Name:

M339J Probability models for actuarial applications
Spring 2022
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. (5 points) Let us denote the claim count r.v. by N. We are given that N is a mixture random variable such that

$$N \mid \Lambda = \lambda \sim Poisson(\lambda)$$

while Λ is Gamma distributed with both its mean and variance equal to 3. How much is $F_N(1)$?

Solution: We have shown in class that the distribution of N is negative binomial. Let us find its parameters r and β. Using the fact that $N \mid \Lambda$ is Poisson, we get

$$r\beta = \mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N \mid \Lambda]] = \mathbb{E}[\Lambda] = 3,$$
$$r\beta(1+\beta) = Var[N] = \mathbb{E}[Var[N \mid \Lambda]] + Var[\mathbb{E}[N \mid \Lambda]] = \mathbb{E}[\Lambda] + Var[\Lambda] = 6.$$

So, $\beta = 1$ and $r = 3$.

Hence, using our tables, we have

$$F_N(1) = p_N(0) + p_N(1) = 0.125 + 0.1875 = 0.3125.$$

Problem 3.2. (5 points) A study is being conducted in which the health of two independent groups of ten policyholders is being monitored over a one-year period of time. Individual participants in the study drop out before the end of the study with probability 0.2 (independently of the other participants). Calculate the probability that at least nine participants complete the study in one of the two groups, but not in both groups?

Solution: Let's label the two groups by $i = 1, 2$. Let X_i denote the number of people who complete the study in group $i = 1, 2$. Then, for $i = 1, 2$, we have that

$$X_i \sim Binomial(m = 10, q = 0.8).$$

For either group, the probability that at least nine people complete the study equals

$$\mathbb{P}[X_i \geq 9] = \mathbb{P}[X_i = 9] + \mathbb{P}[X_i = 10] = \binom{10}{9}(0.8)^9(0.2) + \binom{10}{10}(0.8)^{10} = 0.3758096.$$

The probability that this happens in **exactly** one group is

$$2(0.3758096)(1 - 0.3758096) = 0.4691535. \tag{3.1}$$

Problem 3.3. (5 points) The numer of take-out orders at Tarka in a particular lunch hour is modeled as Poisson with mean 20. Some of these orders contain *mango lassi* and the others do not. The probability that a randomly chosen order includes *mango lassi* is $1/4$. The number of orders in independent from *mango lassi* orders.

Given that there was a total of 16 orders during a particular lunch hour, what's the probability that exactly half of them included *mango lassi*?

Solution: Let N_1 denote the r.v. which stands for the number of orders including *mango lassi*, and let N_2 be the number of orders not including *mango lassi*. According to our extention of the Thinning Theorem, we have the following conditional distribution

$$N_1 \mid N = 16 \sim Binomial(m = 16, q = 1/4).$$

So, we are now ready to calculate the conditional probability

$$\mathbb{P}[N_1 = 8 \mid N = 16] = \binom{16}{8}\left(\frac{1}{4}\right)^8\left(\frac{3}{4}\right)^8 = \binom{16}{8}\frac{3^8}{4^{16}} = 12870 \cdot \frac{3^8}{4^{16}} = 0.019666$$

Problem 3.4. (5 points) Let X have support $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. You are given that its probability (mass) function $\{p_k; k = 0, 1, \dots\}$ satisfies the following recursion:

$$p_k = \frac{5}{k}\, p_{k-1} \qquad k = 1, 2, \dots$$

How much is p_3?

Solution: This particular member of the $(a, b, 0)$ class is Poisson($\lambda = 5$). So,

$$p_3 = e^{-5}\,\frac{5^3}{3!} \approx 0.14037.$$

Problem 3.5. (5 points) Aggregate losses are modeled as follows:

(i) The number of losses has a Poisson distribution with mean 3.
(ii) The amount of each loss has a Burr distribution with parameters $\alpha = 3, \theta = 2$, and $\gamma = 1$.
(iii) The number of losses and the amounts of the losses are mutually independent.

Calculate the variance of aggregate losses.

Solution: In our usual notation,

$$Var[S] = \mathbb{E}[N]Var[X] + Var[N](\mathbb{E}[X])^2.$$

Since the frequency is $Poisson(\lambda)$, we have

$$Var[S] = \lambda Var[X] + \lambda(\mathbb{E}[X])^2 = \lambda\mathbb{E}[X^2].$$

Using the STAM tables, we obtain

$$\mathbb{E}[X^2] = \frac{\theta^2\Gamma(1+\frac{2}{\gamma})\Gamma(\alpha-\frac{2}{\gamma})}{\Gamma(\alpha)} = \frac{2^2\Gamma(1+2)\Gamma(3-2)}{\Gamma(3)} = \frac{2^2\Gamma(3)\Gamma(1)}{\Gamma(3)} = 4.$$

Finally,

$$Var[S] = 3 \times 4 = 12.$$

Problem 3.6. (10 points) Computer maintenance costs for a department are modeled as follows:

- The distribution of the number of maintenance calls **each machine** will need in a year is Poisson with mean 3.
- The cost for a maintenance call has mean 80 and standard deviation 200.
- The number of maintenance calls and the costs of the maintenance calls are all mutually independent.

The department must buy a maintenance contract to cover repairs if there is at least a 10% probability that aggregate maintenance costs in a given year will exceed 120% of the expected costs. Using the normal approximation for the distribution of the aggregate maintenance costs, calculate the minimum number of computers needed to avoid purchasing a maintenance contract (rounding to the nearest integer divisible by 5).

Solution: Let n denote the required number of computers. Let X be the severity random variable, i.e., the cost of *any* maintenance call. As usual, S denotes the aggregate cost. We are looking for the minimal n such that

$$0.1 \geq \mathbb{P}[S > 1.2\mu_S]$$

with $\mu_S = \mathbb{E}[S]$. The total number of maintenance calls with be Poisson with mean $n\lambda$. Hence, S is compound Poisson. So, we have that

$$\mathbb{E}[S] = n\lambda\mathbb{E}[X] = n(3)(80) = 240n,$$
$$\mathrm{Var}[S] = n\lambda\mathbb{E}[X^2] = n(3)(200^2 + 80^2) = 139200n.$$

Therefore, with $\sigma_S = SD[S]$, we gave that

$$0.1 \geq \mathbb{P}[S > 1.2\mu_S] = \mathbb{P}\left[\frac{S-\mu_S}{\sigma_S} > \frac{1.2\mu_S - \mu_S}{\sqrt{139200n}} = \frac{0.2(240n))}{\sqrt{139200n}} = 0.1286535\sqrt{n}\right]$$

The above is equivalent to saying

$$0.9 \leq \Phi\left(0.1286535\sqrt{n}\right) \quad \Leftrightarrow \quad 1.28 \leq 0.1286535\sqrt{n} \quad \Leftrightarrow \quad \left(\frac{1.28}{0.1286535}\right)^2 \leq n.$$

We get $n \geq 98.98667$.

Problem 3.7. (10 pts) We are using the aggregate loss model and our usual notation. The frequency random variable N is assumed to be Poisson distributed with mean equal to 2. The severity random variable is assumed to have the following probability mass function:

$$p_X(100) = 3/5, \quad p_X(200) = 3/10, \quad p_X(300) = 1/10.$$

Find the probability that the total aggregate loss **exactly** equals 300.

Solution: If we focus on the event that $\{S = 300\}$, we know that the number of losses must be $1, 2$ or 3.

$$\begin{aligned}
\mathbb{P}[S = 300] &= \mathbb{P}[S = 300 \,|\, N = 1]\mathbb{P}[N = 1] + \mathbb{P}[S = 300 \,|\, N = 2]\mathbb{P}[N = 2] + \mathbb{P}[S = 300 \,|\, N = 3]\mathbb{P}[N = 3] \\
&= \mathbb{P}[X_1 = 300 \,|\, N = 1]\mathbb{P}[N = 1] + \mathbb{P}[X_1 + X_2 = 300 \,|\, N = 2]\mathbb{P}[N = 2] \\
&\quad + \mathbb{P}[X_1 + X_2 + X_3 = 300 \,|\, N = 3]\mathbb{P}[N = 3] \\
&= p_X(300)p_N(1) + 2p_X(100)p_X(200)p_N(2) + (p_X(100))^3 p_N(3) \\
&= \frac{1}{10}(2)e^{-2} + 2\left(\frac{3}{5}\right)\left(\frac{3}{10}\right)e^{-2}\left(\frac{2^2}{2}\right) + \left(\frac{3}{5}\right)^3 e^{-2}\left(\frac{2^3}{6}\right) = 1.208e^{-2} = 0.163485.
\end{aligned}$$

Problem 3.8. (5 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 3.9. (5 points) For a stop-loss insurance on a three person group, you are given that:

- Loss amounts are independent.
- The distribution of loss amount for each person has the following probability mass function:

$$p_X(0) = 0.4, \quad p_X(100) = 0.3, \quad p_X(200) = 0.2, \quad p_X(300) = 0.1$$

Calculate the probability that the aggregate loss is exactly 300.

Solution:

$$\begin{aligned} \mathbb{P}[S = 300] &= (p_X(100))^3 + 6p_X(0)p_X(100)p_X(200) + 3(p_X(0))^2 p_X(300) \\ &= (0.3)^3 + 6(0.4)(0.3)(0.2) + 3(0.4)^2(0.1) = 0.219. \end{aligned}$$

Problem 3.10. (10 points) The number of claims in a particular time-period, denoted by N, has a geometric distribution with mean 1. The amount of each claim X is uniform on $\{1, 2, 3, 4, 5\}$. The number of claims and the claim amount are independent. Let S be the aggregate claim amount in the period. Calculate $F_S(2)$.

Solution:
Method I. The severity distribution X has the probability mass function given by

$$p_X(x) = \frac{1}{5} \quad \text{for } x \in \{1, 2, 3, 4, 5\}$$

The frequency random variable N is geometric with mean 1. Consulting the STAM tables gives us that $\beta = 1$. Also, it is a member of the $(a, b, 0)$ class with $a = \frac{1}{2}$ and $b = 0$.

Evidently,

$$f_S(0) = f_N(0) = \frac{1}{2}.$$

We can use the recursive method and we obtain the following simplifying formula (in the case of the geometric distribution and the uniform X):

$$f_S(x) = \sum_{y=1}^{x \wedge 5} \left(\frac{1}{2}\right) f_X(y) f_S(x-y) = \frac{1}{2} \sum_{y=1}^{x \wedge 5} f_X(y) f_S(x-y) = \frac{1}{2} \sum_{y=1}^{x \wedge 5} \left(\frac{1}{5}\right) f_S(x-y) = \frac{1}{10} \sum_{y=1}^{x \wedge 5} f_S(x-y).$$

for all $x \geq 0$. Thereafter, we have that

$$f_S(1) = \frac{1}{10} \times \frac{1}{2} = \frac{1}{20},$$
$$f_S(2) = \frac{1}{10}\left(f_S(0) + f_S(1)\right) = \frac{1}{10}\left(\frac{1}{2} + \frac{1}{20}\right) = \frac{11}{200}.$$

Finally,

$$F_S(2) = f_S(0) + f_S(1) + f_S(2) = \frac{1}{2} + \frac{1}{20} + \frac{11}{200} = \frac{121}{200}.$$

<u>*Method II*</u>. Since the support of S is $\{0, 1, 2, \dots\}$, we have that

$$F_S(2) = f_S(0) + f_S(1) + f_S(2).$$

Evidently, $f_S(0) = f_N(0) = \frac{1}{2}$. Looking at all the possible outcomes, we get

$$f_S(1) = p_N(1)p_X(1) = \frac{1}{4} \cdot \frac{1}{5} = \frac{1}{20},$$
$$f_S(2) = p_N(1)p_X(2) + p_N(2)(p_X(1))^2 = \frac{1}{4} \cdot \frac{1}{5} + \frac{1}{8} \cdot \frac{1}{25} = \frac{11}{200}.$$

The final answer is

$$F_S(2) = \frac{1}{2} + \frac{1}{20} + \frac{11}{200} = \frac{121}{200}.$$

Problem 3.11. (10 points) Consider the following collective risk model:
(i) The claim count random variable N is geometric with mean 4.
(ii) The severity random variable has the following probability (mass) function:

$$p_X(1) = 0.6, \; p_X(2) = 0.4.$$

(iii) As usual, individual loss random variables are mutually independent and independent of N.
Assume that an insurance covers **aggregate losses** subject to a deductible $d = 2$.
Find the expected value of aggregate payments for this insurance.

Solution: Total aggregate losses are given by

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

So, the expected value of aggregate payments for this insurance equals

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

Wald's identity gives us

$$\mathbb{E}[S] = \mathbb{E}[N]\mathbb{E}[X] = 4(0.6(1) + 0.4(2)) = 5.6.$$

On the other hand, the distribution of the random variable $S \wedge 2$ is given by

$$S \wedge 2 \sim \begin{cases} 0 & \text{if } N = 0, \\ 1 & \text{if } N = 1 \text{ and } X_1 = 1, \\ 2 & \text{otherwise.} \end{cases}$$

So,

$$\begin{aligned}\mathbb{E}[S \wedge 2] &= p_N(1)p_X(1) + 2(1 - p_N(0) - p_N(1)p_X(1)) = \frac{4}{25}(0.6) + 2\left(1 - \frac{1}{5} - \frac{4}{25}(0.6)\right)\\ &= 0.096 + 2\,(0.704) = 1.504.\end{aligned}$$

So, our answer is $\mathbb{E}[(S-2)_+] = 5.6 - 1.504 = 4.096$.

Problem 3.12. (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 3.13. (7 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 3.14. (8 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

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