# Problem Set #4

MACS 30200, Dr. Evans and Dr. Soltoff
Due Monday, May 28 at 5:30pm

1. **Discrete approximation of an AR(1) process (10 points).** Assume that a random variable evolves according to the following continuous AR(1) process,

$$z_{t+1} = \rho z_t + (1-\rho)\mu + \varepsilon_t \quad \text{s.t.} \quad \rho \in (-1,1) \quad \text{and} \quad \varepsilon_t \sim N(0,\sigma) \tag{1}$$

where $\rho$ governs the persistence of the process, $\mu$ is the long-run average of $z_t$, and $\sigma$ is the standard deviation of the normally distributed error terms. Assume that $\rho = 0.85$, $\mu = 11.4$, and $\sigma = 0.7$.

(a) Assume that $z_0 = \mu$. Simulate a time series of $T = 500$ periods of values of $\{z_t\}_{t=1}^T$ using (1) by drawing a vector of $T$ values from the normal $N(0,\sigma)$ above using the following code (so that all your vectors are the same). Plot the first 100 observations of the resulting simulated time series for $\{z_t\}_{t=1}^{100}$. [Reminder: I want you to plot the $z_t$'s, not the $\varepsilon_t$'s.]

```
import scipy.stats as sts

T = 500
sigma = 0.7
unif_vec = sts.uniform.rvs(loc=0, scale=1, size=T,
                           random_state=25)
eps_vec = sts.norm.ppf(unif_vec, loc=0, scale=sigma)
```

This method of drawing $T$ uniformly distributed values between 0 and 1 and then transforming them into the corresponding $N(0,\sigma)$ values through the inverse CDF function is a little bit indirect of a way to draw normally distributed values. However, we will use the vector of uniforms (`unif_vec`) in part (f) below.

(b) Create a 5-element vector called `z_vals` that represents a discretized version of all the values that $z_t$ can take on. Let `z_vals` be 5 evenly spaced points between $\mu-3\sigma$ and $\mu+3\sigma$. The third element of this vector `z_vals`[2] should equal $\mu$. This vector should nearly span all of the values that $z_t$ takes on in the time series from part (a).

(c) Estimate the probabilities of a $5 \times 5$ Markov transition matrix $\hat{\boldsymbol{P}}$ in the following way. Think of that values in the `z_vals` vector from part (b) as midpoints of bins that divide the whole space of points that $z_t$ can fall in. Define the cutoffs `z_cuts` between the bins as the 4-element vector of midpoints between each of the 5 points in `z_vals`.

```
z_cuts = 0.5 * z_vals[:-1] + 0.5 * z_vals[1:]
```

Then we can classify each data point in our simulated series $\{z_t\}_{t=1}^T$ as being in one of the five bins based on that points relative position to the bin cutoff values in `z_cuts`.

$$z_t \in \begin{cases} \text{bin 1 if} & z_t \leq \texttt{z\_cuts}[0] \\ \text{bin 2 if} & \texttt{z\_cuts}[0] < z_t \leq \texttt{z\_cuts}[1] \\ \text{bin 3 if} & \texttt{z\_cuts}[1] < z_t \leq \texttt{z\_cuts}[2] \\ \text{bin 4 if} & \texttt{z\_cuts}[2] < z_t \leq \texttt{z\_cuts}[3] \\ \text{bin 5 if} & z_t > \texttt{z\_cuts}[3] \end{cases}$$

The first row of the estimated Markov transition matrix $\hat{\boldsymbol{P}}$ represents the probability of moving from state 1 (bin 1) this period to each of the 5 states (bins) next period. More generally, $\hat{\boldsymbol{P}} \equiv Pr\left(z_{t+1} \in \text{bin}_k | z_t \in \text{bin}_j\right)$. Estimate the probabilities in each row of the Markov transition matrix $\hat{\boldsymbol{P}}$ as the empirical probabilities from the simulated data series $\{z_t\}_{t=1}^T$ from part (a). For example, the probability $\pi_{2,4}$ of transitioning from bin 2 in the current period ($\texttt{z\_cuts}[0] < z_t \leq \texttt{z\_cuts}[1]$) to bin 4 in the next period ($\texttt{z\_cuts}[2] < z_t \leq \texttt{z\_cuts}[3]$) is estimated to be the ratio of the number of two-period consecutive data points that start with a value in bin 2 and end with a value in bin 4 divided by the total number of two-period consecutive segments that start in bin 2.

(d) According to your estimated Markov transition matrix $\hat{\boldsymbol{P}}$ from part (c), what is the probability of $z_{t+3}$ being in bin 5 ($z_{t+3} > \texttt{z\_cuts}[3]$) given that $z_t$ is in bin 3 ($\texttt{z\_cuts}[1] < z_t \leq \texttt{z\_cuts}[2]$) today? [Hint: Start with a vector $[0, 0, 1, 0, 0]$.]

(e) According to your estimated Markov transition matrix $\hat{\boldsymbol{P}}$ from part (c), what is the stationary (long-run, ergodic) distribution of $z_t$ (i.e., the percentages of the time that the random variable spends in each of the 5 bins)?

(f) Use the vector of $T$ uniformly distributed variables in `unif_vec` from part (a) to simulate a time series of $T$ values of the discretized version of $z_t \in$ `z_vals` using the estimated transition matrix $\hat{\boldsymbol{P}}$ and an initial value $z_0 = \texttt{z\_vals}[2]$. Plot the time series of this discretized series for $z_t$ versus the continuous version from part (a). Make sure your plot has a legend, title, and labeled axes. To be clear, your discretized time series for $z_t$ should alternate randomly among only the five values in the vector `z_vals`. However, this time series should have many of the same properties as the continuous time series $z_t$ from part (a).