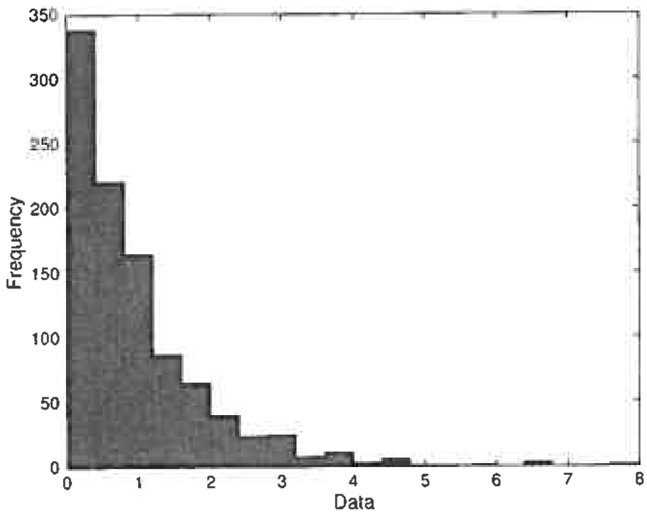
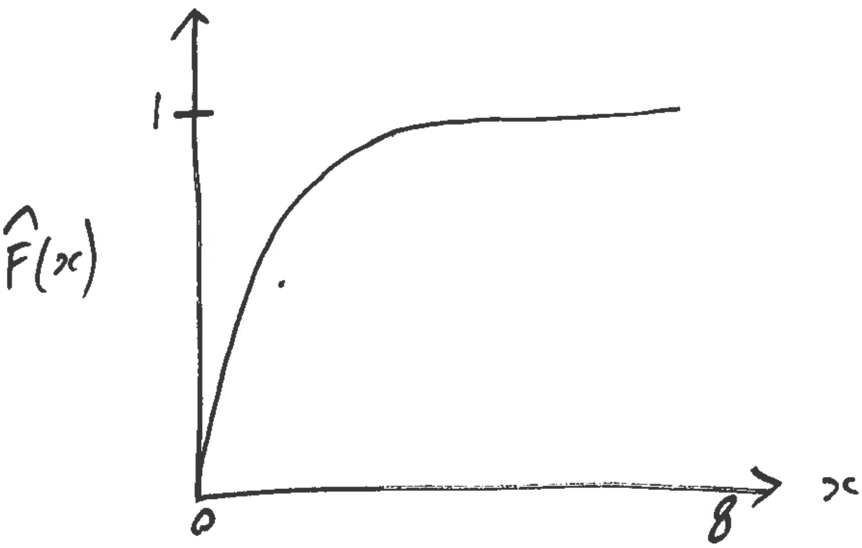
ENGR123 TEST, October 10, 2016

(The test is out of a total of 35 marks-possibly useful fomulae are on the last page)

1. DATA, SUMARY STATISTICS AND GRAPHICS **(5 marks)**

Consider the histogram of a data set given below. Sketch an empirical cumulative distribution function (ECDF) which roughly corresponds to this data set. Show axis labels and scales on your sketch.





2. DISCRETE RANDOM VARIABLES **(10 marks)**

(a) **(7 marks)** The random variable, X, measures the number of times a search algorithm correctly identifies the global maximun of a function. It is suggested that a binomial model (Bin(n,p) where n=20) could be used for X.

(i) What are the four assumptions needed for X to be well modelled as a binomial random variable?

- 20 fixed trials
- binary trials
- indept. trials
- constant prob.

(ii) Find $E(6X+4)$

$6E(X) + 4$

$6 \cdot 20p + 4$

$120p + 4$

(iii) Find $Var(30-4X)$

$4^2 Var(X)$

$16np(1-p)$

$320p(1-p)$

(iv) If $p = 0.95$, what is the probability the algorithm finds all 20 global maxima?

$p^{20} = 0.358$

(v) If $p = 0.95$, what is the probability the algorithm finds none of the global maxima?

$(1-p)^{20} = 9.5 \times 10^{-27}$

()

(b) **(3 marks)** A hospital is assessing whether it has sufficient resources to cope with the demands on its Accident and Emergency Department (A&E). A simulation exercise is planned to investigate what would happen if it didn't increase the number of available beds over the next 5 years. In order to perform the simulation, a model is needed for the number of patient arrivals at A&E per week. One consultant proposes the Poisson model. Explain why the Poisson model is unlikely to be a good model.

Constant rate is very unlikely
ie quiet times – middle of the night
busy times – after sports
– after the pubs close

3. TYPES OF RANDOM VARIABLES **(5 marks)**

Are the following random variables CATEGORICAL, ORDINAL, MEASUREMENT-DISCRETE or MEASUREMENT-CONTINUOUS?

- The error in metres of a GPS location measurement. CTS (OK if DISC. MEAS.)
- The CPU time taken to perform 10^6 simulations. CTS
- The number of blocked calls experienced in a phone network in a one month period. DISCRETE
- The maximum educational level reached by a job applicant. ORDINAL
- The brand of the most expensive coffee on sale at a supermarket. CATEGORICAL

4. CONTINUOUS RANDOM VARIABLES **(15 marks)**

(a) **(8 marks)** Let the random variable, X, have the probability density function, $f(x)$, where:

$$f(x) = A + \frac{x}{2}, \quad 0 \le x \le 1,$$

and $f(x) = 0$ for $x < 0$ and $x > 1$. Answer the following questions:

(i) Find A.

$$\int_0^1 A + x/2 \, dx = 1$$

$$\left[Ax + \frac{x^2}{4}\right]_0^1 = A + \frac{1}{4} = 1 \Rightarrow A = 3/4$$

(ii) Find the CDF, $F(x)$.

$$\int_0^x 3/4 + u/2 \, du = \left[3/4\, u + u^2/4\right]_0^x = 3/4\, x + x^2/4$$

$$F(x) = \begin{cases} 1 & x > 1 \\ 3/4\, x + x^2/4 & 0 \le x \le 1 \\ 0 & x < 0 \end{cases}$$

(iii) Find $P(0.2 < X < 0.5)$.

$F(0.5) - F(0.2)$

$(3/4(0.5) + 0.5^2/4) - (3/4(0.2) + 0.2^2/4)$

0.2775

(iv) Find $E(X)$.

need $\int_0^1 3/4\, x + x^2/2 \, dx = 3/4\, x^2/2 + x^3/6 \Big]_0^1$

$= \frac{3}{8} + \frac{1}{6}$

$\frac{9}{24} + \frac{4}{24} = 13/24$

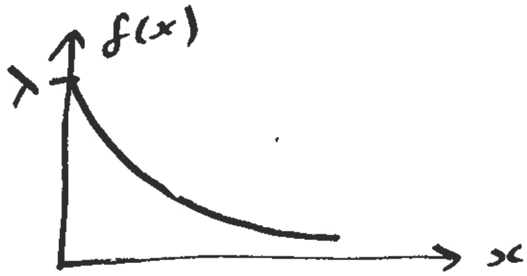
()

(b) **(7 marks)** Let the random variable, X, have an exponential distribution so that

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0,$$

and $f(x) = 0$ for $x < 0$. This model is used by a telecom provider for the times between service requests for YouTube sessions. (in seconds)

(i) Sketch the PDF of X.



(ii) Find the probability that the gap between YouTube sessions is more than one minute.

$$P(X > 60) = \int_{60}^{\infty} \lambda e^{-\lambda x} dx$$

$$= \left[-e^{-\lambda x}\right]_{60}^{\infty}$$

$$= 0 - (-e^{-60\lambda})$$

$$= e^{-60\lambda}$$

(iii) Explain the link between the Poisson model and the exponential model in terms of events occurring in a continuum.

$X \sim Po(\ )$ counts the no.

$Exp(\ )$ models the IATs

Assuming events occur singly, independently, const. rate.

Possibly useful formulae

If X is a binomial random variable with parameters n and p then:

- $E(X) = np$
- $Var(X) = np(1-p)$
- $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

****