

Bayesian Inference for Discrete Random Variables

Statistical Inference II

Mine Dogucu

Noah Johnson

April 17, 2018

Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \implies P(A \cap B) = P(A|B) * P(B)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \implies P(A \cap B) = P(B|A) * P(A)$$

Bayes' Theorem

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A \cap B) + P(A \cap \tilde{B})}$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\tilde{B})P(\tilde{B})}$$

More generally, for a partition of B_1, B_2, \dots, B_M of Ω .

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^M P(A|B_i)P(B_i)}$$

Bayesian Universe

Full Universe

	A	\tilde{A}
B_1		
B_2		

Reduced Universe

	A
B_1	
B_2	

Bayesian Inference by Tables

There is an urn containing a total of 5 balls, some of which are red and the rest of which are green. We don't know how many of the balls are red. Let θ be the number of red balls in the urn.

Bayesian universe of the experiment

	Y = 0 (pull a green ball)	Y = 1 (pull a red ball)
0 red balls in urn		
1 red balls in urn		
2 red balls in urn		
3 red balls in urn		
4 red balls in urn		
5 red balls in urn		

Joint and marginal prior distributions

θ_i	$\pi(\theta_i)$	$P(Y = 0 \theta_i)$	$P(Y = 1 \theta_i)$	$P(Y = 0 \cap \theta_i)$	$P(Y = 1 \cap \theta_i)$
0	1/6	1	0	5/30	0
1	1/6	4/5	1/5	4/30	1/30
2	1/6	3/5	2/5	3/30	2/30
3	1/6	2/5	3/5	2/30	3/30
4	1/6	1/5	4/5	1/30	4/30
5	1/6	0	5/5	0	5/30

We find the marginal distribution by summing the last two columns.

$$P(Y = 0) = \frac{1}{2}$$

$$P(Y = 1) = \frac{1}{2}$$

Finding the posterior probability of $\theta|(Y = 1)$

θ_i	$\pi(\theta_i)$	$P(Y = 1 \theta_i)$	$P(Y = 1 \theta_i) * \pi(\theta_i)$	$\pi(\theta_i Y = 1)$
0	1/6	0	0	0
1	1/6	1/5	1/30	1/15
2	1/6	2/5	2/30	2/15
3	1/6	3/5	3/30	3/15
4	1/6	4/5	4/30	4/15
5	1/6	5/5	5/30	5/15

The posterior probability distribution after second red observation (no replacement).

θ_i	$\pi(\theta_i)$	$P(Y = 1 \theta_i)$	$P(Y = 1 \theta_i) * \pi(\theta_i)$	$\pi(\theta_i Y = 1)$
0	0	0	0	0
1	1/15	0	0	0
2	2/15	1/4	2/60	1/20
3	3/15	2/4	6/60	3/20
4	4/15	3/4	12/60	6/20
5	5/15	4/4	20/60	10/20

Consider an exam with 4 multiple-choice questions. Each question has 3 options. A student who hasn't studied guesses at each question with a probability of $1/3$ of getting it right. Students who have studied might have probabilities $2/3$ or 0.95 of getting questions right. Make a table of likelihoods. You may want to use R for this question.

```
t <- tibble(theta = c(1/3, 2/3, 0.95), prior = rep(1/3, 3))

t$likelihood0 <- dbinom(0, 4, t$theta)
t$likelihood1 <- dbinom(1, 4, t$theta)
t$likelihood2 <- dbinom(2, 4, t$theta)
t$likelihood3 <- dbinom(3, 4, t$theta)
t$likelihood4 <- dbinom(4, 4, t$theta)

t
```

```
## # A tibble: 3 x 7
##   theta prior likelihood0 likelihood1 likelihood2 likelihood3 likelihood4
##   <dbl> <dbl>         <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1 0.333 0.333    0.198          0.395          0.296          0.0988         0.0123
## 2 0.667 0.333    0.0123         0.0988         0.296          0.395          0.198
## 3 0.950 0.333    0.00000625     0.000475      0.0135         0.171          0.815
```

You have observed that a student has responded to two questions correctly. What can you conclude about whether they have studied or not?

```
prior.times.likelihood <- t$prior * t$likelihood2
marginal <- sum(prior.times.likelihood)

posterior <- prior.times.likelihood / marginal
posterior
```

```
## [1] 0.48883284 0.48883284 0.02233431
```

We can't tell whether the student was just guessing or is a medium-strength student ($\theta = 2/3$). But there's only a 2% chance that they are an exceptional student ($\theta = 0.95$).

In another exam, you observe another student who has answered 50 out of 100 questions correctly. What can you conclude about whether they have studied or not?

```
lik50 <- dbinom(50, 100, t$theta)

ptl <- t$prior * lik50
marginal <- sum(ptl)

posterior <- ptl / marginal
posterior
```

```
## [1] 5.000000e-01 5.000000e-01 1.564142e-34
```

Once again, even with more observations, because there was an equal number of successes and failures, there is an equal posterior for not having studied, and for having studied a bit. But the probability of being an

exceptional student ($\theta = 0.95$) who got this unlucky is even smaller now. Really really small.

Part II

Suppose that a factory makes a large number of cars per day. With some small probability, a car is defective. We are going to visit this factory for a week ($N = 7$ days) and record the number of defective cars. Assume that the number of defective car Y_i on Day i follows a Poisson distribution with parameter λ , $Y_i \sim P(\lambda)$, and that the Y_i s are independent. After a week we have a sample Y_1, Y_2, \dots, Y_N . The likelihood of the sample is

$$p(y|\lambda) = p(y_1|\lambda)p(y_2|\lambda) \dots p(y_N|\lambda) = \prod_{i=1}^N p(y_i|\lambda)$$

In real life, we never know the parameter λ but we may have some prior belief about it. For the purposes of this class we will assume on average 1, through 10 cars break and since we do not know much about the car industry we will assume these parameters are equally likely.

In not-so real life, we will set the true $\lambda = 3$ and simulate 7 days of data from this true distribution. Simulate this data using R.

```
set.seed(34243)
data <- rpois(7, 3)
```

Now pretend we never knew the true value of λ and calculate the posterior distribution for estimating λ based on the data we “observed” (i.e. simulated).

```
t <- tibble(lambda = 1:10, prior = rep(1/10, 10))

likelihood <- sapply(t$lambda, function(l) return (prod(dpois(data, l))))

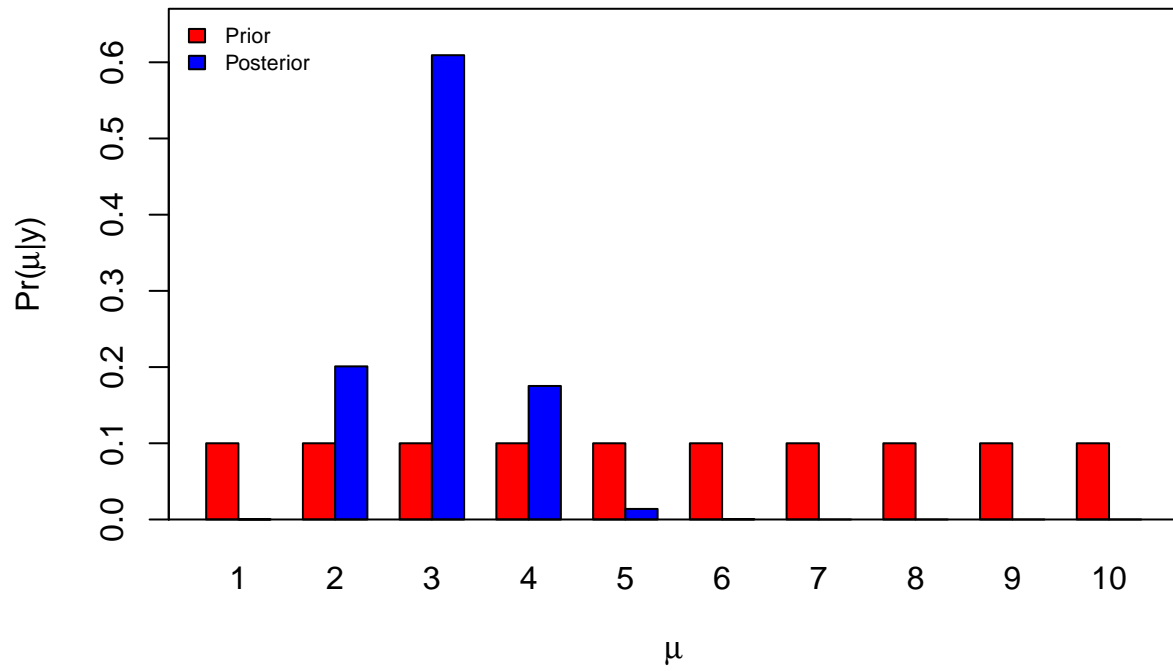
ptl <- t$prior * likelihood
marginal <- sum(ptl)

posterior <- ptl / marginal
posterior

## [1] 2.101484e-04 2.009391e-01 6.092961e-01 1.752031e-01 1.385736e-02
## [6] 4.844446e-04 9.640922e-06 1.270270e-07 1.221477e-09 9.161649e-12
poisd(p(data, t$lambda, t$prior)
```

```
##      Mu Prior   Likelihood   Posterior
## [1,]  1  0.1 3.664649e-09 2.101484e-04
## [2,]  2  0.1 3.504055e-06 2.009391e-01
## [3,]  3  0.1 1.062514e-05 6.092961e-01
## [4,]  4  0.1 3.055259e-06 1.752031e-01
## [5,]  5  0.1 2.416501e-07 1.385736e-02
## [6,]  6  0.1 8.447934e-09 4.844446e-04
## [7,]  7  0.1 1.681222e-10 9.640922e-06
## [8,]  8  0.1 2.215147e-12 1.270270e-07
## [9,]  9  0.1 2.130059e-14 1.221477e-09
## [10,] 10  0.1 1.597644e-16 9.161649e-12
```

Prior and posterior probability for μ given the data y



We can see that $\theta = 3$ is most likely given the data, as expected!