Online sellers example

Joachim Vandekerckhove and Michael Lee

Online Sellers

- This introductory modeling example comes from the YouTube channel 3Blue1Brown, and involves comparing the positive ratings of three online sellers.

Seller	Positive	Total	Percentage
One	10	10	100%
Two	48	50	96%
Three	186	200	93%

Online Sellers

- This introductory modeling example comes from the YouTube channel 3Blue1Brown, and involves comparing the positive ratings of three online sellers.
- They each have different numbers of positive ratings from different numbers of total user evaluations.

Seller	Positive	Total	Percentage
One	10	10	100%
Two	48	50	96%
Three	186	200	93%

Online Sellers

- This introductory modeling example comes from the YouTube channel 3Blue1Brown, and involves comparing the positive ratings of three online sellers.
- They each have different numbers of positive ratings from different numbers of total user evaluations.

Seller	Positive	Total	Percentage
One	10	10	100%
Two	48	50	96%
Three	186	200	93%

- The research question is to infer from these data which seller has the highest probability of generating a positive user evaluation.

Online Sellers (Continued)

- The data are the counts k_i of positive ratings out of n_i total evaluations for the ith seller.

Online Sellers (Continued)

- The data are the counts k_i of positive ratings out of n_i total evaluations for the ith seller.
  - This means that $k_1 = 10$, $k_2 = 48$, $k_3 = 186$ and $n_1 = 10$, $n_2 = 50$, $n_3 = 200$.

Online Sellers (Continued)

- The data are the counts k_i of positive ratings out of n_i total evaluations for the ith seller.
  - This means that $k_1 = 10$, $k_2 = 48$, $k_3 = 186$ and $n_1 = 10$, $n_2 = 50$, $n_3 = 200$.
- A simple statistical model of the rating behavior is to assume the ith seller has some underlying probability θ_i of generating a positive review each time they are evaluated.

Online Sellers (Continued)

- The data are the counts k_i of positive ratings out of n_i total evaluations for the ith seller.
  - This means that $k_1 = 10$, $k_2 = 48$, $k_3 = 186$ and $n_1 = 10$, $n_2 = 50$, $n_3 = 200$.
- A simple statistical model of the rating behavior is to assume the ith seller has some underlying probability θ_i of generating a positive review each time they are evaluated.
  - θ_i are parameters, since they are latent probabilities controlling the process that generates the data.

Online Sellers (Continued)

- The data are the counts k_i of positive ratings out of n_i total evaluations for the ith seller.
  - This means that $k_1 = 10$, $k_2 = 48$, $k_3 = 186$ and $n_1 = 10$, $n_2 = 50$, $n_3 = 200$.
- A simple statistical model of the rating behavior is to assume the ith seller has some underlying probability θ_i of generating a positive review each time they are evaluated.
  - θ_i are parameters, since they are latent probabilities controlling the process that generates the data.
  - Let us assume that these probabilities are equally likely to take any possible value, so $\theta_i \sim \mathrm{uniform}(0, 1)$.

Online Sellers (Continued)

- The data are the counts k_i of positive ratings out of n_i total evaluations for the ith seller.
  - This means that $k_1 = 10$, $k_2 = 48$, $k_3 = 186$ and $n_1 = 10$, $n_2 = 50$, $n_3 = 200$.
- A simple statistical model of the rating behavior is to assume the ith seller has some underlying probability θ_i of generating a positive review each time they are evaluated.
  - θ_i are parameters, since they are latent probabilities controlling the process that generates the data.
  - Let us assume that these probabilities are equally likely to take any possible value, so $\theta_i \sim \text{uniform}(0, 1)$.
- The process that generates the positive counts is then given by $k_i \sim \text{binomial}(\theta_i, n_i)$.

Exercise

Implement the model with these data and assumptions in JAGS. Which seller is better?

- The probabilities are equally likely to take any possible value, so $\theta_i \sim \text{uniform}(0, 1)$.

Seller	Positive	Total	Percentage
One	10	10	100%
Two	48	50	96%
Three	186	200	93%

Exercise

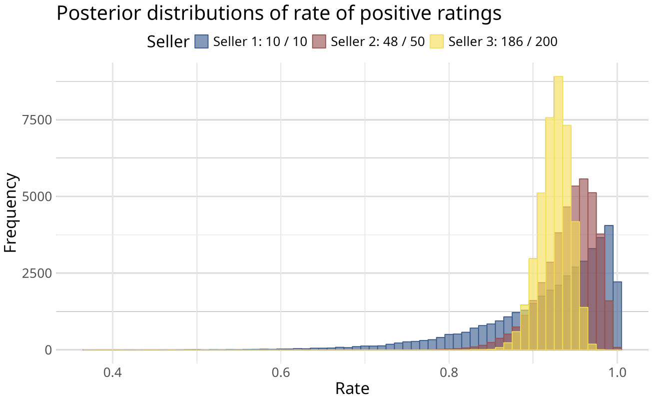
Implement the model with these data and assumptions in JAGS.
Which seller is better?

- The probabilities are equally likely to take any possible value, so $\theta_i \sim \text{uniform}(0, 1)$.
- The process that generates the positive counts is given by $k_i \sim \text{binomial}(\theta_i, n_i)$.

Seller	Positive	Total	Percentage
One	10	10	100%
Two	48	50	96%
Three	186	200	93%

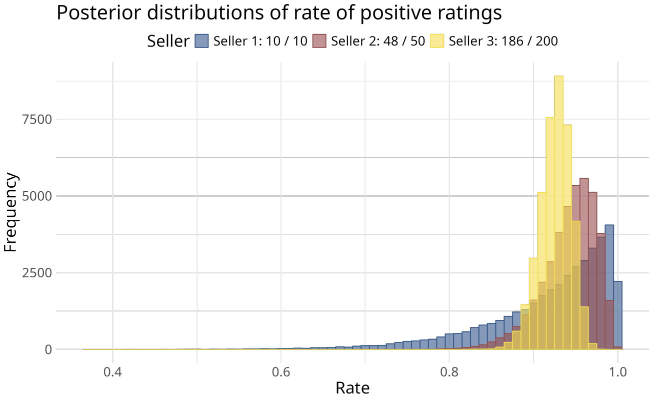
Rate Inferences

- The posterior distributions for the underlying rate of positive ratings show what values are plausible, based on the data and the assumptions of the model.



Rate Inferences (Continued)

- The posterior distributions can be summarized by 95% credible intervals, which are $(0.70, 1.00)$ for Seller 1, $(0.86, 0.99)$ for Seller 2, and $(0.89, 0.96)$ for Seller 3.



Which Seller is Better?

- The posterior distributions represent everything we know about the possible underlying rates of positive reviews.

	Seller 1	Seller 2	Seller 3
Seller 1	–	0.46	0.57
Seller 2	0.54	–	0.70
Seller 3	0.43	0.30	–

Which Seller is Better?

- The posterior distributions represent everything we know about the possible underlying rates of positive reviews.
- But, answering "Which seller is better?" requires being clear on what it means to be better, and applying that criterion to the posterior distributions.

	Seller 1	Seller 2	Seller 3
Seller 1	–	0.46	0.57
Seller 2	0.54	–	0.70
Seller 3	0.43	0.30	–

Which Seller is Better?

- The posterior distributions represent everything we know about the possible underlying rates of positive reviews.
- But, answering "Which seller is better?" requires being clear on what it means to be better, and applying that criterion to the posterior distributions.
    - One approach would be to find the seller who is most likely to have a better positive rate than the other two, which is Seller 2.

	Seller 1	Seller 2	Seller 3
Seller 1	–	0.46	0.57
Seller 2	0.54	–	0.70
Seller 3	0.43	0.30	–

Which Seller is Better?

- The posterior distributions represent everything we know about the possible underlying rates of positive reviews.
- But, answering "Which seller is better?" requires being clear on what it means to be better, and applying that criterion to the posterior distributions.
  - One approach would be to find the seller who is most likely to have a better positive rate than the other two, which is Seller 2.

	Seller 1	Seller 2	Seller 3
Seller 1	–	0.46	0.57
Seller 2	0.54	–	0.70
Seller 3	0.43	0.30	–

- Another approach would be to find the seller with the greatest probability of having at least a 90% positive rate.

Which Seller is Better?

- The posterior distributions represent everything we know about the possible underlying rates of positive reviews.
- But, answering "Which seller is better?" requires being clear on what it means to be better, and applying that criterion to the posterior distributions.
  - One approach would be to find the seller who is most likely to have a better positive rate than the other two, which is Seller 2.

	Seller 1	Seller 2	Seller 3
Seller 1	–	0.46	0.57
Seller 2	0.54	–	0.70
Seller 3	0.43	0.30	–

- Another approach would be to find the seller with the greatest probability of having at least a 90% positive rate.
  - These are 0.68 for Seller 1, 0.89 for Seller 2, 0.92 for Seller 3.

Key Points

- Online sellers is a simple introductory model, but has the basic features of a parameter that controls a data-generating process, and observed data.

Key Points

- Online sellers is a simple introductory model, but has the basic features of a parameter that controls a data-generating process, and observed data.
- Inference about parameters represents uncertainty about their possible values, based on the available data and the assumptions of the model.