

Frequency Modeling

A short course authored by the Actuarial Community

19 Aug 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

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

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

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

- ▶ **Variance**:

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

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

Basic Problem

- ▶ Given a random sample from p on $\{0, 1, \dots\}$
 - ▶ Unknown p
 - ▶ Want to make *good* data-driven decisions
- ▶ Using a statistical model $\{p_\theta | \theta \in \Theta\}$
 - ▶ An indexed set of distributions
- ▶ Assumed to contain p
 - ▶ *Estimating* θ_0 satisfying $p = p_{\theta_0}$
 - ▶ for *optimal* decision making
- ▶ Two simplifying features
 - ▶ Observations in $\{0, 1, \dots\}$
 - ▶ Θ is a Euclidean subset

Compression of Data

- ▶ x_1, \dots, x_n - sample from p on $\{0, 1, \dots\}$
 - ▶ m_k - number of observations equal to k
 - ▶ Note that $\sum_{k \geq 0} m_k = n$
 - ▶ In particular, at most n many m_k 's are non-zero
- ▶ Justification for compressing data to m_k 's
 - ▶ Likelihood Principle
 - ▶ Sufficiency Principle

The Likelihood

- ▶ Likelihood - $L(\cdot)$
 - ▶ Function of parameter (index)
- ▶ $L(\theta) = \prod_{i=1}^n p_{\theta}(x_i)$
 - ▶ Likelihood Principle
- ▶ All information on parameter is in L
 - ▶ $L(\theta) = \prod_{i=1}^n p_{\theta}(x_i) = \prod_{k \geq 0} (p_{\theta}(k))^{m_k}$
 - ▶ Depends only on m_k 's
- ▶ Hence, reduction to m_k 's is lossless

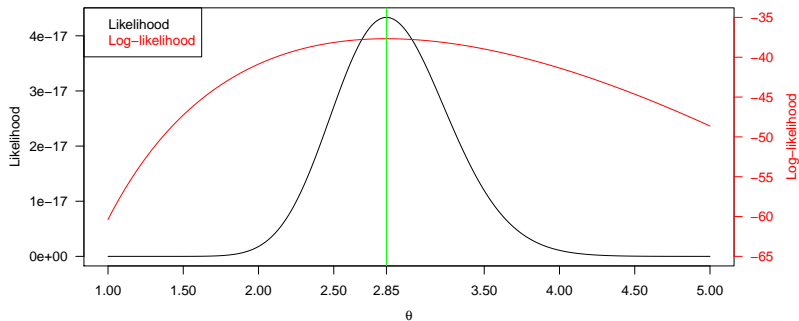
Maximum Likelihood (ML) Estimation

- ▶ A simple statistical model - $\{p_1, p_2\}$
 - ▶ $\Theta = \{1, 2\}$.
 - ▶ $p_1(3) = 0.8; p_1(5) = 0.2$
 - ▶ $p_2(3) = 0.4; p_2(5) = 0.6$
- ▶ Data - $\tilde{x} = (3, 3, 5)$
- ▶ $L(1) = 0.8^2 \cdot 0.2^1 = 0.128; L(2) = 0.4^2 \cdot 0.6^1 = 0.096$
 - ▶ Observations more likely under p_1
 - ▶ MLE (of the parameter) is 1

ML Estimation - General Setup

- ▶ In general, MLE equals $\arg \max L(\theta)$
- ▶ Log-likelihood $l(\cdot)$ given by
- ▶ $l(\theta) := \log L(\theta)$
 - ▶ As log is strictly increasing on $(0, \infty)$
 - ▶ $\arg \max L(\theta) = \arg \max l(\theta)$
 - ▶ MLE computation often involves calculus
- ▶ Convenient to work with log-likelihood $l(\cdot)$
 - ▶ Constant factor drops off upon differentiation
 - ▶ Numerically is more stable than likelihood

Plot of Likelihood and Log-likelihood - Plots



MLE for Binomial and Negative Binomial

Binomial Model

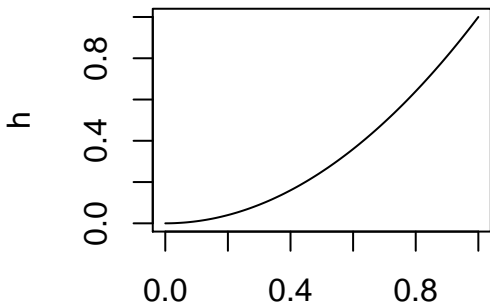
- ▶ In frequency modeling, binomial is a two parameter model
 - ▶ As m , apart from q , is taken to be unknown as well
 - ▶ Unlike as presented in introductory statistics texts
- ▶ Log-likelihood $l(m, q)$ equals
 - ▶ $\sum_{i=1}^n \log \binom{m}{x_i} + n\bar{x} \log(q) + n(m - \bar{x}) \log(1 - q)$
 - ▶ $\bar{x} = n^{-1} \sum_{i=1}^n x_i$
 - ▶ m takes only non-negative integer values
- ▶ For each fixed value of m , maximizing value of q satisfies
 - ▶ $q \cdot m = \bar{x}$
 - ▶ $\hat{m}_{MLE} = \arg \max_{m \geq \max x_i} l(m, \bar{x}/m)$
 - ▶ $\hat{q}_{MLE} = \bar{x} / \hat{m}_{MLE}$
- ▶ Caveat: $\hat{m}_{MLE} = \infty$ if sample variance is at least sample mean
 - ▶ In which case it suggests use of a Poisson model

Negative Binomial Model

- ▶ Negative Binomial model is a two parameter model
 - ▶ Parameter r is taken to be strictly positive valued
 - ▶ Unlike as presented in introductory statistics texts
- ▶ Log-likelihood $l(r, \beta)$ equals
 - ▶ $l(r, \beta) = \sum_{i=1}^n \log \binom{r+x_i-1}{x_i} - n(r + \bar{x}) \log(1 + \beta) + n\bar{x} \log \beta,$
 - ▶ $\bar{x} = n^{-1} \sum_{i=1}^n x_i$
- ▶ r is non-negative
 - ▶ For each fixed value of r , maximizing value of β satisfies
 - ▶ $\beta \cdot r = \bar{x}$
 - ▶ $\hat{r}_{MLE} = \arg \max_r l(r, \bar{x}/r); \hat{\beta}_{MLE} = \bar{x}/\hat{r}_{MLE}$
 - ▶ A good starting point for optimizer from method of moments
 - ▶ $r = \bar{x}^2 / (\text{sample variance} - \bar{x})$
- ▶ Caveat: $\hat{r}_{MLE} = \infty$ if sample variance is at most sample mean
 - ▶ In which case it suggests use of a Poisson model

Reason Behind non-existence of MLE

- ▶ The Poisson is in a sense on the boundary of both
 - ▶ The set of Binomials
 - ▶ m approaching infinity; constant mean
 - ▶ The set of Negative Binomials
 - ▶ r approaching infinity; constant mean
- ▶ But no binomial or a negative binomial is a Poisson
- ▶ So MLE for binomial and negative binomial
 - ▶ Like maximizing an increasing function on an open interval
 - ▶ $h(\cdot)$ on $(0,1)$ with $h(x) = x^2$ does not attain a maximum value



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

Why do we need more frequency distributions?

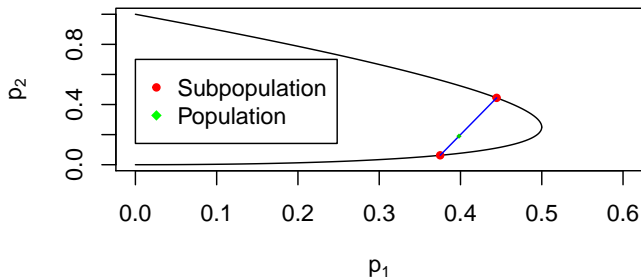
- ▶ $(a, b, 0)$ class consists of three distributions
 - ▶ Binomial, Poisson and the Negative Binomial
- ▶ It is a two parameter family
 - ▶ Hence, there are many more count distributions
- ▶ Two natural ways of extending this class
 - ▶ Both are actuarially driven
- ▶ In a mathematical sense
 - ▶ They are actuarially driven closure requirements

Heterogeneous Population

- ▶ When encountered with a heterogeneous population
 - ▶ Divide them into homogeneous subpopulations
- ▶ Actuaries commonly do so for risk rating
 - ▶ Helps in statistical modeling too
- ▶ $(a,b,0)$ often works well in modeling subpopulations
 - ▶ Even when not at the population level
- ▶ Two distinct $(a,b,0)$ subpopulations do not make a $(a,b,0)$ population

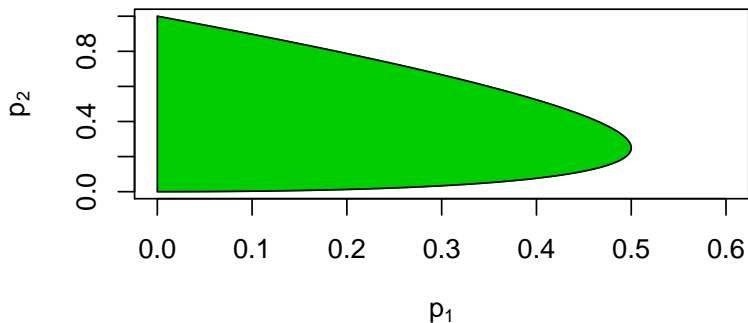
Visualizing Mixtures

- ▶ Consider a portfolio of policies
 - ▶ Two sub-portfolios - sizes in the ratio 2:1
 - ▶ Bin(2,1/4) and Bin(2,2/3) distributed
- ▶ Binomial with $m = 2$ determined by p_1 and p_2
 - ▶ As $p_0 = 1 - p_1 - p_2$



Visualizing Mixtures

- ▶ Consider a portfolio of policies
 - ▶ Two sub-portfolios - sizes in the ratio 2:1
 - ▶ Bin(2,1/4) and Bin(2,2/3) distributed
- ▶ Binomial with $m = 2$ determined by p_1 and p_2
 - ▶ As $p_0 = 1 - p_1 - p_2$
- ▶ Binomial ($m = 2$) mixtures are in the shaded region



Mixtures in General

- ▶ The above was a motivation for extension by mixing
- ▶ Mixture distribution is a convex combination of distributions
 - ▶ Finitely or infinitely many components
- ▶ Section 2.5: mechanics of working with mixtures
- ▶ We next discuss another natural extension

Zero - Modification

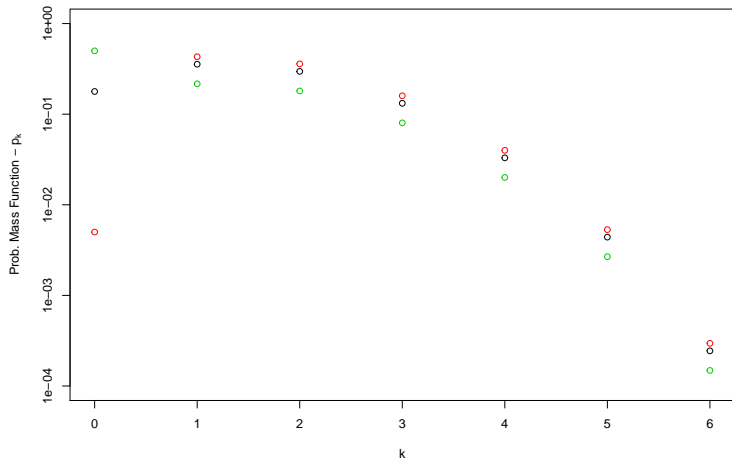
- ▶ Employer provided insurance leads to redundant coverage
 - ▶ An insured may not file claims on a policy
- ▶ Results in outsized proportion of zero claims
 - ▶ ill-fitted by standard parametric families
 - ▶ Like those in the $(a, b, 0)$ -class
- ▶ For $\{p_k\}_{k \geq 0}$ a probability mass function
 - ▶ Zero-modified version $\{q_k\}_{k \geq 0}$

$$q_k = \begin{cases} q_0, & k = 0; \\ (1 - q_0) \frac{p_k}{1 - p_0}, & k \geq 1. \end{cases}$$

- ▶ $q_0 \in [0, 1]$ - adds an extra parameter in the extension
- ▶ Zero-modified versions of $(a, b, 0)$ are practically well motivated
- ▶ Zero-truncated version is zero-modified with $q_0 = 0$
 - ▶ Conditional distribution of claims given at least one claim

Zero - Modification: Log plot of Prob. Mass Function

- ▶ Ratio p_{k+1}/p_k unchanged, for $k \geq 1$



Zero - Modification of $(a, b, 0)$ class

- ▶ $(a, b, 0)$ class defines, for $k \geq 0$, the ratio

$$p_{k+1}/p_k = (1 + b/(k + 1))$$

- ▶ Zero-modification does not alter this ratio for $k \geq 1$
- ▶ Zero-modified version of $(a, b, 0)$ distribution
 - ▶ Satisfies same recurrence but starting from $k \geq 1$
- ▶ $(a, b, 1)$ class of distributions
 - ▶ Those satisfying the previous recurrence for $k \geq 1$
 - ▶ $p_0 \geq 0$ is arbitrary, and p_1 is chosen so that $\sum_{k \geq 0} p_k = 1$
 - ▶ The parameter set equals p_0, a and b

The $(a,b,1)$ Class

- ▶ For the $(a, b, 0)$ -class the valid values for (a, b)
 - ▶ $a < 0$, and $b \in \{-a, -2a, \dots\}$ - Binomial
 - ▶ $a = 0$, and $b \geq 0$ - Poisson
 - ▶ $a > 0$, and $a + b \geq 0$ - Negative Binomial
- ▶ When recurrence relation is restricted to $k \geq 1$
 - ▶ For $a > 0$, constraint $a + b > 0$ relaxes to $a + b/2 > 0$
 - ▶ So $(a,b,1)$ -class apart from zero-modified $(a,b,0)$ -class includes
 - ▶ Extended Truncated Neg. Binomial - $a > 0$, & $-a > b > -2a$
 - ▶ Logarithmic Distribution - $a > 0$, and $b = -a$
 - ▶ And their zero-modified version

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

The Goodness of Fit Problem

- ▶ We discussed a small subset of count distributions
- ▶ None of them may be a good/useful model for the data at hand
 - ▶ *Goodness of Fit Problem* - determining if one is adequate
- ▶ Need for a method to make sound decision on the fit
 - ▶ Introducing one such is the goal of the section
- ▶ We do so via an example

Example: Singapore Automobile Data

- ▶ A 1993 portfolio from a major insurance company in Singapore
- ▶ 7,483 automobile insurance policies
- ▶ Policy level data - count variable is the number of accidents
- ▶ Maximum of 3 accidents per policy observed
- ▶ Average of 69.89 accidents per 1,000 policies ($\bar{N} = 0.06989$)

Observed Accident Counts per Policy

Count (k)	Observed (m_k)
0	6,996
1	455
2	28
3	4
<i>Total</i>	7,483

Fitting a Poisson

- ▶ With the Poisson distribution
 - ▶ The MLE of λ is $\hat{\lambda} = \bar{N}$.
 - ▶ Fitted probabilities \hat{p}_k below use $\hat{\lambda}$
 - ▶ Fitted counts are 7,483 times the fitted probabilities
- ▶ Created a cell for counts ≥ 4
 - ▶ To account for remaining fitted probability

Table. Comparison of Observed to Fitted Counts

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

Adequacy of the Poisson Model

- ▶ For goodness of fit, consider *Pearson's chi-square statistic*

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

- ▶ Has an asymptotic chi-square distribution
 - ▶ If the Poisson distribution is the correct model
- ▶ The degrees of freedom (*df*) equals
 - ▶ the number of cells minus one minus the number of estimated parameters.
- ▶ For the Singapore data
 - ▶ $df = 5 - 1 - 1 = 3$; 99-th %ile equals 11.34487
 - ▶ The Pearson's statistic equals 41.98 (> 11.34487)
 - ▶ The basic Poisson model is **inadequate**
- ▶ In the exercise below, you will fit a zero-inflated Poisson

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