Name:

M339J Probability models for actuarial applications
Spring 2023
University of Texas at Austin
In-Term Exam III
Instructor: Milica Čudina

All written work handed in by the student is considered to be
their own work, prepared without unauthorized assistance.

The University Code of Conduct

"The core values of The University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the university is expected to uphold these values through integrity, honesty, trust, fairness, and respect toward peers and community. As a student of The University of Texas at Austin, I shall abide by the core values of the University and uphold academic integrity."

"I agree that I have complied with the UT Honor Code during my completion of this exam."

Signature:

The maximal score on this exam is 100 points.

Problem 3.1. (10 points) A compound Poisson claim distribution has the parameter λ equal to 4 and individual claim amounts X distributed as follows:

$$p_X(3) = 0.4 \quad \text{and} \quad p_X(9) = 0.6.$$

What is the expected cost of an aggregate stop-loss insurance subject to a deductible of 3?

Solution: With the individual claim amounts $X_j, j \geq 1$ distributed as above, the aggregate claims are

$$S = X_1 + X_2 + \cdots + X_N$$

where $N \sim Poisson(\lambda = 4)$. We are looking for $\mathbb{E}[(S-3)_+]$. The most straightforward approach is to use the following relationship

$$\mathbb{E}[(S-3)_+] = \mathbb{E}[S] - \mathbb{E}[S \wedge 3].$$

By Wald's identity, we know that $\mathbb{E}[S] = \mathbb{E}[N]\mathbb{E}[X]$. We are given that $\mathbb{E}[N] = 4$ and we can calculate

$$\mathbb{E}[X] = 3(0.4) + 9(0.6) = 6.6.$$

So, $\mathbb{E}[S] = 6.6(4) = 26.4$. Focusing on the random variable $S \wedge 3$ and taking into account the support of the claim amounts, we conclude that

$$S \wedge 3 \sim \begin{cases} 0 & \text{if } N = 0, \\ 3 & \text{oterwise.} \end{cases}$$

Hence,

$$\mathbb{E}[S \wedge 3] = 3(1 - p_N(0)) = 3(1 - e^{-4}).$$

Pooling our findings together, we get

$$\mathbb{E}[(S-3)_+] = 26.4 - 3(1 - e^{-4}) = 23.45495.$$

Problem 3.2. (10 points) Aggregate losses, denoted by S, are modeled assuming the number of claims has a negative binomial distribution with mean 4 and variance 8. The amount of each claim is 50. Calculate $\mathbb{E}[(S-75)_+]$.

Solution: The parameters of the negative binomial are $r = 4$ and $\beta = 1$.

We use the interpolation theorem. First, we need to calculate $\mathbb{E}[(S-50)_+]$ and $\mathbb{E}[(S-100)_+]$. Let N denote the number of claims. Then, $S = 50N$. So,

$$\mathbb{E}[(S-50)_+] = \mathbb{E}[(50N-50)_+] = 50\mathbb{E}[(N-1)_+].$$

We have

$$\mathbb{E}[(N-1)_+] = \mathbb{E}[N] - \mathbb{E}[N \wedge 1] = 4 - (1 - p_N(0)) = 3 + p_N(0) = 3 + 0.0625 = 3.0625.$$

So,

$$\mathbb{E}[(S-50)_+] = 50\mathbb{E}[(N-1)_+] = 153.125.$$

Similarly,

$$\mathbb{E}[(S-100)_+] = \mathbb{E}[(50N-100)_+] = 50\mathbb{E}[(N-2)_+].$$

We have

$$\mathbb{E}[(N-2)_+] = \mathbb{E}[N] - \mathbb{E}[N \wedge 2].$$

The distribution of $N \wedge 2$ is

$$N \wedge 2 = \begin{cases} 0, & \text{with probability } p_N(0), \\ 1, & \text{with probability } p_N(1), \\ 2, & \text{with probability } 1 - p_N(0) - p_N(1). \end{cases}$$

Thus,

$$\mathbb{E}[N \wedge 2] = p_N(1) + 2(1 - p_N(0) - p_N(1)) = 0.125 + 2(0.8125) = 1.75$$

Hence,

$$\mathbb{E}[(S-100)_+] = 50\mathbb{E}[(N-2)_+] = 50(\mathbb{E}[N] - \mathbb{E}[N \wedge 2]) = 50(4 - 1.75) = 112.5.$$

By the interpolation theorem,

$$\mathbb{E}[(S-75)_+] = \frac{1}{2}\left(\mathbb{E}[(S-50)_+] + \mathbb{E}[(S-100)_+]\right) = 132.8125.$$

Problem 3.3. (10 points) Consider a discrete random variable X whose probability mass function is of the form provided in this table:

-1	0	1
$1-3p$	$2p$	p

The parameter p is unknown within the admissible set of values.
You observe the following:

$$1, -1, 0, -1, 0, 0.$$

What is the maximum likelihood estimate for the parameter p?

Solution: As usual, first we write down the likelihood function

$$L(p) = (1-3p)^2(2p)^3 p \propto p^4(1-3p)^2.$$

The log-likelihood function is, therefore, equal to

$$\ell(p) = c + 4\ln(p) + 2\ln(1-3p)$$

where c stands for a presently irrelevant constant. The derivative of the log-likelihood function is

$$\ell'(p) = \frac{4}{p} + 2(-3)\frac{1}{1-3p}.$$

We equate the above to zero and get the following condition for p:

$$\frac{4}{p} + 2(-3)\frac{1}{1-3p} = 0 \quad \Rightarrow \quad \frac{2}{p} = \frac{3}{1-3p} \quad \Rightarrow \quad 2(1-3p) = 3p \quad \Rightarrow \quad p = \frac{2}{9}.$$

Problem 3.4. (10 points) Twenty strawberry farms participated in the annual Greater Witshire Strawberry Festival (GWSF). The festival officials were keeping the following (sloppy) track of the strawberry yield in tons:

Interval of yield	Number of farms
$[0, 10)$	10
$[10, \infty)$	10

The yield of a single farm is modeled by a random variable with the following distribution function:

$$F_X(x) = 1 - e^{-\theta x^2} \qquad x > 0$$

with θ unknown. Find the maximum likelihood estimate of the parameter θ based on the above data.

Solution: This is the case of grouped data. So, the likelihood function is of the form:

$$L(\theta) = (F_X(10))^{10}(1 - F_X(10))^{10} = \left(1 - e^{-100\theta}\right)^{10}\left(e^{-100\theta}\right)^{10}.$$

Therefore, the log-likelihood function can be expressed as

$$\ell(\theta) = 10\ln(1 - e^{-100\theta}) + 10\ln(e^{-100\theta}).$$

Even though it is tempting to "cancel" the exponential and the logarithmic functions in the second term, it is actually better to use the substitution $x = e^{-100\theta}$. This is a monotone substitution, so looking for the extremum in terms of θ is equivalent to looking for the extremum in terms of x. We get

$$\tilde{\ell}(x) = 10(\ln(1 - x) + \ln(x)).$$

The derivative of this modified log-likelihood function is

$$\tilde{\ell}'(\theta) = 10\left(-\frac{1}{1-x} + \frac{1}{x}\right).$$

Equating the above to zero, we obtain the following equation in x

$$1 - x = x \quad \Rightarrow \quad x = \frac{1}{2} \quad \Rightarrow \quad e^{-100\theta} = \frac{1}{2} \quad \Rightarrow \quad 100\theta = \ln(2) \quad \Rightarrow \quad \theta = \frac{\ln(2)}{100}.$$

Problem 3.5. (20 points) The number of losses N in a particular insurance period is modeled by a distribution with the following probability mass function:

$$p_N(0) = 0.5, \quad p_N(1) = 0.25, \quad p_N(2) = 0.25.$$

The loss amounts X are modeled by a distribution with the following probability mass function:

$$p_X(10) = 0.25, \quad p_X(20) = 0.75.$$

Assume that N and X are independent random variables. There is a stop-loss insurance on the above losses with a deductible of 20. What is the net stop-loss premium?

Solution: The aggregate losses are

$$S = X_1 + X_2 + \cdots + X_N.$$

We are looking for $\mathbb{E}[(S - 20)_+]$. We know that

$$\mathbb{E}[(S - 20)_+] = \mathbb{E}[S] - \mathbb{E}[S \wedge 20]$$

By Wald's identity, we have $\mathbb{E}[S] = \mathbb{E}[N]\mathbb{E}[X]$ with

$$\mathbb{E}[N] = 1(0.25) + 2(0.25) = 0.75 \quad \text{and} \quad \mathbb{E}[X] = 10(0.25) + 20(0.75) = 17.5.$$

Hence, $\mathbb{E}[S] = 13.125$.

The support of the random variable $S \wedge 20$ is $\{0, 10, 20\}$. We have that

$$\begin{aligned}
p_{S\wedge 20}(0) &= p_S(0) = p_N(0) = 0.5,\\
p_{S\wedge 20}(10) &= p_S(10) = p_N(1)p_X(10) = 0.25(0.25) = 0.0625,\\
p_{S\wedge 20}(20) &= 1 - p_S(0) - p_S(10) = 1 - 0.5 - 0.625 = 0.4375.
\end{aligned}$$

So,

$$\mathbb{E}[S \wedge 20] = 10(0.0625) + 20(0.4375) = 9.375.$$

Finally, our answer is $\mathbb{E}[(S-20)_+] = 13.125 - 9.375 = 3.75$.

Problem 3.6. (15 points) You are modeling the distribution of a continuous random variable X using the probability density function of the following form:

$$f_X(x) = (p+1)x^p, \quad 0 < x < 1,$$

where p stands for a parameter greater than -1. You observe the following values of the random variable X:

$$0.25, 0.37, 0.88, 0.94.$$

Calculate the maximum likelihood estimate of p.

Solution: In general, for the data set $x_1, x_2, \dots, x_n$, the likelihood function will be of the form

$$L(p) = \prod_{i=1}^n f_X(x_i; p) = \prod_{i=1}^n ((p+1)x_i^p) = (p+1)^n \prod_{i=1}^n x_i^p = (p+1)^n (x_1 x_2 \dots x_n)^p.$$

Thus, the loglikelihood function is

$$l(p) = n\ln(p+1) + p\sum_{i=1}^n \ln(x_i).$$

Differentiating the loglikelihood function, we obtain

$$l'(p) = \frac{n}{p+1} + \sum_{i=1}^n \ln(x_i).$$

Equating the above to zero, and solving for p, we get

$$\hat{p}_{MLE} = -\frac{n}{\sum_{i=1}^n \ln(x_i)} - 1.$$

In the present problem, we have

$$\hat{p}_{MLE} = -\frac{4}{\ln(0.25) + \ln(0.37) + \ln(0.88) + \ln(0.94)} - 1 = 0.5562656.$$

Problem 3.7. (15 points) A fleet of vehicles is insured. According to your model, aggregate losses have a compound Poisson distribution with the parameter λ equal to 30. Loss amounts, regardless of the style of car, follow an exponential distribution with mean 5000.

The insurer seeks to reduce the cost of insurance so they decide to do two things:

- They stop insuring convertibles which make up one fifth of the total number of cabins.
- They impose an ordinary deductible of 800 per loss.

What are the expected value and the variance of the total aggregate amount paid by the insurer after the above modifications are implemented?

Solution: The old loss frequency is $N_L \sim Poisson(\lambda_L = 30)$. So, the new loss frequency can be modeled as $\tilde{N}_L \sim Poisson(\tilde{\lambda}_L = 30(4/5) = 24)$.

The loss distribution is given to be $X \sim Exponential(\theta = 5000)$. After the ordinary deductible is imposed, the per loss random variable has the following distribution

$$Y^L \sim \begin{cases} 0 & \text{with probability } F_X(800) \\ Y^P & \text{with probability } S_X(800) \end{cases}$$

with $Y^P \sim Exponential(\theta = 5000)$.

The aggregate claims $S = Y_1^L + Y_2^L + \cdots + Y_{\tilde{N}_L}^L$ have the mean

$$\mathbb{E}[S] = \mathbb{E}[Y^L]\mathbb{E}[\tilde{N}_L] = 5000e^{-800/5000}(24) = 102257.30.$$

The variance of the claims is

$$Var[S] = \mathbb{E}[\tilde{N}_L]Var[Y^L] + Var[\tilde{N}_L](\mathbb{E}[Y^L])^2 = \tilde{\lambda}_L\mathbb{E}[(Y^L)^2] = (24)(2(5000)^2)e^{-800/5000} = 1022572547.$$

Problem 3.8. (10 points) A Poisson distribution is used to fit frequency data. You are given that:

- all the observations greater than or equal to 2 were discarded;
- there was a total of 12 observations equal to 0;
- there were 8 observations equal to 1.

What is the maximum likelihood estimate for the mean of this Poisson distribution?

Solution: The probability that a particular observation is less than or equal to 1 is

$$e^{-\lambda} + \lambda e^{-\lambda} = e^{-\lambda}(1 + \lambda).$$

So, the likelihood function is

$$L(\lambda) = \frac{1}{\left(e^{-\lambda}\left(1 + \lambda\right)\right)^{20}}(e^{-\lambda})^{12}(\lambda e^{-\lambda})^8 \propto \frac{\lambda^8}{(1 + \lambda)^{20}}$$

Thus, the loglikelihood is

$$l(\lambda) = 8\ln(\lambda) - 20\ln\left(1 + \lambda\right).$$

Differentiating with respect to λ, we obtain

$$l'(\lambda) = \frac{8}{\lambda} - \frac{20}{1 + \lambda}.$$

We can equate the above to zero and get

$$8(1 + \lambda) = 20\lambda \quad \Rightarrow \quad 12\lambda = 8 \quad \Rightarrow \quad \lambda = \frac{2}{3}.$$