University of Illinois **Fall 2013**

ECE 313: Problem Set 13
Correlations, LLN, CLT, Minimum Variance Estimation

Due: Wednesday, December 11 at 6 p.m.
Reading: *ECE 313 Course Notes,* Sections 4.3-4.5.

1. [**Correlation and Independence**]
We have seen in the course that independent random variables X and Y are uncorrelated. On the other hand, your course notes observes that independence is a "much stronger" property than correlation. In this exercise we ask you to demonstrate this property by constructing an explicit joint distribution on the random variables (X, Y) such that (X, Y) are uncorrelated but not independent. *Hint*: There are many possible choices. One possibility is to suppose that X and Y are both identically distributed on the support set $\{-1, 0, 1\}$.

2. [**An application in personal finance**]
One of the major industries looking for strong probability skills in their hires is the financial industry. This question is borrowed from a popular economics blogpost, that of Professor N. Gregory Mankiw from Harvard. Writing on July 27, 2013 Professor Mankiw followed up on his NYTimes column on the price of gold with this question for his readers. I quote below.

> The topic [of my NYTimes column] is whether you should invest in gold as part of your portfolio. After you read the column, you might find the following problem of interest. It is based on roughly plausible assumptions.
>
> Imagine that you start off with a portfolio of 60 percent stocks and 40 percent bonds. The returns on stocks, bonds, and gold are uncorrelated. Stocks earn a higher expected return than bonds. Bonds and gold earn the same lower expected return, but gold returns are three times as volatile as bond returns, as measured by the standard deviation. You want to minimize risk, measured by the variance of your portfolio return, without changing the expected return on your portfolio. How much gold should you buy?
>
> I will leave this problem as an exercise for the reader. But I believe you should be able to come up with a precise numerical answer without resorting to a computer.

(a) You are the reader. Find the optimal portfolio (% of stocks, bonds and gold) to minimize risk.

(b) Now suppose you have n equity-types to invest in, each with the same expected return but different standard deviations: equity-type k has a return with standard deviation of σ_k, for $k = 1 \ldots n$. What is the optimal portfolio to minimize risk? Your answer will be in terms of $\sigma_1, \ldots, \sigma_n$.

3. [**Marathon Blackjack**]
In a particular version of the card game blackjack offered by gambling casinos, 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. 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?

4. **[Gaussian approximation for confidence intervals]**
Recall that if X has the binomial distribution with parameters n and p, the Chebychev inequality implies that
$$P\{|X - np| \geq a\sigma\} \leq \frac{1}{a^2}, \tag{1}$$
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.

5. **[All about the inverted L pdf]**
Let $\mathbb{X}$ and $\mathbb{Y}$ be jointly continuous random variables with the pdf given in Example 4.3.3 of the *Course Notes*. Note that the Example finds the marginal density of $\mathbb{X}$ and the conditional density of $\mathbb{Y}$ given the value of $\mathbb{X}$.

   (a) Read Example 4.3.3 carefully and *write down* the marginal density of $\mathbb{Y}$. (You should be able to do this without calculating any integrals etc.). Then find the mean and variance of $\mathbb{Y}$.

   (b) *State* the conditional mean $\mathsf{E}[\mathbb{Y} \mid \mathbb{X} = \alpha]$ and the conditional variance $\mathsf{var}(\mathbb{Y} \mid \mathbb{X} = \alpha)$ of $\mathbb{Y}$ given that $\mathbb{X} = \alpha$. You need to give the answer both for cases: $\alpha \in (0, 0.5)$ and $\alpha \in (0.5, 1)$.

   (c) As part (b) shows, the conditional mean of $\mathbb{Y}$ given that the value of $\mathbb{X}$ is α happens to be a *number* whose value depends on α. Since the value of the conditional mean depends on the value of $\mathbb{X}$, it can *also* be regarded as a *function* of $\mathbb{X}$, that is, a *random variable*. This random variable conventionally denoted as $\mathsf{E}[\mathbb{Y} \mid \mathbb{X}]$. What *kind* of random variable is $\mathsf{E}[\mathbb{Y} \mid \mathbb{X}]$? Continuous? Discrete? Mixed? What is the mean of this random variable? What is its variance?
   Similarly, the conditional variance of $\mathbb{Y}$ given that the value of $\mathbb{X}$ is α is a number $\mathsf{var}(\mathbb{Y} \mid \mathbb{X} = \alpha)$ $= \frac{1}{48}$. This too can be regarded as a random variable that is a function of $\mathbb{X}$, and it is denoted by $\mathsf{var}(\mathbb{Y} \mid \mathbb{X})$ What is the mean of this random variable?

   (d) The (unconstrained) minimum-mean-square-error (MMSE) estimate of $\mathbb{Y}$ given the value of $\mathbb{X}$ is the conditional mean $E[\mathbb{Y} \mid \mathbb{X}]$, that is, if we observe that $\mathbb{X}$ has value α, we *estimate* the value of $\mathbb{Y}$ as $\mathsf{E}[\mathbb{Y} \mid \mathbb{X} = \alpha]$. Sketch this estimator as a function of α. Is it a *linear* estimator? Calculate MSE, the mean-square error of this unconstrained MMSE estimator, using Eq. (4.32) of the *Course Notes*. Hint: the numbers needed have been calculated earlier in the course.

   (e) If $\mathbb{X}$ has value α and we use $\mathsf{E}[\mathbb{Y} \mid \mathbb{X} = \alpha]$ as the estimate for $\mathbb{Y}$ when $\mathbb{X} = \alpha$, the error we incur is $\mathbb{Y} - \mathsf{E}[\mathbb{Y} \mid \mathbb{X} = \alpha]$ and the squared error is $(\mathbb{Y} - \mathsf{E}[\mathbb{Y} \mid \mathbb{X} = \alpha])^2$ which has mean $\mathsf{E}\left[(\mathbb{Y} - \mathsf{E}[\mathbb{Y} \mid \mathbb{X} = \alpha])^2\right] = \mathsf{var}(\mathbb{Y} \mid \mathbb{X} = \alpha)$. Thus, the MSE is the *mean* of this quantity and so we have yet another formula (different looking from Eqs. (4.30)-(4.32)) of the *Course Notes*: $\mathsf{MSE} = \mathsf{E}[\mathsf{var}(\mathbb{Y} \mid \mathbb{X})]$. We calculated $\mathsf{E}[\mathsf{var}(\mathbb{Y} \mid \mathbb{X})]$ in part (c) above. Does it have the same value as we found in part (d)?

   (f) The *linear* MMSE estimator is constrained to be a linear function of $\mathbb{X}$, and as discussed in the *Course Notes*, the linear MMSE estimator is
   $$\hat{\mathbb{Y}} = \mathsf{E}[\mathbb{Y}] + \left(\frac{\mathsf{cov}(\mathbb{X}, \mathbb{Y})}{\mathsf{var}(\mathbb{X})}\right)(\mathbb{X} - \mathsf{E}[\mathbb{X}]) = \frac{7}{12} + \left(\frac{\mathsf{cov}(\mathbb{X}, \mathbb{Y})}{11/144}\right)\left(\mathbb{X} - \frac{7}{12}\right).$$
   Show that $\mathsf{cov}(\mathbb{X}, \mathbb{Y}) = -\frac{1}{36}$ and thus $\hat{\mathbb{Y}} = -\frac{4}{11}\mathbb{X} + \frac{35}{44}$. Sketch a graph of the function $-\frac{4}{11}\alpha + \frac{35}{44}$ and compare it to the graph of $\mathsf{E}[\mathbb{Y} \mid \mathbb{X} = \alpha]$ that you found in part (d).

   (g) Compute the MSE for the linear MMSE estimator and show that it is larger than the MSE of the unconstrained MMSE estimator, and smaller than the MSE if we simply ignored the observed value of $\mathbb{X}$ and use $\mathsf{E}[\mathbb{Y}]$ as the estimator for $\mathbb{Y}$.