# 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), $p_k = \Pr(N = k)$, 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$(\lambda)$ 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 $N \sim \mathsf{NB}(\beta, r)$, 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: $\mathsf{E}(N) = r\beta$
- Variance: $\mathsf{Var}(N) = r\beta(1+\beta)$.
- 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$(\lambda)$ | $a = 0$, $b = \lambda$ |
| NB$(\beta, r)$ | $a = \frac{\beta}{1+\beta}$, $b = (r-1)\frac{\beta}{1+\beta}$ |

# 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 $X_{\text{trunc}}$ is the zero-truncated version of the random variable $X$.

- For a zero-modified distribution, assuming that $X_{\text{mod}}$ 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:

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

# 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.