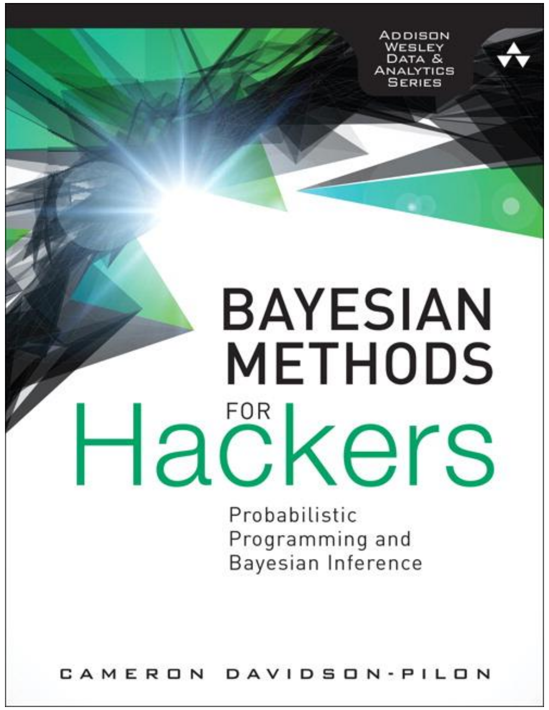
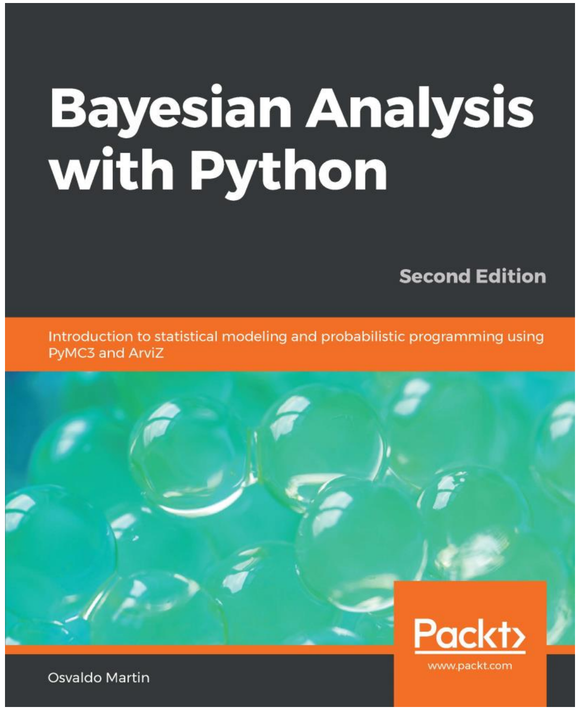
Bayesian modeling

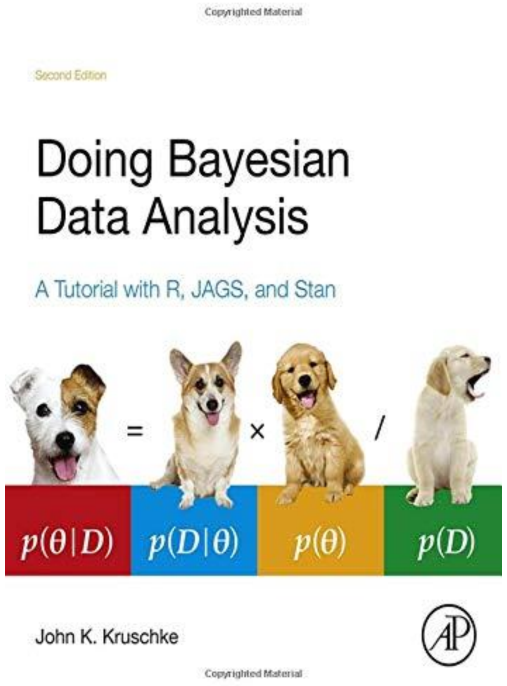
Thinking probalistically

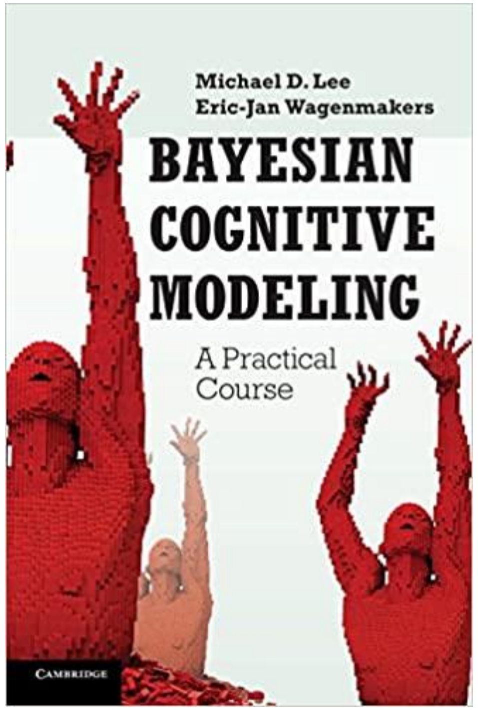
Introduction to PyMC3

Thinking probabilistically









Thinking probabilistically

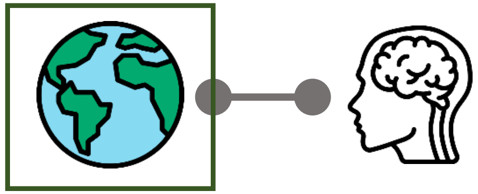
Frequentist

Bayesian

Probability is the long-run frequency of events.

$f(y|\theta)$

Fixed

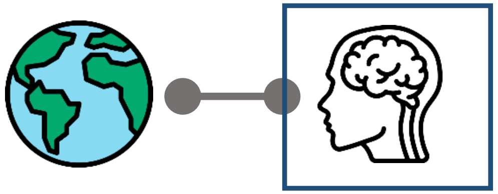


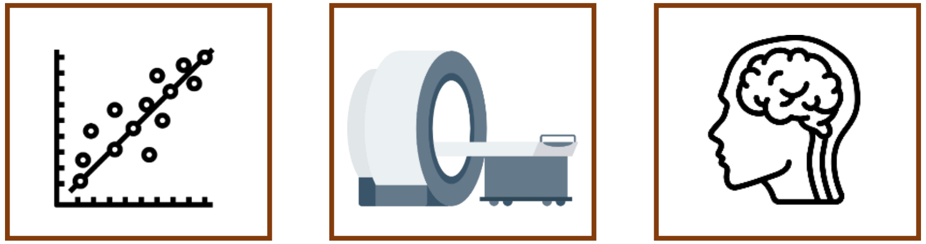
Probability measure the *believability in an event*.

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

Fixed

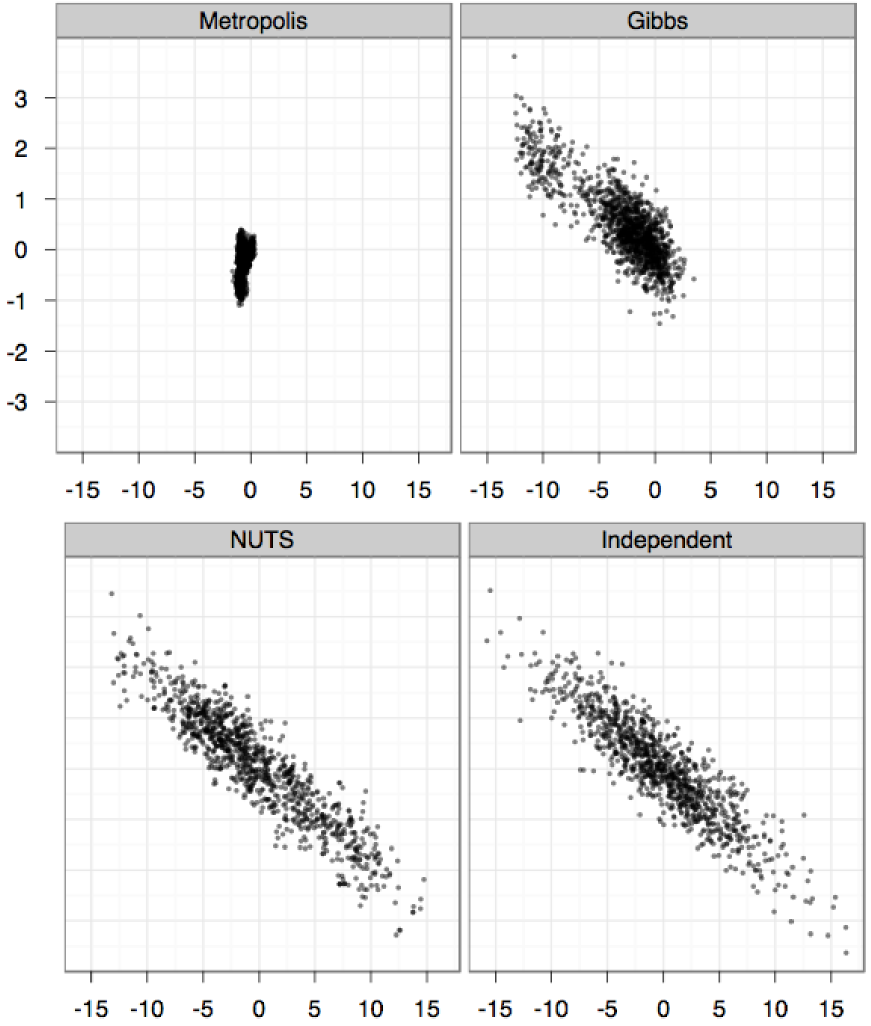
Bayesian inference is simply updating your beliefs after considering new evidence.





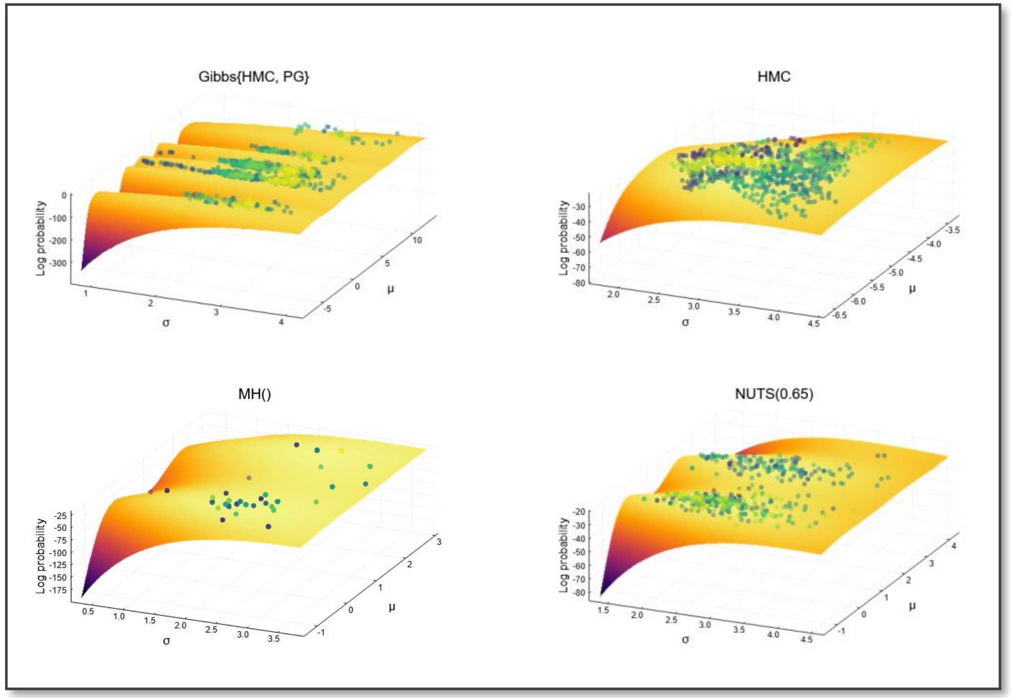
MCMC samplers

- Approximation
  - Variational inference
- Stochastic sampling
  - MCMC methods



Online demo:

https://chi-feng.github.io/mcmc-demo/



https://turing.ml/dev/docs/using-turing/sampler-viz

Probabilistic programming languages

BUGS

JAGS

BAYESIAN MODELING WITH RJAGS

PyMC3

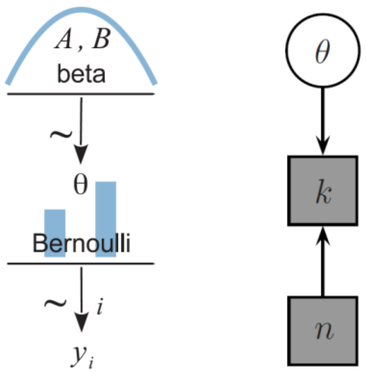
Stan

Turing.jl

TensorFlow

Probabilistic programming languages

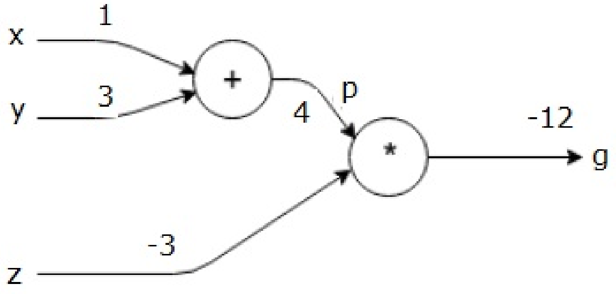
Graphical model



$$\theta \sim \mathrm{Beta}(1, 1)$$
$$k \sim \mathrm{Binomial}(\theta, n)$$

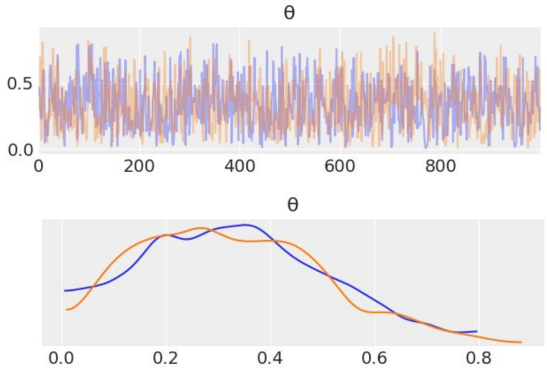
```
with pm.Model() as our_first_model:
  θ = pm.Beta('θ', alpha=1., beta=1.)
  y = pm.Bernoulli('y', p=θ, observed=data)
  trace = pm.sample(1000, random_seed=123)
```

Computational graph



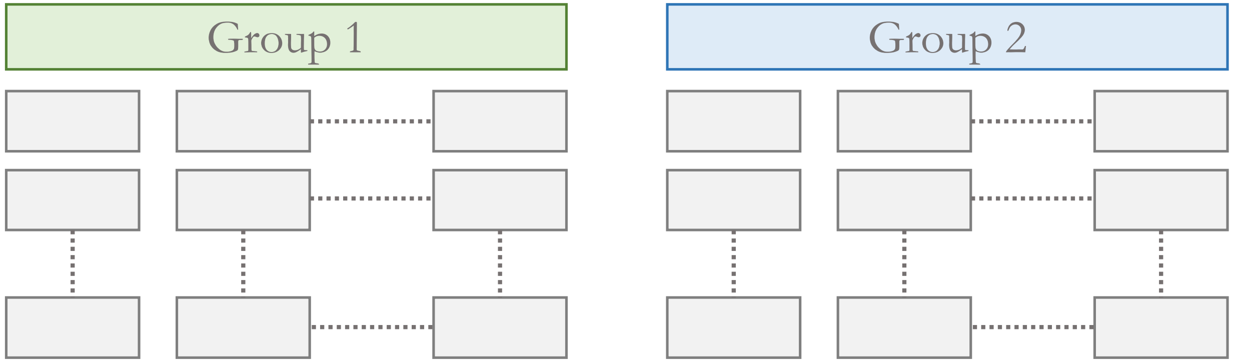
theano

- Automatic differentiation
- GPU computing
- Optimizations



Sampling

Hierarchical/multilevel models

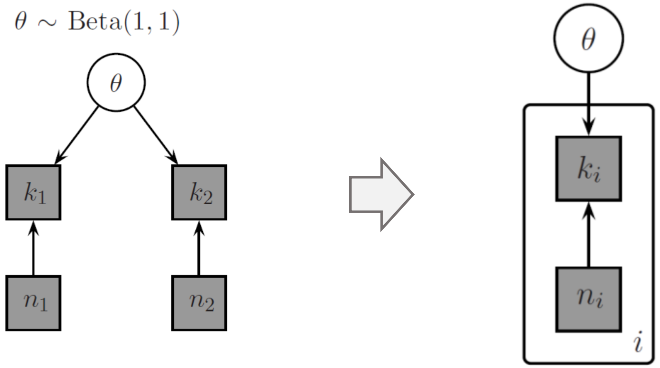


- Hyperpriors
- Hyperparameters

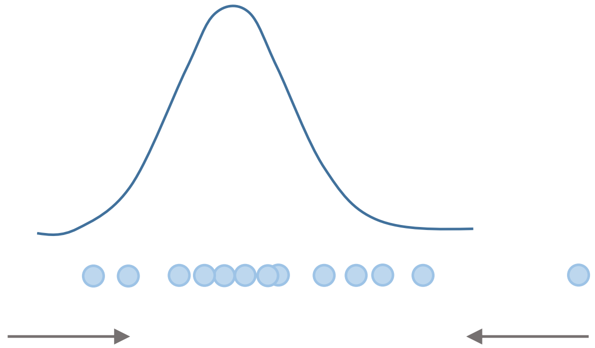
Plate notation

$$k_1 \sim \text{Binomial}(\theta, n_1)$$
$$k_2 \sim \text{Binomial}(\theta, n_2)$$
$$\theta \sim \text{Beta}(1, 1)$$

$$k_i \sim \text{Binomial}(\theta, n_i)$$
$$\theta \sim \text{Beta}(1, 1)$$

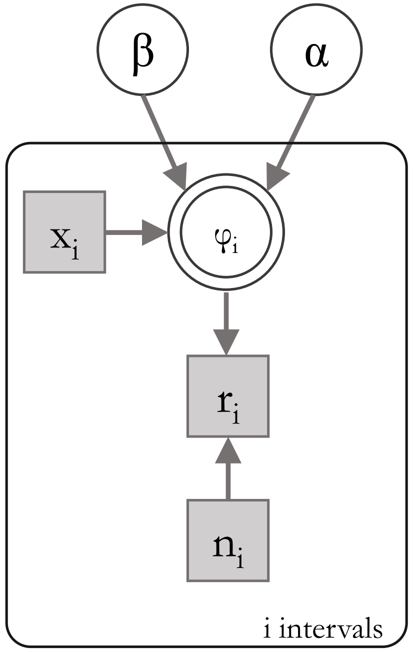


Shrinkage



Psychophysics

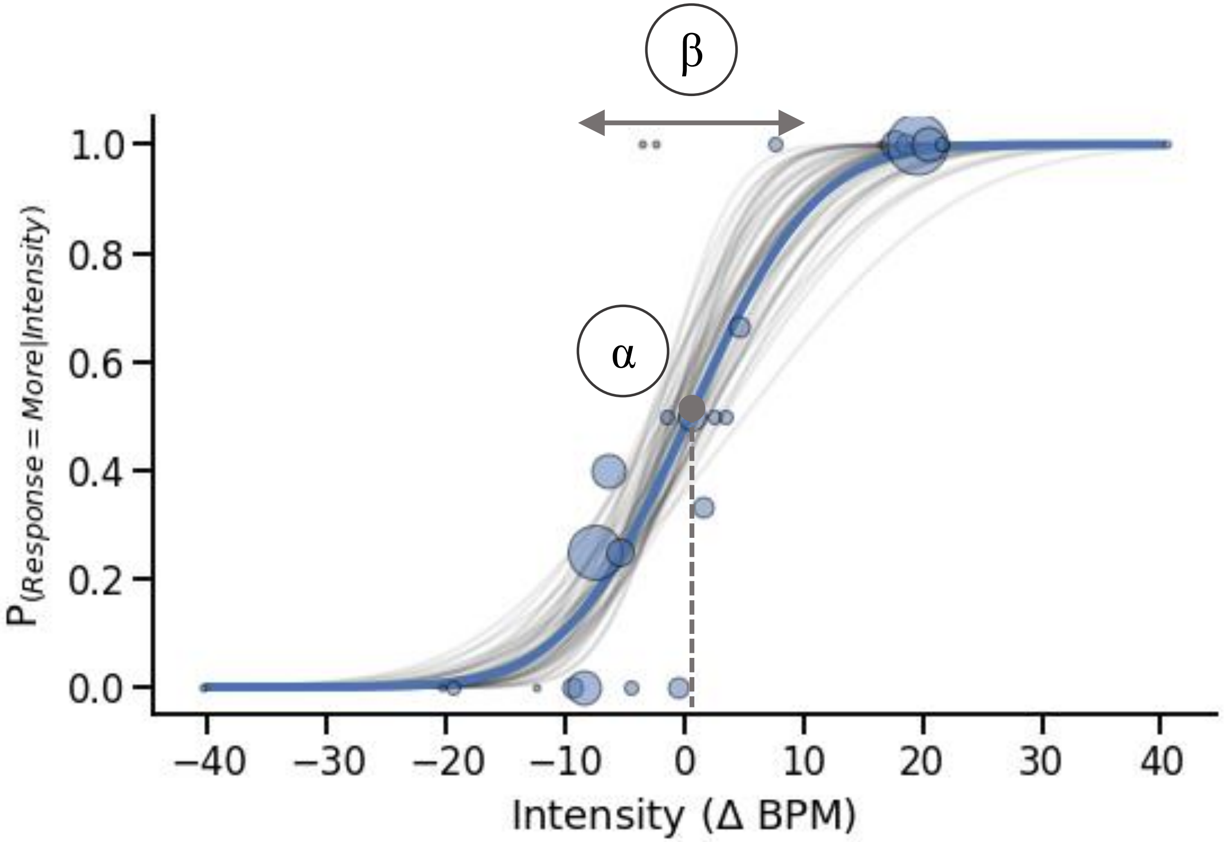
Psychophysics is concerned with measuring how external physical stimuli cause internal psychological sensations.



N responses r_i = 0 3 10 ....

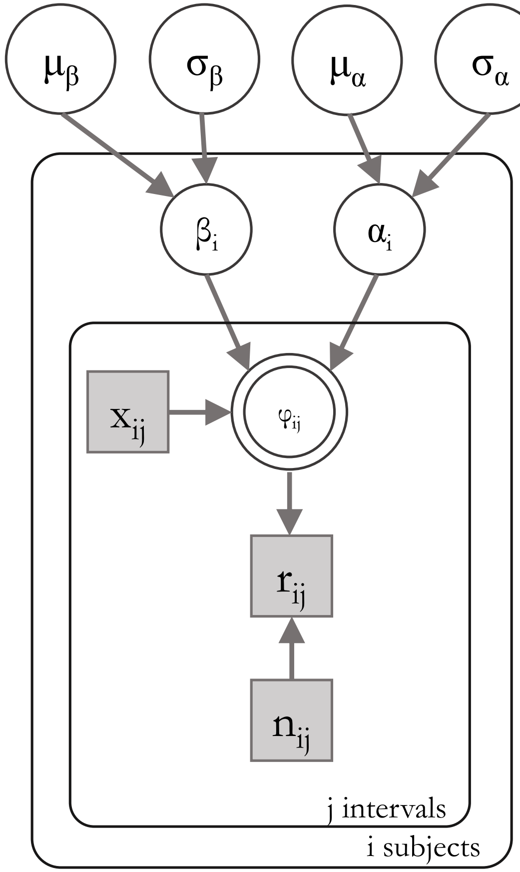
N trials n_i = 5 20 15 ....

Intensities x_i = -20 -10 0 ....



Hierarchial psychophysics

Psychophysics is concerned with measuring how external physical stimuli cause internal psychological sensations.

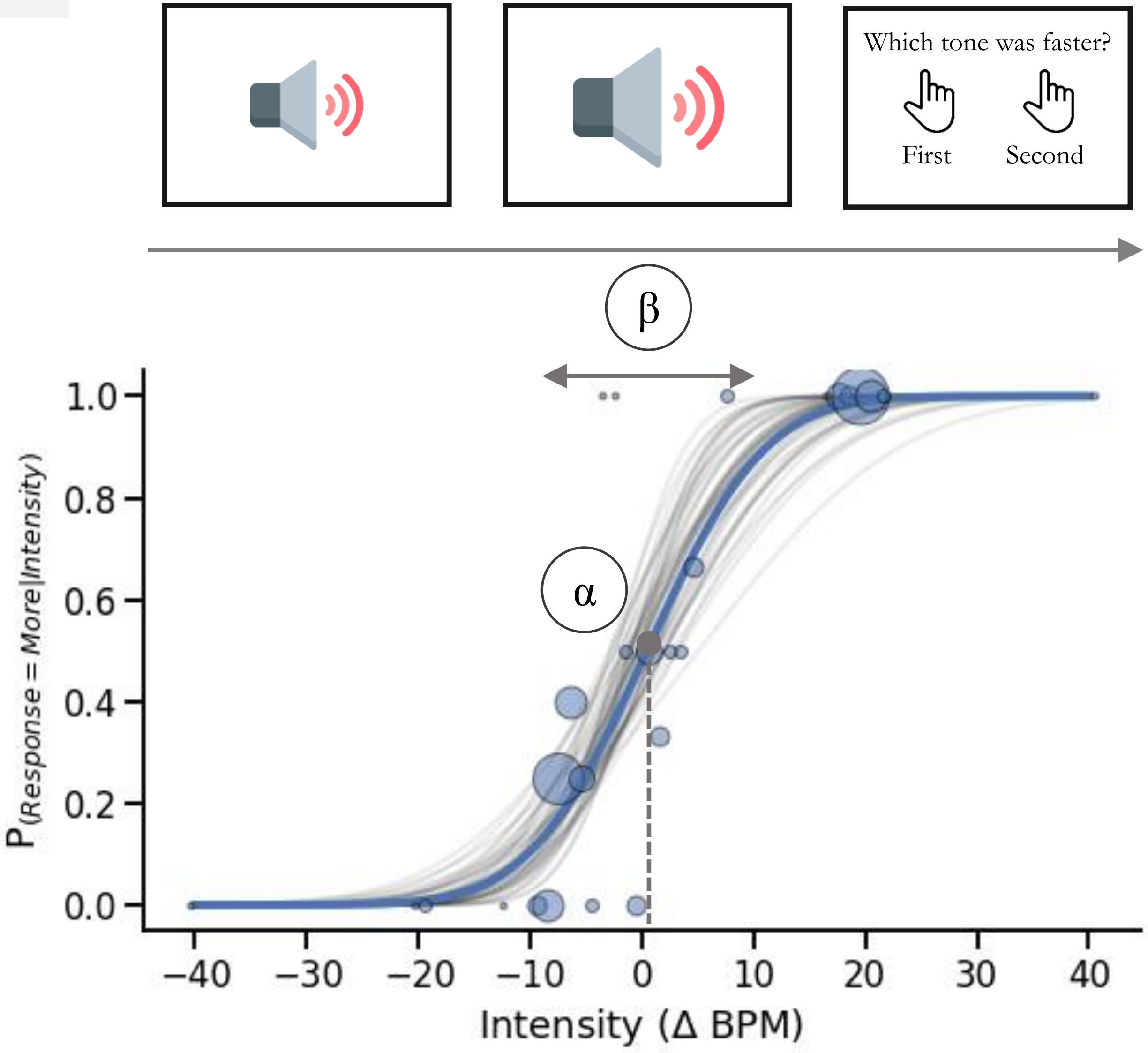


N responses r_i = 0 3 10 ....

N trials n_i = 5 20 15 ....

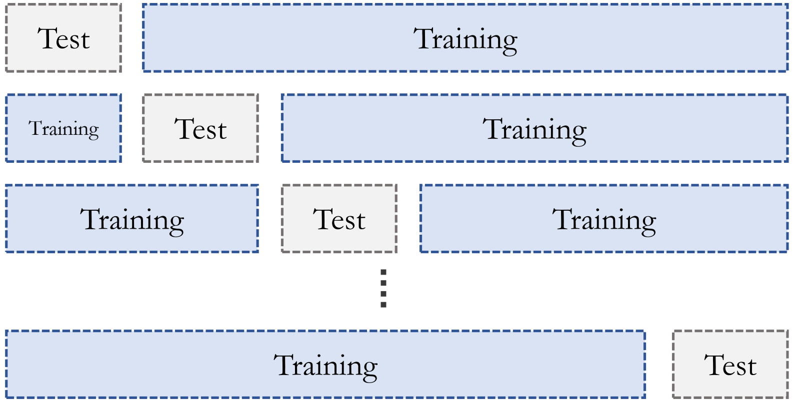
Intensities x_i = -20 -10 0 ....

Subject ID S_i = 0 0 0 ....



Model comparison

Cross-validation



If we have two or more equi valent explanations for the s ame phenomenon, we shoul d choose the simpler one… but also the more accurate.

Information criteria

How well models fit the data while taking into account their complexity through a penalization term

- Mean square error (MSE)

$$\frac{1}{n}\sum_{i=1}^{n}(yi - E(y_i \mid \theta))^2$$

- Log-likelihood

$$\sum_{i=1}^{n}\log p(y_i \mid \theta)$$

- Deviance

$$-2\sum_{i=1}^{n}\log p(y_i \mid \theta)$$

- Akaike information criterion

$$AIC = \underbrace{-2\sum_{i=1}^{n}\log p(y_i \mid \theta_{mle})}_{\text{How well the model fits the data}} + \underbrace{2pAIC}_{\text{Penalizes complex models}}$$

- Widely applicable information Criterion

$$WAIC = \underbrace{-2\mathrm{lppd}}_{\text{How well the model fits the data}} + \underbrace{2p_{WAIC}}_{\text{Penalizes complex models}}$$

The generalized linear model

Interaction term

$$y \;=\; \beta_0 \;+\; \beta_1 x_1 \;+\; \beta_2 x_2$$

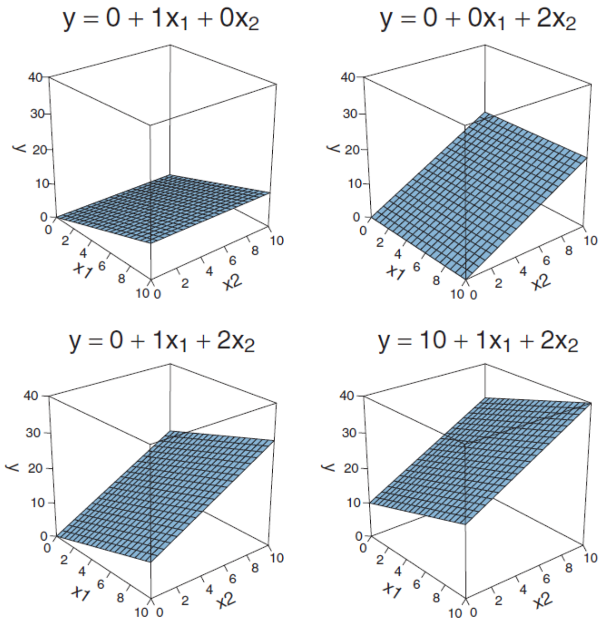


Figure 15.2 Examples of linear functions of two variables, x_1 and x_2. Upper left: Only x_1 has an influence on y. Upper right: Only x_2 has an influence on y. Lower left: x_1 and x_2 have an additive influence on y. Lower right: Nonzero intercept is added.

$$y \;=\; \beta_0 \;+\; \beta_1 x_1 \;+\; \beta_2 x_2 + \boxed{\beta_3 x_1 x_2}$$

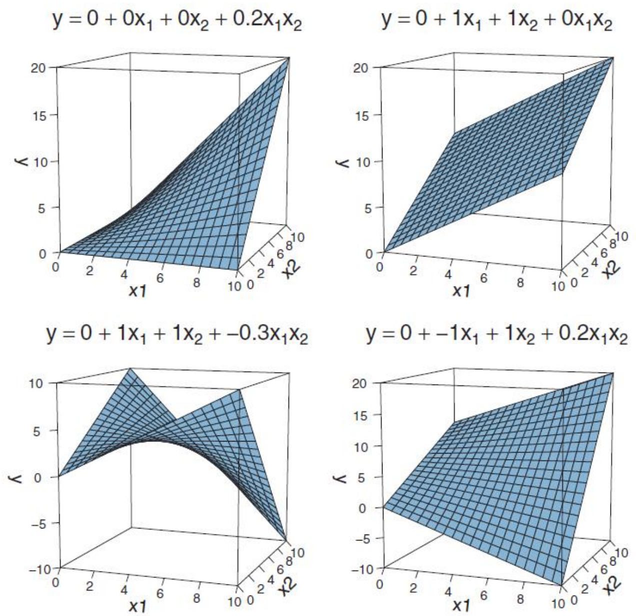


Figure 15.3 Multiplicative interaction of two variables, x_1 and x_2. Upper right panel shows *zero* interaction, for comparison. Figure 18.8, p. 526, provides additional perspective and insight.

$$y = logisitic(x) = \frac{1}{(1+e^{-x})}$$

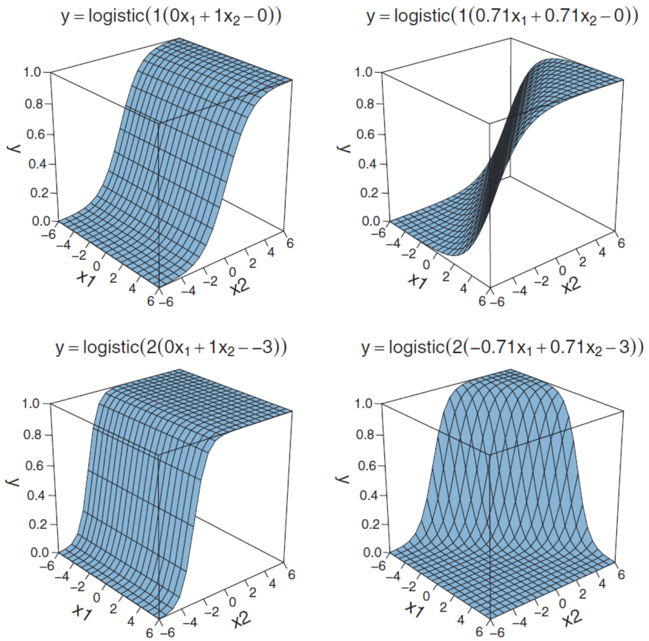
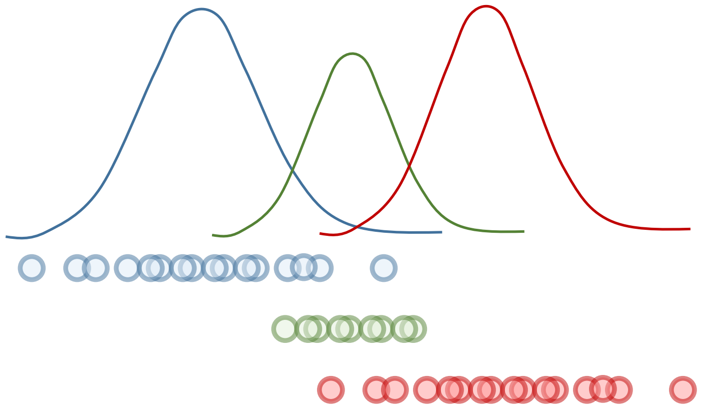
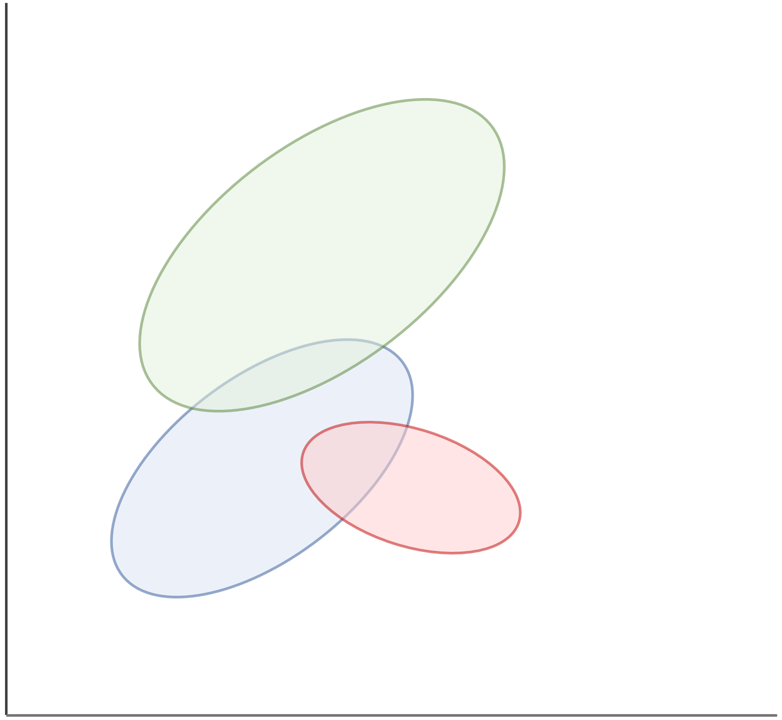


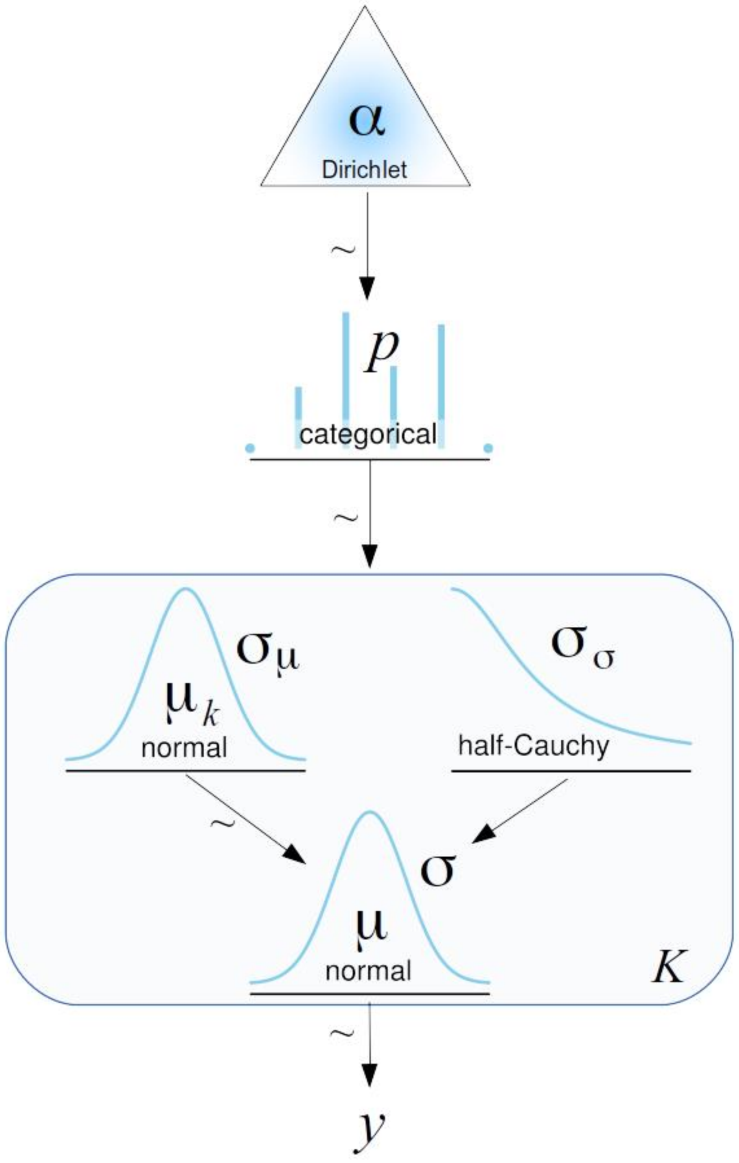
Figure 15.7 Examples of logistic functions of two variables. Top two panels show logistics with the same gain and threshold, but different coefficients on the predictors. The left two panels show logistics with the same coefficients on the predictors, but different gains and thresholds. The lower right panel shows a case with a negative coefficient on the first predictor.

Kruschke, J. (2015). Doing Bayesian data analysis : a tutorial with R, JAGS, and Stan. Boston: Academic Press. Chapter 15.

Mixture models

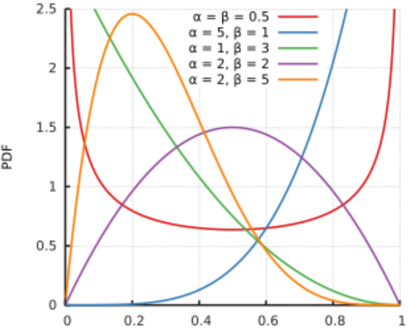






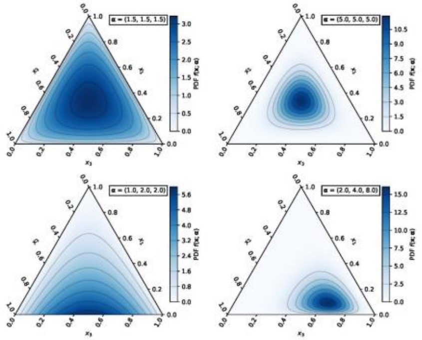
Mixture models

Beta distribution



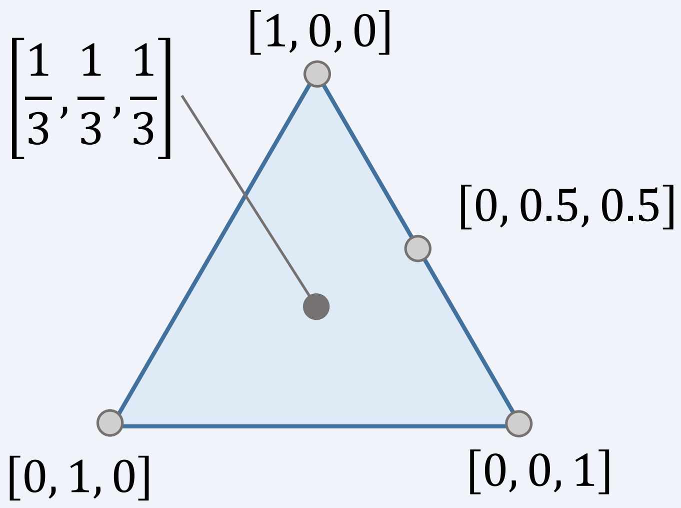
$$[\theta, 1-\theta]$$

Dirichlet distribution



$$\theta_k = [0.2, 0.3, 0.5]$$

The Dirichlet distribution is a generalization of the beta distribution for multiple random variables



The parameters of a Dirichlet distribution is a k dimensional vector of non zero numbers.

$$A = [3, 2, 2]$$

The output of a Dirichlet distribution is also a k dimensional vector of discrete probabilities distribution.

$$\theta_k = [0.2, 0.3, 0.5]$$

Beta distribution → Binomial

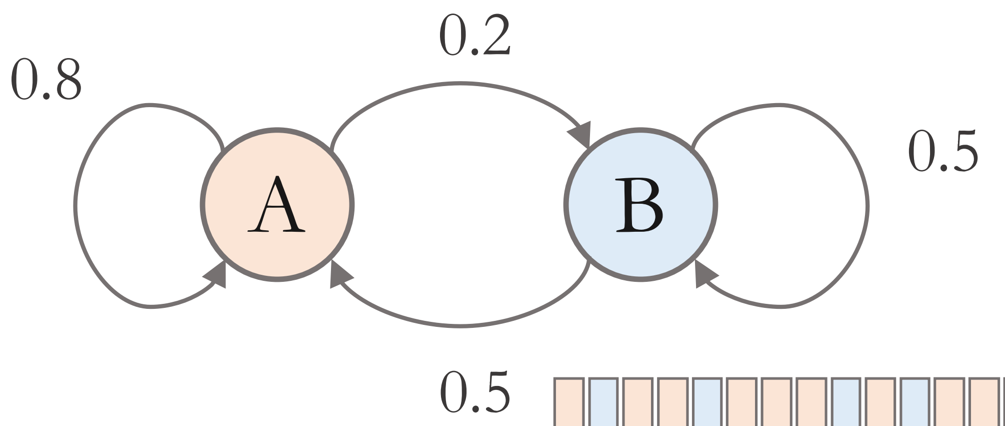
Dirichlet distribution → Multinomial

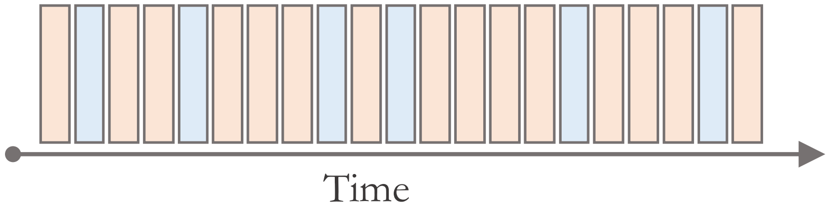
Markov chains

Markov property

$$P(R_{n+1} \,|R_1\,, R_2 \ldots R_n) = P(R_{n+1} \,|\, R_n)$$

Markov chains





Transition probabilities

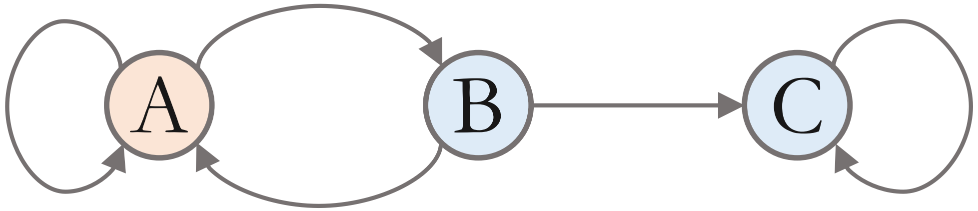
	A	B
A	0.8	0.2
B	0.5	0.5

State transition

$$[1 \quad 0]\begin{bmatrix}0.8 & 0.2\\ 0.5 & 0.5\end{bmatrix} = [0.8\;0.2]$$

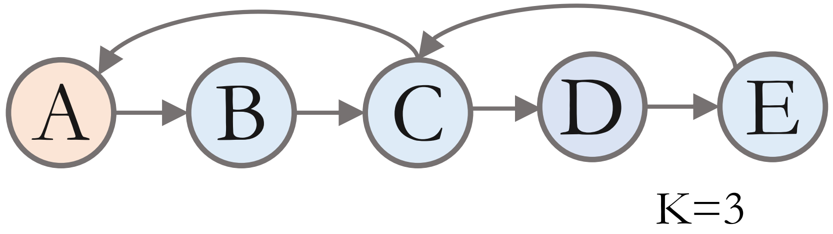
Reducibility

$$P(X_{nij} \,|X_0 = i) > 0$$



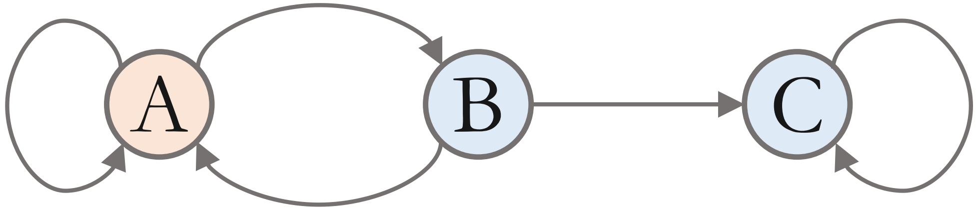
Periodicity

$$k = gcd\{n > 0 : P(X_n = i \,|X_0 = i) > 0\}$$



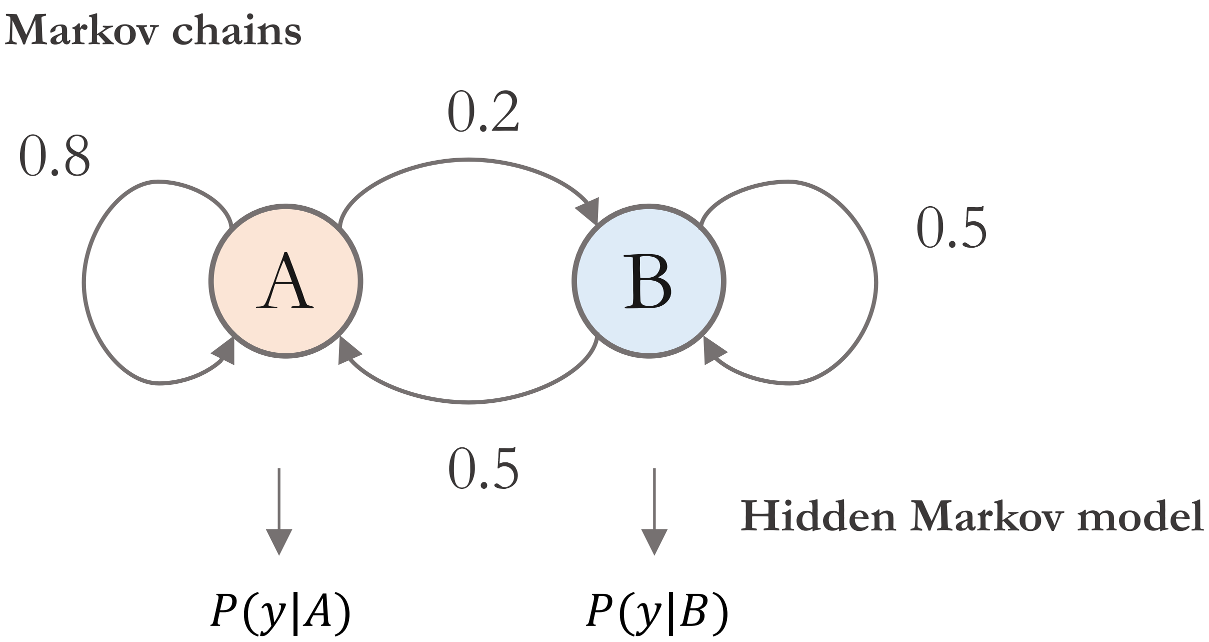
Transience

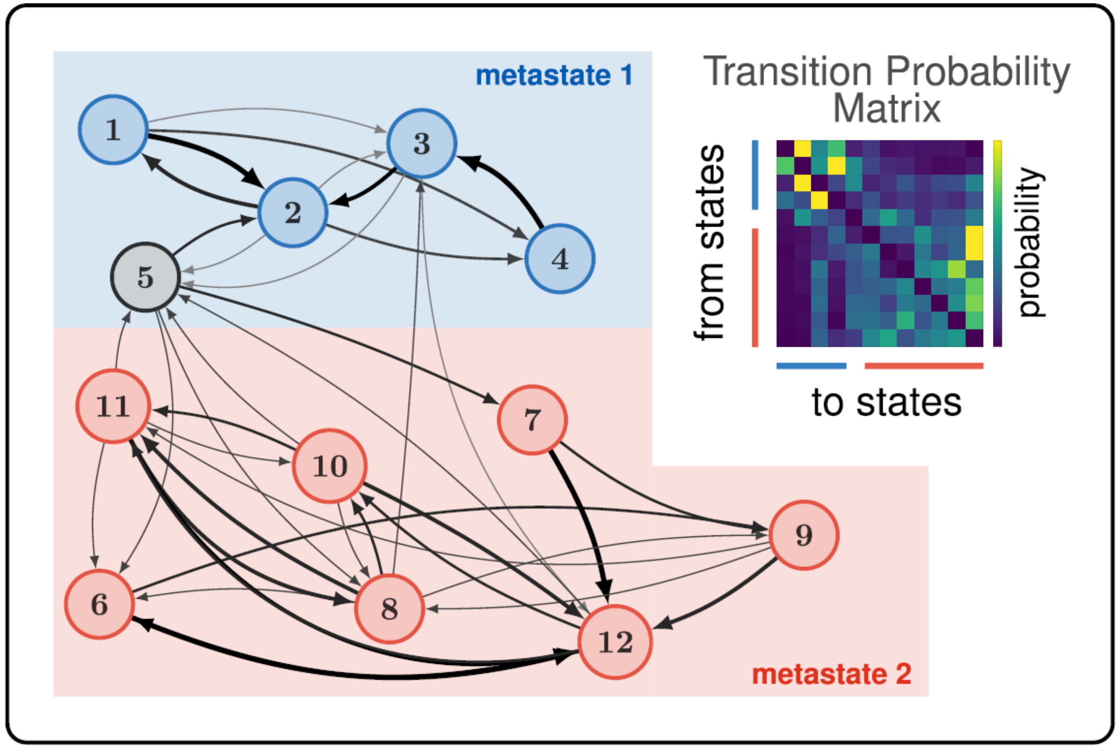
There is a non zero probability that we will never return to *i*.



If *i* is not **transient** it is **recurrent**.

Hidden Markov models





Menara et al. (2021)