Spring 04 | Name: ______________________________
Math 3770 | Third Midterm | Bonetto

1) You play with a friend the following game: you flip a coin and if the result is Head you win 1$ while if it is Tail you lose 1$ (*i.e.* you win -1\$). Let X_i the r.v. that describes the ammount you win at the i-th coin flip. You play the game 2500 times. Let $S = \sum_{i=1}^{2500} X_1$ the r.v. describing the total amount you win after these 2500 plays. Assume that the coin is fair, *i.e.* assume the $P(X_i = 1) = P(X_i = -1) = 0.5$.

a) Write an approximate p.d.f. for S.

We have $E(X_i) = 0$ and $V(X_i) = 1$ so that S is approximatly distributed as $N(0, 2500)$.

b) What is the approximate probability that at the end of the game you lost more than 20$? And more than 100$? (Use the table at the end of the test)

The variable $Z = S/50$ is normal standard so that:

$$P(S < -20) = P\left(Z < -\frac{20}{50}\right) = \Phi(-0.4) =$$
$$P(S < -100) = P\left(Z < -\frac{100}{50}\right) = \Phi(-2.0) =$$

At the end of the game you lost 300$. You suspect that your friend cheated you and the coin is not fair. Assume now that $P(X_i = 1) = p$ and $P(X_i = -1) = 1 - p$

c) Give an estimate on p based on the previous result.

If you lost 300$ it means that the number of $X_i = 1$ was 1100 and the number of $X_i = -1$ was 1400. so that you can estimate:

$$p = \frac{1100}{2500} = 0.44$$

d) Give a 99% CI for p. (Remeber that $x_{0.005} = 2.58$ and that $V(X_i) = 4p(1-p)$)

Let Y_i the r.v. that is 1 if the i-th flip was a Head and 0 if it was a Tail. We have $V(Y_i) = V(X_i)/4 = p(1-p) = s^2$ so that the requested CI is

$$\left[0.44 - \frac{2.58 \cdot 0.496}{50}, 0.44 + \frac{2.58 \cdot 0.496}{50}\right] = [0.414, 0.465]$$

e) (**Bonus**) Do you think the coin was fair? Why?

The coin is probably not fair because 0.5 is not in the 99% CI.

2) After running a sample of size $n = 100$ an experimenter find the following values:

0.985	0.118	0.894	0.784	0.101	0.253	0.020	0.378	0.679	0.681
0.753	0.006	0.624	0.126	0.618	0.771	0.187	0.497	0.510	0.316
0.702	0.367	0.876	0.371	0.525	0.428	0.543	0.668	0.171	0.831
0.862	0.769	0.053	0.967	0.334	0.082	0.270	0.074	0.676	0.810
0.464	0.569	0.512	0.587	0.739	0.872	0.578	0.965	0.334	0.393
0.177	0.112	0.853	0.811	0.955	0.904	0.139	0.686	0.127	0.004
0.001	0.193	0.587	0.375	0.027	0.465	0.785	0.442	0.288	0.853
0.526	0.728	0.266	0.783	0.468	0.197	0.447	0.728	0.966	0.757
0.466	0.869	0.134	0.846	0.018	0.941	0.670	0.800	0.143	0.610
0.908	0.774	0.366	0.944	0.885	0.445	0.163	0.425	0.480	0.031

that ordered in increasing order are:

0.001	0.004	0.006	0.018	0.020	0.027	0.031	0.053	0.074	0.082
0.101	0.112	0.118	0.126	0.127	0.134	0.139	0.143	0.163	0.171
0.177	0.187	0.193	0.197	0.253	0.266	0.270	0.288	0.316	0.334
0.334	0.366	0.367	0.371	0.375	0.378	0.393	0.425	0.428	0.442
0.445	0.447	0.464	0.465	0.466	0.468	0.480	0.497	0.510	0.512
0.525	0.526	0.543	0.569	0.578	0.587	0.587	0.610	0.618	0.624
0.668	0.670	0.676	0.679	0.681	0.686	0.702	0.728	0.728	0.739
0.753	0.757	0.769	0.771	0.774	0.783	0.784	0.785	0.800	0.810
0.811	0.831	0.846	0.853	0.853	0.862	0.869	0.872	0.876	0.885
0.894	0.904	0.908	0.941	0.944	0.955	0.965	0.966	0.967	0.985

a) Compute the median $\tilde{x}$ and the fourth spread f_s of the sample.

The median is

$$\tilde{x} = \frac{0.525 + 0.512}{2} = 0.519$$

The median of the upper part is $(0.774 + 0.783)/2 = 0.779$ while the median of the lower part is $(0.253 + 0.266)/2 = 0.260$ so that the fourth spread is

$$f_s = 0.779 - 0.260 = 0.519$$

He wants to visualize the result with an histogram.

b) How many classes should he use?

The best number of classes is $\sqrt{100} = 10$.

c) Write the boundaries and the value for the second and sixth class.

We can take 0.0 as minimum value and 1.0 as maximum so that the second class is $[0.1, 0.2]$ and the sixth is $[0.5, 0.6]$. The frequency of the second class is 14 while the frequency of the sixth is 9. This implies that the value of the histogram for the second class is 1.4 and for the sixth is 0.9.

The experimenter wants to know the population average μ. He observes that

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{100} x_i = 0.513 \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{100}(x_i^2 - \bar{x})^2 = 0.089$$

d) Write a 95% confidence interval for μ.($z_{0.025} = 1.96$)

The confidence interval is

$$\left[\bar{x} - \frac{1.96s}{\sqrt{100}}, \bar{x} + \frac{1.96s}{\sqrt{100}}\right] = [0.455, 0.571]$$

The experimenter needs to know μ with a confidence level of 95% and a precision of 0.01 (*i.e.* he wants a CI of size at most 0.02).

e) Is the sample he has large enough? If not, how large should the sample be?

The precision achievable with the present sample is just 0.058. He needs a sample of size

$$N = \left(\frac{1.96 * 0.298}{0.01}\right)^2 \simeq 3411$$

3) For theoretical reason we expect that the level of the water in a given river varies from day to day and is distributed according to the following p.d.f.:

$$f(x) = \begin{cases} 0 & x < 0 \\ \\ \lambda^2 x e^{-\lambda x} & x \geq 0 \end{cases}$$

Let X_i, $i = 1, \ldots, n$ be a random sample of size n for this problem, *i.e.* the p.d.f. of X_i is $f(x_i)$ and the X_i are independent.

a) Write the joint p.d.f. of the sample $F(x_1, \ldots, x_n, \lambda)$.

$$F(x_1, \ldots, x_n, \lambda) = \prod_{i=0}^{n} f(x_i, \lambda) = \lambda^{2n} \prod_{i=1}^{n} x_i e^{-\lambda \sum_{i=1}^{n} x_i}$$

b) Derive a MLE for the parameter λ.

You have to solve

$$\frac{d}{d\lambda} F(x_1, \ldots, x_n, \lambda) = 0$$

that implies

$$2n\lambda^{2n-1} \prod_{i=1}^{n} x_i e^{-\lambda \sum_{i=1}^{n} x_i} - \lambda^{2n} \left(\sum_{i=1}^{n} x_i \right) \prod_{i=1}^{n} x_i e^{-\lambda \sum_{i=1}^{n} x_i} = 0$$

The solution is:

$$\lambda = \frac{2n}{\sum_{i=1}^{n} x_i} = \frac{2}{\bar{x}}$$

c) (**Bonus**) Is the above estimator unbaised?

No, because

$$E\left(\frac{2}{\bar{X}}\right) \neq \frac{2}{E(\bar{X})}$$