ACSC/STAT 3703, Actuarial Models I (Further Probability with Applications to Actuarial Science)
Winter 2015
Toby Kenney
Homework Sheet 2
Model Solutions

Basic Questions

1. *Calculate the probability density function of a random variable that is 7 times a beta random variable with $\alpha = 3$ and $\beta = 2$. The density function of this beta random variable is*

$$f_X(x) = \begin{cases} x^2(1-x) & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

   The density function of $7X$ is

$$f_{7X}(x) = \frac{1}{7} f_X\left(\frac{x}{7}\right) = \begin{cases} \frac{1}{7}\left(\frac{x}{7}\right)^2\left(1-\frac{x}{7}\right) & \text{if } 0 \leqslant x \leqslant 7 \\ 0 & \text{otherwise} \end{cases}$$

2. *Calculate the distribution of X^8 when X follows a gamma distribution with $\alpha = 3$ and $\theta = 13$.*

   The density function of X is $\frac{x^2}{2\times 13^3}e^{-\frac{x}{13}}$. The density function of X^8 is $f_{X^8}(a^8) = \frac{a^2}{8a^7\times 2\times 13^3}e^{-\frac{a}{13}}$, so we have $f_{X^8}(x) = \frac{1}{16\times 13^3}x^{-\frac{5}{8}}e^{-\frac{x^{\frac{1}{8}}}{13}}$, so X^8 has a transformed gamma distribution with $\alpha = 3$, $\gamma = \frac{1}{8}$ and $\theta = 13^8$.

3. *X is a random variable with moment generating function $M_X(t) = \frac{1}{(3-t)\left(1-\frac{t}{6}\right)}$. What is the variance of the random variable e^X?*

   We know that $(e^X)^2 = e^{(2X)}$, so $\mathbb{E}((e^X)^2) = M_X(2)$, while $\mathbb{E}(e^X) = M_X(1)$, so we have $\text{Var}(e^X) = M_X(2) - (M_X(1))^2 = \frac{3}{2} - \left(\frac{6}{10}\right)^2 = 1.14$.

4. *X is a mixture of 3 distributions:*

   - With probability 0.2, X follows a gamma distribution with $\alpha = 2$ and $\theta = 2000$.
   - With probability 0.35, X follows a gamma distribution with $\alpha = 3$ and $\theta = 4000$.
   - With probability 0.45, X follows a Weibull distribution with $\theta = 2000$ and $\tau = 4$.

(a) What is the coefficient of variation of X?

The mean of a gamma distribution is $\alpha\theta$, and the mean of a Weibull distribution is $\theta\Gamma\left(1+\frac{1}{\tau}\right)$. The variances are $\alpha\theta^2$ and $\theta^2\left(\Gamma\left(1+\frac{2}{\tau}\right)-\left(\Gamma\left(1+\frac{1}{k}\right)\right)^2\right)$ respectively.

In this case, the first distribution has mean 4000 and variance 8000000; the second distribution has mean 12000 and variance 48000000; and the third distribution has mean 1812.805 and variance 258645.631975. The expected value of the mixture is therefore $0.2\times 4000+0.35\times 12000+0.45\times 1812.805 = 5815.76225$. The variance of the mixture is $0.2\times(4000^2+8000000)+0.35\times(12000^2+48000000)+0.45\times(1812.805^2+258645.631975)-5815.76225^2 = 29772117.871474937$, so the standard deviation is 5456.38, and the coefficient of variation is $\frac{5456.38}{5815.76} = 0.938$.

(b) [bonus] What is the kurtosis of X?

We have the following:

	Distribution 1	Distribution 2	Distribution 3
μ	4000	12000	1812.805
μ_2	80000000	48000000	258645.631975
μ_3	3.2×10^{10}	2.56×10^{11}	11474411.56287975
μ_4	3.84×10^{14}	1.152×10^{16}	183821938794.038572798
μ_2'	2.4×10^{7}	1.92×10^{8}	3544907.60000
μ_3'	2.56×10^{11}	5.44×10^{12}	13309852126.945560125
μ_4'	2.944×10^{15}	1.6896×10^{17}	59198070889950.1844896020

The raw moments of the mixture are therefore

$$\mu_4' = 0.2\times 2.944\times 10^{15}+0.35\times 1.6896\times 10^{17}+0.45\times 59198070889950.1844896020 = 597514391319004$$
$$\mu_3' = 0.2\times 2.56\times 10^{11}+0.35\times 5.44\times 10^{12}+0.45\times 13309852126.945560125 = 1961189433457.1255020$$

$$\mu_2' = 0.2\times 2.4\times 10^{7}+0.35\times 1.92\times 10^{8}+0.45\times 3544907.60000 = 73595208.42000000$$
$$\mu = 0.2\times 4000+0.35\times 12000+0.45\times 1812.805 = 5815.76225$$

We therefore get that the centralised moments of the mixture are

$$\mu_4 = \mu_4' - 4\mu\mu_3' + 6\mu^2\mu_2' - 3\mu^4 = 25631493270307355.9794255146$$
$$\mu_2 = \mu_2' - \mu^2 = 39772117.8714749375$$

so the kurtosis is $\frac{25631493270307355.9794255146}{39772117.8714749375^2} = 16.2037850282$.

5. *For a particular claim, the insurance company has observed the following claim sizes:*

   12.3, 16.8, 24.6, 25.2, 25.4, 25.8, 30.2, and 35.3.

Using a kernel smoothing model with a Gaussian kernel with variance 0.5, calculate the probability that the next claim size is between 22 and 26.

The kernel to be used for each observation a is $\frac{1}{\sqrt{\pi}}e^{-(x-a)^2}$. The kernel distribution is given by taking the average of the kernels from each observation. The probability of being between 22 and 26 is therefore given by

$$\begin{aligned}&\frac{1}{8}\big(\Phi((26-12.3)\sqrt{2})+\Phi((26-16.8)\sqrt{2})+\Phi((26-24.6)\sqrt{2})+\Phi((26-25.2)\sqrt{2})\\ &+\Phi((26-25.4)\sqrt{2})+\Phi((26-25.4)\sqrt{2})+\Phi((26-30.2)\sqrt{2})+\Phi((26-35.3)\sqrt{2})\\ &-\Phi((22-12.3)\sqrt{2})-\Phi((22-16.8)\sqrt{2})-\Phi((22-24.6)\sqrt{2})-\Phi((26-25.2)\sqrt{2})\\ &-\Phi((26-25.4)\sqrt{2})-\Phi((26-25.8)\sqrt{2})-\Phi((26-30.2)\sqrt{2})-\Phi(22-35.3\sqrt{2})\big)\end{aligned}$$

This is

$$\frac{1}{8}\big(\Phi(19.37)+\Phi(13.01)+\Phi(1.98)+\Phi(1.13)+\Phi(0.85)+\Phi(0.28)+\Phi(-5.94)+\Phi(-13.15)$$
$$-\Phi(13.72)-\Phi(7.35)-\Phi(-3.68)-\Phi(-4.53)-\Phi(-4.81)-\Phi(-5.37)-\Phi(-11.60)-\Phi(-18.81)\big)$$

$$\begin{aligned}&=\frac{1}{8}(1+1+0.9761+0.8708+0.8023+0.6103+0+0-1-1-0.0001-0-0-0-0-0)\\ &=\frac{3.2594}{8}=0.4074\end{aligned}$$

Standard Questions

6. *An insurance company finds that the loss experienced by an individual follows an inverse exponential distribution with θ depending on the individual. It models this θ as following a gamma distribution with $\alpha = 3$ and $\theta = 2000$. What is the distribution of the loss of a random individual.*

The density function of the conditional distribution given θ is $f(x) = \frac{\theta e^{-\frac{\theta}{x}}}{x^2}$, and the distribution of θ is $f_\theta(\theta) = \frac{\theta^2 e^{-\frac{\theta}{2000}}}{2\times 2000^3}$. The distribution of the loss for a random individual therefore has

$$f(x) = \int_0^\infty \left(\frac{\theta e^{-\frac{\theta}{x}}}{x^2}\right)\left(\frac{\theta^2 e^{-\frac{\theta}{2000}}}{2\times 2000^3}\right)d\theta = \frac{1}{2\times 2000^3 x^2}\int_0^\infty \theta^3 e^{-\left(\frac{1}{x}+\frac{1}{2000}\right)\theta}d\theta$$

If we let $a = \frac{1}{\frac{1}{x}+\frac{1}{2000}}$, then the integral becomes $\int_0^\infty \theta^3 e^{-\frac{\theta}{a}}\,d\theta$ which is $a^4\Gamma(4) = 6a^4$, so

$$f(x) = \frac{6a^4}{2\times 2000^3x^2} = \frac{3}{8\times 10^9x^2\left(\frac{1}{x}+\frac{1}{2000}\right)^4} = \frac{6000x^2}{(x+2000)^4}$$

7. *A life insurance company models the mortality of an individual as following a Gompertz law with hazard rate given by $\lambda = 0.00001ae^{0.1t}$, where a is the frailty of the individual. It models a as following a gamma distribution with $\alpha = 0.4$ and $\theta = 2$. Calculate the probability that a randomly chosen individual lives to age 100.*

The probability that an individual with frailty a lives to age 100 is $e^{-\int_0^{100} 0.00001ae^{0.1t}\,dt}$. We have $\int_0^{100} e^{-0.1t}dt = 10\left[e^{0.1t}\right]_0^{100} = 10(e^{10}-1)$, so the probability is $e^{-0.0001a(e^{10}-1)} = e^{-2.202547a}$. The probability for a random individual is the expected value of this probability — that is $\mathbb{E}(e^{-2.202547A})$, where A follows a gamma distribution with $\alpha = 0.4$ and $\theta = 2$. This is $M_A(-2.202547) = (1+2\times 2.202547)^{-0.4} = 0.509188664$.

8. *An insurance company wants to model a random variable X. It believes that for large values, it should use a Pareto distribution with $\alpha = 4$ and $\theta = 300$ to model the distribution of values above 5000. For values below 5000, it plans to use an inverse gamma distribution with $\alpha = 3$ and $\theta =$ 800. If 5% of values are above 5000, what is the probability under this model that the value of X is between 3000 and 10000?*

Given that the value is under 5000, the probability that it is above 3000 is the probability that its inverse is above $\frac{1}{3000}$ The inverse of the inverse gamma distribution follows a gamma distribution with $\alpha = 3$ and $\theta =$ 0.00125. The probability that this is below $\frac{1}{3000}$ is given by $\frac{\int_{\frac{1}{5000}}^{\frac{1}{3000}} 800^3x^2e^{-800x}dx}{\int_{\frac{1}{5000}}^{\infty} 800^3x^2e^{-800x}dx}$

Integrating by parts gives

$$\int_a^\infty x^2e^{-800x}dx = \left[-\frac{x^2e^{-800x}}{800}\right]_a^\infty + \int_a^\infty \frac{xe^{-800x}}{400}dx = \frac{a^2e^{-800a}}{800}+\frac{ae^{-800a}}{320000}+\frac{e^{-800a}}{128000000}$$

so the probability above is

$$1-\frac{\left(640000\left(\frac{1}{3000}\right)^2+1600\left(\frac{1}{3000}\right)+2\right)e^{-\frac{800}{3000}}}{\left(640000\left(\frac{1}{5000}\right)^2+1600\left(\frac{1}{5000}\right)+2\right)e^{-\frac{800}{5000}}} = 1-\frac{1.994817806}{1.998788471} = 1-0.998013464 = 0.00199$$

For the Pareto distribution, the probability of being above 5000 is $\left(1+\frac{5000}{300}\right)^{-3} =$ 0.000181358, while the probability of being above 10000 is $\left(1+\frac{10000}{300}\right)^{-3} =$ 0.000024709, so the probability of being below 10000 conditional on being above 5000 is $1-\frac{0.000024709}{0.000181358} = 0.863755666$. The total probability that X is in the interval $(3000, 10000)$ is therefore $0.95\times 0.00199+0.05\times 0.863755666 = 0.0451$.