ACSC/STAT 3703, Actuarial Models I

WINTER 2023

Toby Kenney

Homework Sheet 6

Model Solutions

Basic Questions

1. *Let X follow a negative binomial distribution with $r = 5.2$ and $\beta = 0.9$. What is the probability that $X = 6$?*

The probability is

$$P(X=6) = \binom{4.2+6}{6} \frac{1}{1.9^{5.2}} \left(\frac{0.9}{1.9}\right)^6 = 0.0995432029236$$

2. *The number of claims on each insurance policy over a given time period is observed as follows:*

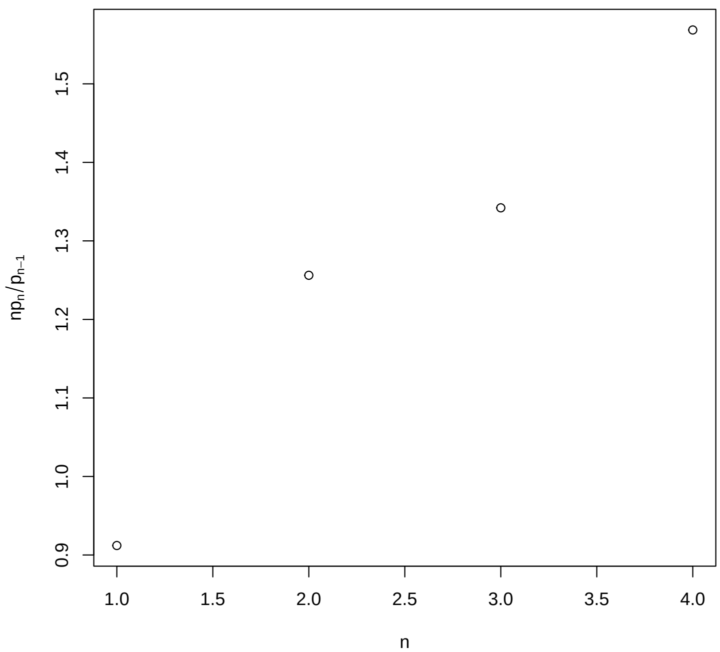
Number of claims	*Number of policies*
0	*398*
1	*363*
2	*228*
3	*118*
4	*40*
5 or more	*13*

Which distribution(s) from the $(a, b, 0)$-class and $(a, b, 1)$-class appear most appropriate for modelling this data?

We estimate

n	$\frac{p_n}{p_{n-1}}$
1	$\frac{363}{398} = 0.912060301508$
2	$\frac{228}{363} = 0.628099173554$
3	$\frac{102}{228} = 0.447368421053$
4	$\frac{40}{102} = 0.392156862745$

We can plot a graph of $n\frac{p_n}{p_{n-1}}$ against n. For a distribution from the $(a, b, 0)$ class, this should be linear with slope a and intercept b. For a distribution from the $(a, b, 1)$ class, all points for $n \neq 1$ should be linear.



On this graph, the slope is positive, so $a > 0$. The points appear to all be approximately linear, so there is not strong evidence for zero modification. This suggests a negative binomial distribution.

3. *X follows an extended modified negative binomial distribution with $r = -0.5$ and $\beta = 1.2$, and $p_0 = 0.3$. What is $P(X = 5)$?*

For the truncated ETNB with $r = -0.5$ and $\beta = 1.2$, we have $p_1 = \frac{r\beta}{(1+\beta)((1+\beta)^r-1)} = \frac{-0.6}{2.2(2.2^{-0.5}-1)} = 0.837099931231$ We also have $a = \frac{\beta}{1+\beta} = \frac{1.2}{2.2} = 0.545454545454545$ and $b = (r-1)a = -1.5 \times 0.545454545454545 = -0.818181818181818181818181818$. This gives us

$$p_2 = \left(\frac{6}{11} - \frac{9}{22}\right) p_1 = 0.114149990622$$
$$p_3 = \left(\frac{6}{11} - \frac{3}{11}\right) p_2 = 0.0311318156242$$
$$p_4 = \left(\frac{6}{11} - \frac{9}{44}\right) p_2 = 0.0106131189628$$
$$p_5 = \left(\frac{6}{11} - \frac{9}{55}\right) p_2 = 0.0040522817858$$

Now for the distribution with $p_0 = 0.3$, we have $P(X = 5) = 0.0040522817858 \times 0.7 = 0.00283659725006$.

4. *Let X follow a mixed negative binomial distribution with $\beta = 2.6$ and r following a gamma distribution with $\alpha = 4$ and $\theta = 3$. What is the probability that $X = 2$?*

   For a given value of r, this probability is

$$\binom{r+1}{2}\left(\frac{2.6}{3.6}\right)^2\left(\frac{1}{3.6}\right)^r = 0.260802469136r(r+1)(3.6)^{-r}$$

   The density of the gamma distribution is $f(r) = \frac{r^3 e^{-\frac{r}{3}}}{81\times 6}$. Thus the marginal probability that $X = 2$ is

$$\begin{aligned}
P(X=2) &= \int_0^\infty \frac{r^3 e^{-\frac{r}{3}}}{81\times 6} 0.260802469136r(r+1)(3.6)^{-r} \\
&= 0.260802469136 \int_0^\infty \frac{(r^5+r^4)e^{-r\left(\frac{1}{3}+\log(3.6)\right)}}{81\times 6}\,dr \\
&= 0.260802469136 \left(\frac{20}{81\left(\frac{1}{3}+\log(3.6)\right)^6} \int_0^\infty \left(\frac{1}{3}+\log(3.6)\right)^6 \frac{r^5 e^{-r\left(\frac{1}{3}+\log(3.6)\right)}}{120}\,dr \right. \\
&\qquad \left. + \frac{4}{81\left(\frac{1}{3}+\log(3.6)\right)^5} \int_0^\infty \left(\frac{1}{3}+\log(3.6)\right)^5 \frac{r^4 e^{-r\left(\frac{1}{3}+\log(3.6)\right)}}{24}\,dr \right) \\
&= 0.260802469136 \left(\frac{20}{81\left(\frac{1}{3}+\log(3.6)\right)^6} + \frac{4}{81\left(\frac{1}{3}+\log(3.6)\right)^5} \right) \\
&= 0.00481410883489
\end{aligned}$$

Standard Questions

5. *An insurance company finds that claim frequency for an individual has mean 0.23 and variance 0.48. They consider modelling this using either a negative binomial distribution or a zero-inflated Poisson distribution. Which of these has a higher probability that the number of claims is 3 or more?*

   For the negative binomial distribution, the mean is $r\beta$ and the variance is $r\beta(1+\beta)$, so to match the values given, we must have $1+\beta = \frac{0.48}{0.23} = 2.08695652174$. Now to match the mean, we have $1.08695652174r = 0.23$, so $r = \frac{0.23}{1.08695652174} = 0.2116$. The zero-inflated Poisson can be viewed as a mixture of a Poisson distribution, and a distribution that is always zero.

Let p_z be the probability of the zero distribution (so $p_0 = p_z + (1-p_z)e^{-\lambda}$). The mean is $(1-p_z)\lambda$ and the variance is $(1-p_z)\lambda + p_z(1-p_z)\lambda^2$. Matching these to the given values gives

$$\begin{aligned}
(1-p_z)\lambda &= 0.23 \\
(1-p_z)\lambda + p_z(1-p_z)\lambda^2 &= 0.48 \\
p_z(1-p_z)\lambda^2 &= 0.25 \\
(1-p_z)^2\lambda^2 &= 0.0529 \\
(1-p_z)\lambda^2 &= 0.3029 \\
\lambda &= \frac{0.3029}{0.23} = 1.31695652174 \\
p_z &= 1 - \frac{0.23}{1.31695652174} = 0.825354902608
\end{aligned}$$

The probabilities of 0, 1 or 2 claims under the two models are

	Negative Binomial	Zero-inflated Poisson
$P(X=0)$	$2.08696^{-0.2116} = 0.855836814896$	$0.82535 + 0.17465^{-1.31695652174} = 0.87215097977$
$P(X=1)$	$2.08696^{-0.2116} \times 0.2116 \times \frac{1.08695652174}{2.08695652174} = 0.0943203489747$	$0.17465 \times 1.31695652174e^{-1.31695652174} = 0.0616283990096$
$P(X=2)$	$2.08696^{-0.2116} \times \frac{0.2116 \times 1.2116}{2} \times \left(\frac{1.08696}{2.08696}\right)^2 = 0.029760035109$	$0.17465 \times \frac{1.31695652174^2}{2} e^{-1.31695652174} = 0.040580961$
Total	0.97991719898	0.97436033978

so the zero-inflated Poisson distribution has a higher probability of more than 2 claims.

6. *If the distribution of X is from the $(a,b,1)$-class and $P(X=2)=0.04$ and $P(X=4)=0.09$, what is the largest possible value of $P(X=3)$?*

We have
$$\left(a+\frac{b}{3}\right)\left(a+\frac{b}{4}\right) = \frac{p_4}{p_2} = 2.25$$
and we want to maximise $p_3 = 0.04\left(a+\frac{b}{3}\right)$. Letting $x = a+\frac{b}{3}$, we have $x\left(x-\frac{b}{12}\right) = 2.25$ which gives $x = \frac{1}{2}\left(\frac{b}{12}+\sqrt{\frac{b^2}{144}+9}\right)$. We see that x is an increasing function of b, so it is maximised by making b as large as possible. If $a>0$, then $x < \frac{4}{3}\left(x-\frac{b}{12}\right) = \frac{4}{3}\frac{2.25}{x}$, so to achieve a larger x, we need $a<0$, which means X follows a zero-modified binomial. In this case, $b=(n+1)a$ for some positive integer n, and since $p_4>0$, we have $n \geqslant 4$. If we substitute $a = -\frac{b}{n+1}$, then we have $b^2\left(\frac{1}{3}-\frac{1}{n+1}\right)\left(\frac{1}{4}-\frac{1}{n+1}\right) = 2.25$, so b is a decreasing function of n. Therefore, b is maximised when $n=4$. In this case $a^2\left(\frac{5}{3}-1\right)\left(\frac{5}{4}-1\right) = 2.25$, so $a^2 = 13.5$, so $a = -\sqrt{13.5} = -3.67423461418$. This gives $p_3 = 3.67423461418\left(\frac{5}{3}-1\right)0.04 = 0.0979795897116$.

7. *(a) Substituting the recurrence* $p_n = \left(a + \frac{b}{n}\right) p_{n-1}$ *for* $n \geqslant 2$ *into the PGF* $P(z) = \sum_{n=0}^{\infty} p_n z^n$ *and its derivatives, write down a differential equation satisfied by* $P(z)$*.*

Differentiating, we have $P'(z) = \sum_{n=1}^{\infty} n p_n z^{n-1}$. Substituting $p_n = \left(a + \frac{b}{n}\right) p_{n-1}$ for $n \geqslant 2$ gives

$$\begin{aligned} P'(z) &= p_1 + \sum_{n=2}^{\infty} n\left(a + \frac{b}{n}\right) p_{n-1} z^{n-1} \\ &= p_1 + az \sum_{n=2}^{\infty} (n-1) p_{n-1} z^{n-2} + (a+b) \sum_{n=2}^{\infty} p_{n-1} z^{n-1} \\ &= p_1 + azP'(z) + (a+b)(P(z) - p_0) \\ (1-az)P'(z) &= (a+b)P(z) - (p_1 - (a+b)p_0) \end{aligned}$$

(b) Show that the PGF of a distribution from the $(a, b, 1)$ *class is*

$$P(z) = \frac{(1-p_0)\left(\frac{1-az}{1-a}\right)^{-\frac{a+b}{a}} + p_0 - (1-a)^{\frac{a+b}{a}}}{1 - (1-a)^{\frac{a+b}{a}}}$$

Writing $c = (p_1 - (a+b)p_0)$, the differential equation satisfied by $P(z)$ is

$$(1-az)P'(z) = (a+b)P(z) - c$$

and we want to show that

$$P(z) = \frac{(1-p_0)\left(\frac{1-az}{1-a}\right)^{-\frac{a+b}{a}} + p_0 - (1-a)^{\frac{a+b}{a}}}{1 - (1-a)^{\frac{a+b}{a}}}$$

is the solution. We calculate

$$P'(z) = \frac{(1-p_0)a\frac{a+b}{a}(1-az)^{-\frac{2a+b}{a}}(1-a)^{\frac{a+b}{a}}}{1 - (1-a)^{\frac{a+b}{a}}}$$

so

$$(1-az)P'(z) = \frac{(1-p_0)(a+b)\left(\frac{1-az}{1-a}\right)^{-\frac{a+b}{a}}}{1 - (1-a)^{-\frac{a+b}{a}}} = (a+b)\left(P(z) + \frac{(1-a)^{\frac{a+b}{a}} - p_0}{1 - (1-a)^{\frac{a+b}{a}}}\right) \quad (1)$$

We also have

$$p_1 = P'(0) = \frac{(1-p_0)(a+b)(1-a)^{\frac{a+b}{a}}}{1 - (1-a)^{\frac{a+b}{a}}}$$

so

$$(a+b)\frac{(1-a)^{\frac{a+b}{a}}-p_0}{1-(1-a)^{\frac{a+b}{a}}} = p_1+\frac{p_0(a+b)(1-a)^{\frac{a+b}{a}}}{1-(1-a)^{\frac{a+b}{a}}}-\frac{p_0(a+b)}{1-(1-a)^{\frac{a+b}{a}}} = p_1-(a+b)p_0$$

Substituting this into (1) shows that $P(z)$ satisfies the required equation.

Bonus Question

8. *Let X be a truncated Poisson distribution with $\lambda = 2$. Is there a non-zero discrete random variable Y independent of X such that $X+Y-1$ is from the $(a,b,1)$ family?*

   [Hint: Use the convolution formula to determine the probability mass function for $X+Y-1$, and apply the recurrence for the $(a,b,1)$ class to get a recursive formula for $P(Y=n)$. You then just need to show that this recurrence gives a probability mass function.]

$$P(X+Y-1=n) = \sum_{i=1}^{n+1} P(X=i)P(Y=n+1-i) = \frac{e^{-2}}{1-e^{-2}}\sum_{i=1}^{n+1}\frac{2^i}{i!}P(Y=n+1-i)$$

Thus

$$a+\frac{b}{n} = \frac{P(X+Y-1=n)}{P(X+Y-1=n-1)} = \frac{\sum_{i=1}^{n+1}\frac{2^i}{i!}P(Y=n+1-i)}{\sum_{i=1}^{n}\frac{2^i}{i!}P(Y=n-i)} = \frac{\sum_{i=0}^{n}\frac{2^{i+1}}{(i+1)!}P(Y=n-i)}{\sum_{i=1}^{n}\frac{2^i}{i!}P(Y=n-i)}$$

$$2P(Y=n) = \left(a+\frac{b}{n}\right)\sum_{i=1}^{n}\frac{2^i}{i!}P(Y=n-i) - \sum_{i=1}^{n}\frac{2^{i+1}}{(i+1)!}P(Y=n-i)$$

$$2P(Y=n) = \sum_{i=1}^{n}\left(a+\frac{b}{n}-\frac{2}{i+1}\right)\frac{2^i}{i!}P(Y=n-i)$$

This recurrence will give a probability distribution provided it never gives a negative value. For a distribution to be from the $(a,b,1)$ class, we must have $a<1$, so for $i=1$, and large enough n, we will have $a+\frac{b}{n}-\frac{2}{i+1}<0$. However, if $a>\frac{2}{3}$, then this will be the only negative term. In particular, for example, if $a=0.9$ and $b>0$ (so the distribution is negative binomial with $\beta=9$ and $r>1$) then $a-\frac{2}{i+1}=-0.1$ when $i=1$ and $a-\frac{2}{i+1}=\frac{7}{30}$ when $i=2$, and is larger for larger i. It therefore follows that $P(Y=n) > -0.1P(Y=n-1)+\frac{7}{30}P(Y=n-2)$. This will be positive unless $P(Y = n-1) > \frac{7}{3}P(Y = n-2)$, so we need to show that $P(Y = n) \leqslant \frac{7}{3}P(Y = n-1)$ for all n.

Now, also using the recurrence, we have

$$\begin{aligned}
P(Y = n) &= \frac{1}{2}\sum_{i=1}^{n}\left(a + \frac{b}{n} - \frac{2}{i+1}\right)\frac{2^i}{i!}P(Y = n-i) \\
&= \left(a + \frac{b}{n} - 1\right)P(Y = n-1) + \frac{1}{2}\sum_{i=2}^{n}\left(a + \frac{b}{n} - \frac{2}{i+1}\right)\frac{2^i}{i!}P(Y = n-i) \\
&= \left(a + \frac{b}{n} - 1\right)P(Y = n-1) + \frac{1}{2}\sum_{j=1}^{n}\left(a + \frac{b}{n} - \frac{2}{j+2}\right)\frac{2^{j+1}}{(j+1)!}P(Y = n-1-j) \\
&\leqslant \left(a + \frac{b}{n} - 1\right)P(Y = n-1) + \frac{1}{2}\sum_{j=1}^{n}\left(a + \frac{b}{n} - \frac{2}{j+1}\right)\frac{2^j}{j!}P(Y = n-1-j) \\
&= \left(a + \frac{b}{n}\right)P(Y = n-1)
\end{aligned}$$

So if $a + b < \frac{7}{3}$, then we have $P(Y = n) \leqslant \frac{7}{3}P(Y = n-1)$ for all n, and thus $P(Y = n) > 0$ for all n, so this does define a probability distribution. (We can rescale $P(Y = 0)$ so that the probabilities sum to 1).