Brief Introduction to Bayesian Inference

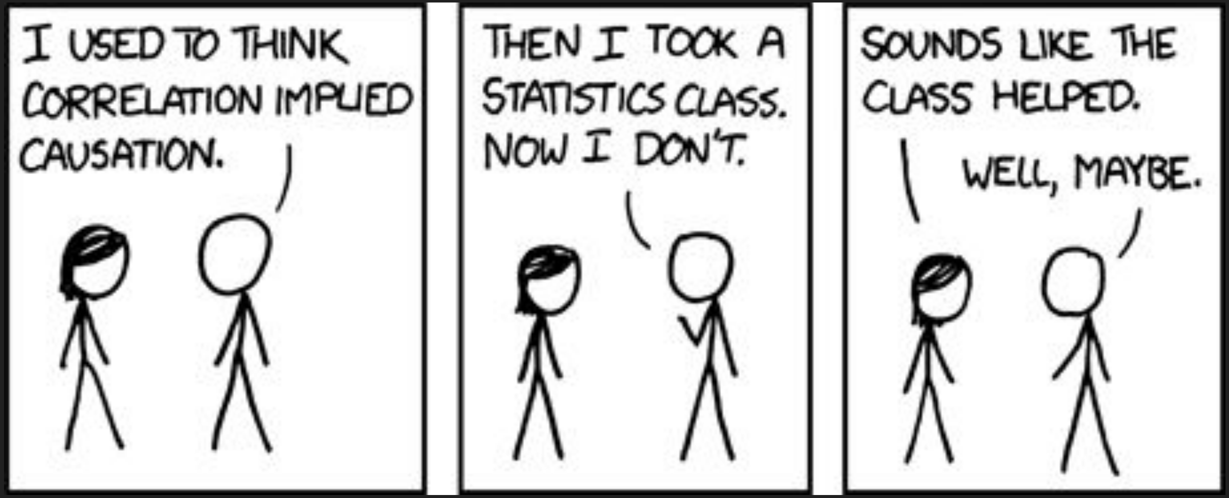
Byron J. Smith

BMI 206

2020-11-19

Slides: https://bit.ly/36GJRrE

Code: https://bit.ly/36M3YVl



Byron J. Smith

BMI 206

2020-11-19

With help from XKCD

Slides: https://bit.ly/36GJRrE

Code: https://bit.ly/36M3YVl

What is science?

- Learn discrete facts about the world.
  - “COVID-19 is caused by SARS-CoV-2.”
  - Karl Popper
  - “Strong Inference”: Rule out all other plausibilities
- Measure (continuous) parameters
  - "Wearing a mask decreases your risk of becoming infected with SARS-CoV-2 by 55.2%."
- ...But what about questions somewhere in between these two extremes?
  - "Does wearing a mask affect your risk of becoming infected?"

"Traditional": Null hypothesis statistical testing

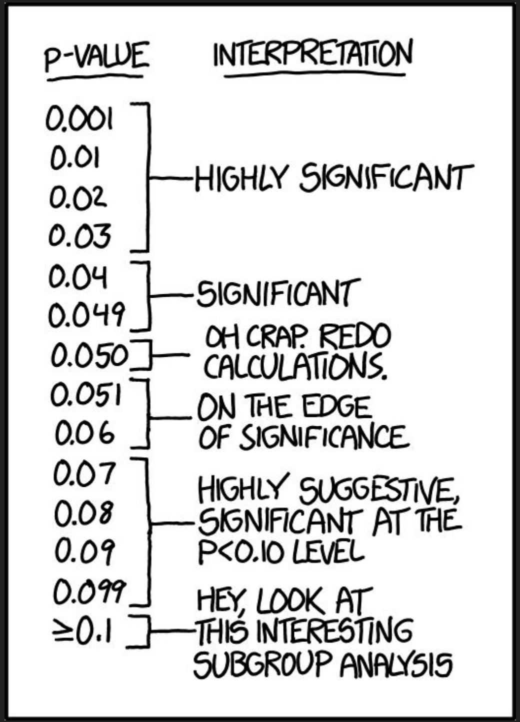
"Does wearing a mask affect your risk of becoming infected?"

- Run an (e.g.) t-test
    - Implies a null hypothesis: $H_0 : \mu_{\text{mask}} = \mu_{\text{not}}$
- Calculate p-value
- If $p < 0.05$: we "reject the null hypothesis"
        - Conclude that wearing a mask *does* affect your infection risk
- Else $p > 0.05$: we cannot reject

What does the p-value represent?

What does the p-value represent?

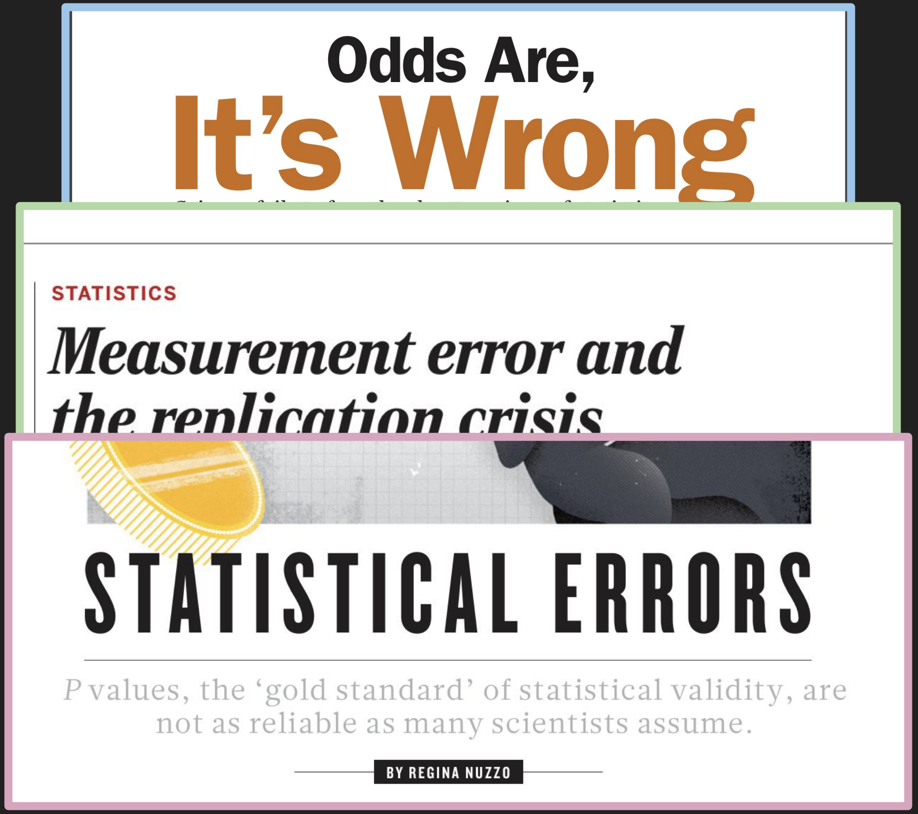
- Probability of seeing a test statistic as extreme as the one we observed in a world where the null hypothesis is true.
  - NOT: The probability of the null hypothesis being true.



“Science by p-value”

<u>Shortcomings:</u>

- Misleading when:
  - Intuition/reality does not match test assumptions (E.g. small sample size)
  - When our null hypothesis is trivially wrong
- Ignores prior information
- Ignores effect size



“Science by p-value”

Shortcomings:

- Misleading when:
  - Intuition/reality does not match test assumptions (E.g. small sample size)
  - When our null hypothesis is trivially wrong
- Ignores prior information
- Ignores effect size

What we would prefer:

"How does wearing a mask affect the risk of becoming infected?"

↓

There is a 95% probability that wearing a mask decreases your risk of becoming infected by 50% or more

Bayes's Rule

$$P(\theta|X) = \frac{P(X|\theta)\,P(\theta)}{P(X)}$$

Bayes's Rule

Parameters

$$P(\theta|X) = \frac{P(X|\theta)\,P(\theta)}{P(X)}$$

Bayes's Rule

Data

$$P(\theta|X) = \frac{P(X|\theta)\, P(\theta)}{P(X)}$$

Bayes’s Rule

“Posterior”

$$\boxed{P(\theta|X)} = \frac{P(X|\theta)\,P(\theta)}{P(X)}$$

Bayes's Rule

"Likelihood"

$$P(\theta|X) = \frac{\boxed{\mathcal{L}(\theta|X)}\,P(\theta)}{P(X)}$$

Bayes's Rule

“Prior”

$$P(\theta|X) = \frac{P(X|\theta)\boxed{P(\theta)}}{P(X)}$$

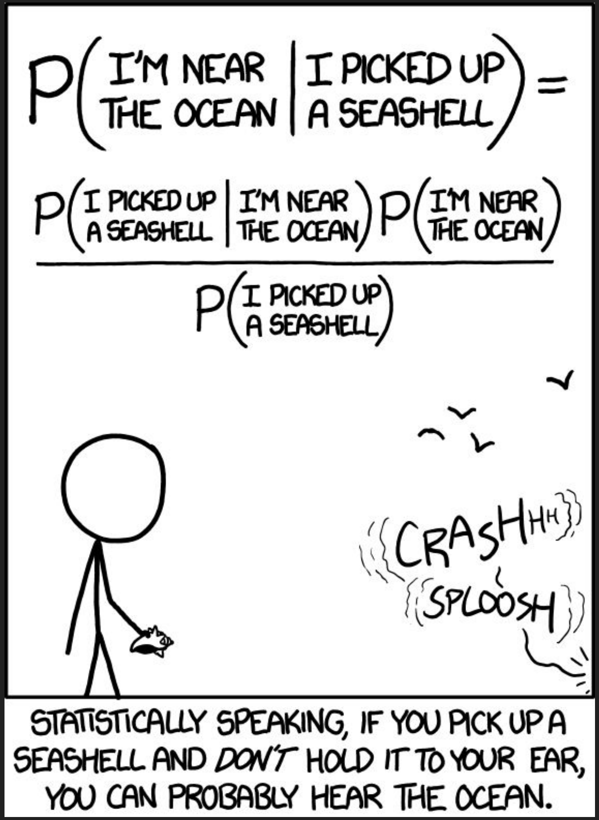
Bayes's Rule

"Normalizing Constant"

$$P(\theta|X) = \frac{P(X|\theta)\,P(\theta)}{\boxed{P(X)}}$$

Bayes's Rule

$$P(\theta|X) = \frac{P(X|\theta)\,P(\theta)}{P(X)}$$



Does wearing a mask affect COVID-19 risk?

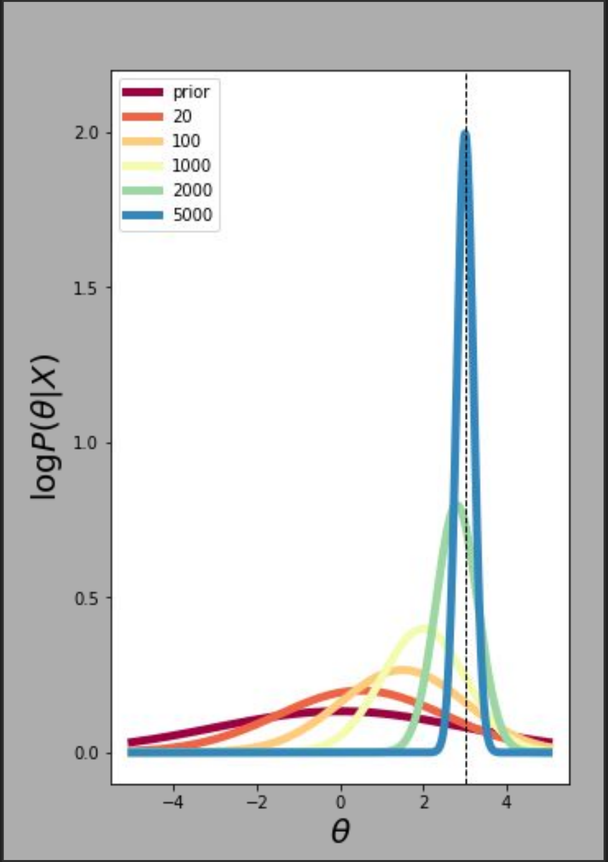
Rephrase: What is the relative risk to wearers vs. non-wearers?

θ	Odds ratio of COVID-19 among wearers and non-wearers
X	Observed individuals and cases among both groups
$P(X\|\theta)$	Probability of the observed numbers, given the odds ratio
$P(\theta)$	Prior probability of the given odds ratio
$P(X)$	Marginal probability of this observation

$$P(\theta|X) = \frac{P(X|\theta)\,P(\theta)}{P(X)}$$

Does wearing a mask affect COVID-19 risk?

- As we collect more data, the posterior distribution converges on the observed odds ratio



How to calculate the posterior probability?

$$P(\theta|X) = \frac{P(X|\theta)\,P(\theta)}{P(X)}$$

How to calculate the posterior probability?

- Hard

$$P(\theta|X) = \frac{P(X|\theta)\, P(\theta)}{\int_\Theta P(X|\theta')\, P(\theta')\, d\theta'}$$

How to calculate the posterior probability?

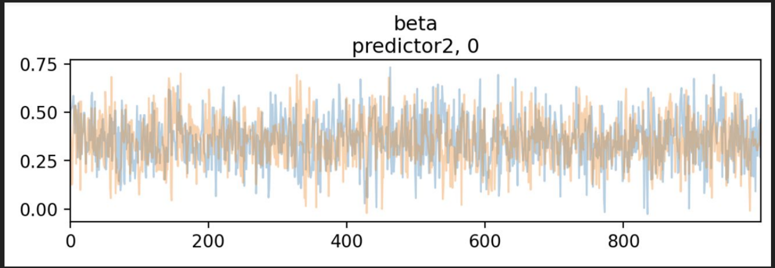
- Hard
- Easier: *Sampling* from the posterior
- Markov-Chain Monte Carlo (Metropolis-Hastings algorithm)

$$P(\theta|X) = \frac{P(X|\theta)\, P(\theta)}{\int_{\Theta} P(X|\theta')\, P(\theta')\, d\theta'} \propto P(X|\theta)\, P(\theta)$$

Sampling θ from the posterior

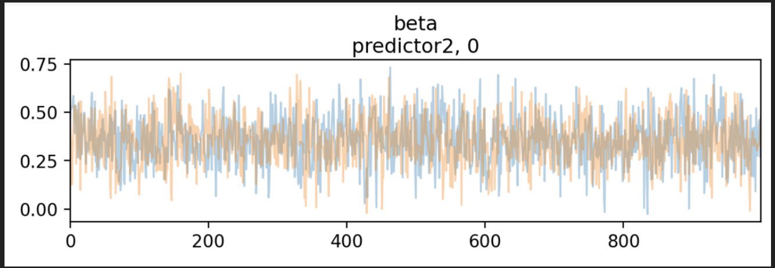
- Markov-Chain Monte Carlo (Metropolis-Hastings algorithm)

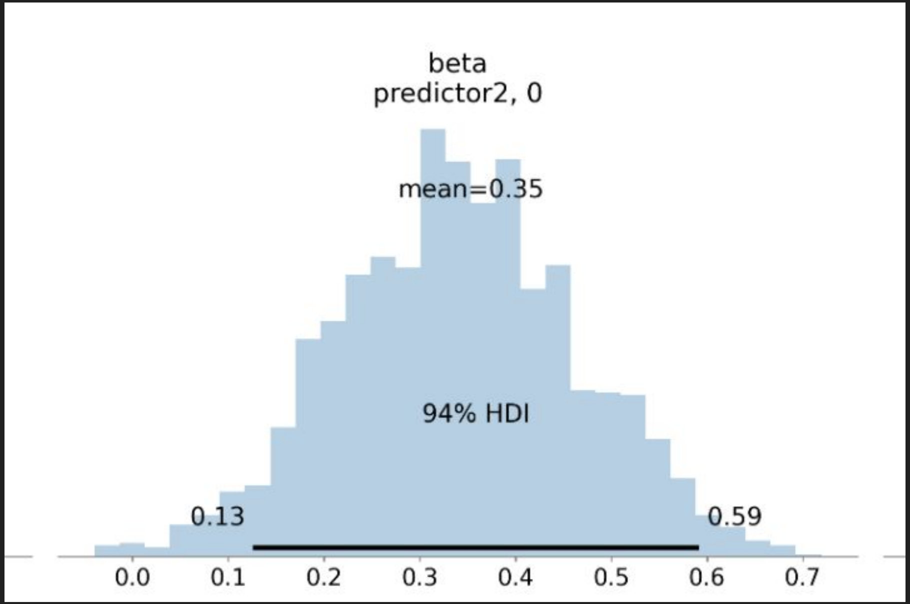
$$\theta_1, \theta_2, \theta_3, \ldots, \theta_n$$



Sampling θ from the posterior

- Markov-Chain Monte Carlo (Metropolis-Hastings algorithm)
- Use samples from the posterior to calculate
  - Expectations
  - Credible intervals





Why doesn’t everyone do it this way?

- Computation
  - Scales unfavorably with number of parameters, size of data
  - May require many samples due to “poor mixing”
- The Prior
  - “Introduces bias”
  - Subject to criticism
- Alternatives
  - Maximum likelihood

Why doesn't everyone do it this way?

- Computation
  - Scales unfavorably with number of parameters, size of data
  - May require many samples due to "poor mixing"
- The Prior
  - "Introduces bias"
  - Subject to criticism
- Alternatives
  - Maximum likelihood

- Computation
  - Huge progress in last 20 years.
  - HMC, Variational Inference
- The Prior
  - Useful for incorporating expert knowledge, constraints, intuition.
  - Makes assumptions explicit
- Alternatives
  - Challenging to assess uncertainty from point estimates

Logistic Regression

$$
\begin{aligned}
y_i &\sim \text{Bernoulli}(p_i) \\
\text{logit}(p_i) &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} \\
\beta_k &\sim \text{Normal}(0, 1)
\end{aligned}
$$

Try it out!

https://bit.ly/36M3YVl

Logistic Regression

$$
\begin{aligned}
y_i &\sim \text{Bernoulli}(p_i) \\
\text{logit}(p_i) &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} \\
\beta_k &\sim \text{Normal}(0, 10)
\end{aligned}
$$

Logistic Regression

$$
\begin{aligned}
y_i &\sim \mathrm{Bernoulli}(p_i) \\
\mathrm{logit}(p_i) &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} \\
\beta_k &\sim \mathrm{Laplace}(1)
\end{aligned}
$$

Democratized statistical modeling

- Emphasizes parameter estimates and uncertainty over p-values
- “Inference Button”: flexible, well built software to sample from and interpret models
  - STAN, PyMC3, Pyro, Turing.jl
- Only scratched the surface in this session

- Every Statistics XKCD (44 and counting): **https://bit.ly/32YH4Jc**