STATISTICS COMPREHENSIVE EXAM
SUMMER 2011

(1) **[10 points]** The random variable X has pdf $f(x) = 2x$ for $0 < x < 1$. $X_1, \ldots, X_n$ is a random sample from the distribution of the continuous random variable X. $\hat{\theta} = \frac{1}{n-1}\sum_{i=1}^{n} X_i$ is taken as an estimate of the distribution mean θ. Find the MSE (mean squared error) of $\hat{\theta}$. Recall $MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias(\hat{\theta})^2$. (NOTE: Your answer should be in terms of n.)

(2) **[20 points]** Let $Y_1, \ldots, Y_n$ be a random sample from the density

$$g(y) = \frac{\theta}{y^2},\ 0 < \theta \leq y.$$

(a) Find a sufficient statistic for θ.
(b) Find the MLE of θ.
(c) Does the method of moment estimate of θ exist? If yes, find it. If no, why?

(3) **[20 points]** Let $X_1, X_2, \ldots, X_n$ be i.i.d. from a distribution having p.d.f. of the form

$$f(x) = \theta x^{\theta-1}, \quad 0 < x \leq 1.$$

(a) Find the Rejection Region of the most powerful test for $H_0 : \theta = 1$ versus $H_1 : \theta = 2$.
(b) Find the likelihood ratio test statistic for $H_0 : \theta = 1$ versus $H_1 : \theta > 1$. When you would reject the null hypothesis?

(4) **[35 points]** Consider 5 i.i.d. observations $X_1, \ldots, X_5 \sim N(0, \sigma^2)$. Consider two estimators, $T_1 = \bar{X}$ (MLE) and $T_2 = X_m$ (median).
   (a) What are the distributions of T_1 and T_2?
   (b) Determine the relative efficiency of T_2 with respect to T_1.
   (c) Does your answer in (b) make sense? (i.e. Which estimator is more efficient and why is this expected?)
   (d) If a sixth observation is added, how does this change the relative efficiency in (b)? Compute the value for the relative efficiency.
   (e) Show that as n increases, the relative efficiency of the mean in comparison to the median approaches 0.
   (f) Introductory statistics classes typically present the sample median as a better measure of central tendency compared to the sample mean due to its robustness. That is, the median is less influenced by outliers than the mean. However, we showed that the median is fairly inefficient compared to the mean in part (e). Does this change your mind about whether the median should be favored over the median? Or, should people use the median in certain situations? State your opinion and explain.
   (g) If you like the median for its robustness, but desire a more efficient estimator, you can try the $p\%$ trimmed mean, which drops the highest $p\%$ and the lowest $p\%$ of the data, and then averages the rest. Confirm that the relative efficiency of the mean compared to the trimmed mean is $1 - 2p\%$. What happens to the variance of the trimmed mean estimator as the amount of trimming increases?

(5) **[15 points]** Consider a One-Way ANOVA for four groups (treatments).
   (a) State the hypotheses.
   (b) Describe, in a meaningful way, what the F statistic represents conceptually. Do not simply state the formula.