Examples of stochastic modeling and analysis in economics

Tore Schweder, March 24, 2011

Contents

1 Introduction

This compendium is a supplement to Taylor and Karlin (1998) which is used in ECON5160 *Stochastic modeling and analysis* as an introduction to dynamic stochastic modeling useful in theoretical economy and econometrics. Reference is made to Taylor and Karlin (1998) throughout in the format TK (section/page/...). There are many good textbooks on probability theory. In addition to those mentioned in Taylor and Karlin (1998), Grimmet and Stirzaker (1992) provides a good and condensed read. Stokey and Lucas (1989) and many other books in economics provides theory and application of probability theory. Wikipedia is useful, and mostly sound for probability theory.

The first Section is a warm up of some calculus on continuous distributions in the context of climate change. Then some material on specific continuous distributions. In the examples to follow I have given my own twist to stuff I have picked up in seminars and papers. The Section on Markov chains with continuous state space and MCMC (Markov Chain Monte Carlo) is included because it really is a beautiful and very influential bit of stochastic analysis – Bayesian methods are in ever more use, also in econometrics, and empirical Bayesian analyzes is often based on running an MCMC. The theory of stochastic differential equations and Ito integration is briefly sketched. The final Section is on some matrix theory. Eigen- and singular value decomposition is useful in many contexts. It provides e^A, the anti-log of a matrix A to be calculated in most cases. This so-called diagonalization is useful for solving systems of differential equations, also stochastic differential equations.

This compendium should be read as an invitation to further reading, and hopefully making a jump into the deep water of stochastic modeling and analysis.

2 Is expected damage due to global temperature rise infinite?

Martin Weitzman (2008, 2009) argues that uncertainty with respect to a key parameter in climate models is so big that levels leading to extreme global temperature are improbable, but not more improbable than making expected damage infinite when the dis-utility of global warming is at least as convex as squared temperature rise. He puts up a very simple deterministic climate model with a closed form solution. The trajectory of global temperature depends on a few parameters, one of which is the positive force of feedback from temperature to radiation. The higher the global temperature, the more of the solar radiation will be absorbed by the earth (oceans, landmasses, atmosphere), and then in turn the higher the global temperature. We shall have a look at his deterministic model, and the feed-back factor. The uncertainty distribution for this parameter is then transformed into an uncertainty distribution in limiting global temperature. We shall also have a look at expected discounted damage due to global temperature rise, and argue that this is infinite. In this simple model the expected damage turn out infinite if man-made greenhouse gases are not entirely removed from the atmosphere.

Let

$R(t)$ = additional (to pre-industrial) solar radiation, and

$T(t)$ = global temperature (degrees Celsius), increase from pre-industrial level at time t measured in years since the industrial revolution.

Global temperature is partially regarded as a sink, determined by the differential equation

$$\frac{d}{dt}T(t) = \frac{1}{c}\left[\lambda_0 R\left(t\right) - T(t)\right] \tag{1}$$

where
$c =$ aggregate thermal inertia (overall planetary capacity of the oceans to take up heat)
$\lambda_0 =$ a feedback-free constant.
The radiation is in turn affected by temperature according to the equation

$$R\left(t\right) = F\left(t\right) + \frac{f}{\lambda_0} T(t) \tag{2}$$

where
$F\left(t\right) =$ exogenously imposed additional radiative forcing (due to greenhouse gas emissions etc.)
$f =$ feedback factor.
With no extra forcing, $F = 0$, $T(t) = 0$ solves the equations when the initial condition is $T(0) = 0$. Weitzman assumes that due to emission of greenhouse gases the additional radiative forcing grows to a level $\overline{F}$ according to the differential equation

$$\frac{d}{dt} F(t) = \beta \left[\overline{F} - F(t)\right] \tag{3}$$

where $\beta > 0$ is a parameter reflecting how fast greenhouse gases are emitted. From the initial condition $F(0) = 0$ it is clear that

$$F(t) = \overline{F}\left(1 - e^{-\beta t}\right). \tag{4}$$

Substituting (2) and (4) in (1),

$$\frac{d}{dt} T(t) = \frac{\lambda_0 \overline{F}}{c}\left(1 - e^{-\beta t}\right) - \frac{1-f}{c} T(t).$$

With $\alpha = \frac{1-f}{c}$ the solution, satisfying the boundary condition, is

$$T(t) = \frac{\lambda_0 \overline{F}}{1-f}\left(1 - e^{-\alpha t}\right) + \frac{\lambda_0 \overline{F}}{c\left(\beta - \alpha\right)}\left(e^{-\beta t} - e^{-\alpha t}\right), \tag{5}$$

which may be written

$$T\left(t\right) = \frac{\lambda_0}{1-f} \overline{F}\left(1 + \frac{1}{\beta - \alpha}\left(\alpha e^{-\beta t} - \beta e^{-\alpha t}\right)\right). \tag{6}$$

The asymptotic global temperature is thus in this model

$$T(\infty) = \frac{\lambda_0 \overline{F}}{1-f}.$$

Global warming will be strong if the feedback coefficient f is close to 1.

Weitzman (2008) admits that his model is extremely simple, and is not by far doing full justice to the knowledge about climate processes possessed by present day climate researchers. He holds however that his model is useful in that it captures the crucial feed-back mechanism from temperature to radiation. He further holds that simple models are useful to "see the forest from the trees", provided that they bring out structural aspects of reality of importance.

The feed-back coefficient f is hard to estimate. Weitzman (2008) argues that current knowledge is incapable of bounding f away from 1 with certainty. He treats the uncertainty concerning f in terms of a probability density ϕ. His results are based on $\phi(f) = 0$ for $f \geq 1$ and $\phi'(1) = -b < 0$. Under this condition, $E\left[T(\infty)^2\right] = \int_0^1 \left[\frac{\lambda_0 \overline{F}}{1-f}\right]^2 \phi(f)df = \infty$, and $E\left[T(\infty)\right]$ is finite. With a damage function at least as convex as the quadratic, the expected long term damage to humankind from continued greenhouse gas emissions is infinite under Weitzman's condition.

Weitzman (2008) argues that reasonable discounting rates can be represented by a probability density ρ on the positive axis such that $\rho'(0) > 0$. He uses this last property to conclude that

$$\int_0^\infty E\left[\int_0^\infty e^{-rt} D(T(t))dt\right] \rho(r)dr = \infty,$$

where expectation is with respect to f or alternatively α. This result is a bit tricky to obtain.

By squaring (6) and integrating,

$$\int_0^\infty e^{-rt}(T(t))^2 dt = \left(\frac{\lambda_0 \overline{F}}{\alpha}\right)^2 \left[\frac{1}{r} + \frac{2\left(\frac{\alpha}{\beta+r} - \frac{\beta}{\alpha+r}\right)}{\beta-\alpha} + \frac{\left(\frac{\alpha^2}{2\beta+r} - \frac{2\beta\alpha}{\alpha+\beta+r} + \frac{\beta^2}{2\alpha+r}\right)}{(\beta-\alpha)^2}\right].$$

Weitzman (2008) shows that the expectation with respect to r and α of this is infinite when the joint density is proportional to αrat the origin. (I have struggled with this but am at a loss.)

This result is indeed troublesome. Weitzman (2008) argues that this result gives strong reasons for preparing for a worst case by researching and establishing an international frame for curtailing temperature rise by fast geo-engineering. Some quotes: "[...] the fat upper tail of the PDF of climate sensitivity lends greater urgency to curtailing GHG emissions [...] a large increase in expected welfare might be gained if some relatively benign form of fast geo-engineering [injecting sunlight-reflective particulates or aerosols [...] into the stratosphere] were deployed in readiness to rapidly derail severe greenhouse heating – should this contingency materialize. [...] as well as there being a strong policy argument that now is the time to learn a lot more about fast geo-engineering – there is an additional strong policy argument that now is also the time to start thinking seriously about an international framework governing the use of this option."

Is Weitzman right? Note that infinite expected dis-utility results under the conditions

- the dynamic model is OK, but of course a gross simplification
- the dis-utility is at least as convex as the quadratic
- greenhouse gas emissions leads to a long-term increase in greenhouse gas in the atmosphere, $\overline{F} > 0$.

The last assumption is perhaps the most interesting. If the two first hold water, it means that we have to move towards zero green house gas emissions in order to force the limiting level back to its "natural" pre-industrial level, $\overline{F} = 0$, if physically possible. A model based on this assumption, rather than that of $F \to \overline{F} > 0$ should have equation (3) replaced by an equation making $F(t)$ have a uni-modal trajectory.

The first assumption may be questioned. It is possible that the linearity in the feed-back (2) is what causes the expected tragedy. With non-linear feed-back, say with f increasing in T, the global temperature might stabilize at a lower level making expected discounted damage finite. See Problem 2.

Another consequence of the analysis, assuming the first two conditions, is that research should be intensified to improve our knowledge about the feed-back factor f. Along with the American Statistical Association, I believe that climate research would improve if statisticians were more engaged in this vital and fascinating area of research much more widely than what is presently the case, see The American Statistical Society. The characterization and quantification of uncertainty is not an easy matter, but sometimes a matter of crucial importance.

3 Some continuous distributions

The gamma distribution

The gamma distribution has its name from the gamma function,

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx.$$

The symbol Γ is the Greek counterpart to the Latin G.

The *gamma density function* is defined for $x > 0$ by

$$g(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}. \tag{7}$$

The parameters α and β are called the *shape parameter* and *scale parameter*, respectively. The *mean* of this distribution is $\alpha\beta$ and the *variance* is $\alpha\beta^2$. Some authors, among them Taylor and Karlin (1998, page 38) (TK), use the *rate parameter* $\lambda = 1/\beta$ instead of the scale parameter. The gamma function is briefly discussed in TK (I.6 (Section 6, Chapter I)).

The gamma distribution is skewed for small values of the shape parameter. It is more and more symmetric as the shape parameter increases. Figure 1 shows five gamma densities. In the limit, as $\alpha \to \infty$ and $\beta = \sigma/\sqrt{\alpha} \to 0$ with the variance σ^2 fixed, the gamma distribution tends to a normal distribution. Its mean is then $\mu = \sigma\sqrt{\alpha}$.

The *Gamma cumulative distribution function* is defined for $x > 0$ by the integral

$$G(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x u^{\alpha-1} e^{-u/\beta} du.$$

The gamma distribution has been found useful in many applications. This is due to its ability to model skewed distributions, its nice mathematical form, its theoretical properties and its relationship to other distributions. One property is that if X_i is gamma distributed with parameters α_i and β (the same scale parameter!) and independent, $X = \sum_{i=1}^n X_i$ is gamma distributed with shape parameter $\alpha = \sum_{i=1}^n \alpha_i$ and scale parameter β. Thus, by the central limit theorem, the gamma distribution approaches the normal distribution as the shape parameter increases.

When $\alpha = 1$, the gamma distribution is called the *exponential distribution*. When $\beta = 2$ and when $\alpha = \frac{1}{2}\nu$ where ν is an integer, the gamma distribution is called the *chi-square distribution*

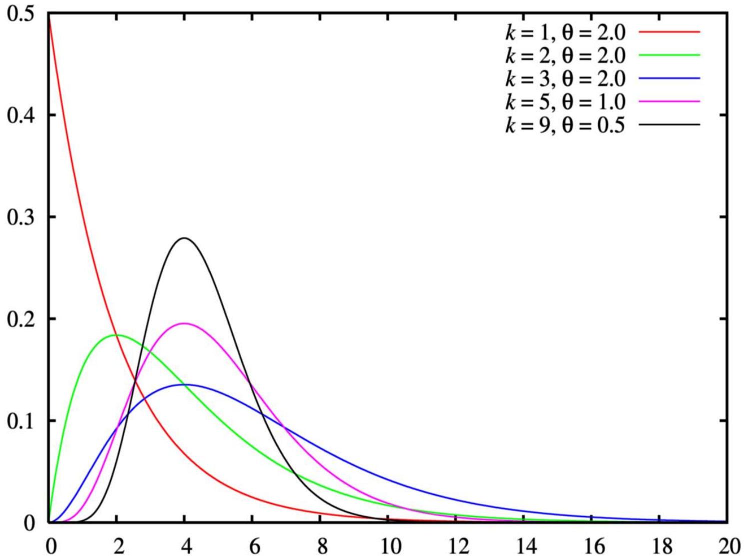

Figure 1: Gamma densities with shape $k = \alpha$ and scale parameter $\theta = \beta$. Wikipedia!

with ν degrees of freedom. This distribution does consequently have mean ν and variance 2ν. When Z has a standard normal distribution, Z^2 has a chi-square distribution with $\nu = 1$ degrees of freedom. This relates the gamma distribution to the normal. The gamma function is readily computed in the system R and in other mathematical and statistical software, and it is easy to simulate gamma distributed data.

Sums of independent gamma distributed variables with the same rate (or scale) parameters are gamma distributed with the same rate parameter, and with shape parameter equal to the sum of the component shape parameters.

The *gamma function* $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx$ is defined for positive real numbers α. By partial integration,

$$\Gamma(\alpha) = (\alpha - 1)\,\Gamma(\alpha - 1) \quad \alpha > 1. \tag{8}$$

Since $\Gamma(1) = 1$, $\Gamma(n) = n!$ for n a natural number. The gamma function interpolates the factorial function! In Problem 4 6 you find $\Gamma(1/2) = \sqrt{\pi}$. Figure 2 sketches the gamma function over(0.2, 4).

The gamma distribution is conjugate to the Poisson distribution in the following sense. When X given λ is Poisson distributed with mean λ, and λ is gamma distributed with shape parameter α and scale parameter β, the the conditional distribution of λ given X is gamma distributed with shape parameter $\alpha + X$ and scale parameter $\beta/(\beta + 1)$. The joint density (which is running over the natural numbers for X and the positive real numbers for λ) is actually proportional to

$$\lambda^{x+\alpha-1} e^{-\lambda(1+1/\beta)},$$

which again is proportional to the mentioned gamma distribution.

The beta distribution

The *Beta distribution* is defined for $0 \le x \le 1$ by the cumulative distribution function

$$F(x; a, b) = \frac{1}{B(a, b)} \int_0^x u^{a-1} (1 - u)^{b-1} du,$$

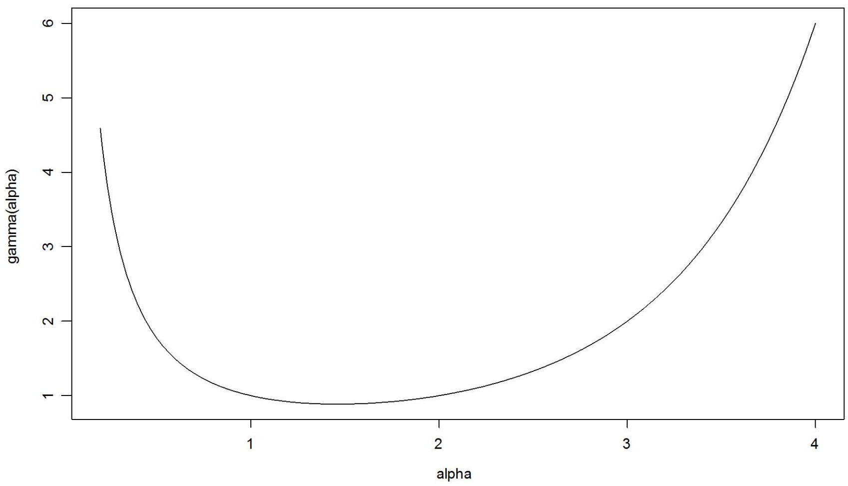

Figure 2: The gamma function over the interval $(0.2, 4)$.

where $B(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\,du$ is the Beta function with parameters a and b. Both parameters are positive. Otherwise, the integral had been infinite. The Beta function satisfies

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

where Γ is the gamma function. Since the gamma function interpolates the factorial function, and satisfies $\Gamma(a+1) = a\Gamma(a)$, the beta function is related to the binomial coefficient for integer-valued parameters:

$$\binom{n+m}{n} = (n+m+1)^{-1}B(n+1,m+1)^{-1}.$$

The density function for the beta distribution is

$$f(u;a,b) = \frac{u^{a-1}(1-u)^{b-1}}{B(a,b)}.$$

The parameters a and b are positive real numbers called *shape parameters*. The density is zero outside the unit interval. When $a = b = 1$, the beta distribution is the uniform distribution. Figure 3 shows three beta densities.

The mean of the beta distribution is $\dfrac{a}{a+b}$.

The beta distribution is obtained from the gamma distribution as "the broken gamma stick": When X and Y are independent and gamma distributed with the same scale parameter, and with shape parameters a and b respectively, the quotient $B = X/(X+Y)$ has a beta distribution with shape parameters a and b. The common scale parameter has no effect on the quotient, and the result holds whatever the scale parameter is.

The beta distribution is conjugated to the binomial distribution in the following sense. Let the conditional distribution of X given $P = p$ be binomially distributed with parameters n and p, and let P have a beta distribution with parameters a and b. The conditional distribution for P given $X = x$ is then a beta distribution with parameters $a + x$ and $b + n - x$.

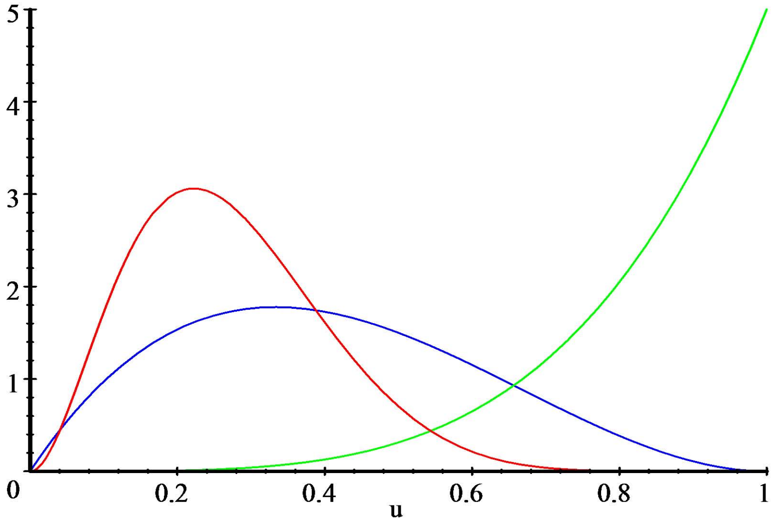

Figure 3: Three beta densities for $(a, b) = (2, 3)\,,\ (5, 1)\,,\ (3, 8)\,.$

The Dirichlet distribution

The beta distribution is a special case of the Dirichlet distribution, which is a multivariate distribution of Y on the subspace of $\mathrm{R}^n : Y > 0$ and $Y_1 + \cdots + Y_n = 1$ with density

$$f(y) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_n)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_n)} \prod_{i=1}^{n} y_i^{\alpha_i - 1}.$$

Independent gamma variables of equal rate divided by their sum is Dirichlet distributed.

The Dirichlet is a natural distribution for a stochastic probability vector (non-negative and summing to one). It is therefore in Bayesian analyses often used as a prior distribution for the probabilities in a multinomial distribution. It is actually conjugated to the multinomial distribution. It is obtained as "the multiply broken gamma stick". Let X_j be independent gamma distributed with shape parameter a_j and scale parameter σ independent of j. Then

$$B_j = \frac{X_j}{\sum_{i=1}^{\nu} X_i} \quad j = 1, \cdots, \nu$$

has a multivariate beta distribution with parameter vector $(a_1, \cdots, a_\nu)$. From the construction, marginal distributions have (multivariate) beta distributions. If $\nu > 3$, (B_1, B_2, B_3) is, for example, beta distributed with parameter vector (a_1, a_2, a_3).

The Dirichlet distribution is conjugate to the multinomial distribution.

4 Tinbergen's model for the formation of income distributions

Tinbergen (1956) sketch a theory of income distributions based on normal distributions of skills in the labor supply and in the labor demand. These distributions will typically not match, and in Tinbergen's theory, income is used to ease the tension between demand and supply of skills.

A question closely related to Tinbergen's is to what extent observed idiosyncratic income risk can account for observed wealth, savings and consumption heterogeneity. The dynamic economic model, proposed by Bewley (1986) and developed further in Aiyagari (1994), has become a leading model to answer these and related questions.

The Aiyagari-Bewley economy is populated with an infinite heterogeneous population of agents who are subject to uninsurable idiosyncratic income risks, and a constant mortality. The income process follow an estimated Markov process for individual income realizations. Since agents' histories of income shocks are different, the model generates endogenous heterogeneity and equilibrium cross-section distributions of wealth, saving and consumption.

In the Aiyagari-Bewley model economies all risk is idiosyncratic, i.e. there is no aggregate risk. Krusell and Smith (1998) extend the framework of Aiyagari-Bewley to ask how aggregate uncertainty interacts with idiosyncratic risk to generate endogenous wealth, saving and consumption heterogeneity.

Storesletten, Telmer and Yaron (2004) asks if individual-specific earnings risk can account for the observed rise inequality by age. In order to answer this question the paper constructs an overlapping generations general equilibrium model in which households face uninsurable earnings shocks over the course of their lifetimes. Earnings inequality is exogenous and is calibrated to match data from the U.S. Panel Study on Income Dynamics. Consumption inequality is endogenous and matches well data from the U.S. Consumer Expenditure Survey.

Returning to the Tinbergen problem, consider first the case of inelastic demand for skills, and skill being one-dimensional. Denote by S the skill on the supply side and D the skill on the demand side. The skill demanded in the labor market has distribution represented by D, which is $N(\mu_D, \sigma_D^2)$. That demand is inelastic means that this distribution is fixed. The supply of skill has distribution represented by $S \sim N(\mu_S, \sigma_S^2)$, and this distribution adjusts through the incentives of the income distribution to that of D. On a scale of additive utilities, income, I, is assumed quadratic in skill,

$$u\left(I(s)\right) = v + \lambda_1 s + \frac{1}{2}\lambda_2 s^2.$$

A mismatch between the skill demanded in a job and the skill supplied, has a negative utility, quadratic in the mismatch. A person with skill s in a job that demands d will thus have utility

$$u(s) = v + \lambda_1 s + \frac{1}{2}\lambda_2 s^2 - \frac{1}{2}w(s-d)^2.$$

Here, v and w are taken for given, while λ_1 and λ_2 will be determined in the economic coordination process.

By translating $s \to s - \mu_S$, the quadratic utility is kept quadratic, but with different coefficients in the transformed skill supplied. There is thus no loss of generality of assuming $\mu_S = 0$, which is done.

The idea is that the income in monetary terms I is a function of s through the quadratic form determined by the λ-coefficients. Thus, with a logarithmic utility,

$$I(s) = \exp\left(v + \lambda_1 s + \frac{1}{2}\lambda_2 s^2\right).$$

If then $\lambda_2 = 0$, the income in monetary terms is log-normally distributed since S is normally distributed.

In equilibrium, this utility is maximized across the distribution. At equilibrium, $u'(s) = 0$:

$$\lambda_1 + \lambda_2 s - w(s - d) = 0.$$

The economic problem is to coordinate skills between supply and demand, and the equation says that a person with skill s finds a job with demand $d = s - (\lambda_1 + \lambda_2 s)/w$. Thus, in distributional terms,

$$\lambda_1 + \lambda_2 S = w(S - D). \tag{9}$$

This equation determines the parameters λ_1 and λ_2 that specifies the income distribution. For these coefficients given, S is thus a linear transform of D, which indeed is possible since both distributions are normal. Taking moments on both side of (9),

$$\begin{aligned} \lambda_1 &= -w\mu_D, \\ (\lambda_2 - w)^2 \sigma_S^2 &= w^2\sigma_D^2. \end{aligned}$$

There are two solution to the second equation,

$$\lambda_2 = w\left(1 \pm \frac{\sigma_D}{\sigma_S}\right).$$

Tinbergen considers the case with skill being two-dimensional. He solves the problem by somewhat cumbersome direct reasoning. In passing, he notes that his equilibrium equations have two solutions, but he gives no argument for his choice. The solution he quotes agrees with the solution $\lambda_2 = w\left(1 - \frac{\sigma_D}{\sigma_S}\right)$. We will keep to this solution. When $\sigma_D = \sigma_S$, $\lambda_2 = 0$. One argument for this solution is that the other solution makes $\lambda_2 > w$, and hence the utility becomes convex in s rather than concave. This discarded solution will thus represent a situation where suppliers minimize their utility when finding a job.

With the chosen solution, supply and demand of skill is coordinated according to the linear equation

$$D = \mu_D + \frac{\sigma_D}{\sigma_S} S.$$

This equation does express more than a relationship between distributions. It should, perhaps, have been phrased in lower case d and s, to highlight that it represents a mapping or coordination between workers with skill s and jobs demanding skill d.

When $\sigma_D = \sigma_S$ and the utility of money is logarithmic and the income distribution is log-normal, $\ln(I) \sim N(v, \lambda_1^2\sigma_S^2)$. The larger the discrepancy in mean skill between supply and demand at the outset, the larger is λ_1 in absolute terms, and the more skewed is the income distribution. If $\mu_D = \mu_S = 0$, the income distribution collapses, and all have the same income $I = \exp(v)$. As an example, consider Figure 4. As another example, consider the skill distributions in Figure 5 which go together with the income distributions in Figure 6 when $w = 1$ and $v = 0$. Note the extreme skewness in income distribution when skills are more dispersed in supply than in demand, $\lambda_2 > 0$, while in the opposite situation the income distribution is virtually restricted to a finite interval, and it is less dispersed and more symmetric (unfortunately, the line types do not match in the two figures).

A special case arise when the distributions of demand and supply are equally centered, making $\lambda_1 = 0$, but with different standard deviations. If $\lambda_2 > 0$, the distribution is very skewed with a long tail to the right. If, however, $\lambda_2 < 0$, the distribution is concentrated on a finite interval, but is still skewed, see Figure 6.

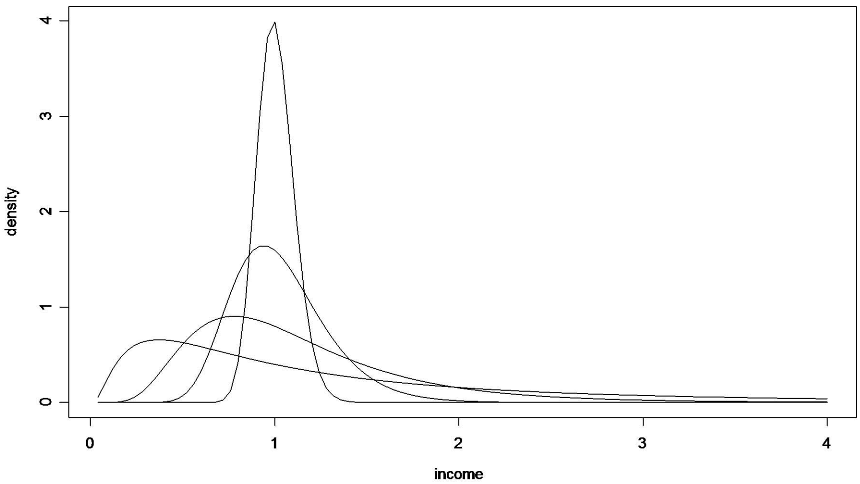

Figure 4: Income densities for $w = \sigma_D = \sigma_S = 1$, $v = \mu_S = 0$ and $\mu_D = 0.1$, 0.25, 0.50 and 1.

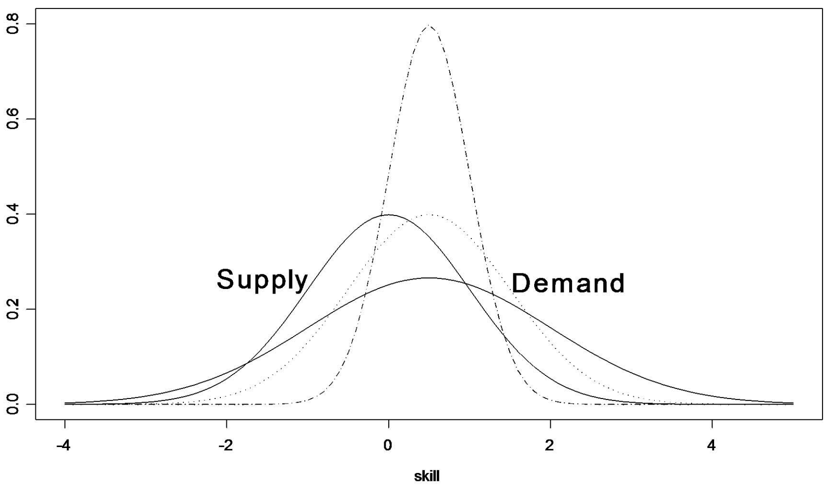

Figure 5: Skill densities. Supply: $\mu_S = 0$, $\sigma_S = 1$. Demand: $\mu_D = 0.5$; $\sigma_D = 1.5$ (solid line), $\sigma_D = 1$ (broken line), $\sigma_D = 0.5$ (dotted line).

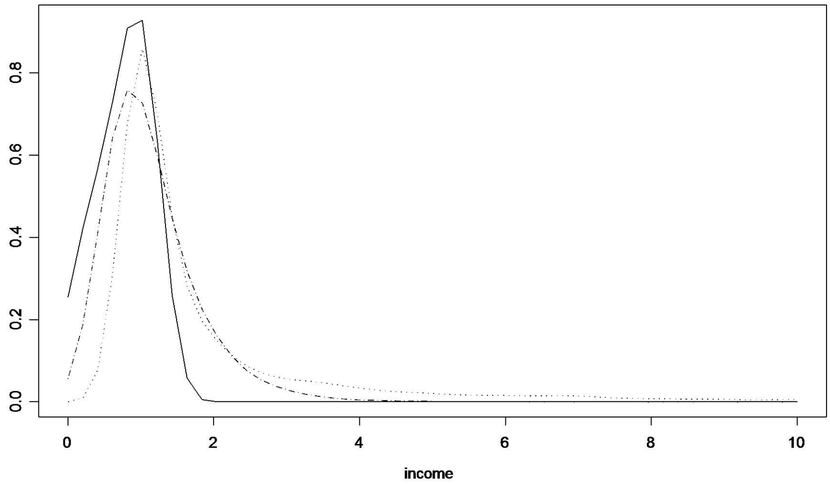

Figure 6: Income densities for $\lambda_1 = -.5$ and $\lambda_2 = -.5$ (solid line), $\lambda_2 = 0$ (broken line) and $\lambda_2 = 0.5$ (dotted line).

5 Transition intensity and competing risk, income assimilation of immigrants

In Norway and in many other countries, immigrants start out with less income than comparable groups of residents. Let us use the term natives for those born in the country and immigrants for those who have immigrated at some point in time. For simplicity, immigrants and natives are equal with respect to age, education, sex and other stratifying variables. Immigrants are assumed to have arrived at the same point in time, $t = 0$. The immigrants tend to have less human capital than the natives (language etc.) as measured by the native standards. Their income will therefore tend to be less than that of the natives.

As time since immigration increases, the income distribution of the immigrants gets more and more similar to that of the natives. This is termed income assimilation. It is of interest to measure the rate of income assimilation. This is not an easy matter since many immigrants emigrate, possibly back to their mother country. One might expect the rate of emigration to be high when the income of the immigrant is unsatisfactory. A competing theory is that immigrants are more prone to move back home when they have acquired a sufficient wealth. With income-specific emigration, the income distribution of the immigrants that stay in the country will have an income distribution affected both by the process of assimilation and by the process of selective emigration. The purpose of this example is to find the structure of the income distribution in a very simple speculative model for selective emigration and assimilation. Borjas and Bratsberg (1996) discuss income assimilation and estimate effects for a number of countries.

Let Y be a stochastic variable representing the income distribution of the natives, and let X_t correspondingly represent the income distribution of the immigrants after t units of time in the country. With no emigration, the assimilation model is

$$X_t = \alpha(t)Y.$$

In terms of the p-quantiles, $y(p)$ and $x_t(p)$ of the two distributions, the model is $x_t(p) = \alpha(t)y(p)$. The assumption is further that an immigrant that starts out with an income $x_0 =$

$y\alpha(0)$ will follow the assimilation path $x_t = y\alpha(t)$ throughout his stay in the country. The assimilation profile $\alpha(t)$ is increasing from $\alpha(0) > 0$ to $\alpha(\infty) = 1$. If immigrants represent a positive selection with respect to income potential, one might have $\alpha(\infty) > 1$.

The emigration intensity (for immigrants), λ_t, depends on current income. Let the cumulative emigration intensity be

$$\Lambda(r; y) = \int_0^r \lambda_t dt.$$

The expected fraction of immigrants at this income path that still are living in the country at time t, and that choose to emigrate before time $t + h$ (h infinitesimally small) is $h\lambda_t$.

In addition to emigration, the immigrant faces the "risk" of dying. Let the mortality intensity be μ_t, independent of income. Emigration and death are competing risks for an immigrant. The two intensities add and the intensity for an immigrant to be removed from the immigrant population is $\mu_t + \lambda_t$.

The cumulative intensity of removal until time r is thus

$$\int_0^r \mu_t dt + \Lambda(r; y) = M(r) + \Lambda(r; y),$$

and the conditional probability of not having emigrated or died before a fixed time r is thus

$$P(r; y) = \Pr(R > r | X_0 = y\alpha(0)) = e^{-M(r)-\Lambda(r;y)} \propto e^{-\Lambda(r;y)}, \tag{10}$$

where R is the time of residence (until emigration or death) of the immigrant and the symbol $\propto$ means proportional to. The proportionality factor $\exp(-M(r))$ depends on the mortality, which is assumed independent of income.

The density at income $X_r = x$ among immigrants still living in the country is proportional to $P(r; y)g(y)$, where g is the income density of natives and $y = x/\alpha(r)$ corresponding to $X_r = x$. Substituting for y in (10) the resulting income density for residing immigrants is

$$g(x; r) \propto g\left(x/\alpha(r)\right) \exp\left(-\Lambda(r; x/\alpha(r))\right).$$

Assume further that

$$\Lambda(r; y) = \max\left\{0, \theta(r)(\tau - y)\right\}$$

for given function θ and constant τ. Under this condition, the intensity of back-emigration decreases with income. This intensity (for immigrants) is then

$$\lambda(r; y) = \max(\theta'(r)(\tau - y), 0), \quad \theta'(r) > 0.$$

Take as a numerical example the case of the income for natives being gamma distributed, say with shape parameter $a = 2$ and scale parameter $b = 5$. The assimilation curve is the logistic $\alpha(t) = e^{\alpha t}/\left(1 + e^{\alpha t}\right)$ starting at $\alpha(0) = 1/2$ and having $\alpha = \ln(3)/5$, corresponding to $\alpha(5) = 3/4$. The income distribution of the residing immigrants at time r after immigration is

$$g(x; r) \propto \gamma\left(x/\alpha(r)\right) \min\left\{1, \exp\left(-\theta(r)(\tau - y)\right)\right\}.$$

Here,

$$\gamma\left(x; a, b\right) = \frac{1}{b^a \Gamma\left(a\right)} x^{a-1} e^{-x/b},$$

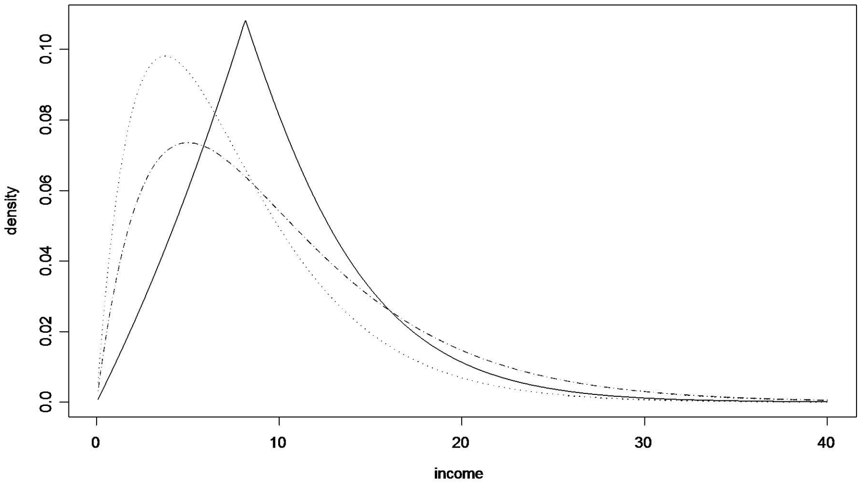

Figure 7: Income densities: after $r = 5$ years; solid line when $\theta(r) = \sqrt{r/100}$, $\tau = 10.86$ and dotted line without emigration; broken line for natives. Gamma model with $a = 2$ and $b = 5$. Immigrants starts at $\alpha(0) = \frac{1}{2}$ and assimilate logistically with $\alpha(5) = \frac{3}{4}$. Mean income is 10 for both natives and residing immigrants.

is the gamma density. When $\theta(t)$ is increasing from 0 to $\theta(r) < 1/b$, the income distribution of the residing immigrants is piecewise gamma distributed:

$$g(x;r) \propto \begin{array}{ll} \gamma\left(x;a,b\alpha(r)\right), & x > \tau\alpha(r) \\ \exp(-\tau\theta(r))\left(1-b\theta(r)\right)^a \gamma\left(x;a,\frac{b\alpha(r)}{1-b\theta(r)}\right), & x \leq \tau\alpha(r). \end{array}$$

This is found by simplifying the above.

For chosen parameter values as in Figures 7 and 8 show income densities for staying immigrants. With these parameter values the immigrants that stay in the country will approach a mean income of 16.2, while the natives will have a mean of 10. Note that this applies to the cohort of immigrants. With a steady rate of immigration, the picture is different.

6 Income distributions and election outcomes

Consider two parties C and L that are up for election to power at times $t = 1, 2, \cdots$. We are interested in the series of election outcomes, and in the distribution of income. Our model is as follows: a voter with attribute x has an economic gain in the period between elections depending on which party is in office. The distribution of attributes over the electorate changes from election to election, partly in response to which party is in office. We shall assume that this response is Markovian in the sense that the distribution at next election only depends on current regime and current distribution. The outcome of an election is determined by the median voter with respect to its gain.

Kundu (2007) studied a related problem in a two-period model, while we consider an infinite sequence of elections. He further assumed income and gain to have finite range, while we allow them to be distributed over infinite ranges.

The attribute of the electorate is distributed over the real numbers with respect to economic gain. A randomly chosen voter has attribute X such that its gain is X when C is in power and

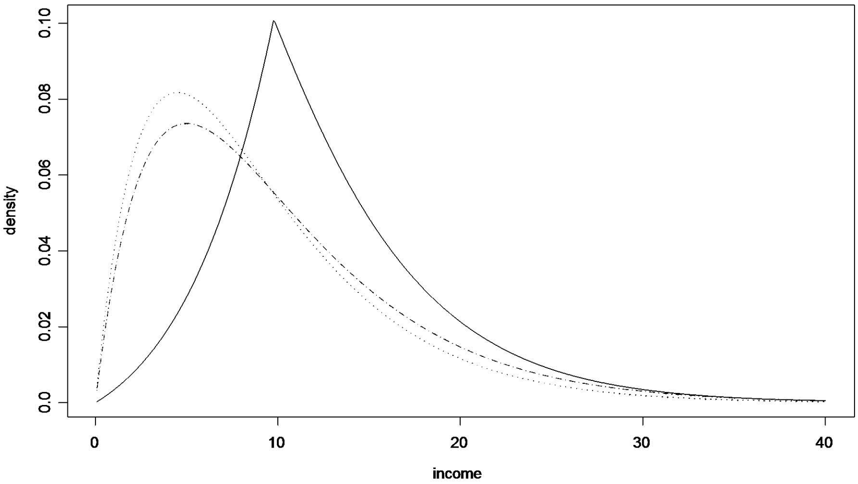

Figure 8: Same as Figure 7, now with $r = 10$ and mean income for residing immigrants is 12.72.

$-X$ when L is in power. X is normally distributed with mean θ_t and variance ω^2 in period t. The probability of the majority voting for C in election t is then $P(X > -X) = \Phi(\theta_t/\omega)$. Thus, C will be in power when $\theta_t > 0$, L wins when $\theta_t < 0$, and the probabilities for winning are $1/2$ when $\theta_t = 0$.

Assuming now that θ_t follows one random walk over the period for which L is in power, and another random walk when C is in power. Let the state space for the combined random walk be $\cdots, -\frac{3}{a}, -\frac{2}{a}, -\frac{1}{a}, 0, \frac{1}{a}, \frac{2}{a}, \cdots$. Assuming that the two separate random walks are reflections of each other, the (central part of the) transition matrix for θ_t is

$$P = \begin{bmatrix} p & q & r & & & & & & \\ & p & q & r & 0 & 0 & 0 & 0 & \\ & 0 & p & q & r & 0 & 0 & 0 & \\ & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & \\ & 0 & 0 & 0 & r & q & p & 0 & \\ & & & & & r & q & p & \\ & & & & & & r & q & p \end{bmatrix}$$

where $q + p + r = 1$.

The political economy of our model is "mean reverting" when $p < r$. If C is in power, there is a shift in the gain distribution to the left, while the gain distribution shifts to the right when L is in power. If $p = r$ the mean gain is a symmetric random walk with probability q of staying put, except for in state 0, and the number of successive periods that one of the parties is in office has infinite mean. If $p > r$ the expected gain radiates out, and the system ends up in one-party rule. In technical terms, the Markov chain $\{\theta_t\}$ is positively recurrent when $r > p$ and null-recurrent when $r = p$.

Assume positive recurrence. Let $\{\pi_i\}$ be the stationary distribution of $\{\theta_t\}$. By direct reasoning

$$\begin{aligned} \pi_i &= \frac{1}{2r}\left(\frac{p}{r}\right)^{|i|-1}\pi_0 \\ \pi_0 &= \frac{r-p}{2r+q} \end{aligned} \tag{11}$$

Assume that the income distribution is lognormal, and is of the form $\exp(Y \pm X) = \exp(Z)$ depending on which party is in power. Note that Y and X might be correlated without harming the model: $Y + X > Y - X \Leftrightarrow X > 0$. Let $Y \sim N(\mu, \sigma^2)$, and $\varrho = cor(X, Y)$. The log income has thus mean

$$\begin{aligned} EZ_t &= EE\left[Z_t|\theta_t\right] = \mu + \sum_{i=-\infty}^{1} -\frac{i}{a}\pi_i + \sum_{i=1}^{\infty}\frac{i}{a}\pi_i \\ &= \mu + 2\sum_{i=1}^{\infty}\frac{i}{a}\pi_i = \mu + 2\frac{\pi_0}{2r}\sum_{i=1}^{\infty}\frac{i}{a}\left(\frac{p}{r}\right)^{i-1} \\ &= \mu + \frac{r}{a\left(r-p\right)\left(2r+q\right)}. \end{aligned}$$

Income, conditional on party when $\theta = 0$ and otherwise on θ, is lognormally distributed. Its marginal distribution is thus a mixture of lognormals. Let us study the income distribution conditional on party C being in power. Under this condition, the equilibrium distribution of θ is

$$\begin{aligned} \pi_i^C &= P\left(\theta = \frac{i}{a}\right) = \frac{1}{r}\left(\frac{p}{r}\right)^{i-1}\pi_0 \quad i > 0 \\ \pi_0^C &= P\left(\theta = 0\right) = \pi_0. \end{aligned}$$

Given C and θ, the log income is normal with mean $\mu + \theta$ and variance $\tau^2 = \sigma^2 + 2\rho\sigma\omega + \omega^2$.

Conditioned on C the probability density for the income $In = \exp(X + Y)$ is a mixture of lognormal densities,

$$f(in|C) = \sum_{j=0}^{\infty}\pi_j^C\varphi\left(\frac{\ln(in) - \left(\mu + \frac{j}{a}\right)}{\tau}\right)\frac{1}{\tau}\frac{1}{in}.$$

Here, φ is the standard normal density. The density for income under L-rule is found the same way, but with τ adjusted according to the reverse correlation.

The income densities are for a given setting of the parameters given in Figure 9. For these parameters income has clearly a more skewed distributed when C is in power than under L. Due to the longer tail in the income distribution under C-rule, mean income greater than when L is in power. Mean income is actually 4330 under C-rule compared to 1760 under L-rule. Median income happens to be the same under the two regimes. From Figure 9 it is seen that under C rule the income distribution is quite a bit more skewed under C than under L in our numerical example. This is more dramatically born out in the PP- and the QQ plot in Figure 10. The inverted S-shaped PP plot shows that cumulative probability at income is higher under C to the left of the median, and lower than for L to the right. Correspondingly, income quantiles at the upper end are progressively higher under C than under L as the probability

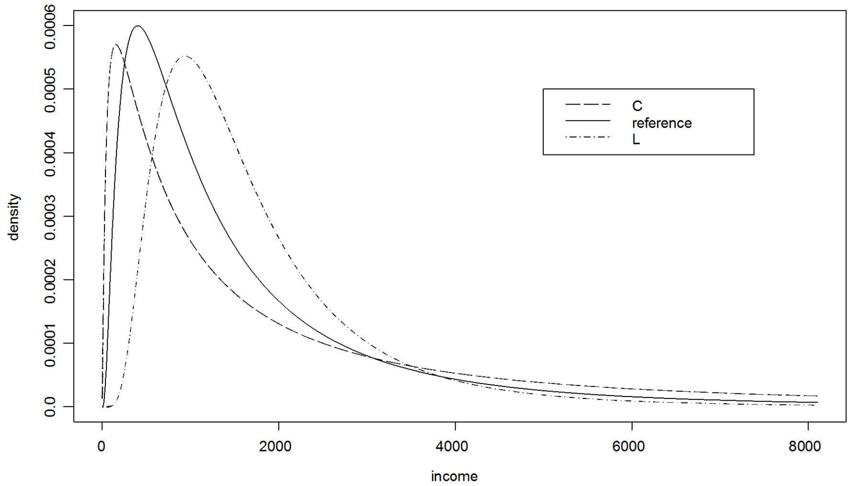

Figure 9: Income distributions (densities) under C- and L rule. The reference distribution is that of $exp(Y)$. Parameter values are $(\mu, \sigma, \omega, \rho, r, p, q, a) = (7, 1, .5, .9, .3, .5, .2, 10)$.

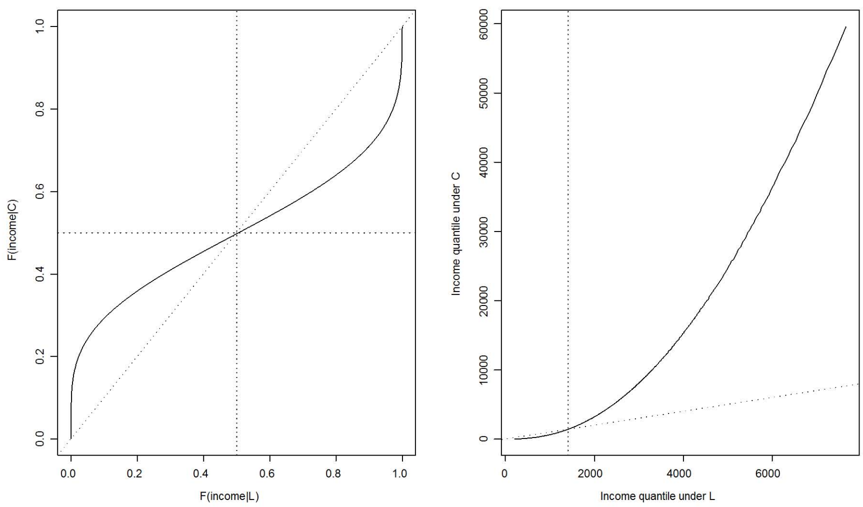

Figure 10: PP-plot (left panel) and QQ-plot (right panel) of income distributions under C- and L rule respectively. The PP-plot is the curve $(F\,(in|C)\,, F\,(in|L))$ over the income range, while the quantile-quantile plot to the right is the curve $\left(F^{-1}\,(p|C)\,, F^{-1}\,(p|L)\right)$ over p in the unit interval. Diagonals are added, and also dotted lines showing median or half-value.

moves above the (common) median. Most of the action in the QQ plot takes place at high incomes, and the effect of increased skewness under C is dramatic.

The conditional mean given state θ and party C is

$$E\left[In|\theta, C\right] = \exp\left(\mu + \theta + \frac{1}{2}\tau^2\right),$$

while

$$var\left[In|\theta, C\right] = \exp\left(2\mu + 2\theta + \tau^2\right)\left(\exp\left(\tau^2\right) - 1\right)$$

Mean income given C, averaged over state, is thus

$$\begin{aligned} E\left[In|C\right] &= E\left[E\left[In|\theta, C\right]|C\right] = \sum_{i=0}^{\infty} \pi_i^C \exp\left(\mu + \frac{i}{a} + \frac{1}{2}\tau^2\right) \\ &= \pi_0 \left[1 + \frac{p\exp(1/a)}{r^2 - rp\exp(1/a)}\right] \exp\left(\mu + \frac{1}{2}\tau^2\right) \end{aligned}$$

provided $p\exp(1/a) < r$. If $p\exp(1/a) \geq r$ mean income given C is infinite.

We can also calculate the conditional variance of income given C.

$$var\left[In|C\right] = E\left[var\left[I|\theta, C\right]|C\right] + var\left[E\left[I|\theta, C\right]|C\right].$$

The first term is simple,

$$\begin{aligned} E\left[var\left[In|\theta, C\right]|C\right] &= \sum_{i=0}^{\infty} \pi_i^C \exp\left(2\mu + 2\frac{i}{a} + \tau^2\right)\left(\exp\left(\tau^2\right) - 1\right) \\ &= \pi_0 \left[1 + \frac{p\exp(2/a)}{r^2 - rp\exp(2/a)}\right] \exp\left(2\mu + \tau^2\right)\left(\exp\left(\tau^2\right) - 1\right). \end{aligned}$$

It is therefore necessary for $var\left[In|C\right] < \infty$ to have $p\exp(2/a) < r$. The second term can also be calculated

$$\begin{aligned} E\left[E\left[In|\theta, C\right]^2|C\right] &= \sum_{i=0}^{\infty} \pi_i^C \exp\left(2\mu + 2\frac{i}{a} + \tau^2\right) \\ &= \pi_0 \left[1 + \frac{p\exp(2/a)}{r^2 - rp\exp(2/a)}\right] \exp\left(2\mu + \tau^2\right), \end{aligned}$$

and is also finite provided $p\exp(2/a) < r$. The remaining term to calculate is

$$\begin{aligned} \left\{E\left[E\left[In|\theta, C\right]|C\right]\right\}^2 &= \left\{E\left[I|C\right]\right\}^2 \\ &= \left\{\pi_0 \left[1 + \frac{p\exp(1/a)}{r^2 - rp\exp(1/a)}\right] \exp\left(\mu + \frac{1}{2}\tau^2\right)\right\}^2. \end{aligned}$$

7 A branching point process, a possible model for market volatility

Engle and Russel (1998) discuss autoregressive conditional duration models for irregularly spaced transaction data in finance. We shall have a look at simple versions of their model.

The model is usually called the Hawkes model, and was originally proposed for earth quake data.

Consider the stock exchange. Focus on a single asset, and let Z_t be its price at time t. The price is constant between trading. Following TK (Example, page 75), the number of tradings in a given period is a discrete stochastic variable. Let the period start at 0 and end at time t, and let the number of tradings be $N(t)$. N is then a stochastic process. It has jumps of unit size at random points in time, and is called a *counting process*. With $Y(T) = Z_T - Z_{T-}$ being the change in price of the asset when there is a trading at time T, the price process can be represented by the stochastic integral

$$Z(t) = \int_0^t Y(u)dN(u) = \sum_{i=1}^{N(t)} Y(T_i),$$

where T_i is the time of the i-th trading.

Assuming price changes to be independent and identically distributed, volatility in the price process stem from clustering in the point process of trading times, $\{T_i\}$, which is represented by the counting process N. When $\{T_i\}$ is a Poisson process, there is no clustering. There are many models for clustered point processes. The Neyman-Scott process have Poisson distributed invisible mother points. Each mother point produce a cluster of a stochastic number of visible daughter points spread around the mother, often normally. Clusters are independent and identically distributed except for their invisible centers. We shall look at a branching point process that is related in construction to the Neyman-Scott process, but which is more interpretable in economic terms.

7.1 The branching point process for a market characterized by "psychology"

The idea is that each trading triggers an interest in the asset, and excites a random number (which might be zero) of succeeding trading. This is modelled through the intensity of the point process. Each trading is actually assumed to produce an intensity component that adds to the intensity of previous points. Assuming this additional intensity component to be a falling exponential, and assuming there also is a constant background intensity α (independent of previous points), the model for the intensity is

$$\lambda(t) = \alpha + \frac{\theta}{\sigma} \sum_{T_i < t} \exp\left(-\frac{t - T_i}{\sigma}\right). \tag{12}$$

Note that this intensity is a stochastic process. Its interpretation is captured in the conditional expectation $E\left(N(t+h) - N(t) | \{T_i \le t\}\right) = \lambda(t)h + o(h)$, where the remainder is negligible: $o(h)/h \to 0$ in probability. The infinitesimal $\lambda(t)h$ is thus the expected number of tradings in a short period of length h, given the history, and $\lambda(t)$ depends on the history as given in (12).

The background Poisson process with intensity α generates initial points in clusters. Each point in a cluster generates a Poisson number of offsprings that all are members of the cluster. The offsprings from point T_i has intensity $I(t > T_i)\frac{\theta}{\sigma}\exp\left(-\frac{t-T_i}{\sigma}\right)$, I being the indicator function. The integral of this intensity function is θ, which thus is the expected number of first-generation offsprings from the point. The cluster is generated as a branching process. Remember that a branching process dies out with probability 1, and has finite mean number, when the expected number of offsprings from one parent is less than 1 (TK III.9).

Let X be the number of points in a cluster and let ξ be the Poisson number of first generation offsprings from a point,

$$E(\xi) = \int_T^\infty \frac{\theta}{\sigma} \exp\left(-\frac{t-T_i}{\sigma}\right) dt = \theta$$

. Since each point in a cluster produces stochastically the same number of points in the cluster as the initial point, and each branch is independent, we have in obvious notation

$$X = 1 + \sum_{i=1}^{\xi} X_i.$$

Thus by conditioning, $E(X) = 1 + \theta E(X)$, and $var(X) = var(X)\theta + (E(X))^2\theta$, yielding

$$E(X) = (1-\theta)^{-1}, \; var(X) = \theta (1-\theta)^{-3}.$$

The generating function for X, $g(s) = E\left(s^X\right)$ satisfies the equation

$$g(s) = s \exp\left((g(s)-1)\theta\right).$$

The clusters are therefore finite with probability 1 if $\theta \le 1$, and they have finite mean number if $\theta < 1$.

With $\theta > 1$, N will explode in finite time. We assume the reverse. The exact moments of $N(t)$ are difficult to compute, but the process is easy to simulate – which is done below. We can, however, calculate approximate moments for large t. The number of clusters initiated before t is Poisson distributed with mean αt, and the asymptotic fraction of these that are completed before time t is 1. Thus by conditional expectation and variance, for large t,

$$\begin{aligned} E(N(t)) &\approx E(\widetilde{N}(t)) = \alpha t (1-\theta)^{-1}, \\ var(N(t)) &\approx var(\widetilde{N}(t)) = \alpha t \theta (1-\theta)^{-3} + \alpha t (1-\theta)^{-2} = \alpha t (1-\theta)^{-3}. \end{aligned}$$

This variance does however not reflect the volatility, which is a local phenomenon.

7.2 Updating equations for the intensity and the cumulative intensity

The model is computationally attractive. The intensity for the next point after T_n is

$$\lambda(t) = \alpha + \left(\lambda(T_n) - \alpha + \frac{\theta}{\sigma}\right) \exp\left(-\frac{t-T_n}{\sigma}\right), \; T_n < t \le T_{n+1}.$$

Letting λ_n being the intensity just after T_n, the updating formula for the intensity is

$$\lambda_{n+1} = \alpha + \frac{\theta}{\sigma} + (\lambda_n - \alpha) \exp\left(-\frac{T_{n+1}-T_n}{\sigma}\right), \; T_n < t.$$

The cumulative intensity for a new point is thus

$$\begin{aligned} \Lambda(T_n, t) &= \int_{T_n}^t \lambda(t) dt \\ &= \alpha (t - T_n) + \theta (\lambda_n - \alpha) \left(1 - \exp\left(-\frac{t-T_n}{\sigma}\right)\right) \end{aligned}$$

for $T_n < t \le T_{n+1}$. The updating equation for the cumulative intensity is

$$\Lambda(T_n, T_{n+1}) = \alpha (T_{n+1} - T_n) + \theta (\lambda_n - \alpha) \left(1 - \exp\left(-\frac{T_{n+1}-T_n}{\sigma}\right)\right)$$

7.3 Likelihood

When estimating the parameters, the likelihood function is of great help. Consider a period of length L. There are n observations in the period. The intensity for new points is λ_0 at the start of the period. The Likelihood for the points, $T_1, \cdots, T_n$ is constructed by sequential conditioning. If $n = 0$, the likelihood is

$$L = \exp\left(-\Lambda(0, L)\right).$$

If $n = 1$, the likelihood is

$$L = \lambda(T_1)\exp\left(-\Lambda(0, T_1) - \Lambda(T_1, L)\right) = (\lambda_1 - \frac{\theta}{\sigma})\exp\left\{-\Lambda(0, T_1) - \Lambda(T_1, L)\right\}$$

The general formula is

$$L = \prod_{i=1}^{n}\left(\lambda_i - \frac{\theta}{\sigma}\right)\exp\left\{-\sum_{i=1}^{n+1}\sigma\left(\lambda_i - \alpha\right)\left(1 - \exp\left(-\frac{T_i - T_{i-1}}{\sigma}\right)\right)\right\},$$

where $T_0 = 0$ at the start of the period, and $T_{n+1} = L$ at the end.

In an econometric setting, there will usually be covariates to be taken into account. One possibility is to have a log-linear regression structure in the background intensity. This is not pursued.

7.4 Simulation and volatility

The points of trading are spread out along the line by simulating waiting distances between them. The distance $D_i = T_{i+1} - T_i$ is simulated by solving

$$\Lambda_i\left(T_i, T_i + D_i\right) = E,$$

where E is a standard exponential variate.

This can better be done by generating the clusters one by one. We use the fact that the union of points generated from independent point processes has intensity the sum of intensities over the separate processes. Let us call this the *additivity property of point processes*. Our first component process is the back ground Poisson process with intensity α. The points generated by this process are called mother points. For each mother point a number, possibly zero, of daughter points are generated. Mother plus daughters constitute clusters. The daughter process of a mother point depends only on the position of the mother, but is otherwise independent of the mother process and other daughter processes. One might worry that the lack of complete independence violates the additivity of component intensities. This is however not a problem since our intensity (12) is dynamically constructed and is for each t based on the history of the process. Our daughter processes will be conditioned on their mothers and are otherwise independent on the history. It is thus a dynamic or conditional version of the additivity property that is used here.

The mother points are generated by drawing independent exponentially distributed waiting times between mother points. The subsequent daughter points of mother poit c come at times $T_{c,k}$. The index i in (12) is replaced by the double index c, k and the sum in the formula for the dynamic intensity is split according to clusters. A cluster points $T_{c,j} < t$ has intensity $\lambda_{c,k}$ of a new daughter point at time t. These cluster-specific intensities sum to

Figure 11: 100 simulated points. $\alpha = 0.01$, $\theta = 0.8$, $\sigma = 20$

$\lambda(t) - \alpha$. Let $\lambda_{c,k} = \lambda_{c,k}(T_{c,k}+)$ be the cluster-specific intensity right after $T_{c,k}$. Then $\lambda_{c,k}(t) = \lambda_{c,k} \exp -\frac{t-T_{c,k}}{\sigma}$ $T_{c,k} < t \leq T_{c,k+1}$, and $\lambda_{c,k+1} = \lambda_{c,k} \exp -\frac{T_{c,k+1}-T_{c,k}}{\sigma} + \frac{\theta}{\sigma}$. The cumulated intensity is $\Lambda_{c,k}(t) = \lambda_{c,k}(1 - \exp -\frac{t-T_{c,k}}{\sigma}/\sigma$, and the next daughter point is simulated by solving $\Lambda_{c,k}(t) = E_{c,k}$ where $E_{c,k}$ is a new draw from the standard exponential distribution. There is no solution if $\frac{\lambda_{c,k}}{\sigma} \leq E_{c,k}$, and there are no more points in the cluster. In the opposite case there is at least one more point, at $T_{c,k+1} = T_{c,k} - \sigma \ln 1 - \frac{\sigma}{\lambda_{c,k}} E_{c,k}$. Clusters consist of a mother point and its daughter points. some clusters overlap.

When the time points of trading $\{T_i\}$ have been simulated, it is no deal to simulate iid price changes $\{z_i\}$. The price process is then

$$Z_t = \sum_{i:T_i \leq t} z_i.$$

The local volatility may be defined as the variance in price change over short periods. Conditioning on the history H_t

$$\begin{aligned} v(h) &= var\left(\sum_{t<T_i\leq t+h} z_i\right) \\ &= Evar\left(\sum_{t<T_i\leq t+h} z_i|H_t\right) + varE\left(\sum_{t<T_i\leq t+h} z_i|H_t\right) \\ &= var(z)E\left(N(t+h) - N(t)|H_t\right) + E(z)^2 var\left(N(t+h) - N(t)|H_t\right) \\ &= \left[var(z) + E(z)^2\right]\lambda(t)h = E(z^2)\lambda(t)h. \end{aligned}$$

The intensity is thus proportional to local volatility.

As an example, let $\alpha = 0.01$, $\theta = 0.8$, $\sigma = 20$. The simulated realization of the process shown in Figure 11. The intensity function shown that is simulated hand-in-hand with the point process is shown in Figure 12. The expected number of points per cluster is 5. We should therefore see about 20 clusters in Figures 11 and 12 . Some of these clusters fall on top of each other, and others might have very few points (one).

As another example, consider the process with parameter values $\alpha = 0.01$, $\theta = 0.8$, $\sigma = 50$. There is now less spacing between clusters relative cluster extensions. Simulating 1000 points yields an intensity as shown in Figure 13.

8 Markov chains with continuous state space

Taylor and Karlin (1998) discuss Markov chains with discrete state space, both with discrete and continuous time, and briefly introduces Brownian motion and related Markov processes in continuous time and with continuous state space. Seierstad (2009) deals in more depth with Markov processes continuous in time and with continuous state space. Here we shall look at the fourth category, i.e Markov chains with discrete time and continuous state space.

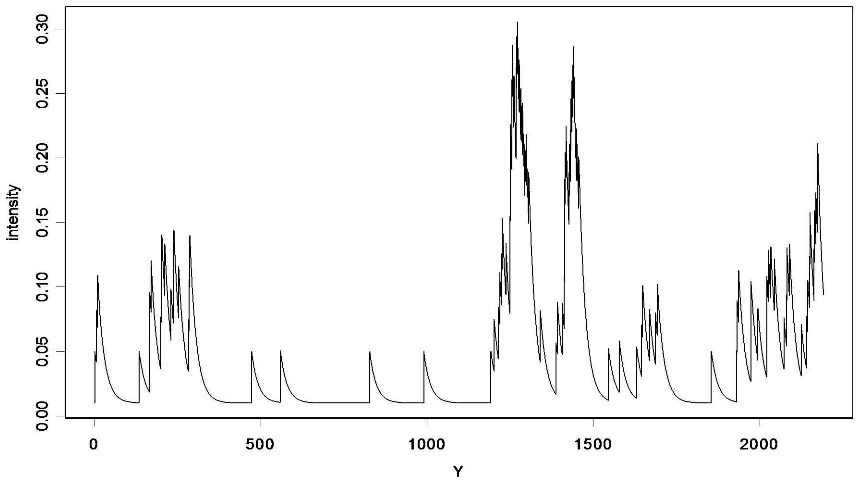

Figure 12: Intensity for a new observation. $\alpha = 0.01$, $\theta = 0.8$, $\sigma = 20$

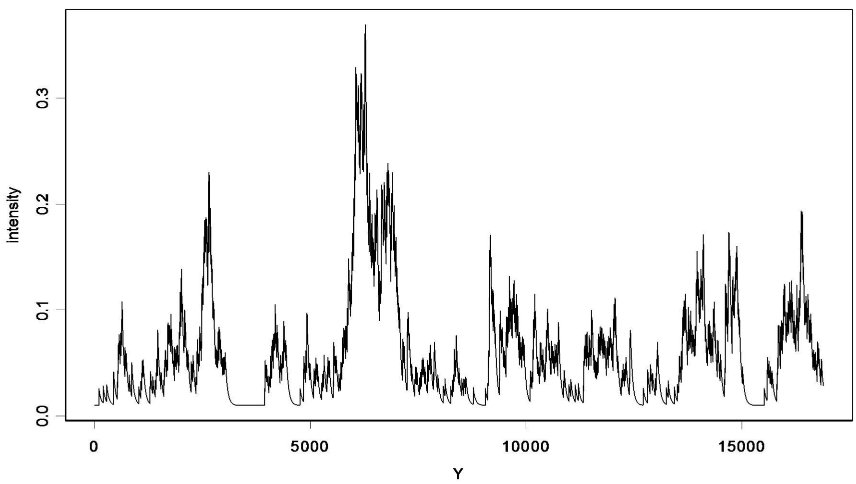

Figure 13: The intensity generated by simulating 1000 points. $\alpha = 0.01, \quad \theta = 0.8, \quad \sigma = 50.$

The theory is no less complex than that for Markov chains with discrete states. Concepts as irreducibility, aperiodicity, strong and weak recurrence, transience, ergodicity, stationarity, recurrence times etc. have the same basic meaning here as when the state space is discrete. The main difference is that summing over states is replaced by integration.

To do integration properly, and to deal properly with the many technical aspects of the theory, measure theory must be mastered. This is outside the scope here, and we will only look at ordinary integration with respect to Lebesgue measure in finite dimensional Euclidean space equipped with the Borel sigma-algebra. The reader unfamiliar with these words should not worry. Just proceed reading, and interpret integration in the familiar way.

Robert and Casella (2004, chapter 4) give a nice introduction to Markov chain theory based on measure theory. The essential for understanding Markov Chain Monte Carlo methods, particularly the Metropolis algorithm, is pulled out from their exposition.

Consider a continuous state space $\mathcal{T}$. A Markov chain $T_1, T_2, \cdots$ is a sequence of stochastic variables with values in $\mathcal{T}$ with the property that the conditional probability of any future event defined by variables $T_{n+1}, T_{n+2}, \cdots$ given the history H_n up until time n (H_n is spanned by T_i $i \le n$), depends only on T_n. The Markov property is as in the discrete case: given the present, the future is independent on the past.

The transition matrix in the discrete case has its counterpart for a continuous state chain in the transition kernel $K(u,t)$. Assume time homogeneity, so that for any n and any (measurable) set $A \subset \mathcal{T}$

$$P\left(T_{n+1} \in A | T_n = u\right) = \int_A K(u,t)dt.$$

It is logical, and indeed customary, to denote the conditional density $K(u,t)$ rather than $K(t|u)$ since the condition $T_n = u$ happens before the consequence in the sequence of the chain.

Example 1. *Let $\mathcal{T}$ be the real line, and consider the AR(1) process $T_{n+1} = \theta T_n + e_n$ where the error terms e_n are iid. This is a Markov chain. If the errors are $N(0,1)$,*

$$K(u,t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\theta u)^2}.$$

What is the transition kernel if the errors are uniformly distributed on the unit interval? The AR(q) process is also a Markov chain. If $X_{n+1} = \theta_0 + \sum_{i=1}^{q} \theta_i T_{n+1-i} + e_n$ the vector process $T_n = (X_n, \cdots X_{n-q+1})$ is Markov on q-dimensional Euclidean space. The transition kernel will in this case be one-dimensional in the sense that given T_n all components of T_{n+1} are given except for the first.

When the kernel is positive, $K(u,t) > 0$ for all t and u in $\mathcal{T}$ the chain is *irreducible*, and $P\left(T_{n+1} \in A | T_n = u\right) > 0$ for all non-null sets A ($\int_A dt > 0$). The chain can also be irreducible under other circumstances, but we shall be content with this simple sufficient condition. If the AR(1) process has normally distributed errors, it is irreducible. Can it be irreducible when the errors are uniform?

If the chain with positive probability can rest in a state, $P(T_{n+1} = T_n) > 0$, the chain must be *aperiodic*.

The chain is *ergodic* if the distribution of T_n converges to a unique distribution regardless of the initial state. A necessary condition for ergodicity is that it is *positive recurrent*, i.e.

that for every non-null A set in $\mathcal{T}$ the chain returns to A infinitively often with probability 1. Recall that the kernel is positive.

A distribution with density π is *stationary* or *invariant* if this is the marginal distribution of T_n when T_1 has this distribution. Breaching conventions and denoting both probability and density by the same symbol ($\pi(A) = \int_A \pi(t)dt,\ K(u,A) = \int_A K(u,t)\,dt$ etc), the defining property of stationarity is

$$\pi(A) = \int K(u,A)\pi(u)du \text{ for all } A \subset \mathcal{T}.$$

Tat this integral is over the whole state space is suppressed in the above notation.

A Markov chain is *reversible* if the conditional distribution of T_n given $T_{n+1} = u$ is identical to that of T_{n+1} given $T_n = u$.

Theorem 8.1. *A chain satisfying the detailed balance condition:*

$$K(u,t)\pi(u) = K(u,t)\pi(t) \text{ for all } t \text{ and } u \text{ in } \mathcal{T} \tag{13}$$

for a probability density π has π as its stationary density and is reversible.

Proof. For any $A \subset \mathcal{T}$ the detailed (because it holds everywhere) balance condition entails, by changing the order of integration,

$$\begin{aligned}
\int_{\mathcal{T}} K(u,A)\pi(u)du &= \int_{\mathcal{T}}\int_A K(u,t)\pi(u)dtdu \\
&= \int_{\mathcal{T}}\int_A K(t,u)\pi(t)dtdu \\
&= \int_A \pi(t)\int_{\mathcal{T}} K(t,u)dudt \\
&= \pi(A).
\end{aligned}$$

The last equation holds because $K(t,u)$ is a probability density in u for each t. That the chain is reversible follows from (13) since it states that the joint distribution of (T_n, T_{n+1}) is identical to that for the reversed pair (T_{n+1}, T_n) in equilibrium. From the identical joint distribution follows identical conditional distributions. □

8.1 MCMC

Markov Chain Monte Carlo (MCMC, or mc^2) is a simulation technique to compute the posterior distribution in a Bayesian setup. Bayes' formula for the posterior density $f(\theta|x)$ from a prior density $f(\theta)$, and a conditional density for x given θ, $f(x|\theta)$ is

$$\pi(\theta) = f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|t)f(t)\,dt},$$

and is used extensively in economics. In some nice models for $f(x|\theta)$, such as the normal and other exponential families, there exist conjugate priors which allow the posterior to be calculated explicitly. In more complex models the denominator $\int f(x|t)f(t)\,dt$ is difficult to calculate when the integral is of high dimension.

The numerator is however given, and can be used to set up a Markov chain with the posterior distribution as its stationary distribution. The MCMC approach is then to simulate this chain in, say, 500000 steps and after a burn-in period, say of 1000 steps to extract the values say 100 steps apart. This sample is then regarded as a sample from the stationary distribution, i.e. the posterior, and statistics of interest are calculated.

In essence, MCMC is an ingenious method to numerically calculate integrals of high dimensions approximately. With T_i $i = 1, \cdots, n$ being a sample from π obtained by an MCMC, the integral $E[h(T)] = \int h(t)\pi(t)\,dt$ is estimated by the mean $\overline{h(T)}$. Importance sampling is another method to obtain a sample from π, and as MCMC it comes in many variants. If the integrand $h(t)\pi(t)$ is close to a Gaussian form, i.e. $\ln(h(t)\pi(t))$ is close to quadratic, the Laplace method is working well.

8.1.1 Bayesian econometrics

Bayesian statistical analysis is more widespread in the natural sciences and medicine than in economics and the social sciences. This is surprising, since the Bayesian approach certainly make more allowance for expert opinion and judgement than the frequentist approach, and judgement has no less place in the social sciences than in the natural sciences. But the Bayesians are on the march also in econometrics, as is evident from the new textbooks (Lancaster 2004; Koop 2003) and from an increasing number of papers. Poirier (1988) provides an opinionated discussion of Bayesianism versus frequentism in economics.

It is the availability of high speed computers and of ingenious algorithms such as importance sampling, Gibbs sampling and the Metropolis algorithm that has made Bayesian calculations feasible also in complex models, see Robert and Casella (2004).

8.1.2 The Metropolis algorithm

N. Metropolis worked with Ulam, Teller and other gifted scientists at the Manhattan project and on developing the hydrogen bomb at Los Alamos. Simulation methods had been used for years with balls in urns and other physical set-ups, but Metropolis was an early user of computer simulation. He is on record for first using the term 'Monte Carlo methods'. In 1953 Metropolis and others published an algorithm for discrete optimization in connection with particle physics. The algorithm was improved by Hastings in 1970, and is in most books referred to as the Metropolis-Hastings algorithm, but only in the 1990s did its usefulness dawn upon the statistical community. See Robert and Casella (2004, section 7.8)

The IEEE computer society list the Metropolis algorithm as one of the 10 algorithms with strongest influence on the practice of science and engineering over the last 100 year. The simplex method for linear programming, the fast Fourier transform and quicksort are other algorithms on the list, see `http://www.computer.org/cise/articles/Top_Algorithms.htm`.

Changing to the notation, but relying on Robert and Casella (2004, section 7.3), let the numerator in Bayes formula be $g(\theta) = f(x|\theta)f(\theta)$. We seek a Markov chain $\{T_n\}$ on the state space $\mathcal{T}$ with $\pi(\theta) = g(\theta) / \int g(t)dt$ as stationary distribution. One such chain is constructed from a proposal kernel

$$q(u,t) > 0 \text{ for all } t \text{ and } u \text{ in } \mathcal{T} .$$

as follows. This kernel is not the kernel for $\{T_n\}$, but is used to construct its kernel $K(u,t)$.

First the so-called acceptance probability

$$\rho(t,u) = \min\left\{\frac{g(u)}{g(t)}\frac{q(u,t)}{q(t,u)}, 1\right\} \tag{14}$$

is constructed. Note that ρ indeed is a probability for all t and u in the state space $\mathcal{T}$ for $\{T_n\}$. Assume $g(t) > 0$ for all $t \in \mathcal{T}$.

The chain is started at some value $t_1 \in \mathcal{T}$. At step n, generate U_n from the distribution with density $q(T_n, u)$, which is the conditional density given T_n. Then generate

$$T_{n+1} = \begin{matrix} U_n & \text{with probability} & \rho(T_n, U_n) \\ T_n & \text{with probability} & 1-\rho(T_n, U_n) \end{matrix} . \tag{15}$$

The proposal value U_n is thus used as the next state with probability ρ, while the chain stays put at its previous state with probability $1-\rho$.

In some cases symmetric proposal kernels are used: $q(t,u) = q(u,t)$. The acceptance probability is then $\min\{g(u)/g(t), 1\}$. A proposal state giving $g(U_n) > g(T_n)$, and consequently higher posterior density, is then accepted with probability 1. The algorithm thus tend to move into regions of high posterior density, while occasionally visiting states with low posterior density. The remarkable fact, to be demonstrated below, is that Metropolis' acceptance mechanism makes the chain visit sets with exactly the desired posterior probabilities in the long run, actually regardless of the proposal kernel q as long as it is everywhere positive.

Is $\{T_n\}$ a Markov chain, and does it really have a unique stationary distribution given by π? First, the Markov property is satisfied since U_n and thus T_{n+1} indeed are conditionally independent of $T_1, \cdots, T_{n-1}$ given T_n. Next, the chain is *aperiodic* except for singular cases, see the Metropolis exercise below. Further, the chain is *irreducible* since $q(u|t) > 0$, and all possible states can thus be reached from any state in one step.

The only remaining question, which actually is the main question, is whether π indeed is the stationary distribution. We proceed to show this by showing that g satisfies the detailed balance equation (13). Since g must have a finite integral for the posterior distribution to exist, we will assume this to be the case. The function $\pi(t) = g(t) / \int g(u)du$ will thus satisfie the detailed balance equation, and is consequently the stationary density for the chain.

The detailed balance equation to be satisfied is

$$g(u)K(u,t) = g(t)K(t,u) \tag{16}$$

for all u and t. By construction,

$$\begin{aligned} K(t,u) &= \rho(t,u)q(t,u) + (1-r(t))\,\delta_t(u), \\ r(t) &= \int \rho(t,u)q(t,u)du. \end{aligned}$$

Here, $\delta_t(u)$ is the Dirach delta function (the limiting density with all mass at t) corresponding to the event that the chain remains in its current state, and $r(t)$ is the probability that it moves to a different state. Since $\rho(t,u) = 1 \Rightarrow \rho(u,t) \leq 1$ the defining relation (14) yields in that case

$$g(t)\rho(t,u)q(t,u) = g(u)\rho(u,t)q(u,t). \tag{17}$$

In the opposite case, $\rho(t,u) < 1$, (14) conversely yields $\rho(u,t) = 1$ and (17) is true. The symmetry relation (17) is therefore true for all pairs of states u and t. Since furthermore

$\delta_u(t) = \delta_t(u) = 0$ for all $u \neq t$, we also have everywhere

$$g(t)\,(1 - r(t))\,\delta_t(u) = g(u)\,(1 - r(u))\,\delta_u(t). \tag{18}$$

Together (17,18) imply the detailed balance equation (16) for the kernel K of $\{T_n\}$. π is indeed the unique stationary distribution of the Metropolis Markov chain. More details on the Metropolis algorithm and the convergence properties of its Markov chain are found in Robert and Casella (2004).

In practice, the proposal kernel q should be chosen to balance two requirements. First, it should be easy to draw values from its conditional distribution. That is, U_n should be computationally cheap to generate regardless of the state T_n. Secondly, q should be chosen to be as proportional to g as possible. This way, the chain will tend to move at each step and the number of steps needed to obtain a reasonable sample will not be excessive.

9 Ornstein–Uhlenbeck processes, expected damage from greenhouse gas emission

Linear differential equations are useful in modeling deterministic processes. The dynamic part of the model of Weitzman (2008) discussed above for how global temperature responds to greenhouse gas emissions is a case of linear differential equation. The evolution of global temperature and radiation is however complex, and the deterministic model is indeed a gross simplification. Without explicit modeling of this complexity, or say the complexity of other processes of interest to economists that are cast in linear differential equations, a possible extension of the model is to let the dynamics be stochastic.

Ornstein–Uhlenbeck processes, or in short OU-processes, are linear stochastic differential equation models with fixed coefficients. The aim of this section is to develop a bit of theory for OU-processes, first in one dimensions. Then a 2-dimensional OU-process model is established for global temperature and radiation, where the deterministic forces of the model is as (1, 2, 3), while these are amended by a continuous stream of stochastic impulses. The question is whether the conclusions of Weitzman (2008) hold up in this more realistic model, or perhaps infinite expected damage is to be expected even under milder conditions. It is actually possible that the more stochasticity there is in the process it will move faster towards higher temperatures.

9.1 One dimension

Ornstein–Uhlenbeck processes are stochastic processes in continuous time and continuous state space. They are Markov processes with linear Gaussian infinitesimal kernel of constant standard deviation. With $X(t)$ a one-dimensional OU-process pulled towards an exogenous level $m(t)$ with force a, the conditional distribution of $X(t + dt)$ given $X(t)$ is in the limit, as $dt \downarrow 0$, normally distributed with mean $a(m(t) - X(t))\,dt$ and variance $\sigma^2 dt$. The motion consists of a drift $a(m(t) - X(t))\,dt$ and a stochastic impulse with distribution $N(0, \sigma^2 dt)$. The stochastic impulses are independent. The standardized cumulative impulses are represented in the so-called Brownian motion process $B(t)$ characterized by having normally and independently distributed differences $B(t_{i+1}) - B(t_i) \sim N(0, t_{i+1} - t_i)$, $t_{i+1} - t_i \geq 0$. Infinitesimally, the motion of $X(t)$ is thus $X(t + dt)\,|X(t) \sim a(m(t) - X(t))\,dt + \sigma(B(t + dt) - B(t)) =$

$a\left(m\left(t\right)-X\left(t\right)\right)dt+\sigma B\left(dt\right).$ This is written as a stochastic differential equation

$$dX\left(t\right)=a\left(m\left(t\right)-X\left(t\right)\right)dt+\sigma B\left(dt\right). \tag{19}$$

Seierstad (2009) discusses Brownian motions and stochastic differential equations, and the OU-process $X\left(t\right)$ pullet towards $m\left(t\right)$ is a special case of the process he discuss in his (4.6) and (4.7). Ito's formula Seierstad (2009;(4.10)) is helpful in "solving" (19). The trick is to consider the process $Y\left(t\right)=g\left(t,X\left(t\right)\right)=X\left(t\right)\exp\left(at\right).$ In Seierstad's formulation, our process has $u\left(t\right)=a\left(m\left(t\right)-X\left(t\right)\right)$ and $v\left(t\right)=\sigma$. Since $g\left(t,x\right)=x\exp\left(-at\right),$

$$\frac{\partial^2}{\partial x^2}g\left(t,x\right)=0,$$

and Ito's lemma yields

$$\begin{aligned}Y\left(t\right) &= X\left(0\right)+\int_0^t\left\{X\left(s\right)ae^{as}+a\left(m\left(s\right)-X\left(s\right)\right)e^{as}\right\}ds+\int_0^t\sigma e^{as}B\left(ds\right)\\ &= X\left(0\right)+\int_0^t am\left(s\right)e^{as}ds+\sigma\int_0^t e^{as}B\left(ds\right).\end{aligned}$$

Substituting for $Y\left(t\right)$ the solution is

$$X\left(t\right)=e^{-at}X\left(0\right)+\int_0^t am\left(s\right)e^{a\left(s-t\right)}ds+\sigma\int_0^t e^{a\left(s-t\right)}B\left(ds\right).$$

The effect of the initial value $X\left(0\right)$ decays exponentially fast, and the effect of the moving target $m\left(s\right)$ is an exponential filter of this process. From the solution,

$$EX\left(t\right)=e^{-at}EX\left(0\right)+\int_0^t am\left(s\right)e^{a\left(s-t\right)}ds$$

since $E\left[\int_0^t e^{a\left(s-t\right)}B\left(ds\right)\right]=\int_0^t e^{a\left(s-t\right)}E\left[B\left(ds\right)\right]=\int_0^t e^{a\left(s-t\right)}0ds=0$. Since further for $u<t$

$$\int_0^t e^{a\left(s-t\right)}B\left(ds\right)=\int_0^u e^{a\left(s-t\right)}B\left(ds\right)+\int_u^t e^{a\left(s-t\right)}B\left(ds\right)$$

where the two pieces are stochastically independent,

$$\begin{aligned}& cov\left[\sigma\int_0^t e^{a\left(s-t\right)}B\left(ds\right),\sigma\int_0^u e^{a\left(s-u\right)}B\left(ds\right)\right]\\ =\ & \sigma^2 cov\left[\int_0^u e^{a\left(s-t\right)}B\left(ds\right),\int_0^u e^{a\left(s-u\right)}B\left(ds\right)\right]\\ =\ & \sigma^2 e^{a\left(u-t\right)}cov\left[\int_0^u e^{a\left(s-u\right)}B\left(ds\right),\int_0^u e^{a\left(s-u\right)}B\left(ds\right)\right]\\ =\ & \sigma^2 e^{a\left(u-t\right)}var\left[\int_0^u e^{a\left(s-u\right)}B\left(ds\right)\right]\\ =\ & \sigma^2 e^{a\left(u-t\right)}\int_0^u e^{2a\left(s-u\right)}var\left[B\left(ds\right)\right]\\ =\ & \sigma^2 e^{a\left(u-t\right)}\int_0^u e^{2a\left(s-u\right)}ds=\frac{\sigma^2}{2a}\left(e^{a\left(u-t\right)}-e^{-a\left(u+t\right)}\right),\end{aligned}$$

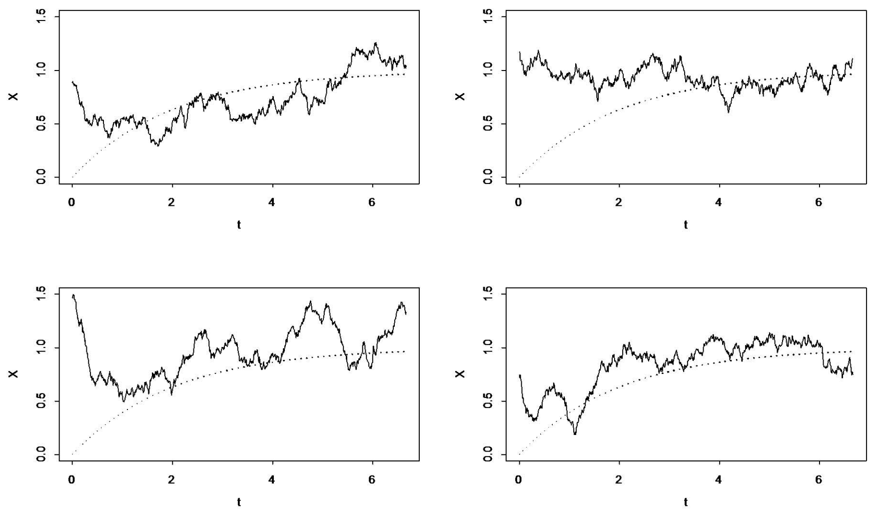

Figure 14: Four replicate simulations (using the method of Problem 21) of an OU-process with $X(0) \sim N\left(1, \frac{\sigma^2}{2a}\right)$, $m(t) = 1 - e^{-t/2}$, $a = 1$ and $\sigma = 1/4$.

and

$$\begin{aligned} cov\left[X(t), X(u)\right] &= e^{-a(u+t)} var\left[X(0)\right] + \frac{\sigma^2}{2a}\left(e^{a(u-t)} - e^{-a(u+t)}\right) \\ var\left[X(t)\right] &= e^{-2at} var\left[X(0)\right] + \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right). \end{aligned}$$

It is also clear that $(X(t_1), X(t_2), \cdots, X(t_n))$ has a multinormal distribution whenever $X(0)$ is normally distributed. The OU-process is thus a Gaussian process in that case.

The OU-process can only be stationary when the deterministic target is not moving, i.e. $m(s) = \mu$. If then $X(0) \sim N\left(\mu, \frac{\sigma^2}{2a}\right)$, $EX(t) = \mu$, $cov\left[X(t), X(u)\right] = \frac{\sigma^2}{2a} e^{a(u-t)}$ $u < t$ and $var\left[X(t)\right] = \frac{\sigma^2}{2a}$. The process is then a stationary Gaussian process in the sense that $(X(t_1 + u), X(t_2 + u), \cdots, X(t_n + u))$ has for all values of u the same multinormal distribution as $(X(t_1), X(t_2), \cdots, X(t_n))$ when all time points are non-negative.

9.2 Many dimensions

Let now $X(t)$ be a vector process of p dimensions, and let $B(t)$ be a p-dimensional Brownian motion. The components of B are then independent one-dimensional Brownian motions. The system of linear stochastic differential equations determining the p-dimensional OU-process X is given by a $p \times p$ matrix of drift coefficients A, a $p \times p$ matrix Σ representing the square root of the infinitesimal covariance matrix Σ^2, and a deterministic (exogenous) p-dimensional process b. The equation taking (19) to p dimensions is

$$dX(t) = (b(t) + AX(t))\, dt + \Sigma B(dt). \tag{20}$$

When A is invertible, the equation can be written

$$dX(t) = -A\left(-A^{-1} b(t) - X(t)\right) dt + \Sigma B(dt),$$

and X is seen to track the p-dimensional trajectory $m(t) = -A^{-1}b(t)$.

Ito's lemma is generalized to p dimensions, see Øksendal (2005; Section 4.2). It will be applied to $Y(t) = g(t, X(t))$ where $g(t,x) = C(t)x$ is a vector-valued function of t (one dimension) and x (p dimensions), and C is a $p \times p$ matrix (see next Section) satisfying

$$\frac{d}{dt}C(t) = -C(t)A = -AC(t). \tag{21}$$

Since $\frac{d^2}{dx'dx}g = 0$, the equation for Y is

$$\begin{aligned} dY(t) &= C'(t)X(t)dt + C(t)dX(t) \\ &= -AC(t)X(t)dt + C(t)(b(t) + AX(t)dt + \Sigma B(dt)) \\ &= C(t)b(t)dt + C(t)\Sigma B(dt). \end{aligned}$$

Integrating this yields the representation

$$X(t) = C(t)^{-1}\left[X(0) + \int_0^t C(s)b(s)ds + \int_0^t C(s)\Sigma B(ds)\right].$$

To enable calculations the matrix C needs to be re-expressed. Diagonalization is always possible if A is symmetric and real (Sydsæter et al. 2005), see below. It is also possible there is a $p \times p$ matrix V holding linearly independent row vectors that are left eigenvectors for A, and Γ is the diagonal $p \times p$ matrix holding the corresponding eigenvalues. That is, $VA = \Gamma V$. There could be multiple eigenvalues, and some could be zero. They appear in Γ with their respective multiplicity. Conditions for the existence of such a non-singular V is known from linear algebra. The drift matrix might then be written

$$A = V^{-1}\Gamma V,$$

and V^{-1}hold the right eigenvectors of A. The matrix

$$C(t) = V^{-1}e^{-t\Gamma}V.$$

is seen to be the $p \times p$ matrix solving the above differential equations. The matrix $e^{-t\Gamma}$ is diagonal, and when γ_k is the k-th eigenvalue of A, the k-th diagonal element of $e^{-t\Gamma}$ is $e^{-t\gamma_k}$.

The resulting representation of our process is consequently

$$X(t) = V^{-1}\left[e^{-t\Gamma}VX(0) + \int_0^t e^{(t-s)\Gamma}Vb(s)ds + \int_0^t e^{(t-s)\Gamma}V\Sigma B(ds)\right],$$

and

$$\begin{aligned} EX(t) &= V^{-1}\left[e^{-t\Gamma}VX(0) + \int_0^t e^{(t-s)\Gamma}Vb(s)ds\right] \\ cov(X(u), X(t)) &= V^{-1}\left[\int_0^u e^{(u-s)\Gamma}V\Sigma^2V'e^{(t-s)\Gamma}ds\right]\left(V^{-1}\right)' \quad u \le t. \end{aligned}$$

The (j,k) element of $\int_0^u e^{(u-s)\Gamma}V\Sigma^2V'e^{(t-s)\Gamma}ds$ is

$$\int_0^u e^{(u-s)\gamma_j}\left[V\Sigma^2V'\right]_{jk}e^{(t-s)\gamma_k}ds = \left[V\Sigma^2V'\right]_{jk}\left(e^{u\gamma_j+t\gamma_k} - e^{(t-u)\gamma_k}\right)\frac{1}{\gamma_j+\gamma_k}.$$

9.3 Stochastic climate dynamics, a simple OU-model

The model of Weitzman(2008) studied above is a system of two linear differential equations for global mean temperature $T(t)$ and amount of greenhouse gas in the atmosphere $F(t)$. Rephrased in parameters $\alpha = (1-f)/c$ and $\lambda = \lambda_0/(1-f)$ they are

$$\begin{aligned} dT(t) &= \frac{1-f}{c}\left(\frac{\lambda_0}{1-f}F(t) - T(t)\right)dt = \alpha(\lambda F(t) - T(t))\,dt \\ dF(t) &= \beta\left(\overline{F} - F(t)\right)dt. \end{aligned}$$

Climate develops according to complex laws of nature. Subsuming the complexity in addition to the linear drift of the system in independent and stochastic impulses modifying the deterministic motion, an OU-model is worth investigating. With

$$X(t) = \begin{bmatrix} T(t) \\ F(t) \end{bmatrix}, \quad b(t) = \begin{bmatrix} 0 \\ \beta\overline{F} \end{bmatrix}, \quad A = \begin{bmatrix} -\alpha & \alpha\lambda \\ 0 & -\beta \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

and $B(t) = \begin{bmatrix} B_1(t) & B_2(t) \end{bmatrix}'$ a 2-dimensional Brownian motion, the model

$$\begin{aligned} dT(t) &= \alpha(\lambda F(t) - T(t))\,dt + \sigma_1 B_1(dt) \\ dF(t) &= \beta\left(\overline{F} - F(t)\right)dt + \sigma_2 B_2(dt) \end{aligned}$$

is a case of (20).

The question posed by Weitzman(2008) is wether $T(t)^2$ increases, and how fast it will grow. Since now temperature is stochastic, the question must be how $E\left[T(t)^2\right] = (ET(t))^2 + var(T(t))$ develops. This is found by running the machinery developed above. And to do that the eigenvalues and eigenvectors of A must be calculated.

Due to the triangular structure of A the model is hierarchical, with $F(t)$ being an autonomous OU-process, and with a causality flow $F \to T$. This hierarchical structure makes it easy to calculate the eigenstructure of the process provided $\alpha \neq \beta$. It is

$$V = \begin{bmatrix} 1 & \frac{\alpha\lambda}{\beta-\alpha} \\ 0 & 1 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} -\alpha & 0 \\ 0 & -\beta \end{bmatrix}, \quad V^{-1} = \begin{bmatrix} 1 & -\frac{\alpha\lambda}{\beta-\alpha} \\ 0 & 1 \end{bmatrix}.$$

With the process representing values in excess of pre industrial levels, the initial value is $X(t) = 0$. By some algebra

$$\begin{aligned} EX(t) &= V^{-1}\int_0^t e^{(t-s)\Gamma} V b(s)\,ds \\ &= \begin{bmatrix} \lambda\overline{F}\left(1 + \frac{1}{\beta-\alpha}\left(\alpha e^{-\beta t} - \beta e^{-\alpha t}\right)\right) \\ \overline{F}\left(1 - e^{-\beta t}\right) \end{bmatrix}. \end{aligned}$$

When $\alpha = \beta$ the drift matrix A cannot be diagonalized. The matrix

$$C(t) = e^{\alpha t}\begin{bmatrix} 1 & -\lambda\alpha t \\ 0 & 1 \end{bmatrix} \tag{22}$$

satisfies the differential equation (21), and the solution can be computed.

In the original parametrization this yields

$$\begin{aligned} ET(t) &= \lambda\overline{F}\left(1+\frac{1}{\beta-\alpha}\left(\alpha e^{-\beta t}-\beta e^{-\alpha t}\right)\right) \\ &= \frac{\lambda_0}{1-f}\overline{F}\left(1+\frac{1-f}{c\beta-(1-f)}e^{-\beta t}-\frac{c\beta}{c\beta-(1-f)}e^{-\frac{1-f}{c}t}\right). \end{aligned} \tag{23}$$

This is the same as the deterministic result (6). Is this a surprise? With a damage function D at least as convex as the quadratic in T Weitzman found the expected discounted damage to be infinite under his conditions on the uncertainties in f and r. Since

$$E\int_0^\infty D(T(t)) \geq E_{f,r}\int_0^\infty E[(T(t)^2)] = E_{f,r}\int_0^\infty [(E(T(t))^2) + var(E(t))] \tag{24}$$

Weitzman's result for the deterministic dynamic model is sufficient to conclude that expected damage is infinite in the stochastic differential equation model for global warming.

Since we were unable to see Weitzman's result in the deterministic case, we might hope that the variance in global increased temperature i sufficient to give infinite expected damage. The variance does require some algebra. With $g=\frac{\alpha\lambda}{\beta-\alpha}$,

$$V\Sigma^2V' = \begin{bmatrix} \sigma_1^2+g^2\sigma_2^2 & g\sigma_2^2 \\ g\sigma_2^2 & \sigma_2^2 \end{bmatrix}.$$

The (j,k) element of $\int_0^t e^{(t-s)\Gamma}V\Sigma^2V'e^{(t-s)\Gamma}ds$ is in the general case $\left[V\Sigma^2V'\right]_{jk}\left(e^{t(\gamma_j+\gamma_k)}-1\right)\frac{1}{\gamma_j+\gamma_k}$. In our case the eigenvalues are $\gamma_1=-\alpha$ and $\gamma_2=-\beta$. The required matrix is thus

$$\begin{aligned} &\int_0^u e^{(u-s)\Gamma}V\Sigma^2V'e^{(t-s)\Gamma}ds \\ =& \begin{bmatrix} \frac{1}{2\alpha}\left(1-e^{-2\alpha t}\right)\left(\sigma_1^2+g^2\sigma_2^2\right) & \frac{1}{\alpha+\beta}\left(1-e^{-(\alpha+\beta)t}\right)g\sigma_2^2 \\ \frac{1}{\alpha+\beta}\left(1-e^{-(\alpha+\beta)t}\right)g\sigma_2^2 & \frac{1}{2\beta}\left(1-e^{-2\beta t}\right)\sigma_2^2 \end{bmatrix}. \end{aligned}$$

From this,

$$\begin{aligned} & var\left[T(t)\right] \\ =& \left[V^{-1}\int_0^u e^{(u-s)\Gamma}V\Sigma^2V'e^{(t-s)\Gamma}ds\left(V^{-1}\right)'du\right]_{1,1} \\ =& \frac{\sigma_1^2}{2\alpha}\left(1-e^{-2\alpha t}\right)+\left(\frac{\alpha\lambda}{\beta-\alpha}\right)^2\sigma_2^2\left[\frac{1}{2\alpha}\left(1-e^{-2\alpha t}\right)-\frac{2}{\alpha+\beta}\left(1-e^{-(\alpha+\beta)t}\right)+\frac{1}{2\beta}\left(1-e^{-2\beta t}\right)\right], \end{aligned} \tag{25}$$

and it is clear that $E_{f,r}\int 0\infty var(T(t)dt$ is finite.

10 Decomposition of matrices

Sydsæter et al. (2005) present some linear algebra in the first chapter, including eigenvalues and diagonalization. below, some supplementary stuff on singular value decomposition and diagonalization.

The Peron-Frobenius theorem allows a rather complete understanding of the limiting behaviour of finite irreducible Markov chains. The theorem provides a decomposition of the transition matrix which has a superficial similarity with singular value decomposition. Since singular value decomposition, SVD, is so useful in econometrics and elsewhere, and since it perhaps is not paid sufficient attention to in economics, a brief description is included despite SVD might not be that central to the study of Markov chains.

10.1 Singular Value Decomposition (SVD)

Any $m \times n$ real matrix A can be factored into a product $A = U\Lambda V^t$, with U and V^t real orthogonal $m \times m$ and $n \times n$ matrices, respectively, and Λ a diagonal matrix with positive numbers in the first rank(A) entries on the main diagonal and zeroes everywhere else. The entries on the main diagonal of Λ are called the *singular values* of A. This factorization $A = U\Lambda V^t$ is called a *singular value decomposition* of A. Singular value decomposition is a direct consequence of the Spectral theorem for symmetric matrices (Sydsæter et al. 2005, theorem 1.7.2). This is so since AA^t and A^tA both are symmetric and real matrices, both with eigenvalues in Λ^2, and with eigenvectors in U and V respectively.

If A has rank $r \leq m \leq n$, say, the zero singular values can be dropped and this is also the rank of $AA^t = U\Lambda^2U^t$ and $A^tA = V\Lambda^2V^t$ where U is $n \times r$ with orthogonal columns and V is $m \times r$ also with orthogonal columns. .Se for example Anderson, TW *An Introduction to multivariate statistical analysis* (1984).

Example 2.

$$\begin{bmatrix} 5 & -5 & -3 \\ -3 & 0 & 5 \\ 1 & 5 & 4 \end{bmatrix}$$
$$= \begin{bmatrix} -.722 & -.191 & -.665 \\ .455 & .593 & -.664 \\ .522 & -.782 & -.341 \end{bmatrix} \begin{bmatrix} 10.1 & 0 & 0 \\ 0 & 4.61 & 0 \\ 0 & 0 & 3.56 \end{bmatrix} \begin{bmatrix} -.443 & .618 & .649 \\ -.763 & -.640 & .0897 \\ -.471 & .456 & -.755 \end{bmatrix}$$

These two outer matrices fail the orthogonality test because they are numerical approximations only. You can check the inner products of the columns to see that they are "approximately" orthogonal.

10.2 Diagonalization

The antilog of a square matix A is $e^A = \sum_{i=0}^{\infty} \frac{1}{i!}(-A)^i$. For linear SDEs the matrix solving $\frac{d}{dt}C(t) = -C(t)A = -AC(t)$ was found useful. When $A = V\Lambda V^{-1}$ the matrix C was found by diagonalization. It is in fact

$$C(t) = e^{-At} = \sum_{i=0}^{\infty} \frac{1}{i!}(-At)^i$$

10.3 Peron-Frobenius theorem

Markov transition matrices have the special property of being non-negative, and having rows that sum to 1. They have a decomposition which only in appearance look like an SVD. Some eigenvalues of $.P$ will be complex when the chain is cyclic. Let $i = \sqrt{-1}$.

Theorem 10.1. *Let P be the transition matrix of a finite irreducible Markov chain with period d and all eigenvalues distinct. Then $P = U\Lambda V^t$, where U is the inverse of V^t, $V^tU = I$ making $P^n = U\Lambda^n V^t$ and where Λ is a diagonal matrix holding the eigenvalues of P. There are d eigenvalues of norm 1, say $\lambda_1 = 1$, $\lambda_k = e^{2\pi i \frac{k-1}{d}}$ $k = 2, \cdots, d$ and the remaining have norm less than 1, $|\lambda_k| < 1$.*

Grimmet and Stritzaker (1992) discuss the Peron-Frobenius theorem, also for the case of multiple eigenvalues.

Example 3. *Let* $P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ .5 & 0 & .5 & 0 \\ 0 & .5 & 0 & .5 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ *The period is 2 and the first two eigenvalues are +1 and -1. The other two are $\pm\frac{1}{2}$. Actually,*

$$\begin{aligned} P &= \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{2}{3} & -\frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} -1 & & & \\ & 1 & & \\ & & -.5 & \\ & & & .5 \end{bmatrix} \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{6} & \frac{1}{3} & -\frac{1}{3} & \frac{1}{6} \end{bmatrix} \\ &\quad + \frac{1}{2} \begin{bmatrix} -\frac{1}{3} \\ -\frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{3} \end{bmatrix} \begin{bmatrix} -1 & -1 & 1 & 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{6} \\ -\frac{1}{6} \\ \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix} \end{aligned}$$

Making

$$P^{2n} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} & 0 \end{bmatrix} + \left(\frac{1}{2}\right)^{2n} \begin{bmatrix} \frac{2}{3} & 0 & -\frac{2}{3} & 0 \\ 0 & \frac{1}{3} & 0 & -\frac{1}{3} \\ -\frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & -\frac{2}{3} & 0 & \frac{2}{3} \end{bmatrix}$$

and

$$P^{2n+1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & \frac{2}{3} & 0 & \frac{1}{3} \end{bmatrix} + \left(\frac{1}{2}\right)^{2n+1} \begin{bmatrix} 0 & \frac{2}{3} & 0 & -\frac{2}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} & 0 \\ 0 & -\frac{1}{3} & 0 & \frac{1}{3} \\ -\frac{2}{3} & 0 & \frac{2}{3} & 0 \end{bmatrix}$$

There is one limit for P^{2n} and another for P^{2n+1} for n running over the integers.

11 Some problems

Problem 1 (Mean and variance). Show by partial integration in the continuous case, and partial summation in the discrete case, that $EX = \int_0^\infty (1 - F(x))\, dx$ when F is the cdf for X and $F(0) = 0$. Why is thus $EX = -\int_0^\infty F(x)dx + \int_0^\infty (1 - F(x))\, dx$ when the expectation exists? Show also that $E\left[X^2\right] = \infty$ is equivalent to $var(X) = \infty$ and $E\left[(X-k)^2\right] = \infty$ for all finite k.

Problem 2 (Infinite damage from global warming?). Assume the feed-back from temperature to radiation, equation (2), to be non-linear, say with f being an increasing function of T making $f(T)T$ convex and with $f(\infty)$ being Weitzman's f. Under Weitzman's distributional assumptions, will then expected discounted damage due to global warming be infinite?

Problem 3. Prove that when X has a gamma distribution with shape parameter α and scale parameter β, the moment EX^c exists and $EX^c = \frac{\Gamma(\alpha+c)}{\Gamma(\alpha)}\beta^c$ whenever $c > -\alpha$. Use (8) to show that $EX = \alpha\beta$, $var(X) = \alpha\beta^2$ and that the third central moment, reflecting skewness, is $E(X - EX)^3 = 2\alpha\beta^3$.

Problem 4. Show that $X = Z^2$ has the density

$$\frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}$$

when Z has the standard normal distribution with density $\varphi(z) = \exp\left(-z^2/2\right)/\sqrt{2\pi}$. Thus, the distribution of Z^2 is indeed the chi-square distribution with 1 degree of freedom, which is the gamma distribution with $\alpha = \frac{1}{2}$ and $\beta = 2$. Comparing with (7), you find $\Gamma(0.5) = \sqrt{\pi}$, which is correct.

Problem 5. Use the convolution formula to prove that the sum of two independent gamma variables with identical scale parameters is gamma distributed.

Problem 6. Show that the gamma distributed is conjugated to the Poisson distribution in the following sense. If the prior distribution for the Poisson parameter μ is gamma with shape parameter α and scale parameter β, and if the result of the Poisson experiment is the number x, the posterior distribution for μ is gamma, actually with shape parameter $x + \alpha$ and scale parameter $1 + 1/\beta$.

Problem 7 (Criminal carriers). Are there distinct groups in the population with respect to criminal carriers (number of criminal charges per year from age 15 to age 24, say)? Moffit (1993) hypothesize two types, the adolescence limited, and the life course persistent criminals – in addition to the non-criminals. Torbjørn Skardhamar at SSB is interested in whether models with latent grouping (distinct groups, but with unobserved group membership) as suggested by Moffit, or latent models with continuously distributed heterogeneity is the more appropriate. Skardhamar (2010), and in a forthcoming paper is looking at the following model. Let a person be faced with Y_a criminal charges from age a to $a+1$. Let it be characterized by observable attributes x. Assume one-dimensional continuous random heterogeneity, and let λ be a stochastic variable representing this latent variation. Given x and λ $\{Y_a\}$ are independent and Poisson distributed, $E[Y_a|x,\lambda] = \mu(a,x)\lambda$. Let λ be gamma distributed with shape- and scale parameter (α,β). Find the joint distribution of $\{Y_a\}$. The covariate vector x might be history-dependent in the sense that x_a is the covariate vector for Y_a, $E[Y_a|x_a,\lambda] = \mu(a,x_a)\lambda$, and x_a might depend on history of criminal charges up to age $a-1$. What would then the joint distribution be? Consider now the discrete latent model. What would the joint distribution be if λ instead has a three-point discrete distribution as suggested by Moffit? Software for fitting latent group models have been developed by Nagin, see Nagin (2005).

Problem 8 (Borel's Paradox). You shall see that the conditional density of a stochastic variable X given an event B of probability zero is ill-defined. The conditional density is in this case really $f(x|B) = \frac{0}{0}$, which certainly can be everything. The "paradox" is that we **define** the conditional density of X given $Z = z$ to be

$$f^{X|Z}(x|z) = \frac{f^{XZ}(x,z)}{f^{Z}(z)}.$$

The denominator represents the event $P(Z = z) = 0$, and the conditional density is thus ill-defined for a given point z. The conventional conditional density is only one of an infinity of possible conditional densities at z. However, regarded also as a function of z, the conventional conditional density has some nice properties. The most important is that it ensures the rule of conditional expectation, $EE(X|Z) = E(X)$.
As an example, let X and Y be iid uniform (0,1). What is the conditional distribution of X given the event $B = \{X = Y\}$? This event, which has probability zero, can be expressed in many ways. Let first $Z_1 = X - Y$. Compute the joint distribution of (X, Z), and use the conventional definition of conditional density $f^{X|Z_1}(x|0)$ that we seek. Then let $Z_2 = \frac{X}{Y} - 1$, and calculate $f^{X|Z_2}(x|0)$ in the same way. To understand what is going on, sketch the two events $B_{1\varepsilon} = \{|X - Y| \leq \varepsilon\}$ and $B_{2\varepsilon} = \left\{\left|\frac{X-Y}{Y}\right| \leq \varepsilon\right\}$ in the unit square. What are then, approximately, the two proper conditional probabilities $P(X \leq x| \quad |X - Y| \leq \varepsilon)$ and $P(X \leq x| \quad \left|\frac{X-Y}{Y}\right| \leq \varepsilon)$? Do these conditional probabilities agree with the two conditional densities you calculated above? You can define an infinity of different neighborhood system that converge to the null event B. Each of these neighborhood systems produce a conditional density through a limiting process, and an infinity of different conditional densities emerge.
The Borel's Paradox played a huge role in the Scientific Committee of the IWC in 1995 when the Bayesian Synthesis method used to assess the Alaskan stock of bowhead whales fell prey to the Borel Paradox (Schweder and Hjort 1996). The problem was that in the Bayesian analysis there were more prior distributions than there were free parameters. With only one parameter α in the unit interval, say, X and Y could carry the two independent prior distributions for α, and the conditional distribution for X given the event $X = Y$ is undetermined – so the Bayesian has a problem in melding his surplus of prior distributions!

Problem 9 (Incomes and elections). Verify the stationary distribution (11). Let the success run T be the number of successive periods that C is in office. You shall determine $E(T)$. Let V_n be the n-th return time to the state $\theta = 0$ for a chain that starts in this state. Why is $T = \sum_{n=1}^{N} V_n$ where N has a geometric distribution with success probability $\frac{1}{2}$, making $E(T) = 2EV$? Use the Basic limit theorem for Markov chains (TK Theorem 4.1) to determine EV from the stationary distribution for θ. What is the expected success run for the party L? Returning to the income distribution, under what conditions is expected income higher when C is in power than under L, and when is the income distribution more dispersed? In the numerical example, median income turned out the same. Is this a coincidence, or will this generally be the case?

Problem 10 (Roots of Markov chains). In his paper "Does permanent income determine the vote?" Jo Thori Lind has the following problem. He has estimated the transition matrix $\mathbf{P}$ of a Markov chain with three states. The time step is one year. He would however like to find the quarterly transition matrix, or more generally $\mathbf{P}^h$ for $h > 0$. For given h Lind fears that there there might be several solutions. Is the solution unique if it is required to be continuous in h and to tend to the identity matrix when h tends to zero? Or is this restriction unnecessary? Lind's solution is to consider a continuous time

Markov chain with infinitesimal generator $\mathbf{A}$ such that $\mathbf{P}^h = \exp(\mathbf{A}h)$. See Taylor and Karlin page 394. Find $\mathbf{A}$ for the $\mathbf{P}$ given below The Peron-Frobenius theorem is useful here. Explain. Lind's estimate is $\mathbf{P} = \begin{bmatrix} 0.33273 & 0.48364 & 0.18364 \\ 0.16167 & 0.64607 & 0.19226 \\ 0.11750 & 0.44125 & 0.44125 \end{bmatrix}$ and his quaerterly solution is $\mathbf{P}^{1/4} = \begin{bmatrix} 0.71650 & 0.21457 & 0.068938 \\ 0.074017 & 0.85003 & 0.075954 \\ 0.038078 & 0.18008 & 0.78184 \end{bmatrix}$. Is he right? Do the numerical calculation!

Problem 11 (Regression paradox). Stigler (1986) tells the story of Francis Galton and his difficulty with the so-called regression paradox: why does not the distribution of heights of males in a population regress towards a distribution concentrated at the mean? Galton was puzzled that we do not see such a regression towards the mean, since tall fathers tend to get shorter sons and short fathers get taller sons. Construct a Markov chain for this process assuming a normal transition kernel making fathers and sons have the same marginal distribution. Show that this marginal distribution indeed is the stationary distribution for the chain.

Problem 12 (The epidemics of rumors). Consider a rumor that is spread by direct contact among N individuals in a population. The process is continuous in time, and time is measured in hours, say. Each individual has a constant intensity λ of making contact with any other individual, i.e. the waiting time until the individual a makes contact with a particular other individual, say b, is exponentially distributed with mean $1/\lambda$. If an individual that has heard the rumor makes a contact with an individual that has not heard the rumor, it is spread on. If however the contacted has already heard the rumor nothing happens. Let $X(t)$ be the number of individuals that has heard the rumor after t units of time. Why is $X(t)$ a Markov process? Is it natural to call it a finite birth process? What is the infinitesimal generator for the process? Why is N an absorbing state? Let $X(0) = 1$. Let T be the time until absorbtion (when everybody has heard the rumor). Show that the expected time until absorbtion is

$$E(T) = \sum_{n=1}^{N-1} \frac{1}{\lambda n (N-n)}.$$

Note that $\frac{1}{n(N-n)} = \frac{1}{N}\left(\frac{1}{n} + \frac{1}{N-n}\right)$ and that $\sum_{n=1}^{N-1} \frac{1}{n} = \ln(N-1) + \gamma + O(N^{-1})$ where $\gamma = 0.5772156649015328606065120900.$ is Euler's constant and $O(N^{-1})N$ converge to some constant as $N \to \infty$. Use this to argue that

$$E(T) = \frac{2(\ln(N-1) + \gamma)}{\lambda N} + \frac{O(N^{-1})}{N}.$$

Is it surprising that a rumor spreads faster to the whole population the more individuals there are?

Problem 13. A random walk with continuously distributed steps is a Markov chain. Show that for no function with finite integral can the detailed balance equation hold for such a chain.

Problem 14 (Metropolis). Show that in the Metropolis Markov chain the event $T_{n+1} = T_n$ has positive probability provided $g(u)q(u,t) \neq g(t)q(t,u)$ for u and t in a set A of positive probability. Show further that the Markov chain $\{U_n\}$ generated by the kernel q is identical to $\{T_n\}$ if $g(u)q(u,t) = g(t)q(t,u)$ everywhere, and thus has the same stationary distribution. The chain might then be periodic.

Problem 15 (MCMC). The yearly number of bankruptcies X in a certain industry is assumed Poisson distributed with rate λ. A Bayesian posterior distribution is sought for λ based on the prior belief that λ is gamma distributed with shape parameter α and scale parameter β. Since the Poisson and the gamma are conjugate distributions, there is really no problem. Do however construct an MCMC to recover the posterior distribution. Use the normal proposal distribution. Make a program and run it to confer that the posterior indeed is a gamma with shape parameter 6 and scale parameter 1/3 when $X = 4$, $\alpha = 2$ and $\beta = 1/2$.

Problem 16. Consider the model for income assimilation of immigrants presented in Section 5. Without back-emigration, immigrants having resided r units of time would have an income distribution with density proportional to $g(x/\alpha(r))$ under the model. Express the mean and variance of X_r for these immigrants in terms of EY and $var(Y)$ in that case? Explain why the income density for immigrants still residing after r units of time is proportional to

$$g\left(x/\alpha(r)\right) \exp\left(-\Lambda(r; x/\alpha(r))\right)$$

when back-emigration follows the proposed model. In Section 5, the intensity of back-emigration was assumed to fall with income. That need not be the case. Assume now the model $\lambda(r) = \theta'(r)\ln(y)$ where θ is positive, increasing and concave. Show that the income for immigrants still residing after r units of time is gamma distributed provided incomes are gamma distributed for natives. It is conceivable that $\alpha(\infty) > 1$. Why is this a necessary condition for immigrants to be fully assimilated with respect to income when back-emigration intensity is increasing in income?

Problem 17 (Random walk?). Campbell et al. (1997) discuss various methods for testing the hypothesis of the return on an asset being a random walk. Consider the following extension of their simple two-state Markov chain model on page 38, let the chain Y_t have states -1, 0, 1, where 0 denotes no change in price, usually because of no trade that day, and the other two states reflect a drop or an increase in price. Let the transition probability matrix be $\begin{bmatrix} v & u & 1-u-v \\ \frac{1-u}{2} & u & \frac{1-u}{2} \\ 1-u-w & u & w \end{bmatrix}$.
What is the range for (u, v, w)? The so-called CJ-statistic is $N_s(n)/N_r(n)$ where $N_s(n)$ is the number of sequences of length 2 in the chain $\{Y_t\}$ over $t = 1, \cdots, n$ of either consistently non-negative or non-positive returns, while $N_r(n)$ is the number of reversals from non-positive to positive or from non-negative to negative. If, for example $n = 10$ and the ten first states happen to be 0,1,1,1,0,1,-1,-1,0,1 making $N_s(10) = 7$ (there are actually 5 non-negative sequences in the first 6 values) and $N_r(10) = 2$. By an alternative definition $N_s(n)$ is the number of sequences in the chain $\{Y_t\}$ over $t = 1, \cdots, n$ of either consistently non-negative or non-positive returns. Then $N_s(10) = 3$. Work out the stationary distribution of the chain Calculate $\mu_s = EN_s(n)/n$ under both definitions. What is $\mu_r = EN_r(n)/n$? Are these numbers independent of n? What is the "theoretical" value of the CJ-statistic, μ_s/μ_r under the two definitions? The value of the asset has steps X_t. The value process is a random walk for some values of (u, v, w). What are these values? What is μ_s/μ_r under the random walk

hypothesis in the two cases? Can the same values be obtained when there are dependencies between consecutive steps? Which version of the CJ-method is useful for testing the random walk hypothesis?

Problem 18 (Random walk?, continued.). In the context of the previous problem, consider the branching point process model for time points of tradings in a specific asset. The branching point process model is discussed in Section 7. There, the value of the asset, Z is thought of as a random walk over time points of trading. Extend this model to the case of the returns (the change in value, possibly on the logarithmic scale) follows the Markov chain $\{Y_t\}$ of the previous problem. A reasonable hypothesis is that the returns are positively related. Why is it reasonable to take that to mean $2v+u>1$, $2w+u>1$? There is a trend in the value of the asset for most values of the parameters u, v, w. Under what condition will there not be a trend in the asset value? Argue that the long-term value in fact is $EZ(t) \approx t\frac{\alpha(1-u)(v-w)}{(1-\theta)(2-u-v-w)}$ for t large. The parameters α and θ are defined in (12). A more appropriate model is to have the return (on the logarithmic scale) normally distributed with mean dependent on the previous return, and perhaps also on the time since the previous trading. Develop a parametric model for such a process. What values of your parameters would yield high volatility in the value process, without inducing a trend? Discuss this intuitively, and simulate the process if you have time.

Problem 19. Originally, 'martingale' denoted the boisterous strategy of doubling the bet if you lose in a coin-tossing game, and to quit at time T when the first gain is made. Let Y_n be the wealth of the player using the martingale strategy. Assume the coin to be fair. Why is $P(Y_T = Y_0 + 1) = 1$ if the first bet is 1, then 2 if loss, then 4 if loss again, etc? The player will thus have a net gain with certainty – even though the game is fair: Y is a martingale! Why is this paradoxical – and what is the root of the problem? Let now X_n be the step in the random walk $S_n = S_{n-1} + X_n$ where $P(X_n = 1) = p > 0$, $P(X_n = -1) = q = 1-p > 0$. Show that $Y_n = (q/p)^{S_n}$ is a martingale. Show that Y_n also is a martingale when the random walk only lives on 1, 2, ..., N-1, and is absorbed in the boundary states 0 and N. The probability of being absorbed in 0 from an initial state i is found in Taylor and Karlin (1998 page 154) and denoted by u_i. Use the martingale property and the fact that the process will eventually be absorbed with certainty in either 0 or N.

Problem 20. Consider the optimal replacement problem discussed in Karlin and Taylor (1998, page 223-228). To simplify matter, assume that without replacement, the system deteriorates monotonically, and that from state i it either stays, moves to state i+1, or moves to the worst state L, with properties p_i, q_i and r_i respectively, $0 = 1, \cdots, L-2$, $p_0 = 0$. From state L-1, the system can stay or move to L, and L is an absorbing state – still when no replacements are done. Write up the transition probability matrix under the no replacement strategy. What are the conditions on p_i, q_i and r_i to secure the stochastic ordering (2) $P\left(X_{n+1} \geq k | X_n = i\right) \leq P\left(X_{n+1} \geq k | X_n = j\right)$ for all k, $i \leq j$ on page 225? Assume this stochastic ordering, and also monotonicity in the maintenance cost, $a_0 \leq a_1 \leq \cdots \leq a_L$. Show that the optimal strategy must have the structure: replace when $X_n \geq k$. Can you characterize the optimal replacement strategy further?

Problem 21 (Simulating an OU-process). Let the 1-dimensional process be given by (19) for given parameters a and σ, and for given moving target $m(t)$. The goal is to simulate the process at discrete points in time $0 < t_1 < t_2 < \cdots < t_n$. Denote $X(t_i)$ by X_i and show that the conditional distribution of X_{i+1} given X_i is normal with mean and variance

$$\begin{aligned} E[X_{i+1}|X_i] &= e^{a(t_i - t_{i+1})} X_i + \int_{t_i}^{t_{i+1}} m(s)\, a e^{a(s-t_{i+1})} \\ var[X_{i+1}|X_i] &= \frac{\sigma^2}{2a}\left(1 - e^{a(t_i - t_{i+1})}\right). \end{aligned}$$

With $Z_{i+1} \sim N(E[X_{i+1}|X_i], var[X_{i+1}|X_i])$ $i = 0, \cdots,$ $Z_0 \sim X_0$ and independent,

$$X_i = \sum_{j=0}^{i} Z_j,$$

and the task is accomplished.

Problem 22. Consider the matrix (22). Show that $C(t) = e^{-At}$ where A is the drift matrix in the stochastic differential equation model for global warming. Check the differential equation for C.

References

[1] Aiyagari, S. R.1994. Uninsured idiosyncratic risk and aggregate savings. *Quarterly Journal of Economics*. pp 659–684.

[2] Bewley, T. F. 1986. Stationary Monetary Equilibrium with a Continuum of Independently Fluctuating Consudmers. In Hildenbrand, W. and Mas-Colell, A. (eds) *Contributions to Mathematical Economics in Honor of Gerard Debreu*. North Holland.

[3] Borjas, George J., and Bernt Bratsberg, 1996: Who Leaves? The Outmigration of the Foreign-Born. *The Review of Economics and Statistics* , **87**(1): 165-176.

[4] Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. 1997. *The econometrics of frinancial markets*. Princeton University Press

[5] Engle, R.F. and Russel, J.R. 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, **66**: 1127-1162.

[6] Grimmet, G.R. and Stirzaker, D.R. 1992. *Probability and random processes*. Claredon Press, Oxford.

[7] Koop, G. 2003. *Bayesian Econometrics*. Wiley.

[8] Krusell, P. and Smith, T. 1998. Income and Wealth Heterogeneity in the Macroeconomy. *Journal of Political Economy* **106:** 867–896.

[9] Kundu, T. 2007. Can democracy always lead to efficient economic transitions? Seminar at ØI 01.02.07, http://www.oekonomi.uio.no/seminar/torsdag-v07/kundu.pdf

[10] Lancaster, T. *An introduction to modern Bayesian econometrics*. Blackwell.

[11] Moffitt, Terrie E. 1993. Adolescence-limited and life-course persistent antisocial behaviour: A developmental taxonomy. *Psychological review* **100**: 674-701.

[12] Nagin, D.S. 2005. Developmental trajectory groups: Fact or a useful statistical fiction? *Criminology* **43**: 873-904.

[13] Poirier. D.J. 1988. Frequentist and subjectivist perspectives on the problem of model building in economics. *Journal of Economic Perspectives*, **2**: 121-144.

[14] Robert, C.P and Casella, G. 2004. *Monte Carlo statistical methods. Second Edition.* Springer.

[15] Schweder, T. and Hjort, N.L. 1996. Bayesian synthesis or likelihood synthesis – what does the Borel paradox say? *Rep.int.Whal.Commn*, **46**: 475-479.

[16] Seierstad, A. 2009. *Stochastic Control in Discrete and Continuous Time.* **48**: 295-320

[17] Skardhamar, T. 2010. Distinguishing facts and artifacts in group-based modeling. *Criminology.* Springer.

[18] Stigler, S.M. 1986. *The history of statistics; the measurement of uncertainty before 1900.* The Belknap Press of Harvard University Press.

[19] Stokey, N.L and Lucas Jr, R.E. 1989. *Recursive methods in economic dynamics.* Harvard University Press

[20] Storesletten, K., Telmer, C. and Yaron, A. 2004. Consumption and Risk Sharing over the Life Cycle. *Journal of Monetary Economics* **51**: 609–633.

[21] Sydsæter, K. Hammond, P., Seierstad, A and Strøm, A. 2005. *Further Mathematics for Economic Analysis.* Prentice Hall.

[22] Taylor, H.M. and Karlin, S. 1998. *An introduction to stochastic modeling. Third edition.* Academic Press.

[23] Tinbergen, J. 1956. *On the theory of income distribution. Weltwirtschafliches Archiv.* **77**: 155-1175 (Reprinted and corrected in *Jan Tinbergen; Selected Papers*, (eds. Klassen, L.H., L.M. Koyck and H.J. Witteveen) North-Holland 1959: 241-263.)

[24] Weitzman, M. (2008) On Fat-Tailed Climate Change, Uncertain Discounting, and the Backstop Role of Fast Geoengineering (Note, Dept. of Economics, Harvard)

[25] Weitzman, M. (2009) On modeling and interpreting the economics of catastophic climate change *The Review of Economics and Statistics*; Vol XCI, pp 1-19. XCI, pp 1-19.

[26] Øksendal, B. (2005). *Stochastic Differential Equations; An Introduction with Applications.* Sixth Edition. Springer.