Problem Set 10 Solutions

1. **A financial parable.**

(a) The bank becomes insolvent if the asset's gain $R \leq -5$ (i.e., it loses more than 5%). This probability is the CDF of R evaluated at -5. Since R is normally distributed, we can convert this CDF to be in terms of a standard normal random variable by subtracting away the mean and dividing by the standard deviation, and then look up the value in a standard normal CDF table.

$$
\begin{aligned}
\mathbf{E}[R] &= 7, \\
\mathrm{var}(R) &= 10^2 = 100, \\
\mathbf{P}(R \leq -5) &= \mathbf{P}\left(\frac{R-7}{10} \leq \frac{-5-7}{10}\right) = \Phi(-1.2) \approx 0.115.
\end{aligned}
$$

Thus, by investing in just this one asset, the bank has a 11.5% chance of becoming insolvent.

(b) If we model the R_i's as **independent** normal random variables, then their sum $R = (R_1 + \cdots + R_{20})/20$ is also a normal random variable (see Example 4.11 on page 214 of the text). Thus, we can calculate the mean and variance of this new R and proceed as in part (a). Note that since the random variables are assumed to be independent, the variance of their sum is just the sum of their individual variances.

$$
\begin{aligned}
\mathbf{E}[R] &= (\mathbf{E}[R_1] + \cdots + \mathbf{E}[R_{20}])/20 = 7, \\
\mathrm{var}(R) &= \frac{1}{20^2}(\mathrm{var}(R_1) + \cdots + \mathrm{var}(R_{20})) = \frac{20 \cdot 100}{400} = 5, \\
\mathbf{P}(R \leq -5) &= \mathbf{P}\left(\frac{R-7}{\sqrt{5}} \leq \frac{-5-7}{\sqrt{5}}\right) = \Phi(-5.367) \approx 0.0000000439 = 4.39 \cdot 10^{-8}.
\end{aligned}
$$

Thus, by diversifying and assuming that the 20 assets have **independent** gains, the bank has seemingly decreased its probability of becoming insolvent to a palatable value.

(c) Now, if the gains R_i are positively correlated, then we can no longer sum up the individual variances; we need to account for the covariance between pairs of random variables. The covariance is given by

$$
\mathrm{cov}(R_i, R_j) = \rho(R_i, R_j)\sqrt{\mathrm{var}(R_i)\mathrm{var}(R_j)} = \frac{1}{2}\sqrt{10^2 \cdot 10^2} = 50.
$$

From page 220 in the text, we know that the variance in this case is

$$
\begin{aligned}
\mathrm{var}(R) &= \mathrm{var}\left(\frac{1}{20}\sum_{i=1}^{20} R_i\right) = \frac{1}{400}\left(\sum_{i=1}^{20}\mathrm{var}(R_i) + \sum_{\{(i,j)|i\neq j\}}\mathrm{cov}(R_i, R_j)\right) \\
&= \frac{1}{400}(20 \cdot 100 + 380 \cdot 50) = 52.5.
\end{aligned}
$$

Since we assume that $R = (R_1 + \cdots + R_{20})/20$ is still normal, we can again apply the same steps as in parts (a) and (b):

$$
\begin{aligned}
\mathbf{E}[R] &= (\mathbf{E}[R_1] + \cdots + \mathbf{E}[R_{20}])/20 = 7, \\
\mathrm{var}(R) &= 52.5, \\
\mathbf{P}(R \leq -5) &= \mathbf{P}\left(\frac{R-7}{\sqrt{52.5}} \leq \frac{-5-7}{\sqrt{52.5}}\right) = \Phi(-1.656) \approx 0.0488.
\end{aligned}
$$

Thus, by taking into account the positive correlation between the assets' gains, we are no longer as comfortable with the probability of insolvency as we thought we were in part (b).

2. Let M and N be the number of males and females, respectively, that cast a vote. We need to find $P(M > N)$, i.e., $P(M - N > 0)$. The central limit theorem does not apply directly to the random variable $M - N$. However, the central limit theorem implies that M and N are well approximated by normal random variables. So, $M - N$ is the difference of two independent approximately normal random variables. Since the difference of two normal random variables is itself normal, it follows that $M - N$ is approximately normal. The mean and variance of $M - N$ are found by

$$\mathbf{E}[M - N] = 300 \cdot 0.4 + 196 \cdot 0.5 = 120 - 98 = 22,$$
$$\operatorname{var}(M - N) = \operatorname{var}(M) + \operatorname{var}(N) = 300 \cdot 0.4 \cdot 0.6 + 196 \cdot 0.5 \cdot 0.5 = 121.$$

Thus, the standard deviation of $M - N$ is 11. Let Z be a standard normal random variable. Using the central limit theorem approximation, we obtain

$$
\begin{aligned}
\mathbf{P}(M - N > 0) &= \mathbf{P}\left(\frac{M - N - 22}{11} > -\frac{22}{11}\right) \\
&\approx \mathbf{P}(Z \geq -2) \\
&= 0.9772.
\end{aligned}
$$

A slightly more refined estimate is obtained by expressing the event of interest as $\mathbf{P}(M - N \geq 1/2)$. We then have

$$
\begin{aligned}
\mathbf{P}(M - N > 1/2) &= \mathbf{P}\left(\frac{M - N - 22}{11} \geq -\frac{21.5}{11}\right) \\
&\approx \mathbf{P}(Z \geq -1.95) \\
&= 0.974.
\end{aligned}
$$

3. (a) Using the Central Limit Theorem, we obtain $\mathbf{P}(\frac{n}{2} - 10 \leq S_n \leq \frac{n}{2} + 10) \approx \Phi(\frac{20}{\sqrt{n}}) - \Phi(-\frac{20}{\sqrt{n}}) \to 0$ as $n \to \infty$.

   (b) The limit is 1, by the weak law of large numbers.

   (c) Using the Central Limit Theorem, we obtain $\mathbf{P}(\frac{n}{2} - \frac{\sqrt{n}}{2} \leq S_n \leq \frac{n}{2} + \frac{\sqrt{n}}{2}) \to \Phi(1) - \Phi(-1) = 0.6826$.

4. (a) Let C denote the coin that Bob received, so that $C = 1$ if Bob received the first coin, and $C = 2$ if Bob received the second coin. Then $\mathbf{P}(C = 1) = p$ and $\mathbf{P}(C = 2) = 1 - p$. Given C, the number of heads Y in 3 independent tosses is a binomial random variable.

   We can find the probability that Bob received the first coin given that he observed k heads using Bayes' rule.

$$
\begin{aligned}
\mathbf{P}(C = 1 \mid Y = k) \;&=\; \frac{\mathbf{P}(Y = k \mid C = 1) \cdot \mathbf{P}(C = 1)}{P(Y = k \mid C = 1) \cdot \mathbf{P}(C = 1) + \mathbf{P}(Y = k \mid C = 2) \cdot \mathbf{P}(C = 2)} \\
&=\; \frac{\binom{3}{k} \cdot (1/3)^k (2/3)^{3-k} p}{\binom{3}{k} \cdot (1/3)^k (2/3)^{3-k} \cdot p + \binom{3}{k} \cdot (2/3)^k (1/3)^{3-k} \cdot (1 - p)} \\
&=\; \frac{2^{3-k} p}{2^{3-k} p + 2^k (1 - p)} = \frac{1}{1 + \frac{1-p}{p} 2^{2k-3}}
\end{aligned}
$$

(b) We want to find k so that the following inequality holds.

$$
\begin{aligned}
\mathbf{P}(C = 1 \mid Y = k) \;&>\; p \\
\frac{2^{3-k} p}{2^{3-k} p + 2^k (1 - p)} \;&>\; p
\end{aligned}
$$

Note that if $p = 0$ or $p = 1$, there is no value of k that satisfies the inequality. We now solve it for $0 < p < 1$:

$$
\begin{aligned}
\frac{2^{3-k}}{2^{3-k} p + 2^k (1 - p)} \;&>\; 1 \\
2^{3-k} \;&>\; 2^{3-k} p + 2^k (1 - p) \\
2^{3-k}(1 - p) \;&>\; 2^k (1 - p) \\
2^{3-k} \;&>\; 2^k \\
2k \;&<\; 3 \\
k \;&<\; 3/2
\end{aligned}
$$

For $0 < p < 1$, $k = 0$ or $k = 1$ the probability that Alice sent the first coin increases. The inequality does not depend on p, and so does not change when p increases. Intuitively, this makes sense: lower values of k increase Bob's belief he got the coin with lower probability of heads.

(c) Given that Bob observes k heads, Bob must decide on whether the first or second coin was used. To minimize the error, he should decide it is the first coin when $\mathbf{P}(C = 1 \mid Y = k) \geq \mathbf{P}(C = 2 \mid Y = k)$. Thus, we have the decision rule given by

$$
\begin{aligned}
\mathbf{P}(C = 1 \mid Y = k) \;&\geq\; \mathbf{P}(C = 2 \mid Y = k) \\
\frac{2^{3-k} p}{2^{3-k} p + 2^k (1 - p)} \;&\geq\; \frac{2^k (1 - p)}{2^{3-k} p + 2^k (1 - p)} \\
2^{3-k} p \;&\geq\; 2^k (1 - p) \\
2^{2k-3} \;&\leq\; \frac{p}{1 - p} \\
k \;&\leq\; \frac{3}{2} + \frac{1}{2} \log_2 \frac{p}{1 - p}
\end{aligned}
$$

(d)   i. If $p = 2/3$, the threshold in the rule above is equal to $\frac{3+\log_2 2}{2} = 2$. Therefore, Bob will decide that he received the first coin when he observes 0, 1 or 2 heads, and will decide that he received the second coin when he observes 3 heads.

We find the probability of a correct decision using the total probability law:

$$
\begin{aligned}
\mathbf{P}(\text{Correct}) &= \mathbf{P}(\text{Correct} \mid C = 1) \cdot p + \mathbf{P}(\text{Correct} \mid C = 2) \cdot (1 - p) \\
&= \mathbf{P}(Y < 3 \mid C = 1) \cdot p + \mathbf{P}(Y = 3 \mid C = 2) \cdot (1 - p) \\
&= (1 - \mathbf{P}(Y = 3 \mid C = 1)) \cdot p + \mathbf{P}(Y = 3 \mid C = 2) \cdot (1 - p) \\
&= (1 - (1/3)^3)(2/3) + (2/3)^3(1/3) = 20/27 \approx .741
\end{aligned}
$$

ii. In absence of any data, all Bob can do is decide he received the first coin with some probability q. Note that this rule includes the deterministic decisions that he received either the first coin ($q = 1$) or the second coin ($q = 0$).

In this case, the probability of correct decision is equal to

$$
\begin{aligned}
\mathbf{P}(\text{Correct}) &= \mathbf{P}(\text{Correct} \mid C = 1) \cdot p + \mathbf{P}(\text{Correct} \mid C = 2) \cdot (1 - p) \\
&= qp + (1 - q)(1 - p) = 1 - p + q(2p - 1) = \frac{1 + q}{3}
\end{aligned}
$$

Clearly, the probability of the correct decision is maximized (or the probability of error is minimized) when $q = 1$, i.e., when Bob deterministically decides he received the first coin. In this case, $\mathbf{P}(\text{Correct}) = 2/3 \approx .667$. Observing 3 coin tosses increases the probability of the correct decision by $2/27 \approx .074$.

(e) If p is increased, the threshold in the decision rule in part (c) goes up, i.e., the range of values of k for which Bob decides he received the first coin can only go up.

(f) Bob will never decide he received the first coin if the threshold in the rule above is below zero:

$$
\begin{aligned}
\frac{3}{2} + \frac{1}{2} \log_2 \frac{p}{1 - p} &< 0 \\
\log_2 \frac{p}{1 - p} &< -3 \\
\frac{p}{1 - p} &< \frac{1}{8} \\
p &< \frac{1}{9}
\end{aligned}
$$

If $p < 1/9$, the prior probability of receiving the first coin is so low that no amount of evidence from 3 tosses of the coin will make Bob decide he received the first coin.

(g) Bob will always decide he received the first coin if the threshold in the rule above is equal to or above 3:

$$
\begin{aligned}
\frac{3}{2} + \frac{1}{2} \log_2 \frac{p}{1 - p} &\geq 3 \\
\log_2 \frac{p}{1 - p} &\geq 3 \\
\frac{p}{1 - p} &\geq 8 \\
p &\geq \frac{8}{9}
\end{aligned}
$$

If $p \geq 8/9$, the prior probability of receiving the first coin is so high that no amount of evidence from 3 tosses of the coin will make Bob decide he received the second coin.

5. (a) Using the total probability theorem, we have

$$p_{T_1}(t) = \int_0^1 p_{T_1|Q}(t, q) f_Q(q) dq = \int_0^1 (1-q)^{t-1} q \, dq = \frac{1}{(t+1)t} \qquad \text{for } t = 1, 2, \ldots$$

(b) The least squares estimate coincides with the conditional expectation of Q given T_1, which is derived as

$$\begin{aligned}
\mathbf{E}[Q \mid T_1 = t] &= \int_0^1 p_{Q|T_1}(q \mid t) q \, dq \\
&= \int_0^1 \frac{p_{T_1|Q}(t \mid q) f_Q(q)}{p_{T_1}(t)} q \, dq \\
&= \int_0^1 t(t+1) q(1-q)^{t-1} q \, dq \\
&= \int_0^1 t(t+1) q^2 (1-q)^{t-1} dq \\
&= t(t+1) \frac{2(t-1)!}{(t+2)!} \\
&= \frac{2}{t+2}
\end{aligned}$$

(c) We write the posterior probability distribution of Q given $T_1 = t_1, \ldots, T_k = t_k$

$$\begin{aligned}
f_{Q|T_1,\ldots,T_k}(q \mid t_1, \ldots, t_k) &= \frac{f_Q(q) \prod_i^k P_{T_i}(T_i = t_i \mid Q = q)}{\int_0^1 f_Q(q) \prod_i^k P_{T_i}(T_i = t_i \mid Q = q) dq} \\
&= \frac{q^k (1-q)^{\sum_i^k t_i - k}}{c} \\
&= \frac{1}{c} q^k (1-q)^{\sum_i^k t_i - k},
\end{aligned}$$

where the denominator integrates out q so it could be viewed as a constant scalar c. To maximize the above probability we set its derivative with respect to q to zero

$$k q^{k-1} (1-q)^{\sum_i^k t_i - k} - (\sum_i^k t_i - k) q^k (1-q)^{\sum_i^k t_i - k - 1} = 0,$$

or equivalently

$$k(1-q) - (\sum_i^k t_i - k) q = 0,$$

which yields the MAP estimate

$$\hat{q} = \frac{k}{\sum_{i=1}^k t_i}.$$

For this part only assume q is sampled from the random variable Q which is now uniformly distributed over $[0.5, 1]$

(d) The LLSE of T_1 given T_2 is

$$\hat{T}_2 = \mathbf{E}[T_2] + \frac{\text{cov}(T_1, T_2)}{\text{var}(T_1)}(T_1 - \mathbf{E}[T_1]),$$

where the coefficients are

$$\mathbf{E}[T_1] = \mathbf{E}[T_2] = \int_{0.5}^1 f_Q(q)\mathbf{E}[T|Q=q]dq = \int_{0.5}^1 2*1/q dq = 2\ln 2,$$

and from the law of total variance

$$\begin{aligned}
\text{var}(T_1) = \text{var}(T_2) &= \mathbf{E}\left[\text{var}(T_1 \mid Q)\right] + \text{var}\left[\mathbf{E}(T_1 \mid Q)\right] \\
&= \mathbf{E}\left[\frac{1-Q}{Q^2}\right] + \text{var}\left[\frac{1}{Q}\right] \\
&= \mathbf{E}[1/Q^2] - \mathbf{E}[1/Q] + \mathbf{E}[1/Q^2] - \mathbf{E}[1/Q]^2 \\
&= \int_{0.5}^2 f_Q(q)\frac{1}{q^2}dq - \int_{0.5}^2 f_Q(q)\frac{1}{q}dq + \int_{0.5}^2 f_Q(q)\frac{1}{q^2}dq - \left(\int_{0.5}^2 f_Q(q)\frac{1}{q}dq\right)^2 \\
&= 2 - 2\ln 2 + 2 - (2\ln 2)^2 \\
&= 4 - 2\ln 2 - (2\ln 2)^2,
\end{aligned}$$

and their covariance

$$\begin{aligned}
\text{cov}(T_1, T_2) &= \mathbf{E}[T_1 T_2] - \mathbf{E}[T_1]\mathbf{E}[T_2] \\
&= \mathbf{E}\left[\mathbf{E}[T_1 T_2 \mid Q]\right] - \mathbf{E}[T_1]\mathbf{E}[T_2] \\
&= \mathbf{E}\left[\mathbf{E}[T_1 \mid Q]\mathbf{E}[T_2 \mid Q]\right] - \mathbf{E}[T_1]\mathbf{E}[T_2] \\
&= \mathbf{E}\left[1/Q^2]\right] - \mathbf{E}[T_1]\mathbf{E}[T_2] \\
&= 2 - 4(\ln 2)^2
\end{aligned}$$

Therefore we have derived the linear least squares estimator

$$\hat{T}_2 = 2\ln 2 + \frac{2 - 4(\ln 2)^2}{4 - 2\ln 2 - (2\ln 2)^2}(T_1 - 2\ln 2) \approx 1.543 + 0.113T_1.$$

6. (a) To find the normalization constant c we integrate the joint PDF:

$$\int_0^1 \int_0^1 f_{X,Y}(x,y)\, dy\, dx = c\int_0^1 \int_0^1 xy\, dy\, dx = c\int_0^1 1/2x\, dx = c/4.$$

Therefore, $c = 4$.

(b) To construct the conditional expectation estimator, we need to find the conditional probability density.

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{4xy}{\int_0^1 4xy\, dx} = \frac{4xy}{2y} = 2x, \quad x \in (0,1]$$

Thus

$$\hat{x}_{\text{CE}}(y) = \mathbf{E}[X|Y=y] = \int_0^1 x \cdot 2x\, dx = 2/3.$$

(c) We first note that the conditional probability does not depend on y. Therefore, X and Y are independent, and whether or not we observe $Y = y$ does not affect the estimate in part (b). Another way to see this is to consider that if we do not observe y, we can compute the marginal $f_X(x) = \int_0^1 4xy \, dy = 2x$ which is equal to the conditional density, and will therefore produce the same estimate.

(d) Since X and Y are independent, no estimator can make use of the observed value of Y to estimate X. The MAP estimator for X is equal to 1, regardless of what value y we observe, since the conditional (and the marginal) density is maximized at 1.

6.041 / 6.431 Probabilistic Systems Analysis and Applied Probability
Fall 2010