# Math 3215 Final Exam, Summer 2009

July 28, 2009

**Instructions:** You will be allowed a simple calculator only – no programmable calculators allowed. You have three hours to complete the following 10 problems, where each will be given equal weight.

**1.** Define the following terms.

a. Say what it means for the events $A_1, ..., A_k$ to be independent with respect to some probability measure $\mathbb{P}$.

b. Say what it means for random variables $X_1, ..., X_k$ to be independent.

c. State the Law of Large Numbers.

d. Unbiased estimator.

e. State the "chain rule" of probability for events $A_1, ..., A_k$.

**2.** In a certain "pick 6" lottery one chooses 6 different numbers from among the numbers $1, 2, 3, ..., 42$. If all 6 of your numbers happen to match the winning 6 (in any order), then you win the "grand prize". But, in fact, if even 4 of your numbers are among 4 of the winning 6, then you win a "small prize". Assuming that every subset of 6 numbers from among 42 is equally likely to be the winning combination, determine the probability of winning the "small prize".

**3.** You have been assigned the task of writing diagnosis software for medical doctors, that given the presence of certain symptoms, must determine the likely illness. [1]

[1]Artificial intelligence, or machine learning, makes heavy use of probability and statistics. Indeed, a very common type of mathematical technique that is used is something called a "Bayesian Network", which uses much more sophisticated versions of the Chain Rule of Probability, and Bayes's Theorem, than we studied in class. Indeed, BN's are

To make things simple, suppose that the software is to only distinguish between whether a visitor to the doctor has flu type A, type B, or does not have the flu at all. Here are the given data around which you build your model:

- Only 10 percent of patients who visit the doctor actually have flu type A.
- And 20 percent have flu type B (assume the patient cannot have both type A and type B).
- Exactly 50 percent of patients with type A flu have a cough, while 40 percent of patients with type B flu have a cough.
- Exactly 20 percent of patients with a cough have type A flu.

Given that a patient does not have a cough, what will (should) the software say is the likelihood that the patient has type A flu? (Hint: Let $A$, $B$ and $C$ be the events that the patient has type A flu, type B flu, or has no flu, respectively, and then let $G$ be the event that the patient has a cough. The given information can be described in terms of conditionals like $\mathbb{P}(G|A) = 0.5$. Note that the last bullet item may be used to deduce $\mathbb{P}(G|C)$, upon applying Bayes's Theorem.)

**4.** You roll a fair 6-sided die twice.

a. Write down a sample space for this experiment, and say what the probability of each event in the sample space is.

b. Compute the probability that the sum of your two rolls is 9, and express this event as a subset of the sample space.

**5.** Prove that

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx \ = \ \sqrt{2\pi}.$$

---

the basis of such cutting-edge AI technologies as Jeff Hawkins's (founder of Palm Inc.) Numeta project, which is an outgrowth of his neuroscience theories presented in his book *On Intelligence*.

**6.** Let $T$ denote the triangle with corners at $(0,0),(0,1)$, and $(1,1)$, and define the function

$$f(x,y) \;=\; \begin{cases} x + cx^2y^2, & \text{if } (x,y) \in T; \\ 0, & \text{otherwise.} \end{cases}$$

FInd the constant $c$ that makes this function into a probability density function.

**7.** Suppose that the number of particles that a certain radioactive sample emits in any given time window obeys a Poisson process, and that the expected number of particles emitted in any given 1 second window is 3. Now suppose you have a very coarse radiation detector, which outputs a signal of $+1$ at the end of the 2 sec. tim window if it detects 3 or more particles in that time, and otherwise outputs a signal of 0 at the end of the time window. What is the expected value of the signal, for 2 second time windows?

**8.** A well-known phenomena in economics and political theory is the "wisdom of the crowd", where the crowd often can make better decisions than a single, albeit very smart, individual. [2] For this problem, you will use the "wisdom of the crowd" to compute a 95% confidence interval of a certain parameter $\mu$, to be described presently: suppose a collection of 5 randomly selected people bet on the weight of an bull at a county fair, and suppose that their guesses are $X_1, ..., X_5$, which we will treat as random variables (since the guessers were randomly chosen). Let us further make the assumptions that each $X_i$ is normally distributed with mean $\mu$ – the true weight of the bull – and variance $\sigma^2$. Suppose that the guesses of the 5 individuals were

$$1000,\ 1005,\ 995,\ 1002,\ 998.$$

Based on these observations, compute a 95% confidence interval for $\mu$. (Hint: Student-$t$.)

[2] Perhaps the first to notice the phenomena was Francis Galton, 19th and early 20th century polymath (and, regrettably, a eugenicist) and disbeliever in the power of democracy, who had once proclaimed, "the stupidity and wrong-headedness of many men and women being so great as to be scarcely credible." However, upon observing a crowd of nearly 1000 people betting on the weight of an ox, he decided to take the average of all their guesses. It turned out that although each person's individual guess was generally far off the mark, the average of *all* guesses came within 1 pound on the true weight! After that, he changed his mind about democracy.

**9.** Recently in Clayton County Georgia, a debate has been raging as to whether to raise taxes or to cut employee hours by a half-day to make up for a shortfall in its \$160 million budget, according to the Atlanta Journal Constitution (June 24, 2009). You decide to do a poll of 100 likely Clayton County voters to get their reaction, and based upon your knowledge of the voter base, you conjecture that 60% of likely voters prefer furloughs over taxes, while 40% prefer taxes over furloughs.

To test your hypothesis, you first let $p$ be the actual percent who prefer furloughs over taxes; and so, your standard hypotheses are

$$\begin{aligned} H_0 &: \quad p = 0.6 \\ H_a &: \quad p \neq 0.6. \end{aligned}$$

Upon conducting your poll of 100 randomly-selected (with replacement) Clayton County likely voters, let $X$ be the number of them who say 'furlough over taxes'. Suppose that your statistical test is such that you reject the null hypothesis whenever $X$ is outside the interval $[55, 65]$. Using the usual normal approximation for the binomial random variables with large $n$ (as afforded to us by a quantitative version of the Central Limit Theorem), determine the probability $\alpha$ of making a type I error.

**10.**

a. Compute the moment generating function for the standard normal random variable $X$, with pdf given by $f(x) = e^{-x^2/2}/\sqrt{2\pi}$.

b. Use your answer from part $a$ to find the 8th moment of $X$.