Statistics 280 - Applied Probability Models - Final Examination - Dec 11, 1999

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

---

1. (28 marks, 4 marks each part)

For each of the random variables X described below, suggest a probability model for the distribution of the random variable (i.e. name the distribution used, even if the distribution of X itself is not one of the named distributions) 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 models which includes all those that we have discussed: Bernoulli, Beta, Binomial, Cauchy, Chisquare, Continuous Uniform, Discrete Uniform, Exponential, Gamma , Geometric , Hypergeometric, Lognormal, Negative Binomial, Normal, Poisson, Weibull, Zeta.

i) Students are being selected at random from the STAT 280 class to answer questions on the readings. No student is expected to answer more than one question. Let X be the number of questions posed to female students.

ii) Students admitted to a dance-party have their hands stamped for later re-admission. Let X be the time it takes to admit 10 students just during the busiest admission time. Assume there is only one person stamping hands.

iii) In the context of example ii), a door-prize is given to every tenth admission. Let X be the number of door-prizes given out in a half-hour (during the busy time).

iv) A light-bulb has the characteristic that the longer it lasts before failing, the longer remaining lifetime it is likely to have. Let X be the total lifetime of the light-bulb.

v) As student registering for a first-year calculus course selects one of the sections at random. Let X be the course selected.

vi) A student answers questions on a multiple choice exam by selecting one of the five answer-choices at random. Let X be the number of questions the student must answer in order to get 3 correct.

vii) A random sample of 400 Canadian university students is selected to answer a survey about their financial support. Let X be the number of students in this sample receiving financial assistance from their parents.

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

Ans:

i) hypergeometric, N=45, p = 22/45 since class size is about 45 and the number of females is close to 50% of 45, and the count is for the new selections making this effectively sampling without replacement.

ii) Gamma(10, rate parameter = 15 per minute) for time in minutes. I am assuming that it arrivals are handled at the rate of one every four seconds, and that interarrival times are exponential.

iii) No of people arriving in a half hour would average 15*30=450 so (1/10)*Poisson(450) would model this well. Of course this will be very close to Normal(45,var=45) which is equally good as an answer.

iv) A weibull or a gamma have hazards that can be increasing. Gamma is easier to calibrate. In the case of gamma the mean of 1000 would be associated with an SD less than 1000 (by comparison with exponential) say 1000 and 500. Gamma mean and SD is s/lambda and sqrt(s)/lambda (see p 360) making s=4 lambda = (1/250).

v) discrete uniform on 1,2,…,9 if these integers label the sections of the course.

vi) neg binomial with p =1/5=.2 and r=3, since the questions are like bernoulli trials.

vii) Bin(400, p) where p would be about .5 if 50% of parents support their kids at university.

2. (20 marks)

i) Draw a freehand graph of a typical Poisson Process outcome for the first minute of an assumed Poisson process for car arrivals at SFU. Assume that cars arrive at the rate of one per six seconds. Be sure to label and scale your axes.

ii) Use your graph to explain the identity $S_n > t \Leftrightarrow N(t) < n$. Here, S_n is the time it takes until the nth car arrives, and $N(t)$ is the number of cars that have arrived in the time interval (0,t).

iii) What two distributions does the identity in ii) link together? (Provide parameters as well as the names.) Assume n>1 for this.

iv) What discrete-time model would you suggest to approximate the Poisson Process model described in part i). Be specific about parameter values.

Ans: iii) Poisson for N and Gamma for S
iv) Binomial with n=60 p=1/6

3. (15 marks)

A Markov Chain is used to construct a discrete time model for the number of customers in a barber shop. The shop can only hold 3 customers and if there are more arrivals when the shop is full, they will simply leave. Each minute, the state space of the Markov chain {0,1,2,3} represents the number of people in the shop, and the one-step transition matrix is

P =

.9	.1	0	0
.1	.8	.1	0
.0	.1	.8	.1
.0	.0	.2	.8

i) Comment on the behavior of the barber as suggested by this transition matrix.

ii) Suggest two methods to compute the proportion of time that the barber has no customers.

iii) Explain how you would simulate a typical hour of experience for this barber. (Be sure to refer to the probabilities in the matrix P.) You do not need any probability symbols or computer code for this - words will do.

Ans. i) He hurries when the shop is full, since the chance of finishing then is larger than when the shop is not full.
ii) n th power of transition matrix for large n and pick off 0 column value, or solve equations from Theorem 2.1.
iii) Just generate one value at a time for the appropriate row of the transition matrix, using random numbers or a computer-based discrete prob generator. (bit more detail needed).

4. (10 marks)

A deck of 4 cards is shuffled and one card dealt and recorded. After replacement of the dealt card, the shuffle and deal is repeated. Call the cards A,K,Q,J. Let X be the be the number of deals required to obtain an A, and let Y be the number of deals required to obtain a K. Compute

i) P(Y=1|X=1)
ii) P(Y=2|X=3)
iii) E(Y|X=3)

Ans i) This is 0 since if an A is first, it cannot also be a K.
ii) prob is 2/3*1/3 since must get Q or J on first one and then K on second.
iv) 1*1/3 + 2*2/9 + 4/9*(3+4) since if it is not obtained in the first two, then it must wait until 4 or more, where 4 is like the first trial of a geometric with prob 1/4.

5. (10 marks)

Problem 36 on p 389 required the derivation of the recursive formula $M_{a,b} = \frac{a}{a+b} M_{a-1,b} + \frac{b}{a+b} M_{a,b-1}$. Explain why this is true in the context of this problem.

Ans. Condition on the first ball drawn, which has probabilities a/(a+b) and b/(a+b). If a white ball is drawn first, then the average number that are left when the process stops will be Ma-1b by the definition of Mab

6. (8 marks) An experienced gambler can win a certain card game with probability 0.5 if he does not cheat, and with probability 0.7 if he does cheat. He is known to cheat about 10 percent of the time. If he wins a particular game, what is the chance that he cheated?

Ans: Use Bayes P(C|W)= .7*.10/(.7*.10+.5*.90)=.07/.52=.135 approx

7. (9 marks) The binomial and hypergeometric distributions model similar random variables since they both describe the probabilities for the number of selections of a certain kind among n selections.

i) Under what circumstances will these probabilities be very close?
ii) Show algebraically why the approximate equality in i) holds.

Ans: i) N much larger than n
ii) it depends on the idea that nCk is approx n^(n-k) when k is close to n

(KLW 99/12/10)