# Chapter 2.2

# Bayes' Theorem

# Bayes' Theorem

- In Bayesian statistics, we select the prior, $\pi(\theta)$, and the likelihood, $f(y|\theta)$
- Based on these two pieces of information, we must compute the posterior $p(\theta|y)$
- Bayes' theorem is the mathematical formula to convert the likelihood and prior to the posterior
- Bayes theorem:

$$p(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{m(y)}$$

- This holds for discrete (PMF) and continuous (PDF) cases

# Bayes' theorem

- Bayes theorem in math:

$$p(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{m(y)}$$

- Bayes theorem in words:

$$p(\theta|y) = \frac{\text{Likelihood} * \text{Prior}}{\text{marginal distribution of Y}}$$

- As in the formula for a conditional distribution, $m(y)$ is just the normalizing constant required so that $\int p(\theta|y)d\theta = 1$
- Most of the time $m(y)$ can be ignored because it doesn't depend on $\theta$ and the objective is to study the posterior of $\theta$

# Derivation of Bayes' theorem

- In lab this week you will prove the theorem
- It follows pretty quickly from the definition of conditional probabilities
- As the proof shows, this is a general theorem about probability that can be applied to non-Bayesian (football) and Bayesian (HIV and Robins) problems

# Football example

- A team plays half its games at home, wins 70% of its home games, and 40% of its road games. Given that the team wins a game, what's the probability it was a home game?

# Football example

- Denote the location as $\theta \in \{H, R\}$ and outcome as $Y \in \{W, L\}$
- The problem gives marginals $\text{Prob}(\theta = H) = \text{Prob}(\theta = R) = 0.5$
- The problem gives conditionals $\text{Prob}(Y = W|\theta = H) = 0.7$ and $\text{Prob}(Y = W|\theta = R) = 0.4$
- Bayes' theorem says

$$\text{Prob}(\theta = H|Y = W) = \frac{\text{Prob}(W|H)\text{Prob}(H)}{\text{Prob}(W)}$$

# Football example

- We are given $\text{Prob}(W|H) = 0.7$ and $\text{Prob}(H) = 0.5$
- We must compute the marginal $\text{Prob}(W)$

$$\begin{aligned}
\text{Prob}(Y = W) &= \sum_{\theta} f(\theta, Y = W) \\
&= \sum_{\theta} f(W|\theta) f(\theta) \\
&= f(W|H)f(H) + f(W|R)f(R) \\
&= 0.7 * 0.5 + 0.4 * 0.5 = 0.55
\end{aligned}$$

- Back to Bayes' theorem:

$$\begin{aligned}
\text{Prob}(H|W) &= \frac{\text{Prob}(W|H)\text{Prob}(H)}{\text{Prob}(H)} \\
&= \frac{0.7 * 0.5}{0.55} = 0.64
\end{aligned}$$

# HIV example

- Let $\theta$ be the parameter of interest with

$$\theta = \begin{cases} 0 & \text{patient does not have HIV} \\ 1 & \text{patient has HIV} \end{cases}$$

- The data is $Y$, defined as

$$Y = \begin{cases} 0 & \text{test is negative} \\ 1 & \text{test is positive} \end{cases}$$

- Objective: Derive the probability that the patient has HIV given the test results

- That is, we want $p(\theta|y)$

# HIV example - Likelihood

- The likelihood describes the distribution of the data as if we knew the parameters
- This is a statistical model for the data
- Since $Y$ is binary, we use a Bernoulli PMF for the likelihood
- We must specify the likelihood for both $\theta = 0$ and $\theta = 1$
- $\text{Prob}(Y = 1|\theta = 0) = q_0$ is the false positive rate
- $\text{Prob}(Y = 1|\theta = 1) = q_1$ is the true positive rate
- How might we select $q_0$ and $q_1$?

# HIV example - Prior

- The prior represents our uncertainty about the parameters before we observe the data
- Since $\theta$ is binary, we use a Bernoulli PMF for the prior
- Prob$(\theta = 1) = p$ is the population prevalence of HIV
- How might we select $p$?

# HIV example - Posterior

- Derive the posterior probability that the patient has HIV given a positive test
- It can be shown that

$$\text{Prob}(\theta = 1 | Y = 1) = \frac{q_1 p}{q_1 p + q_0(1 - p)}.$$

- See "HIV" in the online derivations (same steps as for football)

# HIV example - Posterior

- Derive the posterior probability that the patient has HIV given a negative test

- It can be shown that

$$\text{Prob}(\theta = 1 | Y = 0) = \frac{(1 - q_1)p}{(1 - q_1)p + (1 - q_0)(1 - p)}.$$

# Robins example

- Let $X$ be the number of robins in the forest
- Let $Y$ be the number of robins we observe
- Uniform prior: $\text{Prob}(X = x) = 1/20$ for $x \in \{0, ..., 19\}$
- Likelihood: $Y|X \sim \text{Binomial}(X, 0.2)$ (0.2 is the detection probability)

# Robins example

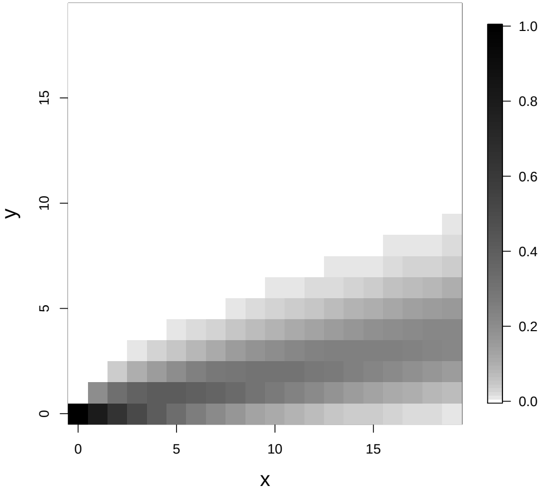
- Given that and we do not observe any birds, what is the probability that no birds are in the forest?

- Translation: What is $\text{Prob}(X = 0 | Y = 0)$?

- Intuitively, how would this change if the prior was $\text{Prob}(X = x) = 1/100$ for $x \in \{0, ..., 99\}$

- Intuitively, how would this change if the detection probability increased from 0.2 to 0.9?

# Joint PMF $f(x, y)$




$$f(x, y) = f(y|x)f_X(x) = \text{dbinom}(y, x, 0.2)(1/20)$$

# Robins example

- Bayes: $\text{Prob}(X = x|Y = y) = \frac{\text{Prob}(Y=y|X=x)\text{Prob}(X=x)}{\text{Prob}(Y=y)}$
- We know $\text{Prob}(Y = y|X = x) = \binom{x}{y}0.2^y 0.8^{x-y}$ is the binomial PMF
- We also know $\text{Prob}(X = x) = 1/20$ for all $x$
- Previously we found that $\text{Prob}(Y = 0) = 1/20 \sum_{x=0}^{19} 0.8^x$
- Bayes: $\text{Prob}(X = 0|Y = 0) = \frac{\text{Prob}(Y=0|X=0)\text{Prob}(X=0)}{\text{Prob}(Y=0)} = \frac{1*(1/20)}{1/20 \sum_{x=0}^{19} 0.8^x} = 0.202$

# Robins example

- Intuitively, how would this change if the prior was $\text{Prob}(X = x) = 1/100$ for $x \in \{0, ..., 99\}$

  It should decrease because the prior probability of $X = 0$ decreases (in fact it is 0.200)

- Intuitively, how would this change if the detection probability increased from 0.2 to 0.9?

  It should increase because with better detection probability we can be more confident in our sample (in fact it is 0.900)