Laws of Probability

Bayesian Modeling for Socio-Environmental Data

Mary B. Collins

August 2016

Outline/Learning Outcomes

- Uncertainty
- Rules of probability
- Factoring joint probabilities
- Marginal distributions

Uncertainty

- Stochastic things are uncertain

sto·chas·tic

/stəˈkastik/

adjective

randomly determined; having a random probability distribution or pattern that may be analyzed statistically but may not be predicted precisely.

Uncertainty

- Bayesian methods are the only branch of statistics that treat all unobserved quantities as random.

Uncertainty

- Bayesian methods are the only branch of statistics that treat all unobserved quantities as random.
- A random variable is a quantity governed by chance (they arise from a probabily distribution).

Sources of Uncertainty

1. Process variance: *error associated with the model itself*

Sources of Uncertainty

1. Process variance: *error associated with the model itself*
2. Observation variance: *imperfect sampling, measurement*

Sources of Uncertainty

1. Process variance: *error associated with the model itself*
2. Observation variance: *imperfect sampling, measurement*
3. Variation among individuals: *genetics, environmental variation*

Sources of Uncertainty

1. Process variance: *error associated with the model itself*
2. Observation variance: *imperfect sampling, measurement*
3. Variation among individuals: *genetics, environmental variation*
4. Model selection uncertainty: *inference is conditional on a model*

Rules of probability

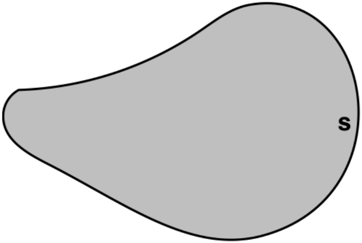
- All random variables have probabilty distributions, which are governed by rules that determine how we gain insight.

Rules of probability

- All random variables have probabilty distributions, which are governed by rules that determine how we gain insight.
- A Bayesian framework treats all unobserved quantities as random variables.

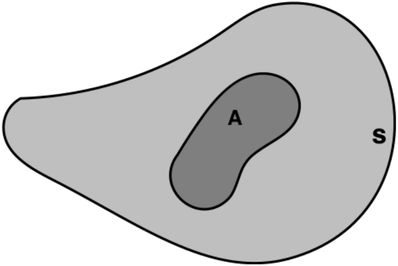
S=Sample Space

- The set of all possible outcomes of an experiment.
- The sample space has a specific area.

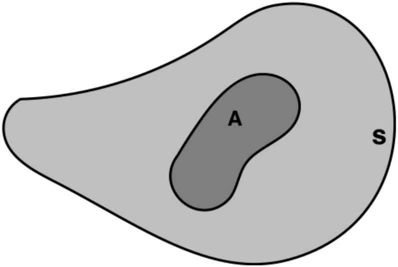


Events in S

- Event A is a set of outcomes with a specific area.
- The area of event A is less than the area of the sample space.



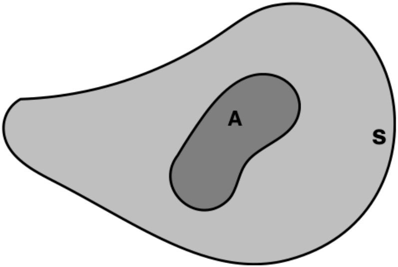
What is the probability of event A?



$\Pr(A) = \frac{\text{Area of} A}{\text{Area of } S}$

What is the probability of event A?

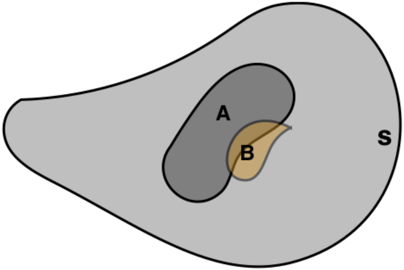
- Area of A = 4
- Area of S = 20



$$\Pr(A) = \frac{\text{Area A}}{\text{Area S}} = \frac{4}{20} = .2$$

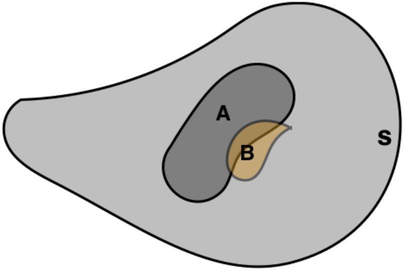
Conditional Probability

What about the case where: when event A happens we learn something about another event, B?



What is the probability of a new event B, given we know that event A has occurred?

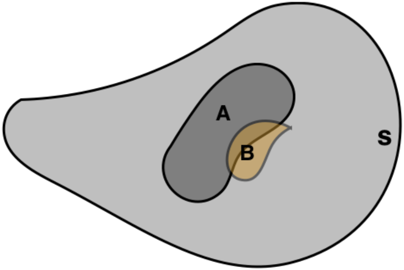
- Area A = 4
- Area B = 2
- Area A $\cap$ B =1
- Area S = 20



$\Pr(B|A)$ = prob' of B conditional on knowing A has occurred

What is the probability of a new event B, given we know that event A has occurred?

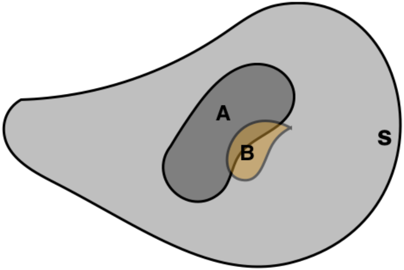
- Area A $=$ 4
- Area B $=$ 2
- Area A $\cap$ B $=$1
- Area S $=$ 20



$$Pr(B|A) = \frac{\text{Joint Prob}}{\text{Prob of A}} = \frac{\Pr(A \cap B)}{\Pr(A)} = \frac{\Pr(A,B)}{\Pr(A)}$$

What is the probability of a new event B, given we know that event A has occurred?

- Area A = 4
- Area B = 2
- Area A $\cap$ B =1
- Area S = 20



$$\Pr(B|A) = \frac{\Pr(A,B)}{\Pr(A)} = \frac{.05}{.2} = .25$$

We will make lots and lots of use of the rearrangement of this equation

$\Pr(B|A) = \frac{\Pr(A,B)}{\Pr(A)}$

$$\Pr(A,B) = \Pr(A|B)\Pr(B) \text{ and equivalently,} \tag{1}$$
$$\Pr(A,B) = \Pr(B|A)\Pr(A) \tag{2}$$

Conditional probabilty

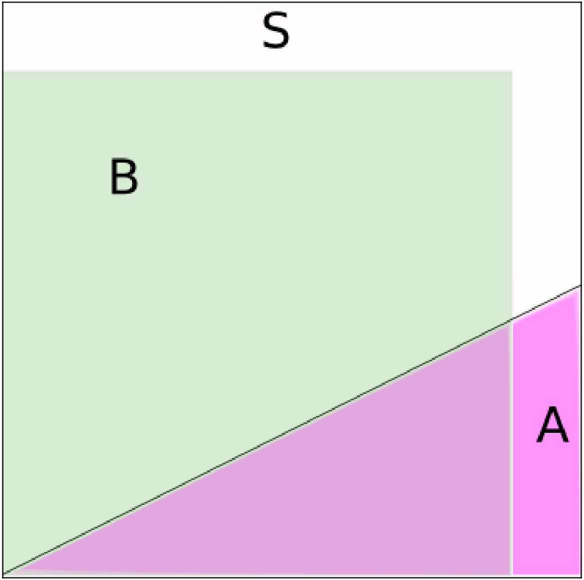
True or False?
$\Pr(B|A) = \Pr(A|B)$

What happens if the occurrence on event A doesn't tell us anything about the occurrence of event B?

In this case, events A and B are said to be ***independent***

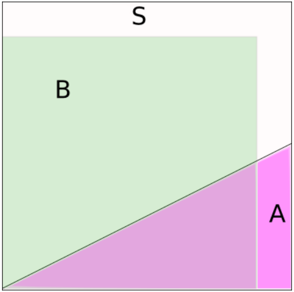
Events are independent if and only if. . .

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


S
B
A

Assuming independence, what is the joint probability of event A and event B?

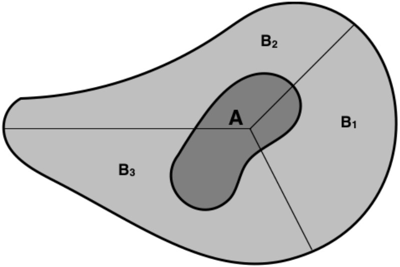
- Area A = 4
- Area B = 3
- Area A $\cap$ B =.6
- Area S = 20



$$\Pr(A, B) = \Pr(A)\Pr(B) = \left(\frac{4}{20}\right)\left(\frac{3}{20}\right) = 0.03$$

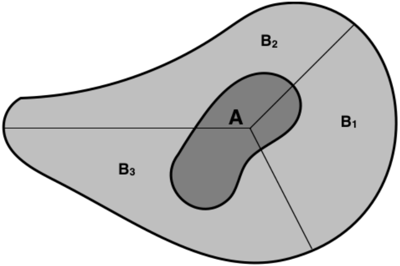
What happens when S (the sample space) is a non-overlapping group of events?

The Law of Total Probability



We can define a set of events $\{B_n : n = 1, 2, 3, ...\}$, which taken together define the entire sample space, $\sum_n B_n = S$.

What is the probability of event A?



$\Pr(A) = \sum_n \Pr(A|B_n)\Pr(B_n)$ (discrete case)
$\Pr(A) = \int \Pr(A|B)\Pr(B)dB$ (continuous case)

Board work on factoring joint probabilities and simplification using knowledge of independence

The Chain Rule

In probability theory, the chain rule (also called the general product rule) permits the calculation of any member of the joint distribution of a set of random variables using only conditional probabilities.

$$\Pr(z_1, z_2, ..., z_n) = \Pr(z_n|z_{n-1}, ..., z_1) ... \Pr(z_3|z_2, z_1) \Pr(z_2|z_1) \Pr(z_1)$$

Notice the pattern here.

- z's can be scalars or vectors.
- Sequence of conditioning doesn't matter.
- When we build models, we choose a sequence that makes sense.

Factoring joint probabilities

- The rules of probability allow us to take complicated joint distributions of random variables and break then down into chunks.

Factoring joint probabilities

- The rules of probability allow us to take complicated joint distributions of random variables and break then down into chunks.
- Chunks can then be analyzed one at a time as if all other random variables were known and constant.

Factoring joint probabilities

- The rules of probability allow us to take complicated joint distributions of random variables and break then down into chunks.
- Chunks can then be analyzed one at a time as if all other random variables were known and constant.
- Provide a usable graphical and mathematical foundation, which is *critical* for accomplishing the model specification step in the general modeling process.

Consider a Bayesian Network (represented by a directed acyclic graph or DAG)

- Bayesian networks specify how joint distributions are factored into conditional distributions using nodes to represent RV's and arrows to represent dependencies.

Consider a Bayesian Network (represented by a directed acyclic graph or DAG)

- Bayesian networks specify how joint distributions are factored into conditional distributions using nodes to represent RV's and arrows to represent dependencies.
- Nodes at the heads of arrows *must* be on the left hand side of the conditioning symbols;

Consider a Bayesian Network (represented by a directed acyclic graph or DAG)

- Bayesian networks specify how joint distributions are factored into conditional distributions using nodes to represent RV's and arrows to represent dependencies.
- Nodes at the heads of arrows *must* be on the left hand side of the conditioning symbols;
- Nodes at the tails of arrows are on the right hand side of the conditioning symbols.

Consider a Bayesian Network (represented by a directed acyclic graph or DAG)

- Bayesian networks specify how joint distributions are factored into conditional distributions using nodes to represent RV's and arrows to represent dependencies.
- Nodes at the heads of arrows *must* be on the left hand side of the conditioning symbols;
- Nodes at the tails of arrows are on the right hand side of the conditioning symbols.
- Any node at the tail of an arrow without and arrow leading into it must be expressed unconditionally.

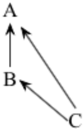
Factoring with DAGs

$\Pr(A, B) =$

Factoring with DAGs

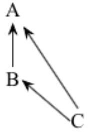
$\Pr(A, B) = \Pr(A|B)\Pr(B)$

Factoring with DAGs



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

Factoring with DAGs



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

Generalizing

$$\Pr(z_1, \ldots, z_n) = \prod_{i=1}^{n} \Pr(z_i | \{P_i\})$$

$\{P_i\}$ is the set of parents of node z_i

Marginal distributions

The marginal distribution of a subset of a collection of random variables is the probability distribution of the random variables contained in the subset.

	x_1	x_2	x_3	x_4	$p_y(Y)\downarrow$
y_1	$4/32$	$2/32$	$1/32$	$1/32$	$8/32$
y_2	$2/32$	$4/32$	$1/32$	$1/32$	$8/32$
y_3	$2/32$	$2/32$	$2/32$	$2/32$	$8/32$
y_4	$8/32$	0	0	0	$8/32$
$p_x(X) \rightarrow$	$16/32$	$8/32$	$4/32$	$4/32$	$32/32$

Joint and marginal distributions of a pair of discrete, random variables X,Y having nonzero mutual information I(X; Y). The values of the joint distribution are in the 4×4 square, and the values of the marginal distributions are along the right and bottom margins.

Marginal Distributions Example

Consider two discrete random variables that are jointly distributed such that, $[x, y] \equiv \Pr(x, y)$

Marginal Distributions Example

We are studying a species for which births occur in pulses. We observe 100 females and record the age of each animal and the number of offspring produced.

$x =$ Age	$y =$ Number offspring 1	2	3	$\sum_y [x, y]$
1	0.1	0	0	0.1
2	0.13	0.12	0.02	0.27
3	0.23	0.36	0.04	0.63
$\sum_x [x, y]$	0.46	0.48	0.06	

When we have a joint distribution of two random variables, we can focus on one by summing over the probabilities of the other.

Marginal Distributions Example

- We care about marginal distributions because they allow us to represent the univariate distribution of unknown quantities that are parts of joint distributions.
- They are a vital tool for simplification.