Solutions

1.5.1 \( P \subseteq Z \subseteq Q \subseteq \mathbb{R} \).

1.5.2 \( A \cap B = \{(i, j) : i^2 + j^2 \leq 9\} \), where \( i, j \) are integers. The set \( A \cap B^c \) looks like a disk of radius 3, centered at the origin, with holes at the locations \((\pm i, \pm j)\) where \( i^2 + j^2 \leq 9 \), and \( i, j \) are integers.

1.5.3 The statement \( x \in (A \cup B)^c \) is equivalent to saying that \( x \not\in A \cup B \). This is equivalent to saying that \( x \not\in A \) and \( x \not\in B \). Or equivalently, \( x \in A^c \) and \( x \in B^c \). This is the same statement as \( x \in A^c \cap B^c \).

1.5.4 \( A \times B \times C = \{(a, u, v), (u, 1, v), (v, 1, v), (v, 1, u), (u, 1, 1), (v, u, 1), (v, 1, 1)\} \).

1.5.5 One partition is \( \{\{1, 2\}, \{4\}\} \). Another is \( \{\{1\}, \{2, 4\}\} \). (Thanks, Mr. Bing Yu.)

1.6.0 \( \mathcal{P}(S) = \{\emptyset, S, \{1\}, \{2\}, \{4\}, \{1, 2\}, \{1, 4\}, \{2, 4\}\} \).

1.6.6 False. For instance, take \( A = [0, 1] \) and \( B = [1, 2] \). Now \( A \cap B = \{1\} \), which is a finite set.

1.6.8 Take \( A_i = \{i\} \) for \( i = 1, 2, \ldots \). Clearly, \( \bigcup_i A_i \) is a countably infinite set.

1.6.9 Take \( A_i = \{1, 2\} \) if \( i = 2, 4, 6, \ldots \), and take \( A_i = \{0, 1\} \) for \( i = 3, 5, \ldots \). Now \( \bigcap_i A_i = \{1\} \).

2.5.1 A Matlab code that simulates the event is given below.

clear;
N=10000; \text{ \% Number of rolls of a pair of fair dice.} 
x=rand([N,12]); \text{ \% 12 columns of uniform deviates} 
x=6*x+1; \text{ \% make rand values fall in [1, 7].} 
z=floor(x); \text{ \% z is the face values of 6 rolls of a pair of dice.} 
w(1)=0;
if ((z(1,1)==6) \& (z(1,2)==6)) | ((z(1,3)==6) \& (z(1,4)==6)) | ... 
((z(1,5)==6) \& (z(1,6)==6)) | ((z(1,7)==6) \& (z(1,8)==6)) | ... 

2.5.2 The sample space consists of sixteen \((2^4)\) elements listed below.

\( (HHHH), (HHHT), (HTHH), (HTHT), (THHH), (HHTT), \)
\( (HTHT), (THTT), (TTHT), (TTTH), (HTTH), (HTTT), \)
\( (THHT), (TTHT), (TTTT) \)

2.5.3 One simple example is the set of all subsets of the sample space (i.e., the power set). Another sigma field is the smallest sigma field, namely the sample
space along with the empty set. Here is another one that lies in between the two extremes.

\[ \{S, A, A^c, \emptyset\} \]

where \( A = \{(THHT), (TTHT), (TTTH), (TTTT)\} \).

2.5.4 If we use the power set, then we may define a probability measure by making each outcome equally likely. If we choose, the sigma field,

\[ \{S, A, A^c, \emptyset\} \]

where \( A = \{(THHT), (TTHT), (TTTH), (TTTT)\} \), then we could define a probability measure by \( P(S) = 1, P(\emptyset) = 0, P(A) = \frac{4}{16} \) and \( P(A^c) = \frac{12}{16} \).

2.5.5 The event, “at least one of the three events will occur”, can be written as \( A \cup B \cup C \). Therefore, by part (iv) of Theorem (2.1.1),

\[ P(A \cup (B \cup C)) = P(A) + P(B \cup C) - P(A \cap (B \cup C)) \]

Applying the same argument again, we see that

\[ P(B \cup C) = P(B) + P(C) - P(B \cap C), \]
\[ P(A \cap (B \cup C)) = P(A \cap B) + P(A \cap C) - P(A \cap B \cap C). \]

Putting all of these together gives the result.

2.5.6 \( P(\{c\}) = \frac{4}{16} \) and \( P(\{a, c\}) = \frac{8}{16} = \frac{2}{3} \).

2.5.7 \( P(\{c\}) = \frac{1}{4} \). If \( P(\{c\}) = \frac{1}{16} \) then the sample space is not simple.

2.5.8 By the multiplication principle, the total number of ways is \( 5 \times 6 \times 7 = 210 \).

2.5.9 By the multiplication principle, in \((365)^7\) ways.

2.5.10 By the multiplication principle, the total number of elements in the equally likely sample space is \((365)^7\). The total number of ways they all could have different birth days is

\[ 365 \times 364 \times 363 \times 362 \times 361 \times 360 \times 359. \]

So, the probability that they all have different birth days is

\[ \frac{365 \times 364 \times 363 \times 362 \times 361 \times 360 \times 359}{(365)^7}. \]

2.5.11 \( \binom{3}{2} \binom{13}{13} + \binom{7}{13} \binom{13}{13} \).

2.5.12 \( \binom{2}{2} \binom{2}{2} + \binom{2}{2} \binom{1}{2} \).

2.5.13 There are \( \binom{3}{13} \) ways of drawing two spades. So, the probability is \( \frac{\binom{3}{13}}{\binom{1}{2}} \).

2.5.14 The exact answer is \( 1 - (35/36)^6 \). The comparison is quite good.

2.5.15 The probability is \( \frac{6}{9} = \frac{1}{3} \). The comparison is quite good.

2.5.16 A Matlab code which does the simulation is as follows.

```matlab
clear;
N=10000;  \% Number of rolls of 6 dice.
x=rand([N,6]); \% 6 columns of uniform deviates
x=6*x+1; \% make rand values fall in [1,7].
z=floor(x); \% z is the face values of 6 rolls of a fair die.
w(1) = 0;
for i=1:n-1
    if((z(i,1) == z(i,2)) \& (z(i,1) == z(i,3)) \& (z(i,1) == z(i,4)) \& ...
        (z(i,1) == z(i,5)) \& (z(i,1) == z(i,6)) \& (z(i,2) == z(i,3)) \& ...
        (z(i,2) == z(i,4)) \& (z(i,2) == z(i,5)) \& (z(i,2) == z(i,6)) \& ...
        (z(i,3) == z(i,4)) \& (z(i,3) == z(i,5)) \& (z(i,3) == z(i,6)) \& ...
        (z(i,4) == z(i,5)) \& (z(i,4) == z(i,6)) \& (z(i,5) == z(i,6)) \& ...
    )
        w(1) = w(1) + 1/i; \% adds up the values.
    else
        w(i) = w(i); \% added value is not used.
    end
end
plot([1:n],w,’x’);
hold on;
legend(’All 6 faces show up in 6 rolls’,1)
ylabel(’Relative Frequencies’);
xlabel(’Number of Trials, n’);
print -dpms dice6.ps;
```

This program gives the results shown in the next figure. The limiting value is very close to the actual value of 0.0154.

2.5.17 Let \( P(\{c\}) = x \). Since all probabilities must add up to one,

\[ 1 - \frac{1}{4} + \frac{1}{4} + x + \frac{x}{2}. \]

Solving for \( x \) gives \( 3x + 1 = 2 \) or \( x = \frac{1}{3} \). That is, \( P(\{c\}) = \frac{1}{3} \) and \( P(\{d\}) = \frac{1}{3}. \)
2.5.18 In the first case take \( x = y = 1 \) in the binomial theorem. In the second case take \( y = 1, \ x = -1 \).

2.5.19 Using Exercise (2.5.19), add \( \sum_{k=0}^{n} \binom{n}{k} = 2^n \), and \( \sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0 \), and then divide by 2 to get the first result. Then subtract, instead, to deduce the second result.

2.5.20 Let \( \mathcal{H} = \mathcal{F} \cap \mathcal{G} \). Since, \( \mathcal{F} \in \mathcal{G} \) and \( \mathcal{G} \in \mathcal{G} \), therefore, \( \mathcal{F} \in \mathcal{G} \). If \( A \in \mathcal{H} \), then \( A \in \mathcal{F} \) and \( A \in \mathcal{G} \). Since both \( \mathcal{F} \), \( \mathcal{G} \) are sigma fields, \( \mathcal{F} \in \mathcal{F} \) and \( \mathcal{G} \in \mathcal{G} \). Hence, \( A \in \mathcal{H} \). Similarly, if \( A_i \in \mathcal{H}, \ i = 1, 2, \cdots \), then \( \bigcup_i A_i \in \mathcal{H} \).

2.5.21 The probability is \( \frac{2 - 2^2}{2} = \frac{2}{3} \).

2.5.22 The area of the rectangle is 2. The area under the curve is \( \int_0^2 (2t - t^2) \, dt = \frac{2}{3} \). So, the desired probability is \( \frac{2}{3} \).

3.5.1 No, because \( 0 = P(A)P(B) \) can happen only if \( P(A) = 0 \) or \( P(B) = 0 \), however, \( P(A \cap B) = 0 \) when \( A, B \) are disjoint.

3.5.2 \( P(A \cap B^c) = P(A) - P(A \cap B) = P(A)(1 - P(B)) \) which finishes the proof.

3.5.3 If \( S_1 = \{ H, T \} \) and \( S_2 = \{ H, T \} \) are the respective outcome sets of the two trials, the sample space is

\[ S = S_1 \times S_2 = \{ (HH), (HT), (TH), (TT) \}. \]

3.5.4 If \( S = \{ (HH), (HT), (TH), (TT) \} \), the probability measure is

\[ P((HH)) = \frac{6}{12}, \ P((HT)) = \frac{4}{12}, \ P((TH)) = \frac{4}{12}, \ P((TT)) = \frac{4}{12}. \]

3.5.5 \( P(\{ \text{at least one head} \}) = 1 - \frac{4}{12} = 0.8 \).

3.5.6 The sample space is the following set

\[ \{ (1HH), (1HT), (1TH), (2HH), (2HT), (2TH), (3HH), (3HT), (3TH), (4HH), (4HT), (4TH), (5HH), (5HT), (5TH), (6HH), (6HT), (6TH), (6TT) \}. \]

3.5.7 For any \( i \in \{ 1, 2, 3, 4, 5, 6 \} \),

\[ P(\{(iHH)\}) = \frac{6}{12}, \ P(\{(iHT)\}) = \frac{4}{12}, \ P(\{(iTH)\}) = \frac{6}{12}, \ P(\{(iTT)\}) = \frac{4}{12}. \]

3.5.8 \( P(\{ \text{even number and at least one head} \}) = \frac{1}{2}(1 - \frac{4}{12}) = 0.4 \).

3.5.9 The first head will occur after \( k \) tails if and only if all the previous \( k \) trials had tails, whose probability is \( (1 - p)^k \). The last trial should result in a head which occurs with probability \( p \).

3.5.10 The first head will occur on the \( k \)-th trial if and only if all the previous \( (k-1) \) trials had tails, whose probability is \( (1 - p)^{k-1} \). The last trial should result in a head which occurs with probability \( p \).

3.5.11 The 13th head will occur after \( k \) tails if and only if all the previous \( k \) trials and the earlier \( 13 - 1 \) heads took place in some order over the \( k + 13 - 1 \) trials. The total number of such possibilities is \( \binom{k + 13 - 1}{13} \), and each has probability \( p^{13}(1 - p)^k \). The last trial should result in a head which occurs with probability \( p \).

3.5.12 The 13th head will occur at the \( k \)-th trial if and only if the earlier \( 13 - 1 \) heads took place in some order over the \( k - 1 \) trials. The total number of such possibilities is \( \binom{13 - 1}{13 - 1} \), and each has probability \( p^{12}(1 - p)^{k-1} \). The last trial should result in a head which occurs with probability \( p \).

3.5.13 This is a binomial experiment, and hence, the required probability is \( \binom{n}{k} p^k(1 - p)^{n-k}, \ k = 0, 1, 2, \cdots, n \).

3.5.14 This is a geometric experiment, and hence, the required probability is \( p(1 - p)^{k-1}, \ k = 0, 1, 2, \cdots \).

3.5.15 This is a shifted geometric experiment, and hence, the required probability is \( p(1 - p)^{k-1}, \ k = 1, 2, 3, \cdots \).

3.5.16 This is a negative binomial experiment, and hence, the required probability is \( \binom{k + 10 - 1}{k} p^k(1 - p)^{10-k}, \ k = 0, 1, 2, \cdots \).

3.5.17 This is a shifted negative binomial experiment, and hence, the required probability is \( \binom{k + 10 - 1}{k} p^k(1 - p)^{10-k}, \ k = 10, 11, 12, \cdots \).
3.5.18 There are \( \binom{n}{k,j} = \frac{n!}{k!(n-k-j)!} \) ways we may have \( k \) number of \( R \)'s and \( j \) number of \( G \)'s (and the rest of them \( B \)'s) over a segment of \( n \) trials. Each such happening has probability \( r^k g^j (1-r-g)^{n-k-j} \). Therefore, the required probability is \( \binom{n}{k,j} r^k g^j (1-r-g)^{n-k-j} \). 

3.5.19 There are \( \binom{n}{k,l} = \frac{n!}{k!(n-k-l)!} \) ways we may have \( k \) number of \( R \)'s, \( j \) number of \( G \)'s and \( l \) number of \( B \)'s (and the rest of them \( U \)'s) over a segment of \( n \) trials. Each such happening has probability \( r^k g^j b^l (1-r-g-b)^{n-k-j-l} \). Therefore, the required probability is \( \binom{n}{k,j,l} r^k g^j b^l (1-r-g-b)^{n-k-j-l} \). 

3.6.20 \( P(A) = \frac{1}{2}, P(B) = \frac{1}{2}, P(C) = \frac{1}{3} \). Also, 
\[
P(A \cap B) = \frac{1}{4}, \quad P(A \cap C) = \frac{1}{4}, \quad P(B \cap C) = \frac{1}{4}.
\]
This makes the three events pairwise independent, but \( P(A \cap B \cap C) = 0 \) which is not the product of \( P(A), P(B), P(C) \).

3.5.21 \( \binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = 2^n - n - 1 \).

4.3.1 The following is a Matlab program that performed the random experiment.

```matlab
clear;
n=20000;  \% Number of rolls of 6 dice.
x=rand([n,4]);  \% 4 columns of uniform deviates.
x=x+0.5;  \% make rand values fall in [0.5,1.5].
z=floor(x);  \% z is now 1 or zero in each column.
zsum=sum(z==2);  \% number of heads in four tosses.
H = histc(zsum,[0,1,2,3,4]);  \% compute histogram.
H = H/n;  \% convert into relative frequencies.
```

This gives the following relative frequency table,

<table>
<thead>
<tr>
<th>Number of Heads</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Frequencies</td>
<td>0.0019</td>
<td>0.2467</td>
<td>0.3743</td>
<td>0.2361</td>
<td>0.0610</td>
</tr>
</tbody>
</table>

4.3.2 The exact density is
\[
P(X = 0) = \frac{1}{16}, \quad P(X = 1) = \frac{4}{16}, \quad P(X = 2) = \frac{6}{16}, \quad P(X = 3) = \frac{4}{16}, \quad P(X = 4) = \frac{1}{16}.
\]
The numbers match with the relative frequencies quite well.

4.3.3 The possible values of \( X \) are 0, 1, 2, 3. The respective probabilities are as follows,
\[
P(X = 0) = \frac{1}{64}, \quad P(X = 1) = \frac{18}{64}, \quad P(X = 2) = \frac{18}{64}, \quad P(X = 3) = \frac{27}{64}.
\]
4.3.4 This cannot be a density since all the probabilities do not add up to one.

4.3.5 This cannot be a density since probabilities cannot be negative.

4.3.6 This is a density since probabilities are nonnegative and add up to one.

4.3.7 The density of \( X \) is
\[
P(X = 0) = \frac{\binom{3}{2}}{\binom{4}{1}}, \quad P(X = 1) = \frac{\binom{3}{2}}{\binom{4}{1}}, \quad P(X = 2) = \frac{\binom{3}{2}}{\binom{4}{1}}, \quad P(X = 3) = \frac{\binom{3}{2}}{\binom{4}{1}}.
\]

4.3.8 Note that \( X \sim B(10, \frac{1}{2}) \). Also,
\[
P(X \leq 3) = \frac{1}{2^{10}} + \frac{10}{2^{10}} + \frac{\binom{10}{3}}{2^{10}}.
\]

4.3.9 This is just the geometric series.

4.3.10 The event is \( A = \{0, 4, 8, 12, 16, \ldots \} \). So, its probability is
\[
P(X \in A) = \sum_{k=0}^{\infty} p(1-p)^{4k} = p \sum_{k=0}^{\infty} (1-p)^{4k} = \frac{p}{1-(1-p)^4}.
\]

4.3.11 The event is \( A = \{0, 3, 6, 9, 12, \ldots \} \). So, its probability is
\[
P(X \in A) = \sum_{k=0}^{\infty} p(1-p)^{3k} = p \sum_{k=0}^{\infty} (1-p)^{3k} = \frac{p}{1-(1-p)^3}.
\]

4.3.12 This is a binomial experiment and hence, \( P(X = k) = \binom{4}{k} p^k (1-p)^{4-k} \), for \( k = 0, 1, 2, \ldots, n \). Here \( p = 0.2 \).

4.3.13 This is a geometric experiment and hence, \( P(Y = k) = p(1-p)^{k-1} \), for \( k = 1, 2, 3, \ldots \). Here \( p = 0.2 \).

4.3.14 This is a shifted geometric experiment and hence, \( P(Z = k) = p(1-p)^{k-1} \), for \( k = 1, 2, 3, \ldots \). Here \( p = 0.2 \).

4.3.15 This is a negative binomial experiment and hence, \( P(W = k) = \binom{k-1}{4-1} p^4 (1-p)^{k-4} \), for \( k = 0, 1, 2, \ldots \). Here \( p = 0.2 \).

4.3.16 This is a shifted negative binomial experiment and hence, \( P(T = k) = \binom{k-1}{3} p^3 (1-p)^{k-3} \), for \( k = 30, 31, 32, \ldots \). Here \( p = 0.2 \).

4.3.17 Roll a loaded die so that faces 1, 2, 3 have one color with combined probability \( \frac{1}{4} \). Put the other probabilities on the remaining faces. Then define the appropriate r.v.

4.3.18 Write down all the outcomes as a 6 x 6 table. Compute the value of \( X \) for each outcome, then write down the density of \( X \).
4.3.19 As shown in Exercise (4.3.13), $Y \sim \text{Geometric}(p)$ where $p = 0.2$. By Remark (4.2.2), $P(Y \geq 20) = (1 - p)^{20} = 0.8^{20}$.

4.3.20 $F(t) = 0$ for $t < -9$. For $t \in [-9,3)$, $F(t) = \frac{t}{3}$. For $t \in [3,6)$, $F(t) = \frac{t}{18}$, and for $t \geq 6$, $F(t) = 1$.

4.3.21 (i) $F(t)$ is nondecreasing in $t$, (ii) $F(t)$ is right continuous, (iii) $F(-\infty) = 0$, $F(\infty) = 1$, and (iv) $F'(x) = F(x) - F(a)$.

4.3.22 The random variable takes values 0, 1, 2, 3 with respective probabilities 0.7, 0.10, 0.01, 0.19.

5.3.1 Since $\int_{-\infty}^{\infty} e^{-|t|} dt = 2$, it is not a density.

5.3.2 Since it is nonnegative and $\int_0^1 (2x) dx = 1$, it is a density.

5.3.3 (i) It is clear that $F(t)$ is a nondecreasing function of $t$. (ii) It is continuous,

5.3.4 The density is $f(t) = 0$ for $t < 0$, and $f(t) = 2te^{-t^2}$ for $t > 0$. We may take $F(0) = 0$.

5.3.5 $F(t)$ is nondecreasing, continuous, with $F(-\infty) = 0$ and $F(\infty) = 1$, therefore, it is a cdf. Its density is $\frac{d}{dt} F(t) = F'(t)$ for $t \in (0,2)$ and zero otherwise.

5.3.6 $F(t)$ is an increasing function of $t$, it is continuous, in fact, differentiable, $F(-\infty) = 0$, $F(\infty) = 1$. So, it is a cdf. Finally, its density is $\frac{d}{dt} F(t) = \frac{1}{e^t}$ for $t < 0$ and $\frac{d}{dt} F(t) = \frac{1}{e^{-t^2}}$ for $t > 0$. That is, $\frac{d}{dt} F(t) = \frac{1}{e^{-t^2}}$ for $t \in \mathbb{R}$, $t \neq 0$. For $t = 0$, we may define the density by any nonnegative value.

5.3.7 It is clear that $F(t)$ is nondecreasing in $t$, $F(-\infty) = 0$, $F(\infty) = 1$, and $F(t)$ is right continuous. So, it is a cdf. However, since it has a point of jump at $t = 0$, it is not a cdf of a continuous random variable. It is a cdf of a discrete random variable either, since the total jump size is not 1. (It is a cdf of a random variable which is partly discrete and partly continuous.)

5.3.8 Clearly, $F(t) = 0$ for $t \leq 0$. For $t \in (0,1)$,

4.3.9 We want $\int_0^1 (cx + x^3) dx = 1$, integrating gives $\frac{c}{2} + \frac{1}{2} = 1$. This gives $c = \frac{1}{3}$.

5.3.10 Clearly, $F(t) = 0$ for $t \leq 0$. For $t \in (0,1)$, we have

4.3.11 We want $\int_0^1 \frac{1}{2} e^{-x^2} dx = 1$, This gives that $c = b$.

5.3.12 Clearly, $F(t) = 0$ for $t \leq b$. For $t > b$, we have

5.3.13 Note that $F(1) = 0, F(3) = \frac{2}{3} = \frac{\Gamma(4)}{\Gamma(1)} = F(3), F(5) = 1 = F(5)$. So, $F(t)$ is a continuous, nondecreasing function of $t$ and $F(-\infty) = 0$ and $F(\infty) = 1$. Hence, $F$ is a cdf of some continuous random variable, whose density must be $f(t) = F'(t) = \frac{1}{2}$ for $1 \leq t < 3$, $f(t) = \frac{1}{2}\sqrt[3]{t}$ for $3 \leq t < 5$, and $f(t) = 0$ for all other values of $t$. Finally,

5.3.14 0.0062 and 0.0062 and 0.9876.

5.3.15 For $t \leq 0$, $F(t) = 0$. For $t > 0$, we have

5.3.16 The density of $X$ is $f(x) = 6x(1 - x)$ when $x \in (0,1)$ and zero otherwise.

5.3.17 The density of $X$ is $\frac{1}{2}e^{-x^2/2}$. Therefore,

5.3.18 The standard normal density has a higher peak and lower tails than the Cauchy(0,1) density.
\[ P(\{Y - 10 < 3\}) = P(7 < e^X < 13) = \Phi(\ln 13) - \Phi(\ln 7), \]
where \( \Phi(t) \) is the distribution function for the standard normal density.

\[ \{(u,v) \in \Delta : |u - v| \leq t\} = \{(u,v) \in \Delta : u - t \leq v \leq u + t\}, \]
and the region is shown as the shaded area in Figure (11.6). The point at which the linear \( v = u - t \) crosses the upper side of the triangle is obtained by equating the equations of the two lines \( v = u - t \) and \( v = 2 - u \). That is, \( u = 1 + \frac{t}{2} \), and \( v = 1 - \frac{t}{2} \). Hence, the area of the unshaded (small) triangle is
\[ \frac{(2 - t)(1 - \frac{t}{2})}{2} = \left(1 - \frac{t}{2}\right)^2. \]
This gives that the shaded area equals
\[ P(X \leq t) = 1 - \left(1 - \frac{t}{2}\right)^2 = 1 - \left(\frac{2 - t}{2}\right)^2. \]

\[ \text{5.3.21} \] It is easy to verify that \( F(t) = 0 \) for \( t \leq 0 \), \( F(t) = 1 \) for \( t \geq 2 \) and \( F(t) = 1 - (1 - \frac{t}{2})^2 \) for \( t \in (0, 2) \).

\[ \text{5.3.4} \] By the definition,
\[ E(X) = \sum_{x} xP(X = x) = cP(X = c) = c. \]

\[ \text{5.5.3} \] \( E(X) = -.1 + 3 + 8 = 11. \) \( E(X^2) = .1 + 3 + 1.6 = 2 \). \( E(X^3) = 2 - 1^2 = 1 \).

\[ \text{5.5.4} \] \( E(X^4) = .1 + 3 + 6.4 = 6.8 \). \( E(X^5) = 6.8 - 2 + 1 = 5.8 \).

\[ \text{5.5.5} \] By the definition, and integration by parts,
\[ E(X) = \int_{0}^{\infty} xe^{-x} dx = 1. \]
Similarly,
\[ E(X^2) = \int_{0}^{\infty} x^2 e^{-x} dx = 2. \]
So, \( \text{Var}(X) = 2 - 1^2 = 1 \).

\[ \text{5.5.6} \] Just note that
\[ E(X) = \sum_{k=0}^{n} \frac{k}{n} \sum_{k=2}^{n} p^k (1 - p)^{n-k} \]
\[ = np \sum_{k=1}^{n} \frac{(n - 1)!}{(k - 1)!(n - 1) - (k - 1)!} p^k (1 - p)^{n - 1} k \]
\[ = np(p + 1 - p)^{n - 1} n = np. \]

Next, by the LST, we have
\[ E(X^2 - 1) = \sum_{k=0}^{n} (k - 1) \frac{n!}{(n - k)!} p^k (1 - p)^{n - k} \]
\[ = n(n - 1)p^2 \sum_{k=2}^{n} \frac{(n - 2)!}{(k - 2)!(n - 2) - (k - 2)!} p^k (1 - p)^{n - 2} k \]
\[ = n(n - 1)p^2 (p + 1 - p)^{n - 2} n = n(n - 1)p^2. \]
Therefore, by the linearity property, \( n(n - 1)p^2 = E(X^2 - 1) = E(X^2) - E(X) = E(X^2) - np \). This gives that \( E(X^2) = n(n - 1)p^2 + np \). \( \text{Var}(X) = n(n - 1)p^2 + np - \frac{n^2p^2}{n} = np(1 - p) \).

\[ \text{5.5.7} \] By linearity, \( E(Y) = E(a + bX) = a + bE(X) \).
\[ \text{Var}(Y) = E(a + bX)^2 - (a + bE(X))^2 \]
\[ = E(a^2 + 2abX + b^2X^2) - a^2 - 2abE(X) - b^2E(X)^2 \]
\[ = b^2E(X^2) - b^2E(X)^2 = b^2\text{Var}(X). \]

\[ \text{5.5.8} \] Just note that
\[ E \left( \frac{1}{X + 1} \right) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k + 1} \lambda^k \]
\[ = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k + 1)!} \]
\[ = e^{-\lambda} \left( e^{\lambda} - 1 \right) = \frac{1 - e^{-\lambda}}{\lambda}. \]
6.5.9  Since \( X^2 + 3X + 2 = (X + 2)(X + 1) \),

\[
E \left( \frac{1}{X^2 + 3X + 2} \right) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{(k+2)(k+1)} \frac{\lambda^k}{k!}
\]

\[
= \frac{e^{-\lambda}}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+2}}{(k+2)!}
\]

\[
= \frac{e^{-\lambda}}{\lambda} (e^\lambda - 1 - \lambda).
\]

(Thanks, Tim Ave'Lallemant.)

6.5.10  By the LST, and using the given hint, we have

\[
E(1/(X + 1)) = \sum_{k=0}^{\infty} \frac{1}{k+1} p(1-p)^{k+1}
\]

\[
= \frac{p}{1-p} \sum_{j=0}^{\infty} \frac{1}{j+1} (1-p)^{j+1}
\]

\[
= -p \ln (1 - 1 + p) - p \ln p = -p \ln \frac{1}{1-p}.
\]

6.5.11  By the LST, we have

\[
E(1/(X + 1)) = \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{n!}{(n-k)!} p^k (1-p)^{n-k}
\]

\[
= \frac{1}{p(n+1)} \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{(n+1)!}{(n-k)!} p^{k+1} (1-p)^{n-k}
\]

\[
= \frac{1}{p(n+1)} \sum_{j=0}^{n+1} \frac{1}{j+1} \frac{(n+1)!}{j!} p^j (1-p)^{n+1-j}
\]

\[
= \frac{1 - (1-p)^{n+1}}{p(n+1)}.
\]

6.5.12  To find \( c \), we must have

\[
1 = \int_{b}^{\infty} f(x) \, dx = c \int_{b}^{\infty} x^{a-1} \, dx = \frac{c}{a} x^a |_{b}^{\infty} = \frac{c}{ab^a}.
\]

Therefore, \( c = ab^a \). Now the expectation of \( X \) is

\[
E(X) = ab^a \int_{b}^{\infty} x \, (x-1)^{a-1} \, dx = \frac{ab^a}{(a-1)b^{a-1}} \quad \text{when} \quad a > 1.
\]

6.5.13  By the definition,

\[
E(X) = \frac{\lambda}{2} \int_{0}^{\infty} x e^{-\lambda x} \, dx.
\]

Since the integrand is an odd function, the expectation is zero. Now the second moment is

\[
E(X^2) = \frac{\lambda}{2} \int_{0}^{\infty} x^2 e^{-\lambda x} \, dx
\]

\[
= \frac{2\lambda}{2} \int_{0}^{\infty} x e^{-\lambda x} \, dx
\]

\[
= \frac{\lambda}{2} \frac{\Gamma(3)}{\lambda^3} \int_{0}^{\infty} x^2 e^{-\lambda x} \, dx
\]

\[
= \frac{\lambda}{2} \frac{\Gamma(3)}{\lambda^3} = \frac{2}{\lambda^2}.
\]

6.5.14  Note that

\[
E(X) = \frac{1}{3\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-10)^2/6} \, dx
\]

\[
= \frac{1}{3\sqrt{2\pi}} \int_{-\infty}^{\infty} (10+u) e^{-u^2/2} \, du,
\]

letting \( u = (x - 10)/3 \),

\[
= \frac{10}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} \, du + \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \, e^{-u^2/2} \, du.
\]

The first integral gives 10, since the integral of any density (in this case the standard normal density) is one. The second integral is zero since the integrand is an odd function. So, \( E(X) = 10 \). Similarly,

\[
E(X^2) = \frac{1}{3\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-(x-10)^2/6} \, dx
\]

\[
= \frac{1}{3\sqrt{2\pi}} \int_{-\infty}^{\infty} (10+u)^2 e^{-u^2/2} \, du,
\]

letting \( u = (x - 10)/3 \),

\[
= \frac{100}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} \, du + \frac{60}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \, e^{-u^2/2} \, du + \frac{9}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 \, e^{-u^2/2} \, du.
\]
The first integral gives 100. The second integral gives a zero. The third integral is obtained by using the fact that the integrand is an even function.

\[
\frac{1}{\sqrt{2\pi}} \int_0^\infty u^2 e^{-u^2/2} \, du = \frac{2}{\sqrt{2\pi}} \int_0^\infty u^2 e^{-u^2/2} \, du
\]

let \( v = u^2/2 \),

\[
= \frac{4}{\sqrt{2\pi}} \int_0^\infty v e^{-v/2} \, dv
\]

= \[
\frac{2}{\sqrt{\pi}} \int_0^\infty v^{3/2} e^{-v} \, dv
\]

= \[
\frac{2\Gamma(3/2)}{\sqrt{\pi}} \Gamma(3/2) \int_0^\infty v^{1/2} e^{-v} \, dv
\]

since the area under any density (in this case the gamma density \( G(1, \frac{3}{2}) \)) is one.

Now \( \Gamma(\frac{3}{2}) = \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2} \). Hence,

\[
E(X^2) = 100 + 9, \quad \text{so} \quad Var(X) = 100 + 9 - (10)^2 = 9.
\]

**6.5.16** Note that

\[
E(Y) = \int_0^\infty e^x \frac{1}{3\sqrt{2\pi}} e^{-\frac{(x-10)^2}{2\times9}} \, dx
\]

= \[
\frac{1}{\sqrt{2\pi}} \int_0^\infty e^{10+3u} e^{-u^2/2} \, du,
\]

letting \( u = (x-10)/3 \),

= \[
\frac{e^{10}}{\sqrt{2\pi}} \int_0^\infty e^{u^2/2} \, du,
\]

complete the square,

= \[
\frac{e^{10+9/2}}{\sqrt{2\pi}} \int_0^\infty e^{-u^2/2} \, du,
\]

= \[
\frac{e^{10+9/2}}{\sqrt{2\pi}} \int_0^\infty e^{-u^2/2} \, du,
\]

= \[
\frac{e^{10+9/2}}{\sqrt{2\pi}} \int_0^\infty e^{-u^2/2} \, du,
\]

The last result comes by the fact that the area under any density (in this case normal density with \( \mu = 3 \) and \( \sigma = 1 \)) is always one. The same trick works for the second moment.

\[
E(Y^2) = \int_0^\infty e^{2x} \frac{1}{3\sqrt{2\pi}} e^{-\frac{(x-10)^2}{2\times9}} \, dx
\]

= \[
\frac{1}{\sqrt{2\pi}} \int_0^\infty e^{10+6u} e^{-u^2/2} \, du,
\]

letting \( u = (x-10)/3 \),

= \[
\frac{e^{10}}{\sqrt{2\pi}} \int_0^\infty e^{6u} e^{-u^2/2} \, du,
\]

complete the square,

\[
= \frac{e^{38}}{\sqrt{2\pi}} \int_0^\infty e^{-u^2/2} \, du,
\]

\[
= \frac{e^{38}}{\sqrt{2\pi}} \int_0^\infty e^{-u^2/2} \, du
\]

Therefore, \( Var(X) = e^{38} - e^{39} \).

**6.5.16** By the linearity property,

\[
h(a) = E(X^2) - 2aE(X) + a^2, \quad a \in \mathbb{R}
\]

is a parabola that opens upwards. Differentiating, we see that the only critical point is the solution of \(-2E(X) + 2a = 0\). That is, \( a = E(X) \). So, the minimum value of \( h(a) \) is \( E(X) - E(X))^2 = Var(X) \).

**6.5.17** In Exercise (6.5.2) we found that \( E(X) = 1 \) and \( Var(X) = 1 \). The exact probability is

\[
P(|X - 1| > 1) = P(X < 0) + P(X > 2) = P(X = -1) = 0.1.
\]

An upper bound is

\[
P(|X - E(X)| > 1) \leq \frac{Var(X)}{1^2} = 1,
\]

which is pretty useless.

**6.5.18** Since \( Var(X) = 10 \), by Chebyshev’s inequality,

\[
P(|X - E(X)| > 4) \leq \frac{Var(X)}{4^2} = \frac{10}{16}.
\]

**6.5.19** Note that

\[
\{|X - 0.1| > 0.001\} = \{X > 0.101\} \cup \{X < 0.099\}.
\]

When \( X \sim Exp(10) \), we have \( P(X > 0.101) = e^{-10(0.101)} \) and \( P(X < 0.099) = 1 - e^{-10(0.099)} \). Therefore, the exact probability is

\[
P(|X - 0.1| > 0.001) = e^{-10(0.101)} + 1 - e^{-10(0.099)} = 0.9926.
\]

To use Chebyshev’s inequality, we note that \( E(X) = 0.1 \) and \( Var(X) = (0.1)^2 = 0.01 \). Hence,

\[
P(|X - 0.1| > 0.001) \leq \frac{Var(X)}{(0.001)^2} = \frac{0.01}{0.000001} = 1.
\]

The right side is much bigger than one. In this case Chebyshev’s inequality is useless.
6.5.20 Note that \( h(t) = -\log(t) \) is a convex function. So, by the Jensen inequality,
\[
E h(X) = -E(\log(X)) \geq h(E(X)) = -\log(E(X)).
\]
Multiplying by \(-1\) gives the result.

6.5.21 Let \( h(t) = t^p \) for \( t \geq 0 \). It is a convex function for \( p \geq 1 \). By Jensen’s inequality, \( E(X^p) \leq h(E(X)) = (E(X))^p \).

7.4.1 The marginal density of \( X \) is as follows,

<table>
<thead>
<tr>
<th>Values of ( X )</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilities</td>
<td>( \frac{7}{16} )</td>
<td>( \frac{9}{16} )</td>
</tr>
</tbody>
</table>

Similar, the marginal density of \( Y \) is

<table>
<thead>
<tr>
<th>Values of ( Y )</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilities</td>
<td>( \frac{4}{16} )</td>
<td>( \frac{4}{16} )</td>
<td>( \frac{8}{16} )</td>
</tr>
</tbody>
</table>

7.4.2 The bivariate pdf of \((X,Y)\), \( F(s,t) \), is obtained by considering all the possible values of \( s \) and \( t \). Certainly, \( F(s,t) = 0 \) if \( s < 0 \) or \( t < 0 \). On the other hand \( F(s,t) = 1 \) if \( s \geq 1 \) and \( t \geq 3 \). So, we consider the other cases. The various cases are as follows, \( 0 \leq s < 1, 1 \leq t < 2, 0 \leq s < 1, 1 \leq t < 3, 0 \leq s < 1, t \geq 3, 1 \leq s \leq 1, t \geq 2, 1 \leq s < 2, t \leq 3 \). It is easy to verify that

\[
F(s,t) = \begin{cases} 
\frac{1}{9} & 0 \leq s < 1, 1 \leq t < 2, \\
\frac{2}{9} & 0 \leq s < 1, 2 \leq t < 3, \\
\frac{1}{3} & 0 \leq s < 1, t \geq 3, \\
\frac{1}{3} & 1 \leq s \leq 1, t \geq 2, \\
\frac{1}{3} & 1 \leq s < 2, t \leq 3. 
\end{cases}
\]

7.4.3 From the solution of Exercise (7.4.1), we see that \( P(Y = 1) = \frac{14}{16} \) and \( P(X = 0) = \frac{7}{16} \). Since, their product is not equal to \( P(X = 0, Y = 1) = \frac{1}{9} \), the two random variables are not independent.

7.4.4 To find the marginal densities, note that

\[
f_X(x) = \int_{-\infty}^{\infty} f(x,y) \, dy = \int_{-x}^{4-x^2} 2(1-x^2) \, dy.
\]

Therefore,

\[
f_X(x) = \frac{1}{4\pi} \sqrt{4 - x^2}, \quad x \in (-2, 2).
\]

Similarly, \( f_Y(y) = \frac{1}{4\pi} \sqrt{4 - y^2}, \quad y \in (-2, 2) \).

7.4.5 Since the product of the marginal densities obtained in Exercise (7.4.4) do not give the joint density, the two random variables are not independent.

7.4.6 In this case, the joint is \( f(x,y) = 2 \) for \( (x,y) \in [0,1] \times [0,1] \), so that \( x + y \leq 1 \), and zero otherwise. So, we see that \( F(s,t) = 0 \) when either \( s < 0 \) or \( t < 0 \). Also, \( F(s,t) = 1 \) if \( s \geq 1 \) or \( t \geq 1 \). The other cases are

\[
F(s,t) = \begin{cases} 
2st & 0 \leq s < 1, 1 \leq t < 1, s + t \leq 1, \\
1 - (1 - s)^2 - (1 - t)^2 & 0 \leq s < 1, 0 \leq t < 1, s + t > 1, \\
1 - 2(1-s)^2 & s \geq 1, 0 \leq t < 1, \\
1 - 2(1-t)^2 & 0 \leq s < 1, t \geq 1. 
\end{cases}
\]

(Thanks to Dr. Vezvai for correcting an error, Nov. 6, 2003.)

7.4.7 It is easy to see that \( X \sim Beta(1,2) \) and \( Y \sim Beta(1,2) \). The joint density of \( X, Y \) is \( f(x,y) = 2 \) for \( x, y \in (0,1) \) with \( x + y \in (0,1) \). Since the product of the marginal densities does not equal the joint density, the two random variables are not independent. (Thanks to Dr. Vezvai for correcting an error, Nov. 6, 2003.)

7.4.8 We just add over \( y \),

\[
f_X(x) = \sum_{y=0}^{3-x} f(x,y) = \frac{(2)}{3} \sum_{y=0}^{3-x} \binom{3-x}{y} (4 - 3 - x - y) = \frac{(2)}{3} \left( \frac{7}{3-x} \right).
\]

This shows that \( X \sim Hypergeometric(2,7,3) \).

7.4.9 The same steps of Exercise (7.4.8) show that \( Y \sim Hypergeometric(3,6,3) \). The two random variables are certainly not independent, since \( P(X = 2) > 0 \) and \( P(Y = 3) > 0 \), but \( P(X = 2, Y = 3) = 0 \).

7.4.10 If we integrate over \( y \), we get

\[
f_X(x) = \int_0^\infty f(x,y) \, dy = 3e^{-2x} \int_0^\infty 4e^{-y} \, dy = 3e^{-2x}.
\]

So, \( X \sim Exp(3) \). Similarly, we get \( Y \sim Exp(4) \). Since \( f_X(x)f_Y(y) = f(x,y) \) for almost all \( x, y \), the random variables are independent.

7.4.11 The marginal densities are, for \( x, y > 0 \),

\[
f_X(x) = e^{-x} \int_0^x 1 \, dy = xe^{-x},
\]
\[
f_Y(y) = \int_y^\infty e^{-x} \, dx = e^{-y}.
\]

So, \( X \sim Gamma(1,2) \) and \( Y \sim Exp(1) \). Since \( f(x,y) \neq f_X(x)f_Y(y) \), the two random variables are not independent.
7.4.12 It is easy to see that \( F(s, t) = 0 \) if \( s < 0 \) or \( t < 0 \). So, let \( s, t \geq 0 \). Then,
\[
F(s, t) = \int_0^s \int_0^t 3e^{-3x} 4e^{-4y} \, dy \, dx
\]
\[
= (1 - e^{-4t}) \int_0^s 3e^{-3x} \, dx = (1 - e^{-4t})(1 - e^{-3s}),
\]
7.4.13 Since the density is zero outside the region \( 0 < y < x < \infty \), the joint cdf is \( F(s, t) = 0 \) for \( s < 0 \) or \( t < 0 \). So, take \( s \geq 0 \), and \( t \geq 0 \). There are two cases, When \( 0 \leq s \leq t \), then
\[
F(s, t) = \int_0^s \int_0^t f(x, y) \, dy \, dx
\]
\[
= \int_0^t \int_0^s e^{-x} e^{-y} \, dx \, dy = -se^{-s} + 1 - e^{-s},
\]
When \( 0 \leq t < s \),
\[
F(s, t) = F(t, t) + \int_t^s e^{-x} \, dx = F(t, t) + (e^{-x} - e^{-t}) t.
\]
7.4.14 Note that for \( 0 < s < t \), we have \( \frac{\partial F}{\partial x} = 0 \). For \( 0 < t < s \), we have
\[
\frac{\partial F}{\partial x} = \frac{\partial (te^{-x})}{\partial t} = e^{-x}.
\]
Hence, the joint density is \( f(x, y) = e^{-x} \) for \( 0 < y < x \) and zero otherwise.
7.4.15 The marginals are obtained by letting the other variable go to infinity. So,
\[
F_X(s) = \lim_{t \to \infty} F(s, t) = 1 - e^{-s} - se^{-s}.
\]
Similarly,
\[
F_Y(t) = \lim_{s \to \infty} F(s, t) = 1 - e^{-t}.
\]
7.4.16 One simple example is given below.

<table>
<thead>
<tr>
<th>( X \setminus Y )</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0/3</td>
<td>1/3</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1/3</td>
<td>1/3</td>
<td>0</td>
</tr>
</tbody>
</table>
7.4.17 Take any two continuous random variables, \( X \) and \( Y \) with respective densities \( f_X(x) \) and \( f_Y(y) \). Define their joint density, \( f(x, y) \), by \( f(x, y) = f_X(x) f_Y(y) \). Thus \( (X, Y) \) have a joint density that makes \( X, Y \) as independent random variables. (Their values lie in the product space.)
7.4.18 Yes, the reason being, when \( \rho = 0 \), the bivariate normal density becomes the product of the marginal densities of \( X \) and \( Y \).

7.4.19 By adding over all values of \( y \), we get the marginal density of \( X \):
\[
P(X = k) = \sum_{j=0}^n P(X = k, Y = j) = \sum_{j=0}^n \left( \frac{\binom{n}{j} \binom{R}{k} \binom{S}{n-k-j}}{\binom{n}{S}} \right)
\]
\[
= \left( \frac{\binom{n}{k}}{\binom{n}{S}} \right) \binom{R}{k} \sum_{j=0}^n \binom{S}{n-k-j}
\]
\[
= \left( \frac{\binom{n}{k}}{\binom{n}{S}} \right) \binom{R+S}{n-k}.
\]
That is, \( X \sim \text{Hypergeometric}(G, R + S, n) \).

8.4.1 By the GLST, the desired expectation is
\[
E(X^2Y^3) = \left( 1^2 \times \frac{3}{5} \right) + \left( 2^3 \times \frac{16}{45} \right) + \left( 3^3 \times \frac{13}{45} \right)
\]
\[
= \frac{9 + 128 + 351}{45} = \frac{488}{45} = 10.8444.
\]
8.4.2 Since we know that \( Z \sim \text{Poisson}(3\lambda) \), \( E(Z) = 3\lambda = \text{Var}(Z) \). Another approach is direct. Just note that
\[
E(Z) = E(X_1) + E(X_2) + E(X_3) = 3\lambda.
\]
By independence,
\[
\text{Var}(Z) = \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) = 3\lambda.
\]
8.4.3 First note that
\[
E(XY) = \left( 1 \times \frac{1}{5} \right) + \left( 2 \times \frac{16}{45} \right) + \left( 3 \times \frac{13}{45} \right)
\]
\[
= \frac{9 + 32 + 39}{45} = \frac{80}{45} = \frac{16}{9}.
\]
The marginal density of \( X \) is \( B(1, \frac{2}{5}) \), So, \( E(X) = \frac{2}{5} \) and \( \text{Var}(X) = \frac{2}{25} \). The marginal density of \( Y \) is as follows,

<table>
<thead>
<tr>
<th>Values of ( Y )</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilities</td>
<td>3/5</td>
<td>2/5</td>
<td>1/5</td>
</tr>
</tbody>
</table>

This gives that \( E(Y) = \frac{14 + 24 + 2}{35} = \frac{50}{35} = \frac{8}{7} \). So, \( \text{Cov}(X, Y) = \frac{16}{9} - \frac{38}{45} = \frac{4}{15} \). Also, note that \( E(Y^2) = \frac{14^2 + 24^2 + 2}{35} = \frac{108}{35} \). So, \( \text{Var}(Y) = \frac{108}{35} - 4 = \frac{4}{35} \). Hence, the correlation is
\[
\text{Corr}(X,Y) = \frac{\frac{4}{15}}{\sqrt{\frac{206}{729} \times \frac{28}{45}}} = 0.31018.
\]
8.4.4 The GLST gives

\[
E(X^3) = \int_0^1 6(1-y) \int_0^y x^3 \, dx \, dy = \int_0^1 (1-y) y^3 \, dy = \frac{1}{7} - \frac{1}{8} = \frac{1}{56}.
\]

Similarly,

\[
E(X^2Y) = \int_0^1 6y(1-y) \int_0^y x^2 \, dx \, dy = \int_0^1 2(1-y) y^4 \, dy = \frac{2}{5} - \frac{2}{6} = \frac{1}{15}.
\]

8.4.5 Here are the necessary calculations.

\[
E(X) = \int_0^1 6(1-y) \int_0^y x \, dx \, dy = 1 - \frac{3}{4} = \frac{1}{4},
\]

\[
E(Y) = \int_0^1 6y(1-y) \int_0^1 dx \, dy = \frac{6}{3} = \frac{1}{2},
\]

\[
E(XY) = \int_0^1 6y(1-y) \int_0^y x \, dx \, dy = \frac{3}{4} - \frac{3}{5} = \frac{3}{20}.
\]

Therefore, \( \text{Cov}(X, Y) = \frac{3}{20} - \frac{1}{5} = -\frac{1}{10} \).

8.4.6 (i) No, (ii) No, (iii) Yes.

8.4.7 Since the joint density is the product of marginal densities of \( X \) and \( Y \), it must be that \( X \) and \( Y \) are independent. Furthermore, \( X \) is \( \text{Exp}(3) \) and \( Y \) is \( \text{Exp}(4) \), both have finite variances. Hence, they must be uncorrelated. That is, \( \text{Cov}(X, Y) = 0 \).

8.4.8 Note that

\[
E(X^2Y^3) = \int_0^1 \int_0^\sqrt{\frac{1-x}{x}} x^3 y^3 \, dy \, dx = \int_0^1 x^2 \, dx \int_0^{\sqrt{\frac{1-x}{x}}} y^3 \, dy = 0,
\]

since the inner integral is zero, (having an odd integrand).

8.4.9 We know \( E(X) = E(Y) = 0 \), due to symmetry. Also,

\[
E(XY) = \int_0^1 \int_0^{\sqrt{\frac{1-x}{x}}} xy \, dy \, dx = \int_0^1 x \, dx \int_0^{\sqrt{\frac{1-x}{x}}} y \, dy = 0.
\]

Therefore, \( \text{Cov}(X, Y) = 0 = \text{Corr}(X, Y) \). However, \( X \) and \( Y \) are certainly not independent, since the region over which the joint density is positive is a circle (not a rectangle).

8.4.10 From Exercise (8.4.3) we know that the exact value is \( \text{Cov}(X, Y) = \frac{4}{21} \). Also, \( \text{Var}(X) = \frac{26}{(24)^2} \), \( \text{Var}(Y) = \frac{25}{4} \). So, the CBS inequality gives an upper bound

\[
|\text{Cov}(X, Y)| \leq \sqrt{\frac{266}{(45)^2} \times \frac{28}{45}} = 0.285889.
\]

8.4.11 From Exercise (8.4.9) we know that the exact value is \( \text{Cov}(X, Y) = 0 \). However, to find \( \text{Var}(X) \), we need the marginal density of \( X \), which is given in Example (7.2.2). That is,

\[
f_X(x) = \frac{2}{\pi} \sqrt{1 - x^2}, \quad x \in (-1, 1).
\]

So,

\[
\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{2}{\pi} \int_0^1 x^2 \sqrt{1 - x^2} \, dx - \left( \frac{2}{\pi} \int_0^1 \sqrt{1 - x^2} \, dx \right)^2
\]

\[
= \frac{2}{\pi} \int_0^1 \sqrt{1 - u^2} \, du, \quad u = x^2,
\]

\[
= \frac{2}{\pi} \int_0^1 u^{\frac{1}{2}} (1 - u)^{\frac{1}{2}} \, du = \frac{2 \Gamma \left( \frac{1}{2} \right) \Gamma \left( \frac{1}{2} \right)}{\pi \Gamma(1)} = \frac{1}{4}.
\]

Here we used that \( \Gamma \left( \frac{1}{2} \right) = \sqrt{\pi} \). Hence, the CBS inequality shows that \( |\text{Cov}(X, Y)| \leq \sqrt{\frac{1}{4}} = \frac{1}{2} \), which is much larger than the actual zero covariance.

8.4.12 (i) \( \text{Var}(Y) = 9 \text{Var}(X) = 9np(1-p) \), (ii) \( \text{Var}(W) = 16 \text{Var}(X) = 16np(1-p) \), (iii) \( \text{Cov}(Y, W) = 12 \text{Var}(X) = 12np(1-p) \), (iv) \( \text{Cov}(Y, W) = 12np(1-p)/(12np(1-p)) = 1 \). This should be obvious since \( Y \) and \( W \) are positively linearly related.

8.4.13 By the GLST and the linearity property of expectations,

\[
E(X_n) = E \left( \frac{X_1 + X_2 + \ldots + X_n}{n} \right) = \frac{E(X_1) + E(X_2) + \ldots + E(X_n)}{n} = \frac{n \mu}{n} = \mu.
\]

8.4.14 Since we know that

\[
\text{Var}(a_1 X_1 + a_2 X_2 + \ldots + a_n X_n) = a_1^2 \text{Var}(X_1) + a_2^2 \text{Var}(X_2) + \ldots + a_n^2 \text{Var}(X_n),
\]

when \( X_i \) are uncorrelated, all we need to do is insert \( a_i = \frac{1}{n} \) for \( i = 1, 2, \ldots, n \) to get the result.

8.4.15 We will prove the result for the discrete case (the other case being similar). The first is just the fact that \( a \leq a_i \) for all \( a_i \in \Delta \). So,

\[
a = \sum_{a \in \Delta} a P(X = a) \leq \sum_{a \in \Delta} a_i P(X = a_i) = E(X).
\]

For monotonicity, use the fact that \( X \leq Y \) implies that \( 0 \leq Y - X \). So, \( 0 \leq E(Y - X) = E(Y) - E(X) \).
Since \( P(X + Y = 12, X - Y = 1) = 0 \) but \( P(X + Y = 12) > 0 \) and \( P(X - Y = 1) > 0 \), the two random variables are not independent. On the other hand, \( \text{Cov}(X + Y, X - Y) = \text{Var}(X) - \text{Var}(Y) + \text{Cov}(X, Y) = \text{Cov}(X, Y) = 0 \), giving that they are uncorrelated.

The CBS inequality says that \( (\text{Cov}(X, Y))^2 \leq \text{Var}(X)\text{Var}(Y) \), which is violated by the statement. Hence, no such pair of random variables can exist.

Note that \( \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) = 12 \). Similarly, \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) = 8 \). Furthermore,
\[
\text{Cov}(X - Y, X + Y) = \text{Var}(X) - \text{Var}(Y) = 1 - 9 = -8.
\]

Therefore,
\[
\text{Corr}(X - Y, X + Y) = \frac{\text{Cov}(X - Y, X + Y)}{\sqrt{\text{Var}(X - Y)\text{Var}(X + Y)}} = \frac{-8}{\sqrt{12 \times 8}} = -0.816.
\]

\[\begin{align*}
\text{Var}(Z) & = p^2 \text{Var}(X) + (1 - p)^2 \text{Var}(Y) + 2p(1 - p)\text{Cov}(X, Y) \\
& = p^2(\text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)) - 2p(\text{Var}(Y) + \text{Cov}(X, Y)) + \text{Var}(Y) \\
& = ap^2 - 2bp + c,
\end{align*}\]

where \( a = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) = \text{Var}(X) \), \( b = \text{Var}(Y) - \text{Cov}(X, Y) \) and \( c = \text{Var}(Y) \). Since \( a > 0 \), this is a convex parabola, its minimum occurs at \( 2bp = 0 \) or \( p = \frac{b}{a} \). Note that for \( p \in [0, 1] \), the value of the parabola at \( p = 0 \) is \( \text{Var}(Y) \) and \( p = 1 \) its value is \( \text{Var}(X) \). The minimum value of \( \text{Var}(Z) \) will occur for \( p \in (0, 1) \) if and only if \( 0 < \frac{b}{a} < 1 \). This is equivalent to saying that \( \text{Cov}(X, Y) < \min\{\text{Var}(X), \text{Var}(Y)\} \). Yet another way of saying this is that diversification reduces risk from both risky assets’ risks if and only if

\[
\text{Corr}(X, Y) < \min\left\{\sqrt[2]{\frac{\text{Var}(X)}{\text{Var}(Y)}}, \sqrt[2]{\frac{\text{Var}(Y)}{\text{Var}(X)}}\right\}.
\]

Otherwise, one of the risky assets is less risky than the portfolio.

Let \((X, Y)\) have the bivariate hypergeometric density
\[
P(X = k, Y = j) = \binom{G}{k} \binom{n}{k} \binom{R}{j} \binom{S}{n - k - j} \binom{N}{n},
\]
where \( N = G + R + S \). Since the marginal densities are ordinary hypergeometric densities, by the help of we can show that
\[
E(X) = np, \ E(Y) = nq, \ \text{Var}(X) = cnp(1 - p), \ \text{Var}(Y) = cnq(1 - q),
\]
where \( c = \frac{N}{N - 1} \), \( p = G/N \), \( q = R/N \). So, we need only find \( E(XY) \). For this purpose, we have
\[
E(XY) = \frac{1}{\binom{n}{k}} \sum_{k=0}^{n} \sum_{j=0}^{k} \binom{G}{k} \binom{R}{j} \binom{S}{n - k - j} / \text{Binomial}(1, p) = \mathbb{E}(X) \mathbb{E}(Y) = npq.
\]

This gives that the correlation is
\[
\text{Corr}(X, Y) = -\frac{npq}{\sqrt{\text{Var}(X)\text{Var}(Y)}},
\]
with
\[
\text{Var}(X) = \frac{RG(n - 1)}{N(N - 1)} - \frac{nGnR}{N^2} = -npq.
\]

\[\begin{align*}
K \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \binom{i + j}{i} p^i q^j & = K \sum_{i=0}^{\infty} p^i \sum_{j=0}^{\infty} \binom{i + j}{i} q^j \\
& = K \sum_{i=0}^{\infty} p^i \frac{1}{(1 - q)^{i + 1}} - K \sum_{i=0}^{\infty} \binom{1}{1 - q}^i \\
& = K \frac{1}{1 - q} - \frac{1 - p - q}{1 - p} = 1.
\end{align*}\]

This gives that \( K = 1 - p - q \). The marginal of \( Y \) is
\[
P(Y = j) = (1 - p - q)q^j \sum_{i=0}^{\infty} \binom{i + j}{i} p^i.
\]
\[
E(XY) = K \sum_{i} \sum_{j} ij \left( \frac{i}{\binom{1-p}{i}} \right) \frac{p^i q^j}{(1-p)^{i+j+1}}
\]

Therefore, \( Y \sim \text{Geometric}(\frac{1-p-q}{1-p}) \). And by a similar argument we see that \( X \sim \text{Geometric}(\frac{1-p-q}{1-p}) \).

\[
E(XY) = Kq \sum_{i} \sum_{j} ij \left( \frac{i}{\binom{1-p}{i}} \right) \frac{p^i q^j}{(1-p)^{i+j+1}} = \frac{Kpq}{(1-p)^2} \int_{0}^{1} v^2 \, dv = \frac{2pq}{K^2}.
\]

Therefore, the correlation is

\[
\text{Cor}(X, Y) = \frac{\rho_{X,Y}}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{\rho_{X,Y}}{\sqrt{\frac{pq}{(1-p)^2} \cdot \frac{pq}{(1-q)^2}}} = \frac{pq}{(1-p)(1-q)}. \]

9.4.1 Recall that the mgf of \( X \) is \((1-2t)^{m/2}\). So, the mgf of \( X + Y \) is

\[
E(e^{iX})E(e^{iY}) = (1-2t)^{m/2} (1-2t)^{n/2} = (1-2t)^{(m+n)/2}, \quad t < \frac{1}{2}.
\]

This is the mgf of another chi square random variable, the one which has \( n + m \) degrees of freedom. Hence, \( X + Y \sim \chi^2_{n+m} \).

(Thanks to Tim A, again!)

9.4.2 The mgf is

\[
E(e^{iX}) = 2 \int_{0}^{1} e^{i} x dx = \frac{2(1 - e^{i} + 1)}{1 - 2i}.
\]

The same argument gives that \( E(e^{iY}) = \frac{2(1 - e^{i} + 1)}{1 - 2i} \).

9.4.3 \( E(e^{iX+Y}) = E(e^{iX})E(e^{iY}) = (1 - p + pe^{i})^n e^{i \lambda (e^{i} - 1)} \).

9.4.4 The mgf of \( X \) is \( e^{2t^2} \). Since \( X \sim N(2, 4) \), its mgf is \( e^{2t+2t^2} \). Also the mgf of \( Y \) is \( e^{3t^2} \). Hence, the mgf of \( Z \) is \( e^{2t+3t^2} \). That is, \( Z \sim N(2, 40) \).

9.4.5 The characteristic function of \( Y \) is

\[
E(e^{iY}) = E \left( e^{iX+Y} \right) = \prod_{j=1}^{n} \left( e^{-p^j} \right)^{n} = e^{-p^j}.
\]

So, \( Y \) is also \( \text{Cauchy}(0,1) \).
9.4.11 Just note that for $n = 2$, $X_n$ is related to $X_1 + X_2$, and hence is normal. And $\frac{(n-1)s^2}{\sigma^2}$ is related to $(X_1 - X_2)^2$, which should produce chi square, and by Exercise (9.4.10) the two should be independent. (The rest is just simple algebra.)

9.4.12 By the result of Exercise (9.4.11), $\frac{X_n - \mu}{\sigma / \sqrt{n}}$ is $N(0,1)$ and $\frac{(n-1)s^2}{\sigma^2}$ is $\chi^2_{n-1}$, and the two random variables are independent, hence by the definition of $l(n-1)$,

$$l(n-1) \sim \frac{X_n - \mu}{\sigma / \sqrt{n}} \sim \chi^2_{n-1}.\frac{S_n}{\sqrt{n}}.$$

9.4.13 By the result of Exercise (9.4.11), $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ and $\frac{(m-1)s^2}{\sigma^2} \sim \chi^2_{m-1}$ and the two random variables are independent, hence by the definition of an $F$ distribution,

$$F(n-1, m-1) \sim \frac{\chi^2_{n-1}}{\chi^2_{m-1}} \sim \frac{S_n^2 / \sigma^2}{S_m^2 / \sigma^2}.$$ 

9.4.14 Since the density of $l(n)$ is symmetric about zero, $E(T) = 0$, and $\text{Var}(T) = E(T^2)$, now for $n > 2$, we have

$$\text{Var}(T) = \frac{\Gamma \left(\frac{n+1}{2}\right)}{\sqrt{n \pi} \Gamma \left(\frac{n}{2}\right)} \int_0^{\infty} t^2 \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \, dt 
= C_n \int_0^{\infty} \left(\frac{t^2}{n} + \frac{n}{n} \right) \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \, dt 
= C_n \int_0^{\infty} \left(n \frac{t^2}{n} + n - n \right) \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \, dt 
= nC_n \int_0^{\infty} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \, dt - nC_n \int_0^{\infty} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \, dt 
= \frac{nC_n \int_0^{\infty} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \, dt}{\sqrt{n}}$$

where $C_n := \frac{\Gamma \left(\frac{n+1}{2}\right)}{\sqrt{n \pi} \Gamma \left(\frac{n}{2}\right)}$. The last integral is one, but the first integral has wrong denominator under the $t^2$ term. To get the correct denominator, let $v = \frac{\sqrt{n} t}{\sqrt{n}}$. This gives that

$$\text{Var}(T) = \frac{nC_n \int_0^{\infty} \left(1 + \frac{v^2}{n-2}\right)^{-(n+1)/2} \, dv}{\sqrt{n-2}} 
= \frac{nC_n \sqrt{n-2}}{\sqrt{n-2}} = n 
= \frac{n\sqrt{n} \Gamma \left(\frac{n+1}{2}\right) \sqrt{n-2} \Gamma \left(\frac{n-1}{2}\right)}{\sqrt{n-2} \Gamma \left(\frac{n}{2}\right) \Gamma \left(\frac{n-1}{2}\right)} 
= \frac{n(n-1)}{n-2} - n = \frac{n}{n-2}.$$ 

9.4.15 The expectation is,

$$E(V) = \int_0^{\infty} x \frac{\Gamma \left(\frac{a+b}{2}\right) x^{a+b/2} / \Gamma \left(\frac{a}{2}\right) \Gamma \left(\frac{b}{2}\right)}{(b+ax)^{(a+b)/2}} \, dx 
= C_{a,b} \int_0^{\infty} \frac{x^{a+b/2}}{(b+ax)^{(a+b)/2}} \, dx$$

where $C_{a,b} := \frac{\Gamma \left(\frac{a+b}{2}\right) x^{a+b/2}}{\Gamma \left(\frac{a+b}{2}\right)}$ is a constant. In order to use the fact that the area under any density is one, we need to correct the constants sitting next to $x$ in the denominator. Had they been $b^* = b + 2$ and $a^* = a + 2$ then it would have worked out since $\frac{\Gamma \left(\frac{a+b}{2}\right) x^{a+b/2}}{\Gamma \left(\frac{a+b}{2}\right)} = \frac{\Gamma \left(\frac{a+b}{2}\right) x^{a+b/2}}{\Gamma \left(\frac{a+b}{2}\right)}$. So, let us first pull $b$ out of the denominator term, and then replace $\frac{\Gamma \left(\frac{a+b}{2}\right) x^{a+b/2}}{\Gamma \left(\frac{a+b}{2}\right)}$ by $\frac{\Gamma \left(\frac{a+b}{2}\right) x^{a+b/2}}{\Gamma \left(\frac{a+b}{2}\right)}$.

$$E(V) = C_{a,b} \int_0^{\infty} \frac{x^{(a+b)/2}}{(b+ax)^{(a+b)/2}} \, dx 
= C_{a,b} \int_0^{\infty} \frac{b^{(a+b)/2}}{(b+ax)^{(a+b)/2}} \frac{x^{(a+b)/2}}{(b+ax)^{(a+b)/2}} \, dx 
= C_{a,b} \int_0^{\infty} \frac{b^{(a+b)/2}}{(b+ax)^{(a+b)/2}} x^{(a+b)/2} \, dx 
= C_{a,b} \int_0^{\infty} \frac{b^{(a+b)/2} x^{(a+b)/2}}{(b+ax)^{(a+b)/2}} \, dx$$

The same trick works for $E(V^2)$. The algebra is grubby. The reader should fill in the details.

9.4.16 The joint density of $X, Y$ is

$$f(x, y) := \frac{1}{2\pi \sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} [x^2 + y^2 - 2\rho xy] \right\},$$

The inverse of the transformation is $X = \frac{U+V}{\sqrt{2}}, Y = \frac{2V - U}{\sqrt{2}}$. The absolute value of the jacobian is $\frac{\sqrt{2}}{2}$. Note that

$$x^2 + y^2 = \frac{u^2 + 2uv + v^2}{9} + \frac{u^2 - 4uv + 4v^2}{81}.$$ 

Hence, we plug this expression into the joint density of $X, Y$ for $x^2 + y^2$ and simplify. After some algebra, one gets that the joint density of $U, V$ is bivariate normal with both respective means 0.

(Thanks Tim!)

9.4.17 The joint density of $X, Y$ is

$$f(x, y) := \frac{1}{2\pi \sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} [x^2 + y^2 - 2\rho xy] \right\},$$
The inverse of the transformation is $X = \frac{U + V}{2}, Y = \frac{2V - U}{3}$. The absolute value of the Jacobian is $\frac{1}{9}$. Note that

$$x^2 + y^2 = \frac{u^2 + 2uv + v^2}{9} + \frac{u^2 - 4uv + 4v^2}{81}.$$ 

Hence, we plug this expression into the joint density of $X, Y$ for $x^2 + y^2$ and simplify. After some algebra, one gets that the joint density of $U, V$ is bivariate normal with both respective means 0. The reader should finish the rest of the algebra.

**9.4.18** The joint density of $X, Y$ is

$$f(x, y) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} [x^2 + y^2 - 2\rho xy] \right\},$$

The inverse of the transformation is $X = \frac{U + V}{3}, Y = \frac{2V - U}{9}$. The absolute value of the Jacobian is $\frac{1}{9}$. Note that

$$x^2 + y^2 = \frac{u^2 + 2uv + v^2}{9} + \frac{u^2 - 4uv + 4v^2}{81}.$$ 

Hence, we plug this expression into the joint density of $X, Y$ for $x^2 + y^2$ and simplify. After some algebra, one gets that the joint density of $U, V$ is bivariate normal with both respective means 0. The reader should finish the rest of the algebra.

**9.4.19** The joint density of $X, Y$ is

$$f(x, y) = \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} x^{a-1} y^{b-1} e^{-\lambda(x+y)}, \quad (x, y > 0).$$

Let $z = x + y$ and $w = x/(x+y) = x/z$. So, $x = wz$ and $y = z(1-w)$. The Jacobian is

$$J = \begin{vmatrix} w & z \\ 1-w & -z \end{vmatrix} = -z.$$ 

So, the joint density of $Z, W$ is

$$f_{Z,W}(z,w) = |J| f(wz, z(1-w)), \quad 0 < w < 1, \quad z > 0,$$

$$= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} (wz)^{a-1} (z(1-w))^{b-1} e^{-\lambda z}$$

$$= \frac{\lambda^{a+b}}{\Gamma(a+b)} z^{a-1} e^{-\lambda z} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} wz^{a-1} (1-w)^{b-1}$$

$$= f_z(z) \cdot f_w(w).$$

Thus, $Z, W$ are independent. $Z$ has a gamma density and $W$ has a beta density.

**9.4.20** Since $X^2, Y^2 \sim G(1/2, 1/2)$, Exercise (9.4.19) gives that $Z$ has beta density with parameters $1/2, 1/2$.
10.4.8 Recall that when $X \sim \text{Geometric}(p)$ then $P(X \leq j) = 1 - (1 - p)^{j+1}$ for any $j = 0, 1, 2, \ldots$. Hence, $P(X \geq j) = 1 - P(X \leq j - 1) = (1 - p)^j$. Using this we get that

$$P(X \geq k + j | X \geq k) = \frac{P(X \geq k + j, X \geq k)}{P(X \geq k)} = \frac{P(X \geq k + j)}{(1 - p)^{k}} = (1 - p)^j.$$ 

So, $P(X \geq k + j | X \geq k) = P(X \geq j)$.

10.4.14 By the TTP,

$$f_X(x) = \int_0^\infty f_{X|Y}(x) f_Y(y) \, dy$$

$$= \int_0^\infty y^\lambda e^{-y} \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha - 1} e^{-\lambda y} \, dy$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha + 1} e^{-(\lambda + 1)y} \, dy$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)(\lambda + 1)^{\alpha + 1}} \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha + 1} e^{-\lambda y} \, dy$$

This is also a Pareto density.

10.4.12 The TTP gives that

$$P\left(\frac{X}{Y} \leq \frac{1}{4}\right) = P(X \leq (Y/4)) = \int_0^1 P(X \leq y/4) \, dy$$

$$= \int_0^1 (1 - e^{-\lambda y/4}) \, dy$$

$$= 1 - \frac{4}{\lambda} \left(1 - e^{-\lambda/4}\right).$$

10.4.13 The bivariate density of $(X, Y)$ is

$$f(x, y) := \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2(1 - \rho^2)} \left[x^2 + 2\rho xy + y^2\right]\right\},$$

The marginal density of $Y$ is $N(0, 1)$. Hence, the conditional density of $X$ given $Y = 2$ is

$$f_{X|Y=2}(x) = \frac{f(x, 2)}{f_Y(2)}$$

$$= \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2(1 - \rho^2)} \left[x^2 + 2\rho xy + y^2\right]\right\}.$$

After simplifying, as predicted by Example (10.3.3), we see that the conditional density of $X$ given $Y = 2$ is normal with mean $2\rho$ and variance $(1 - \rho^2)$.

10.4.15 The marginal density of $Y$ is

$$f_Y(y) = \int_y^1 f(x, y) \, dx = \int_y^1 3x \, dx = \frac{3(1 - y^2)}{2}, \quad y \in (0, 1).$$

Now the conditional density of $X$ given $Y = y$ is

$$f_{X|Y}(x) = \frac{f(x, y)}{f_Y(y)} = \frac{3x}{3(1 - y^2)^2} = \frac{2x}{1 - y^2}, \quad y < x < 1.$$ 

So, we see that

$$P(X \leq 0.5 | Y = 0.25) = \int_{0.25}^{0.5} 3(1 - y^2)^{-2} \, dy = 16 \left(\frac{1}{4} - \frac{1}{16}\right).$$

Now note that

$$P(Y \leq 0.25) = \int_0^{0.25} \frac{3(1 - y^2)}{2} \, dy = 3(0.25 - \frac{1}{16}).$$

Also, note that

$$P(X \leq 0.5, Y \leq 0.25) = \int_0^{0.5} \int_0^{x \min\{x, 0.25\}} 3x \, dy \, dx$$

$$= 3 \int_0^{0.5} x \min\{x, 0.25\} \, dx$$

$$= 3 \int_0^{0.25} x^2 \, dx + \frac{3}{4} \int_0^{0.5} x \, dx$$

$$= \frac{1}{4} + \frac{3}{8} \left(\frac{1}{4} - \frac{1}{16}\right).$$

Therefore,

$$P(X \leq 0.5 | Y \leq 0.25) = \frac{P(X \leq 0.5, Y \leq 0.25)}{P(Y \leq 0.25)}.$$ 

10.4.16 It is easy to see that $X + Y \sim B(8, p)$. The joint density of $X, Y$ is

$$P(X = x, X + Y = y) = P(X = x, Y = y - x) = P(X = x)P(Y = y - x),$$

provided $x \leq y$ and $x \in \{0, 1, 2, 3, 4\}$ and $y = 0, 1, \ldots, 8$. So, the conditional density of $X$ given $X + Y = 3$ is

$$P(X = x | X + Y = 3) = \frac{P(X = x, X + Y = 3)}{P(X + Y = 3)}$$

$$= \frac{P(X = x)P(Y = 3 - x)}{P(X + Y = 3)}$$

$$= \binom{4}{x} p^x (1 - p)^{4 - x}$$

$$= \binom{4}{x} p^x (1 - p)^{4 - x}$$

$$= \binom{4}{x} p^x (1 - p)^{4 - x}.$$ 

This is $\text{Hypergeometric}(4, 4, 3)$. 

10.4.17
10.4.16 Using the result of Exercise 10.4.16,
\[
E(X|X + Y = 3) = \sum_{x=0}^{3} x \binom{4}{x} \binom{4}{3-x} / \binom{8}{3} = 3(4 \times 6) + (2 \times 6 \times 4) + (3 \times 4) / 56 = 84/56 = 3/2.
\]

10.4.17 Let \( R = 25N(10 + X) \) be the daily revenue. By TTE
\[
E(R) = 25E(NE(10 + X|N)) = 25(10 + 16)E(N) = 25 \times 26 \times 50 = \$32,500,000.
\]
By TTV,
\[
Var(R) = 625Var(N(10 + X)) = 625 \{ E(Var(N(10 + X)|N)) + Var(E(N(10 + X)|N)) \} = 625 \{ E(N^2 16(0,2)) + Var(26N) \} = 625 \{ (3,2(50 + 2500) + (676 \times 50)) = \$26,225,000,00.
\]
The standard deviation is \$5,121,035.

10.4.18 Recall that the mgf of \( X \) is \( E(e^{tX}) = e^{t^2/2} \). Therefore, by the TTE,
\[
E(e^{XY/3}) = E\left( E(e^{XY/3}|Y) \right) = E\left( E(e^{Y^2/18}) \right).
\]
Now \( Y^2 \sim \chi^2_1 \) by Example 9.2.2. Hence, using Remark 9.1.1 again, we see that
\[
E(e^{XY/3}) = \left( \frac{1}{1 - 2(1/18)} \right)^{1/2} = \frac{3}{\sqrt{8}}.
\]

11.6.1 \( \sup_{\omega \in S}|U(\omega) - V(\omega)| = \max\{1,1,4\} = 4. \)

11.6.2 Note that \( P(\{c\}) = \frac{3}{12}. \) So,
\[
E[U - V] = \frac{1 \times 1}{3} + \frac{1 \times 1}{4} + \frac{4 \times 5}{12} = \frac{27}{12},
\]
\[
E[U - V]^2 = \frac{1 \times 1}{3} + \frac{1 \times 1}{4} + \frac{4 \times 5}{12} = \frac{87}{12},
\]
\[
E\left( \frac{|U - V|}{1 + |U - V|} \right) = \frac{\frac{1}{3} \times 1}{\frac{1}{3} + \frac{1}{4} + \frac{4 \times 5}{12} - \frac{1}{24}}.
\]

11.6.3 We can bound \( P(Yn - Y \geq \varepsilon) \) via Chebyshev’s inequality, by \( \frac{E(Yn - Y)^2}{\varepsilon} \) or by \( \frac{E(Yn - Y)^2}{\varepsilon^2}. \) This gives the result.
\[ \frac{1}{n^2} \sum_{i \neq j} |\text{Cov}(X_i, X_j)| \leq \frac{1}{n^2} \sum_{i} \sum_{j} |\text{Cov}(X_i, X_j)| = \frac{1}{n^2} \sum_{i} \sum_{j \neq i} |\text{Cov}(X_i, X_j)| \leq \frac{1}{n^2} \sum_{i} \sum_{j \neq i} \sqrt{\text{Var}(X_i) \text{Var}(X_j)} + \frac{\delta}{n^2} \sum_{i} \sum_{j \neq i} 1 \leq 2BK + \frac{\delta(n-1)}{n^2}. \]

Letting \( n \) get large makes \( \text{Var}(X_n) < \delta \), and \( \delta \) is picked arbitrarily small.

**11.6.11** First assume that \( \mu = 0 \) and apply Proposition (11.6.10) to get the result. For the general case verify that

\[ Y_n - \mu = Z_n - \frac{\mu}{2^n} \]

where \( Z_n = \sum_{i=1}^{n} X_{n-i} - \frac{\mu}{2^i} \).

Now justify the following steps.

\[ P(|Y_n - \mu| \geq \varepsilon) = P(|Z_n - (\mu/2^n)| \geq \varepsilon) \leq P(|Z_n| \geq \varepsilon - (\mu/2^n)) \leq P(|Z_n| \geq \varepsilon/2), \]

for all large \( n \).

Now make the conclusion.

**11.6.12** Verify the following integration results and then apply Proposition (11.6.10).

\[ E(X_k) = \frac{1}{2} \int_{-1}^{1} \sin(\pi ku) \, du = 0, \]

\[ E(X_k^2) = \frac{1}{2} \int_{-1}^{1} (\sin(\pi ku))^2 \, du = \frac{1}{2}, \]

\[ E(X_k X_j) = \frac{1}{2} \int_{-1}^{1} \sin(\pi ku) \sin(\pi ju) \, du = 0, \quad \text{when} \ k \neq j. \]

**11.6.13** Since \( E(|X_i|^4) < \infty \), the strong law of large numbers gives that \( Y_n \) converge to \( E(X_i) = \frac{1}{3} = c \) almost surely.

**11.6.14** Since \( E(|X_i|^4) < \infty \), the strong law of large numbers gives that \( Y_n \) converge to \( E(X_i) = r = c \) almost surely.

**11.6.15** Since \( Y_n \sim \text{Cauchy}(0, 1) \) for each \( n \), there cannot exist any constant \( c \) so that \( Y_n \) will converge to that constant almost surely.

**11.6.16** Let \( S \) contain \( m \) elements, and let \( U_n \) converge to \( V \) in \( L^p \) sense. Just note that for any fixed \( \omega_k \in \{1, 2, \cdots, m\} \), we have

\[ P(|\{\omega_k\}|U_n(\omega_k) - V(\omega_k)|^p \leq E|U_n - V|^p \to 0. \]

When \( P(|\{\omega_k\}|) > 0 \), for any \( \varepsilon > 0 \), there exists an \( N(\varepsilon, \omega_k) \), such that \( |U_n(\omega_k) - V(\omega_k)| < \varepsilon \) for all \( n > N(\varepsilon, \omega_k) \). Since we need only prove the result over a set with probability one, we may safely throw away finite many \( \omega \) for which \( P(|\{\omega_k\}|) = 0 \). Therefore, assume without loss of generality that for each \( \omega \in S \), \( P(|\{\omega_k\}|) > 0 \), Take \( M(\varepsilon) = \max_{1 \leq k \leq m} N(\varepsilon, \omega_k) \), So, for all \( n > M(\varepsilon) \), we have

\[ \max_{1 \leq k \leq m} |U_n(\omega_k) - V(\omega_k)| < \varepsilon \].

**11.6.17** and let \( U_n \) converge to \( V \) almost surely. Since we need only prove the result over a set with probability one, we may safely throw away finite many \( \omega \) for which \( P(|\{\omega_k\}|) = 0 \). So, let \( S \) contain \( m \) elements, each of which has positive probability. Just note that for any fixed \( \omega_k \), \( k = 1, 2, \cdots, m \), we are given that for any \( \varepsilon > 0 \), there exists an \( N(\varepsilon, \omega_k) \), such that \( |U_n(\omega_k) - V(\omega_k)| < \varepsilon \) for all \( n > N(\varepsilon, \omega_k) \). Take \( M(\varepsilon) = \max_{1 \leq k \leq m} N(\varepsilon, \omega_k) \), So, for all \( n > M(\varepsilon) \), we have

\[ \max_{1 \leq k \leq m} |U_n(\omega_k) - V(\omega_k)| < \varepsilon \].

**11.6.18** We will construct an example for which \( U_n \) converge to zero almost surely, but \( E|U_n|^p = \infty \) for each \( n \). Let \( S = \{\omega_0, \omega_1, \cdots\} \) so that \( P(|\{\omega_k\}|) > 0 \) for each \( k = 0, 1, 2, \cdots \). For instance, take \( P(|\{\omega_k\}|) = \frac{1}{2^n} \) if you like. Define

\[ U_n(\omega_k) := \begin{cases} \frac{1}{P(|\{\omega_k\}|)} & \text{if} \ k > n, \\ 0 & \text{if} \ k \leq n, \end{cases} \]

Note that for any fixed \( \omega_k \), for \( n > k \), we have \( U_n(\omega_k) = 0 \). Hence, \( U_n(\omega) \to 0 \) for each fixed \( \omega \in S \). That is, \( U_n \) converges to 0 almost surely (and hence in probability as well). On the other hand, for each \( n \geq 1 \),

\[ E|U_n|^p = \sum_{k=0}^{\infty} U_n(\omega_k) P(|\{\omega_k\}|) = \sum_{k=m+1}^{\infty} P(|\{\omega_k\}|) P(|\{\omega_k\}|) = \infty. \]

**11.6.19** We throw away any \( \omega \in S \) for which \( P(|\{\omega_k\}|) = 0 \). The collection of all these points still have probability zero. So, when \( P(|\{\omega_k\}|) > 0 \), and \( U_n(\omega_k) \neq V(\omega_k) \), then there exists an \( \varepsilon > 0 \) and increasing integers \( m_1, m_2, \cdots \) so that \( |U_n(\omega_k) - V(\omega_k)| > \varepsilon \) for all \( n = m_1, m_2, \cdots \). Note that

\[ \{\omega_k\} \subset \{\omega \in S : |U_n(\omega) - V(\omega)| > \varepsilon \}. \]
for all \( n = m_1, m_2, \cdots \). Therefore, we see that

\[
P(|U_n - V| > \varepsilon) \geq P(\omega^*) > 0, \quad \text{for all } n = m_1, m_2, \cdots.
\]

This contradicts the fact that \( U_n \) converges to \( V \) in probability.

**11.6.20** When \( S \) is a finite set, Exercise (11.6.19) shows that convergence almost surely will take place. Then Exercise (11.6.17) gives that almost sure uniform convergence will take place. This in turn (trivially) implies the convergence in the \( L^p \) sense will take place. So, the statement is true.

**11.6.21** Let \( S = \{\omega_1, \omega_2\} \) with each point have the same probability. Define \( U_n(\omega_1) = 2 \) and \( U_n(\omega_2) = 1 \) for all \( n = 1, 2, \cdots \). Also, define \( V(\omega_1) = 1 \) and \( V(\omega_2) = 2 \). Note that all \( U_n \) and \( V \) have the same distribution, so trivially, \( U_n \) converge to \( V \) in distribution (and they are all defined over the same space). However, \( U_n \) does not converge to \( V \) in probability since

\[
E \left( \frac{|U_n - V|}{1 + |U_n - V|} \right) = \frac{1}{1 + 1} = \frac{1}{2}, \quad \text{for all } n.
\]