2/16/93

Course 4.902: Pattern Recognition and Analysis
Solutions to Problem Set 1

Solution to Exercise 1: (2 points)

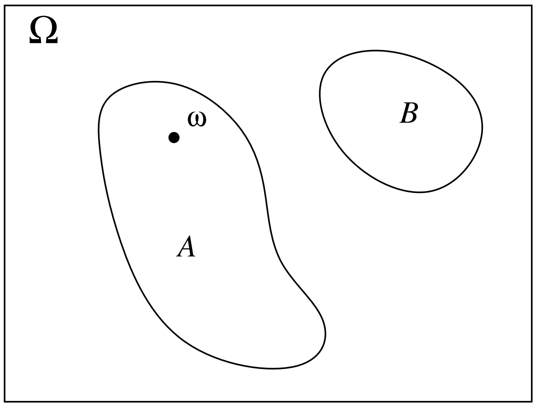
A single outcome of the experiment is a sequence of ten coin flips. Since each coin flip has two possible results (heads or tails), there are $2\times2\times2\times2\times2\times2\times2\times2\times2\times2 = 2^{10} = 1024$ possible outcomes.

Solution to Exercise 2: (3 points)

The answer we were looking for was (b), definitely not independent. This question really should have been stated a bit more carefully to rule out the possibility of one or both events having probability zero, for in that case, the events would in fact be independent (vacuously). Students who chose (c) instead of (b) and cited this explanation will not lose any points.

Suppose $\Pr(A)$ and $\Pr(B)$ are both nonzero. Assume that A and B *are* independent; we will show that this leads to a contradiction. Because A and B are disjoint, B cannot occur when A occurs, so that $\Pr(B|A) = 0$. But by independence, $\Pr(B|A) = \Pr(B)$, so that $\Pr(B) = 0$, which is the desired contradiction.

This can also be seen with the aid of the following Venn diagram.



If event A occurs, it means that some elementary outcome ω has occurred. But since this ω is not in B, event B cannot occur. Thus the occurrence of A tells us a great deal about the probability of B: it tells us that B is impossible. Hence the events are not independent.

Solution to Exercise 3: (5 points)

The probability that the sum is 9 is $25/216 \approx 0.116$; the probability that the sum is 10 is $27/216 = 0.125$. Getting these numbers is a straightforward exercise in counting. First, observe that the number of possible outcomes when three dice are rolled is $6 \times 6 \times 6 = 216$. Now let's count the number of outcomes for which the sums are 9 and 10:

Sum = 9		Sum = 10	
Outcome	Permutations	Outcome	Permutations
1, 2, 6	6	1, 3, 6	6
1, 3, 5	6	1, 4, 5	6
1, 4, 4	3	2, 2, 6	3
2, 2, 5	3	2, 3, 5	6
2, 3, 4	6	2, 4, 4	3
3, 3, 3	1	3, 3, 4	3
total:	25	total:	27

Dividing each of these totals by 216 yields the desired probabilities.

According to Martin Eisen in *Introduction to Mathematical Probability Theory* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969), Galileo solved this problem for a friend in the early seventeenth century. His friend was puzzled after observing that the number 10 seemed to come up more often than 9. He reasoned (erroneously) that since both 9 and 10 can be formed in six different ways, they should be equally likely. Galileo resolved the problem by carefully counting the number of outcomes favorable to 9 and 10, just as we have done above.

Solution to Exercise 4: (10 points)

We are told that $\mathbf{x}$ is a random vector with a continuous and differentiable distribution function $P_{\mathbf{x}}(\hat{\mathbf{x}})$.

a) The density function $p_{\mathbf{x}}(\hat{\mathbf{x}})$ can certainly be greater than 1 for some $\hat{\mathbf{x}}$.

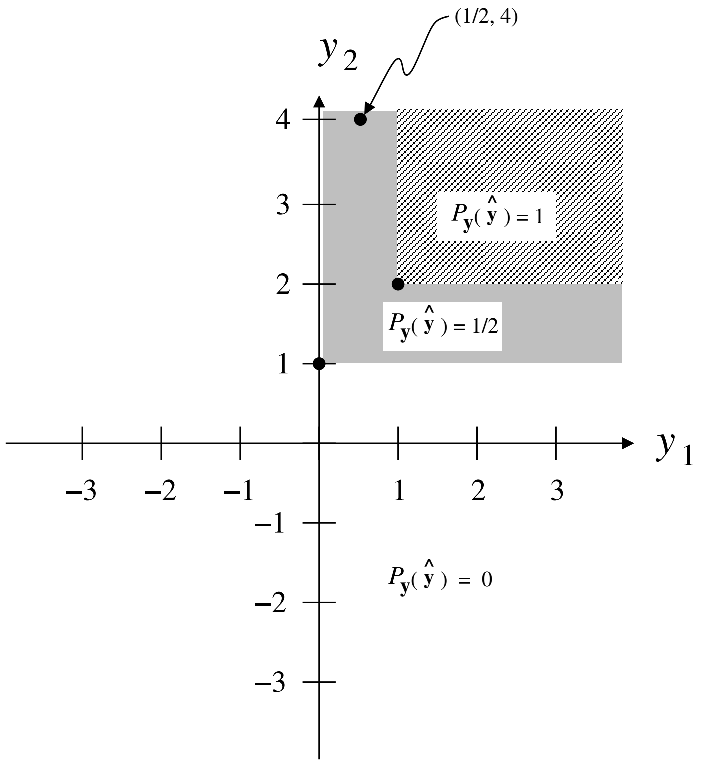
b) On the other hand, $P_{\mathbf{x}}(\hat{\mathbf{x}})$ is defined to be the probability of an event (namely, the event $\mathbf{x} \leq \hat{\mathbf{x}}$), and therefore cannot be greater than 1.

c) Since $\mathbf{x}$ is continuous, the probability that it assumes any particular value is zero. See the end of section 2.2.1 on p. 17 of the text for details.

d) Here, we have a discrete random vector $\mathbf{y}$, which takes on the two values $(1,2)$ and $(0,1)$, each with probability $1/2$. The distribution function, shown below, is obtained by applying the definition

$$\begin{aligned} P_{\mathbf{y}}(\hat{\mathbf{y}}) &= \Pr[\mathbf{y} \leq \hat{\mathbf{y}}] \\ &= \Pr[y_1 \leq \hat{y}_1 \;\text{ and }\; y_2 \leq \hat{y}_2] \end{aligned}$$

to all points in $\hat{y}_1 - \hat{y}_2$ plane. From the graph or the definition, it can be seen that the value of the distribution function $P_{\mathbf{y}}(\hat{\mathbf{y}})$ at the point $\hat{\mathbf{y}} = (1/2, 4)$ is $1/2$.



Solution to Problem 1: (20 points)

In this problem we derive some elementary properties of expectation, correlation, and variance.

We are told that the random variables x and y are discrete, which means that our work will involve sums instead of integrals. However, all of the results hold for continuous random variables as well, and the derivations in that case are analagous. Therefore, no points will be taken off for otherwise correct solutions which tackle the continuous case.

Since x and y are discrete, we can define $X = \{\hat{x}_1, \hat{x}_2, \hat{x}_3, \ldots\}$ to be the set of possible values for x. Similarly, let $Y = \{\hat{y}_1, \hat{y}_2, \hat{y}_3, \ldots\}$ be the corresponding set for

y. For any set A, let $|A|$ denote the number of elements in A. Let $p_x(i)$ denote the probability that $x = \hat{x}_i$; let $p_y(i)$ denote the probability that $y = \hat{y}_i$. Let $p_{x,y}(i,j)$ denote the probability of the joint event, "$x = \hat{x}_i$ and $y = \hat{y}_j$." Then it follows that

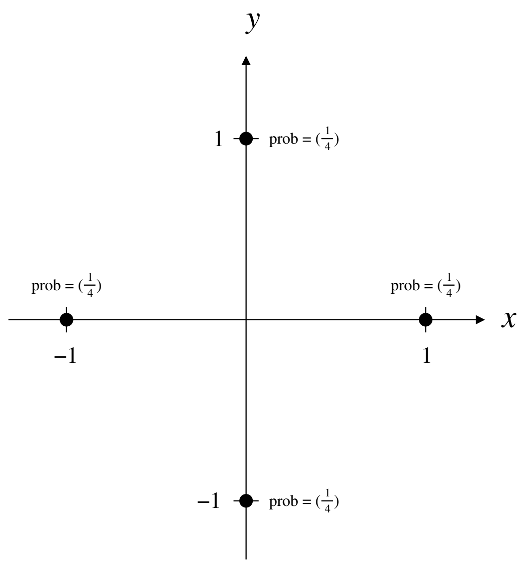
a)

$$\begin{aligned}
E(x+y) &= \sum_{i=1}^{|X|}\sum_{j=1}^{|Y|}(\hat{x}_i + \hat{y}_j)p_{x,y}(i,j) \\
&= \sum_{i=1}^{|X|}\sum_{j=1}^{|Y|}[\hat{x}_i p_{x,y}(i,j) + \hat{y}_j p_{x,y}(i,j)] \\
&= \sum_{i=1}^{|X|}\hat{x}_i[\sum_{j=1}^{|Y|}p_{x,y}(i,j)] + \sum_{j=1}^{|Y|}\hat{y}_j[\sum_{i=1}^{|X|}p_{x,y}(i,j)] \\
&= \sum_{i=1}^{|X|}\hat{x}_i p_x(i) + \sum_{j=1}^{|Y|}\hat{y}_j p_y(j) \\
&= E(x) + E(y).
\end{aligned}$$

b) If x and y are independent, then $p_{x,y}(i,j) = p_x(i)p_y(j)$, so that

$$\begin{aligned}
E(xy) &= \sum_{i=1}^{|X|}\sum_{j=1}^{|Y|}\hat{x}_i\hat{y}_j p_{x,y}(i,j) \\
&= \sum_{i=1}^{|X|}\sum_{j=1}^{|Y|}\hat{x}_i\hat{y}_j p_x(i)p_y(j) \\
&= \sum_{i=1}^{|X|}\hat{x}_i p_x(i)[\sum_{j=1}^{|Y|}\hat{y}_j p_y(j)] \\
&= [\sum_{i=1}^{|X|}\hat{x}_i p_x(i)][\sum_{j=1}^{|Y|}\hat{y}_j p_y(j)] \\
&= E(x)E(y).
\end{aligned}$$

c) Independence implies uncorrelatedness (as we have just shown), but the converse is not true. Independence is a stronger condition than uncorrelatedness; x and y can be uncorrelated without being independent. Many examples are possible. One simple example is given by the following joint probability mass function:



To see that x and y are not independent, note that $\Pr[x = 1|y = 1] = 0$, while $\Pr[x = 1] = 1/4$.

Now we show that x and y are uncorrelated. Consider the expected value of the product xy. By inspection, we have

$$\begin{aligned} E(xy) &= 1/4\big[(-1)(0) + (0)(1) + (0)(-1) + (1)(0)\big] \\ &= 0. \end{aligned}$$

Also by inspection,

$$\begin{aligned} E(x) &= (1/4)(-1) + (1/2)(0) + (1/4)(1) \\ &= 0 \\ &= E(y). \end{aligned}$$

Thus x and y are uncorrelated, but not independent.

d) For independent x and y, we wish to show that $\mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(y)$. From part (a), we have $E(x+y) = E(x) + E(y)$, and by independence and part (b), we have $E(xy) = E(x)E(y)$. Thus

$$\begin{aligned}
\mathrm{Var}(x+y) &= E\big[(x+y)^2\big] - \big[E(x+y)\big]^2 \\
&= E\big[(x+y)^2\big] - \big[E(x) + E(y)\big]^2 \\
&= E(x^2) - 2E(xy) + E(y^2) - \big[E(x)\big]^2 + 2E(x)E(y) - \big[E(y)\big]^2 \\
&= E(x^2) - 2E(x)E(y) + E(y^2) - \big[E(x)\big]^2 + 2E(x)E(y) - \big[E(y)\big]^2 \\
&= E(x^2) - \big[E(x)\big]^2 + E(y^2) - \big[E(y)\big]^2 \\
&= \mathrm{Var}(x) + \mathrm{Var}(y).
\end{aligned}$$

Solution to Problem 2: (20 points)

Let L be the event that the selected die is loaded; let F be the event that it is fair. We are told that the prior probabilities for these events are

$$\Pr(L) = \Pr(F) = 1/2.$$

We are also given the conditional probabilities for a single roll,

$$\begin{aligned}
p(i|F) &= 1/6; \qquad i = 1, \ldots, 6; \\
p(i|L) &= \begin{cases} 1/15 & \text{if } i = 1, \ldots, 5; \\ 2/3 & \text{if } i = 6. \end{cases}
\end{aligned}$$

In part (a), the basic experiment is to roll the die once. Call the outcome of this roll i. We are asked to find $\Pr[L|i=6]$, the conditional probability of L given $i = 6$. We can use Bayes rule if we happen to remember it; if we not, we can derive it very easily from two applications of the definition of conditional probability:

$$\Pr[L|i=6] = \frac{\Pr[L \text{ and } i = 6]}{\Pr[i=6]}$$

$$= \frac{\Pr[i=6|L]\Pr[L]}{\Pr[i=6]}.$$

We already know the two numerator terms; we need to find $\Pr[i=6]$. Since L and F are disjoint, so are the events "L and $i=6$" and "F and $i=6$". The probability of the union of disjoint events is the sum of the probabilities of the events (this is one of the axioms of probability). Also, L and F are collectively exhaustive, so that

$$\begin{aligned}
\Pr[i=6] &= \Pr[(L \text{ and } i=6) \cup (F \text{ and } i=6)] \\
&= \Pr[L \text{ and } i=6] + \Pr[F \text{ and } i=6] \\
&= \Pr[6|L]\Pr[L] + \Pr[6|F]\Pr[F] \\
&= (2/3)(1/2) + (1/6)(1/2) \\
&= 5/12.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pr[L|i=6] &= \frac{\Pr[i=6|L]\Pr[L]}{\Pr[i=6]} \\
&= \frac{(2/3)(1/2)}{5/12} \\
&= 4/5.
\end{aligned}$$

In part (b), the basic experiment is to roll the die twice. Let i denote the outcome of the first roll and let j denote the outcome of the second. We wish to find $\Pr[L|i=6 \text{ and } j \neq 6]$. Following the same approach as we used in part (a), we can write

$$\Pr[L|i=6 \text{ and } j \neq 6] = \frac{\Pr[L \text{ and } i=6 \text{ and } j \neq 6]}{\Pr[i=6 \text{ and } j \neq 6]} \tag{$*$}$$

There is a slight ambiguity here (and in the problem set handout) about whether we are to regard "$j \neq 6$" as the set of all values for j besides 6, or alternatively, as some particular value of j other than six. Fortunately, in this problem it makes no difference in terms of the final answer.

For now, let us assume that "$j \neq 6$" refers to the set of all values for j besides 6. Then

$$\begin{aligned}
\Pr[L \text{ and } i=6 \text{ and } j \neq 6] &= \Pr[L]\Pr[i=6 \mid L]\Pr[j \neq 6 \mid L \text{ and } i=6] \\
&= \Pr[L]\Pr[i=6 \mid L]\Pr[j \neq 6 \mid L] \\
&= (1/2)(2/3)(5/15) \\
&= 1/9.
\end{aligned}$$

This gives us the numerator in the right-hand side of $(*)$. To find the denominator, we use the same reasoning as we used in part (a), and write

$$\begin{aligned}
\Pr[i=6 \text{ and } j \neq 6] &= \Pr[L \text{ and } i=6 \text{ and } j \neq 6] + \Pr[F \text{ and } i=6 \text{ and } j \neq 6] \\
&= (1/2)(2/3)(5/15) + (1/2)(1/6)(5/6) \\
&= 8/72 + 5/72 \\
&= 13/72.
\end{aligned}$$

Thus, we have

$$\Pr[L|i = 6 \text{ and } j \neq 6] = \frac{1/9}{13/72}$$

$$= 8/13.$$

If instead we were to interpret "$j \neq 6$" to mean j equals some particular number not equal to 6, say $j = 1$, then we would obtain the same numerator and denominator in the right-hand side of $(*)$, except that each would be 1/5th as large, giving us precisely the same final answer (as you can easily verify).

Solution to Problem 3: (20 points)

This was in some respects the most difficult problem in the problem set. The main difficulty comes from our having to deal with both discrete and continous quantities in the same experiment. We will address this difficulty in our discussion of part (b).

In part (a) we are asked to find the expected value of y in terms of $\hat{x}$. We know that y can assume one of two values: either $y = 2\hat{x}$ or $y = \hat{x}/2$. The temptation is to say that, since we know nothing more about y, both of the possible values are equiprobable. Then

$$\begin{aligned} E(y|\hat{x}) &= 1/2(2\hat{x} + \hat{x}/2) \\ &= 5/4\hat{x}. \end{aligned}$$

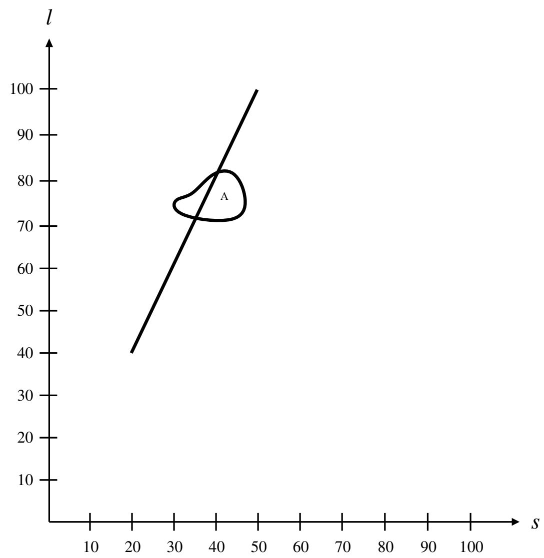
Perhaps this result doesn't bother you. If so, let us give the "expected value" some meaning by considering what happens if we repeat the experiment many times. Collect all of the chosen envelopes into one pile, and all of the unchosen envelopes into another. By a double application of the weak law of large numbers, the total in the unchosen pile will, with near certainty, be close to 5/4 of the total in the chosen pile. This of course is contrary to our intuition, which says that the two totals ought to be nearly the same. One way out of this dilemma is to assign probabilities to $2\hat{x}$ and $\hat{x}/2$ in such a way that $E(y|\hat{x}) = \hat{x}$. This is not to say that the 5/4 answer is wrong. In fact, there is no correct answer to part (a) — our answer will depend on what we think is a "rational" assignment of prior probabilities, and probability theory doesn't tell us how to find one.

The critical step in working out part (b) is obtaining and sketching the joint density for x and y. We need to generalize a bit the notion of "density" to accomodate the discrete aspect of the problem.

Our approach consists of two steps. First, we consider the joint density of the smaller and larger amount, before the envelopes are randomly labeled. We then derive the joint density for x and y from this initial density.

Let s denote the smaller amount, and let $l = 2s$ denote the larger. We know that s is uniformly distributed between 20 and 40, so that the joint density for s and l

is an infinitely thin ridge, located along the line segment indicated in the following picture.



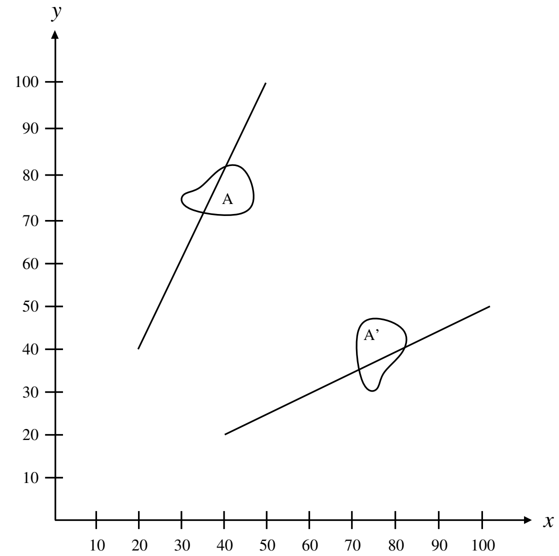
To complete the description of the joint density for s and l, we must determine the "height" of the ridge at every point. To do this, we must first extend our idea of density function to include those which cover zero area, but which still integrate to unity as required. Thus when we speak of "height," we are no longer speaking of the actual value of the function at a point, but rather of its "integration strength" at that point. This is the concept behind an impulse "function" — although we cannot say what value it takes on at every point, we can say what it integrates to over every interval. In this case, since all values of s between 20 and 50 are equally likely, the "height" or "integration strength" of the ridge is uniform along the ridge. Thus, the probability that (s, l) falls into some subset of the plane A, is simply proportional to the length of the intersection of the indicated segment with A.

Now let us consider the joint density for x and y. Again, let A be some subset of the plane, and let $A' = \{(a, b) : (b, a) \in A\}$. Then

$$\begin{aligned}
\Pr[(x, y) \in A] &= \Pr[(s, l) \in A \text{ and } s = x, l = y] + \Pr[(l, s) \in A \text{ and } l = x, s = y] \\
&= \Pr[(s, l) \in A \text{ and } s = x, l = y] + \Pr[(s, l) \in A' \text{ and } l = x, s = y] \\
&= 1/2\Pr[(s, l) \in A] + 1/2\Pr[(s, l) \in A'] \\
&= 1/2\Pr[(s, l) \in A'] + 1/2\Pr[(s, l) \in A] \\
&= \Pr[(x, y) \in A']
\end{aligned}$$

Since the probabilities of landing in A and A' are the same, the pdf for x and y

must look as follows:

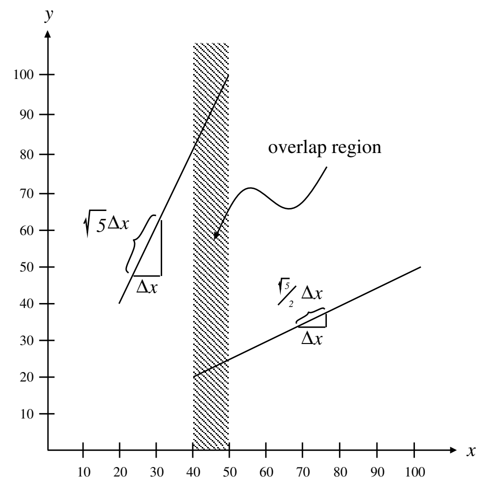


To find the probability of any event in the xy-plane, we need only measure the total length of the intersections of the segments with that event, then divide by the total length of both segments (which is seen to be $60\sqrt{5}$). This picture gives the most basic description of the random experiment, and we can obtain from it whatever specific quantities we might wish to determine about the experiment.

We are asked first to find $p_x(\hat{x})$, the marginal density for x. From the picture that follows, we see that

$$\Pr[\hat{x} < x \leq \hat{x} + \Delta x] = \begin{cases} 1/60\Delta x & \text{if } 20 < \hat{x} \leq 40; \\ (1/60 + 1/120)\Delta x & \text{if } 40 < \hat{x} \leq 50; \\ 1/120\Delta x & \text{if } 50 < \hat{x} \leq 100; \\ 0 & \text{otherwise.} \end{cases}$$

$$\approx p_x(\hat{x})\Delta x$$



Dividing by Δx and noting that the approximation becomes exact as Δx approaches zero, we have

$$p_x(\hat{x}) = \begin{cases} 1/60 & \text{if } 20 < \hat{x} \leq 40; \\ 1/40 & \text{if } 40 < \hat{x} \leq 50; \\ 1/120 & \text{if } 50 < \hat{x} \leq 100; \\ 0 & \text{otherwise,} \end{cases}$$

which is the desired answer.

Next, we are asked to find $p_{y|x}(\hat{y}|\hat{x})$, the conditional probability density function for y given $\hat{x}$. In each of the two regions $20 < \hat{x} \leq 40$ and $50 < \hat{x} \leq 100$, y takes on only one possible value in terms of $\hat{x}$. In the overlap region, y assumes one of two possible values, either $\hat{x}/2$ or $2\hat{x}$. We must determine the probability of each. The crucial observation here is that whenever $\hat{x}$ is in some small interval $(a, b) \subset (40, 50]$, the length of the upper intersected segment is twice that of the lower, so that $\Pr[y \in (2a, 2b)] = 2\Pr[y \in (a/2, b/2)]$. Using the notation $\delta(t)$ to denote an "impulse function," i.e., $\delta(t) = 0$ whenever $t \neq 0$ and $\int_a^b \delta(t)dt = 1$ whenever $a < 0 < b$, we have

$$p_{y|x}(\hat{y}|\hat{x}) = \begin{cases} \delta(\hat{y} - 2\hat{x}) & \text{if } 20 < \hat{x} \leq 40; \\ 2/3\delta(\hat{y} - 2\hat{x}) + 1/3\delta(\hat{y} - \hat{x}/2) & \text{if } 40 < \hat{x} \leq 50; \\ \delta(\hat{y} - \hat{x}/2) & \text{if } 50 < \hat{x} \leq 100. \end{cases}$$

Finally, we are asked to find $E(y|\hat{x})$, the conditional expected value for y given $\hat{x}$. We use the definition of conditional expected value,

$$E(y|\hat{x}) = \int_{-\infty}^{\infty} y p_{y|x}(\hat{y}|\hat{x}) dy,$$

to obtain

$$E(y|\hat{x}) = \begin{cases} 2\hat{x} & \text{if } 20 < \hat{x} \leq 40; \\ 4/3\hat{x} + 1/6\hat{x} = 3/2\hat{x} & \text{if } 40 < \hat{x} \leq 50; \\ \hat{x}/2 & \text{if } 50 < \hat{x} \leq 100. \end{cases}$$

We can check our answer by making sure that $E[E(y|\hat{x})]$ is equal to $E(y)$. By evaluating the integrals in both expressions over each region, it is easy to see that $E[E(y|\hat{x})] = E(y) = E(x) = 52.5$.

Solution to Problem 4: (20 points)

We have a sequence of 12 independent, identical Bernoulli trials, with $\Pr[B] = 2/3$ and $\Pr[G] = 1/3$. In part (a), we are asked which of the following sequences seems most "believable:"

$$\begin{array}{rl} \text{i)} & B\ G\ B\ G\ G\ B\ G\ B\ B\ G\ G\ B \\ \text{ii)} & B\ B\ B\ B\ B\ B\ B\ B\ B\ B\ B\ B \\ \text{iii)} & B\ G\ B\ B\ B\ B\ G\ B\ B\ G\ B\ G \end{array}$$

There is no right answer to this part of the problem; justifications are possible for any of the three choices. For instance, sequence (iii) is consistent with the "relative frequency" notion of probability – the number of B's is 2/3 of the total. However, this might seem a little *too* perfect, and sequence (i) might therefore strike you as more believable. Alternatively, we can argue that *some* particular sequence had to occur, and since (ii) is the most probable particular sequence, it is eminently believable.

In part (b), we are asked to compute the probability of each of the three sequences. Let k denote the number of B's in a sequence (so that $12 - k$ is the number of G's). Since the die rolls are independent, we have

$$\Pr[\text{a particular sequence having exactly } k \ B\text{'s}] = (2/3)^k (1/3)^{12-k}.$$

Applying this formula to each of the three sequences, we get

$$\begin{aligned} \Pr[\text{sequence(i)}] &= (2/3)^6 (1/3)^6 \\ &\approx 0.00012; \\ \Pr[\text{sequence(ii)}] &= (2/3)^{12} (1/3)^0 \\ &\approx 0.00771; \\ \Pr[\text{sequence(iii)}] &= (2/3)^8 (1/3)^4 \\ &\approx 0.00048; \end{aligned}$$

Thus, sequence (ii) has the greatest probability.

In part (c), we are concerned with the probabilities of *classes* of sequences, where sequences are in class k if they have exactly k B's. Since the probability of every sequence in class k is $(2/3)^k (1/3)^{12-k}$, the probability of the whole class is this common sequence probability times the number of sequences in the class. Note that the number of sequences in class k is equal to the number of ways we can select k items from a list of 12 items, without paying attention to ordering. This number, denoted $\binom{12}{k}$, is given by

$$\binom{12}{k} = \frac{12!}{(12-k)!k!}.$$

Thus, the probability of class k is

$$\Pr[\text{set of all sequences having exactly } k\ B\text{'s}] = \frac{12!}{(12-k)!k!}(2/3)^k(1/3)^{12-k}.$$

This probability is tabulated against k in the following table.

Sequence Class Probabilities

k	approx. probability
0	0.00000188
1	0.0000452
2	0.000497
3	0.00331
4	0.0149
5	0.0476
6	0.111
7	0.190
8	0.238
9	0.212
10	0.127
11	0.0462
12	0.00770

From this table, we can see that sequence (iii) is in the most probable class, which one might argue makes it more "believable" than the other two sequences, which are in less probable classes.

Solution to Puzzle: (0 points)

It is better to trade.

To see this, consider first the probability of finding the money if you don't trade. In this case, you stick with your initial choice no matter what Monty says or does, and your chances are 1 in 3 that you have chosen right box.

Now consider the probability of finding the money if your strategy is to trade. If the money happens to be in the box you initially choose, that's bad, because you will trade it away. However, if the money happens to be in one of the other two boxes, you will definitely find it by trading, since Monty will eliminate the empty box for you. Thus, the probability of finding the money is equal to the probability that the money is *not* in the box you initially choose. This probability is 2/3.

So by trading, you actually double your probability of finding the money.