

# Probability Models for Insurance Losses

Stephen J. Mildenhall

2020-10-20

## 1 Probability Models for Insurance Losses

### 1.1 Introduction

Losses have a complicated, dynamic structure. An insurance portfolio's loss depends on the expected loss rate and the length of time it is exposed, analogous to distance equals speed multiplied by time. Total losses  $L = L(e, t)$  over a period  $(0, t]$  are a function of exposure and time. Exposure is measured in homogeneous units, each with the equal expected loss rates. What is the best way to model  $L(e, t)$ ?

Is it reasonable to assume losses occur at a constant rate? Generally, not. Even for earthquake losses, where events appear perfectly random, the occurrence time influences losses. Did the loss occur during the commute, working hours, nighttime, or over a weekend? There are two ways to address this reality. One, over the time frame of an insurance policy the rate averages out and can be treated as constant. Two, use a time change to equalize the claim rate. The time change, called an operational time, runs time faster during high expected loss periods and slower during low. It provides the perfect clock against which to earn premium. As always, what matters is the deviation from expected. We proceed assuming a constant expected rate of claim occurrence, justified by either of these rationales.

If insured risks are independent, we could posit that losses over  $(0, t_1 + t_2]$  should have the same distribution as the sum of independent evaluations over  $(0, t_1]$  and  $(0, t_2]$ . But there could be good reasons why this is not the case. Weather, for example, is strongly autoregressive: weather today is the best predictor of the weather tomorrow. Failure of independence manifests itself in an inter-insured and inter-temporal loss dependency or correlation. However, using common operational time and economic drivers can explain much of this correlation; see [1] and the work of Glenn Meyers on contagion.

Having stipulated caveats for expected loss rate and dependency, various loss models present themselves. A homogeneous model says  $L(e, t) = eR_t$  where  $R_t$  represents a common loss rate for all exposures. This model is appropriate for a stock investment portfolio, with position size  $e$  and stock return  $R_t$ . It implies the CV is independent of volume.

A homogeneous model has  $L(e, t) \sim D(et, \phi)$ , a distribution with mean  $m = et$  and additional parameters  $\phi$ . A GLM then models  $m$  as a function of a linear combination of covariates, often with  $e$  as an offset. We can select the time period so that the expected

loss from one unit of exposure in one period of time equals 1. In this model, time and exposure are interchangeable. The loss distribution of a portfolio twice the size is the same as insuring for twice as long. Thus we can merge loss rate and time into one variable  $\nu$  and consider  $L = L(\nu)$ . If  $\nu$  is an integer then  $L(\nu) = \sum_{1 \leq i \leq \nu} X_i$  where  $X_i \sim L(1)$  are independent. Therefore  $\text{Var}(L(\nu)) = \nu \text{Var}(L(1))$  and the CV of  $L(\nu)$  is  $CV(L(1))/\sqrt{\nu}$ . The CV of losses are not independent of volume, in contrast to the investment modeling paradigm. The actuary knows this: the shape of the loss distribution from a single policy (highly skewed, large probability of zero losses) is very different to the shape for a large portfolio (more symmetric, very low probability of zero losses.)

As always, reality is more complex. [2] discusses five alternative forms for  $L(e, t)$ . It shows that only a model with random operational time is consistent with NAIC Schedule P data, after adjusting for the underwriting cycle.

The focus of this section is to determine which distributions are appropriate for modeling homogeneous blocks with a constant loss rate relative to (operational) time. These distributions are used in predictive modeling for ratemaking and reserving, for example. Other applications, such as capital modeling and Enterprise Risk Management, model heterogeneous portfolios where understanding dependencies becomes critical. These are not our focus and are therefore ignored, despite their importance.

The ability to embed a static loss model  $L$  into a family, or stochastic process, of models  $L(\nu)$  appropriate for modeling different volumes of business over different time frames is an important concept that is introduced in this section. Dynamic modeling leads to a compound Poisson *process* and, more generally, a Lévy process. We begin by recalling the aggregate loss model of collective risk theory.

## 1.2 The Aggregate Loss Model

An **aggregate loss model** random variable has the form  $Y = X_1 + \dots + X_N$  where  $N$  models claim count (frequency) and  $X_i$  are independent, identically distributed (iid) severity variables. Aggregate loss models have an implicit time dimension:  $N$  measures the number of claims over a set period of time, usually one year. Aggregate loss models are the basis of the collective risk model and are central to actuarial science.

Expected aggregate losses equal expected claim count times average severity,  $\mathbf{E}[Y] = \mathbf{E}[N]\mathbf{E}[X]$ . There is a tricky formula for the variance of  $Y$ ; here is an easy way to remember it. If  $X = x$  is constant then  $Y = xN$  has variance  $x^2\text{Var}[N]$ . If  $N = n$  is constant then  $X_1 + \dots + X_n$  has variance  $n\text{Var}(X)$  because  $X_i$  are independent. The variance formula must specialize to these two cases and therefore, replacing  $n$  with  $\mathbf{E}[N]$  and  $x$  with  $\mathbf{E}[X]$ , suggests

$$\text{Var}(Y) = \mathbf{E}[X]^2\text{Var}(N) + \mathbf{E}[N]\text{Var}(X)$$

is a good bet. By conditioning on  $N$  you can check it is correct.

## 1.3 Compound Poisson Distributions

A **compound Poisson** (CP) random variable is an aggregate loss model where  $N$  has a Poisson distribution. Let  $\text{CP}(\lambda, X) = X_1 + \dots + X_N$  where  $N$  is a Poisson with mean  $\lambda$  and  $X_i \sim X$  are iid severities. We shall see that the Tweedie family is a CP where  $X_i$  have a gamma distribution.

By the results in the previous section  $E[\text{CP}(\lambda, X)] = \lambda E[X]$ . Since  $\text{Var}(N) = E[N]$  for Poisson  $N$  and  $\text{Var}(X) = E[X^2] - E[X]^2$ , we get

$$\text{Var}(\text{CP}(\lambda, X)) = E[N]E[X^2].$$

CP distributions are a flexible and tractable class of distributions with many nice properties. They are the building block for almost all models of loss processes that occur at discrete times.

## 1.4 Diversifying and Non-Diversifying Insurance Growth

We want to understand the relationship between the variance of losses and the mean loss for an insurance portfolio as the mean varies, i.e., the variance function. Is risk diversifying, so larger portfolios have relatively lower risk, or are there dependencies which lower the effectiveness of diversification? Understanding how diversification scales has important implications for required capital levels, risk management, pricing, and the optimal structure of the insurance market. Let's start by describing two extreme outcomes using simple CP models. Assume severity  $X$  is normalized so  $E[X] = 1$  and let  $x_2 := E[X^2]$ .

Consider  $\text{CP}_1(\mu, X)$ . The variance function is  $V(\mu) = \text{Var}(\text{CP}_1) = \mu x_2$ , which grows linearly with  $\mu$ .  $\text{CP}_1$  models **diversifying** growth in a line with severity  $X$ . Each risk is independent: increasing expected loss increases the expected claim count but leaves severity unchanged.  $\text{CP}_1$  is a good first approximation to growth for a non-catastrophe exposed line.

Alternatively, consider  $\text{CP}_2(\lambda, (\mu/\lambda)X)$  for a fixed expected event count  $\lambda$ . Now, the variance function is  $V(\mu) = \text{Var}(\text{CP}_2) = \lambda(\mu/\lambda)^2 x_2 = \mu^2(x_2/\lambda)$  grows quadratically with  $\mu$ . The distribution of the number of events is fixed, as is the case in a regional hurricane or earthquake cover. Increasing portfolio volume increases the expected severity for each event.  $\text{CP}_2$  is a model for **non-diversifying** growth in a catastrophe-exposed line. It is the same as the position size-return stock investment model. Risk, measured by the coefficient of variation, is constant, equal to  $\sqrt{x_2/\lambda}$ , independent of volume. This is a homogeneous growth model.

We will see in Part IV that the Tweedie family of compound Poisson distributions interpolates between linear and quadratic variance functions.

A more realistic model adds uncertainty in  $\mu$ , resulting in a variance function of the form  $\mu(a + b\mu)$  which is diversifying for small  $\mu$  and homogeneous for larger  $\mu$ .

## 1.5 The Moment Generating Function of a CP

Many useful properties of CP distributions are easy to see using MGFs. To compute the MFG of a CP we first to compute the MGF of a Poisson distribution. We need two facts: the Poisson probability mass function  $\Pr(N = n) = e^{-\lambda} \lambda^n / n!$  and the Taylor series

expansion of the exponential function  $e^x = \sum_{n \geq 0} x^n/n!$ . Then

$$\begin{aligned}
M_N(t) &= \mathbf{E}[e^{tN}] \\
&= \sum_{n \geq 0} e^{tn} \frac{e^{-\lambda} \lambda^n}{n!} \\
&= e^{-\lambda} \sum_{n \geq 0} \frac{(e^t \lambda)^n}{n!} \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= \exp(\lambda(e^t - 1))
\end{aligned}$$

We can now compute the MGF of  $\text{CP}(\lambda, X)$  using conditional expectations and the identity  $x^n = e^{n \log(x)}$ :

$$\begin{aligned}
M_{\text{CP}(\lambda, X)}(t) &= \mathbf{E}[e^{t \text{CP}(\lambda, X)}] \\
&= \mathbf{E}_N[\mathbf{E}_X[e^{t(X_1 + \dots + X_N)} \mid N]] \\
&= \mathbf{E}_N[(\mathbf{E}_X[e^{tX}])^N] \\
&= \mathbf{E}_N[M_X(t)^N] \\
&= \mathbf{E}_N[e^{N \log(M_X(t))}] \\
&= M_N(\log(M_X(t))) \\
&= \exp(\lambda(M_X(t) - 1)).
\end{aligned}$$

A Poisson variable can be regarded as a CP with degenerate severity  $X \equiv 1$ , which has MGF  $e^t$ . Thus the formula for the MGF of a CP naturally extends the MGF of a Poisson variable.

An example showing the power of MGF methods comes from the problem of **thinning**. We want to count the number of claims excess  $x$  where ground-up claims have a distribution  $\text{CP}(\lambda, X)$ . Take severity to be a Bernoulli random variable  $B$  with  $\Pr(B = 1) = \Pr(X > x) = p$  and  $\Pr(B = 0) = 1 - p$ .  $B$  indicates if a claim exceeds  $x$ .  $\text{CP}(\lambda, B)$  counts the number of excess claims. It is called a thinning of the original Poisson. It is relatively easy to see that  $\text{CP}(\lambda, B) \sim \text{Poisson}(\lambda p)$  using the conditional probability formula  $\Pr(\text{CP} = n) = \sum_{k \geq n} \Pr(\text{CP} = n \mid N = k) \Pr(N = k)$  (exercise). But it is very easy to see the answer using MGFs. By definition  $M_B(t) = (1 - p) + pe^t$ , and therefore

$$\begin{aligned}
M_{\text{CP}}(t) &= M_N(\lambda(M_B(t) - 1)) \\
&= M_N(\lambda((1 - p) + pe^t - 1)) \\
&= \exp(\lambda p(e^t - 1)).
\end{aligned}$$

The MGF of a mixture is the weighted average of the MGFs of the components. If  $\bar{X}$  is a  $p_1, p_2, p_1 + p_2 = 1$ , mixture of  $X_1$  and  $X_2$ , so  $F_{\bar{X}}(x) = p_1 F_1(x) + p_2 F_2(x)$ , then

$$\begin{aligned}
M_{\bar{X}}(t) &= \int e^{tx} (p_1 f_{X_1}(x) + p_2 f_{X_2}(x)) dx \\
&= p_1 \int e^{tx} f_{X_1}(x) dx + p_2 \int e^{tx} f_{X_2}(x) dx \\
&= p_1 M_{X_1}(t) + p_2 M_{X_2}(t).
\end{aligned}$$

This fact allows us to create discrete approximation CPs and convert between a frequency-severity and a jump severity distribution view. It is helpful when working with catastrophe model output.

## 1.6 The Addition Formula

CP distributions satisfy a simple **addition formula**

$$\text{CP}(\lambda_1, X_1) + \text{CP}(\lambda_2, X_2) \sim \text{CP}(\lambda, \bar{X})$$

where  $\lambda = \lambda_1 + \lambda_2$  and  $\bar{X}$  is a mixture of independent  $X_1$  and  $X_2$  with weights  $\lambda_1/\lambda$  and  $\lambda_2/\lambda$ . The addition formula is intuitively obvious: think about simulating claims and remember two facts. First, that the sum of two Poisson variables is Poisson, and second, that to simulate from a mixture, you first simulate the mixture component based on the weights and then simulate the loss from the selected component.

It is easy to demonstrate the the addition formula using MGFs:

$$\begin{aligned} M_{\text{CP}(\lambda_1, X_1) + \text{CP}(\lambda_2, X_2)}(t) &= M_{\text{CP}(\lambda_1, X_1)}(t) M_{\text{CP}(\lambda_2, X_2)}(t) \\ &= \exp(\lambda_1(M_{X_1}(t) - 1)) \exp(\lambda_2(M_{X_2}(t) - 1)) \\ &= \exp(\lambda([\lambda_1/\lambda]M_{X_1}(t) + [\lambda_2/\lambda]M_{X_2}(t)) - 1) \\ &= \exp(\lambda(M_{\bar{X}}(t) - 1)) \\ &= M_{\text{CP}(\lambda, \bar{X})}(t). \end{aligned}$$

The addition formula is the MGF for a mixture in reverse. To sum of multiple CPs you sum the frequencies and form the frequency-weighted mixture of the severities.

The addition formula is handy when working with catastrophe models. catastrophe models output a set of events  $i$ , each with a frequency  $\lambda_i$  and a loss amount  $x_i$ . They model the number of occurrences of each event using a Poisson distribution. Aggregate losses from the event  $i$  are given by  $\text{CP}(\lambda_i, x_i)$  with a degenerate severity.

It is standard practice to normalize the  $\lambda_i$ , dividing by  $\lambda = \sum_i \lambda_i$ , and form the empirical severity distribution  $\hat{F}(x) = \sum_{i: x_i \leq x} \lambda_i/\lambda$ .  $\hat{F}$  is the  $\lambda_i$  weighted mixture of the degenerate distribution  $X_i = x_i$ . By the addition formula, aggregate losses across all events are  $\text{CP}(\lambda, \hat{X})$  where  $\hat{X}$  has distribution  $\hat{F}$ .

## 1.7 Lévy Processes and the Jump Size Distribution

The CP model of insurance losses is very compelling. The split into frequency and severity mirrors the claims process. However, we could take a different approach. We could specify properties an insurance loss model must have, and then try to determine all distributions satisfying those properties. In the process, we might uncover a new way of modeling losses. The latter approach has been used very successfully to identify risk measures. For example, coherent risk measures emerge as those satisfying a set of axioms.

As we discussed in the introduction, in a basic insurance model losses occur homogeneously over time at a fixed expected rate. As a result, losses over a year equal the sum of twelve identically distributed monthly loss variables, or 52 weekly, or 365 daily variables, etc. If the random variable  $Y_1$  represents losses from a fixed portfolio insured for one year, then for any  $n$  there are independent, iid variables  $X_i$ ,  $i = 1, \dots, n$  so that  $Y_1 = X_1 + \dots + X_n$ . Random variables with this divisibility property for every  $n \geq 1$  are called **infinitely divisible** (ID). The Poisson, normal, gamma, negative binomial, lognormal, and Pareto

distributions are all ID. This is obvious for the first four from the form of their MGFs. For example, the Poisson MGF shows

$$e^{\lambda(e^t-1)} = \{e^{\frac{\lambda}{n}(e^t-1)}\}^n$$

so the Poisson is a sum of  $n$  Poisson variables for any  $n$ . The same applies to a CP. Proving the lognormal or Pareto is ID is much trickier. A distribution with finite support, such as the uniform or binomial distributions, is not ID. A mixture of ID distributions is ID if the mixture distribution is ID.

Any infinitely divisible distribution  $Y$  can be embedded into a special type of stochastic process called a Lévy process so that  $Y$  has the same distribution as  $Y_1$ . The process  $Y_t$  shows how losses occur over time, or as volume increases, or both.  $Y_t$  is a **Lévy process** if it has independent and identically distributed increments, meaning the distribution of  $Y_{s+t} - Y_s$  only depends on  $t$  and is independent of  $Y_s$ . Lévy processes are Markov processes: the future is independent of the past given the present. The Lévy processes representation is built-up by determining the distribution of  $Y_\epsilon$  for a very small  $\epsilon$  and then adding up independent copies to determine  $Y_t$ . The trick is to determine  $Y_\epsilon$ .

Lévy processes are very well understood. See [3] for a comprehensive reference. They are the sum of a deterministic drift, a Brownian motion, a compound Poisson process for large losses, and a process that is a limit of compound Poisson processes for small losses. All of four components do not need to be present. Insurance applications usually take  $Y_t$  to be a compound Poisson process  $Y_t = X_1 + \dots + X_{N(t)}$ , where  $X_i$  are iid severity variables and  $N(t)$  is a Poisson variable with mean proportional to  $t$ .

When a normal, Poisson, gamma, inverse Gaussian, or negative binomial distribution is embedded into a Lévy processes, all the increments have the same distribution, albeit with different parameters. This follows from the form of their MGFs, following the Poisson example above. But it is not required that the increments have the same distribution. For example, when a lognormal distribution is embedded the divided distributions are not lognormal—it is well known (and a considerable irritation) that the sum of two lognormals is not lognormal.

The structure of Lévy processes implies that the MGF of  $X_t$  has the form

$$\mathbb{E}[e^{sX_t}] = \exp(t\kappa(s))$$

where  $\kappa(s) = \log \mathbb{E}[e^{sX_1}]$  is cumulant generating function of  $X_1$ . To see this, suppose  $t = m/n$  is rational. Then, by the additive property  $\mathbb{E}[e^{sX_{m/n}}] = \mathbb{E}[e^{sX_{1/n}}]^m$  and  $\mathbb{E}[e^{sX_1}] = \mathbb{E}[e^{sX_{1/n}}]^n$ , showing  $\mathbb{E}[e^{sX_1}]^{1/n} = \mathbb{E}[e^{sX_{1/n}}]$ . Thus  $\mathbb{E}[e^{sX_{m/n}}] = \mathbb{E}[e^{sX_1}]^{m/n}$ . The result follows for general  $t$  by continuity. Therefore, the cumulant generating function of  $X_t$  is  $\kappa_t(s) = t\kappa(s)$ .

## 1.8 Jump Densities and Infinite Activity CPs

Instead of a frequency and severity split, we could focus on the frequency of loss by size. In the CP model, this involves replacing the frequency  $\lambda$  and loss distribution  $F$  with the single loss frequency jump distribution  $\lambda F$ . The expected frequency of a loss in the range  $(x_1, x_2]$  is  $\lambda(F(x_2) - F(x_1))$ . In general, a **jump distribution**  $J(x)$  equals the expected frequency of jumps of size  $> x$ . The jump distribution goes by many names: the loss

frequency curve [4], the event frequency or exceeding probability curve in catastrophe models [5], or a Lévy measure in probability theory [3], after Paul Lévy (1886-1971). Frequency by size of loss is more relevant than event level severity for structuring a reinsurance or risk management program.

We illustrate the idea of jump distributions for  $x \geq 0$ , which is what we need for insurance losses. The theory is easy to extended to positive and negative jumps, but we don't need that generality.

The jump distribution is analogous to a survival function, except  $J(0)$  equals the total number of claims rather than 1. By definition,  $J$  must be a non-increasing function of  $x$ . There is no a priori reason that  $J(0) < \infty$ . If it is, then  $J(x)/J(0)$  is the survival function for a severity distribution and we are back to a CP. Reconstructing frequency and severity is not possible when  $J(0) = \infty$  because it involves dividing by infinity. Can we make rigorous sense of the case where  $J(0) = \infty$ ? To that end, we need to determine if there are any restrictions at all on  $J$ . It turns out, there are.

First, there can be only finitely many jumps of size  $\geq 1$ . If there were an infinite number of such jumps then we would always get an infinite loss—not good for insurance. The value 1 is an arbitrary threshold for **large**; it could be any  $\epsilon > 0$ . Thus we require that  $J(1) < \infty$  for a valid jump distribution: **the expected number of large claims is finite**. There can only be infinitely many claims because of the number of *small* claims. Symbolically,  $\lim_{x \downarrow 0} J(x) \rightarrow \infty$  is valid but  $J(x) < \infty$  for all  $x > 0$ .

The possibility that  $J(0) = \infty$  explains why  $J(x)$  counts the number of jumps greater than  $x$ . We can't define the distribution as the number of losses  $\leq x$  because that could be infinite for all  $x$ .

Second, **the expected aggregate accumulation of all small jumps must be finite**. Suppose  $J$  has a density  $j$ , meaning there is a function  $j$  so that

$$J(x) = \int_x^\infty j(t) dt.$$

Then we require

$$\int_0^1 x j(x) dx < \infty.$$

to ensure accumulated small (attritional) losses are finite. In contrast, the expected *accumulation* of large claims,  $\int_1^\infty x j(x) dx$ , can be infinite provided the *number* of claims  $J(1)$  is finite. For example,  $\text{CP}(1, X)$ , where  $\mathbb{E}[X] = \infty$ , has an unlimited expected accumulation of loss but is still a valid CP model.

Assuming these two conditions hold, here's how to define an **infinite activity compound Poisson** (IACP). If  $J$  has an infinite total number of jumps it is because  $J(x) \rightarrow \infty$ , and hence  $j(x) \rightarrow \infty$ , as  $x \rightarrow 0$ . To manage a potential infinite frequency of small losses, define a capped version of  $j$  by  $j_n(x) = j(x) \wedge n$  for  $n = 1, 2, 3, \dots$  and let  $J_n$  be the corresponding distribution. By construction  $J_n(0) < \infty$  so we can find a loss random variable  $X_n$  with severity distribution  $F_n(x) := 1 - J_n(x)/J_n(0)$ . Then define

$\text{CP}_n := \text{CP}(J_n(0), X_n)$ . We can decompose the MGF of  $\text{CP}_n$  as

$$\begin{aligned} M_n(t) &= \exp\left(J_n(0) \int_0^\infty (e^{tx} - 1) \frac{j_n(x)}{J_n(0)} dx\right) \\ &= \exp\left(\int_0^\infty (e^{tx} - 1) j_n(x) dx\right) \\ &= \exp\left(\int_1^\infty (e^{tx} - 1) j_n(x) dx\right) \times \exp\left(\int_0^1 (e^{tx} - 1) j_n(x) dx\right). \end{aligned}$$

Because there are only finitely many large losses, the first term is a CP for all  $n$  and converges to the  $Y_l := \text{CP}(J(1), X_l)$  where  $X_l$  has density  $j(x)/J(1)1_{[1,\infty)}$ —it is made by disregarding all the losses size  $x \leq 1$ . Because the expected accumulation of small jumps is finite the integral in the second term also converges

$$\int_0^1 (e^{tx} - 1) j_n(x) dx \rightarrow \int_0^1 (e^{tx} - 1) j(x) dx.$$

As a result of the general theory of MGFs<sup>1</sup>, the second term converges to the MGF of a distribution  $Y_s$ .  $Y = Y_l + Y_s$  is the IACP we want. It clearly extends a CP: if  $j(x)$  is bounded then  $j_n(x) = j(x)$  for  $n$  sufficiently large. But it also allows us to define an extended CP with unbounded  $j$ , such as  $j(x) = x^{-3/2}$ . It does not allow  $j(x) = x^{-5/2}$  since then the accumulation of small claims is infinite.

The gamma distribution is an example of an IACP. Since a gamma has density is 0 at 0 it cannot be a CP. A non-negative CP always has a probability mass  $e^{-\lambda} > 0$  at  $x = 0$ . We can derive the jump density for a gamma distribution from the MGF for a gamma using some trickery due to Frullani. A  $\Gamma(\alpha, \beta)$  distribution has density

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where  $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} e^{-x} dx$ . The MGF is

$$M(t) = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x(\beta-t)} dx = \left(\frac{\beta}{\beta-t}\right)^\alpha = \left(1 - \frac{t}{\beta}\right)^{-\alpha}.$$

---

<sup>1</sup>Lévy's Continuity Theorem says that a sequence of random variables converges in distribution iff their characteristic functions converge pointwise to a function that is continuous at the origin. That function is then the characteristic function of the limit.



Therefore the gamma cumulant generating function can be expressed as

$$\begin{aligned}
\kappa(t) &= -\alpha \log \left( 1 - \frac{t}{\beta} \right) \\
&= \alpha \int_0^t \frac{dz}{\beta - z} \\
&= \alpha \int_0^t \int_0^\infty e^{-(\beta-z)s} ds dz \\
&= \alpha \int_0^\infty e^{-\beta s} \int_0^t e^{zs} dz ds \\
&= \alpha \int_0^\infty e^{-\beta s} \frac{e^{st} - 1}{s} ds \\
&= \alpha \int_0^\infty (e^{st} - 1) \frac{e^{-\beta s}}{s} ds
\end{aligned}$$

which is the MGF for an IACP with jump density

$$j(x) = \alpha \frac{e^{-\beta x}}{x}.$$

Since  $e^{-\beta x} \approx 1$  for small  $x$ , the corresponding jump distribution has an infinite expected claim count ( $\int_0^1 \frac{dx}{x} = \infty$ ). We will see related jump distributions again in Part IV; they underlie all the Tweedie and power variance function NEFs. (Note that the last integral in the derivation of  $\kappa$  is only valid for  $0 \leq t < \beta$  and that it can't be separated into two. It relies on the fact that  $e^{st} - 1 \approx st + O(s^2)$  to cancel  $e^{-\beta s} s^{-1} \approx s^{-1}$  as  $s \downarrow 0$ .)

IACP distributions with jump distribution proportional to  $J(x) = x^{-\alpha}$ ,  $0 < \alpha < 1$ , are called **extreme stable** distributions. (The associated jump densities are  $j(x) = \alpha x^{-\alpha-1}$ .) They are extremely thick tailed and have no mean, and hence no variance. They are called **stable** because they are stable under addition: they have the property that  $X_1 + X_2 \sim 2^{1/\alpha} X$  where  $X, X_1, X_2$  are iid. The normal distribution is the best known stable distribution—in fact the only stable distribution with a finite variance. For the normal the scaling constant is  $\alpha = 2$ . They are called **extreme** because they are maximally asymmetric within the class of stable distributions—having only positive jumps—and not because they are related to extreme value theory. There are analogous distributions with both positive and negative jumps, which are not extreme, and another extreme family with only negative jumps. Stable distributions have simple characteristic functions but no simple closed-form density or distribution.

The IACP with jump distribution proportional to  $x^{-1/2}$  is called a Lévy stable  $\alpha = -1/2$  distribution (although it was originally discovered by Holtsmark to solve a problem in astronomy). It has a density

$$f(x) = \frac{1}{\sqrt{2\pi x^3}} e^{-1/2x}.$$

The right tail of the distribution decreases like  $x^{-1/2}$  and the distribution does not have a mean. Its exponential tilts are inverse Gaussian and they have a jump density  $j(x) \propto x^{-3/2} e^{-\beta x}$ . We will meet these distributions again in Section 1.11 when we discuss NEF reciprocity.

## 1.9 Compensated Infinite Activity CPs

For  $\alpha \geq 1$  the IACP construction appears to hit a wall: the small jumps of  $j(x) = x^{-\alpha-1}$  become so frequent they have an infinite expectation. As a result,  $\int_0^\epsilon (e^{sx} - 1)j_n(x)dx \approx \int_0^\epsilon sxj_n(x)dx$  will not converge as  $n \rightarrow \infty$  and the argument we used above to construct an IACP will fail. An extra level ingenuity is required to proceed.

The idea of centering and scaling appears repeatedly in probability. For example, the central limit theorem applies to a sum of random variables after it is suitably shifted and scaled. Here, we have a sequence of smaller and smaller jumps occurring with a Poisson distribution with larger and larger expected frequency. The Poisson distribution with a large mean is approximately normal and its coefficient of variation decreases to zero. Thus the deviation from expected is very well-behaved. These properties suggest that if we subtract the expected value of small losses, the resulting compensated IACP will converge. Specifically, consider the sequence of MGFs

$$\exp\left(\int_0^1 (e^{tx} - 1 - tx)j_n(x) dx\right).$$

If  $\int_0^1 x^2j(x) dx$  exists, then the sequence of integrals will converge too, and the resulting MGF will define a variable  $X_c$ , with mean zero. We can use this new  $X_c$  to define a compensated IACP. Using compensated IACPs we can extend IACPs to  $1 < \alpha < 2$ .

To summarize: a jump density  $j$  with

1. finitely many large jumps  $\int_1^\infty j(x)dx < \infty$  and
2. finite second moment of small jumps  $\int_0^1 x^2j(x)dx < \infty$

defines a **compensated IACP**, whose MGF is the limit  $n \rightarrow \infty$  of

$$\exp\left(\int_1^\infty (e^{tx} - 1)j_n(x) dx\right) \times \exp\left(\int_0^1 (e^{tx} - 1 - tx)j_n(x) dx\right).$$

Our original condition that  $\int_0^1 xj(x)dx < \infty$  implies condition 2, since  $x^2 < x$  for  $0 < x < 1$ . The two enumerated conditions are usually combined into one, stated as

$$\int_0^\infty (x^2 \wedge 1)j(x) dx < \infty.$$

A jump distribution satisfying this condition is called a **Lévy measure**.

The behavior of the generalized CP defined by jump distribution  $J$  falls into one of three cases.

1. When  $\int_0^1 j(x) dx < \infty$  it is a finite activity CP variable, taking values  $x \geq 0$ .
2. When  $\int_0^1 j(x) dx = \infty$  but  $\int_0^1 xj(x) dx < \infty$  it is an IACP, also taking values  $x \geq 0$ .
3. When  $\int_0^1 xj(x) dx = \infty$  but  $\int_0^1 x^2j(x) dx < \infty$  it is a compensated IACP. Although this distribution only has positive jumps, the presence of the compensation terms means it can take any real value, positive or negative.

When  $J(x) = x^{-\alpha}$ , case 2 occurs for  $0 < \alpha < 1$  and case 3 occurs for  $1 < \alpha < 2$ . In case 3, the process can *creep* down and assume negative value, even though it only has positive jumps. Its left and right tail behaviors are very different. Negative outcomes are the result of an infinite sum of high-frequency Poisson distributions having below-average outcomes, and have a very thin tailed distribution, decaying like  $e^{-k(x/\alpha)^{\alpha/(\alpha-1)}} < e^{-k(x/\alpha)^2}$  for  $1 < \alpha < 2$ , see [3], part 14. Large positive outcomes, on the other hand, are the result of a power-law jump density and have a tail like  $x^{-\alpha}$ . In case 2, the mean does not exist. In case 3, the right tail is thin enough for the mean to exist, even though the variance does not. We will revisit these distribution in Part IV.

A realization of  $X_c$  is determined as follows. Imagine a small particle that jumps forward on a conveyor belt moving in the opposite direction at a constant speed equal to the expected forward movement of the jumps. Thus on average the particle stays at the origin. The variable  $X$  gives the position after one unit of time. In case 2, the conveyor belt and expected forward speed are both finite, so we can split the sum of the compensated variable into an IACP and a separate sum of offsets, and both sums converge separately. But case 3, the conveyor belt and the expected forward speed are both infinite: the sum cannot be separated. The behavior is analogous to the fact that  $1 - 1/2 + 1/3 - 1/4 + \dots$  converges (to  $\log(2)$ ) but neither  $1 + 1/3 + 1/5 + \dots$  nor  $1/2 + 1/4 + 1/6 + \dots$  converges.

Finally, what happens at  $\alpha = 2$ ? A miracle: we get the normal distribution. Refining the conveyor belt analogy, the particle moves forward *continuously* at infinite speed, with the belt moving in the opposite direction at infinite speed.  $X$  measures its position after one unit of time. This dynamic underlies continuous Brownian motion.

But  $\alpha = 2$  really is the wall: there is no way to piece together a jump density more explosive at  $x = 0$  than  $x^{-3}$ . As result, there is no Tweedie family member for  $0 < p < 1$ , as we will see in Part IV.

We brushed over one case:  $\alpha = -1$ . Here we have the worst of both worlds. The large jumps create right tail so thick the distribution does not have a mean. And the small jumps are so frequent we need to use the compensated-sum trick.  $Y$  is an extreme Cauchy distribution. The symmetric Cauchy, with jump size density proportional to  $|x|^2$  for  $x \neq 0$  is not an exponential family distribution because its density involves a term  $1 + x^2$ .

The famous Lévy-Khintchine theorem and associated Lévy-Itô decomposition say that every Lévy process is determined by a drift, a Lévy distribution  $J$  satisfying the two conditions above, and the standard deviation of a Brownian motion component. These components are essentially unique.

If all of this might seem abstract and far fetched, remember that the gamma and inverse Gaussian distributions are IACPs, and we have identified their jump densities. The lognormal, Pareto, inverse gamma, Weibull ( $0 < \alpha \leq 1$ ), and many other well-known families of random variables, are also IACPs. The extreme stable distributions with  $1 < \alpha < 2$  have been used in finance to model log stock returns. If  $X$  is such a variable, then it has a very thick right tail, with distribution like  $x^{-\alpha}$  and a very thin left tail. As a result  $Y = e^{-X}$  is a well behaved distribution with all moments. It is called a **finite moment log stable** distribution and has been used to model asset returns in markets where prices drift up but crash down, see [6].

## 1.10 Exponential Dispersion Models

An **exponential dispersion model** (EDM) is a collection of NEFs whose generator densities form a Lévy process. There are two kinds of EDM.

Let  $Z_\nu$  be a Lévy process with density  $c(y; \nu)$ . The variable  $\nu$  is called the **index parameter**.  $\nu$  is used in place of  $t$  to be agnostic: it can represent expected loss volume or time or a combination. In the finite variance case, the distribution of  $Z_\nu$  becomes less dispersed as  $\nu$  increases. Therefore it is natural to define  $1/\nu$  as a **dispersion parameter**. For Brownian motion the dispersion parameter is usually denoted  $\sigma^2$ . Throughout  $\sigma^2 = 1/\nu$  and the two notations are used inter-changeably.  $Z_\nu$  and  $\nu Z_1$  have the same mean, but the former has varying coefficient of variation  $\text{CV}(Z_1)/\sqrt{\nu}$  and the latter a constant CV.  $Z_\nu$  is more ( $\nu < 1$ ) or less ( $\nu > 1$ ) dispersed than  $Z_1$ . The difference between  $Z_\nu$  and  $\nu Z_1$  are illustrated in the lower two panels of Figure 1.

The first kind of EDM is an **additive exponential dispersion** model. It comprises the NEFs associated with the generator distribution  $Z_\nu$ . It has generator density  $c(z; \nu)$  and cumulant generator  $\kappa_\nu(\theta) = \nu \kappa(\theta)$ . Its members are all exponential tilts of  $c(z; \nu)$  as  $\nu$  varies. A representative distribution  $Z \sim \text{DM}^*(\theta, \nu)$  has density

$$c(z; \nu) e^{\theta z - \nu \kappa(\theta)}.$$

As  $\nu$  increases, the mean of the generator distribution increases and it becomes less dispersed. The same is true for  $Z$ . The cumulant generating function of  $Z$  is

$$K(t; \theta, \nu) = \nu(\kappa(t + \theta) - \kappa(\theta)).$$

If  $Z_i \sim \text{DM}^*(\theta, \nu_i)$  and  $Z = \sum Z_i$ , then the cumulant generating function for  $Z$  is  $(\sum_i \nu_i)(\kappa(t + \theta) - \kappa(\theta))$  showing  $Z$  has distribution  $\text{DM}^*(\theta, \sum_i \nu_i)$  within the same family, explaining why the family is called *additive*. Additive exponential dispersions are used to model total losses.

The second kind of EDM is a **reproductive exponential dispersion** model. It comprises the NEFs associated with the generator distribution of  $Y_\nu = Z_\nu/\nu$ .  $Y_\nu$  has a mean independent of  $\nu$ , but this independence is illusionary since the mean can be adjusted by re-scaling  $\nu$ . In the finite variance case,  $Y_\nu$  has decreasing CV as  $\nu$  increases. The densities of  $Y$  and  $Z$  are related by  $f_Y(y) = \nu f_Z(\nu y)$ . Hence a representative has density

$$f_Y(y) = \nu c(\nu y, \nu) e^{\nu(\theta y - \kappa(\theta))}.$$

The cumulant generating function of  $Y$  is

$$K(t; \theta, \nu) = \nu(\kappa(t/\nu + \theta) - \kappa(\theta)).$$

To end this section, let's work out the impact of tiling on the Lévy measure  $J$ . In the last section, we saw that the cumulant generator of  $Z$  has the form

$$\kappa(\theta) = a\theta + \int_0^1 (e^{\theta x} - 1 - \theta x) j(x) dx + \int_1^\infty (e^{\theta x} - 1) j(x) dx$$

where  $a$  is a location (drift) parameter. Rather than split the integral in two, write it as

$$\int_0^\infty (e^{\theta x} - 1 - \theta x \mathbf{1}_{(0,1)}(x)) j(x) dx.$$

where  $1_{(0,1)}$  is the indicator function on  $(0, 1)$ . The  $\theta$ -tilt of  $Z$  has cumulant generating function

$$\begin{aligned} K_\theta(s) &= \kappa(s + \theta) - \kappa(\theta) \\ &= as + \int_0^\infty (e^{(s+\theta)x} - e^{\theta x} - sx1_{(0,1)}(x)) j(x) dx \\ &= as + \int_0^\infty (e^{sx} - 1 - se^{-\theta x}x1_{(0,1)}(x)) e^{\theta x} j(x) dx. \end{aligned}$$

The factor  $e^{-\theta x}$  in the compensation term can be combined with  $1_{(0,1)}$ , resulting in an adjustment to the drift parameter. Thus the Lévy distribution of the tilted distribution is  $e^{\theta x} j(x)$ .

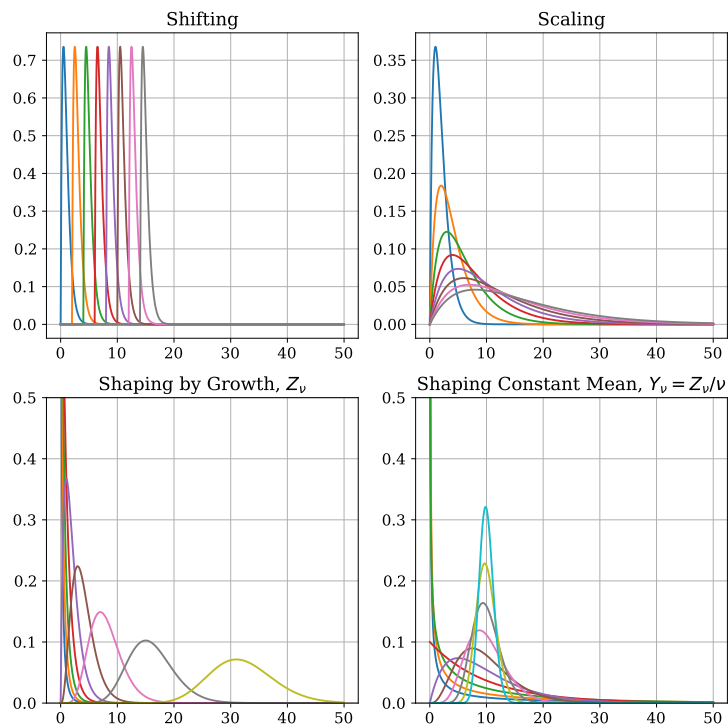


Figure 1: The impact of shifting, scaling, growth and shaping on a gamma density.

## 1.11 Building New Models From Old

Building complex models from simple ones is a time-honored method. It keeps the number of fundamentally different models to a minimum. For example, all normal distributions are shifted and scaled versions of a standard normal. Scaling and shaping are the same operation for a normal distribution, but in general they are not. Lognormal and gamma distributions have an additional shape parameter. Various programming languages exploit the shift/scale/shape paradigm to order the menagerie of probability distributions.

In this section we describe four ways to make a new NEF from an old one:

- shifting,
- scaling,
- shaping, and
- reciprocity.

The effect of the first three of these transformations is illustrated in Figure 1. In every case the underlying distribution is a gamma.

By definition, exponential tilting operates *within* a NEF, c.f., the NEF Circle. It does not create a new NEF.

A distortion operator, such as the Wang and proportional hazard transform, is also familiar to actuaries. The impact of most distortions on NEFs is unclear. There are exceptions. For example, the Wang transform acts as an exponential tilt on normal variables.

This section describes the impact of each transformation on the variance function. It gives several examples of reciprocity, which turns out to be related to surplus growth or probability of default.

Throughout this section,  $X$  and  $X_1$  have means  $m$  and  $m_1$  and variance functions  $V$  and  $V_1$ . They generate NEFs  $F$  and  $F_1$  with **mean domains**, the interior of the set of possible means,  $M$  and  $M_1$ . The NEF is uniquely specified by the distribution of  $X$  or by  $M_F$  and  $V$ .

### 1.11.1 Affinities: Shift and Scale

Transformations such as  $X_1 = aX + b$  are called **affinities**.

For a **shift**  $X_1 = X + b$  and  $m_1 = m + b$ . Therefore

$$V_1(m_1) = V(m_1 - b).$$

For a **scale**  $X_1 = aX$ ,  $m_1 = am$  and  $m = m_1/a$ . Therefore

$$V_1(m_1) = a^2V(m_1/a)$$

since  $\text{Var}(aX) = a^2\text{Var}(X)$ .

For a general affinity  $X_1 = aX + b$

$$V_1(m_1) = a^2V\left(\frac{m_1 - b}{a}\right).$$

The mapping  $X \mapsto aX + b$  gives a one-to-one mapping between  $F$  and  $F_1$ .

### 1.11.2 Shape, Power, or Division

If  $X_\nu$ ,  $\nu \geq 0$ , is a Lévy process then  $E[X_\nu] = \nu E[X_1]$  and  $\text{Var}(X_\nu) = \nu \text{Var}(X_1)$ . Therefore, if  $V_\nu$  and  $V_1$  are the respective variance functions of  $X_\nu$  and  $X_1$  then we get the **shape** transformation

$$V_\nu(\mu) = \text{Var}(X_\nu) = \nu \text{Var}(X_1) = \nu V_1\left(\frac{\mu}{\nu}\right)$$

The shape transformation can be used in another way. Rather than consider  $X_\nu$  as a family with increasing means we can consider  $X_\nu/\nu$ , which has a constant mean as  $\nu$  varies. If  $Y_\nu = X_\nu/\nu$  then

$$V_\nu(\mu) = \text{Var}(Y_\nu) = \text{Var}(X_\nu)/\nu^2 = \text{Var}(X_1)/\nu = V_1(\mu)/\mu.$$

$X_\nu$  corresponds to the additive EDM that models total losses from a growing portfolio, and  $Y_\nu$  to the reproductive EDM that models the pure premium or loss ratio.

The power or division terminology arises by writing  $X_\nu$  as the  $\nu$ -fold convolution of  $X_1$  using the  $\nu$  power of the MGF. Division arises from the ability to “take roots” of an infinitely divisible distribution. The insurance meanings are losses over multiple periods and losses over sub-periods.

### 1.11.3 Reciprocity

Reciprocity is a more profound way to relate two NEFs. We define it first and then illustrate with several examples.

Given a NEF, define  $\tilde{\Theta}$  to be those  $\theta \in \Theta$  so that  $\kappa'(\theta) > 0$ , i.e., canonical parameters corresponding to distributions with a positive mean, and define  $M_F^+$  to be the image of  $\tilde{\Theta}$  under the mean value mapping  $\kappa'$ . Thus  $\kappa'$  is a bijection between  $\tilde{\Theta}$  and  $M_F^+$ . NEFs  $F$  and  $F_1$  define a **reciprocal pair** if

1.  $\theta \mapsto -\kappa(\theta)$  maps  $\tilde{\Theta}$  to  $\tilde{\Theta}_1$ ,
2.  $\theta \mapsto -\kappa_1(\theta)$  maps  $\tilde{\Theta}_1$  to  $\tilde{\Theta}$ , and
3.  $-\kappa_1(-\kappa(\theta)) = \theta$  for all  $\theta \in \tilde{\Theta}$ .

For a reciprocal pair, the left-hand diagram commutes. The meaning, starting from  $\theta$  and  $\theta_1$ , is illustrated in the two right hand side diagrams.

$$\begin{array}{ccccc}
 M_F^+ & \xleftarrow{\kappa'} & \tilde{\Theta} & & \kappa'(\theta) & \longleftarrow & \theta & & 1/\kappa'_1(\theta_1) & \longleftarrow & -\kappa_1(\theta_1) \\
 \uparrow_{m^{-1}} & & \downarrow_{-\kappa_1} & & \downarrow & & \downarrow & & \uparrow & & \uparrow \\
 M_{F_1}^+ & \xleftarrow{\kappa'_1} & \tilde{\Theta}_1 & & 1/\kappa'(\theta) & \longleftarrow & -\kappa(\theta) & & \kappa'_1(\theta_1) & \longleftarrow & \theta_1
 \end{array}$$

Differentiating  $-\kappa_1(-\kappa(\theta)) = \theta$  shows  $\kappa'_1(-\kappa(\theta))\kappa'(\theta) = 1$ , i.e., the diagram commutes.

The variance functions of a reciprocal pair satisfy the important **reciprocity formula**

$$V(m) = m^3 V_1\left(\frac{1}{m}\right),$$

which follows by differentiating  $\kappa'_1(-\kappa(\theta))\kappa'(\theta) = 1$  again and noting  $m = \kappa'(\theta)$  and  $\kappa'_1(-\kappa(\theta)) = 1/m$  to obtain

$$\kappa''_1(-\kappa(\theta))(\kappa'(\theta))^2 = \kappa'_1(-\kappa(\theta))\kappa''(\theta) \implies V_1\left(\frac{1}{m}\right)m^2 = \frac{1}{m}V(m).$$

Unlike the formulas for affinities and shaping, the reciprocal variance transformation involves the mean outside  $V(\cdot)$ .

The variance function reciprocity formula shows that the set of NEFs with polynomial variance functions of order less than or equal to three is closed under reciprocity. This fact allowed such NEFs to be completely classified by [7]. Classifying all NEFs with a polynomial variance function is a difficult problem, though it is known that any polynomial with positive coefficients that vanishes at  $m = 0$  is the variance function of some NEF with

positive mean domain [7, Corollary 3.3]. If  $V(0) = 0$  then the NEF must be supported on a subset of the positive or negative reals because of the non-degenerate requirement.

If  $V(0) = 0$  let  $a := \lim_{m \downarrow 0} V(m)/m$  be the right derivative of  $V$  at zero. If  $a = 0$  then the corresponding NEF is continuous on the positive reals, but can have a mass at zero (Tweedie). If  $a > 0$  then the NEF is discrete and  $a$  is the smallest positive value with positive probability. For standard counting distributions  $a = 1$ . Thus polynomial variance functions of the form  $V(m) = mW(m)$ ,  $W(0) \neq 0$ , correspond to counting distributions (Poisson  $V(m) = m$ , negative binomial  $m(m-1)$ , binomial  $m(m+1)$ , etc.) and those of the form  $m^2W(m)$  to continuous distributions (gamma  $V(m) = m^2$ , inverse Gaussian  $V(m) = m^3$  and Kendall-Ressel  $m^2(1+m)$ ). The Tweedie family shows that fractional powers are also possible.

In general, reciprocity is mysterious. But in some cases it has a beautiful interpretation as the first hitting time distribution for a Lévy process, as follows. In order to use standard notation we use  $t$  in place of  $\nu$  to index processes, and we change signs of jumps so they are negative rather than positive—premium and income are positive, losses are negative.

Let  $X_t$  be a **spectrally negative** Lévy process, with no positive jumps. The Lévy measure of  $X_t$  is supported on the negative reals.  $X_t$  could be any of the following plus a positive drift: Brownian motion, an infinite activity process, or a compound Poisson process.  $X_t$  models the surplus process: cumulative premium minus cumulative losses. Losses are negative jumps and premium is a continuous positive trend.

Define the first hitting time for level  $x$  to be  $T_x = \inf\{t > 0 \mid X_t \geq x\}$ . Since  $X_t$  has no upward jumps,  $X_{T_x} = x$ : it can't jump over  $x$ , it can only jump downwards. Next,  $T_x$  is infinitely divisible. For any  $n$ ,  $T_x$  is the sum of  $n$  independent copies of  $T_{x/n}$ , because the Markov property of  $X_t$  applies at a stopping time. We can reset  $X_t$  at the random time  $T_{x/n}$ . As a result,  $T_x$ ,  $x \geq 0$ , is a strictly increasing Lévy process. Such processes are called **subordinators**.

We can identify the distribution of  $T_x$ . By assuming  $\mathbf{E}[X_1] \geq 0$  (i.e., premium exceeds expected claims, so the surplus process is increasing on average) it is guaranteed that  $X_t$  will hit any positive level  $x$  in finite time,  $\Pr(T_x < \infty) = 1$ . Since  $X_t$  is a Lévy process, its moment generating function  $\mathbf{E}[e^{\theta X_t}] = e^{t\kappa_X(\theta)}$  where  $\kappa_X$  is the cumulant generating function of  $X_1$ , independent of  $t$ .

The exponential tilt of  $X_t$  has density  $e^{\theta x - t\kappa_X(\theta)} f_t(x)$ , where  $f_t$  is the density of  $X_t$ . Since a density integrates to 1, we have

$$\mathbf{E}[e^{\theta X_t - t\kappa_X(\theta)}] = \int e^{\theta x - t\kappa_X(\theta)} f_t(x) dx = 1$$

for all  $t \geq 0$ . Therefore the process  $e^{\theta X_t - t\kappa_X(\theta)}$ ,  $t \geq 0$ , is a martingale. Martingales are a generalization of a constant process. They have no drift and a constant mean expectation. The lack of drift also holds if we stop the process at a random stopping time<sup>2</sup> Stopping

---

<sup>2</sup>The Optional Stopping Theorem says you can't beat a series of fair bets by choosing a clever stopping time. The famous doubling strategy shows there are limitations: the sequence of bets must be bounded. If  $X_{t \wedge T}$  is a bounded and  $T$  is a stopping time with finite expectation then  $\mathbf{E}[X_T] = \mathbf{E}[X_0]$ . In the present case, the process  $\exp(\beta X_t - \lambda t)$  is bounded for  $0 \leq t \leq T_x$  (because  $B_t \leq x$ ). However,  $\mathbf{E}[T]$  could be infinite. To get around that, work with  $T_n = T \wedge n$  and let  $n \rightarrow \infty$ .



$e^{\theta X_t - t\kappa_X(\theta)}$  at  $T_x$  and remembering  $X_{T_x} = x$  produces an expression for the MGF of  $T_x$

$$\mathbb{E}[e^{\theta x - T_x \kappa_X(\theta)}] = 1 \implies M_{T_x}(-\kappa_X(\theta)) := \mathbb{E}[e^{-\kappa_X(\theta) T_x}] = e^{-\theta x}.$$

Taking logs of both sides gives

$$\kappa_{T_x}(-\kappa_X(\theta)) = -\theta x$$

where  $\kappa_{T_x}$  is the cumulant generating function of  $T_x$ . Since  $T_x$  is infinitely divisible we know that  $\kappa_{T_x} = x \kappa_T$ , where  $\kappa_T$  is the cumulant generating function of  $T_1$ . As a result

$$-\kappa_T(-\kappa_X(\theta)) = \theta,$$

showing  $T_x$  and  $X_t$  define reciprocal NEFs, and explaining the name. The formula relies on the fact that  $X_t$  is a Lévy process with no positive jumps and  $\mathbb{E}[X_1] \geq 0$ .

### 1.11.3.1 Example: Normal-Inverse Gaussian.

Here is the classic example of reciprocity. Let  $X_t = ct + \sigma B_t$  be a Brownian motion with a positive trend, so  $T_x$  is guaranteed to be finite for  $x \geq 0$ .  $X_t$  is normal with mean  $ct$  and variance  $t$ , by definition of a Brownian motion. The cumulant generator  $\kappa_X(\theta) = \log \mathbb{E}[e^{\theta X_t}] = \log \mathbb{E}[e^{\theta \sigma B_t + ct\theta}] = t(c\theta + \sigma^2 \theta^2 / 2)$  by the well-known formula for the mean of a lognormal distribution.

Reciprocity shows

$$\kappa_T(-\kappa_X(\theta)) = \kappa_T\left(-c\theta - \frac{\sigma^2 \theta^2}{2}\right) = -\theta.$$

Inverting, using the quadratic formula, shows

$$\frac{\sigma^2 \theta^2}{2} + c\theta + y = 0 \implies \theta = \frac{-c + \sqrt{c^2 - 2y\sigma^2}}{\sigma^2},$$

since the argument for the Laplace transform must be positive and the other root is negative. Therefore the cumulant generating function of  $T_1$  is

$$\kappa_T(y) = \frac{c - \sqrt{c^2 - 2y\sigma^2}}{\sigma^2} = \frac{c}{\sigma^2} \left(1 - \sqrt{1 - \frac{2\sigma^2 y}{c^2}}\right) = \frac{1}{\mu\sigma^2} \left(1 - \sqrt{1 - 2\mu^2\sigma^2 y}\right)$$

where  $\mu = 1/c$ . Consulting a table of Laplace transforms, reveals this is the cumulant generating function for an inverse Gaussian distribution with mean  $\mu$  and variance  $\mu^3\sigma^2$ .  $\lambda$  is used for  $1/\sigma^2$  in [8]. This reciprocal relationship between cumulant generating functions is why Tweedie chose the named inverse Gaussian (op. cite p. 250). The mean makes sense: if we are traveling at speed  $c$  then it should take time  $x/c$  to travel distance  $x$ .

If  $c = 0$ , so there is no drift, and  $\sigma^2 = 1$ , then  $\kappa_T(y) = -\sqrt{-2y}$ , which is the cumulant generating function for a Lévy stable distribution with  $\alpha = 1/2$ . In the absence of drift, the waiting time  $T_1$  has infinite expectation: you are guaranteed to get to level 1, just not quickly!

How are these facts related to exponential family distributions? By the inverse relationship of cumulant generating functions, the normal and inverse Gaussian are reciprocal distributions. Therefore the variance functions are related by  $V_1(m) = m^3 V(m) = \sigma^2 m^3$ , as expected.

### 1.11.3.2 Example. Poisson-Exponential.

If  $X_t$  is a Poisson process with rate  $\lambda$ , meaning  $\mathbf{E}[X_t] = \lambda t$ , then the waiting time  $T_1$  is exponential, as is well known. The Poisson cumulant generating function is  $\kappa_X(\theta) = \lambda(e^\theta - 1)$  and the exponential's is  $K_T(\theta) = -\log(1 - \theta/\lambda)$ . We can check

$$-\kappa_T(-\kappa_X(\theta)) = \log\left(1 + \frac{\lambda(e^\theta - 1)}{\lambda}\right) = \theta.$$

Variance reciprocity says  $V_T(m) = m^3 V_X(1/m) = m^2$ .

It is also possible to give an explanation of reciprocity as a first hitting time for a random walk, indexed by  $t = 0, 1, 2, \dots$ , rather than by continuous  $t \geq 0$ . For  $X_i$  iid discrete with  $0 < \mathbf{E}[X_i] \leq 1$  and  $\Pr(X_i = 0) > 0$  define  $S_n = X_1 + \dots + X_n$  and  $Y_n = n - S_n$ . Then  $Y_n$  is a discrete version of a spectrally negative Lévy process, it models the cumulative surplus process of a single policy with annual premium 1 and integer valued losses given by  $X$ . We can consider the distribution of  $T_1$  the first hitting time for level 1. It is possible to show that  $T_1$  is infinitely divisible and to write down its distribution, see [9] and [7]. It is more natural to consider  $G := T_1 - 1$ , which is supported on  $0, 1, 2, \dots$ . The discrete reciprocity formula is

$$V_G(m) = (m + 1)^3 V_X\left(\frac{m}{m + 1}\right).$$

$G$  is infinitely divisible since  $T_1$  is, and so using the shape transformation there is a  $G_t$  with variance function

$$V_{G_t}(m) = t V_G\left(\frac{m}{t}\right) = t \left(\frac{m}{t} + 1\right)^3 V_X\left(\frac{m/t}{m/t + 1}\right) = \frac{(m + t)^3}{t^2} V_X\left(\frac{m}{m + t}\right).$$

This approach yields distributions including the generalized Poisson (Abel) and generalized negative binomial distributions.

### 1.11.3.3 Example. Poisson-Generalized Poisson Discrete Model

Here is how random walk reciprocity works when  $X$  is Poisson.  $X$  models an integer valued claim *amount* process. Assume  $\lambda = \mathbf{E}[X] < 1$  and the annual premium is 1. At each  $n = 1, 2, 3, \dots$  premium is collected and  $X$  is paid, resulting in a net surplus change of  $1 - X$ . Initial surplus is zero.  $T_1$  is the waiting time until accumulated surplus first equals 1 and we want to determine the distribution of the shifted distribution  $T_1 - 1$ , with support  $0, 1, 2, \dots$ . We drop subscripts when equal to 1. This problem is opposite to the usual problem of time until default. By definition, the cumulant generating function for the surplus  $Y$  is

$$\kappa_Y(\theta) = \log \mathbf{E}[e^{\theta t - \theta S_t}] = t(\theta + \lambda(e^{-\theta} - \lambda)).$$

A discrete version of the first hitting time argument says

$$-\kappa_T(-\kappa_Y(\theta)) = -\kappa_T(-\theta - \lambda(e^{-\theta} - 1)) = \theta.$$

Therefore  $-\kappa_T(y)$  is given by the solution to  $-\theta - \lambda(e^{-\theta} - 1) = y$  in  $\theta$ . Set

$$w = e^{-\theta}, \quad z = e^y, \quad \text{and} \quad g(w) = e^{\lambda(w-1)}.$$

Then

$$w e^{-\lambda(w-1)} = e^y \iff w = z g(w),$$

which can be solved using the Lagrange inversion formula, [10],

$$w(z) = \sum_{n \geq 1} \frac{z^n}{n!} \left\{ \frac{d^{n-1}}{dw^{n-1}} g(w)^n \right\}_{w=0} = \sum_{n \geq 1} \frac{z^n}{n!} (\lambda n)^{n-1} e^{-n\lambda}.$$

Thus  $\kappa_T(y) = -\theta = \log(w(y))$ , implying  $M_T(y) = w(e^y)$ . As a result,  $w(z)$  is the probability generating function<sup>3</sup> of  $T$  and  $\Pr(T = n)$  is the coefficient of  $z^n$  in  $w(z)$

$$\Pr(T = n) = \frac{e^{-\lambda n}}{n!} (\lambda n)^{n-1}.$$

$T$  is supported on  $1, 2, \dots$ . For the shifted distribution  $T_0 = T - 1$

$$\Pr(T_0 = n) = \frac{e^{-\lambda(n+1)}}{(n+1)!} (\lambda(n+1))^n = e^{-\lambda} \frac{(\lambda e^{-\lambda})^{n+1}}{n!} \lambda^n (n+1)^{n-1}.$$

More generally, let the premium rate equal  $r$  and suppose  $\mathbb{E}[X] = \lambda < r$ . By considering time steps of size  $1/r$  we can reduce to the case where the premium rate is 1 and the frequency is  $\lambda/r < 1$ , showing there is no loss of generality assuming  $r = 1$ . Let  $\nu > 0$  be an arbitrary surplus level and  $T_\nu$  be its first hitting time. The standard cumulant generating function argument is the same, showing

$$\kappa_{T_\nu}(-\theta - \lambda(e^{-\theta} - 1)) = -\nu\theta.$$

To evaluate  $\kappa_{T_\nu}(y)$  solve  $-\theta - \lambda(e^{-\theta} - 1) = y$  for  $\theta$ . Exponentiate to convert the original equation into  $w(z) = zg(w(z))$ . Once we have a solution, it follows that

$$\kappa_{T_\nu}(y) = -\nu\theta = \nu \log(w(e^y)).$$

Hence the probability generating function of the  $T_\nu$  is

$$P_\nu(z) = e^{\kappa_{T_\nu}(\log(z))} = e^{\nu \log(w(z))} = w(z)^\nu.$$

Finally, since we are interested in the left-shift  $T_0 = T - \nu$ , apply the advanced form of the Lagrange Inversion Formula<sup>4</sup> with  $F(z) = g(z)^\nu$ . Observe that

$$g^n \frac{\partial F}{\partial w} = g^n \frac{\partial g^\nu}{\partial w} = \nu g^{n+\nu-1} g' = \frac{\nu}{n+\nu} \frac{\partial g^{n+\nu}}{\partial w}.$$

Applying the advanced Lagrange inversion formula, these ingredients serve up

$$\begin{aligned} P(z) &= g(0)^\nu + \sum_{n \geq 1} \frac{z^n}{n!} \frac{\nu}{n+\nu} \left\{ \frac{\partial^n}{\partial w^n} g(w)^{n+\nu} \right\}_{w=0} \\ &= e^{-\lambda\nu} + \sum_{n \geq 1} \frac{z^n}{n!} \frac{\nu}{n+\nu} (\lambda(n+\nu))^n e^{-\lambda(n+\nu)} \\ &= e^{-\lambda\nu} \sum_{n \geq 0} z^n \frac{\nu(n+\nu)^{n-1}}{n!} (\lambda e^{-\lambda})^n \\ &= e^{-\theta} \sum_{n \geq 0} z^n \frac{\theta(n\lambda + \theta)^{n-1}}{n!} e^{-\lambda n}. \end{aligned}$$

<sup>3</sup>The probability generating function of a discrete distribution with  $\Pr(N = n) = p_n$  is  $P(z) = \sum_{n \geq 0} p_n z^n$ .

<sup>4</sup> $g(z)$  achieves the left-shift and  $\nu$  power is convolution, giving the hitting probability to the level  $\nu$  rather than 1. Since  $T_\nu \geq \nu$  the net effect of this transformation is to shift left by  $\nu$ , resulting in a distribution with support on  $0, 1, 2, \dots$ .

The  $e^{-\lambda\nu}$  term gives the probability of moving directly from 0 to  $\nu$  with no claims at all, and  $\lambda e^{-\lambda}$  is the probability of one claim. In the last line  $\theta = \nu\lambda$ . This distribution is called Consul's generalized Poisson, a Lagrangian Poisson, or a shifted Tanner-Borel distribution. It is a two parameter distribution, depending on  $\nu > 0$  and  $0 \geq \lambda \leq 1$ . The mean and variance are

$$\mathbb{E}[T_\nu] = \frac{\nu\lambda}{1-\lambda} \quad \text{and} \quad \text{Var}[T_\nu] = \frac{\nu\lambda}{(1-\lambda)^3}.$$

As  $\lambda \uparrow 1$  it becomes very thick tailed: the loss ratio is close to 100% and surplus accretes very slowly.

When  $\nu = 1$  we can determine the variance function following the path of the derivation from Poisson  $X$  with  $V_X(m) = m$ , to change in surplus  $1 - X$  with  $V(m) = 1 - m$ , to reciprocal  $G_1$  with  $V_1(m) = m^3 V(1/m)$ , and finally shifted  $G = G_1 - 1$  with

$$V_G(m) = V_1(m+1) = (m+1)^3 \left(1 - \frac{1}{m+1}\right) = m(m+1)^2.$$

For  $\nu > 1$  use the fact that  $T_\nu$  is infinitely divisible to apply the shape transformation to  $G$  to obtain

$$V_{G_\nu}(m) = \nu V_G\left(\frac{m}{\nu}\right) = m \left(1 + \frac{m}{\nu}\right)^2,$$

consistent with the stated mean and variance. Notice that for very small  $m$ ,  $V_{G_\nu}$  behaves like a Poisson, as does the underlying distribution.

If  $\lambda \downarrow 0$  and  $\nu \uparrow \infty$  so that  $\theta = \lambda\nu$  is constant then density shows the generalized Poisson reduces to a Poisson( $\theta$ ). For example, if  $\lambda = 1/m$  and  $\nu = \theta m$  then although there will be hardly any periods with claims, there are a very large number of periods and on average  $\theta$  claims. The waiting time is a sum of  $\nu$  Poisson  $\lambda$  variables, a Poisson  $\theta$ .

There is a recursive formula for the event probabilities.

The generalized Poisson is related to queuing theory. A single-server queue has  $\nu$  customers. New customers arrive randomly at rate  $\lambda$  per unit time and are served in a constant unit time. If  $\lambda < 1$  the queue shrinks on average.  $T_\nu$  is the distribution of the number of customers who are served before the queue vanishes or, equivalently, how long until the queue vanishes.

In the derivation  $\nu$  is an integer, but we can re-interpret the problem in continuous time to remove this constraint. Simply replace  $Y_n = n - S_n$  with  $Y_t = t - N_t$  where  $N_t$  is a Poisson process with intensity  $\lambda$ . The same derivation applies, allowing  $\nu > 0$ . In insurance terms, the discrete model applies when solvency regulation and accounting is performed at the end of each period and the continuous model applies when there is continuous solvency monitoring. The UK regulatory regime requires management monitor solvency continuously.

#### 1.11.3.4 Example. Kendall-Ressel Family

The Kendall-Ressel family has variance function

$$V(m) = \frac{m^2}{\lambda} \left(1 + \frac{m}{\lambda}\right).$$

It is the reciprocal of  $X_t = t - G_t$  where  $G_t$  is a gamma process with density  $x^{t-1}e^{-x}/\Gamma(t)$ . Insurance interpretation:  $X_t$  is accumulated profit from an inflow of premium 1 per unit time and a cumulative claims process  $G_t$ . Let  $T_x$  be the waiting time until accumulated surplus equals  $x > 0$ . As usual, reduce to premium of 1 per unit time by adjusting the time period.  $\mathbb{E}[G_1] = 1$  is allowable (oscillating) but tilts with mean  $< 1$  are better. By a shaping argument, assume  $x = 1$ . The variance function is derived

$$\begin{aligned} V_G(m) &= m^2 \mapsto V_X(m) = (1 - m)^2 \\ \mapsto V_T(m) &= m^3 V_X\left(\frac{1}{m}\right) = m(m - 1)^2 \\ \mapsto V_{T-1}(m) &= m^2(1 + m) \end{aligned}$$

with the  $X$  to  $T$  step coming from reciprocity. The density can be derived using a clever formula due to Borovkov and Zolotarev, see [11]. If  $X_t$  is a spectrally negative Lévy process and  $T_x$  is the first hitting time of level  $x > 0$ , and if both  $X_t$  and  $T_x$  have densities  $f_{X_t}, f_{T_x}$ , then

$$f_{T_x}(t) = \frac{x}{t} f_{X_t}(x).$$

There is also a discrete version of this formula, where  $f$  is interpreted as the probability mass function, [10].

**Example.** For  $X_t = ct + \sigma B_t$ , a Brownian motion with drift, the Borovkov-Zolotarev formula gives an expression for the density of the inverse Gaussian waiting time:

$$f_{T_x}(t) = \frac{x}{t} f_{X_t}(x) = \frac{x}{\sqrt{2\pi t^3} \sigma} \exp\left(-\frac{(x - ct)^2}{2\sigma^2 t}\right).$$

Returning to  $X_t = t - G_t$  we have

$$f_{X_t}(x) = f_{G_t}(t - x) = \frac{(t - x)^{t-1}}{\Gamma(t)} e^{-(t-x)}.$$

The process  $T_x \geq x$  ( $T_x = x$  occurs when there are no claims) and so it is more natural to consider  $R_x := T_x - x$  which defines the **Kendall-Ressel** distribution, [7], [12]. Applying Borovkov-Zolotarev gives its density

$$f_{R_x}(t) = f_{T_x}(x + t) = \frac{x}{t} f_{X_t}(x) = \frac{x}{x + t} \frac{t^{x+t-1}}{\Gamma(x + t)} e^{-t} = \frac{xt^{x+t-1}}{\Gamma(x + t + 1)} e^{-t}.$$

Remember  $(x + t)\Gamma(x + t) = \Gamma(x + t + 1)$ .

The Kendall-Ressel distribution is infinitely divisible. Opposite to the stable and Tweedie families, that have a simple cumulant generating function but no elementary expression for their densities, the Kendall-Ressel has a simple density but no simple expression for its cumulant generating function and therefore no simple expression for the density in the NEF it generates. Solving  $\kappa'(\theta) = \mu$  necessitates solving the equation  $\theta - \log(1 + \theta) = y$  which has no closed form. Thus, there is no simple form for the NEF it generates in the canonical parameterization. However, [12], shows there is a simple expression in terms of the mean value parameterization.

Table 1 tabulates various common variance functions and how they are related through reciprocity.

Table 1: Relationship of common NEFs in terms of their variance functions. FHT = first hitting time. LM refers to the numbering scheme in [7].

LM#	Base Distribution, $X$	$V_X(m)$	$\lambda V_X(m/\lambda)$	$m^3 V_{X_\lambda}(1/m)$	$V_{R_1}(m+1)$	$\nu V_{G_1}(m/\nu)$	Comments
	<i>Variable</i>	$X = X_1$	$X_\lambda$	$R_\lambda$	$G_1 = R_1 - 1$	$G_\nu$	
	<i>Transformation</i>		shape	reciprocal	shift	shape	
1	Gaussian	1	$\lambda$	$\lambda m^3$			$\lambda = \sigma^2$ , inverse Gaussian FHT
2	Poisson	$m$	$m$	$m^2$			Generated by PGF $e^t$ . Gamma FHT; no shape, ODP is <i>scale</i>
3	Binomial	$m(1-m)$	$m(1-\frac{m}{\lambda})$				$N$ must be integer, $0 \leq m \leq N$ , not ID
4	Negative binomial	$m(1+m)$	$m(1+\frac{m}{\lambda})$				Generated by PGF $(1-z)^{-a}$ . Gamma FHT; no shape, ODP is <i>scale</i>
5	Gamma	$m^2$	$m^2/\lambda$				FHT for Poisson process
6	GHS	$1+m^2$	$\lambda(1+m^2/\lambda^2)$				
n/a	Tweedie $p$	$m^p$	$\lambda^{1-p} m^p$				$p = 3/2$ is self-reciprocal
7	1-Poisson	$1-m$		$m^2(m-1)$	$m(m+1)^2$	$m(1+m/\lambda)^2$	Abel, generalized Poisson, Borel-Tanner
8	1-Negative binomial	$(1-m)(1+\frac{1-m}{a})$		$m(m-1)(m\frac{a+1}{a}-\frac{1}{b})$	$m(1+m)(1+\frac{a+1}{a}m)$		Takács
9	Strict arcsine	$m(1+\frac{m^2}{a^2})$					Generated by PGF $\exp(a \arcsin z)$
10	1-Strict arcsin	$(1-m)(1+\frac{(1-m)^2}{a^2})$		$(m-1)(\frac{a^2+1}{a^2}-\frac{2m}{a^2}+\frac{m^2}{a^2})$	$m(1+2m+\frac{a^2+1}{a^2}m^2)$		Large arcsine
11	(1-Gamma)/ $\lambda$	$(1-\lambda m)^2/\lambda^2$		$m(\frac{m}{\lambda}-1)^2$	$\frac{m^2}{\lambda}(1+\frac{m}{\lambda})$		Kendall Ressel, first step scales by $\lambda$ , last step translates
12	Inverse Gaussian	$m^3$	$m^3/\lambda^2$	$1/\lambda^2$			Normal

## 1.12 Variance Functions for Actuaries

GLMs allow the modeler to select a family of distributions for modeling losses based a variance function. The variance function and unit mean then uniquely specify the loss distribution, within exponential families. Moreover, the exponential family distribution has minimal information of any distribution with the posited mean-variance relationship, [13]. Thus selecting a variance function is a relatively safe way to incorporate known properties into the modeling process. This section summarizes facts about variance functions that are germane to their selection for actuarial modeling.

A model of a positive (or negative) quantity will have a variance function satisfying  $V(0) = 0$ . The value of  $\theta$  corresponding to zero mean occurs at an endpoint of  $\Theta$ .

Any polynomial with positive coefficients satisfying  $V(0) = 0$  is the variance function of a NEF. In general, if  $V$  is a non-zero entire series (complex differentiable, infinite degree polynomial) with nonnegative coefficients,  $V(0) = 0$ , and positive radius of convergence, then it is the variance function of a NEF, [7].

Jorgensen, Martinez, and Tsao, [14], show that for a NEF with variance function  $V$  and support  $S$  satisfying  $\inf S = 0$  then the following hold:

1.  $\inf \Omega = 0$ ,
2.  $\lim_{\mu \rightarrow 0} V(\mu) = 0$ ,
3.  $\lim_{\mu \rightarrow 0} V(\mu)/\mu = \delta := \inf \{S \setminus \{0\}\}$ ,
4. if  $\Pr(0) > 0$  then  $\lim_{\mu \rightarrow 0} V'(\mu) = \delta$ , and
5. if  $\Pr(0) > 0$  then  $\lim_{\mu \rightarrow 0} V(\mu)/\mu^2 = \infty$ .

They comment that l'Hospital's rule implies  $\lim_{\mu \rightarrow 0} V'(\mu) = \delta$  when  $\nu\{0\} = 0$  ( $\nu$  is the carrier measure) provided the limit of  $V'$  exists. However, they are unable to prove the limit exists in general in spite of the fact that it exists in all examples. Similarly, they

expect the limit in (5) to be zero or finite when  $\nu\{0\} = 0$ . Applying l'Hospital shows  $2 \lim_{\mu \rightarrow 0} V(\mu)/\mu^2 = \lim_{\mu \rightarrow 0} V''(\mu)$ .

If  $V(m) \sim c_0 m^p$  for  $m \rightarrow 0$  or  $m \rightarrow \infty$  then  $V$  is called **regular of order  $p$**  at 0 or  $\infty$ . The previous result shows that discrete distributions supported on  $x \geq 0$  are regular of order 1. A Tweedie is regular of order  $1 + \epsilon$  for  $0 < \epsilon < 1$  and is mixed. By the above remarks, we expect the distribution to be continuous it is regular of order  $\geq 2$  at zero.

We can tell the asymptotic small and large mean behavior from  $V$ .  $ED(\mu, \sigma^2)$  (resp.  $Tw_p(\mu, \sigma^2)$ ) refers to a reproductive exponential dispersion model with mean  $\mu$  and variance  $\sigma^2 V(\mu)$  (resp. Tweedie distribution with variance  $\sigma^2 \mu^p$ ). If  $V$  is regular of order  $p$  then  $p \notin (0, 1)$ , and for any  $\mu, \sigma$

$$c^{-1} ED(c\mu, \sigma^2 c^{2-p}) \rightarrow_d Tw_p(\mu, c_0 \sigma^2)$$

as either  $\mu \rightarrow 0$  or  $\mu \rightarrow \infty$ , where convergence occurs through possible values of the parameters. If  $c^{p-2} \rightarrow \infty$  the model must be ID, [14]. This result says that a suitably scaled version of a reproductive EDM behaves like a Tweedie PVF distribution for small or large means. For example, a negative binomial has  $V(m) = m(1 + m)$ . For small means it behaves like a Poisson  $V(m) = m$  and for large means like a gamma  $V(m) = m^2$ . This is consistent with its description as a gamma-mixed Poisson distribution. For small means the mixing is irrelevant, the Poisson process risk dominates. For very large means, the uncertainty across the population dominates and the process risk diversifies away, see [2], fig. 4 and 5.

Table 1 and #tbl-esoteric display details of some common and less-common variance functions. The references refer to [14].

Table 2: Assorted variance functions, references to [14].

Variance function $V(\mu)$	Distribution Name	Support	Reference
$1 + \mu^2 + \sqrt{1 + \mu^2}$	Laplace exponential		p. 68
$(1 + \mu^2)^{3/2}$	Hyperbolic distribution	$(-\infty, \infty)$	p. 121
$\frac{1}{4}(1 + 8\mu - \sqrt{1 + 8\mu})$	Hermite	discrete	p. 123
$1 + \mu^2 - \sqrt{1 + \mu^2}$	$\nu(dy) = (e^{2y} - 1)dy$	$(0, \infty)$	p. 156
$\mu\sqrt{\mu^2 + 4\mu}$	$\nu(dy) = \delta_0(dy) + 1_{(0, \infty)} dy$	$[0, \infty)$	p. 157
$\frac{\mu\sqrt{1 + 2\mu}}{\mu^3 + 2\mu^2 + \mu^2\sqrt{\mu^2 + 2\mu}}$	<i>nu</i> known	discrete	p. 157
	Lévy Stable times	$(0, \infty)$	p. 158
	exponential		
$\frac{\mu + \mu^3/2 + (\mu^2/2)\sqrt{2 + \mu^2}}{\mu + \mu\sqrt{\mu + 2 - 2\sqrt{1 + \mu}}}$	Poisson-inverse Gaussian	discrete	p. 169
	Poisson-Poisson mixture	discrete	p. 170

## References

1. Avanzi, B., Taylor, G., Wong, B.: Correlations between insurance lines of business: An illusion or a real phenomenon? some methodological considerations. *ASTIN Bulletin*. 46, 225–263 (2016). <https://doi.org/10.1017/asb.2015.31>
2. Mildenhall, S.J.: Actuarial Geometry. *Risks*. 5, (2017). <https://doi.org/10.3390/risks5020031>
3. Sato, K.-I.: *Levy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge (1999)

4. Patrik, G., Bernegger, S., Rüegg, M.: [The use of risk adjusted capital to support business decision-making](#). Casualty Actuarial Society Forum. Spring, (1999)
5. Mitchell-Wallace, K., Jones, M., Hillier, J., Foote, M.: Natural Catastrophe Risk Management and Modeling - A Practitioner's Guide. Wiley-Blackwell (2017)
6. Carr, P., Wu, L.: The Finite Moment Log Stable Process and Option Pricing. *Journal of Finance*. 58, 753–778 (2003). <https://doi.org/10.1111/1540-6261.00544>
7. Letac, G., Mora, M.: Natural Real Exponential Families with Cubic Variance Functions. *The Annals of Statistics*. 18, 1–37 (1990). <https://doi.org/10.1214/aos/1176347491>
8. Johnson, N.L., Kotz, S., Balakrishnan, N.: *Continuous Univariate Distributions*. John Wiley; Sons (1994)
9. Letac, G.: [The Rovinj summer school on natural exponential families](#). 1–15 (2004)
10. Wendel, J.G.: Left-Continuous Random Walk and the Lagrange Expansion. *The American Mathematical Monthly*. 82, 494 (1975). <https://doi.org/10.2307/2319745>
11. Pakes, A.G.: A hitting time for Lévy processes, with application to dams and branching processes. *Annales de la faculté des sciences de Toulouse Mathématiques*. 5, 521–544 (1996). <https://doi.org/10.5802/afst.838>
12. Bar-Lev, S.K., Boukai, B., Landsman, Z.: The Kendal-Ressel Exponential Dispersion Model: Some Statistical Aspects and Estimation. *International Journal of Statistics and Probability*. 5, 32 (2016). <https://doi.org/10.5539/ijsp.v5n3p32>
13. Wedderburn, R.W.M.: Quasi likelihood functions, generalized linear models, and the Gauss Newton method. *Biometrika*. 61, 439–447 (1974). <https://doi.org/10.2307/2334725>
14. Jørgensen, B.: *The theory of dispersion models*. CRC Press (1997)