Math 586. Problem Set 2
Due Feb. 13, 2009

1. The relationship between air quality and health-related problems is one that, at times, has been controversial. Suppose as a scientist, you devised an air-quality index (A) and you also examined the number of asthma cases (N) (rounded to the nearest 5) that appeared in a certain citys hospitals. Both A and N were measured daily. Based on data collected over a reasonable period of time (say several hundred days) you estimated the following joint probability function.

		A				
		1	2	3	4	5
	0	0.1	0.05	0.05	0.01	0
	5	0.02	0.05	0.1	0.02	0
N	10	0	0.03	0.15	0.08	0.01
	15	0.01	0	0.03	0.03	0.05
	20	0	0.01	0.05	0.05	0.1

   Find:

   (a) The joint probability that the index A is strictly greater than 2 and N is less than or equal to 5.

   (b) Given that A=5, find the expected value of N.

   (c) Given that A=3, find the expected value of N.

2. In the attached Table, there are some values of permeabilities (in Millidarcies) measured from cores taken in a particular region. Also shown are the corresponding natural logarithm values (you should check the logs to see if they are correct). Carry out the analyses indicated below. You can write some simple matlab programs to assist you, or use standard software packages, but it is also easy to do these "by hand".

   (a) Do boxplots and histograms for both the original 40 values and also for the natural logs. Plot the empirical distribution functions for each of these. Do you notice anything unusual in these plots? Comment on the results. For your information, the means for the data and the logs are 4.00 and 0.74 respectively.

   (b) Suppose that now you find out that the cores really didnt all come from the same region. Instead, the first 20 values (left hand columns) came from one fairly homogeneous area and the second 20 values (right hand columns) came from another homogeneous area that differs substantially from the first region. Do histogram and boxplots for the raw data for these two sets. Compare to each other and to the plots from (a). For your information, the means for the data and their logs for the first 20 obsevations are 1.73 and 0.05 respectively.

(c) Repeat part (b) using the natural logarithm values in Table 1. Comment.

(d) Find the empirical cumulative distributions for the two data sets used in (b) and (c) and plot them on the same graph. Compare the two distributions using Q-Q plots and using the Kolmogorov-Smirnov test. Comment on similarities and differences.

(e) Consider just the first 20 values. Do they appear to come from a normal distribution? What about the natural logarithms – do they appear to be normal? Justify your answers in various ways.

Permeability	Log-Permeability	Permeability	Log-Permeability
1.35	0.3	6.61	1.89
1.52	0.42	4.06	1.4
1.96	0.67	5.99	1.79
0.81	-0.21	5.97	1.79
1.86	0.62	0.91	-0.09
0.32	-1.12	1.54	0.43
0.44	-0.81	4.29	1.46
0.63	-0.46	5.98	1.79
0.36	-1.03	0.62	-0.48
0.2	-1.6	22.25	3.1
1.36	0.31	14.85	2.7
2.75	1.01	5.35	1.68
6.11	1.81	20.38	3.01
1.32	0.28	4.19	1.43
3.08	1.13	1.1	0.1
0.68	-0.39	5.29	1.67
0.7	-0.36	4.88	1.59
1.47	0.38	1.19	0.18
7.46	2.01	5.99	1.79
0.15	-1.92	4.02	1.39

3. Suppose X is a positive continuous random variable with positive density. That is, $P(X > 0) = 1$ with probability density $f(x) > 0$ for all $x > 0$. (This implies that $F^{-1}(p)$ always exists.) The expected value of X is

$$\mathbb{E}[X] = \int_0^\infty x\, f(x)\, dx$$

Let $q(p)$ be the p-th quantile of $F(x)$. Show that

$$\mathbb{E}[X] = \int_0^1 q(p)\, dp$$

(Hint: change of variables.)