

BAYESIAN INFERENCE

2 NOVEMBER 2018

LIKELIHOOD

Compute the probability of observing this dataset given a proposed set of parameters.

```
y <- c(4, 0, 1)
likelihood <- function (lambda) {
  probs <- dpois(y, lambda)
  prod(probs)
}
```

```
likelihood(1)
```

```
## [1] 0.002074461
```

```
likelihood(0.5)
```

```
## [1] 0.0002905341
```

MAXIMUM LIKELIHOOD

Keep trying new parameters until you find the most likely set

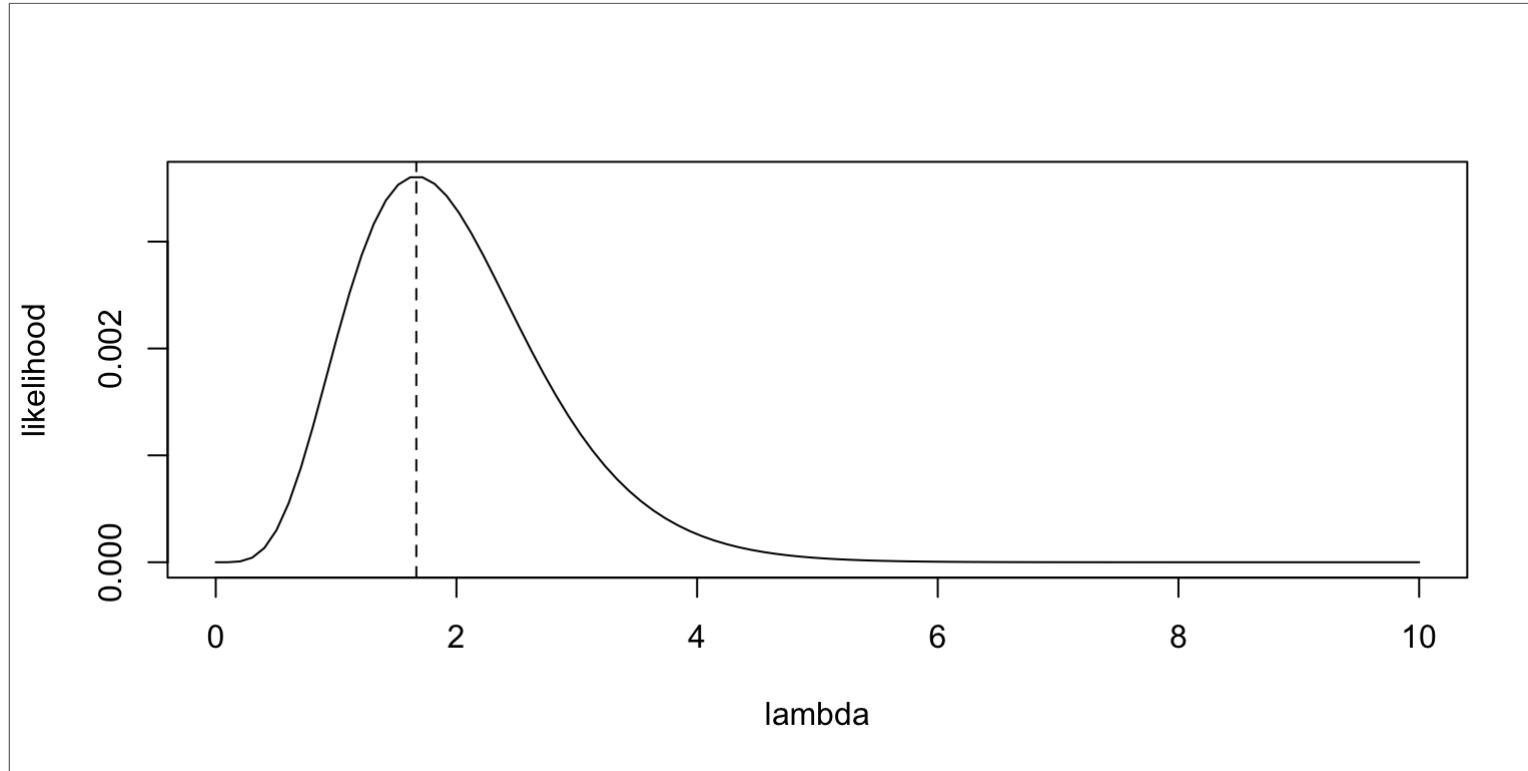
```
objective <- function(lambda) {  
  -1 * likelihood(lambda)  
}  
optim(c(lambda = 0), objective)$par
```

```
##   lambda  
## 1.666667
```

LIKELIHOOD SURFACE

Compute the likelihood at a range of possible values

```
lambda <- seq(0, 10, length.out = 100)
likelihood <- sapply(lambda, likelihood)
```



LIKELIHOOD-BASED INFERENCE

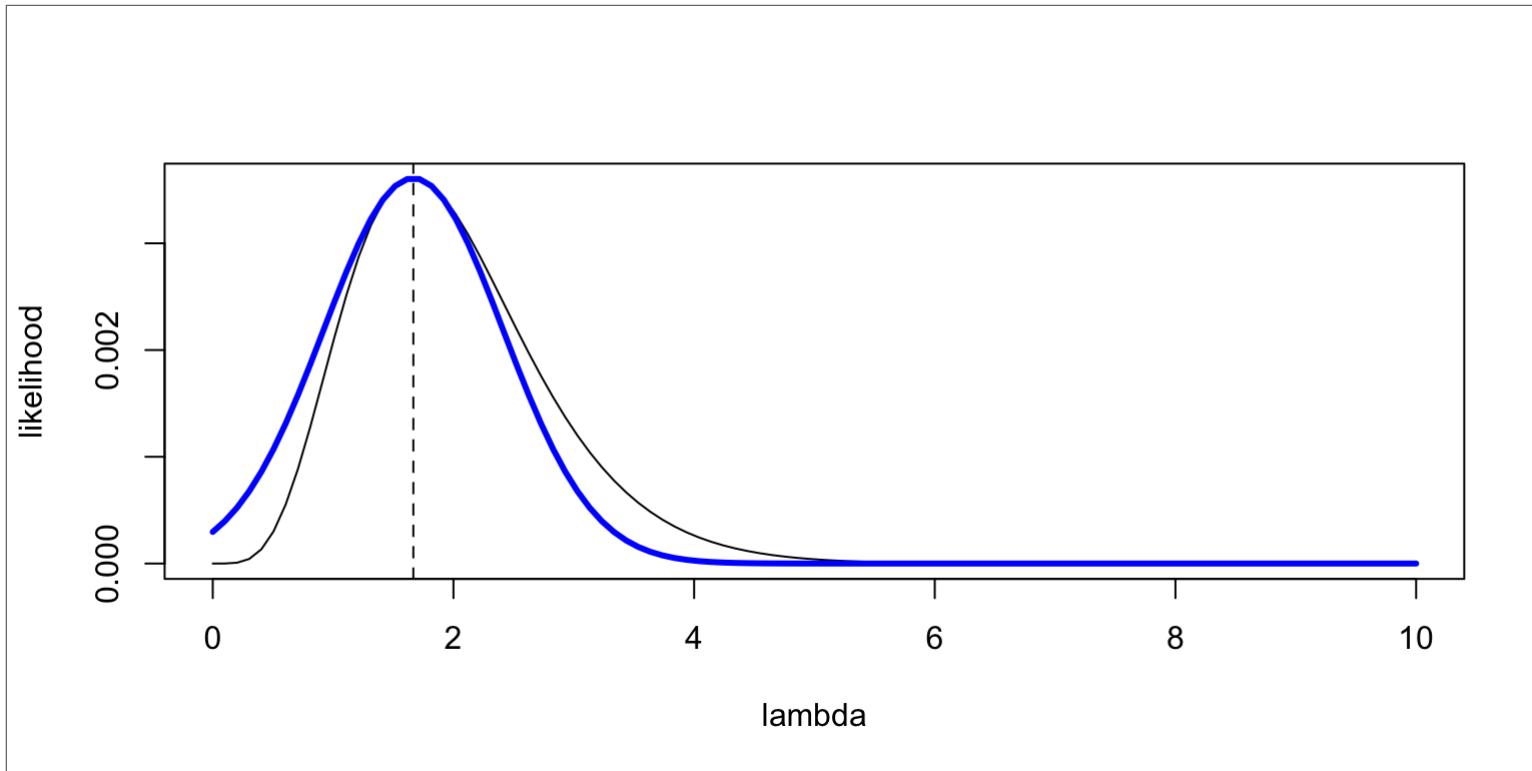
Considers the probability distribution over *data*, given parameters, but then treats this surface *kind-of* like a probability distribution over parameters

LIKELIHOOD-BASED INFERENCE

- **estimates & standard errors**
- confidence intervals
- p-values

these assume the likelihood surface is normally distributed (it isn't)

LIKELIHOOD-BASED INFERENCE



LIKELIHOOD-BASED INFERENCE

- estimates & standard errors
- **confidence intervals**
- p-values

“Were this procedure to be repeated on numerous samples, the fraction of calculated confidence intervals (which would differ for each sample) that encompass the true population parameter would tend toward 90%.”

LIKELIHOOD-BASED INFERENCE

- estimates & standard errors
- confidence intervals
- **p-values**

- “the probability that the observed effects were produced by random chance alone”
- “the probability under the null hypothesis of obtaining a result equal to or more extreme than what was actually observed”
- “the probability that the null hypothesis is true, or the probability that the alternative hypothesis is false”

BAYES THEOREM

Considers the probability distribution over **parameters**, given
parameters **data**

$$p(\text{parameters}|\text{data}) = \frac{p(\text{data}|\text{parameters}) \times p(\text{parameters})}{p(\text{data})}$$

BAYES THEOREM

Considers the probability distribution over ~~data parameters~~, given
~~parameters data~~

$$p(\text{parameters}|\text{data}) = \frac{p(\text{data}|\text{parameters}) \times p(\text{parameters})}{\textcircled{?}}$$

THE MOST COLOSSAL CONFLICT THE SCREEN HAS EVER KNOWN

JOHN BECK
presents

KING KONG vs. GODZILLA

PRINT BY
TECHNICOLOR

SCOPE







LIKELIHOOD VS PRIOR

[click me!](#)

PRIOR ELICITATION

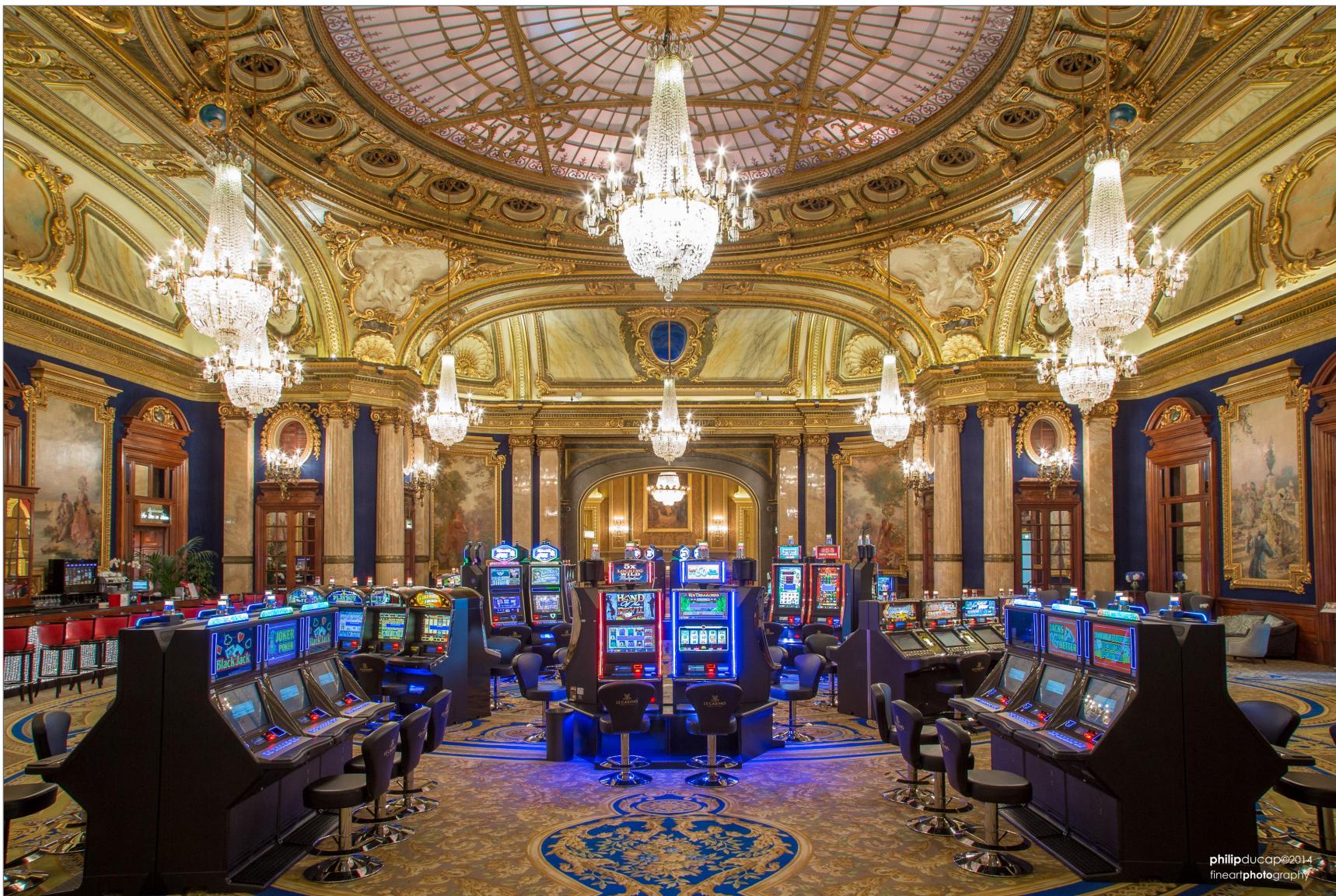
go to the whiteboard!

CALCULATING THE POSTERIOR

$$p(\text{parameters}|\text{data}) = \frac{p(\text{data}|\text{parameters}) \times p(\text{parameters})}{\textcircled{\text{S}}}$$

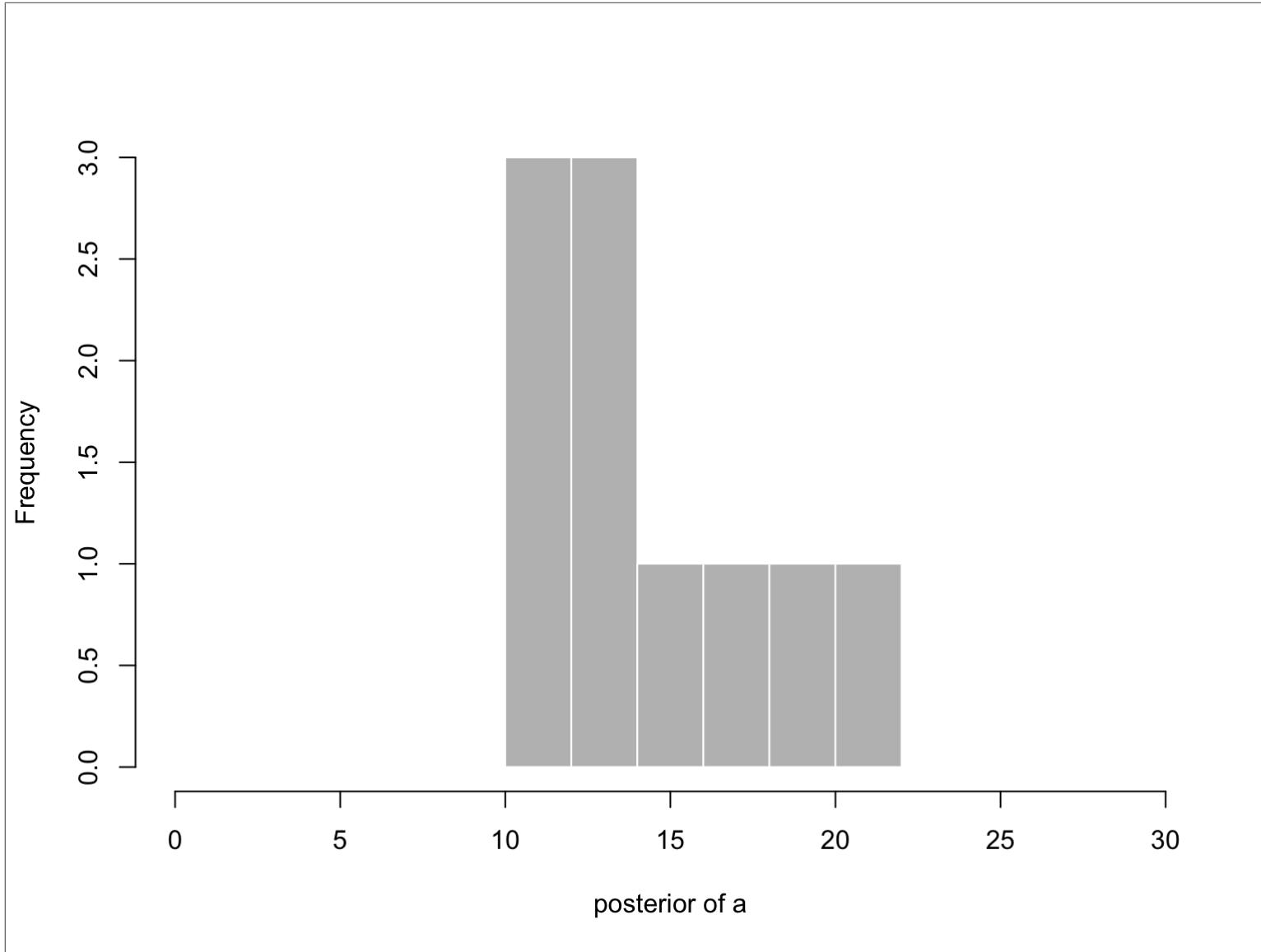
Because of the $\textcircled{\text{S}}$ it's bit tricky estimate the parameters of the posterior

Instead, we can draw random samples from the posterior. With enough samples, we can estimate the parameters

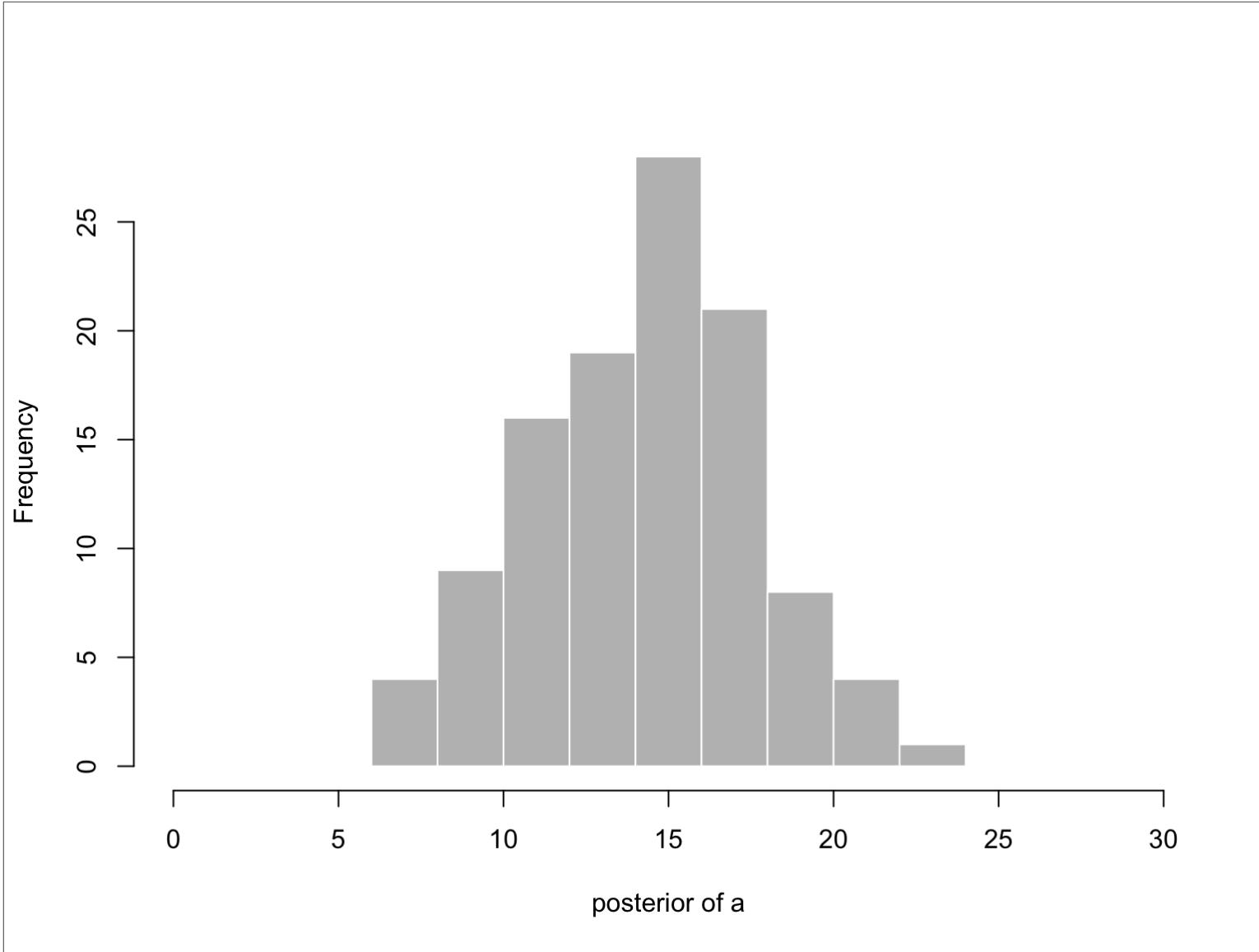


philipducamp©2014
fineartphotography

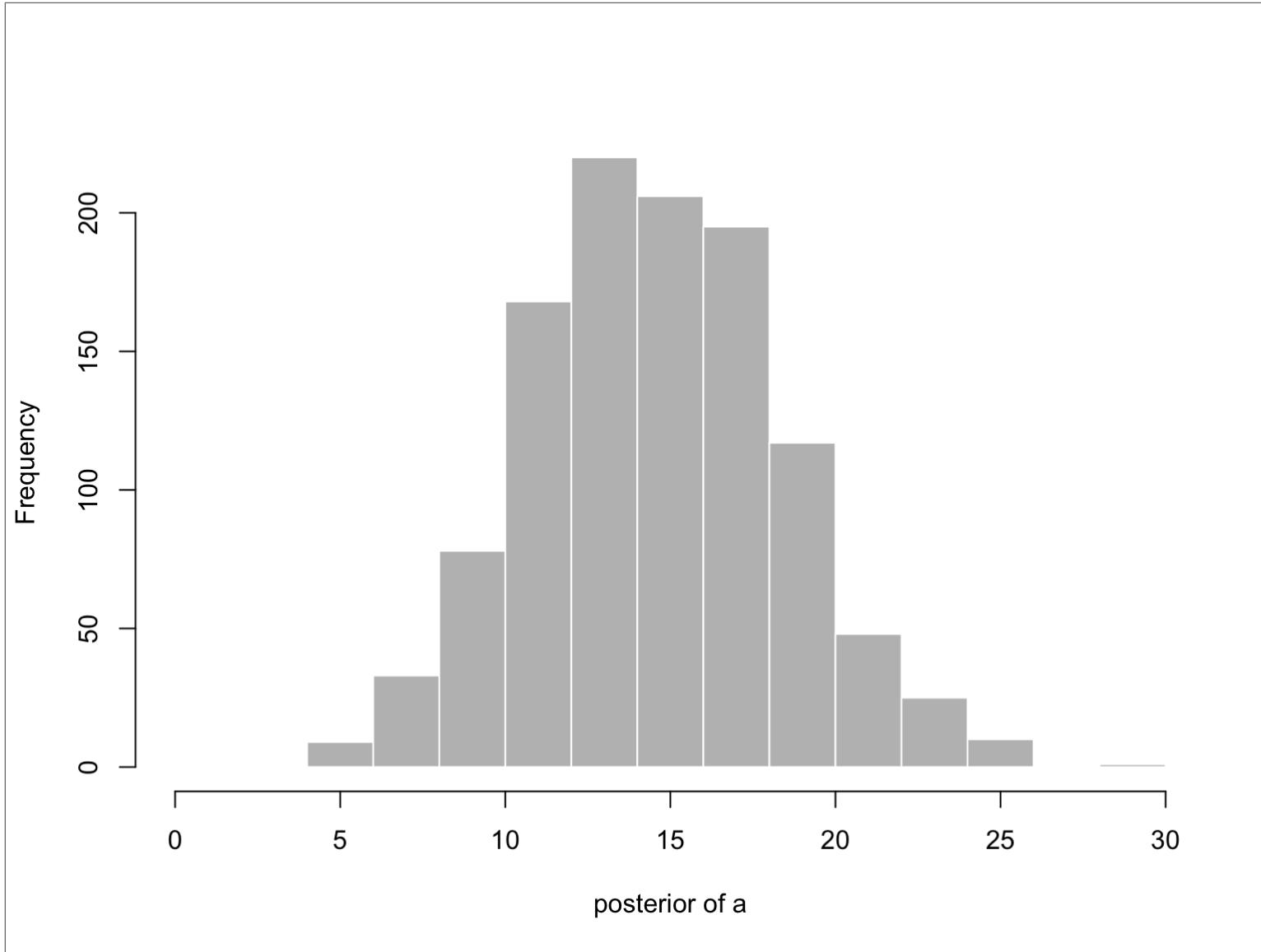
POSTERIOR SAMPLES



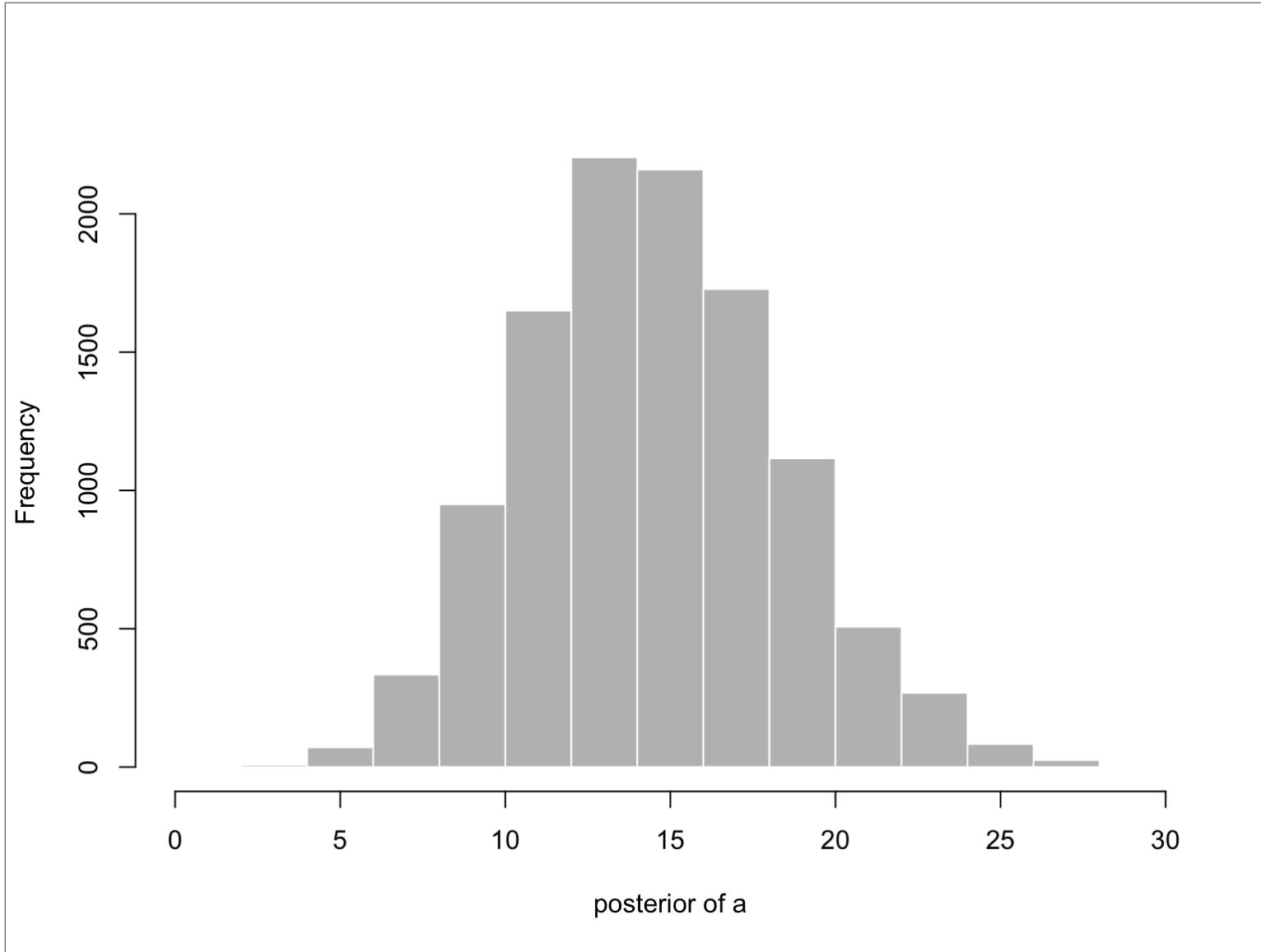
POSTERIOR SAMPLES



POSTERIOR SAMPLES



POSTERIOR SAMPLES



MCMC

Unfortunately, we don't have a random number generator for every model.

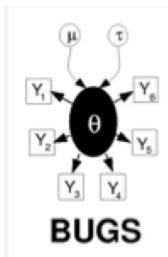
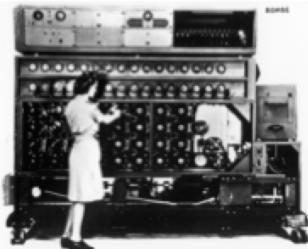
Markov chain Monte Carlo gives us a time series of correlated random numbers from our distribution

click me!

MCMC SOFTWARE



1953



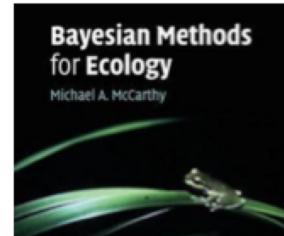
1997



JAGS



2007



2012

MODEL AVERAGING

One approach to dealing with multiple candidate models is to average them, based on how good they are

Bayesian inference does this automatically!

We get lots of different models, weighted according to how probable they are

WHEN TO DO BAYES?

- when we have prior information to use
- when we don't have much data
- when our model is too complicated for maximum likelihood
- when we really care about uncertainty

WHEN NOT TO DO BAYES?

- when we have vague priors,
- when we plenty of data
- when there's a maximum likelihood method that works
- when we don't care about uncertainty that much