We have seen the types of statistical animals we will be dealing with, both from a data collection perspective (agricultural models, for example) and as mathematical objects like random walks. We will now start putting these time series objects into a mathematical framework so that we can analyze, predict, and estimate.

Stochastic Processes

When you studied homoscedasticity in probability or regression you probably said to yourself: "You don't expect me to spell that, do you? Spelling Connecticut is hard enough." You also may have wondered why people don't just say "constant variance".

It's the same here- *stochastic processes* are really just *random processes*.

> A *discrete stochastic process* is a family of random variables structured as a *sequence* (finite or infinite) and has a discrete time index, denoted X_t . Discrete stochastic processes may model, for instance, the recorded daily high temperatures in Melbourne, Australia.
>
> A *continuous stochastic process* is also a family of random variables, but is indexed by a continuous variable and denoted as $X(t)$. A commonly encountered continuous process is the *Weiner Process*, having nothing to do with hot dogs, but instead describing a particle's position as a function of time as it floats on the surface of a liquid (Brownian Motion). Another commonly encountered continuous process is the *Poisson Process*.

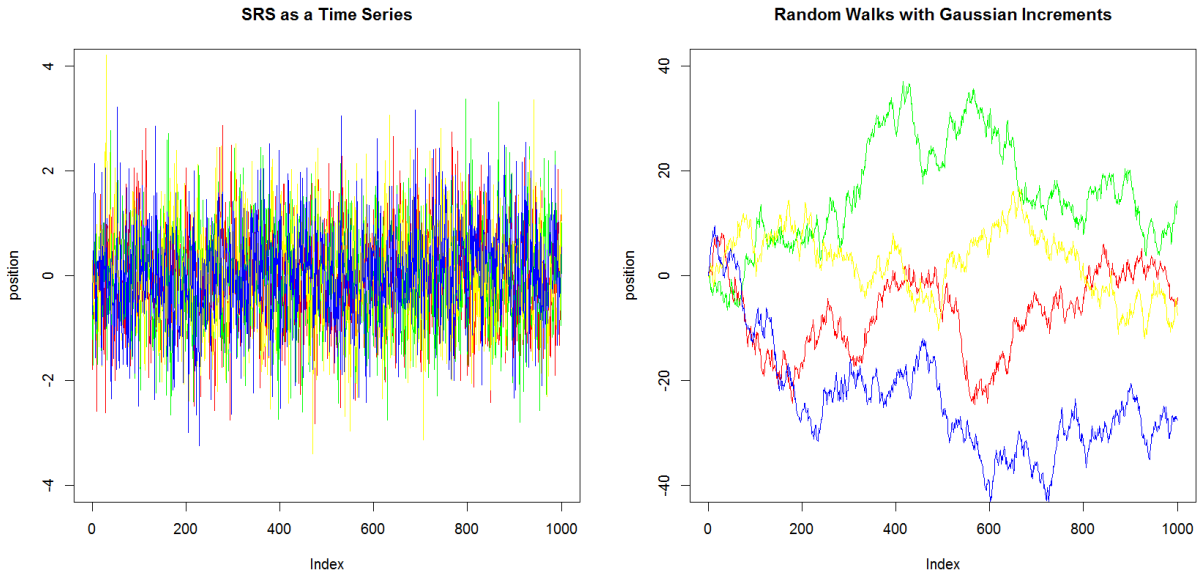
Ensembles and Realizations

We can create a very simple stochastic process if we think about sampling.

In elementary statistics courses, we often wish to estimate the mean of a population. Suppose you think that a population of Math SAT scores are approximately normally distributed. We might form a simple random sample (SRS) and construct a sample mean $\bar{x} = \frac{\sum x_i}{n}$ as an estimator of the population. We don't usually think of this as a random process, but we could, with an indexed set of random variables corresponding to the observations $X_1, X_2, \ldots, X_N$. There's no interesting structure here, the observations are all independent.

Since addition is commutative, we probably don't care about the order in which the SAT scores come at us, we just want to add them up. Four different samples are plotted below…what a mess!

Contrast this with the example of the random walk we have already discussed. We keep track of where we are (our position up/down) at a set of time steps. At each new time step we take our previous location and add a random amount. The trajectories taken by a particle will be quite different if we run the experiment a few times. Each of these trajectories is another time series.



When we acquire time series data in the field, we don't usually have the luxury of observing multiple trajectories.

> *We usually just have one sequentially observed data set and must infer the properties of the generating process from this single trajectory.*

The usual terminology here is that an individual trajectory corresponds to a *realization* of a stochastic process. This is what we have been calling a time series. The set of all possible trajectories is called the *ensemble*.

> *A stochastic process is a rather complicated thing. To fully specify its structure we would need to know the joint distribution of the full set of random variables.*

Mean, Variance, and Autocovariance Functions

Since the joint distribution is really too much to think about, and is also usually unknown, we try to summarize instead. To start off, think of a discrete stochastic process, and mentally create a table where we keep track of each random variable's mean and variance.

Table 1 Mean and Variance Function for a generic discrete stochastic process

X_1	X_2	X_3	...	X_N
$E[X_1] = \mu_1$	$E[X_2] = \mu_2$	$E[X_3] = \mu_3$		$E[X_N] = \mu_N$
$V[X_1] = \sigma_1^2$	$V[X_2] = \sigma_2^2$	$V[X_3] = \sigma_3^2$		$V[X_N] = \sigma_N^2$

We have obviously worked to define a mean function and a variance function, indexed consistently with the appropriate random variable. For the simple random sample example, this would be a very boring function!

Table 2 Boring Mean and Variance Function for the SRS example

X_1	X_2	X_3	...	X_N
$E[X_1] = \mu$	$E[X_2] = \mu$	$E[X_3] = \mu$		$E[X_N] = \mu$
$V[X_1] = \sigma^2$	$V[X_2] = \sigma^2$	$V[X_3] = \sigma^2$		$V[X_N] = \sigma^2$

In both the discrete setting and the continuous setting we say

$$Mean\ Function: \quad \mu(t) \equiv \mu_t \equiv E[X(t)] \text{ also written } E[X_t]$$

$$Variance\ Function: \quad \sigma^2(t) \equiv \sigma_t^2 \equiv V[X(t)] \text{ also written } V[X_t]$$

We also care about how the random variables *relate* to one another. We have been looking at the correlation and covariance functions, and continue in the spirit above by defining, for random variables $X(t_1)$ and $X(t_2)$

$$Autocovariance\ Function: \ \gamma(t_1, t_2) \equiv E\big[\big(X(t_1) - \mu(t_1)\big)\big(X(t_2) - \mu(t_2)\big)\big]$$

All of these functions; the mean, the variance, and the covariance functions depend upon the particular random variables we are considering (indexed by time). Can we make more global statements about the process itself (and lose the time index)?

Strict Stationarity

The problem here is that, for a given stochastic process we only have one observed time series. This is a bit like asking you whether a coin is fair by letting you toss it just once. It's just a silly question.

To get some traction, we often assume a property called *stationarity*. This will help us to essentially "pool" our data to in order to say something useful about the process

As we said above, to completely know the stochastic process we need to know more than the individual probability densities of our random variables, we also need to know the joint distribution of the entire family. In your introductory probability course you probably considered 2-d random variables (remember the Calculus III style iterated-integrals?) We now have much larger systems on hand, with thousands or even an infinite number of random variables. Performing iterated integrals isn't going to work! Let's make the first, very restrictive simplification of *strict stationarity*. It essentially says that the joint distribution doesn't really depend upon where you are looking along the stochastic process. More formally, we say a stochastic process is strictly stationary if

$$\begin{matrix} Joint\ Distribution\ of \\ X(t_1),\ X(t_2), \dots,\ X(t_k) \end{matrix} \quad is\ the\ same\ as \quad \begin{matrix} Joint\ Distribution\ of \\ X(t_1+\tau),\ X(t_2+\tau), \dots,\ X(t_k+\tau) \end{matrix}$$

In this setting, if we shift our attention by a distance τ we will still have the same joint distribution. In particular, the joint distribution *only depends upon our lag spacings*. This is actually quite profound and yields beautifully simplifying results.

First, consider the simple case of $k = 1$. Then immediately

$$\begin{matrix} The\ distribution\ of \\ X(t_1) \end{matrix} \quad \text{is the same as} \quad \begin{matrix} The\ distribution\ of \\ X(t_1+\tau) \end{matrix}$$

Since τ is arbitrary, this means that all of the underlying random variables have the *same distribution*. They are all identically distributed (though not necessarily *IID*, emphasis on *independent*). Try to think through whether you believe a time series with a trend can be stationary. Support your reasoning by making a statement on $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$.

Now since each random variable $X(t)$ has the same distribution, our two functions (mean and variance) must be *constant* and we can say (when they exist!)

$$Mean\ Function: \quad \mu(t) = \mu \qquad Variance\ Function: \quad \sigma^2(t) = \sigma^2$$

How about a function which says something about the relationship between the random variables? Let $k = 2$ and say

$$\begin{matrix} The\ joint\ Distribution\ of \\ X(t_1),\ X(t_2) \end{matrix} \quad \text{is the same as} \quad \begin{matrix} The\ joint\ Distribution\ of \\ X(t_1+\tau),\ X(t_2+\tau) \end{matrix}$$

The joint distribution depends only on the *lag spacing*, so

$$Autocovariance\ Function: \ \gamma(t_1, t_2) = \gamma(t_2 - t_1) = \gamma(\tau)$$

It doesn't matter where you are sitting along the process, we just care about the distance between the random variables. Strict stationarity is a very strong condition, but as you can see it gives us some extremely simplifying results. Just stating the obvious, we can also scale the covariance to obtain the *autocorrelation function*

$$\rho(\tau) \equiv \frac{\gamma(\tau)}{\gamma(0)}$$

Strict stationarity is very strong but also hard to determine. How about a weaker assumption?

Weak (also called second-order) Stationarity

There is a less burdensome condition we can impose which will still allow us to "pool" our results across a realization in order to infer properties of the ensemble. Rather than getting our results as a consequence of strict stationarity, we can just let these *results* define our characterization by saying that a process is *weakly stationary* if

$$Mean\ Function: \quad \mu(t) = \mu$$

$$Autocovarince\ Function: \ \gamma(t_1, t_2) = \gamma(t_2 - t_1) = \gamma(\tau)$$

We are being efficient in our definition and are including the variance *en passant*. For an easy result, note that if $\tau = 0$ we also immediately have constant variance:

$$\gamma(0) = constant = \ E[(X(t) - \mu)(X(t) - \mu)]$$