1) The following numbers x_i, $i = 1, \ldots, 18$, represent a sample of size $n = 18$ from a given population.

2.1389	2.8132	2.4451	2.4660	2.6038	2.4186
3.8592	2.1988	2.3529	2.2028	2.7468	1.5104
2.1987	2.5252	2.8462	2.2722	2.2026	2.0153

a) Compute the sample median and fourth spread and find eventual outliers.

The data, once ordered, are:

1.5104	2.0153	2.1389	2.1987	2.1988	2.2026
2.2028	2.2722	2.3529	2.4186	2.4451	2.4660
2.5252	2.6038	2.7468	2.8132	2.8462	3.8592

so that

$$\tilde{x} = (2.3529 + 2.4186)/2 = 2.3858$$
$$lf = 2.1988 \quad uf = 2.6038 \quad fs = 2.6038 - 2.1988 = 0.4050$$

Since $uf + 1.5 * fs = 3.2113$ *and* $lf - 1.5 * fs = 1.59130$ *we have that* 1.5104 *and* 3.8592 *are outliers. Finally since* $uf + 3 * fs = 3.8188$ *we have that* 3.8592 *is an extreme outlier.*

b) Knowing that $\sum_{i=1}^{18} x_i = 43.8166$ and , $\sum_{i=1}^{18} x_i^2 = 110.5081$ compute the sample mean and variance.

We have

$$\bar{x} = 43.8166/18 = 2.4343 \qquad \sigma_x^2 = \frac{1}{17}\left(110.5081 - \frac{43.8166^2}{18}\right) = 0.2263$$

d) Draw a box plot of the data.

2) The number of cars that arrive at a control station every day is described by a random variable X with a Poisson p.d.f. with parameter 10, *i.e.* $P(X = x) = \frac{10^x}{x!}e^{-10}$. Assume that 40% of all the cars that arrive need service.

a) Find the expected value and variance of the number of cars that arrive at the control station every day.

$$E(X) = 10 \qquad V(X) = 10$$

b) Find the probability that exactly N cars arrive and exactly n of these cars need service.

The required probability is:

$$P(N \text{ cars arrive } \& \ n \text{ cars need service}) =$$
$$P(N \text{ cars arrive})P(n \text{ cars need service } |N \text{ cars arrive})$$

We have

$$P(N \text{ cars arrive}) = \frac{10^N}{N!}e^{-N}$$
$$P(n \text{ cars need service}|N \text{ cars arrive}) = \binom{N}{n} 0.4^n 0.6^{N-n}$$

so that the probability is:

$$P(N \text{ cars arrive } \& \ n \text{ cars need service}) = \frac{10^N}{N!}e^{-N}\binom{N}{n} 0.4^n 0.6^{N-n} =$$
$$= e^{-10}\frac{4^n 6^{N-n}}{n!(N-n)!}$$

2) Continued

c) **Bonus** Prove that the number of cars that need service that arrive in a given day is described by a r.v. Y with Poisson distribution with parameter 4.

We have to compute $p(y) = P(Y = y)$. *This is given by:*

$$P(Y = y) = \sum_{N=y}^{\infty} P(N \text{ cars arrive \& } y \text{ cars need service}) =$$

$$= \sum_{N=y}^{\infty} e^{-10} \frac{4^y 6^{N-y}}{y!(N-y)!} = e^{-4} \frac{4^y}{y!}$$

c) Using the result of points b) and c) find the probability that exactly N cars arrived in a given day given that exacly n cars needing service arrived that day. Interpret your result in term of the number of car not needing service that arrive in a day.

$$P(N \text{ cars arrive } | n \text{ cars need service }) =$$

$$= \frac{P(N \text{ cars arrive \& } n \text{ cars need service })}{P(n \text{ cars need service })} =$$

$$= e^{-10} \frac{4^n 6^{N-n}}{n!(N-n)!} e^4 \frac{n!}{4^n} = e^{-6} \frac{6^{N-n}}{(N-n)!}$$

Since $N - n$ *is the number of car that do not need service that arrive that day, yhis result tell has that the number of car that do not need service that arrive in a given day is a Poisson variable with parameter 6.*

3) In Atlanta there are 2,000,000 families. Among them 40,000 do not report correctly their incomes. The IRS select a sample of 200 families and controlls their tax returns. Let X be the number of incorect reports among these 200.

a) What is the probability distribution of X? Write a formula for the probability that $X = 4$.

X is an hypergeometric r.v. with $N = 2.000.000$, $M = 40.000$ and $n = 200$. We than have:

$$P(X = 4) = \frac{\binom{40,000}{4}\binom{1,960,000}{196}}{\binom{2,000,000}{200}}$$

b) Use a binomial approximation to compute the average and variance of X. Justify the approximation.

Since $200 << 40,000$ and $200 << 2,000,000$ awe can use a binomial approximation. X has an approximate p.d.f of a binomial with parameters $p = 0.02$ and $n = 200$. This implyes that:

$$E(X) = 200 \cdot 0.02 = 4 \qquad V(X) = 200 \cdot 0.02 \cdot 0.98 = 3.92$$

4) Let X be a continuous r.v. with p.d.f. $f(x)$ given by:

$$f(x) = \begin{cases} \frac{3}{x^4} & \text{if } x > 1 \\ 0 & \text{otherwise} \end{cases}.$$

Compute:

a) The expected value and variance of X.

$$E(X) = \int_1^\infty x \frac{3}{x^4} dx = \int_1^\infty \frac{3}{x^3} dx = -\frac{3}{2} x^{-2} \Big|_1^\infty = \frac{3}{2}$$

$$E(X) = \int_1^\infty x^2 \frac{3}{x^4} dx = \int_1^\infty \frac{3}{x^2} dx = -3x^{-1} \Big|_1^\infty = 3$$

so that $E(X) = 3/2$ *and* $V(X) = 3 - (3/2)^2 = 3/4$.

b) The c.d.f. of X and the $100p$-percentile.

The c.d.f. is given by:

$$F(x) = \int_1^x \frac{3}{y^4} dy = -y^{-3} \Big|_1^x = 1 - \frac{1}{x^3}$$

The 100p-percentile $\eta(p)$ *satisfyes:*

$$p = F(\eta(p)) = 1 - \frac{1}{\eta(p)^3}$$

so that

$$\eta(p) = \sqrt[3]{\frac{1}{1-p}}$$