Frequency Distributions

Fall 2017

Outline

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?

The Importance of Frequency

- Insurers pay claims in monetary units, e.g., US dollars. So, why should they care about how frequently claims occur?
- Setting the price of an insurance good can be problematic:
  - In manufacturing, the cost of a good is (relatively) known
  - In other areas of financial services, market prices are available
  - Price of an insurance good?: start with an expected cost, add "margins" to account for riskiness, expenses, and a profit/surplus allowance
- We can 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 I

- **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
- **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

Foundations

- We focus on claim counts N with support on the non-negative integers $k = 0, 1, 2, \ldots$
- The probability mass function is denoted as $\Pr(N = k) = p_k$
- We can summarize the distribution through its moments:
  - The mean, or first moment, is

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

  - More generally, the rth moment is

$$\mathrm{E}\ N^r = \mu_r' = \sum_{k=0}^{\infty} k^r p_k$$

  - It is common to use the variance, which is the second moment about the mean,

$$\mathrm{Var}\ N = \mathrm{E}\ (N - \mu)^2 = \mathrm{E}\ N^2 - \mu^2$$

- Also recall the moment generating function (mgf):

$$M(t) = \mathrm{E}\ e^{tN} = \sum_{k=0}^{\infty} e^{tk} p_k$$

Probability Generating Function

- The probability generating function (pgf) is:

$$\begin{aligned} \mathrm{P}(z) &= \mathrm{E}\, z^N = \mathrm{E}\, \exp(N \ln z) = M(\ln z) \\ &= \sum_{k=0}^{\infty} z^k p_k \end{aligned}$$

- By taking the mth derivative, we see that the pgf "generates" the probabilities:

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

- Further, the pgf can be used to generate moments:

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

and

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

Important Frequency Distributions

- The three important (in insurance) frequency distributions are:
  - Poisson
  - Negative binomial
  - 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 to us

Poisson Distribution

- This distribution has parameter λ, probability mass function

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

and pgf

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

- The expectation is $\mathrm{E}\,N = \lambda$, which is the same as the variance, $\mathrm{Var}\,N = \lambda$

Negative Binomial Distribution

- This distribution has parameters (r, β), probability mass function

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

and pgf

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

- The expectation is $\mathrm{E}\, N = r\beta$ and the variance is $\mathrm{Var}\, N = r\beta(1+\beta)$
- If r = 1, this distribution is called the geometric distribution
- As $\beta > 0$, we have $\mathrm{Var}\, N > \mathrm{E}\, N$. This distribution is said to be overdispersed (relative to the Poisson)

Binomial Distribution

- This distribution has parameters (m, q), probability mass function

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

and pgf

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

- The mean is $\mathrm{E}\, N = mq$ and the variance is $\mathrm{Var}\, N = mq(1-q)$

The $(a, b, 0)$ Class

- Recall the notation: $p_k = \Pr(N = k)$
- *Definition*. A count distribution is a member of the $(a, b, 0)$ class if the probabilities p_k satisfy

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

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

  - There are only three distributions that are members of the $(a, b, 0)$ class. They are the Poisson $(a = 0)$, binomial $(a < 0)$, and negative binomial $(a > 0)$
  - The recursive expression provides a computationally efficient way to generate probabilities

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

- *Example: Poisson Distribution*.
  - Recall $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$. Examining the ratio,

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

Thus, the Poisson is a member of the $(a, b, 0)$ class with $a = 0$, $b = \lambda$, and initial starting value $p_0 = e^{-\lambda}$

Other special cases (Please check)

- *Example: Binomial Distribution*. Member of the $(a, b, 0)$ class with $a = \frac{-q}{1-q}$, $b = \frac{(m+1)q}{1-q}$, and $p_0 = (1-q)^m$
- *Example: Negative Binomial Distribution*. Member of the $(a, b, 0)$ class with $a = \frac{\beta}{1+\beta}$, $b = \frac{(r-1)\beta}{1+\beta}$, and $p_0 = (1+\beta)^{-r}$

The $(a, b, 0)$ Class - Exercises

Exercise. A discrete probability distribution has the following properties:

$$\begin{aligned} p_k &= c\left(1+\frac{1}{k}\right)p_{k-1}, \quad k = 1, 2, 3, \ldots \\ p_0 &= 0.5 \end{aligned}$$

Determine the expected value of this discrete random variable

Exercise. A discrete probability distribution has the following properties:

$$\Pr(N = k) = \left(\frac{3k+9}{8k}\right)\Pr(N = k-1), \quad k = 1, 2, 3, \ldots$$

Determine the value of $\Pr(N = 3)$

Parameter Estimation

- The customary method of estimation is **maximum likelihood**
- To provide intuition, we outline the ideas in the context of Bernoulli distribution
  - This is a special case of the binomial distribution with $m = 1$
  - For count distributions, either there is a claim ($N = 1$) or not ($N = 0$). The probability mass function is:

$$p_k = \Pr(N = k) = \begin{cases} 1 - q & \text{if } k = 0 \\ q & \text{if } k = 1 \end{cases} .$$

- The Statistical Inference Problem
  - Now suppose that we have a collection of independent random variables. The ith variable is denoted as N_i. Further assume they have the same Bernoulli distribution with parameter q
  - In statistical inference, we assume that we observe a sample of such random variables. The observed value of the ith random variable is n_i. Assuming that the Bernoulli distribution is correct, we wish to say something about the probability parameter q

Bernoulli Likelihoods

- *Definition*. The likelihood is the observed value of the mass function
- For a single observation, the likelihood is:

$$\begin{cases} 1-q & \text{if } n_i = 0 \\ q & \text{if } n_i = 1 \end{cases} .$$

- The objective of maximum likelihood estimation (MLE) is to find the parameter values that produce the largest likelihood
  - Finding the maximum of the logarithmic function yields the same solution as finding the maximum of the corresponding function
  - Because it is generally computationally simpler, we consider the logarithmic (log-) likelihood, written as:

$$\begin{cases} \ln(1-q) & \text{if } n_i = 0 \\ \ln q & \text{if } n_i = 1 \end{cases} .$$

Bernoulli MLE I

- More compactly, the log-likelihood of a single observation is:

$$n_i \ln q + (1 - n_i) \ln(1 - q)$$

- Assuming independence, the log-likelihood of the data set is:

$$L_{Bern}(q) = \sum_i \{n_i \ln q + (1 - n_i) \ln(1 - q)\}$$

  - The (log) likelihood is viewed as a function of the parameters, with the data held fixed
  - In contrast, the joint probability mass function is viewed as a function of the realized data, with the parameters held fixed
- The method of maximum likelihood means finding the values of q that maximize the log-likelihood

Bernoulli MLEII

- We began with the Bernoulli distribution in part because the log-likelihood is easy to maximize
- Take a derivative of $L_{Bern}(q)$ to get

$$\frac{\partial}{\partial q} L_{Bern}(q) = \sum_i \left\{ n_i \frac{1}{q} - (1 - n_i) \frac{1}{1-q} \right\}$$

  and solving the equation $\frac{\partial}{\partial q} L_{Bern}(q) = 0$ yields

$$\hat{q} = \frac{\sum_i n_i}{\text{sample size}}$$

  or, in words, the MLE $\hat{q}$ is the fraction of ones in the sample
- Just to be complete, you should check, by taking derivatives, that when we solve $\frac{\partial}{\partial q} L_{Bern}(q) = 0$ we are maximizing the function $L_{Bern}(q)$, not minimizing it

Frequency Distributions MLE I

- We can readily extend this procedure to all frequency distributions
- For notation, suppose that θ ("theta") is a parameter that describes a given frequency distribution $\Pr(N = k; \theta) = p_k(\theta)$
  - In later developments we will let θ be a vector but for the moment assume it to be a scalar
- The log-likelihood of a a single observation is

$$\begin{cases} \ln p_0(\theta) & \text{if } n_i = 0 \\ \ln p_1(\theta) & \text{if } n_i = 1 \\ \vdots & \vdots \end{cases} .$$

that can be written more compactly as

$$\sum_k I(n_i = k) \ln p_k(\theta).$$

This uses the notation $I(\cdot)$ to be the indicator of a set (it returns one if the event is true and 0 otherwise)

Frequency Distributions MLE II

- Assuming independence, the log-likelihood of the data set is

$$L(q) = \sum_i \left\{ \sum_k I(n_i = k) \ln p_k(\theta) \right\} = \left\{ \sum_k m_k \ln p_k(\theta) \right\}$$

  where we use the notation m_k to denote the number of observations that are observed having count k

  Using notation: $m_k = \sum_i I(n_i = k)$

- **Special Case**. *Poisson*. A simple exercise in calculus yields

$$\hat{\lambda} = \frac{\sum_k k m_k}{\text{sample size}}$$

  the average claim count

Other Frequency Distributions

- Naturally, there are many other count distributions needed in practice
- For many insurance applications, one can work with one of our three basic distributions (binomial, Poisson, negative binomial) and allow the parameters to be a function of known explanatory variables
  - This allows us to explain claim probabilities in terms of known (to the insurer) variables such as age, sex, geographic location, etc.
  - This field of statistical study is known as **regression analysis** - it is an important topic that we will not pursue in this course
- To extend our basic count distributions to alternatives needed in practice, we consider two approaches:
  - Zero truncation or modification
  - Mixing

Zero Truncation or Modification

- Why truncate or modify zero?
  - If we work with a database of claims, then there are no zeroes!
  - In personal lines (like auto), people may not want to report that first claim (why?)
- Let's modify zero probabilities in terms of the $(a, b, 0)$ class
- *Definition*. A count distribution is a member of the $(a, b,$ 1) class if the probabilities p_k satisfy

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

  for constants a, b and for $k = 2, 3, \ldots$
- Note that this starts at $k = 2$, not $k = 1$ (starts at p_1, not p_0)
- Thus, all distributions that are members of the $(a, b,$ 0) are members of the $(a, b,$ 1) class. Naturally, there are additional distributions that are members of this wider class

Zero Truncation or Modification

- Pick a specific distribution in the $(a, b$, 0) class:
  - Consider p_k^0 to be a probability for this member of $(a, b, 0)$
  - Let p_k^M be the corresponding probability for a member of $(a, b, 1)$, where the M stands for "modified"
  - Pick a new probability of a zero claim, p_0^M, and define

$$c = \frac{1 - p_0^M}{1 - p_0^0}.$$

  - We then calculate the rest of the modified distribution as

$$p_k^M = c p_k^0$$

- *Special Case: Poisson Truncated at Zero.* For this case, we assume that $p_0^M = 0$, so that the probability of $N = 0$ is zero, hence the name "truncated at zero"
- For this case, we use the letter T to denote probabilities instead of M, so we use p_k^T for probabilities. Thus,

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

Modified Poisson Example

Example: Zero Truncated/Modified Poisson. Consider a Poisson distribution with parameter $\lambda = 2$

We show how to calculate $p_k, k = 0, 1, 2, 3$, for the usual (unmodified), truncated, and a modified version with $p_0^M = 0.6$

Solution. For the Poisson distribution as a member of the $(a, b$,0) class, we have $a = 0$ and $b = \lambda = 2$

Thus, we may use the recursion $p_k = \lambda p_{k-1}/k = 2p_{k-1}/k$ for each type, after determining starting probabilities

k	p_k	p_k^T	p_k^M
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 $k = 0, 1, 2, \ldots$
- $\frac{p_n}{p_m} = \frac{m!}{n!}, m \geq 0, n \geq 0$

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

Mixtures of Finite Populations

- Suppose that our population consists of several subgroups, each having their own distribution
- We randomly draw an observation from the population, without knowing from which subgroup we are drawing
- For example, suppose that N_1 represents claims from "good" drivers and N_2 represents claims from "bad" drivers. We draw

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

- Here, α represents the probability of drawing a "good" driver
- "Mixture" of two subgroups

Finite Population 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

The distribution of the population and the mean number of colds are as follows:

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

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

Mixtures of Infinitely Many Populations

- We can extend the mixture idea to an infinite number of populations (subgroups)
- To illustrate, suppose we have a population of drivers. The ith person has their own (personal) Poisson distribution with expected number of claims, λ_i
- For some drivers, λ is small (good drivers), for others it is high (not so good drivers). There is a distribution of λ
- A convenient distribution for λ is a gamma distribution with parameters (α, θ)
- Then, one can check that if $N|\Lambda \sim$ Poisson(Λ) and if $\Lambda \sim$ gamma(α, θ):

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

Negative Binomial as a Gamma Mixture of Poissons

Example. Suppose that $N|\Lambda \sim$ Poisson(Λ) and that $\Lambda \sim$ gamma with mean of 1 and variance of 2. Determine the probability that $N = 1$

Solution. For a gamma distribution with parameters (α, θ), we have that mean is $\alpha\theta$ and the variance is $\alpha\theta^2$. Thus:

$$\alpha = \frac{1}{2} \text{ and } \theta = 2$$

Now, one can directly use the negative binomial approach to get $r = \alpha = \frac{1}{2}$ and $\beta = \theta = 2$. Thus:

$$\begin{aligned} \Pr(N = 1) = p_1 &= \binom{1 + r - 1}{1} (\frac{1}{(1+\beta)^r})(\frac{\beta}{1+\beta})^1 \\ &= \binom{1 + \frac{1}{2} - 1}{1} \frac{1}{(1+2)^{1/2}} (\frac{2}{1+2})^1 \\ &= \frac{1}{3^{3/2}} = 0.19245 \end{aligned}$$

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 ($n\widehat{p}_k$)
0	6,996	6,977.86
1	455	487.70
2	28	17.04
3	4	0.40
4	0	0.01
Total	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 $\widehat{\lambda} = \overline{N}$
  - Estimated probabilities, using $\widehat{\lambda}$, are denoted as $\widehat{p}_k$

Singapore Data: Adequacy of the Poisson Model

- With the Poisson distribution:
  - The maximum likelihood estimator of λ is $\widehat{\lambda} = \overline{N}$
  - Estimated probabilities, using $\widehat{\lambda}$, are denoted as $\widehat{p}_k$
- For goodness of fit, consider *Pearson's chi-square statistic*

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

  - 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	3	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