Probability and Statistics

DR. AHMED TAYEL

Department of Engineering Mathematics and Physics, Faculty of Engineering, Alexandria University

ahmed.tayel@alexu.edu.eg

Outline

Part 2 – Random variables

2.3 Some important discrete PMF's

- Bernoulli PMF → 1 trial
  - success $x=1$
  - fail $x=0$
- Binomial PMF → indep. Bernoulli trials "n trial"
  - X: no. of successes out of n
- Geometric PMF → indep. Bernoulli trials until success
  - X: no. of trials until success
- Poisson PMF → $\lim$ Binomial PMF
  - no. of trials → ∞
  - $P_{success} \to 0$

Binomial PMF

indep. Bern. trials

$n=3$

Event of interest

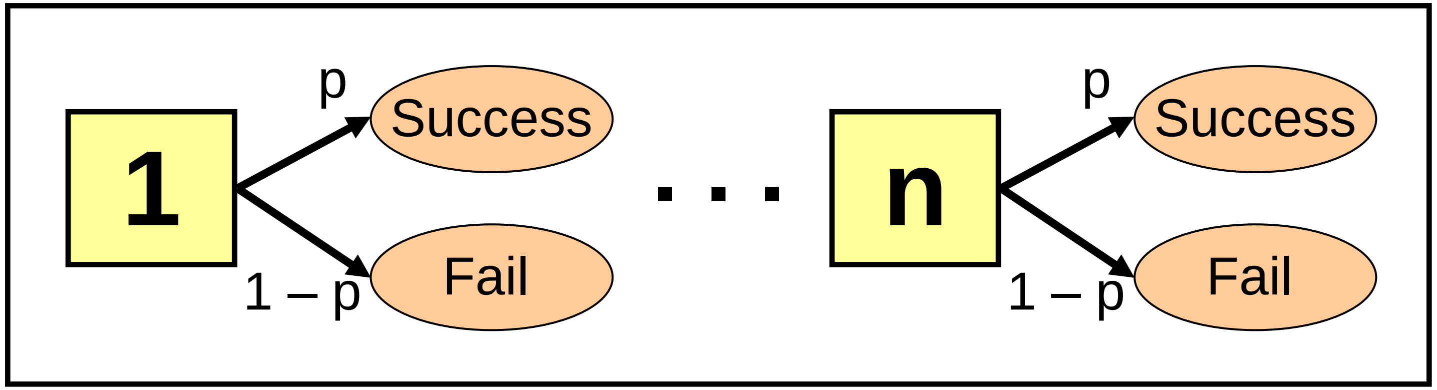
$P_{sucess} = 1/2$

Binomial random variable

Motivation

- A coin is tossed 3 times. Find the probability of getting 2 heads.
- A coin is tossed 10 times. Find the probability of getting 7 heads.
- A biased coin is tossed 10 times. Find P(7 heads).

Binomial experiment



Independent Bernoulli trials

Binomial RV

X is the *number of successes* $\Rightarrow$ **X ~ Binomial (n, p)**

$x = 0, 1, 2, \ldots, n$

Deriving the Binomial PMF

$$P_X(x) = {}^nC_x \, p^x (1-p)^{n-x}$$

x success out of n ; $x = 0, 1, 2, \ldots, n$

1 : success
0 : fail
p : prob. of success

Assume 3 trials
req. prob. (1 success)

$\xrightarrow{\text{Stages}}$

1 0 0 $\to p(1-p)^2$
or 0 1 0 $\to p(1-p)^2$
or 0 0 1 $\to p(1-p)^2$

Alternatives

$$P(1 \text{ success}) = p(1-p)^2 + p(1-p)^2 + p(1-p)^2$$

$$= 3\, p(1-p)^2$$

${}^3C_1$

Validity of the PMF

Check $P_X(x)$ is a proper pmf

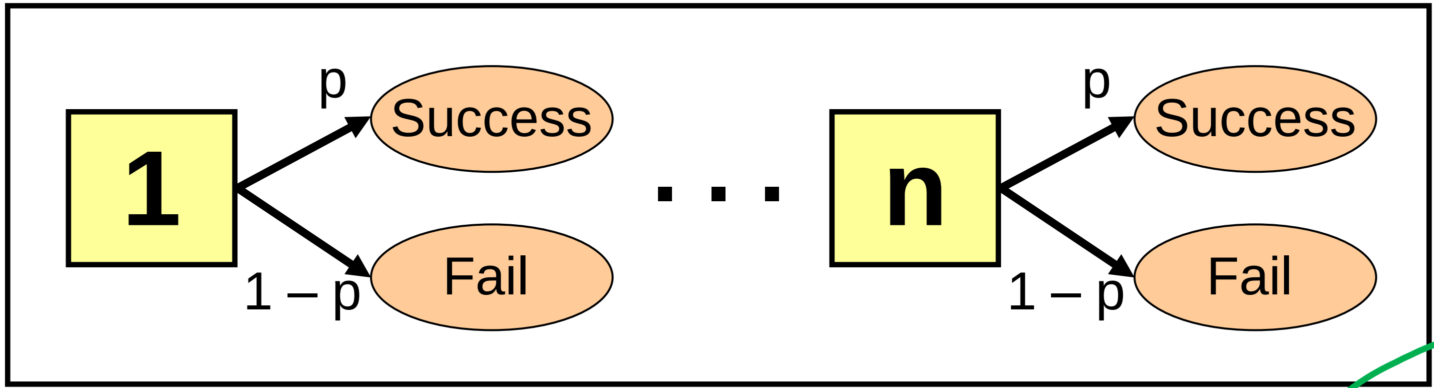
$$\sum_{\forall x = 0,1,2,\ldots,n} {}^nC_x \, p^x \underbrace{(1-p)}_{q}{}^{n-x} = {}^nC_0 \, p^0 q^n + {}^nC_1 \, p^1 q^{n-1} + \cdots + {}^nC_n \, p^n q^0$$

$$= (p+q)^n = (p+1-p)^n = 1^n = 1$$

Binomial random variable

$V(X_1 - X_2) = V(X_1) + \underbrace{(-1)^2}_{+} V(X_2)$

Binomial experiment



success

fail

1

0

X_i : Bernoulli trial no. i

Binomial RV

X is the *number of successes* ⇒ **X ~ Binomial (n, p)**

# of trials

$P_{success}$ in one trial

$$P_X(x) = \underbrace{\binom{n}{x}}_{{}^nC_x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n$$

$$n = 1, 2, 3, \ldots, \quad 0 < p < 1$$

E(X) = n p, #

Var(X) = n p (1 – p) #

Bin.

$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = \underbrace{p + p + \cdots + p}_{n \text{ times}} = np$

$Var(X) = Var(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n) = \underbrace{p(1-p) + \cdots + p(1-p)}_{n \text{ time}}$

$= np(1-p)$

Example

A die is thrown 7 times. Find the probability of getting $\leftarrow n$

(a) a six twice.

(b) an odd number at least twice.

(a) $\underbrace{\text{Event of interest}}_{\text{Event to be counted}}$ "success" 6 appears $\dashrightarrow P = 1/6$

X: no. of times 6 appears $\sim \text{Binomial}(7, 1/6)$

$$P_X(x) = {}^7C_x \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{7-x} \quad ; x = 0, 1, 2, \ldots, 7$$

$$P(X=2) = P_X(2) = {}^7C_2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{7-2} = \checkmark$$

very imp. Starting pt.

(b) Success: odd no. $\dashrightarrow P = \frac{3}{6} = \frac{1}{2}$ $\quad Y \sim \text{Bin}(7, \frac{1}{2})$; $P_Y(y) = {}^7C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{7-x}$

Y: no. of times an odd no. appears

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - [P_Y(0) + P_Y(1)]$$

Example

An arriving call is *equally likely* to be voice or data. Observe 10 calls which are assumed to be independent.

What is the probability of

(a) only 3 voice calls?

(b) at most one voice call?

n

success $\quad P = 1/2$

X: no. of voice calls $\sim Bin(10, \frac{1}{2})$

$$P_X(x) = {}^{10}C_x \left(\tfrac{1}{2}\right)^x \left(\tfrac{1}{2}\right)^{10-x} \; ; \; x = 0, 1, 2, \ldots, 10$$

(a) $P(X=3) = P_X(3)$ direct sub. $= {}^{10}C_3 \left(\tfrac{1}{2}\right)^3 \left(\tfrac{1}{2}\right)^7 = \checkmark$

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

Example

A box contains 4 white balls and 6 green balls. A ball is selected randomly, its color is recorded, and then the ball drawn is returned back to the box. Suppose this experiment is repeated 5 times. Find the expected number of white balls in the 5 trials and compute its variance.

indep.

n

success

X: no. of white balls $\sim Bin(5, 0.4)$

$\boxed{\begin{matrix} 4W \\ 6G \end{matrix}}$ $P(w) = 4/10$

$E(X) = np = 5 * 0.4 = 2$

Example

Data packets containing 100 bits are transmitted over a communication link. P(A bit in error) = q. Bits are independent.

(a) What is the probability that exactly two bits are received in error in a packet.

(b) A packet can be recovered if at most 2 bits are in error.

What is the probability that a packet can't be recovered?

X: $\underbrace{\text{no. of bits in error}}_{\text{to be counted}}$ "success" $\sim \text{Bin}(100, q)$

$$P_X(x) = {}^{100}C_x\, q^x (1-q)^{100-x}$$

(a) $P(X=2) = P_X(2)$

(b) $P_{Recov.} = P(X \le 2)$

$= P_X(0) + P_X(1) + P_X(2) = \square$

$P_{Not\ Recov.} = P(X > 2)$

$= 1 - P_{Recov.} = 1 - \square = \checkmark$

Example

A multiple choice test consists of 10 questions, each with 4 choices. If you answer the questions randomly.

(a) What is the probability that you correctly answer at most 7 questions?

(b) What is the expected number of questions answered wrong?

n

X: no. of questions answered correctly $\sim Bin(10, \underset{1/4}{p_{correct}})$

$$P_X(x) = {}^{10}C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x}$$

(a) $P(X \leq 7) = 1 - P(X > 7) = 1 - [P_X(8) + P_X(9) + P_X(10)]$

(b) Y: no. of ques. ans. wrong $\sim Bin\left(10, \frac{3}{4}\right)$

$$E(Y) = np = 10 * \frac{3}{4} = 7.5$$

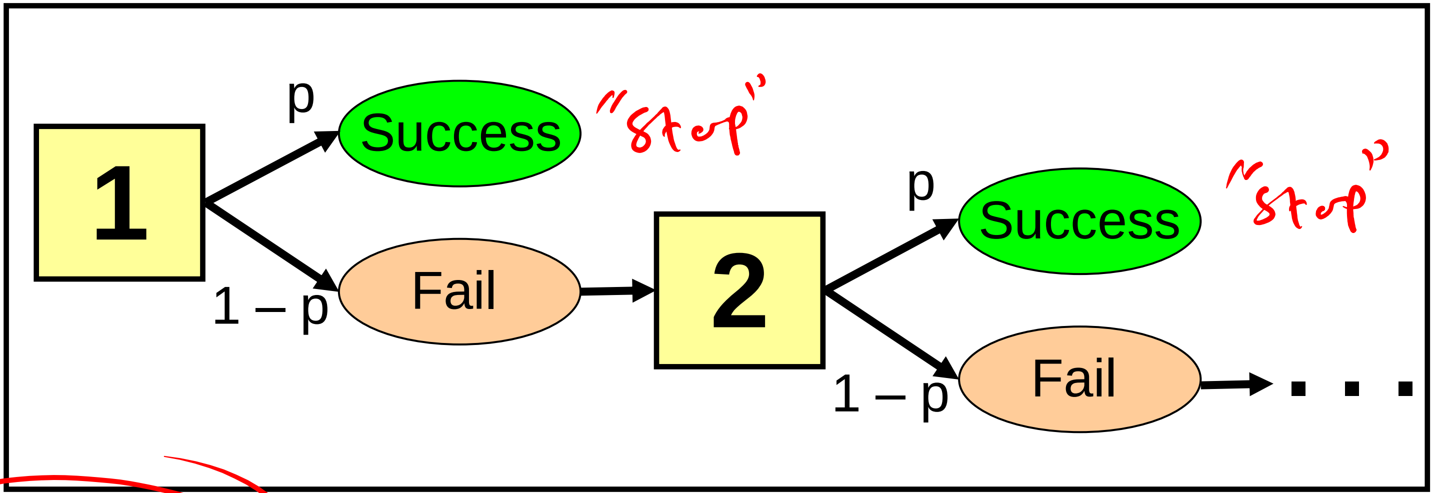
or

$$\begin{aligned} E(Y) &= E(10 - X) \\ &= 10 - E(X) \\ &= 10 - 10 * \frac{1}{4} = 10 * \frac{3}{4} \\ &= 7.5 \end{aligned}$$

Geometric PMF

Geometric random variable

Geometric experiment



Independent Bernoulli trials until success

Geometric RV

X is the *number of trials* ⇒ **X ~ Geometric (p)**

$x = 1, 2, 3, \ldots, \infty$

Deriving the Geometric PMF

X	1^{st} trial	2	3	4	...	x
	S	FS	FFS	FFFS		
	p	$(1-p)p$	$(1-p)^2 p^1$	$(1-p)^3 p^1$		

$$p_X(x) = (1-p)^{x-1} p \; ; \; x = 1, 2, 3, \ldots$$

Validity of the PMF

Skip

Geometric random variable

PMF

$$P_X(x) = p\,(1-p)^{x-1}, \quad x = 1, 2, 3, \ldots$$

$$0 < p < 1$$

Mean & varaince

$$E(X) = \frac{1}{p}, \qquad Var(X) = \frac{1-p}{p^2}$$

Example

Assume that X ~ Geometric (0.1).

p

Compute P(X = 3), P(X ≥ 3)

$$P_X(x) = (1-p)^{x-1}\, p \quad ; \quad x = 1, 2, 3, \ldots$$

$$= (0.9)^{x-1}\, 0.1$$

$$P(X=3) = P_X(3) = (0.9)^{3-1}(0.1) = \checkmark$$

$$P(X \geq 3) = 1 - P(X<3) = 1 - \left[P_X(1) + P_X(2) \right]$$

Starting pt.

Example

A die is thrown until a six appears. success $\to p = 1/6$

What is the probability it is thrown 3 times?

X: no. of trials until six appears $\sim \mathrm{Geom}(1/6)$

$$p_X(x) = \frac{1}{6}\left(\frac{5}{6}\right)^{x-1}$$

$$P(X=3) = p_X(3) =$$

Example

At busy time of a telephone exchange, people find it difficult to use a line. It may be of interest to know the expected number of attempts necessary in order to gain a connection. Suppose that probability of connecting during a busy time is 0.05.

Exercise

(a) Find the expected number of attempts.

(b) Find the probability that 5 attempts are necessary.

Example

An arriving call is *equally likely* to be voice or data.

What is the probability that $\underbrace{\text{the first data call is the tenth one}}_{\text{success at the }10^{th}\text{ trial}}$.

Data $\equiv$ success

X: no. of trials until data

$X \sim \text{Geom.}(1/2)$

$$P_X(x) = \frac{1}{2}\left(\frac{1}{2}\right)^{x-1}$$

$$P_X(10) = \frac{1}{2}\left(\frac{1}{2}\right)^{9}$$

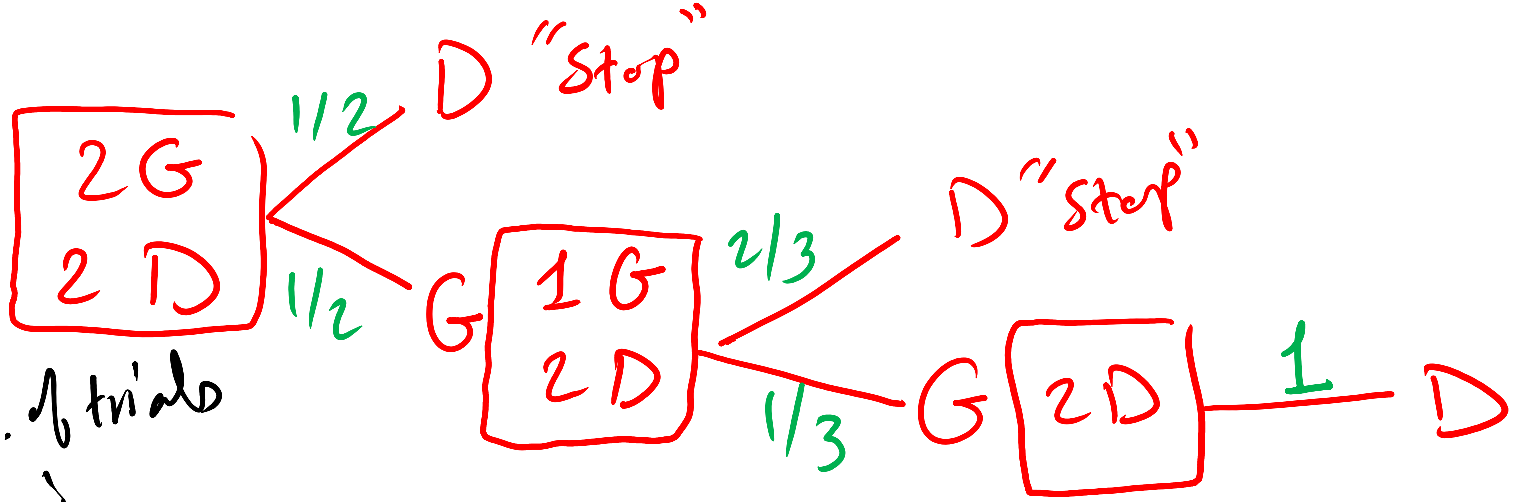
Example

A lot containing two good items and two defective items. The items are tested one after the other until a defective item is found.

Find the expected number of trials to find out defective items?

Not indep.

→ Geom. is not applicable



no. of trials

x	1	2	3
$P_X(x)$	$1/2$	$\frac{1}{2} * \frac{2}{3} = \frac{1}{3}$	$\frac{1}{2} * \frac{1}{3} * 1$ or $1 - [\frac{1}{2} + \frac{1}{3}] = \frac{1}{6}$

$$E(X) = \sum_{\forall x} x \, P_X(x)$$

$$= 1 * \frac{1}{2} + 2 * \frac{1}{3} + 3 * \frac{1}{6}$$

$$= \frac{1}{2} + \frac{2}{3} + \frac{1}{2} = 1\frac{2}{3}$$

Poisson PMF

Poisson random variable

Theorem

$$\lim_{\substack{n\to\infty,\ p\to 0,\\ np\to\lambda}} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda}\frac{\lambda^x}{x!}$$

$E(X) = np \to \lambda$

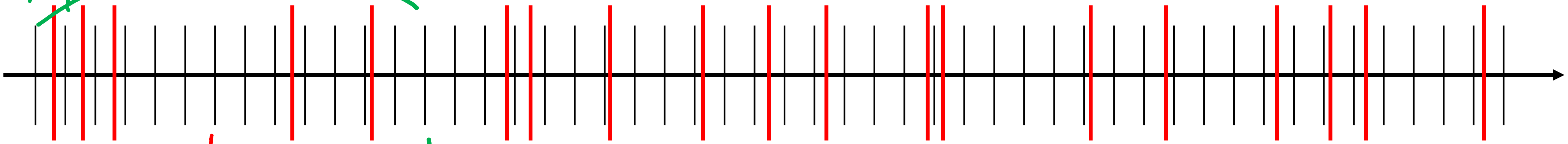
Conds. to apply Poisson

success
p
1-p
fail
trial

Case I To approximate binomial RV ($p < 0.1$ and $n\,p \le 10$)

in 2 hours
$2\lambda = \lambda + \lambda$

λ: Average number of successes in a given interval



no. of event in 1 hour
λ / 1 hour
λ / 1 hr

Case II To *model* number of events that occur ***randomly*** in a unit of *time or space*.

Poisson random variable

Examples of possible Poisson distributions

1) Number of messages arriving at a telecommunications system in a day

2) Number of flaws in a meter of fibre optic cable

3) Number of customers arriving at a Bank in a given time

4) Number of users in a communication cell in a given area

Sum of Poisson variables

If X is Poisson with average number λ_{X} and Y is Poisson with average number λ_Y

Then $X + Y$ is Poisson with average number $\lambda_X + \lambda_Y$

Poisson random variable

PMF

$$P_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

$$\lambda > 0$$

Validity of the PMF

skip

Mean & varaince

$$E(X) = \lambda, \qquad Var(X) = \lambda$$

Example

Assume that X ~ Poisson (3). λ

Bin $n \to \infty$

Compute P(X = 3), P(X ≥ 2)

$$P_X(x) = e^{-\lambda} \frac{\lambda^x}{x!} \; ; \; x = 0, 1, 2, \ldots$$

$$= e^{-3} \frac{3^x}{x!}$$

sub.

$$P(X=3) = P_X(3)$$

Start from 0 similar to Bin

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [\, P_X(0) + P_X(1) \,]$$

Example

Percentage of defective items produced by a machine is 0.005. p

In a lot of 1000 items, what is the probability of getting at most 2 defective items?

X : no. of defec. items

$X \sim Bin(1000, 0.005)$

$$P_X(x) = {}^{1000}C_x \, (0.005)^x (0.995)^{1000-x}$$

$$P(X \le 2) = P_X(0) + P_X(1) + P_X(2)$$

$$= \square$$ دقیق

Check

$p = 0.005 < 0.1$

$np = 1000 * 0.005 = 5 < 10$

$X \sim Poisson(\lambda = np = 5)$

$$P_X(x) = e^{-5} \frac{5^x}{x!}$$

$$P(X \le 2) = P_X(0) + P_X(1) + P_X(2)$$

$$\overset{\text{تقریبی}}{=} \Delta \simeq \square$$

Example

The number of telephone calls per minute at a switchboard has a Poisson distribution with rate 3 calls/min. λ

What is the probability of 5 calls per minute?

What is the probability of 8 calls during a 2-min interval?

X: no. of calls /min $\sim Pois(3)$

$$P_X(x) = e^{-3}\frac{3^x}{x!}$$

$$P(X=5) = P_X(5) = e^{-3}\frac{3^5}{5!} = \checkmark$$

Y: no. of calls/2min $\sim Pois(6)$

$$P_Y(y) = e^{-6}\frac{6^y}{y!}$$

$$P(Y=8) = P_Y(8) = \checkmark$$

λ	duration (min)
3	1
$\boxed{?!}$	2

$$\lambda_{new} = \frac{2*3}{1} = 6$$

Example

The number of misprints on a document page typed by a secretary has a Poisson distribution with average 2.

What is the probability that a page contains more than 3 misprints? Exercise

Example

Suppose that the number of flaws along a thin copper wire follows a Poisson distribution with a mean 2.3 flaws per mm.

Determine the probability of at least one flaw in 2 mm of wire?

λ	mm
2.3	1
$\boxed{\checkmark}$	2

Example

The number of particles of contamination that occur on an optical disk has Poisson distribution with average 0.1 particles per cm^2, the area of the disk is 100 cm^2.

(a) Find the probability that at most two particles exist in the area of an optical disk?

(b) If you purchase 5 disks, what is the probability that exactly 3 of them have at most 2 particles on their surface?

$$\lambda = \frac{100 \times 0.1}{1} = 10$$

λ	cm^2
0.1	1
$\square$	100

X: no. of part. / disk $\sim Pois(10)$

$$P_X(x) = e^{-10}\frac{10^x}{x!}$$

$$P(X \le 2) = P_X(0) + P_X(1) + P_X(2)$$

$$= \square$$

Y: no. of disks with at most 2 part. $\sim Bin(5, \square)$

$$= {}^5C_y\, \square^y (1-\square)^{5-y}$$

$$P(Y=3) = P_Y(3) = \checkmark$$

Example

The number of accidents per hour at a certain intersection is assumed to have Poisson. The rate of this distribution varies with time. During the interval from 8 AM to 9 AM, the rate equal 1 accident/hr, while it equals 3 accidents/hr during the interval from 9 AM to 10 AM.

What is the probability that during the interval from 8 AM to 10 AM there is at most 2 accidents?

$$\underbrace{\lambda_1 \qquad \lambda_2}_{\lambda_1+\lambda_2}$$

Summary

Distribution	X	x	PMF	E(X)	V(X)
Bernoulli	Success/fail $X \sim Ber(p)$	$x = 1,0$	$P_X(x) = \begin{cases} p, x = 1 \\ 1-p, x = 0 \end{cases}$	p	$p(1-p)$
Binomial	No. of successes during n trials $X \sim Bin(n, p)$	$x = 0,1,2,\ldots,n$	$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$
Geometric	No. of trials until first success $X \sim Geo(p)$	$x = 1,2,3,\ldots$	$P_X(x) = p(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson	- Limit of Binomial. - No. of success per specific interval. $X \sim Pois(\lambda)$	$x = 0,1,2,\ldots$	$P_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}$	λ	λ