ACSC/STAT 4703, Actuarial Models II

Fall 2016

Toby Kenney

Homework Sheet 8

Model Solutions

For each question that asks you to simulate a small number of samples from a distribution, use the following simulated uniform values, starting from the first, and using as many numbers as needed for the question. Go back to the first value at the start of each part question.

0.84099888	0.70862315	0.77035929	0.10697185	0.50703901	0.33571753
0.97884700	0.82590460	0.32134825	0.17141305	0.06130620	0.64046291
0.20002273	0.86746587	0.46627231	0.53668803	0.41593316	0.86867032
0.81685865	0.88962299	0.90098504	0.01358413	0.80805736	0.17980851
0.29944468	0.35325712	0.57054569	0.22218065	0.72032510	0.17402895
0.08618214	0.86043490	0.83286257	0.75478531	0.60834949	0.08403240
0.81798767	0.19805759	0.34754212	0.36734934		

Basic Questions

1. *Use the method of inversion to simulate two random samples from*

   (a) a Weibull distribution with $\tau = 3$, $\theta = 70$.

   The distribution function is $F(x) = 1 - e^{-\left(\frac{x}{70}\right)^3}$. If U is uniformly distributed on $[0,1]$, we can simulate from this Weibull distribution by setting

$$\begin{aligned}1 - e^{-\left(\frac{X}{70}\right)^3} &= U\\ e^{-\left(\frac{X}{70}\right)^3} &= 1-U\\ \left(\frac{X}{70}\right)^3 &= -\log(1-U)\\ \frac{X}{70} &= \left(-\log(1-U)\right)^{\frac{1}{3}}\\ X &= 70\left(-\log(1-U)\right)^{\frac{1}{3}}\end{aligned}$$

   Substituting $U = 0.84099888$ and $U = 0.70862315$ gives $X = 85.75899$ and $X = 75.06461$.

   (b) a Pareto distribution with $\alpha = 4$, $\theta = 1200$.

   The distribution function is $F(x) = 1 - \left(\frac{1200}{1200+x}\right)^4$. If U is uniformly distributed on $[0,1]$, we can simulate from this Weibull distribution by setting

$$
\begin{aligned}
1-\left(\frac{1200}{1200+X}\right)^4 &= U\\
\left(\frac{1200}{1200+X}\right)^4 &= 1-U\\
\frac{1200}{1200+X} &= (1-U)^{\frac{1}{4}}\\
\frac{1200+X}{1200} &= (1-U)^{-\frac{1}{4}}\\
\frac{X}{1200} &= (1-U)^{-\frac{1}{4}}-1\\
X &= 1200\left((1-U)^{-\frac{1}{4}}-1\right)
\end{aligned}
$$

Substituting $U_1 = 0.84099888$ and $U_2 = 0.70862315$ gives $X_1 = 700.33952$ and $X_2 = 433.30573$.

2. *An insurance company classifies individuals into three classes, each with a different claim frequency distribution, as shown in the following table:*

Class	*Probability*	*Frequency Distribution*	*Parameters*
1	*0.60*	*Binomial*	$n = 29, p = 0.046$
2	*0.25*	*Poisson*	$\lambda = 0.08$
3	*0.15*	*Poisson*	$\lambda = 0.18$

(a) Simulate 3 claim frequencies from 3 random individuals.

For the first individual we take the uniform random variable $U_1 = 0.84099888$. Since this is between 0.6 and 0.85, this individual is simulated from class 2, so follows a Poisson distribution with $\lambda = 0.08$. The probability of being zero is $e^{-0.08} = 0.9231163$, which is larger than $U_2 = 0.70862315$, so the simulated claim frequency is zero. Our next uniform random variable is $U_3 = 0.77035929$, which is between 0.6 and 0.85, so this is another individual from class 2, and since $U_4 = 0.10697185 < 0.9231163$, the simulated claim frequency is 0. Finally, since $U_5 = 0.50703901 < 0.6$, the third individual is simulated from Class 1, so follows a binomial distribution with $n = 22, p = 0.016$ The probability $p_0 = 0.954^{29} = 0.255213$ and $p_1 = 29 \times 0.954^{28} \times 0.046 = 0.356870$, so we have $p_0 < U_6 = 0.33571753 < p_0 + p_1$, so the simulated value is 1.

(b) Simulate 3 claim frequencies from a single individual.

For a single individual, as in part (a), we use U_1 to simulate the class, giving Class 2. Now we use U_2, U_3, U_4 to simulate claim frequencies from this Poisson distribution. Since these are all less than $p_0 = 0.9231163$, the simulated claim frequencies are all 0.

3. *A home insurance policy has three types of claim with probabilities in the table below:*

Claim Type	*Probability*
Fire	*0.07*
Theft	*0.55*
Other	*0.38*

Simulate the number of each type from a sample of 744 claims.

The number of fire claims follows a binomial distribution with $n = 744$ and $p = 0.07$. We approximate this as a normal with mean $744 \times 0.07 = 52.08$ and variance $744 \times 0.07 \times 0.93 = 48.4444$. Inverting the random variable $U_1 = 0.84099888$ gives us $\Phi^{-1}(U_1) = 0.9985716$, so the random variable is $0.9985716 \times sqrt48.4444 + 52.08 = 59.03$, so our simulated value is 59. [We get the same value if we calculate the exact binomial probabilities, using for example the `pbinom` function in `R`.]

Having simulated 59 fires, there are 685 other claims. The number of thefts is therefore binomial with $n = 685, p = \frac{0.55}{0.93}$. This has mean $685 \times \frac{55}{93} = 405.107526882$ and variance $685 \times \frac{55}{93} \times \frac{38}{93} = 165.527806683$. Using $U_2 = 0.70862315$, we have $\Phi^{-1}(U_2) = 0.5493669$ so our simulated value is $405.107526882 + 0.5493669 \times \sqrt{165.527806683} = 412.17555$, which we round to 412. [Again, using the exact probabilities gives the same answer.]

In summary, our simulated frequencies are:

Claim Type	Frequency
Fire	59
Theft	412
Other	273

4. *Use a stochastic process method to simulate 3 samples from each of the following distributions:*

(a) A binomial distribution with $n = 8$ and $p = 0.08$.

For the binomial, we have $c = -8\log(0.92)$ and $d = \log(0.92)$, so we get

	Simulation 1			Simulation 2			Simulation 3		
λ_i	U_n	T_n	$\sum_{i=1}^n T_i$	U_n	T_n	$\sum_{i=1}^n T_i$	U_n	T_n	$\sum_{i=1}^n T_i$
0.66705287	0.840999	2.756669	2.756669	0.708623	1.848636	1.848636	0.770359	2.205581	2.205581

So the simulated values are all 0.

(b) A negative binomial distribution with $r = 3$ and $\beta = 1.1$.

For the negative binomial, we have $c = 3\log(1 + 1.1)$ and $d = \log(1 + 1.1)$. This gives the following table:

	Simulation 1			Simulation 2			Simulation 3		
λ_i	U_n	T_n	$\sum_{i=1}^n T_i$	U_n	T_n	$\sum_{i=1}^n T_i$	U_n	T_n	$\sum_{i=1}^n T_i$
2.225812	0.840999	0.8261452	0.8261452	0.770359	0.6609899	0.6609899	0.825905	0.7853996	0.7853996
2.967749	0.708623	0.4155128	1.2416580	0.106972	0.0381222	0.6991121	0.321348	0.1306199	0.9160195
3.709687				0.507039	0.1906698	0.8897819	0.171413	0.0506872	0.9667067
4.451624				0.335718	0.0918873	0.9816692	0.061306	0.0142119	0.9809185
5.193561				0.978847	0.7424527	1.7241219	0.640463	0.1969627	1.1778813

So the simulated values are 1, 4 and 4.

5. *Simulate 4 samples from a normal distribution with $\mu = -3$ and $\sigma = 2$ using*

(a) A Box-Muller transformation.

We have $U_1 = 0.84099888$ and $U_2 = 0.70862315$ so $Z_1 = \sqrt{-2\log(0.84099888)}\cos(2\pi \times 0.70862315) = -0.151279116562$ and $Z_2 = \sqrt{-2\log(0.84099888)}\sin(2\pi \times 0.70862315) = -0.568721839222$

We have $U_3 = 0.77035929$ and $U_4 = 0.10697185$ so $Z_3 = \sqrt{-2\log(0.77035929)}\cos(2\pi \times 0.10697185) = 0.56524335544$ and $Z_4 = \sqrt{-2\log(0.77035929)}\sin(2\pi \times 0.10697185) = 0.44977380346$

The normal random variables are therefore

$$X_1 = 2 \times (-0.151279116562) - 3 = -3.30255823312$$
$$X_2 = 2 \times (-0.568721839222) - 3 = -4.13744367844$$
$$X_3 = 2 \times (0.56524335544) - 3 = -1.86951328912$$
$$X_4 = 2 \times (0.44977380346) - 3 = -2.10045239308$$

(b) The polar method.

We calculate

$$X_1 = 2 \times 0.84099888 - 1 = 0.68199776$$
$$X_2 = 2 \times 0.70862315 - 1 = 0.41724630$$
$$X_3 = 2 \times 0.77035929 - 1 = 0.54071858$$
$$X_4 = 2 \times 0.10697185 - 1 = -0.78605630$$

We let $W_1 = {X_1}^2 + {X_2}^2 = 0.639215419509$ and $W_2 = {X_3}^2 + {X_4}^2 = 0.910261089527$ and $Y_1 = \sqrt{\frac{-2\log(W_1)}{W_1}} = 1.183299208406$ and $Y_2 = \sqrt{\frac{-2\log(W_2)}{W_2}} = 0.454517834787$.

Then our standard normal random variables are

$$
\begin{aligned}
Z_1 &= Y_1X_1 && = 0.807007409542\\
Z_2 &= Y_1X_2 && = 0.493727216500\\
Z_3 &= Y_2X_3 && = 0.245766238211\\
Z_4 &= Y_2X_4 && = -0.357276607497
\end{aligned}
$$

The simulated normal random variables are then

$$
\begin{aligned}
0.807007409542 \times 2 - 3 &= -1.38598518092\\
0.493727216500 \times 2 - 3 &= -2.01254556700\\
0.245766238211 \times 2 - 3 &= -2.50846752358\\
-0.357276607497 \times 2 - 3 &= -3.71455321499
\end{aligned}
$$

Standard Questions

6. *An insurance company models total claim frequency as following a Poisson distribution with $\lambda = 4.2$. Claim severity is independent of frequency and follows a Weibull distribution with $\tau = 0.6$ and $\theta = \$2,100$. They calculate the mean of the aggregate claim distribution as $13,270.36 and the variance as 171523504.582. They are interested in how much would be saved by introducing a deductible of $2000 per claim and a policy limit of $10,000 per claim, with a maximum out of pocket cost of $5000. (This means that once the policyholder has paid a total of $5,000, either from deductibles or from losses exceeding the policy limit, all future deductibles and policy limits are waived). They will use a simulation to calculate the new mean aggregate claims.*

   (a) What sample size should they use so that the relative error in the estimated mean is less than 1% with probability 0.95 [You may assume that the new mean aggregate claims are close to the current mean of $13,270.36, for the purposes of determining the size of the 1% relative error.]

   Based on the current estimate, the error in the estimated mean should be less than $132.7026. For a sample of size n, the variance of the estimated value is $\frac{171523504.582}{n}$. A 95% confidence interval is within 1.96 standard deviations, so we want

$$
\begin{aligned}
1.96\sqrt{\frac{171523504.582}{n}} &\leqslant 132.7026\\
\frac{171523504.582}{n} &\leqslant \left(\frac{132.7026}{1.96}\right)^2\\
n &\geqslant \frac{171523504.582}{\left(\frac{132.7026}{1.96}\right)^2}\\
&= 37417.6855
\end{aligned}
$$

So they should simulate at least 37,418 aggregate losses.

(b) Use the random numbers at the top of this sheet to simulate the first 5 aggregate losses with the new modifications.

For the Poisson distribution with $\lambda = 4.2$, we calculate the following distribution function:

n	$F(n)$
0	0.0149955768205
1	0.0779769994665
2	0.2102379870231
3	0.3954033696024
4	0.5898270213106
5	0.7531428887455
6	0.8674639959499
7	0.9360566602726

We use this lookup table to simulate the loss frequencies:

Simulation	random variable	Frequency	Random variables for simulating losses
1	$U_1 = 0.84099888$	6	$U_2, U_3, U_4, U_5, U_6, U_7$
2	$U_8 = 0.82590460$	6	$U_9, U_{10}, U_{11}, U_{12}, U_{13}, U_{14}$
3	$U_{15} = 0.46627231$	4	$U_{16}, U_{17}, U_{18}, U_{19}$
4	$U_{20} = 0.88962299$	7	$U_{21}, U_{22}, U_{23}, U_{24}, U_{25}, U_{26}, U_{27}$
5	$U_{28} = 0.22218065$	3	U_{29}, U_{30}, U_{31}

For each simulated loss, we simulate the loss by inversion:

$$
\begin{aligned}
e^{-\left(\frac{X}{2100}\right)^{0.6}} &= 1 - U\\
\left(\frac{X}{2100}\right)^{0.6} &= -\log(1-U)\\
\frac{X}{2100} &= (-\log(1-U))^{\frac{1}{0.6}}\\
X &= 2100\left(-\log(1-U)\right)^{\frac{1}{0.6}}
\end{aligned}
$$

This gives us the following losses:

Simulation	Loss 1	Loss 2	Loss 3	Loss 4	Loss 5	Loss 6	Loss 7
Simulation 1	2977.87	3996.61	55.58	1179.19	473.34	19911.75	
Simulation 2	432.79	129.60	21.09	2180.90	172.43	6783.69	
Simulation 3	1356.53	746.75	6834.84	5072.64			
Simulation 4	8492.14	1.64	4841.03	141.51	375.31	526.07	1586.79
Simulation 5	3144.66	133.25	38.04				

After applying the deductible and policy limit to each claim, we get the following claims:

Simulation	Loss 1	Loss 2	Loss 3	Loss 4	Loss 5	Loss 6	Loss 7
Simulation 1	977.87	1996.61	0.00	0.00	0	13000.00	0
Simulation 2	0.00	0.00	0.00	180.90	0	4783.69	0
Simulation 3	0.00	0.00	4834.84	3072.64	0	0.00	0
Simulation 4	6492.14	0.00	2841.03	0.00	0	0.00	0
Simulation 5	1144.66	0.00	0.00	0.00	0	0.00	0

This gives the following aggregate losses and aggregate claims:

Simulation	Aggregate losses	Aggregate claims before out-of-pocket limit	Aggregate claims after out-of-pocket limit
Simulation 1	$28,594.34	$15,974.48	$23,594.34
Simulation 2	$ 9,720.50	$ 4,964.59	$ 4,964.59
Simulation 3	$14,010.76	$ 7,907.48	$ 9,010.76
Simulation 4	$15,964.49	$ 9,333.17	$10,964.49
Simulation 5	$ 3,315.95	$ 1,144.66	$ 1,144.66

7. *An insurance company is estimating the VaR of a new policy. They simulate 100 aggregate losses, and estimate the following VaR at various levels:*

level	*VaR*
92%	*$71,300*
93%	*$71,875*
94%	*$72,125*
95%	*$72,500*
96%	*$72,975*
97%	*$73,225*
98%	*$73,750*
99%	*$74,750*

Based on these estimates, how many aggregate losses do the need to simulate, so that the probability of an error greater than 1% in their VaR estimate at the 95% level is at most 0.01? [Hint: to obtain a confidence interval for the VaR, consider instead estimating the percentile of a given point — what is the distribution of the sample percentile of the true VaR?]

The estimate for the VaR is the 95th percentile of the sample. If the sample is of size n, then the number of points above the 95th percentile of the distribution follows a binomial distribution with parameters n and $p =$

0.05. Therefore, the sample percentile at which the true 95th percentile lies is approximately normally distributed with mean 0.95 and variance $\frac{0.05\times0.95}{n}$. A 95% confidence interval for the value at risk is therefore between the $0.95-1.96\times\sqrt{\frac{0.05\times0.95}{n}}$ and $0.95+1.96\times\sqrt{\frac{0.05\times0.95}{n}}$ percentiles of the sample. We want this confidence interval to have width 1% of the VaR. Based on the current estimates, this 1% width should be 725. We see that based on current estimates, an interval of this width would go from the 93rd percentile to the 97th percentile of the distribution. We therefore need

$$
\begin{aligned}
1.96\times\sqrt{\frac{0.05\times0.95}{n}} &= 0.02\\
\sqrt{\frac{0.05\times0.95}{n}} &= \frac{0.02}{1.96}\\
\frac{0.05\times0.95}{n} &= \left(\frac{0.02}{1.96}\right)^2\\
n &= \frac{0.05\times0.95}{\left(\frac{0.02}{1.96}\right)^2}\\
&= 456.19
\end{aligned}
$$

so they need to simulate at least 457 aggregate losses.