STAT230

PROBABILITY

Chapter 5 (b)

UNIVERSITY OF WATERLOO

The Negative Binomial Distribution

Based on an experiment satisfying the following conditions:

1. The experiment consists of a sequence of independent trials.
2. Each trial can result in either a success (S) or a failure (F).
3. The probability of success p is constant from trial to trial
4. The experiment continues until a total of k successes have been observed, where k is a specified positive integer.

Binomial vs. Negative Binomial distributions

Binomial

- We know we have n trials, but we DON'T know the number of successes that we will obtain.

Negative Binomial

We know the number of successes, k, but we DON'T know the number of trials that we need to obtain k successes.

UNIVERSITY OF WATERLOO

The random variable of interest is

X = the number of failures obtained before the k^{th} successes occur.

X is called a Negative Binomial variable, X ~ NB (k, p)

Note

The number of success is fixed, while the number of trials is random.

Example

If a fair coin is tossed until we get our 5^{th} head, the number of tails we obtain has a Negative Binomial distribution with k = 5 and p = 1/2 .

UNIVERSITY OF WATERLOO

- If X is a Negative Binomial r.v with p.f. NB(k, p), then

$$f(x) = P(X = x) = \binom{x+k-1}{x} p^k (1-p)^x \qquad x = 0, 1, 2, \ldots.$$

In all there will be x+k trials (x F's and k S's) and the <u>last trial must be success.</u>

In the first x + k -1 trials we therefore need x failures and (k -1) successes, in any order.

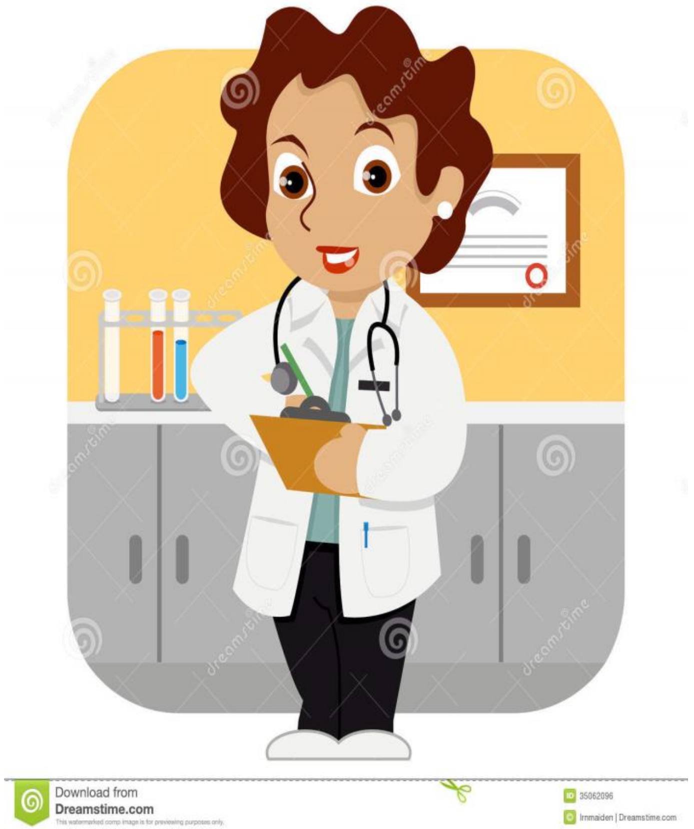
X = the total number of trials needed to get the k^{th} success.

Example

Asking for the probability of getting 3 tails before the 5^{th} head is exactly the same as asking for a total of 8 tosses in order to get the 5^{th} head.

Example

A Pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural child birth regimen.



Let

p = P(a randomly selected couple agree to participate).

Let $p = 0.2$.

(a) What is the probability that a total of 15 couples must be asked to get 5 who agree to participate?

(What is the probability that 10 F's occur before the 5th S)

$$P(X = x) = \binom{x + k - 1}{x} p^k (1 - p)^x \qquad x = 0,1,2, \ldots$$

k= 5 p = 0.2 x = 10

$$P(X=10) = \binom{10+5-1}{10} (0.2)^5 (0.8)^{10}$$

$$= 0.034$$

UNIVERSITY OF WATERLOO

(b) What is the probability that at most 10 F's are observed (asked) before the 5th S (at most 15 couples are asked).

$$P(X \leq 10) = \sum_{x=0}^{10} \binom{x+5-1}{x} (0.2)^5 (0.8)^x$$

Example

The fraction of a large population that has a specific blood type T is 0.08 (8%). For blood donation purposes it is necessary to find 5 people with type T blood. If randomly selected individuals from the population are tested one after another, then what is the probability that y persons have to be tested to get 5 type T persons.

- Think of a type T person as a success (S) and a non-type T as an F.
- let X = number of non-type T persons in order to get 5 S's.
- Let Y = number of persons who have to be tested.
- $X \sim NB(k = 5; p = 0.08)$
- We are actually asked here about $Y = X + 5$

$$P(X = x) = \binom{x+k-1}{x} p^k (1-p)^x \qquad x = 0,1,2, \ldots.$$

$p = 0.08$

$k=5$

$x = y-5$

$$P(Y=y) = \binom{y-5+5-1}{y-5} (0.08)^5 (0.92)^{y-5} \qquad y = 5, 6, 7, \ldots.$$

(b) What is the probability that over 80 people have to be tested?

$$P(Y > 80) = P(X+5 > 80)$$

$$= P(X > 80-5)$$

$$= P(X > 75)$$

$$= 1 - P(X \leq 75)$$

$$= 1 - \sum_{x=0}^{75} f(x)$$

Geometric Random Variable

A special case of NB (x; 1, p) when k =1.

Suppose that independent trials, each having a probability p, of being a success, are performed until a success occurs. If we let X equal the number of failures obtained before the **first success** then

$$X \sim \textit{Geometric}(p)$$

$$f(x) = P(X = x) = p(1-p)^x \qquad x = 0, 1, 2, \ldots$$

$$0 < p < 1$$

UNIVERSITY OF WATERLOO

Example

The probability you win a lottery prize in any given week is a constant p. The number of weeks **before you win a prize for the first time has a Geometric distribution.**

UNIVERSITY OF WATERLOO

Example

A representative from the National Football League's Marketing Division randomly selects people on a random street in Toronto City until he finds a person who attended the last home football game. Let *p*, the probability that he succeeds in finding such a person, equal 0.20. *And, let X denote the number of people he selects before he finds his first success*. What is the probability that the marketing representative must select 3 people before he finds one who attended the last home football game?

To find the desired probability, we need to find $P(X = 3)$, which can be determined readily using the p.f. of a geometric random variable with

$p = 0.20,$

$1-p = 0.80,$

$x = 3$

$$f(x) = P(X = x) = p\ (1-p)^{x}$$

$$P(X = 3) = (0.80)^3\ \ 0.20 = 0.1024$$

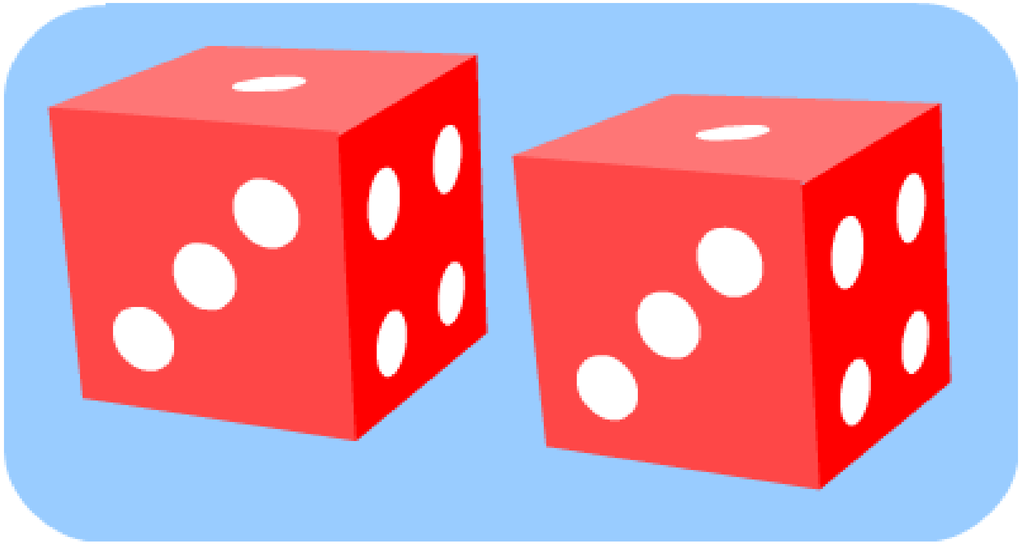
Example

In a particular game you may only begin if you roll a double to start.

Find the probability that:

(a) You start on your first go?

(b) You need a total of 4 attempts before you start?



(a) $X \sim$ *Geometric* (1/6) as probability of double = 1/6

Starting on your first go requires rolling a double on your first attempt.

$f(x) = P(X = x) = p\,(1-p)^x$

$P(X = 0) = 1/6$

(b) You need a total of 4 attempts to start, requires you to fail for the first 3 attempts.

$f(x) = P(X = x) = p\ (1-p)^{x}$

$P(X = 3) = (1/6)(5/6)^3$

$= 0.0964$

An Extra Examples

Example (1)

A shipment of 25 car headlights contains 20 which are defective. You choose from this shipment without replacement until you have 3 which are defective.

(a) What is the probability it takes exactly 6 items?

(b) Instead, you select 8 items what is the probability that you have exactly 5 defective.

(a)

20 are defective (D).

5 are not defective (ND).

Total of 6 items (3 D + 3 ND).

P(6 trials required) = P(2 D in the first 5 items drawn, then D last)

$$= \frac{\binom{20}{2}\binom{5}{3}}{\binom{25}{5}} \quad \frac{20 - 2}{25 - 5}$$

(b) Let X = number of defective selected.

X ~ Hypergeometric distribution

$$f(x) = P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \qquad x = 0, 1, \ldots, n$$

N=25, n=8, r = 20, x=5

$$P(X = 5) = \frac{\binom{20}{5}\binom{5}{3}}{\binom{25}{8}}$$

UNIVERSITY OF WATERLOO

Example (2)

x	**-1**	**0**	**1**	**2**
$f(x)=P(X=x)$	2/6	1/6	2/6	1/6
$F(x) =$ $P(X \leq x)$??				

UNIVERSITY OF WATERLOO

(a) Find the c.d.f?

x	-1	0	1	2
$f(x)=P(X=x)$	2/6	1/6	2/6	1/6
$F(x) = P(X \leq x)$	2/6	3/6	5/6	6/6

(b) $P(-3 \leq X < 2.5)$

$$= F(2) = 1$$

OR

$P(X= \{-3 \text{ or } -1 \text{ or } 0, \text{ or } 1, \text{ or } 2\}) = f(-3) + f(-1) + f(0) + f(1) + f(2)$

(c) $P(0.5 < X \leq 2) =$

$$F(2) - F(0.5) = F(2) - F(0) = 6/6 - 3/6 = 3/6$$

OR

$$f(1) + f(2) = 2/6 + 1/6 = 3/6$$

(d) $P\,(X=2 \cap 1< X \leq 2)$

$$= f(2) = 1/6$$

(e) $P\,(X=2 \mid 1< X \leq 2) =$

$$\frac{P\,(X=2 \cap 1< X \leq 2)}{P(1< X \leq 2)}$$

$P(1< X \leq 2) = F(2) - F(1) = 6/6 - 5/6 = 1/6$

$P\,(X=2 \mid 1< X \leq 2) = \dfrac{1/6}{1/6}$

UNIVERSITY OF WATERLOO

Example (3)

A student is getting ready to take an important oral examination and is concerned about the possibility of having an "on" day or an "off" day. He figures that if he has an on day, then each of his examiners will pass him, independently of one another, with probability 0.8, whereas if he has an off day, this probability will be reduced to 0.4. Suppose that the student will pass the examination if a majority of the examiners pass him. I the student believes that he is twice as likely to have an off day as he is to have an on day, should he request an examination with 3 examiners or with 5 examiners?

UNIVERSITY OF WATERLOO

$$P(\text{Pass exam}) = P(\text{Pass} \cap \text{"on day"}) + P(\text{Pass} \cap \text{"off day"})$$

$$= P(\text{Pass} \mid \text{"on day"})\,P(\text{"on day"}) + P(\text{Pass} \mid \text{"off day"}) \times P(\text{"off day"})$$

Case 1 if 3 examiners.

$$= \left[\binom{3}{2}(0{\cdot}8)^2(0{\cdot}2)^1 + \binom{3}{3}(0{\cdot}8)^3(0{\cdot}2)^0\right] \times \tfrac{1}{3}$$

$$+ \left[\binom{3}{2}(0{\cdot}4)^2(0{\cdot}6)^1 + \binom{3}{3}(0{\cdot}4)^3(0{\cdot}6)^0\right] \times \tfrac{2}{3}$$

$$= 0{\cdot}29867 + 0{\cdot}23467 = 0{\cdot}5333$$

Case 2 if 5 examiners

$$= \left[\binom{5}{3}(0.8)^3(0.2)^2 + \binom{5}{4}(0.8)^4(0.2)^1 + \binom{5}{5}(0.8)^5(0.2)^0\right] \times \frac{1}{3}$$

$$+ \left[\binom{5}{3}(0.4)^3(0.6)^2 + \binom{5}{4}(0.4)^4(0.6) + \binom{5}{5}(0.4)^5(0.6)^0\right] \times \frac{2}{3}$$

$$= 0.314 + 0.2116 = 0.5256.$$

$\therefore$ the P(Pass exam) is greater if he has 3 examiners

Example (4)

Three times per day, a student flips a fair coin, and if it is Heads they drink a coffee. We say a day is a "no-coffee" day if they do not drink any coffees that day.

a) What is the probability that a day is a "no-coffee" day?

b) What is the probability that in a week (7 days), there are 2 "no-coffee" days?

c) What is the probability that it takes 30 days to have 6 "no-coffee" days?

a) Let X = # of Heads obtained when we flip a coin 3 times.

$X \sim Bin(n=3; p=0.5)$

$P(\text{No coffee day}) = P(\text{No Heads}) = P(X=0)$

$= \binom{3}{0}(0.5)^0(0.5)^3 = 1/8$

UNIVERSITY OF WATERLOO

(b)

$\therefore$ Let Y = # of "no coffee days" in a week i.e 7 days

$Y \sim Bin(n=7, p=1/8)$

$\therefore P(Y=2) = \binom{7}{2}(1/8)^2(7/8)^5 = 0.168$

UNIVERSITY OF WATERLOO

(c)

Z = # of days with "at least one coffee" (failures) before the 6^{th} day with "no coffee" (success).

$Z \sim NB(k=6; p=1/8)$ $\quad P(Z=24) = \binom{24+6-1}{24}(1/8)^6(7/8)^{24}$

$= 0.018$.

Example (5)

A purchaser of electrical components buys them in lots of size 10. It is his policy to inspect 3 components randomly from a lot and to accept the lot only if all 3 are nondefective. If 30 percent of the lots have 4 defective components and 70 percent have only 1, what proportion of lots does the purchaser reject?

Let,

A: denotes the event that the purchaser accepts a lot.

A^c: denotes the event that the purchaser rejects a lot.

$P(A) + P(A^c)=1$

P(A) = P(A and from the lot with 4 defectives) +
P(A and from the lot with 1 defective)

= P(A|lot has 4 defectives) (lot with 4 defectives) +
P(A|lot has 1 defective)(lot with 1 defective)

$$\frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}}(3/10) + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}}(7/10) = 54/100$$

$P(A^c)$= 1- 54/100 = 46 % of the lots are rejected.

Example (6)

Suppose X is a random variable having a Geometric distribution with probability function

$$f(x) = (1-p)^x \, p \qquad \text{for } x=0,1,2,\ldots$$

(a) Show that $P(X < x) = 1- (1- p)^x$ for $x=0,1,2,\ldots.$

(b) Show that $P(X \geq x+y \ \cap X \geq x) = P(X \geq x) \, P(Y \geq y)$

(a) $P(X < x) = 1 - P(X \geq x)$

$$P(X \geq x) = \sum_{y=x}^{\infty} (1-p)^y p$$

$$= p \sum_{y=x}^{\infty} (1-p)^y$$

$$= p \{ (1-p)^x + (1-p)^{x+1} + (1-p)^{x+2} + \dots$$

$$= p \{ (1-p)^x + (1-p)^x (1-p)^1 + (1-p)^x (1-p)^2 + (1-p)^x (1-p)^3 + \dots \}$$

Using the Geometric Series

$$a + ar + ar^2 + \dots = \frac{a}{1-r}$$

$$P(X \geq x) = p \frac{(1-p)^x}{1-(1-p)}$$

Using the Geometric Series

$$= (1-p)^x \qquad \text{for } x = 0, 1, 2, \ldots.$$

$$P(X < x) = 1-(1-p)^x \quad \text{for } x = 0, 1, 2, \ldots.$$

(b) Show that $P(X \geq x+y \cap X \geq x) = P(X \geq x)\,P(Y \geq y)$

$P(X \geq x+y \cap X \geq x) = P(X \geq x+y)$

Using the result from (a)

$$
\begin{aligned}
P(X \geq x+y \cap X \geq x) &= (1-p)^{x+y} \\
&= (1-p)^{x}(1-p)^{y} \\
&= P(X \geq x)\,P(Y \geq y)
\end{aligned}
$$

UNIVERSITY OF WATERLOO

Poisson Distribution

$$f(x) = P\,(X = x) = \frac{e^{-\mu}\ \mu^{x}}{x!}$$

for $x = 0, 1, 2, \ldots, \infty$

where:

X = represents the number of events of some type.

μ = expected (average) number of events ($\mu > 0$).

e = base of the natural logarithm system (2.71828..)

UNIVERSITY OF WATERLOO

The distribution was derived by the French mathematician Siméon Poisson in 1837.

http://1.bp.blogspot.com/-OJ8j43iurXE/UUdDRddWQaI/AAAAAAAAAFE/3keKfmcyNe0/s1600/Poisson_2.jpeg

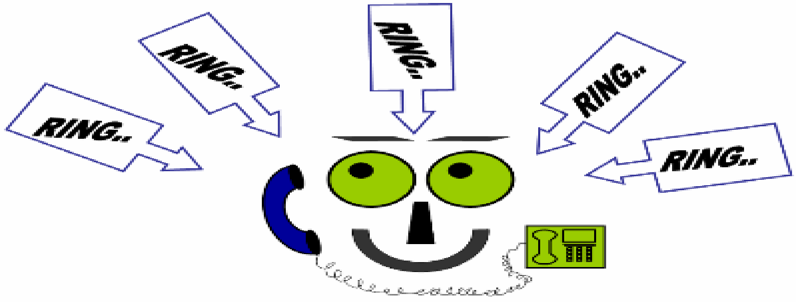
Examples

- The number of customers entering a post office on a given day.

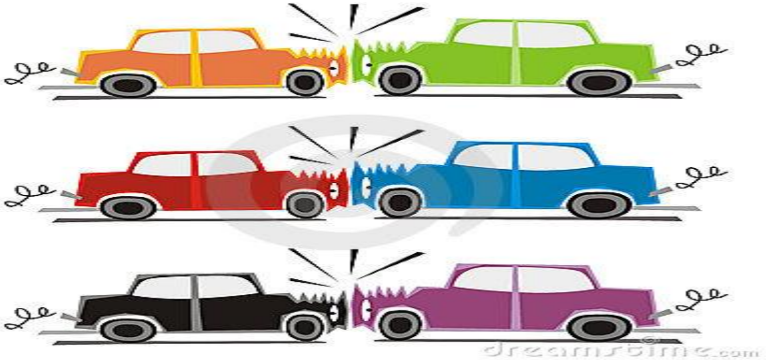
- The number of misprints on a page (or a group of pages) of a book.

- The number of people in a community who survive to age 100.

- The number of telephone calls at a business per hour.



- The number of accidents at an intersection (in a specific time period).



Checking that for the Poisson distribution $\sum_{\text{all x}} f(\mathrm{x}) = 1.$

Poisson Distribution from Binomial.

The Poisson random variable has a tremendous range of applications in diverse areas because it may be used as an approximation for a binomial random variable with parameters (n, p) when n is large and p is small enough so that np is of moderate.

UNIVERSITY OF WATERLOO

- Suppose that in the Binomial p.f Binomial$(x; n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\mu > 0$ ($\mu = np$)

Then Binomial $(x; n, p) \to p(x; \mu)$

According to this proposition, in any Binomial experiment in which n is large and p is small,

Binomial $(x; n, p) \approx p(x; \mu)$

Example

If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is 0.005 and errors are independent from page to page, what is the probability that

(a) One of its 400-page novels will contain exactly one page with errors?

(b) At most three pages with errors?

Let

S: denotes a page containing at least one error

F: denotes an error-free page,

(a)The number X of pages containing exactly one error is a Binomial r.v. with n = 400 and p = 0.005, so np=2.

$$P(X=1) = b(1;400,0.005) \approx p(1;2) = \frac{e^{-2}(2)^1}{1!} = 0.271$$

$$P(X \le 3) \approx \sum_{x=0}^{3} p(x;2) = \sum_{x=0}^{3} e^{-2}\frac{2^x}{x!} = 0.135+0.271+0.271+0.180 = 0.857$$

Poisson Distribution from the Poisson Process

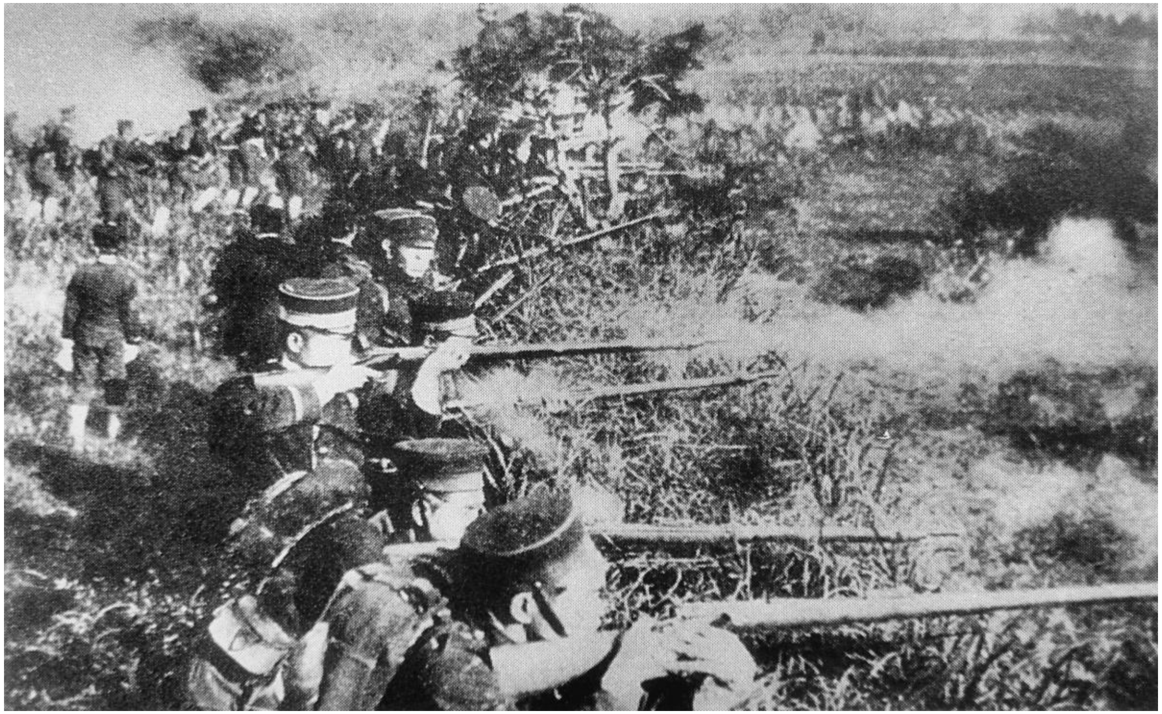
Another use of the Poisson probability distribution arises in situations where "events" occur at certain points in time or in space.

The number of earthquakes occurring per year.

http://www.sellcell.com/blog/wp-content/uploads/2014/05/earthquake.jpg

UNIVERSITY OF WATERLOO

The number of wars per year.



http://upload.wikimedia.org/wikipedia/commons/b/b2/Sino_Japanese_war_1894.jpg

Web site hits in an hour.

for some positive constant λ, the following assumptions hold :

1. *Independence*: The number of occurrences in non-overlapping intervals are independent. Whatever occurs in one interval (no-overlapping) has no effect on what will occur in other.

2. *Individuality*: for sufficiently short time periods of length Δt; the probability that 2 or more events occurring in the interval is equal to zero, that is, events occur singly not in clusters.

3. *<u>Homogeneity or Uniformity</u>*:

Events occur at a uniform or homogeneous rate λ over time.

Note:

In a Poisson process with rate of occurrence λ, the number of event occurrences X in a time interval of length t has a Poisson distribution with $\mu = \lambda t$.

Interpretation of μ and λ

- λ: is the rate of occurrence parameter for the events.

  It represents the average rate of occurrence of events per unit of time (or area or volume).

- $\lambda t = \mu$: is represents the average number of occurrences in t units of time.

Note

It is important to note that the value of λ depends on the units used to measure time.

Example

Suppose pulses arrive at the counter individually and independently at a constant rate with an average rate of six per minute. Find the probability that in a 0.5-min interval at least one pulse is received.

$$\lambda = 6$$

The number of pulse in 0.5-min interval has a Poisson distribution with parameter

$$\mu = \lambda t = 6(0.5) = 3.$$

Then with X = the number of pulses received in the 30-sec interval,

$P(X \geq 1) = 1 - P(X < 1)$

$$= 1 - p(X = 0) = 1 - \frac{e^{-3}(3)^0}{0!} = 0.950$$

UNIVERSITY OF WATERLOO

Example

Suppose that earthquakes recorded in Ontario each year follow a Poisson process with an average of 2 per year.

What is the probability that at least 3 earthquakes occur during the next 2- year period?

$$P(X \geq 3) = 1 - P(X < 3)$$

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

$$= 1 - e^{-4} - 4e^{-4} - (4^2 / 2)\, e^{-4}$$

UNIVERSITY OF WATERLOO

Example

Suppose that a cases of West Nile Virus in Canada occur individually and independently at a constant rate, has an incidence of 1 in 1000 person-years.

What is the probability that exactly one case of West Nile Virus occur during the next 1-year period in a population of 10,000?

$$\lambda = 1/1000 = 0.001$$

$$\mu = 0.001(10{,}000) = 10$$

10 new cases expected in this population per year

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

7

Distinguishing Poisson from Binomial and Other Distributions

- **Can we specify in advance the maximum value which X can take (n) ?**

- If we can, then the distribution is not Poisson.

- If there is no fixed upper limit, the distribution might be Poisson, but is certainly not Binomial or Hypergeometric.

UNIVERSITY OF WATERLOO

- **Does it make sense to ask how often the event did not occur?**

  - If it does make sense, the distribution is not Poisson.

  - If it does not make sense, the distribution might be Poisson.

Examples

(1) It does not make sense to ask how often a person did not hiccup during an hour. So the number of hiccups in an hour might have a Poisson distribution. It would certainly not be Binomial, Negative Binomial, or Hypergeometric.

(2) If a coin were tossed until the 3rd head occurs it does make sense to ask how often heads did not come up. So the distribution would not be Poisson.

Combining Other Models with the Poisson Process

A very large (essentially infinite) number of butterflies is released in a large field. They scatter randomly so that on average of 6 butterflies on a tree. Trees are all the same size (any tree has a volume of 1 unit).

(a) Find the probability a tree has > 3 butterflies on it.

(b) When 10 trees are picked at random, what is the probability 8 of these trees have > 3 butterflies on them?

(c) Trees are checked until 5 with > 3 butterflies are found. Let X be the total number of trees checked. Find the probability function, f(x).

(d) Find the probability a tree with > 3 butterflies on it has exactly 6.

(e) On 2 trees there are a total of t butterflies.

Find the probability that x of these are on the first of these 2 trees.

Example

Flaws occur in glass for widescreen TVs according to a Poisson process with an average of 2 flaws per 1000 in^2 .

(1) List the three conditions for a general Poisson process and what they mean in words.

(1) Independence: events in non-overlapping intervals are independent.

(2) Individuality: the probability that 2 or more events occurring in the interval is equal to zero, that is, events occur singly not in clusters.

(3) Homogeneity or Uniformity: events occur at a uniform or homogeneous rate λ over time.

UNIVERSITY OF WATERLOO

(2) A 42-in widescreen TV needs 750 in^2 of glass. The TV is rejected at the factory if it has 2 or more flaws.

(i) Find the probability a TV is rejected.

$$\mu = \lambda t = \frac{2}{1000} \times 750 = 1.5 \text{ per } 750 \text{ in}^2$$

$$P(X \geq 2) = 1 - P(X < 2)$$
$$= 1 - [P(X=0) + P(X=1)]$$
$$= 1 - e^{-1.5}\left[\frac{(1.5)^0}{0!} + \frac{(1.5)^1}{1!}\right]$$
$$= 0.44.$$

(ii) What is the probability that after 30 such TVs are produced by the factory, exactly 8 are rejected?

Y = # rejected among 30.

$$Y \sim Bin(n = 30; p = 0.44)$$

$$P(Y = 8) = \binom{30}{8}(0.44)^8(1-0.44)^{22}$$

$$\simeq 0.024.$$

(3) A store orders 20 TVs (the store will only be sent satisfactory TVs). Find the probability function (pf) of Y, the total number of TVs **produced** to fill the order. Don't forget the range of Y.

$\therefore Y = W + 20$ where W = # of failures (TVs rejected) before the 20^{th} satisfactory one.

$\therefore W \sim NB(k=20, p=0.56)$

$$\underbrace{\_\ \_\ \_\ \_\ \cdots\ \_}_{w+19}\ \Big|\ \_$$

20^{th} satisfactory one

$$P(Y=y) = \binom{y-1}{y-20}(0.56)^{20}(0.44)^{y-20} \quad y = 20, 21, \ldots$$

(4) A 32-in widescreen TV only needs 450 in^2 of glass, and the factory produces these too. A quality-control expert examines one 32-in and one 42-in, and notes that there are t flaws in total. Given this, what is the probability that x of the flaws are on the larger TV? Show your working clearly.

$$P(x \text{ flaws on } 42\text{in} \mid t \text{ flaws on both})$$

$$= \frac{P(x \text{ flaws on } 42\text{in} \cap t \text{ flaws on both})}{P(t \text{ flaws on both})}$$

$Poi(1.5)$ $\quad$ $Poi\left(\frac{2}{1000} \times 450 = 0.9\right)$

$$= \frac{P(x \text{ flaws on } 42\text{in})\, P(t-x \text{ on } 32\text{in})}{P(t \text{ flaws on both})} \quad \cdots$$

Independence
of non-overlapping
interval.

$Poi(1.5 + 0.9 = 2.4)$

$$= \frac{\dfrac{\cancel{e^{-1.5}}(1.5)^x}{x!} \, \dfrac{\cancel{e^{-0.9}}(0.9)^{t-x}}{(t-x)!}}{\dfrac{\cancel{e^{-2.4}}(2.4)^t}{t!}}$$

$$= \frac{t!}{x!\,(t-x)!}\left(\frac{1.5}{2.4}\right)^x\left(\frac{0.9}{2.4}\right)^{t-x}.$$

$$= \binom{t}{x}(0.625)^x(0.375)^{t-x}.$$