University of Illinois **Spring 2012**

ECE 313: Problem Set 13
LLN, CLT, and joint Gaussian distribution

Due:	Wednesday, May 2, at 4 p.m.
Reading:	*ECE 313 Notes* Sections 4.10 & 4.11.
Reminder:	The final exam will be held Monday, May 7, 1:30-4:30 p.m. Sections E&C (9 am & 10 am sections): Room 150 Animal Sciences Lab Sections D&F (12 pm & 1 pm sections): Room 116 Roger Adam Lab Two two-sided 8.5”×11” sheets of notes allowed, with font size no smaller than 10 pt or equivalent handwriting. Bring a picture ID. No calculators. The exam consists of a sampling of problems related to the lectures, course notes, and problems sets from throughout the semester. It will have about twice as many problems as an hour exam plus ten true or false questions. Material for problem sets 11-13 will be covered slightly more heavily because it wasn't covered in the hour exams. You are responsible for knowing the forms of all the key discrete and continous-type distributions listed on the inside covers of the notes (same as Appendix 6.3).

1. **[Rate of convergence in law of large numbers for independent Gaussians]**
By the law of large numbers, if $\delta > 0$ and $S_n = X_1 + \cdots + X_n$, where $X_1, X_2, \ldots$ are uncorrelated with mean zero and bounded variance, $\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n}\right| \geq \delta\right\} = 0$.

   (a) Express $P\left\{\left|\frac{S_n}{n}\right| \geq \delta\right\}$ in terms of n, δ and the Q function, for the special case $X_1, X_2, \ldots$ are independent, $N(0,1)$ random variables.

   (b) An upper bound for $Q(x)$ for $x > 0$, which is also a good approximation for x at least moderately large, is given by

   $$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \leq \int_x^\infty \frac{1}{\sqrt{2\pi}} \left(\frac{u}{x}\right) e^{-u^2/2} du = \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}.$$

   Use this bound to obtain an upper bound on the probability in part (a) in terms of δ and n but not using the Q function.

2. **[Marathon blackjack]**
In a particular version of the card game *blackjack* offered by gambling casinos[1], if a player uses a particular optimized strategy, then in one game with one unit of money initial bet, the net return is -0.0029 and the standard deviation of the net return is 1.1418 (which can be squared to get the variance). Suppose a player uses this strategy and bets \$100 on each game, regardless of how much the player won or lost in previous games.

   (a) What is the expected net gain of the player after 1000 games? (Answer should be a negative dollar amount.)

   (b) What is the probability the player is ahead after 1000 games? (Use the Gaussian approximation suggested by the central limit theorem for this and other parts below.)

   (c) What is the probability the player is ahead by at least \$1000 after 1000 games?

   (d) What value of n is such that after playing n games (with the same initial bet per game), the probability the player is ahead after n games is about 0.4?

3. **[Gaussian approximation for confidence intervals]**
Recall that if X has the binomial distribution with parameters n and p, the Chebychev inequality implies that

   $$P\{|X - np| \geq a\sigma\} \leq \frac{1}{a^2}, \tag{1}$$

[1] See http://wizardofodds.com/games/blackjack/appendix/4/

where σ^2 is the variance of X: $\sigma = \sqrt{np(1-p)} \leq \frac{\sqrt{n}}{2}$. If n is known and p is estimated by $\widehat{p} = \frac{X}{n}$, it follows that the confidence interval with endpoints $\widehat{p} \pm \frac{a}{2\sqrt{n}}$ contains p with probability at least $1 - \frac{1}{a^2}$. (See Section 2.9.) A less conservative, commonly used approach is to note that by the central limit theorem,

$$P\{|X - np| \geq a\sigma\} \approx 2Q(a). \tag{2}$$

Example 2.9.2 showed that $n = 625$ is large enough for the random interval with endpoints $\widehat{p} \pm 10\%$ to contain the true value p with probability at least 96%. Calculate the value of n that would be sufficient for the same precision (i.e. within 10% of p) and confidence (i.e. 96%) based on (2) rather than (1). Explain your reasoning.

4. **[Jointly Gaussian Random Variables I]**
   Suppose X and Y are jointly Gaussian with $\mu_X = 1$, $\mu_Y = 2$, $\sigma_X^2 = 9$, $\sigma_Y^2 = 16$, and $\mathrm{Cov}(X, Y) = 6$.

   (a) Describe the marginal distribution of X in words and write the explicit formula for its pdf, $f_X(u)$.

   (b) Describe the conditional distribution of Y given $X = 5$ in words, and write the explicit formula for the conditional pdf, $f_{Y|X}(v|5)$.

   (c) Find the numerical value of $P(Y \geq 2|X = 5)$.

   (d) Find the numerical value of $E[Y^2|X = 5]$.

5. **[Jointly Gaussian Random Variables II]**
   Suppose Y and W are jointly Gaussian random variables with $E[Y] = 2$, $E[W] = 0$, $\mathrm{Var}(Y) = 16$, $\mathrm{Var}(W) = 4$, and $\rho = 0.25$. Let $X = 3Y + W + 3$.

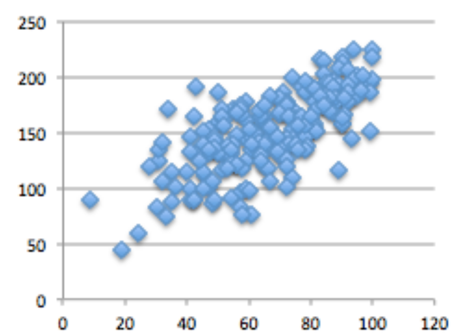
   (a) Find $E[X]$ and $\mathrm{Var}(X)$.

   (b) Calculate the numerical value of $P\{X \geq 20\}$.

   (c) Find $E[Y|X]$. Your answer should be a function of X.

   (d) Find the mean square error, $E[(Y - E[Y|X])^2]$.

6. **[Joint empirical distribution of ECE 313 scores]**
   The scatterplot below shows 205 points, (u_i, v_i), where u_i is the score on exam two, and v_i is the score on the final exam, for the i^{th} student in ECE313 in a recent semester. The empirical mean and standard deviation for exam two are $\mu_X = 67$ and $\sigma_X = 19$, the empirical mean and standard deviation for the final exam are $\mu_Y = 152$ and $\sigma_Y = 35$, and the empirical correlation coefficient (computed using a spreadsheet function) is $\rho = 0.71$. Visual inspection of the data suggests it is reasonable to assume that the joint distribution of the two scores is jointly normal. Let (X, Y) be jointly normal random variables with the above parameter values.



   (a) Find $E[Y|X = u]$ as a function of u.

   (b) Describe in words the conditional distribution of Y given $X = u$.