Project: Step 11.

N_{ij} = # 2-step transitions $i \to * \to j$ in data set.

$N_i = \sum_j N_{ij}$ = # 2-step transitions $i \to * \to *$ in data set.

$M_{ij} = N_i \left((\hat{P})^2 \right)_{ij}$

$= N_i \sum_k \hat{P}_{ik} \hat{P}_{kj}$

$\hat{P}$ = empirical 1-step transition matrix for data set (Step 6).

N_{ij} : observed frequency $i \to * \to j$

M_{ij} : expected frequency $i \to * \to j$.

Outcomes	$1 \to * \to 1$	$1 \to * \to 2$ ...	$i \to * \to j$ ...
Observed	N_{11}	N_{12} ...	N_{ij} ...
Expected	M_{11}	M_{12} ...	M_{ij} ...

Goodness of fit test:

$$d = \sum_{ij} \frac{\left((\text{Observed})_{ij} - (\text{Expected})_{ij}\right)^2}{(\text{Expected})_{ij}}$$

H_0: $\hat{P}$ is good

H_1: $\hat{P}$ is not good.

Reject H_0 if $d > \chi^2_{1-\alpha,\, df.}$

df = number of degrees of freedom.

= number of parameters to be estimated

− number of constraints.

parameters to estimate = $\#\{i \to * \to j\}$

$= m^2$ where m is the number of states in chain

this is reduced if any cells are combined because expected frequencies are too small.

constraints = number of rows in the 2-step matrix = m.

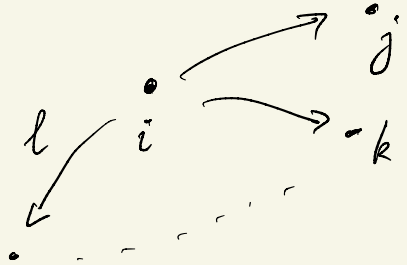
because the row sums of N_{ij} and M_{ij} must be equal.

m equations → $\sum_j N_{ij} = \sum_j M_{ij}$ all i.

Step 10.

$\hat{p}_{ij}$ = empirical transition matrix

= prob. $(i \to j)$



m states.

$\hat{P}$ m×m matrix.

Simulate Markov chain:

generate sequence: $S_0 \to S_1 \to S_2 \to S_3 \to \cdots \to S_N$

S_0: synthetic time series

(N = size of data set)

S_0: initial state

S_1: state at $t=1$

S_2: state at $t=2$.

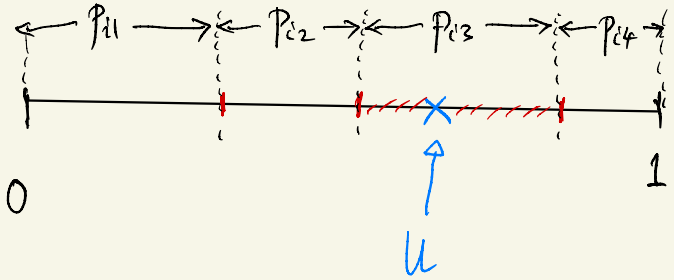
⋮

At each step, generate a random variable in set of states, $X_n \in \{1, 2, \ldots, m\}$

Prob. dist. : current state is i, then

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \hat{P}_{ij}$$

prob. to jump next to j, given that you are in state i.



$m = 4$:

$P_{i1}, P_{i2}, \ldots, P_{i4}$

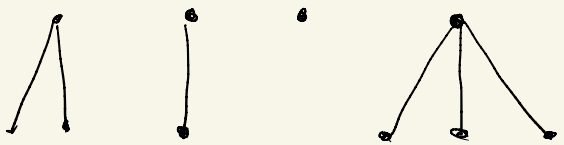
$P_{i1} + P_{i2} + P_{i3} + P_{i4} = 1$.

Generate $U \sim$ uniform on $(0,1]$.

$\mathbb{P}(U$ lies in P_{i3} interval$) = \dfrac{P_{i3}}{1} = P_{i3}$.

In this case $X_{n+1} = 3$.

Branching Process.



X_n = pop. after n steps

$\downarrow$

X_{n+1}

$$X_{n+1} = \sum_{i=1}^{X_n} Z_i$$

Z_i = number of offspring of i^{th} individual.

$X_n = 0$ is the absorbing state. = extinction.

ρ = "rho" = prob. to reach extinction.

$= \mathbb{P}(X_n = 0 \text{ for some } n)$.

Theorem: define $m = \mathbb{E}[Z]$ = mean number of offspring.

① Then if $m \leq 1 \Rightarrow \rho = 1$ (extinction certain).

② If $m > 1 \Rightarrow \rho$ is the smallest positive solution of the equation

$$\phi(s) = s$$

where

$$\phi(s) = \mathbb{E}[s^Z]$$

$$= p_0 + sp_1 + s^2 p_2 + s^3 p_3 + \cdots$$

and $p_k = \mathbb{P}(Z = k)$.

Example 1

Suppose that the number of offspring has the pdf

Z	0	1	2	3
prob.	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{12}$	$\frac{1}{12}$

Compute ρ.

First find $m = \mathbb{E}[Z]$

$$= \frac{1}{2} + 2\left(\frac{1}{12}\right) + 3\left(\frac{1}{12}\right)$$

$$= \frac{11}{12}$$

since $m < 1 \Rightarrow \rho = 1$

Example 2

Suppose the pdf of Z is

Z	0	1	2	3
prob	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{12}$	$\frac{1}{6}$

Compute ρ.

First find $m = \mathbb{E}[Z]$

$$= \frac{1}{2} + 2\left(\frac{1}{12}\right) + \left(\frac{2}{6}\right)$$

$$= \frac{7}{6}$$

since $m > 1 \Rightarrow \rho < 1$.

Solve $\phi(s) = s$.

$$\frac{1}{4} + \frac{1}{2}s + \frac{1}{12}s^2 + \frac{1}{6}s^3 = s.$$

$$\Leftrightarrow 2s^3 + s^2 - 6s + 3 = 0.$$

Since $\phi(1) = \mathbb{E}[1^Z] = 1$, we know

$s=1$ is always a solution.

$$2s^3 + s^2 - 6s + 3 = (s-1)(2s^2 + 3s - 3)$$

$$= 0$$

$$\Rightarrow 2s^2 + 3s - 3 = 0.$$

$$s = \frac{-3 \pm \sqrt{9+24}}{4}.$$

since $0 < \rho < 1 \Rightarrow$ choose $+$ sign

$$\rho = \frac{\sqrt{33} - 3}{4}$$

MTH 7241 Fall 2022: Prof. C. King

Notes 7: Bayesian Inference

0.1 Bayesian vs frequentist

The distinction arises from the interpretation of the probability of an event. We write $\mathbb{P}(A)$ in both cases, but the meaning is a little different.

0.1.1 Frequentist 'classical' meaning

$\mathbb{P}(A)$ is the long-run fraction of occurrences of the event A under repeated independent sampling. This leads to, for example, the weak law of large numbers: make independent measurements of a random variable $X_1, \ldots, X_n, \ldots$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}[X]$$

$\mathbb{P}(A|B)$ is the long-run fraction of occurences of event A under repeated independent sampling when restricted to outcomes where B occurs.

Ex: toss a fair coin, $\mathbb{P}(H) = \frac{1}{2}$.

This means: toss many times then it comes up Heads 50% of the time.

repeat trial many times, keep the results when event B occurs.

0.1.2 Bayesian meaning

$\mathbb{P}(A)$ is your subjective belief of how likely it is that event A will occur.

$\mathbb{P}(A|B)$ is your updated belief of how likely it is that event A will occur given that event B has occurred.

$\mathbb{P}(A|B)$: updated belief of likelihood of A, given that event B happened.

: incorporate data into your probability models.

Classical frequentist viewpoint of random variable:

X has some pdf which we may or may not know exactly.

Ex: $X = \begin{cases} 1 & \text{toss Heads} \\ 0 & \text{toss Tails} \end{cases}$

X	0	1
prob.	$1-p$	p

p is a number which we may or may not know.

Bayesian viewpoint of random variable:

X : we have some model for its pdf.

Make measurements of X, and we update the parameters of the pdf of X based on measurements.

Bayesian: parameters of pdf become random variables.

Recall Bayes rule:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)} \tag{1}$$

0.1.3 Frequentist meaning

A and B are two random events, and it makes sense to consider both $\mathbb{P}(A|B)$ and $\mathbb{P}(B|A)$ (so long as both $\mathbb{P}(A)$ and $\mathbb{P}(B)$ are not zero). Bayes rule is the formula that gives the relation between these conditional probabilities. The order of events A, B in the conditional probabilities has no particular significance.

0.1.4 Bayesian meaning

$\mathbb{P}(A)$ is your belief in the likelihood of event A, and $\mathbb{P}(A|B)$ is your updated belief based on the fact that event B has occurred. Bayes rule tells you how to update your belief based on the measured data. The order of events A, B in the conditional probabilities is significant.

Bayesian application of Bayes rule:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)} \tag{2}$$

In most applications, the event B is measurement of some data, for example some independent measurements of a random variable $X_1, \ldots, X_n$. So we use $\mathcal{D}$ to indicate the data measured. The event A is our guess about the value of some set of parameters θ which determine the distribution of X. So $\mathbb{P}(A) = \mathbb{P}_0(\theta)$ is called the *prior* distribution and represents our belief about θ before the measurements are taken. Then $\mathbb{P}(B|A) = \mathbb{P}(\mathcal{D}|\theta)$ is called the *likelihood* and $\mathbb{P}(A|B) = \mathbb{P}_1(\theta|\mathcal{D})$ is the *posterior*. The normalizing constant $\mathbb{P}(B) = \mathbb{P}(\mathcal{D})$ is called the *evidence*. The formula is used in cases where distributions are discrete, continuous, or a mixture of the two. So the formula is

$$\mathbb{P}_1(\theta|\mathcal{D}) = \frac{\mathbb{P}(\mathcal{D}|\theta)\,\mathbb{P}_0(\theta)}{\mathbb{P}(\mathcal{D})}$$

The normalizing constant can be computed by summing over possible values for θ:

$$\mathbb{P}(\mathcal{D}) = \sum_j \mathbb{P}(\mathcal{D}|\theta_j)\,\mathbb{P}_0(\theta_j)$$

$$\mathbb{P}_1(\theta|\mathcal{D}) = \frac{\mathbb{P}(\mathcal{D}|\theta)\,\mathbb{P}_0(\theta)}{\mathbb{P}(\mathcal{D})}$$

θ: parameters you want to know in the pdf of a random variable, or hypotheses you want to decide between.

$\mathcal{D}$: data from measurements.

$\mathbb{P}_0(\theta)$: prior distribution (your pdf for θ before data is measured).

$\mathbb{P}_1(\theta|\mathcal{D})$: posterior distribution (your pdf for θ after data is measured)

$\mathbb{P}(\mathcal{D} \mid \theta)$: likelihood

$\mathbb{P}(\mathcal{D})$: evidence.

Example: Let's look at an example where we want to decide between two competing hypotheses. Suppose your friend claims that she can predict the outcome of a coin toss with 100% certainty.
H_0: your friend is lying, and she has 50% probability of guessing the outcome.
H_1: your friend is telling the truth.
We want to decide between H_0 and H_1. Our prior distribution is

$$\mathbb{P}(H_0) = 0.99 = 1 - \mathbb{P}(H_1)$$

We measure data: a coin is tossed 5 times, and your friend predicts the outcome every time. What is your updated belief in H_0?

Prior distribution: $\mathbb{P}(H_0) = 0.99$

$\mathbb{P}(H_1) = 0.01$

Data: toss coin 5 times, friend predicts correctly every time.

$\mathcal{D} = \{5$ out of 5 correct predictions$\}$.

Find posterior distribution:

$\mathbb{P}_1(H_0 | \mathcal{D}) = ?$, $\mathbb{P}_1(H_1 | \mathcal{D}) = ?$

Apply Bayesian inference:

$$\mathbb{P}_1(H_0 | \mathcal{D}) = \frac{\mathbb{P}(\mathcal{D} | H_0)\, \mathbb{P}_0(H_0)}{\mathbb{P}(\mathcal{D})}$$

Prior: $\mathbb{P}_0(H_0) = 0.99$

Likelihood: $\mathbb{P}(\mathcal{D} | H_0) = \mathbb{P}($correctly predicted 5 times $| H_0)$

$= \left(\frac{1}{2}\right)^5$

Evidence: $\mathbb{P}(\mathcal{D}) = \mathbb{P}(\mathcal{D} \mid H_0)\,\mathbb{P}_0(H_0) + \mathbb{P}(\mathcal{D} \mid H_1)\,\mathbb{P}_0(H_1)$

$= \left(\frac{1}{2}\right)^5 (0.99) + (1)(0.01).$

$$\Rightarrow \mathbb{P}_1(H_0 \mid \mathcal{D}) = \frac{\left(\frac{1}{2}\right)^5 (0.99)}{\left(\frac{1}{2}\right)^5 (0.99) + (0.01)} = 0.756$$

Based on data, your updated prob. for H_0 has decreased from 99% to 75.6%.

Bayesian inference provides a systematic way to answer this question. Let $\mathcal{D}$ denote the outcomes of the 5 coin tosses, then

$$\mathbb{P}(\mathcal{D}|H_0) = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

$$\mathbb{P}(\mathcal{D}|H_1) = 1$$

Using Bayes rule we get

$$\mathbb{P}(H_0|\mathcal{D}) = \frac{\mathbb{P}(\mathcal{D}|H_0)\,\mathbb{P}(H_0)}{\mathbb{P}(\mathcal{D})}$$

To compute $P(\mathcal{D}$ we use the total probability formula:

$$P(\mathcal{D}) = \mathbb{P}(\mathcal{D}|H_0)\,\mathbb{P}(H_0) + \mathbb{P}(\mathcal{D}|H_1)\,\mathbb{P}(H_1)$$

Putting it together we find

$$\mathbb{P}(H_0|\mathcal{D}) = \frac{0.99/32}{0.99/32 + 0.01} = 0.756$$

So our confidence that your friend is lying has been reduced to 75.6% based on her successes.