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

H1: 10% yellow M&Ms

H2: 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

$$\begin{aligned} P(H_1 \mid k=1) &= \frac{P(H_1) \times P(k=1 \mid H_1)}{P(k=1)} \\ &= \frac{0.5 \times 0.33}{0.5 \times 0.33 + 0.5 \times 0.41} \\ &\approx 0.45 \end{aligned}$$

$$\begin{aligned} &P(H_2 \mid k=1) \\ &= 1 - 0.45 = 0.55 \end{aligned}$$

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