m5f22ins.tex

Chapter VII. INSURANCE MATHEMATICS

I. Insurance Background

The idea of insurance is simple: it is the *spreading*, or *pooling*, of risk. The relevant theory is that of *collective risk*.

History.

Insurance can be traced back to antiquity (Greek and Roman times). Like much else, it disappeared, to be re-developed in Renaissance Italy (Genoa, 14th C.). It received a great impetus in the UK from the Great Fire of London in 1666; fire insurance had started there by 1681. Property insurance had begun in London by 1710, and in Philadelphia (Benjamin Franklin) in 1752.

Shipping insurance grew in London around Edward Lloyd's coffee house in the 1680s. He died in 1710; Lloyd's of London had developed by 1774.

John Graunt (1620-74) published his *Bills of Mortality* in 1662 (breaking down London deaths by cause, age etc.). This was followed by the first life table (Edmund Halley, 1693). Mutual life insurance had begun by 1762. One of the earliest such companies is Scottish Widows (1815) (founded to look after the widows of Presbyterian ministers who died in office, and had to leave the manse – the minister's house).

At a national level, national insurance began in Germany with Bismarck in the 1880s. It developed here with e.g. Lloyd George (pre-WWI), Beveridge and the Beveridge Report (1942), and the founding of the Welfare State post-WWII.

Limited liability.

Lloyd's of London pre-dates limited liability (which developed in the mid-19th C.). The Lloyd's participants, or *names*, had unlimited liability, and were liable for the full extent of losses, irrespective of their investment or their assets. This changed, following the Lloyd's scandal of the 1990s.

Insurance is now done (and most was before the Lloyd's scandal) by limited liability companies. So for these, the possibility of *ruin* is crucial. Not only would this wipe out the company, its assets and expertise, the jobs of its employees etc., but it would leave policy-holders without cover.

Reinsurance.

Because a run of large claims could bankrupt an insurance company, companies seek to lay off large risks – to reinsure – insure themselves – with larger, specialist reinsurance companies.

The question arises as to where reinsurane companies re-reinsure themselves ... This raises the modern form of Juvenal's question (*Satires*, c. 80 AD): *Quis custodiet ipsos custodes* – Who guards the guards? Who polices the police? Reinsurers reinsure insurers, but – who reinsures the reinsurers? – etc.

Regulation.

It is in the interest of some industries to agree to cover each other's liabilities in the event of a bankruptcy. For instance, this happens with *travel firms*. If a travel firm goes bust, leaving large numbers of people stranded abroad, or unable to travel on a foreign holiday booked and paid for, this would destroy public confidence in the whole industry – *unless* other firms, by prior agreement, step in to cover. This is what happens, and works well.

As motor insurance is compulsory by law, motor insurance companies are regulated by the state, and again, this provides a degree of protection in case of bankruptcy.

The actuarial profession.

People involved in the insurance industry have been known as *actuaries* from the early days of insurance. Companies offering insurance employ actuaries, and these need to be qualified. Actuaries become qualified by passing exams set by the *Institute of Actuaries*. London is an important centre for the actuarial/insurance industry, and so is Edinburgh. The mathematics involved is interesting, and useful. Those taking this course would be well advised to consider an actuarial career as one of their career possibilities.

Life v. non-life.

The usual way the modern insurance industry splits is between life and non-life. Life insurance is payable on death, and/or as an annuity ceasing on death. Life insurance is often combined with a mortgage (so that the mortgage is paid if one dies before it expires). Naturally, assessing premiums here depends on a detailed knowledge of mortality rates over ages, etc. The relevant mathematics is largely *Survival Analysis* – hazard rates, etc. Much use is made here nowadays of *martingale methods* (Ch. IV). Non-life splits again into categories: motor; house; (house) contents (these are the only three kinds of insurance ordinary people take out); (personal) accident (the next commonest); travel; commercial property; industrial; ... There are even catastrophe insurance, weather insurance etc. nowadays.

2. The Poisson Process; Compound Poisson Processes

The Poisson distribution.

This is defined on $\mathbb{N}_0 := \{0, 1, 2, \cdots\}$ for a parameter $\lambda > 0$ by

$$p_k := e^{-\lambda}\lambda^k/k! \quad (k = 0, 1, 2, \cdots).$$

From the exponential series, $\sum_k p_k = 1$, so this does indeed give a probability distribution (or law, for short) on $\mathbb{N}_0$. It is called the *Poisson distribution* $P(\lambda)$, with *parameter* λ, after S.-D. Poisson (1781-1840) in 1837.

The Poisson law has *mean* λ. For if N is a random variable with the Poisson law $P(\lambda)$, $N \sim P(\lambda)$, N has mean

$$E[N] = \sum kP(N = k) = \sum kp_k = \sum k.e^{-\lambda}\lambda^k/k! = \lambda \sum e^{-\lambda}\lambda^{k-1}/(k-1)! = \lambda,$$

as the sum is 1 (exponential series – $P(\lambda)$ is a probability law). Similarly,

$$E[N(N-1)] = \sum k(k-1)e^{-\lambda}\lambda^k/k! = \lambda^2 \sum e^{-\lambda}\lambda^{k-2}/(k-2)! = \lambda^2 :$$

$$var(N) = E[N^2] - (E[N])^2) = E[N(N-1)] + E[N] - (E[N])^2) = \lambda^2 + \lambda - (\lambda)^2 = \lambda :$$

the Poisson law $P(\lambda)$ with parameter λ has mean λ and variance λ.

Note. 1. The Poisson law is the commonest one for *count data* on $\mathbb{N}_0$.

2. This property – that the mean and variance are equal (to the parameter, λ) is very important and useful. It can be used as the basis for a test for Poissonianity, the *Poisson dispersion test*. Data with variance greater than the Poisson are called *over-dispersed*; data with variance less than Poisson are *under-dispersed*.

3. The variance calculation above used the (second) *factorial moment*, $E[N(N-1)]$. These are better for count data than ordinary moments.

The Exponential Distribution

A random variable T on $\mathbb{R}_+ := (0, \infty)$ is said to have an *exponential distribution* with *rate* (or parameter) λ, $T \sim E(\lambda)$, if

$$P(T \le t) = 1 - e^{-\lambda t} \qquad \text{for all } \; t \ge 0.$$

So this law has density

$$f(t) := \lambda e^{-\lambda t} \quad (t > 0), \quad 0 \quad (t \le 0)$$

(as $\int_{-\infty}^{t} f(u)du = P(T \le t)$, as required). So the mean is

$$E[T] = \int tf(t)dt = \int_0^\infty \lambda te^{-\lambda t}dt = 1/\lambda. \int_0^\infty ue^{-u}du = 1/\lambda$$

(putting $u := \lambda t$). Similarly,

$$E[T^2] = \int t^2 f(t)dt = \int_0^\infty \lambda t^2 e^{-\lambda t}dt = 1/\lambda^2 \int_0^\infty u^2 e^{-u}du = 2/\lambda^2,$$

$$var(T) = E[T^2] - (E[T])^2 = 2/\lambda^2 - (1/\lambda)^2 = 1/\lambda^2.$$

The Lack-of-Memory Property.

Imagine components – lightbulbs, say – which last a certain *lifetime*, and are then discarded and replaced. Do we expect to see *aging*? With human lifetimes, of course we do! On average, there is much less lifetime remaining in an old person than in a young one. With some machine components, we also see aging. This is why parts in cars, aeroplanes etc. are replaced after their expected (or 'design') lifetime, at routine servicing. But, some components do *not* show aging. These things change with technology, but in the early-to-mid 20th C. lightbulbs typically didn't show aging. Nor in the early days of television did television tubes (not used in modern televisions!). In Physics, the atoms of radioactive elements show lack of memory. This is the basis for the concept of *half-life*: it takes the same time for half a quantity of radioactive material to decay as it does for half the remaining half, etc.

We can find *which* laws show no aging, as follows. The law F has the *lack-of-memory property* iff the components show no aging – that is, if a component still in use behaves as if new. The condition for this is

$$P(X > s + t | X > s) = P(X > t) \qquad (s, t > 0):$$

$$P(X > s + t) = P(X > s)P(X > t).$$

Writing $\overline{F}(x) := 1 - F(x)$ $(x \ge 0)$ for the *tail* of F, this says that

$$\overline{F}(s + t) = \overline{F}(s)\overline{F}(t) \qquad (s, t \ge 0).$$

Obvious solutions are

$$\overline{F}(t) = e^{-\lambda t}, \qquad F(t) = 1 - e^{-\lambda t}$$

for some $\lambda > 0$ – the exponential law $E(\lambda)$. Now

$$f(s+t) = f(s)f(t) \qquad (s, t \geq 0)$$

is a 'functional equation' – the *Cauchy functional equation* – and we quote that these are the *only* solutions, subject to minimal regularity (such as one-sided boundedness, as here – even on an interval of arbitrarily small length!).

So the exponential laws $E(\lambda)$ are *characterized* by the lack-of-memory property. Also, the lack-of-memory property corresponds in the renewal context to the *Markov property*. The renewal process generated by $E(\lambda)$ is called the *Poisson (point) process* with *rate* λ, $Ppp(\lambda)$. So: among renewal processes, the only Markov processes are the Poisson processes. We meet Lévy processes below: among renewal processes, the only Lévy processes are the Poisson processes.

It is the lack of memory property of the exponential distribution that (since the inter-arrival times of the Poisson process are exponentially distributed) makes the Poisson process the basic model for events occurring 'out of the blue'. Typical examples are accidents, insurance claims, hospital admissions, earthquakes, volcanic eruptions etc. So it is not surprising that Poisson processes and their extensions (compound Poisson processes) dominate in the actuarial and insurance professions, as well as geophysics, etc.

Gamma distributions.

Recall the *Gamma function*,

$$\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt, \qquad (x > 0)$$

($x > 0$ is needed for convergence at the origin). One can check (integration by parts, and induction) that

$$\Gamma(x+1) = x\Gamma(x) \quad (x > 0), \qquad \Gamma(n+1) = n! \quad (n = 0, 1, 2, \cdots);$$

thus Gamma provides a *continuous extension to the factorial.* One can show

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}$$

(the proof is essentially that $\int_{\mathbb{R}} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$, i.e. that the standard normal density integrates to 1). The Gamma function is needed for Statistics, as it commonly occurs in the normalisation constants of the standard densities.

The *Gamma distribution* $\Gamma(\nu, \lambda)$ with parameters $\nu, \lambda > 0$ is defined to have density

$$f(x) = \frac{\lambda^\nu}{\Gamma(\nu)}.x^{\nu-1}e^{-\lambda x} \qquad (x > 0).$$

This has MGF

$$\begin{aligned} M(t) := \int e^{tx} f(x)dx &= \frac{\lambda^\nu}{\Gamma(\nu)}.\int_0^\infty e^{tx}.x^{\nu-1}e^{-\lambda x}dx \\ &= \frac{\lambda^\nu}{\Gamma(\nu)}.\int_0^\infty x^{\nu-1}e^{-(\lambda-t)x}dx \\ &= \frac{\lambda^\nu}{\Gamma(\nu)}.\frac{1}{(\lambda-t)^\nu}\int_0^\infty u^{\nu-1}e^{-u}du \\ &= \Big(\frac{\lambda}{\lambda-t}\Big)^\nu \qquad (t < \lambda). \end{aligned}$$

Sums of exponential random variables.

Let $W_1, W_2, \dots W_n$ be independent exponentially distributed random variables with parameter λ ('W for waiting time' – see below): $W_i \sim E(\lambda)$. Then

$$S_n := W_1 + \cdots + W_n \sim \Gamma(n, \lambda).$$

For, each W_i has *moment-generating function* (MGF)

$$M(t) := E[e^{tW_i}] = \int_0^\infty e^{tx} f(x)dx = \int_0^\infty e^t x.\lambda e^{-\lambda x}dx$$

$$= \lambda.\int_0^\infty e^{-(\lambda-t)}dx = \lambda/(\lambda-t) \qquad (t < \lambda).$$

The MGF of the sum of independent random variables is the product of the MGFs (same for characteristic functions, CFs, and for probability generating functions, PGFs – check). So $W_1 + \cdots + W_n$ has MGF $(\lambda/(\lambda-t))^n$, the MGT of $\Gamma 9, n, \lambda)$ as above:

$$S_n := W_1 + \cdots W_n \sim \Gamma(n, \lambda).$$

The Poisson Process

Definition. Let $W_1, W_2, \dots W_n$ be independent exponential $E(\lambda)$ random variables, $T_n := W_1, + \ldots + W_n$ for $n \geq 1$, $T_0 = 0$, $N(s) := \max\{n : T_n \leq s\}$.

Then $N = (N(t) : t \geq 0)$ (or $(N_t : t \geq 0)$) is called the *Poisson process* (or *Poisson point process*) with *rate* λ, $Pp(\lambda)$ (or $Ppp(\lambda)$).

Interpretation: Think of the W_i as the waiting times between arrivals of events, then T_n is the arrival time of the nth event and $N(s)$ the number of arrivals by time s. Then $N(s)$ has a Poisson distribution with mean λs:

Theorem. If $\{N(s), s \geq 0\}$ is a Poisson process, then
(i) $N(0) = 0$;
(ii) $N(t+s) - N(s)$ is Poisson $P(\lambda t)$. In particular, $N(t) \sim P(\lambda t)$;
(iii) $N(t)$ has independent increments.
Conversely, if (i),(ii) and (iii) hold, then $\{N(s), s \geq 0\}$ is a Poisson process.

Proof. Part (i) is clear: the first lifetime is positive (they all are).

The link between the Poisson *process*, defined as above in terms of the exponential distribution, and the Poisson *distribution*, is as follows. First,

$$P(N_t = 0) = P(t < X_1) = e^{-\lambda t}.$$

This starts an induction, which continues (using integration by parts):

$$\begin{aligned} P(N_t = k) &= P(S_k \leq t < S_{k+1}) = P(S_k \leq t) - P(S_{k+1} \leq t) \\ &= \int_0^t \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} dx - \int_0^t \frac{\lambda^{k+1}}{\Gamma(k+1)} x^k e^{-\lambda x} dx \\ &= \int_0^t \Big[\frac{\lambda^k}{\Gamma(k+1)}.x^k - \frac{\lambda^{k-1}}{\Gamma(k)}.x^{k-1}\Big] d(e^{-\lambda x}) \\ &= \Big[\frac{\lambda^k}{\Gamma(k+1)}.t^k - \frac{\lambda^{k-1}}{\Gamma(k-1)}.t^{k-1}\Big] e^{-\lambda t} - \int_0^t e^{-\lambda x} \Big[\frac{\lambda^k}{\Gamma(k)}.x^{k-1} - \frac{\lambda^{k-1}}{\Gamma(k-1)}.x^{k-2}\Big] dx \\ &= \Big[\frac{\lambda^k}{\Gamma(k+1)}.t^k - \frac{\lambda^{k-1}}{\Gamma(k-1)}.t^{k-1}\Big] e^{-\lambda t} + \int_0^t e^{-\lambda x} \Big[\frac{\lambda^{k-1}}{\Gamma(k-1)}.x^{k-2} - \frac{\lambda^k}{\Gamma(k)}.x^{k-1}\Big] dx. \end{aligned}$$

But the integral here is $P(N_t = k-1)$. So (passing from Gammas to factorials)

$$P(N_t = k) - e^{-\lambda t}\frac{(\lambda t)^k}{k!} = P(N_t = k-1) - e^{-\lambda t}\frac{(\lambda t)^{k-1}}{(k-1)!},$$

completing the induction. This shows that

$$N(t) \sim P(\lambda t).$$

This gives (ii) also: re-start the process at time t, which becomes the new time-origin. The re-started process is a new Poisson process, by the lack-of-memory property applied to the current item (lightbulb above); this gives (ii) and (iii). Conversely, independent increments of N corresponds to the lack-of-memory property of the lifetime law, and we know that this characterises the exponential law, and so the Poisson process. //

Time-dependent rates.

The parameter λ is called the *rate* or *intensity* of the Poisson process. Think of it as the rate at which accidents happen (or telephone calls arrive at an exchange), or the intensity of a bombardment, etc. The above extends to include time-dependent intensities. We say that $\{N(s), s \geq 0\}$ is a *Poisson process* with *rate* $\lambda(r)$ if
(i) $N(0) = 0$,
(ii) $N(t+s) - N(s)$ is Poisson with mean $\int_s^t \lambda(r)dr$, and
(iii) $N(t)$ has independent increments.

Limit Theory.

For independent, identically distributed (iid for short) random variables $X_1, X_2, \cdots$, the *sample mean* (a *statistic*: a function of the data – random, as the data is, but known, after sampling, when you have the data) is

$$\overline{X} := \frac{1}{n} \sum_1^n X_k.$$

The *mean*, or *population mean*, $E[X]$ is defined as in Measure Theory, though we can restrict here to the discrete and density cases – a weighted average $\sum x_k f(x_k)$ in the discrete case where X takes values x_k with probability $f(x_k)$, and in the density case by the continuous analogue $\int x f(x)dx$ when X has density f. Always, the sum or integral is *absolutely convergent*:

$$E[|X|] < \infty; \qquad \sum |x_k| f(x_k) < \infty; \qquad \int |x| f(x)dx < \infty.$$

One would expect that $\overline{X}$ would tend to $E[X]$ as the sample size n increases. This is exactly right. By Kolmogorov's Strong Law of Large Numbers of 1933 (SLLN, or just LLN for short), convergence takes place *with probability one* (*almost surely*, or a.s. for short):

$$\overline{X} \to E[X] \qquad (n \to \infty) \qquad a.s.$$

For renewal theory (in particular, for the Poisson process), this gives another LLN.

Theorem (LLN for Renewal Theory). For X_i (positive) iid with mean μ, the renewal process $N = (N(t))$ satisfies

$$\frac{N(t)}{t} \to \frac{1}{\mu} \qquad (t \to \infty) \qquad a.s.$$

Proof. By definition of $N(t)$ and $S_n := \sum_1^n X_k$,

$$S_{N(t)} \leq t < S_{N(t)+1}.$$

So as soon as $N(t) > 0$,

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)+1} . \frac{N(t)+1}{N(t)}.$$

As $t \to \infty$, $N(t) \to \infty$ a.s. So the LLN the left tends to μ a.s.; so does the first term on the right, while the second term on the right tends to 1. This gives

$$t/N(t) \to \mu \qquad (t \to \infty) \qquad a.s.$$

The result follows by inverting this. //

The Conditional Mean Formula

Theorem (Conditional Mean Formula. For $\mathcal{B}$ any σ-field,

$$E[E[X|\mathcal{B}]] = E[X].$$

Proof. Take $\mathcal{C}$ the trivial σ-field $\{\emptyset, \Omega\}$. This contains no information, so an expectation conditioning on it is the same as an unconditional expectation. The first form of the tower property now gives

$$E[E[X|\mathcal{B}] \,|\{\emptyset, \Omega\}] = E[X|\{\emptyset, \Omega\}] = E[X]. \qquad //$$

The Conditional Variance Formula

Theorem (Conditional Variance Formula).

$$var(Y) = E[var(Y|X)] + var(E[Y|X]).$$

Proof. Recall $varX := E[(X - EX)^2]$. Expanding the square,

$$varX = E[X^2 - 2X.(EX) + (EX)^2] = E(X^2) - 2(EX)(EX) + (EX)^2 = E(X^2) - (EX)^2.$$

Conditional variances can be defined in the same way. Recall that $E(Y|X)$ is constant when X is known ($= x$, say), so can be taken outside an expectation over X, E_X say. Then

$$var(Y|X) := E(Y^2|X) - [E(Y|X)]^2.$$

Take expectations of both sides over X:

$$E_X var(Y|X) = E_X[E(Y^2|X)] - E_X[E(Y|X)]^2.$$

Now $E_X[E(Y^2|X)] = E(Y^2)$, by the Conditional Mean Formula, so the right is, adding and subtracting $(EY)^2$,

$$\{E(Y^2) - (EY)^2\} - \{E_X[E(Y|X)]^2 - (EY)^2\}.$$

The first term is $varY$, by above. Since $E(Y|X)$ has E_X-mean EY, the second term is $var_X E(Y|X)$, the variance (over X) of the random variable $E[Y|X]$ (random because X is). Combining, the result follows. //

Interpretation.
$varY$ = total variability in Y,
$E_X var(Y|X)$ = variability in Y not accounted for by knowledge of X,
$var_X E(Y|X)$ = variability in Y accounted for by knowledge of X.
In words:
variance = mean of conditional variance + variance of conditional mean, with these interpretations. This is extremely useful in Statistics, in breaking down uncertainty, or variability, into its contributing components. There is a whole area of Statistics devoted to such Components of Variance.

Compound Poisson Processes

We now associate i.i.d. random variables X_i with each arrival and consider

$$S(t) = X_1 + \ldots + X_{N(t)}, \qquad S(t) = 0 \text{ if } N(t) = 0.$$

Thus $S(t)$ is a *random sum* – a sum of a random number of random variables.

A typical application in the insurance context is a Poisson model of claim arrivals with random claim sizes The claims arrive at the epochs of a Poisson process with rate λ. The claims are independent (different motor accidents are independent; so are different house-insurance claims for fire damage, burglary etc.). Then the claim-total mean is the claim-number mean times the claim-amount mean. This is a special case of *Wald's identity* (below).

Theorem. (i) For N Poisson distributed with parameter λ and $X_1, X_2, \ldots$ independent of each other and of N, each with distribution F with mean μ, variance σ^2 and characteristic function $\phi(t)$, the compound Poisson distribution of

$$Y := X_1 + \ldots + X_N$$

has characteristic function $\psi(u) = \exp\{-\lambda(1-\phi(u))\}$, mean $\lambda\mu$ and variance $\lambda E[X^2]$.
(ii) For $N = (N_t)$ a compound Poisson process with rate λ and jump-distribution F with mean μ and variance σ^2, N_t has CF $\psi(u) = \exp\{-\lambda t(1-\phi(u))\}$, mean $\lambda t\mu$ and variance $\lambda t E[X^2]$.

Proof. (i) The characteristic function (CF) follows from

$$\begin{aligned}
\psi(t) = E[e^{itY}] &= E[\exp\{it(X_1 + \ldots + X_N)\}] \\
&= \sum_n E[\exp\{it(X_1 + \ldots + X_N)\}|N = n].P(N = n) \\
&= \sum_n e^{-\lambda}\lambda^n/n!.E[\exp\{it(X_1 + \ldots + X_n)\}] \\
&= \sum_n e^{-\lambda}\lambda^n/n!.(E[\exp\{itX_1\}])^n \\
&= \sum_n e^{-\lambda}\lambda^n/n!.\phi(t)^n \\
&= \exp\{-\lambda(1 - \phi(t))\}.
\end{aligned}$$

We give two proofs for the mean and variance, (a) by differentiating the CF, (b) from the Conditional Mean and Conditional Variance Formulae. Recall that if X has CF ϕ,

$$\phi(t) = E[e^{iXt}].$$

Differentiating formally (this is justified here – we quote this),

$$\phi'(t) = E[iXe^{iXt}]: \qquad \phi'(0) = iE[X]; \qquad E[X] = -i\phi'(0);$$

$$\phi''(t) = E[-X^2 e^{iXt}]: \qquad \phi''(0) = -E[X^2]; \qquad E[X^2] = -\phi''(0).$$

(a) Differentiate the CF:

$$\psi'(t) = \psi(t).\lambda\phi'(t),$$

$$\psi''(t) = \psi'(t).\lambda\phi'(t) + \psi(t).\lambda\phi''(t).$$

By above, ($\phi(0) = 1$ and) $\phi'(0) = i\mu$, $\phi''(0) = -E[X^2]$,

$$\psi'(0) = \lambda\phi'(0) = \lambda.i\mu,$$

and as also $\psi'(0) = iEY$, this gives

$$E[Y] = \lambda\mu.$$

Thus the mean of the random sum $Y := X_1 + \cdots + X_N$ is the product of the means of X (short for a typical X_i) and N:

$$E[Y] := E[X_1 + \cdots + X_N] = E[X].E[N].$$

This is (a special case of) *Wald's identity* (Abraham Wald (1902-1950) in 1944). Similarly,

$$\psi''(0) = i\lambda\mu.i\lambda\mu + \lambda\phi''(0) = -\lambda^2\mu^2 - \lambda E[X^2],$$

and also ($\psi(0) = 1$, $\psi'(0) = i\lambda\mu$ and) $\psi''(0) = -E[Y^2]$. So

$$var\ Y = E[Y^2] - [EY]^2 = \lambda^2\mu^2 + \lambda E[X^2] - \lambda^2\mu^2 = \lambda E[X^2].$$

(b) Given N, $Y = X_1 + \ldots + X_N$ has mean $NEX = N\mu$ and variance $N\ var\ X = N\sigma^2$. As N is Poisson with parameter λ, N has mean λ and variance λ. So by the Conditional Mean Formula,

$$EY = E[E(Y|N)] = E[N\mu] = \lambda\mu.$$

By the Conditional Variance Formula,

$$\begin{aligned} var\ Y &= E[var(Y|N)] + var\ E[Y|N] \\ &= E[Nvar\ X] + var([N\ E[X]) \\ &= E[N].var\ X + var\ N.(EX)^2 \\ &= \lambda[E[X^2] - (E[X])^2] + \lambda.(E[X])^2 \\ &= \lambda E[X^2] = \lambda(\sigma^2 + \mu^2). \end{aligned}$$

(ii) Apply (i): N_t has mean λt and variance λt. //

In the insurance context (below), the Poisson points represent the claim arrivals, so the Poisson rate λ is the rate at which claims arrive; μ is the mean claim size. So $\lambda\mu$ has the interpretion of a *claim rate* – rate at which money goes *out* of the company in claims.

Just as the mathematics of the Black-Scholes model (Ch. VI) is dominated by *Brownian motion*, that of insurance is dominated by the *Poisson* and *compound Poisson* processes. These are the basic prototypes, and all we have time to cover in detail in this course. However, these are models, of reality, and reality is always more complicated than any model! Box's dictum (George Box, British statistician, 1919-2013): *All models are wrong. Some models are useful.* In more advanced work, more complicated and detailed models are needed. So there is plenty of scope for useful applications in the real world of any probability or statistics you know, or will learn! At the end of the course (VII.5), we discuss briefly some generalisations. But to note for now: the principal weakness of our assumptions here is the *independence* of claims. This is reasonable under normal conditions, but not during a crisis. Think of natural disasters such as major hurricanes, etc.

§3. Renewal theory

Renewal Processes

Suppose we use components – light-bulbs, say – whose lifetimes $X_1, X_2, \ldots$ are independent, all with law F on $(0, \infty)$. The first component is installed new, used until failure, then replaced, and we continue in this way. Write

$$S_n := \sum_1^n X_i, \qquad N_t := \max\{k : S_k \leq t\}.$$

Then $N = (N_t : t \geq 0)$ is called the *renewal process* generated by F; it is a *counting process*, counting the number of failures seen by time t. Note that

$$S_{N(t)} \leq t.$$

Note. For stochastic processes, notations such as N_t and $N(t)$ are used interchangeably.

Renewal processes are often used, but the only ones we need here are the

Poisson processes – those for which the lifetime law is exponential.

The renewal function

We saw above that

$$N_t/t \to 1/\mu \qquad (t \to \infty), \ a.s.$$

If we apply the expectation operator $E[.]$ formally, this suggests that

$$E[N_t]/t \to 1/\mu \qquad (t \to \infty).$$

This is indeed true, but although its conclusion seems weaker than that of the a.s. result, its proof if harder (though not as hard as that of the SLLN!).

Theorem. If the mean lifetime length μ is finite, the renewal function $E[N_t]$ satisfies

$$E[N_t]/t \to 1/\mu \quad (t \to \infty).$$

Proof. The conclusion with $\geq$ in place of $=$ does indeed follow from the a.s. result by taking expectations. This is by *Fatou's lemma*, which we quote from Measure Theory. [For proof, see e.g. a book on Measure Theory, or my homepage, Stochastic Processes, I.5 Lecture 8.] For the $\leq$ part, choose $a > 0$, and truncate the X_n at level a:

$$\tilde{X}_n := min(X_n, a).$$

Write $\tilde{N}_t$, $\tilde{\mu}$ for the 'tilde' analogues of N_t, μ. By Wald's identity,

$$E[\tilde{X}_1 + \cdots + \tilde{X}_{\tilde{N}_t}] = E[\tilde{X}].E[\tilde{N}_t] = \tilde{\mu}.E[\tilde{N}_t].$$

Now $\tilde{N}_t \geq N_t$ (because of the truncation, there will be more renewals if anything), and $\tilde{S}_{\tilde{N}_t - 1} + \tilde{X}_{\tilde{N}_t} \leq t + a$ (the 't' from the first term, the 'a' from the second). So

$$\begin{aligned}
E[N_t]/t &\leq E[\tilde{N}_t]/t && (N_t \leq \tilde{N}_t) \\
&= \tilde{\mu}^{-1} E[\tilde{X}_1 + \cdots + \tilde{X}_{\tilde{N}_t}]/t && \text{(above – Wald's identity)} \\
&= \tilde{\mu}^{-1} E[\tilde{S}_{\tilde{N}_t}]/t && \text{(definition of } \tilde{S}_n) \\
&\leq \tilde{\mu}^{-1} && (S_{N(t)} \leq t, \text{ and similarly for } \tilde{S}_n,\ \tilde{N}_t).
\end{aligned}$$

So

$$\limsup E[N_t]/t \leq \tilde{\mu}^{-1}.$$

Now let $a \uparrow \infty$: $\tilde{\mu} \to \mu$, giving the $\leq$ part and the result. //

With F the lifetime distribution function – that of each X_i – the distribution function of $S_n := X_1 + \cdots + X_n$ is $F * \cdots * F$ (n Fs), the n-fold convolution of F with itself, written F^{*n}. Define

$$U(t) := \sum_{n=0}^{\infty} F^{*n}(t).$$

This is called the *renewal function* of F. For, it gives the mean number $E[N_t]$ of renewals up to time t:

Theorem. The renewal function gives the mean number of renewals:

$$U(t) = E[N_t].$$

So if the mean lifetime is μ,

$$U(t)/t \to 1/\mu \qquad (t \to \infty).$$

Proof.

$$\begin{aligned} E[N_t] &= \sum_0^{\infty} nP(N_t = n) \\ &= \sum n[P(N_t \geq n) - P(N_t \geq n+1)] \\ &= \sum P(N_t \geq n), \end{aligned}$$

by *partial summation* (or *Abel's lemma*). [This is the discrete analogue of integration by parts. See e.g. a book on Analysis, or my homepage, M3P16 Analytic Number Theory, I.3.] But $\{N_t \geq n\} = \{S_n \leq t\}$, so

$$E[N_t] = \sum P(S_n \leq t) = \sum F_n^*(t) = U(t),$$

giving the first part; the second part follows from the result above. //

The renewal theorem

Renewal theory needs a distinction between two cases. If the X_i are integer-valued (when so are the S_n), or are supported by an arithmetic progression (AP), we are in the *lattice case*, otherwise in the *non-lattice case*.

The next result looks like a differenced form of the last one. It is due to David Blackwell (1919-2010) in 1953. We state it for the non-lattice case and $\mu < \infty$, but it extends to the lattice case and $\mu = \infty$ also.

Theorem (Blackwell's renewal theorem). In the non-lattice case,

$$U(t+h) - U(t) \to h/\mu \quad (t \to \infty) \quad \forall h > 0.$$

This famous result has a number of different proofs, but we do not include one here (my favourite is only a few lines, but needs a prerequisite beyond our scope here).

Blackwell's theorem has a number of variants. The one we need (which we also quote) is due to W. L. Smith and W. Feller. Recall the *Riemann integral* (defined for functions on a finite interval), and the *Lebesgue integral* which generalises it (defined for functions on e.g. the line, plane etc.). We need a new concept.

Definition. Divide the line into intervals $I_{n,h} := (nh, (n+1)h]$. For a function z on $\mathbb{R}$ and $x \in I_{n,h}$, write

$$\overline{z}_h := sup\{z(y) : y \in I_{n,h}\}, \qquad \underline{z}_h := inf\{z(y) : y \in I_{n,h}\}.$$

Call z *directly Riemann integrable (dRi)* if $\int \overline{z_h} := \int_{-\infty}^{\infty} \overline{z}_h(x)dx$ is finite for some (equivalently, for all) $h > 0$, and similarly for $\int \underline{z_h}$, and

$$\int \overline{z}_h - \int \underline{z}_h \to 0 \qquad (h \to 0).$$

This is the same as Riemann integrability if z is supported on some finite interval, but for z of unbounded support is stronger than Lebesgue integrability: z is dRi iff it is Lebesgue integrable, and both $\int \overline{z}_h$ and $\int \underline{z}_h$ have a common limit $\int z$ as $h \to 0$. Condition dRI will hold whenever we need it. We quote that dRi needs z bounded and a.e. continuous (w.r.t. Lebesgue measure), and that this plus z of bounded support implies dRi. Also, z non-increasing and Lebesgue integrable imples dRi.

The *renewal equation* for F and z (both known) is the integral equation

$$Z(t) = z(t) + \int_0^t Z(t-u)dF(u) \quad (t \geq 0): \quad Z = z + F * Z. \qquad (RE)$$

Here F (for us, the lifetime distribution above) and z are given, and (RE) is to be solved for Z.

Theorem (Key Renewal Theorem). If z in (RE) is dRi, then for U the renewal function of F as above,

$$\lim_{t\to\infty} Z(t) = \lim_{t\to\infty} U * z(t) = \frac{1}{\mu}\int_0^\infty z(x)dx.$$

The proof of the Key Renewal Theorem from Blackwell's Renewal Theorem is not long or hard, but as it is Analysis rather than probability or insurance mathematics, we omit it. For a proof, see e.g. [RSST, 6.1.4 p216-219.

§4. The Ruin Problem.

Consider the cash flow of an insurance company. The premium income comes in from the policy holders at constant rate, c say (to a first approximation: the company hopes to attract more policy holders, and premium rates will typically vary on renewal – but are constant during the lifetime of the policy). So income over time t is ct. If the company has initial capital u, its capital at time t is thus $u+ct$. Meanwhile, claims occur. We model these as occurring at the instants of a Poisson process of rate λ, the claims being independent and identically distributed (iid) with claim distribution F, with CF ϕ, mean μ and variance σ^2. So the number of claims over the interval $[0,t]$ is $N(t)$, which is Poisson distributed with parameter λt: $N(t) \sim P(\lambda t)$. So by the Theorem of VII.2 above, the total claim has mean $\lambda\mu t$. Thus cash comes in at rate c, but goes out at rate $\lambda\mu$. This simple argument suggests – what is indeed true – that a *necessary* condition for the company to avoid bankruptcy is

$$c > \lambda\mu :$$

money should come in *faster* than it goes out. The proof is by the Strong Law of Large Numbers (LLN, as above). In the critical case $c = \lambda\mu$ the company is 'balanced on a knife-edge', and will soon go bankrupt.

The company thus *must* have $c > \lambda\mu$, so we assume this from now on. But, any insurance company has only finite funds; it may face arbitrarily severe runs of bad luck; combining these, bankruptcy is always a possibility. (Indeed, this is true for all companies, not just insurance companies! This is why bankruptcy needs to be recognised as a possibility, and governed by bankruptcy law. This varies from time to time and from country to country – a very interesting and important subject, but not one we can pursue here.)

Clearly the company's best defence against bankruptcy is to have a large cash reserve u, to act as a buffer, or 'insurance policy', against such runs of bad luck. Clearly the probability of ruin – ruin probability – decreases with u. How fast? The classical *ruin problem* is to investigate this question, to which we return below.

Note. We may if we wish take $c = 1$ for convenience. This (slightly) simplifies the formulae. It amounts to changing from real time to *operational* or *business* time – looking at the situation in the time-scale most natural to it. Recall that there are *no natural units of time or space* (except the Planck scale, at subatomic level, for those with a background in Physics!): time is measured in seconds, minutes, hours, days (60 s to the m, 60 m to the h, 24 h to the day – pre-decimal), and length in metres (metric system – mm, cm, m, km) or inches/feet/yards/miles (Imperial measure) – neither is natural, both are conventional.

The Net Profit Condition (NPC)

With c the premium rate, X_i the claim sizes and W_i the inter-claim waiting times, write

$$Z_i := X_i - cW_i.$$

Then

$$E[Z_i] := E[X_i] - cE[W_i] = \mu - c/\lambda.$$

The first term on the right measures money *out* (of the company), the second measures money *in*. As we have seen, to avoid bankruptcy we need ('more in than out')

$$E[Z_i] := E[X_i] - cE[W_i] = \mu - c/\lambda < 0 : \qquad c > \mu\lambda. \qquad (NPC)$$

This is called the *net profit condition (NPC)*. For as we have seen, $\lambda\mu$ is the claim rate (rate at which cash goes *out* to claims); c is the *premium rate* (rate at which cash comes *in*, through premiums); we need (NPC) – 'more in than out' for survival.

Safety loading and premium calculation

The first duty of any company is to stay solvent – to avoid bankruptcy.

To do this, an insurance company has to have its premium rate $c > \mu\lambda$ so as to satisfy (NPC).

But, like any other business, the insurance business is competitive. If premiums are too *low*, the firm goes bankrupt (above) because its premium income fails to meet its outgoings on claims. But if premiums are too *high*, the firm will not be competitive with other firms; over time, it will lose market share to them, and will eventually go bankrupt (or otherwise go out of business – e.g., be taken over) as premium income declines to be too small to meet overheads. So the firm needs to take a policy decision as to how much to charge in premiums. This is measured by the *safety loading (SL)*, ρ, defined by

$$c = (1+\rho)\frac{E[X_i]}{E[W_i]} = (1+\rho)\lambda\mu : \qquad \rho := \frac{c - \lambda\mu}{\lambda\mu}. \qquad (SL)$$

Thus $\rho > 0$ in (SL) is equivalent to (NPC).

Lundberg's inequality

Before, we used the characteristic function (CF), defined for a random variable X by $\phi(t) := E[e^{itX}]$, for t real. The reason for using complex numbers here – for the $i := \sqrt{-1}$ – is to ensure that the CF *always exists*. It does, because

$$|\phi(t)| = |E[e^{itX}]| \le E[|e^{itX}|] = E[1] = 1.$$

(Recall Euler's formula: for θ real, $e^{i\theta} = cos\,\theta + i sin\,\theta$, so $|e^{i\theta}| = \sqrt{cos^2\theta + sin^2\theta} = 1$. Recall also that expectation is integration (w.r.t. a probability distribution), so 'mod of integral $\le$ integral of mod'.) But we now find it convenient to use real numbers, and switch to the *moment-generating function (MGF)*,

$$M(s) := E[e^{sX}].$$

This is certainly defined for $s = 0$: $M(0) = E[e^0] = E[1] = 1$. But it may not be defined (finite) for all (or even any) $s \neq 0$. (Example: the *exponential* distribution $E(\lambda)$ with parameter λ has MGF $\lambda/(\lambda - s)$, but this is only finite for $s < \lambda$.) We now assume the *small claim condition (SCC)*,

$$M(s) := E[e^{sX_1}] < \infty \quad \forall s \in (-s_0, s_0), \quad \text{for some } s_0 > 0. \qquad (SCC)$$

This implies that the tail of X_1 decays exponentially. For (*Markov's Inequality*): for $s \in (0, s_0)$ and $x > 0$,

$$M(s) = E[e^{sX_1}] \ge E[e^{sX_1}; X_1 > x] \ge e^{sx}E[1; X_1 > x] = e^{sx}P(X_1 > x) :$$

$$P(X_1 > x) \leq e^{-sx} M(s) \qquad \forall x > 0.$$

Differentiating the MGF twice (and writing X for X_1 for convenience):

$$M(s) = E[e^{sX}], \qquad M'(s) = E[Xe^{sX}], \qquad M''(s) = E[X^2 e^{sX}] \geq 0.$$

Also, the MGF M(s) is smooth (we can differentiate it as often as we like, where it is defined). So its graph has a tangent, and as $M'' \geq 0$, the *tangent is increasing* – the *graph bends upwards*. Such functions are called *convex*. Also, as $M(0) = 1$, the graph goes through 1 at the origin. Now smooth convex functions can intersect any line at most twice (e.g., a parabola may not cut a line, or can cut it once (double point of contact), or twice, but not more).

The crucial assumption is that $M(s)$ cuts the line $y = 1$ *twice*, once (necessarily) at the origin and once at a positive point r.

Definition.

The *Lundberg coefficient* (or *adjustment coefficient*) r, which we assume to exist in what follows, is the point $r > 0$ (we assume r exists; it is then unique) such that $r = s$ satisfies

$$M_{Z_1}(s) := E[\exp\{s(X_1 - cW_1)\}] = 1. \qquad (LC)$$

The right is (writing X, W for X_1, W_1) $M_X(s).M_W(-cs)$. Now as $W \sim E(\lambda)$, W has Laplace-Stieltjes transform (LST) $E[e^{-tW}] = M_W(-t) = \int_0^\infty e^{-tx}.\lambda e^{-\lambda x} dx = \lambda/(\lambda + t)$. So the defining property of the Lundberg (adjustment) coefficient is (writing M for M_X for short)

$$M(r).\frac{\lambda}{\lambda + cr} = 1: \qquad M(r) = \frac{\lambda + cr}{\lambda} = 1 + \frac{cr}{\lambda}. \qquad (LC')$$

Theorem (Lundberg's Inequality). Assuming that the Small-Claims Condition (SCC) holds and that the Lundberg coefficient r in (LC) exists, the ruin probability $\psi(u)$ with initial capital u and over all time satisfies

$$\psi(u) \leq e^{-ru}.$$

Proof. Write

$$S_n := Z_1 + \cdots + Z_n, \qquad Z_i := X_i - cW_i.$$

Then $S = (S_n)$ is a *random walk*, with *step-lengths* $Z_i := X_i - cW_i$. As the ruin probability increases with time, the ruin probability $\psi(u)$ is the increasing limit of the ruin probability $\psi_n(u)$ with just the first n claims X_i and waiting times W_i involved:

$$\psi_n(u) = P(\max_{1 \le k \le n} S_k > u) = P(S_k > u \text{ for some } k \in \{1, \cdots, n\}\ ).$$

We prove that

$$\psi_n(u) \le e^{-ru} \qquad \forall n \in \mathbb{N}, u > 0. \qquad (*)$$

The result follows from this by letting $n \to \infty$; we prove $(*)$ by induction (on n).

The induction starts, by Markov's Inequality:

$$\psi_1(u) \le e^{-ru} M_{Z_1}(r) = e^{-ru},$$

by definition of the Lundberg coefficient: $M_{Z_1}(r) = 1$.

Assume that $(*)$ holds for n, and write F for F_{Z_1}, the distribution function of Z_1. Then

$$\begin{aligned}
\psi_{n+1}(u) &= P(\max\{S_k : 1 \le k \le n+1\} > u) \\
&= P(Z_1 > u) + P(Z_1 \le u, \max\{Z_1 - (S_k - Z_1) : 2 \le k \le n+1\} > u) \\
&= p_1 + p_2,
\end{aligned}$$

say.

We now make our first use of the *renewal argument*, which will allow us to reduce the proof of our main results to an application of the Key Renewal Theorem. The idea is to *condition* on the value of the first claim Z_1, and let the process 'renew itself' with the first claim, starting afresh thereafter. So, starting the random walk after $Z_1 = x$ in the p_2-term above and conditioning on the value x of Z_1,

$$p_2 = \int_{(-\infty, u]} P(\max_{1 \le k \le n}(x + S_k) > u) dF(x).$$

In full, this is a use of the Conditional Mean Formula. For an event A, the random variable I_A (its indicator function: 1 if $\omega \in A$, 0 if not) has mean

$$E[I_A] = P(A).$$

Then conditioning on information $\mathcal{B}$ (size of first claim here),

$$P(A) = E[I_A] = E[E[I_A|\mathcal{B}]].$$

Now

$$p_1 = \int_{(u,\infty)} dF(x) \leq \int_{(u,\infty)} e^{r(x-u)} dF(x),$$

as $r > 0$, while

$$\begin{aligned} p_2 &= \int_{(-\infty,u]} P(\max_{1\leq k\leq n}(x+S_n) > u) dF(x) \\ &= \int_{(-\infty,u]} \psi_n(u-x) dF(x) \\ &\leq \int_{(-\infty,u]} e^{r(x-u)} dF(x) \quad \text{(by the induction hypothesis).} \end{aligned}$$

Combining the domains $(-\infty, u]$ and (u, ∞) of integration here,

$$p_1 + p_2 \leq \int_{-\infty}^{\infty} e^{r(x-u)} dF(x) = e^{-ru} \int e^{rx} dF(x) = e^{-ru} M(r) = e^{-ru},$$

as $M(r) = 1$ by definition of the Lundberg coefficient r, completing the induction. //

Example: Exponential claims.

Recall the exponential distribution $E(\lambda)$ with parameter λ, which has mean $1/\lambda$ and MGF $\lambda/(\lambda - s)$. With the arrival process Poisson with rate λ as above (so the inter-claim waiting times are $E(\lambda)$), consider now the simplest case, when the claim sizes are also exponential, $E(\gamma)$ say. So W_i has MGF $\gamma/(\gamma - s)$, cW_i has MGF $\gamma/(\gamma - cs)$, and $Z_i = X_i - cW_i$ has MGF

$$M_Z(s) = M_X(s) M_{cW}(-s) = \frac{\gamma}{\gamma - s} . \frac{\lambda}{\lambda + cs}.$$

As usual, we assume the Net-Profit Condition (NPC):

$$E[X]/E[W] = \lambda/\gamma < c.$$

Then the Lundberg coefficient r is the (unique, positive) root of

$$M_Z(r) = \frac{\gamma}{\gamma - r} . \frac{\lambda}{\lambda + cr} = 1.$$

This is a quadratic,

$$Q(r) := -[(cr+\lambda)(-r+\gamma) - \lambda\gamma] = cr^2 + (\lambda - c\gamma) r = r(cr + \lambda - c\gamma) = 0,$$

with positive root

$$r = \gamma - \frac{\lambda}{c} > 0,$$

by (NPC). In terms of the safety loading ρ,

$$c = \frac{E[X]}{E[W]}(1+\rho) = \frac{\lambda}{\gamma}(1+\rho).$$

So in terms of the safety loading ρ rather than the premium rate c,

$$r = \gamma \frac{\rho}{(1+\rho)},$$

and the Lundberg inequality is

$$\psi(u) \leq \exp\{-u\gamma\rho/(1+\rho)\}.$$

This is nearly exact: in this case, there is a constant C with

$$\psi(u) = C\exp\{-u\gamma\rho/(1+\rho)\}.$$

Note. This example is unusually simple: in general, there is no closed form for r, and we have to find it by numerical methods. This is typically the case for solutions of transcendental (rather than algebraic) equations.

Cumulant-generating function (CGF)
Definition. The cumulant-generating function (CGF) $\kappa(s)$ of a distribution is the logarithm of the MGF M:

$$\kappa(s) := \log M(s).$$

Thus the Lundberg (adjustment) coefficient may also be defined by

$$\kappa_{Z_1}(s) = \log M_{Z_1}(s) := \log E[\exp\{s(X_1 - cW_1)\}] = 0. \qquad (LC')$$

Like the MGF, the CGF is also *convex.* For,

$$\kappa = \log M, \qquad \kappa' = M'/M, \qquad \kappa'' = [MM'' - (M')^2]/M^2.$$

By the Cauchy-Schwarz inequality,

$$(M')^2 = (E[Xe^{sX}])^2 \leq MM'' = E[e^{sX}].E[X^2e^{sX}]$$

($E[.]$ is an integral, over the probability space Ω w.r.t. the probability measure P, or $dP(\omega)$; here we apply C-S for the measure $e^{sX(\omega)}dP(\omega)$). So $\kappa'' \geq 0$. So κ is convex. The graph of $\kappa(s)$ has two roots, $s = 0$ and $s = r$, the Lundberg (adjustment) coefficient.

The ruin problem and the renewal equation

With initial capital u, the company's capital at time t is

$$C_t = u + ct - \sum_1^{N_t} X_i.$$

The probability of (ultimate) ruin and of survival are

$$\psi(u) = P(\inf_{0<t<\infty} C_t < 0 | C_0 = u),$$

$$\overline{\psi}(u) := 1 - \psi(u) = P(C_t \geq 0 \ \forall t | C_0 = u).$$

The key to the relevance of renewal methods here – the *renewal argument* we used before – is that the capital process *renews itself at the time of the first claim*: if this is at time $W_1 = s$ and of size $X_1 = x$, it begins again, with initial capital $u + cs - x$ (of course if this is negative, the company goes bankrupt when it receives its first claim!). We can *condition* (as above) on the time W_1 (density $\lambda e^{-\lambda s}$) and size X_1 (distribution F) of first claim:

$$\overline{\psi}(u) = \int_0^\infty \lambda e^{-\lambda s} ds \int_0^{u+cs} dF(x) \overline{\psi}(u + cs - x).$$

Change variable from s to $t := u + cs$: the limits $0 < x < u + cs$, $s > 0$ become $0 < x < t$, $t > u$:

$$\overline{\psi}(u) e^{-\lambda u/c} = \frac{\lambda}{c} \int_u^\infty \lambda e^{-\lambda t/c} dt \int_0^t dF(x) \overline{\psi}(t - x).$$

This shows that $\overline{\psi}$ is differentiable (as the exponential and the integral are). Differentiating w.r.t. u:

$$e^{-\lambda u/c}(\overline{\psi}'(u) - \frac{\lambda}{c}\overline{\psi}(u)) = -\frac{\lambda}{c} e^{-\lambda u/c} \int_0^u \overline{\psi}(u - x) dF(x) :$$

$$\overline{\psi}'(u) = \frac{\lambda}{c}\overline{\psi}(u) - \frac{\lambda}{c} \int_0^u \overline{\psi}(u - x) dF(x).$$

Integrate over $u \in [0,t]$, and write

$$h(y) := \int_0^{t-y} \overline{\psi}(u)du \quad (0 \leq y \leq t), \quad 0 \quad (y > t).$$

The first term on the right integrates to $h(0)$, so

$$\begin{aligned}
\overline{\psi}(t) - \overline{\psi}(0) - \frac{\lambda}{c}h(0) &= -\frac{\lambda}{c}\int_0^t du \int_0^u \overline{\psi}(u-x)dF(x) \\
&= -\frac{\lambda}{c}\int_0^t duh(x)dF(x) \quad (\text{def. of } h(.) \text{ on } [0,t]) \\
&= -\frac{\lambda}{c}\int_0^\infty duh(x)dF(x) \quad (h(.) = 0 \text{ on } [t,\infty]) \\
&= \frac{\lambda}{c}\int_0^\infty duh(x)d(1-F)(x).
\end{aligned}$$

Integrating by parts, the integrated term on the right cancels with the last term on the left:

$$\overline{\psi}(t) - \overline{\psi}(0) = -\frac{\lambda}{c}\int_0^\infty h'(x)(1-F(x))dx :$$

$$\overline{\psi}(t) = \overline{\psi}(0) + \frac{\lambda}{c}\int_0^\infty \overline{\psi}(t-x)(1-F(x))dx.$$

This integral equation for $\overline{\psi}$ translates into one for ψ itself:

$$\psi(u) = \frac{\lambda}{c}\Big(\int_u^\infty (1-F(x))dx + \int_0^\infty \psi(u-x)(1-F(x))dx\Big). \qquad (*)$$

Note that F has mean

$$\mu := \int_0^\infty xdF(x) = -\int_0^\infty xd(1-F)(x).$$

Integrating by parts (rather as above), the integrated term vanishes, giving

$$\mu = \int_0^\infty (1-F(x))dx.$$

Thus $(1-F(x))/\mu$ is a probability density on $(0,\infty)$, and the integral equation $(*)$ above is of *renewal-equation* type. This is crucial: it reduces the proof of the main result (Cramér's estimate of ruin, below) to an application of the Key Renewal Theorem.

Multiplying both sides of $(*)$ by e^{ru} gives

$$\psi(u)e^{ru} = \frac{\lambda}{c}e^{ru}\int_u^\infty (1-F(x))dx + \int_0^\infty \psi(u-x)e^{r(u-x)}.\Big[\frac{\lambda}{c}(1-F(x)e^{rx}\Big]dx. \qquad (**)$$

Cramér's estimate of ruin

Theorem (Cramér's estimate of ruin).

For the Cramér-Lundberg model, under the Net Profit Condition (NPC) and the Lundberg condition (LC), with r the Lundberg coefficient and $\psi(u)$ the probability of ruin with initial capital u,

$$e^{ru}\psi(u) \to C: \qquad \psi(u) \sim Ce^{-ru} \qquad (u \to \infty),$$

where the constant C is given by

$$C = \frac{c-\lambda\mu}{cr\int_0^\infty xe^{rx}(1-F(x))dx}.$$

Proof. From the existence of the Lundberg coefficient $r > 0$ in (LC),

$$M(r) := \int_0^\infty e^{rx}dF(x) = -\int_0^\infty e^{rx}d(1-F)(x) = 1 + \frac{cr}{\lambda}.$$

Integrating by parts (again as above!), the integrated term is 1, giving

$$\int_0^\infty (1-F(x))e^{rx}dx = \frac{c}{\lambda}:$$

$$\frac{\lambda}{c}(1-F(x))e^{rx}$$

is a probability density on $(0,\infty)$, which shows that $(**)$ is an integral equation of renewal type (RE). So by the Key Renewal Theorem, its solution $\psi(u)e^{ru}$ has a limit, C say, as $u \to \infty$, giving the first (and more important) part.

To identify the limit C: from the Key Renewal Theorem, C is the integral of the first (z-) term on the right, divided by the mean of the probability distribution in the convolution. The integral term is λ/c times

$$\begin{aligned}\int_0^\infty e^{ru}du\int_u^\infty (1-F(x))dx &= \frac{1}{r}\int_0^\infty [\int_u^\infty (1-F(x))dx]d(e^{ru}) \\ &= \frac{1}{r}[e^{ru}\int_u^\infty (1-F(x))dx]_0^\infty + \frac{1}{r}\int_0^\infty e^{ru}(1-F(u))du \\ &= -\frac{\mu}{r}+\frac{c}{r\lambda} = \frac{c-\lambda\mu}{cr},\end{aligned}$$

by the calculation above. So, in the notation of the Key Renewal Theorem,

$$\int_0^\infty z(x)dx = \frac{\lambda}{c}.\frac{c-\lambda\mu}{cr}.$$

The mean of this density (the 'μ' term in the Key Renewal Theorem) is

$$\frac{\lambda}{c}.\int_0^\infty xe^{rx}(1-F(x))dx.$$

So C is their ratio:

$$C = \frac{c-\lambda\mu}{cr\int_0^\infty xe^{rx}(1-F(x))dx}. \qquad //$$

Note. 1. The argument above draws on several sources: [Mik 4.2.2, 166-171], [AA, 4.5a p90], [A, IV.2 Ex. 2.3; IV.4, 5], [RSST, 5.3.2, 5.4.2].
2. In addition to the Key Renewal Theorem, the crux in the above is the *change of measure*

$$F = F(dx) \mapsto \frac{\lambda}{c}(1-F(x))e^{rx}dx.$$

This is also called *exponential tilting* and the *Esscher transform*, after the Swedish actuary Fredrik Esscher in 1932. (It also occurs in *large deviations*, important in many areas of probability, statistics and statistical mechanics.) This change-of-measure technique is of course also related to that in *Girsanov's theorem* in mathematical finance (Ch. VI).

Filip Lundberg

Filip Lundberg (1876-1965) was a Swedish actuary and pioneer of the theory of collective risk. His work in actuarial mathematics goes back to 1903, long before probability theory as we know it existed. He is credited by Cramér (1969, 1976) as initiating the theory of collective risk, in a series of papers in the late 1920s. Here, as in the work of Cramér below, one sees the modern formulation: the income stream of an insurance company, from premiums, is deterministic and linear; the outgoings, to meet claims, form a compound Poisson process, from the claims process (a Poisson process, of rate or intensity λ say) and the claim-size distribution (F say). Given the company's initial capital, u say, one studies the dependence of the probability of ruin (clearly positive) as a function of u and the current time, obtaining the familiar exponential estimate.

Lundberg may be regarded as having introduced the *Poisson process*, the foundation stone of actuarial mathematics. But one must bear in mind that the very term stochastic process is anachronistic here: the term was coined by Khinchin in the 1920s, and the necessary mathematical underpinning had to wait for Kolmogorov's *Grundbegriffe* of 1933.

Cramér (1969) draws attention to the implications of Lundberg's work for *reinsurance*. This field is of ever-growing importance, as the financial world becomes larger and more complicated, as it poses in modern form Juvenal's famous question (VII.1): *quis custodiet ipsos custodes*? Who guards the guards? Who insures the insurers? Who reinsures the reinsurers?

Harald Cramér

Harald Cramér (1893-1985) was a Swedish mathematician and probabilist of great distinction. In his personal recollections (Cramér, Half a century with probability, *Annals of Prob.*(1976)) he writes, of the period after he obtained his PhD (in 1917, in analytic number theory, under Marcel Riesz): "For a young Swedish mathematician of my generation, who wanted to find a job that would enable him to support a family, it was quite natural to turn to insurance. It was a tradition for Swedish insurance companies to employ highly qualified mathematicians as actuaries ..." (he continues to describe how his actuarial and insurance work led him into probability theory). It is by no means unusual for people to be drawn into a field for such reasons (Doob in probability in the US, and Bartlett and Cox in statistics in the UK, come to mind). In 1929 Cramér became the first holder of the chair in Actuarial Mathematics and Mathematical Statistics at the University of Stockholm – an important event in the development of actuarial mathematics in Scandinavia, and indeed more generally.

The Cramér estimate of ruin (above) of 1930 is perhaps Cramér's most prominent contribution to actuarial and insurance mathematics, and with it the now-standard *Cramér-Lundberg model* in insurance, as we will now call the model above.

§5. Complements

More general processes

The classical Cramér-Lundberg model above is the basic prototype in insurance mathematics, but it is by no means the only one, and is not general enough for all purposes.

1. Non-homogeneous Poisson processes.

These we have met before. Here the Poisson rate $\lambda(t)$ may vary with time. Matters become more complicated, but the theory may be carried through much as before.

2. Cox processes.

These were introduced by D. R. (Sir David) Cox (1924 -) in 1955, under the name *doubly stochastic Poisson process* or *mixed Poisson process*. Here the Poisson rate is *random*. This makes things more flexible and realistic, as well as more complicated.

Perhaps the most important case of a Cox process is where the rate has a *Gamma* distribution, when it is called a *Pólya process*. Recall that the Gamma distribution is the prototype of an error (or noise) distribution on the positive half-line, just as the Normal is on the line. For background here, see Generalised Linear Models (GLMs) in regression, in statistics.

3. Lévy processes.

The compound Poisson process models a situation where we can clearly identify the jumps. But what matters to the company is the flow of cash. For a large company, claims of small (or even ordinary) size may be so numerous as to be treated as 'small change'; it is the *large claims* that predominate, as these can be lethal. Allowing for this, it makes sense to generalise to *Lévy processes* (named after the great French probabilist Paul Lévy (1886 - 1971) for his pioneering work on them in the 1930s). These are stochastic processes with *stationary independent increments*. By the *Lévy-Khintchine formula* and the *Lévy-Itô decomposition*, they may be decomposed into three independent components: (i) a linear deterministic drift (trivial); (ii) a Brownian-motion component; (iii) a sum of jumps (any of these may be absent). The jumps case splits, into (a) only finitely many jumps in finite time (*finite ac-*

tivitiy, FA – the *compound Poisson* case above); (b) infinitely many jumps in finite time (*infinite activity, IA*). The theory can be extended to the Lévy case; for details, see e.g. [Kyp].

4. Gerber-Shiu theory.

This theory (Hans Gerber and Elias Shiu, 1997 and 1998) looks at the financial situation of a company *at ruin* or bankruptcy. This is an important matter!:

(i) The size of the cash reserve just before failure governs how much in the pound (dollar, euro, ...) the creditors will receive.

(ii) The *overshoot* – amount of the deficit which triggers failure – will be used by the liquidators, creditors, regulators etc. to determine whether or to what extent the company was negligent. This has important legal implications. Never forget that it is *illegal* under the Companies Act to trade while insolvent – or to enter into a transaction without the capacity to carry it through. A transaction needs two counter-parties, each willing to trade, and each able to do so. Each has to trust the other here, and inability to complete a deal is a breach of trust here. See e.g. [Kyp, Ch. 10].

Related problems and processes

Fluctuation theory.

The ruin problem above involves the infimum (minimum) over time of the cash balance of the company, or equivalently the supremum (maximum) of the liability. Thus for a process $X := \{X_t\}$, the *supremum* and *infimum* processes are relevant:

$$\overline{X}(t) := \sup\{X_s : s \in [0,t]\}, \qquad \underline{X}(t) := \inf\{X_s : s \in [0,t]\}.$$

Related to these is the *reflected* process, $\overline{X} - X$:

$$(\overline{X} - X)(t) := \overline{X}(t) - X(t) \geq 0.$$

The study of these and related functionals is called *fluctuation theory* – cf. the title of Kyprianou's book [Kyp], where ruin problems are indeed studied.

Queues and dams.

Other areas of Applied Probability involve such functionals and processes, which have to be non-negative, for example, *storage* processes [Kyp Ch. 4] (one cannot store a negative quantity of a commodity, etc.). The classical example here is a *dam*, whose reservoir may run dry – become empty – but which cannot store a negative amount of water. Here the input process of water is often modelled by a Lévy process. Now consider a *queue* – for simplicity, a *single-server* queue. The server is initially idle (say). This *idle period* ends when the first customer arrives for service; the server then works non-stop to serve him, and continues in the same way with any customers who arrive during this *busy period*, and so on. The analogue of the content of the dam (or storage model) is the *workload* facing the server – alternatively, from the customer's point of view, the *virtual waiting time* – the time a customer arriving at time t would have to wait to begin service (a.s. no customer *does* arrive at any given t – but busy facilities such as restaurants, exhibitions etc. often post how long one *would* have to wait if one *did*).

Random walks.

The subject lurking in the background here is that of *random walks* – sums $S_n = X_1 + \cdots + X_n$ of iid random variables. These have an extensive and interesting *fluctuation theory*, developed in the 1950s by Spitzer, Baxter, Sparre Andersen and others.

Duality.

The link between ruin problems and queues etc. lies in *duality* for random walks. In brief, this involves reversing both time and space – looking at the steps of a random walk backwards in time and 'upside down' (see e.g. [Kyp, 3.2]). One can often then transfer from one of the above problem areas to another. This is the most efficient, attractive and modern way to handle the material. Before, results that are now handled easily by duality as above had to be discovered at least twice, and it to be noticed that the relevant distributions were the same. Duality enables one to proceed 'pathwise' – by looking at the random quantities themselves, not just their distributions.

In the ruin problem here, the Net Profit Condition means: to avoid ruin, 'more money in than out'. In queues, this corresponds to the *stability condition*: the server can handle work faster than it comes in (mean service time less than mean inter-arrival time – without this, the server is overwhelmed and the queue 'blows up'). For such a *stable* queue, the workload (or virtual waiting time) has a limit distribution as time increases – the queue settles down to a steady state. The most important case is the $M/G/1$ queue (M: Markov arrival process – Poisson; G: general service-time distribution; 1: single server). Here the limiting waiting-time distribution is given by the classical *Pollaczek-Khintchine formula*. This corresponds to our main result above, the *Cramér estimate of ruin*. For background and details, see [Kyp, 1.3.1, 1.3.2].

Splitting times.

The time at which the maximum over $[0, t]$ is attained is *not a stopping time*: one cannot 'peep into the future' to decide when to quit the gambling table, etc! But, the duality results above show that such times have special properties. If one decomposes the path over $[0, t]$ into the pre- and post-maximum fragments, these are *independent* given where and when the maximum occurs, etc.: the path *splits* at the maximum, in this sense. Such splitting-time arguments are very useful and powerful, and have been systematically exploited by David Williams and others.

Stochastic calculus for jump processes

In Ch. V we developed stochastic (Itô) calculus based on Brownian motion, and applied it in Ch. VI to mathematical finance (Black-Scholes theory). It turns out that this calculus can be extended to the processes with jumps relevant here in Ch. VII on insurance, where the jumps represent the claims. This is technically easier (at least for the Poisson process), but actually came later. It was developed in the context of queueing theory, where the jumps represent customers arriving (or departing). We will be brief; for background and details, see e.g.
D. Applebaum, *Lévy processes and stochastic calculus*, 2nd ed., CUP, 2009 [1st ed. 2004],
P. Brémaud, *Point processes and queues: martingale dynamics*, Springer, 1981.

Recall that the essence of Brownian-based stochastic calculus is captured in the simple equation

$$(dB_t)^2 = dt.$$

The essence of Poisson-based stochastic calculus is similarly captured in

$$(dN_t)^2 = dN_t.$$

For, the change dN_t in a Poisson process $N = (N_t)$ at time t is 0 or 1, and the above expresses that these are the only roots of $x^2 = x$, i.e. $x^2 - x = x(x - 1) = 0$.

The context of Lévy processes in [App] is the simplest natural one containing both the Brownian and the Poisson/compound Poisson cases. But the natural context for stochastic integration is (a lot) more general still – that of *semi-martingales*. These are processes expressible as the sum of a local martingale and a process of (locally) finite variation (FV). The theory here was developed by Paul-André Meyer (1934-2003) and the French (Strasbourg, Paris) school – the 'general theory of processes'.

Non-life insurance: regression and covariates
House insurance

If one insures a house's *contents*, one of the the principal risk factors the insurance company will consider (and the easiest one to measure) is the risk of *burglary*. This varies greatly according to the nature of the area: affluent areas have more to attract a burglar, but tend to have better burglar alarms; poorer areas tend to have higher crime rates, etc. If one insures a house as a *building*, the principal risk factor is *subsidence*. This depends largely on the geological conditions in the area (and so are indicated by the postal code), but also on the quality of the building at the time the area was developed (which can be assessed from past claims). Risk of *fire* is important in both, but harder to assess (it depends on people not leaving chip-pans on the cooker when called to the door or the phone, etc.). These subsidiary bits of information are called *covariates*; the way to use them is called *regression*. The areas of statistics involved are very useful in the actuarial/insurance profession.

Motor insurance

Motor insurance rates vary widely. Of course, the most important single thing is the claims record of the insuring motorist – a good record is worth money, in a no-claims bonus. But, the type of car is also relevant (sports cars are penalised); so is the type of driver (young men are penalised), the annual mileage, the type of use (private or for hire), etc.

Life insurance

Eventual death is certain, so life insurance is largely a matter of covariates such as: age, sex, medical record, profession etc. The tools involved come under Survival Analysis: hazard rates, etc. Following the introduction of the *proportional hazards model* by Cox in 1972, martingale methods have been widely used. This is a very interesting and useful area, but not one we can pursue further here.

To give some flavour of Survival Analysis: suppose that a person survives for time t. What is the chance that he dies by time $t + dt$? With T as the lifetime, with distribution function F on $(0, \infty)$, density f and tail $\overline{F}(x) =$

$1 - F(x)$, this is

$$\begin{aligned} P(T \leq x + dx | T > x) &= P(x < T \leq x + dx)/P(T > x) \\ &= (F(x + dx) - F(x))/(1 - F(x)) \\ &\sim f(x)dx/(1 - F(x)) \\ &= h(x)dx, \end{aligned}$$

say, where $h(x)$ has the interpretation of a *hazard rate.* So

$$h(x) = f(x)/(1 - F(x)).$$

Integrating,

$$1 - F(x) = \exp\{\int_0^x h(u)du\} : \qquad F(x) = 1 - \exp\{\int_0^x h(u)du\}.$$

The simplest case is *constant* hazard rate, λ say, leading to the *exponential* distribution $E(\lambda)$, and so to the *Poisson process* $Ppp(\lambda)$ of VII.2:

$$h(x) \equiv \lambda, \qquad F(x) = 1 - e^{-\lambda x}, \quad (x > 0) : \qquad F = E(\lambda).$$

Now hazard rates vary according to many factors, or covariates: age (older people die out faster than younger ones); medical history; weight, smoking status, occupation, marital status (married people live longer!), etc. So applicants for life insurance will be asked to fill out a form detailing the covariates the insurance company deems relevant; assessing the premium depending on these covariates involves regression, as with the non-life examples above.

Reinsurance

Reinsurers play a major role, in the modern economy, beyond insuring insurers. Reinsurance companies act as de facto *regulators*: they monitor insurers and put a price on their heads. The government need have no say, as 'it's money that talks here'. A good reinsurance premium implies confidence, and makes it easier for the primary insurer to raise capital on the open market. Insurers hold, to cover losses, a mix of cash reserve, investment reserve and reinsurance. (It used to be that the reinsurance pot was biggest, but that is changing as investment becomes more affordable.) The basic fact is that the balance of the three sources of capital is important, and precarious: the reinsurance company watches the cash position of the client like a hawk.

Lender of last resort

Companies may fail, and disappear (leaving debts behind them, as well as lost jobs, etc.). But countries cannot disappear (even though sovereign states have on occasion defaulted on debt, split up, etc.). The ultimate underpinning (in so far as there is one) here is provided by the state, in the form of the central bank – the Bank of England (BoE) in the UK, the Federal Reserve Bank (Fed) in the USA, the European Central Bank (ECB) in the EU, and indeed the World Bank at UN level. The phrase 'lender of last resort' is used to convey this.

Postscript to Ch. VII, Insurance Mathematics

As noted in VII.1, the actuarial profession regulates itself carefully. The Institute of Actuaries sets professional exams, which intending actuaries must pass in order to become qualified. In order to earn exemption by passing a course at university, the university course (particularly its syllabus) must be accredited (validated) by the Institute. (The situation is similar in the accountancy profession.)

The two main centres for actuarial work in the UK are London and Edinburgh. In London, the City University was an early centre, followed later by the London School of Economics (LSE). The LSE's Risk and Stochastics MSc has now become a major producer of actuaries. In Edinburgh, a similar role has long been played by Heriot-Watt University.

As a glance at the skyline in the City of London reveals, London is a major world financial centre. The financial services industry is one of the UK's major industries (thirty years ago manufacturing industry predominated – recall that the UK pioneered the Industrial Revolution – but this is no longer so). Most of the leading UK Mathematics Departments have MSc programmes in Financial Mathematics. I think it is fair to say that UK academia provides well for the needs of the financial services industry. I think it is also fair to say that it provides less well for the needs of the actuarial profession and the insurance industry. This is a great pity (recall from VII.1 the UK's historic leading role here).

I am very pleased that Insurance Mathematics is included in the syllabus for this course. I would urge anyone taking this course who does not already have a clear career path mapped out ahead of them to consider actuarial work (which I would probably have gone into myself had I not been sucked into academia). The work is very useful, and very interesting.

It is worth noting that the boundary between the mathematics of finance (Ch. I-VI) and insurance (Ch. VII) has become quite blurred in recent years. The two areas are no longer separate, as they once were, and the trend towards further interaction will no doubt continue. So it does not have to be an 'either or' choice for you!

NHB