1B Discrete random variables

First familiriaze yourself with the notions "random variable" and "distribution", e.g. from lecture 1B slides and from Ross.

Class problems

1B1 (Two wolves) On a farm there are four ducks, four geese and two hens. During the night two wolves arrive. Each wolf catches one bird at random. Let us denote

$$\begin{aligned} X &= \text{the number of ducks caught,} \\ Y &= \text{the number of geese caught.} \end{aligned}$$

(a) Determine the distribution of the random variable X, by creating a table of its possible values and their probabilities.

(b) Determine the distribution of Y.

(c) Determine the *joint distribution* of the random variables X and Y by creating a 3×3 table, with cells containing the probabilities of values of the pair (X, Y).

(d) Calculate the row sums and column sums of the table in (c). Compare them to the tables in (a) and (b).

(e) Using the table in (c), calculate the probability for the event that the number of ducks caught equals the number of geese caught.

(f) Using the table in (c), determine whether the random variables X and Y are stochastically dependent or independent.

Solution.

(a) The number of ducks caught X can have one of the values 0,1,2. There are at least two ways for determining the distribution. Both methods give the same result:

k	0	1	2
$\mathrm{P}(X = k)$	$\frac{15}{45}$	$\frac{24}{45}$	$\frac{6}{45}$

One can also simplify the fractions. On the other hand, it is easier to compare and add them if they have a common denominator (divisor).

Method 1: Two-bird subsets. A *subset* of 2 birds is chosen at random from the set of the 10 birds on the farm. (Let us mentally earmark the birds with 10 different names so we can identify the birds.) The number of different 2-bird subsets is

$$\binom{10}{2} = 45.$$

(i) Let us find out how many different 2-bird subsets there are that contain *zero* ducks. Such a subset contains two birds from the six *nonducks*, so there are

$$\binom{6}{2} = 15$$

such subsets. Thus $\mathrm{P}(X = 0) = \binom{6}{2}/\binom{10}{2} = \frac{15}{45} = \frac{1}{3}$.

(ii) Let us find out how many 2-bird subsets there are that contain exactly *one* duck. Such a subset contains *one* duck out of the four ducks, and *one* nonduck out the six nonducks. By the multiplication rule, the number of choices is

$$\binom{4}{1} \times \binom{6}{1} \;=\; 4 \times 6 \;=\; 24$$

Thus $\mathrm{P}(X = 1) = \frac{24}{45} = \frac{8}{15}$.

(iii) Let us find out how many 2-bird subsets there are that contain *two* ducks. Such a subset contains two ducks our of the four ducks on the farm, so the number of such subsets is

$$\binom{4}{2} \;=\; 6$$

Thus $\mathrm{P}(X = 2) = \frac{6}{45} = \frac{2}{15}$. (Alternatively, by the additivity of probability, you could calculate $\mathrm{P}(X = 2) = 1 - \mathrm{P}(X = 0) - \mathrm{P}(X = 1)$, because the three probabilities must add up to one.)

Method 2: One bird at a time. Let us mentally earmark the wolves by numbers 1 and 2, in the order that they catch the birds. Consider

$$\begin{aligned} I_1 &= \text{number of ducks caught by wolf 1,} \\ I_2 &= \text{number of ducks caught by wolf 2.} \end{aligned}$$

For zero ducks to be caught, we need the *first* wolf to catch a nonduck (which happens with probability $6/10$). Then there are 9 birds left, of which 5 are nonducks. So the *second* wolf has a $5/9$ chance of catching a nonduck. Note that this is a conditional probability. By the chain rule (e.g. Ross eq. 3.6.2),

$$\begin{aligned} \mathrm{P}(X = 0) &= \mathrm{P}(I_1 = 0, I_2 = 0) \\ &= \mathrm{P}(I_1 = 0)\,\mathrm{P}(I_2 = 0 \mid I_1 = 0) \\ &= \frac{6}{10} \times \frac{5}{9} \\ &= \frac{30}{90}. \end{aligned}$$

For exactly one duck to be caught, we need to consider two (mutually exclusive) possibilities: either the first wolf catches a duck and the second does not, or vice versa. Thus

$$\begin{aligned}
\mathrm{P}(X=1) &= \mathrm{P}(I_1=0, I_2=1) + \mathrm{P}(I_1=1, I_2=0) \\
&= \mathrm{P}(I_1=0)\,\mathrm{P}(I_2=1 \mid I_1=0) \;+\; \mathrm{P}(I_1=1)\,\mathrm{P}(I_2=0 \mid I_1=1) \\
&= \frac{6}{10} \times \frac{4}{9} \;+\; \frac{4}{10} \times \frac{6}{9} \\
&= \frac{48}{90}.
\end{aligned}$$

Finally, for two ducks to be caught, the first wolf must catch a duck (probability $4/10$), and then the second wolf must catch a duck (probability $3/9$, because there are only three ducks left, and nine birds in total). Thus

$$\begin{aligned}
\mathrm{P}(X=2) &= \mathrm{P}(I_1=1, I_2=1) \\
&= \mathrm{P}(I_1=1)\,\mathrm{P}(I_2=1 \mid I_1=1) \\
&= \frac{4}{10} \times \frac{3}{9} \\
&= \frac{12}{90}.
\end{aligned}$$

Observe that the three probabilities add up to 1, as required: $\frac{30}{90} + \frac{48}{30} + \frac{12}{90} = \frac{90}{90}$.

(b) By symmetry, it is clear that the number of geese caught Y has the same distribution as X, so it is

k	0	1	2
$\mathrm{P}(Y=k)$	$\frac{15}{45}$	$\frac{24}{45}$	$\frac{6}{45}$

Observe that although X and Y have the same *distribution*, they are not the *same* random variable. It is quite possible that X and Y have different values. "Same distribution" means they are equally probable to have a given value, for example, the value 0.

(c) The possible values of (X,Y) are $(0,0), (0,1), (0,2), (1,1), (1,0), (2,0)$. (Note that $X+Y$ must be something between zero and two, because two birds are caught in total.) Let us calculate the probabilities of each possible pair, using the subset method (see above).

Event $(X,Y) = (0,0)$ means that the two hens are caught. There is just one such possibility, out of the 45 different two-bird subsets, so

$$\mathrm{P}(X=0, Y=0) \;=\; \frac{\binom{2}{2}}{\binom{10}{2}} \;=\; \frac{1}{45}.$$

Event $(X,Y) = (0,1)$ means 1 goose and 1 hen are caught (and no ducks), so

$$\mathrm{P}(X=0, Y=1) \;=\; \frac{\binom{4}{1}\binom{2}{1}}{\binom{10}{2}} \;=\; \frac{8}{45}.$$

Event $(X, Y) = (0, 2)$ means 2 geese are caught (no ducks, no hens), so

$$\mathrm{P}(X = 0, Y = 2) \;=\; \frac{\binom{4}{2}}{\binom{10}{2}} \;=\; \frac{6}{45}.$$

Event $(X, Y) = (1, 1)$ means 1 duck and 1 goose are caught (and no hens), so

$$\mathrm{P}(X = 1, Y = 1) \;=\; \frac{\binom{4}{1}\binom{4}{1}}{\binom{10}{2}} \;=\; \frac{16}{45}.$$

By symmetry, the probabilities for (X, Y) being $(1, 0)$ and $(0, 1)$ are equal. Likewise, the probabilities for $(2, 0)$ ja $(0, 2)$ are equal. Further observe that $X + Y$ cannot be greater than two (since only two birds are caught), so the probabilities for $(1, 2), (2, 1), (2, 2)$ are zeros. Now we can fill in the table for the joint distribution of X and Y:

		Y		
		0	1	2
	0	$\frac{1}{45}$	$\frac{8}{45}$	$\frac{6}{45}$
X	1	$\frac{8}{45}$	$\frac{16}{45}$	0
	2	$\frac{6}{45}$	0	0

(d) Augmenting the table with row and column sums:

		Y			
		0	1	2	sum
	0	$\frac{1}{45}$	$\frac{8}{45}$	$\frac{6}{45}$	$\frac{15}{45}$
X	1	$\frac{8}{45}$	$\frac{16}{45}$	0	$\frac{24}{45}$
	2	$\frac{6}{45}$	0	0	$\frac{6}{45}$
	sum	$\frac{15}{45}$	$\frac{24}{45}$	$\frac{6}{45}$	1

Observe that the row sums contain the distribution of X. Because these are on the "margin" of the table, this is also called the "marginal distribution" of X. Similarly, the column sums contain the "marginal" distribution of Y.

(e) In the joint distribution table, let us color red the cases that correspond to the event $\{X = Y\}$:

		Y		
		0	1	2
	0	$\frac{1}{45}$	$\frac{8}{45}$	$\frac{6}{45}$
X	1	$\frac{8}{45}$	$\frac{16}{45}$	0
	2	$\frac{6}{45}$	0	0

Adding them up we get $\mathrm{P}(X = Y) = 17/45$.

(f) The random variables X and Y are stochastically dependent. From the joint distribution table we can observe (for example)

$$P(Y = 2 \mid X = 2) \;=\; \frac{P((X,Y) = (2,2))}{P(X = 2)} \;=\; \frac{0}{6/45} \;=\; 0,$$

while

$$P(Y = 2) \;=\; \frac{6}{45}.$$

1B2 (Unknown die.) In a box there are four dice: one 4-sided, one 6-sided, and two 8-sided dice. Each die has its sides numbered as usual, with integers starting from 1. Your friend picks one die from the box randomly, and rolls that die once, without showing it to you. Let us denote

$$\begin{aligned} S &= \text{number of sides on the chosen die,} \\ T &= \text{result of the roll.} \end{aligned}$$

(a) Determine the distribution of S by creating a table with its possible values and their probabilities.

(b) Determine the joint distribution of S and T by creating a 3×8 table of the probabilities for the values of the pair (S, T).

(c) From the table in (b), determine the (marginal) distribution of T.

(d) Your friend tells you that the result of the roll was 3. Determine the *conditional distribution* of S using this information. That is, for each integer i that is a possible value for S, determine the conditional probability $P(S = i \mid T = 3)$.

Solution.

(a) Each of the four dice is equally probable (probablity 1/4 each). But $S = 8$ can happen in two ways, so $P(S = 8) = \frac{1}{4} + \frac{1}{4} = 1/2$. So the distribution of S is

i	4	6	8
$P(S = i)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

(b) S can have values in $\{4, 6, 8\}$, and T can have values in $\{1, 2, \ldots, 8\}$. So the possible values of the pair (S, T) are in

$$\{(i, j) : i = 4, 6, 8,\; j = 1, 2, \ldots, 8\}.$$

The distribution of (S, T) can be presented as a table with rows $4, 6, 8$ and columns $1, 2, \ldots, 8$. Let us determine the first row first.

If $S = 4$, then the rolling result T is uniformly distributed in $\{1, 2, 3, 4\}$, thus

$$\mathrm{P}(T = j \mid S = 4) \;=\; \frac{1}{4}$$

for $j = 1, \ldots, 4$. Since $\mathrm{P}(S = 4) = \frac{1}{4}$, and applying the chain rule, we get

$$\mathrm{P}(S = 4, T = j) \;=\; \mathrm{P}(S = 4)\,\mathrm{P}(T = j \mid S = 4) \;=\; \frac{1}{16},$$

when $j \in \{1, 2, 3, 4\}$. For other values of j we have $\mathrm{P}(S = 4, T = j) = 0$ because the 4-sided die does not give results bigger than 4. This gives the first row of the table.

The other two rows are filled in the same fashion.

		T 1	2	3	4	5	6	7	8
	4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	0	0	0	0
S	6	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	0	0
	8	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

(c) Augmenting the table with row and column sums,

		T 1	2	3	4	5	6	7	8	sum
	4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	0	0	0	0	$\frac{1}{4}$
S	6	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	0	0	$\frac{1}{4}$
	8	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{2}$
	sum	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{5}{48}$	$\frac{5}{48}$	$\frac{1}{16}$	$\frac{1}{16}$	

(We did not really need the row sums here.) From the column sums we get the distribution of T:

j	1	2	3	4	5	6	7	8
$\mathrm{P}(T = j)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{5}{48}$	$\frac{5}{48}$	$\frac{1}{16}$	$\frac{1}{16}$

(d) From item (c) we know that $\mathrm{P}(T = 3) = \frac{1}{6}$. The required conditional probability is

$$\mathrm{P}(S = i \mid T = 3) \;=\; \frac{\mathrm{P}(S = i, T = 3)}{\mathrm{P}(T = 3)}$$

for $i \in \{4, 6, 8\}$. The joint probabilities can be read from the joint distribution table (here in red)).

$S \backslash T$	1	2	3	4	5	6	7	8	sum
4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	0	0	0	0	$\frac{1}{4}$
6	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	0	0	$\frac{1}{4}$
8	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{2}$
sum	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{5}{48}$	$\frac{5}{48}$	$\frac{1}{16}$	$\frac{1}{16}$	

Dividing the joint probabilities by $\mathrm{P}(T = 3) = \frac{1}{6}$ we get the conditional probabilities:

i	4	6	8
$\mathrm{P}(S = i \mid T = 3)$	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{3}{8}$

Home problems

1B3 (Family planning) A couple decides to have children until at least one of the following conditions is fulfilled:

- They have two girls.
- They have four children in total.

Children are born one at a time, so that each child is a girl or a boy with equal probabilities, independent of the previous children. Let us inspect the situation after the couple has stopped having children (so at least one of the above conditions has been fulfilled). Let the total number of children be $C = G + B$, where

$$\begin{aligned} G &= \text{the number of girls,} \\ B &= \text{the number of boys.} \end{aligned}$$

(a) Determine the distribution of the random variable C. Hint: Consider the possible child sequences in birth order, such as BBG (two boys and then a girl). Think how the sequences are generated, starting from the first child. Determine the probabilities of the sequences, and stop when appropriate.

(b) Determine the joint distribution of G and B.

(c) With what probability does the family have more girls than boys?

(d) Determine the joint distribution of C and G.

Grading. 0.5 points for each item (a)–(d), then rounded up to an integer.

Solution.

(a) First we observe that C can take only values 2, 3 and 4. (It cannot be 1, because neither condition can be fulfilled with one child only. It cannot exceed 4, because after four children no more are born.)

Value $C = 2$ occurs only with two girls (sequence GG), that is, with probability $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Value $C = 3$ occurs if the third child is a girl, and she is the second girl in the family; that is, with child sequences GBG and BGG, with total probability $(1/2)^3 + (1/2)^3 = 1/4$. Value $C = 4$ occurs if the first three children include 0 or 1 girls, that is, sequences BBB, BBG, BGB, GBB, with total probability $4 \times (1/2)^3 = 1/2$. (Then the gender of fourth child does not matter; the second condition is fulfilled in either case.)

So we have obtained the distribution of C as

j	2	3	4
$\mathrm{P}(C = j)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

Observe that the probabilities of 2,3,4 add up to one, as they should because these are all possible values of C.

(b) In the previous item, we already listed all possible child sequences up to three children. Add the fourth child in the case of four children. So the sequences are:

- GG (two children, $G = 2$ and $B = 0$); happens with probability $\frac{1}{4}$
- GBG and BGG (three children, $G = 2$ and $B = 1$); happens with probability $\frac{1}{8} + \frac{1}{8} = \frac{1}{4}$
- BBBB, BBBG, BBGB, BBGG, BGBB, BGBG, GBBB, GBBG (four children). Each of these eight sequences has probability $1/16$. Now one of the sequences has *no* girls $(B = 4, G = 0)$ so this happens with probability $1/16$. Four sequences have *one* girl $(B = 3, G = 1)$ so this happens with probability $4 \times (1/16) = 1/4$. Three sequences have *two* girls $(B = 2, G = 2)$, total probability $3 \times (1/16) = 3/16$.

With careful thinking, we can see that there are no other possibilities for the pair (B, G). For example, if the girl count is not two, then the total child count must be four; also, no matter what happens, there will not be more than four children, so $B + G \leq 4$. Impossible events have probability zero. Collecting our observations as a table,

		B					
		0	1	2	3	4	Sum
	0	0	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$
G	1	0	0	0	$\frac{1}{4}$	0	$\frac{1}{4}$
	2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	0	0	$\frac{11}{16}$
	Sum	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{16}$	1

After populating the joint distribution table, it is a good idea to calculate both marginals and the grand total from the table; if this is not 1, something is wrong.

(c) In the joint distribution, color the cases where $G > B$.

		B=0	1	2	3	4	Sum
G	0	0	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$
	1	0	0	0	$\frac{1}{4}$	0	$\frac{1}{4}$
	2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	0	0	$\frac{11}{16}$
	Sum	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{16}$	1

Adding up the colored probabilities, we observe that there are more girls than boys with probability 50%.

(d) Let us use the table in item (b). Whenever the pair (G, B) has a certain value, we also know the value of C (because it is $G + B$). Collect every possible event from the (G, B) table and add its probability to the corresponding place in the (G, C) table, observing that $C = G + B$, and also that one cell in the (G, C) may receive more than one addition.

		C=0	1	2	3	4	Sum
G	0	0	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$
	1	0	0	0	0	$\frac{1}{4}$	$\frac{1}{4}$
	2	0	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{11}{16}$
	Sum	0	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	1

The bottom marginal contains the column sums, which are the distribution of C. We observe it matches what we calculated in (a). If it didn't, something would be wrong.

1B4 (Extreme values) An ordinary six-sided die is rolled five times. Let the results be X_1, X_2, X_3, X_4, X_5, in that order. Let the greatest result be G. For example, if the results are $1, 4, 3, 2, 2$, then $G = 4$.

(a) What are the probabilities for the events $\{G \leq k\}$, for each of the integers $k = 1, 2, 3, 4, 5, 6$?

It would be possible to list all $6^5 = 7776$ possible result sequences, and collect the cases where, say, $G \leq 3$, but we really do not want to do that by hand. Instead, use this hint. What is the condition that each individual roll has to fulfill so that $G \leq 3$? What is the probability that this condition is fulfilled (for all five rolls)? Do the same for each k.

(b) Using the results from the previous item, and the additive rule of probability, determine the probabilities for the events $\{G = k\}$, for all $k = 1, 2, 3, 4, 5, 6$. Now present the distribution of G as a table. Explain the rough shape of the distribution in words (or use a picture).

(c) Show that G and X_1 are *dependent* (i.e. not independent), e.g. by showing that equation (4.3.8) [Ross section 4.3.1] is not fulfilled.

Hint: Consider, for example, the events $\{G = 1\}$ and $\{X_1 = 1\}$.

Grading. 0.5 points for each item (a)–(c), then rounded up to an integer.

Solution.

(a) Following the hint, we observe that the event "maximum result is at most 3" is exactly the same event as "all five results are at most 3", that is, $\{X_1 \leq 3 \textbf{ and } \ldots \textbf{ and } X_5 \leq 3\}$. Each roll has a $3/6 = 1/2$ probability of being at most three, so (by independence) $\mathrm{P}(G \leq 3) = (3/6)^5 = (1/2)^5 = 1/32$.
Similarly, for any k, we have $G \leq k$ exactly if all five rolls are at most k, with probability $(k/6)^5$. So we have the probabilities

k	1	2	3	4	5	6
$\mathrm{P}(G \leq k)$	$\frac{1}{7776}$	$\frac{32}{7776}$	$\frac{243}{7776}$	$\frac{1024}{7776}$	$\frac{3125}{7776}$	$\frac{7776}{7776}$

This is a table of the CDF of G. Note that these are *not* the densities.

(b) Case $k = 1$ is easy: The events $\{G = 1\}$ ja $\{G \leq 1\}$ are the same, so $\mathrm{P}(G = 1) = \mathrm{P}(G \leq 1) = 1/7776$.
Case $k = 2$: Following the hint, we observe that $\mathrm{P}(G \leq 2) = \mathrm{P}(G \leq 1) + \mathrm{P}(G = 2)$, because there are two ways for the integer G to be at most two: either it is at most one, or it equals two. Thus $\mathrm{P}(G = 2) = \mathrm{P}(G \leq 2) - \mathrm{P}(G \leq 1) = 32/7776 - 1/7776 = 31/7776$.
In a similar fashion, for all $k = 2, 3, 4, 5, 6$ we must have $\mathrm{P}(G = k) = \mathrm{P}(G \leq k) - \mathrm{P}(G \leq k - 1) = k^5/7776 - (k-1)^5/7776$. As a table:

k	1	2	3	4	5	6
$\mathrm{P}(G = k)$	$\frac{1}{7776}$	$\frac{31}{7776}$	$\frac{211}{7776}$	$\frac{781}{7776}$	$\frac{2101}{7776}$	$\frac{4651}{7776}$

(c) We do not need the full joint distribution table. It is enough to find *one* case where the values differ, breaking the independence. Following the hint, we observe if $G = 1$ occurs, then *all* five results are ones, in particular $X_1 = 1$. (More generally we could observe that surely always $X_1 \leq G$.) So $\mathrm{P}(G = 1) = \mathrm{P}(G = 1 \text{ and } X_1 = 1) = 1/7776$. However, $\mathrm{P}(G = 1) \cdot \mathrm{P}(X_{=}1) = (1/7776) \cdot (1/6) \neq 1/7776$.

There are many other ways of observing that independence does not hold here. Perhaps the events $G = 1$ and $X_1 = 2$ would have been even easier, because their intersection event is impossible, so it has zero probability! Also, one can consider conditional probabilities instead of joint probabilities.