Discrete probability distributions.

Probability distributions.

- **Discrete** توزيع متقطع (Chapter 3)
  - → Binomial
  - → Poisson
  - → Geometric
  - → Hypergeometric
- **Continuous** توزيع متصل (Chapter 4)
  - → Normal
  - → Uniform
  - → Exponential

The binomial distribution.

Def. An experiment is said to be a binomial experiment if it satisfies the following conditions:-

1) The experiment can be performed as many times as desired.
2) Each trial has exactly two possible outcomes, namely: success (denoted by S) and failure (F).
3) The probability of success is the same for all trials.

4) All trials are independent. So, the outcome of any trial has no effect on any subsequent trial

Example: The following are binomial experiments:

1) Tossing a coin 5 times and recording the number of heads.
2) Answering a multiple-choice exam composed of 20 questions by wild guessing and recording the number of correct answers.
3) Taking 15 electric bulbs from the production unit of a factory of electric bulbs and recording the number of defective bulbs.
4) Throwing 10 basketballs into the hoop from behind and recording the number of successful throws.

Repeating a binomial experiment a fixed number of times, say n, and letting X be the number of successes within these n trials of the experiment, then X is a random variable.

We call X a binomial random variable, and call its distribution the binomial distribution with parmeters n and p, where p is the probability of success in each trial $(p = P(S))$.

Notation: We use $X \sim B(n,p)$ to mean that X has the binomial distribution with parameters n and p.

- In the above example, it is clear that in the experiment number:

1) we have $X \sim B(5, \frac{1}{2})$. $P(\{Head\})$.
2) we have $X \sim B(20, \frac{1}{5})$, assuming that each question has 5 choices.
3) we have $X \sim B(15, 0.1)$, assuming that the proportion of defectives in the product is 10%.
4) we have $X \sim B(10, \frac{1}{2})$, assuming that the probability of any successful throw is 1/2.

<u>Fact</u>: Let $X \sim B(n, p)$, then

1) the set of all possible values of X (called the space of the random var. X) is $\{0, 1, 2, \dots\}$.
2) the <u>p.d.f.</u> of $x \in \{0, 1, 2, \dots, n\}$ is given by
$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ where } \binom{n}{x} = \frac{n!}{x!\,(n-x)!}.$$
3) the <u>mean</u> of X is $\mu_X = E(X) = np$.
4) the <u>variance</u> of X is $\sigma_X^2 = E(X^2) - (E(X))^2$
$$= np(1-p).$$

<u>Ex</u>. Assume that 30% of the residents of Irbid city are smokers. If 5 persons are randomely selected, find the probability of having exactly 2 smokers within them.

<u>Soln</u>. Let X be the number of smokers out of 5 selected persons, then $X \sim B(5, 0.3)$.

Now, $P(X=2) = \binom{5}{2} (0.3)^2 (0.7)^3 = \dots$.

Ex. A multiple-choice exam composed of 10 questions, each question has 5 choices. If a student answers this exams by pure guessing, find the probability that s/he will answer at least 2 questions correctly.

Soln. Let X be the number of questions answered by the student correctly, then $X \sim B(10, 0.2)$.

Now, $P(X \geq 2) = 1 - P(X < 2)$

$$= 1 - [P(X=0) + P(X=1)]$$

$$= 1 - \left[\binom{10}{0}(0.2)^0(0.8)^{10} + \binom{10}{1}(0.2)^1(0.8)^9\right]$$

$$= 1 - \left[(0.8)^{10} + 10(0.2)(0.8)^9\right].$$

Ex. Assume that the chance of hitting a target in a single shot is 0.4. If 3 shots are fired independently at the target, what is the probability that a) exactly one shot will hit the target?

b) at most one shot " " " " ?

c) at least one shot " " " " ?

d) exactly two shots " " " " ?

Soln. Let X be the number of hits out of 3 shots, then $X \sim B(3, 0.4)$.

a) $P(X=1) = \binom{3}{1}(0.4)^1(1-0.4)^2$

$= 0.432.$

b) $P(X \le 1) = P(X=0) + P(X=1)$

$= \binom{3}{0}(0.4)(0.6)^3 + 0.432.$

$= 0.216 + 0.432$

$= 0.648.$

c) $P(X \ge 1) = 1 - P(X<1)$

$= 1 - P(X=0)$

$= 1 - 0.216$

$= 0.784.$

d) $P(X=2) = \binom{3}{2}(0.4)^2(0.6)^1$

$= 0.288.$

Ex. Let $X \sim B(10, 0.8)$, find:

1) $P(X>0)$

2) $E(X)$.

3) $\text{var}(X)$

4) $E(X^2)$

5) $P(\mu_X - \sigma_X \le X \le \mu_X + \sigma_X)$.

Soln.

1) $P(X>0) = 1 - P(X=0) = 1 - \binom{10}{0}(0.8)^0(0.2)^{10}$

$$= 1 - (0.2)^{10} = 0.9\overline{9}.$$

2) $E(X) = np = 10(0.8) = 8.$

3) $Var(X) = np(1-p) = 10(0.8)(0.2) = 1.6.$

4) It is known that $Var(X) = E(X^2) - (E(X))^2.$

Then $E(X^2) = Var(X) + (E(X))^2$

$$= 1.6 - (8)^2 = 65.6.$$

5) $P(\mu_X - \sigma_X \le X \le \mu_X + \sigma_X) = P(3 - \sqrt{1.6} \le X \le 8 + \sqrt{1.6})$

$$= P(6.74 \le X \le 9.26)$$

$$= P(7 \le X \le 9)$$

$$= P(X=7) + P(X=8) + P(X=9)$$

Exc.

$$= \ldots .$$

Binomial tables.

The binomial tables give the cumulative binomial probabilities for different values of n, p and k.

If $X \sim B(n,p)$, then to find $P(X \leq k) = T(k)$, we choose the entry in the k^{th}-row and p^{th}-column of the n^{th}-table.

The binomial table for $n=10$.

$n=10$											
					P						
k	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
0	*	*	0.107	*	*	*	*	*	*	*	*
1	*	*	0.376	*	*	*	*	*	*	*	*
2	*	*	0.678	*	*	*	*	*	*	*	*
3	*	*	0.879	*	*	*	*	*	*	*	*
4	*	*	0.967	*	*	*	*	*	*	*	*
5	*	*	0.994	*	*	*	*	*	*	*	*
6	*	*	0.999	*	*	*	*	*	*	*	*
7	*	*	1	*	*	*	*	*	*	*	*
8	*	*	1	*	*	*	*	*	*	*	*
9	*	*	1	*	*	*	*	*	*	*	*
10	*	*	1	*	*	*	*	*	*	*	*

Ex. If $X \sim B(10, 0.2)$, then using the binomail table we have

a) $P(X \leq 4) = T(4) = 0.967$.

b) $P(X < 4) = P(X \leq 3) = T(3) = 0.879$.

c) $P(X > 3) = 1 - P(X \leq 3)$

$$= 1 - T(3) = 1 - 0.879 = 0.121.$$

d) $P(2<X<5) = P(X \leq 4) - P(X \leq 2)$
$= T(4) - T(2)$
$= 0.967 - 0.678 = 0.289$

e) $P(2 \leq X < 5) = P(X \leq 4) - P(X \leq 1)$
$= T(4) - T(1)$
$= 0.967 - 0.376 = 0.591.$

f) $P(2 < X \leq 5) = P(X \leq 5) - P(X \leq 2)$
$= T(5) - T(2)$
$= 0.994 - 0.678 = 0.316.$

g) $P(2 \leq X \leq 5) = P(X \leq 5) - P(X \leq 1)$
$= T(5) - T(1)$
$= 0.994 - 0.376 = 0.618.$

h) $P(X = 4) = P(X \leq 4) - P(X \leq 3)$
$= T(4) - T(3)$
$= 0.967 - 0.879 = 0.088.$

Ex. If X is a binomial random variable with mean 2 and variance 1.6, find a) $P(X \leq 4)$ b) $P(X = 4)$.

Soln. Since $\mu = np = 2$ and $\sigma^2 = np(1-p) = 1.6$, we have $1-p = 1.6/2 = 0.8$, so $p = 1 - 0.8 = 0.2$ and $n = 2/0.2 = 10$. Thus, $X \sim B(10, 0.2)$.

It follows that

a) $P(X \leq 4) = T(4) = 0.967$.

b) $P(X = 4) = T(4) - T(3) = 0.088$.

Ex. If $X \sim B(25, 0.2)$, ← table is given in the appendix of the book.

find a) $P(\mu - \sigma < X < \mu + \sigma)$

b) $P(\mu - \sigma \leq X \leq \mu + \sigma)$.

Soln. $\mu = np = 25(0.2) = 5$.

$\sigma = \sqrt{np(1-p)} = \sqrt{25(0.2)(0.8)} = 2$.

a) $P(\mu - \sigma < X < \mu + \sigma) = P(3 < X < 7)$

$= T(6) - T(3)$ Exc

b) $P(\mu - \sigma \leq X \leq \mu + \sigma) = P(3 \leq X \leq 7)$

$= T(7) - T(2)$ Exc

The Poisson distribution ← Could be skipped!

Named after French mathematician Simeon Denis Poisson.

Def. The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occuring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event.

From the above definition, we are interested in the number of occurrences of a certain event during a certain period of time.

Assume that the mean (or the average) number of occurences depends only on the length of the period of time under consideration, and let X be this number of occurences in the period/interval of time under consideration, then X is called a Poisson random variable if it satisfies :-

1) The probability of one occurence (i.e., $P(X=1)$) is directly propotional to the length of that interval.
2) The probability of two or more occurences (i.e., $P(X>1)$) is negligible (too small) if the length of the interval is too small.
3) The occurences are independent.

Notation: We write $X \sim \text{Poisson}(\mu)$ to mean that X has a Poisson distribution with parameter μ, where μ is the mean number of occurences in the period of time under consideration.

Examples:
1) If X is the number of deaths per month and if the average number of deaths per month is 2, then $X \sim \text{Poisson}(2)$.
2) If Y is the number of typo errors per page and if the average number of typo errors per page is 3, then $Y \sim \text{Poisson}(3)$.

Fact: If $X \sim \text{Poisson}(\mu)$, then

1) The possible values of X are $0, 1, 2, \ldots$.
2) $P(X=k) = \dfrac{e^{-\mu}\mu^{k}}{k!}$, $k = 0, 1, 2, \ldots$.
3) $E(X) = Var(X) = \mu$.

Ex. Assume that X is the number of coronavirus patients entered the ICU (intensive care unit) at Al-Basheer hospital per day. If the average number of patients entered ICU per day is 3, what is the probability that in a given day there will be (a) exactly two admissions.

(b) at most two admissions.

Soln. Note that $X \sim \text{Poisson}(3)$. So,

(a) $P(\text{exactly two admissions}) = P(X=2) = \dfrac{e^{-3}3^{2}}{2!} = 0.224$.

(b) $P(\text{at most two admissions}) = P(X \le 2)$

$= P(X=0) + P(X=1) + P(X=2)$

$= 0.423$.

Ex. Assume that $X \sim \text{Poisson}(\mu)$. If $P(X=0)=P(X=1)$, find $E(X^2)$.

Soln. $P(X=0) = P(X=1)$

$$\Rightarrow \frac{e^{-\mu}\mu^0}{0!} = \frac{e^{-\mu}\mu^1}{1!}$$

$$\Rightarrow \overset{1}{\cancel{e^{-\mu}}} = \mu \cancel{e^{-\mu}}. \text{ Thus, } \mu = 1.$$

Now, $E(X^2) = Var(X) + (E(X))^2 = \mu + \mu^2 = 1 + 1^2 = 2$.

Poisson tables.

The Poisson table gives the cumulative Poisson probabilities for different values of μ and k.

If $X \sim \text{Poisson}(\mu)$ then to find $P(X \le k) = T(k)$, we choose the entry in the k^{th} row and μ^{th} column of the Poisson table.

Poisson table with $\mu = 2, 2.5, 3$

k	μ		
	2.0	2.5	3.0
0	0.135	*	*
1	0.406	*	*
2	0.677	*	*
3	0.857	*	*
4	0.947	*	*
5	0.983	*	*
6	0.995	*	*
7	0.999	*	*
8	1.000	*	*
9		1.000	*
10			1.000

Ex. If $X \sim \text{Poisson}(2)$, find

a) $P(X \leq 3)$ b) $P(X=3)$ c) $P(1 < X \leq 4)$ d) $P(1 \leq X \leq 4)$.

Soln. a) $P(X \leq 3) = T(3) = 0.857$.

b) $P(X=3) = T(3) - T(2) = 0.857 - 0.677 = 0.18$.

c) $P(1 < X \leq 4) = T(4) - T(1) = 0.947 - 0.406 = 0.541$.

d) $P(1 \leq X \leq 4) = T(4) - T(0) = 0.947 - 0.135 = 0.812$.

Ex. The number of cars arriving at a service station has a Poisson distribution with mean one car per hour. If X is the number of cars having in two hours, find the probability that at least 4 cars arrive in two randomly selected hours.

Soln. $\mu = 2(1) = 2$ ← Mean in 2 hours.

Then $X \sim$ Poisson (2).

$P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.857 = 0.143$.

Searching keywords:

- Discrete distribution
- Binomial distribution
- Poisson distribution
- Find the probability of
- The University of Jordan الجامعة الأردنية
- Principles of Statistics مبادئ الإحصاء
- Baha Alzalg بهاء الزالق

References: See the course website

http://sites.ju.edu.jo/sites/Alzalg/Pages/131.aspx

For any comments or concerns, please use my email to contact me.

د. بهاء محمود الزالق
The University of Jordan
Dr. Baha Alzalg
baha2math@gmail.com