Examples of discrete probability distributions:

The binomial and Poisson distributions

Binomial Probability Distribution

- A fixed number of observations (trials), n
  - e.g., 15 tosses of a coin; 20 patients; 1000 people surveyed
- A binary random variable
  - e.g., head or tail in each toss of a coin; defective or not defective light bulb
  - Generally called “success” and “failure”
  - Probability of success is p, probability of failure is 1 – p
- Constant probability for each observation
  - e.g., Probability of getting a tail is the same each time we toss the coin

Binomial example

Take the example of 5 coin tosses. What's the probability that you flip exactly 3 heads in 5 coin tosses?

Binomial distribution

Solution:

One way to get exactly 3 heads: HHHTT

What's the probability of this exact arrangement?

P(heads)xP(heads) xP(heads)xP(tails)xP(tails)
$=(1/2)^3 \; x \; (1/2)^2$

Another way to get exactly 3 heads: THHHT

Probability of this exact outcome = $(1/2)^1 \; x \; (1/2)^3 \; x \; (1/2)^1 = (1/2)^3 \; x \; (1/2)^2$

Binomial distribution

In fact, $(1/2)^3 \, x \, (1/2)^2$ is the probability of each unique outcome that has exactly 3 heads and 2 tails.

So, the overall probability of 3 heads and 2 tails is:

$(1/2)^3 \, x \, (1/2)^2 + (1/2)^3 \, x \, (1/2)^2 + (1/2)^3 \, x \, (1/2)^2 + \ldots\ldots$ for as many unique arrangements as there are—but how many are there??

$\binom{5}{3}$ ways to arrange 3 heads in 5 trials

Outcome	Probability
THHHT	$(1/2)^3 \ x \ (1/2)^2$
HHHTT	$(1/2)^3 \ x \ (1/2)^2$
TTHHH	$(1/2)^3 \ x \ (1/2)^2$
HTTHH	$(1/2)^3 \ x \ (1/2)^2$
HHTTH	$(1/2)^3 \ x \ (1/2)^2$
THTHH	$(1/2)^3 \ x \ (1/2)^2$
HTHTH	$(1/2)^3 \ x \ (1/2)^2$
HHTHT	$(1/2)^3 \ x \ (1/2)^2$
THHTH	$(1/2)^3 \ x \ (1/2)^2$
HTHHT	$(1/2)^3 \ x \ (1/2)^2$

10 arrangements $x \ (1/2)^3 \ x \ (1/2)^2$

The probability of each unique outcome (note: they are all equal)

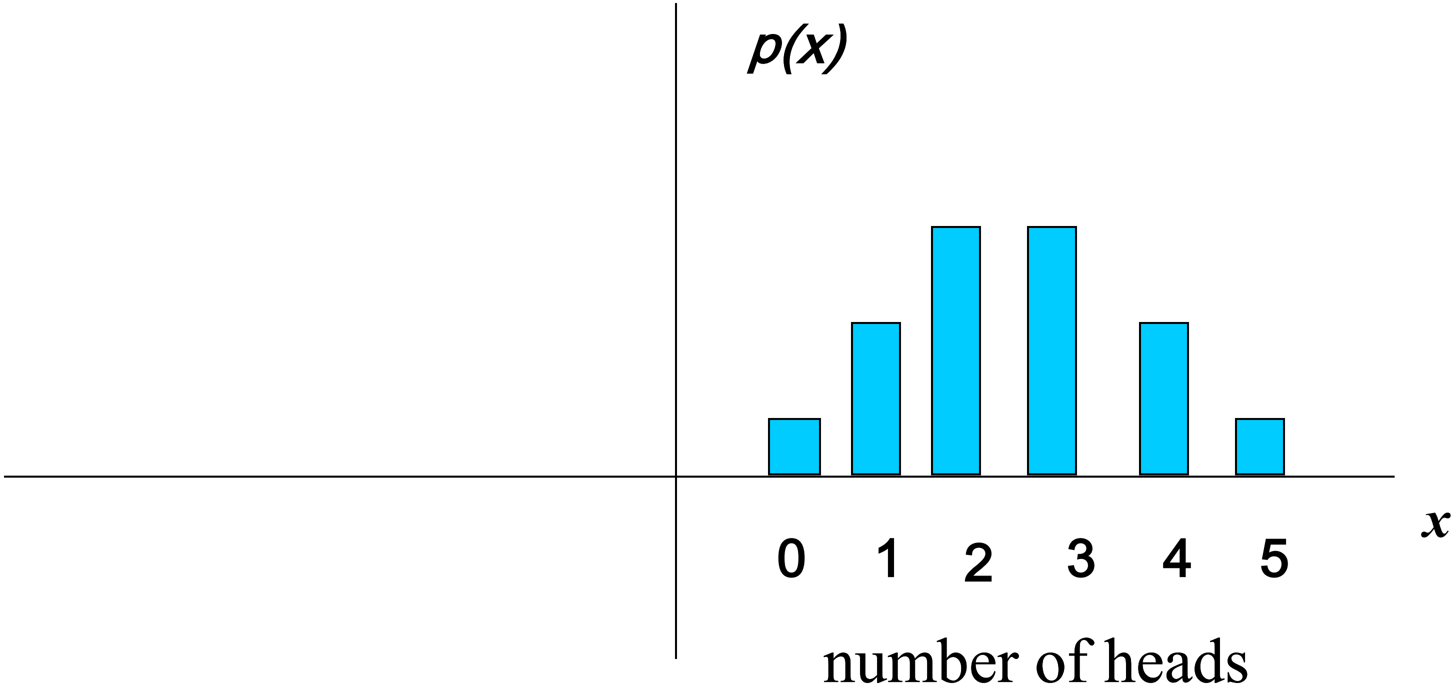
$_5C_3 = 5!/3!2! = 10$

$\therefore$ **P(3 heads and 2 tails) =** $\binom{5}{3}$ ***x P(heads)^3 x P(tails)^2 =***

10 x (½)^{5}=31.25%

Binomial distribution function:

X= the number of heads tossed in 5 coin tosses



Example 2

As voters exit the polls, you ask a representative random sample of 6 voters if they voted for proposition 100. If the true percentage of voters who vote for the proposition is 55.1%, what is the probability that, *in your sample,* exactly 2 voted for the proposition and 4 did not?

Solution:

Outcome		Probability
YYNNNN		$= (.551)^2 x (.449)^4$
NYYNNN	$(.449)^1 x (.551)^2 x (.449)^3$	$= (.551)^2 x (.449)^4$
NNYYNN	$(.449)^2 x (.551)^2 x (.449)^2$	$= (.551)^2 x (.449)^4$
NNNYYN	$(.449)^3 x (.551)^2 x (.449)^1$	$= (.551)^2 x (.449)^4$
NNNNYY	$(.449)^4 x (.551)^2$	$= (.551)^2 x (.449)^4$
.		
.		

$\binom{6}{2}$ ways to arrange 2 Obama votes among 6 voters

15 arrangements $x\,(.551)^2 x\,(.449)^4$

$\therefore$ P(2 yes votes exactly) = $\binom{6}{2} x\,(.551)^2 x\,(.449)^4 = 18.5\%$

Binomial distribution, generally

Note the general pattern emerging → if you have only two possible outcomes (call them 1/0 or yes/no or success/failure) in n independent trials, then the probability of exactly X "successes"=

n = number of trials

$$\binom{n}{X} p^{X}(1-p)^{n-X}$$

X = # successes out of n trials

p = probability of success

$1\text{-}p$ = probability of failure

Definitions: Binomial

- **Binomial:** Suppose that n independent experiments, or trials, are performed, where n is a fixed number, and that each experiment results in a "success" with probability p and a "failure" with probability $1-p$. The total number of successes, X, is a binomial random variable with parameters n and p.
- We write: ***X ~ Bin (n, p)*** *{reads: "X is distributed binomially with parameters n and p}*
- And the probability that *X=r* (i.e., that there are <u>exactly</u> *r* successes) is:

$$\boxed{P(X=r)=\binom{n}{r}p^r(1-p)^{n-r}}$$

Definitions: Bernouilli

Bernouilli trial: If there is only 1 trial with probability of success p and probability of failure 1-p, this is called a Bernouilli distribution. (special case of the binomial with n=1)

Probability of success: $$P(X=1)=\binom{1}{1}p^{1}(1-p)^{1-1}=p$$

Probability of failure: $$P(X=0)=\binom{1}{0}p^{0}(1-p)^{1-0}=1-p$$

Binomial distribution: example

- If I toss a coin 20 times, what's the probability of getting exactly 10 heads?

$$\binom{20}{10}(.5)^{10}(.5)^{10}=.176$$

Binomial distribution: example

- If I toss a coin 20 times, what's the probability of getting of getting 2 or fewer heads?

$$\binom{20}{0}(.5)^0(.5)^{20} = \frac{20!}{20!0!}(.5)^{20} = 9.5x10^{-7} +$$

$$\binom{20}{1}(.5)^1(.5)^{19} = \frac{20!}{19!1!}(.5)^{20} = 20x9.5x10^{-7} = 1.9x10^{-5} +$$

$$\binom{20}{2}(.5)^2(.5)^{18} = \frac{20!}{18!2!}(.5)^{20} = 190x9.5x10^{-7} = 1.8x10^{-4}$$

$$= 1.8x10^{-4}$$

**All probability distributions are characterized by an expected value and a variance:

<u>If</u> X follows a binomial distribution with parameters n and p: $\boldsymbol{X \sim Bin\ (n, p)}$

<u>Then:</u>

$\mu_x = E(X) = np$

$\sigma_x^2 = Var\ (X) = np(1-p)$

$\sigma_x = SD\ (X) = \sqrt{np(1-p)}$

Note: the variance will always lie between

0*N-.25 *N

p(1-p) reaches maximum at p=.5

P(1-p)=.25

Characteristics of Bernouilli distribution

For Bernouilli ($n=1$)

$E(X) = p$

$Var(X) = p(1-p)$

Variance Proof (optional!)

For Y~Bernouilli (p)

$$\begin{cases} Y=1 \text{ if yes} \\ Y=0 \text{ if no} \end{cases}$$

$$Var(Y) = E(Y^2) - E(Y)^2$$

$$= [1^2 p + 0^2(1-p)] - [1p + 0(1-p)]^2$$

$$= p - p^2$$

$$= p(1-p)$$

For X~Bin (N,p)

$$X = \sum_{i=1}^{n} Y_{Bernouilli}; Var(Y) = p(1-p)$$

$$= Var(X) = Var(\sum_{i=1}^{n} Y) = \sum_{i=1}^{n} Var(Y) = np(1-p)$$

Recall coin toss example

- X= number of heads in 100 tosses of a coin
- $X \sim Bin\ (100, .5)$
- $E(x) = 100*.5=50$
- $Var(X) = 100*.5*.5 = 25$
- $SD(X) = 5$

Things that follow a binomial distribution…

Cohort study (or cross-sectional):

- The number of exposed individuals in your sample that develop the disease
- The number of unexposed individuals in your sample that develop the disease

Case-control study:

- The number of cases that have had the exposure
- The number of controls that have had the exposure

Practice problems

- 1. You are performing a cohort study. If the probability of developing disease in the exposed group is .05 for the study duration, then if you sample (randomly) 500 exposed people, how many do you expect to develop the disease? Give a margin of error (+/- 1 standard deviation) for your estimate.

- 2. What's the probability that **at most** 10 exposed people develop the disease?

Answer

1. You are performing a cohort study. If the probability of developing disease in the exposed group is .05 for the study duration, then if you sample (randomly) 500 exposed people, how many do you expect to develop the disease? Give a margin of error (+/- 1 standard deviation) for your estimate.

$$X \sim \text{binomial}\ (500, .05)$$

$$E(X) = 500\ (.05) = 25$$

$$\text{Var}(X) = 500\ (.05)\ (.95) = 23.75$$

$$\text{StdDev}(X) = \text{square root}\ (23.75) = 4.87$$

$$\therefore 25 \pm 4.87$$

Answer

2. What's the probability that **<u>at most</u>** 10 exposed subjects develop the disease?

This is asking for a CUMULATIVE PROBABILITY: the probability of 0 getting the disease or 1 or 2 or 3 or 4 or up to 10.

$P(X \leq 10) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + \ldots + P(X=10) =$

$$\boxed{\binom{500}{0}(.05)^0(.95)^{500} + \binom{500}{1}(.05)^1(.95)^{499} + \binom{500}{2}(.05)^2(.95)^{498} + \ldots + \binom{500}{10}(.05)^{10}(.95)^{490} < .01}$$

(we'll learn how to approximate this long sum next week)

A brief distraction: Pascal's Triangle Trick

You'll rarely calculate the binomial by hand. However, it is good to know how to …

Pascal's Triangle Trick for calculating binomial coefficients

Recall from math in your past that Pascal's Triangle is used to get the coefficients for binomial expansion…

For example, to expand: $(p + q)^5$

The powers follow a set pattern: $p^5 + p^4q^1 + p^3q^2 + p^2q^3 + p^1q^4 + q^5$

But what are the coefficients?

- Use Pascal's Magic Triangle…

Pascal's Triangle

Edges are all 1's

To get the coefficient for expanding to the 5^{th} power, use the row that starts with 5.

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1

Add the two numbers in the row above to get the number below, e.g.: 3+1=4; 5+10=15

$(p + q)^5 = \mathbf{1}p^5 + \mathbf{5}p^4q^1 + \mathbf{10}p^3q^2 + \mathbf{10}p^2q^3 + \mathbf{5}p^1q^4 + \mathbf{1}q^5$

Same coefficients for X~Bin(5,p)

For example, X=# heads in 5 coin tosses:

X	P(X)
0	$\binom{5}{0}(.5)^0(.5)^5$
1	$\binom{5}{1}(.5)^1(.5)^4$
2	$\binom{5}{2}(.5)^2(.5)^3$
3	$\binom{5}{3}(.5)^3(.5)^2$
4	$\binom{5}{4}(.5)^4(.5)^1$
5	$\binom{5}{5}(.5)^5(.5)^0$

$\binom{5}{0}$=5!/0!5!=1 $\binom{5}{1}$=5!/1!4! = 5 $\binom{5}{2}$ = 5!/2!3!=5x4/2=10 $\binom{5}{3}$=5!/3!2!=10

$\binom{5}{4}$=5!/4!1!= 5 $\binom{5}{5}$=5!/5!1!=1 (Note the symmetry!)

X	P(X)
0	$\mathbf{1}(.5)^5$
1	$\mathbf{5}(.5)^5$
2	$\mathbf{10}(.5)^5$
3	$\mathbf{10}(.5)^5$
4	$\mathbf{5}(.5)^5$
5	$\mathbf{1}(.5)^5$
	$32(.5)^5 = 1.0$

From line 5 of Pascal's triangle!

Relationship between binomial probability distribution and binomial expansion

If $p + q = 1$ (which is the case if they are binomial probabilities)

then: $(p + q)^5 = (1)^5 = 1$ or, equivalently:

$$1p^5 + 5p^4q^1 + 10p^3q^2 + 10p^2q^3 + 5p^1q^4 + 1q^5 = 1$$

(the probabilities sum to 1, making it a probability distribution!)

P(X=0) P(X=1) P(X=2) P(X=3) P(X=4) P(X=5)

Practice problems

If the probability of being a smoker among a group of cases with lung cancer is .6, what's the probability that in a group of 8 cases you have less than 2 smokers? More than 5?

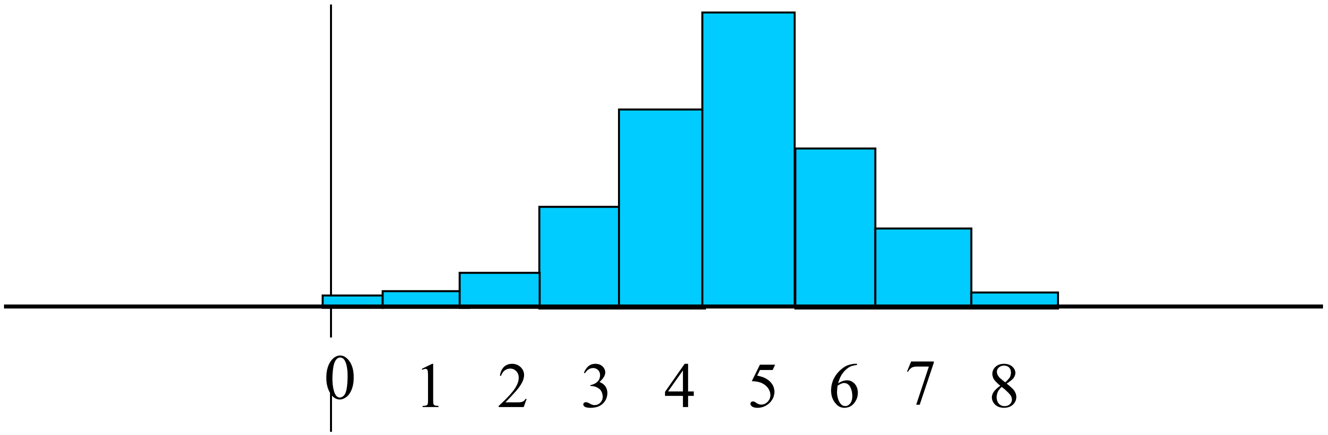
What are the expected value and variance of the number of smokers?

Answer

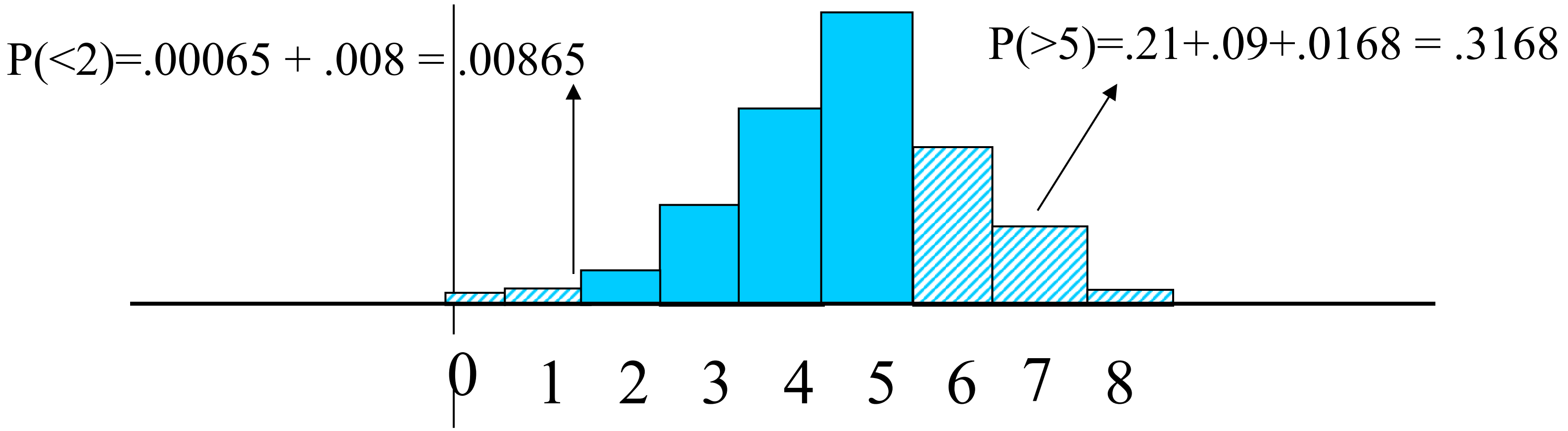
X	P(X)
0	$1(.4)^8 = .00065$
1	$8(.6)^1 (.4)^7 = .008$
2	$28(.6)^2 (.4)^6 = .04$
3	$56(.6)^3 (.4)^5 = .12$
4	$70(.6)^4 (.4)^4 = .23$
5	$56(.6)^5 (.4)^3 = .28$
6	$28(.6)^6 (.4)^2 = .21$
7	$8(.6)^7 (.4)^1 = .090$
8	$1(.6)^8 = .0168$

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1

Answer, continued



Answer, continued



E(X) = 8 (.6) = 4.8
Var(X) = 8 (.6) (.4) =1.92
StdDev(X) = 1.38

Practice problem

If Stanford tickets in the medical center 'A' lot approximately twice a week (2/5 weekdays), if you want to park in the 'A' lot twice a week for the year, are you financially better off buying a parking sticker (which costs $726 for the year) or parking illegally (tickets are $35 each)?

Answer

If Stanford tickets in the medical center 'A' lot approximately twice a week (2/5 weekdays), if you want to park in the 'A' lot twice a week for the year, are you financially better off buying a parking sticker (which costs \$726 for the year) or parking illegally (tickets are \$35 each)?

Use Binomial→

Let X be a random variable that is the number of tickets you receive in a year.

Assuming 2 weeks vacation, there are 50x2 days (twice a week for 50 weeks) you'll be parking illegally. p=.40 is the chance of receiving a ticket on a given day:

X~bin (100, .40)

E(X) = 100x.40 = 40 tickets expected (with std dev of about 5)

40 x \$35 = \$1400 in tickets (+/- \$200); better to buy the sticker!

Multinomial distribution
(beyond the scope of this course)

The multinomial is a generalization of the binomial. It is used when there are more than 2 possible outcomes (for ordinal or nominal, rather than binary, random variables).

- Instead of partitioning n trials into 2 outcomes (yes with probability p / no with probability $1-p$), you are partitioning n trials into 3 or more outcomes (with probabilities: $p_1, p_2, p_3, ..$)
  - General formula for 3 outcomes:

$$P(D = x, R = y, G = z) = \frac{n!}{x!\,y!\,z!} p_D^x p_R^y (1 - p_D - p_R)^z$$

Multinomial example

Specific Example: if you are randomly choosing 8 people from an audience that contains 50% democrats, 30% republicans, and 20% green party, what's the probability of choosing exactly 4 democrats, 3 republicans, and 1 green party member?

$$P(D=4, R=3, G=1) = \frac{8!}{4!\,3!\,1!}(.5)^4(.3)^3(.2)^1$$

You can see that it gets hard to calculate very fast! The multinomial has many uses in genetics where a person may have 1 of many possible alleles (that occur with certain probabilities in a given population) at a gene locus.

Introduction to the **Poisson Distribution**

- Poisson distribution is for counts—if events happen at a constant rate over time, the Poisson distribution gives the probability of X number of events occurring in time T.

Poisson Mean and Variance

- Mean

$$\boxed{\mu = \lambda}$$

- Variance and Standard Deviation

$$\boxed{\sigma^2 = \lambda}$$

$$\boxed{\sigma = \sqrt{\lambda}}$$

For a Poisson random variable, the variance and mean are the same!

where λ = expected number of hits in a given time period

Poisson Distribution, example

The Poisson distribution models counts, such as the number of new cases of SARS that occur in women in New England next month.

The distribution tells you the probability of all possible numbers of new cases, from 0 to infinity.

If X= # of new cases next month and $X \sim$ Poisson (λ), then the probability that $X=k$ (a particular count) is:

$$\boxed{p(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}}$$

Example

- For example, if new cases of West Nile Virus in New England are occurring at a rate of about 2 per month, then these are the probabilities that: 0,1, 2, 3, 4, 5, 6, to 1000 to 1 million to… cases will occur in New England in the next month:

Poisson Probability table

X	P(X)
0	$\frac{2^0 e^{-2}}{0!}$ =.135
1	$\frac{2^1 e^{-2}}{1!}$ =.27
2	$\frac{2^2 e^{-2}}{2!}$ =.27
3	$\frac{2^3 e^{-2}}{3!}$ =.18
4	=.09
5	
…	…

Example: Poisson distribution

Suppose that a rare disease has an incidence of 1 in 1000 person-years. Assuming that members of the population are affected independently, find the probability of k cases in a population of 10,000 (followed over 1 year) for k=0,1,2.

The expected value (mean) =λ = .001*10,000 = 10
10 new cases expected in this population per year→

$$P(X=0)=\frac{(10)^0 e^{-(10)}}{0!}=.0000454$$

$$P(X=1)=\frac{(10)^1 e^{-(10)}}{1!}=.000454$$

$$P(X=2)=\frac{(10)^2 e^{-(10)}}{2!}=.00227$$

more on Poisson…

"Poisson Process" (rates)

Note that the Poisson parameter λ can be given as the mean number of events that occur in a defined time period OR, equivalently, λ can be given as a rate, such as λ=2/month (2 events per 1 month) that must be multiplied by t=time (called a "Poisson Process") →

X ~ Poisson (λ)

$$P(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

E(X) = λt

Var(X) = λt

Example

For example, if new cases of West Nile in New England are occurring at a rate of about 2 per month, then what's the probability that exactly 4 cases will occur in the next 3 months?

X ~ Poisson (λ=2/month)

$$\mathrm{P(X = 4 \text{ in } 3 \text{ months})} = \frac{(2*3)^4 e^{-(2*3)}}{4!} = \frac{6^4 e^{-(6)}}{4!} = 13.4\%$$

Exactly 6 cases?

$$\mathrm{P(X = 6 \text{ in } 3 \text{ months})} = \frac{(2*3)^6 e^{-(2*3)}}{6!} = \frac{6^6 e^{-(6)}}{6!} = 16\%$$

Practice problems

1a. If calls to your cell phone are a Poisson process with a constant rate $\lambda=2$ calls per hour, what's the probability that, if you forget to turn your phone off in a 1.5 hour movie, your phone rings during that time?

1b. How many phone calls do you expect to get during the movie?

Answer

1a. If calls to your cell phone are a Poisson process with a constant rate λ=2 calls per hour, what's the probability that, if you forget to turn your phone off in a 1.5 hour movie, your phone rings during that time?

X ~ Poisson (λ=2 calls/hour)

P(X≥1)=1 – P(X=0)

$$P(X=0)=\frac{(2*1.5)^0 e^{-2(1.5)}}{0!}\frac{(3)^0 e^{-3}}{0!}=e^{-3}=.05$$

$\therefore$P(X≥1)=1 – .05 = 95% chance

1b. How many phone calls do you expect to get during the movie?

E(X) = λt = 2(1.5) = 3