ACSC/STAT 4703, Actuarial Models II
Fall 2015
Toby Kenney
Homework Sheet 4
Model Solutions

Basic Questions

1. *An insurance company models number of claims an individual makes in a year as following a negative binomial distribution with $\beta = 1.4$, and R an unknown parameter with prior distribution a gamma distribution with $\alpha = 3$ and $\theta = 0.04$.*

(a) What is the probability that a random individual makes exactly 3 claims?

If $R = r$, then the conditional probability of making exactly 3 claims is $\frac{r(r+1)(r+2)}{6}\left(\frac{1}{2.4}\right)^r\left(\frac{1.4}{2.4}\right)^3 = 0.03308256\frac{r(r+1)(r+2)}{2.4^r} = 0.03308256r(r+1)(r+2)e^{-\log(2.4)r}$.

The probability is the expected value of the conditional probability, which is given by

$$\begin{aligned}
&\int_0^\infty \left(\frac{r^2}{0.04^3\Gamma(3)}e^{-\frac{r}{0.04}}\right)\left(0.03308256r(r+1)(r+2)e^{-\log(2.4)r}\right)\,dr\\
&= \frac{0.03308256}{0.04^3\Gamma(3)}\int_0^\infty e^{-25r}r^3(r+1)(r+2)e^{-\log(2.4)r}\,dr\\
&= 258.4575\int_0^\infty e^{-(25+\log(2.4))r}(r^5+3r^4+2r^3)\,dr\\
&= 258.4575\left(\int_0^\infty r^5e^{-(25+\log(2.4))r}\,dr + 3\int_0^\infty r^4e^{-(25+\log(2.4))r}\,dr + 2\int_0^\infty r^3e^{-(25+\log(2.4))r}\,dr\right)\\
&= 258.4575\left((25+\log(2.4))^{-6}\Gamma(6) + (25+\log(2.4))^{-5}\Gamma(5) + (25+\log(2.4))^{-4}\Gamma(4)\right)\\
&= 0.004097388
\end{aligned}$$

(b) The company now observes the following claim frequencies:

Number of claims	*Frequency*
0	*584*
1	*90*
2	*36*
3	*12*
4	*3*
5	*3*
6	*1*

What is the probability that $R > 0.4$?

The posterior probability density is

$$\frac{\pi(r)r^{145}(r+1)^{55}(r+2)^{19}(r+3)^7(r+4)^4(r+5)\left(\frac{1}{2.4}\right)^{729r}}{\int_0^\infty \pi(r)r^{145}(r+1)^{55}(r+2)^{19}(r+3)^7(r+4)^4(r+5)\left(\frac{1}{2.4}\right)^{729r}\,dr}$$

The probability that $R > 0.4$ is therefore

$$\frac{\int_{0.4}^\infty r^2e^{-25r}r^{145}(r+1)^{55}(r+2)^{19}(r+3)^7(r+4)^4(r+5)\left(\frac{1}{2.4}\right)^{729r}}{\int_0^\infty r^2e^{-25r}r^{145}(r+1)^{55}(r+2)^{19}(r+3)^7(r+4)^4(r+5)\left(\frac{1}{2.4}\right)^{729r}\,dr}$$

Numerically, we calculate this is 1.328473×10^{-11}

R-code:

```
y<-(40001:10000000)/100000
x<-(1:10000000)/100000
sum(y^2*exp(-25*y)*y^145*(y+1)^55*(y+2)^19*(y+3)^7*(y+4)^4*(y+5)/2.4^(729*y))/sum(x^2*exp(-
```

(c) Calculate the predictive probability that an individual makes 5 claims next year.

The probability that an individual makes 5 claims next year conditional on $R = r$ is $\frac{r(r+1)(r+2)(r+3)(r+4)}{5!}\left(\frac{1.4}{2.4}\right)^5\left(\frac{1}{2.4}\right)^r$

The predictive probability is therefore

$$\frac{\left(\frac{1.4}{2.4}\right)^5}{5!}\frac{\int_0^\infty r^2e^{-25r}r^{146}(r+1)^{56}(r+2)^{20}(r+3)^8(r+4)^5(r+5)\left(\frac{1}{2.4}\right)^{730r}\,dr}{\int_0^\infty r^2e^{-25r}r^{145}(r+1)^{55}(r+2)^{19}(r+3)^7(r+4)^4(r+5)\left(\frac{1}{2.4}\right)^{729r}\,dr}$$

We calculate this numerically:

R-code:

```
sum(x^2*exp(-25*x)*x^146*(x+1)^56*(x+2)^20*(x+3)^8*(x+4)^5*(x+5)/2.4^(729*x))/sum(x^2*exp(-
```

This gives the answer as

$$9.39272 \times \frac{\left(\frac{1.4}{2.4}\right)^5}{5!} = 0.005286815$$

2. *An insurance company models loss sizes as following a Pareto distribution with $\alpha = 3$, and finds that the posterior distribution for Θ is a Gamma distribution with $\alpha = 4$ and $\theta = 1400$. Calculate the Bayes estimate for Θ based on a loss function:*

   (a) $l(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$

   The value which minimises this expected loss function is the posterior mean of Θ, which is $4 \times 1400 = 5600$.

   (b) $l(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^4$

   The expected value of this loss function is

$$
\begin{aligned}
\int_0^\infty \pi_{\Theta|X}(\theta)(\hat{\theta}-\theta)^4\,d\theta &= \int_0^\infty \frac{\theta^3}{1400\Gamma(4)} e^{-\frac{\theta}{1400}}(\hat{\theta}-\theta)^4\,d\theta \\
&= \frac{1}{1400\Gamma(4)}\int_0^\infty (\theta^7 - 4\theta^6\hat{\theta} + 6\theta^5\hat{\theta}^2 - 4\theta^4\hat{\theta}^3 + \theta^3\hat{\theta}^4)e^{-\frac{\theta}{1400}}\,d\theta \\
&= 840\times 1400^4 - 4\times 120\times 1400^3\hat{\theta} + 6\times 20\times 1400^2\hat{\theta}^2 - 4\times 4\times 1400\hat{\theta}^3 + \hat{\theta}^4
\end{aligned}
$$

Numerically, we find that this is minimised by $\hat{\theta} = 6502.119$.

3. *An insurance company models claim amounts as following an exponential distribution with mean Θ, where the prior distribution for Θ is a Gamma distribution with $\alpha = 701$ and $\theta = 600$. They observe 700 claims, with mean claim amount \$3,742. Calculate a 95% credibility interval for Θ.*

The likelihood of the data is

$$
\Theta^{-700}e^{-\frac{2619400}{\Theta}}
$$

The prior distribution of Θ is $\pi(\theta) = \frac{\theta^{700}e^{-\frac{\theta}{600}}}{600^{701}\Gamma(701)}$. The posterior distribution is therefore given by $\pi_{\Theta|X}(\theta|X) = \frac{\theta^{700}e^{-\frac{\theta}{600}}\theta^{-700}e^{-\frac{2619400}{\theta}}}{\int_0^\infty \theta^{700}e^{-\frac{\theta}{600}}\theta^{-700}e^{-\frac{2619400}{\theta}}\,d\theta} = \frac{e^{-\frac{\theta}{600}-\frac{2619400}{\theta}}}{\int_0^\infty e^{-\frac{\theta}{600}-\frac{2619400}{\theta}}\,d\theta}$.

(a) Using an HPD interval.

The HPD interval is the interval between points θ_1 and θ_2 with equal posterior density. That is

$$
\begin{aligned}
e^{-\frac{\theta_1}{600}-\frac{2619400}{\theta_1}} &= e^{-\frac{\theta_2}{600}-\frac{2619400}{\theta_2}} \\
-\frac{\theta_1}{600} - \frac{2619400}{\theta_1} &= -\frac{\theta_2}{600} - \frac{2619400}{\theta_2} \\
{\theta_1}^2\theta_2 + 2619400\times 600\theta_2 &= {\theta_2}^2\theta_1 + 2619400\times 600\theta_1 \\
(\theta_1-\theta_2)(\theta_1\theta_2 - 2619400\times 600) &= 0
\end{aligned}
$$

Which means either $\theta_1 = \theta_2$ or $\theta_1\theta_2 = 1571640000$. We want to find the solution to $\theta_1\theta_2 = 15716400002245200$ that has a 95% confidence interval.

Substituting $z = \frac{\theta}{600} + \frac{2619400}{\theta}$, we see that z is minimised by $\theta = \sqrt{1571640000}$, and takes the value $z_0 = 2\sqrt{\frac{2619400}{600}}$.

We also see that $\theta = 300z - \sqrt{90000z^2 - 1571640000}$.

Also, we obtain $\frac{dz}{d\theta} = \frac{1}{600} - \frac{2619400}{\theta^2} = \frac{1}{600} - \frac{2619400}{\left(300z-\sqrt{90000z^2-1571640000}\right)^2} = \frac{1}{600} - \frac{2619400}{180000z^2-1571640000-600z\sqrt{90000z^2-1571640000}}$.

Letting $\theta_1 < \theta_2$ satisfy $\theta_1\theta_2 =$, we find

$$\left.\frac{dz}{d\theta}\right|_{\theta_2} - \left.\frac{dz}{d\theta}\right|_{\theta_1} = \frac{2619400}{180000z^2 - 1571640000 - 600z\sqrt{90000z^2 - 1571640000}} - \frac{2619400}{180000z^2 - 1571640000 + 600z\sqrt{90000z^2 - 1571640000}}$$

We are therefore looking to find the value t such that

$$\frac{\int_{z_0}^{t}\left(\frac{2619400}{180000z^2-1571640000-600z\sqrt{90000z^2-1571640000}} - \frac{2619400}{180000z^2-1571640000+600z\sqrt{90000z^2-1571640000}}\right)e^{-z}\,dz}{\int_{z_0}^{\infty}\left(\frac{2619400}{180000z^2-1571640000-600z\sqrt{90000z^2-1571640000}} - \frac{2619400}{180000z^2-1571640000+600z\sqrt{90000z^2-1571640000}}\right)e^{-z}\,dz} = 0.95$$

Numerically, this is solved by $t = 136.090$. The values θ_1 and θ_2 are then the solutions to $\theta = 300 \times 136.090 \pm \sqrt{90000 \times 136.090^2 - 1571640000} = [31069.75, 50584.25]$.

(b) With equal probability above and below the interval.

With equal probability above and below the interval, the interval is between the 2.5th percentile and the 97.5th percentile of the posterior distribution. That is, we find the solutions to

$$\frac{\int_0^t e^{-\frac{\theta}{600}-\frac{2619400}{\theta}}\,d\theta}{\int_0^\infty e^{-\frac{\theta}{600}-\frac{2619400}{\theta}}\,d\theta} = 0.025$$
$$\frac{\int_0^t e^{-\frac{\theta}{600}-\frac{2619400}{\theta}}\,d\theta}{\int_0^\infty e^{-\frac{\theta}{600}-\frac{2619400}{\theta}}\,d\theta} = 0.975$$

Numerically, the interval is

$$[33693.97, 47350.64]$$

4. *Calculate a conjugate prior distribution for the variance of a normal distribution with mean 0.*

The log-likelihood of the data for a normal distribution with mean 0 and variance σ^2 is

$$-\sum_{i=1}^{n}\frac{{x_i}^2}{2\sigma^2} - n\log(\sigma) = -\frac{\sum_{i=1}^{n}{x_i}^2}{2\sigma^2} - n\log(\sigma)$$

If the prior distribution is then $\pi(\theta)$, where $\theta = \sigma^2$, then the posterior distribution is proportional to $\pi(\theta)\theta^{-\frac{n}{2}}e^{-\frac{\sum_{i=1}^{n}{x_i}^2}{2\theta}}$.

We see that an inverse gamma distribution is therefore a conjugate prior.

Standard Questions

5. *An insurance company models number of claims made by an individual in a year as following a Poisson distribution and finds that the posterior distribution for Λ is a Gamma distribution with $\alpha = 4$ and $\theta = 0.02$. The company decides to use an estimate $\hat{\lambda}$ such that the probability of 3 or more claims using $\hat{\lambda}$ is the same as the probability of 3 or more claims under the predictive distribution. Find the value of this $\hat{\lambda}$.*

   The predictive distribution is a Gamma mixture of Poisson distributions. This is a negative binomial with $r = \alpha$, $\beta = \theta$. The probability of 3 or more claims is therefore $1 - \left(\frac{1}{1.02}\right)^4 \left(1 + \binom{4}{1}\left(\frac{0.02}{1.02}\right) + \binom{5}{2}\left(\frac{0.02}{1.02}\right)^2\right) = 0.0001442237$.

   For the Poisson distribution, we have that the probability of 3 or more claims is $1 - e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2}\right)$, so we need to solve

$$e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2}\right) = 1 - 0.0001442237$$

   Numerically, we find the solution to this is $\lambda = 0.0976423$.

Bonus Question

6. *An insurance company models loss amounts as following a Weibull distribution with $\tau = 3$. It uses the inverse gamma prior for the unknown parameter Θ, with parameters $\alpha = 3$ and $\theta = 6000$. This is a conjugate prior, and the posterior distribution after observing N observations $X_1, \ldots, X_n$ with $\sum_{i=1}^{n} {X_i}^\tau = t$ is inverse gamma with $\alpha = 3 + N$, and $\theta = \frac{1}{\frac{1}{6000} + t}$. Calculate the probability that the posterior probability of $\Theta > 3000$ is more than 0.05, after a sample of 10 observations.*

   For given values of N and t, the posterior probability that $\Theta > 3000$ is the probability that an inverse gamma distribution with $\alpha = 3 + N$ and $\theta = \frac{1}{\frac{1}{6000} + t}$ is more than 3000. This is the probability that a Gamma distribution with $\alpha = 3 + N$ and $\theta = \frac{1}{6000} + t$ is less than $\frac{1}{3000}$. We have that $N = 10$, so this probability is the quantile of Gamma distribution with $\alpha = 13$ corresponding to the value $\frac{\left(\frac{1}{3000}\right)}{\left(\frac{1}{6000} + t\right)} = \frac{2}{1 + 6000t}$. That is, the overall probability is

$$\begin{aligned}
&\int_0^\infty f_T(t) \int_{\frac{2}{1+6000t}}^\infty \frac{x^{12} e^{-x}}{12!}\, dx\, dt \\
&= \int_0^\infty \int_{\frac{1}{3000x} - \frac{1}{6000}}^\infty \frac{x^{12} e^{-x}}{12!} f_T(t)\, dt\, dx
\end{aligned}$$

   for each X_i, we have that ${X_i}^\tau$ follows an exponential distribution, with mean Θ. This means that T follows a gamma distribution with parameters 10 and Θ. This gives a probability of

$$\int_0^\infty f_T(t) \int_{\frac{2}{1+6000t}}^\infty \frac{x^{12}e^{-x}}{12!}\,dx\,dt$$
$$= \int_0^\infty \int_{\frac{1}{3000x}-\frac{1}{6000}}^\infty \frac{x^{12}e^{-x}}{12!}\frac{t^9 e^{-\frac{t}{\theta}}}{\theta^{10}10!}\,dt\,dx$$
$$= \int_0^\infty \int_{\frac{1}{3000x\theta}-\frac{1}{6000\theta}}^\infty \frac{x^{12}e^{-x}}{12!}\frac{u^9 e^{-u}}{10!}\,du\,dx$$
$$= \int_0^\infty \int_{\frac{1}{3000x\theta}-\frac{1}{6000\theta}}^\infty \frac{x^{12}u^9 e^{-(x+u)}}{12!\times 10!}\,du\,dx$$

Substituting $a = x + u$, this becomes

$$\int_l^\infty \frac{e^{-a}}{12!\times 10!}\int_{h(a)}^a (a-u)^{12}u^9\,du\,da$$
$$= \int_l^\infty \frac{e^{-a}}{12!\times 10!}\int_0^{k(a)} v^{12}(a-v)^9\,dv\,da$$
$$= \int_l^\infty \frac{e^{-a}}{12!\times 10!}\left(\frac{a^9k(a)^{13}}{13} - 9\frac{a^8k(a)^{14}}{14} + \cdots - \frac{k(a)^{22}}{22}\right)da$$

where $l = \frac{2}{\sqrt{3000\theta}} - \frac{1}{6000\theta}$ and $k(a) = a - \frac{(a-c)+\sqrt{(a-c)^2+4(ac-2c)}}{2} = \frac{(a+c)-\sqrt{(a+c)^2-8c}}{2}$ where $c = \frac{1}{6000\theta}$