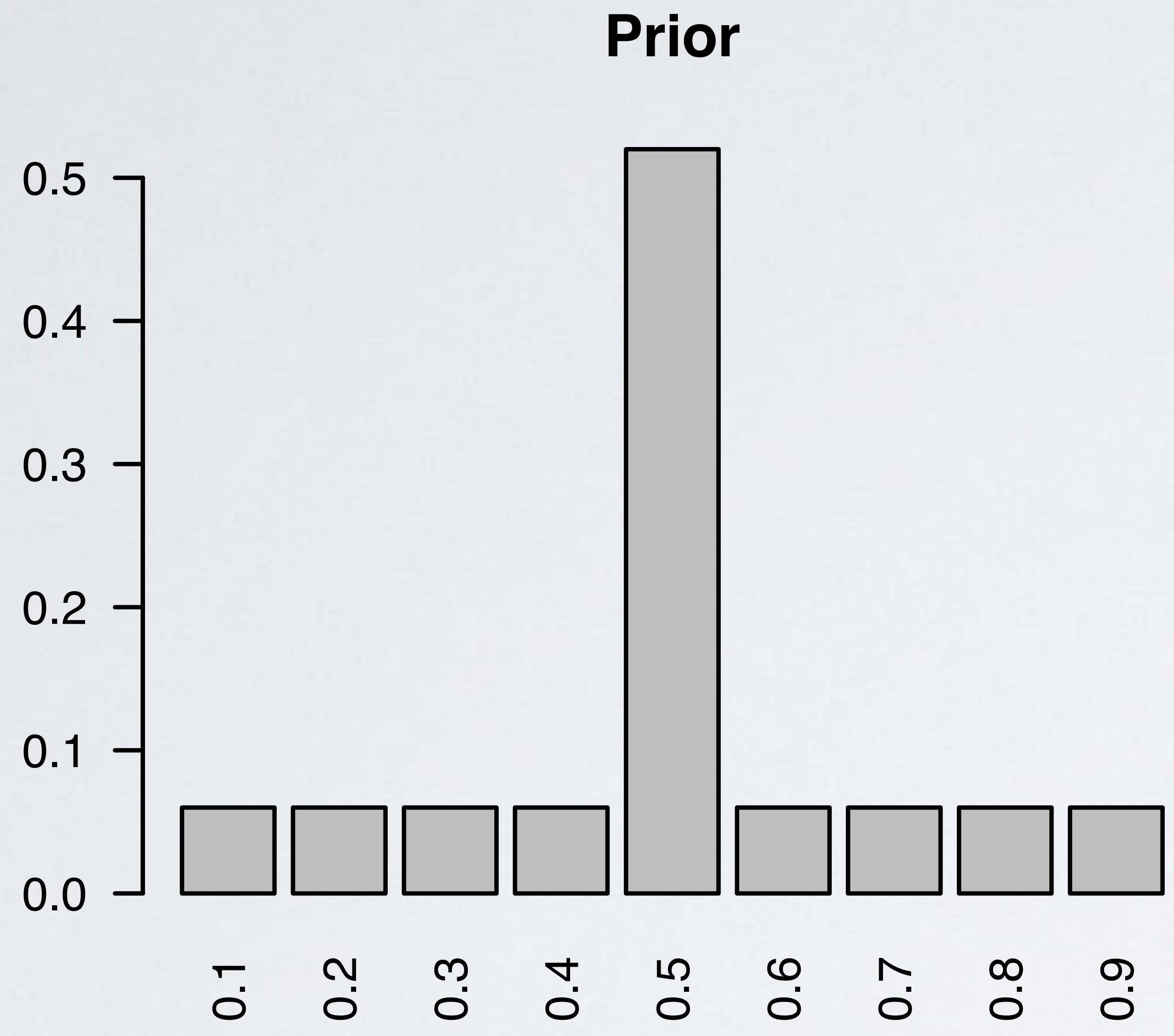
original results

$n = 20, k = 4$



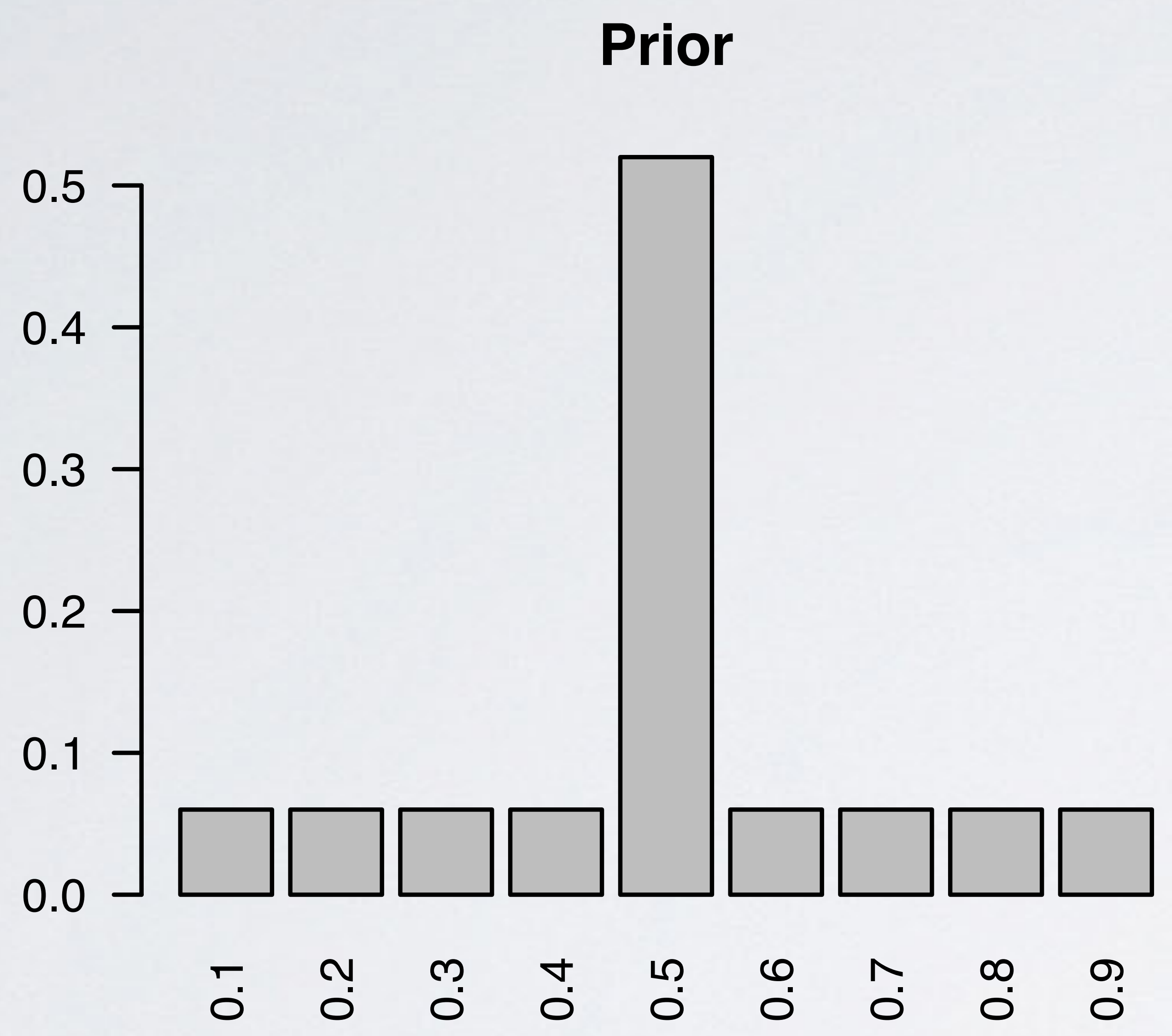
what if we had more data

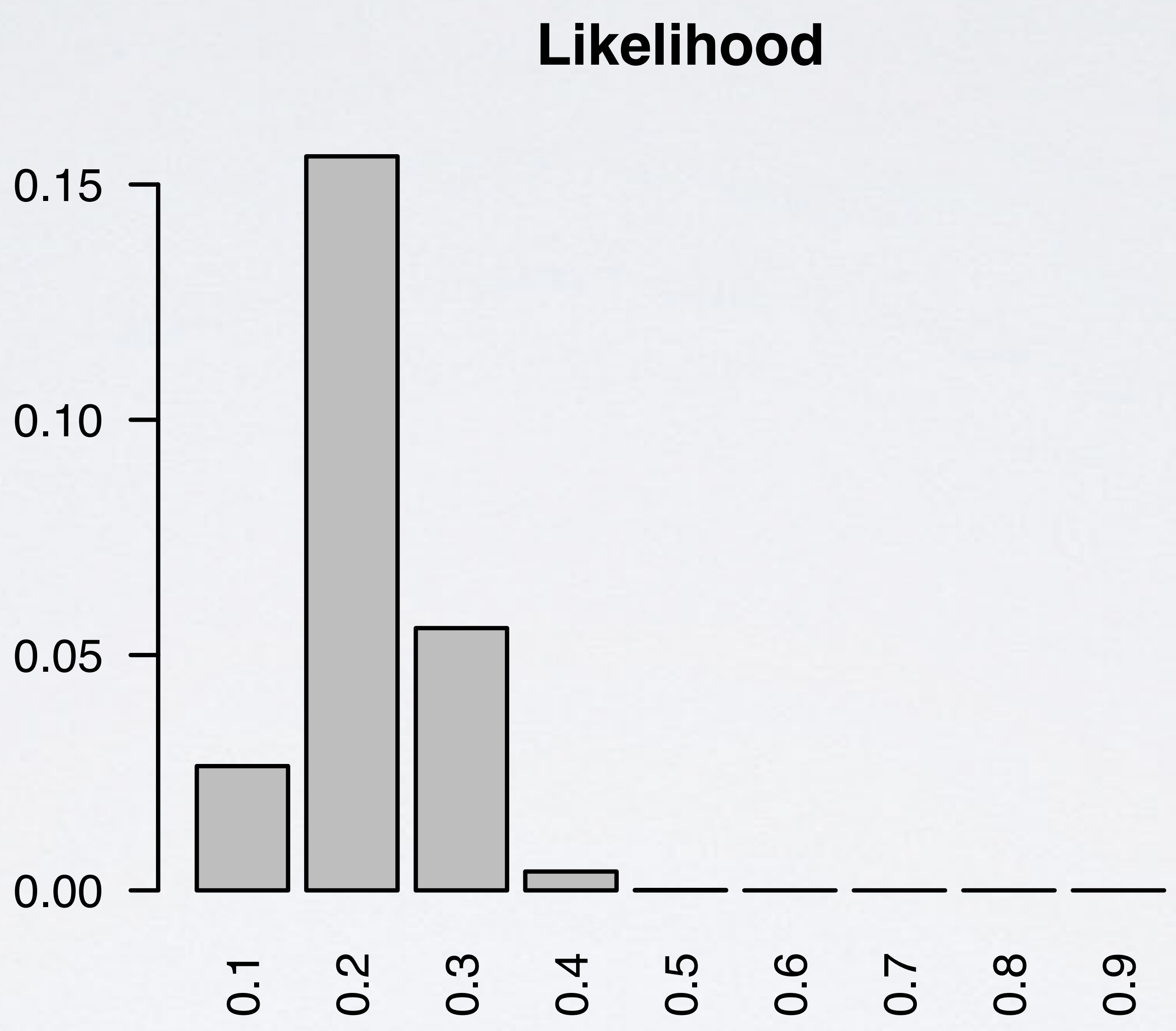
$n = 40, k = 8$



what if we had more data

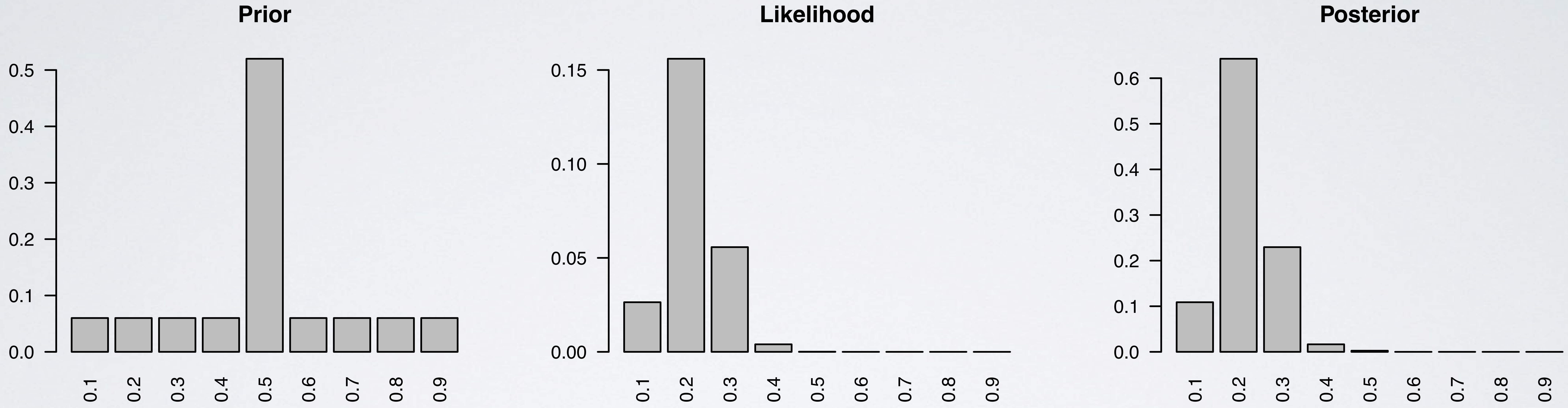
$n = 40, k = 8$





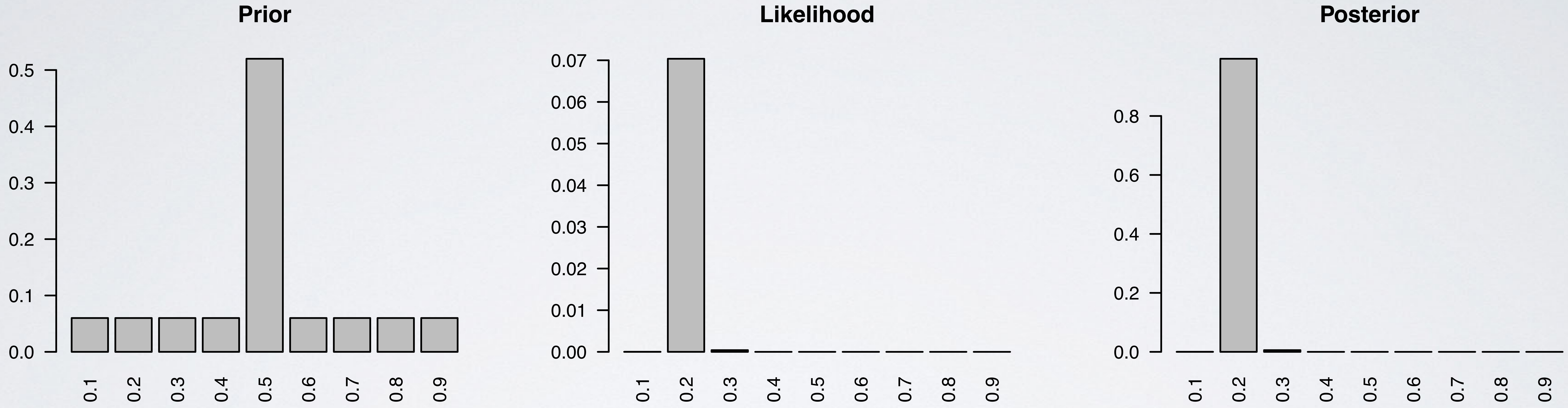
what if we had more data

$n = 40, k = 8$



what if we had *even* more data

$n = 200, k = 40$



frequentist vs. bayesian inference

Duke UNIVERSITY

Dr. Mine Çetinkaya-Rundel

M&Ms

- we have a population of M&Ms
- percentage of yellow M&Ms is either 10% or 20%
- you have been hired as a statistical consultant to decide whether the true percentage of yellow M&Ms is 10%
- you are being asked to make a decision, and there are associated payoff/losses that you should consider

payoffs / losses

	True state of the population	
Decision	% yellow = 10%	%yellow = 20%
% yellow = 10%	Your boss gives you a bonus :)	You lose your job :(
%yellow = 20%	You lose your job :(	Your boss gives you a bonus :)

data

- you can “buy” a random sample from the population
- you pay $200 for each M&M, and you must buy in $1,000 increments (5 M&Ms at a time)
- you have a total of $4,000 to spend (you may buy 5, 10, 15, or 20 M&Ms)

frequentist inference

hypotheses

H_0: 10% yellow M&Ms

H_A: >10% yellow M&Ms

sig. level

$\alpha = 0.05$

sample

RG<u>Y</u>BO

obs. data

$k = 1, n = 5$

p-value

$$P(K \geq 1 \mid n = 5, p = 0.10)$$
$$= 1 - P(k = 0 \mid n = 5, p = 0.10)$$
$$= 1 - 0.90^5 \approx 0.41$$

→ Fail to reject H_0

bayesian inference

hypotheses

H_1: 10% yellow M&Ms

H_2: 20% yellow M&Ms

prior

$$P(H_1) = 0.5$$

$$P(H_2) = 0.5$$

sample

RGYBO

obs. data

$k = 1, n = 5$

likelihood

$$P(k = 1 \mid H_1) = \binom{5}{1} 0.10 \times 0.90^4 \approx 0.33$$

$$P(k = 1 \mid H_2) = \binom{5}{1} 0.20 \times 0.80^4 \approx 0.41$$

posterior

$$P(H_1 \mid k = 1) = \frac{P(H_1) \times P(k = 1 \mid H_1)}{P(k = 1)}$$

$$= \frac{0.5 \times 0.33}{0.5 \times 0.33 + 0.5 \times 0.41}$$

$$\approx 0.45$$

$$P(H_2 \mid k = 1)$$

$$= 1 - 0.45 = 0.55$$

bayesian vs. frequentist inference

	FREQUENTIST	BAYESIAN	
obs. data	P(k or more | 10% yellow)	P(10% yellow | n,k)	P(20% yellow | n,k)
$n = 5, k = 1$	0.41	0.45	0.55
$n = 10, k = 2$	0.26	0.39	0.61
$n = 15, k = 3$	0.18	0.34	0.66
$n = 20, k = 4$	0.13	0.29	0.71