# Journal Club: Statistical Rethinking

## Chapter 2: Small Worlds and Large Worlds

Carles Milà

2022-03-03

# The title's metaphore

Small world:

- The world of our analysis
- Our data/sample realization

- The fitted model in the terms of our data

- In this chapter, we'll focus on this

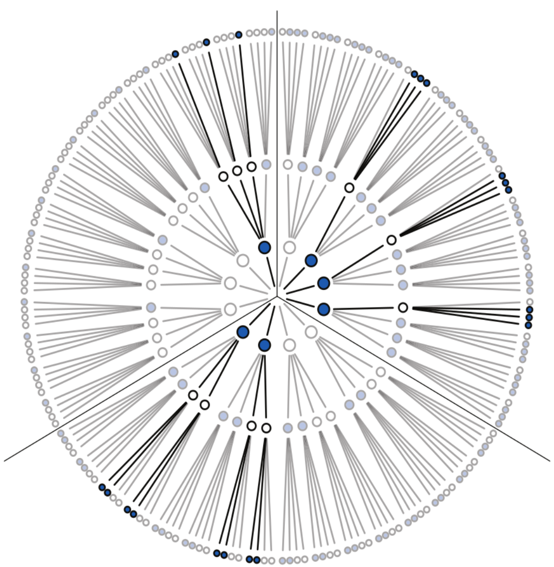
Big world:

- The phenomenon we are modelling

- The inference process

# The garden of forking paths

- Bayesian inference simplified: counting all possibilities
- The marble example: 4 marbles, they can be blue or white, but not how many of each colour. Therefore, there are 5 possibilities:
  (1) [○○○○], (2) [●○○○], (3) [●●○○], (4) [●●●○], (5) [●●●●]
- We sample with replacement 3 marbles and get: ●○●
- In how many ways can this sample be produced? Let's list them all!



| Conjecture | Ways to produce ●○● |
|---|---|
| [○○○○] | $0 \times 4 \times 0 = 0$ |
| [●○○○] | $1 \times 3 \times 1 = 3$ |
| [●●○○] | $2 \times 2 \times 2 = 8$ |
| [●●●○] | $3 \times 1 \times 3 = 9$ |
| [●●●●] | $4 \times 0 \times 4 = 0$ |

# The garden of forking paths

- Let's sample a fourth marble: blue. We can update as follows:

| Conjecture | Ways to produce ● | Prior counts | New count |
|---|---|---|---|
| [○○○○] | 0 | 0 | $0 \times 0 = 0$ |
| [●○○○] | 1 | 3 | $3 \times 1 = 3$ |
| [●●○○] | 2 | 8 | $8 \times 2 = 16$ |
| [●●●○] | 3 | 9 | $9 \times 3 = 27$ |
| [●●●●] | 4 | 0 | $0 \times 4 = 0$ |

- New update from factory: All bags have at least 1 blue/white marble, and the proportion of bags with 1/2/3 marbles are 3/2/1:

| Conjecture | Prior count | Factory count | New count |
|---|---|---|---|
| [○○○○] | 0 | 0 | $0 \times 0 = 0$ |
| [●○○○] | 3 | 3 | $3 \times 3 = 9$ |
| [●●○○] | 16 | 2 | $16 \times 2 = 32$ |
| [●●●○] | 27 | 1 | $27 \times 1 = 27$ |
| [●●●●] | 0 | 0 | $0 \times 0 = 0$ |

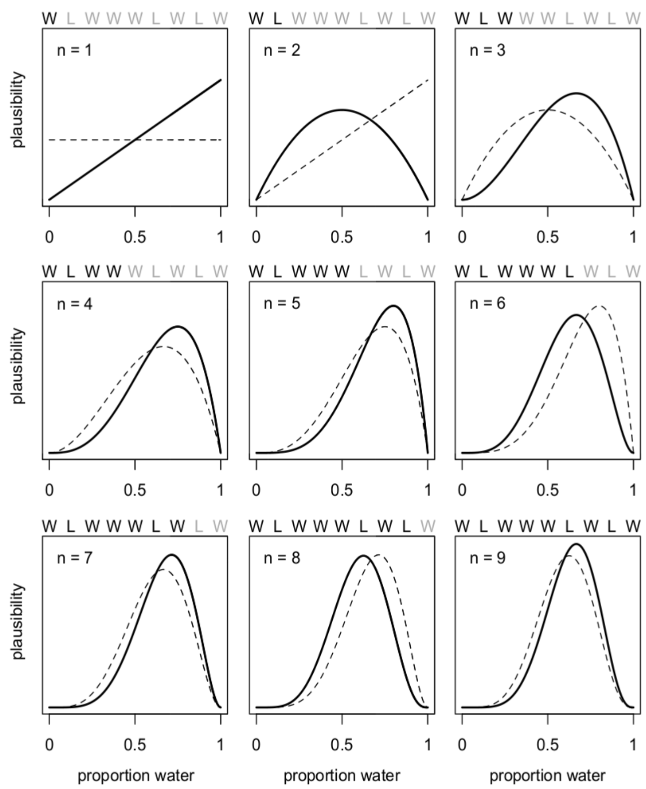
# The garden of forking paths

- We can summarise the updating process as:
  $$\text{plausibility of } p \text{ after } D_{\text{new}} \propto \text{ways } p \text{ can produce } D_{\text{new}} \times \text{prior plausibility of } p$$
- Turning this into probabilities by making sure that the sum of plausabilites adds up to 1:
  $$\text{plausibility of } p \text{ after } D_{\text{new}} = \frac{\text{ways } p \text{ can produce } D_{\text{new}} \times \text{prior plausibility } p}{\text{sum of products}}$$
- This is Bayes theorem! Some definitions:
  - Parameter: *p* (in this example)
  - Likelihood: relative number of ways that a parameter can produce the data
  - Prior: prior plausability of the parameter value
  - Posterior: updated plausability of parameter after observing the data

| Possible composition | $p$ | Ways to produce data | Plausibility |
|---|---|---|---|
| [○○○○] | 0 | 0 | 0 |
| [●○○○] | 0.25 | 3 | 0.15 |
| [●●○○] | 0.5 | 8 | 0.40 |
| [●●●○] | 0.75 | 9 | 0.45 |
| [●●●●] | 1 | 0 | 0 |

# Building a model

- Earth (independent) tossing example: W L W W W L W L W
- *p* parameter: proportion of water covering the globe
- Let's see how the plausabilites are updated with each toss!




- Here *p* is continuous, though grid-discretised
- More data, less uncertainty
- Bayesian inference perfect in the small world... but in the large world:
  - We are conditioning on a model. What if it is wrong?
  - We need to do checks (later chapters)!

# Components of a model

1. Variables:

    1. Observed: data. tosses in our example: $N = W + L$
    2. Unobserved: parameters. $p$ in our example.

2. Definitions:

    1. Observed variables:
    LIKELIHOOD: A mathematical function that gives us the plausability of the data given a parameter value. In our example we have the binomial distribution with PMF:

    $$Pr(W, L|p) = \frac{(W+L)!}{W!L!} p^W (1-p)^L$$

    2. Unobserved variables:
    PRIOR: Initial plausability assignment for each value of the parameter. They can be informative, or flat using a uniform distribution in [a,b]. For p, a=0 and b=1, and therefore:

    $$Pr(p) = \frac{1}{b-a} = \frac{1}{1-0} = 1$$

# Components of a model

1. Model: Let's summarise steps 1 and 2:

$$W \sim \text{Binomial}(N, p)$$

$$p \sim \text{Uniform}(0, 1)$$

2. Bayes' theorem: A bit of math and using conditional probability properties

$$Pr(W, L, p) = Pr(W, L|p)Pr(p)$$

$$Pr(W, L, p) = Pr(p|W, L)Pr(W, L)$$

Now we can equate the two sides of the expressions:

$$Pr(W, L|p)Pr(p) = Pr(p|W, L)Pr(W, L)$$

And get to Bayes theorem!

$$Pr(p|W, L) = \frac{Pr(W, L|p)Pr(p)}{Pr(W, L)}$$

# Components of a model

In word format:

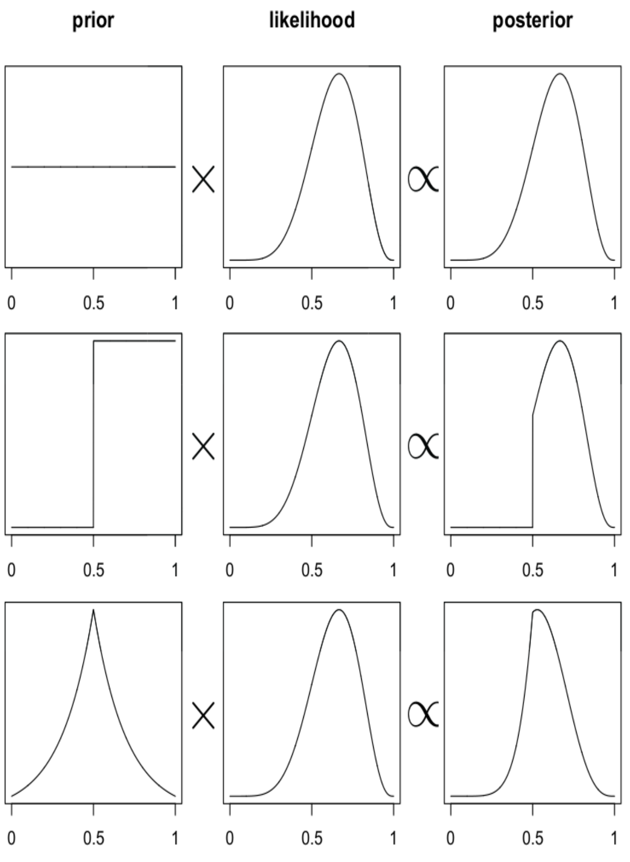
$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Average probability of data}}$$

Where the normalization constant is defined as:

$$Pr(W, L) = E[Pr(W, L|p)] = \int Pr(W, L|p) Pr(p) dp$$

# Components of a model

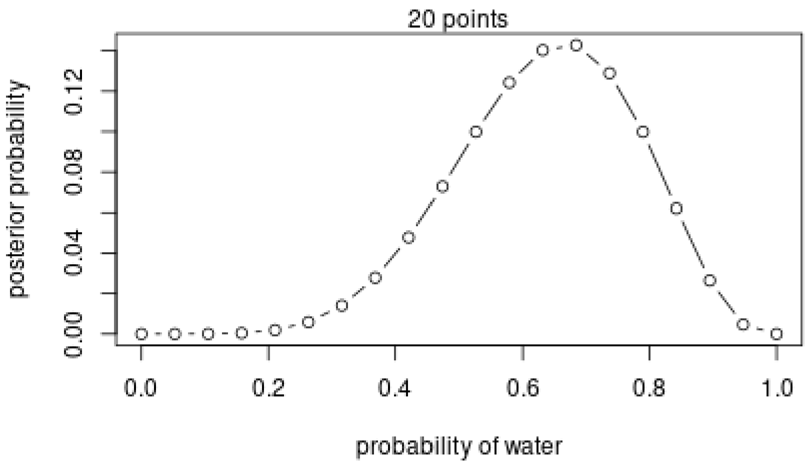
In graphic format:



# Estimation

How to perform inference? Option 1: Grid approximation.

```
# define grid
p_grid <- seq( from=0 , to=1 , length.out=20 )
# define prior
prior <- rep( 1 , 20 )
# compute likelihood at each value in grid
likelihood <- dbinom( 6 , size=9 , prob=p_grid )
# compute product of likelihood and prior
unstd.posterior <- likelihood * prior
# standardize the posterior, so it sums to 1
posterior <- unstd.posterior / sum(unstd.posterior)
# Plot
plot( p_grid , posterior , type="b" ,
xlab="probability of water" , ylab="posterior probability" )
mtext( "20 points" )
```



# Estimation

Option 2: quadratic inference: Find posterior mode + Approximate curvature near the peak with a Gaussian distribution.

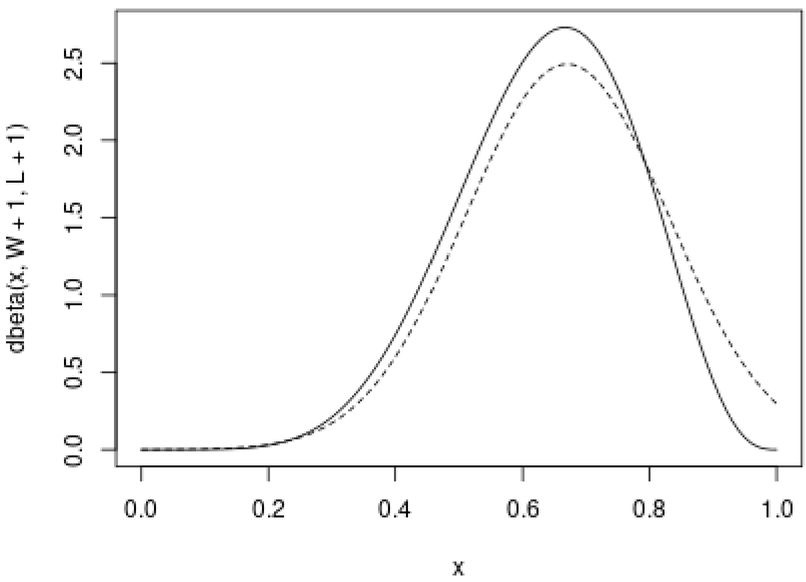
```r
library(rethinking)
globe.qa <- quap(
alist(
W ~ dbinom( W+L ,p) , # binomial likelihood
p ~ dunif(0,1)
# uniform prior
) ,
data=list(W=6,L=3) )
# display summary of quadratic approximation
precis( globe.qa)
```

```
##        mean        sd      5.5%     94.5%
## p 0.6666671 0.1571337 0.4155371 0.9177971
```

Limitations: this approximation may not always work, but other function can be tried.

# Estimation

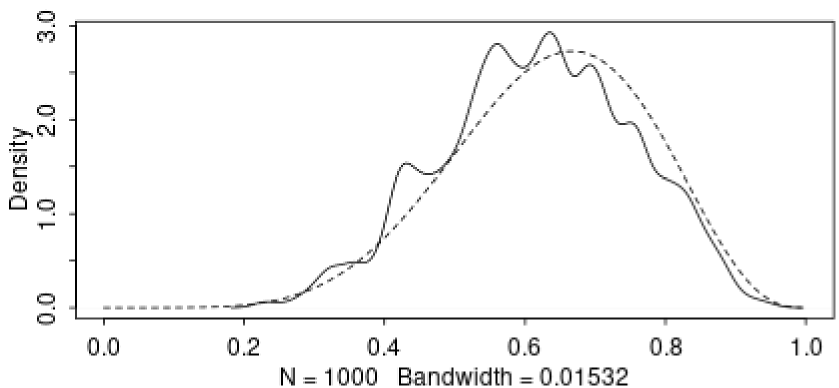
Option 3: Analytical: In some cases (conjugate priors) an analytical exact solution is possible.



# Estimation

Option 4: MCMC: We simulate from the posterior rather than approximate the function and work with these samples

```
n_samples <- 1000 # Number of simulations
p <- rep( NA , n_samples ) # Posterior simulations
p[1] <- 0.5 # Starting value
W <- 6 # Data: water tosses
L <- 3 # Data: land tosses
for ( i in 2:n_samples ) {
  p_new <- rnorm( 1 , p[i-1] , 0.1 ) # Simulate p based on previous value
  if ( p_new < 0 ) p_new <- abs( p_new ) # bound between 0 and 1
  if ( p_new > 1 ) p_new <- 2 - p_new
  q0 <- dbinom( W , W+L , p[i-1] ) # Get likelihood for p at step i-1 and i
  q1 <- dbinom( W , W+L , p_new )
  p[i] <- ifelse( runif(1) < q1/q0 , p_new , p[i-1] ) # decide whether to take the new p
  # or stay with the old one based on the likelihood ration and some chance
}
dens( p , xlim=c(0,1) )
curve( dbeta( x , W+1 , L+1 ) , lty=2 , add=TRUE )
```



# Practice

**2E1.** Which of the expressions below correspond to the statement: *the probability of rain on Monday*?

(1) $\Pr(\text{rain})$
(2) $\Pr(\text{rain}|\text{Monday})$
(3) $\Pr(\text{Monday}|\text{rain})$
(4) $\Pr(\text{rain}, \text{Monday}) / \Pr(\text{Monday})$

**2E2.** Which of the following statements corresponds to the expression: $\Pr(\text{Monday}|\text{rain})$?

(1) The probability of rain on Monday.
(2) The probability of rain, given that it is Monday.
(3) The probability that it is Monday, given that it is raining.
(4) The probability that it is Monday and that it is raining.

**2E3.** Which of the expressions below correspond to the statement: *the probability that it is Monday, given that it is raining*?

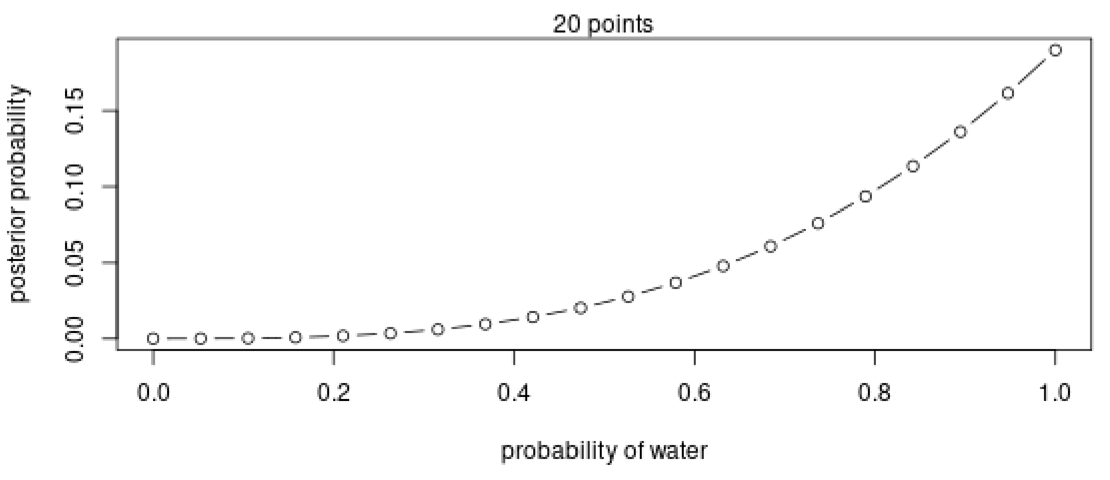
(1) $\Pr(\text{Monday}|\text{rain})$
(2) $\Pr(\text{rain}|\text{Monday})$
(3) $\Pr(\text{rain}|\text{Monday}) \Pr(\text{Monday})$
(4) $\Pr(\text{rain}|\text{Monday}) \Pr(\text{Monday}) / \Pr(\text{rain})$
(5) $\Pr(\text{Monday}|\text{rain}) \Pr(\text{rain}) / \Pr(\text{Monday})$

**2M1.** Recall the globe tossing model from the chapter. Compute and plot the grid approximate posterior distribution for each of the following sets of observations. In each case, assume a uniform prior for $p$.

(1) W, W, W
(2) W, W, W, L
(3) L, W, W, L, W, W, W

## For (1): W, W, W

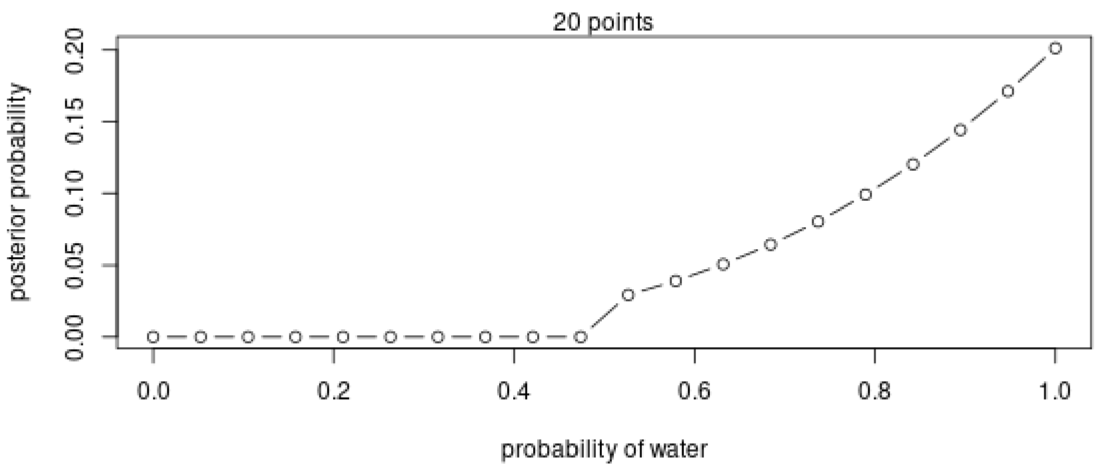
```
p_grid <- seq(from=0 , to=1 , length.out=20)
prior <- rep(1 , 20)
likelihood <- dbinom(3 , size=3, prob=p_grid ) # The only thing I've changed
unstd.posterior <- likelihood * prior
posterior <- unstd.posterior / sum(unstd.posterior)
plot( p_grid , posterior , type="b" ,
xlab="probability of water" , ylab="posterior probability" )
mtext( "20 points" )
```



**2M2.** Now assume a prior for $p$ that is equal to zero when $p < 0.5$ and is a positive constant when $p \geq 0.5$. Again compute and plot the grid approximate posterior distribution for each of the sets of observations in the problem just above.

For (1): W, W, W

```
p_grid <- seq(from=0 , to=1 , length.out=20)
prior <- ifelse(p_grid < 0.5, 0, 1) # The only thing I've changed
likelihood <- dbinom(3 , size=3, prob=p_grid )
unstd.posterior <- likelihood * prior
posterior <- unstd.posterior / sum(unstd.posterior)
plot( p_grid , posterior , type="b" ,
xlab="probability of water" , ylab="posterior probability" )
mtext( "20 points" )
```



**2M3.** Suppose there are two globes, one for Earth and one for Mars. The Earth globe is 70% covered in water. The Mars globe is 100% land. Further suppose that one of these globes—you don't know which—was tossed in the air and produced a "land" observation. Assume that each globe was equally likely to be tossed. Show that the posterior probability that the globe was the Earth, conditional on seeing "land" (Pr(Earth|land)), is 0.23.

$$Pr(Earth|land) = \frac{Pr(land|Earth) * Pr(Earth)}{Pr(land)}$$

```
prior_earth <- 0.5
likelihood_land_earth <- 0.3
likelihood_land_mars <- 1
denominator <- prior_earth*likelihood_land_earth + (1-prior_earth)*likelihood_land_mars
likelihood_land_earth*prior_earth/denominator
```

```
## [1] 0.2307692
```

**2M4.** Suppose you have a deck with only three cards. Each card has two sides, and each side is either black or white. One card has two black sides. The second card has one black and one white side. The third card has two white sides. Now suppose all three cards are placed in a bag and shuffled. Someone reaches into the bag and pulls out a card and places it flat on a table. A black side is shown facing up, but you don't know the color of the side facing down. Show that the probability that the other side is also black is 2/3. Use the counting method (Section 2 of the chapter) to approach this problem. This means counting up the ways that each card could produce the observed data (a black side facing up on the table).

Conditional probability:

$$Pr(backblack|frontblack) = \frac{Pr(backblack) \cap Pr(frontblack)}{Pr(frontblack)}$$

```
numerator <- 1*1/3 + 0*1/3 + 0*1/3
denominator <- 1*1/3 + 0.5*1/3 + 0*1/3
numerator/denominator
```

```
## [1] 0.6666667
```

**2M5.** Now suppose there are four cards: B/B, B/W, W/W, and another B/B. Again suppose a card is drawn from the bag and a black side appears face up. Again calculate the probability that the other side is black.

Conditional probability:

$$Pr(backblack|frontblack) = \frac{Pr(backblack) \cap Pr(frontblack)}{Pr(frontblack)}$$

```
numerator <- 1*1/4 + 0*1/4 + 0*1/4 + 1*1/4
denominator <- 1*1/4 + 0.5*1/4 + 0*1/4 + 1*1/4
numerator/denominator
```

```
## [1] 0.8
```

**2M6.** Imagine that black ink is heavy, and so cards with black sides are heavier than cards with white sides. As a result, it's less likely that a card with black sides is pulled from the bag. So again assume there are three cards: B/B, B/W, and W/W. After experimenting a number of times, you conclude that for every way to pull the B/B card from the bag, there are 2 ways to pull the B/W card and 3 ways to pull the W/W card. Again suppose that a card is pulled and a black side appears face up. Show that the probability the other side is black is now 0.5. Use the counting method, as before.

Conditional probability:

$$Pr(backblack|frontblack) = \frac{Pr(backblack) \cap Pr(frontblack)}{Pr(frontblack)}$$

```
numerator <- 1*1/6 + 0*1/3 + 0*1/2
denominator <- 1*1/6 + 0.5*1/3 + 0*1/2
numerator/denominator
```

```
## [1] 0.5
```