Bayes’ rule, Monte Carlo Simulation and Markov Chain

Outline

- Bayes' rule
- Monte Carlo Simulation
- Markov Chain
- Markov Chain Monte Carlo (MCMC)

1 Bayes Theorem

- Bayes' theorem uses the concept of conditional probability.

$$\begin{aligned}\mathrm{P(A\ and\ B)} &= \mathrm{P(A)*P(B\ after\ A)} = \mathrm{P(A)*P(B|A)} \\ &= \mathrm{P(B)*P(A\ after\ B)} = \mathrm{P(B)*P(A|B)}\end{aligned}$$

$$\Downarrow$$

$$P[B|A] = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$$

1 Bayes Theorem

- Bayes' theorem uses the concept of conditional probability.

$$P(A \text{ and } B) = P(A)*P(B \text{ after } A) = P(A)*P(B|A)$$
$$= P(B)*P(A \text{ after } B) = P(B)*P(A|B)$$

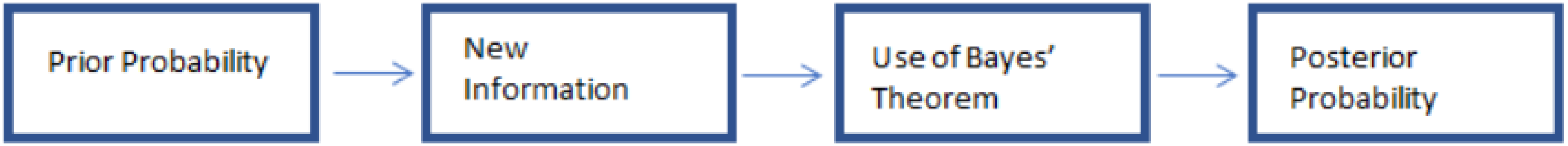
$$\Downarrow$$

$$P[B|A] = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$$

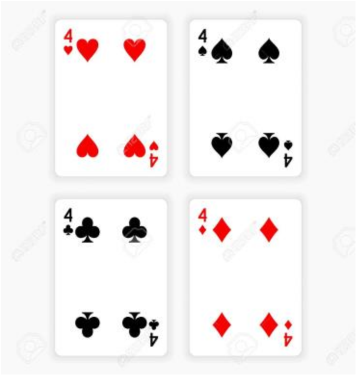
$$P[B|A] = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)\,P(B)}{P(A)}$$

- P(B) is the **prior probability** or marginal probability of B. It is ”prior” in the sense that it does not take into account any information about A.
- P(B|A) is the conditional probability of B, given A. It is also called the **posterior probability** because it is derived from or depends upon the specified value of A.
- P(A|B) is the conditional probability of A given B.
- P(A) is the prior or marginal probability of A, and **acts as a normalizing constant**.

- Bayes’ theorem calculates the posterior probability of a new event using a prior probability of some events.
- Bayes’ theorem, also calculates the probability of some future events.



B is red card, A is four card



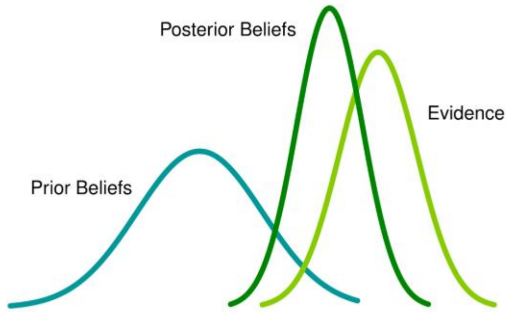
- **Marginal probability**: the probability of an event occurring (p(A)), it may be thought of as an unconditional probability. It is not conditioned on another event. Example: the probability that a card drawn is red (p(red) = 0.5). Another example: the probability that a card drawn is a 4 (p(four)=1/13).
-
- **Joint probability**: p(A and B). The probability of event A **and** event B occurring. It is the probability of the intersection of two or more events. The probability of the intersection of A and B may be written p(A ∩ B). Example: the probability that a card is a four and red =p(four and red) = 2/52=1/26. (There are two red fours in a deck of 52, the 4 of hearts and the 4 of diamonds).
-
- **Conditional probability**: p(A|B) is the probability of event A occurring, given that event B occurs. Example: given that you drew a red card, what's the probability that it's a four (p(four|red))=2/26=1/13. So out of the 26 red cards (given a red card), there are two fours so 2/26=1/13.

Bayes's Theorem: Let A be any event. Then for any $1 \leq k \leq K$ we have

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{P(A)} = \frac{P(A \mid B_k)P(B_k)}{\sum_{j=1}^{K} P(A \mid B_j)P(B_j)}.$$

Of course there is also a continuous version of Bayes's Theorem with sums replaced by integrals.

Bayes's Theorem provides us with a simple rule for updating probabilities when new information appears
- in Bayesian modeling and statistics this new information is the observed data
- and it allows us to update our prior beliefs about parameters of interest which are themselves assumed to be random variables.



Let $\boldsymbol{\theta}$ be some unknown parameter vector of interest. We assume $\boldsymbol{\theta}$ is random with some distribution, $\pi(\boldsymbol{\theta})$

- this is our prior distribution which captures our prior uncertainty regarding $\boldsymbol{\theta}$.

There is also a random vector, $\mathbf{X}$, with PDF (or PMF) $p(\mathbf{x} \mid \boldsymbol{\theta})$

- this is the likelihood.

The joint distribution of $\boldsymbol{\theta}$ and $\mathbf{X}$ is then given by $p(\boldsymbol{\theta}, \mathbf{x}) = \pi(\boldsymbol{\theta})p(\mathbf{x} \mid \boldsymbol{\theta})$

- we can integrate to get the marginal distribution of $\mathbf{X}$

$$p(\mathbf{x}) = \int_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta})p(\mathbf{x} \mid \boldsymbol{\theta})\, d\boldsymbol{\theta}$$

Bayes's Theorem: Let A be any event. Then for any $1 \leq k \leq K$ we have

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{P(A)} = \frac{P(A \mid B_k)P(B_k)}{\sum_{j=1}^{K} P(A \mid B_j)P(B_j)}.$$

We can compute the posterior distribution via Bayes's Theorem:

$$\pi(\boldsymbol{\theta} \mid \mathbf{x}) = \frac{\pi(\boldsymbol{\theta})p(\mathbf{x} \mid \boldsymbol{\theta})}{p(\mathbf{x})} = \frac{\pi(\boldsymbol{\theta})p(\mathbf{x} \mid \boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta})p(\mathbf{x} \mid \boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

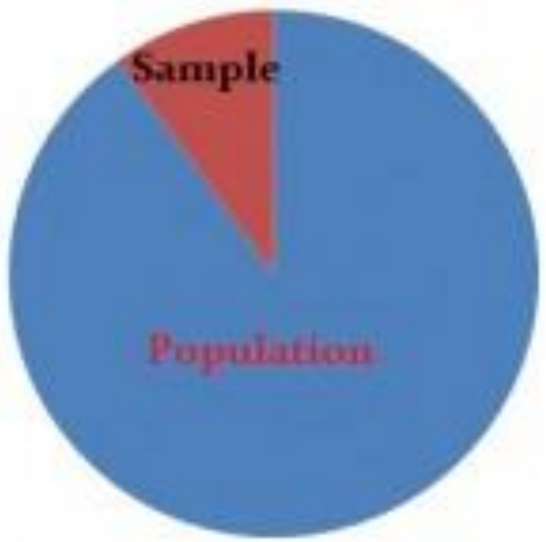
$B_1 \ldots B_{K,}$ population and B_k is a subset in population

(1)

2 Monte Carlo Simulation

- **Monte Carlo simulation** (also called the *Monte Carlo Method* or *Monte Carlo sampling*) is a way to account for risk in decision making and quantitative analysis. The method **finds all possible outcomes of your decisions and assesses the impact of risk.**

- You'll come across many terms in statistics that define **different sampling methods**: simple random sampling, systematic sampling, stratified random sampling and cluster sampling. How to tell the difference between the different sampling methods can be a challenge.



Different Sampling Methods: How to Tell the Difference: Steps

- **Step 1:** Find out if the study sampled from individuals (for example, picked from a pool of people). You'll find **simple random sampling** in a school lottery, where individual names are picked out of a hat. But a more "systematic" way of choosing people can be found in "systematic sampling," where every nth individual is chosen from a population. For example, every 100th customer at a certain store might receive a "doorbuster" gift.
- **Step 2:** Find out if the study picked groups of participants. For large numbers of people (like the number of potential draftees in the Vietnam war), it's much simpler to pick people by groups (**simple random sampling**). In the case of the draft, draftees were chosen by birth date, "simplifying" the procedure.
- **Step 3:** Determine if your study contained data from more than one carefully defined group ("strata" or "cluster"). Some examples of strata could be: Democrats and Republics, Renters and Homeowners, Country Folk vs. City Dwellers, Jacksonville Jaguars fans and San Francisco 49ers fans. If there are two or more very distinct, clear groups, you have a **stratified sample** or a "cluster sample."
- If you have data about the individuals in the groups, that's a stratified sample. In order to perform stratified sampling on this sample, you could perform random sampling of each strata independently.
- If you only have data about the groups themselves (you may only know the location of the individuals), then that's a **cluster sample**.
- **Step 4:** Find out if the sample was easy to get. **Convenience samples** are like convenience stores: why go out of your way to get samples, when you can nip out to the corner store? A classic example of convenience sampling is standing at a shopping mall, asking passers by for their opinion.

Quantified Probability and Real-Life Uses

- Analyzing radiative heat transfer problems (Wang et.al),
- Estimating the transmission of particles through matter (Biersack & Haggmark),
- Calculating the probability of cost overruns in large projects (McCabe),
- Foreseeing where prices of securities are likely to move (Boyle et. al),
- Analyzing how a network or electric grid will perform in different scenarios. For example, Sortomme et. al ran simulations for how electric vehicle charging will affect the electric drig in the future.
- Assessing risk for credit or insurance (Gordy).
- Simulating proteins in biology (Earl et. al)

Accuracy

- While a Monte Carlo simulation provides some good accuracy, it is unlikely to hit the “exact” mark for several reasons:
- Vast amounts of data are usually involved.
- There are usually several unknowns in the system.
- As it is probabilistic (i.e. randomness plays a role in predicting future events), there will always be a margin of error related to the results.

In fact, **it can be quite easy to run a “bad” Monte Carlo simulation** (Brandimarte, 2014). This can happen for a variety of reasons, including:

- Use of an incorrect model or an unrealistic probability distribution,
- The underlying risk factors aren’t complete (i.e. you haven’t specified them well enough),
- The choice of Monte Carlo (which uses a stochastic model) isn’t suited to your data,
- The random number generator chosen for the method isn’t good enough,
- Computer bugs, which you may not be aware of if your area of expertise is statistics (as opposed to programming).

3 Markov Chain

Dependent Models: Markov Chain

- In many cases we can assume that the random value depends only on recent past values, instead of the entire history. For example, the rainfall today may depend on the rainfall in the past several days, but we could model it assuming that is only dependent on whether it rained yesterday. This is the basis for using **Markov** models.

- Consider time steps separated by Δt and denote a time sequence as $0 \times \Delta t$, $1 \times \Delta t$, $2 \times \Delta t$, …, $n \times \Delta t$, … or simply 0, 1, 2, …, n,… for short. Note that any t is given by $t = n \times \Delta t$.

- We stated above that the Markov property is memoryless. It implies that the future movement of the object depends only on its present state. And that is the most important concept to understand.

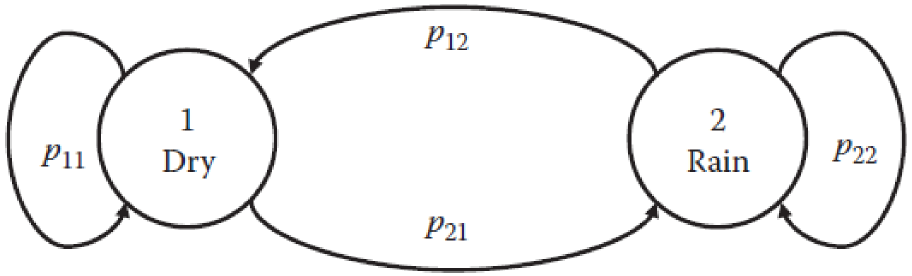
Markov Property Is Memoryless

- This brings us to the concept of Markov chains.

A Markov chain is a random process that has a Markov property

A Markov model consists of a transition probability matrix **P** that projects a vector of probabilities **X**(*t*) through time. Each entry of the vector **X**(*t*) is a probability of being in a given state and therefore the entries of **X** are the values of a pmf. The entries of **P** are probabilities of transition from state to state. Therefore,

$$\mathbf{X}(t) = \mathbf{P}\mathbf{X}(t-1) \quad (11.1)$$



A Markov chain diagram for the sequence of dry or rainy days.

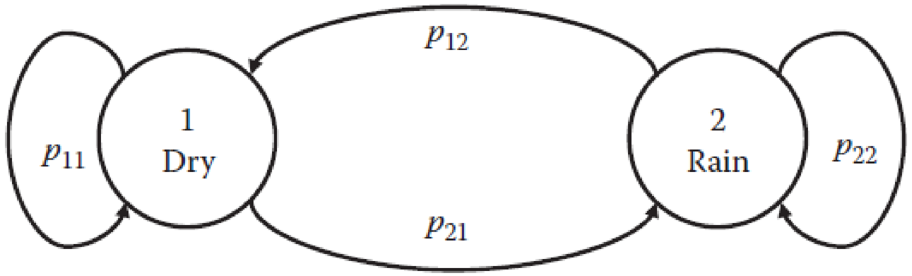
For example, tomorrow is Dry Day $X_{1,\,t+1}$, if tomorrow is Wet day, then $X_{2,t+1}$;
Today, the Dry day probability is $X_{1,\,t}$, and Wet day probability is $X_{2,t}$
$X_{1,\,t}+X_{2,\,t}=1$
(Each day, the system can be in either a dry or a wet state)

$X_{1,t+1}= P_{11} *X_{1,t}+ P_{21} *X_{2,\,t}$
P_{11} is the probability when today Dry day and tomorrow also dry day,;
P_{21} is the probability when today Wet day, but tomorrow dry day.
$X_{2,t+1}=P_{12}*X_{1,t}+Pc*X_{2,t}$
P_{22} is the probability when today Wet day and tomorrow wet day;
P_{12} is the probability when today Dry day, but tomorrow wet day.

A Markov chain is a random process that has a Markov property

A Markov model consists of a transition probability matrix **P** that projects a vector of probabilities **X**(t) through time. Each entry of the vector **X**(t) is a probability of being in a given state and therefore the entries of **X** are the values of a pmf. The entries of **P** are probabilities of transition from state to state. Therefore,

$$\mathbf{X}(t) = \mathbf{P}\mathbf{X}(t-1) \tag{11.1}$$



A Markov chain diagram for the sequence of dry or rainy days.

For example, tomorrow is Dry Day $X_{1,\ t+1}$, if tomorrow is Wet day, then $X_{2,t+1}$;
Today, the Dry day probability is $X_{1,\ t}$, and Wet day probability is $X_{2,t}$
$X_{1,\ t}+X_{2,\ t}=1$
(Each day, the system can be in either a dry or a wet state)

$X_{1,t+1}= P_{11} *X_{1,t}+ P_{21} *X_{2,\ t}$
P_{11} is the probability when today Dry day and tomorrow also dry day,;
P_{21} is the probability when today Wet day, but tomorrow dry day.
$X_{2,t+1}=P_{12}*X_{1,t}+Pc*X_{2,t}$
P_{22} is the probability when today Wet day and tomorrow wet day;
P_{12} is the probability when today Dry day, but tomorrow wet day.

Each day, the system can be in either a dry or a wet state.

Dry is 1 and Wet is 2

$X_1^* + X_2^*$ =1 (neither today or tomorrow)

- If today is Dry day (X_1^*), tomorrow

Dry day, P11 (Dry to Dry)

Wet day, P21 (Dry to Wet)

P11+P21=1

- If today is Wet day (X_2^*), , tomorrow

Dry day, P12 (Wet to Dry)

Wet day, P22 (Wet to Wet)

P21+P22=1

Tomorrow P11 Today

$$\begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix} = \begin{bmatrix} 1-p_{21} & p_{12} \\ p_{21} & 1-p_{12} \end{bmatrix} \begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix}$$

P22

- An example: consider that at a site, as a long-term average there are about 50% percent of rainy days in a year, and 20% of the dry days are followed by rainy days.
- Question: How often are dry days followed by dry days and how often are wet days followed by dry days?
- Solution: In this case 50% of rainy days in a year means that $p_2^*=0.5$, whereas 20% of the dry days are followed by rainy days means that $p_{21} = 0.2$.

$$p_2^* = \frac{p_{21}}{p_{21} + p_{12}} = \frac{0.2}{0.2 + p_{12}} = 0.5$$

- Rearrange and then solve for p_{12} from $p_{12} + 0.2 = 0.2/0.5=0.4$ to get $p_{12} = 0.2$, then $p_{11} = 1 - p_{21} = 0.8$ and $p_{22}= 1-p_{21}=0.8$, $p_1^*=0.5$
- The answer is that 80% of the time dry days are followed by dry days.
- The answer is that 20% of the time wet days are followed by dry days.

$$\begin{array}{ccc} & P11 & \\ \begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix} = & \begin{bmatrix} 1-p_{21} & p_{12} \\ p_{21} & 1-p_{12} \end{bmatrix} & \begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix} \\ & P22 & \end{array}$$

and $X_1^* + X_2^* = 1$ ⇒

$$X_1^* = \frac{p_{12}}{p_{21}+p_{12}} \quad \text{and} \quad X_2^* = \frac{p_{21}}{p_{12}+p_{21}}$$

- An example: consider that at a site, as a long-term average there are about 50% percent of rainy days in a year, and 20% of the dry days are followed by rainy days.
- Question: How often are dry days followed by dry days and how often are wet days followed by dry days?
- Solution: In this case 50% of rainy days in a year means that $p_2^*=0.5$, whereas 20% of the dry days are followed by rainy days means that $p_{21} = 0.2$.

$$p_2^* = \frac{p_{21}}{p_{21}+p_{12}} = \frac{0.2}{0.2+p_{12}} = 0.5$$

- Rearrange and then solve for p_{12} from $p_{12} + 0.2 = 0.2/0.5=0.4$ to get $p_{12} = 0.2$, then $p_{11} = 1 - p_{21} = 0.8$ and $p_{22}= 1-p_{21}=0.8$, $p_1^*=0.5$
- The answer is that 80% of the time dry days are followed by dry days.
- The answer is that 20% of the time wet days are followed by dry days.

4 What is MCMC and when would you use it?

- MCMC is simply an algorithm for sampling from a distribution.
- It’s only one of many algorithms for doing so. The term stands for “Markov Chain Monte Carlo”, because it is a type of “Monte Carlo” (i.e., a random) method that uses “Markov chains” (we’ll discuss these later). MCMC is just one type of Monte Carlo method, although it is possible to view many other commonly used methods as simply special cases of MCMC.