Homework 3 **Yen-Lin Chen**
ECE 5412 yc2253@cornell.edu
Fall 2019 Due: 11/30/19

Problem 49

Let $f = \mathbf{E}[(X - g(Y))'R(X - g(Y))]$ and we would like to find $g(Y)$ that minimizes f.

$$\begin{aligned} f =& \mathbf{E}[X'RX - g(Y)'RX - X'Rg(Y) + g(Y)'Rg(Y)] \\ \frac{\partial f}{\partial g(Y)} =& 0 = \mathbf{E}[-2RX + 2Rg(Y)] \\ \Longrightarrow& \mathbf{E}[RX] = \mathbf{E}[Rg(Y)] \end{aligned} \tag{1}$$

Using the iterated expectations: $\mathbf{E}[g(X, Y)] = \mathbf{E}_Y\{\mathbf{E}_X[g(X, Y)|Y]\}$

$$\mathbf{E}_Y\left\{\mathbf{E}_X[RX|Y]\right\} = \mathbf{E}_Y\left\{\mathbf{E}_X[Rg(Y)|Y]\right\} = \mathbf{E}_Y[Rg(Y)] \tag{2}$$

$$R\mathbf{E}_Y\{\mathbf{E}_X[X|Y]\} = R\mathbf{E}_Y[g(Y)] \tag{3}$$

Since R is positive definite, R^{-1} exists. Therefore, $\mathbf{E}_X[X|Y] = g(Y)$

Problem 51

Suppose $\pi_1 = P'\pi_0 = (p_1, p_2)'$. Since $y_1 = x_1 + v_1$ where $v_1 \sim N(0, 1)$, we have $y_1|x_1 \sim N(y_1 - x_1, 1)$.

$$P(y_1|x_1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(y_1 - x_1)^2\right] \tag{4}$$

From the Bayes rule,

$$P(x_1|y_1) = \frac{P(y_1|x_1)P(x_1)}{\sum_{x_1} P(y_1|x_1)P(x_1)} \tag{5}$$

where $P(x_1 = 1) = p_1$ and $P(x_2 = 2) = p_2$.

$$\begin{aligned} P(x_1 = 1|y_1) &= \frac{p_1 \exp\left[-\frac{1}{2}(y_1 - 1)^2\right]}{p_1 \exp\left[-\frac{1}{2}(y_1 - 1)^2\right] + p_2 \exp\left[-\frac{1}{2}(y_1 - 2)^2\right]} \\ P(x_1 = 2|y_1) &= \frac{p_2 \exp\left[-\frac{1}{2}(y_1 - 2)^2\right]}{p_1 \exp\left[-\frac{1}{2}(y_1 - 1)^2\right] + p_2 \exp\left[-\frac{1}{2}(y_1 - 2)^2\right]} \end{aligned} \tag{6}$$

We can simplify Eq. (6) further as the following

$$\begin{aligned} P(x_1 = 1|y_1) &= \frac{p_1 \exp\left[-\frac{1}{2}(2y - 3)\right]}{p_1 \exp\left[-\frac{1}{2}(2y - 3)\right] + p_2} \\ P(x_1 = 2|y_1) &= \frac{p_2}{p_1 \exp\left[-\frac{1}{2}(2y - 3)\right] + p_2} \end{aligned} \tag{7}$$

Problem 54

The transition probability of the Markov Chain is

$$P_{ij} = P(X = j|X = i) = \frac{b_j}{b_i + b_j} Q_{ij} \tag{8}$$

Since $Q_{ij} = Q_{ji}$ and $b_i > 0$ $\forall i = 1, 2, \ldots, N$, we want to find the stationary distribution π_∞ that satisfies $P_{ij}\pi_\infty(i) = P_{ji}\pi_\infty(j)$, i.e.

$$\frac{b_j}{b_i + b_j} Q_{ij}\pi_\infty(i) = \frac{b_i}{b_i + b_j} Q_{ji}\pi_\infty(j) \tag{9}$$

Suppose $Q_{ij} = Q_{ji} \neq 0$, we have

$$\frac{\pi_\infty(i)}{b_i} = \frac{\pi_\infty(j)}{b_j} \tag{10}$$

Therefore, with stationary distribution $\pi_\infty(i) \propto b_i$, this Markov Chain satisfies the balanced equation and is reversible.

Problem 55

Let $X \in \mathbb{R}^n$ be the data to be classified into binary class $Y \in \{0, 1\}$. The naive Bayes classifier is

$$g(X) = \arg\max_Y P(Y|X) \tag{11}$$

Given data X, the naive Bayes classifier outputs 1 if $P(Y = 1|X) > P(Y = 0|X)$. The linear discriminant analysis (LDA) is a classifier based on constructing a hyper-plane in $\mathbb{R}^n$.

We first explore the parametric case where $P(X|Y) \sim N(\mu_Y, \Sigma)$ with prior $\pi(Y = 0)$ and $\pi(Y = 1)$. Then the naive Bayes classifier is comparing the conditional probability of $P(Y = 0|X)$ and $P(Y = 1|X)$, i.e.

$$\begin{aligned} g(X) &= \arg\max_{Y\in\{0,1\}} \left[\log \pi(Y) - \frac{1}{2}(X - \mu_Y)'\Sigma^{-1}(X - \mu_Y)\right] \\ &= \arg\max_{Y\in\{0,1\}} \left[\log \pi(Y) + \frac{1}{2}(2\mu_Y\Sigma^{-1}X - \mu_Y'\Sigma^{-1}\mu_Y)\right] \end{aligned} \tag{12}$$

The LDA hyper-plane is $P(Y = 0|X) = P(Y = 1|X) = \frac{1}{2}$ and substituting μ, Σ and π with the empirical estimators $\hat{\mu}$, $\hat{\Sigma}$ and $\hat{\pi}$:

$$\log \hat{\pi}(Y = 0) + \frac{1}{2}(2\hat{\mu_0}'\hat{\Sigma}^{-1}X - \hat{\mu_0}'\hat{\Sigma}^{-1}\hat{\mu_0}) = \log \hat{\pi}(Y = 1) + \frac{1}{2}(2\hat{\mu_1}'\hat{\Sigma}^{-1}X - \hat{\mu_1}'\hat{\Sigma}^{-1}\hat{\mu_1}) \tag{13}$$

We use the *fisheriris.mat* data set to illustrate the performance of parametric LDA. The data set contains three classes: *setosa*, *versicolor* and *virginica* with $n = 4$ and 50 observation for each class. We remove the *virginica* data to make it a binary classification problem with $Y = 0$ for *setosa* and $Y = 1$ for *versicolor*. Therefore, $\hat{\pi}(Y = 0) = \hat{\pi}(Y = 1) = \frac{1}{2}$. Eq. (13) is simplified as follows

$$\begin{aligned} &2\hat{\mu_0}'\hat{\Sigma}^{-1}X - \hat{\mu_0}'\hat{\Sigma}^{-1}\hat{\mu_0} = 2\hat{\mu_1}'\hat{\Sigma}^{-1}X - \hat{\mu_1}'\hat{\Sigma}^{-1}\hat{\mu_1} \\ \iff\ &2(\hat{\mu_0} - \hat{\mu_1})'\hat{\Sigma}^{-1}X = \hat{\mu_0}'\hat{\Sigma}^{-1}\hat{\mu_0} - \hat{\mu_1}'\hat{\Sigma}^{-1}\hat{\mu_1} \end{aligned} \tag{14}$$

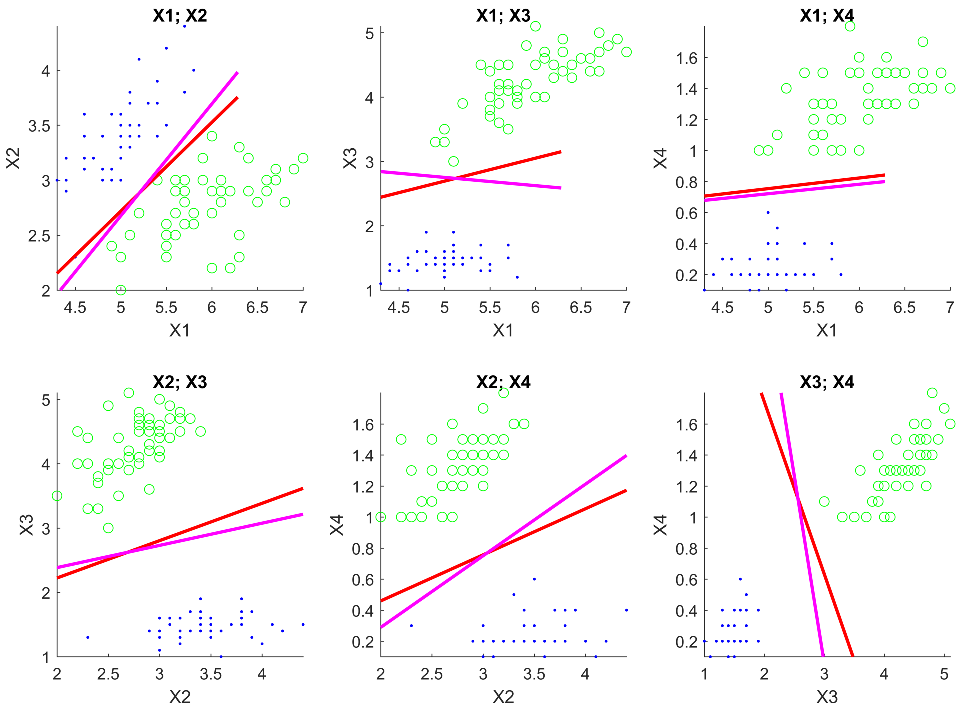


Figure 1: Problem 55: The naive Bayes and logistic classifiers are shown as red and magenta lines.

Secondly, we can use a semi-parametric model:

$$P(Y=1|X=x)=\frac{e^{\alpha+\beta' x}}{e^{\alpha+\beta' x}+1}=\frac{1}{1+e^{-(\alpha+\beta' x)}} \tag{15}$$

We apply the maximum likelihood to estimate α and β using the data. The likelihood function is

$$L=\prod_{i=1}^{n} P(x_i)^{y_i}(1-P(x_i))^{(1-y_i)} \tag{16}$$

$$l=\log L=\sum_{i=1}^{n}\log\left(1-P(x_i)\right)+\sum_{i=1}^{n} y_i \log\left(\frac{P(x_i)}{1-P(x_i)}\right) \tag{17}$$

where $P(x_i)=P(Y=1|X=x_i)$. Eq. (17) does not have a close form expression so we have to use numerical optimization to estimate α and β. Fortunately, we have the gradient of the target function l:

$$\begin{aligned}\frac{\partial l}{\partial \alpha}&=\sum_{i=1}^{n}(y_i-P(x_i))\\ \frac{\partial l}{\partial \beta_j}&=\sum_{i=1}^{n}(y_i-P(x_i))x_{ij}\end{aligned} \tag{18}$$

The binary classification results are shown in Fig. 1 with the naive Bayes and logistic classifiers being the red and magenta lines respectively.

patient #	Re-admission Time (t_i)	Censored	Gender	Smoker
1	5	0	0	1
2	3	0	1	1
3	19	1	1	0
4	17	1	1	0

Table 1: A snippet of the *arrhythmia.mat* data set.

Problem 57

Please see the final section for the short report.

Problem 58

In this problem we explore the effects of gender and smoking on the re-admission times of patients with cardiac arrhythmia. The data set *arrhythmia.mat* has 100 observations and contains the time between the first and second admission to the hospital due to cardiac arrhythmia. Some of the data is right-censored with different censoring cutoff because contact to the patient might be lost at different time. The data is further labeled with the gender and smoker. A snippet of the data is shown in Table 1.

Let t_i be the re-admission time observation of the i^{th} patient and y_i be the real readmission time. For uncensored data, $t_i = y_i \ \forall i = 1, 2, \ldots, n_{uc}$ while for censored data with cutoff c_j, $t_j = c_j \ \forall j = 1, 2, \ldots, n_c$. We formulate the Poisson model with parameter x to be estimated as the following.

$$P(t_i = y_i|x) = \frac{e^{-x}x^{t_i}}{t_i!} \quad \forall i = 1, 2, \ldots, n_{uc} \tag{19}$$

$$P(t_j = c_j|x) = \left(1 - \sum_{k=0}^{t_j} \frac{e^{-x}x^k}{k!}\right) \quad \forall j = 1, 2, \ldots, n_c \tag{20}$$

The likelihood for the observed uncensored data $\{t_i\}_{i=1}^{n_{uc}}$ and censored data $\{t_j\}_{j=1}^{n_c}$ is

$$P(\{t_i\}_{i=1}^{n_{uc}}; \{t_j\}_{j=1}^{n_c}|x) = \left(\prod_{i=1}^{n_{uc}} \frac{e^{-x}x^{t_i}}{t_i!}\right) \prod_{j=1}^{n_c} \left(1 - \sum_{k=0}^{t_j} \frac{e^{-x}x^k}{k!}\right) \tag{21}$$

We further assume the improper prior for the parameter x: $P(x) \sim \frac{1}{x}$ to evaluate the posterior distribution $P(x|\{t_i\}_{i=1}^{n_{uc}}; \{t_j\}_{j=1}^{n_c})$. We use Gibbs sampling to simulate a *cloud* of the x estimates and access the mean and variance of the estimate. For Gibbs sampling, we need to augment $\{y_j\}_{j=1}^{n_c}$ to the censored data and marginalize on $\{y_j\}_{j=1}^{n_c}$. We simulate the augmented data $\{y_j\}_{j=1}^{n_c}$ using the following probability distribution.

$$P\left(y_j|\{t_i\}_{i=1}^{n_{uc}}, \{t_j\}_{j=1}^{n_c}, x\right) \propto \left(\frac{e^{-x}x^{y_j}}{y_j!}\right) I(y_j \geq t_j) \tag{22}$$

Note that sampling from Eq. (22) is equivalent to sampling from a Poisson with mean x and than adding a constant t_j. With all the augmented $\{y_j\}_{j=1}^{n_c}$, the posterior of x is then

$$\begin{aligned} P\left(x|\{t_i\}_{i=1}^{n_{uc}}, \{t_j\}_{j=1}^{n_c}\right) &= P\left(x|\{t_i\}_{i=1}^{n_{uc}}, \{y_j\}_{j=1}^{n_c}\right) \\ &\propto \left(\prod_{i=1}^{n_{uc}} \frac{e^{-x}x^{t_i}}{t_i!}\right) \left(\prod_{j=1}^{n_c} \frac{e^{-x}x^{y_i}}{y_i!}\right) \left(\frac{1}{x}\right) \end{aligned} \tag{23}$$

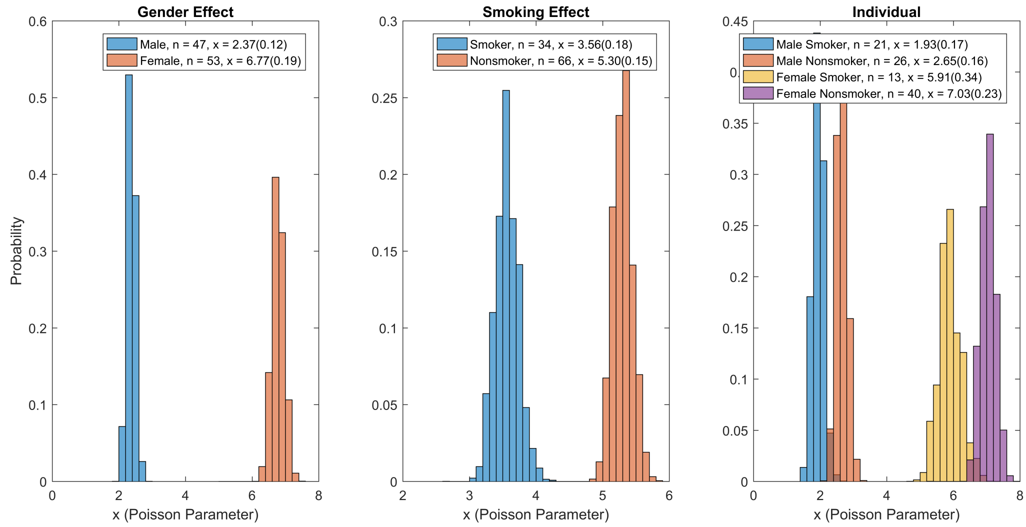


Figure 2: Problem 58: The histograms of x estimate using Gibbs sampling to investigate the (left) gender effect (middle) smoking effect and (right) combined effect of the re-admission time for 100 patients with cardiac arrhythmia. The mean and standard deviation of x estimate is reported in the legend where n is the number of observations in the data set.

One way is to we simulate $\{y_j^{(k)}\}_{j=1}^{n_c}$ using Eq. (22) and $x^{(k)}$ using Eq. (23). The sequence $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$ converges to the posterior distribution $P(x|\{t_i\}_{i=1}^{n_{uc}}, \{t_j\}_{j=1}^{n_c})$. The other way (given the form of the posterior) is to use the MAP estimator derived as the following.

$$L = \log \left[P\left(x|\{t_i\}_{i=1}^{n_{uc}}, \{y_j\}_{j=1}^{n_c} \right) \right] \tag{24}$$

$$\frac{dL}{dx} = -(n_{uc} + n_c) + \frac{1}{x}\left(\sum_{i=1}^{n_{uc}} t_i + \sum_{j=1}^{n_c} y_j - 1 \right) = 0 \tag{25}$$

$$\hat{x}_{MAP} = \frac{\sum_{i=1}^{n_{uc}} t_i + \sum_{j=1}^{n_c} y_j - 1}{n_{uc} + n_c} \tag{26}$$

As one side note, the prior $P(x) \propto \frac{1}{x}$ comes in as the -1 term. Given large number of observations, n, the effect of the prior dies with $\mathcal{O}(1/n)$.

The distributions of x estimate from Gibbs sampling is shown in Fig. 2 and we compare the gender, smoking effects as well as combined effects. From the data, the gender plays a much more important role in the re-admission time for the patients; male patients tend to have re-admission time of 2.37 compare to 6.77 of female patients. As expected smokers have shorter re-admission time. Therefore, using the Poisson likelihood and Gibbs sampling, we report that

$$x(\text{Female Nonsmoker}) > x(\text{Female Ssmoker}) > x(\text{Male Nonsmoker}) > x(\text{Male Smoker}) \tag{27}$$

Problem 59

The nonlinear stochastic system is defined as the following.

$$x_{k+1} = \sin(x_k) + x_k \quad w \sim N(0, \Sigma) \tag{28}$$

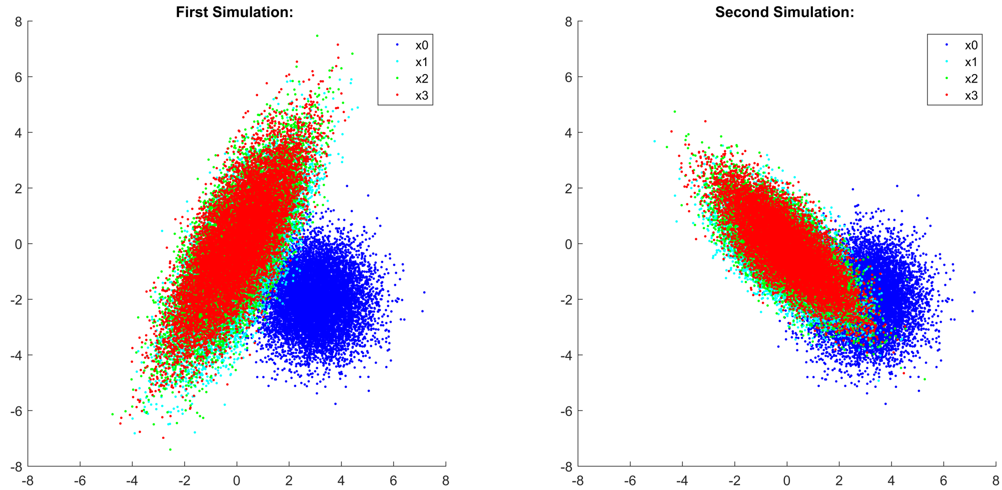


Figure 3: Problem 59: The densities of x_0, x_1, x_2 and x_3 from 2 simulations with Σ_1 and Σ_2.

Let the initial density of x_0 be $\pi_0(x)$ and we can write the transition probability from time k to time $k+1$ as

$$P(x_{k+1}|x_k) = N(\sin(x_k), \Sigma) \tag{29}$$

By Chapman-Komogorov equation, the density of x_{k+1} is $\pi_{k+1}(x_{k+1})$:

$$\pi_{k+1}(x_{k+1}) = \int P(x_{k+1}|x_k)\pi_k(x_k)dx_k \tag{30}$$

Instead of doing the integration, we can simulate x_{k+1}'s using the composite method by knowing both $P(x_{k+1}|x_k)$ and $\pi_k(x_k)$.

1. Generate $x_k^* \sim \pi_k(x_k)$
2. Generate $x_{k+1}^* \sim P(x_{k+1}|x_k^*)$

The density of simulated x_{k+1}^*'s, $\hat{\pi}_{k+1}(x_{k+1})$ will converge to $\pi_{k+1}(x_{k+1})$. Using this simulated density, we can repeat the same procedure for $k+2$, $k+3$, etc. Note that we will have to generate x_{k+1}^* from the empirical density $\hat{\pi}_{k+1}(x_{k+1})$ but the simulated x_{k+1}^* is already determined. Therefore, to simulate this system for $k = 1, 2, 3$, we use the following simplified procedure.

1. Generate $x_0^* \sim \pi_0(x_0)$.
2. Generate $x_1^* \sim P(x_1|x_0^*) \rightarrow$ find density, mean and covariance, $k = 1$.
3. Generate $x_2^* \sim P(x_2|x_1^*) \rightarrow$ find density, mean and covariance, $k = 2$.
4. Generate $x_3^* \sim P(x_3|x_2^*) \rightarrow$ find density, mean and covariance, $k = 3$.

To run the simulation, suppose that x is a normally distributed two-dimensional random vector with mean μ and covariance I_2, i.e. $x \sim N_2(\mu, I_2)$, where $\mu = (3, -2)'$ to demonstrate the non-linearity of the problem. We choose the two-dimensional example for the ease of visualization. We ran two simulations with different Σ's:

$$\Sigma_1 = \begin{bmatrix} 1 & 1.5 \\ 1.5 & 3 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 0.9 \end{bmatrix} \tag{31}$$

and the simulated densities are shown in Fig. 3. We report the statistics as follows.

For the first simulation,

$$\bar{x_1} = \begin{bmatrix} 0.0923 \\ -0.5368 \end{bmatrix} \quad \bar{x_2} = \begin{bmatrix} 0.0527 \\ -0.1024 \end{bmatrix} \quad \bar{x_3} = \begin{bmatrix} 0.0281 \\ -0.0127 \end{bmatrix} \tag{32}$$

$$\mathrm{cov}(x_1) = \begin{bmatrix} 1.42 & 1.49 \\ 1.49 & 3.21 \end{bmatrix} \quad \mathrm{cov}(x_2) = \begin{bmatrix} 1.48 & 1.67 \\ 1.67 & 3.49 \end{bmatrix} \quad \mathrm{cov}(x_3) = \begin{bmatrix} 1.49 & 1.72 \\ 1.72 & 3.50 \end{bmatrix} \tag{33}$$

For the second simulation,

$$\bar{x_1} = \begin{bmatrix} 0.1034 \\ -0.5611 \end{bmatrix} \quad \bar{x_2} = \begin{bmatrix} 0.0506 \\ -0.3136 \end{bmatrix} \quad \bar{x_3} = \begin{bmatrix} 0.0418 \\ -0.1974 \end{bmatrix} \tag{34}$$

$$\mathrm{cov}(x_1) = \begin{bmatrix} 1.46 & -0.82 \\ -0.82 & 1.13 \end{bmatrix} \quad \mathrm{cov}(x_2) = \begin{bmatrix} 1.47 & -1.01 \\ -1.01 & 1.29 \end{bmatrix} \quad \mathrm{cov}(x_3) = \begin{bmatrix} 1.50 & -1.10 \\ -1.10 & 1.36 \end{bmatrix} \tag{35}$$

Problem 61

Denote the state in the phase space as $z_k = (x_k, \dot{x}_k, y_k, \dot{y}_k)'$ and acceleration $r_{k+1} = (\ddot{x}_k, \ddot{y}_k)'$, which will be deemed as a constant r. The state and the observation models are the following.

$$z_{k+1} = Az_k + fr_{k+1} + w_k \quad w_k \sim (0, Q) \quad k = 0, 1, \ldots \tag{36}$$

$$y_k = Cz_k + v_k \quad v_k \sim N(0, R) \tag{37}$$

where

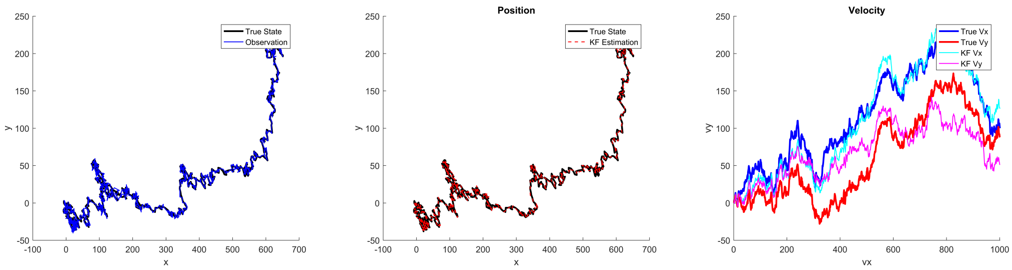
$$A = \begin{bmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad f = \begin{bmatrix} \Delta^2/2 & 0 \\ \Delta & 0 \\ 0 & \Delta^2/2 \\ 0 & \Delta \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{38}$$

and Δ is the sampling period. We will use Kalman Filter (KF) to estimate the state x_k given observations y_k. We write the recursive KF equations as the following.

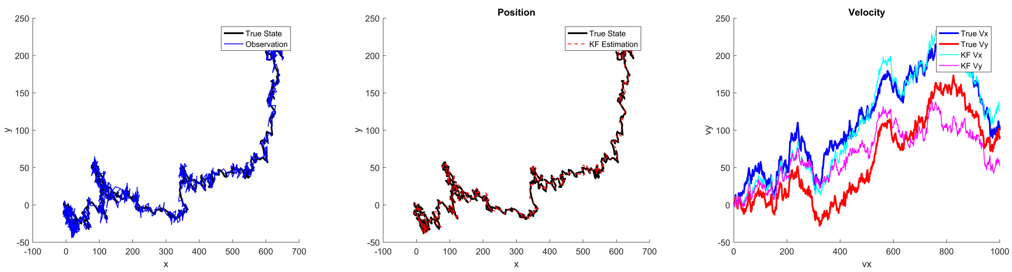
$$\begin{aligned} \hat{z}_{k+1|k} &= A\hat{z}_k + fr \\ y_{k+1|k} &= C\hat{z}_{k+1|k} \\ \Sigma_{k+1|k} &= A\Sigma_k A' + Q \\ S_{k+1} &= C\Sigma_{k+1|k}C' + R \\ \hat{z}_{k+1} &= \hat{z}_{k+1|k} + \Sigma_{k+1|k}C'S_{k+1}^{-1}(y_{k+1} - y_{k+1|k}) \\ \Sigma_{k+1} &= \Sigma_{k+1|k}C'S_{k+1}^{-1}C\Sigma_{k+1|k} \end{aligned} \tag{39}$$

We simulate the observations using three different noise levels $\sigma^2 R$ where $\sigma^2 R$ with $\sigma^2 = 1, 5, 10$. The results are shown in Fig. 4. The KF filter works very well even in the most noisy case; the KF estimates still recapitulate the positions of the target. It is interesting to note that the KF estimates for the velocities are not as good as the positions because in the observation matrix C, we do not observe the velocities. However, the KF estimates of the velocities seem to be insensitive to the observation noise because the velocities are inferred instead of being corrupted measurements. This problem demonstrate decent performance of Kalman Filter even though the observation is highly noisy.

(a)



(b)



(c)

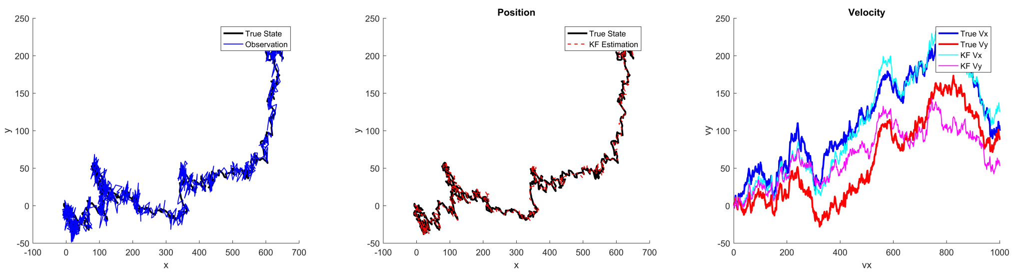


Figure 4: Problem 61: Simulated positions (black traces in the left 2 panels), velocities traces (blue and red in the right panel) and noisy observations (blue in the left panel) using Eq. (36) and (37). The position traces of KF estimate are shown as red dashed lines in the middle panel while the KF estimate of velocities are shown as cyan and magenta in the right panel. We use three different noise levels: $\sigma^2 R$ with (a) $\sigma^2 = 1$, (b) $\sigma^2 = 5$ and (c) $\sigma^2 = 10$.

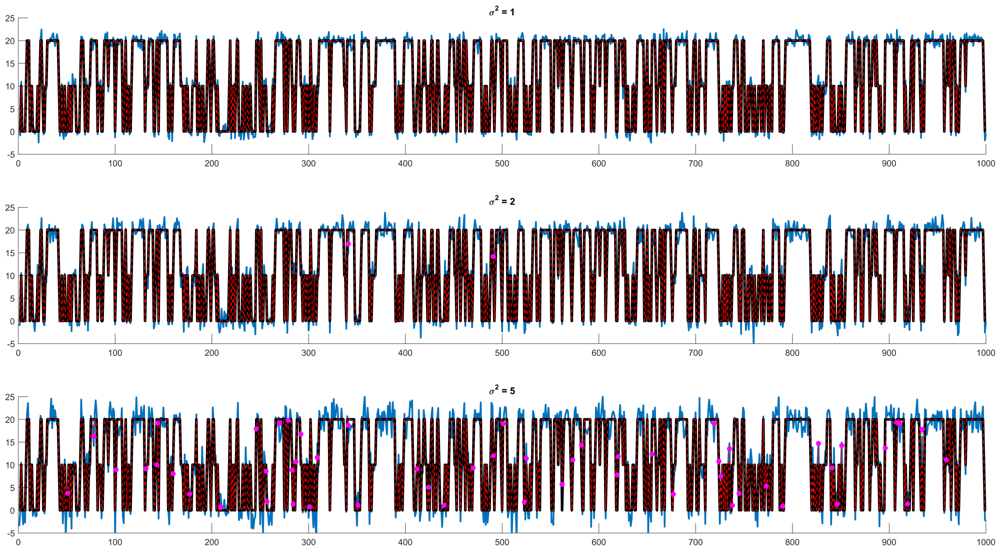


Figure 5: Problem 62(a): The simulated noisy measurements y_k (blue) of Markov Chain states x_k (black) using $\sigma^2 = 1$, $\sigma^2 = 2$ and $\sigma^2 = 5$. The red dashed traces represents the estimates from HMM filter. The magenta dots indicate the time k when the HMM estimate $\hat{x}_k$ does not agree with the true state x_k.

Problem 62

Suppose we have the transition matrix $p \in \mathbb{R}^{3\times 3}$ for the Markov Chain x_k in state space $\{0, 10, 20\}$ and let $\pi_k(i) = P(x_k = i|y_{1:k})$. Now the observation of the Markov Chain is noisy:

$$y_k = x_k + w \quad w \sim N(0, \sigma^2) \tag{40}$$

(a)

The probability of observation y_k given state at x_k is

$$P(y_k|x_k) = N(x_k, \sigma^2) \tag{41}$$

Note that x_k is discrete from a 3-state Markov Chain while y_k is a continuous observation. We apply the Hidden Markov Model (HMM) Filter using the following state estimate π_{k+1} given observation $y_{1:k+1}$:

$$\pi_{k+1} = \frac{B_{y_{k+1}} P' \pi_k}{\mathbf{1}' B_{y_{k+1}} P' \pi_k} \quad B_{y_k} = \begin{bmatrix} P(y_k|x_k = 0) & 0 & 0 \\ 0 & P(y_k|x_k = 10) & 0 \\ 0 & 0 & P(y_k|x_k = 20) \end{bmatrix} \tag{42}$$

Let $S = (0, 10, 20)'$ be the state vector. The optimal estimate from the HMM filter is

$$\hat{x}_{k+1} = S' \pi_{k+1} \tag{43}$$

The traces of the simulated Markov Chain x_k and the noisy observations y_k with $\sigma^2 = 1, 2, 5$ are shown as blue and black respectively in Fig. 5. The trace of estimate from HMM filter, $\hat{x}_k$ is

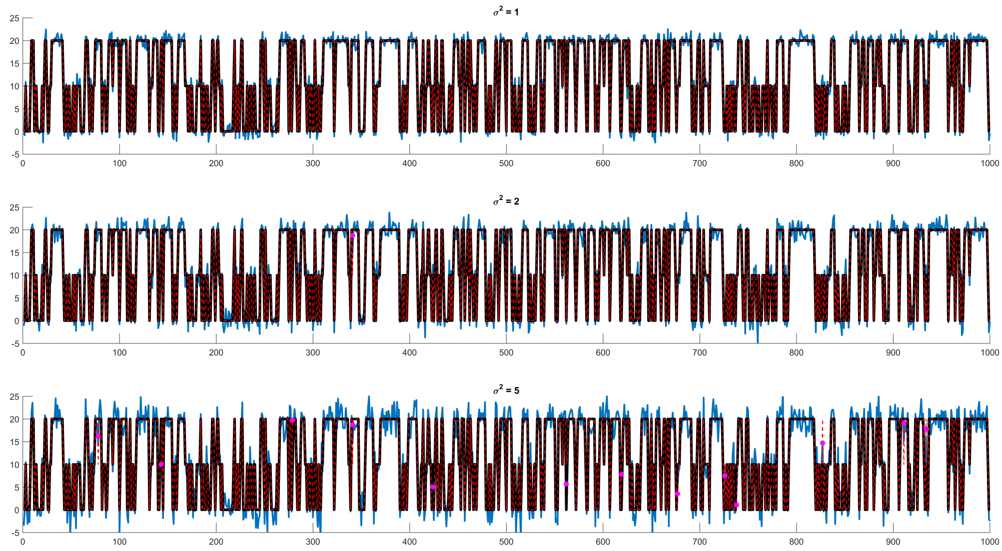


Figure 6: Problem 62(b): The simulated noisy measurements y_k (blue) of Markov Chain states x_k (black) using $\sigma^2 = 1$, $\sigma^2 = 2$ and $\sigma^2 = 5$. The red dashed traces represents the estimates from fixed interval smoothed estimate (FISE) using the Viterbi algorithm. The magenta dots indicate the time k when the FISE estimate $\hat{x}_k$ does not agree with the true state x_k.

shown in the red dashed line. In general, HMM filter performs very well in this case, where for different observation noises, the most of the estimates recapitulate the hidden Markov Chain. We indicate the disagreement using the magenta dots in Fig. 5 and with increasing σ^2, the error occurs more frequently due to the high noisy once in a while. We report the error rate r for three different cases as the following.

$$\begin{aligned} \sigma^2 &= 1 \quad r = 0.00\% \\ \sigma^2 &= 2 \quad r = 0.20\% \\ \sigma^2 &= 5 \quad r = 5.50\% \end{aligned} \tag{44}$$

The good performance is also due to the fact that the differences between the Markov states are big enough that the noises do not corrupt the signal significantly even for the case where $\sigma^2 = 5$.

(b)

We further use a fixed interval smoothed estimate (FISE) for the Markov state x_k using the Viterbi algorithm. In the Viterbi algorithm, we have to maximize the likelihood $P(y_{1:k}, x_{1:k})$ and since the Markov state x_k is equally spaced with the same observation noise variance, the maximum-likelihood estimate for state at time k given observation y_k is essentially rounding the observation to the closest state. The red dashed lines in Fig. 6 demonstrate the traces FISE from the same observations $y_{1:1000}$. Using the same plotting scheme, we label the disagreement between FISE and real state by the magenta dots and report the error rates as the following.

$$\begin{aligned} \sigma^2 = 1 \quad r = 0.00\% \\ \sigma^2 = 2 \quad r = 0.10\% \\ \sigma^2 = 5 \quad r = 1.30\% \end{aligned} \tag{45}$$

Attachments

acceptance_rejection.m, armc.m, gibbs.m, p55.m, p58.m, p59.m, p61.m, p62.m

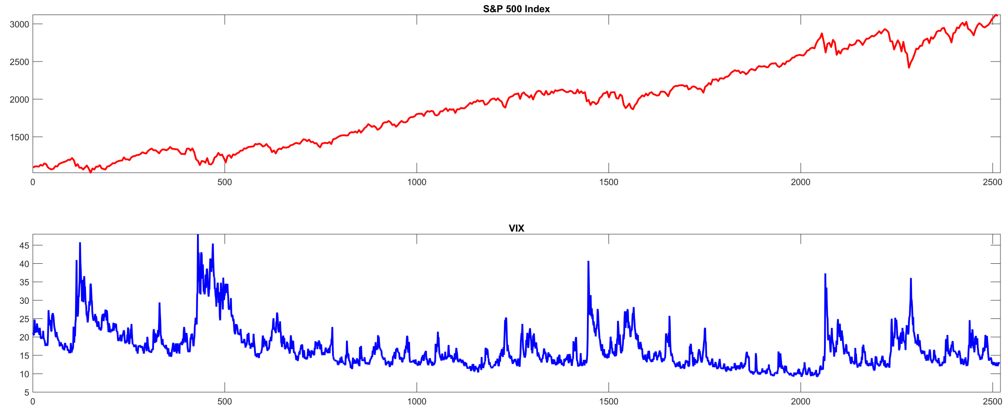


Figure 7: The historical SP500 Index and VIX.

Metropolis-Hastings Algorithm on Hierarchical Bayesian Models

In this section, we will explore and model the volatility in the U.S. stock market. In finance, the volatility is defined as the variation of a trading price or composite index time series as measured by the standard deviation of the logarithmic returns[1]. If the stock price fluctuates significantly, the volatility is high and any investment in such stock incurs high risk. On the other hand, if the stock price does not have a great swing, the volatility is low. According to Wikipedia, it is still unclear how the volatility comes to exist in the first place and therefore, it is hard to model. One of the most popular and consistent measure of stock market volatility is the CBOE Volatility Index (VIX). The VIX is the future implied volatility of the Standard and Poor 500 (SP500) Index based on the current price[2]. We download the daily 10-year historical data for VIX and SP500 Index from Yahoo Finance[3]. Fig. 7 shows the VIX and SP500 Index of the last 10 years from 2009/11/22 to 2019/11/22. The x axes are the number of trading days since 2009/11/22. It is clear that whenever the SP500 experienced a plunge, the implied volatility, VIX spiked.

Inspired by the concept of Markov Modulated Poisson Model mentioned very briefly in the class. We came up with this original and simple model for the VIX using the combination of Poisson distribution and hierarchical Bayesian model. Suppose the market volatility is a Markov Chain x_k at time k and assume that this Markov Chain changes state on a weekly instead of daily basis with transition matrix P. We have no information about the structure of P but we will use the VIX index to estimate P. During the time period k and $k+1$, we have l observations $y_{k,1}, y_{k,2}, \ldots, y_{k,l}$ of the market volatility by VIX. We formulate the hierarchical Bayesian model as the following. The VIX from time k and $k+1$ follows the Poisson distribution, from the real market volatility x_k.

$$P(y_{k,i}|x_k) = \frac{e^{-x_k} x_k^{y_{k,i}}}{y_{k,i}!} \quad \Longrightarrow \quad P(y_{k,1:l}|x_k) = \prod_{i=1}^{l} \frac{e^{-x_k} x_k^{y_{k,i}}}{y_{k,i}!} \tag{46}$$

Formally, we have to model x_k using the intrinsic transition probability P such that $\pi_k = P'\pi_{k-1}$, but P is unknown to us. Instead, we use the fact that the state x_k is not far away from the previous state x_{k+1}. Therefore, we model x_k as a normal distribution centered on the previous state x_{k-1} with some unknown variance θ.

$$x_k|\theta \sim N(x_{k-1}, \sigma^2 = \theta) \quad \Longrightarrow \quad P(x_k|\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x_k - x_{k-1})^2} \tag{47}$$

Finally, assume that the variance θ has the improper prior distribution $1/\theta$ and we have the posterior distribution as follows.

$$\begin{aligned} P(x_k|y_{k,1:l}) &\propto \int P(y_{k,1:l}|x_k) P(x_k|\theta) p(\theta) d\theta \\ &\propto \int \prod_{i=1}^{l} \left(\frac{e^{-x_k} x_k^{y_{k,i}}}{y_{k,i}!} \right) \left[\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x_k - x_{k-1})^2} \right] \left(\frac{1}{\theta} \right) d\theta \end{aligned} \tag{48}$$

The aim is to obtain the estimate $\hat{x}_k = E(x_k|y_{k,1:l})$ but Eq. (48) is difficult to integrate to get the posterior distribution. Therefore, we run the Metropolis-Hastings Algorithm described as follows to obtain the empirical posterior distribution. Suppose we have a initial sample $\bar{z} = (\bar{x}, \bar{\theta})'$.

1. Generate the new sample z from $q(z|\bar{z})$, where $q(z|\bar{z}) = q(\bar{z}|z)$ is symmetric.

$$q(z|\bar{z}) \sim N(\bar{z}, \sigma^2 I_2) \tag{49}$$

2. Calculate the acceptance probability $\alpha_{\bar{z},z}$.

$$\alpha_{\bar{z},z} = \min \left(1, \frac{P(y_{k,1:l}|x) P(x|\theta) p(\theta)}{P(y_{k,1:l}|\bar{x}) P(\bar{x}|\bar{\theta}) p(\bar{\theta})} \right) \tag{50}$$

3. Generate $u \sim \mathrm{U}[0, 1]$ and accept z if $u < \alpha_{\bar{z},z}$.

We tested the Metropolis-Hastings algorithm using 2 periods of the VIX and 50000 samples, shown in the top panel of Fig. 8. The MH algorithm performs well in estimating the market volatility using the hierarchical Bayesian model. It is worth noting that for the high-volatility period (blue), the previous Markov state is at about 30 but given enough number of MH samples, the true posterior distribution is still recapitulated.

Now we can estimate the state of hidden Markov Chain that determines the Poisson distribution of the VIX. Our goal is to estimate the transition probability of market volatility. We can do this by estimating the states for the full VIX time series using a period of two weeks (10 trading days). The result is shown in Fig. 9 where we show the estimated market volatility along with the corresponding standard deviations. Using this estimated state of the Markov Chain, be can further estimate the transition probability P.

The estimated market volatility state and the transition probability of the underlying Markov Chain P is shown in Fig. 10. The states of $1, 2, \ldots, 9$ correspond to the low volatility states. The states of $10, 11, \ldots, 14$ are medium volatility whereas state $15, 16$ are high volatility states where the investor's panic permeated the market. We see that in the low-volatility states, investors are confident about the market and the market remains in the low-volatility states more often. However, once in a while, the panic kicked in and the volatility went up step by step until hitting state $10, 11, 12$, where there is a high chance that the market makes transition to the high-volatility states, $15, 16$. And market plunged. These high-volatility states are very unstable in the sense that it is difficult for the market to keep its high volatility.

The current stock market is in the low-volatility state 7. The probability of the market going to state 6 (low volatility) is 0.0861 while high-volatility with probability is 0.1030 with

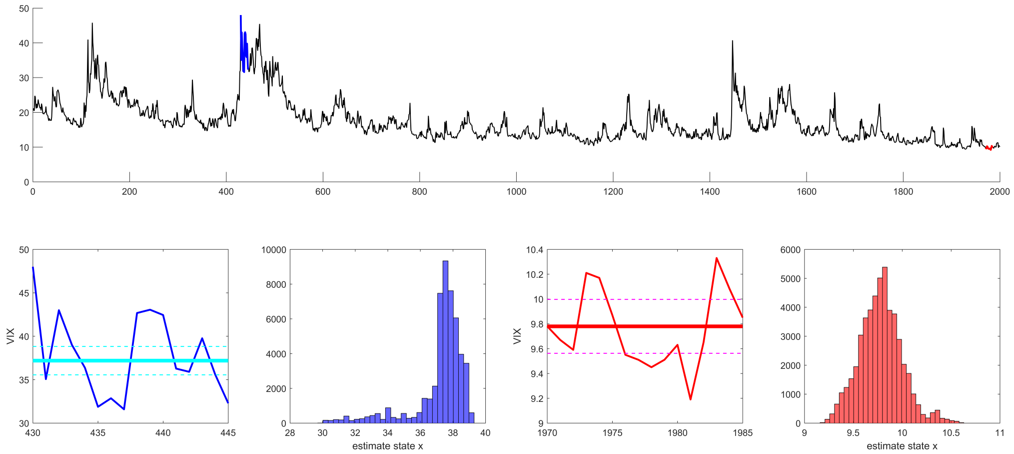


Figure 8: Application of Metropolis-Hastings algorithms for estimating the market volatility using 2 short periods (blue and red) of 14 trading days. The cyan line represent the estimated state $\hat{x}_k$ using the observed VIX and the cyan dashed lines represent the standard deviation in the $\hat{x}_k$. The histogram shows the empirical distribution of samples from the MH algorithm. The same plotting scheme applies to the red period during the low volatility era.

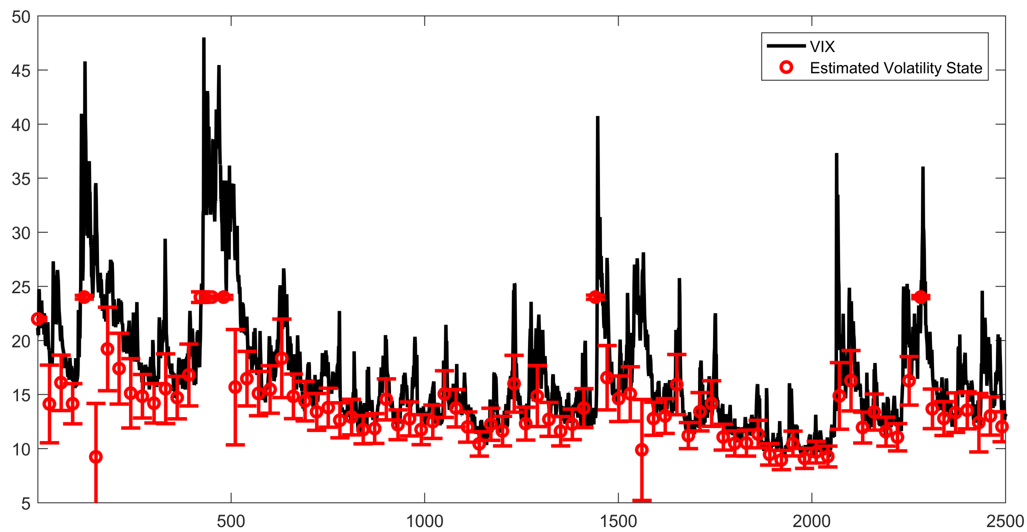


Figure 9: Estimated market volatility using VIX observations within a 2-week window. This estimated market volatility can be used to estimated the transition probability of underlying Markov Chain, P.

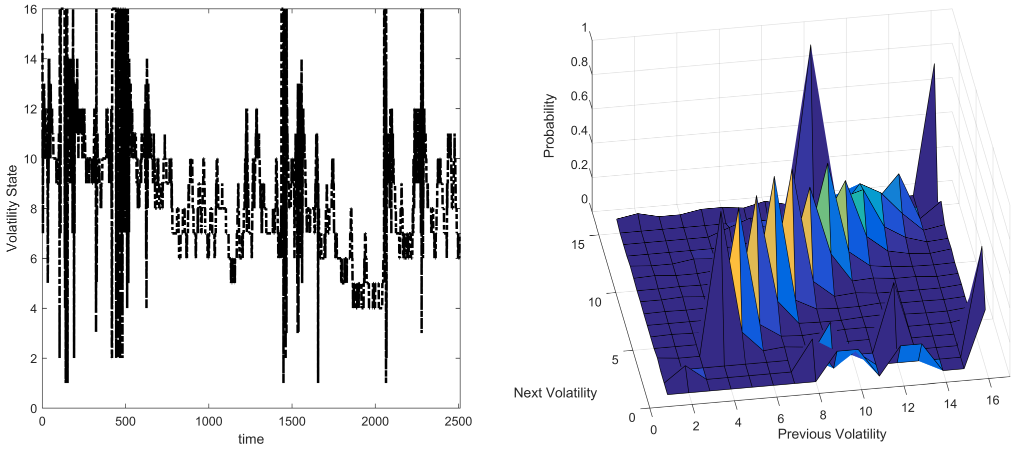


Figure 10: Estimated market volatility states and estimated the transition probability of underlying Markov Chain.

$P(\text{future} = 10|\text{now} = 7) = 0.0150$. The expected future volatility is 12.2, at the verge of high-volatility.

Interestingly, the stationary distribution of this simplified model for the market volatility is

$$\pi_\infty = \begin{bmatrix} P(\text{low}) \\ P(\text{medium}) \\ P(\text{high}) \end{bmatrix} = \begin{bmatrix} 0.6494 \\ 0.2673 \\ 0.0833 \end{bmatrix} \tag{51}$$

References

1. https://en.wikipedia.org/wiki/Volatility_(finance)

2. https://en.wikipedia.org/wiki/VIX

3. https://finance.yahoo.com/

Disclaimer

This is a purely simple demonstration and academic analysis. The reader may use the content at his/her/its own risk and the reader shall have no claims whatsoever against the author.