Modified on Monday 14th December, 2020, 12:22:0712:22

Dec 14, 2020

- Let's say that we are interested in the average and variance of the ages of the students who have registered for MAS109. Let N be the number of students and sequence $X = x_1, x_2, \dots, x_N$ denote the ages of all students.
- The mean $\bar{x}$ of the ages in X is

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i,$$

and the variance of the ages in X is

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2$$

- We want to estimate $\bar{x}$ and σ^2 by taking n samples from X. We choose a student r at random from $1, \dots, N$, and let $s_1 = x_r$. Repeat this process independently n times to have $s_1, \dots, s_n$. Students can be chosen more than once.
- The sample mean $\bar{s}$ is

$$\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i,$$

and the sample variance is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(s_i - \bar{s})^2.$$

We use $\bar{s}$ and S^2 as the estimators of the average and variance of student's ages.

- The expected value, the expectation of $\bar{s}$ is $\bar{x}$. We can also show that the expected value of S^2 is σ^2. Why do we divide it with $n-1$ to obtain S^2, while we divide it with N for σ^2?
- Let's begin with known facts.

$$E[\bar{s}] = \frac{1}{n}\left(\sum_{i=1}^{n} E[s_i]\right) = \bar{x}$$

$$E[S^2] = \frac{1}{n-1}\left(\sum_{i=1}^{n} E[(s_i - \bar{s})^2]\right) = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(s_i - \bar{s})^2\right]$$

Note that for every i

$$\sigma^2 = E[(s_i - \bar{x})^2].$$

$$E\left[\sum_{i=1}^{n}(s_i - \bar{x})^2\right] = n\sigma^2$$

We identify n samples as a point in $\mathbb{R}^n$.

$$\mathbf{s} = (s_1, s_2, \dots, s_n) \in \mathbb{R}^n$$

$$\mathbf{1} = (1, 1, \dots, 1) \in \mathbb{R}^n, \quad \bar{\mathbf{x}} = \bar{x}\mathbf{1}, \quad \bar{\mathbf{s}} = \bar{s}\mathbf{1}$$

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}(s_i-\bar{x})^2\right] &= E\left[\|\mathbf{s}-\bar{\mathbf{x}}\|^2\right] \\
&= E\left[\|\mathbf{s}-\bar{\mathbf{s}}\|^2+\|\bar{\mathbf{s}}-\bar{\mathbf{x}}\|^2\right], \quad \text{since } (\mathbf{s}-\bar{\mathbf{s}}) \perp \mathbf{1} \\
&= E[\|\mathbf{s}-\bar{\mathbf{s}}\|^2] + nE[(\bar{s}-\bar{x})^2] \\
&= E[\|\mathbf{s}-\bar{\mathbf{s}}\|^2] + \frac{1}{n}E[((s_1-\bar{x})+\cdots+(s_n-\bar{x}))^2] \\
&= E[\|\mathbf{s}-\bar{\mathbf{s}}\|^2] + \sigma^2 \\
&= E\left[\sum_{i=1}^{n}(s_i-\bar{s})^2\right] + \sigma^2
\end{aligned}
$$

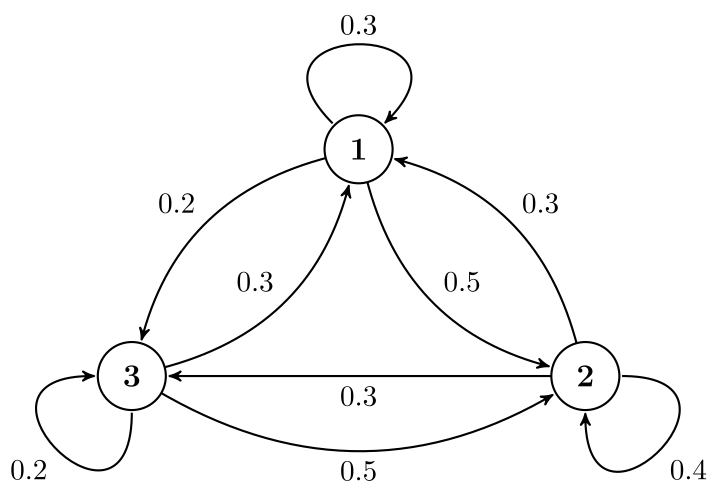
So we have

$$E\left[\sum_{i=1}^{n}(s_i-\bar{s})^2\right] = (n-1)\sigma^2,$$

which gives

$$E[S^2] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(s_i-\bar{s})^2\right] = \sigma^2.$$

- Finite Markov chain, finite state Markov chain
- There are three ride-free bicycle places. Initially the number of bicycles in place i is b_i, $y_0=(b_1,b_2,b_3)$. Everyday the number of bicycles in each place changes: $y_{n+1}=Py_n$. Does $\lim_{n\to\infty} y_n$ exist?



$$P = \begin{bmatrix} 0.3 & 0.3 & 0.3 \\ 0.5 & 0.4 & 0.5 \\ 0.2 & 0.3 & 0.2 \end{bmatrix}$$

Characteristic polynomial: $\lambda(\lambda-1)(\lambda+\frac{1}{10})$

Right eigenvectors: $\lambda_1=1:(33,50,27)$; $\lambda_2=0:(1,0,-1)$; $\lambda_3=-\frac{1}{10}:(0,1,-1)$

Left eigenvectors: $\lambda_1=1:(1,1,1)$; $\lambda_2=0:(7,-3,-3)$; $\lambda_3=-\frac{1}{10}:(5,-6,5)$

Recall that for any square matrix A, $A\mathbf{u}=\lambda\mathbf{u}$ and $A^T\mathbf{v}=\mu\mathbf{v}$ with $\lambda\neq\mu$ imply $\mathbf{u}^T\mathbf{v}=0$.

$$X = \begin{bmatrix} 33 & 1 & 0 \\ 50 & 0 & 1 \\ 27 & -1 & -1 \end{bmatrix}, \quad X^{-1} = \begin{bmatrix} \frac{1}{110} & \frac{1}{110} & \frac{1}{110} \\ \frac{7}{10} & -\frac{3}{10} & -\frac{3}{10} \\ -\frac{5}{11} & \frac{6}{11} & -\frac{5}{11} \end{bmatrix}$$

$$P = XDX^{-1} = X\,\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)X^{-1}$$

$$P^n = XD^nX^{-1} = X\,\mathrm{diag}(\lambda_1^n, \lambda_2^n, \lambda_3^n)X^{-1} \to \begin{bmatrix} \frac{33}{110} \\ \frac{50}{110} \\ \frac{27}{110} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

Columns of X are (right) eigenvectors of P

$$PX = X\,\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$$

Rows of X^{-1} are left eigenvectors of P, equivalently, columns of $(X^{-1})^T$ are eigenvectors of P^T.

$$X^{-1}P = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)X^{-1}$$

- There are N states: $1, \ldots, N$. Let p_{ij} denote the transition probability from state j at time n to state i at time $n+1$. Let $P = (p_{ij})$, an $N \times N$ matrix, be the transition probability matrix. For each i,

$$\sum_{i=1}^{N} p_{ij} = 1.$$

- p_{ij} is the probability that $x(n+1) = i$ if $x(n) = j$.
- Let $y_0 \in \mathbb{R}^N$ be the vector of the initial probabilities for states. $y_{n+1} = Py_n = \cdots = P^n y_0$. Does $\lim_{n\to\infty} y_n$ exist? It may or may not exist depending on P. If exists, let y_∞ denote the limit $\lim_{n\to\infty} y_n$.
- For $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, the limit doesn't exist.
- The limit exists if P is a positive Markov matrix: All p_{ij} are positive, each column of P sums to 1. In this case $\lambda = 1$ is the largest eigenvalue and it has a positive eigenvector, i.e., every entry is positive. For other eigenvalues λ, $|\lambda| < 1$.
- **Perron-Frobenius theorem**

  Let $A = (a_{ij})$ be an $n \times n$ matrix with $a_{ij} > 0$ for all i, j. Then the largest eigenvalue λ_{max} is positive and there is an eigenvector corresponding to λ_{max} with all positive components. Moreover, the algebraic multiplicity of λ_{max} is one and for any other eigenvalue λ, possibly complex, $|\lambda| < \lambda_{max}$ holds.