Introduction to Bayesian inference for evolutionists & ecologists

Kent E. Holsinger

Department of Ecology & Evolutionary Biology, U-3043

University of Connecticut

Storrs, CT 06269-3043

Why be Bayesian?

How worried should you be if you get a positive test for COVID-19?

- Abbott Alinity SARS-CoV2 Assay

SARS-CoV-2 concntration	Number tested	Number detected
1X to 2X LOD	20	20
20X LOD	20	20
Negative	31	0

- Do you believe it's perfect?
  - $\mathrm{P}(\text{true positive}) = 0.976(0.913, 1.000)$
  - $\mathrm{P}(\text{true negative}) = 0.970(0.888, 0.999)$

Why be Bayesian?

- Positive test $\neq$ carrying virus
  - Positive and carrying virus (true positive)
  - Positive and not carrying virus (false positive)
- Assume prevalence (proportion of population carrying virus) is 3% and imagine that we test 1000 people

How much should I worry?

$$\begin{aligned}
N_{\text{infected}} &= 1000 \times 0.03 = 30 \\
N_{\text{not infected}} &= 1000 \times 0.97 = 970 \\
N_{\text{infected and positive}} &= N_{\text{infected}} \times 0.976 \\
N_{\text{not infected and positive}} &= N_{\text{not infected}} \times 0.030
\end{aligned}$$

How much should I worry?

$$
\begin{aligned}
N_{\text{infected}} &= 1000 \times 0.03 = 30 \\
N_{\text{not infected}} &= 1000 \times 0.97 = 970 \\
N_{\text{infected and positive}} &= N_{\text{infected}} \times 0.976 \\
N_{\text{not infected and positive}} &= N_{\text{not infected}} \times 0.030 \\
\frac{N_{\text{infected and positive}}}{N_{\text{positive}}} &= \frac{30 \times 0.976}{30 \times 0.976 + 970 \times 0.030} \\
&= \frac{29.3}{29.3 + 29.1} \\
&= 0.502
\end{aligned}
$$

Coin flip on whether you have COVID

Bayes' Rule

$$\begin{aligned}
\mathrm{P}(\text{infected}|\text{positive}) &= \frac{\mathrm{P}(\text{positive}|\text{infected})}{\mathrm{P}(\text{positive})}\mathrm{P}(\text{infected}) \\
&= \frac{0.976}{(0.976)(0.03) + (0.030)(0.97)}(0.03) \\
&= 0.502
\end{aligned}$$

$$\mathrm{P}(X|Y) = \frac{\mathrm{P}(Y|X)}{\mathrm{P}(X)}\mathrm{P}(Y)$$

Bayes' Rule for inference

$$
\begin{aligned}
\mathrm{P}(\theta|X) &= \frac{\mathrm{P}(X|\theta)}{\mathrm{P}(X)}\mathrm{P}(\theta) \\
\theta &= \text{parameter} \\
X &= \text{data}
\end{aligned}
$$

Bayes' Rule for inference

$$\begin{aligned}
\mathrm{P}(\theta|X) &= \frac{\mathrm{P}(X|\theta)}{\mathrm{P}(X)}\mathrm{P}(\theta) \\
\theta &= \text{parameter} \\
X &= \text{data} \\
\mathrm{P}(X|\theta) &= \text{likelihood}
\end{aligned}$$

Maximum likelihood estimate: value of θ that maximizes $\mathrm{P}(X|\theta)$

Bayes' Rule for inference

$$\begin{aligned}
\mathrm{P}(\theta|X) &= \frac{\mathrm{P}(X|\theta)}{\mathrm{P}(X)}\mathrm{P}(\theta) \\
\theta &= \text{parameter} \\
X &= \text{data} \\
\mathrm{P}(X|\theta) &= \text{likelihood} \\
\mathrm{P}(\theta) &= \text{prior distribution of } \theta \\
\mathrm{P}(\theta|X) &= \text{posterior distribution of } \theta
\end{aligned}$$

Bayesian inference: based on posterior distribution, $\mathrm{P}(\theta|X)$

Stan

A probabilistic language for Bayesian analysis

```
data {
  int<lower=0> k;       // number of positives observed
  int<lower=0> N;       // number in sample
}

parameters {
  real<lower=0, upper=1> p;   // frequency of positives in the sample
}

model {
  // likelihood
  //
  k ~ binomial(N, p);

  // prior
  //
  p ~ uniform(0.0, 1.0);
}
```