Modelling of uncertainty and probability: Bayesian analysis and networks

Jürgen Scheffran
CliSAP Research Group Climate Change and Security
Institute of Geography, Universität Hamburg

Models of Human-Environment Interaction
Lecture 5, May 13, 2015
Probabilistic nature of ecology

Ecology studies how states of biological systems interact with environment and how this results in temporal dynamics and spatial patterns that we observe.

State $S$: individual state such as “alive” or collection of individuals, such as “occurrence” or “local abundance, $N$”

Arrows connect states in successive time periods and govern state changes; may represent coin-flip-like survival, e.g. of demographic rates of survival, fecundity, immigration, and emigration, on which the ecological mechanisms act to determine how a population is distributed in space or evolves over time.

Hierarchical scales of organization: genes are nested within individuals, individuals within populations, populations within metapopulations or communities, and communities within metacommunities.

Source: Kery/Schaub 2012
### Ecological scales of organization

<table>
<thead>
<tr>
<th>Scale of Organization = Type of Population</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Static (State Variable)</strong></td>
<td></td>
</tr>
<tr>
<td>(1) One Site, One Species: Classic Population = Population of Individuals</td>
<td>Abundance $N$</td>
</tr>
<tr>
<td>(2) Multiple Sites, One Species: Metapopulation = Population of (Local) Populations</td>
<td>$N_s$, $z_s$, Occupancy $\psi = \Pr(N &gt; 0)$</td>
</tr>
<tr>
<td>(3) One Site, Multiple Species: Community = Population of Species</td>
<td>$N_k$, $z_k$, Species Richness</td>
</tr>
<tr>
<td>(4) Multiple Sites, Multiple Species: Metacommunity = Population of Communities</td>
<td>$N_{k,s}$, $z_{k,s}$, Species Richness</td>
</tr>
<tr>
<td><strong>Dynamic (Vital Rates)</strong></td>
<td></td>
</tr>
<tr>
<td>(1) One Site, One Species: Survival Probability ($\phi$), Recruitment Rate ($\gamma$)</td>
<td></td>
</tr>
<tr>
<td>(2) Multiple Sites, One Species: Extinction Probability ($1 - \phi_s$), Colonization Rate ($\gamma_s$), Dispersal Rates</td>
<td></td>
</tr>
<tr>
<td>(3) One Site, Multiple Species: Extinction Probability ($1 - \phi_k$), Colonization Rate ($\gamma_k$),</td>
<td></td>
</tr>
<tr>
<td>(4) Multiple Sites, Multiple Species: Extinction Probability ($1 - \phi_{k,s}$), Colonization Rate ($\gamma_{k,s}$), Dispersal Rates</td>
<td></td>
</tr>
</tbody>
</table>

Source: Kery/Schaub 2012
The estimator model

Bayesian inference

The basis for Bayesian inference is the so-called Bayes rule, attributed to the English minister and mathematician Thomas Bayes (Bayes, 1763).

Recipe for every Bayesian analysis about any uncertain quantity is quite simple and mechanical:

- What is uncertain and of interest to you? Call it \( \theta \).
- What do you know? Call it \( D \) […].
- Then calculate \( p(\theta|D) \).
- How? Using the rules of probability, nothing more, nothing less.

To describe how Bayesians learn from data using the rules of probability, we will introduce Bayes rule in the context of two sets of mutually excluding, observable events, \( A \) and \( B \). Remember that a vertical bar (\(|\)) means “conditional on” and is read as “given”.
Bayesian modeling

Bayes’ theorem: gives the probability of a random event $A$ occurring given that we know a related event $B$ occurred. This probability is noted $P(A|B)$ and is read "probability of $A$ given $B$“ (posterior).

Bayesian belief network (BBN): graphical tool for decisions under uncertainty. --Can be used to build a Decision Support System (e.g. a Bayesian Expert System).
--Bayesian networks are composed of three elements: nodes representing system variables, links representing causal relationships between nodes, and a probabilities,
--Each node specifies the belief that a node will be in a particular state given the states of those nodes that can affect it directly.

Bayesian learning: the process by which a BBN updates its set of probabilities as a result of receiving case data about variables in the table.

Conditional probability tables (CPT) specify belief that a node will be in a particular state given the states of those nodes affecting it directly (its parents).
Mathematical representation of Bayesian learning

- Assume two states of the world $s1$ and $s2$ and agents being unsure which of them is the actual or true state of the world.

- At time $t$, the $i$-th agent attaches probability $z_i(t)$ to $s2$ being the true state of the world and thus believes $s1$ to be true state with the probability $1-z_i(t)$.

- In the light of their beliefs $z_i(t)$, agents choose a particular course of action.

- They observe a result based on which they update the probabilities of $s2$ being the true state of the world.

- The Bayesian mechanism provides a plausible way in which beliefs can change over time (belief updating).
Bayes Theorem

\[ \Pr(B, A) = \Pr(B|A) \Pr(A). \]  
(1.1)

Bayes’ theorem is easily established by fiddling around with Eq. (1.1). First, by simply switching the roles of \( A \) and \( B \), we have

\[ \Pr(A, B) = \Pr(A|B) \Pr(B). \]  
(1.2)

Then, since \( \Pr(B, A) \) is the same as \( \Pr(A, B) \), we may substitute the right-hand side of Eq. (1.1) in the left-hand side of Eq. (1.2), obtaining

\[ \Pr(B|A) \Pr(A) = \Pr(A|B) \Pr(B); \]

finally, we divide both sides by \( \Pr(A) \), obtaining

\[ \Pr(B|A) = \frac{\Pr(A|B) \Pr(B)}{\Pr(A)}. \]  
(1.3)

Events \( B_i \) being mutually exclusive and exhaustive, from the basic notion of probability and (1.2) follows

\[ \Pr(A) = \sum_j \Pr(A, B_j) = \sum_j \Pr(A|B_j) \Pr(B_j). \]  
(1.4)

Bayes’ theorem is established by substituting a specific event \( B_i \) for \( B \) in (1.3), and then substituting (1.4) in the denominator of the right-hand side of (1.3). Thus Bayes’ theorem is that

\[ \Pr(B_i|A) = \frac{\Pr(A|B_i) \Pr(B_i)}{\sum_j \Pr(A|B_j) \Pr(B_j)}. \]  
(1.5)

Source: Link/Barker 2010
Bayesian modeling

- Probabilistic/statistical graphical model representing a set of random variables and conditional dependencies via **directed acyclic graph** (DAG).

- Bayesian networks are DAGs whose nodes represent random variables in the Bayesian sense: they may be observable quantities, latent variables, unknown parameters or hypotheses.

- **Edges** represent conditional dependencies; nodes that are not connected represent variables that are conditionally independent of each other.

- **Nodes** are associated with a probability function that takes (as input) the node's parent variables, and gives (as output) the probability (or probability distribution, if applicable) of the variable represented by the node.

**Example:** a Bayesian network could represent the probabilistic relationships between diseases and symptoms.

→ Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.
Graphical models and Bayesian models

- Graphical models are a marriage between probability and graph theory.
- Fundamental to the idea of a graphical model is the notion of modularity -- a complex system is built by combining simpler parts.
- Probability theory provides the glue whereby the parts are combined, ensuring that the system as a whole is consistent, and providing ways to interface models to data.
- The graph theoretic side of graphical models provides both an intuitive interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to the design of efficient general-purpose algorithms.
- Many of the classical multivariate probabilistic systems studied in fields such as statistics, systems engineering, information theory, pattern recognition and statistical mechanics are special cases of the general graphical model formalism -- examples include mixture models, factor analysis, hidden Markov models, Kalman filters and Ising models.

Source: Murphy 1998
Probabilistic graphical models are graphs in which nodes represent random variables, and the (lack of) arcs represent conditional independence assumptions. Hence they provide a compact representation of joint probability distributions.

*Undirected graphical models*, also called Markov Random Fields (MRFs) or Markov networks, have a simple definition of independence: two (sets of) nodes A and B are conditionally independent given a third set, C, if all paths between the nodes in A and B are separated by a node in C.

*Directed graphical models* also called Bayesian Networks or Belief Networks (BNs), have a more complicated notion of independence, which takes into account the directionality of the arcs, as we explain below.

Source: Murphy 1998
Conditional Probability Distribution (CPD) table: Example

Chain rule and joint probability of all nodes
\[ P(C, S, R, W) = P(C) \times P(S | C) \times P(R | C, S) \times P(W | C, S, R) \]

Conditional independence relationships:
- \( R \) is independent of \( S \) given its parent \( C \), and the last term because \( W \) is independent of \( C \) given its parents \( S \) and \( R \).

Source: Murphy 1998
Bayesian inference for the example

- Solve using Bayesian networks in probabilistic inference.
- For the water sprinkler network, we observe the fact that the grass is wet. There are two possible causes: either it is raining, or the sprinkler is on.
- We can use Bayes' rule to compute the posterior probability of each explanation (where 0==false and 1==true).

\[
\Pr(S = 1 | W = 1) = \frac{\Pr(S = 1, W = 1)}{\Pr(W = 1)} = \frac{\sum_{c,r} \Pr(C = c, S = 1, R = r, W = 1)}{\Pr(W = 1)} = \frac{0.2781}{0.6471} = 0.430
\]

\[
\Pr(R = 1 | W = 1) = \frac{\Pr(R = 1, W = 1)}{\Pr(W = 1)} = \frac{\sum_{c,s} \Pr(C = c, S = s, R = 1, W = 1)}{\Pr(W = 1)} = \frac{0.4581}{0.6471} = 0.708
\]

where

\[
\Pr(W = 1) = \sum_{c,r,s} \Pr(C = c, S = s, R = r, W = 1) = 0.6471
\]

is a normalizing constant, equal to the probability (likelihood) of the data. So we see that it is more likely that the grass is wet because it is raining: the likelihood ratio is \(0.7079/0.4298 = 1.647\).

Source: Murphy 1998
Bayesian population analysis estimation of survival from capture–recapture data

Capture and recapture is a method to estimate an animal population size. A portion of the population is captured, marked, and released. Later, another portion is captured and the number of marked individuals within the sample is counted. The number of marked individuals within the second sample should be proportional to the number of marked individuals in the whole population, an estimate of the total population size can be obtained by dividing the number of marked individuals by the proportion of marked individuals in the second sample.
Analysis of population dynamics

Ibex population dynamics, Switzerland

Source: Kery/Schaub 2012
Bayesian population analysis estimation of survival from capture–recapture data

Focus on **survival probability** as key demographic parameter and main components of population dynamics

Focus on **estimation** (what is the survival in population?) and modeling, e.g. to test whether survival changes with age or group differences

→ Estimate how strongly it varies over time or what proportion of temporal variability can be explained by an external covariate such as weather.

Survival estimation; count the number of individuals alive at time \( t \) (\( C_t \)), and keep track of how many die (\( D\Delta t \)) during a period \( \Delta t \). Alternatively count the number of people alive at \( t + \Delta t \) and survived period \( \Delta t \) (\( L\Delta t \)).

Then survival probability is

\[
    s_t = \frac{C_t - D\Delta t}{C_t} = \frac{L\Delta t}{C_t}
\]

**Detection of individuals** is usually difficult: when individual is not seen, we don’t know whether it is dead or still alive.

**Capture–recapture models:** Most common statistical method to jointly estimate recapture and survival probabilities in animal and plant

→ Cormack–Jolly–Seber (CJS) model

Source: Kery/Schaub 2012
Cross-classification of vessels fishing off the coast of New Zealand


<table>
<thead>
<tr>
<th></th>
<th>Japan</th>
<th>NZ</th>
<th>Russia</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Bycatch</td>
<td>18</td>
<td>90</td>
<td>100</td>
<td>208</td>
</tr>
<tr>
<td>Bycatch</td>
<td>1</td>
<td>6</td>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>96</td>
<td>123</td>
<td>238</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{Pr(Nation)} & = 0.080 & 0.403 & 0.517 \\
\text{Pr(Bycatch|Nation)} & = 0.053 & 0.063 & 0.187 \\
\text{Pr(Nation|Bycatch)} & = 0.033 & 0.200 & 0.767 \\
\text{Odds Ratio} & = 0.397 & 0.370 & 3.072 \\
\end{align*}
\]

\[
\frac{\text{Pr(Russian, Bycatch)}}{\text{Pr(Russian|Bycatch)}} = \frac{\frac{23}{238}}{\frac{23}{30}} = \frac{30}{238}
\]

Link/Barker 2010
Exemplary posterior density plots of survival and recapture probabilities
Example of the state and observation process of a marked individual over time for the multistate model.

Source: Kery/Schaub 2012
Capture–recapture and observation data of ringed plovers at two sites

<table>
<thead>
<tr>
<th>True State at Time $t$</th>
<th>Site A</th>
<th>Site B</th>
<th>Dead</th>
</tr>
</thead>
<tbody>
<tr>
<td>True State at Time $(t+1)$</td>
<td>$\phi_A(1 - \psi_{AB})$</td>
<td>$\phi_A\psi_{AB}$</td>
<td>$1 - \phi_A$</td>
</tr>
<tr>
<td>Site A</td>
<td>$\phi_B\psi_{BA}$</td>
<td>$\phi_B(1 - \psi_{BA})$</td>
<td>$1 - \phi_B$</td>
</tr>
<tr>
<td>Site B</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dead</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observation at Time $t$</th>
<th>Seen in A</th>
<th>Seen in B</th>
<th>Not seen</th>
</tr>
</thead>
<tbody>
<tr>
<td>True State at Time $t$</td>
<td>$p_A$</td>
<td>0</td>
<td>$1 - p_A$</td>
</tr>
<tr>
<td>Site A</td>
<td>0</td>
<td>$p_B$</td>
<td>$1 - p_B$</td>
</tr>
<tr>
<td>Site B</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dead</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Kery/Schaub 2012 p. 20
Posterior distributions of survival, movement, and recapture probabilities

Red lines give the values of the data generating parameters

Source: Kery/Schaub 2012
Nodes show two age classes (\(N1\): 1-year-old females; \(N\text{ad}\): females older than 1 year) and the arrows the transition probabilities based on vital rates (\(S\text{juv}\): juvenile survival; \(S\text{ad}\): adult survival; \(f\): productivity).

Source: Kery/Schaub 2012
Graphical representation of different integrated population models

Small squares represent data, circles parameters (blue: target parameters, green: nuisance parameters), large squares individual submodels, arrows flux of information. Circles in two submodels indicate that they are informed from two data sources.

Source: Kery/Schaub 2012
Posterior means of demographic parameters in a hoopoe population in SW Switzerland

Source: Kery/Schaub 2012
Posterior means of demographic parameters in hoopoe population in SW Switzerland

Source: Kery/Schaub 2012
A simple Bayesian belief network

Conditional probability table (CPT) of an imaginary stakeholder.

<table>
<thead>
<tr>
<th>Intensity of industrial fishing</th>
<th>Status of seafloor habitats:</th>
<th>Sustainability of fish stocks:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Good</td>
<td>0.60 0.40 0.00</td>
</tr>
<tr>
<td>Low</td>
<td>Poor</td>
<td>0.00 0.10 0.90</td>
</tr>
<tr>
<td>High</td>
<td>Good</td>
<td>0.40 0.60 0.00</td>
</tr>
<tr>
<td>High</td>
<td>Poor</td>
<td>0.00 0.00 1.00</td>
</tr>
</tbody>
</table>

Source: Stoll/Welp 2006
Advantages and challenges of Bayesian analyses

Advantages:
- Even difficult models can be fit, beyond the classical framework.
- Derived quantities may be computed trivially easily, with full propagation of the uncertainty in the components that make up the derived quantity. This can be a very hard problem in the classical framework.
- Results are exact, no asymptotics in the estimates, which may be of questionable value in small-data situations so typical of ecological studies.
- Software tools allows quantitative ecologist to understand construction of complex models so that the code can be modified to fit one’s own purposes.

Challenges:
- Bayesian statistics may appear difficult.
- Choice of priors and sensitivity of estimates need some thinking.
- Solution engines are blackboxes and are hard to understand, leaving a certain uneasiness with people who like to understand most of what they do.
- Convergence of Markov chains may be difficult to assess.
- Analyses can be slow compared to other ways of model fitting.