

Rate of Lactose Intolerance

Aleksa Kostic

2/20/2022

```
library(bayesrules)
```

$\pi \sim \text{Beta}$

$\mathcal{L}(\pi|y) \sim \text{Binom}(\pi)$

```
pi <- c(0.4,0.5,0.6,0.7)
f_pi <-c(0.1,0.2,0.44,0.26)
```

	π	0.4	0.5	0.6	0.7
prior	$f(\pi)$	0.1	0.2	0.44	0.26
guessed posterior	$f(\pi y)$	0.05	0.1	0.75	0.1

Since the proportion of people in the sample who are lactose intolerant is 0.5875, that most closely matches 0.6, and so the posterior model would likely increase the probability of the true value of π as being 0.6 and decrease it for the other values (0.4, 0.5, 0.7)

```
likelihood <- dbinom(47, 80, pi)
```

	π	0.4	0.5	0.6	0.7
prior	$f(\pi)$	0.1	0.2	0.44	0.26
likelihood	$\mathcal{L}(\pi y)$	0.0003	0.0264	0.0880	0.0093

```
normalizing_const <- sum(f_pi*likelihood)
```

```
posterior <- (f_pi*likelihood)/normalizing_const
```

	π	0.4	0.5	0.6	0.7
posterior	$f(\pi y)$	0.00064	0.11354	0.83370	0.05201

The change from the prior to the posterior here is greater than the one that I had guessed. While the trends of change I anticipated were correct, the posterior for $\pi = 0.6$ was greater than my guess, and the posteriors for $\pi = 0.4, 0.5, 0.7$ were less than I guessed.

If she'd gotten a sample of 800 people and 470 were lactose intolerant, the posterior value for $\pi = 0.6$ given $y = 470/800$ would be very high, around 0.99

```
likelihood_large <- dbinom(470, 800, pi)
```

	π	0.4	0.5	0.6	0.7
prior	$f(\pi)$	0.1	0.2	0.44	0.26
likelihood	$\mathcal{L}(\pi y)$	$4.92e-27$	$1.29e-07$	0.0221	$3.80e-12$

```
normalizing_const_large <- sum(f_pi*likelihood_large)
```

```
posterior_large <- (f_pi*likelihood_large)/normalizing_const_large
```

	π	0.4	0.5	0.6	0.7
posterior	$f(\pi y)$	$5.055567e-26$	$2.646640e-06$	$9.999974e-01$	$1.015611e-10$

My guess was correct. If we were to round these values to four decimal places, the posterior values for 0.4, 0.5, 0.7 would be 0.0000, and the posterior value for 0.6 would be 1.0000.