# Chapter 8.3 Bayesian Inference with Discrete Priors

Jim Albert and Monika Hu

Chapter 8 Modeling Measurement and Count Data

# Example: Roger Federer's time-to-serve

- Want to measure efficiency of a server in tennis
- Consider the time-to-serve which is the measured time in seconds between the end of the previous point and the beginning of the current point.
- How long, on average, is Roger Federer's time-to-serve?

# Data

- Suppose one collects a single time-to-serve measurement in seconds denoted as $Y$.
- Assume $Y$ is Normally distributed with unknown mean $\mu$ and standard deviation $\sigma = 4$.
- Assume that the standard deviation $\sigma$ of the measurement distribution is known and the objective is to learn about the single mean measurement $\mu$.

# Normal density

- Recall the Normal probability curve has the general form

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}, -\infty < y < \infty.$$

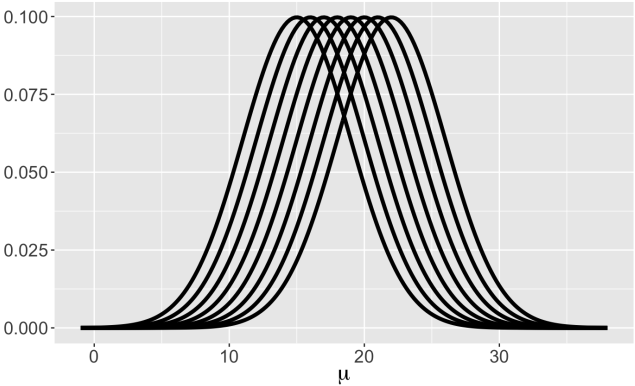
- Since $\sigma$ is known, the only parameter in Equation (8.3) is $\mu$.
- We are interested in learning about the mean time-to-serve $\mu$.

# A discrete prior

- Specify a subjective discrete prior for Federer's mean time-to-serve by specifying a list of plausible values for $\mu$ and assigning a probability to each of these values.
- Suppose one thinks that values of the equally spaced values $\mu = 15, 16, \ldots, 22$ are plausible.
- Assign a Uniform prior will be assigned where each value of $\mu$ is assigned the same probability $1/8$.

# The prior

- Each value of $\mu$ corresponds to a particular Normal sampling curve for the time-to-serve measurement. Each sampling curve has the same prior probability.



# Data

- One collects a single time-to-serve measurement for Federer, and suppose it is $Y$ =15.1 seconds.
- The likelihood function is the Normal density of the actual observation $u$ viewed as a function of the mean $\mu$.
- By substituting in the observation $y$ =15.1 and the known value of $\sigma$=4, one obtains

$$L(\mu) = \frac{1}{\sqrt{2\pi}4} \exp\left\{-\frac{1}{2(4)^2}(15.1-\mu)^2\right\}.$$

- This calculation is repeated for each of the eight values $\mu$ =15,16, ..., 22, obtaining eight likelihood values.

# Posterior

- One applies Bayes' rule to obtain the posterior distribution for $\mu$.
- The posterior probability of the value $\mu = \mu_i$ given the data $y$ for a discrete prior has the form

$$\pi(\mu_i \mid y) = \frac{\pi(\mu_i) \times L(\mu_i)}{\sum_j \pi(\mu_j) \times L(\mu_j)},$$

- where $\pi(\mu_i)$ is the prior probability of $\mu = \mu_i$ and $L(\mu_i)$ is the likelihood function evaluated at $\mu = \mu_i$.

# Table of posterior probabilties

| $\mu$ | Prior | Data/Likelihood | Posterior |
|---|---|---|---|
| 15 | 0.125 | 0.0997 | 0.1888 |
| 16 | 0.125 | 0.0972 | 0.1842 |
| 17 | 0.125 | 0.0891 | 0.1688 |
| 18 | 0.125 | 0.0767 | 0.1452 |
| 19 | 0.125 | 0.0620 | 0.1174 |
| 20 | 0.125 | 0.0471 | 0.0892 |
| 21 | 0.125 | 0.0336 | 0.0637 |
| 22 | 0.125 | 0.0225 | 0.0427 |

# Inference

- The posterior distribution for $\mu$ favors values $\mu = 15$, and 16.
- The posterior probabilities decrease as a function of $\mu$.
- The sample mean is $y$ =15.1 and the $\mu$ value closest to the sample mean ($\mu$=15) is assigned the highest posterior probability.

# Multiple measurements

- Suppose one collects $n$ time-to-serve measurements, denoted as $Y_1, ..., Y_n$, that are Normally distributed with mean $\mu$ and fixed standard deviation $\sigma = 4$.
- Each observation follows the same Normal density

$$f(y_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-(y_i - \mu)^2}{2\sigma^2}\right\}, -\infty < y_i < \infty.$$

- Since $\sigma = 4$ is known, the only parameter is $\mu$ and we are interested in learning about this mean parameter $\mu$.
- The same discrete Uniform prior is used for $\mu$.

# The data

- Suppose one collects a sample of 20 times-to-serve for Federer:

15.1 11.8 21.0 22.7 18.6 16.2 11.1 13.2 20.4 19.2 21.2 14.3 18.6 16.8 20.3 19.9 15.0 13.4 19.9 15.3

- The likelihood function is the joint density of the actual observed values $y_1, ..., y_n$ viewed as a function of the mean $\mu$.
- After some algebra, one obtains

$$L(\mu) = \exp\left\{-\frac{20}{2(4)^2}(\bar{y} - \mu)^2\right\}$$

# The likelihood

- Substitute the known values $n$ =20 and the standard deviation $\sigma$=4.
- Compute the sample mean $\bar{y} = (15.1 + 11.8 + ... + 15.3)/20 = 17.2$
- For each possible value of $\mu$, we substitute the value to find the corresponding likelihood.
- For example, the likelihood of $\mu$ =15 is equal to

$$L(15) = \exp\left\{-\frac{20}{2(4)^2}(17.2 - 15)^2\right\} \approx 0.022.$$

- This calculation is repeated for each of the eight values $\mu$ =15,16,...,22, obtaining eight likelihood values.

# The posterior

- Apply Bayes' rule to obtain the posterior distribution for $\mu$.
- The posterior probability of $\mu = \mu_i$ given the sequence of recorded times-to-serve $y_1, ..., y_n$

$$\pi(\mu_i \mid y_1, \cdots, y_n) = \frac{\pi(\mu_i) \times L(\mu_i)}{\sum_j \pi(\mu_j) \times L(\mu_j)},$$

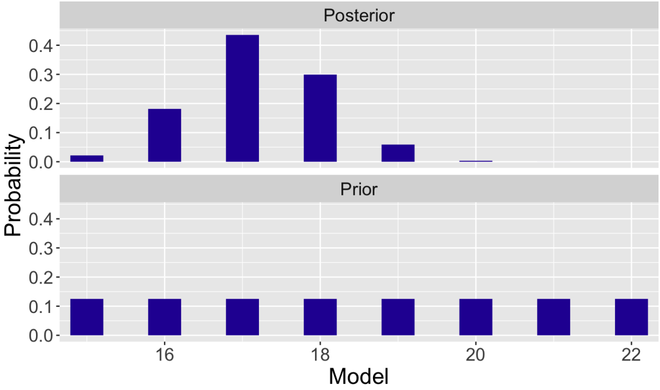
- where $\pi(\mu_i)$ is the prior probability of $\mu = \mu_i$ and $L(\mu_i)$ is the likelihood function evaluated at $\mu = \mu_i$.

# Table of posterior probabilities

| $\mu$ | Prior | Data/Likelihood | Posterior |
|---|---|---|---|
| 15 | 0.125 | 0.0217 | 0.0217 |
| 16 | 0.125 | 0.1813 | 0.1815 |
| 17 | 0.125 | 0.4350 | 0.4353 |
| 18 | 0.125 | 0.2990 | 0.2992 |
| 19 | 0.125 | 0.0589 | 0.0589 |
| 20 | 0.125 | 0.0033 | 0.0033 |
| 21 | 0.125 | 0.0001 | 0.0001 |
| 22 | 0.125 | 0.0000 | 0.0000 |

# Graph of prior and posterior

- Posterior for $\mu$ favors the values $\mu = 16$, 17, and 18 seconds.



# Inference: Federer's time-to-serve

- Our prior said that any of the eight possible values of $\mu$ were equally likely with probability 0.125.
- After observing the sample of 20 measurements, one believes $\mu$ is most likely 16, 17, and 18 seconds, with respective probabilities 0.181, 0.425, and 0.299.
- $\mu$is in the set $\{16, 17, 18\}$ seconds with probability 0.915.

$$P(16 \leq \mu \leq 18) = 0.181 + 0.435 + 0.299 = 0.915$$

- This region of values of $\mu$ is called a 91.5% posterior probability region for the mean time-to-serve $\mu$.