MTH5126 - Statistics for Insurance

Worksheet 1 - Solutions

Question 1.

Derive the formula for the MGF of the standard normal distribution.

Hint: Complete the square in the exponent.

Answer:
The PDF of $N(0,1)$ is:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

So the MGF is:

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 - 2tx)} dx$$

Completing the square gives:

$$M_X(t) = e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} dx$$

But this integral is just the area under the graph of the PDF of a $N(t,1)$ distrubtion. Therefore it is equal to 1.

So the MGF of the standard normal distribution is $e^{\frac{1}{2}t^2}$

Question 2.

A random variable X follows a gamma distribution with parameters α and λ.

(i)   Derive the moment generating function (MGF) of X.

(ii)   Derive the coefficient of skewness of X.

Answer:

(i)

$$M(t) = E[e^{tx}] = \int_0^\infty e^{tx} * \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx$$

$$= \int_0^\infty \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-(\lambda-t)x} dx$$

$$= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \int_0^\infty \frac{1}{\Gamma(\alpha)} (\lambda-t)^\alpha x^{\alpha-1} e^{-(\lambda-t)x} dx$$

The integral is PDF of $\Gamma(\alpha, \lambda - t) \Rightarrow = 1$

$$\Rightarrow M(t) = \left(\frac{\lambda}{\lambda-t}\right)^\alpha = \left(1-\frac{t}{\lambda}\right)^{-\alpha}, t < \lambda.$$

(ii)

$$\text{Coefficient} = \frac{E[(x-\mu)^3]}{\left(\sqrt{\text{Var}(x)}\right)^3}, \text{ given } E[x] - \mu$$

$$E[(x-\mu)^3] = E[x^3 - 3x^2\mu + 3x\mu^2 - \mu^3]$$

$$= E[x^3] - 3\mu \, E[x^2] + 3\mu^2 \, E[x] - \mu^3$$

$$= E[x^3] - 3\mu \, E[x^2] + 2\mu^3$$

$$= E[x^3] - 3 * \frac{\alpha}{\lambda} * E[x^2] + 2(E[x])^3$$

$$M(t) = \left(1-\frac{t}{\lambda}\right)^{-\alpha} \text{ from (i)}$$

$$M'(t) = -\alpha * \left(-\frac{1}{\lambda}\right)^1 * \left(1-\frac{t}{\lambda}\right)^{-\alpha-1} = \frac{\alpha}{\lambda} * \left(1-\frac{t}{\lambda}\right)^{-\alpha-1}$$

$$M''(t) = \frac{\alpha}{\lambda} * [-(\alpha+1)] * \left(-\frac{1}{\lambda}\right)^1 * \left(1-\frac{t}{\lambda}\right)^{-\alpha-2}$$

$$= \frac{\alpha*(\alpha+1)}{\lambda^2} * \left(1-\frac{t}{\lambda}\right)^{-\alpha-2}$$

$$M'''(t) = \frac{\alpha*(\alpha+1)}{\lambda^2}*[-(\alpha+2)]*\left(-\frac{1}{\lambda}\right)^1*\left(1-\frac{t}{\lambda}\right)^{-\alpha-3}$$

$$= \frac{\alpha*(\alpha+1)*(\alpha+2)}{\lambda^3}*\left(1-\frac{t}{\lambda}\right)^{-\alpha-3}$$

$$E[x^3] = M'''(0) = \frac{\alpha*(\alpha+1)*(\alpha+2)}{\lambda^3}$$

$$E[x^2] = M''(0) = \frac{\alpha(\alpha+1)}{\lambda^2}$$

So

$$E[(x-\mu)^3] = \frac{\alpha*(\alpha+1)*(\alpha+2)}{\lambda^3} - \frac{3\alpha}{\lambda}*\frac{\alpha(\alpha+1)}{\lambda^2} + 2*\left(\frac{\alpha}{\lambda}\right)^3$$

$$= \frac{1}{\lambda^3}\left(\alpha^3+3\alpha^2+2\alpha-3\alpha^3-3\alpha^2+2\alpha^3\right) = \frac{2\alpha}{\lambda^3}$$

$$\text{Coefficient} = \frac{\dfrac{2\alpha}{\lambda^3}}{\left(\sqrt{\dfrac{\alpha}{\lambda^2}}\right)^3} = \frac{2\alpha}{\alpha^{3/2}} = \frac{2}{\sqrt{\alpha}}$$

Question 3. MLE, Method of moments, Method of percentiles

A random sample $x_1, x_2, \ldots, x_{20}$ is taken from a distribution having the density function:

$$f(x) = \frac{k}{5} x^{-\frac{4}{5}} e^{-kx^{\frac{1}{5}}}, x > 0$$

For this sample:

$$\sum_{i=1}^{20} x_i = 247,360$$

$$\sum_{i=1}^{20} x_i^{\frac{1}{5}} = 102.778$$

and the median is 10,000.

Determine the:
(a) maximum likelihood estimate of k.
(b) method of moments estimate of k.
(c) method of percentiles estimate of k.

Hint: For part b, compare the density function in the question with the PDFs of the distributions you saw in lectures.

Answer:

(a) **Maximum Likelihood Estimation**
The likelihood function is:

$$L = \prod_{i=1}^{n} \frac{k}{5} x_i^{-\frac{4}{5}} e^{-kx_i^{\frac{1}{5}}} = k^n e^{-k\sum x_i^{\frac{1}{5}}} \times constant$$

So the log-likelihood is:

$$logL = nlogk - k\sum x_i^{\frac{1}{5}} + constant$$

Differentiating with respect to k:

$$\frac{\partial}{\partial k} logL = \frac{n}{k} - \sum x_i^{\frac{1}{5}}$$

Equating this to zero, we find that:

$$\hat{k} = \frac{n}{\sum x_i^{\frac{1}{5}}} = \frac{20}{102.778} = 0.1946$$

Differentiating again:

$$\frac{\partial^2}{\partial k^2} logL = -\frac{n}{k^2} < 0$$

So ^k = 0.1946 is the maximum likelihood estimate of k.

(b) Method of moments

Comparing the density function given in the question with the PDF's of known distributions, we see that this is a Weibull distribution with $c = k$ and $\gamma = \frac{1}{5}$, which has mean:

$$E(X) = \frac{\Gamma(1+\frac{1}{\gamma})}{c^{\frac{1}{\gamma}}} = \frac{\Gamma(6)}{k^5} = \frac{120}{k^5}$$

Equating this to the sample mean:

$$\frac{120}{k^5} = \frac{1}{20} \times 247,360 \implies \hat{k} = 0.3957$$

(c) Method of percentiles

The CDF of this Weibull distribution is:

$$F(x) = 1 - e^{-kx^{\frac{1}{5}}}$$

Substituting in the sample median of 10,000 and equating this to 0.5 gives:

$$1 - e^{-k \times 10,000^{\frac{1}{5}}} = 0.5$$

$$\implies \tilde{k} = -\frac{\ln 0.5}{10,000^{\frac{1}{5}}} = 0.1099$$

Exercise: You may want to check that you can indeed get the CDF, $F(x)$ by integrating the pdf, $f(u)$ from u=0 to u=x.

Question 4. R

In-built functions can be found in R for the distributions in the table below. The table below shows the R functions you need for a series of calculation.

(i) Complete the table below.

Hint: Feel free to search online. Or type and run `?dgamma` **and Help notes will appear. Similarly for the other functions.**

Calculation required	$X{\sim}Exp(\lambda)$	$X{\sim}Gamma(\alpha, \lambda)$	$X{\sim}logN(\mu, \sigma^2)$	$X{\sim}Weibull(c,\gamma)$
$f(x)$	`dexp(x, λ)`			
$F(q)$	`pexp(q, λ)`			
Find q for $P(X< q) = p$	`qexp(p, λ)`			
Simulate n random variates	`rexp(n, λ)`			

(ii) Use R to find the median for the following distribution: *Weibull*(0.1099, 0.2).

Answer:

(i)

Calculation required	$X{\sim}Exp(\lambda)$	$X{\sim}Gamma(\alpha, \lambda)$	$X{\sim}logN(\mu, \sigma^2)$	$X{\sim}Weibull(c,\gamma)$
$f(x)$	`dexp(x, λ)`	`dgamma(x, α,λ)`	`dlnorm(x,μ, σ)`	`dweibull(x, γ, c`$^{-(1/\gamma)}$`)`
$F(q)$	`pexp(q, λ)`	`pgamma(q, α,λ)`	`plnorm(q,μ, σ)`	`pweibull(q, γ, c`$^{-(1/\gamma)}$`)`
Find q for $P(X< q) = p$	`qexp(p, λ)`	`qgamma(p, α,λ)`	`qlnorm(p,μ, σ)`	`qweibull(p, γ, c`$^{-(1/\gamma)}$`)`
Simulate n random variates	`rexp(n, λ)`	`rgamma(n, α,λ)`	`rlnorm(n,μ, σ)`	`rweibull(n, γ, c`$^{-(1/\gamma)}$`)`

Note that this website gives examples of how to plot graphs for the lognormal distribution but by changing some of the commands, you will be able to plot graphs for other distributions too:

Log Normal Distribution in R (4 Examples) | dlnorm, plnorm, qlnorm, rlnorm (statisticsglobe.com)

(ii)

```
> qweibull(0.5,0.2, 0.1099^(-5))

[1] 9980.191
```

The median for the *Weibull*(0.1099, 0.2) distribution is 9980.191.

Question 5. R

There are no in-built functions in R for the *Pareto(α, λ)* distribution. So we have to define the functions dpareto, ppareto, qpareto and rpareto from first principles. This has been done for dpareto as follows.

```
dpareto <- function(x,a,lambda){
  a*lambda^(a)/( (lambda+x)^(a+1) )
}
```

(i) Write down the R code for ppareto, qpareto and rpareto, and paste your coding into your answer.

Hint: runif generates random numbers which are between 0 and 1. Use this when you define rpareto.

(ii) Simulate, using rpareto, 1,000 values from a Pareto distribution with parameters α = 3, λ = 1, assigning the simulation to a variable called Pareto_vector and calculate the mean and variance of the simulated values.

Hint: Use the mean and var functions in R.

Answer:

(i)

```
ppareto <- function(q,a,lambda){
  1 - (lambda/(lambda+q))^a
}


qpareto <- function(p,a,lambda){
  lambda*( (1-p)^(-1/a) -1)
}


rpareto <- function(n,a,lambda){
  lambda*( (1-runif(n))^(-1/a) -1)
}
```

(ii)

```
> set.seed(42)
> Pareto_vector <- rpareto(1000,3,1)
> mean(Pareto_vector)
[1] 0.4658703
> var(Pareto_vector)
[1] 0.4288426
```

The mean and variance of the simulated values are 0.4658703 and 0.4288426 respectively.

Note: The random number generator functions simulate a sample of n random variates from a specified distribution. You are advised to specify an initial value for the recursion, using the `set.seed()` function, so you can reproduce your numbers. Otherwise, you'll generate a different set of numbers every time you run your code. It doesn't matter what seed value you use. The seed value in the solution has been chosen arbitrarily.