

**General Insurance Convention,
Blackpool, 18-21 October 2005**

A MATHS TOOLKIT FOR ACTUARIES

PART II

Trevor Maynard (trevor.maynard@lloyds.com)

James Orr (james.orr@lcp.uk.com)

John Berry (john.berry@emb.co.uk)

Andreas Tsanakas (andreas.tsanakas@lloyds.com)

John Harnett (john.harnett@lloyds.com)

Will Bateman (will@codecogs.com)

CONTENTS

1. Foreword
2. Some interesting areas of research
3. The IT toolkit
4. Survey of software used at universities
5.  and why GI Actuaries might need it!
6.  **CODECOGS** and its use within Excel
SHARING IDEAS
A member of Zyba LTD.
7. What next?

1. Foreword

Author: Trevor Maynard

This paper is not finished. Depending on support and interest at GIRO we may widen the target audience to actuaries in all practice areas (we have contacted some like-minded life and pensions actuaries who are interested to take part). But, rather than cautiously waiting for the final product, polished, without flaw or risk of error, we thought we would illustrate further some of the themes raised in Part I to keep the ball rolling. Hopefully there is more to come, in time.

As you can see this “paper” is really a collection of articles written on a number of subjects that relate (in the widest sense) to a maths toolkit for actuaries.

The article in section 2 is written by Andreas Tsanakas and covers the topics of Risk Measures, Copulas and Recursions.

- Those of you involved in ICA work are using risk measures all the time (VaR is a risk measure for example); there is a general theory about these which sets out a framework for comparing them and highlights the good points and bad points about some of them. We suggest that knowledge of the overarching theory would be helpful to actuaries in all practice areas.
- Anyone carrying out Monte Carlo simulations is using a copula, did you know this? It is good to be aware of hidden assumptions and this section is definitely worth a read to get a feel for what copulas are, and to find out where to read more.
- Do we need to do Monte Carlo simulations? Sometimes we do; but there are other methods that in some cases can be quicker and more precise; recursions may be one of those fields that many actuaries are not using day to day (or at least we don't think they are...) because of not knowing about them.

The article in section 3 is what you might call a “meta” level discussion. How can we share information? How can we keep up with the overwhelming amount of information on the internet these days? Where could the maths toolkit be stored? How can it be kept up to date? John Harnett's article isn't about tools in the toolkit, it is about toolkits themselves. Many readers may have heard of Wikipedia? Have you heard of Wiki's as a general concept? If not, take a look at the article as they are a truly modern way of collaborating on work. Lloyd's (where some of the authors work) has started to use them in-house in a couple of departments; the actuarial profession could definitely use them for working parties. If you want to know more about Blogging, RSS feeds and tagging (we advise strongly that you do!) then read on. The article is hosted on a Wiki and the best way to experience it is to take a look in its natural habitat rather than in the two dimensional format in this paper; it is much more enjoyable in hyperspace! We give instructions on how to visit the wiki later on; you can even update the text yourself – an evolving article! So don't be surprised if the version you take a look at doesn't resemble the article in this paper (any graffiti will be deleted unless it is funny).

The 4th section shows the preliminary results of a survey the authors have been doing of British universities and the software they teach to their students. Are you using this software? Have you heard of it?

In the article in section 5, John Berry (who now works for EMB but wrote the article as an MSc student at Oxford University) explains to us how useful the statistical software “R” can be. As he explains R is **FREE** not just to individuals but to business too. It is used by a huge number of statisticians around the world (as evidenced by the survey results in section 4) who continually refine it, for **FREE** (are you getting the point?!). It is easy to download (though your IT department may need to be gently massaged to make it happen) – if you have a

home PC why not take a look first? It produces some world class graphics as standard, Excel just can't compete and it can be linked in with Excel, matlab etc. It is very flexible but the price you pay for this is that it isn't that pretty to look at; some effort will be required to learn how to use it. But what else are you going to do on the train? (OK there are lots of things you could do, but this is certainly a productive use of time).

In the final article (section 6) Will Bateman explains how his company Code Cogs can help with numerical analysis. Code Cogs has already worked with Hat¹ – a consultancy specialising in catastrophe modeling – to help them build an online DFA model, which enables their clients to price complex reinsurance contracts using probability loss distributions from catastrophe models. The article explains what they do better than I can! It illustrates for one how Excel can be customised to make it much more powerful (this functionality will be available shortly after GIRO). A small amount of outlay is required for some of the options (they have to make some money!) but not much.

This note concludes in section 7 with a discussion of how the maths toolkit could be taken forward in the future, this will hopefully form part of the discussion at the GIRO workshop.

¹ www.hat-consulting.com.

2. Some interesting areas of (academic actuarial) research

Author: Dr Andreas Tsanakas

In this note three areas of academic research are briefly discussed, that are of potential interest to practising actuaries. The selection of topics is dictated primarily by the author's own research interests.

2.1 Risk measures

A risk measure is a function that maps any portfolio of random assets and liabilities to a corresponding level of capital required to support it. Risk measures were introduced in the actuarial literature in the 1970s under the name 'premium calculation principles'. Though they were initially meant as insurance pricing tools, they entered the consciousness of practising actuaries for good with the advent of risk sensitive regulation. While pricing and setting risk capital are very different exercises, their respective mathematics are quite similar. So while the percentile-VaR is currently the widest known risk measure, the simple (re)insurance pricing rule "mean plus a standard deviations" is certainly another.

When deciding which risk measure to use, for either pricing or capitalisation purposes, the properties of alternative risk measures have to be considered. This has until recently been a fairly academic debate, initially occupied with the formal characterisation of different risk measures via their sets of properties. More recently the debate tended to focus on the weaknesses of the percentile-based VaR. These weaknesses, stemming from VaR's 'blindness' with respect to the extreme tails of probability distributions, are meanwhile better understood by the actuarial community. It is for example the case that VaR has inconsistent diversification properties, especially in the case of aggregating high reinsurance layers.

Tail-VaR (also called C[onditional]-VaR or Expected Shortfall or T[ail]C[onditional]E[xpectation]) has been proposed as a corrective to the weaknesses of VaR. While Tail-VaR is effective as an alternative to VaR, the issue of risk measures' properties is a much wider one. Different risk measures (with their respective sets of properties) are appropriate for different problems – questions one could ask are:

- Is the risk measure used in capital setting, pricing or performance measuring? Different purposes may require focus on different areas of the probability distribution.
- Should the risk measure be indifferent to the aggregation of risk, i.e. is there always a diversification benefit from pooling risks? If aggregation is an issue one should be aware that both VaR and TailVaR fail to recognise this.
- If the risk measure is going to be used for the rebalancing of portfolios, do we understand the dynamics of that process? For example how does the optimal portfolio behave in the presence of unhedgeable (background) risk?

This is naturally far from exhaustive.

Some links:

<http://www.casact.org/library/astin/vol26no1/71.pdf>

<http://www.ingentaconnect.com/content/fia/baj/2003/00000009/00000004/art00009>

<http://www.wiley.co.uk/eoas/pdfs/TAP027-.pdf>

<http://www.econ.kuleuven.be/tew/academic/actuawet/pdfs/RiskShorti.pdf>

2.2 Copulas

Copulas are functions summarising the dependence structure between risks. The term 'copula' has become a buzzword in actuarial circles in the last few years. Again, capital setting has provided much of the motivation for the concept of copulas becoming wider known (if not necessarily understood). It is well understood that dependency between risks makes portfolios more variable and therefore risky. Specifically dependency in the area of extreme losses can have a profound effect on, say, the VaR of the aggregate portfolio. In other words, if someone has modelled dependencies in the context of a capitalisation project, they will have noticed that copulas amount to more than an academic construct. They cost money.

Copulas are simple abstractions from the joint probability distribution between risks. They enable the separation of the dependency structure from the marginal behaviour, that is, from the probability distributions of the individual risks. While the concept of the copula is not particularly difficult to explain, a fair amount of mystique around copulas persists. A usual misconception is that a copula is an exotic add-on to any modelling exercise. In reality, any multivariate risk model, however constructed has a copula, regardless of whether that copula is explicitly defined. In whatever way the dependency structure has been constructed (explicit copula assumption, correlation matrix, risk drivers or just plain independence) it corresponds to a copula. If you want to understand your portfolio, try understanding your copula.

The most widely used copula is the one corresponding to independent risks. If one wants to introduce correlations, the Gaussian copula is usually the obvious choice², as it allows a correlation matrix specification. The Gaussian copula corresponds to the dependence structure of a multivariate normal distribution. As such it inherits the multi-normal's tractability, but also its asymptotic independence properties for the extreme tails. The latter can be a significant weakness when joint extreme losses are of interest. Generalisations of the Gaussian copula, such as the t-copula, to some extent correct this. Another useful copula for the modelling of extremal dependencies is the Gumbel copula. This has been popular because of its highly conservative properties, but is rather inflexible and not always easy to work with.

Some links:

www.casact.org/pubs/proceed/proceed98/980848.pdf

<http://www.risklab.ch/ftp/papers/DependenceWithCopulas.pdf>

<http://www.math.ethz.ch/%7Estrauman/preprints/pitfalls.pdf>

http://library.soa.org/library/naaj/1997-09/naaj9801_1.pdf

² It is that obvious in fact, that often one omits the copula specification and just relates the correlation matrix. This relates to the popular misconception that copulas are in some sense antithetical to the use of correlations as dependency metrics. Any portfolio has both a copula and a correlation matrix, with the latter effectively being a property of the former.

2.3 Recursions

When modelling a (re)insurance portfolio, a typical challenge is the determination of the aggregate loss distribution, that is, the probability distribution of the sum of all risks contained in the portfolio. It is usually not feasible to determine such aggregate distributions by hand and a numerical method has to be adopted.

The wide availability of software implementing Monte-Carlo simulation based methods, combined with Monte-Carlo's great flexibility and relative freedom from the 'curse of dimensionality', has made simulation the standard approach for evaluating aggregate loss distributions. There are however situations where alternative algorithms can be applied, which often significantly outperform Monte-Carlo. One group of such algorithms is known under the banner of 'recursions'.

The best known recursive algorithm is by Panjer and calculates the aggregate probability distribution in the collective risk model. The collective risk model considers portfolios exposed to a random number of losses (frequency), with each loss following the same probability distribution (severity). The collective risk model is commonly used for the modelling of an insurance company's exposure to large losses and for the modelling of reinsurance portfolios. Panjer's algorithm is quite fast (generally faster than Monte-Carlo) and can be generalised to include features such as XL reinsurance layers and reinstatements.

A different, lesser known, family of recursions are used to solve the 'individual risk model'. The individual risk model considers a portfolio consisting of a fixed number of risks, each with a probability of producing a loss and a loss distribution conditional on the loss occurring. The loss probabilities and conditional distributions can be different for each risk. A variety of algorithms exists for the calculation of the aggregate loss distribution in the individual loss model, with the ones by De Pril and Dhaene & Vandebroek being among the better known. The speed of these algorithms varies; they tend to perform better when the probabilities of a loss occurring are low³. Sometimes the collective risk model is used to approximate the individual one, so as to take advantage of the speed of Panjer recursions.

It is noted that the above-mentioned recursive methods rely on the assumption of independence between risks. While some efforts have been made for introducing dependencies, this is very much still an open field for research. We note that other non-Monte-Carlo methods for the aggregation of loss distributions exist, such as the Fast Fourier Transform and the Heckman-Meyers method. These tend to require more complex programming, but are more amenable to the introduction of dependencies.

Links:

<http://www.casact.org/library/astin/vol12no1/22.pdf>

www.casact.org/pubs/proceed/proceed98/980848.pdf

<http://www.casact.org/library/astin/vol30no2/349.pdf>

[http://www.econ.kuleuven.ac.be/tew/academic/actuawet/pdfs/DV\(recur\).pdf](http://www.econ.kuleuven.ac.be/tew/academic/actuawet/pdfs/DV(recur).pdf)

³ In fact, while recursive methods perform best for low loss frequencies, these are exactly the situations where Monte-Carlo simulation becomes inefficient. Hence, rather than viewing recursive and Monte-Carlo based methods as antagonistic, one could consider them complementary.

3. IT - toolkit

Author: John Harnett

If you are reading this off a hard copy then you won't be getting the most out of this page (but please keep reading!). This short note attempts to illustrate some IT methods that may be a great tool in the actuary's toolkit. The note is written in the way it is meant to be used (i.e. with lots of embedded links), it is also hosted on a "Wiki" (see below). I suggest that you visit the wiki to experience this as it was intended. You can even update the Wiki to reflect your own views on anything below, and this is greatly encouraged!!

Text in **bold** shows where a hyperlink is embedded in the Wiki.

3.1 How to visit

If you want to just read the article and follow the links the just follow this link <http://jamestrevorjohn.pbwiki.com/> and you're in....

If you would like to add your own comments to the wiki then just click "Edit" (once you are in the wiki); you will be asked for:

- User Name - Just put in your name
- Email address - Put in your email address
- Password: GIRO - All capital letters

3.2 Overview

For knowledge based professions the ability to continue to learn is linked to the accessibility of relevant information. The services offered by the internet are continually evolving. It is interesting to look at the services that may play in CPD

Most internet users still treat the web as a read-only medium. **However, it was not conceived in that way.** The common read-only web meme is founded on the concepts of "Search and Browse". Web 2.0, as some have named it, adds "Subscription" to this picture. With Subscription, the user starts to automate the process of information gathering that the internet greatly simplified. Subscription begins to enable staying on top of what's new in areas that are of interest to the user.

Given the nature of internet services (often described as "small pieces that are loosely coupled"), many are involved in the discussion of **Web2.0**. For an overview of Web2.0 more comprehensive than the one I attempt to give here check out **this** article (if you are reading this from a hard copy of course you cant!). It's more comprehensive and a better read than this one, don't worry about the jargon, there are parts that will resonate with you.

For the layman there follows a brief introduction to some of the internet's newest services to the user. We talk about Blogging, wiki's, Tagging, and RSS; terms that may seem strange to the uninitiated. But remember the time when you hadn't heard of DVD's, email, Google, Yahoo! and even emoticons.

3.3 Blogging

Personal publishing to the internet has taken off because it's become simple and cheap/free. The act of personal web publishing is most commonly called blogging. The "right" blogs are compelling, informative and written with a voice that we recognise as authentic. They compare favourably with the company brochures transferred to the web or online banking. Early adopters from IT communities took up publishing to the web very quickly. For Actuaries, blogging is a means of having a conversation with their own and other communities. The personal nature of the blog enables discovery of new services and content from a source that is trusted by the subscriber. The way people find out about these services has spread beyond the search model offered by Google. Finding new stuff now includes the "word of mouth" provided by the bloggers.

Search services are being developed that add richness to the blogging ecosystem (or blogosphere). In the way that the google search engine provided an index for the static webpage version of the internet, new services are supporting the conversations that are taking place on blogs. What has this to do with CPD? Well blogs are used for many things but one of their functions is to support community learning and collaboration. They are achieving this through some interesting techniques.

This chap (again, click on the link to find out who I am talking about) identifies 3 cornerstones of participation in blogging. I would add a zeroth pillar, the link. When you find some content that you think will be of interest to your readers, link to it. When authoring web based content, creating a link is simple. It is realising and measuring the value of the link that has powered **Google**.

3.4 BlogSearch

Most knowledge workers are aware of the Google search engine. And those that aren't have people that work for them that are dependent on it. The **google-ranking** of a site, whilst still important, is not the be all and end all of searching. Other offerings, servicing niche markets, are finding their place. The search engine **Technorati** searches over 15.4 million blogs (as of 18 August) and provides "a real-time search engine that keeps track of what is going on in the **blogosphere** — the world of weblogs.". Not that technorati is the only blog search engine. We're in a period of pre-cambrian natural selection relating to **blog search**. **Google are in the blog search space too.**

3.5 RSS Feeds

An enabling feature of blogs is that they can be subscribed to, not via submission of subscribers email, but through an anonymous mechanism that informs the subscriber of change.

So lets say an actuarial expert starts a blog and produces content that makes readers want to find out more and new information. Let's say the blogger produces an ongoing set of thoughts on a Maths toolkit for actuaries. Now occasionally there will be something new that that blogger wants everyone to know about. The heavyweight solution to this is have the user's submit email addresses and for the blogger to manage another mini-spamming industry. The solution that is driving the blogosphere forward is the idea of subscribing to content via a technology called "RSS" (Really Simple Syndication). This is really only another internet file format. But that file format describes the "state" of a site. It enables anonymous subscription to sources around the internet. And it is **already widespread**. And **simple to deploy**. RSS already enables the user to stay on top of their internet sources. Users can subscribe to many different RSS sources through an application called an RSS Aggregator. This program or service shows the user what they have yet to read. In this way it is possible to create your own personalised "newspaper" filled with just the content you are interested in and from the sources you know to be reliable/useful.

3.6 Tagging

Tagging is the latest attempt to categorise web content. Its strengths based on the scale of the internet and the need to find methods of categorisation of data that are not based on physical characteristics of the medium. Tags are metadata about a tag that describe the content of the data. Tags can be added by the author of a page or by the reader of a page. In aggregate, and on an internet scale, tags assist in the categorisation of web data. Here's a link to a good attempt at a beginner's **tutorial**. Tags are something that become easier to understand when used and added to. **Actuarial tags are already out there!** These represent channels for finding new data. The **first** shows how he uses tags within the social bookmarking service del.icio.us. He uses some IT only jargon in there. But if you can just imagine him talking about Maths or Statistics instead of Healthcare, the benefits of operating on an internet scale should be apparent. Make sure you have your speaker or headphones on!

3.7 Wikis and Wikipedia - an Example of Radical Trust

The online encyclopedia "wikipedia" includes many **articles** relevant to **Actuaries**. **Wikipedia** is a free-content encyclopedia, written collaboratively by people from around the world. Wikipedia is an example of a wiki. Wiki is Hawaiian for "fast". It's a web site you, the reader, can edit. This leads to a coherence and structure that is a useful resource.

In an 8.5 minute screencast you can see a demonstration of the **development** of a page in **Wikipedia**. Wiki's are defined **here**. Regardless of the content, the impressive and affirming thing about wikipedia is that in under 5 years it's got a wealth of **coherent content**.

3.8 In summary

It's useful to remember that all the content in wikipedia can be

- changed by the reader,
- "read" by the user,
- linked to,
- tagged,
- and blogged about.

Blogs can be subscribed to by people looking for guidance (aka CPD) in their profession.

Staying on top of new information in the topics you are interested in is greatly facilitated by RSS Subscription. It's being **pushed** by mainstream **organisations**. The **beeb** talk about blogs and techorati.

Finding out about new information is aided by technologies such as RSS for subscription. Social bookmarks (tagging) are another means to determining what's out there. Wiki's work because our constructive instincts outnumber our destructive ones. They provide a way to openly collaborate and continue to learn. See the following link for my subscriptions:

<http://www.bloglines.com/public/sebastiansuncle>

Here's a link to an article describing, in fairly esoteric language, the components of Web2.0. It's particularly interesting in that it has a good graphic of Web2.0's features.

4 Survey of software used at universities

Author: Trevor Maynard

4.1 University questionnaire

We emailed some of our contacts at universities and asked the following questions:

- What are the main software packages that you use in your undergraduate and post-graduate training? Please include a short description, if it is not a mainstream package; an internet link would be very useful!
- What other software packages do you use in your research, which should be of interest to actuaries?
- Do you have any further thoughts on what else might be in a "Maths Toolkit for Actuaries"?

We have had some responses from Oxford, City, Heriot Watt, Imperial, Kings, Warwick and UCD and the results so far are shown below. The questionnaires were completed by individuals and probably largely reflect their own views only.

	system	Website	Use
Used in teaching	R Maple Winbugs S Plus OX Mathematica MuPAD Minitab	www.oxmetrics.net/index.html http://www.wolfram.com/ http://research.mupad.de/home.html	Stats Stats Stats Finance
Used in research	Many of the above PAJEK C++ (MSc and above) TeX	www.vlado.fmf.uni-lj.si/pub/networks/pajek/ http://tug.ctan.org/	network visualisation tool Writing mathematical papers
Other of interest	Many of the above Past papers on Matlab	http://xxx.lanl.gov	preprint server

I suspect that many of the systems above are used in both of teaching and research in practice – this was the case in the responses we received; I have shown each system once only, the allocation is somewhat arbitrary.

Gesine Reinert from Oxford suggested the Pajek system and commented:

"A network approach could be useful to understand how companies are inter-related, say, via management board memberships. Such networks might then be used to explain the creation and the propagation of policies and strategic decisions."

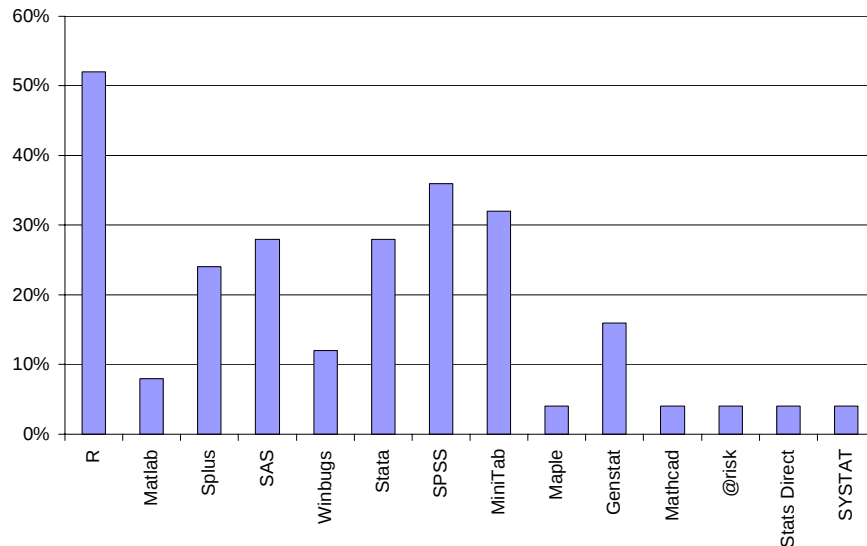
Richard Verrall from CASS business school (CITY) commented:

"Ox is much liked by people in econometrics and has replaced Gauss"

Robin Reed from Warwick made the point that spreadsheets are easy to use to get an answer but it may not be the right answer and more structured languages can often cut down the risk of error. He suggested the following excellent link; it may seem a bit anti-microsoft but makes some good points: http://www.burns-stat.com/pages/Tutor/spreadsheet_addiction.html

4.2 Statisticians

We also emailed the AllStats mailing list⁴ which is subscribed to by statisticians; so the software is likely to be biased towards statistical applications. We've had 25 responses so far, from a whole range of disciplines, though this is only a small percentage of the subscribers. Preferences from this group are



Clearly R is a great favourite, with over 50% using it. Many of those returning the questionnaire added comments and some of the things they had to say about R were:

I always use R unless I'm forced to do otherwise.

I could not be more satisfied with a statistical software package.

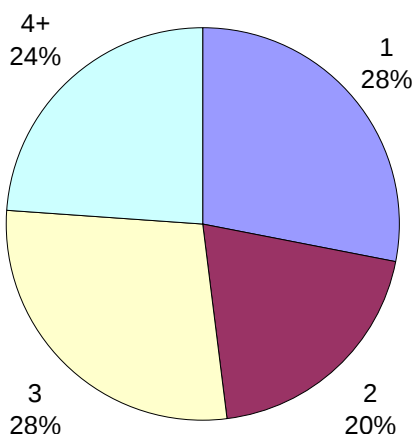
I use R for the tough stuff

They appear to like it! One other very telling favourable comment for R was that, despite it being freeware, it is better supported than many proprietary systems. This is because, a bit like a wiki (see John Harnett's article in section 3), users are encouraged to produce and share bug fixes and write new functionality. Another example of the global mind at work. Actuaries need to be a part of this, surely?

⁴ those of you interested in maintaining your toolkit may be interested to join this, if so use this link <http://www.ltsn.gla.ac.uk/allstat/> and follow the instructions

The next most used system is SPSS (the Statistical Package for Social Sciences), this does not appear to be by choice in many cases and some comments were not favourable. Matlab was used “when the problem involves some heavy duty maths”. SAS was described to be “good for manipulating large data sets and multiple files”. R does all its manipulations in memory, but with RAM being so cheap these days this is hardly a constraint for much of the data we encounter.

Another insight from the survey is that many statisticians use *several* of the software choices day to day. From the responses received we find 24% of them use 4 or more systems (see chart below).



4.3 Summary

Clearly there are a lot of systems out there. But there is also a clear favourite, R (if you note that Splus is a commercial application of the S language and R is a freeware version of the same language then even more people are using this underlying language). The fact that R is free may not be a coincidence; the fact that it fits with the modern collaborative use of the internet (i.e. is open source) is surely not a coincidence.

The table below shows you where you can find out more about some of the software above.

R	http://www.r-project.org/
Matlab	http://www.mathworks.com/
Splus	http://www.insightful.com/products/splus/default.asp
SAS	www.sas.com
Winbugs	http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml
Stata	www.stata.com
SPSS	www.spss.com
MiniTab	
Maple	http://www.maplesoft.com/
Genstat	www.vsn-intl.com
Mathcad	
@risk	www.palisade-europe.com/
Stats Direct	http://www.statsdirect.com/

The next article gives some further insights into why R is such an excellent package.

5. (and why GI Actuaries might need it!)

Author: John Berry, MSc. student in Applied Statistics, University of Oxford (now at EMB)

5.1 What is R?

R is a language and environment for statistical computing and graphics. It is *hugely* powerful. It can be downloaded from <http://www.r-project.org>.

5.2 How much will it cost me?

Nothing. R is released under the [GNU General Public License \(GPL\)](http://www.gnu.org/licenses/gpl.html). The following is a quote taken from <http://www.r-project.org>.

"It is the opinion of the R Core Team that one can use R for commercial purposes (e.g., in business or in consulting)."

Speak to your legal department before using it for commercial purposes. Check licenses of specific add-on packages before using them for commercial benefit.

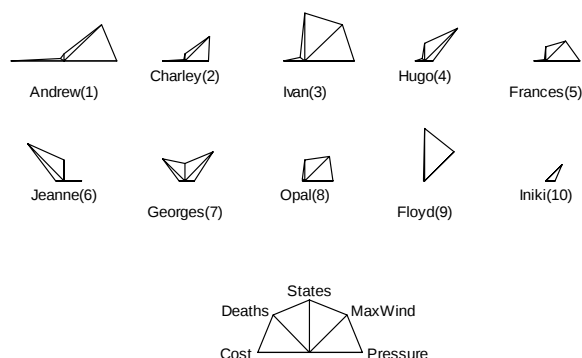
5.3 Who uses it?

Academics, Banks, Hedge Funds, Government, Pharmaceuticals, Marketing firms.

The following examples will demonstrate the graphical capability of R. In all cases, the user has a high level of control over the display.

Example 1 (trivial)

Graphical display of the ten most costly⁵ hurricanes since 1980 (and before 2005!!). Known as a 'stars' plot. Numbers in brackets are ranks by cost.

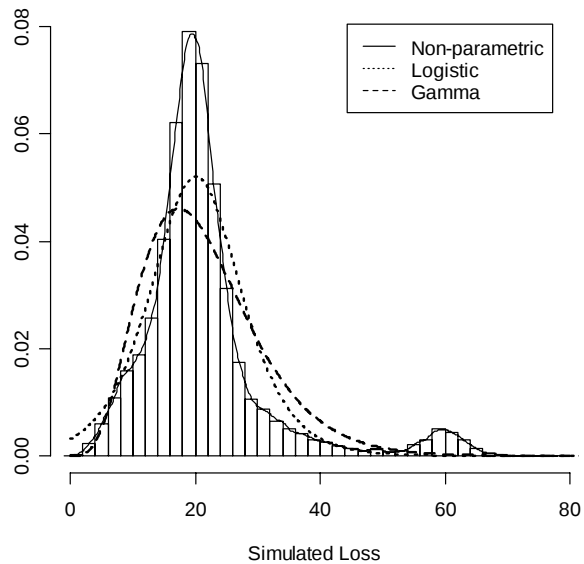


Very useful for multivariate data with up to about 50 individuals and perhaps 8-10 variables.

⁵ data for illustration only, from <http://en.wikipedia.org>

Example 2 (almost trivial)

This shows the dangers of assuming a standard parametric form for loss data which come from a mixture distribution. The data are simulated from a mixture of two gammas and a normal distribution. Superimposed are a gamma density and a logistic density fit by maximum likelihood. The fitting of the gamma density is upset by the weight in the tail of the simulated data.



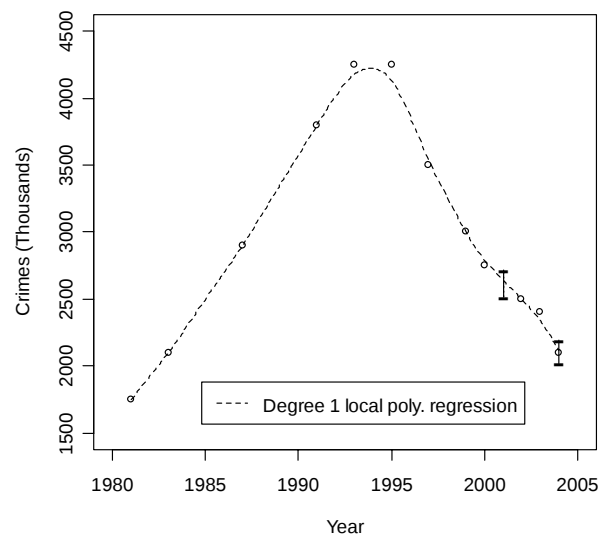
A non-parametric kernel density estimate is shown to provide a much improved fit. It took five minutes to perform this exercise from start to finish.

Example 3 (perhaps non-trivial)

Suppose a motor insurer has an interest in the number of *actual* car thefts in the UK in recent years. Suppose further that she is unwilling to accept figures based on the number of thefts *reported* to the police. Instead, she decides to use the British Crime Survey⁶ as a basis for her analysis. The survey interviews individuals in an attempt to obtain a better picture of the number of *actual* thefts of cars. A simple analysis would tabulate or plot the data against the year in question and use these methods as a basis for inference.

Suppose however that the insurer believes that there may be some estimation error in the figures compiled by the BCS. Perhaps she wishes to obtain a 95% confidence interval for the actual number of cars stolen in, say, 2001 and 2004.

A bootstrap method for obtaining a confidence interval from such data is described in Davison and Hinkley (1997). The following plot displays the confidence intervals along with a local polynomial regression.



⁶ www.crimestatistics.org.uk

5.4 Getting Started

Although very versatile, it takes a little patience to get to grips with R. When compared with the off-the-shelf statistical packages it does not offer a particularly user-friendly interface. As such, it is well worth using an introductory text as a reference. Reference II is probably the most suitable in this regard.

Once the initial familiarisation stage is over, the real benefits of R come from the analyst having more control over the analyses and the format of their output. The user is free to write code to run in the R environment, or he/she can download packages (libraries) to perform specific tasks.

5.5 Technical Notes

The R language is very similar to the S language originally developed at Bell Laboratories (formerly AT&T, now Lucent Technologies). Code written in S will usually run unaltered in R. Reference II explicitly notes areas where R will produce different output.

A plug-in by Erich Neuwirth is available which provides an interface between Excel and R, so that statistical analysis can be performed in R and returned to an Excel spreadsheet. This requires the DCOM interface to be installed. The plug-in brings the advantage that the 'R-literate analyst' can produce spreadsheets for use by 'non-R-literate' analysts.

There is a package RODBC which provides a link to data sources which support an ODBC interface. In addition, there are substantial help files for importing data from various other sources.

5.6 References

- I. R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- II. Venables, W.N. and Ripley, B.D. (2003). Modern Applied Statistics with S. Springer, New York.
- III. Davison, A.C. and Hinkley, D.V. (1997). Bootstrap Methods and Their Applications. Cambridge University Press, Cambridge.

The R libraries locpoly, MASS, and boot were used in the production of this document.

The author can be contacted at John.Berry@emb.co.uk

6. An Introduction to CodeCogs.com and its use within Excel

Author: Dr Will Bateman of Zyba Ltd

6.1 In Brief

CodeCogs (www.codecogs.com) is a scientific numerical library of C/C++ code that has been created for technical users. It is in every sense an “open” system, which gives users full access to the raw source code of anything they licence, while also allowing anyone to submit new or improved solutions on their terms (licensing and price).

Currently the library has over 500 numerical functions, approximately 200 of which are in statistics and 50 in finance. The majority of these can be used with Excel, indeed all Excel functions are also provided as C/C++ equivalents within the library. Shortly you will be able to use these functions with VB, C#, R, PHP and Fortran.

The CodeCogs system is designed to be transparent and quick to use. Considerable effort has been invested in a documentation wizard to facilitate the creation of high quality documentation, with the aim of making the site a technical reference as well as a code library.

6.2 Motivation

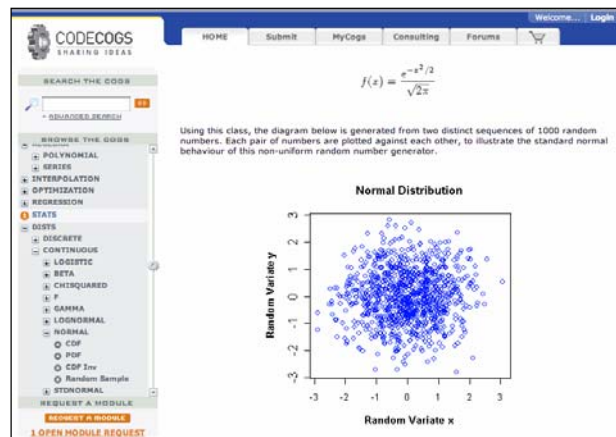
CodeCogs was designed in response to the problems its founders had in finding well-documented, easily-navigable and affordable numerical C/C++ code. In particular we wanted high-quality software that could be used by both commerce (where it can be incorporated into proprietary software with commercial support) and in open-source software systems. Furthermore, we wanted to be able to download only the code we needed, at the point of use, rather than an entire library.

6.3 Pricing

There is no fixed price structure, with the cost of commercial licences set by the contributing developers. As such, almost 80% of all components on CodeCogs are free to use under the GNU GPL licence, with most functions being available under a commercial licence for an average of only £2. Excel integration for a function costs an additional £1.

6.4 Navigating CodeCogs

On entry to the website you are immediately presented with a browser on the left side that contains a complete list of all published modules. Along the top are the key section tags, which allow you to make new submissions, view your own activities or enter a discussion through the forum. The central area of the site is dedicated to displaying the documentation and code.



At the top of each module page is a dropdown menu listing all dependencies required by each module. To the right of this is the licensing area, which summarises the licences you own and the corresponding download options available.



The source code behind a module can either be viewed on screen; downloaded as a single file or compressed zip (which also contains the complete directory paths) or, when available, downloaded as an Excel extension.

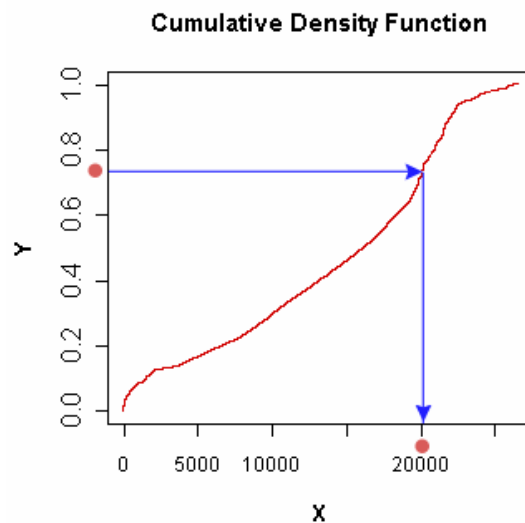
6.5 Installing a module from CodeCogs into Excel

There are two quick steps to install a module from CodeCogs into Excel. First download the extension (as described above). Next open Microsoft Excel with your chosen workbook and click on the **Add-Ins** option located within the **Tools** menu. Finally choose **Browse**, locate the extension (.xll) file that you downloaded from CodeCogs and select it to install.

Example: Simulating Losses to a Treaty

Suppose you want to simulate the payout on an insurance treaty due to a sequence of randomly generated events. Using your historical data (perhaps previous payouts) and other observations, you compiled a list of independent events with a particular return period (or rate) and a mean cost/loss. There are many systems that may fit this model, including losses from river flooding, hurricanes, or even ships lost at sea. All can be characterised by a rate and associated with a loss.

Next you assume that the Poisson distribution is representative of the number of events you expect in any given period (say 1 year). So in your model, we'll use a randomly generated Poisson⁷ number to give you the number of events in each year you simulate. For each of these events, you now randomly select events from your event list, in proportion to their rate. Obviously events with a high rate of occurrence should be selected more often, so we use a Discrete⁸ random generator, which essentially converts a uniformly generated number from 0-1 (Y) into a corresponding event number (X), using the normalised cumulative rates from the rates you've supplied (see illustration below).



⁷ Available from CodeCogs.com in section **stats > dists > discrete > poisson > RandomSample**

⁸ Available from CodeCogs.com in section **stats > dists > discrete > discrete > RandomSample**

With the main numerical calculations being handled by the CodeCogs add-ins, setting up an efficient system within Excel is comparatively simple:

- Download and install the random sampling routines you need from CodeCogs.
- Define your event sets, with rates and losses – for example use columns A and B, respectively, and we'll assume you have 100 unique events.
- Calculate the total rate by entering “=sum(A1:A100)” into cell C1.
- In D1 calculate the number of events in the first simulated year, with “=cc_Poission(\$C\$1)”
- In E1 enter “=if(\$D\$1>(column(E1)-5), index(\$B\$1:\$B\$100,cc_Discrete(\$A\$1:\$A\$100)),”)”. The calculated value is the loss from a newly sampled event.
- Repeatedly copy the last equation into cells F1, G1, H1, I1, J1, etc, up to the maximum number of events you may possibly expect in any one year.
- Add your treaty structure and compute the losses due to the events in cells F1 to J1.
- Copy everything from cell D1 to J1+ “treaty calculations”, downwards, creating a new row for each simulation you require.
- Finally sum up the treaty losses across all simulations to calculate the overall simulated loss of your insurance treaty.

You can rapidly expand this basic model to include many other alterations. For example, if you have an idea about the variance in your loss estimates for each event, then you can add randomness about the mean - perhaps through the use of normal or beta random sampling function.

6.6 Summary

Adding CodeCogs functions to Excel (available shortly after GIRO), means you get the ease of use and flexibility of Excel, combined with the performance and accuracy of the CodeCogs software.

In addition, you get a much broader range of mathematical, statistical and engineering methods at your disposal.

CodeCogs have also emulated all the standard Excel functions in C/C++. So if you decide to turn your Excel solution into a standalone package, then you can download all the source code behind any CodeCogs or Excel functions that you use.

Thanks to Lucian Bentea for producing the figures and setting up the example Excel sheet. A copy of this can be downloaded from: www.codecogs.com/pages/excel/example1.html.

The author may be contacted at will@codecogs.com

7. What next?

Author: Trevor Maynard

Many of you will have read the strategy review of the actuarial profession. It is our hope that we will opt for a hybrid of strategies 2 and 3; a sort of international careers option. This will fit well with the opinions expressed in this paper and Part I. There is an excellent phrase in the paper "life long learning" which is used in place of CPD. Expressed in this way CPD seems much more fun! It also seems more natural. Most of the actuaries I have met are genuinely interested in their work and many like to understand the mathematical/scientific fundamentals. The career options (aka life long learning) for future strategy seem to be aligned with the inherent nature of the actuary and we urge you to show your support for these options on-line.

Some options for taking this forward follow:

- **Do nothing; let it evolve** – it is possible that wikis, bloggs, R discussion groups will spontaneously form due to the enthusiasm generated by this paper! It is a nice thought, but, in our experience, some structure is required; at least initially. Lets see.
- **Continue with the working party** - we could continue with the working party; hopefully enlist some new members and aim to produce an actual toolkit next year.
- **Set up a wiki** – we could set up an actuarial toolkit wiki. As we urged in part I, we hope this would be a conduit for links to other sites etc. We don't want to reinvent the wheel. This should be open to all; subscribers would be encouraged to link to examples of their tools (e.g. vb code, R workspaces etc) and interesting articles they have read. We could contact the universities and encourage them to participate. Maybe such forums exist already, in which case we could join in.
- **SIAS paper or equivalent** – since the authors found out we have to personally pay for printing of SIAS papers we aren't so keen! However some sort of electronic version of the paper could be produced and distributed to the whole profession. We could aim to present it at Staple Inn in the new year.

It would be great to hear your views at the workshop and any other suggestions. We think the working party should continue and set up a wiki as a forum; working party members will be encouraged to Blogg!

A colleague of ours, Markus Gesmann is planning to write a longer paper on the use of R in insurance which is due to be completed shortly before Christmas. We will email links to this once it is available. Perhaps one task for the working party could be to find out a bit more about each of the software choices in section 4 and consider whether they would be useful for actuaries?