STAT 321 / ELEC 321

HOMEWORK 2

Problems marked with (*) have a numerical component. For these problems, computations can be done using R or Matlab. Please, submit a copy of your computer script and display your results using tables, pictures, etc. when convenient.

Problem 1*: $10,000$ independent items may be checked using a non-destructive test. The test sensitivity and specificity are 0.90 and 0.98, respectively. Each test costs \$5 and the test can be applied to each item individually or to several items pooled together. It is known that on average a fraction $p = 0.01$ of the items are defective. The items can be pooled into k groups of size m. If a pool fails the test, then each item in that pool is tested individually. Consider the following pooling strategies:

k (number of pools)	m (pool size)
1	1000
2	500
5	200
10	100
20	50
40	25
50	20
100	10
125	8
200	5

a) Let X_m, represent the testing cost if we use pools of size m. Calculate the mean and the standard deviation for X_m, $m = 5, 8, ..., 1000$

b) Derive the random variable, T_j, $j = 1, 2, ..., 10$, that represents the total testing cost for each of the 10 strategies described above. Calculate the mean and the standard deviation for T_j, $j = 1, 2, ..., 10$.

c) What is the best strategy (among the 10 considered above) from the expected cost point of view?

Problem 2: Consider a sequence of independent trials with identical probability $p = 0.10$ of "success".

(a) Let S_i be the time of the i^{th} success and T_j the time of the j^{th} failure. Show that

$$P(T_j < S_i) = P(Bin(i + j - 1, p) \geq i)$$

where $Bin\,(i+j-1,p)$ represents a binomial random variable with $i+j-1$ trials and probability p of success.

(b) Calculate $P\,(T_j < S_i)$ for the cases $(i,j) = (1,2)\,,(2,1),(5,7),(7,5)$.

(b) Suppose that trials are continued until we obtain 20 successes. Estimate, using simulation, the expected value and the standard deviation of the number of failures.

Problem 3: Suppose that number of traffic accidents in a city follows a Poisson distribution with rate $\lambda = 5$ per day.

(a) What is the expected number of accidents in a given week? The variance?

(b) What is the probability of more than 40 accidents in a given week?

(c) What is the probability that the waiting time for the next accident is less than 4 hours?

(d) What is the expected waiting time (in hours) for the fourth accident?

Problem 4: (Method of Moments Estimation) Suppose a random variable X has distribution $F\,(x)$, which depends on unknown parameters $\theta_1,...,\theta_m$. Suppose that we have independent measurements of X, denoted $X_1, X_2, ..., X_n$. A simple method for estimating $\theta_1,...,\theta_m$ is known as *"the method of moments"* which are the solution $\widehat{\theta}_1,...,\widehat{\theta}_m$ to the simultaneous equations

$$\frac{1}{n}\sum_{i=1}^{n} X_i^m = E\,(X^m) = g_m\,(\theta_1,...,\theta_m)$$

Apply the method of moment to estimate the unknown parameters values in the following situations.

a) The voltage of a given electrical circuit is independently measure 15 times, resulting in

$$\overline{x} = \frac{1}{15}\sum_{i=1}^{15} x_i = 11.96 \text{ volts}$$

$$sd = \sqrt{\frac{1}{15}\sum_{i=1}^{15}(x_i - \overline{x})^2} = 0.21 \text{ volts}$$

If the voltage is modeled as a normal random variable with mean μ and variance σ^2. Estimate the true value for the voltage μ. Estimate the standard error of your estimate, i.e. $\sigma/\sqrt{15}$.

b) Suppse now that the voltage X is modeled as a Gamma random variable with density

$$f\,(x) = \frac{\lambda^\alpha}{\Gamma\,(\alpha)} x^{\alpha-1} e^{-\lambda x},\ x > 0.$$

b1) Show that the moment generating fucntion for X is

$$M(t) = \left(1 - \frac{1}{\lambda}\right)^{-\alpha}, \ t < \lambda$$

b2) Calculate the method of moment estimates $\widehat{\lambda}$ and $\widehat{\alpha}$.

b3) **Bonus:** Estimate the standard errors for $\widehat{\lambda}$ and $\widehat{\alpha}$ using the parametric bootstrap: generate 1000 samples of size 15 from a Gamma$\left(\widehat{\alpha}, \widehat{\lambda}\right)$ and compute the bootstrap estimates $\left(\widehat{\alpha}_b, \widehat{\lambda}_b\right)$ for $b = 1, 2, ..., 1000$. These values emulate the distribution of $\widehat{\alpha}_b$ and $\widehat{\lambda}_b$.

Problem 5: Let U be a uniform random variable on the interval $(0, 1)$. Let

$$X = -\ln(1 - U)/\lambda$$

(a) Show that $F_X(x) = 1 - e^{-\lambda x}$, $E(X) = 1/\lambda$ and $Var(X) = 1/\lambda^2$.

(b) Set now

$$Y = (X - 1/2)^2.$$

What is the range of Y? Derive the probability density anction (pdf) and cumulative distribution function (cdf) for Y. Calculate the mean, median and standard deviation of Y.

Problem 6: Suppose that the lifetime Y of a system has failure rate

$$h(y) = (y - 5)^2, \quad 0 < y < 10$$

(a) Does this system gets weaker or stronger as it ages?
(b) Find the distribution function and density function for Y.
(c) Find the median life of the system, that is the value m such that $F(m) = 1/2$.

Problem 7: A large group of students took a test in Stats and the final grades have a mean of 70 and a standard deviation of 10. If we can approximate the distribution of these grades by a normal distribution, what percent of the students

a) scored higher than 80?
b) should pass the test (grades ≥ 60)?

c) should fail the test (grades < 60)?

Problem 8: An article reports that 30% of 100 watt GE light bulbs run at at least 105 Watts, and that 10% run at at least 110 Watts. If wattage is normally distributed, what are the mean and variance?

Problem 9: The thickness of silicon wafers is normally distributed with mean 1mm, standard deviation 0.1mm. A wafer is acceptable if it has thickness between 0.85 and 1.1.

a) What is the probability that a wafer is acceptable?

b) If 200 wafers are selected, estimate the probability that between 140 and 160 wafers are acceptable.

Problem 10*: (i) Show that if $U \sim Unif\,(0,1)$ and $F\,(x)$ is invertible [that is, $F^{-1}\,(\alpha)$ is well defined for all $0 < \ \alpha \ < 1$] then

$$P\left(F^{-1}\,(U) \leq x\right) = F\,(x)\,, \text{ for all } x$$

$$Y = F_X^{-1}\,(U)$$

has distribution function $F(y)$. That is, show that $P(Y \leq y) = F(y)$.

This technique can be used to simulate engineering processes with random components. First generate $U \sim Unif\,(0,1)$ and set $X = F^{-1}\,(U)$.

(ii) Generate a sample of 1000 independent Pareto random variables with cdf

$$F\,(x) = 1 - \left(\frac{1}{x}\right)^5, \quad x > 1. \tag{1}$$

(iii) Display your sampling results using a histogram (e.g. use the command **hist** in R). Compare this histogram with the Pareto density $f\,(x) = F'\,(x)$ (iv) Use a quantile-quantile plot (a q-q plot) to check if your sample seems to come from the Pareto distribution (1). **Hint:** a q-q plot is a plot of a set of theoretical quantiles (x-axis) versus the corresponding set of empirical quantiles. If the sample comes from the theoretical distribution, the q-q plot will approximately follow a straight line. Given $0 < \alpha < 1$, the theoretical $\alpha-$quantile, $q\,(\alpha)$ for the Pareto distribution (1) satisfies the equation

$$F\,(q\,(\alpha)) = \alpha.$$

That is, $q\,(\alpha)$ is obtained from the equation

$$1 - \left(\frac{1}{q\,(\alpha)}\right)^5 = \alpha.$$

Notice that $P\left(X \leq q\left(\alpha\right)\right) = \alpha$. The empirical $\alpha-$quantile $\widehat{q}\left(\alpha\right)$ for your sample $\mathbf{x} = \left(x_1, x_2, ..., x_{1000}\right)$ is a number such that $\alpha 100\%$ of the sample values do not excede $\widehat{q}\left(\alpha\right)$. The empirical quantile, $\widehat{q}\left(\alpha\right)$, may be obtained using the R-function **quantile**($\mathbf{x}$,α).

You may use the grid $\alpha = 0.01, 0.02, ..., 0.99$ for your q-q plot.