Conditional Probability and the Collective Risk Model

Leigh J. Halliwell, FCAS, MAAA

Abstract. One of the most powerful and profound tools of casualty actuarial science is the collective risk model $S = X_1 + \ldots + X_N$. It is widely used by casualty actuaries, especially by those in the field of reinsurance. Nearly one hundred pages of one standard textbook (Klugman, [1998], Chapter 4) hardly suffice to survey the ingenuity with which actuaries and scholars have analyzed it. Much of their analysis proceeds from the application of conditional probability to the so-called individual risk model $S = X_1 + \ldots + X_n$. This paper penetrates deeper into both conditional probability and the collective risk model, deriving new insights into higher moments and their generating functions. Particular attention is devoted to the fourth moment of the collective risk model, for which no formula seems previously to have been published. An appendix extends conditional probability to a novel technique of loss development.

Keywords: conditional probability, moments, cumulants, collective risk model.

1. INTRODUCTION

This paper applies conditional probability to the moments of the collective risk model. In the next section we will set forth definitions of conditional moments and co-moments, and in third section will derive formulas in which unconditional moments are expressed in terms of conditional ones. Next, in the fourth section, after explaining why moments higher than the third are not additive, we will introduce an additivity-restoring adjustment known as a cumulant. In the fifth section we will apply conditional cumulant formulas to the collective risk model to seize the prize of a manageable formula for its fourth cumulant. Finally, in the sixth section we will explain the cumulant generating function, and show its usefulness in relating cumulants to moments and in deriving cumulants of the collective risk model.

2. DEFINITIONS

Let X be a random variable with finite mean $E[X] = \mu$. For positive integer *n*, define the n^{th} moment of X as $M_n[X] = E[(X-\mu)^n]$.[1] Here we have defined what probability theory calls the *central* moments of X. Of course, the first central moment is zero. The second moment is the variance, and the third is the skewness. We shall call the fourth moment the kurtosis.[2] Now let Θ denote an event which conditions the probability distribution of X. We can then speak of the conditional n^{th} moment $M_n[X|\Theta] = E[(X-\mu_\Theta)^n|\Theta]$, where $\mu_\Theta = E[X|\Theta]$.[3]

Furthermore, as a multivariate extension, we can define the n^{th} co-moment as $CM[X_1, \ldots, X_n] = E\left[\prod_{k=1}^{n}(X_k - \mu_k)\right]$. Since the order of the co-moment is the number of its arguments, it is superfluous to subscript the definition as CM_n.[4] The first co-moment is zero; the second is the covariance. We shall call the third and the fourth co-moments the co-skewness and the co-kurtosis. A co-moment of random variables is zero, if any of them is constant, since in that case one of the factors in the Π operator will always be zero. The same random variable may appear in the argument list more than once; as a special

---

[1] The reader should not confuse $M_n[X]$ for the n^{th} moment with $M_X(t)$, the moment generating function of X.

[2] Some define kurtosis as the fourth cumulant, $E[(X-\mu)^4] - 3E[(X-\mu)^2]^2$, also known as excess kurtosis because the kurtosis of the normal distribution is three times the square of its variance. Sometimes (e.g., Daykin [1994], 24) skewness and kurtosis are defined as what we would call *coefficients* of skewness and kurtosis, i.e., the moments or cumulants stripped of dimension by dividing them by the third and fourth powers of the standard deviation.

[3] In general, $E[(X - E[X])^n|\Theta] \neq E[(X - E[X|\Theta])^n|\Theta]$. Conditioning at one level of expectation should by default cascade into the next or nested level, and so on. The tendency to disregard this inequality may indicate a defect in the accepted notation. It helps (at least it helps this author) to regard unconditional expectation as conditional upon the universal event V: $E[X] = E[X|V]$.

[4] One must be wary of such mistakes as equating $CM[X, X, Y]$ and $CM[X^2, Y]$, which confuses a third co-moment with a second.

case, $CM\left[\overbrace{X,\ldots,X}^{n\text{ times}}\right]=E\left[\prod_{k=1}^{n}(X-\mu)\right]=E\left[(X-\mu)^n\right]=M_n[X]$. The conditional co-moment is $CM\left[X_1,\ldots,X_n\middle|\Theta\right]=E\left[\prod_{k=1}^{n}\left(X_k-E\left[X_k\middle|\Theta\right]\right)\middle|\Theta\right]$.

3. UNCONDITIONAL MOMENTS IN TERMS OF CONDITIONAL

Our purpose here is to derive formulas that express unconditional moments in terms of moments and co-moments conditional upon Θ. This begins with the key sequence:

$$M_n[X]=E\left[(X-\mu)^n\right]=\underset{\Theta}{E}\left[E\left[(X-\mu)^n\middle|\Theta\right]\right]=\underset{\Theta}{E}\left[E\left[\{(X-\mu_\Theta)+(\mu_\Theta-\mu)\}^n\middle|\Theta\right]\right]$$

Next we expand this according to the binomial theorem:

$$\begin{aligned}
M_n[X]&=\underset{\Theta}{E}\left[E\left[\{(X-\mu_\Theta)+(\mu_\Theta-\mu)\}^n\middle|\Theta\right]\right]\\
&=\underset{\Theta}{E}\left[E\left[\sum_{k=0}^{n}\binom{n}{k}(X-\mu_\Theta)^{n-k}(\mu_\Theta-\mu)^k\middle|\Theta\right]\right]\\
&=\sum_{k=0}^{n}\binom{n}{k}\underset{\Theta}{E}\left[E\left[(X-\mu_\Theta)^{n-k}(\mu_\Theta-\mu)^k\middle|\Theta\right]\right]\\
&=\sum_{k=0}^{n}\binom{n}{k}\underset{\Theta}{E}\left[E\left[(X-\mu_\Theta)^{n-k}\middle|\Theta\right](\mu_\Theta-\mu)^k\right]\\
&=\sum_{k=0}^{n}\binom{n}{k}\underset{\Theta}{E}\left[M_{n-k}\left[X\middle|\Theta\right](\mu_\Theta-\mu)^k\right]
\end{aligned}$$

The fourth line follows from the third because $(\mu_\Theta-\mu)^k$ behaves as a constant within the nested expectation, and so can be taken outside it. Of these $n+1$ terms, the $(n\text{–}1)^{th}$ is zero, since $M_{n-(n-1)}\left[X\middle|\Theta\right]=M_1\left[X\middle|\Theta\right]=0$. Hence, in this binomial form, the n^{th} moment has n non-vanishing terms, namely:

$$M_n[X] = \sum_{k=0}^{n} \binom{n}{k} \underset{\Theta}{E}\left[M_{n-k}[X|\Theta](\mu_\Theta - \mu)^k\right]$$

$$= \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{E}\left[M_{n-k}[X|\Theta](\mu_\Theta - \mu)^k\right] + \underset{\Theta}{E}\left[(\mu_\Theta - \mu)^n\right]$$

However, at this point we have not expressed the unconditional moment in terms of conditional moments and co-moments; we must express the expectation within the Σ operator as a co-moment. Letting $\zeta_{n-k} = \underset{\Theta}{E}\left[M_{n-k}[X|\Theta]\right]$, and remembering that first central moments are zero, we derive:

$$M_n[X] = \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{E}\left[M_{n-k}[X|\Theta](\mu_\Theta - \mu)^k\right] + \underset{\Theta}{E}\left[(\mu_\Theta - \mu)^n\right]$$

$$= \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{E}\left[\left\{M_{n-k}[X|\Theta] - \zeta_{n-k} + \zeta_{n-k}\right\}(\mu_\Theta - \mu)^k\right] + \underset{\Theta}{E}\left[(\mu_\Theta - \mu)^n\right]$$

$$= \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{E}\left[\left\{M_{n-k}[X|\Theta] - \zeta_{n-k}\right\}(\mu_\Theta - \mu)^k\right]$$

$$+ \sum_{k=1}^{n-2} \binom{n}{k} \zeta_{n-k} \underset{\Theta}{E}\left[(\mu_\Theta - \mu)^k\right] + \underset{\Theta}{E}\left[(\mu_\Theta - \mu)^n\right]$$

$$= \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{CM}\left[M_{n-k}[X|\Theta], \overbrace{\mu_\Theta, \ldots, \mu_\Theta}^{k \text{ times}}\right]$$

$$+ \sum_{k=1}^{n-2} \binom{n}{k} \zeta_{n-k} \underset{\Theta}{M_k}[\mu_\Theta] + \underset{\Theta}{M_n}[\mu_\Theta]$$

$$= \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{CM}\left[M_{n-k}[X|\Theta], \overbrace{E[X|\Theta], \ldots, E[X|\Theta]}^{k \text{ times}}\right]$$

$$+ \sum_{k=2}^{n-2} \binom{n}{k} \underset{\Theta}{E}\left[M_{n-k}[X|\Theta]\right] \underset{\Theta}{M_k}\left[E[X|\Theta]\right] + \underset{\Theta}{M_n}\left[E[X|\Theta]\right].$$

Hence, for $n \geq 2$, to express the n^{th} unconditional moment in the desired conditional form requires $n + \max(0, n-3)$ non-vanishing terms. The first Σ operator does not come into play until $n \geq 3$, and the second until $n \geq 4$. At least the complexity does not increase after the fourth moment.

The second, third, and fourth moments follow readily from the general formula:

$$\begin{aligned}
Var[X] &= M_2[X] \\
&= \underset{\Theta}{E}\left[M_2[X|\Theta]\right] + \underset{\Theta}{M_2}\left[E[X|\Theta]\right] \\
&= \underset{\Theta}{E}\left[Var[X|\Theta]\right] + \underset{\Theta}{Var}\left[E[X|\Theta]\right] \qquad \text{cf. Klugman [1998], 393}
\end{aligned}$$

$$\begin{aligned}
Skew[X] &= M_3[X] \\
&= \underset{\Theta}{E}\left[M_3[X|\Theta]\right] + 3\underset{\Theta}{CM}\left[M_2[X|\Theta], E[X|\Theta]\right] + \underset{\Theta}{M_3}\left[E[X|\Theta]\right] \\
&= \underset{\Theta}{E}\left[Skew[X|\Theta]\right] + 3\underset{\Theta}{Cov}\left[Var[X|\Theta], E[X|\Theta]\right] + \underset{\Theta}{Skew}\left[E[X|\Theta]\right]
\end{aligned}$$

$$\begin{aligned}
Kurt[X] &= M_4[X] \\
&= \underset{\Theta}{E}\left[M_4[X|\Theta]\right] + 4\underset{\Theta}{CM}\left[M_3[X|\Theta], E[X|\Theta]\right] + 6\underset{\Theta}{CM}\left[M_2[X|\Theta], E[X|\Theta], E[X|\Theta]\right] \\
&\quad + 6\underset{\Theta}{E}\left[M_2[X|\Theta]\right]\underset{\Theta}{M_2}\left[E[X|\Theta]\right] + \underset{\Theta}{M_4}\left[E[X|\Theta]\right] \\
&= \underset{\Theta}{E}\left[Kurt[X|\Theta]\right] + 4\underset{\Theta}{Cov}\left[Skew[X|\Theta], E[X|\Theta]\right] + 6\underset{\Theta}{Coskew}\left[Var[X|\Theta], E[X|\Theta], E[X|\Theta]\right] \\
&\quad + 6\underset{\Theta}{E}\left[Var[X|\Theta]\right]\underset{\Theta}{Var}\left[E[X|\Theta]\right] + \underset{\Theta}{Kurt}\left[E[X|\Theta]\right]
\end{aligned}$$

4. MOMENTS VERSUS CUMULANTS

That the conditional expression of the n^{th} moment requires $n + \max(0, n-3)$ terms indicates a "bend in the road" between $n = 3$ and $n = 4$. It is hardly coincidental that moments beyond the third are not additive. If X and Y are independent random variables with means μ and ν, the n^{th} moment of their sum is:

$$\begin{aligned}
M_n[X+Y] &= E\left[\{(X+Y)-(\mu+\nu)\}^n\right] = E\left[\{(X-\mu)+(Y-\nu)\}^n\right] \\
&= \sum_{k=0}^{n}\binom{n}{k}E\left[(X-\mu)^{n-k}(Y-\nu)^k\right] \\
&= \sum_{k=0}^{n}\binom{n}{k}E\left[(X-\mu)^{n-k}\right]E\left[(Y-\nu)^k\right] \\
&= \sum_{k=0}^{n}\binom{n}{k}M_{n-k}[X]M_k[Y] \\
&= M_n[X]+M_n[Y]+\sum_{k=1}^{n-1}\binom{n}{k}M_{n-k}[X]M_k[Y] \\
&= M_n[X]+M_n[Y]+\sum_{k=2}^{n-2}\binom{n}{k}M_{n-k}[X]M_k[Y]
\end{aligned}$$

The Σ operator disrupts the additivity when its range is non-empty, i.e., when $n-2 \geq 2$, or $n \geq 4$. So, with independence, the first three moments are additive; in the fourth moment the term $\binom{4}{2}M_{4-2}[X]M_2[Y] = 6M_2[X]M_2[Y] = 6Var[X]Var[Y]$ disrupts the additivity, a term analogous to $6\underset{\Theta}{E}\left[Var[X|\Theta]\right]\underset{\Theta}{Var}\left[E[X|\Theta]\right]$ in the kurtosis formula.

Nevertheless, adjustments to moments higher than the third can obviate the disruption and restore additivity. Such adjusted moments are known as cumulants. The fourth-order adjustment is:

$$M_4[X+Y]-3M_2[X+Y]^2 = M_4[X]+M_4[Y]+6M_2[X]M_2[Y]-3M_2[X+Y]^2$$
$$= M_4[X]+M_4[Y]+6M_2[X]M_2[Y]-3(M_2[X]+M_2[Y])^2$$
$$= M_4.[X]-3M_2[X]^2+M_4[Y]-3M_2[Y]^2.$$

Thus, the fourth cumulant, $\kappa_4[X]=M_4[X]-3M_2[X]^2 = Kurt_4[X]-3Var[X]^2$ is additive.[5]

Since the collective risk model involves sums of independent random variables, the fourth cumulant, which we shall call the excess kurtosis, will prove more useful than the fourth moment. Its conditional expression is:

$$XsKurt[X]=Kurt[X]-3Var[X]^2$$
$$=\underset{\Theta}{E}[Kurt[X|\Theta]]+4\underset{\Theta}{Cov}[Skew[X|\Theta],E[X|\Theta]]+6\underset{\Theta}{Coskew}[Var[X|\Theta],E[X|\Theta],E[X|\Theta]]$$
$$+6\underset{\Theta}{E}[Var[X|\Theta]]\underset{\Theta}{Var}[E[X|\Theta]]+\underset{\Theta}{Kurt}[E[X|\Theta]]-3\left(\underset{\Theta}{E}[Var[X|\Theta]]+\underset{\Theta}{Var}[E[X|\Theta]]\right)^2$$
$$=\underset{\Theta}{E}[Kurt[X|\Theta]]+4\underset{\Theta}{Cov}[Skew[X|\Theta],E[X|\Theta]]+6\underset{\Theta}{Coskew}[Var[X|\Theta],E[X|\Theta],E[X|\Theta]]$$
$$+\underset{\Theta}{Kurt}[E[X|\Theta]]-3\underset{\Theta}{E}[Var[X|\Theta]]^2-3\underset{\Theta}{Var}[E[X|\Theta]]^2$$
$$=\underset{\Theta}{E}[Kurt[X|\Theta]]-3\underset{\Theta}{E}\left[Var[X|\Theta]^2\right]+4\underset{\Theta}{Cov}[Skew[X|\Theta],E[X|\Theta]]$$
$$+6\underset{\Theta}{Coskew}[Var[X|\Theta],E[X|\Theta],E[X|\Theta]]+\underset{\Theta}{Kurt}[E[X|\Theta]]-3\underset{\Theta}{Var}[E[X|\Theta]]^2$$
$$+3\underset{\Theta}{E}\left[Var[X|\Theta]^2\right]-3\underset{\Theta}{E}[Var[X|\Theta]]^2$$
$$=\underset{\Theta}{E}[XsKurt[X|\Theta]]+4\underset{\Theta}{Cov}[Skew[X|\Theta],E[X|\Theta]]$$
$$+6\underset{\Theta}{Coskew}[Var[X|\Theta],E[X|\Theta],E[X|\Theta]]+\underset{\Theta}{XsKurt}[E[X|\Theta]]+3\underset{\Theta}{Var}[Var[X|\Theta]]$$

[5] The formulas for higher-order cumulants become increasingly more complicated. The reader can verify the additivity of the next two cumulants according to the definitions (cf. Section 6):

$$\kappa_5[X]=M_5[X]-10Var[X]Skew[X]$$

$$\kappa_6[X]=M_6[X]-15Var[X]\kappa_4[X]-10Skew[X]^2-15Var[X]^3.$$

5. MOMENTS OF THE COLLECTIVE RISK MODEL

The collective risk model, which casualty actuaries must study for their examinations, is stock-in-trade, especially in the field of reinsurance. It considers aggregate loss S as the sum of a random number N of independent, identically distributed claims: $S = X_1 + \ldots + X_N$. Here we will apply our conditional formulas[6] to derive the first four cumulants of S in terms of those of X and N. Because the X_i are independent and identically distributed, as well as due to the additivity of cumulants, $E[S|N] = N\,E[X]$, $Var[S|N] = NVar[X]$, $Skew[S|N] = N\,Skew[X]$, and $XsKurt[S|N] = N\,XsKurt[X]$. Because $E[X]$, $Var[X]$, $Skew[X]$, and $XsKurt[X]$ are constants, we may remove them from moments conditional upon N, being careful to raise them to the power of the conditional moments.

The first cumulant, the mean, is trivial: $E[S] = \underset{N}{E}\big[E[S|N]\big] = \underset{N}{E}\big[N\,E[X]\big] = E[N]E[X]$. For the second, the variance:

$$\begin{aligned}
Var[S] &= \underset{N}{E}\big[Var[S|N]\big] + \underset{N}{Var}\big[E[S|N]\big] \\
&= \underset{N}{E}\big[NVar[X]\big] + \underset{N}{Var}\big[N\,E[X]\big] \\
&= E[N]Var[X] + Var[N]E[X]^2.
\end{aligned}$$

Every actuary at some time learned this formula; to many it remains familiar.

However, the third moment is not studied, and hence, not commonly known:

$$\begin{aligned}
Skew[S] &= \underset{N}{E}\big[Skew[S|N]\big] + 3\underset{N}{Cov}\big[Var[S|N], E[S|N]\big] + \underset{N}{Skew}\big[E[S|N]\big] \\
&= \underset{N}{E}\big[N\,Skew[X]\big] + 3\underset{N}{Cov}\big[N\,Var[X], N\,E[X]\big] + \underset{N}{Skew}\big[N\,E[X]\big] \\
&= E[N]Skew[X] + 3Cov[N, N]Var[X]E[X] + Skew[N]E[X]^3 \\
&= E[N]Skew[X] + 3Var[N]Var[X]E[X] + Skew[N]E[X]^3
\end{aligned}$$

---

[6] We will change the nomenclature of these formulas so as to agree with that of the collective risk model, i.e., S will appear instead of X, and N instead of Θ. In this section X will represent the severity of a claim.

Nonetheless, this formula appears in Patrik [1996, 377] and Klugman [1998, 298].

Last, we derive the excess kurtosis, whose formula we have not seen in print before:

$$\begin{aligned}
XsKurt[S] &= \underset{N}{E}\big[XsKurt[S|N]\big] + 4\underset{N}{Cov}\big[Skew[S|N], E[S|N]\big] \\
&\quad + 6\underset{N}{Coskew}\big[Var[S|N], E[S|N], E[S|N]\big] + \underset{N}{XsKurt}\big[E[S|N]\big] + 3\underset{N}{Var}\big[Var[S|N]\big] \\
&= \underset{N}{E}\big[N\,XsKurt[X]\big] + 4\underset{N}{Cov}\big[N\,Skew[X], N\,E[X]\big] \\
&\quad + 6\underset{N}{Coskew}\big[N\,Var[X], N\,E[X], N\,E[X]\big] + \underset{N}{XsKurt}\big[N\,E[X]\big] + 3\underset{N}{Var}\big[N\,Var[X]\big] \\
&= E[N]XsKurt[X] + 4Cov[N,N]Skew[X]E[X] \\
&\quad + 6Coskew[N,N,N]Var[X]E[X]^2 + XsKurt[N]E[X]^4 + 3Var[N]Var[X]^2 \\
&= E[N]XsKurt[X] + 4Var[N]Skew[X]E[X] \\
&\quad + 6Skew[N]Var[X]E[X]^2 + XsKurt[N]E[X]^4 + 3Var[N]Var[X]^2.
\end{aligned}$$

6. THE CUMULANT GENERATING FUNCTION

The moment generating function of a sum of independent random variables equals the product of their moment generating functions. Since logarithms convert multiplication into addition, it is natural to consider the logarithm of the moment generating function, which has come to be known as the cumulant generating function ψ (c.g.f.), i.e., $\psi_X(t) = \ln E\big[e^{tX}\big]$. Its derivatives at zero are called cumulants:[7] $\kappa_i[X] = \psi_X^{[i]}(0)$. If the X_i are independent of one another:

$$\psi_{\sum_i X_i}(t) = \ln E\left[e^{t\sum_i X_i}\right] = \ln E\left[\prod_i e^{tX_i}\right] = \ln \prod_i E\big[e^{tX_i}\big] = \sum_i \ln E\big[e^{tX_i}\big] = \sum_i \psi_{X_i}(t).$$

---

[7] In Section 4 we introduced cumulants as "moments adjusted to restore additivity." This hardly suffices for a definition, and we have not proven the existence and the uniqueness of the adjustment. The derivatives of the c.g.f. at zero constitute a proper definition of cumulant, and the Taylor-series argument of this section can be made into a rigorous proof of the uniqueness of the adjustment.

Since differentiation is a linear operator, the cumulant of a sum of independent random variables equals the sum of the cumulants of the random variables. The first three cumulants equal the mean, the variance, and the skewness. But equality ceases with the fourth cumulant: $\kappa_4[X] = M_4[X] - 3M_2[X]^2 = E\left[(X-\mu)^4\right] - 3Var[X]^2$ (Daykin [1994, 23] and Halliwell [2003, 65]). Here we will show the relevance of the c.g.f., to (1) the expression of cumulants in terms of moments and (2) the moments of the collective risk model.

First, the Taylor-series expansion of the c.g.f. embeds the cumulants:

$$\psi_X(t) = \psi_X(0) + \sum_{j=1}^{\infty} \psi_X^{[j]}(0)t^j/j! = 0 + \sum_{j=1}^{\infty} \kappa_j[X]t^j/j!.$$

The central moments of X, $M_n[X] = E\left[(X-\mu)^n\right]$, are similar coefficients in the Taylor-series expansion of the moment generating function of $X - \mu$:

$$E\left[e^{t(X-\mu)}\right] = 1 + \sum_{j=1}^{\infty} M_j[X]t^j/j! = 1 + 0 + \sum_{j=2}^{\infty} M_j[X]t^j/j!.$$

We can combine these two equations to relate the cumulants and the moments:

$$\begin{aligned}\sum_{j=1}^{\infty} \kappa_j[X]t^j/j! &= \psi_X(t) = \ln E\left[e^{tX}\right] = \ln E\left[e^{t(X-\mu)+t\mu}\right] = t\mu + \ln E\left[e^{t(X-\mu)}\right] \\ &= t\mu + \ln\left(1 + \sum_{j=2}^{\infty} M_j[X]t^j/j!\right).\end{aligned}$$

But the logarithm has its own Taylor-series expansion for $-1 < x \le 1$, viz.:

$$\ln(1+x) = x - x^2/2 + x^3/3 - x^4/4 + \ldots = \sum_{k=1}^{\infty} (-1)^{k-1} x^k/k.$$

So the relationship can be expressed as two polynomials in t:

$$\sum_{j=1}^{\infty}\kappa_j[X]t^j/j! = t\mu + \ln\left(1+\sum_{j=2}^{\infty}M_j[X]t^j/j!\right)$$
$$= t\mu + \sum_{k=1}^{\infty}(-1)^{k-1}\left(\sum_{j=2}^{\infty}M_j[X]t^j/j!\right)^k \Big/ k$$
$$= \mu t^1/1! + \sum_{j=2}^{\infty}M_j[X]t^j/j! + \sum_{k=2}^{\infty}(-1)^{k-1}\left\{\sum_{j=2}^{\infty}M_j[X]t^j/j!\right\}^k \Big/ k.$$

Matching coefficients of identical polynomials must be equal. It is the last expression on the right side of the final equation that complicates the matching; however, it is quartic and higher in *t*. Hence, the first three cumulants must be the mean, the variance, and the skewness. And the formula for higher cumulants begins as $\kappa_j[X] = M_j[X] + \dots$.

As an example of higher-order matching, we will derive the kurtosis formula. A fourth power of *t* arises in the last expression only from $k = 2$ powers of two, or as 2+2:

$$\kappa_4[X]t^4/4! = M_4[X]t^4/4! + (-1)^{2-1}\{(M_2[X]t^2/2!)(M_2[X]t^2/2!)\}/2$$
$$= M_4[X]t^4/4! - M_2[X]^2 t^4/8$$
$$\kappa_4[X] = M_4[X] - 3M_2[X]^2.$$

The fifth cumulant is a little more complicated, still involving $k = 2$, but obtained twice as 2+3 and 3+2. The sixth cumulant involves $k = 2$ as 2+4, 3+3, and 4+2, as well as $k = 3$ as 2+2+2. This c.g.f. technique is arguably the easiest way to derive the formulas of footnote 5.

Second, we will derive the c.g.f. of the collective risk model $S = X_1 + \dots X_N$, being mindful of the change in nomenclature (cf. footnote 6):

$$\psi_S(t) = \ln e^{\psi_S(t)} = \ln \underset{N}{E}\left[e^{\psi_{S|N}(t)}\right] = \ln \underset{N}{E}\left[e^{\psi_X(t)N}\right] = \psi_N(\psi_X(t)) = (\psi_N \circ \psi_X)(t).$$

So the c.g.f. of the aggregate loss is the composition of the cumulant generating functions of frequency and severity (Daykin [1994, 59]). This is the most elegant way to derive the aggregate cumulants, and it is more efficient than the conditional-moment technique of Section 5. To show this, we will derive the first two moments.

$$\psi_S(t) = \psi'_N(\psi_X(t))\psi'_X(t)$$
$$\therefore E[S] = \psi'_S(0) = \psi'_N(\psi_X(0))\psi'_X(0) = \psi'_N(0)\psi'_X(0) = E[N]E[X]$$

$$\psi''_S(t) = \psi'_N(\psi_X(t))\psi''_X(t) + \psi''_N(\psi_X(t))\psi'_X(t)^2$$
$$\therefore Var[S] = \psi''_S(0) = \psi'_N(0)\psi''_X(0) + \psi''_N(0)\psi'_X(0)^2 = E[N]Var[X] + Var[N]E[X]^2.$$

Curious and ambitious readers, performing the third and fourth derivatives, can verify the formulas in Section 5 for the aggregate skewness and excess kurtosis.

7. CONCLUSION

We have shown how unconditional moments can be expressed in terms of conditional moments and co-moments. Adjusting moments into cumulants allowed us to form fairly simple formulas for the skewness and the excess kurtosis of the collective risk model. These formulas can also be derived directly from the cumulant-generating function. Actuaries who have been reluctant to apply the method of moments to just the first two moments of the collective risk model can now with these formulas fit more versatile distributions to more than two moments. One ought to be more comfortable with extrapolations into the right tail of an aggregate loss distribution after having considered its skewness and kurtosis.

Aside from the collective risk model, a conditioning partition Θ can change the moments of a sum of independent random variables without changing their unconditional moments. The appendix shows how this can be done in loss reserving. Moreover, if some amount of capital or risk margin were allocated to a moment, conditioning would allow a sub-allocation to the partitions.

Acknowledgment

The author gratefully acknowledges the help of reviewers Jason Cook, FCAS, and Evan Glisson, FCAS, from whose comments this paper has much improved.

REFERENCES

[1.] Daykin, C. D., T. Pentikäinen, and M. Pesonen, *Practical Risk Theory for Actuaries*, London: Chapman & Hall, 1994.

[2.] Halliwell, Leigh J., "The Valuation of Stochastic Cash Flows," *CAS Forum* (Reinsurance Call Papers, Spring 2003), pp. 1-68, www.casact.org/pubs/forum/03spforum/03spf001.pdf.

[3.] Klugman, Stuart A., H. Panjer, and G. Willmot, *Loss Models: From Data to Decisions*, New York: John Wiley & Sons, 1998.

[4.] Patrik, Gary S., "Reinsurance," *Foundations of Casualty Actuarial Science*, Third Edition, Casualty Actuarial Society, 1996, pp. 283-380.

APPENDIX A

Conditional Probability and Claim Development

A recent assignment spurred our interest in the subject of this paper. We had a list of the case reserves of about 200 claims, and were satisfied that the total IBNER[8] for them was zero, i.e., $E[\text{IBNER}] = 0$. But in addition, we wanted some measure of the variance. The claims stemmed from an unusual exposure, and we deemed no other data sources appropriate. Since we assumed the average development to be zero, the claim list itself could serve as an empirical distribution f_X with moments $\mu \pm \sigma$. Regarding the *n* claims as independent, we might decide the moments of the total unpaid loss (i.e., case plus IBNER) to be $n\mu \pm \sqrt{n}\sigma$, or total IBNER to be $0 \pm \sqrt{n}\sigma$. But this ignores the likelihood of rank correlation, i.e., that after development large claims tend to stay large, and small claims tend to stay small. Hence, $\sqrt{n}\sigma$ is a maximal value.

Therefore, we decided to order the claims by their case reserves and to stratify them into 10 groups of approximately 20 claims. Belonging to a stratum is the event Θ that conditions a claim's probability density as $f_{X|\Theta}$. Since stratification provides no new information, $f_X(x) = \underset{\Theta}{E}\left[f_{X|\Theta}(x)\right]$. Then we assumed that each claim would develop as follows: with probability *p* its distribution would remain that of its stratum and with probability $q = 1 - p$ it would migrate randomly.

Consequently, the distribution of a developed claim is a mixture of distributions; with probability *p* the developed claim is distributed as $f_{X|\Theta}$ and with probability *q* as f_X. Let *Y* be the developed amount of claim *X*. Mixing is easy with moment generating functions.

[8] IBNER means "Incurred But Not Enough Reported (or Reserved)." Cf. Patrik [1996], 350.

The moment generating function of Y conditional upon the stratum of X is $M_{Y|\Theta}(t) = pM_{X|\Theta}(t) + qM_X(t)$. The overall, or unconditional, moment generating function of a developed claim is:

$$\begin{aligned}
M_Y(t) &= \underset{\Theta}{E}\left[M_{Y|\Theta}(t)\right] \\
&= \underset{\Theta}{E}\left[pM_{X|\Theta}(t) + qM_X(t)\right] \\
&= p\,\underset{\Theta}{E}\left[M_{X|\Theta}(t)\right] + qM_X(t) \\
&= pM_X(t) + qM_X(t) \\
&= M_X(t).
\end{aligned}$$

Since equality of moment generating functions implies identical distributions, Y is distributed as X. Since we have provided no new information, this "conservation of distribution" is fitting.

But the reader may now be wondering how the variance of total IBNER can change despite the conservation of the overall distribution. The paradox is resolved with a distinction: variance pertains to the *sum* of claims, whereas conservation pertains to their *mixture*, more accurately, to the mixture of their distributions. The overall or unconditional variance $Var[X]$ is conserved, but its apportionment between $\underset{\Theta}{E}\left[Var[X|\Theta]\right]$ and $\underset{\Theta}{Var}\left[E[X|\Theta]\right]$ depends on Θ. At the one extreme, a blunt or non-discriminating stratification Θ tells nothing about X: $Var[X|\Theta] = Var[X]$. In this case:

$$\begin{aligned}
\underset{\Theta}{Var}\left[E[X|\Theta]\right] &= Var[X] - \underset{\Theta}{E}\left[Var[X|\Theta]\right] \\
&= Var[X] - \underset{\Theta}{E}\left[Var[X]\right] \\
&= Var[X] - Var[X] \\
&= 0.
\end{aligned}$$

Conversely, if the variance of the conditional mean is zero, $E[X|\Theta]$ must be constant, or $E[X|\Theta] = E[X]$. So the conditional distributions of a blunt stratification tend to be indistinguishable as to their first two moments. In this case the variance of the sum tends toward the maximal $\sqrt{n}\sigma$. At the opposite extreme, Θ is so fine or discriminating that $Var[X|\Theta] = 0$. Then:

$$\begin{aligned} \underset{\Theta}{Var}[E[X|\Theta]] &= Var[X] - \underset{\Theta}{E}[Var[X|\Theta]] \\ &= Var[X] - \underset{\Theta}{E}[0] \\ &= Var[X]. \end{aligned}$$

This means that the all the variance is between the strata, no variance is within a stratum. In this case the variance of the sum tends toward the minimal value of zero. To borrow and mix notions from optics and credibility, the blunt stratification passes the white light of zero credibility; the fine stratification like a prism refracts light into the spectrum of full credibility.

Since we will be conditioning on migration M, we will drop Θ and speak of the i^{th} stratum. Let there be s strata, and let $\pi_i > 0$ be the probability for a claim to be in the i^{th} stratum, as determined by the actual portion of claims in that stratum. Though the strata need not to be balanced, or of equal population, $\sum_{i=1}^{s} \pi_i = 1$. We may model developed claim Y_i of the i^{th} stratum as follows. Randomly draw one undeveloped claim from each stratum's distribution; these $X_1, \ldots, X_s$ are independent. Then form an "unstratified" or average claim X as the choice of X_j with probability π_j. Finally, flip a "Bernoulli coin" with probability p of heads. If the coin lands heads, let Y_i equal X_i; otherwise, let it equal X.

The mean of the developed claim is $E[Y_i] = \underset{\mathrm{M}}{E}\left[E[Y_i|\mathrm{M}]\right] = pE[X_i] + qE[X]$. According to the formula of Section 3, the variance is:

$$\begin{aligned}
Var[Y_i] &= \underset{\mathrm{M}}{E}\left[Var[Y_i|\mathrm{M}]\right] + \underset{\mathrm{M}}{Var}\left[E[Y_i|\mathrm{M}]\right] \\
&= pVar[X_i] + qVar[X] \\
&\quad + p(E[X_i] - E[Y_i])^2 + q(E[X] - E[Y_i])^2 \\
&= pVar[X_i] + qVar[X] \\
&\quad + p(qE[X_i] - qE[X])^2 + q(pE[X] - pE[X_i])^2 \\
&= pVar[X_i] + qVar[X] + pq(E[X_i] - E[X])^2.
\end{aligned}$$

Since considerations of rank correlation drew us to this model, we should also determine the covariance of Y_i with X_i. Taking the next formula without proof,[9] we have:

$$Cov[Y_i, X_i] = \underset{\mathrm{M}}{E}\left[Cov[Y_i, X_i|\mathrm{M}]\right] + \underset{\mathrm{M}}{Cov}\left[E[Y_i|\mathrm{M}], E[X_i|\mathrm{M}]\right].$$

The second term on the right side of the equation is zero. For the migration M does not affect the expectation of X_i, and the covariance of something with a constant is zero. Hence, $Cov[Y_i, X_i] = \underset{\mathrm{M}}{E}\left[Cov[Y_i, X_i|\mathrm{M}]\right]$. In the following reduction, we must consider that the random migration can return (P) with probability π_i to the i^{th} stratum. Again, a covariance term becomes zero due to the immunity of X_i to **P**:

---

9 The proof hinges on $Cov[X, Y|\Theta] = \underset{\Theta}{E}\left[E\left[\{(X - \mu_\Theta) - (\mu_\Theta - \mu)\}\{(Y - \nu_\Theta) - (\nu_\Theta - \nu)\}|\Theta\right]\right]$.

$$\begin{aligned}
Cov[Y_i, X_i] &= \underset{\mathrm{M}}{E}\left[Cov[Y_i, X_i|\mathrm{M}]\right] \\
&= pCov[X_i, X_i] + qCov[X, X_i] \\
&= pCov[X_i, X_i] + q\left\{\underset{\mathrm{P}}{E}\left[Cov[X, X_i|\mathrm{P}]\right] + \underset{\mathrm{P}}{Cov}\left[E[X|\mathrm{P}], E[X_i|\mathrm{P}]\right]\right\} \\
&= pCov[X_i, X_i] + q\underset{\mathrm{P}}{E}\left[Cov[X, X_i|\mathrm{P}]\right] \\
&= pCov[X_i, X_i] + q\sum_{j=1}^{s}\pi_j Cov[X_j, X_i] \\
&= pCov[X_i, X_i] + q\pi_i Cov[X_i, X_i] \\
&= (p + q\pi_i)Var[X_i].
\end{aligned}$$

The covariance of the developed claim amount with the undeveloped is positive, but the correlation coefficient is more informative:

$$Corr[Y_i, X_i] = \frac{Cov[Y_i, X_i]}{\sqrt{Var[Y_i]Var[X_i]}} = \frac{(p + q\pi_i)Var[X_i]}{\sqrt{Var[Y_i]Var[X_i]}} = (p + q\pi_i)\sqrt{\frac{Var[X_i]}{Var[Y_i]}}.$$

The correlation increases with respect to *p*, the probability that the distribution of a developed claim remains that of its stratum, from a minimum of $\pi_i\sqrt{\frac{Var[X_i]}{Var[X]}}$ for $p = 0$ to a maximum of 1 for $p = 1$. It seems that $Corr[Y_i, X_i] \approx p$. However, this is Pearson correlation, whereas we are concerned with rank, or Spearman, correlation.

Because an analytic answer eluded us, we resorted to simulation. Keeping with the assignment that spurred our interest, we simulated 1,000 iterations of the "development" of the integers from 1 to 200 in $s = 10$ groups of 20 consecutive integers over a range of non-migration probabilities *p* from 0% to 100% in steps of 5%. We randomly permuted the integers within each group – this alone would suffice if $p = 1$ and inter-group migrations were impossible. But then we flipped the Bernoulli coin for each integer, marked which places were the migrating "tails," and randomly permuted among those places their integers.

Then we calculated the rank correlation for that iteration, and averaged it over all the iterations. The table below contains the result:

#Iter	#Goups	#InGrp	*p*	RankCorr
1,000	10	20	0%	0.000
1,000	10	20	5%	0.046
1,000	10	20	10%	0.100
1,000	10	20	15%	0.146
1,000	10	20	20%	0.201
1,000	10	20	25%	0.250
1,000	10	20	30%	0.302
1,000	10	20	35%	0.349
1,000	10	20	40%	0.396
1,000	10	20	45%	0.446
1,000	10	20	50%	0.497
1,000	10	20	55%	0.550
1,000	10	20	60%	0.601
1,000	10	20	65%	0.648
1,000	10	20	70%	0.696
1,000	10	20	75%	0.747
1,000	10	20	80%	0.797
1,000	10	20	85%	0.844
1,000	10	20	90%	0.895
1,000	10	20	95%	0.944
1,000	10	20	100%	0.990

Indeed, it seems that the rank correlation approximates p. Nonetheless, it cannot exactly equal p. For at a near 100% probability of not migrating, permutation within each group still disrupts a perfect correlation. Therefore, we suspect RankCorr(p) to start out at zero with a slope of unity, but to be slightly concave (i.e., to have a negative second derivative) so that it loses ground to p as p increases to one.

In sum, as p, the probability of not migrating (i.e., the probability for the distribution of a claim to remain that of its stratum) approaches zero, stratification becomes irrelevant. Regardless of how the claims are stratified, they will all develop according to the overall distribution. This will produce an aggregate standard deviation approaching the maximal $\sqrt{n}\sigma$. And if there were only one stratum, migration would be from overall to overall, and the aggregate standard deviation again would be $\sqrt{n}\sigma$. But as *p* increases, and as the strata become narrower, the aggregate standard deviation decreases. In the extreme, with one claim per stratum (better, with zero variance within each stratum) and $p = 1$, the aggregate standard deviation is zero.

Pondering these relations with two moments led us to the idea of adding higher moments to the conditional distributions, and thence to treating the higher moments of the collective risk model. Although we do not intend for this to be a paper on a new development method, the reader can see how this claim-by-claim method can be employed to apportion moments of loss that mesh with any desired aggregate moments, as well as to obtain useful subtotals, e.g., by accident year.