Solutions for Homework 1, due January 29

Problems assigned ∼1.2.6, 1.2.9, 1.2.10(iii), 1.2.12. Each problem in this and all subsequent assignments is worth 9 pts.

1.2.6. Observe that $I_{E^c} = 1 - I_E$. We are given that $EI_E I_H = EI_E I_H$. Then $P(E^c H) = EI_E I_H = E(1 - I_{C})I_H = EI_H - EI_E I_H = P(H) - P(E)P(H) = P(E^c)P(H)$. Thus, if $P(EH) = P(E)P(H)$ we can put $c$ sign on top of one of the events. Obviously on top of any of them. Since $P(E^c H) = P(E^c)P(H)$, this equality holds if we replace $H$ by $H^c$ and we are done.

1.2.9. We are given that 
\[ P(X = a, Y = b, Z = c) = P(X = a)P(X = b)P(Z = c) \] (1) for all $a \in A, b \in B,$ and $c \in C$, where $A, B, C$ are discrete sets containing all values of $X, Y, Z$, respectively. By summing up (1) with respect to $c \in C$ and observing that, for instance, 
\[ \sum_c P(X = a, Y = b, Z = c) = P(X = a, Y = b, Z \in C) = P(X = a, Y = b), \] we conclude that $X$ and $Y$ are independent. Then (1) is rewritten as 
\[ P((X, Y) = (a, b), Z = c) = P(X = a)P(X = b)P(Z = c) \]
\[ = P(X = a, Y = b)P(Z = c) = P((X, Y) = (a, b))P(Z = c) \]
and we are done.

1.2.10(iii). We are given that 
\[ P(X^1 = a_1, X^2 = a_2, Y^1 = b_1, Y^2 = b_2) \]
\[ = P(X = (a_1, a_2), Y = (b_1, b_2)) = P(X = (a_1, a_2))P(Y = (b_1, b_2)) \]
\[ = P(X^1 = a_1)P(X^2 = a_2)P(Y^1 = b_1)P(Y^2 = b_2). \]
Sum up the equality between the extreme terms with respect to all possible values of $a_2$ and $b_1$ and use that 
\[ \sum_{a_2, b_1} P(X^1 = a_1, X^2 = a_2, Y^1 = b_1, Y^2 = b_2) = P(X^1 = a_1, Y^2 = b_2) \]

1.2.12. Particular case from which it is clear what to do in the general case. Assume that $X_1, X_2, X_3$ are independent and let us prove that $X_1$ and $X_3$ are independent. We are given that 
\[ P(X_1 = a, X_2 = b, X_3 = c) = P(X_1 = a)P(X_2 = b)P(X_3 = c) \]
for all \(a \in A, b \in B, c \in C\), where \(A, B, C\) are the sets of values of \(X_1, X_2, X_3\). Then we observe that 
\[
\{X_1 = a, X_3 = c\} = \{X_1 = a, X_2 \in B, X_3 = c\} = \bigcup_{b \in B} \{X_1 = a, X_2 = b, X_3 = c\}
\]
and the events in the union are disjoint. Hence 
\[
P(X_1 = a, X_2 = b) = \sum_{b \in B} P(X_1 = a, X_2 = b, X_3 = c)
\]
\[
= \sum_{b \in B} P(X_1 = a) P(X_2 = b) P(X_3 = c)
\]
\[
= P(X_1 = a) P(X_3 = c) \sum_{b \in B} P(X_2 = b) = P(X_1 = a) P(X_3 = c) P(X_2 \in B)
\]
and the result follows.

**Comments.** Some of you used the notation \(P(X)\) and \(P(X^c)\), where \(X\) is a random variable. I don’t know what these symbols mean. One person used the theory of characteristic functions. This is not allowed. You have to use only what was already in the book before the homework is assigned. Otherwise you could just find the facts, I ask you to prove, in a book and refer to it. This is also inappropriate because our random variables are not necessarily real or vector valued.

One person didn’t do HW 1. This can cost dearly in the final count of grades (see the syllabus).

Ordered grades for HW1:

---

**Solutions for Homework 2, due February 1**

Problems assigned 1.3.8, 1.3.12, 1.3.13, kind of 1.2.5.

1.3.8. We have 
\[
P(u(Z) = v(Z)) = \sum_{c \in C} P(u(Z) = v(Z), Z = c)
\]
\[
= \sum_{c \in C} P(u(c) = v(c), Z = c).
\]
Here 
\[
P(u(c) = v(c), Z = c) = P(Z = c)
\]
if \( P(Z = c) > 0 \) because for those \( c \) we are given that \( u(c) = v(c) \). On the other hand, if \( P(Z = c) = 0 \), then both sides of (3) equal zero. Thus (3) always holds and coming back to (2) we conclude

\[
P(u(Z) = v(Z)) = \sum_{c \in C} P(Z = c) = 1
\]

which is exactly what we need.

1.3.12. Owing to Problem 8, it suffices to check that \( E(X \mid Z = c) = E(X) \), whenever \( P(Z = c) > 0 \). It is easy to see or we just know that if random variables are independent, any functions of them are independent. The random variable \( I_{Z = c} \) is a function of \( Z \), therefore, \( X \) and \( I_{Z = c} \) are independent and

\[
E_X I_{Z = c} = (EX)EI_{Z = c} = P(Z = c)EX.
\]

Upon dividing both parts by \( P(Z = c) \) we find \( E(X \mid Z = c) = EX \).

1.3.13. Very tempting is to use Theorem 1.2.14 with \( f \equiv 1 \). However this theorem comes later and you are not allowed to use its result. First we consider the case that \( X \geq 0 \). This assumption allow us to change the order of summation in the following computations:

\[
E(E(X \mid Y)) = \sum_b E(X \mid Y = b)P(Y = b)
\]

\[
= \sum_b \sum_a aP(X = a \mid Y = b)P(Y = b)
\]

\[
= \sum_a \sum_b P(X = a, Y = b) = \sum_a aP(X = a) = EX.
\]

By the way, we also used the fact that we can sum positive numbers in any order. Generally interchanging the order of summation is a delicate matter, see Problem 1.1.3.

In the general case we have \( E|X| < \infty \) and

\[
E(X \mid Y = b) = \frac{1}{P(Y = b)} EXI_{Y = b} = \frac{1}{P(Y = b)} (EX + EXI_{Y = b} = EXI_{Y = b})
\]

where we used the fact that \( E(U - V) = EU - EV \), whenever \( E|U| < \infty \) and \( E|V| < \infty \), and the fact that \( |X \pm I_{Y = b}| \leq |X| \) implying that \( E|X \pm I_{Y = b}| \leq E|X| < \infty \). We conclude that

\[
U := E(X \mid Y) = E(X_+ \mid Y) - E(X_- \mid Y) =: V - W.
\]

Furthermore, by the above

\[
EE(X_\pm \mid Y) = EX_\pm < \infty.
\]
Therefore, \( EU = EV - EW \), which is rewritten as
\[
EE(X \mid Y) = EE(X_+ \mid Y) - EE(X_- \mid Y) = EX_+ - EX_- = EX,
\]
where the second equality is obtained from (4).

Kind of 1.2.5. Set \( a = EX^2 \), \( b = EXY \), \( c = EY^2 \) and first observe that \( a \) and \( c \) are finite by the assumption and \( E|XY| < \infty \) because \( |XY| \leq X^2 + Y^2 \). Then for real \( \lambda \) write
\[
0 \leq E(X + \lambda Y)^2 = E(X^2 + 2\lambda XY + \lambda^2 Y^2)
\]
We may write the last expectation of a sum as the sum of expectations due to the fact that the terms in the sum have finite expectations. Then we get
\[
0 \leq a + 2\lambda b + \lambda^2 c.
\]
Since this quadratic polynomial is nonnegative for any \( \lambda \), by an elementary fact, \( b^2 - ac \leq 0 \), which is exactly what we need.

Ordered grades for HW2:
36, 36, 36, 36, 36, 32, 30, 30, 20.

Ordered grades for HW1+HW2:
71, 71, 71, 66, 60, 59, 50, 49, 44.

### Solutions for Homework 3, due February 14

Problems assigned 2.1.7, 2.2.2, 2.2.3, 2.2.5

2.1.7. If \( P(X_1 = 0) < 1 \) and \( P(Y_1 = 0) < 1 \), then for some \( i, j = 0, 1 \) we have \( P(X_1 = (-1)^i) > 0 \), \( P(Y_1 = (-1)^j) > 0 \) and \( ((-1)^i, (-1)^j) \not\in \{0, \pm e_1, \pm e_2 \} \) and \( P(Z_1 - Z_0 = ((-1)^i, (-1)^j)) > 0 \). This is impossible because even without this probability the sum of probabilities that \( Z_1 - Z_0 \) took a value from the set \( \{0, \pm e_1, \pm e_2 \} \) is one. Thus either \( P(X_1 = 0) = 1 \) or \( P(Y_1 = 0) = 1 \). In the first case, since the increments \( Z_{n+1} - Z_n \) have the same distribution so do the increments \( X_{n+1} - X_n \) and since \( X_1 = X_1 - X_0 = 0 \) we have \( X_{n+1} - X_n = 0 \), \( X_{n+1} = X_n = \ldots = X_1 = 0 \). Similar argument works in the second case.

2.2.2. We are given that
\[
P(X_{n+1} = y, X_{[0,n]} = x_{[0,n]}) = p_n(x_n, y)P(X_n = x_n, X_{[0,n-1]} = x_{[0,n-1]}).
\]
By summing with respect to all values of \( x_{[0,n-1]} \), we get
\[
P(X_{n+1} = y, X_n = x_n) = p_n(x_n, y)P(X_n = x_n).
\]
Finally, if \( P(X_{[0,n]} = x_{[0,n]}) > 0 \), then also \( P(X_n = x_n) > 0 \) and
\[
P(X_{n+1} = y \mid X_n = x_n) = p_n(x_n, y),
\]
which after being compared with what is given yields
\[ P(X_{n+1} = y \mid X_{[0,n]} = x_{[0,n]}) = P(X_{n+1} = y \mid X_n = x_n) \]
as long as \( P(X_{[0,n]} = x_{[0,n]}) > 0 \). Thus \( X_n \) qualifies to be called a Markov chain.

2.2.3. We have
\[ P(Z_n = z_n, ..., Z_0 = z_0) = P(Z_n = z_n \mid Z_{n-1} = z_{n-1}, ..., Z_0 = z_0)P(Z_{n-1} = z_{n-1}, ..., Z_0 = z_0) \]
in both cases where \( P(Z_{n-1} = z_{n-1}, ..., Z_0 = z_0) \) is bigger than zero or not because in the latter case the probability we started with is zero. Next if \( P(Z_{n-1} = z_{n-1}) > 0 \), the last quantity is
\[ = p(z_n \mid z_{n-1})P(Z_{n-1} = z_{n-1}, ..., Z_0 = z_0) \]
by what is said before the problem. However if \( P(Z_{n-1} = z_{n-1}) = 0 \), then \( P(Z_{n-1} = z_{n-1}, ..., Z_0 = z_0) = 0 \). We see that always
\[ P(Z_n = z_n, ..., Z_0 = z_0) = p(z_n \mid z_{n-1})P(Z_{n-1} = z_{n-1}, ..., Z_0 = z_0). \]
By using induction we get the result.

2.2.5. By Problem 2.2.4
\[ P(Z_{[n+1,m]} = z_{[n+1,m]}, Z_{[0,n]} = z_{[0,n]}) = p_m(z_m \mid z_{m-1}) \cdot \cdots \cdot p_{n+1}(z_{n+1} \mid z_n)P(Z_{[0,n]} = z_{[0,n]}) \]
if \( P(Z_{[0,n]} = z_{[0,n]}) > 0 \). Yet the above equality is trivial if \( P(Z_{[0,n]} = z_{[0,n]}) = 0 \). Next, if \( \pi_n(z_n) > 0 \), by Problem 2.2.4
\[ p_m(z_m \mid z_{m-1}) \cdot \cdots \cdot p_{n+1}(z_{n+1} \mid z_n)P(Z_{[0,n]} = z_{[0,n]}) = P(Z_{[n+1,m]} = z_{[n+1,m]} \mid Z_n = z_n)P(Z_{[0,n]} = z_{[0,n]}). \]
Again if \( \pi_n(z_n) = 0 \), equality still trivially holds. Thus, in all cases
\[ P(Z_{[n+1,m]} = z_{[n+1,m]}, Z_{[0,n]} = z_{[0,n]}) = P(Z_{[n+1,m]} = z_{[n+1,m]} \mid Z_n = z_n)P(Z_{[0,n]} = z_{[0,n]}) \]
and it only remains to use that
\[ P(Z_{[0,n]} = z_{[0,n]}) = P(Z_{[0,n-1]} = z_{[0,n-1]} \mid Z_n = z_n)\pi_n(z_n). \]

Ordered grades for HW3:
36, 36, 36, 36, 36, 35, 34, 34.

Ordered grades for HW1+HW2+HW3:
107, 107, 106, 102, 95, 94, 85, 84, 44.
Solutions for Midterm February 28

Problems assigned 3.1.6, 3.1.7, 3.1.9.

3.1.6. We have

\[ E(\xi_{n+1}^{-1}I_{Y_{n+1}=b} \mid Y_{[0,n]} = b_{[0,n]}) = E(\xi_{n+1}^{-1}(b_{[0,n]}, b)I_{Y_{n+1}=b} \mid Y_{[0,n]} = b_{[0,n]}), \]

where \( \xi_{n+1}^{-1}(b_{[0,n]}, b) \) is a constant so that the last expression equals

\[ \xi_{n+1}^{-1}(b_{[0,n]}, b)E(I_{Y_{n+1}=b} \mid Y_{[0,n]} = b_{[0,n]}), \]

and the desired result follows.

3.1.7. For the transition function of \((X_n, Y_n)\) we have

\[ p((r,s), (u,v)) = \frac{1}{4} (I_{|r-u|=1, s=v} + I_{r=u, |s-v|=1}) = q^{ra}(s, v). \]

Furthermore, with probability one \(|Y_n - Y_{n+1}| = 1\) if and only if \(Y_{n+1} \neq Y_n\). Therefore,

\[ q^{ra}(Y_n, Y_{n+1}) = \frac{1}{4} (I_{|r-a|=1, Y_{n+1}=Y_n} + I_{r=a, Y_{n+1} \neq Y_n}) \]

and by Theorem 3.1.2

\[ \pi_{n+1}^a = \frac{1}{4\xi_{n+1}^a} \sum_{r \in \mathbb{Z}} \pi_n^r (I_{|r-a|=1, Y_{n+1}=Y_n} + I_{r=a, Y_{n+1} \neq Y_n}) \]

\[ = \frac{1}{4\xi_{n+1}^a} [(\pi_n^{a+1} + \pi_n^{a-1})I_{Y_{n+1}=Y_n} + \pi_n^a I_{Y_{n+1} \neq Y_n}]. \]  \( (5) \)

We will return to (5) after we find \(\xi_{n+1}^a\). According to Theorem 3.1.2

\[ \xi_{n+1} = \frac{1}{4} \sum_{r, s \in \mathbb{Z}} \pi_n^r (I_{|r-s|=1, Y_{n+1}=Y_n} + I_{r=s, Y_{n+1} \neq Y_n}). \]

Obviously, for each \(r \in \mathbb{Z}\)

\[ \sum_{s \in \mathbb{Z}} I_{|r-s|=1} = 2, \quad \sum_{s \in \mathbb{Z}} I_{r=s} = 1. \]

Therefore

\[ \xi_{n+1} = \frac{1}{2} I_{Y_{n+1}=Y_n} \sum_{r \in \mathbb{Z}} \pi_n^r + \frac{1}{4} I_{Y_{n+1} \neq Y_n} \sum_{r \in \mathbb{Z}} \pi_n^r = \frac{1}{2} I_{Y_{n+1}=Y_n} + \frac{1}{4} I_{Y_{n+1} \neq Y_n}, \]

(a.s.) since we know that

\[ \sum_{r \in \mathbb{Z}} \pi_n^r = 1 \]

(a.s.). This is one of the formulas we needed to derive.
Finally, if $Y_n = Y_{n+1}$, then $\xi_{n+1} = 1/2$ and by (5) we get

$$\pi_{n+1} = \frac{1}{2}(\pi_{n+1}^a + \pi_{n+1}^{a-1})$$

and if $Y_n \neq Y_{n+1}$, then $\xi_{n+1} = 1/4$ and $\pi_{n+1}^a = \pi_n^a$. This yields the desired equation for $\pi_{n+1}^a$.

3.1.9. (i) We have

$$E|X_n| \leq E|X_n - X_0| + E|X_0| \leq n + E|X_0| \leq \infty,$$

$EX_n^2 = E(X_n - X_0 + X_n)^2 \leq 2E(X_n - X_0)^2 + 2EX_0^2 \leq 2n^2 + 2EX_0^2 < \infty,$

so that $m_n$ and $d_n$ are well defined and finite (a.s.) and

$$m_n = \sum_{a \in \mathbb{Z}} a\pi_n^a, \quad d_n = \sum_{a \in \mathbb{Z}} a^2\pi_n^a$$

with the sums converging absolutely (a.s.).

Next by Problem 3.1.7

$$m_{n+1} = (1/2)\sum_{a \in \mathbb{Z}} ((a-1)1\pi_{n+1}^{a-1} + (a+1)1\pi_{n+1}^{a+1})I_{Y_n = Y_{n+1}}$$

$$m_n I_{Y_n \neq Y_{n+1}}.$$  

Since by Remark 1.3.25,

$$\sum_a \pi_n^a = 1$$

almost surely we have

$$m_{n+1} = (1/2)\sum_{a \in \mathbb{Z}} (a-1)1\pi_{n+1}^{a-1}I_{Y_n = Y_{n+1}} + (1/2)\sum_{a \in \mathbb{Z}} (a+1)1\pi_{n+1}^{a+1}I_{Y_n = Y_{n+1}}$$

$$+ m_n I_{Y_n \neq Y_{n+1}} = (1/2)m_n I_{Y_n = Y_{n+1}} + (1/2)m_n I_{Y_n = Y_{n+1}} + m_n I_{Y_n \neq Y_{n+1}} = m_n.$$  

It follows that $m_n = E(X_0 \mid Y_0)$.

Next,

$$d_{n+} = d_n I_{Y_n \neq Y_{n+1}}$$

$$(1/2)(\sum_a [(a-1)^2 + 2(a-1) + 1]1\pi_n^{a-1} + \sum_a (a+1)^2 - 2(a+1) + 1]1\pi_n^{a+1})I_{Y_n = Y_{n+1}}$$

$$= d_n I_{Y_n \neq Y_{n+1}} + (1/2)(2 + \sum_a (a-1)^2 \pi_n^{a-1} + \sum_a (a+1)^2 \pi_n^{a+1})I_{Y_n = Y_{n+1}}$$

$$= d_n I_{Y_n \neq Y_{n+1}} + (1/2)(2d_n + 2)I_{Y_n = Y_{n+1}}.$$  

Therefore, (a.s.)

$$d_{n+1} = d_n + I_{Y_n = Y_{n+1}}$$

and, for $n \geq 1$, $d_n$ is the sum of $E(X_0^2 \mid Y_0)$ and the number of $i = 0, 1, ..., n-1$ such that $Y_{i+1} = Y_i$. 
(ii) By Corollary 1.3.24
\[ I := E|X_2^2 - E(X_2^2 | Y_{[0,2]}^2)|^2 = EX_2^4 - E|I_{Y_0=0} + I_{Y_1-Y_0=0}|^2. \]

Upon finding the distribution of \(X_2\) or using \(X_2 = (X_2-X_1)+(X_1-X_0)\) and the fact that the increments are independent, one finds that the first term on the right is 2.5. In the second term the indicators are independent (since that increments of \(Y_n\) are independent) and \(Y_0 = 0\), so that
\[ E|I_{Y_0=1} + I_{Y_1-Y_0=0}|^2 = P(Y_1 = 0) + 2P(Y_1 = 0)P(Y_1 = Y_2) + P(Y_1 = Y_2) = 1/2 + 2/4 + 1/2 = 3/2, \quad I = 1. \]

Ordered grades for Midterm:
36, 36, 36, 36, 34, 34, 34, 33, 20.

Ordered grades for HW1+HW2+HW3+Midterm:
143, 140, 140, 138, 131, 128, 121, 118, 64.

Solutions for Homework 4, due March 7
Problems assigned: 2.2.4, 2.2.36. 4.1.6, 4.1.8

2.2.4. By Problem 2.2.3
\[ P(Z_{[n+1,m]} = z_{[n+1,m]}, Z_{[0,m]} = z_{[0,n]}) = [p_m(z_{m-1} | z_m) \cdot \ldots \cdot p_{n+1}(z_{n+1} | z_n)] \times \]
\[ \times [p_n(z_n | z_{n-1}) \cdot \ldots \cdot p_1(z_1 | z_0) \pi_0(z_0)] = p_m(z_{m-1} | z_m) \cdot \ldots \cdot p_{n+1}(z_{n+1} | z_n)P(Z_n = z_n, \ldots, Z_0 = z_0). \]

Upon dividing through by \(P(Z_n = z_n, \ldots, Z_0 = z_0)\) you get the first needed formula. For fixed \(z_{[n+1,m]}\) its right-hand side is independent of \(z_{[0,n]}\) and the second formula follows directly from Problem 1.3.5.

2.2.36. The entry \(p_{d+1}^{d-1}\) is the probability to be at time \(d-1\) at state \(j\) starting at time 0 from state \(i\). If \(i = 1, j = 1\), then this probability is not less than the probability to go from 1 to itself \(d-1\) times, which happens with probability \(2^{-(d-1)}\). If \(i = 1, j > 1\), then \(p_{1j}^{d-1}\) is not less than the probability to stay at 1 first \(d-j\) times and then go one unit at a time all the way to the right to state \(j\), for which you will need the remaining \(j-1\) steps.

Generally, if \(i + j \leq d + 1\), then with nonzero probability making \(i-1\) first steps to the left you will reach state 1, then you can stay at 1 for the next \(d + 1 - i - j\) steps and after that go straight to \(j\) in \(j-1\) steps. Altogether we need \(d-1\) steps and the whole trajectory
has probability $2^{-(d-1)}$, implying that in this case $p_{ij}^{d-1} \geq 2^{-(d-1)}$. If $i + j > d + 1$, the argument is similar, but you go first straight to state $d$ in $d - i$ steps, stay there for the next $i + j - d - 1$ (> 0) steps and then go down to state $j$ in the remaining $d - j$ steps. We see that $p_{ij}^{d-1} \geq 2^{(d-1)}$ for all $i, j$.

For the second part the answer is obviously 0.

4.1.6. Set $X = \alpha Y + \beta Z$. Then, obviously, $X \in \mathcal{L}$ and

$$X = [\alpha a_1(X) + \beta a_1(Y)]\ell_1 + \ldots + \alpha a_k(X) + \beta a_k(Y)\ell_k.$$  

We know that the representation of $X$ as a linear combination of $\ell_i$ is unique and the coefficient in this representation were called $a_i(X)$. The above formula gives a representation of $X$. By uniqueness of coefficients

$$a_i(\alpha Y + \beta Z) = \alpha a_i(Y) + \beta a_i(Z).$$

4.1.8. Was solved in class.

Ordered grades for HW4:
36, 36, 36, 34, 34, 33, 32, 30, 7.

Ordered grades for HW1+...+HW4+MidT:
179, 176, 174, 174, 163, 158, 155, 151, 71.

Solutions for Homework 5, due March 28

Problems assigned 3.3.1, 3.3.3, 3.3.17, 4.1.10

3.3.1. The first conditional probability is the ratio of

$$p_1 = P(X_2 = 2, Y_0 = 1, Y_1 = 2, Y_2 = 1)$$

and

$$p_2 = P(Y_0 = 1, Y_1 = 2, Y_2 = 1).$$

Observe that

$$\{Y_0 = 1, Y_1 = 2, Y_2 = 1, X_2 = 2\}$$

$$= \{W_0 = 1, X_1 = 1, W_1 = 1, X_2 = 2, W_2 = -1\}$$

$$= \{W_0 = 1, W_1 = 1, W_2 = -1\} \cap \{X_1 = 1, X_2 = 2\},$$

where the event is the intersection of two independent ones. Therefore,

$$p_1 = P(W_0 = 1, W_1 = 1, W_2 = -1)P(X_1 = 1, X_2 = 2) = 2^{-5}.$$  

Next,

$$\{Y_0 = 1, Y_1 = 2, Y_2 = 1\} = \{W_1 = 0, X_1 = 1, W_1 = 1, X_2 = 2, W_2 = -1\}$$

$$\cup \{W_1 = 0, X_1 = 1, W_1 = 1, X_2 = 0, W_2 = 1\}.$$
It follows that
\[ p_2 = 2 \cdot 2^{-5}, \quad \frac{p_1}{p_2} = \frac{1}{2}. \]

Concerning the second conditional probability in question, notice that
\[ \{Y_0 = 1, Y_1 = 2, Y_2 = 1, X_2 = 2, Y_3 = 4\} \]
\[ = \{W_0 = 1, X_1 = 1, W_1 = 1, X_2 = 2, W_2 = -1, X_3 = 3, W_3 = 1\}. \]
In addition,
\[ \{Y_0 = 1, Y_1 = 2, Y_2 = 1, Y_3 = 4\} \]
\[ = \{W_0 = 1, X_1 = 1, W_1 = 1, X_2 = 2, W_2 = -1, X_3 = 3, W_3 = 1\}, \]
so that the second conditional probability is one.

3.3.3. Notice that
\[ p((a_n, b_n), (a_{n+1}, b_{n+1}))P(X_n = a_n, Y_n = b_n) \]
\[ = P(X_{n+1} = a_{n+1}, Y_{n+1} = b_{n+1} | X_n = a_n, Y_n = b_n)P(X_n = a_n, Y_n = b_n) \]
\[ = P(X_{[n,n+1]} = a_{[n,n+1]}, Y_{[n,n+1]} = b_{[n,n+1]}). \]
Similarly,
\[ p((a_{n+1}, b_{n+1}), (a_{n+2}, b_{n+2}))p((a_n, b_n), (a_{n+1}, b_{n+1}))P(X_n = a_n, Y_n = b_n) \]
\[ = p((a_{n+1}, b_{n+1}), (a_{n+2}, b_{n+2}))P(X_{[n,n+2]} = a_{[n,n+2]}, Y_{[n,n+2]} = b_{[n,n+2]}), \ldots \]
\[ P(X_n = a_n, Y_n = b_n)p((a_n, b_n), (a_{n+1}, b_{n+1}))p((a_{n+1}, b_{n+1}), (a_{n+2}, b_{n+2})) \times \ldots \]
\[ \times p((a_{m-1}, b_{m-1}), (a_m, b_m)) = P(X_{[n,m]} = a_{[n,m]}, Y_{[n,m]} = b_{[n,m]}). \]
Now it is elementary that formula (3.3.4) will be proved if we prove the representation for \( \kappa_{nm}^a \) from the hint.

We will proceed by backward induction on \( n \). Notice that we need the representation only for \( n < m \) (see the statement of the problem). Thus, \( m \geq 1 \). Next we observe that the representation holds for \( n = m - 1 \). Indeed, by definition (3.3.5)
\[ \kappa_{n,m}^a(b_{[n,m]}) = \kappa_{n-1,m}^a(b_{[n-1,m]}) = \sum_{a_m \in A} q_{a_{m-1}a_m}^m(b_{m-1}, b_m) \kappa_{nm}^r(b_{[m,m]}) \]
\[ = \sum_{a_m \in A} q_{a_{m-1}a_m}^m(b_{m-1}, b_m) = \sum_{a_m \in A} p((a_{m-1}, b_{m-1}), (a_m, b_m)). \]

Finally, assume that we have the representation for \( \kappa_{nm}^a \) with an \( n < m \) and \( n \geq 1 \) and we want to get the representation for \( \kappa_{n-1,m}^a \). Then it suffices to use again definition (3.3.5) and the induction hypothesis allowing one to replace \( \kappa_{nm}^a \) with an appropriate sum.
3.3.17. I repeat what was done in class. Notice that for \( i \leq n \)

\[
\pi_{in}^a = P(X_i = a \mid Y_{[0,n]}) = \sum_r \pi_{in}^{ar},
\]

where

\[
\pi_{in}^{ar} = P(X_i = a, X_n = r \mid Y_{[0,n]}).
\]

By Problem 3.3.9

\[
\pi_{i,n+1}^{ar} = \frac{1}{\xi_{n+1}} \sum_s \pi_{in}^{as} q^{sr}(Y_n, Y_{n+1}).
\]

Hence,

\[
g_{n+1}^{ar} := \sum_{i \leq n+1} \pi_{i,n+1}^{ar} = \pi_{n+1,n+1}^{ar} + \sum_{i \leq n} \pi_{i,n+1}^{ar}
\]

\[
= \pi_{n+1,n+1}^{ar} + \frac{1}{\xi_{n+1}} \sum_s \left( \sum_{i \leq n} \pi_{i,n}^{as} \right) q^{sr}(Y_n, Y_{n+1})
\]

\[
= \pi_{n+1,n+1}^{ar} + \frac{1}{\xi_{n+1}} \sum_s g_{n}^{as} q^{sr}(Y_n, Y_{n+1}).
\]

(6)

Observe that

\[
\pi_{n+1,n+1}^{ar} = P(X_{n+1} = a, X_{n+1} = r \mid Y_{[0,n+1]})
\]

\[
= \delta^{ar} P(X_{n+1} = a \mid Y_{[0,n+1]}) = \delta^{ar} \pi_{n+1}^a.
\]

Since with probability one we have \( \xi_k > 0 \), we can multiply the extreme terms in (6) by \( \xi_0 \cdot \ldots \cdot \xi_{n+1} \) and cancel \( \xi_{n+1} \) on the right. Then we remember the definition of \( \rho_k^a \) and see that

\[
\hat{g}_{n+1}^{ar} = \xi_0 \cdot \ldots \cdot \xi_n g_n^{ar}
\]

satisfies

\[
\hat{g}_{n+1}^{ar} = \delta^{ar} \rho_n^a + \sum_{s \in A} \hat{g}_{n}^{as} q^{sr}.
\]

Furthermore, \( \hat{g}_0^{sr} = \delta^{sr} \pi_0^s \). We conclude that \( \hat{g}_n^{as} \) and \( \bar{g}_n^{as} \) satisfy the same recursive equation and since they coincide for \( n = 0 \), by induction, they coincide for all \( n \).

Now we forget about the equations for \( \hat{g}_n^{as} \) and \( \bar{g}_n^{as} \) and observe that

\[
g_n^a = \sum_{i=0}^n \pi_{in}^a = \sum_{r \in A} \sum_{i=0}^n \pi_{in}^{ar} = \sum_{r \in A} g_n^{ar},
\]

\[
g_n^a = \frac{1}{\xi_0 \cdot \ldots \cdot \xi_{n+1}} \sum_{r \in A} \hat{g}_n^{ar}.
\]
We sum up both parts of this equality with respect to \( a \in A \) and notice that
\[
\sum_{a \in A} g_n^a = \sum_{a \in A} E\left( \sum_{i=0}^{n} I_{X_i=a} \mid Y_{[0,n]} \right) = \sum_{a \in A} E\left( \sum_{i=0}^{n} I_{X_i=a} \mid Y_{[0,n]} \right)
\]
\[
= \sum_{i=0}^{n} E(1 \mid Y_{[0,n]}) = n + 1.
\]
Therefore,
\[
\frac{1}{\xi_0 \cdot \ldots \cdot \xi_{n+1}} \sum_{a,r \in A} \bar{g}_{ar} = n + 1,
\]
and we are done.

4.1.10. Introduce
\[
Z = a\Pi\mathcal{L}X + b\Pi\mathcal{L}Y.
\]
Then \( Z \in \mathcal{L} \) and for any \( W \in \mathcal{L} \) we have
\[
\langle W, aX + bY - Z \rangle = a\langle W, X - \Pi\mathcal{L}X \rangle + b\langle W, Y - \Pi\mathcal{L}Y \rangle = 0.
\]
Hence by definition \( Z = \Pi\mathcal{L}(aX + bY) \).

Ordered grades for HW5:
36, 36, 36, 36, 36, 36, 36, 34.

Ordered grades for HW1+...+HW5:
179, 178, 177, 174, 163, 160, 155, 151, 51.

**Solutions for Homework 6, due April 11**

Problems assigned 4.1.12, 4.1.16, 4.1.19, 4.1.24

4.1.12(ii). We know that \( \Pi\mathcal{L}X \) exists and
\[
\Pi\mathcal{L}X = b_1Y_1 + \ldots + b_mY_m
\]
for some constants \( b_i \) since \( \Pi\mathcal{L}X \in \text{Span}(Y_{[1,k]}) \). By definition \( \langle X - \Pi\mathcal{L}X, Y_i \rangle = 0 \), \( \langle \Pi\mathcal{L}X, Y_i \rangle = \langle X, Y_i \rangle \) for \( i = 1, \ldots, m \), and since
\[
\langle \Pi\mathcal{L}X, Y_i \rangle = b_1\langle Y_1, Y_i \rangle + \ldots + b_m\langle Y_m, Y_i \rangle,
\]
we get all the system, which by this argument has at least one solution. On the other hand, if we take any solution \((b_1, \ldots, b_m)\) of the system, then \( Z := b_1Y_1 + \ldots + b_mY_m \in \mathcal{L} \) and the system tells us that \( Z \perp Y_i \) for all \( i \). Then \( Z \) is perpendicular to any linear combination of \( Y_i \) the set of which is \( \mathcal{L} \), so that \( Z \perp \mathcal{L} \), which proves that indeed \( \Pi\mathcal{L}X = b_1Y_1 + \ldots + b_mY_m \) where \((b_1, \ldots, b_m)\) is any solution of the system.
4.1.19. To show that the limit exists it suffices to show that

\[ S_n = \sum_{k=1}^{n} a_n X_n \]

form a Cauchy sequence, that is

\[ \|S_m - S_n\| \to 0 \]

as \( n, m \to \infty \). However, by (4.1.5) for \( m \geq n \)

\[ \|S_m - S_n\|^2 = \| \sum_{r=n+1}^{m} a_r X_r \|^2 = \sum_{r=n+1}^{m} |a_r|^2 \|X_r\|^2 \]

\[ \leq \sum_{r=n+1}^{m} |a_r|^2 \leq \sum_{r=n+1}^{\infty} |a_r|^2 = \sum_{r=1}^{\infty} |a_r|^2 - \sum_{r=1}^{n} |a_r|^2 \to 0 \]

since

\[ \lim_{m \to \infty} \sum_{r=1}^{n} |a_r|^2 = \sum_{r=1}^{\infty} |a_r|^2. \]

4.1.16 (i) You just treat 1 as a random variable, say \( Y_{k+1} \).
(ii) First we show that \( V = \text{EX}_2 \) equals \( \Pi_{\text{Span} \{1,Y_1,Y_2\}} \text{X}_2^2 \). For that it suffices to show that \( V \in \text{Span} \{1,Y_1,Y_2\} \) and \( X_2^2 - V \perp \text{Span} \{1,Y_1,Y_2\} \).

That the constant \( V \) can be represented as a linear combinations of \( Y_1,Y_2 \) is obvious. Next,

\[ \langle X_2^2 - V, 1 \rangle = E(X_2^2 - V) = 0, \]
\[ \langle X_2^2 - V, Y_1 \rangle = E(X_2^2 Y_1) - VEY_1, \]
\[ \langle X_2^2 - V, Y_2 \rangle = E(X_2^2 Y_2) - VEY_2. \]

Here \( EY_1 = 0 \), \( EY_2 = E(Y_2 - Y_1) + EY_1 = 0 \),

\[ E(X_2^2 Y_1) = E(X_1^2 Y_1) + 2E(X_2 - X_1)X_1 Y_1 + E(X_2 - X_1)^2 Y_1. \]

Since \( X_1 Y_1 = 0 \) and by the independence of increments \( E(X_2 - X_1)^2 Y_1 = (E(X_2 - X_1)^2)EY_1 = 0 \), we have from the above that \( X_2^2 - V \perp 1, Y_1 \).

Finally,

\[ E(X_2^2 Y_2) = E(X_2^2 (Y_2 - Y_1)) = EX_1^2 (Y_2 - Y_1) + 2EX_1 (X_2 - X_1)(Y_2 - Y_1) \]
\[ + E(X_2 - X_1)^2 (Y_2 - Y_1), \]

where \( (X_2 - X_1)(Y_2 - Y_1) = 0 \) and \( EX_1^2 (Y_2 - Y_1) = (EX_1^2)E(Y_2 - Y_1) = 0 \), so that \( X_2^2 - V \perp Y_2 \) and we are done.

4.1.24. In the notation from the hint

\[ Y_{k+i} \in \text{Span} \{Y_{[1,k+n]} \}, \quad \bar{Y}_{k+i} \in \text{Span} \{Y_{[1,k]} \} \subset \text{Span} \{Y_{[1,k+n]} \}. \]
It follows that
\[ Y_{k+i}, \bar{Y}_{k+i} \in \text{Span} (Y_{[1,k+n]}), \]
so that \( Z \in \text{Span} (Y_{[1,k+n]}) \). Then the only thing to check is that
\( X - Z \perp \text{Span} (Y_{[1,k+n]}) \), which by Problem 4.1.23 happens iff \( X - Z \perp Y_j \) for all \( j = 1, \ldots, k + n \), that is
\[ E(X - Z)Y_j = 0 \]
for all \( j = 1, \ldots, k + n \).

By the definition of \( Z \) and \( \bar{Y}_{k+i} \) we have that
\[ Z = b_1(Y_{k+1} - \bar{Y}_{k+1}) + \ldots + b_n(Y_{k+n} - \bar{Y}_{k+n}) \]
for some constants \( b_i \), and \( E(Y_{k+1} - \bar{Y}_{k+1})Y_j = 0 \) for \( j \leq k \). It follows that \( EYZ_j = 0 \) for \( j \leq k \) and since by assumption \( EYZ_j = 0 \) for \( j \leq k \), we have (7) for \( j \leq k \).

In particular, since \( \bar{Y}_{k+i} \in \text{Span} (Y_{[1,k]}) \) we have that \( E(X - Z)\bar{Y}_{k+i} = 0 \). Now
\[ \langle X - Z, Y_{k+i} \rangle = \langle X - Z, Y_{k+i} - \bar{Y}_{k+i} \rangle = 0, \]
where the last equality follows by the definition of \( Z \). Thus (7) holds for all \( j \) and we are done.