The `sleepstudy` dataset

J. Sleep Res. (2003) **12,** 1–12

Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study

GREGORY BELENKY, NANCY J. WESENSTEN, DAVID R. THORNE, MARIA L. THOMAS, HELEN C. SING, DANIEL P. REDMOND, MICHAEL B. RUSSO and THOMAS J. BALKIN
Division of Neuropsychiatry, Walter Reed Army Institute of Research, Silver Spring, MD, USA

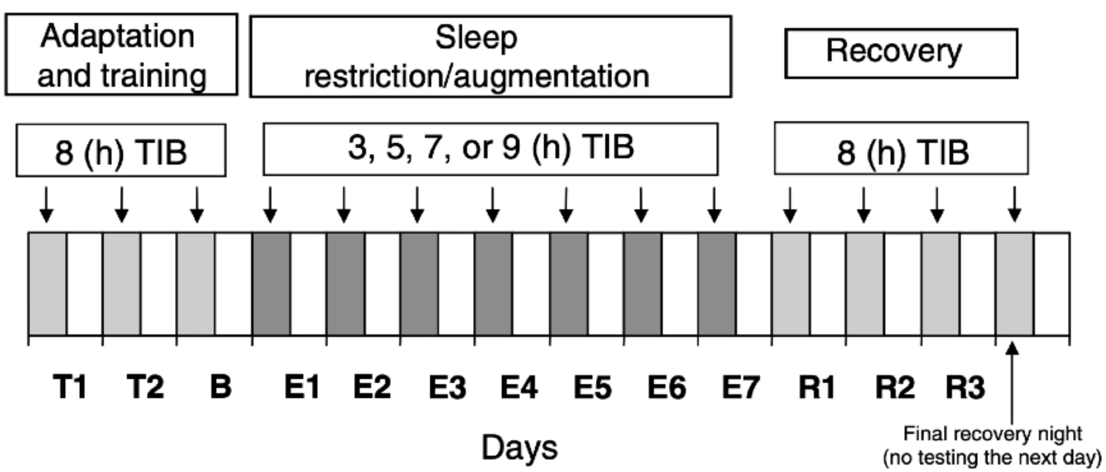


Figure 1. Study experimental design, showing nightly time in bed across days (adaptation/training, baseline, experimental phase, recovery phase).

The `sleepstudy` dataset

J. Sleep Res. (2003) **12,** 1–12

Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study

GREGORY BELENKY, NANCY J. WESENSTEN, DAVID R. THORNE, MARIA L. THOMAS, HELEN C. SING, DANIEL P. REDMOND, MICHAEL B. RUSSO and THOMAS J. BALKIN
Division of Neuropsychiatry, Walter Reed Army Institute of Research, Silver Spring, MD, USA

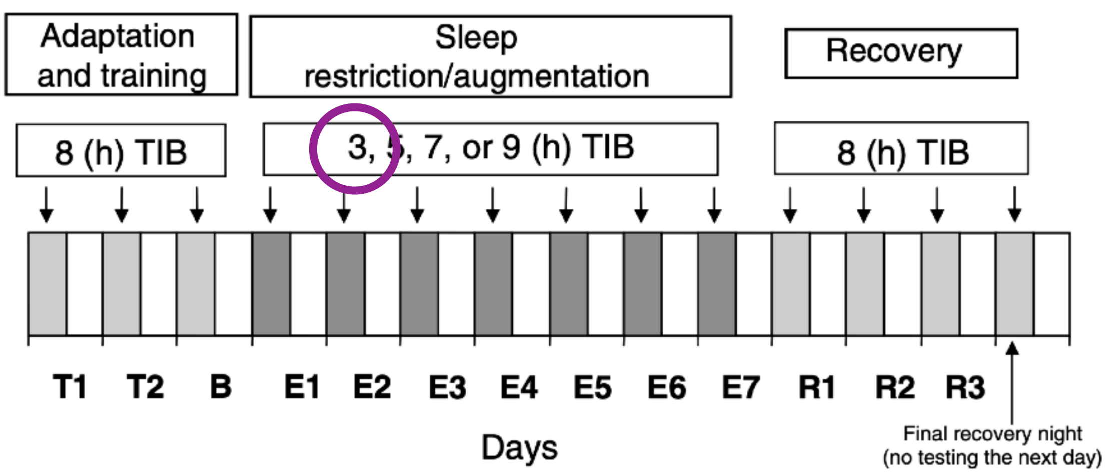
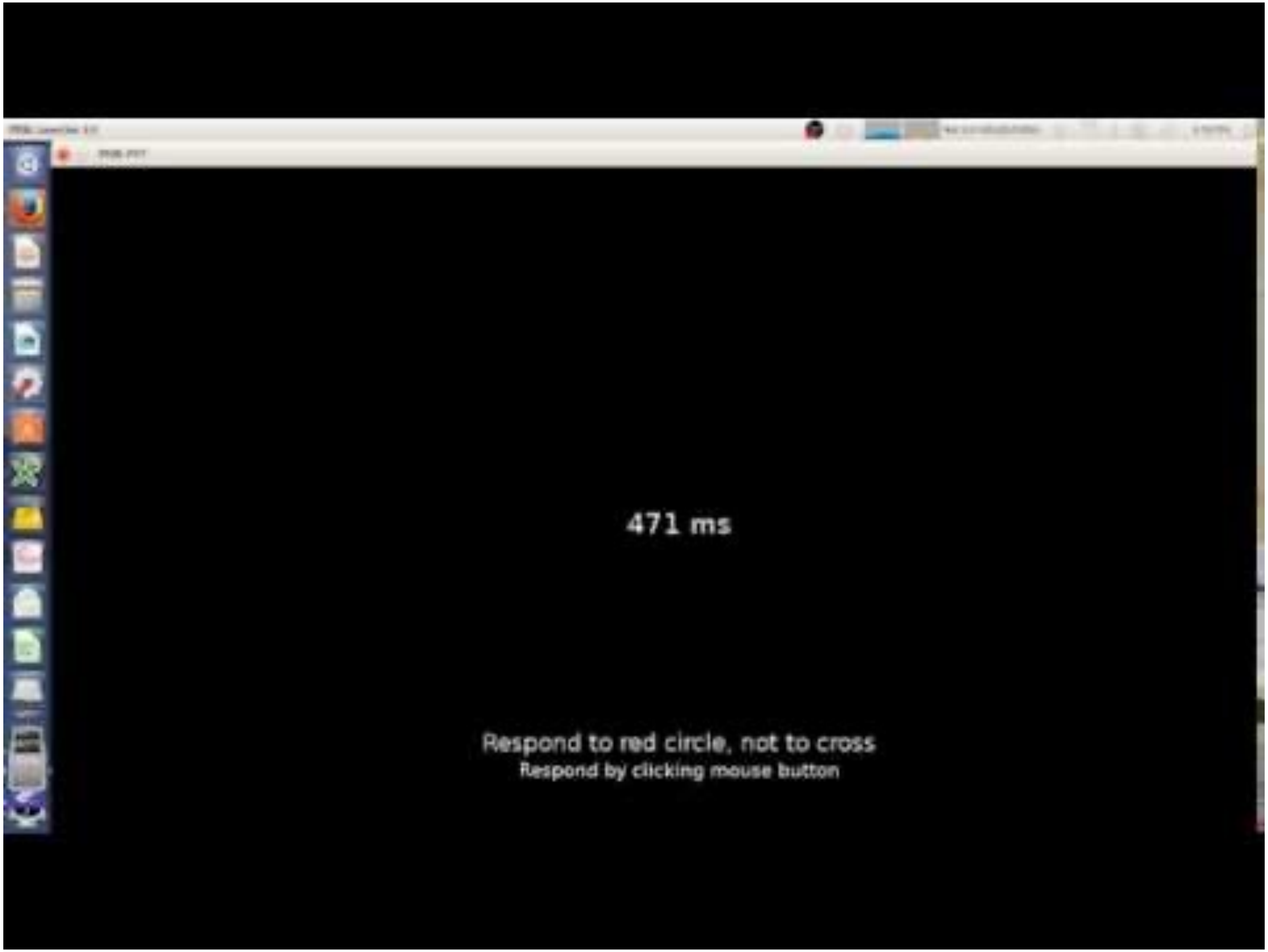


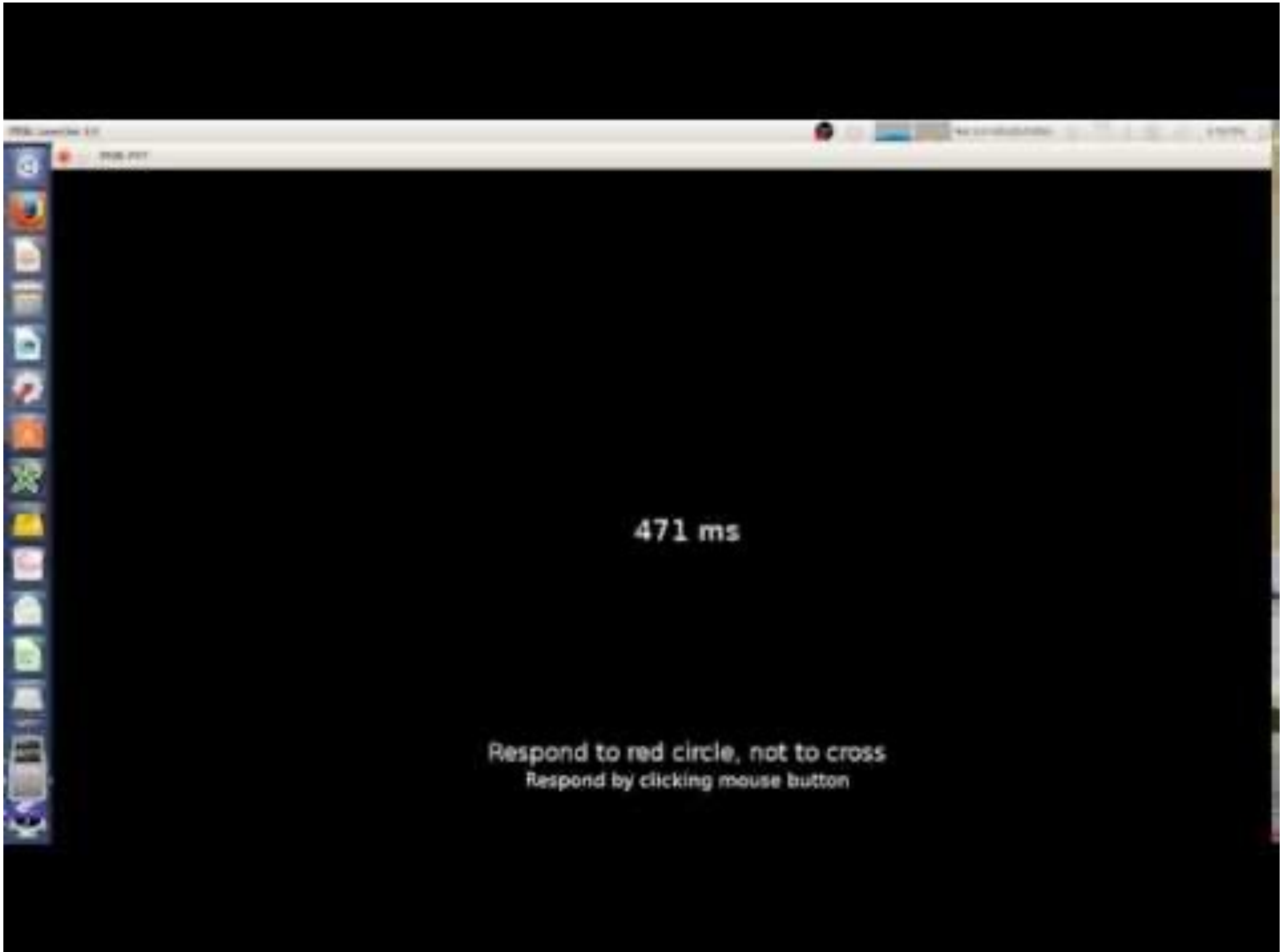
Figure 1. Study experimental design, showing nightly time in bed across days (adaptation/training, baseline, experimental phase, recovery phase).

Psychomotor vigilance test (PVT)



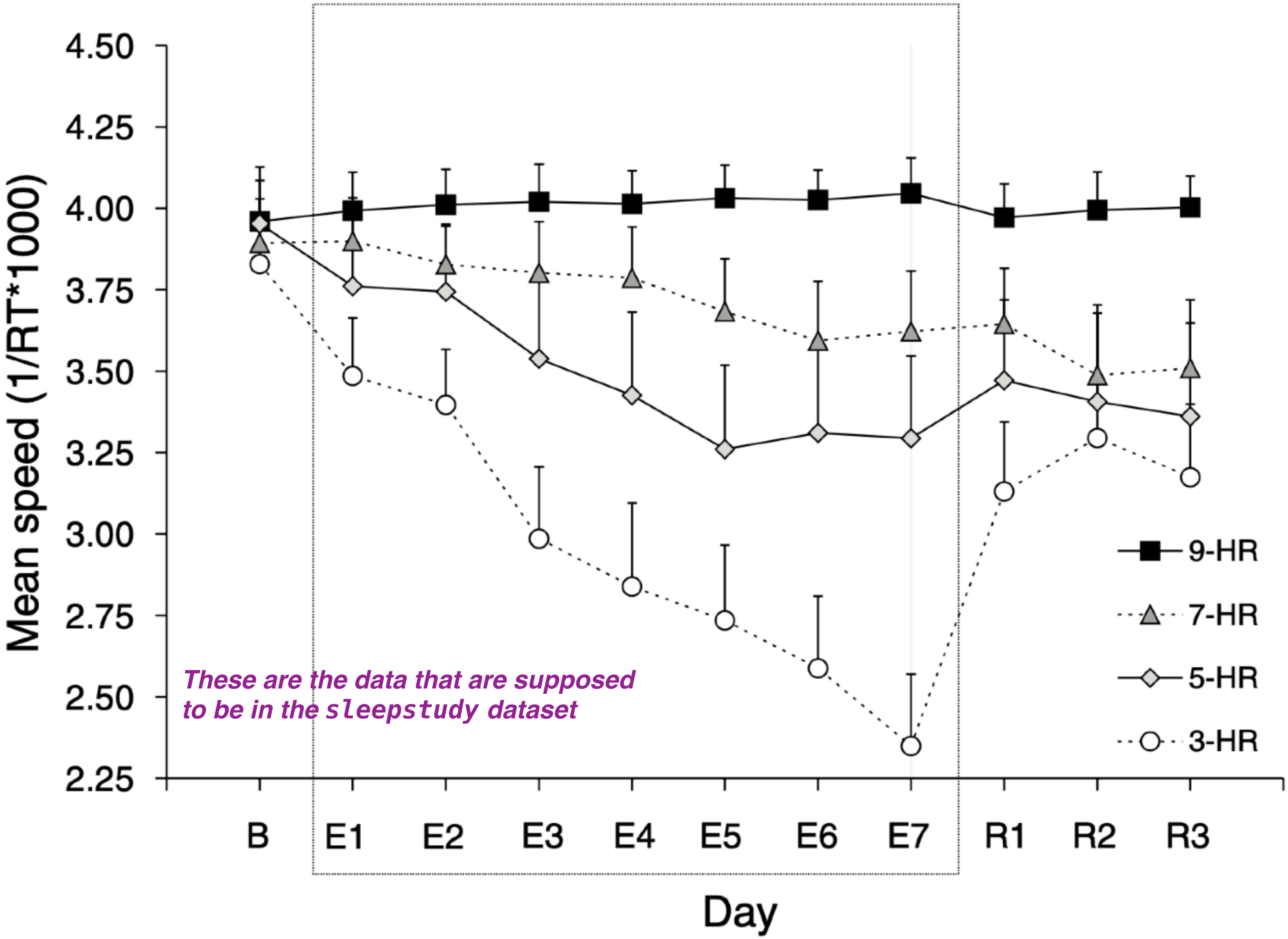
https://www.youtube.com/watch?v=eG4K4t6RweQ

Psychomotor vigilance test (PVT)



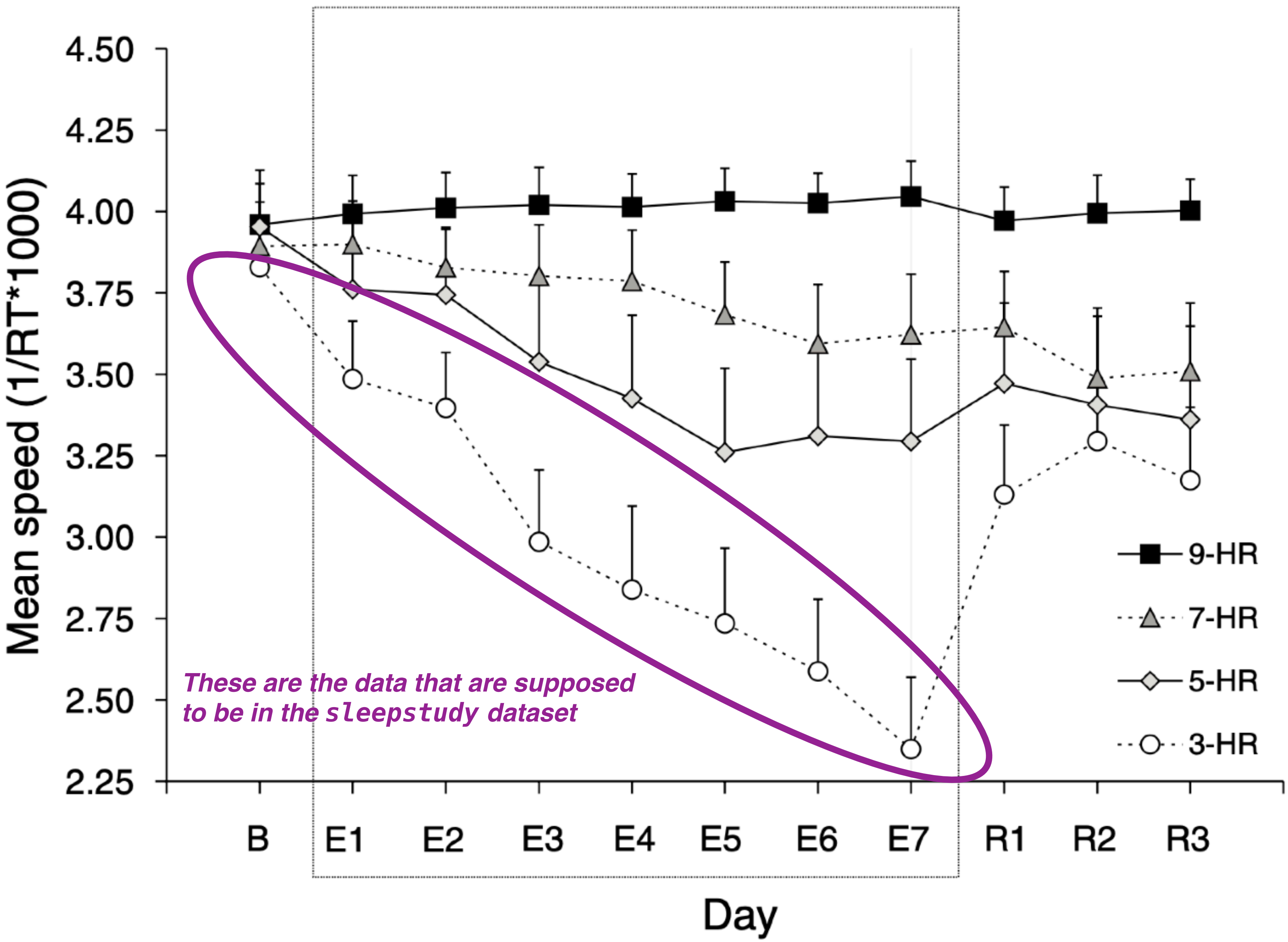
https://www.youtube.com/watch?v=eG4K4t6RweQ

Results



(error bars are standard error of the mean)

Results



(error bars are standard error of the mean)

Our scientific question

- What is the **distribution** of effects of sleep deprivation on vigilance (as measured by the PVT) across individuals?

Our scientific question

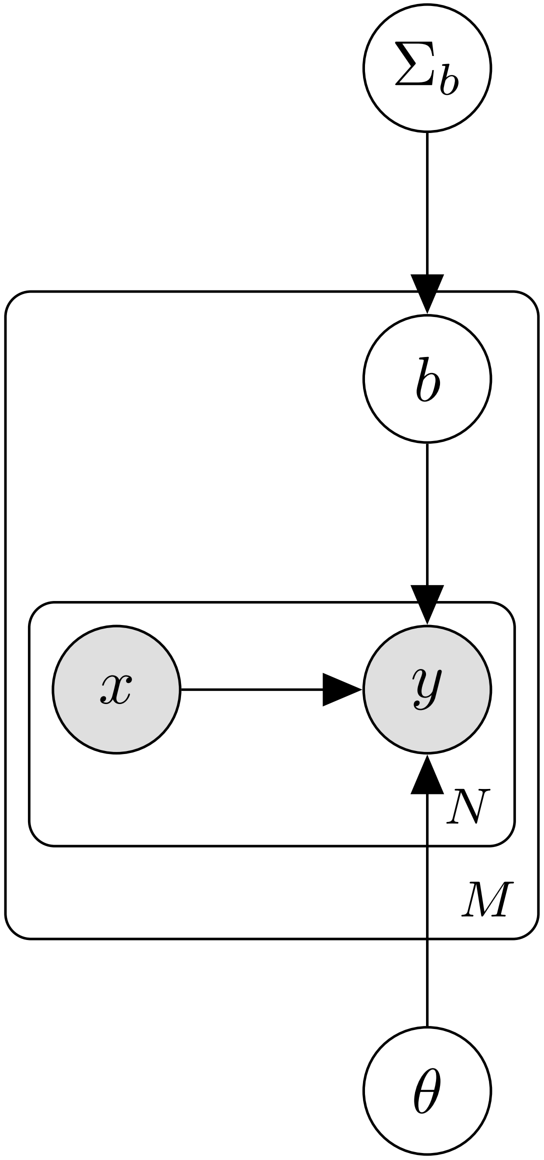
- What is the **distribution** of effects of sleep deprivation on vigilance (as measured by the PVT) across individuals?

METHODS

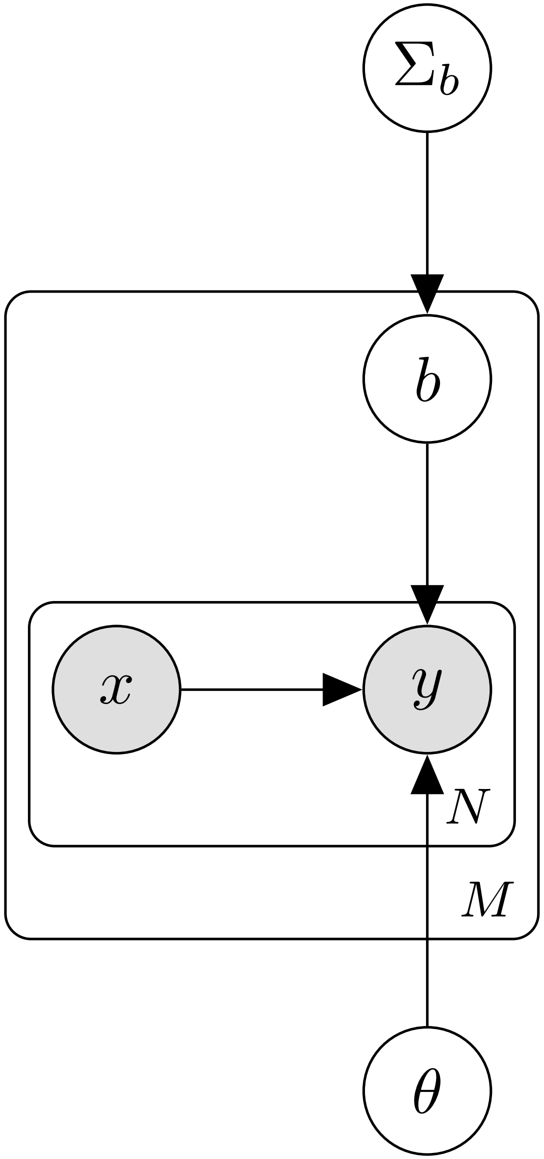
Subjects

Sixty-six volunteers (16 women, age 24–55, mean = 43 years; and 50 men, age 24–62, mean = 37 years) participated. All subjects held valid Commercial Motor Vehicle (CMV) drivers' licenses. Subjects were in good general health as determined by medical history and medical examination and were free of neurological diseases, psychiatric disorders, sleep disorders, and drug or alcohol addiction. They did not use nicotine in any form and reported consuming no more than 300–400 mg caffeine per day. Subjects were medication-free (including over-the-counter medications) beginning 48 h prior to the study, with the exception that female subjects could continue birth control medications.

Mixed linear model assumption

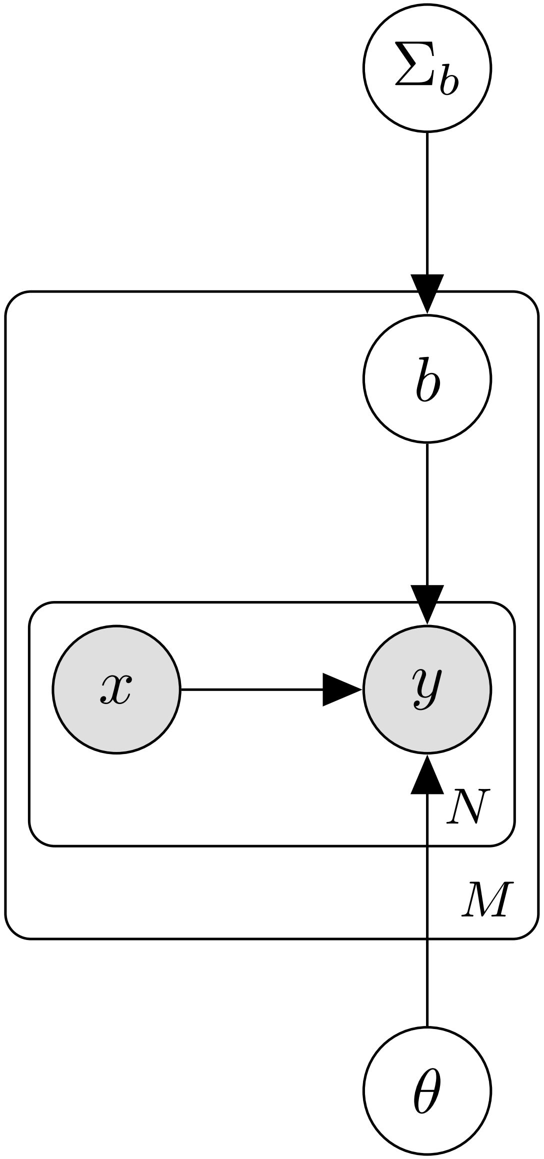


Mixed linear model assumption



$$\theta = \langle \beta, \sigma^2 \rangle$$

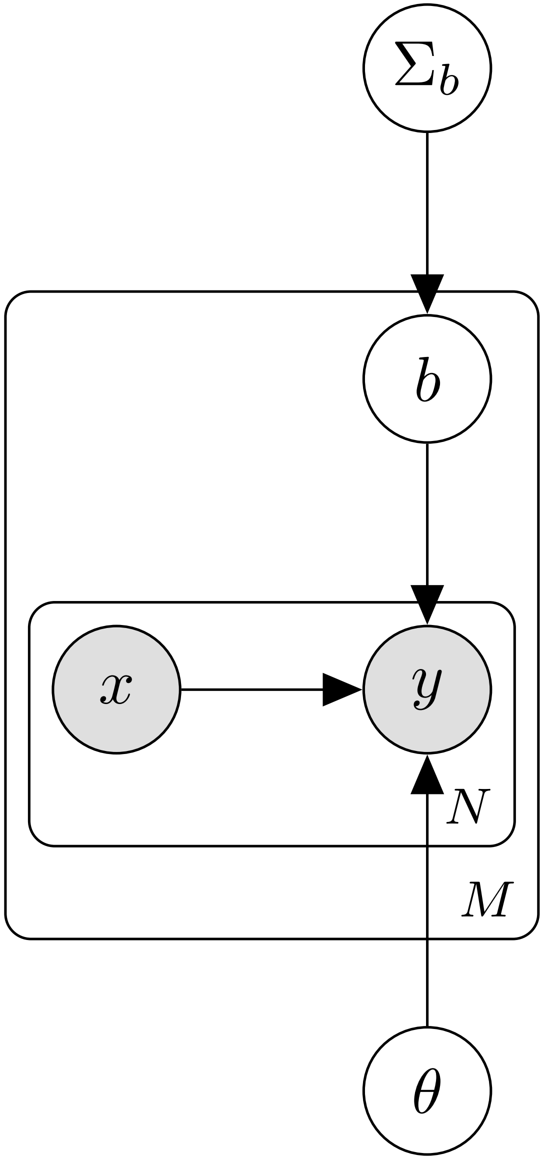
Mixed linear model assumption



$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

Mixed linear model assumption

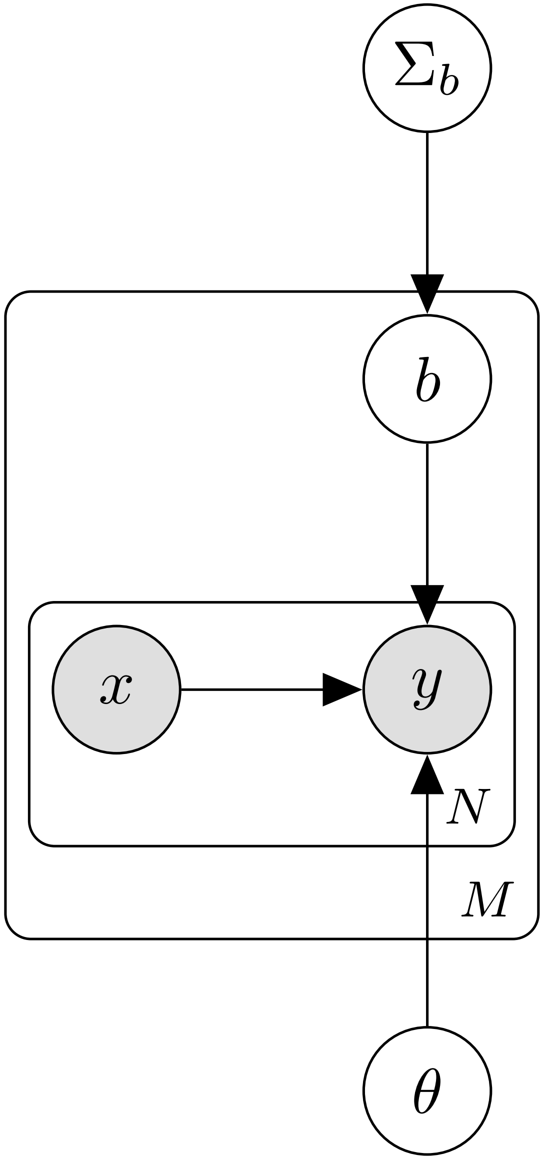


$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

$$\eta = \left(\beta + b\right) x$$

Mixed linear model assumption



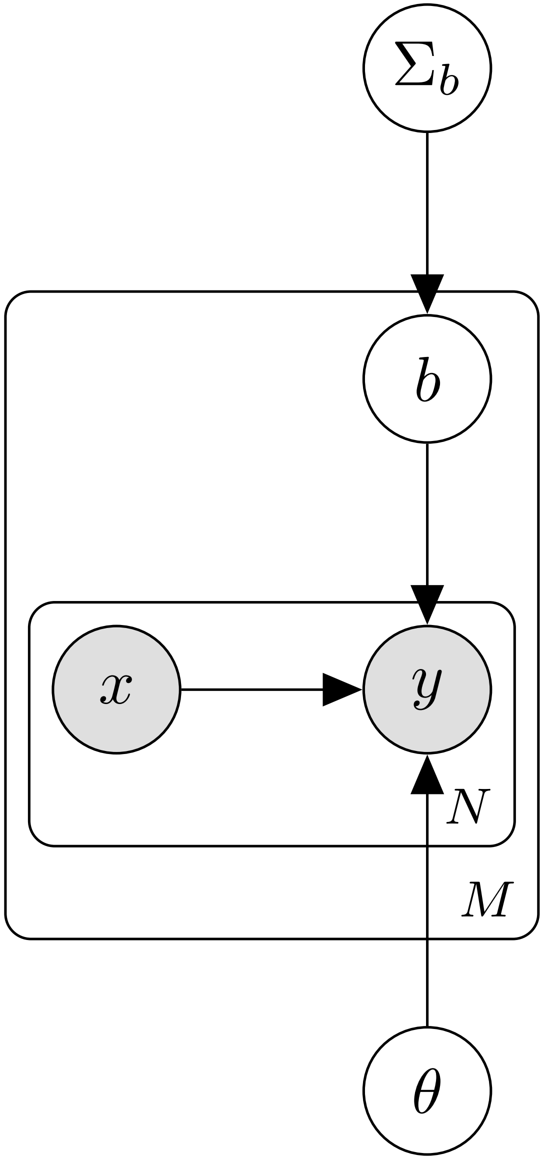
$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

$$\eta = \left(\beta + b\right) x$$

$$\hat{y} = \eta$$

Mixed linear model assumption



$$\theta = \langle \beta, \sigma^2 \rangle$$

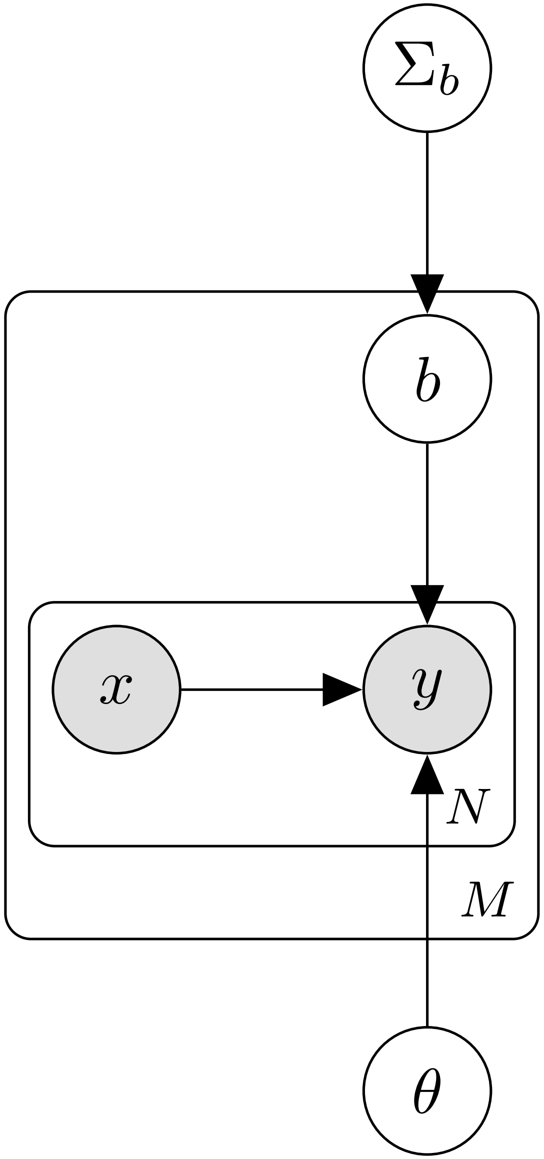
$$b \sim N(0, \Sigma_b)$$

$$\eta = \left(\beta + b\right) x$$

$$\hat{y} = \eta$$

$$P(y \mid \hat{y}) \sim N(\hat{y}, \sigma^2)$$

Mixed linear model assumption



$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

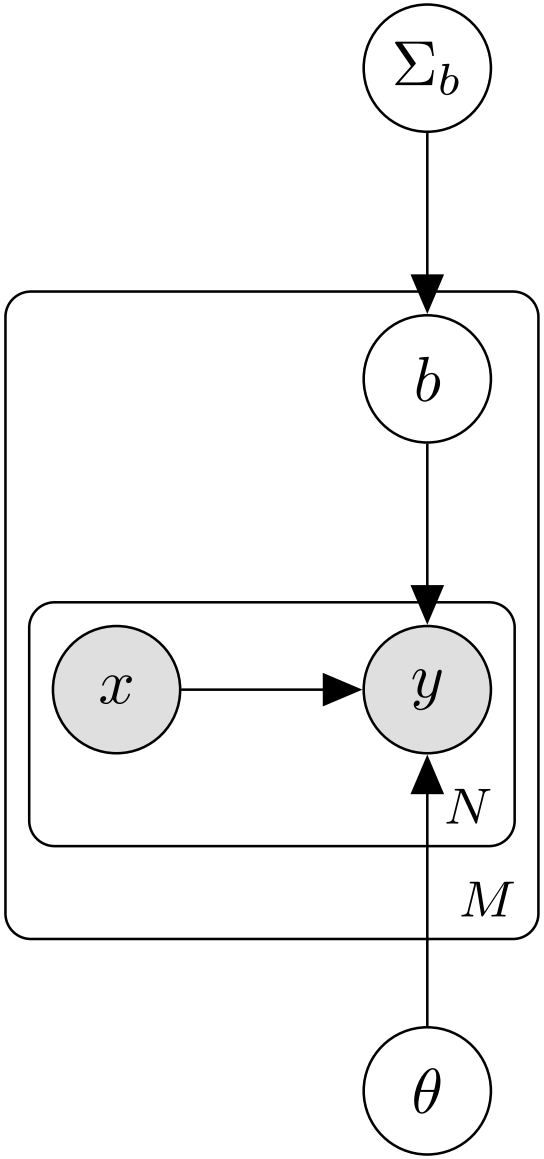
$$\eta = \left(\beta + b\right) x$$

$$\widehat{y} = \eta$$

$$P(y \mid \widehat{y}) \sim N(\widehat{y}, \sigma^2)$$

Given our data visualization, we will treat the predictor X (# days of sleep deprivation) as a single scalar

Mixed linear model assumption



$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

$$\eta = \left(\beta + b\right) x$$

$$\widehat{y} = \eta$$

$$P(y \mid \widehat{y}) \sim N(\widehat{y}, \sigma^2)$$

Given our data visualization, we will treat the predictor X (# days of sleep deprivation) as a single scalar

Question: What should our overall model look like, and once fitted how will we use it to answer our scientific question?

Maximum-likelihood linear mixed model fit

```
> summary(m.lme4 <- lmer(Response ~ Days + (Days | Subject),data=sleepstudy,REML=F))
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Response ~ Days + (Days | Subject)
   Data: sleepstudy

     AIC      BIC   logLik deviance df.resid 
   137.0    156.1    -62.5    125.0      174 

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-3.7431 -0.5376 -0.0467  0.5156  3.8957 

Random effects:
 Groups   Name        Variance Std.Dev. Corr 
 Subject  (Intercept) 0.175580 0.41902       
          Days        0.003202 0.05659  -0.18
 Residual             0.072816 0.26984       
Number of obs: 180, groups:  Subject, 18

Fixed effects:
            Estimate Std. Error t value
(Intercept)  3.96581    0.10560  37.554
Days        -0.11099    0.01506  -7.368

Correlation of Fixed Effects:
     (Intr)
Days -0.288
```

Maximum-likelihood linear mixed model fit

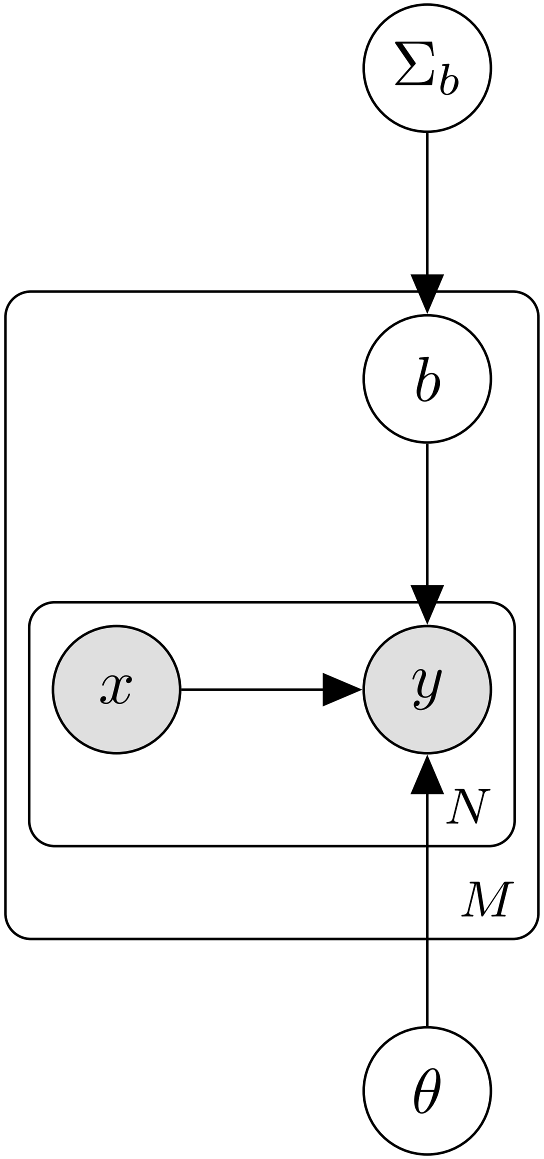
```
> summary(m.lme4 <- lmer(Response ~ Days + (Days | Subject),data=sleepstudy,REML=F))
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Response ~ Days + (Days | Subject)
   Data: sleepstudy

     AIC      BIC   logLik deviance df.resid
   137.0    156.1    -62.5    125.0      174

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7431 -0.5376 -0.0467  0.5156  3.8957

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept) 0.175580 0.41902
          Days        0.003202 0.05659  -0.18
 Residual             0.072816 0.26984
Number of obs: 180, groups:  Subject, 18

Fixed effects:
            Estimate Std. Error t value
(Intercept)  3.96581    0.10560  37.554
Days        -0.11099    0.01506  -7.368

Correlation of Fixed Effects:
     (Intr)
Days -0.288
```

Maximum-likelihood linear mixed model fit

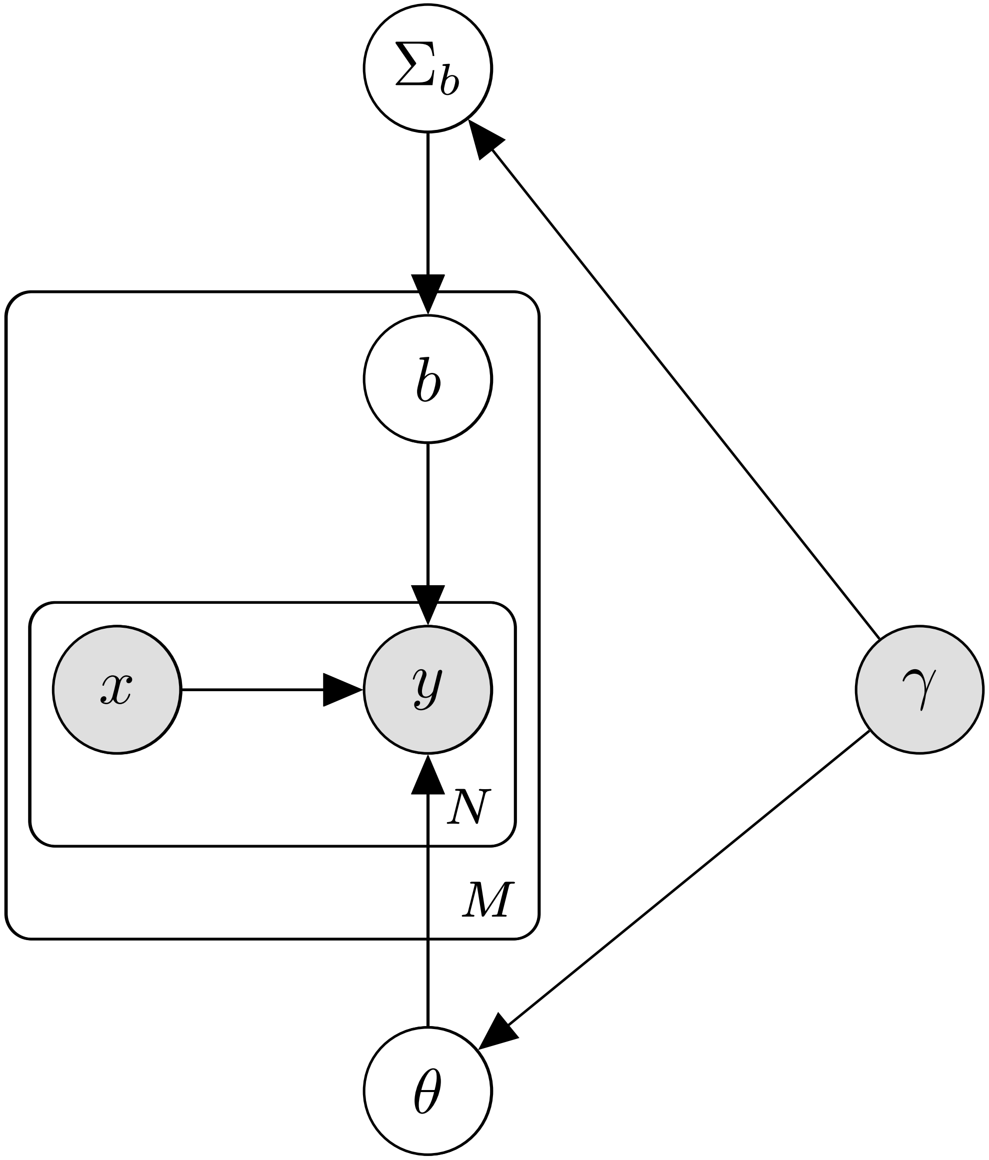
```
> summary(m.lme4 <- lmer(Response ~ Days + (Days | Subject),data=sleepstudy,REML=F))
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Response ~ Days + (Days | Subject)
   Data: sleepstudy

     AIC      BIC   logLik deviance df.resid
   137.0    156.1    -62.5    125.0      174

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7431 -0.5376 -0.0467  0.5156  3.8957

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept) 0.175580 0.41902
          Days        0.003202 0.05659  -0.18
 Residual             0.072816 0.26984
Number of obs: 180, groups:  Subject, 18

Fixed effects:
            Estimate Std. Error t value
(Intercept)  3.96581    0.10560  37.554
Days        -0.11099    0.01506  -7.368

Correlation of Fixed Effects:
     (Intr)
Days -0.288
```

We don't get any uncertainty bounds on this!

Prior on model parameters



Prior on model parameters



Unnormalizable posteriors

*For now, θ denotes **all** model parameters, not just the fixed effects*

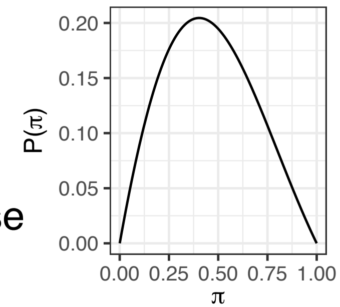
- Our motivation: Bayesian posterior inference

$$P(\theta \mid \mathbf{y}, I) = \frac{P(\mathbf{y} \mid \theta, I) P(\theta \mid I)}{P(\mathbf{y} \mid I)}$$

- Sometimes $P(\mathbf{y} \mid I)$ can't be calculated exactly. Example

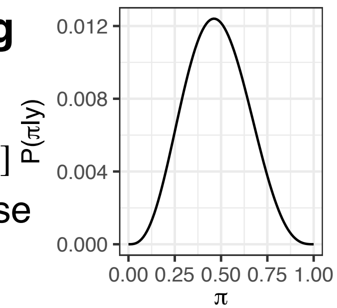
Bernoulli data with non-conjugate prior:

$$P(\pi) \propto \begin{cases} \pi(1-\pi)e^{-\pi^2} & \pi \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

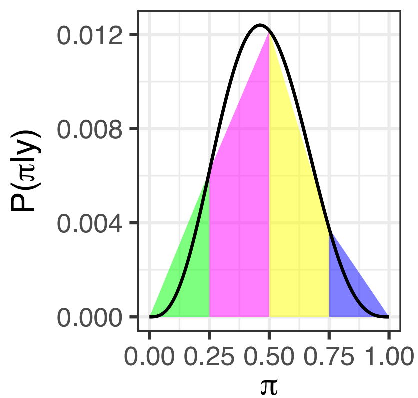


Posterior after observing 2 heads, 2 tails:

$$P(\pi) \propto \begin{cases} \pi^3(1-\pi)^3 e^{-\pi^2} & \pi \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$



- In simple cases like this, we can numerically approximate the integral:



- But in high dimension and/or unbounded ranges, difficult or even impossible!

Unnormalizable posteriors

*For now, θ denotes **all** model parameters, not just the fixed effects*

- Our motivation: Bayesian posterior inference

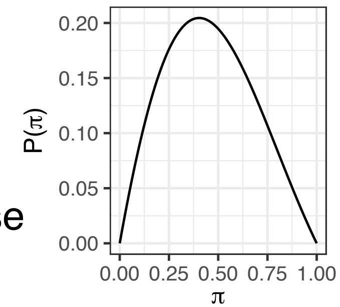
$$P(\theta \mid \mathbf{y}, I) = \frac{P(\mathbf{y} \mid \theta, I) P(\theta \mid I)}{P(\mathbf{y} \mid I)}$$

Model parameters

- Sometimes $P(\mathbf{y} \mid I)$ can't be calculated exactly. Example

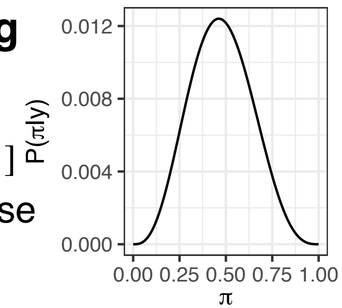
Bernoulli data with non-conjugate prior:

$$P(\pi) \propto \begin{cases} \pi(1-\pi)e^{-\pi^2} & \pi \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

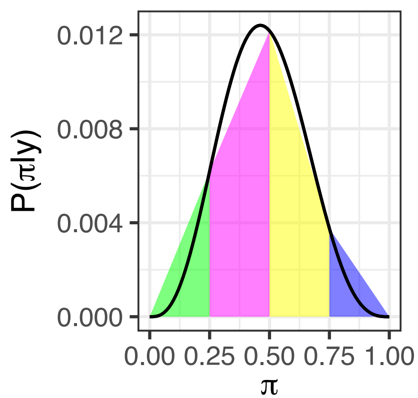


Posterior after observing 2 heads, 2 tails:

$$P(\pi) \propto \begin{cases} \pi^3(1-\pi)^3 e^{-\pi^2} & \pi \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$



- In simple cases like this, we can numerically approximate the integral:



- But in high dimension and/or unbounded ranges, difficult or even impossible!

Unnormalizable posteriors

*For now, θ denotes **all** model parameters, not just the fixed effects*

- Our motivation: Bayesian posterior inference

$$P(\theta \mid \mathbf{y}, I) = \frac{P(\mathbf{y} \mid \theta, I)P(\theta \mid I)}{P(\mathbf{y} \mid I)}$$

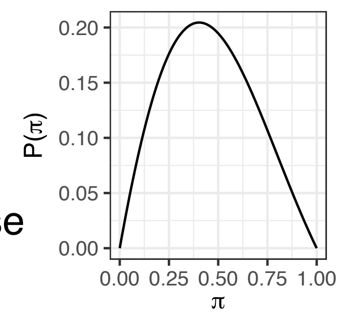
Model parameters

Background knowledge

- Sometimes $P(\mathbf{y} \mid I)$ can't be calculated exactly. Example

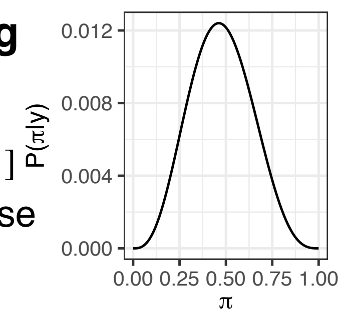
Bernoulli data with non-conjugate prior:

$$P(\pi) \propto \begin{cases} \pi(1-\pi)e^{-\pi^2} & \pi \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

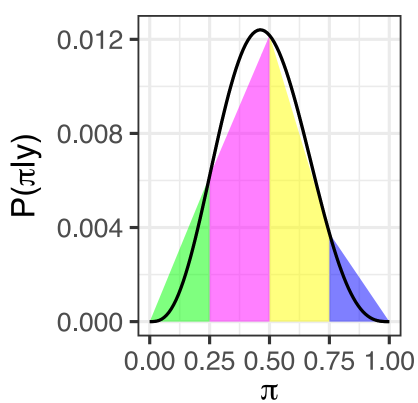


Posterior after observing 2 heads, 2 tails:

$$P(\pi) \propto \begin{cases} \pi^3(1-\pi)^3e^{-\pi^2} & \pi \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$



- In simple cases like this, we can numerically approximate the integral:



- But in high dimension and/or unbounded ranges, difficult or even impossible!

Unnormalizable posteriors

*For now, θ denotes **all** model parameters, not just the fixed effects*

- Our motivation: Bayesian posterior inference

Observed data

$$P(\theta \mid \mathbf{y}, I) = \frac{P(\mathbf{y} \mid \theta, I)P(\theta \mid I)}{P(\mathbf{y} \mid I)}$$

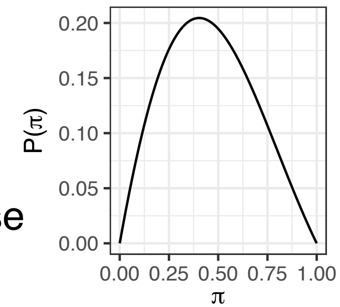
Model parameters

Background knowledge

- Sometimes $P(\mathbf{y} \mid I)$ can't be calculated exactly. Example

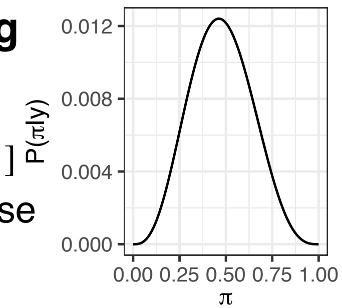
Bernoulli data with non-conjugate prior:

$$P(\pi) \propto \begin{cases} \pi(1-\pi)e^{-\pi^2} & \pi \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

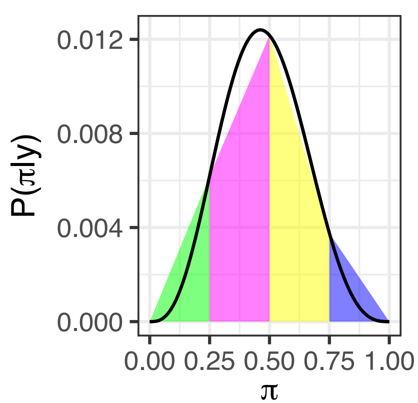


Posterior after observing 2 heads, 2 tails:

$$P(\pi) \propto \begin{cases} \pi^3(1-\pi)^3e^{-\pi^2} & \pi \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$



- In simple cases like this, we can numerically approximate the integral:



- But in high dimension and/or unbounded ranges, difficult or even impossible!

Unnormalizable posteriors

*For now, θ denotes **all** model parameters, not just the fixed effects*

- Our motivation: Bayesian posterior inference

Observed data

$$P(\theta \mid \mathbf{y}, I) = \frac{P(\mathbf{y} \mid \theta, I) P(\theta \mid I)}{P(\mathbf{y} \mid I)}$$

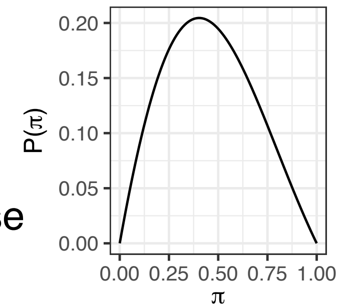
Model parameters

Background knowledge

- Sometimes $P(\mathbf{y} \mid I)$ can't be calculated exactly. Example

Bernoulli data with non-conjugate prior:

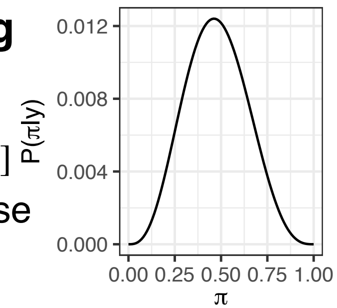
$$P(\pi) \propto \begin{cases} \pi(1-\pi)e^{-\pi^2} & \pi \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$



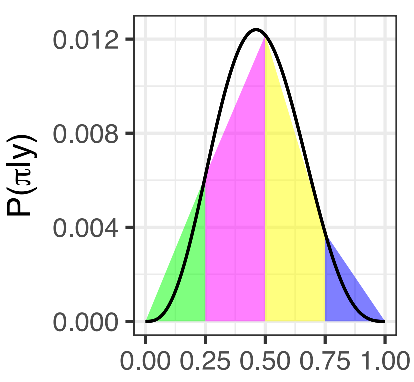
Posterior after observing 2 heads, 2 tails:

$$P(\pi) \propto \begin{cases} \pi^3(1-\pi)^3 e^{-\pi^2} & \pi \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

No closed form!



- In simple cases like this, we can numerically approximate the integral:



- But in high dimension and/or unbounded ranges, difficult or even impossible!

MCMC for posterior sampling

- We do a random walk on the *unnormalized* posterior:

$$P(\theta \mid \mathbf{y}, I) \propto P(\mathbf{y} \mid \theta, I) P(\theta \mid I)$$

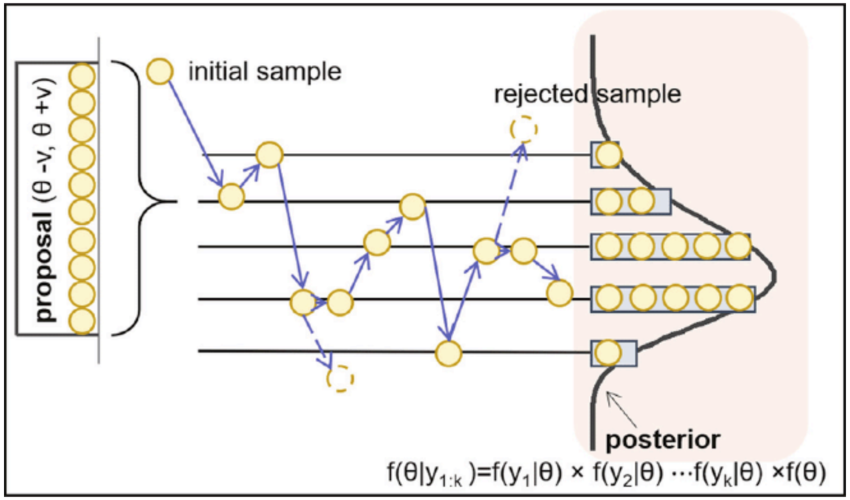
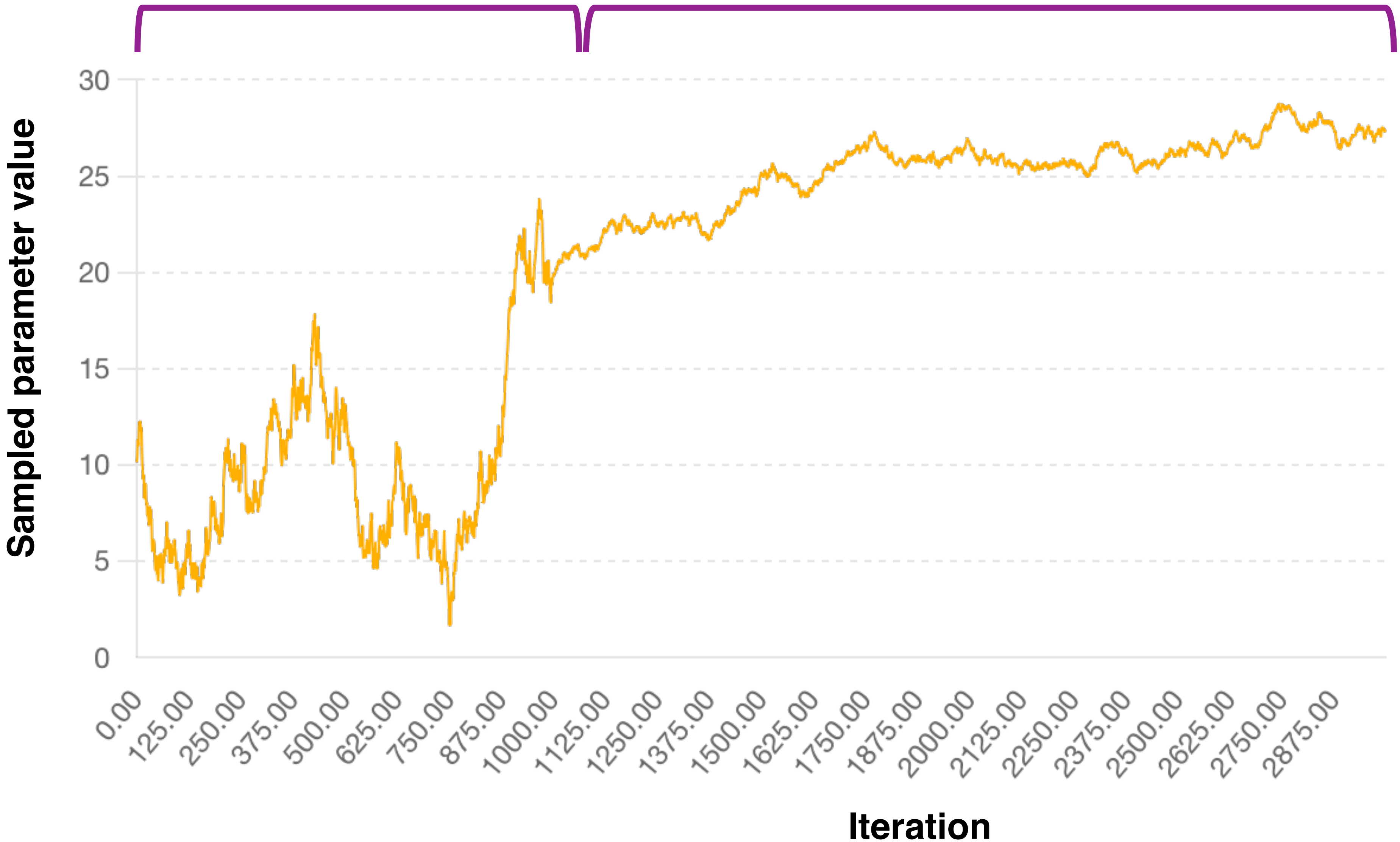
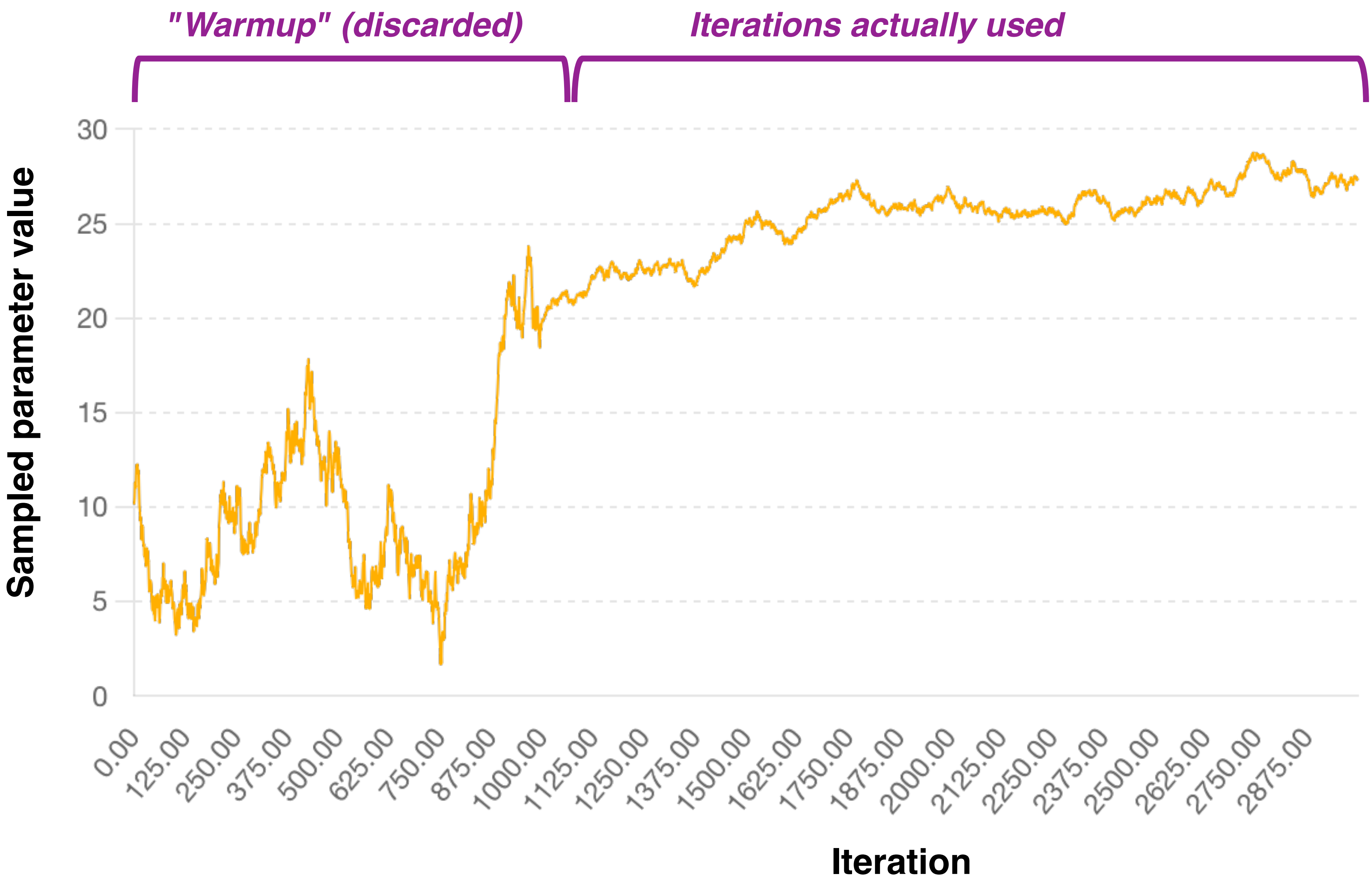


Image credit: Dong et al., 2019, http://dx.doi.org/10.5772/intechopen.82781

Burn-in/warmup

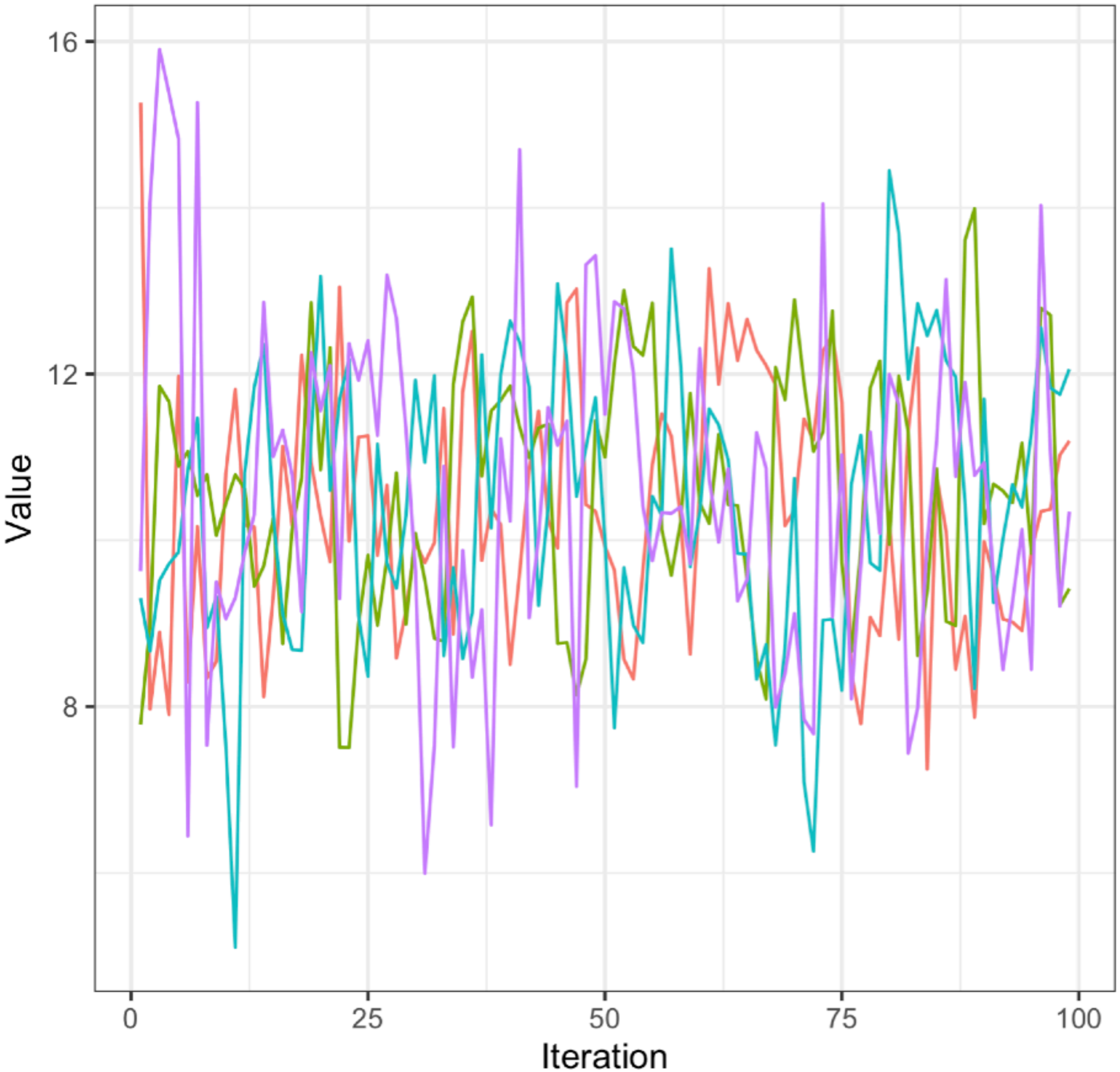


Burn-in/warmup



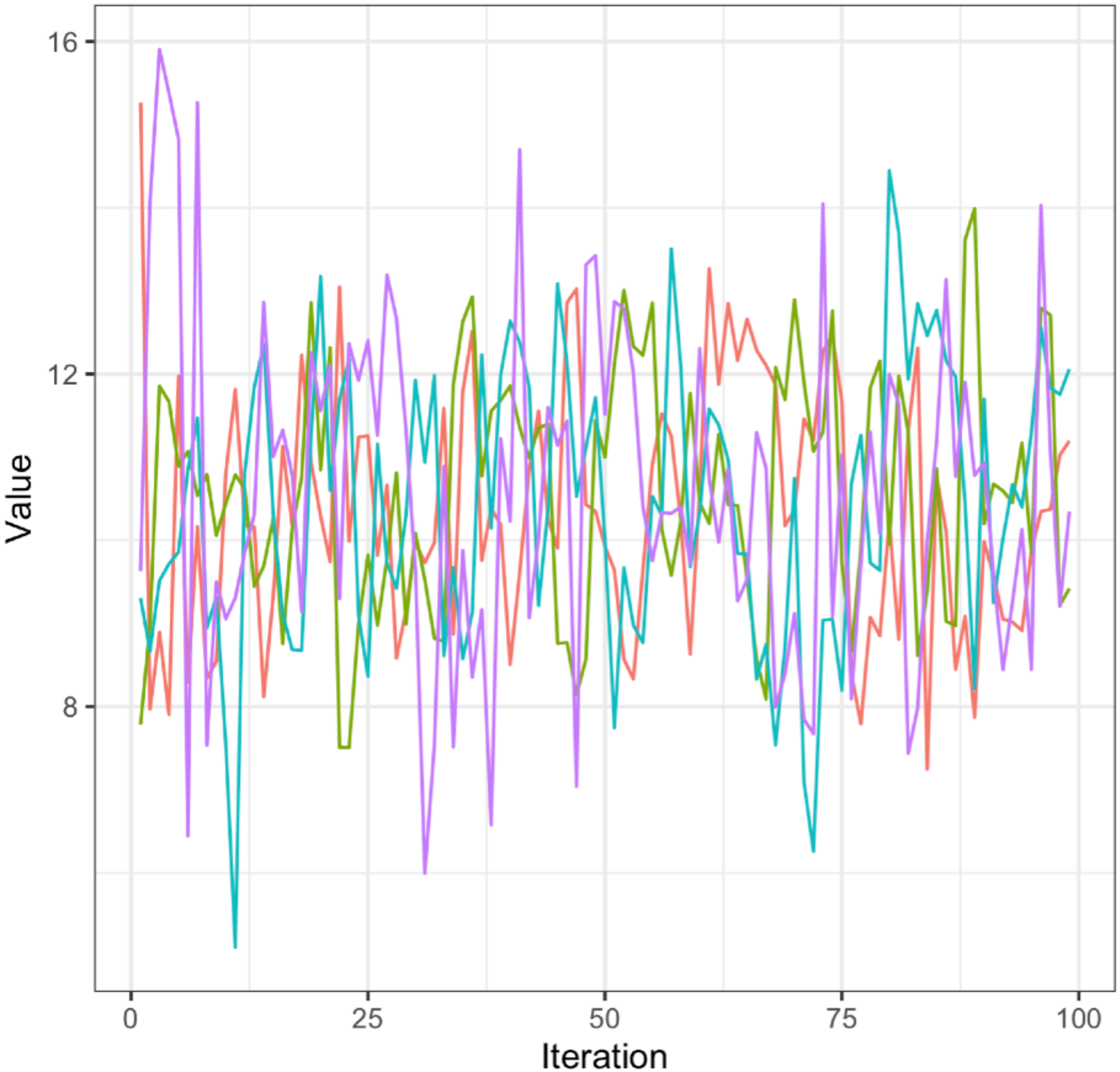
How well have we explored the posterior?

Traceplot for β



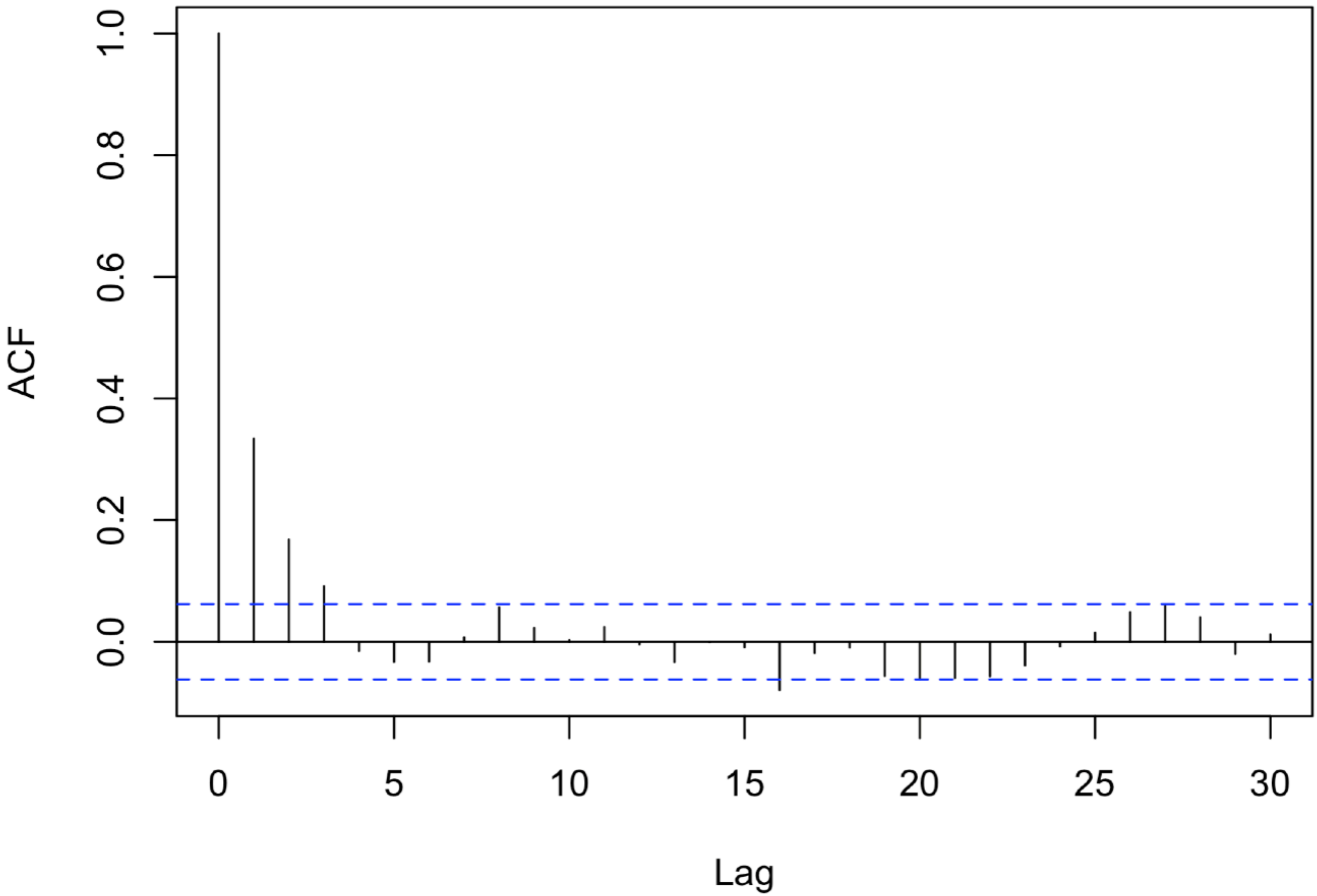
How well have we explored the posterior?

Traceplot for β



The classic regression diagnostic is (some version of) the ratio of **between-chain** versus **within-chain** variances, called $\widehat{R}$

Autocorrelation and effective sample size



Therefore, the *effective number of samples* from the Markov Chain is lower than the total number of samples