Tuesday, 29th April 2014
2.00 pm – 3.30 pm

EXAMINATION FOR THE DEGREES OF M.A., M.SCI AND B.SC. (SCIENCE)

Probability

Hand calculators with simple basic functions (log, exp, square root, etc.) may be used in examinations. No calculator which can store or display text or graphics may be used, and any student found using such will be reported to the Clerk of Senate".

Note: Candidates should attempt THREE out of the FOUR questions. If more than three questions are attempted please indicate which questions should be marked; otherwise, the first three questions will be graded.

CONTINUED OVERLEAF/

The Standard Normal Distribution Function

When Z is a N(0,1) random variable, this table gives $\Phi(z) = P(Z \leq z)$ for values of z from 0.00 to 3.67 in steps of 0.01.

When $z < 0$, $\Phi(z)$ can be found from this table using the relationship $\Phi(-z) = 1-\Phi(z)$.

z	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9999		

CONTINUED OVERLEAF/

1. **(a)** Suppose that the continuous random variable U has the Beta distribution with probability density function

$$f_U(u) = \frac{(m+n-1)!}{(m-1)!(n-1)!} u^{m-1}(1-u)^{n-1}, \qquad 0 < u < 1,$$

where m and n are positive integers. Derive E(U) and var(U).

You may use without proof the result that, for non-negative integers r and s,

$$\int_0^1 u^r (1-u)^s \, du = \frac{r!\, s!}{(r+s+1)!}.$$

[7 MARKS]

(b) Suppose that the continuous random variables X and Y have joint probability density function

$$f_{XY}(x,y) = \begin{cases} 60x^2 y, & 0 < x < 1,\ 0 < y < 1,\ 0 < x + y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Derive the marginal probability density functions of X and Y. Use the results proved in part (a) to obtain their expected values and variances. Find the correlation between X and Y.

[13 MARKS]

CONTINUED OVERLEAF/

2. Suppose that the discrete random variables X and Y are independent with $X \sim \text{Bi}(n, \theta)$ and $Y \sim \text{Bi}(m, \theta)$, where n and m are positive integers and θ is a value in the range (0, 1). Let $Z = X + Y$.

(i) Write down the range space, R_Z, of Z. Explain why, for any z in R_Z,

$$\mathrm{P}(Z = z) = \sum_{x=0}^{z} \mathrm{P}(X = x)\mathrm{P}(Y = z - x).$$

Hence show that the random variable Z also has a Binomial distribution, and state its parameters.

You may use without proof the result that

$$\sum_{x=0}^{z} \binom{n}{x}\binom{m}{z-x} = \binom{m+n}{z}$$

[9 MARKS]

(ii) Suppose that Z is found to have the value z, for some z in R_Z. List all the values that X can take and obtain the conditional probability mass function, $\mathrm{P}(X = x \mid Z = z)$. Identify this conditional probability distribution.

[7 MARKS]

(iii) A network consists of two sub-networks, the first consisting of 10 identical components and the second consisting of another 30 identical components. Each component has probability 0.1 of failing within one year, independently of all the other components. After one year, it is found that exactly five components have failed. Find the conditional probability that exactly two components in the first sub-network have failed.

[4 MARKS]

CONTINUED OVERLEAF/

3. A bag contains m red beads and $n - m$ blue beads (where $m \geq 1$ and $n - m \geq 1$). $k \geq 1$ beads are chosen at random without replacement from all the n beads in the bag. Let the random variable X_i ($i = 1, 2, \ldots, k$) take the value 1 if the i'th bead chosen is red and the value 0 if it is blue.

(i) For $i = 1, 2, \ldots, k$, explain why $P(X_i = 1) = \frac{m}{n}$. Find $E(X_i)$ and $\mathrm{var}(X_i)$.

[4 MARKS]

(ii) For $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, k$, where $i \neq j$, explain why

$$P(X_i = 1, X_j = 1) = \frac{m(m-1)}{n(n-1)}.$$

Justify the statement that $E(X_i X_j) = P(X_i = 1, X_j = 1)$. Hence show that the covariance between X_i and X_j ($i \neq j$) is

$$\mathrm{cov}(X_i, X_j) = -\frac{m(n-m)}{n^2(n-1)}.$$

[8 MARKS]

(iii) Let the random variable S be the total number of red beads chosen from the bag. Write S in terms of $X_1, \ldots, X_k$. Hence find $E(S)$ and show that

$$\mathrm{var}(S) = k\frac{m}{n}\left(1 - \frac{m}{n}\right)\left(\frac{n-k}{n-1}\right).$$

You may use without proof the result that, for any random vector $\underline{U}$ and vector of constants $\underline{a}$, both of length p (≥ 1),

$$E\left(\underline{a}^{\mathrm{T}}.\underline{U}\right) = \underline{a}^{\mathrm{T}}.E\left(\underline{U}\right) \qquad \mathrm{var}\left(\underline{a}^{\mathrm{T}}.\underline{U}\right) = \underline{a}^{\mathrm{T}}.\mathrm{cov}\left(\underline{U}\right).\underline{a}$$

[8 MARKS]

CONTINUED OVERLEAF/

4. **(a)** Suppose that the discrete random variable X has the Geometric distribution with probability mass function

$$P(X = x) = \theta^{x-1} (1 - \theta), \qquad x = 1, 2, \ldots .$$

Show that X has moment generating function (m.g.f.)

$$M_X(t) = \frac{(1-\theta)e^t}{1-\theta e^t}, \qquad t < -\log_e \theta .$$

Use the m.g.f. to show that $E(X) = \frac{1}{1-\theta}$ and $\text{var}(X) = \frac{\theta}{(1-\theta)^2}$.

[9 MARKS]

(b) Consider a (potentially infinite) sequence of Bernoulli trials, each with success probability $1 - \theta$. Let the discrete random variable Y be the number of trials up to and including the trial on which the n'th success occurs ($n \geq 1$). Explain why Y may be written in the form

$$Y = X_1 + X_2 + \ldots + X_n,$$

where $X_1, X_2, \ldots, X_n$ are independent random variables, each with the Geometric distribution described in part (a). Referring to the Central Limit Theorem, write down an approximation to the distribution of Y for large values of n.

[6 MARKS]

(c) In an investigation of Senile Dementia, it is required to recruit 400 elderly people from the general population as 'controls'. The investigators have access to a large pool of elderly people, through social clubs, but they believe that only 80% of all the elderly people in this pool are both suitable for inclusion in the study and willing to take part. Find the approximate probability that they will require to contact at least 520 elderly people in order to recruit 400 suitable 'controls'.

[5 MARKS]

/END OF QUESTION PAPER