Computational Methods in Stochastics

Lecture IV

Markov Chains and Stochastic Models

In what follows, the basis for doing Markov chain Monte Carlo simulations (MCMC) is given.

We will cover the basic formulation of a Markov chain as a dynamic system in continuous time.

The concrete example of how to simulate a stochastic Markov chain process in continuous time will be the *inhomogeneous* Poisson process.

We introduce the notation used by Wilkinson, which is better suited for continuous state spaces and restate everything to hopefully make things more easily understandable.

The concepts of the following Markov chain part are necessary for understanding and justifying e.g. Hamiltonian MC method.

Markov Processes and Stochastic Models (discrete state-space)

Markov Chains

Discrete state space: Online book, Ch 11.2.

Recap: The probability of the process initially being in state j is p_j, i.e. the distribution law of X_0 is $\Pr\{X_0 = j\} = p_j$, then the probability of the process being in state k at time n is

$$p_k^{(n)} = \sum_{j=0}^{\infty} p_j P_{jk}^{(n)} = \Pr\{X_n = k\}.$$

For a process with r states,

$p^{(n)} = (p_1^{(n)}, p_2^{(n)}, \dots, p_r^{(n)})$ is an r-dimensional row vector.

The above provides the procedure for a Monte Carlo simulation of the Markov process.

Markov Chains

The time dependence of the Markov process can be mapped from successive steps:

$$p_k^{(n+1)} = \sum_{j=0}^{\infty} p_j^{(n)} P_{jk}.$$

Transition, of course, can be decomposed by any number of steps: $\mathbf{P}^{m+n} = \mathbf{P}^m \mathbf{P}^n$. The process can accordingly be propagated by any number of time steps starting from the states at any time.

MC simulation proceeds as:

$$p^{(1)} = p^{(0)}\mathbf{P} \rightarrow p^{(2)} = p^{(1)}\mathbf{P} \rightarrow \ldots \rightarrow p^{(n)} = p^{(n-1)}\mathbf{P}.$$

Q: Where do we stop, in other words, how large should *n* be?

A: Until the process is stationary, if you are interested in dynamics. Beyond that, if you want to simulate the distribution (in the stationary state).

Markov Chains

Stationary distributions

A distribution is said to be a *a stationary distribution* of the homogeneous Markov chain governed by the matrix **P** if

$$p = p\mathbf{P}.$$

So, p is a row eigenvector of the transition matrix with eigenvalue equal to 1.

If at some time n we have $p^{(n)} = p$, then $p^{(n+1)} = p^{(n)}\mathbf{P} = p\mathbf{P} = p$.

Similarly, $p^{(n+k)} = p, \forall k \geq 0$.

This is saying that, **if a chain has a stationary distribution, it retains that distribution for all future time**.

Markov Chains

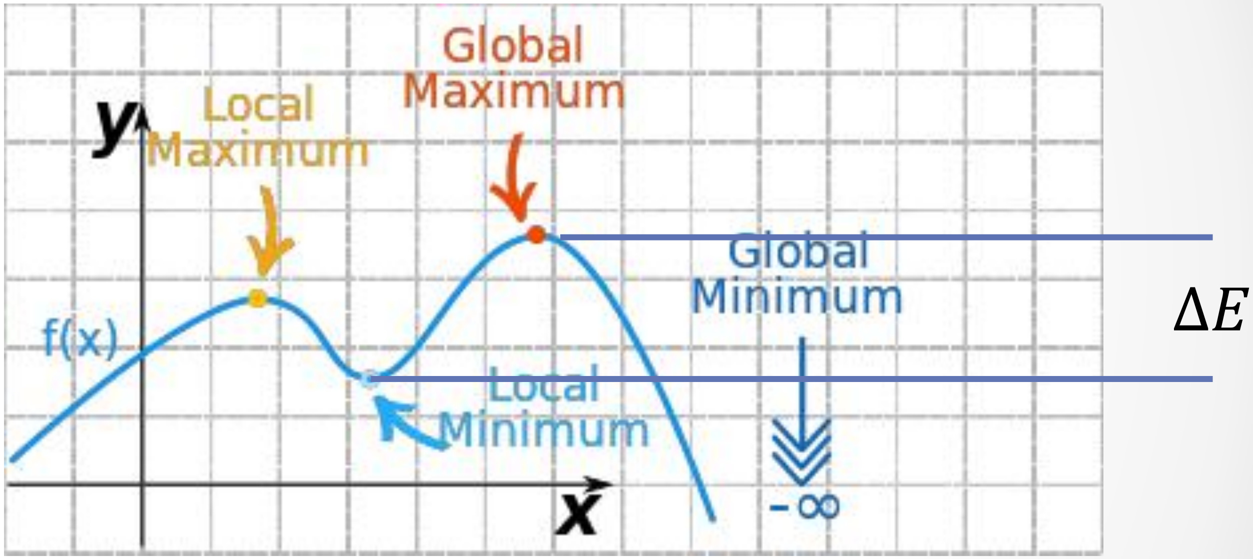
Convergence

Convergence to a stationary distribution is a rather technical subject. We won't go into that. Suffice it to say that the hope is to find *equilibrium distribution,* i.e. a stationary distribution to which p converges. Convincing oneself that a found stationary distribution is also the equilibrium distribution can be hard, to say the least.

Markov Chains

Finding equilibrium by MC means to a physicist or chemist finding *global energy minimum* and to a statistician finding *maximum likelihood*.

If $y = f(x)$ is proportional to the energy at x, then the likelihood for corresponding stationary distributions has maxima at the energy minima.



ΔE

Detailed balance

In equilibrium the detailed balance holds:

$$p_i P_{ij} = p_j P_{ji}$$

In physics etc.:

$P_{ij} \propto e^{-\Delta E/k_B T}$. Boltzmann weight – everyone should know this.

Markov Chains and MC

Notation in Wilkinson:

The set $\{\theta^{(t)}|t = 0, 1, 2, \dots\}$ is a discrete time stochastic process. The state space is such that $\theta^{(t)} \in S, \forall t$, and may be discrete or continuous.

A **Markov chain**: For $A \subseteq S, t = 0, 1, 2, \dots,$

$$\mathrm{P}(\theta^{(t+1)} \in A|\left(\theta^{(t)} = x, \theta^{(t-1)} = x_{t-1}, \dots, \theta^{(0)} = x_0\right)$$
$$= \mathrm{P}\left(\theta^{(t+1)} \in A\middle|\theta^{(t)} = x\right), \forall x, x_{t-1}, \dots, x_0 \in S$$

In the case of no *t* dependence, $\mathrm{P}\left(\theta^{(t+1)} \in A\middle|\theta^{(t)} = x\right) = \mathrm{P}(x, A), \forall t$ and the Markov chain is homogeneous; the *transition kernel* $\mathrm{P}(x, A)$ (previously one-step transition probability $P_{ij}^{n,n+1} = P_{ij}$) determines the behaviour of the chain. $\mathrm{P}(x,\cdot)$ is a *probability measure* over S. (So, P(·) = Pr{·}.)

Markov Chains and MC

Writing $P(x, y) = P(\theta^{(t+1)} = y | \theta^{(t)} = x)$, the **stochastic transfer matrix** becomes

$$P = \begin{pmatrix} \mathrm{P}(x_1, x_1) & \cdots & \mathrm{P}(x_1, x_r) \\ \vdots & \ddots & \vdots \\ \mathrm{P}(x_r, x_1) & \cdots & \mathrm{P}(x_r, x_r) \end{pmatrix}$$

The ith row is the probability distribution of the values of $\theta^{(t+1)}$ under the condition that $\theta^{(t)} = x_i$ (note that here the index i starts from 1 whereas it started from 0 in the previous strictly discrete notation).

We denote:

$$\begin{gathered} \mathrm{P}(\theta^{(t)} = x_1) = \pi^{(t)}(x_1) \\ \mathrm{P}(\theta^{(t)} = x_2) = \pi^{(t)}(x_2) \\ \vdots \\ \mathrm{P}(\theta^{(t)} = x_r) = \pi^{(t)}(x_r) \end{gathered}$$

Markov Chains and MC

→ As an r-dimensional row vector

$$\pi^{(t)} = \left(\pi^{(t)}(x_1), \pi^{(t)}(x_2), \dots, \pi^{(t)}(x_r)\right).$$

The probability density at time $t + 1$ can be computed as (a total probability)

$$\mathrm{P}\left(\theta^{(t+1)} = x_1\right) = \mathrm{P}(x_1, x_1)\, \pi^{(t)}(x_1) + \mathrm{P}(x_2, x_1)\, \pi^{(t)}(x_2) + \cdots$$
$$+\mathrm{P}(x_r, x_1)\, \pi^{(t)}(x_r)$$

Similarly for $\mathrm{P}\left(\theta^{(t+1)} = x_2\right)$ etc. →

$$\left(\pi^{(t+1)}(x_1), \pi^{(t+1)}(x_2), \dots, \pi^{(t+1)}(x_r)\right) =$$

$$=\left(\pi^{(t)}(x_1), \pi^{(t)}(x_2), \dots, \pi^{(t)}(x_r)\right) \times \begin{pmatrix} \mathrm{P}(x_1, x_1) & \cdots & \mathrm{P}(x_1, x_r) \\ \vdots & \ddots & \vdots \\ \mathrm{P}(x_r, x_1) & \cdots & \mathrm{P}(x_r, x_r) \end{pmatrix}$$

$$\Leftrightarrow \pi^{(t+1)} = \pi^{(t)} P$$

Markov Chains and MC

The way an MC simulation would proceed:

$$\pi^{(1)} = \pi^{(0)}P$$

$$\pi^{(2)} = \pi^{(1)}P = \pi^{(0)}PP = \pi^{(0)}P^2$$

$$\vdots$$

$$\pi^{(t)} = \pi^{(0)}P^t$$

So, if the one-step transition matrix is P, then the m-step transition matrix is P^m.

Also, the $(m + n)$-step transition matrix is $P^m P^n = P^{m+n}$.

The set of linear equations corresponding to the above statement are called the *Chapman-Kolmogorov equations.*

Markov Chains and MC

Derivation of Chapman-Kolmogorov equations

Define $P(n)$ to be the n-step transition matrix with elements $p_{ij}(n)$.

Due to the theorem of total probability,

$$p_{ij}(m+n) = \mathrm{P}\left(\theta^{(m+n)} = j | \theta^{(0)} = i\right)$$

$$= \sum_{k=1}^{r} \mathrm{P}\left(\theta^{(m+n)} = j | \theta^{(m)} = k\right) \mathrm{P}\left(\theta^{(m)} = k | \theta^{(0)} = i\right)$$

Chapman-Kolmogorov equations

$$\Leftrightarrow p_{ij}\ (m+n) = \sum_{k=1}^{r} p_{kj}(n) p_{ik}(m).$$

Markov Chains and MC

Stationary distributions

… in this new notation: $\pi = \pi P$ is a row eigenvector of P with corresponding eigenvalue $\lambda = 1$ (and a fixed point of the linear map induced by P).

$\pi^{(n+1)} = \pi^{(n)} P = \pi P = \pi$, or $\pi^{(n+k)} = \pi, \forall k$.

Since, $\pi = \pi P \Leftrightarrow \pi - \pi P = 0$,

the stationary distribution can be solved from $\pi(I - P) = 0$.

Markov Chains and MC

Some Important Properties and Special Cases

Reversibility and Detailed Balance

If $\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(N)}$ is a Markov chain, then so is $\theta^{(N)}, \theta^{(N-1)}, \dots, \theta^{(0)}$. (For proof, see Wilkinson, 5.2.5.)

The transition kernel for the reversed chain:

$$\mathrm{P}_t^*(x, y) = \mathrm{P}\big(\theta^{(t)} = y | \theta^{(t+1)} = x\big)$$

$$= \frac{\mathrm{P}\big(\theta^{(t+1)} = x | \theta^{(t)} = y\big)\mathrm{P}\big(\theta^{(t)} = y\big)}{\mathrm{P}(\theta^{(t+1)} = x)} \quad \text{(Bayes Theorem)}$$

$$= \frac{\mathrm{P}(y, x)\pi^{(t)}(y)}{\pi^{(t+1)}(x)}.$$

→ **The reversed chain is**, in general, **not homogeneous**, that is, transition probabilities depend on time.

Markov Chains and MC

However, in equilibrium $\pi^{(t+1)}(x) = \pi^{(t)}(x) = \pi$, so

$$\mathrm{P}^*(x, y) = \frac{\mathrm{P}(y, x)\pi(y)}{\pi(x)}.$$

→ **In equilibrium the reversed chain is homogeneous.**

We can write P^* as

$$P^* = \mathrm{diag}\{\pi\}^{-1} P^{\mathrm{T}} \mathrm{diag}\{\pi\}$$

If $P^*(x, y) = P(x, y),\ \forall x, y$, then the chain is said to be (time) *reversible*. For reversible chains *detailed balance* holds:

$$\pi(x)\mathrm{P}(x, y) = \pi(y)\mathrm{P}(y, x), \qquad \forall x, y.$$

Detailed balance states the symmetry in the flow of probability between two states: Probability of a transition from x to y = probability of a transition from y to x $\forall (x, y)$.

Markov Chains and MC

Stochastic simulation and analysis

Simulation of a Markov chain = simulation of a new state randomly, when probabilities are given by a probability vector p.

An MC simulation of a Markov chain proceeds as:

1. A transition matrix is given or deduced for the system.
2. The initial distribution $\pi^{(0)} = \left(\pi^{(0)}(x_1), \pi^{(0)}(x_2), \dots, \pi^{(0)}(x_r)\right)$ is given.
3. Sample an initial state $\theta^{(0)}$ from $\pi^{(0)}$ using a lookup method.
4. Sample the state for $\theta^{(t)}$ using the set of probabilities from the $\theta^{(t-1)}$the row of P for $t = 1, 2, \dots$ (P$\left(\theta^{(t)} = x_i\right) = \pi^{(t)}(x_i),\ i \in [1, r]$, and $\pi^{(t)} = \pi^{(t-1)}P$).

Markov Chains and MC

A layman's statement on simulations of a Markov process:

When the distribution changes in time, which means the system is out of equilibrium and the states are non-stationary, one can simulate starting from an initial distribution $\pi^{(0)}$ simply using $\pi^{(t)} = \pi^{(t-1)}P$ to explore how (and if) the system approaches equilibrium. In some cases there may be several equilibrium states and where one ends up depends on $\pi^{(0)}$. This is the "dynamical aspect".

When we are in the stationary (that may be an equilibrium) state, we are interested in sampling different realisations of the distribution in that state. Then we apply the method outlined on the previous slide. In order to reach the stationary state, we may first do the above procedure sufficiently long so that $\pi^{(t)} \to \pi$. Then we can use $\pi^{(0)} = \pi$ to start sampling $\theta^{(t)}$. (Continues ⟶)

Markov Chains and MC

Sampling $\theta^{(t)}$ for a Markov process means doing Monte Carlo using the probabilities in the $\theta^{(t-1)}$th row of P. (This means the row corresponding the value $\theta^{(t-1)} = x$; the x^{th} row includes transition probabilities from x to all states $(\mathrm{P}(x, y) = \mathrm{P}\left(\theta^{(t)} = y | \theta^{(t-1)} = x\right))$.

Markov Processes and Stochastic Models (continuous state-space)

Markov Chains and MC

Markov chains with continuous state-spaces

When, for example, sampling from continuous distributions using a Markov chain (MCMC), one is working with a continuous state-space.

A commonly used model for such simulations in time-series analysis is the *first-order auto-regressive model*, AR(1).

The model for an AR(1) process $\{Z_t | t = 1, 2, \ldots\}$can be summarised as

$$Z_t = \alpha Z_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma^2).$$

This means: The value of the stochastic process at time t, Z_t, depends on the value of the the stochastic process at time $t - 1$.

Markov Chains and MC

The non-deterministic part of the AR(1) process is given by the noise term ε_t. The noise process $\{\epsilon_t | t = 1, 2, \dots\}$ is assumed to be independent: the pair of random quantities ε_i and ε_j are independent of one another $\forall i \neq j$.

The conditional distribution of Z_t given $Z_{t-1} = z_{t-1}$:

$$Z_t | (Z_{t-1} = z_{t-1}) \sim N(\alpha z_{t-1}, \sigma^2)$$

AR(1) is Markov chain and its state-space $S = \mathbb{R}$.

Markov Chains and MC

Transition kernels

For a *homogeneous chain*, define $\mathrm{P}(x, A) = \mathrm{P}\left(\theta^{(t+1)} \in A | \theta^{(t)} = x\right)$.

$\mathrm{P}(x, \{y\}) = 0$ for continuous state-spaces → define

$$\begin{aligned} \mathrm{P}(x, y) &= \mathrm{P}\left(\theta^{(t+1)} \leq y | \theta^{(t)} = x\right) \\ &= \mathrm{P}\left(\theta^{(1)} \leq y | \theta^{(0)} = x\right), \ \forall x, y \in S. \end{aligned}$$

This is a **conditional cumulative distribution function** (CDF) and distributional form of the *transition kernel*.

The corresponding **conditional density** defines the density form of the transition kernel (useful in cases of multiple dimensions, e.g. $S \subseteq \mathbb{R}^n$):

$$p(x, y) = \frac{\partial}{\partial y} \mathrm{P}(x, y) = p(y|x), \qquad x, y \in S.$$

Markov Chains and MC

Example

AR(1): $\theta^{(t+1)} = \alpha\theta^{(t)} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2)$

$$\Rightarrow \left(\theta^{(t+1)} | \theta^{(t)} = x\right) \sim N(\alpha x, \sigma^2)$$

Transition kernel

$$\Rightarrow p(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y - \alpha x}{\sigma}\right)^2\right\}$$

(density for y for a given fixed value of x)

Markov Chains and MC

Stationarity and reversibility

The state $\theta^{(t)}$ is a realisation from the probability density function $\pi^{(t)}(x),\ x \in S$. Total probability:

$$\pi^{(t+1)}(y) = \int_S p(x,y)\pi^{(t)}(x)dx$$

For the stationary distribution,

$$\pi(y) = \int_S p(x,y)\pi(x)dx$$

The transition density for the reversed chain (from Bayes)

$$p_t^*(x,y) = \frac{p(y,x)\pi^{(t)}(y)}{\pi^{(t+1)}(x)}$$

Markov Chains and MC

… homogenises in the stationary limit

$$p^*(x,y) = \frac{p(y,x)\pi(y)}{\pi(x)}.$$

→ Again, for a (time) reversible chain, (the continuous form) of the detailed balance equations hold:

$$\pi(x)p(x,y) = \pi(y)p(y,x), \qquad \forall x,y \in S.$$

Integrating both sides of the detailed balance condition gives $\pi(y) = \int_S p(x,y)\pi(x)dx$ → **Any homogeneous Markov chain satisfying detailed balance is reversible with stationary distribution $\boldsymbol{\pi}(\cdot)$.**

Markov Chains and MC

Stochastic simulation and analysis

1. Sample $\theta^{(0)}$ from $\pi^{(0)}(\cdot)$ using an appropriate method.

2. Simulate (the realisation) $\theta^{(t+1)}$ from $p\left(\theta^{(t)},\cdot\right)$ for $t = 1, 2, \ldots$ using one of the standard techniques. (So, accept $\theta^{(t+1)}$ according to $\mathrm{P}(x, y) = P\left(\theta^{(t+1)} \leq y | \theta^{(t)} = x\right)$.)

Computing $\mathrm{P}(x, y)$ is not trivial. For the normal distribution, Python has in the library scipy.stats norm.cdf(x) that computes the value of CDF for the value x for the *standard normal distribution* $N(0, 1)$. CDF for $N(\mu, \sigma^2)$ can then be computed using $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$, where $F(\cdot)$ is CDF of $N(\mu, \sigma^2)$ and $\Phi(\cdot)$ is CDF of $N(0, 1)$.

Markov Chains in Continuous Time and Discrete Space

****

The last variation of Markov chains. And we get to the beef (simulation) at the end

Markov Chain Gang of the 30's:

Markov Chains and MC

Markov Chains in Continuous Time Online book, Ch 11.3

Let’s state how the Markov chains are constructed in continuous time.

Markov process in continuous time:

$$\mathrm{P}(X(t+dt)=x|\{X(t)=x(t)|t\in[0,t]\})=$$
$$=\mathrm{P}(X(t+dt)=x|X(t)=x(t)),\ \forall t\in[0,\infty),x\in S.$$

Finite state-space

The process can take on one of r states, $S=\{1,2,\dots,r\}$.

Markov Chains and MC

If at time t the process is in state $x \in S$, its future behaviour can be characterised by *the transition kernel*

$$p(x,t,x',t') \equiv P(X(t+t') = x' | X(t) = x).$$

If this function does not depend explicitly on t, the process is said to be *homogeneous* and the kernel can be written $p(x,x',t')$. For each value of t' this kernel can be expressed as an $r \times r$ transition matrix, $P(t')$. ($P(0) = I$.)

Just as in the discrete time case

$$P(t+t') = P(t)P(t') = P(t')P(t)$$

Chapman-Kolmogorov equations:

$$p(i,j,t+t') = \sum_{k=1}^{r} p(i,k,t)p(k,j,t'), \qquad i,j = 1,\dots,r.$$

Markov Chains and MC

Define the *transition matrix* Q as a derivative of $P(t')$ at t'

$$Q = \frac{d}{dt'} P(t')|_{t'=0} = \lim_{\delta t \to 0} \frac{P(\delta t) - P(0)}{\delta t} = \lim_{\delta t \to 0} \frac{P(\delta t) - I}{\delta t}.$$

The elements of Q give the **"hazards"** of moving to different states.

→ The infinitesimal transition matrix

$$P(dt) = I + Qdt. \ \Rightarrow P(t + dt) = P(dt)P(t) = [I + Qdt]P(t).$$

→ The stationary distribution π:

$$\pi P(dt) = \pi$$

$$\pi(I + Qdt) = \pi$$

$$\pi Q = 0.$$

Markov Chains and MC

$$\frac{d}{dt}P(t) = \frac{P(t+dt) - P(t)}{dt} = \frac{P(dt)P(t) - P(t)}{dt}$$
$$= \frac{P(dt) - I}{dt}P(t) = QP(t).$$

$\Rightarrow \frac{d}{dt}P(t) = QP(t)$; the initial condition $P(0) = I$.

$\Rightarrow P(t) = \exp\{Qt\}$. (exp$\{\cdot\}$ is the matrix exponential function , see e.g. Golub & van Loan: Matrix Computations .)

We can avoid working with differential equations in matrix form by writing in the component form, *Kolmogorov's backward equations*:

$$\frac{d}{dt}p(i,j,t) = \sum_{k=1}^{r} q_{ik}p(k,j,t), \qquad i,j = 1,2,\dots,r.$$

Markov Chains and MC

Had we expanded $P(t+dt)$ as $P(t)P(dt)$ rather than $P(dt)P(t)$, we would have ended up with

$$\frac{d}{dt}P(t) = \mathrm{P(t)Q}$$

$\Longrightarrow$ *Kolmogorov's forward equations:*

$$\frac{d}{dt}p(i,j,t) = \sum_{k=1}^{r} q_{kj}p(i,k,t), \qquad i,j = 1,2,\dots,r.$$

Markov Chains and MC

Example

A simple model for the activation of a single prokaryotic gene. The gene will be activated unless a repressor protein is bound to its regulatory region. Two states: state 0 (inactive) and state 1 (active). (→ The state vector is (0 1).)

In the inactive state there's a a constant hazard of $\alpha > 0$ for activation.
In the active state there's a a constant hazard of $\beta > 0$ for inactivation.

The rows of Q must sum to zero → $Q = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$

Solving $\pi Q = 0$ gives the stationary distribution

$\pi = \left(\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right)$. ($\pi$ is a PDF, so obviously $\sum_i \pi_i = 1$.)

Markov Chains and MC

The infinitesimal transition matrix

$$P(dt) = I + Qdt = \begin{pmatrix} 1 - \alpha dt & \alpha dt \\ \beta dt & 1 - \beta dt \end{pmatrix}.$$

This process is also known as the *telegraph process* that can form a basis for a simple model of ion channels.

****

Looks simple enough, right? In fact, you have to do some tedious calculations to get π. If you don't impose any constraints, any π will satisfy $\pi Q = 0$. (So, any π would appear to be a stationary distribution, which, of course, cannot be true.) The book Stochastic Modelling for Systems Biology (Chapter 5) is on this point misleading.

****

Markov Chains and MC

The following drill is just to show you how tedious even simple exact calculations for MC models are.
If you don't like it, put it in your pipe and smoke it.

To get the stationary distribution you need to also use Kolmogorov's (backward) eqs.

$\frac{d}{dt}P(t) = QP(t)$. Solving from this, e.g. , notation $p(i,j,t) = P_{ij}(t)$.

$$\frac{d}{dt}P_{00}(t) = -\alpha P_{00}(t) + \beta P_{01}(t) \text{ and } P_{01}(t) = 1 - P_{00}(t)$$

$$\Rightarrow \frac{d}{dt}P_{00}(t) = \beta - (\alpha + \beta)P_{00}(t).$$

Then let $Q_{00}(t) = e^{(\alpha+\beta)t}P_{00}(t)$…

Markov Chains and MC

Finally, $Q_{00}(t) = \left(\frac{\beta}{\alpha+\beta}\right) e^{(\alpha+\beta)t} + \left(\frac{\alpha}{\alpha+\beta}\right)$, and

$P_{00}(t) = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}\, e^{-(\alpha+\beta)t}.$

$P_{01}(t) = 1 - P_{00}(t) = \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}\, e^{-(\alpha+\beta)t}.$

By symmetry,

$$P_{11}(t) = \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}\, e^{-(\alpha+\beta)t},$$

$$P_{10}(t) = \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta}\, e^{-(\alpha+\beta)t}.$$

For details, see ISM, Chapter 6.3.3. – if interested.

From the previous eqs, the stationary state:

$$\lim_{t\to\infty} P_{11}(t) = \pi_1 = \frac{\alpha}{\alpha+\beta}$$

$$\lim_{t\to\infty} P_{00}(t) = \pi_0 = \frac{\beta}{\alpha+\beta}$$

$$\Rightarrow \pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$$

Conclusion: it makes sense to simulate & compute the expectation values numerically.

Markov Chains and MC

Stochastic Simulation

Given the infinitesimal transition matrix $P(dt) = I + Qdt$, we have for small time steps Δt

$$P(\Delta t) \approx I + Q\Delta t.$$

$P(\Delta t)$ is regarded as the transition matrix of a discrete time Markov chain → simulated sequence at times $0, \Delta t, 2\Delta t, 3\Delta t, \ldots$

The precision can be improved by using the exact value

$$P(\Delta t) = \exp\{Q\Delta t\}. \qquad \text{(We started from } \frac{d}{dt}P(t) = QP(t); P(0) = I.\text{)}$$

Δt must be chosen sufficiently small for not to miss interesting transitions.

Markov Chains and MC

This is derivation/justification of the simulation algorithm for discrete events on page 16.

Discrete event simulation

The process is currently in state x → the xth row of Q gives the hazards for the transitions to other states. The row sums to zero → $-q_{xx}$ gives the combined hazard for moving away from the current state (q_{xx} is non-positive) → the time to a transition event is exponential with rate $-q_{xx}$.

When this transition event occurs, the new state will be random with probabilities proportional to the xth row of Q

Markov Chains and MC

The probability that the next event will be in the time interval $(t + t', t + t' + dt]$ and will consist of a move to state j. Let this probability divided by dt be $f(t', j|t, i)$. The probability $f(t', j|t, i)dt$:

$$\begin{aligned} f(t', j|t, i)dt &= \mathrm{P}(\text{Next event in } (t + t', t + t' + dt]|t, i) \\ &\times \mathrm{P}(j|\text{Next event in } (t + t', t + t' + dt], t, i) \end{aligned}$$

The first term: the hazards for the individual transitions are given by the off-diagonal elements of the ith row of Q; the combined hazard is $-q_{ii}$. (The probability of two events occurring in the interval $(t, t + dt]$ is $O(dt^2)$ and can be neglected, $dt \ll 1$.)

Markov Chains and MC

The time to the next event is $Exp(-q_{ii})$ → the first term is $-q_{ii}e^{q_{ii}t'}dt$.

The second term is

$$P(X(t+t'+dt)=j|[X(t+t')=i]\cap[X(t+t'+dt)\neq i])$$

$$=\frac{P(X(t+t'+dt)=j|X(t+t')=i)}{P(X(t+t'+dt)\neq i|X(t+t')=i)}=\frac{q_{ij}dt}{\sum_{k\neq i}q_{ik}dt}=\frac{q_{ij}}{-q_{ii}}.$$

The first term times the second term:

$$f(t',j|t,i)=-q_{ii}e^{q_{ii}t'}\times\frac{q_{ij}}{-q_{ii}}$$

Markov Chains and MC

$f(t', j|t, i)$ factorises into the form of a probability density for the time to the next event and a probability mass function for the type of that event. → The next event can be simulated with the generation of two random variables. In addition the two random variates can be simulated independently. →

The standard discrete event simulation:

1. Initialise the process at $t = 0$ with initial state i.
2. Call the current state i. Simulate the time to the next event, t', as an $Exp(-q_{ii})$ random quantity.
3. Put $t \coloneqq t + t'$.
4. Simulate new state j as a discrete random quantity with PMF $-q_{ij}/q_{ii},\ j \neq i$.
5. Output the time t and state j.
6. If $t < T_{max}$, return to step 2.

To get a distribution of of values at different times many realisations of the process must be simulated.

Markov Chains and MC

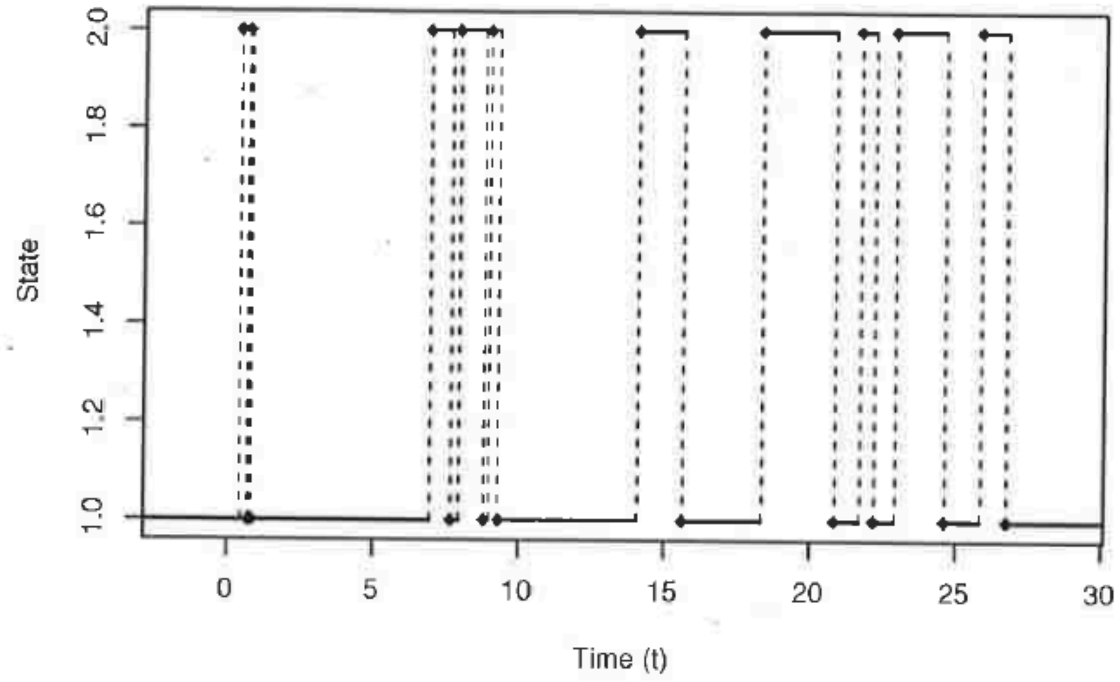


Figure 5.4 *A simulated realisation of the simple gene activation process with* $\alpha = 0.5$ *and* $\beta = 1$.

(In the figure 1 corresponds to state 0 and 2 corresponds to state 1.)

Markov Chains and MC

Countable state-space; immigration-death process

The model: Individuals arrive into the population with constant hazard λ. Each individual dies independently with constant hazard μ. There is no reproduction.

The key transition equations:

$$P(X(t+dt) = x+1 | X(t) = x) = \lambda dt$$

$$P(X(t+dt) = x-1 | X(t) = x) = x\mu dt$$

$$P(X(t+dt) = x | X(t) = x) = 1 - (\lambda + x\mu)dt$$

$$P(X(t+dt) = y | X(t) = x) = 0, \qquad \forall y \notin \{x-1, x, x+1\}.$$

These equations define a homogeneous Markov process with infinite state-space $S = 0, 1, 2, \ldots$

Markov Chains and MC

It is straightforward to write down the infinite-dimensional Q matrix (tridiagonal form):

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ \mu & -\lambda-\mu & \lambda & 0 & 0 & \dots \\ 0 & 2\mu & -\lambda-2\mu & \lambda & 0 & \dots \\ 0 & 0 & 3\mu & -\lambda-3\mu & \lambda & \dots \\ \vdots & \vdots & \vdots & & & \end{bmatrix}$$

The (infinite dimensional) stationary distribution $\pi = (\pi_0, \pi_1, \pi_2, \dots)$:

$$\pi Q = 0 \Rightarrow \pi_k = \frac{\lambda^k}{k!\,\mu^k}\pi_0, \qquad k = 1, 2, \dots$$

Note: The reason for writing the discrete event simulation as a group of difference equations instead of a single matrix is obvious, although mentally one formulates the model with matrices.

Markov Chains and MC

The terms are of the expansion $\pi_0 e^{\lambda/\mu}$. Imposing the unit-sum constraint, $\pi_0 = e^{-\lambda/\mu}$, so

$$\pi_k = \frac{(\lambda/\mu)^k}{k!} e^{-\lambda/\mu}, \qquad k = 1, 2, \ldots$$

So, the stationary distribution of this process is a Poisson random quantity with mean λ/μ.

The process can be simulated by discrete event simulation.

Markov Chains and MC

Inhomogeneous Poisson process

In the homogeneous (stationary) Poisson process the event hazard λ is constant.

Previously, N_t was defined as the number of events of the Poisson process in the interval $(0, t]$; $N_t \sim Po(\lambda t)$. The process N_t is the *counting process*.

The counting process is a Markov process governed by the *homogeneous Markovian transition equations*

$$P(N_{t+dt} = x + 1 | N_t = x) = \lambda dt$$

$$P(N_{t+dt} = x | N_t = x) = 1 - \lambda dt$$

$$P(N_{t+dt} = y | N_t = x) = 0, \forall y \notin \{x, x + 1\}.$$

Markov Chains and MC

Poisson process can be generalised such that the event hazard is not a constant but a function $\lambda(t)$. → The probability of an event in the interval $(t, t + dt]$ is $\lambda(t)dt$.

The *inhomogeneous Markovian transition equations* for the associated counting process N_t are

$$P(N_{t+dt} = x + 1 | N_t = x) = \lambda(t)dt$$

$$P(N_{t+dt} = x | N_t = x) = 1 - \lambda(t)dt$$

$$P(N_{t+dt} = y | N_t = x) = 0, \forall y \notin \{x, x + 1\}.$$

Markov Chains and MC

As for a sufficiently small interval $(t, t+dt]$ $\lambda(t)$ is constant, the number of events in that interval is approximately Poisson $\sim Po(\lambda(t)dt)$, independent of all other intervals.

→ The number of events in the interval $(0,t], N_t$ is the sum (integral) over all intervals.

The sum of independent Poissons is Poisson, so

$$N_t \sim Po\left(\int_0^t \lambda(s)ds\right) = Po(\Lambda(t)),$$

where $\Lambda(t) = \int_0^t \lambda(s)ds$ is the *cumulative hazard.*

The number of events in the interval $(s,t], 0 < s < t$ is $Po\big(\Lambda(t) - \Lambda(s)\big)$.

Markov Chains and MC

Time-change of a Poisson process

To illustrate, take a unit Poisson process (whose rate is 1) having a counting process $M(t)$. Then define a new process that speeds up time by a factor of λ: $N(t) = M(\lambda t)$.

Obviously, $N(t)$ is - and so the time-change $t' = \lambda t$ gives - the counting process of a homogeneous Poisson process of rate λ. This generalises to non-linear time-changes.

Time-changes must be monotonic; to ensure this we write

$$t' = \int_0^t \lambda(s) ds, \quad \text{where } \lambda(t) > 0, \forall t > 0.$$

Markov Chains and MC

Again starting with a unit process $M(t)$, define

$$N(t) = M(t') = M\left(\int_0^t \lambda(s)ds\right) \sim Po\left(\int_0^t \lambda(s)ds\right).$$

$N(t)$ is an *inhomogeneous Poisson process* with rate $\lambda(t)$.

Slogan: A unit Poisson process can be transformed to an arbitrary Poisson process by distorting the time appropriately.

Markov Chains and MC

Proposition. For the inhomogeneous Poisson process with rate function $\lambda(t)$ the time T to the *first* event has distribution function

$F(t) = 1 - \exp\{-\Lambda(t)\}$, and hence the density function

$$f(t) = \lambda(t)\exp\{-\Lambda(t)\}, \; t > 0.$$

Proof.

$$F(t) = P(T \le t) = 1 - P(T > t) = 1 - P(N_t = 0)$$
$$= 1 - \frac{\Lambda(t)^0 \exp\{-\Lambda(t)\}}{0!} = 1 - \exp\{-\Lambda(t)\}$$

Markov Chains and MC

In *stochastic simulation* the motivation often is to simulate the first/next event of an inhomogeneous process. One way to do this is to use the **transformation method**. We can simulate $u \sim U(0,1)$ and solve $F(t) = u$. Rearranging gives $\Lambda(t) = -\log(1-u)$. Since u and $1-u$ have the same distribution, we can solve $\Lambda(t) = -\log u$ for t.

If $\Lambda(t)$ is analytically invertible, then $t = \Lambda^{-1}(-\log u)$.

Since $\Lambda(t)$ is by construction monotonically increasing, $\Lambda(t) = -\log u$ will have at most one solution for $u \in (0,1)$. If $\Lambda(t)$ does not tend to infinity as $t \to \infty$, there may not be a solution; in other words, the first event may not happen at all. To avoid this, in practise the event hazard $\lambda(t)$ is not allowed to decay faster than $1/t$.

Markov Chains and MC

Example

Consider the inhomogeneous Poisson process with rate function $\lambda(t) = \lambda t$ for some constant $\lambda > 0$.

The cumulative hazard $\Lambda(t) = \lambda t^2/2$.

$\Rightarrow N_t \sim Po(\lambda t^2/2) \Rightarrow$ the number of events in the interval $(s, t]$ is $Po(\lambda(t^2 - s^2)/2)$.

The PDF for the time to the first event

$$f(t) = \lambda t \exp\left\{-\frac{\lambda t^2}{2}\right\}, \qquad t > 0.$$

Solve
$\lambda t^2/2 = -\log u$ for t.
$\downarrow$

To simulate:

1. Sample $u \sim U(0,1)$.
2. Obtain the time to the first event as $t = \sqrt{-\frac{2 \log u}{\lambda}}$. Return to 1.

Markov Chains and MC

Simulation of inhomogeneous Poisson processes

In the previous example, where transformation could be made analytically, simulation is effective: no wasted random numbers.

When $\lambda(t)$ is not this nice we need to do rejection-type simulation. The derivation of this method is typical in stochastics/statistics: Theoretically the justification of the method is not that straightforward, but intuitively – as an algorithm – it seems obvious.

An inhomogeneous Poisson process can be simulated by first simulating from a unit homogeneous Poisson process defined on particular region $\mathbb{R}^2$.

Markov Chains and MC

The **2d Poisson process** is defined on a set $\Omega \subseteq \mathbb{R}^2$ and has the property that for any measurable subset $A \subseteq \Omega$ with area $\mu(A)$ the number of events in A will have a $Po(\mu(A))$ distribution.

→ The number of events in an infinitesimal square $(x, x + dx] \times (y, y + dy]$ will have a $Po(dxdy)$ distribution; the probability of an event in this region will be $dxdy$, independently of all other regions.

The *uniformity property* of the (2d) homogeneous Poisson process: Conditional on the number of points in a region Ω, those points are uniformly distributed over that region.

To simulate a 2d Poisson process on a finite region Ω:

1. Sample the *number of points* from $Po(\mu(\Omega))$.
2. Scatter these points uniformly over Ω.

Markov Chains and MC

If it is possible to simulate a unit Poisson process on a region $\overline{\Omega} \supseteq \Omega$, then the points of the originally sampled Poisson process that lie within Ω form a Poisson process on Ω. → Rejection sampling can be applied on the bounding region $\overline{\Omega}$.

1d inhomogeneous Poisson process out of 2d (unit) Poisson process:

Consider an inhomogeneous Poisson process with bounded, non-negative hazard function $\lambda(t)$ satisfying $0 \leq \lambda(t) \leq U_\lambda$ for some upper-bound U_λ. We wish to simulate from this process on $(0, T]$ for some $T > 0$.
Let Ω be the region of $\mathbb{R}^2$ bounded by $x = 0, x = T, y = 0$, and $y = \lambda(x)$ and let $\overline{\Omega}$ be the rectangle $[0, T]\times[0, U_\lambda]$. → By construction, $\Omega \subseteq \overline{\Omega} \subseteq \mathbb{R}^2$.

Markov Chains and MC

To sample an inhomogeneous 1d Poisson process with $\lambda(t)$:

- sample a Poisson process on $\bar{\Omega}$ by first simulating the number of points $m \sim Po(U_\lambda T)$
- for $i = 1, \ldots, m$ independently sample an x-coordinate from $U(0, T)$ and y-coordinate from $U(0, U_\lambda)$
- keep only the points in Ω, i.e. the points (x, y) satisfying $y \leq \lambda(x)$
- from the above-obtained Poisson process on Ω marginally the x-coordinates give the event times of the 1d inhomogeneous Poisson process

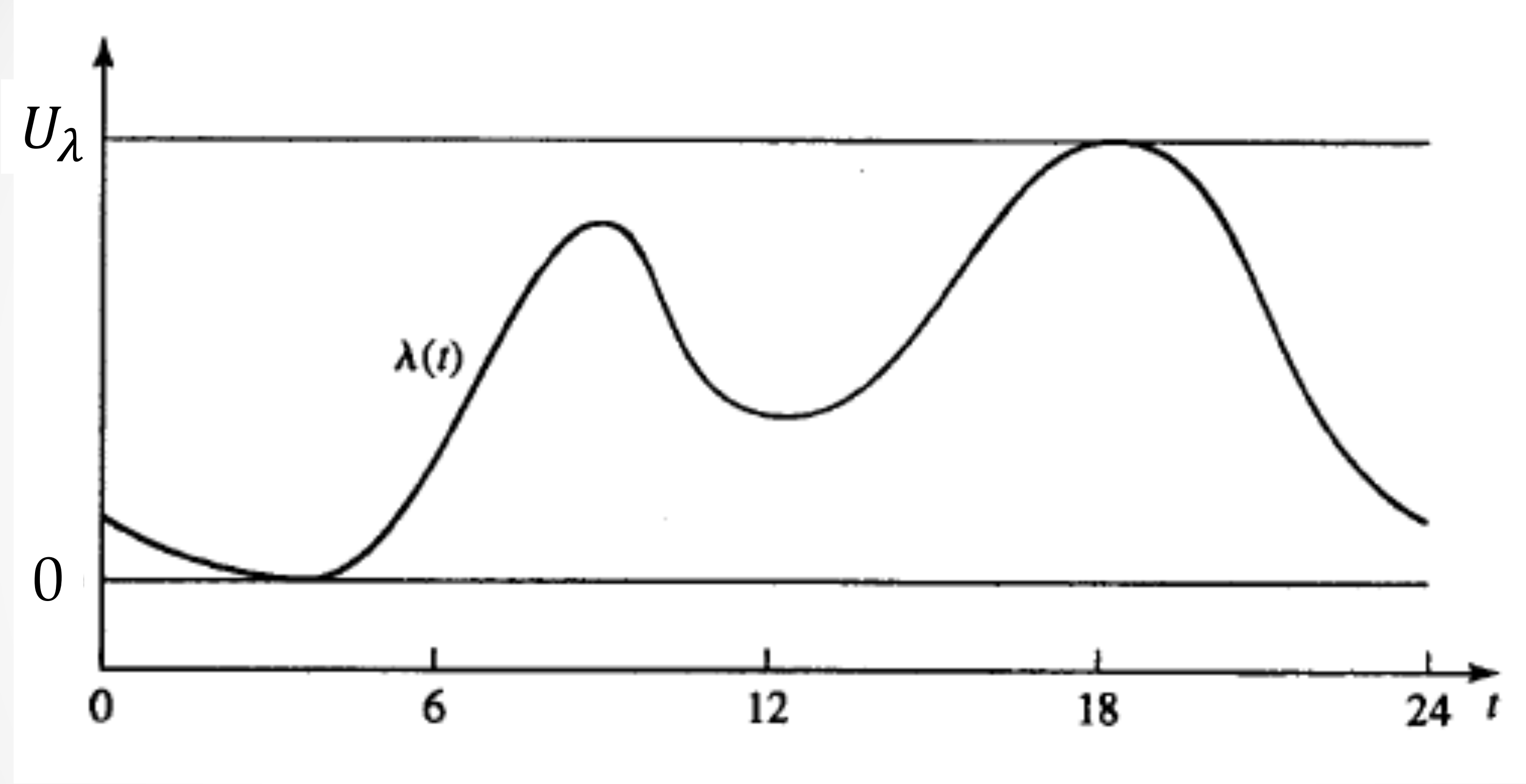
Markov Chains and MC

The algorithm for sampling event times from an inhomogeneous Poisson process:

```
1. Sample $m \sim Po(U_\lambda T)$
2. For $i := 1, \dots, m$
      (a) Sample $x \sim U(0, T)$
      (b) Sample $y \sim U(0, U_\lambda)$
      (c) If $y \le \lambda(x)$, print $x$
```

The printed x represent **a single realisation** of the inhomogeneous Poisson process on the time interval $(0, T]$ **in any order**, i.e. not chronologically.

Markov Chains and MC



Markov Chains and MC

So, the point of drawing m (page 62) from the Poisson distribution is this: If one does not do it in the beginning, the number of simulated events in the interval T is not Poissonian. So, one needs to do it. After the simulation we will have a number of events at different x proportional to $\left(\frac{\lambda(x)}{U_\lambda}\right) * Po(U_\lambda T)$ as we should. In the basic rejection without this part the number of events would not have any relation to Poissonian.

If we only do one simulation - with one number m of events, then it really makes no difference if we just set this number m to whatever we like and do the rejection sampling. So, as said in the lecture, this m drawn from Poissonian only has meaning when performing the simulation several times with different m. Only several simulations with several m from the Poissonian constitutes a full Poissonian, process. However, often just one simulation with m runs is performed and it gives the right time dependence. **This is just one realisation of the Poisson process**, but to get complete statistics, in fact, several batches over different numbers m of runs drawn from $m \sim Po(U_\lambda T)$ should be done and results averaged over them. Only this way we can obtain statistics for the process.

Markov Chains and MC

If we are looking for the time of the first event, the previous algorithm is inefficient: multiple realisations have to be run until the end, find the minimum times and average over them.

One approach **to sample the event times in order** is to use properties of uniform order statistics: Denoting the sorted values $X_{(i)}$ for $X_i \sim U(0, T), i = 1, \dots, m$, so that $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(m)}$,

$$X_{(1)} = \min\{X_i\},$$

which has a CDF $F_{(1)}(x) = (x/T)^m$; simulate it by sampling $u \sim U(0, 1)$ and setting $x_{(1)} = Tu^{1/m}$. → Conditional on $x_{(i-1)}$, $X_{(i)}$ is the minimum of $m - i + 1$ uniforms on $[x_{(i-1)}, T]$. This can be simulated by sampling $u \sim U(0, 1)$ and setting $x_{(i)} = (T - x_{(i-1)})u^{1/(m-i+1)}$.

Markov Chains and MC

The revised algorithm for sampling events in order:

```
1. Set $x_{(0)} := 0$
2. Sample $m \sim Po(U_\lambda T)$
3. For $i := 1, \dots, m$
    (a) Sample $u \sim U(0,1)$
    (b) Set $x_{(i)} := \left(T - x_{(i-1)}\right)u^{1/(m-i+1)}$
    (c) Sample $y \sim U(0, U_\lambda)$
    (d) If $y \leq \lambda\left(x_{(i)}\right)$, print $x_{(i)}$
```

In fact, **only every other $x_{(i)}$ are in increasing order**. Starting with $x_{(0)} := 0$, $x_{(1)}$ will be large, $x_{(2)}$ will be the minimum and $x_{(2)}, x_{(4)}, x_{(6)}, \dots$ will be ordered. (Wilkinson book is very sloppy in this Poisson part.)

To simulate the process correctly, every $x_{(i)}$ has to be used, of course.

Markov Chains and MC

THE PREVIOUS DERIVATION of the revised algorithm from the Wilkinson book IS COMPLETELY WRONG!

In Wilkinson’s derivation there are two errors, which happily rectify each other so that every other sample appears in order and the whole range $(0, T]$ is sampled.

The first mistake: Wilkinson used CDF for maximum values instead of minimum values.

The second mistake: Wilkinson did not take into account the fact that for each recursive interval $(x_{(i-1)}, x_{(i)}]$, the starting point $x_{(i-1)}$ needs to be taken into account.

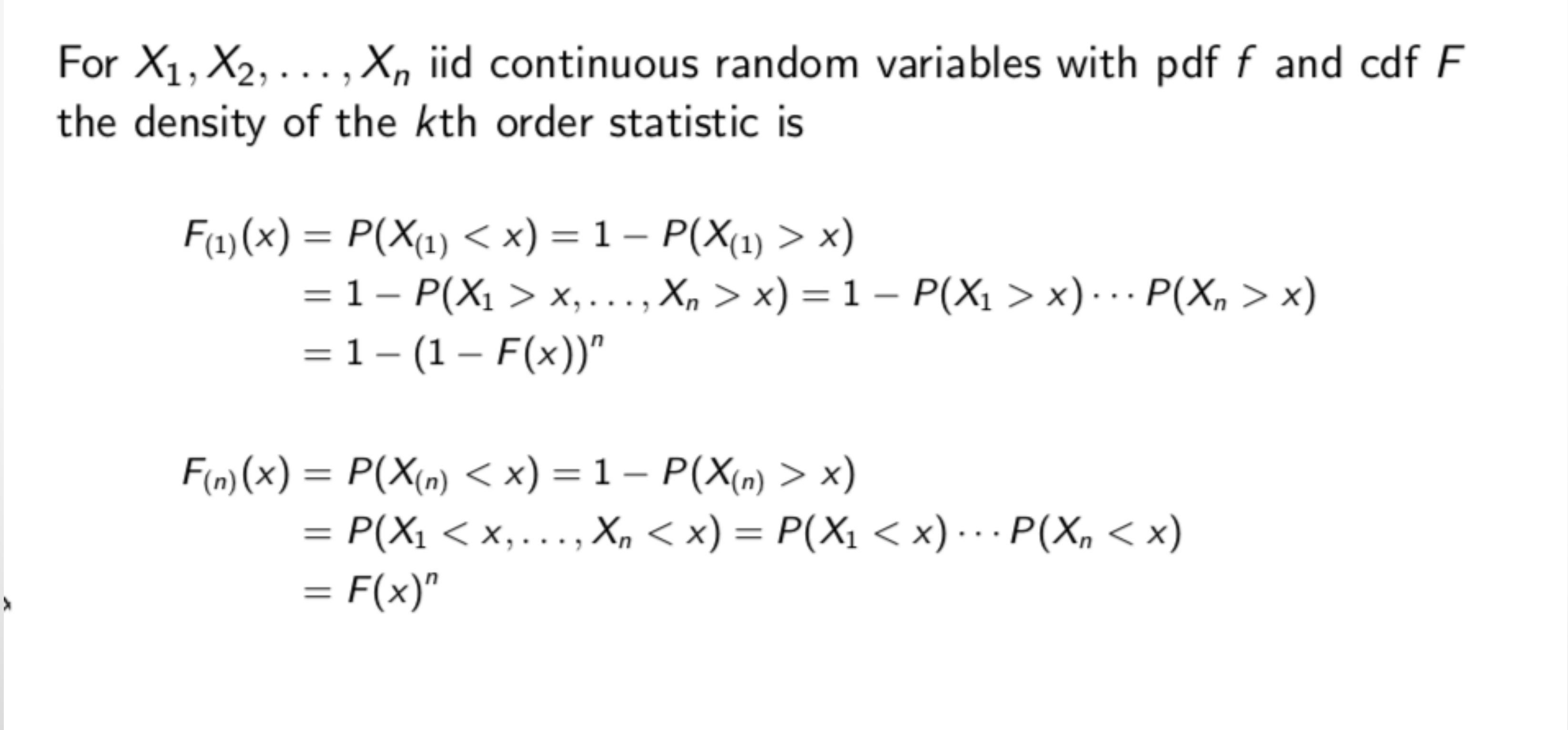
(Notice how I enjoy repeating the author’s name here ☺)

As a result, some ordering was obtained due to the correct condition $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(m)}$. However, this condition is not completely fulfilled due to the first mistake. Due to the second mistake the whole range $(0, T]$ is sampled despite the first mistake.

Markov Chains and MC

THE FOLLOWING SHOULD BE CORRECT. If you spot a glitch somewhere, please let me know:

The CDF's for the first (and for reference, the last) event (see Online Statistics Course); here, $n = m$:



For $X_1, X_2, \ldots, X_n$ iid continuous random variables with pdf f and cdf F the density of the kth order statistic is

$$\begin{aligned} F_{(1)}(x) &= P(X_{(1)} < x) = 1 - P(X_{(1)} > x) \\ &= 1 - P(X_1 > x, \ldots, X_n > x) = 1 - P(X_1 > x) \cdots P(X_n > x) \\ &= 1 - (1 - F(x))^n \end{aligned}$$

$$\begin{aligned} F_{(n)}(x) &= P(X_{(n)} < x) = 1 - P(X_{(n)} > x) \\ &= P(X_1 < x, \ldots, X_n < x) = P(X_1 < x) \cdots P(X_n < x) \\ &= F(x)^n \end{aligned}$$

Markov Chains and MC

In the book, CDF for the last event was picked by mistake. When we take the first-event CDF:

$$F_{(1)}(x) = 1 - \left[1 - F_{U(0,T]}(x)\right]^m = 1 - \left[1 - \frac{x}{T}\right]^m$$

Calculate **the inverse for the first event** in the interval $(0, T]$:

(the 1st event)

$$u = 1 - \left[1 - \frac{x}{T}\right]^m \Leftrightarrow x = \left[1 - (1-u)^{1/m}\right]T \equiv x_{(1)} \swarrow$$

Then we calculate the time for the ith event $x_{(i)}$ conditional on $x_{(i-1)}$ - meaning that $i - 1$ events have occurred before.

Clearly, $x_i \sim U(x_{(i-1)}, T]$, and there will be $m - i + 1$ events in the interval $(x_{(i-1)}, T]$.

Markov Chains and MC

Accordingly, CDF for $X(i)$:

$$F_{X(i)} = 1 - \left\{1 - \left[1 - F_{U(x_{(i-1)},T]}\right]^{m-i+1}\right\}$$
$$= 1 - \left[1 - \frac{x_{(i)}}{T - x_{(i-1)}}\right]^{m-i+1}$$

In the **inverse transformation** we need to take into account that we are relating $u \sim U(0,1]$ to $x_{(i)} \sim F_{X(i)}$, where $x_{(i)} \in (x_{(i-1)}, T]$; accordingly, $x_{(i)} \rightarrow x_{(i)} - x_{(i-1)}$ in the CDF.

$$u = 1 - \left[1 - \frac{x_{(i)} - x_{(i-1)}}{T - x_{(i-1)}}\right]^{m-i+1} \Leftrightarrow$$
$$x_{(i)} = x_{(i-1)} + \left[1 - (1-u)^{\frac{1}{m-i+1}}\right]\left(T - x_{(i-1)}\right) =$$
$$= x_{(i-1)} + \left[1 - u^{\frac{1}{m-i+1}}\right]\left(T - x_{(i-1)}\right)$$

Markov Chains and MC

From the previous derivation, the revised algorithm for sampling events in order reads as:

```
1. Set $x_{(0)} := 0$
2. Sample $m \sim Po(U_\lambda T)$
3. For $i := 1, \dots, m$
    (a) Sample $u \sim U(0,1)$
    (b) Set $x_{(i)} := x_{(i-1)} + \left(T - x_{(i-1)}\right)\left(1 - u^{1/(m-i+1)}\right)$
    (c) Sample $y \sim U(0, U_\lambda)$
    (d) If $y \leq \lambda\left(x_{(i)}\right)$, print $x_{(i)}$
```

This algorithm gives all events in order.

Markov Chains and MC

Thinning approach to jump time simulation

In simulating the process described above one can sample the *inter-event times* from an $Exp(U_\lambda)$ distribution to get the x-coordinates, then sample the corresponding y-coordinates from a $U(0, U_\lambda)$ distribution and reject the points outside Ω.

The thinning algorithm:

```
1. Set $t_0 := 0$
2. Set $i := 1$
3. Repeat
    (a) Sample $t \sim Exp(U_\lambda)$
    (b) Set $t_i := t_{i-1} + t$
    (c) If $t_i > T$, stop
    (d) Sample $u \sim U(0,1)$
    (e) If $u \leq \lambda(t_i)/U_\lambda$, print $t_i$
    (f) Set $i := i + 1$
```

Markov Chains and MC

The thinning algorithm only requires knowledge of the function $\lambda(t)$ – and only at a finite number of times (although time is continuous here) – and not on its integral or inverse.

Hence, the algorithm can be used when a hazard function is not analytically convenient or even deterministic; for example $\lambda(t)$ may be determined by a stochastic process. Only the upper bound has to be available.

The name comes from the fact that we are using rejection to *thin* a homogeneous Poisson process to give a target inhomogeneous process.

The algorithm can be modified to use any inhomogeneous Poisson process with a rate function $\nu(t)$ such that $\nu(t) \geq \lambda(t)\ \forall \in [0, T]$ and which can be directly simulated. Often linear and quadratic forms for $\nu(t)$ are used. This is analogous to *envelope rejection*.