Prelim "Probability /Statistics " UH Mathematics

Four examples of prelim problems based on the MATH6382 probability course

Problem 1
Let S and T be two independent random variables having the same uniform distribution on the interval [0,4].
Consider the quadratic polynomial q(x)= x2 – S x + T which has random coefficients S and T.
1.1.The discriminant of this quadratic equation is D= S^2 – 4T. Compute the probability Prob(D >= 0) by
first computing p(t) = Prob(D >= 0 | T= t) for any given t in [0,4]. Interpret concretely the event {D >= 0 }.
1.2. Compute the CDF G(r) of the random variable R= S^2 and the CDF H(v) of the random variable V= - 4T.
Compute then the pdf g(r) of R and the pdf h(v) of V.
Compute the joint density function of the random vector W = (R, V).
1.3. use the relation D = R + V to compute the density function of D.

Problem 2
Let A ={1,2,3, ... ,k} be a finite set, and let Xn , n=0 ,1, 2, ... , be an ergodic Markov chain with state space A.
Denote Q(i,j) be the k x k transition matrix of the Markov chain X n .
2.1 Explain what is the stationary probability M of this Markov chain and how one can practically compute M.
2.2 From now on , assume that M is known , and that the probability distribution of X0 is equal to M.
What is then the probability distribution of Xn ?
2.3 Compute P[(Xn = 1) & (Xn+1 = k) & (Xn+2 = 1) & (Xn+3 = k)] as well as P((Xn = i) & (Xn+4 = j) for all i,j in A.

Problem 3
A cosmic ray detector is installed on an earth orbiting satellite. Let T1 < T2 < ... < Tn < ... be the successive cosmic rays arrival times recorded by the detector . Assume that these random arrival times
can be modeled by a Poisson process for which the mean rate of random arrivals per hour is an unknown parameter m .
3.1 Let K(J) be the random number of cosmic ray detected within a fixed time interval J of duration = 1
hour . Explain what is the probability distribution of K(J), the mean value and the variance of K(J)
3.2 Let J1 , J2 , ..., J400 be 400 disjoint time intervals of duration 1 hour . Denote U1 = K(J1), ... , U400 = K(J400)
the 400 random numbers of cosmic rays detected during these time intervals .
One decides to estimate the unknown parameter m by

$Z = (1/400) [U1 + U2 +... + U400]$.
Compute the mean and variance of Z . Explain why one can apply the central limit theorem to Z.
3.4 Explain what is the probability density function of the interarrival time $W_n = (T_{n+1} - T_n)$.
Compute
the CDF $F(w)$ of W_n and the 1% quantile q of W_n defined by $P(W_n \le q) = 0.01$.

Problem 4
X and Y are independent variables having the same exponential density with mean value 2
4.1 write the joint density $f(x,y)$ of (X,Y) and a double integral formula to compute $E(g(X,Y)]$ for any function $g(x,y)$
4.2 let $Z = \max(X,Y)$; give a double integral formula to compute the expected value $E(Z)$ and try to compute that double integral.

Examples of prelim problems based on the MATH 6383 statistics course

1. Suppose we have the following data, $(x_1, y_1), \ldots, (x_n, y_n)$ for some positive integer n. Here y_i's are counts of a certain events and x_i's are covariates related to y_i's. We assume that $x_1, \ldots, x_n$ are fixed, and each y_i's are assumed to follow a Poisson distribution with the rate

$$\lambda = \eta \cdot |x_i|,$$

for $\eta > 0$. We also assume the y_i's are independent.

(a) Using the *Least Squares principle*, describe the steps to estimate the parameter, η, using the data, and give forms of your estimator for the parameter. If you use a strategy other than the least squares principle, you will not get points.

(b) With your estimator from above, we estimate the rate of the Poisson distribution, say $\hat{\lambda}$. Is your $\hat{\lambda}$ unbiased? Show your work.

(c) Calculate the variance of your estimator for λ, $\hat{\lambda}$.

(d) We now estimate the parameter η using the *Maximum Likelihood Method* using the data. Give the form of the estimator for η.

2. Let $X_1, \cdots, X_n$ be iid with the pdf:

$$f(x|\theta) = \theta x^{\theta-1}, x \in (0, 1), \theta > 0.$$

(a) Find the method of moment estimator for θ.

(b) Find the MLE of θ.

(c) Prove that $-\log(X_i)$ follows an Exponential distribution with mean $\frac{1}{\theta}$.

(d) Is the MLE of θ an unbiased estimator? Calculate the expected value of the estimator to show whether it is an unbiased estimator or not, and if it is not, calculate the bias.

(e) Calculate the variance of the MLE of θ and show that the variance goes to zero as $n \to \infty$.

3. Suppose there is a mayor election in a large city. We are interested in the chance of a candidate, Mary Jane Smith, to win the election. In a recent survey performed by a renowned organization, 965 residents responded that they would vote for her among 2,125 participants. Answer the following questions with this information.

(a) Construct a 95% confidence interval for p, the proportion of voters who support Mary Jane Smith. You may use the following output from R:

```
> qnorm(0.9,0,1)
[1] 1.281552
> qnorm(0.95,0,1)
[1] 1.644854
> qnorm(0.975,0,1)
[1] 1.959964
> qnorm(0.99,0,1)
[1] 2.326348
> qnorm(0.995,0,1)
[1] 2.575829
```

(b) Suppose, now, we perform the following hypothesis test using the z test:

$$H_0 : p = 0.5 \quad vs \quad H_a : p < 0.5$$

What is the test statistic, p-value, and your conclusion of the test if you set $\alpha = 0.05$? For the p-value, you may express the value using Φ. In making a conclusion of the hypothesis test, you may use the R output given above.

(c) Suppose, now, we learned (after the election) that 55% of voters supported Mary Jane Smith. Retrospectively, we consider the power of the test done in question (b). What is the power of this test? You may express the value using Φ notation.

(d) Suppose, now, we want to do a likelihood ratio test (LRT) (using the distribution of the data themselves rather than normal approximation) with the same hypothesis given in question (b). What is the value of the LRT statistic?

4. Answer the following questions. For each question, answers should be concise.

(a) Suppose $X_1, \ldots, X_n$ are random sample from a population with mean μ and variance 1. It is known that the "best" linear estimator for μ in this case is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Here, "best" means the estimator has the smallest variance among all linear estimators. What is the variance of $\bar{X}$? Show your work.

(b) Suppose there is a small town in the middle of the desert, and the yearly rainfall amount follows an exponential distribution, with the mean rain amount of 10 inches per year. For the next 50 years, what is the probability that the minimum yearly rain amount would exceed 15 inches per year?