HMC Math 62- Probability & Statistics

Final: Due Friday (midnight), June 25, 2004 (3 Hour Takehome)

1 (Short answer): *(30 Points)*

(i): State the three Axioms that define a probability $P(A)$ for $A \subseteq S$ (the sample space).

(ii): State (some/any version of) the Central Limit Theorem.

(iii): The city of Townsville was attacked by 12 giant monsters in the month of February. Given this monster rate, what probability distribution might we use to model the number of monsters that will attack in the coming week? (Name a distribution and parameter value(s).)

(iv): Certain authors of statistics text books have been known to fabricate some of the alleged "real world" examples and their journal references. If a certain author fabricates, on average, two-fifths of his examples, what are the chances that at least half of the 40 examples in Chapter 4 will be fabrications? Use a normal approximation, and give a numeric value for your answer (use the normal tables to look up Φ).

(v): In poker, a "flush" is a hand where all five cards have the same suit (like Hearts). If you are dealt five cards from a standard deck, what is the probability that you get a flush?

(vi): What does it mean when we say that S^2 is an *unbiased estimator* for σ^2?
(Get it? "What does it *mean*"? Get it? Ha!)

2: *(12 Points)* Suppose that a particle starts at the origin of the real line and moves along the line in jumps of one unit (where jumps are independent). For each jump, the probability is p that it jumps one unit to the right and $1-p$ that it jumps one unit to the left. Let X_n be the position of the particle after n jumps.

Hint: You can approach this problem in one of two ways:
(1) Represent X_n as $J_1 + J_2 + \cdots + J_n$ for appropriately defined r.v.s J_i ; or
(2)Let Y be an appropriate Binomial r.v. and express X_n in terms of Y.

(i): Find $E[X_n]$ and $\mathrm{Var}(X_n)$.

(ii): If $n = 10$, what is $P(X_{10} = 4)$?

3: *(12 Points)* Suppose continuous r.v. X has density function $f(x) = \begin{cases} 1/3 & , \quad -1 \le x \le 0 \\ 2(x-1)^2 & , \quad 1 \le x \le 2 \\ 0 & , \quad \text{otherwise} \end{cases}$

(i): What is $E[X]$?

(ii): Compute $P(X > -\frac{1}{2} \mid X < 0)$.

BONUS: *(3 Points)* Compute and graph the cdf $F(X)$.

4 ("The Little Statistics Problem"): *(10 Points)* Suppose we want to perform an ANOVA test to determine if $I = 4$ types of sleep inducers have the same mean time-till-sleep. We time how long (in minutes) it takes each sleep inducer to put $J = 5$ people to sleep, and our observed data is summarized in the table below:

INDUCER	Sleeping Pills	Sloths	Hairy Math Professors	Statistics Text Books
Sample Mean	5.08	5.14	5.71	5.47
Sample SD	.463	.405	.372	.399

Use this data to test the null hypothesis "all sleep inducers are equally effective" at an $\alpha = 0.05$ significance level. Remember, when using the F-distribution table, $\nu_1 = I - 1$ and $\nu_2 = I(J-1)$.

5 ("The Big Statistics Problem"): *(36 Points)* The Splatt Dining Hall at Hairy Dude College adds the Goo© nutritional suppliment (filled with Vitamin Q) to all its food for bigger brains and pastier skin in its students. The makers of Stuph© are trying to underbid Goo© and win the nutritional suppliment contract for Splatt, but Splatt will only switch if Stuph© is found to have a higher concentration of Vitamin Q. Splatt has sent 49 samples of Stuph© off to the Chemistry Department for testing, and they have measured a sample mean concentration of .113 g/oz of Vitamin Q ("Keeping Your Geeks Sharp and Pale," *J. Improb. Stats.*, 1991: 201-207).

For parts *(i), (ii)* and *(iii)* (*not (iv)!*), assume that the underlying distribution has a standard deviation of .026 g/oz.

(i): Construct a 95% (2-sided) confidence interval for the true mean concentration of Vitamin Q in Stuph© .

(ii): Goo© is known to have a Vitamin Q concentration of .107 g/oz. Use the Chemistry Department's findings to test the hypothesis "Goo© is better". First, state mathematically what our null and alternative hypotheses are to be. Next, using the standard test statistic Z_0 for hypothesis testing on a mean, what would be our rejection region be if we wish to test our null hypothesis at a 5% significance level? What does our test conclude given the observed sample mean?

(iii): Suppose now that Stuph© actually has a Vitamin Q concentration of .116 g/oz. What were the chances (prior to running our test), that our test would fail to indicate (at a 5% significance level) that Stuph© was the better suppliment?

(iv): Suppose that another Vitamin Q suppliment, Braneslyme© , is also bidding for the contract, but only $n = 6$ samples of it were available to test. The observed Vitamin Q concentrations were:

.100 .121 .108 .103 .116 .124

Suppose that these values are coming from a normal distribution. Compute the sample mean and variance for these observations, and give a 99% confidence interval for Braneslyme© 's Vitamin Q concentration.
(Remember: small sample from normals means t-distribution!)

BONUS: *(2 Points)* What should you have written on your hand?