Econophysics - written exam 23 January 2013 (with solutions)

Docent: Dr. Diego Garlaschelli

You should return these pages completed. Please write your answers to all questions by filling the boxes. Motivate each of your answers. Answers are ignored if there is no explanation. Please write clearly and in English. Please don't forget to write your name and student ID below. Good luck!

NAME: . . . .

SURNAME: . . . .

STUDENT ID: . . . .

PROBLEM 1 - WITH ANSWERS

A stochastic process is defined as the temporal evolution of a random variable. Consider the stochastic process defined in terms of the random variable X that evolves as follows:

$$X(0) = 0$$

$$X(t) = X(t-1) + \eta(t)$$

where $t = 0, 1, 2, \ldots$ is (discrete) time, and η is a continuous random variable. Values of η at different times are drawn independently from the same probability density. We assume that this distribution has zero mean and standard deviation equal to $\sigma = 1/a$, where a is a real number.

Question 1.1

Consider a particular time $T > 0$. Write the quantity $X(T)$ as a function only of the values of η.

$X(T) = \eta(1) + \eta(2) + \cdots + \eta(T) = \sum_{t=1}^{T} \eta(t)$

Question 1.2

Write under which conditions for T and a the Central Limit Theorem can be applied to find the distribution of $X(T)$, and explain why in each case.

Condition(s) for T:

Since $X(T)$ is the sum of T independent and indentically distributed random variables $\eta(1), \ldots, \eta(T)$, the Central Limit Theorem can be applied to find the distribution of $X(T)$ if T is large.

Condition(s) for a:

Moreover, the variance $\sigma^2 = 1/a^2$ of each η must be finite, i.e. $a \neq 0$.
But, since the standard deviation σ cannot be negative, we must also have $\sigma = 1/a > 0$, i.e. $a > 0$.

Question 1.3

When the above conditions hold, write the approximate probability density for $X(T)$, i.e. the asymptotic probability that $X(T) = x$, in terms of a and T alone.

$P(x) = \frac{1}{\sqrt{2\pi Var[X(T)]}} e^{-\frac{(x-E[X(t)])^2}{2Var[X(T)]}} = \frac{a}{\sqrt{2\pi T}} e^{-\frac{(ax)^2}{2T}}$

Question 1.4

Now consider N independent stochastic processes, each defined as for $X(T)$ above:

$$X_i(0) = 0 \quad \forall i$$

$$X_i(t) = X_i(t-1) + \eta_i(t) \quad \forall i$$

where $i = 1, \dots N$ with N large and where all $\eta_i(t)'s$ are independent and identically distributed as discussed above. Based on the above results, discuss whether $X_i(t)$ can be considered a good model for how an individual person i accumulates their wealth through many random increments in real societies. Explain why, separately for the bulk and for the tail of the resulting distribution of wealth across all the N individuals.

Bulk (low and medium wealth):

$X_i(t)$ is a bad model, since it predicts a Gaussian wealth distribution while real wealth distributions have a log-normal bulk (Gibrat's distribution).

Tail (high wealth):

$X_i(t)$ is a bad model, since it predicts a Gaussian wealth distribution while real wealth distributions have a power-law tail (Pareto's distribution).

Question 1.5

Discuss whether $X_i(t)$, defined as above, can be considered a good model for how the log-price of a stock i evolves in a real market with N stocks. Note that, if $X_i(t)$ is the log-price, then $\eta_i(t)$ is the log-return. Answer separately (explain why, and if necessary under which conditions) for the shape of the distribution of log-returns, for the auto-correlation of the log-returns and absolute log-returns of each stock, and for the cross-correlations among the N stocks.

Distribution of log-returns (whether and why):

$X_i(t)$ can be a good model, if each η is drawn from a power-law distribution with finite variance $\sigma^2 = 1/a^2 < \infty$, as in real data (this also reproduces aggregate normality).

Autocorrelation of log-returns of a single stock (whether and why):

$X_i(t)$ is a good model, because it predicts that the log-returns $\eta(t)$ are independent and thus uncorrelated, in accordance with real data.

Autocorrelation of absolute log-returns of a single stock (whether and why):

$X_i(t)$ is a bad model, because it predicts that the absolute log-returns $|\eta(t)|$ are also independent and thus uncorrelated, in contrast with real data.

Cross-correlations (whether and why):

$X_i(t)$ is a bad model, since the independence of all η's implies no correlation among stocks (the cross-correlation matrix converges to random matrix theory), in contrast with real data.

PROBLEM 1 - EXPLANATIONS

Answer to question 1.1

Recursively,

$$X(T) = X(T-1) + \eta(T) = X(T-2) + \eta(T-1) + \eta(T) = \cdots = \eta(1) + \eta(2) + \cdots + \eta(T) = \sum_{t=1}^{T} \eta(t) \tag{1}$$

Answer to question 1.2

From the above formula, it is clear that $X(T)$ is the sum of T independent and indentically distributed random variables $\eta(1), \ldots, \eta(T)$. The Central Limit Theorem can be applied to find the distribution of $X(T)$ if T is large and if the variance $\sigma^2 = 1/a^2$ of each η is finite, i.e. if $a \neq 0$. Note however that, since the standard deviation σ cannot be negative, we must also have $\sigma = 1/a > 0$, i.e. $a > 0$.

Answer to question 1.3

When the Central Limit Theorem can be applied, it predicts that the asymptotic distribution of $X(T)$ is Gaussian with mean

$$E[X(T)] = \langle X(T) \rangle = \sum_{t=1}^{T} \langle \eta(t) \rangle = 0$$

(since all η's have zero mean) and variance

$$Var[X(T)] = \sum_{t=1}^{T} Var[\eta(t)] = \sum_{t=1}^{T} \sigma^2 = \frac{T}{a^2}$$

(since all covariances are zero, the η's being independent). Thus the probability that $X(T) = x$ is

$$P(x) = \frac{1}{\sqrt{2\pi Var[X(T)]}} e^{-\frac{(x-E[X(t)])^2}{2Var[X(T)]}} = \frac{a}{\sqrt{2\pi T}} e^{-\frac{(ax)^2}{2T}}$$

Answer to question 1.4

$X_i(t)$ is a bad model for how people accumulate their wealth, because the above results show that it leads to a Gaussian distribution $P(x)$ of wealth in the population. By contrast, in real societies the bulk of the wealth distribution is found to be log-normal (Gibrat's distribution) and the tail is found to be power-law (Pareto's distribution). So the model is bad in both respects.

Answer to question 1.5

With respect to the resulting distribution of log-returns, $X_i(t)$ can be a good model, if each η is drawn from a power-law distribution with finite variance $\sigma^2 = 1/a^2 < \infty$ as in real data (this also reproduces aggregate normality).

With respect to the autocorrelation of log-returns of each stock i, $X_i(t)$ is a good model, because it predicts that the log-returns $\eta(t)$ are independent and thus uncorrelated, in accordance with real data.

With respect to the autocorrelation of absolute log-returns of each stock i, $X_i(t)$ is a bad model, because it predicts that the absolute log-returns $|\eta(t)|$ are also independent and thus uncorrelated, in contrast with real data.

With respect to cross-correlations among different stocks, $X_i(t)$ is a bad model, since the independence of all η's implies that, for large T and N (assuming $T/N > 1$), the expected correlation matrix becomes in accordance with random matrix theory, in contrast with real data e.g. about the deviating eigenvalues or the market mode.

PROBLEM 2 - WITH ANSWERS

N firms produce goods by buying and selling intermediate products from and to each other. The resulting inter-firm transaction network is modelled as an (undirected) Erdos-Renyi random graph. Based on this model, the expected average nearest-neighbour degree is estimated as $\langle k^{nn} \rangle \approx 2$ and the expected average clustering coefficient as $\langle C \rangle \approx 0.01$.

Question 2.1

Determine the total number N of firms and the probability p that any two firms are connected.

$N = \frac{\langle k^{nn} \rangle}{\langle C \rangle} \approx \frac{2}{0.01} = 200$

$p = \langle C \rangle = 0.01$

Question 2.2

What can be concluded about the connected components of the transaction network?

In a large ($N \to \infty$) Erdos-Renyi random graph there exists a critical connection probability $p_c \approx 1/N$ such that, for $p < p_c$, the network is subdivided into many small connected components, and for $p > p_c$ there exists a *giant connected component* containing a finite fraction of all vertices. In this case, $p_c \approx 1/200$ while $p = 1/100$, so $p > p_c$ and there must be a giant connected component (the graph is above the so-called *percolation threshold*).

Question 2.3

Determine the average number $\langle k \rangle$ of firms each firm is connected to, the total expected number $\langle L \rangle$ of links, and the average firm-firm distance $\langle D \rangle$ in the network.

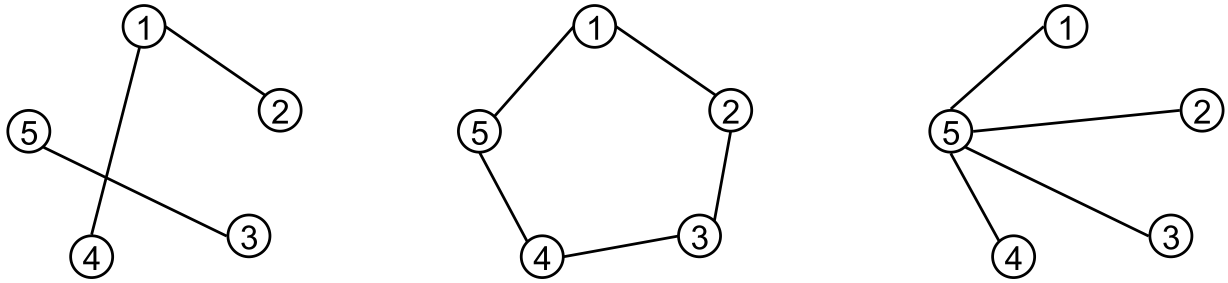
$\langle k \rangle = (N-1)p \approx 2$

$\langle L \rangle = \frac{N(N-1)}{2} p = \frac{N}{2} \langle k \rangle = 200$

$\langle D \rangle \approx \frac{\log(N)}{\log(\langle k \rangle)} = \frac{\log(200)}{\log(2)} \approx 7.64$

Question 2.4

Consider the following three possible configurations for the subgraph involving the first 5 firms only:



Subgraph 1 *Subgraph 2* *Subgraph 3*

Calculate the probability of occurrence for each of the three possibilities in the inter-firm network considered above.

$P(\text{subgraph 1}) = p^3(1-p)^{10-3} = \left(\frac{0.01}{0.99}\right)^3 \cdot 0.99^{10} \approx 9.32 \cdot 10^{-7}$

$P(\text{subgraph 2}) = p^5(1-p)^{10-5} = \left(\frac{0.01}{0.99}\right)^5 \cdot 0.99^{10} \approx 9.51 \cdot 10^{-11}$

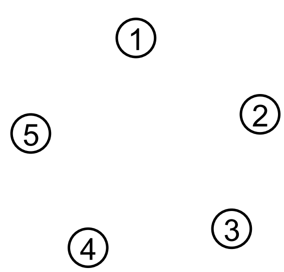
$P(\text{subgraph 3}) = p^4(1-p)^{10-4} = \left(\frac{0.01}{0.99}\right)^4 \cdot 0.99^{10} \approx 9.41 \cdot 10^{-9}$

Question 2.5

Now draw the most probable configuration for the subgraph involving the first 5 firms only, and calculate its probability.

Most probable subgraph:

Since, for $p < 1/2$, the probability decreases as the number of links in the subgraph increases, the most probable subgraph has zero links and is therefore empty:



$P(\text{most probable subgraph}) = p^0(1-p)^{10-0} = 0.99^{10} \approx 0.904$

PROBLEM 2 - EXPLANATIONS

Answer to question 2.1

In an Erdos-Renyi random graph, if p denotes the connection probability and N the number of vertices, the expected average clustering coefficient is $\langle C\rangle = p$ and the expected average nearest neighbour degree is $\langle k^{nn}\rangle = (N-1)p \approx Np$. Therefore the number of vertices can be obtained from these two quantities as follows:

$$N = \frac{\langle k^{nn}\rangle}{\langle C\rangle} \approx \frac{2}{0.01} = 200 \tag{2}$$

Similarly, the connection probability is simply

$$p = \langle C\rangle = 0.01 \tag{3}$$

Answer to question 2.2

In a large ($N \to \infty$) Erdos-Renyi random graph there exists a critical connection probability $p_c \approx 1/N$ such that, for $p < p_c$, the network is subdivided into many small connected components, and for $p > p_c$ there exists a *giant connected component* containing a finite fraction of all vertices. In this case, $p_c \approx 1/200$ while $p = 1/100$, so $p > p_c$ and there must be a giant connected component (the graph is above the so-called *percolation threshold*).

Answer to question 2.3

The average number $\langle k\rangle$ of firms each firm is connected to is the *average degree* of the network. For the Erdos-Renyi graph it can be obtained as

$$\langle k\rangle = (N-1)p \approx 2 \tag{4}$$

The total expected number $\langle L\rangle$ of links is simply

$$\langle L\rangle = \frac{N(N-1)}{2}p = \frac{N}{2}\langle k\rangle = 200 \tag{5}$$

The average firm-firm distance $\langle D\rangle$ is

$$\langle D\rangle \approx \frac{\log(N)}{\log(\langle k\rangle)} = \frac{\log(200)}{\log(2)} \approx 7.64 \tag{6}$$

Answer to question 2.4

In a subgraph with n vertices, there are $n(n-1)/2$ possible links. Therefore the probability of occurrence of a subgraph with n vertices and l links is the product of the probability p^l of l successes (existing links) and of the probability $(1-p)^{n(n-1)/2-l}$ of $n(n-1)/2-l$ failures (missing links):

$$P(subgraph) = p^l(1-p)^{n(n-1)/2-l} = \left(\frac{p}{1-p}\right)^l (1-p)^{n(n-1)/2} \tag{7}$$

Note that, for n fixed, this probability decreases as l increases if $p/(1-p) < 1$ (i.e. if $p < 1/2$), and it increases as l increases if $p/(1-p) > 1$ (i.e. if $p > 1/2$). For the three subgraphs considered here (where $p = 0.01$, $1-p = 0.99$, $n = 5$, and $n(n-1)/2 = 10$) the above probability reads

$$P(subgraph \quad 1) = \left(\frac{0.01}{0.99}\right)^3 \cdot 0.99^{10} \approx 9.32 \cdot 10^{-7} \tag{8}$$

$$P(subgraph \ 2) = \left(\frac{0.01}{0.99}\right)^5 \cdot 0.99^{10} \approx 9.51 \cdot 10^{-11} \tag{9}$$

$$P(subgraph\ \ 3) = \left(\frac{0.01}{0.99}\right)^4 \cdot 0.99^{10} \approx 9.41 \cdot 10^{-9} \tag{10}$$

Answer to question 2.5

Since in our case $p < 1/2$, the most probable subgraph has the minimum value of l, i.e. $l = 0$. This subgraph is therefore the empty graph. Its probability is

$$P(empty\ \ subgraph) = 0.99^{10} \approx 0.904 \tag{11}$$