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

---

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.

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 $P(X < 2)$.

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.

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.

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.

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?

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:

$X \backslash Y$	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.

Bonus Problems:

8. Let X and Y be two independent random variables such that X has a discrete Uniform distribution with parameter 2 and Y has a Bernoulli distribution with parameter $\frac{1}{3}$. Let $U = X + Y$ and $V = X - Y$.
a) Find the joint pdf of (U, V).
b) Find the marginal pdfs of U and V.
c) Are U and V independent? Justify the answer.

9. A point in plane has coordinates (X, Y), where the values of X and Y are determined by rolling two dice (one for X and one for Y). Find the probability that the point (X, Y) is on the circle $x^2 + y^2 = 10$.

16.11.2023

input params. $\longrightarrow$ $\boxed{}$ $\longrightarrow$ output rand. nr.

$$X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I} \quad I \subseteq \mathbb{N}$$

$\longrightarrow$ distribution (pdf)

ex: • $X = \begin{cases} 1 & \text{"heads"} \\ 0 & \text{"tails"} \end{cases}$

p = prob. of "heads" Bernoulli

$1-p$ = prob. of "tails"

• X = result of rolling a dice with $\underline{m}$ faces

$$X\begin{pmatrix} 1 & 2 & \dots & m \\ \frac{1}{m} & \frac{1}{m} & \dots & \frac{1}{m} \end{pmatrix} \longrightarrow \text{distribution of } X$$

<u>Discrete uniform</u>

name_prefix/pdf(x, param.) $= P(X = x)$

name_prefix/cdf(x, param) $= P(X \leq x)$

name_prefix/rnd (param)
= generate a val
for X

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.

X = nr of corrupted files out of 2 files

prob of 0.4 to be corrupted

prob of 0.3 ... 2nd ...

$X\begin{pmatrix} ? & ? & ? \\ ? & ? & ? \end{pmatrix}$ distribution (independent)

M1

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

$P(X=0) = P(\overline{F_1}) \cdot P(\overline{F_2}) = 0.6 \cdot 0.7 = 0.42$

F_i : " file i is corrupted " $i = 1,2$

$P(X=2) = P(F_1) \cdot P(F_2) = 0.4 \cdot 0.3 = 0.12$

M2

$\frac{0.4}{1}$ $\frac{0.3}{2}$ $\longrightarrow$ Poisson Model

$(0.4 \cdot x + 0.6) \cdot (0.3 \cdot x + 0.7) = \ldots$

$= 0.12 \cdot x^2 + 0.46 \cdot x + 0.42 \cdot x^0$

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 $P(X < 2)$.

X = nr. of heads by coin $\times$ 3 times
a) name of the distribution, pdf of x
b) $P(X \le 2)$ & $P(X < 2)$

$q = 1-p \rightarrow$ prob of failure
$n=3$
$p=0.5$

Binomial $\rightarrow$ (same prob of success)

$$X \begin{pmatrix} k \\ C_3^k \frac{1}{2^3} \end{pmatrix}_{k=\overline{0,3}}$$

$$C_n^k \cdot p^k \cdot q^{n-k}$$
$$= C_3^k \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{3-k}$$
$$= C_3^k \cdot \left(\frac{1}{2}\right)^3$$

b) $P(X \le 2)$ = binocdf(2, 3, 0.5)
$P(X < 2)$ = binocdf(1, 3, 0.5)

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.

X_i = new accounts in i days $\quad i = 1,2$
avg. rate = 10 acc/day = λ
a) $P(X_1 > 8)$
b) $P(X_2 \le 16)$

a) Poisson distr.

$$P(X_1 > 8) \overset{\lambda_1=10}{=\!=} \sum_{k=9}^{\infty} \frac{10^k}{k!} \cdot e^{-10} = 1 - \sum_{k=0}^{8} \frac{10^k}{k!} \cdot e^{-10}$$

b) $$P(X_2 \leq 16) \overset{\lambda_2=20}{=\!=} \sum_{k=0}^{16} \frac{20^k}{k!} \cdot e^{-20}$$

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.

X = nr. of attempts to log on

a) pdf $X \begin{pmatrix} \cdots \\ \cdots \end{pmatrix}$

P("succesful attempt") = 0.7

b) $P(X \leq 4)$

c) $P(X \geq 3)$

a) independence $\Rightarrow$ " $\cdot$ "

$P(X < 1) = 0.7$

$X \begin{pmatrix} k \\ 0.7 \cdot 0.3^{k-1} \end{pmatrix} \quad k = 1, 2, \ldots$

$\hookrightarrow$ P("failure")

y = nr of failures before the success

Geom. dist.

$y + 1 = X \begin{pmatrix} k+1 \\ 0.7 \cdot 0.3^k \end{pmatrix} \quad k = 0, 1, \ldots$

$y \begin{pmatrix} k \\ 0.7 \cdot 0.3^k \end{pmatrix} \quad k = 0, 1 \ldots$

b) $P(X \leq 4) = \text{geocdf}(3, 0.7)$

$P(y \leq 3)$ → 3

c) $P(X \geq 3) = 1 - \text{geocdf}(1, 0.7)$

$P(y \geq 2)$ → 1

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.

X = rand nr from $\{1,2,3,4\}$

$$y = \begin{cases} 1 & x = 1 \\ 2 & x = \text{prime} \\ 3 & \text{otherwise} \end{cases}$$

a) pdf $y(\;)$

b) pdf $x+y(\;)$ $\quad x \cdot y(\;)$

a) pdf $y\begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \end{pmatrix}$

$x = 1 \Rightarrow y = 1$

$x = 2 \Rightarrow y = 2$

$x = 3 \Rightarrow y = 2$

$x = 4 \Rightarrow y = 3$

$x = 5 \Rightarrow y = 2$

$P(y=1) = P(x=1)$

$P(y=2) = P(x \in \{2,3,5\}) = \frac{3}{5}$