ERG2040A Probability Models and Applications
Example Sheet 3

1. (The Geometric Distribution) Find the mode, median, mean and variance of the geometrically distributed random variable T, viz. $P(T = k) = q^{k-1}p, \forall k \in \mathcal{N}$ where $p + q = 1$ and $p, q > 0$. Show that the geometric distribution is memoryless, i.e. $P(T = k + l \,|\, T > l) = P(T = k)$.

2. (The Exponential Density) The exponential density function, viz.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

where λ is a positive constant, is commonly used in modelling first occurrence and inter-arrival times of unscheduled rare events (particle emissions, phonecalls, valentines, disasters, etc.; c.f. geometric distribution).

   (a) Find its mode, median, mean and variance from first principle.

   (b) Find its generating function, and thereby its skewness and kurtosis.

   (c) Show that the exponential density is memoryless. Can you express the memoryless property as an independence relation between appropriately defined random variables?

3. (Binomial and Poisson Distributions) It is instructive to consider a binomial distributed random variable $S \sim b(n, p, k)$ as the number of successes out of n Bernoulli trials with success rate p.

   (a) Find the mode, median, mean and variance of S.

   (b) Point events (e.g. births, deaths, arrivals) occur randomly in the interval $(-T/2, T/2)$ such that (i) the number of point events in any two non-overlapping sub-intervals are independent, and (ii) the average rate is uniform and equals λ per unit length. By treating the resulting *Poisson points* as successes among a large number of Bernoulli trials conducted uniformly in the interval, show that the point spacings are exponentially distributed. Derive also the resulting *Poisson distribution* of X, the total number of points on the interval. Find its mode, median, mean and variance.

   (c) Learn Matlab and plot $b(n, p, k)$, its Poisson approximation and the percentage error for the following sets of (n, p) values: (i) (100,.1); (ii) (100,.01); (iii) (1000,.1).

4. (Justifying the Poisson Assumption) The number of fouled toilet incidents in the HSH Engineering Building is found to follow a Poisson distribution with a mean rate of once a day.

   (a) What is the chance that more than two foul toilets are found on a particular day?

   (b) Which do you think explains the Poisson property better: a small number of frequent culprits or a large number of occasional absentminded users?

5. (Scalar Quantization) A random analogue signal is sampled and uniformly quantized to N bits. Assuming the analogue samples are uniformly distributed over the dynamic range, show that the average error is 0 while the root mean squared (RMS) error is inversely proportional to 2^N. Convince yourself that as N increases, the uniformity assumption becomes less essential for the results to hold.

6. (Variance) X and Y are two random variables with means μ_X and μ_Y, and variances σ_X^2 and σ_Y^2 respectively.

   (a) Derive the variances of cX, cXY and $cX + dY$ where c, d are constants and X, Y are independent.

   (b) Construct an example for which the variance of $X + Y$ is smaller than both σ_X^2 and σ_Y^2.

7. (N draws with replacement from a set of N) A generous professor teaches a class of N and gives a random student a red packet in each lecture. What is the expected number of miserable students (without a single red packet) after N lectures? (Practise the indicator variable method.)

8. (Card Collection) How many cards on average need be collected for a full set of N different designs available with equal probabilities?

9. (Summing a Random Number of Random Variables) What is the expectation of $S = X_1 + X_2 + \ldots + X_N$ where the X_i's are uniformly distributed in $[0, 1]$ and N is the the outcome of a fair die roll?

10. (Income Distribution and Inequality) Assume the proportion of the working population in Hong Kong with income more than HK\$$x$ is well approximated by the Gamma distribution

$$G(x; \lambda, n) = \begin{cases} e^{-\lambda x} \sum_{k=0}^{n-1} (\lambda x)^k / k! & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

   where λ, n are positive constants.

   A recent survey shows that while the average income has remained constant at HK\$12k per month, the modal income has dropped from HK\$9k in 1997 to HK\$8k in 1999 due to economic setback.

   (a) Plot the two income distributions using Matlab.

   (b) Assuming the working population is three millions, how many more people are now below the HK\$5,000 per month mark?

   (c) Estimate the change in the median income.

   (d) Estimate the values of the Gini index before and after the setback.

Wai-Yin Ng
Febrary 2000