Lab 1 Directed acyclic graphs (DAGs)

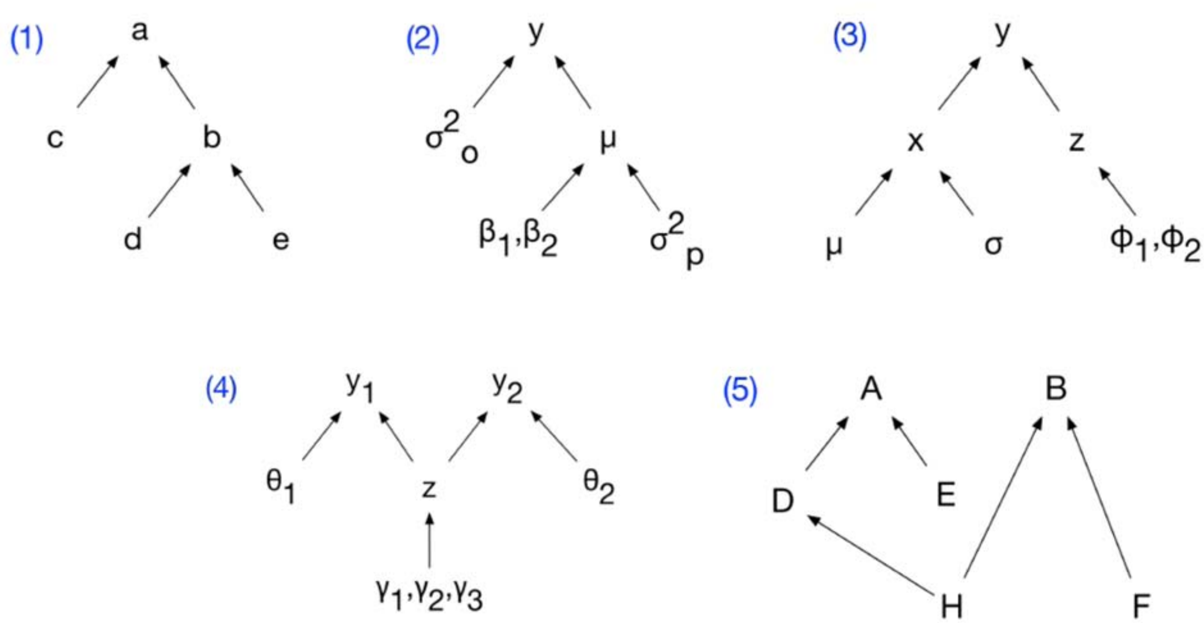
The goal of Bayesian analysis is to discover the characteristics of probability distributions that govern the behavior of random variables of interest, for example, the size of a population, the rate of nitrogen accumulation in a stream, the diversity of plants on a landscape, the change in lifetime income that occurs with changing level of education.

It follows that understanding the laws of probability and statistical distributions provides the foundation for Bayesian analysis. Keep in mind the following learning objectives

- Understand the concepts of conditional and independent random variables.
- Understand discrete and continuous marginal distributions.
- Be able to write out joint distributions of random variables given Bayesian networks (directed acyclic graphs).
- Become familiar with frequently used statistical distributions representing discrete and continuous random variables.
- Learn R functions for calculating properties of distributions and for sampling from them.
- Use moment matching, a procedure key to linking models to data in the Bayesian framework.

Problem 1: Converting DAGs to joint distributions

Write out the joint and conditional distributions for the following Bayesian networks. For discrete random variables, [A]is equivalent to Pr(A). For continuous random variables [A]is the probability density of A.



Problem 2: Converting joint distributions to DAGs

Draw Bayesian networks (DAGs) for the joint and conditional distributions below.

$Pr(A,B)=Pr(A|B)Pr(B)$

$Pr(A,B,C)=Pr(A|B,C)Pr(B|C)Pr(C)$

$Pr(A,B,C,D)=Pr(A|C)Pr(B|C)Pr(C|D)Pr(D)$

$Pr(A,B,C,D,E)=Pr(A|C)Pr(B|C)Pr(C|D,E)Pr(D)Pr(E)$

$Pr(A,B,C,D)=Pr(A|B,C,D)Pr(B|C,D)Pr(C|D)Pr(D)$

$Pr(A,B,C,D)=Pr(A|B,C,D)Pr(C|D)Pr(B)Pr(D)$

Problem 3: Simplifying

Simplify the expression below, given that z2 and z3 are independent random variables.

$$Pr(z1,z2,z3)=Pr(z1|z2,z3)Pr(z2|z3)Pr(z3)$$

Problem 4: Interpreting and factoring

The probability of a vector of observations y depends on a vector of true ecological states of interest, z, and the parameters in a data model: θd and σd. The probability of the true states z depends on the parameters in an ecological process model: θp and σp. We know that θd, θp, σp, and σd are all independent. Write out a factored expression for the joint distribution, $Pr(y,z,\theta d,\theta p,\sigma p,\sigma d)$. Drawing a Bayesian network will help.

Problem 5: Diamond's pigeons

Jared Diamond studied the distribution of fruit pigeons *Ptilinopus rivoli* and *P. solomonensis* on 32 islands in the Bismark archipelago northeast of New Guinea (Table 1). Define the event R as an island being occupied by *P. rivoli*, and the event S as an island being occupied by *P. solomonensis*. The complementary events are that an island is not occupied by *P. solomonensis* (S^c) and not occupied by *P. rivoli* (R^c).

Table 1: Data on distribution of species of fruit pigeons on islands

Status	**Number of Islands**
P. rivoli present, *P. solomonensis* absent	9
P. solomonensis present, *P. rivoli* absent	18
Both present	2
Both absent	3
Total	32

1. Fill in Table 2 to estimate the *marginal* probabilities of presence and absence of the two species. The cells show the joint probability of the events specified in the row and column. The right column and the bottom row show the marginal probabilities.

a. What is the sum of the marginal rows?
b. What is the sum of the marginal columns?
c. Why? Note, when we marginalize over R we are effectively eliminating S and vice versa.

Table 2: Estimates of marginal probabilities for island occupancy

Events	S	S^c	Marginal
R	$\Pr\left(S, R\right) =$	$\Pr\left(S^c, R\right) =$	$\Pr\left(R\right) =$
R^c	$\Pr\left(S, R^c\right) =$	$\Pr\left(S^c, R^c\right) =$	$\Pr\left(R^c\right) =$
Marginal	$\Pr\left(S\right) =$	$Pr\left(S^c\right)$	$=$

2. Use the data in Table 1 and the probabilities in Table 2 to illustrate the rule for the union of two events, the probability that an island contains either species, Pr(R∪S).

3. Use the marginal probabilities in Table 2 to calculate the probability that an island contains both species i.e., Pr(R,S), assuming that R and S are independent. Compare the results from those calculations with the probability that both species occur on an island calculated directly from the data in Table 1. Interpret the results ecologically. What is Pr(R|S)? What is Pr(S|R).

4. Based on the data in Table 1, the probability that an island is occupied by both species is 2/32 = .062. Diamond interpreted this difference as evidence of niche separation resulting for interspecific competition, an interpretation that stimulated a decade of debate. What are the conditional probabilities, Pr(R|S) and Pr(S|R)?