Seminar Nr. 4, Discrete Random Variables and Discrete Random Vectors

Theory Review

Bernoulli Distribution with parameter $p \in (0,1)$ pdf: $X \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$

Binomial Distribution with parameters $n \in \mathbb{N}, p \in (0,1)$ pdf: $X \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}$

Discrete Uniform Distribution with parameter $m \in \mathbb{N}$ pdf: $X \begin{pmatrix} k \\ \dfrac{1}{m} \end{pmatrix}_{k=\overline{1,m}}$

Hypergeometric Distribution with parameters $N, n_1, n \in \mathbb{N}$ $(n_1 \le N)$ pdf: $X \begin{pmatrix} k \\ \dfrac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n} \end{pmatrix}_{k=\overline{0,n}}$

Poisson Distribution with parameter $\lambda > 0$ pdf: $X \begin{pmatrix} k \\ \dfrac{\lambda^k}{k!} e^{-\lambda} \end{pmatrix}_{k=0,1,\dots}$

X represents the number of "rare events" that occur in a fixed period of time; λ represents the frequency, the average number of events during that time.

(Negative Binomial) Pascal Distribution with parameters $n \in \mathbb{N}, p \in (0,1)$ pdf:

$X \begin{pmatrix} k \\ C_{n+k-1}^k p^n q^k \end{pmatrix}_{k=0,1,\dots}$

Geometric Distribution with parameter $p \in (0,1)$ pdf: $X \begin{pmatrix} k \\ pq^k \end{pmatrix}_{k=0,1,\dots}$

Cumulative Distribution Function (cdf) $F_X : \mathbb{R} \to \mathbb{R}$, $F_X(x) = P(X \le x) = \sum\limits_{x_i \le x} p_i$

$(X,Y) : S \to \mathbb{R}^2$ **discrete random vector**:
− **(joint) pdf** $p_{ij} = P\left(X = x_i, Y = y_j\right)$, $(i,j) \in I \times J$,
− **(joint) cdf** $F = F_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R},\ F(x,y) = P(X \le x, Y \le y) = \sum\limits_{x_i \le x} \sum\limits_{y_j \le y} p_{ij},\ \forall (x,y) \in \mathbb{R}^2$,
− **marginal densities** $p_i = P(X = x_i) = \sum\limits_{j \in J} p_{ij},\ \forall i \in I,\ q_j = P(Y = y_j) = \sum\limits_{i \in I} p_{ij},\ \forall j \in J.$

For $X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$, $Y \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j \in J}$,
X and Y are **independent** $<=>$ $p_{ij} = P\left(X = x_i, Y = y_j\right) = P\left(X = x_i\right) P\left(Y = y_j\right) = p_i q_j$.
$X+Y \begin{pmatrix} x_i + y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}$, $\alpha X \begin{pmatrix} \alpha x_i \\ p_i \end{pmatrix}_{i \in I}$, $XY \begin{pmatrix} x_i y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}$, $X/Y \begin{pmatrix} x_i / y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}$ $(y_j \ne 0)$

---

1. A computer virus is trying to corrupt two files. The first file will be corrupted with probability 0.4. Independently of it, the second file will be corrupted with probability 0.3. Find the probability distribution function (pdf) of X, the number of corrupted files.

Solution:
How many files can be corrupted by the virus? $0, 1$ or 2. So X can take the values $0, 1, 2$. With what

probability each?

$$\begin{aligned}
P(X=0) &= P(1^{\text{st}} \text{ no and } 2^{\text{nd}} \text{ no}) = P(1^{\text{st}} \text{ no})P(2^{\text{nd}} \text{ no}) = (1-0.4)(1-0.3) = 0.42;\\
P(X=2) &= P(1^{\text{st}} \text{ yes})P(2^{\text{nd}} \text{ yes}) = 0.4 \cdot 0.3 = 0.12;\\
P(X=1) &= 1 - P(X=0) - P(X=2)\\
&\overset{\text{OR}}{=} P(\{1^{\text{st}} \text{ yes},\ 2^{\text{nd}} \text{ no}\} \cup \{1^{\text{st}} \text{ no},\ 2^{\text{nd}} \text{ yes}\}) = 0.4 \cdot 0.7 + 0.6 \cdot 0.3 = 0.46.
\end{aligned}$$

Actually, the *easiest* way to compute *all* at once, is to use the Poisson model, with "success" meaning a file is corrupted, so parameters are $n = 2$, $p_1 = 0.4$, $p_2 = 0.3$, X is the number of successes, i.e. takes values $X = 0, 1, 2$, with probabilities from

$$\begin{aligned}
(p_1x + q_1)(p_2x + q_2) &= (0.4x + 0.6)(0.3x + 0.7)\\
&= 0.12x^2 + 0.46x + 0.42\\
&= P(X=2)x^2 + P(X=1)x + P(X=0).
\end{aligned}$$

So, the pdf of X is

$$X \begin{pmatrix} 0 & 1 & 2 \\ 0.42 & 0.46 & 0.12 \end{pmatrix}$$

Careful!! This is NOT a Poisson variable! A Poisson random variable has NOTHING to do with the Poisson model!!

2. A coin is flipped 3 times. Let X denote the number of heads that appear.
a) Find the pdf of X. What type of distribution does X have?
b) Find $P(X \le 2)$ and $P(X < 2)$.

Solution:
a) X denotes the number of successes in $n = 3$ trials, where "success" means "heads", so probability of success is $p = 1/2$. Thus, X has a Binomial $B(3, 1/2)$ distribution. Its pdf is

$$X \begin{pmatrix} k \\ C_3^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{3-k} \end{pmatrix}_{k=\overline{0,3}} = \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}.$$

b)

$$\begin{aligned}
P(X \le 2) &= F_X(2) = \frac{7}{8} = \mathit{binocdf}(2, 3, 1/2) = 0.875,\\
P(X < 2) &= P(X \le 1) = \frac{1}{2} = F_X(1) = \mathit{binocdf}(1, 3, 1/2) = 0.5.
\end{aligned}$$

3. (New Accounts) Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day.
a) Find the probability that more than 8 new accounts will be initiated today;
b) Find the probability that at most 16 new accounts will be initiated within 2 days.

Solution:
a) New account initiations qualify as rare events, i.e. discrete events observed over a period of time. Then X, the number of today's new accounts has a Poisson distribution. What is the parameter? The parameter λ is the average number of new accounts initiated per day, thus, $\lambda = 10$. So

$$\begin{aligned}
P(A) &= P(X > 8) = 1 - P(X \le 8)\\
&= 1 - \sum_{k=0}^{8} \frac{10^k}{k!} e^{-10} = 1 - F_X(8)\\
&= 1 - \mathit{poisscdf}(8, 10) = 0.6672.
\end{aligned}$$

b) We argue similarly for Y, the number of new accounts opened within 2 days. But it **IS NOT** $2X$!! Instead, it's like this: if 10 new accounts are opened daily, then within 2 days, on average, 20 new accounts will be opened. Thus Y has a Poisson $\mathcal{P}(20)$ distribution. So

$$\begin{aligned} P(B) &= P(Y \le 16) = \sum_{k=0}^{16} \frac{20^k}{k!} e^{-20} \\ &= F_Y(16) = \mathit{poisscdf}(16, 20) \\ &= 0.2211. \end{aligned}$$

4. It was found that the probability to log on to a computer from a remote terminal is 0.7. Let X denote the number of attempts that must be made to gain access to the computer:
a) Find the pdf of X;
b) Find the probability (express it in terms of the cdf F_X) that at most 4 attempts must be made to gain access to the computer;
c) Find the probability that at least 3 attempts must be made to gain access to the computer.

Solution:
a) What values can X take? The values $1, 2, \ldots$ We compute each probability:

$$\begin{aligned} P(X=1) &= 0.7; \\ P(X=2) &= P(1^{\text{st}} \text{ no and } 2^{\text{nd}} \text{ yes}) = P(1^{\text{st}} \text{ no}) P(2^{\text{nd}} \text{ yes}) = (1-0.7) \cdot 0.7 = 0.7 \cdot 0.3; \end{aligned}$$

Note: One might argue here that the two events ("1^{st} no" and "2^{nd} yes") are *not* independent, since the very fact that we got to the second attempt to log on was the effect of failing in the first attempt. While that is true, still, the probability to log on is *the same*, no matter how many times we failed before. So, even if we considered them *not* to be independent, we would have

$$P(1^{\text{st}} \text{ no} \cap 2^{\text{nd}} \text{ yes}) = P(1^{\text{st}} \text{ no}) P(2^{\text{nd}} \text{ yes} \mid 1^{\text{st}} \text{ no}) = 0.3 \cdot 0.7.$$

Now, proceed further to notice a pattern:

$$\begin{aligned} P(X=3) &= P(1^{\text{st}} \text{ no and } 2^{\text{nd}} \text{ no and } 3^{\text{rd}} \text{ yes}) = (1-0.7)(1-0.7)0.7 = 0.7 \cdot (0.3)^2; \\ \ldots & \ldots \\ P(X=k) &= 0.7 \cdot (0.3)^{k-1} \\ \ldots & \ldots \end{aligned}$$

So, the pdf of X is

$$X \left(\begin{array}{c} k \\ (0.7)(0.3)^{k-1} \end{array} \right)_{k=1,2,\ldots} \quad \text{and} \quad X-1 \left(\begin{array}{c} l \\ (0.7)(0.3)^{l} \end{array} \right)_{l=0,1,\ldots}$$

The variable $X-1$ (the number of *failures* that occurred before the first success) has a Geometric distribution with parameter $p = 0.7$. Variable X (the number of *trials* needed to get the first success) has an "almost" $Geo(0.7)$ distribution. This is known as the *Shifted Geometric* distribution.
Note: In some books, X is called a Geometric random variable.

b)

$$P(X \le 4) = P(X-1 \le 3) = F_X(4) = F_{X-1}(3) = \mathit{geocdf}(3, 0.7) = 0.9919.$$

c)

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \;=\; 1 - P(X \leq 2) \;=\; 1 - F_X(2) \\ &= 1 - F_{X-1}(1) \;=\; 1 - geocdf(1, 0.7) \;=\; 0.09. \end{aligned}$$

5. A number is picked randomly out of 1, 2, 3, 4 and 5. Let X denote the number picked. Let Y be 1 if the number picked was 1, 2 if the number was prime and 3, otherwise.
a) Find the pdf's of X, Y;
b) Find the pdf's of $X + Y$, XY.

Solution:
a) X is the number picked, so it can take the values $1, 2, \ldots, 5$, all equally probable. So the pdf is

$$X \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix},$$

which is a discrete Uniform $U(5)$ distribution.

Y takes values 1 (if $X = 1$), 2 (if $X = 2, 3$ or 5) and 3 (if $X = 4$). So its pdf is

$$Y \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \end{pmatrix}.$$

b) Recall, when doing operations with *discrete* random variables, we operate on their *values*.
So, we get all the possible values of $X + Y$ by adding each possible value of X to each possible value of Y. Thus, $X + Y$ can take values: $2, 3, 4, 5, 6, 7$ and 8. Now, for the probabilities:

$$P(X + Y = 2) \;=\; P(X = 1, Y = 1),$$

where the comma means $\cap$. Now, are the events $(X = 1)$ and $(Y = 1)$ independent? **Not at all!** The value of Y is obviously very much depending on the value of X, in fact, it is *completely* determined by the value of X. Once the value of X is known, we automatically have the value of Y. As events,

$$(X = 1) \text{ implies (induces) } (Y = 1), \text{i.e.}$$

$$\begin{aligned} (X = 1) &\subseteq (Y = 1), \text{i.e.} \\ (X = 1) \cap (Y = 1) &= (X = 1). \end{aligned}$$

So, $P(X = 1, Y = 1) = P(X = 1) = 1/5$. Similarly we compute the other probabilities. Notice that, because of the dependence of X and Y, some combinations of values are *impossible*.

$$\begin{aligned} P(X + Y = 2) &= P(X = 1) \;=\; \frac{1}{5}, \\ P(X + Y = 3) &= P\Big((X = 1, Y = 2) \cup (X = 2, Y = 1)\Big) \;=\; 0 + 0 \;=\; 0 \text{ (both are impossible)}, \\ P(X + Y = 4) &= P\Big((X = 1, Y = 3) \cup (X = 2, Y = 2) \cup (X = 3, Y = 1)\Big) \\ &= P(X = 2, Y = 2) \;=\; P(X = 2) \;=\; \frac{1}{5}, \\ \ldots \quad & \ldots \quad \ldots \end{aligned}$$

In the end (recall that we *do not* list in the pdf values with probability 0), the pdf of $X + Y$ is

$$X + Y \begin{pmatrix} 2 & 4 & 5 & 7 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \end{pmatrix}.$$

In a similar fashion, we find the pdf of $X \cdot Y$. The variable $X \cdot Y$ can take values $1, 2, 3, 4, 5, 6, 8, 9, 10, 12$ and 15. Again, some combinations are impossible. For example,

$$\begin{aligned} P(X \cdot Y = 4) &= P\Big((X = 2, Y = 2) \cup (X = 4, Y = 1)\Big) \\ &= P(X = 2, Y = 2) \;=\; P(X = 2) \;=\; \frac{1}{5}. \end{aligned}$$

In the end, the pdf of $X \cdot Y$ is

$$X \cdot Y \begin{pmatrix} 1 & 4 & 6 & 10 & 12 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}$$

6. Same problem with 2 numbers being picked randomly. Variable X refers to the 1^{st} number, variable Y to the 2^{nd}. Is there a difference in the answers, from the previous problem?

Solution:

Part a) is the same.

b) The difference is that now X and Y are **independent**, so *all* combinations of values are possible. For example,

$$\begin{aligned} P(X + Y = 7) &= P(X = 4, Y = 3) + P(X = 5, Y = 2) \\ &\stackrel{\text{ind}}{=} P(X = 4)P(Y = 3) + P(X = 5)P(Y = 2) \\ &= \frac{1}{25} + \frac{3}{25} \;=\; \frac{4}{25}. \end{aligned}$$

Thus, now the pdf's are

$$X + Y \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \frac{1}{25} & \frac{4}{25} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{4}{25} & \frac{1}{25} \end{pmatrix}$$

$$X \cdot Y \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 8 & 9 & 10 & 12 & 15 \\ \frac{1}{25} & \frac{4}{25} & \frac{2}{25} & \frac{4}{25} & \frac{1}{25} & \frac{4}{25} & \frac{3}{25} & \frac{1}{25} & \frac{3}{25} & \frac{1}{25} & \frac{1}{25} \end{pmatrix}$$

7. An internet service provider charges its customers for the time of the internet use. Let X be the used time (in hours, rounded to the nearest hour) and Y the charge per hour (in cents). The joint pdf for (X, Y) is given in the following table:

X \ Y	1	2	3
1	0	0.10	0.40
2	0.06	0.10	0.10
3	0.06	0.04	0
4	0.10	0.04	0

Find

a) the marginal pdf's of X and Y;

b) the probability that a customer will be charged only 1 cent per hour when being online for 2 hours (event B);

c) the probability that a customer will be charged at most 2 cents per hour when being online for at least 3 hours (event C);

d) the pdf of Z, the total charge for a customer.

Solution:

a) To get the *marginal* pdf's, i.e. the pdf's of the components X and Y, for each fixed value of one of them, we add all the values on that row or column, hence, getting them on the *margins*.

$X \backslash Y$	1	2	3	
1	0	0.10	0.40	0.50
2	0.06	0.10	0.10	0.26
3	0.06	0.04	0	0.10
4	0.10	0.04	0	0.14
	0.22	0.28	0.50	

So, the marginal pdf's are

$$X\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0.5 & 0.26 & 0.1 & 0.14 \end{pmatrix} \text{ and } Y\begin{pmatrix} 1 & 2 & 3 \\ 0.22 & 0.28 & 0.5 \end{pmatrix}.$$

b) We get the value from the table.

$$P(B) = P\Big((X,Y) = (2,1)\Big) = 0.06$$

c) Same here, but we have more combinations.

$$\begin{aligned} P(C) &= P(X \geq 3, Y \leq 2) \\ &= P\Big((X = 3 \cup X = 4) \cap (Y = 1 \cup Y = 2)\Big) \\ &= P(X = 3, Y = 1) + P(X = 3, Y = 2) + P(X = 4, Y = 1) + P(X = 4, Y = 2) \\ &= 0.06 + 0.04 + 0.1 + 0.04 = 0.24. \end{aligned}$$

d) The total charge for a customer is given by the number of total hours spent online multiplied by the price per each hour, i.e. $Z = X \cdot Y$. Z can take the values $\{1, 2, 3, 4, 6, 8, 9, 12\}$. We have:

$$\begin{array}{lclclcl} P(Z = 1) &=& P(X = 1, Y = 1) &=& 0, && \\ P(Z = 2) &=& P(X = 1, Y = 2) &+& P(X = 2, Y = 1) &=& 0.16, \\ P(Z = 3) &=& P(X = 1, Y = 3) &+& P(X = 3, Y = 1) &=& 0.46, \\ \ldots & \ldots & & & & & \end{array}$$

So, the pdf is

$$Z\begin{pmatrix} 2 & 3 & 4 & 6 & 8 \\ 0.16 & 0.46 & 0.2 & 0.14 & 0.04 \end{pmatrix}.$$