

Frequency Modeling

A short course authored by the Actuarial Community

19 Jan 2021

How Frequency Augments Severity Information

Basic Terminology

- ▶ **Claim** - indemnification upon the occurrence of an insured event
- ▶ **Loss** - some authors use claim and loss interchangeably, others think of loss as the amount suffered by the insured whereas claim is the amount paid by the insurer
- ▶ **Frequency** - how often an insured event occurs, typically within a policy contract
- ▶ **Count** - In this chapter, we focus on count random variables that represent the number of claims, that is, how frequently an event occurs
- ▶ **Severity** - Amount, or size, of each payment for an insured event

The Importance of Frequency

- ▶ Insurers pay claims in monetary units, e.g., US dollars. So, why should they care about how frequently claims occur?
- ▶ Many ways to use claims modeling – easiest to motivate in terms of pricing for personal lines insurance
 - ▶ Recall from Chapter 1 that setting the price of an insurance good can be a perplexing problem.
 - ▶ In manufacturing, the cost of a good is (relatively) known
 - ▶ Other financial service areas, market prices are available
- ▶ Insurance tradition: Start with an expected cost. Add "margins" to account for the product's riskiness, expenses incurred in servicing the product, and a profit/surplus allowance for the insurance company.
- ▶ Think of the expected cost as the expected number of claims times the expected amount per claims, that is, expected *frequency times severity*.
 - ▶ Claim amounts, or severities, will turn out to be relatively homogeneous for many lines of business and so we begin our investigations with frequency modeling.

Other Ways that Frequency Augments Severity Information

- ▶ **Contractual** - For example, deductibles and policy limits are often in terms of each occurrence of an insured event
- ▶ **Behaviorial** - Explanatory (rating) variables can have different effects on models of how often an event occurs in contrast to the size of the event.
 - ▶ In healthcare, the decision to utilize healthcare by individuals is related primarily to personal characteristics whereas the cost per user may be more related to characteristics of the healthcare provider (such as the physician).

Other Ways that Frequency Augments Severity Information

II

- ▶ **Databases.** Many insurers keep separate data files that suggest developing separate frequency and severity models. This recording process makes it natural for insurers to model the frequency and severity as separate processes.
 - ▶ Policyholder file that is established when a policy is written. This file records much underwriting information about the insured(s), such as age, gender and prior claims experience, policy information such as coverage, deductibles and limitations, as well as the insurance claims event.
 - ▶ Claims file, records details of the claim against the insurer, including the amount.
 - ▶ There may also be a "payments" file that records the timing of the payments although we shall not deal with that here.

Other Ways that Frequency Augments Severity Information

III

▶ **Regulatory and Administrative**

- ▶ Regulators routinely require the reporting of claims numbers as well as amounts.
- ▶ This may be due to the fact that there can be alternative definitions of an "amount," e.g., paid versus incurred, and there is less potential error when reporting claim numbers.

REVIEW

In this section, you learn how to summarize the importance of frequency modeling in terms of

- ▶ contractual,
- ▶ behavioral,
- ▶ database, and
- ▶ regulatory/administrative motivations.

Basic Frequency Distributions

Frequency Distributions

- ▶ **Frequency** - how often an insured event occurs, typically within a policy contract
- ▶ Discrete probability distributions, called **count distributions**, model the number of losses to a policyholder or the number of claims to an insurance company

Foundations

- ▶ Focus on frequency random variable N with support on $k = 0, 1, 2, \dots$
- ▶ **Probability mass function** (pmf) is denoted as $\Pr(N = k) = p_k$
- ▶ **Cumulative distribution function** (cdf) is denoted as

$$\Pr(N \leq x) = F(x) = \begin{cases} \sum_{k=0}^{[x]} p_k, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

- ▶ Summarize distribution through **moments**:
 - ▶ **Mean**:

$$E(N) = \mu = \sum_{k=0}^{\infty} k p_k$$

- ▶ **Variance**:

$$\text{Var}(N) = E(N - \mu)^2 = E(N^2) - \mu^2 = \sum_{k=0}^{\infty} k^2 p_k - \left(\sum_{k=0}^{\infty} k p_k \right)^2$$

Probability Generating Function

- ▶ Probability generating function (pgf) is:

$$P(z) = E(z^N) = \sum_{k=0}^{\infty} z^k p_k$$

- ▶ Taking the m th derivative, pgf “generates” probabilities:

$$P^{(m)}(0) = \frac{\partial^m}{\partial z^m} P(z)|_{z=0} = p_m m!$$

- ▶ Further, pgf can generate moments:

$$P^{(1)}(1) = \sum_{k=0}^{\infty} k p_k = E(N)$$

and

$$P^{(2)}(1) = E[N(N-1)]$$

Important Frequency Distributions

- ▶ Three important frequency distributions are:
 - ▶ Poisson
 - ▶ Binomial
 - ▶ Negative binomial
- ▶ They are important because:
 - ▶ They fit well many insurance data sets of interest
 - ▶ They provide the basis for more complex distributions that even better approximate real situations of interest

Poisson Distribution

- ▶ This distribution has positive parameter λ , pmf

$$p_k = \frac{e^{-\lambda} \lambda^k}{k!}$$

and pgf

$$P(z) = M(\ln z) = \exp(\lambda(z - 1))$$

- ▶ Expectation is $E(N) = \lambda$, which is same as variance, $Var(N) = \lambda$

Binomial Distribution

- ▶ This distribution has parameters m (positive integer) and $0 < q < 1$, pmf

$$p_k = \binom{m}{k} q^k (1 - q)^{m-k}$$

and pgf

$$P(z) = (1 + q(z - 1))^m$$

- ▶ $k = 0, 1, 2, \dots, m$
- ▶ Mean is $E(N) = mq$ and variance is $\text{Var}(N) = mq(1 - q)$
- ▶ If $m = 1$, called **Bernoulli distribution**
- ▶ As $0 < q < 1$, we have $\text{Var}(N) < E(N)$

Negative Binomial Distribution

- ▶ This distribution has positive parameters (r, β) , pmf

$$p_k = \binom{k+r-1}{k} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^k$$

and pgf

$$P(z) = (1 - \beta(z - 1))^{-r}$$

- ▶ Expectation is $E(N) = r\beta$ and variance is $\text{Var}(N) = r\beta(1 + \beta)$
- ▶ If $r = 1$, called **geometric distribution**
- ▶ $\text{Var}(N) > E(N)$

REVIEW

In this section, we learned how to:

- ▶ Determine quantities that summarize a distribution such as the distribution and survival function, as well as moments such as the mean and variance
- ▶ Define and compute the moment and probability generating functions
- ▶ Describe and understand relationships among three important frequency distributions, the binomial, Poisson, and negative binomial distributions

$(a, b, 0)$ Class

$(a, b, 0)$ Class

- ▶ *Definition.* A count distribution is a member of the $(a, b, 0)$ class if probabilities p_k satisfy

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k},$$

for constants a, b and for $k = 1, 2, 3, \dots$

- ▶ Only three distributions are members of the $(a, b, 0)$ class:
 - ▶ Poisson ($a = 0$),
 - ▶ binomial ($a < 0$), and
 - ▶ negative binomial ($a > 0$)
- ▶ Recursive expression provides a computationally efficient way to generate probabilities

$(a, b, 0)$ Class - Special Cases

► *Example: Poisson Distribution.*

► Recall $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$

$$\frac{p_k}{p_{k-1}} = \frac{\lambda^k/k!}{\lambda^{k-1}/(k-1)!} \frac{e^{-\lambda}}{e^{-\lambda}} = \frac{\lambda}{k}$$

$a = 0$, $b = \lambda$, and $p_0 = e^{-\lambda}$

► *Example: Binomial Distribution.* $a = \frac{-q}{1-q}$, $b = \frac{(m+1)q}{1-q}$, and $p_0 = (1-q)^m$

► *Example: Negative Binomial Distribution.* $a = \frac{\beta}{1+\beta}$, $b = \frac{(r-1)\beta}{1+\beta}$, and $p_0 = (1+\beta)^{-r}$

REVIEW

In this section, you learn how to:

- ▶ Define the $(a, b, 0)$ class of frequency distributions
- ▶ Discuss the importance of the recursive relationship underpinning this class of distributions
- ▶ Identify conditions under which this general class reduces to each of the binomial, Poisson, and negative binomial distributions

Estimating Frequency Distributions

Maximum Likelihood Estimation

- ▶ Consider a random sample of claim frequency data
 - ▶ There are a total of n observations
 - ▶ There are n_k observations each with a claim count of k , where $k = 0, 1, 2, \dots$
 - ▶ $\sum_{k=0}^{\infty} n_k = n$
- ▶ Want to fit a count distribution to these data (Poisson, negative binomial, binomial, etc.); we want to estimate the parameter(s) of a count distribution using these data
- ▶ Objective of **maximum likelihood estimation**: find the parameter value(s) that produce largest likelihood of observing data

Likelihood and Log-Likelihood Functions

- ▶ Let p_k denote the pmf for a count distribution
- ▶ **Likelihood**: function of estimable parameter(s) of the distribution (θ) with data (measured by n_k values, $k = 0, 1, 2, \dots$) fixed
- ▶ **Likelihood function**,

$$L(\theta) = (p_0)^{n_0} (p_1)^{n_1} (p_2)^{n_2} \cdots = \prod_{k=0}^{\infty} (p_k)^{n_k},$$

evaluated at observed data

- ▶ Alternatively, the **log-likelihood function** is,

$$l(\theta) = \ln L(\theta) = \sum_{k=0}^{\infty} n_k \ln(p_k),$$

- ▶ Maximize either the likelihood or log-likelihood function with respect to the parameter(s) to get the most likely parameter value(s) given observed data: the **maximum likelihood estimators (MLEs)**

REVIEW

In this section, you learned how to:

- ▶ Define a likelihood for a sample of observations from a discrete distribution
- ▶ Define the maximum likelihood estimator for a random sample of observations from a discrete distribution
- ▶ Calculate the maximum likelihood estimator for the binomial, Poisson, and negative binomial distributions

Other Frequency Distributions

Other Frequency Distributions

- ▶ Many other count distributions are needed in practice
- ▶ Three basic distributions (Poisson, binomial, negative binomial) can explain claim probabilities in terms of known (to insurer) variables such as age, sex, geographic location, etc.
- ▶ This field of statistical study is known as **regression analysis**
- ▶ To extend basic count distributions, consider two approaches:
 - ▶ Zero truncation or modification
 - ▶ Mixtures

Zero Truncation or Modification

Zero Truncation or Modification

- ▶ Modify probability of zero claims in terms of the $(a, b, 0)$ class
- ▶ *Definition.* A count distribution is a member of the $(a, b, 1)$ class if probabilities p_k satisfy

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k},$$

for constants a, b and for $k = 2, 3, \dots$

- ▶ Note this starts at $k = 2$, not $k = 1$ (starts at p_1 , not p_0)
- ▶ All distributions in $(a, b, 0)$ are part of $(a, b, 1)$

Zero Truncation or Modification

- ▶ Pick a specific distribution in $(a, b, 0)$ class, and consider p_k for this member:
- ▶ Let p_k^M be corresponding probability for a member of $(a, b, 1)$ class, where M stands for “modified”
- ▶ Pick a new probability of zero claims, p_0^M , and define

$$c = \frac{1 - p_0^M}{1 - p_0}.$$

- ▶ Calculate the rest of the modified distribution as

$$p_k^M = cp_k, \quad k = 1, 2, \dots$$

- ▶ *Special Case: Truncated at Zero.* Assume probability of $N = 0$ is zero:

$$p_k^T = \begin{cases} 0 & k = 0 \\ \frac{1}{1-p_0} p_k & k \geq 1 \end{cases}$$

Modified Poisson Example

Example: Zero Truncated/Modified Poisson. For a Poisson distribution with mean 2, as a member of the $(a, b, 0)$ class, $a = 0$ and $b = \lambda = 2$

Use recursion $p_k = \lambda p_{k-1}/k = 2p_{k-1}/k$, after determining starting probabilities

k	p_k	p_k^T	p_k^M (where $p_0^M = 0.6$)
0	$p_0 = e^{-\lambda} = 0.135335$	$p_0^T = 0$	$p_0^M = 0.6$
1	$p_1 = p_0(0 + \frac{\lambda}{1}) = 0.27067$	$p_1^T = \frac{p_1}{1-p_0} = 0.313035$	$p_1^M = \frac{1-p_0^M}{1-p_0} p_1 = 0.125214$
2	$p_2 = p_1\left(\frac{\lambda}{2}\right) = 0.27067$	$p_2^T = p_1^T\left(\frac{\lambda}{2}\right) = 0.313035$	$p_2^M = p_1^M\left(\frac{\lambda}{2}\right) = 0.125214$
3	$p_3 = p_2\left(\frac{\lambda}{3}\right) = 0.180447$	$p_3^T = p_2^T\left(\frac{\lambda}{3}\right) = 0.208690$	$p_3^M = p_2^M\left(\frac{\lambda}{3}\right) = 0.083476$

Modified Distribution Exercise

Exercise: Course 3, May 2000, Exercise 37. You are given:

- ▶ p_k denotes the probability that the number of claims equals k for $k = 0, 1, 2, \dots$
- ▶ $\frac{p_n}{p_m} = \frac{m!}{n!}, m \geq 0, n \geq 0$

Using the corresponding zero-modified claim count distribution with $p_0^M = 0.1$, calculate p_1^M

REVIEW

In this section, you learned how to:

- ▶ Define the $(a, b, 1)$ class of frequency distributions and discuss the importance of the recursive relationship underpinning this class of distributions
- ▶ Interpret zero truncated and modified versions of the binomial, Poisson, and negative binomial distributions
- ▶ Compute probabilities using the recursive relationship

Mixture Distributions

Discrete/Finite Mixtures

- ▶ Suppose a population consists of several subgroups, each having their own distribution
- ▶ Randomly draw an observation from the population, without knowing from which subgroup we are drawing
- ▶ Suppose N_1 represents claims from “good” drivers (GD) and N_2 claims from “bad” drivers (BD). We draw:

$$N = \begin{cases} N_1 & \text{with prob } \alpha \\ N_2 & \text{with prob } (1 - \alpha). \end{cases}$$

- ▶ Here, α represents probability of drawing a “good” driver
- ▶ “Mixture” of two subgroups

Discrete Mixture Probability Mass Function

- For pmf

$$\begin{aligned}\Pr(N = k) &= \Pr(N = k, \text{GD}) + \Pr(N = k, \text{BD}) \\&= \Pr(N = k | \text{GD}) \Pr(\text{GD}) + \Pr(N = k | \text{BD}) \Pr(\text{BD}) \\&= \Pr(N_1 = k) \Pr(\text{GD}) + \Pr(N_2 = k) \Pr(\text{BD}) \\&= \alpha p_{N_1}(k) + (1 - \alpha) p_{N_2}(k)\end{aligned}$$

- Similar argument can be made for cdf

Discrete Mixture Example

Exercise. Exam "C" 170. In a certain town the number of common colds an individual will get in a year follows a Poisson distribution that depends on the individual's age and smoking status:

	Proportion of population	Mean number of colds
Children	0.3	3
Adult Non-Smokers	0.6	1
Adult Smokers	0.1	4

- ▶ Calculate the probability that a randomly drawn person has 3 common colds in a year
- ▶ Calculate the conditional probability that a person with exactly 3 common colds in a year is an adult smoker

Mixture Moments

- Start with the mean. Using [law of iterated expectations](#):

$$E(N) = \alpha E(N_1) + (1 - \alpha)E(N_2).$$

We can also write

$$N^2 = \begin{cases} N_1^2 & \text{with probability } \alpha \\ N_2^2 & \text{with probability } 1 - \alpha \end{cases}$$

Thus

$$E(N^2) = \alpha E(N_1^2) + (1 - \alpha)E(N_2^2).$$

Same argument holds for any moment

Continuous Mixtures

- ▶ Can extend mixture idea to an infinite number of subgroups
- ▶ Consider a population of drivers. The i th person has their own Poisson distribution with expected number of claims, λ_i
 - ▶ For some drivers, λ is small (better drivers), for others it is high (worse drivers). There is a distribution of λ
- ▶ A convenient distribution for λ is a **gamma distribution** with parameters (α, θ)
- ▶ One can check that if $N|\Lambda \sim \text{Poisson}(\Lambda)$ and if $\Lambda \sim \text{gamma}(\alpha, \theta)$:

$$N \sim \text{Negative Binomial}(r = \alpha, \beta = \theta)$$

Continuous Mixtures II

- ▶ Consider a general framework for a continuous mixture:
 - ▶ Let $(N|\Lambda = \lambda)$ have pmf $p_{N|\Lambda}(k|\lambda) = \Pr(N = k|\Lambda = \lambda)$
 - ▶ Let Λ have pdf $f_{\Lambda}(\lambda)$
 - ▶ Random draw:

$$p_k = \Pr(N = k) = E_{\Lambda}[p_{N|\Lambda}(k|\Lambda)] = \int_{\lambda} p_{N|\Lambda}(k|\lambda) f_{\Lambda}(\lambda) d\lambda$$

- ▶ Idea of above: first determine claim count pmf given a specific value λ , then take expectation over all possible values of λ to get claim count pmf for random draw
- ▶ Use the [law of iterated expectations](#) to calculate raw moments of N , the [law of total variance](#) to calculate variance of N

REVIEW

In this section, you learned how to:

- ▶ Define a mixture distribution when the mixing component is based on a finite number of sub-groups
- ▶ Compute mixture distribution probabilities from mixing proportions and knowledge of the distribution of each subgroup
- ▶ Define a mixture distribution when the mixing component is continuous

Goodness of Fit

Example: Singapore Automobile Data

- ▶ A 1993 portfolio of $n = 7,483$ automobile insurance policies from a major insurance company in Singapore.
- ▶ The count variable is the number of automobile accidents per policyholder.
- ▶ There were on average \$ 0.06989\$ accidents per person.

Table. Comparison of Observed to Fitted Counts

Based on Singapore Automobile Data

Count (k)	Observed (m_k)	Fitted Counts using the Poisson Distribution($n\hat{p}_k$)
0	6,996	6,977.86
1	455	487.70
2	28	17.04
3	4	0.40
4	0	0.01
<i>Total</i>	7,483	7,483.00

The average is $\bar{N} = \frac{0 \cdot 6996 + 1 \cdot 455 + 2 \cdot 28 + 3 \cdot 4 + 4 \cdot 0}{7483} = 0.06989$.

Singapore Data: Adequacy of the Poisson Model

- ▶ With the Poisson distribution
 - ▶ The maximum likelihood estimator of λ is $\hat{\lambda} = \bar{N}$.
 - ▶ Estimated probabilities, using $\hat{\lambda}$, are denoted as \hat{p}_k .
 - ▶ Fitted counts are 7,483 times the fitted probabilities ($n\hat{p}_j$).
- ▶ For goodness of fit, consider *Pearson's chi-square statistic*

$$\sum_k \frac{(m_k - n\hat{p}_k)^2}{n\hat{p}_k}.$$

- ▶ Assuming that the Poisson distribution is a correct model; this statistic has an asymptotic chi-square distribution
- ▶ The degrees of freedom (df) equals the number of cells minus one minus the number of estimated parameters.
- ▶ For the Singapore data, this is $df = 5 - 1 - 1 = 3$.
- ▶ The statistic is 41.98; the basic Poisson model is inadequate.

Example. Course C/Exam 4. May 2001, 19.

During a one-year period, the number of accidents per day was distributed as follows:

Number of Accidents	0	1	2	3	4	5
Number of Days	209	111	33	7	5	2

You use a chi-square test to measure the fit of a Poisson distribution with mean 0.60. The minimum expected number of observations in any group should be 5. The maximum number of groups should be used.

Determine the chi-square statistic.

REVIEW

In this section, you learned how to:

- ▶ Calculate a goodness of fit statistic to compare a hypothesized discrete distribution to a sample of discrete observations
- ▶ Compare the statistic to a reference distribution to assess the adequacy of the fit