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*: 1000 independent items may be checked using a non-destructive test. Each test costs \$5 and the test can be applied to each item individually or to several items pooled together. The pool fails the test if one or more of the items is faulty. It is known that on average a fraction $p = 0.01$ of the items will fail the test. 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) Describe the random variable, X_i, $i = 1, 2, ..., k$, that represents the testing cost for a pool of size m. Calculate the mean and the standard deviation for this random variable. **Note:** "Describing" in this case means giving the random variable's possible values and respective probabilities.

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: 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 3: 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

$$Y = e^{\lambda X},$$

What is the range of Y? Derive the probability density anction (pdf) and cumulative distribution function (cdf) for Y.

Problem 4: Let Z be the standard normal random variable with density

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\left(z^2/2\right)}, \qquad -\infty < z < \infty$$

(a) Show that

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(z^2/2\right)} dz = 1$$

HINT: Notice that

$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\left(x^2+y^2\right)/2} dx dy$$

and use the polar coordinate transformation

$$x = r\cos(v)$$
$$y = r\sin(v)$$

with $r > 0$ and $-\pi < v < \pi$.

(b) Show that $E(Z) = 0$ and $Var(Z) = 1$.

HINT: Notice that

$$\varphi'(z) = -z\varphi(z)$$

and

$$z\varphi'(z) = -z^2\varphi(z).$$

Also recall that $\int udv = uv - \int vdu$ [integration by part].

(c) Let

$$X = \sigma Z + \mu, \qquad 0 < \sigma < \infty \quad , \quad -\infty < \mu < \infty$$

What are the mean and variance of X ?

(d) Derive the distribution function and failure rate function for

$$Y = |Z|$$

Hint: Let $\Phi(z) = P(Z \leq z)$. Show first that $F_Y(y) = 2\Phi(y) - 1$.

Problem 5: Suppose that the lifetime Y of a system has failure rate

$$h(y) = 2y, \quad y > 0$$

(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 6: In an industrial process the diameter of a ball bearing, 5 ± 0.02 cm, is an important specification. Ball bearings falling outside this range are scrapped. Suppose the diameter of a ball bearing has a normal distribution with mean 5 cm and standard deviation σ cm. For what values of σ at most 1% of the ball bearings will be scrapped?

Problem 7: The temperature reading from a thermocouple placed in a constant-temperature medium is normally distributed with mean μ, the actual temperature of the medium, and standard deviation equal to 0.1μ. For what values of μ there is a probability larger than or equal to 0.95 that the reading is within 0.1^o from μ?

Problem 8: Suppose the continuous random variable X has an invertible cdf $F_X(x)$. Show that the new random variable U defined as

$$U = F_X(X)$$

has uniform distribution on the interval $(0, 1)$.

Problem 9*: A very useful result: suppose that $F_X(x)$ is as in Problem 8 and that U has uniform distribution in the interval $(0,1)$. (i) Show that the random variable

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

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

NOTE: This result can be used to generate random variables with distribution F_X as follows: First generate $U \sim Unif(0,1)$. Then set $X = F_X^{-1}(U)$. This technique can be used to simulate engineering processes with random components.

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

$$F(x) = 1 - \left(\frac{1}{x}\right)^4, \quad x > 1. \tag{1}$$

(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

$$1 - \left(\frac{1}{q(\alpha)}\right)^4 = \alpha.$$

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

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

Problem 10: Let X_1, X_2 and X_3 be three independent normal random variables with means 90, 100 and 110, and variances 10, 12 and 14, respectively. Compute (using R or Matlab)

$$P\left(-9 \leq X_2 - \frac{1}{2}(X_1 + X_3) \leq 9\right)$$

Problem 11: Let X be a random variable with mean μ and variance σ^2.

(a) (Chevychev's Inequality) Show that

$$P\left(|X-\mu| \leq \varepsilon\right) \geq 1 - \frac{\sigma^2}{\varepsilon^2}, \qquad \text{for all } \varepsilon > 0.$$

Hint: Notice that

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

$$\geq \int_{-\infty}^{\mu-\varepsilon} (x-\mu)^2 f(x)\, dx + \int_{\mu+\varepsilon}^{\infty} (x-\mu)^2 f(x)\, dx$$

(b) Let $X_1, X_2, ..., X_n$ be independent measurements of the random variable X. Let

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Show that

$$E\left(\overline{X}\right) = \mu \quad \text{and} \quad Var\left(\overline{X}\right) = \frac{\sigma^2}{n}$$

(c) Use the results in (a) and (b) to show that

$$\lim_{n\to\infty} P\left(\left|\overline{X}-\mu\right| \leq \varepsilon\right) = 1, \qquad \text{for all } \varepsilon > 0.$$

Briefly discuss why this result proves the Law of Large Numbers.