# Exam 1
# Answers to selected problems

1. Let us write the likelihood function:

$$L(x_1, x_2, ..., x_n \mid \theta) = \frac{\theta^n}{\prod_{i=1}^{n}(1+x_i)^{\theta+1}}, \theta > 0, 0 < x_i < \infty$$

We then find the MLE of $\theta$ by maximizing the log-likelihood function:

$$\frac{d}{d\theta}\ln(L(x_1, x_2, ..., x_n \mid \theta)) = \frac{n}{\theta} - \sum_{i=1}^{n}\ln(1+x_i) = 0, \theta = \frac{n}{\sum_{i=1}^{n}\ln(1+x_i)}.$$

Now, since $L(x_1, x_2, ..., x_n \mid \theta) = \frac{\theta^n}{\prod_{i=1}^{n}(1+x_i)^{\theta+1}} = \frac{\theta^n}{\exp\left((\theta+1)\sum_{i=1}^{n}\ln(1+x_i)\right)}$, in view of Theorem 5.6.1 (Larsen and Marx) it follows that $\sum_{i=1}^{n}\ln(1+x_i)$ is a sufficient statistic for $\theta$.

2.

3. Since the populations are independent, the sample means will be independent as well. Moreover, recall from Theorem 7.3.2 that the sample mean and the sample variance of a sample taken from a normal population are independent, and we have $\frac{(n_j - 1)S_j^2}{\sigma^2} \sim \chi^2(n_j - 1)$. We then have the following:

a) Since $\hat{\theta} = c_1\overline{X}_1 + c_2\overline{X}_2 + ... + c_k\overline{X}_k$ is a linear combination of independent random variables that are normally distributed $N(\mu_j, \sigma^2 / n_j)$, $\hat{\theta}$ is normally distributed with mean $E(\hat{\theta}) = c_1\mu_1 + c_2\mu_2 + ... + c_k\mu_k = \theta$ and variance $V(\hat{\theta}) = \sum_{i=1}^{k}\frac{c_i^2\sigma^2}{n_j}$.

b) In view of what was said above, $\frac{SSE}{\sigma^2} = \frac{\sum_{j=1}^{k}(n_j - 1)S_j^2}{\sigma^2}$ is a sum of independent Chi-square variables, so the result will be a Chi-square variable whose degrees of freedom equal the sum of the degrees of the variables that have been summated. Therefore, $\frac{SSE}{\sigma^2} \sim \chi^2(n_1 + n_2 + ... + n_k - k)$.

c) Since the sample means are independent, and the sample variances are independent variables, using again Theorem 7.3.2, it follows that $\hat{\theta}, SSE$ are independent random variables. Now, $\frac{\hat{\theta}-\theta}{V(\hat{\theta})} \sim N(0,1)$. Then, by Definition 7.3.3

$$\frac{\hat{\theta}-\theta}{V(\hat{\theta})} \Big/ \sqrt{\frac{SSE}{\sigma^2(n_1+n_2+...+n_k-k)}} \sim T(n_1+n_2+...+n_k-k)$$

But the above is nothing but $\frac{\hat{\theta}-\theta}{\sqrt{\left(\frac{c_1^2}{n_1}+\frac{c_2^2}{n_2}+...+\frac{c_k^2}{n_k}\right)MSE}}$, so its distribution is Student with $n_1+n_2+...+n_k-k$ degrees of freedom.

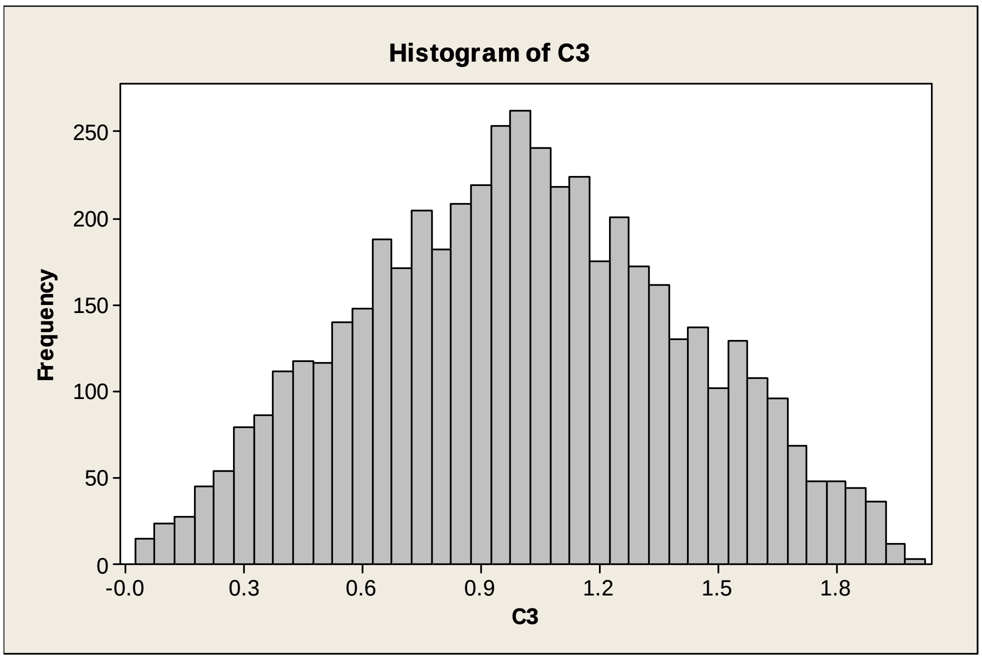
4.
5.

6. a) The level of significance of test (1) is $\alpha = P(X_1 > 0.95 \mid \theta = 0) = .05$.

Now, from the distribution of $X_1 + X_2 : f(t) = \begin{cases} t-2\theta, & 2\theta < t < 2\theta+1 \\ 2\theta+2-t, & 2\theta+1 \le t < 2\theta+2 \end{cases}$

$P(X_1 + X_2 > C \mid \theta = 0) = \int_C^2 (2-t)dt = 2(2-C) - \frac{t^2}{2}\Big|_C^2 = \frac{C^2-4C+4}{2} = .05$ (assuming that C will have to be greater than 1). We have $(C-2)^2 = .1 \Rightarrow C = 2-\sqrt{.1} \approx 1.684$.

Below is a histogram of 5000 simulated values for $X_1 + X_2$.



b) In order to calculate the power of the first test we need to calculate

$h_1(\theta) = P(X_1 > 0.95 \mid \theta > 0)$. We have $h_1(\theta) = \begin{cases} \theta + 0.05, \theta \le 0.95 \\ 1, \theta > 0.95 \end{cases}$.

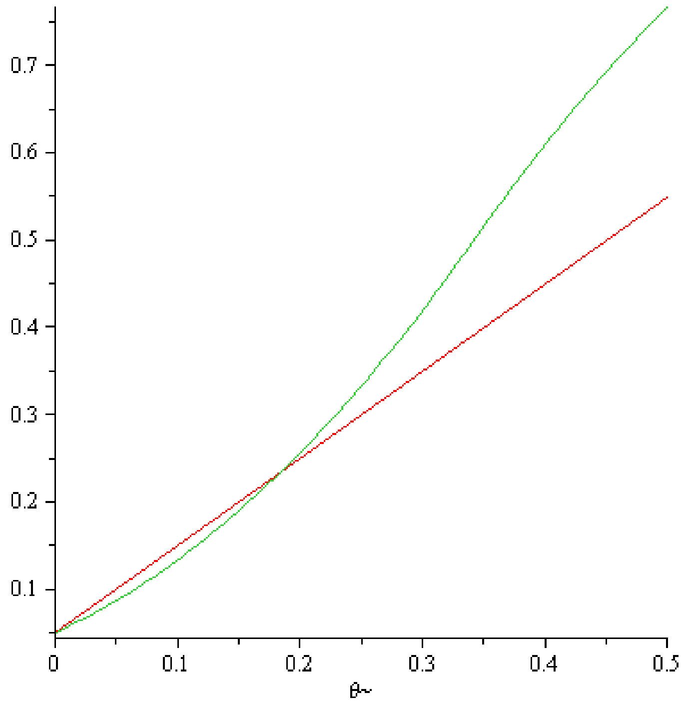
For the second test we have: $h_2(\theta) = P(X_1 + X_2 > 2 - \sqrt{.1} \mid \theta > 0) = \int_{2-\sqrt{.1}}^{2\theta+2} f(t)dt$.

For $\theta : 2\theta + 1 < 2 - \sqrt{.1}, h_2(\theta) = \int_{2-\sqrt{.1}}^{2\theta+2} (2\theta + 2 - t)dt = \frac{(2\theta + \sqrt{.1})^2}{2}$

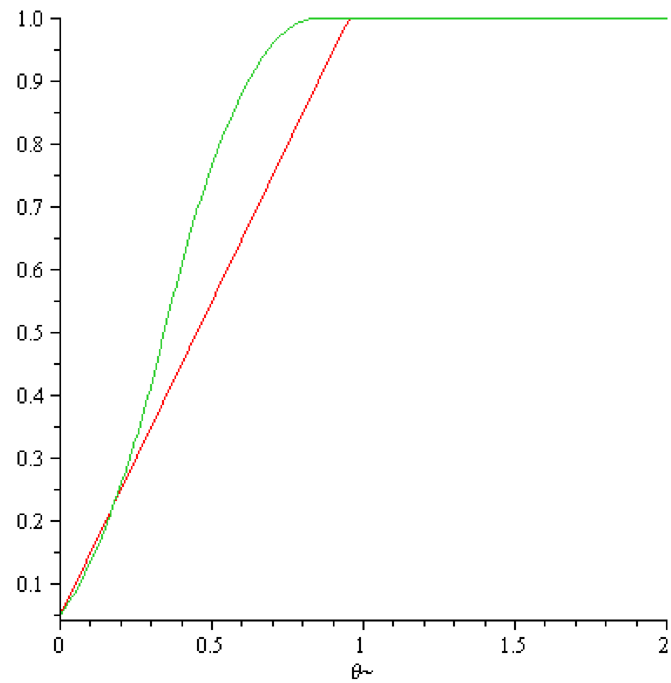
For $\theta : 2\theta + 1 \ge 2 - \sqrt{.1}$, but $2\theta < 2 - \sqrt{.1}, h_2(\theta) = 1 - \int_{2\theta}^{2-\sqrt{.1}} (t - 2\theta)dt = 1 - \frac{(2 - \sqrt{.1} - 2\theta)^2}{2}$

For $2\theta \ge 2 - \sqrt{.1}, h_2(\theta) = 1$

Below is a plot of the two power functions for $0 \le \theta \le 0.5$. The graph in red is $h_1(\theta)$ and the graph in green is $h_2(\theta)$. It can be seen that on [0,0.2] (approximately) the first test is more powerful that the second, and after that,the second test is more powerful.




Here i a graph of the two power functions on the interval [0,2].




7.