STAT 8003 Group K:
Homework 3

Lu Li, Andrew Powell, and Abdullah Aljebreen

September 17, 2014

1. Consider a bivariate distribution with $P(X = 1, Y = 2) = 0.4$, $P(X = 2, Y = 3) = 0.6$. Find the correlation coefficient p_{XY} between X and Y.

$$\begin{aligned}
E\{XY^*\} &= \sum_{\text{all } x} \sum_{\text{all } y} XY^* P(X = x, Y = y) = 4.4 \\
E\{|X|^2\} &= \sum_{\text{all } x} |X|^2 P(X = x) = 2.8 \\
E\{|Y|^2\} &= \sum_{\text{all } y} |Y|^2 P(Y = y) = 7 \\
E\{X\} &= \sum_{\text{all } x} XP(X = x) = 1.6 \\
E\{Y\} &= \sum_{\text{all } y} YP(Y = y) = 2.6 \\
\text{Cov}\{X, Y\} &= E\{XY^*\} - E\{X\}E\{Y\} = 0.24 \\
\text{Var}\{X\} &= E\{|X|^2\} - E\{X\}^2 = 0.24 \\
\text{Var}\{Y\} &= E\{|Y|^2\} - E\{Y\}^2 = 0.24 \\
p_{XY} &= \frac{\text{Cov}\{X, Y\}}{\sqrt{\text{Var}\{X\}\text{Var}\{Y\}}} = 1
\end{aligned} \tag{1}$$

2. Find two random variables X and Y, such that $\text{Cov}\{X, Y\} = 0$ but X and Y are not independent.

   Let $X = \{-1, 0, 1\}$ with equal probability and $Y = X^2$. Since the relationship between Y and X are defined, X and Y are not independent.

$$\begin{aligned}
\text{Cov}\{X, Y\} &= E\{XY^*\} - E\{X\}E\{Y\} \\
&= E\{X(X^2)^*\} - E\{X\}E\{(X^2)\} \\
&= E\{X^3\} - 0(0) \\
&= 0
\end{aligned} \tag{2}$$

3. In the Example of GDP. Assume that the data follows a gamma distribution $\Gamma(\alpha, \beta)$

   (a) Derive the estimator of shape α and rate β using the methods of moments.

   Since there are two parameters for which need calculation, two moments are necessary. The first two moments, m_1 and m_2, are calculated as follows.

$$
\begin{aligned}
m_1 &= E\{X\} \\
m_2 &= \mathrm{Var}\{X\} + E\{X\}^2
\end{aligned} \tag{3}
$$

Substituting the variance and mean for a gamma distribution, the equations above result in the following.

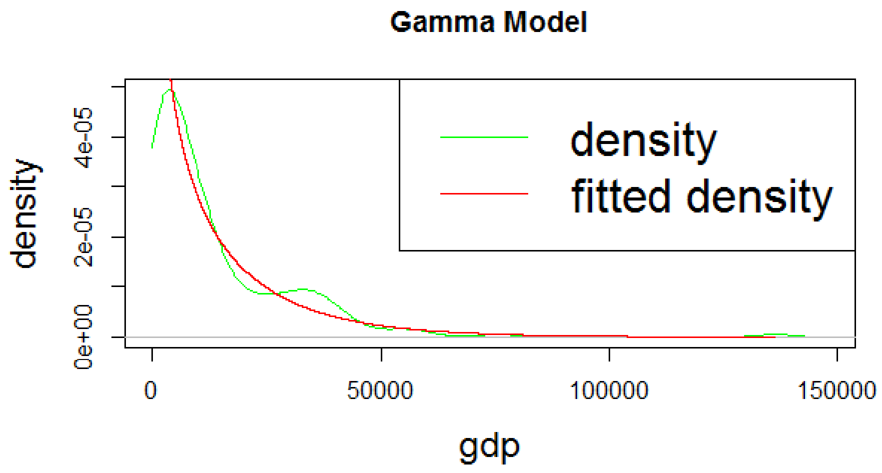
$$
\begin{aligned}
\mathrm{Var}\{X\} &= \frac{\alpha}{\beta^2} \\
\mathrm{Mean}\{X\} = E\{X\} &= \frac{\alpha}{\beta} \\
m_1 &= \frac{\alpha}{\beta} \\
m_2 &= m_1\beta^{-1} + m_1^2 \\
\beta &= \frac{m_1}{m_2 - m_1^2} \\
\alpha &= m_1\beta
\end{aligned} \tag{4}
$$

(b) Compare the density of the data versus the fitted curve.

```r
## load the data
gdp.all <- read.csv("http://goo.gl/PLH4ek")
gdp <- gdp.all[,62]
gdp[ is.na(gdp) ] <- NULL


## No. 3
## The method of moment based on the Gamma distribution

m1 <- mean( gdp )
m2 <- mean( gdp^2 )

 ################################
 # mu = m1 = alpha/beta         #
 # var = m2-m1^2 = alpha/beta^2 #
 # alpha = m1^2/(m2-m1^2 )      #
 # beta = m1/(m2-m1^2 )         #
 ################################

alpha.hat <- m1^2/(m2-m1^2 )
beta.hat <- m1/(m2-m1^2 )

 #################
 # > alpha.hat   #
 #   0.6523107   #
 # > beta.hat    #
 #   4.83963e-05 #
 #################

plot( density( gdp, from=0 ), col='green', xlab="gdp", ylab="density",
      main="Gamma Model", cex.lab=1.5, cex=2)
x=c(1:max(gdp))
points( x, dgamma( x, shape = alpha.hat, rate = beta.hat ), 'l',
        col='red' )
legend("topright", c("density", "fitted density"), lty=c(1, 1),
       col=c('green', 'red'), cex=2 )

```

Listing 1: Problem 3b Source Code

Figure 1



4. Use the given relationships to answer the following questions.

   (a) Assume that X follows a gamma distribution with parameters shape α and rate β. Calculate the cumulant generating function $S_X(t)$ of $\ln(X)$.

$$
\begin{aligned}
S_{\ln(X)}(t) &= \ln(M_{\ln(X)}(t)) \\
&= \ln(E\{e^{\ln(X)t}\}) \\
&= \ln(E\{X^t\}) \\
&= \ln\left[\int_0^\infty x^t f_x(x)dx\right] \\
&= \ln\left[\int_0^\infty x^t \left(\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x}\right) dx\right] \\
&= \ln\left[\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha+t-1}e^{-\beta x}dx\right] \\
c_0 &= \alpha + t \\
v_0 &= \beta x \\
dv_0 &= \beta dx \\
S_{\ln(X)}(t) &= \ln\left[\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty \left(\frac{v_0}{\beta}\right)^{c_0-1} e^{-v_0}\frac{dv_0}{\beta}\right] \\
&= \ln\left[\frac{\beta^{\alpha-c_0-1}}{\Gamma(\alpha)}\int_0^\infty v_0^{c_0-1}e^{-v_0}dv_0\right] \\
&= \ln\left[\frac{\beta^{\alpha-(c_0-1)-1}}{\Gamma(\alpha)}\Gamma(c_0)\right] \\
&= \ln\left[\frac{\beta^{\alpha-(\alpha+t)}}{\Gamma(\alpha)}\Gamma(\alpha+t)\right] \\
&= \ln\left[\beta^{-t}\frac{\Gamma(\alpha+t)}{\Gamma(\alpha)}\right]
\end{aligned} \tag{5}
$$

   (b) Calculate $E\{\ln(X)\}$ and $\text{Var}\{\ln(X)\}$. Write your final result by using the digamma function $\psi(x)$ and trigamma function ψ_1(x).

$$\begin{aligned} E\{\ln(X)\} &= \frac{d}{dt}[S_{\ln(X)}(t)]|_{t=0} \\ &= \frac{d}{dt}[-t\ln(\beta) + \ln(\Gamma(\alpha+t) - \ln(\Gamma(\alpha))]|_{t=0} \\ &= [-\ln(\beta) + \frac{d}{dt}(\ln(\Gamma(\alpha+t)))]|_{t=0} \\ &= \psi(\alpha) - \ln(\beta) \end{aligned} \tag{6}$$

$$\begin{aligned} \mathrm{Var}\{\ln(X)\} &= \frac{d}{dt}[\psi(\alpha+t) - \ln(\beta)]|_{t=0} \\ &= [\frac{d}{dt}(\psi(\alpha+t))]_{t=0} \\ &= \psi_1(\alpha) \end{aligned} \tag{7}$$

(c) Match the first and second moment of $E\{\ln(X)\}$, and derive the MOM estimator of α and β.

$$\begin{aligned} m_1 &= E\{\ln(X)\} \\ m_2 &= \mathrm{Var}\{\ln(X)\} + E\{\ln(X)\}^2 \\ m_1 &= \psi(\alpha) - \ln(\beta) \\ m_2 &= \psi_1(\alpha) + m_1^2 \\ \alpha &= \psi_1^{-1}[m_2 - m_1^2] \\ \beta &= e^{\psi(\alpha) - m_1} \end{aligned} \tag{8}$$

(d) Apply your estimator to the GDP dataset and estimate the parameters of α and β.

```
## No. 4
## The method of moment based on the Log-Gamma distribution
m1.log <- mean( log(gdp) )
m2.log <- mean( (log(gdp))^2 )

 ###########################################
 # mu = m1.log = digamma(alpha) - ln(beta) #
 # var = m2.log-m1.log^2 = trigamma(alpha) #
 # alpha = trigammaInverse(m2.log-m1.log^2)#
 # beta = exp(digamma(alpha) - m1.log)     #
 ###########################################

alpha.log.hat <- limma::trigammaInverse(m2.log-m1.log^2)
beta.log.hat <- exp(digamma(alpha.log.hat) - m1.log)

 ###################
 # > alpha.log.hat #
 #   0.9575706     #
 # > beta.log.hat  #
 #   8.029302e-05  #
 ###################

```

Listing 2: Problem 3b Source Code