University of Regina

Statistics 471/871 – Time Series Analysis

Lecture Notes

Winter 2005

Michael Kozdron

kozdron@math.uregina.ca
www.math.uregina.ca/~kozdron

Version: January 5, 2005

In The Very Beginning . . .

The textbook [2] is *extremely* well-written and is very easy to read. There is also some very good software for analyzing time series called ITSM 2000 which is packaged with this book.

Homework Assigned on Friday, January 7: Make sure you obtain a copy of the textbook *with* the required software. If your text copy does not have the software, please see me. Read the Preface on pages vii–viii. Read the two handouts:

- Handout #1: Course Outline
- Handout #2: Syllabus

Look at Appendices A, B, and C on pages 369–394 as a review of your earlier mathematical statistics courses. Make sure you can get your copy of ITSM 2000 to run. Read Appendix D.1 on page 396 and do Exercise 1.16 on page 43.

Homework Assigned Each and Every Class: Read the appropriate sections in the textbook, and do any related exercises from the textbook. Also, read the appropriate sections of the notes, and do any exercises in those sections.

On-going Assignment: Let me know of any errors in the notes, or of places that could use an improved exposition.

In The Beginning . . .

CAVEAT: I DON'T CARE IF YOU UNDERSTAND ALL OF THIS OR NOT, ESPECIALLY IF YOU DON'T HAVE SUFFICIENT BACKGROUND. HOWEVER, IT IS WORTH HEARING HOW BEAUTIFULLY INTERTWINED THIS STUFF IS! AND WHEN YOU DO LEARN IT, MAYBE YOU'LL LOOK BACK ON TIME SERIES AND SAY, "YEAH, THAT'S PRETTY NEAT."

Since the title of this course is *Time Series Analysis*, it might be useful if we know what a time series is!

Definition. A *time series* is simply a set of observations $\{x_t\}$, with each data point being observed at a specific time t.

However, this definition is severely inadequate. We will soon see how we can improve this definition, and give some substance to the subject of time series analysis.

Notation

The symbol $A := B$ means A *is defined to equal* B, whereas $C = D$ by itself means simply that C and D are equal. This is an important distinction because if you write $A := B$, then there is no need to verify the equality of A and B. They are equal by definition. However, if $C = D$, then there *IS* something that needs to be proved, namely the equality of C and D (which might not be obvious). Exercise 3.1.8 illustrates this subtle difference.

1 Preliminary Remarks and an Overview of the Course

1.1 Introduction to Random Variables

Suppose that Ω is the sample space of outcomes of an experiment.

Example 1.1.1. Flip a coin once: $\Omega = \{H, T\}$.

Example 1.1.2. Toss a die once: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Example 1.1.3. Toss a die twice: $\Omega = \{(i, j) : 1 \leq i \leq 6,\ 1 \leq j \leq 6\}$.

Note that in each case Ω is a *finite* set. (That is, the *cardinality* of Ω, written $|\Omega|$, is finite.)

Example 1.1.4. Consider a needle attached to a spinning wheel centred at the origin. When the wheel is spun, the angle ω made with the tip of the needle and the positive x-axis is measured. The possible values of ω are $\Omega = [0, 2\pi)$.

In this case, Ω is an *uncountably infinite* set. (That is, Ω is uncountable with $|\Omega| = \infty$.)

Definition 1.1.5. A *random variable* X is a function from the sample space Ω to the real numbers $\mathbb{R} = (-\infty, \infty)$. Symbolically, $X : \Omega \to \mathbb{R}$ via

$$\omega \in \Omega \mapsto X(\omega) \in \mathbb{R}.$$

Example 1.1.1 (continued). Let X denote the number of heads on a single flip of a coin. Then, $X(H) = 1$ and $X(T) = 0$.

Example 1.1.2 (continued). Let X denote the upmost face when a die is tossed. Then, $X(i) = i$, $i = 1, \ldots, 6$.

Example 1.1.3 (continued). Let X denote the sum of the upmost faces when two dice are tossed. Then, $X((i, j)) = i + j$, $i = 1, \ldots, 6$, $j = 1, \ldots, 6$. Note that the elements of Ω are ordered pairs, so that the function $X(\cdot)$ acts on (i, j) giving $X((i, j))$. We will often omit the inner parentheses and simply write $X(i, j)$.

Example 1.1.4 (continued). Let X denote the cosine of the angle made by the needle on the spinning wheel and the positive x-axis. Then $X(\omega) = \cos(\omega)$ so that $X(\omega) \in [-1, 1]$.

Remark. The use of the notation X and $X(\omega)$ is EXACTLY analogous to elementary calculus. There, the function f is described by its action on elements of its domain. For example, $f(x) = x^2$, $f(t) = t^2$, and $f(\omega) = \omega^2$ all describe EXACTLY the same function, namely, the function which takes a number and squares it.

Remark. For historical reasons, the term *random variable* (written X) is used in place of *function* (written f) and generic elements of the domain are denoted by ω instead of by x.

Remark. It was A.N. Kolmogorov in the 1930's who formalized probability and realized the need to treat random variables as *measurable* functions. See Math 810: Analysis I or Stat 851: Probability.

1.2 Discrete and Continuous Random Variables

There are two extremely important classes of random variables, namely the so-called discrete and continuous. In a sense, these two classes are the same since the random variable is described in terms of a density function. However, there are slight differences in the handling of sums and integrals so these two classes are often taught separately in undergraduate courses.

Important Observation. Recall from elementary calculus that the Riemann integral $\int_a^b f(x)\,dx$ is *defined* as an appropriate limit of Riemann sums $\sum_{i=1}^N f(x_i^*)\,\Delta x_i$. Thus, you are ALREADY FAMILIAR with the fact that SOME RELATIONSHIP exists between integrals and sums.

Definition 1.2.1. Suppose that X is a random variable. Suppose that there exists a function $f:\mathbb{R}\to\mathbb{R}$ with the properties that $f(x)\geq 0$ for all x, $\int_{-\infty}^{\infty} f(x)\,dx = 1$, and

$$P(\{\omega\in\Omega : X(\omega)\leq t\}) =: P(X\leq t) = \int_{-\infty}^{t} f(x)\,dx.$$

We call f the *(probability) density (function)* of X and say that X is a *continuous* random variable. Furthermore, the function F defined by $F(t) := P(X\leq t)$ is called the *(probability) distribution (function)* of X.

Fact. By the Fundamental Theorem of Calculus, $F'(x) = f(x)$.

Exercise 1.2.2. Prove the fact that $F'(x) = f(x)$, being sure to carefully state the necessary assumptions on f. Convince me that you understand the use of the dummy variables x and t in your argument.

Remark. There exist continuous random variables which *do not* have densities. For our purposes, though, we will always assume that our continuous random variables are ones with a density.

Example 1.2.3. A random variable X is said to be *normally distributed with parameters* μ, σ^2, if the density of X is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \quad -\infty<\mu<\infty,\ 0<\sigma<\infty.$$

This is sometimes written $X\sim\mathcal{N}(\mu,\sigma^2)$. In Exercises 1.3.4 and 3.1.9, you will show that the mean of X is μ and the variance of X is σ^2, respectively.

Definition 1.2.4. Suppose that X is a random variable. Suppose that there exists a function $p:\mathbb{Z}\to\mathbb{R}$ with the properties that $p(k)\geq 0$ for all k, $\sum_{k=-\infty}^{\infty} p(k) = 1$, and

$$P(\{\omega\in\Omega : X(\omega)\leq N\}) =: P(X\leq N) = \sum_{k=-\infty}^{N} p(k).$$

We call p the *(probability mass function or) density* of X and say that X is a *discrete* random variable. Furthermore, the function F defined by $F(N) := P(X\leq N)$ is called the *(probability) distribution (function)* of X.

Example 1.1.3 (continued). If X is defined to be the sum of the upmost faces when two dice are tossed, then the density of X, written $p(k) := P(X=k)$, is given by

$p(2)$	$p(3)$	$p(4)$	$p(5)$	$p(6)$	$p(7)$	$p(8)$	$p(9)$	$p(10)$	$p(11)$	$p(12)$
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

and $p(k) = 0$ for any other $k \in \mathbb{Z}$.

Remark. There do exist random variables which are neither discrete nor continuous; however, such random variables will not concern us.

1.3 Law of the Unconscious Statistician

Suppose that $X : \Omega \to \mathbb{R}$ is a random variable (either discrete or continuous), and that $g : \mathbb{R} \to \mathbb{R}$ is a (piecewise) continuous function. Then $Y := g \circ X : \Omega \to \mathbb{R}$ defined by $Y(\omega) = g(X(\omega))$ is also a random variable.

We now define the expectation of the random variable Y, distinguishing the discrete and continuous cases.

Definition 1.3.1. If X is a discrete random variable and g is as above, then the *expectation* of $g \circ X$ is given by

$$\mathbb{E}(g \circ X) := \sum_k g(k)\, p(k)$$

where p is the probability mass function of X.

Definition 1.3.2. If X is a continuous random variable and g is as above, then the *expectation* of $g \circ X$ is given by

$$\mathbb{E}(g \circ X) := \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$

where f is the probability density function of X.

Notice that if $g(x) = 1$ for all x, then the expectation of X itself is

- $\mathbb{E}(X) := \sum_k k\, p(k)$, if X is discrete, and
- $\mathbb{E}(X) := \int_{-\infty}^{\infty} x\, f(x)\, dx$ if X is continuous.

Remark. The first place I read of the title *Law of the Unconscious Statistician* was in [1]. There doesn't seem to be much history of this funny title, other than that it appears to have been coined by Paul Halmos.

Exercise 1.3.3. Suppose that X is a Bernoulli(p) random variable. That is, $P(X = 1) = p$ and $P(X = 0) = 1 - p$ for some $p \in [0, 1]$. Carefully verify that

- $\mathbb{E}(X) = p$,
- $\mathbb{E}(X^2) = p$, and
- $\mathbb{E}(e^{-\theta X}) = 1 - p(1 - e^{-\theta})$, for $0 \le \theta < \infty$.

Exercise 1.3.4. The purpose of this exercise is to make sure you can compute some straightforward (but messy) integrals. Suppose that $X \sim \mathcal{N}(\mu, \sigma^2)$; that is, X is a normally distributed random variable with parameters μ, σ^2. (See Example 1.2.3 for the density of X.) Show directly (without using any unstated properties of expectations or distributions) that

- $\mathbb{E}(X) = \mu$,
- $\mathbb{E}(X^2) = \sigma^2 + \mu^2$, and
- $\mathbb{E}(e^{-\theta X}) = \exp\left(-\theta\mu - \frac{\sigma^2\theta^2}{2}\right)$, for $0 \leq \theta < \infty$.

Together with Exercise 3.1.9, this is the reason that if $X \sim \mathcal{N}(\mu, \sigma^2)$, we say that X is normally distributed with mean μ and variance σ^2.

1.4 Stochastic Processes

Definition 1.4.1. A *stochastic process* is simply an infinite collection of indexed random variables.

Definition 1.4.2. A *discrete time stochastic process* is a sequence of random variables $\{X_n,\ n = 1, 2, 3, \ldots\}$ indexed by the positive integers.

Definition 1.4.3. A *continuous time stochastic process* is a family of random variables $\{X_t,\ 0 \leq t < \infty\}$ indexed by the positive real numbers.

Important Language Remark. The adjectives discrete time and continuous time refer to the indexing set of the collection of random variables and NOT to the individual random variables themselves.

Example 1.4.4. Suppose that for each $n \in \mathbb{N}$ the random variable X_n is normally distributed with mean 0 and variance n. Then the discrete time stochastic process $\{X_n, n = 1, 2, 3, \ldots\}$ consists of individual continuous random variables.

Example 1.4.5. Suppose that for each $t \in [0, \infty)$ the random variable X_t is normally distributed with mean 0 and variance t. Then the continuous time stochastic process $\{X_t, 0 \leq t < \infty\}$ consists of individual continuous random variables.

1.5 Realizations and Time Series

If X is a random variable, then we call $X(\omega)$ a *realization* of the random variable. The physical interpretation is that if X denotes the UNKNOWN outcome (*a priori*) of the experiment before it happens, then $X(\omega)$ represents the realization or observed outcome (*a posterior*) of the experiment after it happens.

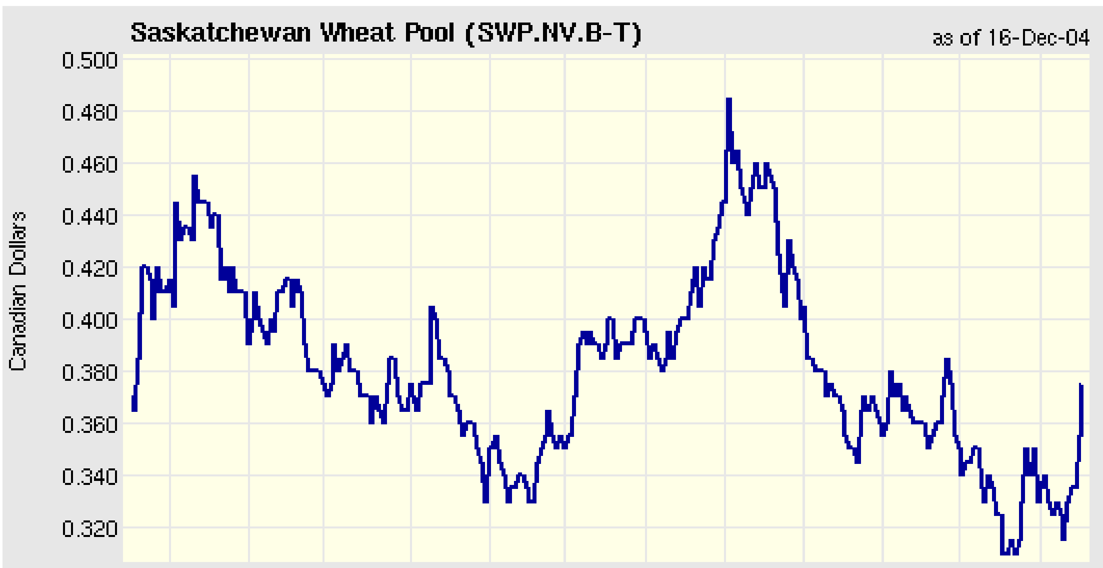
We are now sufficiently prepared to state an improvement of our original definition of time series.

Definition 1.5.1. A *time series* is a realization of a stochastic process.

In a sense, this is saying the same thing as our original definition. However, much more utility is gained via the realization of a stochastic process point of view, than if we limit ourselves to the view that a time series is only a collection of data increasingly ordered by time.

Example 1.5.2. Consider the stock SWP (Saskatchewan Wheat Pool) which is traded on the Toronto Stock Exchange. Consider the value X_n, $n = 1, \ldots, 365$ of the stock at the end of each trading day from January 1 to December 31, 2005. At the moment the values of all of these random variables are not known. Thus, we can view $\{X_n\}$ as a *stochastic process*. However, if we consider the value Y_n, $n = 1, \ldots, 365$ of the same stock at the end of each trading day from January 1 to December 31, 2004, then all those values are known so we can view $\{Y_n\}$ as a *time series*.

Remark. It is often useful to plot a "time series plot" of the time series (or particular realizations of a stochastic process). That is, plot n vs. X_n or t vs. X_t. The resulting graph is known as a *sample path*, and is also sometimes called the *trajectory* of the time series or stochastic process.



This is a "time series plot" of the closing prices of the stock SWP from December 16, 2003, until December 16, 2004.

1.6 Analyzing Random Variables and Stochastic Processes

If we know the distribution of a random variable, then we know all of the information about that random variable. For example, if we know that X is a normal random variable with mean 0 and variance 1, then we know everything possible about X without actually realizing it.

Unfortunately, it is not so easy to analyze stochastic processes. Recall that the *joint dsistribution function* of a random vector $(X_1, X_2, \ldots, X_d)$ is the (d-dimensional) function

$$F(t_1, t_2, \ldots, t_d) = P(X_1 \le t_1, X_2 \le t_2, \ldots, X_d \le t_d).$$

Hence, the joint distribution function for a discrete time stochastic process $(X_1, X_2, \ldots)$ is the (infinite-dimensional) function

$$F(t_1, t_2, \ldots) = P(X_1 \le t_1, X_2 \le t_2, \ldots).$$

But what about the joint distribution function for a continuous time stochastic process $\{X_t,\ 0 \le t < \infty\}$? Now, we CANNOT specify all of the possible values of the constituent random variables since there are uncountably many of them.

(That is, with an infinite sequence we can at least visualize all the random variables: X_1, X_2, X_3, X_4, etc. But with an uncountable collection of times, the problem is that for every pair of real numbers $a < b$, there exists a c with $a < c < b$. Essentially, this is the Intermediate Value Theorem in disguise.)

For a continuous time stochastic process, the best we can do is specify the *finite-dimensional distributions*, namely the following: for each N, and for each sequence of times $0 \leq s_1 \leq s_2 \leq \cdots \leq s_N$, the distribution of $(X_{s_1}, X_{s_2}, \ldots, X_{s_N})$.

YUCK! That is way too much to keep track of; it is untenable!

INSTEAD: If we impose additional structure on the stochastic process, in much the same way that we study particular random variables (uniform, normal, Poisson, binomial, Bernoulli, Cauchy, exponential, geometric, beta, gamma, Chi-squared), then the stochastic process becomes amenable to analysis.

Some particular "types" of stochastic processes are: random walks, martingales, renewal sequences, exchangeable sequences, point processes, Lévy processes, Markov processes, interacting particle systems, branching processes, diffusions, Wiener processes (aka Brownian motion). Whole volumes are written on particular aspects of each one of these!

Fortunately for us, our main focus will be on *stationary processes.* The analysis of time series is made possible by our understanding of stationary processes.

1.7 This Course

Stat 471/871 is an introduction to the analysis of time series. It is worth noting that mathematicians and statisticians are particularly keen to call things "introductions" when, in fact, they require anything but introductory or elementary techniques. Perhaps instead of "An Introduction to the Analysis of Time Series" this course would be better titled "The Analysis of Univariate Time Series."

As noted in the Preface of our textbook [2], there is sufficient material for a full year introduction to univariate and multivariate time series and forecasting. We will follow their advice, and cover chapters 1 through 6 in this one-semester course on univariate time series. My plan is to proceed linearly through the text, except that I will cover the chapters in the order 1, 2, 3, 5, 6, 4.

One reason for studying stationary processes is that the machinery of complex analysis and Fourier analysis make it possible to perform a *spectral analysis* which completely characterizes stationary processes. Chapter 4 in [2] provides an explanation of these methods, but I believe that it is best left for a final topic.

2 The *Statistics* of Time Series Analysis

2.1 Estimating Parameters

Recall that the overarching goal of Statistics is to estimate population *parameters.* This is done by calculating *statistics*, which are simply numbers computed from data, and using them as *point estimates* of the appropriate parameter.

As you learned in Stat 151, if you have a population with an unknown mean μ and unknown