Contents

Question 1 3
(a) .......................................................... 3
(b) .......................................................... 3

Question 2 4
(a) .......................................................... 4
(b) .......................................................... 5
(c) .......................................................... 5
(d) .......................................................... 5

Question 3 12
(a) .......................................................... 12
(b) .......................................................... 12
(c) .......................................................... 12
Question 1

You move into a new neighborhood, and there are three other houses in your cul-de-sac. When talking to your new neighbors, let's call them Alice, Bob, and Cathy, you find out that Alice has three pets. Since you don't know if Alice is a cat or a dog lover, you can assume that the probability that Alice has a cat is $\frac{1}{2}$ and the probability that she has a dog is $\frac{1}{2}$. Alice reveals to you that one of her pets is a cat.

(a) What is the probability that at least one of Alice's pets is a dog?
After getting to know your neighbors more, you find out the following: Alice has two dogs and a cat, Bob has a cat, a dog, and a hamster, while Cathy has four dogs. However, since you are busy with school, you never got a chance to see any of the pets. During your evening walk, you see a dog without a leash, and you want to alert your neighbors.

(b) Whose door should you knock first and why?

Answer:

(a) Before Alice give us the information about her first pet, the sample space of our problem can be given from the following set, where $D$ stands for dog and $C$ for cat:

$$A = \{(D, D, D), (D, D, C), (D, C, D), (D, C, C), (C, D, D), (C, D, C), (C, C, D), (C, C, C)\}$$

Also from our hypothesis it is clear that we make the assumption that the above space is homogeneous, i.e. all the simple events in the above set occur with the same probability $\frac{1}{8}$. The existence of one cat in her pets makes us exclude the case $(D, D, D)$, reducing our sample space to:

$$B = \{(D, D, C), (D, C, D), (D, C, C), (C, D, D), (C, D, C), (C, C, D), (C, C, C)\}$$

This new space is also homogeneous. Indeed, the updated probability of each of the above simple events $b \in B$, can be computed directly from the definition of conditional probability:

$$\mathbb{P}(b|B) = \frac{\mathbb{P}(b, B)}{\mathbb{P}(B)} = \frac{1/8}{7/8} = \frac{1}{7}.$$ 

The event in which at least one of Alice's pets is a dog, given $B$, is precisely:

$$G = \{(D, D, C), (D, C, D), (D, C, C), (C, D, D), (C, D, C), (C, C, D)\} = B \setminus (C, C, C)$$

so

$$\mathbb{P}(G) = \mathbb{P}(B \setminus (C, C, C)) = 1 - \mathbb{P}((C, C, C)) = \frac{6}{7}.$$
(b) It is easy to see that the door we should knock first is that of Cathy’s, as she has more dogs than the other two neighbors together. But let’s make the above statement rigorous. First, we will sum up the information given on a table:

<table>
<thead>
<tr>
<th></th>
<th>Cats</th>
<th>Dogs</th>
<th>Hamsters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Bob</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cathy</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

To model our problem we think as follows: During our evening walk we see a pet, which belongs to someone of our neighbors, and we consider this event as a random one. By this we mean that we assume that it is equally likely to see any of the 10 neighbors’ pets. We use two random variables $X, Y$ where $X$ gives us the owner and $Y$ the kind of the observed pet. According to our hypothesis, we have already observed that the pet is a dog. That is we know $Y = \text{Dog}$ and the question is what is the most likely value of $X$. That is we want to find for which value $x$ of $X$, $P(X = x | Y = \text{Dog})$ is maximized. We just compute the $P(X = x | Y = \text{Dog})$ for all different scenarios:

$$P(X = \text{Alice} | Y = \text{Dog}) = \frac{P(X = \text{Alice}, Y = \text{Dog})}{P(Y = \text{Dog})} = \frac{2/10}{7/10} = \frac{2}{7}$$  \hspace{1cm} (1)

$$P(X = \text{Bob} | Y = \text{Dog}) = \frac{P(X = \text{Bob}, Y = \text{Dog})}{P(Y = \text{Dog})} = \frac{1/10}{7/10} = \frac{1}{7}$$  \hspace{1cm} (2)

$$P(X = \text{Cathy} | Y = \text{Dog}) = \frac{P(X = \text{Cathy}, Y = \text{Dog})}{P(Y = \text{Dog})} = \frac{4/10}{7/10} = \frac{4}{7}$$  \hspace{1cm} (3)

We see that (3) gives us the maximum value thus, the answer is that the dog most likely belongs to Cathy as predicted.

**Question 2**

Recall the 1-d Gaussian distribution that we studied in the class.

(a) Prove that the Gaussian distribution is well normalized. In other words show that

$$\int_{-\infty}^{\infty} N(x \mid \mu, \sigma^2) = 1$$

Question 2 continued on next page...
(b) Formally show that for the Gaussian distribution
\[ E[x] = \mu \]
\[ Var[x] = \sigma^2 \]

(c) Show that the maximum likelihood (ML) estimate for the Gaussian is given by
\[
\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} x_n \\
\sigma_{ML}^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu_{ML})^2
\]

(d) Show that the ML estimate of the mean is unbiased but the variance is biased:
\[ E[\mu_{ML}] = \mu \]
\[ E[\sigma^2_{ML}] = \left( \frac{N - 1}{N} \right) \sigma^2 \]

Answer:
(a) To prove that the Gaussian is well normalized we simply need to show that
\[
\int_{-\infty}^{\infty} N(x \mid \mu, \sigma^2) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = 1.
\]

First we make the substitution \( z = \frac{x-\mu}{\sqrt{2\pi\sigma^2}} \Rightarrow dx = \sqrt{2\sigma^2} \; dz \), then we get:
\[
\int_{-\infty}^{\infty} N(z \mid \mu, \sigma^2) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-z^2} \; dz. \tag{4}
\]

So in order the above to be well normalized we need to show that:
\[
I = \int_{-\infty}^{+\infty} e^{-z^2} \; dz = \sqrt{\pi} \tag{5}
\]

We shall first show that \( I \) is bounded. We notice that:
\[
\forall z \in [-1, 1] \Rightarrow 0 < e^{-z^2} \leq 1 \tag{6}
\]
\[
\forall z \in (-\infty, -1) \Rightarrow 0 < e^{-z^2} \leq e^z \tag{7}
\]
\[ \forall z \in (1, \infty) \implies 0 < e^{-z^2} < e^{-z} \quad (8) \]

Since \( e^{-z^2} > 0 \ \forall z \in \mathbb{R} \) and by taking into account equations (6), (7), (8) we get the following:

\[ 0 < \int_{-1}^{1} e^{-z^2} \, dz \leq \int_{-1}^{1} 1 \, dz \quad (9) \]

\[ 0 < \int_{-\infty}^{1} e^{-z^2} \, dz \leq \int_{-\infty}^{1} e^z \, dz \quad (10) \]

\[ 0 < \int_{1}^{\infty} e^{-z^2} \, dz \leq \int_{1}^{\infty} e^{-z} \, dz \quad (11) \]

Adding inequalities (9), (10), (11) we get:

\[ 0 < \int_{-\infty}^{+\infty} e^{-z^2} \, dz \leq \int_{-1}^{1} 1 \, dz + \int_{-\infty}^{1} e^z \, dz + \int_{1}^{\infty} e^{-z} \, dz = 2 + 2e^{-1} < \infty \quad (12) \]

Since the integral \( I \) is bounded we can calculate \( I^2 \):

\[ I^2 = \int_{-\infty}^{+\infty} e^{-z^2} \, dz \int_{-\infty}^{+\infty} e^{-z^2} \, dz \]

By doing a variable change \( y = z \) and since both integrals are bounded we get:

\[ I^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-z^2} e^{-y^2} \, dz \, dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(z^2+y^2)} \, dz \, dy \]

Switch to polar coordinates: \( r = z^2 + y^2, z = r \cos \theta, y = r \sin \theta \). The Jacobian for this transformation is:

\[
\det J(r, \theta) = \begin{vmatrix} \frac{\partial (z, y)}{\partial (r, \theta)} \end{vmatrix} = \begin{vmatrix} \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r \cdot (\sin^2 \theta + \cos^2 \theta) = r
\]

Also since we need to cover all of the \( r - \theta \) plane we require that \( 0 \leq r < \infty \) and that \( 0 \leq \theta \leq 2\pi \). Hence the resulting integral is:

\[ I^2 = \int_{0}^{+\infty} \int_{0}^{2\pi} e^{-r^2} r \, d\theta \, dr = 2\pi \int_{0}^{+\infty} e^{-r^2} r \, dr \]
Doing a simple substitution \( u = r^2 \quad \Rightarrow \quad du = 2r dr \) we get:

\[
I^2 = \pi \int_0^{+\infty} e^{-u} du = \pi \lim_{R \to \infty} (-e^{-R} + 1) = \pi(0 + 1) = \pi
\]

Since \( I \geq 0 \quad \Rightarrow \quad I = \sqrt{\pi}. \) Summing up:

\[
\int_{-\infty}^{+\infty} N(z | \mu, \sigma^2) = \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{\pi}} \sqrt{\pi} = 1.
\]

(b) First show that \( E[x] = \mu, \) that is \( \int_{-\infty}^{+\infty} x N(x | \mu, \sigma^2) dx = \mu. \)

Again make the substitution \( z = \frac{x-\mu}{\sqrt{2}\sigma^2} \) then \( x = z\sqrt{2}\sigma^2 \) and \( dx = \sqrt{2}\sigma^2 \, dz. \)
So we get:

\[
E[z] = \frac{1}{\sqrt{2\sigma^2 \pi}} \int_{-\infty}^{+\infty} (z\sqrt{2\sigma^2} + \mu) e^{-\frac{z^2}{2}} \, dz = \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} \, dz + \sqrt{\frac{2\sigma^2}{\pi}} \int_{-\infty}^{+\infty} ze^{-\frac{z^2}{2}} \, dz = \mu + \sqrt{\frac{2\sigma^2}{\pi}} \int_{-\infty}^{+\infty} ze^{-\frac{z^2}{2}} \, dz.
\]

The above integral equals 0, as the integrand is an odd function. The only thing remaining to show is that \( \int_0^{+\infty} ze^{-\frac{z^2}{2}} \, dz \) is bounded.

Indeed, by making a change of variables, \( u = z^2 \quad \Rightarrow \quad du = 2z \, dz, \) we get:

\[
\int_0^{+\infty} e^{-u} \left( \frac{du}{2} \right) = -\frac{1}{2} \lim_{R \to \infty} [e^{-u}]_0^R = 0 - 0 = 0.
\]

Thus

\[
E[u] = \mu + \sqrt{\frac{2\sigma^2}{\pi}} \cdot 0 = \mu.
\]

Now we show that \( Var[x] = E[(x-\mu)^2] = \sigma^2, \) i.e. \( \int_{-\infty}^{+\infty} (x-\mu)^2 N(x | \mu, \sigma^2) dx = \sigma^2. \)

In a similar fashion let \( z = \frac{x-\mu}{\sqrt{2\sigma^2}} \) then \( x = z\sqrt{2}\sigma^2 \) and \( dx = \sqrt{2\sigma^2} \, dz. \) So we get:
$$Var[x] = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} z^2 e^{-z^2} \, dz$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} z(z e^{-z^2}) \, dz$$

Now integrate by parts. Let \( u = z \) and \( v = \frac{-1}{2} e^{-z^2} \) then:

$$Var[x] = \frac{2\sigma^2}{\sqrt{\pi}} \left[ -\lim_{R \to \infty} \left[ \frac{z e^{-z^2}}{2} \right]_{-R}^{R} + \frac{1}{2} \int_{-\infty}^{+\infty} e^{-z^2} \, dz \right]$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \left[ -\lim_{R \to \infty} \left[ \frac{R}{2eR^2} + \frac{R}{2eR^2} \right] + \frac{\sqrt{\pi}}{2} \right]$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \left[ -\lim_{R \to \infty} \left[ \frac{R}{2eR^2} + \frac{R}{2eR^2} \right] + \sigma^2 \right].$$

In order to determine the limit we use L’Hopital’s Rule since \( R \to \infty \) and \( e^{R^2} \to \infty \). Hence:

$$\lim_{R \to \infty} \frac{R}{2eR^2} = \lim_{R \to \infty} \frac{1}{2ReR^2} = 0$$

As a result we get that:

$$Var[x] = 0 + \sigma^2 = \sigma^2$$

(c) Let \( X = (X_1, X_2, \cdots, X_N) \) be i.i.d. Gaussians. We want to find the values of \( \mu, \sigma^2 \) that maximize the likelihood function: \( \mathcal{N}(X | \mu, \sigma^2) \). First, as \( X_i \) are i.i.d. we have that

$$\mathcal{N}(X | \mu, \sigma^2) = \mathcal{N}(X_1, X_2, \cdots, X_N | \mu, \sigma^2)$$

$$= \prod_{i=1}^{N} \mathcal{N}(X_i | \mu, \sigma^2)$$
So
\[
\max_{\mu, \sigma^2} \arg \mathcal{N}(X | \mu, \sigma^2) = \max_{\mu, \sigma^2} \prod_{i=1}^{N} \mathcal{N}(X_i | \mu, \sigma^2)
\]
\[
= \max_{\mu, \sigma^2} \log \left( \prod_{i=1}^{N} \mathcal{N}(X_i | \mu, \sigma^2) \right)
\]
\[
= \max_{\mu, \sigma^2} \sum_{i=1}^{N} \log \left( \mathcal{N}(X_i | \mu, \sigma^2) \right)
\]
\[
= \min_{\mu, \sigma^2} \sum_{i=1}^{N} \log \left( \mathcal{N}(X_i | \mu, \sigma^2) \right)
\]

Substituting now the formula for the Gaussian pdf we get
\[
\min_{\mu, \sigma^2} - \sum_{i=1}^{N} \log \left( \mathcal{N}(X_i | \mu, \sigma^2) \right) = \min_{\mu, \sigma^2} - \sum_{i=1}^{N} \log \left( \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right)
\]
\[
= \min_{\mu, \sigma^2} - \left( \sum_{i=1}^{N} \log \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right) \right) - \left( \frac{(x_i - \mu)^2}{2\sigma^2} \right)
\]
\[
= \min_{\mu, \sigma^2} N \log(2\pi\sigma^2) + \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma^2}
\]

To find \(\mu_{ML}\) we take derivative with respect to \(\mu\) to be equal to 0:
\[
\frac{\partial}{\partial \mu} \left( N \log(2\pi\sigma^2) + \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma^2} \right) = \left( - \left( \sum_{i=1}^{N} \frac{2(x_i - \mu)}{2\sigma^2} \right) \right) = 0 \iff \sum_{i=1}^{N} (x_i - \mu) = 0 \iff \mu = \frac{\sum_{i=1}^{N} x_i}{N}
\]

Thus, \(\mu_{ML} = \frac{\sum_{i=1}^{N} x_i}{N}\). Now we take the derivative with respect to \(\sigma^2\) to be equal to 0, with \(\mu_{ML}\) in place of \(\mu\):
\[
\frac{\partial}{\partial \sigma^2} \left( N \log(\sqrt{2\pi\sigma^2}) + \sum_{i=1}^{N} \frac{(x_i - \mu_{ML})^2}{2\sigma^2} \right) = 2\pi N \frac{\sigma}{4\pi^2} + \left( - \frac{\sum_{i=1}^{N} (x_i - \mu_{ML})^2}{2\sigma^4} \right) = 0 \iff 
\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu_{ML})^2}{N}.
\]

Hence \( \sigma_{ML}^2 = \frac{\sum_{i=1}^{N} (x_i - \mu_{ML})^2}{N} \).

(d) First show that \( E[\mu_{ML}] = \mu \). Starting by the definition of \( \mu_{ML} \) we get:

\[
E[\mu_{ML}] = E\left[ \frac{1}{N} \sum_{n=1}^{N} x_n \right] = \frac{1}{N} E \left[ \sum_{n=1}^{N} x_n \right] = \frac{1}{N} \sum_{n=1}^{N} E[x_n] = \frac{1}{N} N \mu = \mu
\]

where we used the linearity of expected value.

Now we need to show that \( E[\sigma_{ML}^2] = \left( \frac{N-1}{N} \right) \sigma^2 \). Starting again with the definition we get:

\[
E[\sigma_{ML}^2] = E\left[ \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu_{ML})^2 \right] \\
= \frac{1}{N} E \left[ \sum_{n=1}^{N} (x_n - \mu_{ML})^2 \right] \\
= \frac{1}{N} E \left[ \sum_{n=1}^{N} x_n^2 - 2\mu_{ML} \sum_{n=1}^{N} x_n + \sum_{n=1}^{N} \mu_{ML}^2 \right] \\
= \frac{1}{N} E \left[ \sum_{n=1}^{N} x_n^2 - 2\mu_{ML}(N\mu_{ML}) + N\mu_{ML}^2 \right] \\
= E\left[ \frac{1}{N} \sum_{n=1}^{N} x_n^2 - \mu_{ML}^2 \right] \\
= E\left[ \frac{1}{N} \sum_{n=1}^{N} x_n^2 - \left( \frac{1}{N} \sum_{n=1}^{N} x_n \right)^2 \right]
\]

At this point we have to use the well-known identity, \( \forall a \in \mathbb{R} \):

\[
\left( \sum_{i=1}^{N} a_i \right)^2 = \sum_{i=1}^{N} a_i^2 + \sum_{i=1}^{N} \sum_{j \neq i} a_i a_j
\] (13)
Using (13) we get:

\[
E\left[\frac{1}{N} \sum_{n=1}^{N} x_n^2 - \left( \frac{1}{N} \sum_{n=1}^{N} x_n \right)^2 \right] = \\
E\left[\frac{1}{N} \sum_{n=1}^{N} x_n^2 - \frac{1}{N^2} \sum_{n=1}^{N} x_n^2 - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \right] = \\
\frac{1}{N} \sum_{n=1}^{N} E[x_n^2] - \frac{1}{N^2} \sum_{n=1}^{N} E[x_n^2] - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} E[x_i x_j] = \\
\frac{1}{N} \sum_{n=1}^{N} E[x_n^2] - \frac{1}{N^2} \sum_{n=1}^{N} E[x_n^2] - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mu^2 = \\
\frac{1}{N} \sum_{n=1}^{N} E[x_n^2] - \frac{1}{N^2} \sum_{n=1}^{N} E[x_n^2] - \frac{N^2 - N}{N^2} \mu^2 = \\
\left[\frac{1}{N} - \frac{1}{N^2}\right] \sum_{n=1}^{N} E[x_n^2] - \frac{N - 1}{N} \mu^2 = \\
\left[\frac{1}{N} - \frac{1}{N^2}\right] N E[x_1^2] - \frac{N - 1}{N} \mu^2 = \\
\frac{N - 1}{N} E[x_1^2] - \frac{N - 1}{N} \mu^2
\]

Since all of \(x_i\) are i.i.d. \(\implies E(x_i x_j) = E(x_i)E(x_j)\). Therefore it follows that:

\[
\frac{1}{N} \sum_{n=1}^{N} E[x_n^2] - \frac{1}{N^2} \sum_{n=1}^{N} E[x_n^2] - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \mu^2 = \\
\frac{1}{N} \sum_{n=1}^{N} E[x_n^2] - \frac{1}{N^2} \sum_{n=1}^{N} E[x_n^2] - \frac{N^2 - N}{N^2} \mu^2 = \\
\frac{N - 1}{N} E[x_1^2] - \frac{N - 1}{N} \mu^2
\]

Since \(\sigma^2 = E[x_n^2] - \mu^2 \iff E[x_n^2] = \sigma^2 + \mu^2\) for all \(n \in \{1, 2, \cdots, N\}\) we finally get:

\[
\sigma^2_{ML} = \frac{N - 1}{N} E[x_n^2] - \frac{N - 1}{N} \mu^2 = \frac{N - 1}{N} \left[\sigma^2 + \mu^2\right] - \frac{N - 1}{N} \mu^2 = \frac{N - 1}{N} \sigma^2
\]
Question 3

(a) Show that if $X$ and $Y$ are independent random variables, then

$$
P(X,Y) = P(X) \cdot P(Y)
$$

$$
E[XY] = E[X] \cdot E[Y]
$$

(b) Write some code to simulate a large number (say 10,000) of coin tosses with two independent coins, and empirically verify the above results. Write a short (5-10 sentence) summary of your experiment and what you observed.

(c) Now let your coin tosses depend on each other (e.g., the value that coin 2 takes depends on the value of coin 1) and empirically verify that the above results do not hold. Write a short (5 - 10 sentence) summary of your experiment and what you observed.

Answer:

(a) For this question we will assume that our random variables are discrete (in the continuous case instead of probabilities we will have pdfs and instead of sums integrals, and with these modifications the proof is exactly the same). We denote with $D_X, D_Y$ the domain of definition of $X, Y$, respectively. Let also $X(D_X), Y(D_Y) \subset \mathbb{R}$ be the image of $X, Y$, respectively. The random variables $X, Y$ are called independent iff:

$$
P(X = x | Y = y) = P(X = x) \text{ and } P(Y = y | X = x) = P(Y = y), \forall x \in X(D_X), y \in Y(D_Y)
$$

Let $x \in X(D_X), y \in Y(D_Y)$ we have:

$$
P(X = x) = P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} \iff
$$

$$
P(X = x, Y = y) = P(X = x)P(Y = y).
$$

That is:

$$
P(X, Y) = P(X)P(Y).
$$

The expected values of $X, Y$ are given by:

$$
E[X] = \sum_{x \in X(D_X)} xP(X = x), \quad E[Y] = \sum_{y \in Y(D_Y)} yP(Y = y)
$$

Question 3 continued on next page...
and the expected value of $XY$ is given by:

$$E[XY] := \sum_{r \in (D_X \times D_Y)} r \mathbb{P}(XY = r) = \sum_{x,y \in (D_X \times D_Y)} xy \mathbb{P}(X = x, Y = y)$$

$$= \sum_{x \in D_X} \sum_{y \in D_Y} xy \mathbb{P}(X = x, Y = y) = \sum_{x \in D_X} \sum_{y \in D_Y} xy \mathbb{P}(X = x) \mathbb{P}(Y = y)$$

$$= \sum_{x \in D_X} x \mathbb{P}(X = x) \sum_{y \in D_Y} y \mathbb{P}(Y = y) = E[X]E[Y]$$

(b) With the random variables $X$ and $Y$ we are simulating the toss of the first and second coin, respectively. More specifically each random variable takes values 0 and 1 corresponding to TAILS and HEADS, respectively. The large number $n$ of coin flips allows the relative frequencies $P_{\text{head}_1}$, $P_{\text{head}_2}$ to approach the theoretical success probabilities of each coin $\text{prob}_1$, $\text{prob}_2$. As a result, we can also expect convergence of the mean values $M_x$, $M_y$, $M_{xy}$ to their respective expected values $E_x$, $E_y$, $E_{xy}$. We attach our code:

```python
from __future__ import division
import random

# Definition of coin with success probability p:
def coin_flip(prob):
    return 1 if random.random() < prob else 0

if __name__ == "__main__":
    head_c1, head_c2 = 0, 0
    head_c1notc2, head_c2notc1, head_both, head_none = 0, 0, 0, 0
    prob1 = 0.3 # Success Probability of X
    prob2 = 0.9 # Success Probability of Y
    n = 1000000 # Number of tosses

    print("Running %d coin flips...") % n)
    for i in range(1, n):
        # Generating two random coins
        first = coin_flip(prob1) # Random Variable X
        second = coin_flip(prob2) # Random Variable Y
        # Keep track of number of successes
        if (first == 1) and (second == 1):
            head_both += 1 # num of tosses with X=Y=heads
            head_c1 += 1 # num of tosses with X=head
            head_c2 += 1 # num of tosses with Y=head
        if (first == 1) and (second == 0):
            head_c1notc2 += 1 # num of tosses with X=heads, Y=tails
            head_c1 += 1 # num of tosses with X=head
        if (first == 0) and (second == 1):
            head_c2notc1 += 1 # num of tosses with Y=heads, X=tails
            head_c2 += 1 # num of tosses with Y=head
        if (first == 0) and (second == 0):
            head_none += 1 # num of tosses with X=tails, Y=tails
    print("... done")
```

Question 3 continued on next page...
head_none += 1  # num of tosses with X=Y=tails

# Relative frequency of the event: X=1, Y=1
P_head_both = head_both/n

# Relative frequency of the event: X=1, Y=0
P_head_1not2 = head_c1notc2/n

# Relative frequency of the event: Y=1, X=0
P_head_2not1 = head_c2notc1/n

# Relative frequency of the event: Y=0, X=0
P_head_none = head_none/n

# Relative frequency of heads for the first coin (X=1)
P_head1 = head_c1 /n

# Relative frequency of heads for the second coin (Y=1)
P_head2 = head_c2 /n

# Calculate Expectations

# THEORY
E_x = prob1  # Expected Value of X (E(X))
E_y = prob2  # Expected Value of Y (E(Y))
E_xy = E_x * E_y  # Expected Value of XY (E(XY))

# EXPERIMENT
M_x = P_head1  # Mean Value of X (M(X))
M_y = P_head2  # Mean Value of Y (M(Y))
M_xy = head_both /n  # Mean Value of XY (M(XY))

print ("M(X) = %f" % M_x)
print ("M(Y) = %f" % M_y)
print ("P(X=1 ,Y =1) - P(X =1) *P(Y =1) = %f" % (P_head_both - P_head1 * P_head2))
print ("P(X=1 ,Y =0) - P(X =1) *P(Y =0) = %f" % (P_head_1not2 - P_head1 * (1 - P_head2)))
print ("P(X=0 ,Y =1) - P(X =0) *P(Y =1) = %f" % (P_head_2not1 - (1 - P_head1) * P_head2))
print ("P(X=0 ,Y =0) - P(X =0) *P(Y =0) = %f" % (P_head_none - (1 - P_head1) * (1 - P_head2)))
print ("M(XY) - M(X)*M(Y)= %f" % (M_xy - M_x * M_y))

We had the following output for n = 1000000:

Running 1000000 coin flips...
M(X) = 0.300611
M(Y) = 0.900021
P(X=1,Y=1) - P(X=1) * P(Y=1) = -0.000050
P(X=1,Y=0) - P(X=1) * P(Y=0) = 0.000050
P(X=0,Y=1) - P(X=0) * P(Y=1) = 0.000050
P(X=0,Y=0) - P(X=0) * P(Y=0) = -0.000051
M(XY) - M(X) * M(Y) = -0.000050

The lines 2 and 3, just give us a measure of convergence, i.e. if the number of iterations chosen is big enough in order to obtain small errors. As we can see, in our case the difference between the mean and expected value is of order 0.001, a considerably small number. The lines 4-7 justify the first result of 3(a), as the maximum difference obtained between P(X, Y) and P(X) * P(Y) is just of order 0.0001. Finally, the last line illustrates the second result of 3(a) again with error 0.0001.

Question 3 continued on next page...
(c) This time we rewrite the above code to make $Y$ depend on $X$, (lines 23-27):

```python
from __future__ import division
import random

# Definition of coin with success probability p:
def coin_flip(prob):
    return 1 if random.random() < prob else 0

if __name__ == '__main__':
    head_c1, head_c2 = 0, 0
    head_c1notc2, head_c2notc1, head_both, head_none = 0, 0, 0, 0
    prob1 = 0.3 # Success Probability of X
    prob2a = 0.9 # Success Probability of Y given X=heads
    prob2b = 0.2 # Success Probability of Y given X=tails
    n = 1000000 # Number of tosses

    print("Running %d coin flips..." % n)
    for i in range(1, n):
        # Generating two random coins
        first = coin_flip(prob1) # Random Variable X
        if first == 1:
            second = coin_flip(prob2a) # Random Variable Y|X=heads
        if first == 0:
            second = coin_flip(prob2b) # Random Variable Y|X=tails

        # Keep track of number of successes
        if (first == 1) and (second == 1):
            head_both += 1 # num of tosses with X=Y=heads
            head_c1 += 1 # num of tosses with X=head
            head_c2 += 1 # num of tosses with Y=head
        if (first == 1) and (second == 0):
            head_c1notc2 += 1 # num of tosses with X=heads, Y=tails
            head_c1 += 1 # num of tosses with X=head
        if (first == 0) and (second == 1):
            head_c2notc1 += 1 # num of tosses with Y=heads, X=tails
            head_c2 += 1 # num of tosses with Y=heads
        if (first == 0) and (second == 0):
            head_none += 1 # num of tosses with X=Y=tails

    # Relative frequency of the event: X=1, Y=1
    P_head_both = head_both / n

    # Relative frequency of the event: X=1, Y=0
    P_head_1not2 = head_c1notc2 / n

    # Relative frequency of the event: Y=1, X=0
    P_head_2not1 = head_c2notc1 / n

    # Relative frequency of the event: Y=0, X=0
    P_head_none = head_none / n

    # Relative frequency of heads for the first coin (X=1)
    P_head1 = head_c1 / n
```

Question 3 continued on next page...
# Relative frequency of heads for the second coin (Y=1)
P_head2 = head_c2/n

# Calculate Expectations

# THEORY
E_x = prob1   # Expected Value of X (E(X))
E_y = prob2a*prob1+prob2b*(1-prob1)   # Expected Value of Y (E(Y))
E_xy = prob2a*prob1   # Expected Value of XY (E(XY))

# EXPERIMENT
M_x = P_head1   # Mean Value of X (M(X))
M_y = P_head2   # Mean Value of Y (M(Y))
M_xy = P_head_both   # Mean Value of XY (M(XY))

print ("M(X) = %f" % M_x)
print ("E(Y) - M(Y) = %f" % (E_y - M_y))
print ("P(X=1,Y=1) - P(X=1)*P(Y=1) = %f" % (P_head_both - P_head1*P_head2))
print ("P(X=1,Y=0) - P(X=1)*P(Y=0) = %f" % (P_head_1not2 - P_head1*(1-P_head2)))
print ("P(X=0,Y=1) - P(X=1)*P(Y=0) = %f" % (P_head_2not1 - (1-P_head1)*P_head2))
print ("P(X=0,Y=0) - P(X=0)*P(Y=0) = %f" % (P_head_none - (1-P_head1)*(1-P_head2)))
print ("E(XY) - E(X)E(Y) = %f" % (E_xy - E_x*E_y))
print ("M(XY) - M(X)*M(Y) = %f" % (M_xy - M_x*M_y))

And we get the following output for n = 1000000:

Running 1000000 coin flips...
M(X) = 0.300500
E(Y) - M(Y) = -0.000789
P(X=1,Y=1) - P(X=1)*P(Y=1) = 0.146769
P(X=1,Y=0) - P(X=1)*P(Y=0) = -0.146769
P(X=0,Y=1) - P(X=1)*P(Y=0) = -0.146769
P(X=0,Y=0) - P(X=0)*P(Y=0) = 0.146768
E(XY) - E(X)E(Y) = 0.147000
M(XY) - M(X)*M(Y) = 0.146769

As in the previous question 3(b), line 2 is gives us the level of convergence. Line 3 can be seen as a validation of the total probability rule. In lines 4-7, 8 we observe that the first and second formula of 3(a) no longer hold.