STA 4103/5107 Computational Methods in Statistics II

Department of Statistics
Florida State University

Class 5
January 24, 2017

Review: Time Rescaling Theorem

- **Time Rescaling Theorem** Suppose $\lambda(t) > 0$ in $[0, T]$. If $\{s_1, \ldots, s_n\}$ is a random sample from a Poisson process with rate $\lambda(t)$, then $\{t_1, \ldots, t_n\}$, with

$$t_k = F(s_k) \text{ and } F(s) = \int_0^s \lambda(t)dt$$

is a Poisson process with constant rate 1 from $[0, F(T)]$.

- **Simulation using Time Rescaling Theorem**

Simulation of a Poisson process with rate function $\lambda(t)$ on $[0, T]$:

Step 1: Sample $\{t_1, \ldots, t_n\}$ from Poisson with constant rate 1 on $[0, F(T)]$.

Step 2: Output $\{F^{-1}(t_1), \ldots, F^{-1}(t_n)\}$.

Review: Simulation by Thinning

- Simulation of a Poisson process with rate function $\lambda(t)$ on $[0, T]$:

  Step 1: Sample $\{s_1, \ldots, s_n\}$ from Poisson process with constant rate $M = \max(\lambda(t))$ on $[0, T]$.

  Step 2: For each s_i, we delete it with probability $1 - \lambda(s_i)/M$.

  Step 3: Output the remaining sample.

- **Example:** Simulate an inhomogeneous Poisson process over the interval [0, 10] where the rate function

$$\lambda(t) = 3 + 3\sin(2t)$$

  1. Plot the rate function versus time t.
  2. Generate 30 sample paths for this process.

Chapter 7

Markov Chain Monte Carlo Methods

Review: Monte Carlo Method

- In Chapter 6, we study the classical formulation of the Monte Carlo methods and their applications.
- In cases where it is not possible to directly sample from a distribution, one resorts to sampling in an "approximate" fashion.
- The idea is to sample from a distribution which converges to the desired distribution asymptotically.
- In Chapter 7, we will present an important tool commonly used for this approximate sampling, namely the **Markov Chain Monte Carlo (MCMC)** technique.

Basic Idea

- We focus on a family of discrete time stochastic processes that provide a tool to sample from probability distributions.
- Let X be a continuous random variable with a probability density function $g(x)$, our goal is to generate samples from g.
- One such method is the use of Markov chains to sample from complicated probability density functions, when the direct methods do not apply.
- The basic idea is to construct a Markov chain, starting from almost arbitrary initial conditions, but satisfying a set of conditions relating to g, so that as time n gets longer the values of the sequence start resembling the samples of g.

7.1 Introduction to Markov Chain

Discrete Time Stochastic Processes

- Denote a discrete time stochastic process as: $\{X_t, t = t_1, t_2, \ldots\}$. Such a process can be characterized by n^{th}-order joint probability density function, for any $n \geq 1$,

$$f_{X_{t_1}, X_{t_2}, \cdots, X_{t_n}}(x_1, x_2, \cdots, x_n).$$

- When the sequence times are given, the density can be simplified as:

$$f(x_1, x_2, \cdots, x_n).$$

- Using the rule for total probability, we can factor the joint density function as:

$$f(x_1, x_2, \cdots, x_n) = f(x_1) f(x_2 \mid x_1) \cdots f(x_n \mid x_{n-1}, \cdots, x_2, x_1)$$

Markov Process

- **Definition 16** A stochastic process is called a Markov process if

$$f(x_n \mid x_{n-1}, \cdots, x_2, x_1) = f(x_n \mid x_{n-1})$$

- A Markov process is a stochastic process whose past has no influence on the future if its present is specified.
- This definition implies that joint density function can be written as a product of one-step conditional densities as follows:

$$f(x_1, x_2, \cdots, x_n) = f(x_1) f(x_2 \mid x_1) \cdots f(x_n \mid x_{n-1})$$

- To specify a Markov process one needs only a marginal density and a sequence of one-step transition densities.

Properties with Markov Processes

- **1.** The conditional expectation of the future, given the past and the present is equal to the conditional expectation of the future given only the present. That is,

$$E[X_{t_n} \mid X_{t_{n-1}}, \cdots, X_{t_1}] = E[X_{t_n} \mid X_{t_{n-1}}].$$

Proof: By definition,

$$
\begin{aligned}
& E[X_{t_n} \mid X_{t_{n-1}}, \cdots, X_{t_1}] \\
&= \int x_n f(x_n \mid x_{n-1}, \cdots, x_1)\, dx_n \\
&= \int x_n f(x_n \mid x_{n-1})\, dx_n \\
&= E[X_{t_n} \mid X_{t_{n-1}}].
\end{aligned}
$$

Properties with Markov Processes

- **2.** A Markov process remains Markov if the time index is reversed. That is,

$$f(x_1 \mid x_2, \cdots, x_n) = f(x_1 \mid x_2)$$

Proof: The left side is:

$$f(x_1 \mid x_2, \cdots, x_n) = \frac{f(x_1, x_2, \cdots, x_n)}{f(x_2, \cdots, x_n)}$$

$$= \frac{f(x_1) f(x_2 \mid x_1) \cdots f(x_n \mid x_{n-1})}{f(x_2) f(x_3 \mid x_2) \cdots f(x_n \mid x_{n-1})}$$

$$= \frac{f(x_1) f(x_2 \mid x_1)}{f(x_2)} = f(x_1 \mid x_2)$$

Properties with Markov Processes

- **3.** Conditioned on a given value of the present, the past and the future are statistically independent of each other. That is,

$$f(x_1, x_3 \mid x_2) = f(x_1 \mid x_2) f(x_3 \mid x_2)$$

Proof:

$$f(x_1, x_3 \mid x_2) = \frac{f(x_1, x_2, x_3)}{f(x_2)}$$

$$= \frac{f(x_1 \mid x_2) f(x_2 \mid x_3) f(x_3)}{f(x_2)}$$

$$= f(x_1 \mid x_2) f(x_3 \mid x_2)$$

Properties with Markov Processes

- **4.** A Markov process satisfies a condition, called Chapman-Kolmogorov condition, that will be useful later. That is, for a Markov process,

$$f(x_3 \mid x_1) = \int f(x_3 \mid x_2) f(x_2 \mid x_1) dx_2$$

Proof:

$$f(x_3 \mid x_1) = \int f(x_3, x_2 \mid x_1) dx_2$$

$$= \int f(x_3 \mid x_2) f(x_2 \mid x_1) dx_2$$

Stationary Stochastic Processes

- **Definition 17** A stochastic process is called **stationary** if its n^{th}-order joint density function is translation invariance, for all $n \geq 1$. In other words, for any collection of times $\{t_1, t_2, \ldots, t_n\}$, we have
$$f_{X_{t_1}, X_{t_2}, \cdots, X_{t_n}}(x_1, x_2, \cdots, x_n) = f_{X_{t_1+c}, X_{t_2+c} \cdots, X_{t_n+c}}(x_1, x_2, \cdots, x_n),$$
for all $n > 0$ and c.

- In particular, for a stationary process the marginal density associated the process does not depend upon the time t; i.e.
$$f_{X_t}(x) = f_{X_{t+k}}(x),$$
for all t.

Homogeneous Markov process

- **Definition 18** A Markov process is called **homogeneous** if the conditional density is invariant to a time shift. That is, for all n,

$$f_{X_{t_n}|X_{t_{n-1}}}(x_n \mid x_{n-1}) = f_{X_{t_2}|X_{t_1}}(x_n \mid x_{n-1}),$$

- It is important to note that the marginal density may be, and often is, dependent upon the time shift, i.e. for some n,

$$f_{X_{t_n}}(x) \neq f_{X_{t_1}}(x).$$

- A stationary Markov process is always homogeneous. However, a homogeneous process, in general, is not stationary.

- When is a homogeneous Markov process stationary?

Stationary Probability Density

- If there exists a density function g such that:

$$g(y) = \int f_{X_{t_2}|X_{t_1}}(y\,|\,x)g(x)\,dx,$$

  then the resulting Markov chain is a stationary process. The density function g is called the **stationary probability density** of that Markov chain.

- Our goal is to construct such homogeneous Markov processes that are not stationary to start with, but converge to stationary processes as the process is followed for a long time.

- If the marginal density of this stationary process matches the density g, then the sample paths of this process sample from g.