Math 3215: Final Exam Sample Questions

Will Perkins

April 26, 2012

1 Basic Probability

1. Write a formula for $\Pr[A \cup B \cup C]$ in terms of probabilities of intersections of events.

2. What is the difference between an outcome and an event?

3. You roll a single die, and let Y be the number that comes up (1..6 with equal probability). Then you roll Y spearate dice, and let X be the sum of the numbers that come up.

   - What is $\mathbb{E}X$?
   - Are X and Y independent?
   - What is $cov(X, Y)$?
   - What is the conditional expectation of X given that $Y \geq 3$: $\mathbb{E}[X|Y \geq 3]$?

2 General Random Variables

1. Can $cov(X, Y)$ ever be larger than $\max\{var(X), var(Y)\}$? Why or why not? Can it ever be larger than the minimum of the variances?

2. Is the CDF of a random variable always continuous? Why or why not?

3. Is the CDF always non-decreasing? Why or why not?

3 Discrete Random Variables

1. I have m marbles and I throw each one into one of n buckets with equal probability, independently at random. Let X_i be the number of marbles that land in bucket number i.

   - What is the expectation of X_1?
   - What is the distribution of X_1?
   - What is the covariance of X_1 and X_2?
   - Are X_1 and X_2 independent? Why or why not?
   - What is the distribution of $X_1 + X_2$?
   - What is the distribution of $X_1 + \cdots + X_{n-1}$?
   - What is the distribution of $X_1 + \cdots + X_n$?
   - If $m = 2n$, then what is probability that bucket 1 is empty? (lim as $n \to \infty$)

4   Continuous Random Variables

1. Let X and Y be iid unform $[0, 2]$ random variables.

   - Calculate the density functions and CDF's for $X + Y$ and $X - Y$.
   - What is $\Pr[X > 1 | X + Y \geq 1]$?
   - What is $\mathbb{E}[\min\{X, Y\}]$?

2. Let X have an exponential distribution with mean 1. What is $\Pr[3 \leq X < 7]$?

3. Let X have the CDF $F(t) = t^2$ for $t \in [0, 1]$.

   - What is $\mathbb{E}X$?
   - What is $var(X)$?
   - What is the density function for X?

5   More probability

1. If $\mathbb{E}X = 0$ and $var(X) = 4$, give an upper bound on the probability that $|X| > 3$.

2. Let $X_1, \ldots X_n$ be iid with mean 1 and variance 9.

   - What is the mean and variance of $Y = \frac{X_1 + \ldots X_n}{\sqrt{n}}$?
   - To what distribution (precisely) does Y converge?
   - What is $\mathbb{E}[Y^2]$?
   - To what distribution does $\frac{X_1 + \cdots + X_n}{n^{2/3}}$ converge? Prove it.

3. In the random graph $G(n, p)$ with $p = c/n$:

   - What is the expected number of triangles in the graph? (limit as $n \to \infty$)
   - What is the variance of the number of triangles?
   - Can you guess the approximate distribution of the number of triangles? What, roughly, is the probability that there are no triangles?

4. Does the weak law of large numbers hold if $X_1, \ldots X_n$ have pairwise 0 covariance, and each has mean 1 and variance 1 (but not identically distributed)? why or why not?

6   Statisitical Method

1. For each of the following situations, give the probability model that best fits, the statistic you would use to answer the question, and the distribtuion you would use to calculate the p-value.

   - Estimate the fraction of voters who support a certain bill.
   - Determine the relationship between income and education in the US.
   - Determine if a new drug is effective in curing a disease.
   - Determine which of 2 drugs is more effective.
   - Determine if religious affiiliation is independent from political preference.
   - Test whether the number of daily errors by a computer server follows a Poisson distribution.

2. If you expect to make one spelling error every hundred words, how many errors would you have to make in a 1200 word essay to change your mind about your error rate?

3. When testing whether a new teaching method is effective or not, describe how you would set up the experiment, and what your null and alternate hypotheses would be.

7   Estimators

1. A baseball player hits .300 for a season with 400 at bats. (i.e. 120 hits in 400 chances). Give a 95% confidence interval for his 'true' batting average.

2. You collect 100 samples of a rare flower. The average height of the sample of flowers is 10.7 inches, with a standard deviation of 1.8 inches. Give a 95% confidence interval for the true mean of this breed of flowers. Discuss any assumptions you are making.

3. Derive the maximum liklihood estimator for T for a sequence of independent random variables $X_1, \ldots X_n$ that come from a uniform $[0, T]$ distribution.

4. I have three coins, with different probabilities of heads. Coin 1 has $p = .5$, coin 2 has $p = .6$, coin 3 has $p = .8$. I then choose one coin at random: coin 1 with probability $1/2$, coin 2 with probability $1/4$ and coin 3 with probability $1/4$.

   - What is the (unconditional) probability that you flip the selected coin 10 times and get 10 heads?
   - What is the (unconditional) probability that you flip the selected coin 10 times and get 6 heads?
   - If you flip the selected coin 10 times and get 6 heads, what is the maximum liklihood estimator of the number of the coin picked? (prove it)

5. Describe precisely when you would use the Student's T-distribution in a statistical test.

6. Give a mathematical reason for why the student's t-distribution converges to the Normal distribution as $n \to \infty$.

7. Given an instance of $G(n, p)$ what is the maximum likelihood estimator for p? What is the mean and variance of this MLE?

8 Goodness of Fit

1. Set up the test, and calculate the value of the statistic (but not the p-value) to test whether the following data came from a discrete uniform distribution on $\{1, \ldots 10\}$:

1	2	3	4	5	6	7	8	9	10
15	8	5	14	6	9	17	10	8	8

2. Set up the test and calculate the statistic to test whether this data came from a Poisson distribution:

0	1	2	3	4	5	6
10	15	15	21	18	12	2

3. For the hypothesized distribution, $p_1 = .3, p_2 = .5, p_3 = .2$, and the observed data $20, 60, 20$, calculate the chi square test statistic for a goodness-of-fit test.

9 Regression

1. Write down the probability model used in a linear regression, with y_i's as the dependent variables and x_i's as the independent variable.

2. Derive the MLE for α in the above model (the constant term). What is the variance of this MLE?