Math 3070 § 1.	**Failure Time Example: Hypothesis**	Name: Example
Treibergs	**test for Exponential Variables**	July 10, 2011

How to test hypotheses with exponential variables is discussed in Problem 8.85 of Devore, *Probability and Statistics for Engineering and the Sciences, 8th ed.,* Brooks Cole, 2012. Devore applies it to fake data of time to failure of identical components.

Assume that we have a random sample $X_1, X_2, \ldots, X_n \sim \text{Exp}(\lambda)$. Since $\text{E}(X) = \mu = 1/\lambda$, then an estimator for μ is the mean.

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The MLE for λ turns out to be $\hat{\lambda} = 1/\hat{\mu}$.

Note that if $X_i \sim \text{Exp}(\lambda)$, then its pdf for $x \ge 0$ is

$$f(x) = \lambda e^{-\lambda x}.$$

Hence, the pdf for $Y = 2\lambda X_i$ is

$$f_Y(x) = \frac{1}{2\lambda} f\left(\frac{x}{2\lambda}\right) = \frac{1}{2}e^{-x/2} = f(x;2)$$

which is exactly the pdf for the χ^2 distribution with two degrees of freedom. For ν degrees of freedom, the χ^2 pdf is

$$f(x,\nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, & \text{if } x \ge 0; \\ 0, & \text{otherwise.} \end{cases}$$

It is a fact that the sum of χ^2-variables is also a χ^2 variable. To see this, suppose $Y \sim \text{ChiSq}(\nu)$ and $Z \sim \text{ChiSq}(\kappa)$ so if $W = Y + Z$ then for $w \ge 0$ the cdf

$$\begin{aligned} \text{P}(W \le w) &= \text{P}(X + Y \le w) \\ &= \iint_{x+y\le w} f(x;\lambda)\, f(y;\mu)\, dx\, dy \\ &= \frac{1}{2^{(\nu+\kappa)/2}\Gamma(\nu/2)\Gamma(\kappa/2)} \int_0^w \int_0^{w-x} x^{\nu/2-1} y^{\kappa/2-1} e^{-(x+y)/2}\, dy\, dx \end{aligned}$$

Changing variables $u = x - y$, $v = x + y$ we find

$$\text{P}(W \le w) = \frac{1}{2^{(\nu+\kappa)/2}\Gamma(\nu/2)\Gamma(\kappa/2)} \int_{v=0}^{w}\int_{u=-v}^{v} \left(\frac{u+v}{2}\right)^{\nu/2-1} \left(\frac{v-u}{2}\right)^{\kappa/2-1} e^{-v/2}\, \frac{du\, dv}{2}$$

Hence, the pdf for $w \ge 0$ is

$$\begin{aligned} f_W(w) &= \frac{d}{dw}\text{P}(W \le w) \\ &= \frac{1}{2^{(\nu+\kappa)/2+1}\Gamma(\nu/2)\Gamma(\kappa/2)} \int_{u=-w}^{w} \left(\frac{u+w}{2}\right)^{\nu/2-1} \left(\frac{w-u}{2}\right)^{\kappa/2-1} e^{-w/2}\, du \\ &= \frac{w^{(\nu+\kappa)/2-2}e^{-w/2}}{2^{(\nu+\kappa)/2+1}\Gamma(\nu/2)\Gamma(\kappa/2)} \int_{u=-w}^{w} \left(\frac{u+w}{2w}\right)^{\nu/2-1} \left(\frac{w-u}{2w}\right)^{\kappa/2-1} du \\ &= \frac{w^{(\nu+\kappa)/2-2}e^{-w/2}}{2^{(\nu+\kappa)/2+1}\Gamma(\nu/2)\Gamma(\kappa/2)} \cdot \frac{2w\Gamma(\nu/2)\Gamma(\kappa/2)}{\Gamma((\nu+\kappa)/2)} \\ &= \frac{1}{2^{(\nu+\kappa)/2}\Gamma((\nu+\kappa)/2)} w^{(\nu+\kappa)/2-1} e^{-w/2} \end{aligned}$$

where we have used the fact that the integrand is up to gamma function factors the pdf of the beta distribution with $\alpha = \nu/2$, $\beta = \kappa/2$, $A = -w$ and $B = w$. Note that the pdf of the sum is also χ^2 with $f_W(w) = f(w; \nu + \kappa)$. By induction, it follows that

$$2\lambda \sum_{i=1}^{n} X_i \sim \text{ChiSq}(2n).$$

To test the hypothesis that $\lambda < \lambda_0$ it is equivalent to test $\mu > \mu_0$ where $\mu_0 = 1/\lambda_0$. Under the null hypothesis $\mu = \mu_0$, the statistic

$$X^2 = 2\lambda_0 \sum_{i=1}^{n} X_i \sim \text{ChiSq}(2n).$$

Thus there is evidence at the $\alpha = .10$ level that $\lambda < \lambda_0$ if $X^2 > \chi^2_\alpha$, where the critical value $\text{P}(\chi^2 > \chi^2_\alpha) = \alpha$. In case $n = 10$, so for 20 degrees of freedom, $\chi^2_{.10} = 28.41$. Because the statistic works out to be $X^2 = 12.44$, we cannot reject the null hypothesis: there is no strong evidence that $\lambda < \lambda_0$. We can also compute the p-value. For the one-sided upper tailed test, the p-value is $\text{P}(\chi^2 \geq X^2) = .480$. We plot the distribution with 20 degrees of freedom, the observed X^2, the critical value $\chi^2_{.10}$, the p-value (which is the entire yellow tail area under the pdf to the right of the observed X^2, and the pink area equal to .10 of the whole tail to the right of $\chi^2_{.10}$.

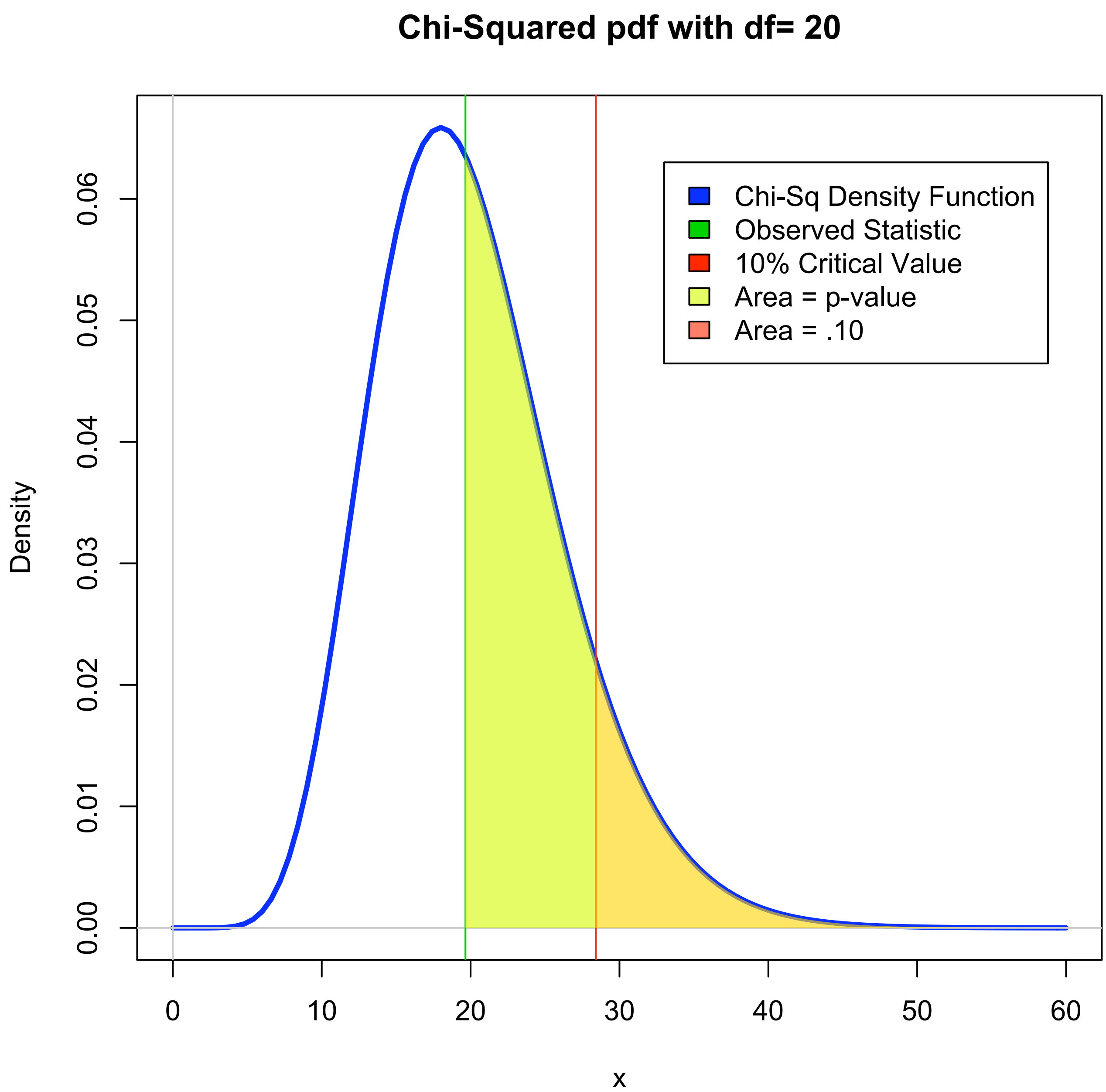
We simulate $B = 10,000$ random samples of size $n = 10$ taken from Exp(1). We plot the histogram of twice the sums, and the theoretical χ^2 distribution with 20 degrees of freedom. Note that the empirical distribution follows the theoretical closely.

We also test if the data is exponential and the simulated sums follow χ^2. To do this, we make QQ-plots of the data and the simulation. If there are n observations, we plot the theoretical quantile $q_i = \text{P}(p_1)$ with the observed quantile, which is the ith point y'_i, after sorting the data $y_1, \ldots, y_n$. The percentage is computed by the function $p_i =$ `ppoints(n)[i]`, which gives the fraction corresponding to the ith of n points. Devore uses the function $p_i = (i - .5)/n$ which is the same as `ppoints(n)[i]` if $n > 10$ but `ppoints(n)[i]` $= (i - .375)/(n + .25)$ if $n \leq 10$. We plot the data vs. exponential quantiles. The agreement is pretty good for such a small sample so that exponentialness is not contradicted. We also run $B = 500$ sums of exponentials and plot them against χ^2 quantiles. The points line up reasonably well, so χ^2-ness is not contradicted.

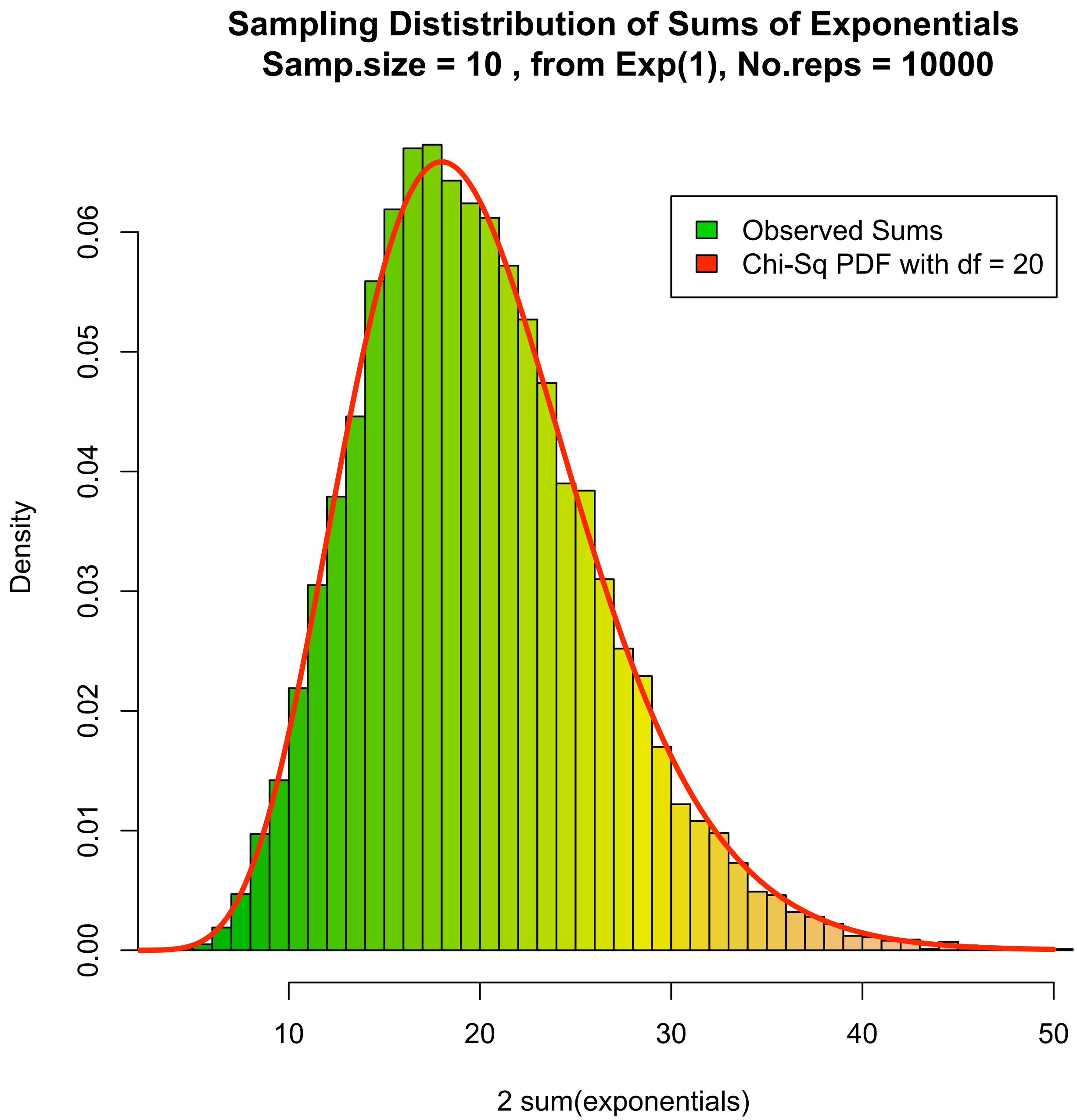
R Session:

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
  Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.31 (5538) powerpc-apple-darwin8.11.1]
[Workspace restored from /Users/andrejstreibergs/.RData]
```

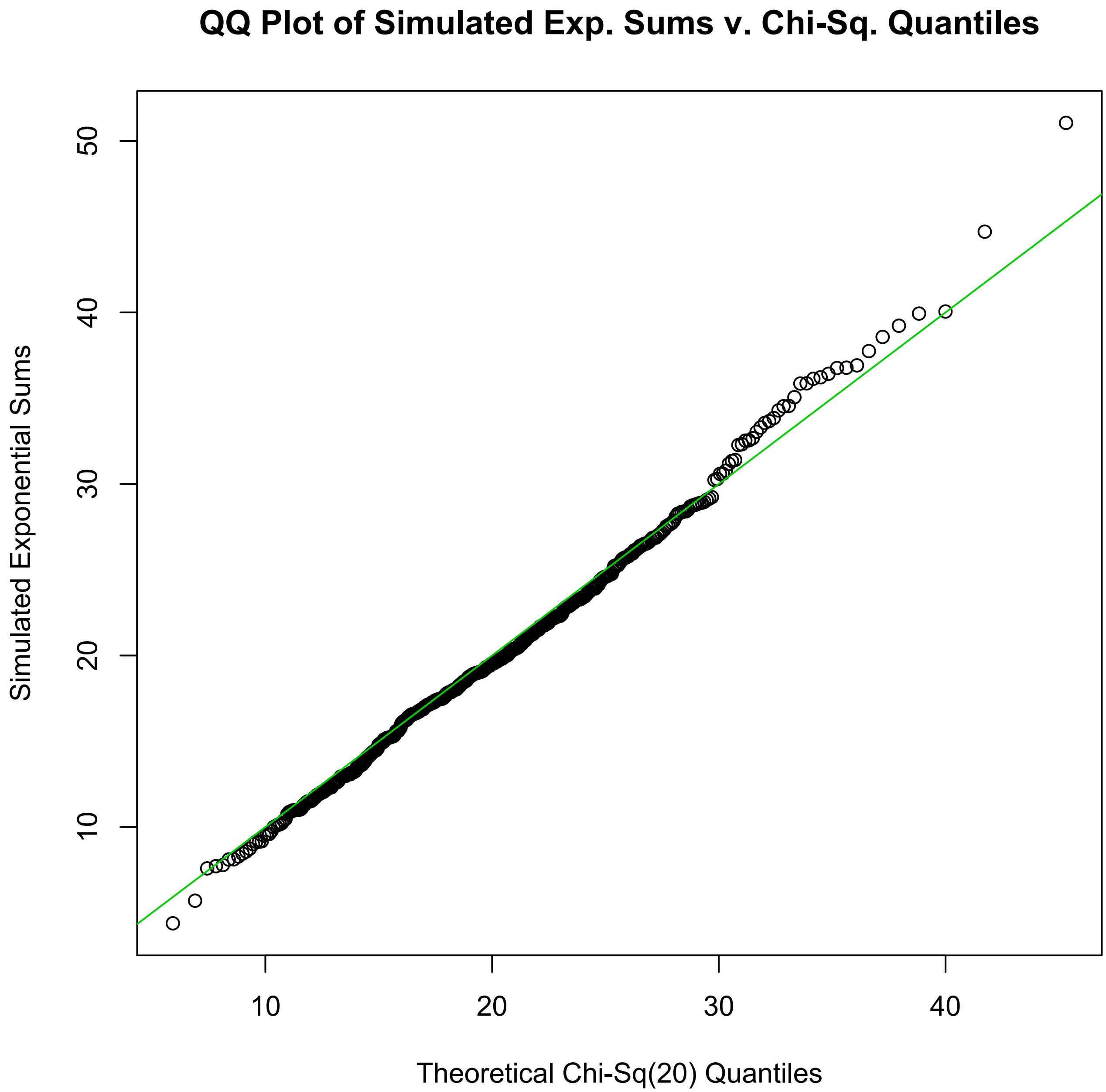
```
> ############### ENTER FAILURE TIME DATA. TEST IF LAMBDA < 75 ##################
> # Devore 8.85
> tf <- scan()
1: 95 16 11 3 42 71 225 64 87 123
11:
Read 10 items
> tf
 [1]  95  16  11   3  42  71 225  64  87 123

> # estimator for  mu
> mu <- mean(tf);mu
[1] 73.7
> lambda <- 1/mu
> mu0 <- 75
> lambda0 <- 1/mu0
> # test if mu < 75
> # equiv test if lambda > lambda0
> n <- length(tf)
> # In H0 lambda=lambda0
> # Compute the statistic
> chi2 <- 2*lambda0*sum(tf)
> chi2
[1] 19.65333

> # p - value
> pvalue <- pchisq(chi2,df=2*n,lower.tail=F); pvalue
[1] 0.4797945

> # Cannot reject niull hypothesis.
> # Find alpha = .10 critical value of statistic
> alpha <- .10
> qcrit <- qchisq(alpha,df=2*n,lower.tail=F); qcrit
[1] 28.41198

> # Larger than statistic. We cannot reject H0.
>
> ############### PLOT CHI SQ DISTRIBUTION WITH DF=20 ############################
> seq1 <- seq(chi2,qcrit,(qcrit-chi2)/117)
> seq2 <- seq(qcrit,60,(60-qcrit)/223)
> clr <- rainbow(15,alpha=.6)
> curve( dchisq(x, df = 2*n), 0, 60, lwd = 3, col = 4, ylab = "Density",
+ main = paste("Chi-Squared pdf with df=", 2*n))
> abline(h = 0, col = 8); abline(v = 0, col = 8)
> abline(v = chi2, col = 3); abline(v = qcrit, col = 2)
> polygon(c(chi2, seq1, qcrit, chi2), c(0, dchisq(seq1, df = 2*n), 0, 0),
+ col = clr[4], border = NA)
> polygon(c(qcrit, seq2, 60, qcrit), c(0, dchisq(seq2, df = 2*n), 0, 0),
+ col = clr[3], border = NA)
> legend(33, .063, legend=c("Chi-Sq Density Function", "Observed Statistic",
+ "10% Critical Value", "Area = p-value", "Area = .10"),
+ fill = c(4, 3, 2, clr[4], clr[1]), bg = "white")
> # M3074FailTime1.pdf
```


Chi-Squared pdf with df= 20
Chi-Sq Density Function
Observed Statistic
10% Critical Value
Area = p-value
Area = .10
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0
10
20
30
40
50
60
x

```
> ################ SAMPLING DISTRIBUTION OF SUMS OF EXPONENTIALS #################
>
> tn <- 2*n
> B <- 10000
> v <- replicate(B,2*sum(rexp(n,1)))
> hist(v, breaks = 40, freq = F, col=terrain.colors(50), xlab="2 sum(exponentials)",
+ main = paste("Sampling Dististribution of Sums of Exponentials\n Samp.size =", n,
+ ", from Exp(1), No.reps =", B))
> lines(xs, dchisq(xs, df = tn), lwd = 3, col = 2)
> legend(30, .063, legend = c("Observed Sums",paste("Chi-Sq PDF with df =", tn)),
+ fill = c(3, 2))
> # M3074FailTime2.pdf
>
> ################ EXPONENTIAL PP-PLOT TO TEST IF DATA IS EXPONENTIAL #############
>
> ppoints(n)
 [1] 0.06097561 0.15853659 0.25609756 0.35365854 0.45121951 0.54878049 0.64634146
 [8] 0.74390244 0.84146341 0.93902439

> qqplot(qexp(ppoints(n), lambda), tf, ylab="Observed Waiting Times",
+ xlab = "Theoretical Exponential Quantiles",
+ main = "QQ Plot of Waiting Times v. Exponential Quantiles")
> abline(0,1,col=2)
> # M3074FailTime3.pdf
>
> ############# CHI-SQ PP-PLOT OF SAMPLING DISTRIBUTION ###########################
>
> B <- 500
> v <- replicate(B, 2*sum(rexp(n, 1)))
> qqplot(qchisq(ppoints(B), df = tn), v, ylab="Simulated Exponential Sums",
+ xlab = "Theoretical Chi-Sq(20) Quantiles",
+ main = "QQ Plot of Simulated Exp. Sums v. Chi-Sq. Quantiles")
> abline(0,1,col=3)
> # M3074FailTime4.pdf
```


Sampling Dististribution of Sums of Exponentials
Samp.size = 10 , from Exp(1), No.reps = 10000
Observed Sums
Chi-Sq PDF with df = 20
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
10
20
30
40
50
2 sum(exponentials)


QQ Plot of Waiting Times v. Exponential Quantiles
Observed Waiting Times
Theoretical Exponential Quantiles
0
50
100
150
200
0
50
100
150
200


QQ Plot of Simulated Exp. Sums v. Chi-Sq. Quantiles
Simulated Exponential Sums
10
20
30
40
50
Theoretical Chi-Sq(20) Quantiles