m3f33chI

Chapter I: PROBABILITY BACKGROUND

1. Area: Prelude to measure.

We shall mainly deal with area, as this is two-dimensional. We can draw pictures in two dimensions, and our senses respond to this; paper, whiteboards and computer screens are two-dimensional. By contrast, one-dimensional pictures are much less vivid, while three-dimensional ones are harder (they need the mathematics of perspective – and are called sculptures) – and dimensions higher than four are harder still.

1. *Rectangles*, base b, height h: area $A := bh$.

2. *Triangles*. $A = \frac{1}{2}bh$.

Proof: Drop a perpendicular from vertex to base; then extend each of the two triangles formed to a rectangle and use 1. above.

3. *Polygons*. Triangulate: choose a point in the interior and connect it to the vertices. This reduces the area A to the sum of areas of triangles; use 2. above.

4. *Circles*. We have a choice:

(a) Without calculus. Decompose the circle into a large number of equiangular sectors. Each is approximately a triangle; use 2. above [the approximation boils down to $\sin\theta \sim \theta$ for θ small: Archimedes].

(b) With calculus and plane polar coordinates. Use $dA = dr.rd\theta = rdrd\theta$: $A = \int_0^r \int_0^{2\pi} rdrd\theta = \int_0^r rdr. \int_0^{2\pi} d\theta = \frac{1}{2}r^2.2\pi = \pi r^2$.

Note. The ancient Greeks essentially knew integral calculus – they could do this, and harder similar calculations [volume of a sphere $V = \frac{4}{3}\pi r^3$; surface area of a sphere $S = 4\pi r^2 dr$, etc.; note $dV = Sdr$].

What the ancient Greeks did not have is differential calculus [which we all learned first!] With this, they would have had the idea of velocity, and differentiating again, acceleration. Then they might well have got Newton's Law of Motion, Force = mass × acceleration. This triggered the Scientific Revolution. Had this happened in antiquity, the world would have been spared the Dark Ages and world history would have been completely different!

5. *Ellipses*, semi-axes a, b. Area $A = \pi ab$ (w.l.o.g., $a > b$) [Archimedes].

Proof: cartesian coordinates: $dA = dx.dy$.

Reduce to the circle case: compress ['squash'] the x-axis in the ratio b/a [so $dx \mapsto dx.b/a$, $dA \mapsto dA.b/a$]. Now the area is $A = \pi b^2$, by 4. above. Now 'un-

squash': dilate the x-axis in the ration a/b. So $A \mapsto A.a/b = \pi b^2.a/b = \pi ab$.

Fine – what next? We have already used *both* the coordinate systems to hand. There is no general way to continue this list. Indeed, I don't know another example of comparable neatness and importance to the above.

The only general procedure is to superimpose finer and finer sheets of graph paper on our region, and count squares ('interior squares' and 'edge squares'). This yields numerical approximations – which is all we can hope for, and all we need, in general.

The question is whether this procedure always works. Where it is clearly most likely to fail is with highly irregular regions: 'all edge and no middle'.

It turns out that this procedure does *not* always work; it works for *some but not all* sets – those whose structure is 'nice enough'. This goes back to the 1902 thesis of Henri LEBESGUE (1875-1941):
H. Lebesgue: Intégrale, longueur, aire. *Annali di Mat.* **7** (1902), 231-259.
Similarly in other dimensions. So: some but not all sets have a length/area/volume. Those which do are called *(Lebesgue) measurable*; length/area/volume is called *(Lebesgue) measure*; this subject is called Measure Theory.

We first meet integration in just this context – finding areas under curves (say). The 'Sixth Form integral' proceeds by dividing up the range of integration on the x-axis into a large number of small subintervals, $[x, x + dx]$ say. This divides the required area up into a large number of thin strips, each of which is approximately rectangular; we sum the areas of these rectangles to approximate the area.

This informal procedure can be formalised, as the *Riemann integral* (G. F. B. RIEMANN (1826-66) in 1854). This (basically, the Sixth From integral formalised in the language of epsilons and deltas) is part of the undergraduate Mathematics curriculum.

We see here the essence of calculus (the most powerful single weapon in mathematics, and indeed in science). If something is reasonably smooth, and we break it up finely enough, curves look straight, so we can handle them. We make an error by this approximation, but when calculus applies, this error can be made arbitrarily small, so the approximation is effectively exact.
Example: We do this sort of thing automatically. If in a discussion of global warming we hear an estimate of polar ice lost, this will translate into an estimate of increase in sea level (neglecting the earth's curvature).
Note. The 'squashing' argument above was deliberately presented informally. It can be made quite precise – but this needs the mathematics of *Haar measure*, a fusion of Measure Theory and Topological Groups.

§2. Probability basics: Recapitulation from Years 1 and 2

Think of some economic fundamental — a stock price, or an option on it. These things are difficult to predict — say, in a year's time, with any accuracy. But, successful predictions (at least, more successful than the competition) would make money, so such predictions are much studied. There is uncertainty present. The mathematical machinery we have to model uncertainty, or randomness, is probability; the application area one meets first is statistics. These are important enough to be core material in Years 1 and 2. We will need background knowledge of such things as:

Terminology and notation
Random variables; notation such as X.
Expectation (mean), $E[X]$, as a measure of *location*;
variance (variability), $varX$, as a measure of *randomness, dispersion or spread*.
Distribution function, $F(x) = F_X(x) := P(X \leq x)$ (distribution; d/n; law);
densities, $f(x)$: $F(x) = \int_{-\infty}^{x} f(u)du$.
Expectation

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)dF(x).$$

This is true quite generally (as a Lebesgue-Stieltjes integral: see Ch. III). If F has density f, this is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

— so '$dF(x) = f(x)dx$' in the density case). In the discrete case, if X takes values x_n with probability a_n, this is a summation:

$$E[g(X)] = \sum a_n g(x_n).$$

These formulae will be familiar to you from Years 1 and 2; we will deal with both together, and pass to a much better integral (there are several!) in Ch. III.
Several random variables: (X, Y), etc. Joint and marginal distributions: $F(x, y) = F_{X,Y}(x, y) := P(X \leq x, Y \leq y)$ (joint: governs behaviour of X, Y

together; marginals: govern their behaviour *separately*).
Covariances: $cov(X,Y) := E[(X - E[X])(Y - E[Y])]$;
correlation $\rho = \rho(X,Y) := cov(X,Y)/\sqrt{var\ X.var\ t.}$
So: $E[X]$ has the dimensions of X; $var\ X$ has the dimensions of X^2; $cov(X,Y)$ has those of XY; $corr(X,Y$ is dimensionless, and takes values in $[-1,1]$ (Cauchy-Schwarz inequality).
$cov(X,Y) = E[XY] - E[X].E[Y]$; $var\ X = cov(X,X) = E[X^2] - (E[X])^2$, $var\ X \geq 0$; $var\ X = 0$ iff X is constant a.s. (almost surely — with probability one), and then the constant is $E[X]$.

Independence

A family of random variables $X_1, \cdots, X_n$ are *independent* if, for any subfamily $X_{r_1}, \cdots, X_{r_k}$, and any intervals I_i (more generally, measurable sets: Ch. III),

$$P(X_{r_1} \in I_1, \cdots, X_{r_k} \in I_k) = P(X_{r_1} \in I_1) \cdots P(X_{r_k} \in I_k).$$

That is, knowing how the Xs behave *separately* (as on the RHS) tells us how they behave *together* (as on the LHS).
Taking the $I_i = (-\infty, x_i]$:
the Xs are independent iff their *joint* distribution function *factorises* into the product of the *marginal* distribution functions.
When the densities exist, differentiating wrt $x_1, \cdots, x_n$ gives: the Xs are independent iff their *joint* density function *factorises* into the product of the *marginal* density functions.
Similarly for mass functions, in the discrete case.
Similarly for the various transforms we shall need; see below.

Multiplication Theorem: if X, Y are independent and have means,

$$E[XY] = E[X].E[Y].$$

Proof.

$$\begin{aligned} E[XY] &= \int\int xy dF_{X,Y}(x,y) = \int\int xyd(F_X(x)F_Y(y)) \\ &= \int xdF_X(x) \int ydF_Y(y) = E[X].E[Y]. \end{aligned}$$

Note the notation, and compare it with what you may have met in calculus of several variables. We use $\int\int$ rather than $\int$, to emphasise that this is

(like) a double integral in calculus courses: we usually evaluate such a double integral by integrating out first over one variable and then over the other, so a double integral is reduced to two repeated single integrals. We will see how to do this using Lebesgue-Stieltjes integrals (better!) in Ch. III. But on the first RHS above, the one d means 'integrate w.r.t.' one two-dimensional function (or, in Ch. III, *measure*).

Discrete and continuous

The most basic case is when random variables X take only discrete values (finitely or countably infinitely many — it usually doesn't matter which), x_n say. Then if

$$a_n := P(X = x_n),$$

the sequence $a = (a_n)$ is called the (probability) *mass function*. Think of probability as mass; there is mass one altogether; in the discrete case, the mass function divides it up over the possible values x_n; we can calculate all probabilties by it summing:

$$P(A) = P(X \in A) = \sum_{n:x_n \in A} a_n = \sum_{n:x_n \in A} P(X = x_n).$$

Summation is easy! All this is completely elementary. But, it largely suffices for the mathematics of insurance (Ch. VIII), essentially because we can *count* insurance claims, and also for mathematical finance in discrete time (Ch. V), when we can usually take the state space discrete also.

3. Transforms

Characteristic functions (CFs).
Definition. For a random variable X with distribution function F, the *characteristic function (CF)* is

$$\phi(t), \text{ or } \phi_X(t), := E[e^{itX}]$$

(here t, X are real and $i = \sqrt{-1}$; everything is real unless we say otherwise). The great thing about the CF is that it *always exists* (converges): $|e^{itX}| \leq 1$, and $|\int .| \leq \int |.|$, so also $|E[.]| \leq E[|.|]$, so

$$|\phi(t)| = |E[e^{itX}]| \leq E[|e^{itX}|] = E[1] = 1 \ (< \infty!)$$

If X, Y are independent with d/ns F, G, their sum $X+Y$ has d/n H, where H is the *convolution*

$$H = F * G :$$

$$\begin{aligned} H(z) &:= P(X+Y \leq z) = \int_{x+y \leq z} dF(x)dG(y) = \int_{-\infty}^{\infty} dF(x) \int_{\infty}^{z-x} dG(y) \\ &= \int_{-\infty}^{\infty} G(z-x)dF(x) = \int_{-\infty}^{\infty} F(z-y)dG(y), \end{aligned}$$

by symmetry. If F, G have densities f, g, H has density $h := f * g$, the *convolution* of f, g, where

$$h(z) := (f*g)(z) = \int_{-\infty}^{\infty} g(z-x)f(x)dx = \int_{-\infty}^{\infty} f(z-y)g(y)dt.$$

Technically, the CF is the *Fourier-Stieltjes transform* of the d/n F, and the *Fourier transform* of the density f when this exists.

We are constantly *averaging* independent readings, so *summing*. Summing numbers (e.g. values of random variables) is easy, even for lots of them. But this corresponds to *convolution* for distributions, and for n summands, this involves $n-1$ integrations – awkward for large n! By contrast, the CF makes things easy:
the CF of an independent sum is the product of the CFs.
Proof. The CF of $X+Y$ is

$$\begin{aligned} \phi_{X+Y}(t) &:= E[e^{it(X+Y)}] \\ &= E[e^{itX}.e^{itY}] \quad \text{(property of exponentials)} \\ &= E[e^{itX}].E[e^{itY}] \quad \text{(independence: Multiplication Theorem)} \\ &= \phi_X(t).\phi_Y(t). \end{aligned}$$

There are several variants of this:

Moment-generating functions (MGFs).

The MGF of a rv X is

$$M(t) := E[e^{tX}].$$

This looks just like the CF, except that we have no i. This looks simpler – but it is in fact more complicated. For, $|e^{itX}| \leq 1$, so all the expectations

above converge. But, as e^{tX} may grow exponentially, the MGF may fail to be defined for all real t – may diverge to $+\infty$ for some t. However, when it does exist, we can expand the exponential and take expectations termwise: with

$$\mu_n := E[X^n]$$

the *nth moment* of X,

$$M(t) = E[e^{tX}] = E[\sum_0^\infty t^n X^n/n!] = \sum_0^\infty t^n E[X^n]/n! = \sum_0^\infty t^n \mu_n/n!.$$

The one function $M(t)$ on the left is said to *generate* the infinitely many moments μ_n on the right, hence the name MGF.

Moments.

Recall from Complex Analysis (M2P3) that an analytic function determines and is determined by its power-series expansions. So if power-series expansion of $M(t)$ above about the origin has radius of convergence $R > 0$, we can find the moments by differentiation:

$$\mu_n = E[X^n] = M^{(n)}(0).$$

Similarly for CFs: remembering the i: $i^n \mu^{(n)} = \phi^{(n)}(0)$ (here things always converge, so we don't have to worry about R).

Laplace-Stieltjes transform (LST).

When $X \geq 0$, we know that $e^{-sX} \leq 1$ for $s \geq 0$, so $E[e^{-sX}] \leq 1$; in particular, the expectation (= integral) always exists, unlike with the MGF. So, the LST of $X \geq 0$ is defined as

$$\psi(s) := E[^{-sX}] \quad (s \geq 0).$$

Note. The distinction between Fourier and Laplace transforms vanishes if one uses complex variables – which is what one should do anyway. This goes back to the classic book of Paley and Wiener:
R. E. A. C. PALEY & N. Wiener, *Fourier transforms in the complex domain*, AMS, 1934.

Probability generating functions (PGFs).

If X is not only non-negative but also integer-valued, we can specialise

the LST above to the PGF (replace e^{-s}, $s \geq 0$, by $s \in [0,1]$: if X takes values in $\mathbb{N}_0$, with

$$P(X = n) = p_n \quad (n = 0, 1, 2, \cdots),$$

the PGF of X is

$$P(s) := E[s^X] = \sum_0^\infty p_n s^n;$$

as $P(1) = \sum_0^\infty p_n = 1$ ($< \infty$!), the radius of convergence of the power series here is always at least 1.
As with CF, so with MGF, LST and PGF: the transform (convolution) of an independent sum is the product of the transforms, and similarly for moments.

4. 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

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

So this law has density

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

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

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

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

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

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.

The Gamma function.

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.

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 $W_1, W_2, \ldots W_n$ 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 $W_1, W_2, \ldots W_n$ 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):

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

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. //

The compound Poisson process.

If in an insurance company (Ch. VIII) claims arrive in a Poisson process of rate λ, and are independent, of each other and the arrival process, with distribution F, the random sums

$$Y_t := X_1 + \cdots + X_{N(t)}$$

form a process $Y = (Y_t)$, called a *compound Poisson process*, $CP(\lambda, F)$. For the distribution, mean and variance of Y_t, see Problems 1.