CME 308 Final Exam
Spring 2009

Peter Glynn

June 11, 2009

Please sign the Stanford Honor Code on your blue book.

This exam is open notes and open book. You have three hours. There are a total of 57 points. Good luck!

Problem 1: 25 Points

Suppose that we collect scalar observations $Y_1, Y_2, \ldots, Y_n$ at (scalar) values $x_1, x_2, \ldots, x_n$ of the independent variable, respectively. We assume that

$$Y_i = g(x_i) + \epsilon_i,$$

where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are i.i.d. $\mathcal{N}(0, \sigma_*^2)$ r.v.'s. The function $g(\cdot)$ is assumed to be *non-decreasing*.

(a) (2 points) Write down the likelihood function for this problem. (Hint: View the $g(x_i)$'s as parameters.)

(b) (5 points) Show that the maximum likelihood estimator is the solution of a convex optimization problem.

(c) (5 points) Suppose that we need a 95% (approximate) confidence interval for $g(x_i)$. Provide a sampling based algorithm for computing such a confidence interval.

(d) (5 points) Suppose that we require an (approximate) 90% prediction interval for Y at an x-value for which no observations have yet been taken and $x \in [\min_{1\le i\le n} x_i, \max_{1\le i\le n} x_i]$. Provide an algorithm for computing such a prediction interval.

(e) (5 points) Suppose that we change the model so that $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are i.i.d. r.v.'s with common density

$$f(x) = \frac{\lambda^*}{2}\exp(-\lambda^*|x|).$$

Show that the maximum likelihood estimator is the solution of a linear program.

(f) (3 points) Suppose that we change the original formulation so that

$$Y_i = g(x_i) Z_i,$$

where Z_i is a log-normal r.v. with parameters 0 and σ_*^2. The Z_i's are assumed i.i.d. and g is non-decreasing. How would you compute the maximum likelihood estimator in this context?

Solution:

(a) The model states that

$$Y_i = g*(x_i) + \epsilon_i,$$

where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $\mathcal{N}(0, \sigma_*^2)$ r.v.'s. The unknown parameters are $g^*(x_1), \ldots, g^*(x_n)$, and σ_*^2. The likelihood function is

$$L_n(g(x_1), \ldots, g(x_n), \sigma^2 | Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(Y_i - g(x_i))^2)$$

and the log-likelihood is

$$\mathcal{L}_n(g(x_1), \ldots, g(x_n), \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(Y_i - g(x_i))^2}{2\sigma^2}$$

(b) Let $y_i = g(x_i)$, for $1 \le i \le n$. Note that because g^* is non-decreasing, the appropriate parameter space over which to maximize $\mathcal{L}_n$ is

$$\mathcal{P} = \{(y_1, \ldots, y_n, \sigma^2) : y_1 \le y_2 \le \cdots \le y_n, \sigma^2 \ge 0\}.$$

Hence, we minimize

$$-\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(Y_i - y_i)^2}{2\sigma^2}$$

subject to $(y_1, \ldots, y_n, \sigma^2) \in \mathcal{P}$. This is equivalent to solving the minimization problem

$$\begin{array}{rrcl} \text{minimize} & \multicolumn{3}{l}{\sum_{i=1}^{n}(Y_i - y_i)^2} \\ \text{s.t.} & y_1 & \le & y_2 \\ & y_2 & \le & y_3 \\ & & \vdots & \\ & y_{n-1} & \le & y_n \end{array}$$

If $(\hat{y}_1, \ldots, \hat{y}_n)$ is a minimizer of the above convex optimization problem (actually, a quadratic programming problem), the MLE's for $g^*(x_i)$, $1 \le i \le n$ are $\hat{y}_i$ and the MLE for σ_*^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - y_i)^2.$$

(c) Parametric Bootstrap:

1. Sample n i.i.d. normal r.v.'s $\epsilon_1^*, \ldots, \epsilon_n^*$ with mean zero and variance $\hat{\sigma}^2$.
2. Set $Y_i^* = \hat{y}_i + \epsilon_i^*$ for $1 \le i \le n$.
3. Compute the minimizer $(\hat{y}_1^*, \ldots, \hat{y}_n^*)$ of

$$\begin{array}{rrcl} \text{minimize} & \multicolumn{3}{l}{\sum_{i=1}^{n}(Y_i^* - y_i)^2} \\ \text{s.t.} & y_1 & \le & y_2 \\ & y_2 & \le & y_3 \\ & & \vdots & \\ & y_{n-1} & \le & y_n \end{array}$$

   set $\hat{g}^*(x_i) = \hat{y}_i^*$.
4. Repeat steps 1 to 3 m i.i.d. times, thereby producing m i.i.d. copies $\hat{g}_1^*(x_i), \ldots, \hat{g}_m^*(x_i)$ of the r.v. $\hat{g}^*(x_i)$.

5. Let
$$\widehat{F}_m(x) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{\{\hat{g}_j^*(x_i)-\hat{y}_i \le x\}}$$
and find z_1, z_2 such that
$$\widehat{F}_m(z_1) = 0.025, \quad \widehat{F}_m(z_2) = 0.975.$$
6. Then, an approximate 95% condfidence interval for $g^*(x_i)$ is
$$[\hat{y}_i - z_2, \hat{y}_i - z_1].$$

(d) We adopt the reasonable assumption that g^* is piecewise linear over each of the $n-1$ intervals $[x_i, x_{i+1}]$, $1 \le i \le n-1$. (You may also reasonably assume that g^* is piecewise constant over each such interval.) Note that any Y observed at x will then satisfy
$$Y = g^*(x) + \sigma_* \epsilon,$$
where ϵ is a normal r.v. with mean zero and variance 1 (and independent of $Y_1, \dots, Y_n$). The r.v. Y has distribution
$$\mathbf{P}(Y \in A) = \int_A \frac{1}{\sigma_*}\phi\left(\frac{y - g^*(x)}{\sigma_*}\right) dy, \tag{1}$$
where ϕ is the density of a $\mathcal{N}(0,1)$ r.v. The prediction interval should take into account the sampling error in estimating $g^*(x)$ and σ_* from the (finite) sample $Y_1, \dots, Y_n$, so we replace (1) by
$$\iiint_{\mathbb{R}\times\mathbb{R}_+\times A} \frac{1}{\sigma}\phi\left(\frac{y-z}{\sigma}\right) \mathbf{P}(\hat{g}(x) \in dz, \hat{\sigma} \in d\sigma). \tag{2}$$
To compute (2), we run steps 1-4 of the parametric bootstrap algorithm of part (c). Let $\hat{g}_j^*(x)$ be the linear interpolant at point x that interpolates between the neighboring x_i's and let
$$\hat{\sigma}_j^* = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(Y_{ij}^* - \hat{y}_{ij}^*)^2}$$
(where Y_{ij}^* is the j-th bootstrap copy of Y_i^* and $\hat{y}_{ij}^*$ is the j-th bootstrap copy of $\hat{y}_i^*$). The 90% prediction interval is the interval $[y_1, y_2]$, where
$$\frac{1}{m}\sum_{j=1}^{m}\int_{y_2}^{\infty} \frac{1}{\hat{\sigma}_j^*}\phi\left(\frac{y - \hat{g}_j^*(x)}{\hat{\sigma}_j^*}\right) dy = 0.05$$
and
$$\frac{1}{m}\sum_{j=1}^{m}\int_{-\infty}^{y_1} \frac{1}{\hat{\sigma}_j^*}\phi\left(\frac{y - \hat{g}_j^*(x)}{\hat{\sigma}_j^*}\right) dy = 0.05$$

(e) Here, the likelihood function becomes
$$L_n(g(x_1), \dots, g(x_n), \lambda) = \prod_{i=1}^{n} \frac{\lambda}{2}\exp(-\lambda|Y_i - g(x_i)|)$$
and the log-likelihood is
$$\mathcal{L}_n(g(x_1), \dots, g(x_n), \lambda) = n\log(\lambda/2) - \lambda\sum_{i=1}^{n}|Y_i - g(x_i)|.$$

As in the Gaussian case, we solve

$$\begin{array}{rrcl} \text{minimize} & \sum_{i=1}^{n} |Y_i^* - y_i| & & \\ \text{s.t.} & y_1 & \leq & y_2 \\ & y_2 & \leq & y_3 \\ & & \vdots & \\ & y_{n-1} & \leq & y_n \end{array} \tag{3}$$

The MLE's, for $g^*(x_1), \ldots, g^*(x_n)$ is the minimizer $(\hat{y}_1, \ldots, \hat{y}_n)$ of (3). Of course, (3) can be reformulated as

$$\begin{array}{rrcll} \text{minimize} & \sum_{i=1}^{n} (w_i^+ + w_i^-) & & & \\ \text{s.t.} & w_j^+ - w_j^- & \geq & w_{j-1}^+ - w_{j-1}^-, & 1 \leq j \leq n \\ & w_i^+, w_i^- & \geq & 0, & 1 \leq i \leq n \end{array} \tag{4}$$

where we recover y_i by setting $y_i = w_i^+ - w_i^-$.

(f) Note that

$$\log Y_i = \log g(x_i) + \log z_i$$

for $1 \leq i \leq n$. Set $\widetilde{Y}_i = \log Y_i$, $\hat{g}(x_i) = \log g(x_i)$ and $\epsilon_i = \log z_i$. Then,

$$\widetilde{Y}_i = \hat{g}(x_i) + \epsilon_i,$$

where $\hat{g}$ is non-decreasing and the ϵ_i's are i.i.d. normal r.v.'s with mean 0 and variance σ_*^2. So, we have reduced the problem to the original formulation of part (a). Compute the MLE's for the transformed problem, and exponentiate the estimated $\hat{g}(x_i)$'s to estimate the $g(x_i)$'s.

Problem 2: 12 Points

Suppose we are given the autoregressive model

$$Y_n = -\frac{1}{12} Y_{n-1} + \frac{1}{2} Y_{n-2} + 1 + \epsilon_n,$$

where the ϵ_i's are i.i.d. $\mathcal{N}(0,1)$ r.v.'s. Suppose also that $Y_0 = 3$, $Y_1 = 3$, and $Y_2 = 2$.

(a) (3 points) Is $(Y_n : n \geq 0)$ a stable autoregressive process in the sense that there exists a (finite-valued) r.v. Y_∞ such that

$$Y_n \xrightarrow{\mathcal{D}} Y_\infty,$$

as $n \to \infty$?

(b) (3 points) Compute the conditional distribution of Y_3, given Y_0, Y_1, and Y_2 as above.

(c) (2 points) Compute $\mathbf{E}[Y_4 | Y_0 = 3, Y_1 = 3, Y_2 = 2]$.

(d) (4 points) Compute the conditional distribution of $Y_3 + Y_4$, given Y_0, Y_1, and Y_2 as above.

Solution:

(a) Put it in vector form:

$$\begin{pmatrix} Y_{n-1} \\ Y_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{1}{12} \end{pmatrix} \begin{pmatrix} Y_{n-2} \\ Y_{n-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ \epsilon_n \end{pmatrix}.$$

For stability, we need the coefficient matrix F to have spectral radius less than 1. But the eigenvalues of F are $\frac{2}{3}$ and $-\frac{3}{4}$. So, the spectral radius is less than one and $Y_n \xrightarrow{\mathcal{D}} Y_\infty$ as $n \to \infty$.

(b) Conditoinal on $Y_0 = 3, Y_1 = 3$, and $Y_2 = 2$,

$$Y_3 = -\frac{1}{12}(2) + \frac{1}{2}(3) + 1 + \epsilon_3 = \frac{7}{3} + \epsilon_3,$$

so the conditional distribution of Y_3 is normal with mean $\frac{7}{3}$ and variance 1.

(c) We need to compute $\widehat{Y}_{4|2}$. Note that

$$\begin{aligned}
\widehat{Y}_{4|2} &= -\frac{1}{12}\widehat{Y}_{3|2} + \frac{1}{2}\widehat{Y}_{2|2} + 1 \\
&= -\frac{1}{12}\frac{7}{3} + \frac{1}{2}(2) + 1 \\
&= \frac{65}{36}
\end{aligned}$$

(d) We know that the conditional distribution will be Gaussian. So, we only need to compute the conditional expectation and conditional variance to characterize the conditional distribution. But

$$\widehat{Y}_{3|2} = \frac{7}{3}$$

and

$$\widehat{Y}_{4|2} = \frac{65}{36}$$

so

$$\widehat{Y}_{3|2} + \widehat{Y}_{4|2} = \frac{84 + 65}{36} = \frac{149}{36}.$$

Conditional on $Y_0 = 3, Y_1 = 3$, and $Y_2 = 2$,

$$Y_4 = -\frac{1}{12}Y_3 + \frac{1}{2}(2) + 1 + \epsilon_4$$

so, conditional on $Y_0 = 3, Y_1 = 3$, and $Y_2 = 2$,

$$Y_3 + Y_4 = \frac{11}{12}Y_3 + 2 + \epsilon_4$$

Hence, the conditional variance of the sum is

$$\left(\frac{11}{12}\right)^2 \mathrm{Var}(Y_3|Y_0 = 3, Y_1 = 3, Y_2 = 2) + 1 = \frac{121}{144}(1) + 1 = \frac{265}{144}.$$

So, $Y_3 + Y_4$ is normal with mean 149/36 and variance 265/144, conditional on $Y_0 = 3, Y_1 = 3$, and $Y_2 = 2$.

Problem 3: 30 Points

Consider a continuous time system in which:

- arrivals to the system occur at rate λ
- two servers are available
- the system has an infinite capacity waiting room
- customers are served in the order in which they arrive
- customers choose the first available server
- if both servers are available, the arriving customer chooses each server with equal probability 1/2.

- customers that wait each abandon the system at rate γ

(a) (4 points) Formulate this model as a Markov jump process when both servers have identical service rate μ. What is the state space of your model?

(b) (2 points) (Continuation of (a).) Is your Markov chain irreducible?

(c) (3 points) (Continuation of (a).) What is the rate matrix for your model?

(d) (4 points) (Continuation of (a).) Compute the equilibrium distribution for your model when $\gamma = 0$.

(e) (4 points) (Continuation of (d).) What is the expected number of customers waiting in equilibrium?

(f) (5 points) (Continuation of (d).) Suppose there are currently three customers in the system. What system of equations would you solve in order to compute the expected time for the system to empty?

(g) (4 points) (Continuation of (d).) Suppose that server 1 is faster than sever 2, so that $\mu_1 > \mu_2$. Formulate this as a Markov jump process by providing the state space and transition rate graph for your new model.

(h) (4 points) (Continuation of (g).) Suppose that server 1 fails at rate β when it is operational. Once it fails, it is out of service forever. Provide the state space and transition rate graph of your new model.

Solution:

(a) We model the number-in-system process as a Markov jump as a Markov jump process with states $0, 1, 2, \ldots$ and transition rate graph

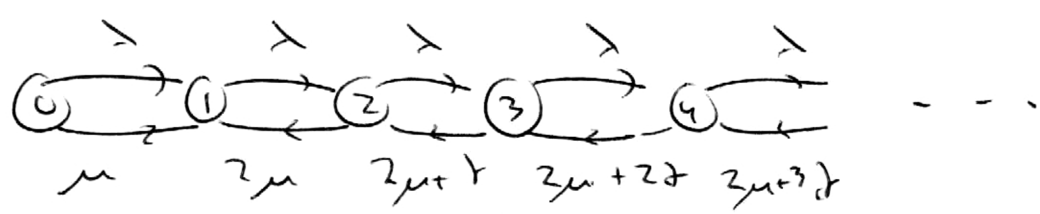

Figure 1: Transition Graph for Problem 3, part (a)

(b) Yes

(c) The rate matrix is

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdots \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & \cdots \\ \vdots & & & \ddots & \end{pmatrix}$$

(d) When $\gamma = 0$,

$$\pi(x) = \pi(0)\frac{\lambda^x}{2^{x-1}\mu^x}$$

for $x \geq 1$. In order for the $\pi(x)$'s to sum to one, λ must be less than 2μ, in which case

$$\begin{aligned}\pi(0) &= \left(1 + \sum_{x=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^x \left(\frac{1}{2}\right)^{x-1}\right)^{-1} \\ &= \left(1 + \frac{\lambda}{\mu}\left(1 - \frac{\lambda}{2\mu}\right)^{-1}\right)^{-1} \\ &= \left(1 + \frac{2\lambda}{2\mu - \lambda}\right)^{-1} \\ &= \frac{2\mu - \lambda}{2\mu + \lambda}.\end{aligned}$$

Hence, if $\lambda < 2\mu$,

$$\pi(x) = \left(\frac{2\mu - \lambda}{2\mu + \lambda}\right)\left(\frac{\lambda}{\mu}\right)^x \left(\frac{1}{2}\right)^{x-1}$$

for $x \geq 1$; if $\lambda \geq 2\mu$, there is no equilibrium.

(e) We need to compute

$$\begin{aligned}\sum_{x=2}^{\infty}(x-2)\pi(x) &= \sum_{x=2}^{\infty} 2(x-2)\left(\frac{2\mu - \lambda}{2\mu + \lambda}\right)\left(\frac{\lambda}{2\mu}\right)^x \\ &= 2\left(\frac{2\mu - \lambda}{2\mu + \lambda}\right)\left(\frac{\lambda}{2\mu}\right)^2 \sum_{y=0}^{\infty} y \left(\frac{\lambda}{2\mu}\right)^y \\ &= \frac{1}{2\mu + \lambda}\frac{\lambda^2}{\mu}\sum_{y=0}^{\infty} y\left(\frac{\lambda}{2\mu}\right)^y \left(1 - \frac{\lambda}{2\mu}\right) \\ &= \frac{1}{2\mu + \lambda}\frac{\lambda^2}{\mu}\frac{\lambda}{2\mu - \lambda}\end{aligned}$$

where the last line comes from the mean of a geometric r.v.

(f) Let $\tau = \inf\{x \geq 0 : X(t) = 0\}$ we need to compute $\mathbf{E}[\tau | X(0) = 3]$. Set

$$u^*(x) = \mathbf{E}_x T.$$

The function $(u^*(x) : x \geq 0)$ satisfies the linear system

$$Qu = -1$$

subject to $u(0) = 0$. In this setting, this linear system becomes

$$\lambda u(x+1) - (\lambda + 2\mu)u(x) + 2\mu u(x-1) = -1$$

for $x \geq 2$ and

$$\lambda u(2) - (\lambda + \mu)u(1) = -1.$$

(g) When there are two or more customers in the system, both servers are utilized and the aggregate departure rate with $x \geq 2$ customers in the system is $\mu_1 + \mu_2$. When there is exactly one customer in the system, we need to differentiate on the basis of which server is being utilized by that remaining customer. So, call $(1, 1)$ the state in which there is 1 customer in the system being served by the fast (μ_1) server and $(1, 2)$ the state in which there is 1 customer in the system being served by the slow (μ_2) server. The transition rate graph is:

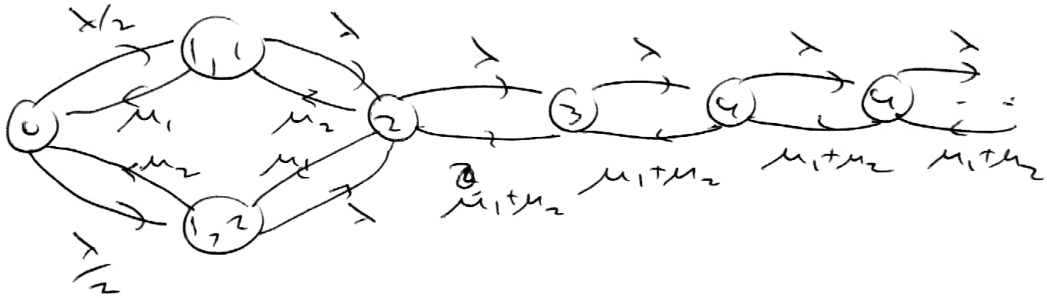

Figure 2: Transition Graph for Problem 3, part (g)

(h) We now need to know whether the first server is available/operational or not. Here are our new states:

$(0, A)$:	No customers in the in system; server 1 available.
$(0, N)$:	No customers in the system; server 1 not available.
$(1, 1)$:	1 customer in the system being served by the available server 1.
$(1, 2)$:	1 customer in the system being served by server 2 with server 1 available.
$(1, N)$:	1 customer in the system with server 1 not available.
(k, A) :	k customers in the system with server 1 available.
(k, N) :	k customers in the system with server 1 not available.

The transition rate graph is

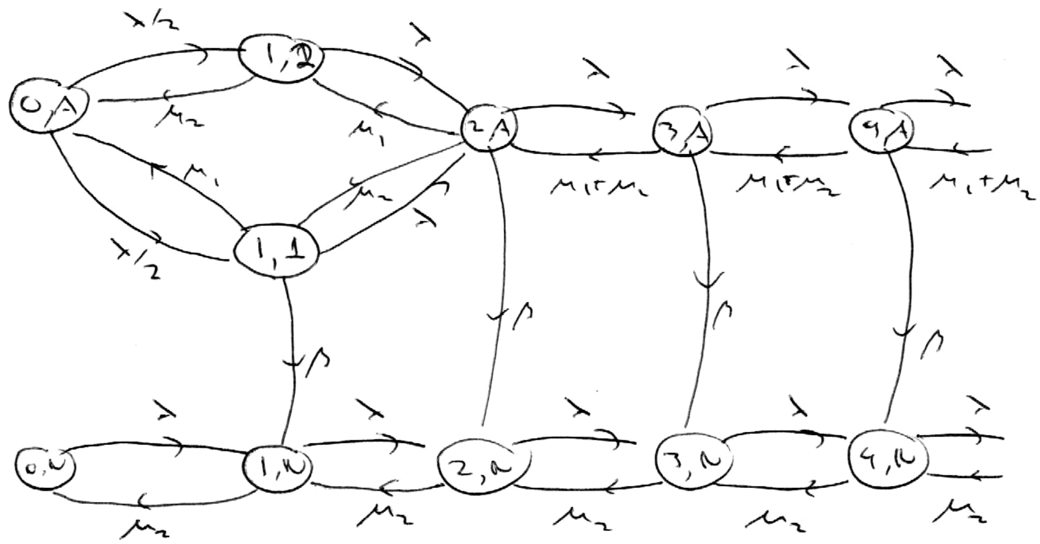

Figure 3: Transition Graph for Problem 3, part (h)

Problem 4: 16 Points

Consider an n-period decision problem in which we have the option, at the start of period i, $i = 0, 1, \ldots, n-1$, to either invest our wealth in the equity markets or to invest in treasuries. If we invest in treasuries, the log of the return in period i is 0.01. The one period log-return in equities is described by a 2-state Markov chain

$(R_i : i \geq 0)$ with states $x_1 = -0.01$ and $x_2 = 0.02$. The transition matrix for the equity log-returns is

$$P = \begin{array}{cc} \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} & \begin{array}{c} x_1 \\ x_2 \end{array} \\ \begin{array}{cc} x_1 & x_2 \end{array} & \end{array}$$

In any given time period, we invest either entirely in equities or entirely in treasuries. Furthermore, there is a transaction cost that is levied in proportion to the amount traded. In particular, if your current wealth at the beginning of period i is W_i and you trade from one asset to another, you pay $(1 - e^{-0.005})W_i$ in transaction costs so you will be left with wealth $e^{-0.005}W_i$. You will then make returns in period i from the new asset on $e^{-0.005}W_i$ instead of W_i.

Finally, the log-return for period $i-1$ is observable at the beginning of period i, in congruence with the Markov chain $(R_i : i \geq 0)$. Suppose that $R_{-1} = 0.02$. Our goal is to maximize our expected log-wealth at the end of period $n-1$.

(a) (3 points) What discrete state space Markov chain can you use to model the log-wealth?

(b) (6 points) Write down a backwards recursion that will allow you to compute an optimal investment policy.

(c) (1 point) How would you compute the optimal control for period 0, based on the solution computed from part 2's backwards recursion?

(d) (6 points) Adjust the formulation so that we maximize expected wealth instead.

Solution:

(a) This problem can be modeled in several different ways. But one particularly attractive approach is to recognize that when one is maximizing the log-wealth at the end of period n, this "final-value" problem is equivalent to an "additive reward" problem in which we cumulate the changes in the log-wealth over each of the n periods. Also, since the appreciation on the assets and the trading costs are proportional to the wealth, we can assume (without loss of generality) that we start with a wealth of \$1 (and then just scale the policy based on the magnitude of the actual wealth).

One possible choice of state at the beginning of period i is (x, j) where x is the equity return in period $i-1$ and j is the asset class in which wealth was invested in period $i-1$, where $j = 1$ corresponds to equities and $j = 2$ corresponds to treasuries.

(b) Let $V_i(x, j)$ be the value function corresponding to the start of period i, assuming that equity return in the previous period was x and wealth invested in asset class j. Then, if $x_1 = -0.01$ and $x_2 = 0.02$

$$\begin{aligned} V_i(x,1) = \max\{&P(x,x_1)[(-0.01) + V_{i+1}(x_1,1)] \\ &+ P(x,x_2)[(0.02) + V_{i+1}(x_2,1)], \\ &P(x,x_1)[(0.005) + V_{i+1}(x_1,2)] \\ &+P(x,x_2)[(0.005) + V_{i+1}(x_2,2)]\} \end{aligned}$$

$$\begin{aligned} V_i(x,2) = \max\{&P(x,x_1)[(0.01) + V_{i+1}(x_1,2)] \\ &+ P(x,x_2)[(0.01) + V_{i+1}(x_2,2)], \\ &P(x,x_1)[(0.015) + V_{i+1}(x_1,1)] \\ &+P(x,x_2)[(0.015) + V_{i+1}(x_2,1)]\} \end{aligned}$$

with $V_n(x, j) = 0$.

(c) Having computed $V_{n-1}, V_{n-2}, \ldots, V_0$, choose the asset class which maximizes the right hand side of the equation defining V_0. For example, if we are invested in equities at the start of period 0, stay in equities for period 0 if

$$P(x_2, x_1)[-(0.01) + V_1(x_1, 1)] + P(x_2, x_2)[(0.02) + V_1(x_2, 1)] \geq \\ P(x_2, x_1)[(0.005) + V_1(x_1, 2)] + P(x_2, x_2)[(0.005) + V_1(x_1, 2)]$$

(d) In maximizing expected wealth rather than log-wealth, we now need to track our cumulative log-wealth over time and then exponentiate at the end of the n-th period to obtain the final period wealth. We now need to add an additional state variable to accumulate the cumulative log-wealth obtained so far. Since log-wealth changes in increments of 0.005, we use the state description (s, x, j) where s is cumulative log-wealth accumulated to the end of period $i-1$ where s is such that $s \times 0.005$ cumulative log-wealth, x is equity return in period $i-1$, and j is the asset class in which wealth invested in period $i-1$.

$$\begin{aligned} V_i(s, x, 1) = \max\{&P(x, x_1)[(s - 0.01) + V_{i+1}(x_1, 1)] \\ &+ P(x, x_2)[(s + 0.02) + V_{i+1}(x_2, 1)], \\ &P(x, x_1)[(s + 0.005) + V_{i+1}(x_1, 2)] \\ &+P(x, x_2)[(s + 0.005) + V_{i+1}(x_2, 2)]\} \end{aligned}$$

and similarly for $V_i(s, x, 2)$. Here, $V_n(s, x, j) = \exp(s)$.

Problem 5: 13 Points

Consider a system in which two classes of customers are processed: high priority and low priority. There is a single server that can process work at unit rate. When customers arrive, they join one of two queues; the high priority queue or the low priority queue. Customers in each class are served in the order in which they arrive. When a customer completes service, the server serves the next high priority customer, if there is one. Otherwise, the server serves the next low priority customer. Suppose that high priority customers arrive at $t = 4.7, 5.9, 12.3, 15.9, 27.1$, with service time requirements $2.6, 3.1, 2.9, 8.1$, and 1.7. The low priority customers arrive at times $2.9, 5.3, 7.2, 19.1, 30.4$, with service time requirements $4.9, 11.1, 0.7, 0.5$, and 10.4.

(a) (7 points) How many customers are in the system at time 14.73?

(b) (3 points) We run 4 independent simulations of the system. The number of customers in the system at $t = 14.73$ in the 4 simulations are 1, 3, 2, and 2, respectively. Give an approximate 90% confidence interval for the mean number of customers in the system at $t = 14.73$. ($\mathbf{P}(\mathcal{N}(0, 1) > 1.96) = 0.025$, $\mathbf{P}(\mathcal{N}(0, 1) < -1.64) = 0.05$).

(c) (3 points) Suppose that the expected value of the high priority service time requirements is 4. Explain how you would take advantage of this information to create a new estimator for the mean number of customers in the system at $t = 14.73$ having a lower variance than the estimator above.

Solution:

(a) There are 3 customers in the system at $t = 14.73$.

(b) The sample mean is $\overline{X}_4 = \frac{1}{4}(1 + 3 + 2 + 2) = 2$. The sample variance is $s^2 = \frac{1}{3}((1 - 2)^2 + (3 - 2)^2 + (2 - 2)^2 + (2 - 2)^2) = \frac{2}{3}$. The 90% CI is $[1.33, 2.67]$.

(c) (From David Fong) Use a control variate:

For the i-th simulation, we record the sample mean service time of high priority customers

$$\overline{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_i j,$$

where Y_{ij} is the service time of the j-th high priority customer in the i-th simulation and N_i is the number of high priority customers. Clearly $\mathbf{E}[\overline{Y}_i] = 4$ for all i, so we set the control variate as $C_i = \overline{Y}_i - 4$ and let $\widehat{X}_i = X_i - \lambda C_i$, where X_i is the number of customers in the system at $t = 14.73$ for the i-th simulation, be the new estimator. The optimal parameter λ is given by

$$\lambda = \frac{\sigma_{XC}}{\sigma_C^2}$$

where σ_{XC} is the covariance of X and C and σ_C^2 is the variance of C. Thus, we know that $\widehat{X}_i$ is an unbiased estimator and that $\mathrm{Var}\,(\widehat{X}_i) \leq \mathrm{Var}\,(X_i)$.

To employ this technique, we shall use the first simulation to compute an estimate

$$\hat{\lambda} = \frac{S_{XC}}{S_C^2},$$

of λ, where S_{XC} is the sample covariance of X and C and S_C^2 is the sample variance of C. Discarding this simulation from estimating the number-in-system (to avoid estimation bias), we use $\hat{\lambda}$ in all subsequent simulations to compute $\widehat{X}_i$.