Introduction to Modelling

A model is an imitation of a real world system, used to investigate effects of changes without the disadvantages of actually changing the system

⊕ Less risky, expensive or slow than trying it in the real world.
⊕ Ability to try different scenarios to find the best one
⊖ Creation requires high investment in time and expertise
⊖ Reliance on input data and model accuracy (future events may invalidate it)

When investigating a models suitability, we need to consider

- Validity of a model for its objectives
- Errors caused by parameters not perfectly representing the real world.
- Credability of the data input and thus the data output.
- Impact of correlation between random variables driving the model
- Spurious accuracy and ease of communicating results.

The process of creating a model goes as follows.

- Develop a set of objectives to be met by the model
- Plan the modelling process and how the model will be validated
- Collect and analyse the necessary data
- Define the model, capturing the essence of the real world system
- Involve experts of the real world model to give feedback
- Decide the language / package, write it and debug it.
- Test for reasonableness (Turing test, sensitivity analysis)
- Analyse the data and communicate the results.

Basic Stochastic Concepts

This course deals with modelling systems with stochastic processes $\{X_t : t \in J\}$. These are sets of random variables $X_t \in S$ indexed by time $t \in J$. The nature of the state space S and time J give different classes of processes.

Cts. space,	Disc time	AMIRA
Disc space,	Disc time	Markov chains
Disc space,	Cts time	Markov jump processes
Cts space,	Cts time	Brownian motion, diffusions

The process is defined by the joint distributions of $X_{t_1}, \ldots, X_{t_n}$ $\forall t_1, \ldots, t_n \forall n$. Typically we work with those expressible in simpler terms. In particular, they have some of the following characteristic properties.

- (Weak) Stationarity. - $E[X_t] = k$ $\forall t$ and $cov(s,t) = E[X_s X_t] - E[X_s][X_t]$ depends only on the lag $t-s$. This is a weaker version of stationarity which is sufficient for many practical purposes

- Strict Stationarity - The joint distributions of $X_{t_1}, \ldots, X_{t_n}$ and $X_{t_1+k}, \ldots, X_{t_n+k}$ a identical for all $t_1, \ldots, t_n$ and $t_1+k, \ldots, t_n+k$ in J and integers n.

- Independent Increments - The increment $X_{t+u} - X_t$ is independent of $\{X_s : 0 \leq s \leq$

- Markov Property. - $P[X_t \in A \mid X_{s_1} = x_1, \ldots, X_{s_n} = x_n] = P[X_t \in A \mid X_{s_n} = x_n]$

- Martingale Property - $E[X_{t+u} \mid X_t] = X_t$

- Purely Indeterministic - Knowledge of the values of $X_1, \ldots, X_n$ is p progressively less useful at predicting X_N as $N \to \infty$.

Classifying ARIMA Series

Firstly, let $\{e_t : t \in \mathbb{Z}\}$ be a sequence of independant normal random variables with common mean 0 and variance σ^2. Secondly, define the backwards shift operator B such that $(BX)_t = X_{t-1}$.

When given a time series to classify as ARIMA, replace all X_{t-n} with B^n, all e_{t-n} with Be_t and cast into the following form, dividing factors thro

$$(1 + \ldots + \alpha_p B^p)(1 - B)^d (X - M) = (1 + \ldots + \beta_q B^q) e$$

This process is characterised as ARIMA (p, d, q) with difference $I(d)$.

To better understand what an ARIMA process is, we split it into AR, I and MA

$I(d)$ This part specifies that the $d-1^{th}$ set of differences is not stationa but that the d^{th} set is. In the context of the ARIMA model, this d^{th} set is modelled by an ARMA process.

$AR(p)$ The equations (normal and with B) of an autoregressive model are

$$X_t = \mu + \alpha_1 (X_{t-1} - \mu) + \ldots + \alpha_p (X_{t-p} - \mu) + e_t$$
$$(1 - \alpha_1 B - \ldots - \alpha_p B^p)(X - \mu) = e$$

If the process is to be stationary, the roots of the equatio $1 - \alpha_1 z - \ldots - \alpha_p z^p = 0$ must all be greater than 1 in absolute value.

$MA(q)$ The equations (normal and with B) of a moving average model are

$$X_t = \mu + e_t + \beta_1 e_{t-1} + \ldots + \beta_q e_{t-q}$$
$$X - \mu = (1 + \beta_1 B + \ldots + \beta_q B^q) e$$

If the process is to be invertible, the roots of the equation $1 + \beta_1 z + \ldots + \beta_q z^q = 0$ must all be greater than 1 in absolute value

The covariance between pairs of elements a set distance apart is a useful way of characterising ARMA models.

Define the auto covariance as $\gamma_k = cov(X_t, X_{t+k}) = E(X_t X_{t+k}) - E(X_t)E(X_{t+k})$
Note that $\gamma_0 = cov(X_t, X_t)$ is the common variance of all X_t.

Finding autocovariance can be easy - MA(q) has $\gamma_k = \sigma^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} \quad \forall k \leq q$.
For AR it is harder, though - AR(3) gives us the relationship
$\gamma_k = \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \alpha_3 \gamma_{k-3} + \sigma^2 1_{\{k=0\}}$ which we have to solve using simultaneous equations to get the autocovariance.

Define the auto correlation function (ACF) as $\rho_k = \frac{\gamma_k}{\gamma_0}$

Define the partial autocorrelation function (PACF) as the coefficient $\phi_{k,k}$ when choosing coefficients to minimise $E((X_t - \phi_{k,1} X_{t-1} - \ldots - \phi_{k,k} X_{t-k})^2)$.
For example $\phi_1 = \rho_1$, $\phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$. In practice, we don't calculate this.

When calculating these for AR(p) and MA(q) we find the last nonzero PACF is p and the last nonzero ACF is q respectively.

Having found the ACF and PACF, we often will want to plot them on a correlogram.

Fitting ARIMA processes

To fit an ARIMA process we must first ensure that the data is stationary. Typically looking at a plot against time is sufficient. If it is not, there are a number of ways to transform it.

- Differencing (although this is implicit in the I)
- Least squares linear regression

- Seasonal differencing (including method of seasonal means)
- Method of moving averages

- Some nonlinear transform (log is common)

Now we have a stationary distribution. Estimate the mean $\hat{\mu} = \frac{1}{n}\sum x_t$ and the auto covariance $\hat{\gamma}_k = \frac{1}{n}\sum_{t=k+1}^{n}(x_t - \hat{\mu})(x_{t-k} - \hat{\mu})$ to give $r_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$. This set is the sample autocorrelation function (SACF). The sample partial autocorrelation functi is worked out just as the PACF was.

If this sample was from white noise, then r_k and ϕ_k are approximately normal with mean 0 and variance $1/n$. If over 1 in 20 values fall outside $\pm 1.96/\sqrt{n}$ then the test fails.

Alternatively, a 'portmanteau' test from Ljung and Box states if the white noise model is correct $n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k} \sim \chi^2_m$.

If the sample was from MA(q), then r_k is close to 0 for $k > q$; indeed these val will be roughly normally distributed with mean 0, variance $\frac{1}{n}\left(1 + \sum_{i=1}^{q}\rho_i^2\right)$. We can use the same test as for white noise. If the sample was from AR(p we use the same test on the PACF.

Box - Jenkins

This is one method for finding a suitable ARIMA process.

- Take differences until the SACF r_k decays rapidly with k. (alternatively, take the number of differences that minimises the sample variance)
- Set $\hat{\mu}$ to be the sample mean.
- If either the SACF or SPACF is close to zero for large k, fit an AR(p) or MA(q) model. Otherwise exhaustively search ARMA(p,q)
- Assume $e_t, e_{t-1}, \ldots$ to be zero and use MLE to pick coefficients. Backforecast using these to get new $e_t, e_{t-1}, \ldots$ and iterate until estimates conve Alternatively calculate the theoretical ACF in terms of parameters and solve for sample values.
- Finally estimate $\hat{\sigma}^2$ using the values generated.

We now diagnose this to ensure it is a good model.

- Visual inspection
- Check $\hat{e}_t, \ldots$ form a good white noise approximation
- Count turning points - we expect $\frac{2}{3}(N-2)$ with variance $(16N-29)/90$

Forecasting is relatively easy. If we know, have estimated or have forecast a given parameter, use that. For others use the expected value (ie $e_{n+1} = 0$). For an ARIMA with I nonzero undo the differencing.

Misc ARIMA.

Exponential smoothing

This is a much simpler method of estimation, using a single parameter α either chosen by the user or estimated from previous data

$$\hat{x}_n = \alpha (x_n + (1-\alpha) x_{n-1} + (1-\alpha)^2 x_{n-2} + \dots)$$

Vector autoregression

This is a time series acting on a vector of values. - instead of coefficients we have matrices, whose eigenvalues must be less than one (although I do not believe this is sufficient)

Most single variable results have appropriate analogues - for example the process is weakly stationary if $E[\underline{X}_t]$ and the covariance matrix $\text{cov}(\underline{X}_t, \underline{X}_{t+k})$ are independant of t.

Cointegrated time series.

Two time series are cointegrated if

- X and Y are $I(1)$ processes
- $\exists (a,b)$ such that $aX + bY$ is stationary.

This happens when one process drives the other, or if both are being driv by the same underlying process

Misc Time Series

Bilinear models $X_n + \alpha(X_{n-1} - \mu) = \mu + e_n + \beta e_{n-1} + b(X_{n-1} - \mu)e_{n-1}$

Exhibits large fluctuations when away from the mean.

Threshold autoregressive $X_n = \mu + \begin{cases} \alpha_1(X_{n-1} - \mu) + e_n & X_{n-1} \leq d \\ \alpha_2(X_{n-1} - \mu) + e_n & X_{n-1} > d \end{cases}$

Random coefficient autoregressive $X_n = \mu + \alpha_n(X_{n-1} - \mu) + e_n$ α_n random.

Generally less regular than AR(1)

ARCH(p) $X_n = \mu + e_n \sqrt{\alpha_0 + \sum \alpha_k (X_{n-k} - \mu)^2}$

The autoregressive with conditional heteroscedasity model gives a period of volatility following a significant change.

Conditional Probability Proofs

Claim $E(E(X|Y)) = E(X)$

Proof
$$E(X) = \sum_x x \, P(X=x)$$
$$= \sum_{x,y} x \, P(X=x|Y=y) P(Y=y)$$
$$= \sum_y E(X|Y) P(Y=y) \quad = E(E(X|Y))$$

Claim $E((X - E(X|Y))^2) \leq E((X - h(Y))^2)$

Proof
$$E((X - h(Y))^2) = E((X - E(X|Y) + E(X|Y) - h(Y))^2)$$
$$= E((X - E(X|Y))^2)$$
$$+ 2E((X - E(X|Y))(E(X|Y) - h(Y)))$$
$$+ E((E(X|Y) - h(Y))^2)$$

But $E((X - E(X|Y))(E(X|Y) - h(Y)))$
$$= E(E((X - E(X|Y))(E(X|Y) - h(Y)) \,|\, Y))$$
$$= (E(X|Y) - h(Y))(E(X|Y) - E(X|Y)) = 0$$

And $E((E(X|Y) - h(Y))^2)$ is positive so the result holds □

Properties of Martingales

We say that Y_n is a martingale with respect to the process X_n if

- $E(|Y_n|) < \infty \quad \forall n$
- $E(Y_{n+1} \mid X_0, \ldots, X_n) = Y_n \quad \forall n$

We say that T is a stopping time for X_n if $I_{\{T=n\}} = f(X_0, \ldots, X_n)$. Thus 'a process reaching its maximum value' would not be a stopping time.

These are used in the optional stopping theorem

Let X_n be a martingale. Then $E[X_T] = E[X_0]$ for the stopping time T if either X is bounded or T is bounded.

If neither are bounded, use the truncated stopping time $\min(T, m)$ and let $m \to \infty$.

One way of using this is to find facts about random walks - given a random walk X_n, both $(\frac{q}{p})^{X_n}$ and $X_n + n(q-p)$ are martingales. Expanding gives the probability of each side being hit and the expected length.

Continuous time martingales have the same properties as discrete time ones, except a filtration F_t replaces the vector $X_1, \ldots, X_n$.

When proving something is a martingale, take $E(Y_{n+1} \mid X_0, \ldots, X_n)$, split Y_{n+1} into $(Y_{n+1} - Y_n) + Y_n$ and fiddle to get rid of the difference.

Markov Chains

Markov chains are fully specified by their initial probability distribution and the transition probabilities. $P(X_n = j \mid X_m = i) = p_{ij}^{(m,n)}$. We will assume time invariance $p_{ij}^{(1)} = p_{ij}^{(m,m+1)}$

The Chapman Kolmogorov equations $p_{ij}^{(m,n)} = \sum_{k \in S} p_{ik}^{(m,l)} p_{kj}^{(l,n)}$ $m \leq l \leq n$ are the driving force behind most markov chain results. To prove them, write the probabilities out in full and use the markov property.

It is clear that the probabilities $p_{ij}^{(1)}$ can be written as a matrix P, and a concequence of probability is that each row sums to one. The CK equations further give us that $p_{ij}^{(k)}$ is represented by the matrix P^k.

Stationary distributions are those that satisfy $X = PX$. If the state space is finite and the chain is irreducible (each state can be reached by each other) then the stationary distribution is unique. Moreover, if it is aperiodic, then all distributions will tend to it.

As implied, markov chains can be infinite. These will typically be simple models like random walks. Random walks have $p_i = A + B^i/N$ if symmetric and $p_i = A + B(q/p)^i$ where A, B must be calculated from boundary conditions

When modelling using markov chains, the best estimate of p_{ij} is $\hat{p}_{ij} = \frac{n_{ij}}{n_i}$. We test the fit either by looking at graphs to see if the shapes match or by sampling triplets n_{ijk} and using the formula

$$\sum_{i,j,k} \frac{(n_{ijk} - n_{ij}\hat{p}_{jk})^2}{n_{ij}\hat{p}_{jk}} \sim \chi^2_{\text{num}(n_{ijk}) - \text{num}(n_{ij}) + \text{num}(n_i) - 1}$$

Poisson Proofs

Let N_t be a process such that

$$P(N_{t+h} - N_t = 0 \mid F_t) = 1 - \lambda h + o(h)$$
$$P(N_{t+h} - N_t = 1 \mid F_t) = \lambda h + o(h)$$
$$P(N_{t+h} - N_t \neq 0,1 \mid F_t) = o(h)$$

Theorem N_t is a poisson variable with mean λt.

Proof If $p_i(t) = P(N_t = 1)$

Then $p_i(t+h) = p_i(t)(1-\lambda h) + p_{i-1}(t)\lambda h + o(h)$

Thus $dp_i/dt = -\lambda p_i(t) + \lambda p_{i-1}(t)$

But this is satisfied by $p_i(t) = \frac{e^{-\lambda t}(\lambda t)^i}{i!}$ □

Theorem $T_0, \ldots, T_j$ is a sequence of independent exponential variables.

Proof
$$P(T_{j+1} > t+s \mid T_j = s) = P(N_{t+s} - N_s = 0 \mid T_j = s)$$
$$= p_i(t)$$
$$= e^{-\lambda t}$$

Repeat a non conditional version for T_0 and we're done □

Markov Jump processes

A markov jump process is a continuous time markov process with a discrete state space.

It has transition probabilities $p_{ij}(t) = P(X_t = j \mid X_0 = i)$, but more useful are the transition rates σ_{ij} given by $p_{ij} = I_{\{i=j\}} + h\sigma_{ij} + o(h)$. The matrix of these is A. Differentiating $\sum p_{ij}(t) = 1$ gives $\sum \sigma_{ij} = 0$, so each row of A must sum to zero.

The Chapman - Kolmogorov $p_{ij}(t+s) = \sum_{k \in S} p_{ik}(s)\, p_{kj}(t)$ hold and lead to the forward differential equation $\frac{d}{dt} P(t) = P(t)A$ and the backward equation $\frac{d}{dt} P(t) = A P(t)$ [which is slightly more useful]

In the time inhomogeneous case, these change to $p_{ij}(s,t) = \sum_{k \in S} p_{ik}(s,u)\, p_{kj}(u,t)$, $\frac{d}{dt} P(s,t) = P(s,t) A(t)$ and $\frac{d}{dt} P(s,t) = A(s) P(s,t)$, we also get

$$p_{ij}(s,t) = \sum_{k \neq i} \int_0^{t-s} e^{\int_s^{s+w} \sigma_{ii}(u)\,du}\, \sigma_{ik}(s+w)\, p_{kj}(s+w,t)\, dw$$

$$= \sum_{k \neq i} \int_0^{t-s} p_{ik}(s, t-w)\, \sigma_{kj}(t-w)\, e^{\int_{t-w}^{t} \sigma_{ii}(u)\,du}\, dw$$

Finally, including duration dependance (and thus no longer having a true markov process), we get (for a well/ill/dead model)

$$p_{SH}(s,t) = \int_s^t e^{-\int_s^v (\rho(u, w-s+u) + \nu(u, w-s+u))\,du}\, \rho(v, w-s+v)\, p_{HH}(v,t)\, dv$$

Splitting Jump Processes

Theorem The first holding time of a markov jump process with rates σ_{ij} is exponentially distributed with parameter $-\sigma_{ii}$.

Proof Let $B_n = \{ X_{\frac{k}{2^n}} = X_0 \quad k = 1, 2, 3, \ldots, 2^n \}$

so $P(B_n \mid X_0 = i) = (1 + \sigma_{ii} \frac{t}{2^n} + o(\frac{1}{2^n}))^{2^n}$

But $P(T_0 > t \mid X_0 = i) = \lim_{n \to \infty} P(B_n \mid X_0 = i)$

$= e^{\sigma_{ii} t}$ □

An immediate concequence is that the mean holding time is $1/-\sigma_{ii}$

By selectively ignoring transitions, we can easily see that the probability of jumping to a state is proportional to the transition rate to that state

Brownian Motion

Brownian motion is the continuous time version of a random walk. It has the following properties

- Independant stationary gaussian increments $[B_t - B_s \sim N(0, t-s)]$
  Assuming the first two, gaussian $\Leftrightarrow$ continuous sample path

- Markov property $\Leftarrow$ independant increments

- $\operatorname{cov}(B_t, B_s) = \operatorname{cov}(B_t, B_t) + \operatorname{cov}(B_s - B_t, B_t)$
  $= t + 0$

- $E(B_t \mid F_t) = E(B_t \mid F_t) + E(B_s - B_t \mid F_t)$
  $= B_t + 0 \Rightarrow$ Martingale
  In addition $B_t^2 - t$ and $e^{\lambda B_t - \frac{1}{2}\lambda^2 t}$ is a martingale

- $B'_t = \frac{1}{\sqrt{c}} B_{ct} \Leftarrow \operatorname{cov}(B'_t, B'_s) = \frac{1}{c}\operatorname{cov}(B_{ct}, B_{cs}) = \frac{c}{c} t = t$
  $B''_t = t\, B_{1/t} \Leftarrow \operatorname{cov}(B''_t, B''_s) = ts \operatorname{cov}(B_{1/t}, B_{1/s}) = ts\, 1/s = t$

One constructs brownian motion from a random walk by taking the limit $k \to \infty$ of $X_t^{(k)} = \sum_{j=1}^{2^k t} Y_j^{(k)}$ with $P(Y_j^{(k)} = \pm \frac{1}{\sqrt{2^k}}) = \frac{1}{2}$. This can be useful because of the functional central limit theorem

Theorem For every continuous functional $\lim_{n\to\infty} [H(X_t^{(n)}, t \geq 0)] = E(H(B_t, t \geq 0))$

The continuous functionals include $\max B_t$ and $\int_0^T B_t$. Another important functional is $E(e^{-\lambda \tau_k}) = \left(2p / e^{\lambda} + \sqrt{e^{2\lambda} - 4p(1-p)}\right)^k$. This gives the result

- $E(e^{-\lambda \tau_k}) = e^{-y\sqrt{2\lambda}}$ — Standard Brownian Motion
- $E(e^{-\lambda \tau_k}) = e^{-y(\mu + \sqrt{2\lambda + \mu^2})}$ — Brownian Motion with Drift
- $P(\tau_y < \infty) = \lim_{\lambda \to 0} E(e^{-\lambda \tau_y}) = e^{-2y\mu}$

Problems using Brownian Motion

Suppose we wish to calculate $P[B_2 < 3 \mid B_1 = 1, B_4 = 4]$. In terms of increments

$$f_{B_2 \mid B_1=1,\, B_4=4} = \frac{f_{B_4-B_2}(4-x)\, f_{B_2-B_1}(x-1)}{f_{B_4-B_1}(4-1)}$$

Since we know the increments are normal $f_{B_2 \mid B_1=1,\, B_4=4} = K e^{-(4-x)^2} e^{-\frac{1}{2}(x-1)^2}$

$$= L e^{-\frac{1}{2}\left(\frac{x-2}{\frac{1}{\sqrt{2}}}\right)^2}$$

So $f_{B_2 \mid B_1=1,\, B_4=4}$ is normal with mean 2 and variance $\frac{1}{1+\sqrt{2}}$

Theorem Let τ be a stopping time for standard Brownian motion where $P(\tau < \infty) = 1$. Then

$$X_t = \begin{cases} B_t & t \leq \tau \\ 2B_\tau - B_t & t > \tau \end{cases}$$

has the same probability distribution as B_t $t \geq 0$.

Let $M_t = \max_{0 \leq s \leq t} (B_s)$. We can calculate this using the reflection principle

$$\begin{aligned} P[M_t \geq y] &= P[M_t \geq y,\, B_t > y] + P[M_t \geq y,\, B_t < y] \\ &= 2P[M_t \geq y,\, B_t > y] \\ &= 2P[B_t > y] \end{aligned}$$

Sending $t \to \infty$ for $P(\tau_y \leq t) = P(M_t \geq y)$ shows $P(\tau_y < \infty) = 1$

Stochastic Calculus

This describes how to integrate functions with respect to brownian motion. The main things we need to know are

- $E\left(\int_0^t f(s)\, dB_s\right) = 0$
- $E\left(\int_0^t f(s)\, dB_s\right)^2 = E\left(\int_0^t f^2(s)\, ds\right)$

- $\int_0^t f(s)\, dB_s$ is a martingale with continuous sample paths.

There is also an important result called Ito's Lemma.

Theorem Let $dX_t = Y_t\, dB_t + Z_t\, dt$ and let $f = f(t, X_t)$. Then

$$df = \frac{\partial f}{\partial x}\, dB_t + \left[\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x} Z_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} Y_t^2\right] dt$$

One example of this in use is with $X_t = B_t$, $f = X_t^2 - t$. Then

$$d(B_t^2 - t) = 2B_t\, dB_t + \left[-1 + \tfrac{1}{2} 2\right] dt = 2B_t\, dB_t$$

Which gives us that $\int B_t\, dB_t = B_t^2 - t$

Diffusion Processes

A time homogeneous diffusion process has the following properties

- It is a markov process
- It has continuous sample paths
- There are functions $\mu(x)$ and $\sigma^2(x) > 0$ such that as $h \to 0^+$

$$E(X_{t+h} - X_t \mid X_t = x) = h\mu(x) + o(h)$$
$$E((X_{t+h} - X_t)^2 \mid X_t = x) = h\sigma^2(x) + o(h)$$
$$E((X_{t+h} - X_t)^3 \mid X_t = x) = o(h)$$

The Ornstein Uhlenbeck process is one example with $\mu = -\gamma x$, $\sigma^2 = \sigma^2$. The density function obeys $\frac{d}{dy}(\mu(y)\pi(y)) = \frac{1}{2}\frac{d^2}{dy^2}(\sigma^2(y)\pi(y))$ which gives the long term distribution $\pi(y) = N(0, \sigma^2/2\gamma)$

Geometric Brownian motion $S_t = S_0 e^{\sigma B_t + \mu t}$ is another example. Solve the expectations to get $\mu(x) = x(\mu + \frac{1}{2}\sigma^2)$ and $\sigma^2(x) = \sigma^2 x^2$.

We can get stochastic differential equations from these by substituting mean and variance into $dX_t = Y_t\, dB_t + Z_t\, dt$ - for Ornstein Uhlenbeck we have $dX_t = \sigma\, dB_t - \gamma X_t\, dt$.

Finally, a Levy process $X_t = \theta t + \sigma B_t + J_t$ is a mixture of discrete and continuous time processes.

Generating Random Numbers

The most successful random number generators are linear congruential generators. These generate sequences of integers using the recursive rule

$$x_n = (a x_{n-1} + c) \pmod m$$

The parameters a, c and m must be carefully chosen. Often $c=0$, $m = 2^{31}-1$. The latter causes problems in 32 bit machines due to overflow, so use $m = aq+r$

$$ax_n \pmod m = a(x_n \pmod q) - r\lfloor x_n/q \rfloor + m\, I_{\{a(x_n \pmod q) < r\lfloor x_n/q \rfloor\}}$$

This gives random numbers uniformly distributed between 0 and $m-1$. There are two methods to get this into a specified distribution.

If the distribution function $F(x)$ is invertible, then $X = F^{-1}(U[0,1])$ has the desired distribution. Unfortunately, often finding F^{-1} is non-trivial.

Another method is to generate a known distribution, then carve out the distribution desired by rejecting a certain proportion of values for each point. Supposing $f(x)$ is the known and $h(x)$ is the desired distribution, for a given value x reject it if $\frac{f(x)}{C h(x)} = g(x) < u \sim U(0,1)$ where $C = \max_x \frac{f(x)}{h(x)}$

Given $u_1, u_2 \sim U(0,1)$ there are two algorithms to find normal distribution

Box Muller $\quad z_1 = \sqrt{-2\ln u_1}\cos(2\pi u_2) \qquad z_2 = \sqrt{-2\ln u_1}\sin(2\pi u_2)$

Polar $\quad$ Set $v_1 = 2u_1 - 1$, $v_2 = 2u_2 - 1$, $s = v_1^2 + v_2^2$ if $s > 1$ restart

$z_1 = \sqrt{-2\ln s / s}\, v_1 \qquad z_2 = \sqrt{-2\ln s / s}\, v_2$

Alternatively, set $z = \sum_1^{12} u_i - 6$ which is approximately normal by the central limit theorem.

Monte Carlo Simulation

The reason we are generating these numbers is for Monte Carlo simulation. Pseudorandom numbers repeatability is especially useful in the following areas.

- Sensativity analysis
- Numerical evaluation of derivatives
- Comparative simulations

Where otherwise the random numbers would distort the results sought.

We usually use Monte Carlo simulation to estimate $E(X)$. That is, we want to ensure that $|\hat{\theta} - \theta| < \varepsilon$ with probability $> 1-\alpha$.

As $\hat{\theta} - \theta \sim N(0, \tau^2/n)$ and we can estimate $\hat{\tau}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\theta})^2$, we can find $z_{\alpha/2}$ as the value satisfying $P[-z_{\alpha/2} \leq N(0,1) \leq z_{\alpha/2}] = 1-\alpha$ and use it to give $n > z_{\alpha/2}^2 \hat{\tau}^2/\varepsilon^2$, so we need that many runs.

Similarly, the relative error gives $n > z_{\alpha/2}^2 \hat{\tau}^2/\varepsilon^2\hat{\theta}^2$.