

Approximate Bayesian Computation

Kai Liu & Kavitha N.

Why ABC ?

Bayesian statistics:

$$f(\theta|y) \propto f(y|\theta)f(\theta)$$

- The likelihood is **intractable**



How does work?

- Step 1: construct a generative model M that produces the same type of data as your observed data
- Step 2: sample tentative parameter values θ from the prior distributions, plug these into the generative model and simulate a dataset \hat{y}

How does work?

- Step 3: check if the simulated data set/summary statistics matches the actual data/actual summary statistics, if yes, keep them, if no, throw them away. i.e. $d(\mathbf{y}_i, \mathbf{y}) \leq \epsilon$ or $d(S(\mathbf{y}_i), S(\mathbf{y})) \leq \epsilon$
- Step 4: repeat Step 2-3 a large number of times building up the list of probable parameter values
- Step 5: the distribution of the probable parameter values represents the posterior information

Other algorithms

1. ABC MCMC
2. ABC partial rejection control (ABC PRC)
3. ABC population Monte Carlo (ABC PMC)
4. ABC sequential Monte Carlo (ABC SMC)

ABC MCMC

- acceptance probability:

$$r = \min\left(1, \frac{\pi(\theta_{i+1})J(\theta_i|\theta_{i+1})}{\pi(\theta_i)J(\theta_{i+1}|\theta_i)}\right) \quad \text{if} \quad d(S(y_{i+1}), S(y)) \leq \epsilon$$

Pros and cons of ABC MCMC

- Pros: very efficient even the prior distribution differs substantially from the posterior distribution
- Cons: chains are in danger of getting stuck in the low probability regions of posterior distribution if proposal distribution is poorly chosen

ABC PRC

- Step 1: generate a pool of N candidate values for θ (called particles) from the prior distribution

ABC PRC

- Step 2: choose the distribution of a random variable η (i.e. $\eta \sim N(0, \sigma^2)$) to perturb θ to θ^* , that is $\theta^* = \theta + \eta$. $q_f(\cdot|\theta)$ is called forward transition kernel and $q_b(\cdot|\theta^*)$ is called backward transition kernel

ABC PRC

- Step 3: simulate data given the perturbed particles. If the simulated data satisfies the condition that $d(S(y_{sim}), S(y_{obs})) \leq \epsilon$, keep the particle, otherwise, discard it and repeat Step 2 until we obtain N particles with weights

$$w_i = \frac{\pi(\theta_i^*)q_b(\theta_i|\theta_i^*)}{\pi(\theta_i)q_f(\theta_i^*|\theta_i)} \text{ for } i = 1, \dots, N$$

ABC PRC

- Step 4: repeat Step 1-3 by sampling particles with probabilities of normalized weights in last iteration.

Pros and cons of ABC PRC

- Pros: discard particles from the pool in low-probability regions and increase the number of particles in high-probability regions
- Cons: the efficiency of the sampler relies heavily on the choice of the transition kernels and prior

ABC PMC

- Given the weight $w_{i,t-1}$ for particle $\theta_{i,t-1}$ on iteration $t-1$, the new weight $w_{i,t}$ for particle $\theta_{i,t}$ on iteration t is given by

$$w_{i,t} = \frac{\pi(\theta_{i,t})}{\sum_{j=1}^N w_{j,t-1} q(\theta_{j,t-1} | \theta_{i,t}, \sigma_{t-1})}$$

where $q(\cdot | \theta_{i,t}, \sigma_{t-1})$ is a Gaussian kernel and

$$\sigma_{t-1}^2 = \frac{2}{N} \sum_{i=1}^N \left(\theta_{i,t-1} - \frac{\sum_{j=1}^N \theta_{j,t-1}}{N} \right)^2$$

ABC SMC

- The weights in ABC SMC are given by

$$w_{i,t} = \frac{\pi(\theta_{i,t})}{\sum_{j=1}^N w_{j,t-1} q(\theta_{j,t-1} | \theta_{i,t})}$$

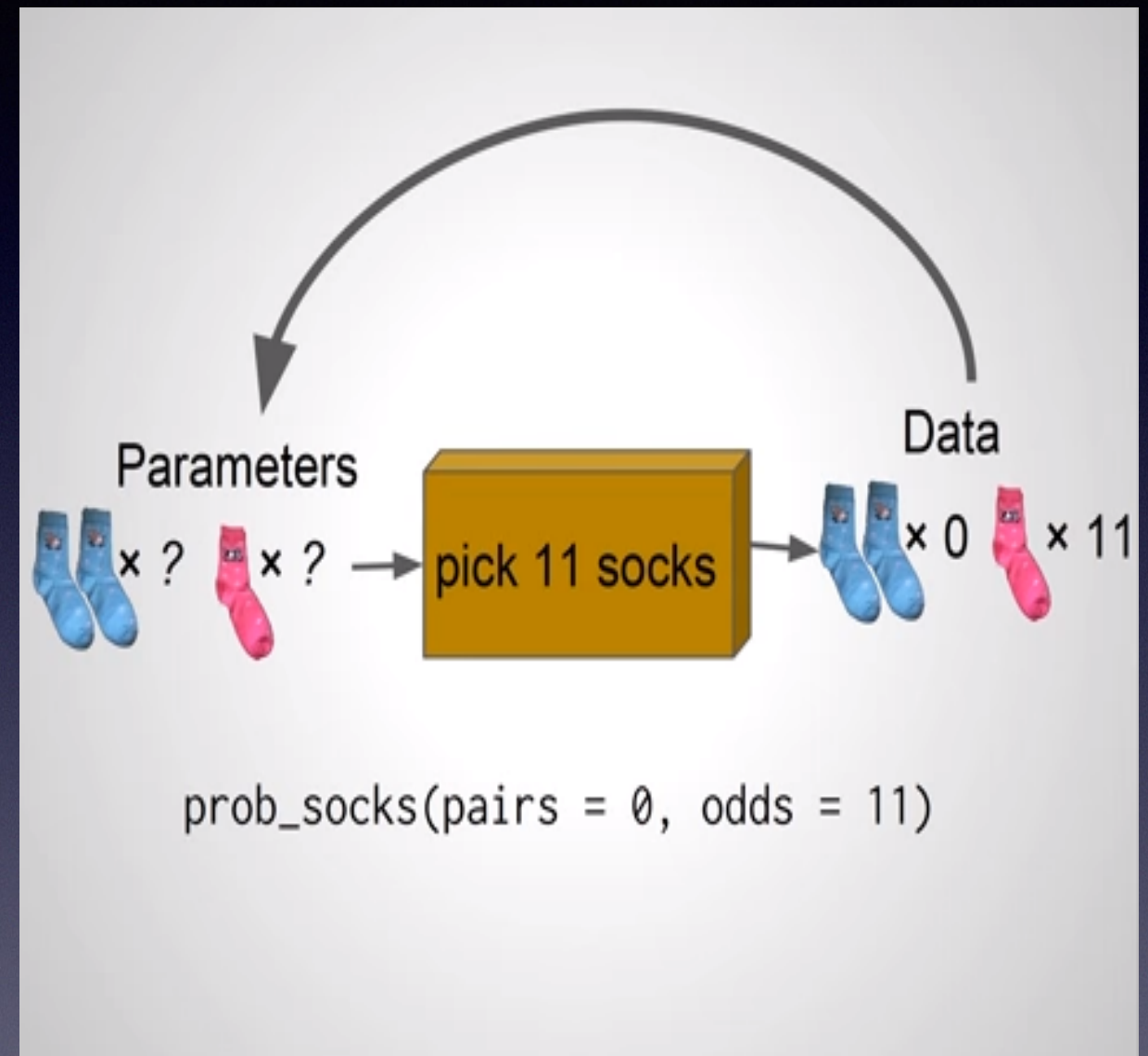
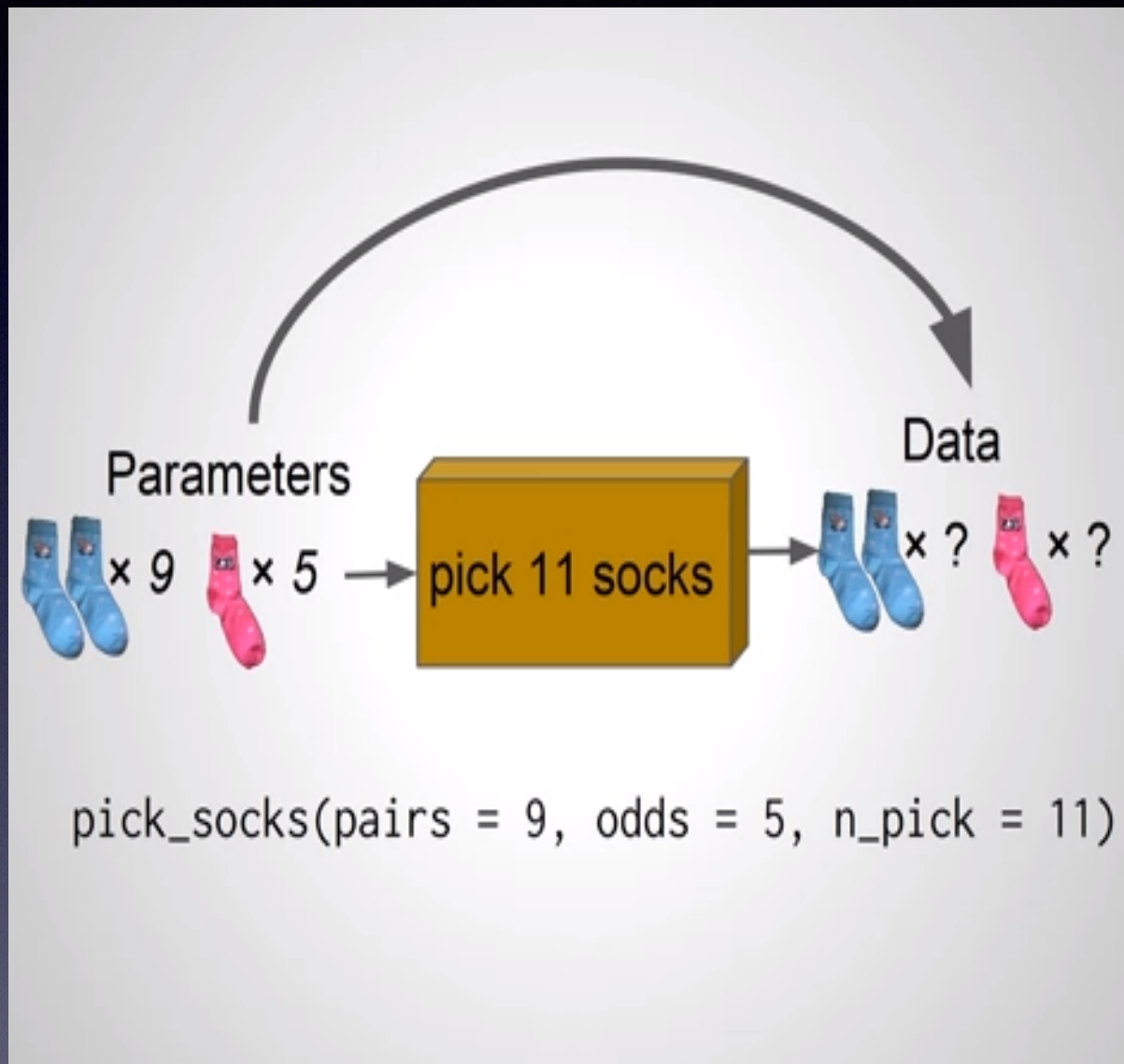
Question: Given the tiny dataset of eleven unique socks, how many socks does Karl Broman have in his laundry in total?



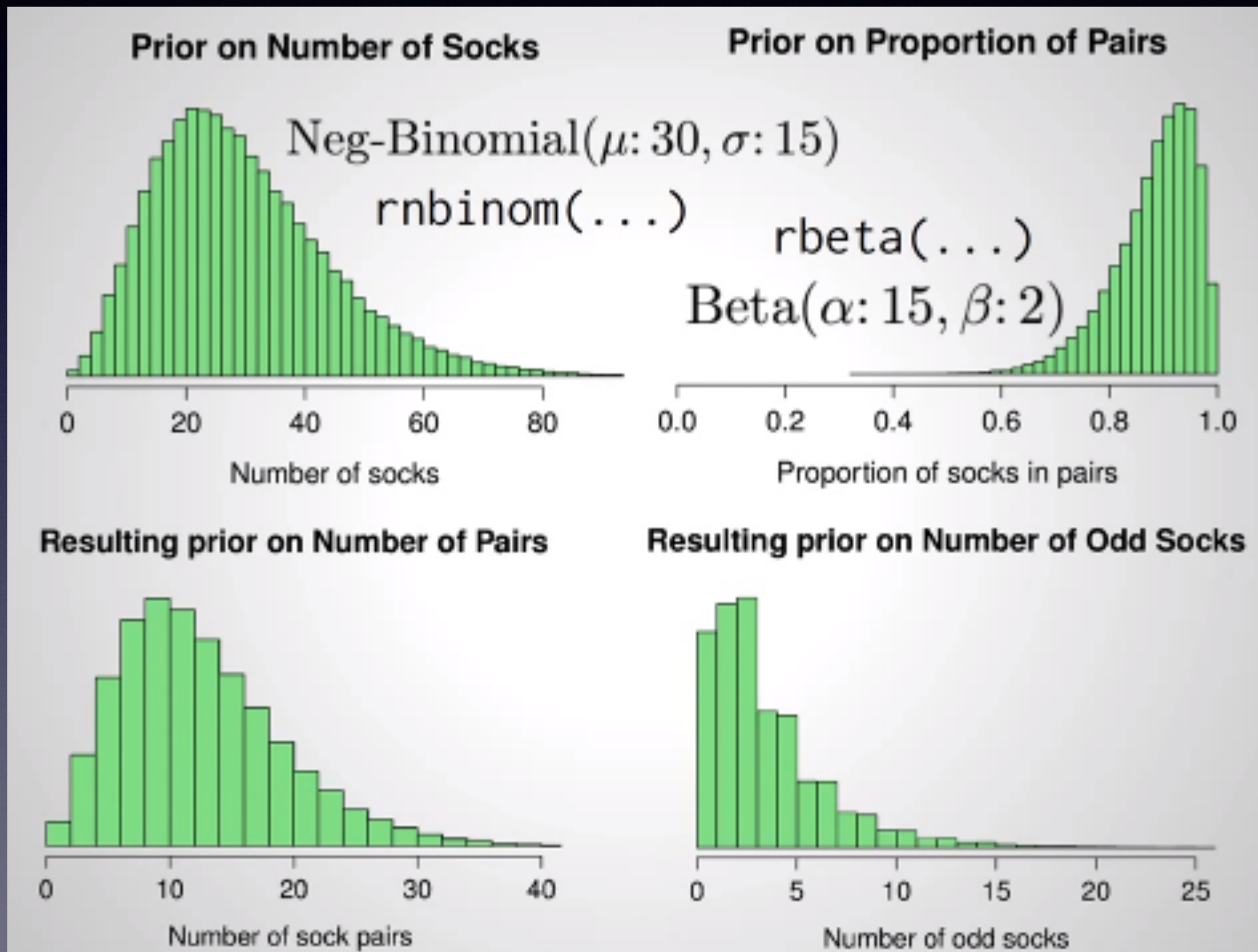
Lets assume we have some pairs of socks, say 9 and some odd socks, say 5.



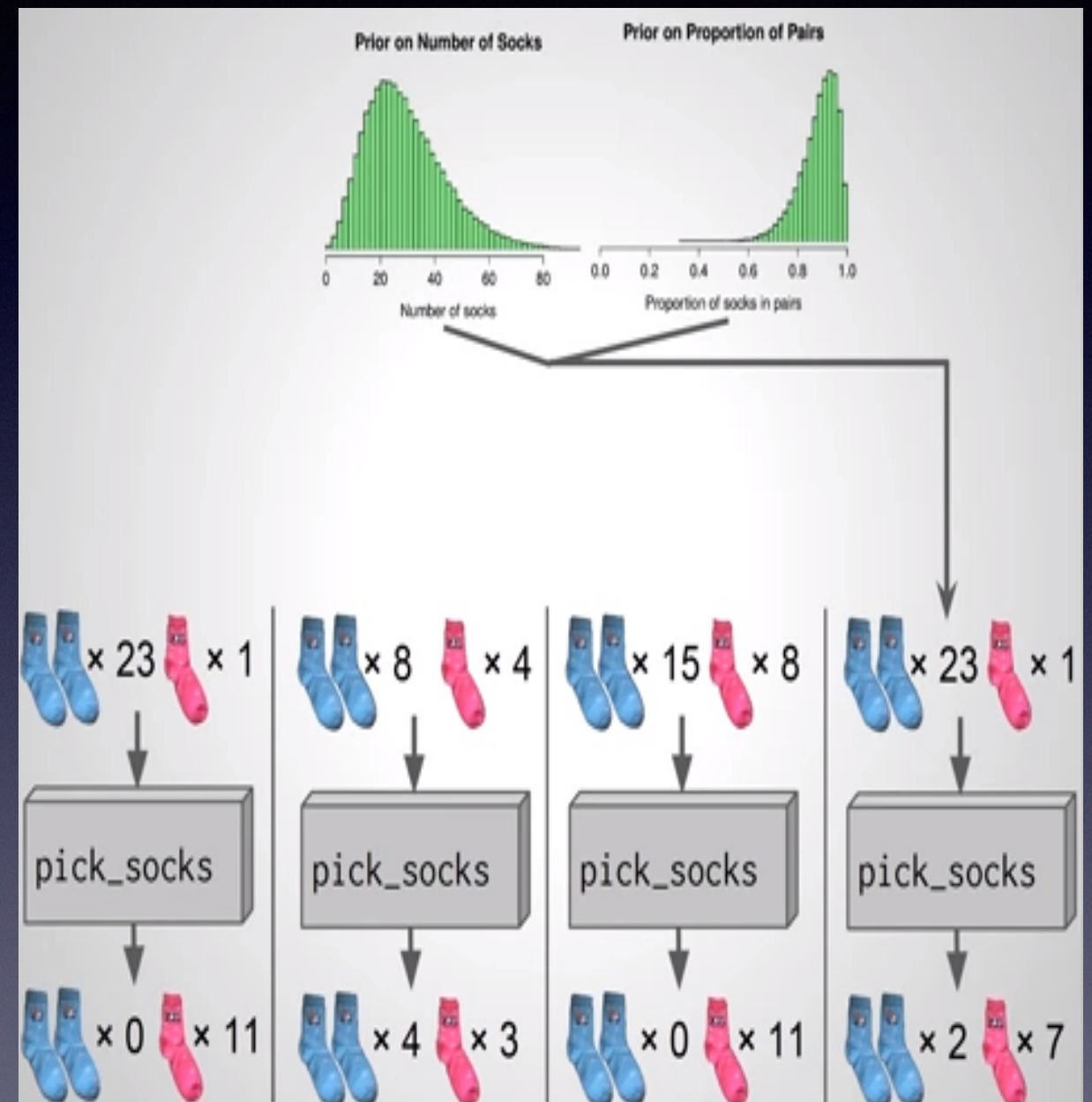
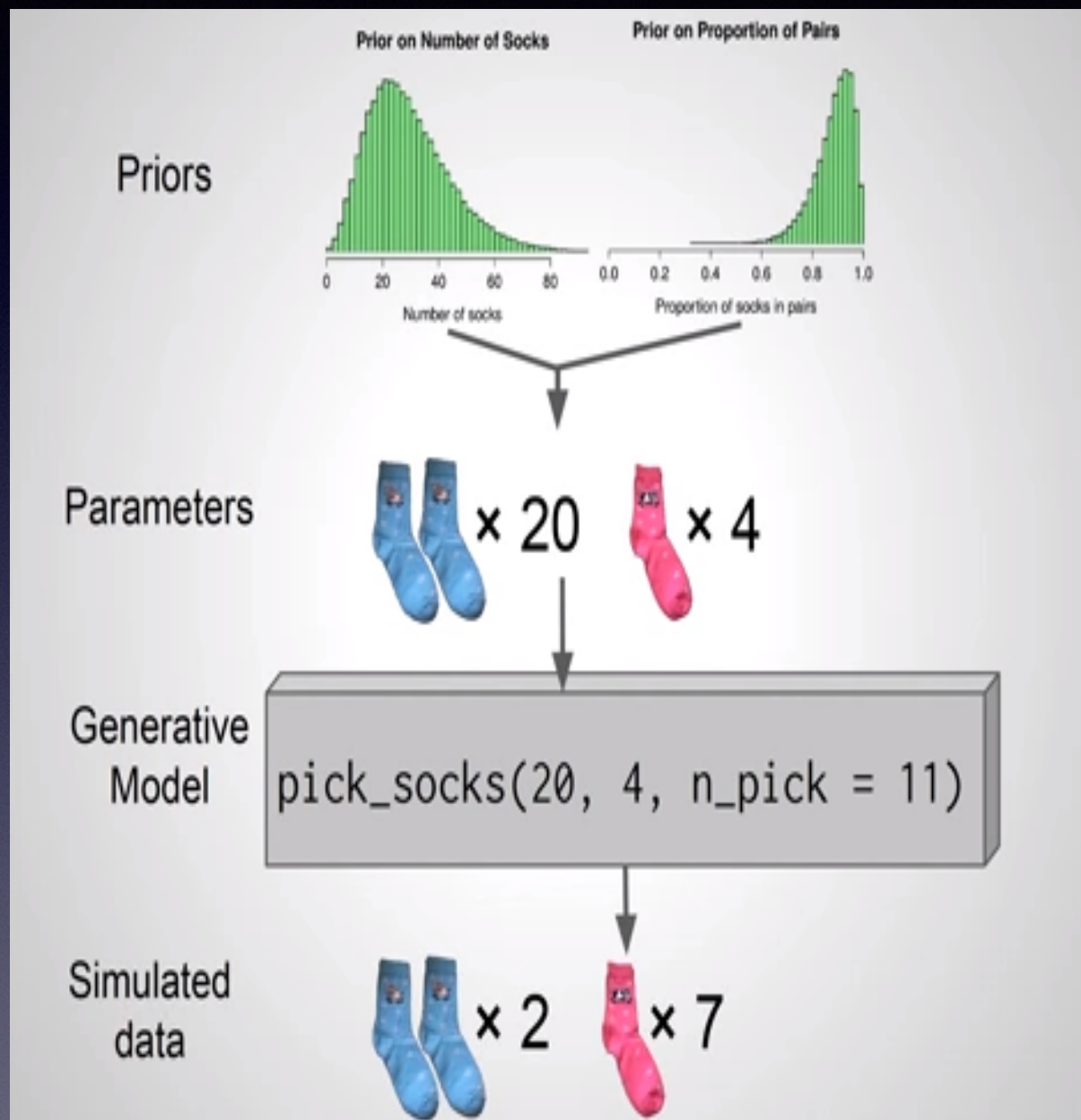
Criteria.



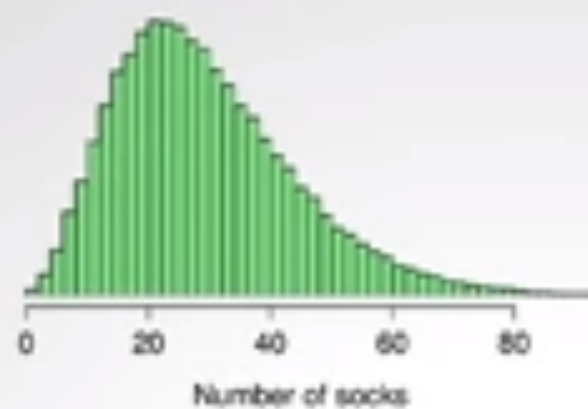
Choosing the priors.



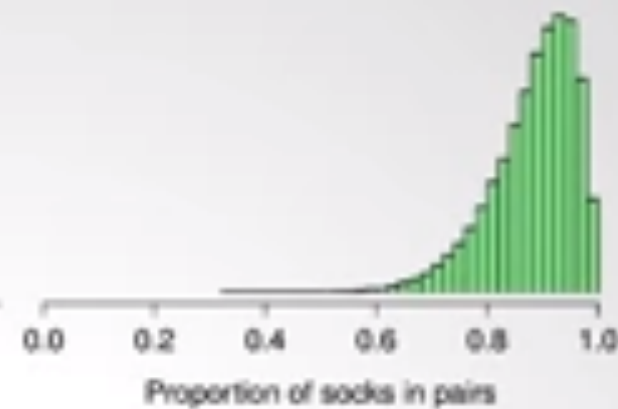
Simulation





Prior on Number of Socks



Prior on Proportion of Pairs







 $\times 23$  $\times 1$



pick_socks





 $\times 0$  $\times 11$



 $\times 8$  $\times 4$



pick_socks





 $\times 4$  $\times 3$


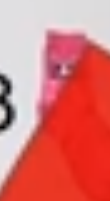
 $\times 15$  $\times 8$



pick_socks



 $\times 0$  $\times 11$

 $\times 23$  $\times 1$



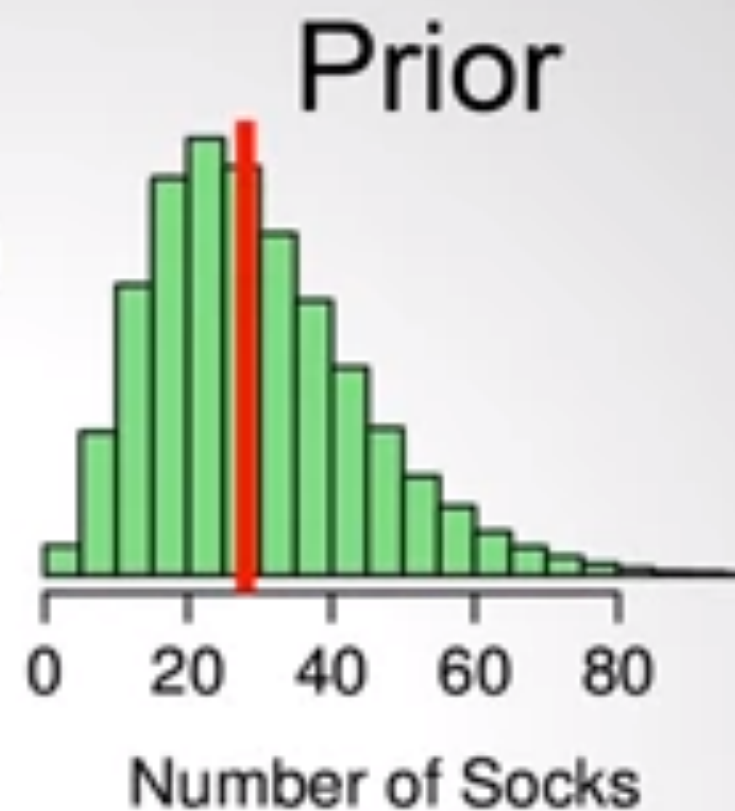
pick_socks



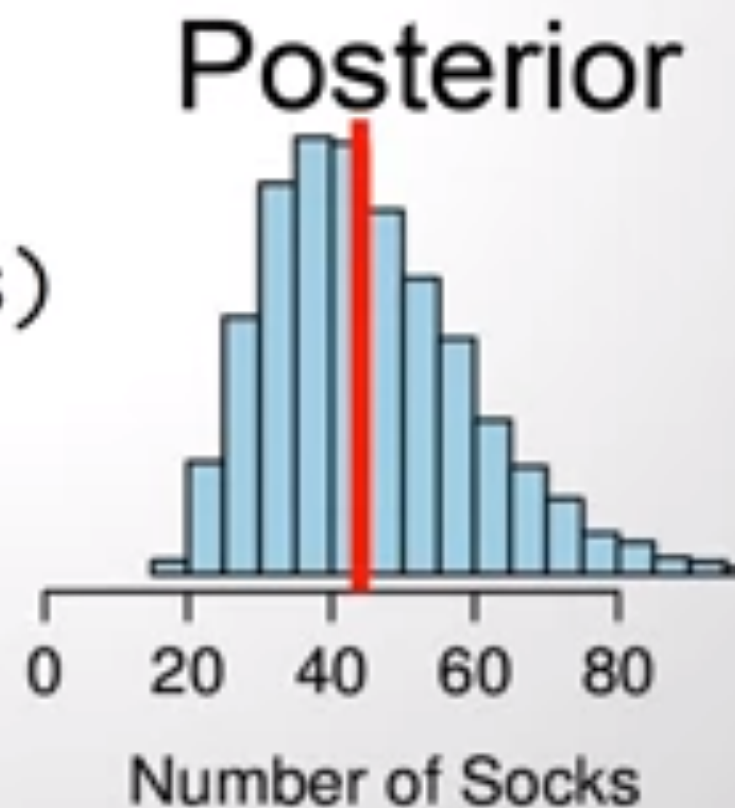
 $\times 2$  $\times 7$

Actual data:  $\times 0$  $\times 11$

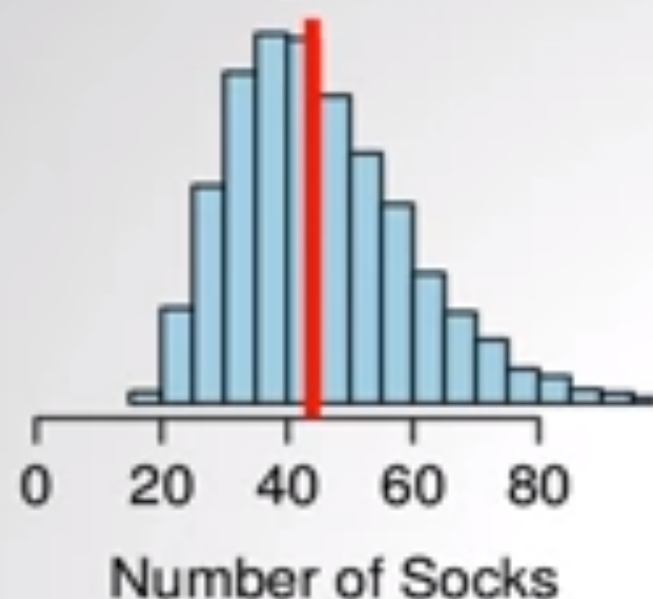

```
hist(sock_sim$n_socks)  
median(sock_sim$n_socks)
```



```
hist(posterior$n_socks)  
median(posterior$n_socks)
```



Our best guess: 44



Actual number of socks:
 $21 \times 2 + 3 = 45$

Error:



Karl Broman
@kwbroman



Following

[@rabaath](#) [@sgrifter](#) There were 21 pairs and 3 singletons. Will spend the rest of the evening working out what my est would have been.

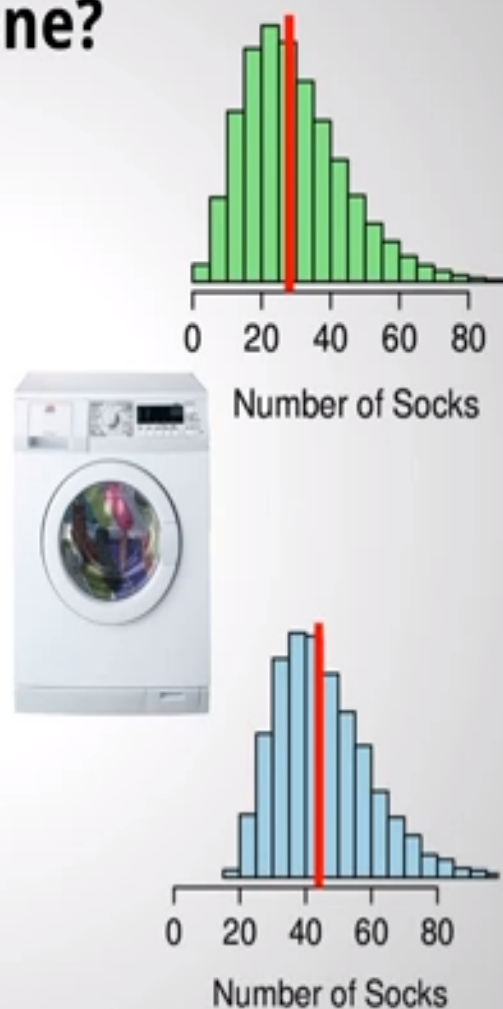


3:00 PM - 17 Oct 2014

What have we done?

So, what have we done?

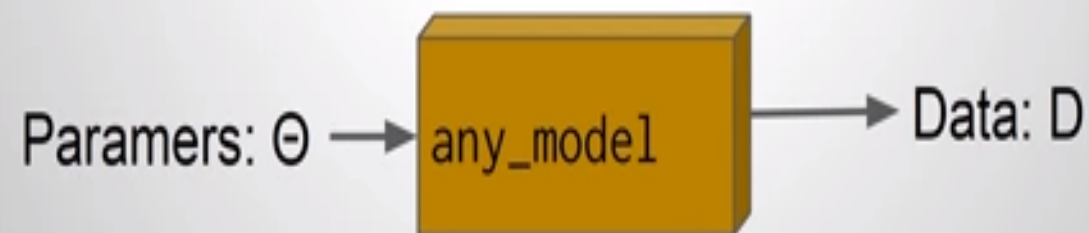
- We have specified prior information
- A generative model
- And got out the probability of different parameter values using ABC.



- The example we used was about socks in Karl Broman's washing machine.

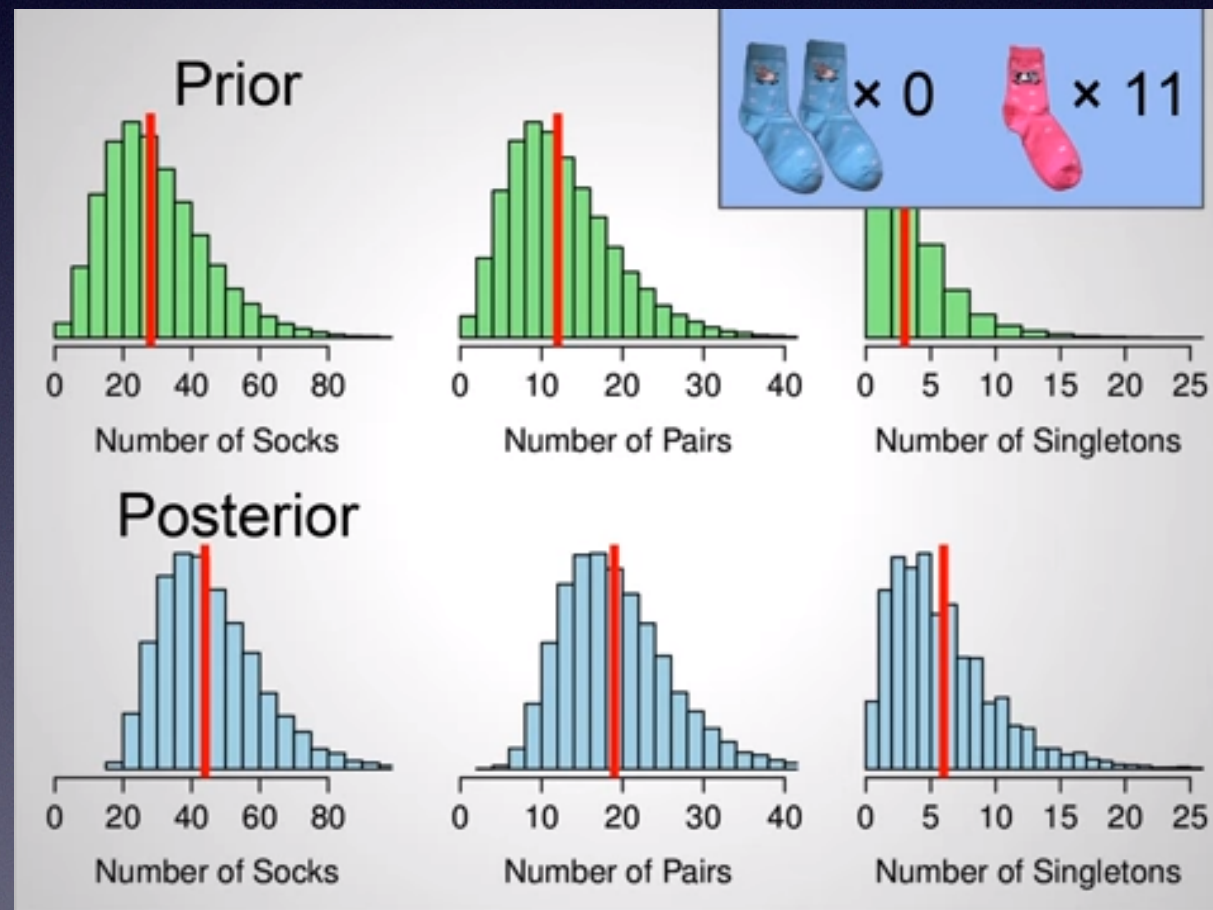


- But the general method works on *any* generative model.

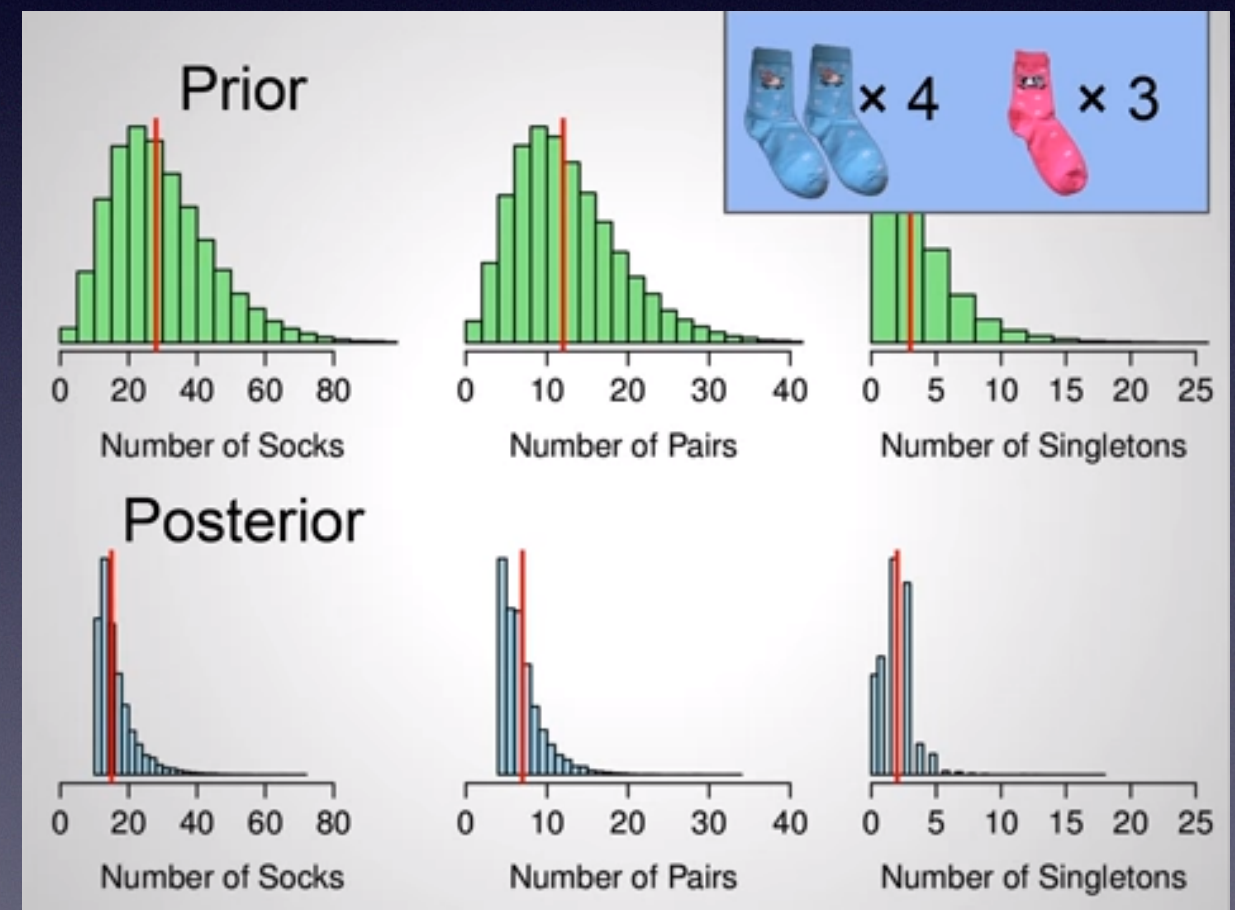


Do we need a prior?

USING PRIOR.



WITHOUT USING PRIOR.



THE EXPONENTIAL MODEL

The exponential distribution $\text{Exp}(\lambda)$ has the probability density function

$$f(y|\lambda) = \begin{cases} 0 & \text{if } y < 0 \\ \lambda \exp(-\lambda y) & \text{if } y \geq 0, \end{cases}$$

The gamma distribution $\Gamma(\alpha, \beta)$ has probability density function

$$f(y|\alpha, \beta) = \begin{cases} 0 & \text{if } y < 0 \\ \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-y\beta) & \text{if } y \geq 0, \end{cases}$$

where the hyperparameters $\alpha > 0$ and $\beta > 0$ are usually called the shape and rate parameters, respectively. The exponential distribution is a special case of the gamma distribution with $\alpha = 1$ and $\beta = \lambda$. The gamma distribution is a conjugate prior for exponential likelihood, so for observed data $Y = \{Y_1, Y_2, \dots, Y_n\}$, and a gamma prior with $\alpha = \alpha_0$ and $\beta = \beta_0$, the posterior distribution of λ is

$$\lambda | \alpha, \beta, Y \sim \Gamma \left(\alpha = \alpha_0 + n, \beta = \beta_0 + \sum_{i=1}^n Y_i \right).$$

We will use this posterior to evaluate the accuracy of the ABC PMC algorithm. The values of the hyperparameters α_0 and β_0 were fixed at 0.1.

Estimating the posterior using ABC PMC

- We face two problems at this point.
 1. Y is continuous.
 2. The proposal distribution may be very far from the posterior.

How can we solve them?

Solution for the problems.

1. Rather than taking the exact match, we can take some amount of tolerance threshold, ϵ_t .
2. Gradually reduce the value of ϵ_t , so that we “move” efficiently from the prior(proposal) to the desired posterior.

The distance function.

$$\rho_1(X, Y) = \frac{1}{n} \left| \sum_{i=1}^n X_i - \sum_{i=1}^n Y_i \right| = |\bar{X} - \bar{Y}|.$$

We also examined

$$\rho_2(X, Y) = |\text{median}(X) - \text{median}(Y)|$$

and

$$\begin{aligned} \rho_3(X, Y) &= |[F^{-1}(0.75, X) - F^{-1}(0.25, X)] \\ &\quad - [F^{-1}(0.75, Y) - F^{-1}(0.25, Y)]| \\ &= |\text{IQR}(X) - \text{IQR}(Y)|, \end{aligned}$$

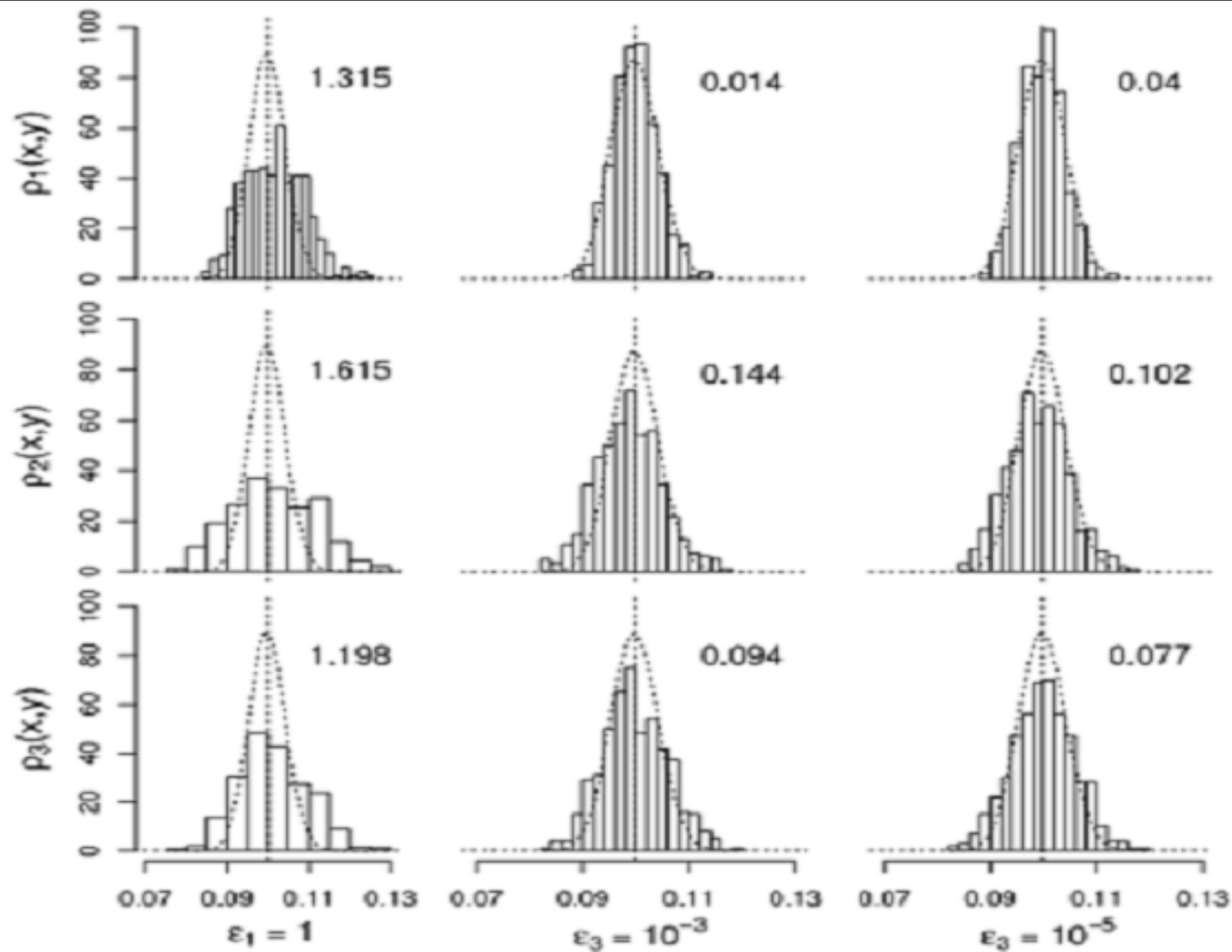
Tolerance

- To generate the data, we took $n = 500$ samples from $\text{Exp}(\lambda = 0.1)$, so the observation ranged from 0 to 70, with mean = 10, sd = 10 and IQR approximately 11.
- Decreasing set of tolerance is

$$\epsilon = \{3, 1, 10^{-1}, 10^{-3}, 10^{-4}, 10^{-5}\}$$

Ultimately, only the smallest element of ϵ matters.

Results



Interpretation.

- The dashed curves shows the true posterior and the histogram show the estimated posterior obtained using ABC PMC.
- As ϵ_t decreases, the accuracy of the posterior increases.
- If further reduction in ϵ_t do not produce changes, then the estimate has converged to its final target.

Computation time.

Table 1

Computation times for the exponential example (in minutes).

Iteration	Tolerance ϵ	Mean $\rho_1(X, Y)$	Median $\rho_2(X, Y)$	IQR $\rho_3(X, Y)$
1	3	0.07	2.74	0.12
2	1	0.07	1.18	0.08
3	10^{-1}	0.13	3.19	0.68
4	10^{-3}	0.42	2.96	4.84
5	10^{-4}	3.72	16.21	45.76
6	10^{-5}	34.40	143.40	454.71
Total		38.81	169.67	506.18

Algorithm of ABC PMC

```
1: Given data  $Y$  and model  $Y \sim \text{Model}(\theta)$ , a set of tolerance thresholds  $\epsilon$ , and  
   prior distribution  $\pi(\theta)$  :  
2: At iteration  $t = 1$ ,  
3: for  $1 \leq i \leq N$  do  
4:   while  $\rho(X, Y) > \epsilon_1$  do  
5:     Sample  $\theta^*$  from the prior:  $\theta^* \sim \pi(\theta)$   
6:     Generate data  $X$  from  $\theta^*$ :  $X \sim \text{Model}(\theta^*)$   
7:     Calculate discrepancy  $\rho(X, Y)$   
8:   end while  
9:   Set  $\theta_{i,1} \leftarrow \theta^*$   
10:  Set  $w_{i,1} \leftarrow \frac{1}{N}$   
11: end for  
12: Set  $\sigma_1^2 \leftarrow 2\text{Var}(\theta_{1:N,1})$   
13: At iteration  $t > 1$ ,  
14: for  $2 \leq t \leq T$  do  
15:   for  $1 \leq i \leq N$  do  
16:     while  $\rho(X, Y) > \epsilon_t$  do  
17:       Sample  $\theta^*$  from the previous iteration:  $\theta^* \sim \theta_{1:N,t-1}$  with probabilities  $w_{1:N,t-1}$   
18:       Perturb  $\theta^*$  by sampling  $\theta^{**} \sim N(\theta^*, \sigma_{t-1}^2)$   
19:       Generate data  $X$  from  $\theta^{**}$ :  $X \sim \text{Model}(\theta^{**})$   
20:       Calculate discrepancy  $\rho(X, Y)$   
21:     end while  
22:     Set  $\theta_{i,t} \leftarrow \theta^{**}$   
23:     Set  $w_{i,t} \leftarrow \frac{\pi(\theta_{i,t})}{\sum_{j=1}^N w_{j,t-1} q_f(\theta_{j,t-1} | \theta_{i,t}, \sigma_{t-1})}$   
24:   end for  
25:   Set  $\sigma_t^2 \leftarrow 2\text{Var}(\theta_{1:N,t})$   
26: end for
```


Reference

Brandon, M. Turner and Trisha Van Zandt (2012). A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*. 56, 69-85

THE END