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

1. **[Rate of convergence in law of large numbers for independent Gaussians]**

   (a) The sums of independent normal random variables is normal so S_n has a normal distribution. Also, $E[S_n] = E[X_1] + \cdots + E[X_n] = 0$, and since the X's are independent, hence uncorrelated, $\text{Var}(S_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n) = n$. That is, S_n has the $N(0, n)$ distribution. Thus,

   $$P\left\{\left|\frac{S_n}{n}\right| \geq \delta\right\} = P\left\{\left|\frac{S_n}{\sqrt{n}}\right| \geq \delta\sqrt{n}\right\} = 2Q(\delta\sqrt{n}).$$

   (b) Taking $x = \delta\sqrt{n}$, yields

   $$P\left\{\left|\frac{S_n}{n}\right| \geq \delta\right\} = 2Q(\delta\sqrt{n}) \leq \frac{2}{\delta\sqrt{2\pi n}}e^{-n\delta^2/2}$$

   Thus, for δ fixed, the probability converges to zero exponentially fast as $n \to \infty$.

2. **[Marathon blackjack]**

   (a) The expected net gain is $(1000)(\$100)(-0.0029) = -\$290$

   (b) The net gain, in dollars, is the sum of the gains for the 1000 games, $S = X_1 + \cdots + X_n$, where each X_i has mean $100(0.0029) = 0.29$ and standard deviation $(100)(1.141) = 114$. As already mentioned, S has mean $1000(-0.29) = -290$, and its standard deviation is $114\sqrt{1000} = 3605$. By the Gaussian approximation backed by the central limit theorem,

   $$P\{S > 0\} = P\left\{\frac{S + 290}{3605} > \frac{290}{3605}\right\} \approx Q\left(\frac{290}{3605}\right) = 0.4679$$

   (c)

   $$P\{S > 1000\} = P\left\{\frac{S + 290}{3605} > \frac{1290}{3605}\right\} \approx Q\left(\frac{1290}{3605}\right) = 0.3602$$

   (d) Replacing 1000 be a general integer n, we have

   $$P\{S > 0\} = P\left\{\frac{S + (0.29)n}{114\sqrt{n}} > \frac{(0.29)n}{114\sqrt{n}}\right\} \approx Q\left(\frac{(0.29)n}{114\sqrt{n}}\right)$$

   Since $Q(0.253) = 0.4$, we solve $\frac{(0.29)n}{114\sqrt{n}} = 0.253$ or

   $$n = \left(\frac{(0.253)(114)}{0.29}\right)^2 = 9891$$

   (That is a lot of blackjack!)

3. **[Gaussian approximation for confidence intervals]**

   Using the Chebychev inequality, the 96% confidence level is achieved using $a = 5$. If instead the Gaussian approximation based on the central limit theorem is used, we select a so that $2Q(a) = 0.04$, or $Q(a) = 0.02$, or $a = 2.0538$. As before, the half-width of the interval is $\frac{a}{2\sqrt{n}}$, which should be less than or equal to 0.1. This requires $n \geq \left(\frac{a}{0.2}\right)^2 = 25a^2 = 105.4$, so $n = 105$ should do. (Note: One can get the best of both approaches–i.e. use a smaller value of n and have the same type of guarantee as produced by the Chebychev bound, by calculating $P\{|X - np| \geq a\sigma\}$ exactly using the binomial distribution.)

4. **[Jointly Gaussian Random Variables I]**

   (a) X has the $N(1,9)$ distribution; $f_X(u) = \frac{1}{\sqrt{18\pi}} e^{-(u-1)^2/18}$.

   (b) By the formula for wide sense conditional expectation, $\widehat{E}[Y|X=5] = 2 + \frac{6}{9}(5-1) = 4.667$, and by the formula for the corresponding MSE, $\sigma_e^2 = \sigma_Y^2 - \frac{\text{Cov}(X,Y)^2}{\sigma_X^2} = 16 - \frac{6^2}{9} = 12$. So, the conditional distribution of Y given $X = 5$ is the $N(4.667, 12)$ distribution; $f_{Y|X}(v|5) = \frac{1}{\sqrt{24\pi}} e^{-(v-4.667)^2/24}$.

   (c) By part (b), this is the probability that a $N(4.667, 12)$ random variable is greater than or equal to 2, which is $Q\left(\frac{2-4.667}{\sqrt{12}}\right) = Q(-0.9698) = 0.7793$.

   (d) By part (b), this is the second moment of a $N(4.667, 12)$ random variable, which is $(4.667)^2 + 12 = 33.7778$.

5. **[Jointly Gaussian Random Variables II]**

   (a) $E[X] = E[3Y + W + 3] = 3E[Y] + E[W] + 3 = 9$.
   We use the fact $\text{Cov}(Y,W) = \rho\sigma_Y\sigma_W = (0.25)(4)(2) = 2$ to get

$$\begin{aligned}\text{Var}(X) &= \text{Var}(3Y + W + 3) \\ &= 9\text{Var}(Y) + \text{Var}(W) + 2 \cdot 3\text{Cov}(Y,W) \\ &= 144 + 4 + 12 = 160\end{aligned}$$

   (b) Since X is a linear combination of jointly Gaussian random variables, X is Gaussian. So

$$P\{X \geq 20\} = P\left\{\frac{X-9}{\sqrt{160}} \geq \frac{20-9}{\sqrt{160}}\right\} = Q\left(\frac{20-9}{\sqrt{160}}\right) = Q(0.8696) = 0.192$$

   (c) Since X and Y are linear combinations of the jointly Gaussian random variables X and W, the variables X and Y are jointly Gaussian. Therefore,

$$E[Y|X] = \widehat{E}[Y|X] = E[Y] + \frac{\text{Cov}(Y,X)}{\text{Var}(X)}(X - E[X]).$$

   Using

$$\text{Cov}(Y,X) = \text{Cov}(Y, 3Y + W + 3) = 3\text{Var}(Y) + \text{Cov}(Y,W) = 48 + 2 = 50,$$

   we find

$$E[Y|X] = \widehat{E}[Y|X] = 2 + \frac{50}{160}(X - 9)$$

   (d)

$$\text{MSE} = \text{Var}(Y) - \frac{(\text{Cov}(Y,X))^2}{\text{Var}(X)} = 16 - \frac{(50)^2}{160} = 0.375.$$

6. **[Joint empirical distribution of ECE 313 scores]**

   (a) By the joint Gaussian assumption,
   $E[Y|X=u] = \widehat{E}[Y|X=u] = 152 + \frac{(35)(0.71)}{19}(u - 67) = 152 + 1.31(u - 67)$

   (b) By the formula for minimum MSE for linear estimation,
   $\sigma_e = \sigma_X\sqrt{1-\rho^2} = 35\sqrt{1 - 0.71^2} = 24.65$. Thus, given $X = u$, the conditional distribution of Y is normal with mean $152 + 1.31(u - 67)$ and standard deviation 24.65 (i.e. variance 607.5).