Statistics 280 - Applied Probability Models - Final Examination - April 21, 1998

Instructions: Attempt all questions. Your will need your course text for this examination. You may use any books or notes. Calculators are allowed. The marks total 100. You have three hours to complete the exam.

---

1. (42 marks, 3 marks each part)

For each of the random variables X, Y described below, suggest a probability model for the distribution of the random variable (i.e. name the distribution) and suggest parameter values that would be reasonable in the context described. Justify your model choices and your suggested parameter values. Assume independence whenever necessary to relate to a named distribution. Here is a list of possible models: Bernoulli, Beta, Binomial, Cauchy, Chisquare, Continuous Uniform, Discrete Uniform, Exponential , Gamma , Geometric , Hypergeometric, Lognormal, Negative Binomial, Normal, Poisson, Weibull. See also p 358 and p 474 in your text for details.

Marking: For this question, use
correct model name 1.5 marks
correct justification of model name 0.5 marks
correct parameter values 0.5 marks
correct justification of parameter choices 0.5 marks.
(or mark out of 6 and divide by 2!)
Also, note that some justification might include the asumption of independence, but I don't require this comment .

a) One thousand cedar trees are planted in a woodlot in 1978, and in 1998 a random sample of 25 trees is selected and the heights measured. Let X represent the total height of the sampled trees, in meters.

Ans a) $N(\mu, \sigma)=N(175,2.5)$ where I guess the mean to be 7 m and the SD to be 0.5 m. The mean is 25 x 7 and the SD is $\sqrt{25}$ x 0.5. Normality from CLT.

b) In the situation described in a), let Y be the quantity $(\frac{X}{25} - \mu)/\sigma$, where μ is the mean tree height and σ is the standard deviation of tree height.

Ans b) N(0,sigma=1/5)) since SD of sample mean is $\sigma/\sqrt{25}$. Lin trans does not change normality.

c) A store clerk spends idle moments looking through the coins in the till for rare coins that can be sold to a coin collector. Only coins that can be sold for $1.00 or more are kept for this purpose. Over several weeks the clerk looks through 10,000 coins and 17 of these rare coins are found. Let X+17 be the number of coins searched until the number of rare coins found (including the 17 already found) is 25.

Ans c) Negative Binomial (r=8, p=.0017) Bernoulli trials. .0017 is the estimate of the probability of success. And 8 is the number of successes needed.

d) 100 pound bags of whole wheat (unground kernels) contain some grains that are unacceptable for making grade A flour. About 20 percent of the grains are “defective” and the weight of these defective grains tends to be less than the weight of the acceptable grains. Let X be the proportion of the bag's weight that is composed of the defective grains of wheat.

Ans d) Beta has mean about .17 (something less than .20) which must be r/(r+s) so try r=1, s=5. The case r=2, s=10 is shown on p 459 in the text, which would be OK too.

e) A sample of 25 radio stations across North America are surveyed to determine the proportion of their program time that is devoted to music. It is noted that some stations play no music at all. Let X be the number of stations that play some music.

Ans e) If one knew the number of stations sampled, Hypergeometric would be OK. This answer using a guess of N is OK. But a very adequate approx in this case is Binomial n=25 p=.95 if 1 in 20 stations is a no-music station.

f) A package of 100 firecrackers is tested by igniting a random selection of five of them. Let X be the number tested firecrackers that ignite properly.

Ans f) This is hypergeometric N=100, n=5, p = .01 since sampling is definitely without replacement. p=.01 means 1% fail to ignite.

g) In mid-August, "shooting stars" appear about 2 per hour on clear nights. Let X be the length of time in hours that you have to wait to see your first shooting star, on a clear night in August.

Ans g) Exponential with rate parameter 2. The number of shooting stars in any time interval (0,t) is Poisson so the interarrival time is Exponential with the same rate parameter.

h) Young male drivers are involved in automobile accidents at the rate of one per year. Let X be the number of accidents for such a driver in an 18 month time interval.

Ans h) Poisson with mean 1.5, since rate * time is the mean.

i) IQ tests produce a score that, over a large population, is supposedly Normal with mean 100 and standard deviation 15. Let X be the number of people tested before one achieves a score of 130 or more.

Ans i) Geometric with p=1-.97725=.02275 (.02 or .025 is OK here) since the large population allows independence to be assumed, and we have Bernoulli trials, and we are counting the number of trials til the first success.

j) Let X be the number of hours that a certain brand new car engine will last, with normal maintenance.

Ans j) Weibull or Gamma would do here, since both can have increasing failure rates (Gamma does when r>1, Weibull does when β >1.) We want the mean to be say 15 years or possibly more. For Gamma mean is r/θ and for Weibull with ν=0 it is $\alpha\Gamma(1+1/\beta)$ so r=6, θ=0.4 works for Gamma and, using Γ(n)=(n-1)! when n is an interger, α=15, β=10 should be about right since 1+1/10 =approx = 1. We do want parameter values with increasing failure rate.

k) An aircraft control computer is provided with three "mother-boards": only one is needed for normal operation but if some shock kills the one in use the next one will automatically take over. The computer will operate as long as the number of shocks is less than three. Shocks occur on average about 1 every two years. Let X be the length of time in years the computer operates without maintenance.

Ans k) Gamma ($r=3$, $\theta = 0.5$) based on Poisson Process.

l) Consider the distribution of word length in English prose. Let X be the continuous random variable that has a shape similar to the probability mass function of English word length.

Ans l) Strongly right skewed, and not exponential since 1 letter words are not common. Lognormal or Gamma are OK. Mean word length is about 5. Gamma $r=4$ $\theta=.8$ would work. Lognormal is more difficult. ln 5 is $1.6 = \xi + (1/2)\delta^2$ so $\xi=1$ and $\delta = 1.1$ would do it.

m) Take your age, multiply by your birth year, and divide by your social insurance number, and record the 10th decimal place digit. Let X be that digit. (Consider the distribution that would arise for all SFU students.)

Ans m) Discrete uniform on 0,1, 9 since any digit is equally likey to appear in that position.

n) A computer produces pseudo-random numbers that purport to be indistinguishable from independent, standard normal deviates. Consider two such to be called X_1 and X_2 (These are two random variables, not merely numbers). Let $Y=X_1/X_2$.

Ans m) Cauchy with $\theta=0$, since the inverse must have the same distribution and the distribution will be symmetric about 0.

Note: Did you remember to provide numerical values of the parameters in each case, with justifications?

2. (8 marks)

A man is trapped in a mine by a rockfall, but there are three possible escape routes. One leads to freedom, but takes 1 hour, one leads back to the rockfall, and takes two hours, and a third route takes three hours but also leads back to the rockfall. The man selects one of the three routes at random each time he sets out from the rockfall (because he does not remember the routes that were unsuccessful).

Let T be the time it takes the man to reach freedom. By conditioning, find an expression for E(T) and use it to solve for E(T).

Ans 2. $E(T) = E(T|\text{door}=1)\ 1/3 + E(T|\text{door}=2)\ 1/3 + E(T|\text{door}=3)\ 1/3$
$= 1.1/3 + (1+E(T))\ 1/3 + (2+E(T))\ 1/3 = 4/3 + (2/3)E(T)$
so $E(T) = 4$ hours on average.

3. (18 marks)

Suppose two hockey teams, A and B, are evenly matched, in the sense that, at any time during a game between the two teams, each team is equally likely to score a goal.

Moreover suppose that in games between these two teams, goals are scored at the rate of two per period, and a game is three periods.

a) What is the distribution of the total number of goals scored in a game (total goals of both teams)?

Ans a) Poisson with mean 6

b) Guess the distribution of the total number of goals scored in a game by team A. (Warning: it is not likely that you know this fact, but you should be able to guess it anyway.)

Ans b) Poisson with mean 3

c) To verify (or determine) the distribution in b), how would you do a simulation that would generate game scores in accord with the specification in the preamble of this question? You can make whatever assumptions you think are reasonable in order to get a simple model.

Ans c) Generate number of goals with Poisson 6, then, conditional on this number X use Binomial (X, 0.5) to generate score of game.

d) How would you summarize your simulation results from c)?

Ans d) A table whose rows are the the total number of goals, and whose entries are the relative frequency of the scores for A. Or a scatter plot of these numbers.

e) In a game with 7 goals, how would you simulate the actual times during the game in which the goals were scored? (Use "period" as a unit of time).

Ans e) Use uniform (0,3) and generate 7 random values from it. The sorted sample will be the goal times in order of play.

f) In a game with 7 goals, what is the distribution of the number of goals scored in the first period?

Ans f) Bin (n=7;p=1/3)

4. (9 marks)

If X_1 and X_2 are IID Exponential with rate parameter λ, then $Y = X_1+X_2$ is Gamma with shape parameter 2 and scale parameter λ. Describe three approaches to the proof of this fact. (I don't need the details - just one sentence indicating the basic technique in each case, is all that is required).

Ans 4. i) moment generating functions - product of three exponental mgfs gives gamma mgf

ii) Poisson Process relationship between Gamma and Poisson probability

$$(S_2 \leq t) \Leftrightarrow (N(t) \geq n)$$

iii) By integration of joint density function over the region Y<y and then differentiate to show gamma density.

5. (15 marks)

The Bivariate Normal distribution is a distribution for a vector (X_1, X_2). Suppose for a certain population X_1 is height in cms and X_2 is weight in kgms, and the following facts are given:

(X_1, X_2).has a Bivariate Normal distribution
mean height = 160 cm SD of height= 10 cms
mean weight= 65 kg SD of weight = 5 kg
correlation of height and weight = .7

i) What is the distribution of weight for people with height 170 cms?

ans I) Normal mean = 65+0.7*1*5 = 68.5 kg and SD = $5*(1-0.7^2)^{1/2}$ = 3.6 kg

ii) Draw a scatter diagram showing a typical sample of ten pairs of values from this population.

Ans ii) anything in the box: height in (140, 180) and weight in (55, 75) with some indication of a .7 correlation will do.

iii) Would ordinary (Euclidean) distance from the mean be a good way to describe how unusual a particular weight-height combination was? Explain.

Ans iii) No. because the part of the box above that is not occupied by points is just as close as the points in the other quadrants .

6. (8 marks)

Describe the long run behavior of the Markov Chain with the following 1-step transition matrix:

0.1	0.1	0.1	0.1	0.1	0.5
0.0	0.0	0.5	0.0	0.5	0.0
0.2	0.2	0.0	0.2	0.2	0.2
0.5	0.2	0.0	0.0	0.0	0.3
0.0	0.3	0.2	0.5	0.0	0.0
0.2	0.2	0.2	0.2	0.2	0.0

By considering likely paths for this chain, guess the long run distribution. (That is, for each of states 0,1,2,3,4,5, what will be the relative frequency of each state? Your answer does not have to be exact ly correct for full marks).

Ans 6. The chain will circulate through all the states . The long run distribution is 1/6 for each state. If the guesses are within .1 of this, give them full marks. Any 0's are badly mistaken and should get no marks for this question. Give a bonus mark if they guess 1/6 for each.