Claims Frequency Distribution Models

Chapter 6

Stat 346 - Short-term Actuarial Math

Introduction

- Here we introduce a large class of counting distributions, which are discrete distributions with support consisting of non-negative integers.

- Generally used for modeling number of events, but in an insurance context, the number of claims within a certain period, e.g. one year.
- We call these claims frequency models.
- Let N denote the number of events (or claims). Its probability mass function (pmf), $p_k = \text{Pr}(N = k)$, for $k = 0, 1, 2, \ldots$, gives the probability that exactly k events (or claims) occur.

Some familiar discrete distributions

Some of the most commonly used distributions for number of claims:

- Binomial (with Bernoulli as special case)
- Poisson
- Geometric
- Negative Binomial
- The $(a, b, 0)$ class
- The $(a, b, 1)$ class

Bernoulli random variables

- N is Bernoulli if it takes only one of two possible outcomes:

$$N = \begin{cases} 1, & \text{if a claim occurs} \\ 0, & \text{otherwise} \end{cases} .$$

- q is the standard symbol for the probability of a claim, i.e. $\Pr(N = 1) = q$.
- We write $N \sim \text{Bernoulli}(q)$.
- Mean $\mathrm{E}(N) = q$ and variance $\mathrm{Var}(N) = q(1 - q)$

Binomial random variables

- We write $N \sim \text{Binomial}(m, q)$ if N has a Binomial distribution with pmf:

  $$p_k = \Pr(N = k) = \binom{m}{k} q^k (1-q)^{m-k} = \frac{m!}{k!(m-k)!} q^k (1-q)^{m-k},$$

  for $k = 0, \ldots, m$.
- Binomial r.v. is also the sum of independent Bernoulli's with $N = \sum_{k=1}^{m} N_k$ where each $N_k \sim$ Bernoulli(q).
- Mean $\text{E}(N) = mq$ and variance $\text{Var}(N) = mq(1-q)$

Poisson random variables

- $N \sim$ Poisson(λ) if pmf is

$$p_k = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \text{for } k = 0, 1, 2, \ldots$$

- Mean and variance are equal: $\mathsf{E}(N) = \mathsf{Var}(N) = \lambda$
- Sums of independent Poissons: If $N_1, \ldots, N_n$ be n independent Poisson variables with parameters $\lambda_1, \ldots, \lambda_n$, then the sum

$$N = N_1 + \cdots + N_n$$

has a Poisson distribution with parameter $\lambda = \lambda_1 + \cdots + \lambda_n$.

Negative binomial random variable

- N has a Negative Binomial distribution, written $N \sim \text{NB}(\beta, r)$, if its pmf can be expressed as

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

  for $k = 0, 1, 2, \ldots$ where $r > 0, \beta > 0$.

- Mean: $\text{E}(N) = r\beta$
- Variance: $\text{Var}(N) = r\beta(1+\beta)$.
- Clearly, since $\beta > 0$, the variance of the NB exceeds the mean.

Geometric random variable

- The Geometric distribution is a special case of the Negative Binomial with $r = 1$.
- N is said to be a Geometric r.v. and written as $N \sim \text{Geometric}(p)$ if its pmf is therefore expressed as

$$p_k = \Pr(N = k) = \frac{1}{1+\beta}\left(\frac{\beta}{1+\beta}\right)^k, \quad \text{for } k = 0, 1, 2, \ldots.$$

- Mean is $\mathrm{E}(N) = \beta$ and variance is $\mathrm{Var}(N) = \beta(1+\beta)$.

Special class of distributions

- The $(a, b, 0)$ class of distributions satisfies the recursion equations of the general form:

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad \text{for} \quad k = 1, 2, \ldots.$$

- The three distributions (including Geometric as special case of NB) are the only distributions that belong to this class: Binomial, Poisson, and Negative Binomial.

- It can be shown that the applicable parameters a and b are:

Distribution	Values of a and b
Binomial(m, q)	$a = -\frac{q}{1-q}$, $b = (m+1)\frac{q}{1-q}$
Poisson(λ)	$a = 0$, $b = \lambda$
NB(β, r)	$a = \frac{\beta}{1+\beta}$, $b = (r-1)\frac{\beta}{1+\beta}$

Example

Suppose N is a counting distribution satisfying the recursive probabilities:

$$\frac{p_k}{p_{k-1}} = \frac{4}{k} - \frac{1}{3},$$

for $k = 1, 2, \ldots$

Identify the distribution of N.

Solution to the Recursive Probability Example

- Given recursive probabilities:

$$\frac{p_k}{p_{k-1}} = \frac{4}{k} - \frac{1}{3},$$

  for $k = 1, 2, \ldots$

- Comparing with the general form of $(a, b, 0)$ class:

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

  we have $a = -\frac{1}{3}$ and $b = 4$.

- Comparing with the known distributions, only the binomial distribution has a negative a. Set $a = -\frac{1}{3} = -\frac{q}{1-q}$ and $b = 4 = (m+1)\frac{q}{1-q}$

SOA question

The distribution of accidents for 84 randomly selected policies is as follows:

Number of Accidents	Number of Policies
0	32
1	26
2	12
3	7
4	4
5	2
6	1

Identify the frequency model that best represents these data.

Analyzing the Accident Data

- Given accident data for 84 policies:

Number of Accidents (k)	Number of Policies	Ratio $\frac{p_k}{p_{k-1}}$
0	32	-
1	26	$\frac{26}{32} \approx 0.81$
2	12	$\frac{12}{26} \approx 0.46$
3	7	$\frac{7}{12} \approx 0.58$
4	4	$\frac{4}{7} \approx 0.57$
5	2	$\frac{2}{4} = 0.50$
6	1	$\frac{1}{2} = 0.50$

- Analyzing the ratio $\frac{p_k}{p_{k-1}}$ for each k to identify a possible distribution from the $(a, b, 0)$ class.

Truncation and modification at zero

- The $(a, b, 1)$ class of distributions satisfies the recursion equations of the general form:

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad \text{for} \quad k = 2, 3, \ldots.$$

- Only difference with the $(a, b, 0)$ class is the recursion here begins at p_1 instead of p_0. The values from $k = 1$ to $k = \infty$ are the same up to a constant of proportionality. For the class to be a distribution, the remaining probability must be set for $k = 0$.
  - zero-truncated distributions: the case when $p_0 = 0$
  - zero-modified distributions: the case when $p_0 > 0$
- The distributions in the second subclass is indeed a mixture of an $(a, b, 0)$ and a degenerate distribution. A zero-modified distribution can be viewed as a zero-truncated by setting $p_0 = 0$.

Expectation and Variance

Zero-Modified and Zero-Truncated Distributions

- For a zero-truncated distribution, the expected value (mean) and variance are given by:

$$\mathsf{E}[X_{\text{trunc}}] = \frac{\mathsf{E}[X]}{1 - p_0},$$
$$\mathsf{Var}[X_{\text{trunc}}] = \frac{\mathsf{Var}[X]}{1 - p_0} - \frac{p_0 \cdot [\mathsf{E}[X]]^2}{(1 - p_0)^2},$$

  where X_{trunc} is the zero-truncated version of the random variable X.

- For a zero-modified distribution, assuming that X_{mod} is the modified variable and p_0' is the modified probability at zero:

$$\mathsf{E}[X_{\text{mod}}] = (1 - p_0') \cdot \mathsf{E}[X_{\text{trunc}}],$$
$$\mathsf{Var}[X_{\text{mod}}] = (1 - p_0') \cdot \mathsf{Var}[X_{\text{trunc}}] + p_0' \cdot (1 - p_0') \cdot [\mathsf{E}[X_{\text{trunc}}]]^2.$$

Zero-Modified and Zero-Truncated Distributions

- Zero-Modified Distributions:
  - In zero-modified distributions, the probability at zero, p_0, is artificially altered.
  - This modification changes the probabilities p_k for $k \geq 1$.
  - The adjusted probabilities for $k \geq 1$ are:

$$p'_k = \frac{(1 - p'_0) \cdot p_k}{1 - p_0} \quad \text{for } k \geq 1,$$

    where p_k are the original probabilities, p_0 is the original probability at zero, and p'_0 is the modified probability at zero.
- Zero-Truncated Distributions:
  - In a zero-truncated distribution, occurrences at zero are removed (i.e., $p'_0 = 0$).
  - The probabilities p_k for $k \geq 1$ are scaled up so that the distribution sums to 1.
  - The adjusted probabilities are:

$$p'_k = \frac{p_k}{1 - p_0} \quad \text{for } k \geq 1,$$

The Role of $C = \frac{1-p_0'}{1-p_0}$

- Define the constant:
$$C = \frac{1 - p_0'}{1 - p_0}$$
- This constant establishes a proportionality between the probabilities of the zero-modified and zero-truncated distributions:
$$p_k' = C \cdot p_k, \quad \text{for } k \geq 1.$$
- Additionally, this proportionality extends to the moments of the distribution:
$$\mathsf{E}[X_{\text{mod}}^k] = C \cdot \mathsf{E}[X^k].$$
- C plays a key role in linking the expectations and variances of zero-modified and zero-truncated distributions.

Zero-Modified Poisson Example ($p_0' > 0$)

- Consider a Poisson distribution with rate parameter $\lambda = 3$:

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

- Let the original $p_0 = e^{-\lambda} \approx 0.0498$, and modify it to $p_0' = 0.1$.
- Compute C:

$$C = \frac{1 - p_0'}{1 - p_0} = \frac{1 - 0.1}{1 - 0.0498} \approx 0.947.$$

- Adjusted probabilities:

$$p_k' = C \cdot p_k, \quad \text{for } k \geq 1.$$

- Example for $k = 1, 2$:

$$p_1' = 0.947 \cdot 0.1494 \approx 0.1415, \quad p_2' = 0.947 \cdot 0.224 \approx 0.212.$$

- Adjusted expectation:

$$\mathsf{E}[X_m od] = C \cdot \mathsf{E}[X] = 0.947 \cdot 3 \approx 2.841.$$

Variance of the Zero-Modified Poisson Distribution

- To compute the variance of the zero-modified Poisson distribution, we use:

$$\mathrm{Var}[X_{\text{mod}}] = \mathrm{E}[X_{\text{mod}}^2] - \left(\mathrm{E}[X_{\text{mod}}]\right)^2 .$$

- Start by finding $\mathrm{E}[X_{\text{mod}}^2]$:

$$\mathrm{E}[X_{\text{mod}}^2] = C \cdot \mathrm{E}[X^2],$$

- For a Poisson distribution with $\lambda = 3$:

$$\mathrm{Var}[X] = \lambda = 3, \quad \mathrm{E}[X] = \lambda = 3.$$

- Thus:

$$\mathrm{E}[X^2] = 3 + 3^2 = 12.$$

- Compute $\mathrm{E}[X_{\text{mod}}^2]$:

$$\mathrm{E}[X_{\text{mod}}^2] = C \cdot \mathrm{E}[X^2] = 0.947 \cdot 12 \approx 11.364.$$

- Finally, compute $\mathrm{Var}[X_{\text{mod}}]$:

$$\mathrm{Var}[X_{\text{mod}}] = 11.364 - (2.841)^2 \approx 11.364 - 8.070 = 3.294.$$

Zero-Truncated Poisson Example ($p'_0 = 0$)

- For a zero-truncated Poisson distribution ($p'_0 = 0$):

$$p'_k = \frac{p_k}{1 - p_0}, \quad \text{for } k \geq 1,$$
$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}.$$

- Let $\lambda = 3$, so $p_0 = e^{-\lambda} \approx 0.0498$ and $1 - p_0 \approx 0.9502$.
- Example for $k = 1, 2$:

$$p'_1 = \frac{p_1}{1 - p_0} = \frac{0.1494}{0.9502} \approx 0.1572, \quad p'_2 = \frac{p_2}{1 - p_0} = \frac{0.2240}{0.9502} \approx 0.2357.$$

- Expected value for the zero-truncated Poisson:

$$\mathsf{E}[X_{\text{trunc}}] = \frac{\mathsf{E}[X]}{1 - p_0} = \frac{3}{0.9502} \approx 3.157.$$

Illustrative example

Consider the zero-modified Geometric distribution with probabilities

$$\begin{aligned} p_0 &= \frac{1}{2} \\ p_k &= \frac{1}{6}\left(\frac{2}{3}\right)^{k-1}, \quad \text{for} \quad k = 1, 2, 3, \ldots \end{aligned}$$

Derive the mean and the variance of this distribution.