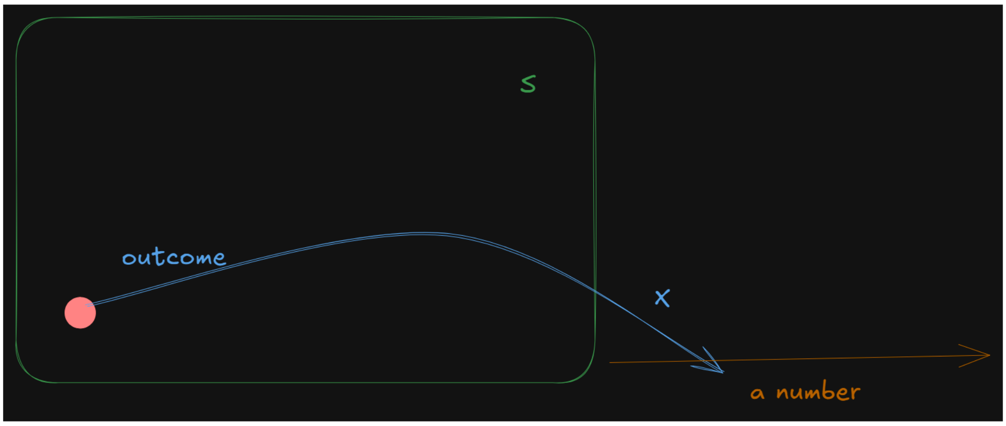
Chapter 3

3.1, 3.2 - Random Variables (rv)

Definition

For a given sample space S of an experiment, a rv X is a function that assigns every outcome in S to exactly 1 real number



Example 6

Left (L), Right (R), Straight (A)
Experiment terminates when car turns left
Let $X =$ the number of cars observed
Possible X values?

- The set of possible X values is $\{1, 2, 3, 4, ...\}$
- Outcomes/X values
  - $L; 21$
  - $AL, RL; 2$
  - $RAL, RRL, ARL, AAL; 3$
-

Notation

X, Y, Z, U, V are random variables
x, y, z, u, v are the values of the random variables

Bernoulli Random Variables

Definition

A rv that has only 2 possible values

Example

Flip a coin once

- $X = \{^{1, Head}_{0, Tail}$

Discrete Random Variable

Definition

A rv X is discrete if its set of possible values is finite or countable infinite

Probability Distribution

Definition

The probability distribution or the probability mass function (pmf) of a discrete rv X is the function $p_X(x) = P(X = x)$ for any real number x

Warning

$\sum_x p_X(x) = 1$

Example

Find the pmf:

x	2	5	-10	otherwise
$p_X(x)$	$\frac{3}{36} = \frac{1}{12}$	$\frac{18}{36} = \frac{1}{2}$	$\frac{15}{36} = \frac{5}{12}$	0

$p(x) = \{^{\frac{1}{12}, x=2}_{\dots}$

Example 12

1. $P(y \leq 50) = 1 - P(y > 50) = 1 - 0.17 = 0.83$
2. $P(y > 50) = 0.17$
3. $P(y \leq 49) = P(y \leq 50) - P(y = 50) = 0.83 - 0.17 = 0.66$
4. $P(y \leq 47 = 0.27)$

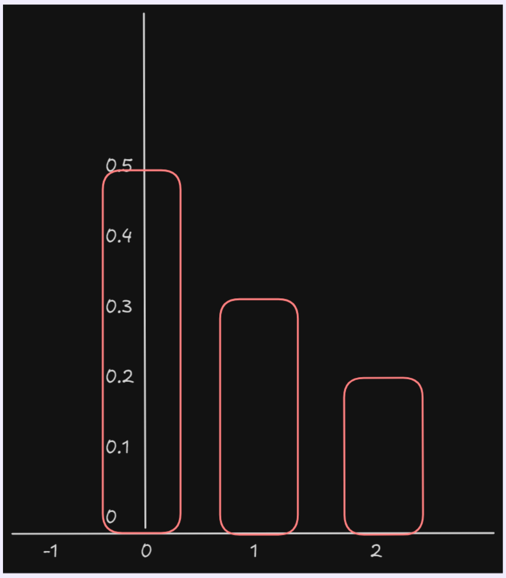
Example

The pmf of a rv X (probability distribution)

x	0	1	2	**otherwise**
$p(x)$	$.5$	$.3$	$.2$	0

The graph of the pmf

- Histogram $\Rightarrow$



The Cumulative Distribution Function (cdf)

Definition

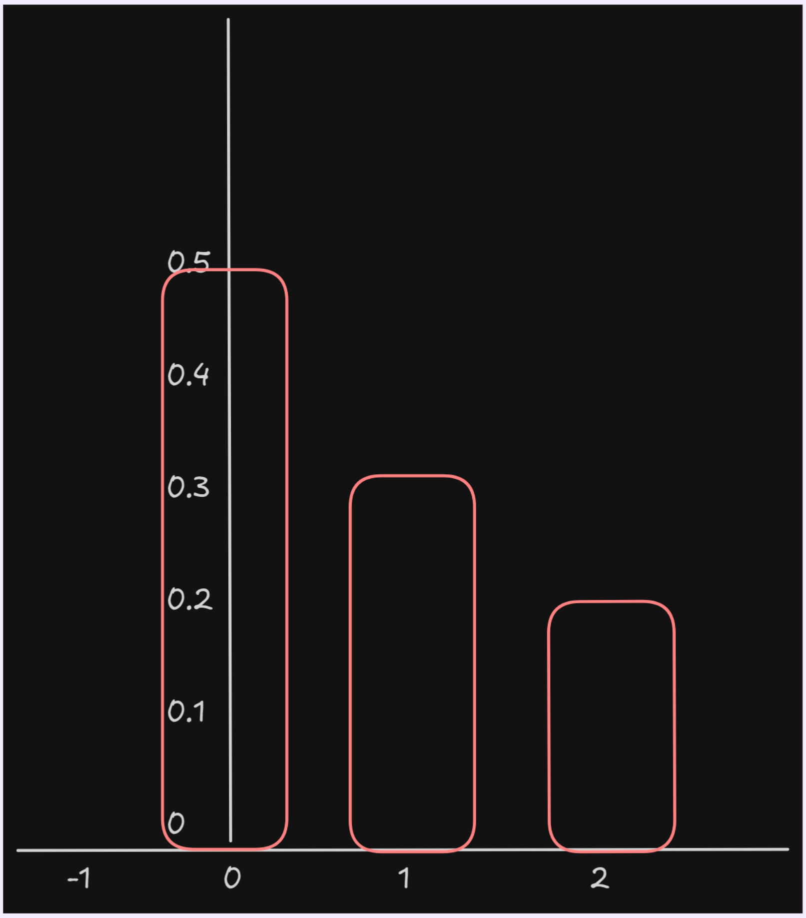
Let X be a discrete rv with the pmf $p(x)$ and the set of possible values D. The cdf of X, denoted by $F(x)$, is the function $F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$, where x is any real number

Example

x	**0**	**1**	**2**	**otherwise**
$p(x)$	.5	.3	.2	0

$F(0) = P(X \leq 0) = p(0) = .5$
$F(1) = P(X \leq 1) = p(0) + p(1) = .8$
$F(2) = P(X \leq 2) = p(0) + p(1) + p(2) = 1$

Example 18



Proposition

Let X be a discrete rv with the cdf $F(x)$. Then for any real numbers a and b with $a \leq b$,
$P(X \leq b) = F(b)$
$P(a \leq X \leq b) = F(b) - F(a^-)$ where a^- is the largest possible value of X that is less than a

Example 18 (continued)

$P(3 \leq M \leq 6) = p(3) + p(4) + p(5) + p(6)$
$P(3 \leq M \leq 6) \neq F(6) - F(3)$

$P(3 \leq M \leq 6) = F(6) - F(2) = 1 - \frac{1}{9} = \frac{8}{9}$
$P(3 < M \leq 6) = F(6) - F(3) = 1 - \frac{1}{4} = \frac{3}{4}$
$P(4 < M) = 1 - P(M \leq 3) = 1 - F(3) = 1 - \frac{1}{4} = \frac{3}{4}$

3.3 - Expected Values

Definition

Let X be a discrete rv with the pmf $p(x)$ and the set of possible values D
The expected value of X, also known as the mean of X is the following number
$E(x) = \mu_x = \sum\limits_{x \in D} xp(x)$ provided that the sum exists

Example Game

Roll a fair six faced die
Let $X =$ profit/loss per game

outcome	x	$p(x)$
$1, 2, 3, 4$	-3	$\frac{4}{6} = \frac{2}{3} = .667$
$5, 6$	$+6$	$\frac{1}{3} = .333$

$E(X) = -3(\frac{2}{3}) + 6(\frac{1}{3}) = 0$
Play the game 100 times

outcome	**count**	**relative freq.**	x
$1, 2, 3, 4$	68	$.68$	-3
$5, 6$	32	$.32$	$+6$

Loss/profit per game:
$\frac{-3(68)+6(32)}{100} = -3(.68) + 6(.32) = -.12$
$E(x)$ is the long-run average value of X when the experiment is performed repeatedly

Example 32

3 capacities: 16, 18, and 20 ft^3
Suppose X has pmf

x	**16**	**18**	**20**
$p(x)$	0.2	0.5	0.3

- Compute $E(x), E(x^2)$, and $V(X)$
  - $E(X)$
    - $16 * .2 + 18 * .5 + 20 * .3 = 18.2 ft^3$
    - The expected capacity of a freezer sold at a store
  - $E(X^2)$
    - $16^2 * .2 + 18^2 * .5 + 20^2 * .3 = 333.2 ft^6$
  - $V(X)$
    - $(16 - 18.2)^2(.2) + (18 - 18.2)^2(.5) + (20 - 18.2)^2(.3)$
    - $333.2 - 18.2^2 = 1.96 ft^6$
    - $\sigma = \sqrt{1.96} = 1.4 ft^3$
- Price $Pr(X) = 70X - 650$
  - $E(Pr(X)) = (16(70) - 650.2 + (18(70) - 650) * .5 + (20(70) - 650) * .3$
  - $E(Pr) = E(70X - 650) = 70E(X) - 650 = 70(18.2) - 650 = 624$
- Calculate variance in price
  - $V(70X - 650) = 70^2 V(X) = 70^2(1.96) = \$^2 9604$
  - $\sigma_{70X-650} = 70(1.4) = \$98$
- Rated capacity is X, actual capacity is $h(X) = X - .008X^2$. What is expected actual capacity
  - $E(X - .008X^2) = \sum_{x \in \{16,18,20\}} (x - .008x^2)p(x)$
  - $= \sum_{x \in \{16,18,20\}} xp(x) - .008 \sum_{x \in \{16,18,20\}} x^2 p(x)$
  - $E(X) - .008E(X^2) = 18.2 - .008(333.2) = 15.5 ft^3$

Proposition

If $h(x)$ is a function of rv X with the pmf $p_X(x)$ and the set of possible values D, then $E(h(X)) = \sum_{x \in D} h(x) p_X(x)$ provided that the sum exists

Proposition

Suppose a rv X has $E(X)$. Let a, b be any real numbers. Then, $E(aX + b) = aE(X) + b$

Variance

Definition

Suppose rv X with the pmf $p(x)$ and the set of possible values D has the mean μ. The variance of X, denoted by by $V(X)$ or σ^2, is $V(X) = \sigma^2 = \sum_{x \in D} (x - \mu)^2 p(x)$

$V(X) = E((X-\mu)^2)$ provided that the sum exists

Definition

The standard deviation of X is $\sigma = \sqrt{\sigma^2} = \sqrt{V(X)}$

Proposition

Let X be a rv with the mean μ and variance $V(X)$. For any real numbers a and b,
$V(aX+b) = a^2V(X)$, $\sigma_{aX+b} = |a|\sigma_X$

Shortcut Formula

Proposition

for $V(X)$:
$V(X) = E(X^2) - (E(X))^2$

Example 36

damage	0	1000	5000	10000
chance	.8	.1	.08	.02

$\$500$ deductible policy
Expected profit should be $\$100$
What premium should they charge?

- $Y =$ the payment made per customer
Y	0	500	4500	9500
$p(y)$	.8	.1	.08	.02
- $E(Y) = 50 + 360 + 190 = 600$
- Premium should be $600 + 100 = \$700$

3.4 - The Binomial Probability Distribution

Binomial Experiment

1. Sequence of n smaller experiments called trials
2. Each trials can result in one of the same two possible outcomes (dichotomous trials) which are denoted as S, success, or F, failure. The assignment of these variables are arbitrary
3. Trials are independent
4. The probability of success $P(S)$ is constant from trial to trial, we denote this probability as p
5. The probability of failure $P(F)$ is constant from trial to trial, we denote this probability as $q = 1 - p$
   The goal of such an experiment is to count the number of successes X

Examples

1. Toss a coin (fair or unfair) 10 times and count the number of heads
2. The National Statistics claims that 30% of Americans can raise one eyebrow at a time. Ask any ten people whether they can lift one eyebrow at a time and record the number of those who answer yes

Non Examples

- A deck of 20 cards contains 10 red and 10 black cards
    - 5 cards are randomly chosen and the number of red cards is recorded
- Roll a six faced die until a 6 appears

Binomial Distribution

Definition

Consider a binomial experiment with n trials where
$X =$the number of successes;
$p =$the probability of a success in a single trial.
Then X has a binomial distribution and we write

$$X \ Bin(n, p)$$

More notations
$b(x; n, p)$, the pmf of X
$B(x; n, p)$, the cdf of x

Example

A student is given a 3-questions multiple choice quiz. Each question has 4 possible answers of which only one is correct. Since the student has not been attending the class recently, he or she will be guessing on all 3 questions.
Let X be the number of correct answers. Construct the pmf of the RV

- $n = 3$
- Let $S = success =$the student picks the correction answer to a question
- Then $p = P(S) = \frac{1}{4} = 0.25$ and $1 - p = P(F) = \frac{3}{4} = 0.75$
- Thus $X\ Bin(3, 0.25)$
- The pmf of X

Outcome	x	$b(x; n, p) = b(x; 3, .25)$
FFF	0	$.75^3$
SFF, FSF, FFS	1	$3 * (.25)(.75^2)$
SSF,SFS,FSS	2	$3 * (.25^2)(.75)$
SSS	3	$.25^3$

x	**0**	**1**	**2**	**3**	**otherwise**
$b(x; 3, .25)$	$.75^3$	$3 * (.25)(.75^2)$	$3 * (.25^2)(.75)$	$.25^3$	0

$PMF : b(x; n, p) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$,
$\frac{S}{1}, \frac{F}{2}, \frac{S}{3} \dots \frac{S}{n}$
$x = 0, 1, 2, \dots, n$

Example 50

25% of calls involve fax, sample of 25 calls
Given: $n = 25, S =$"fax message",$p = .25$
$X =$the number of fax messages
$X\ Bin(n = 25, p = .25)$

- Probability that exactly 6 involve a fax message?
  - $P(X = 6) = b(6, 25, .25) = \binom{25}{6} * (.25)^6 (.75)^{19}$
  - $= \frac{25!}{6!19!} (.25^6)(.75^19) =$
  - $P(X = 6) = B(6; 25, .25) - B(5; 25, .25) = .561 - .378 = .183$
- Probability that at most 6 involve a fax message
  - $CDF : B(x; n, p) = P(X \leq x; n, p)$
  - $= \sum_{y=0}^{x} \binom{n}{y} p^y (1 - p)^(n - y), y = 1, 2, \dots, n$
  - $P(X \leq 6) = B(6; n = 25, p = .25) = .561$
- Probability that at least 6 involve a fax message

- $1 - P(X \leq 5) = 1 - B(5; 25, .25) = 1 - .378 = .622$
- Probability that more than 6 involve fax message
  - $P(X > 6) = 1 - P(X \leq 6) = 1 - B(6, 25, .25) = 1 - .561 = .439$
- Probability that between 4 and 9 involve a fax message
  - $P(4 \leq X \leq 9) = B(9, 25, .25) - B(3, 25, .25) = .929 - .096 = .833$

Proposition

If $X\ Bin(n, p)$, then $E(X) = \mu = np$, $V(X) = np(1-p), \sigma = \sqrt{np(1-p)}$

Refer to the previous exercise

- Given: $X\ Bin(25, .25)$
- What is the expected number of calls among the 25 that involve a fax message?
  - $E(X) = 25(.25) = 6.25$
- What is the standard deviation of the number among the 25 calls that involve a fax message?
  - $V(X) = np(1-p) = 6.25(.75) \Rightarrow \sigma = \sqrt{6.25(.75)} = 2.165$
- What is the probability that the number of calls among the 25 that involve a fax transmission exceeds the expected number by more than 2 standard deviations?
  - $P(X > 6.25 + 2(2.165)) = P(X > 10.58)$
  - $= 1 - P(X \leq 10) = 1 - B(10, 25, .25) = 1 - .97 = .03$

Example 60

$\$1.00$ for passenger cars and $\$2.50$ for other vehicles, 60% are passenger cars, 25 vehicles cross the bridge, what is the resulting expected toll revenue?

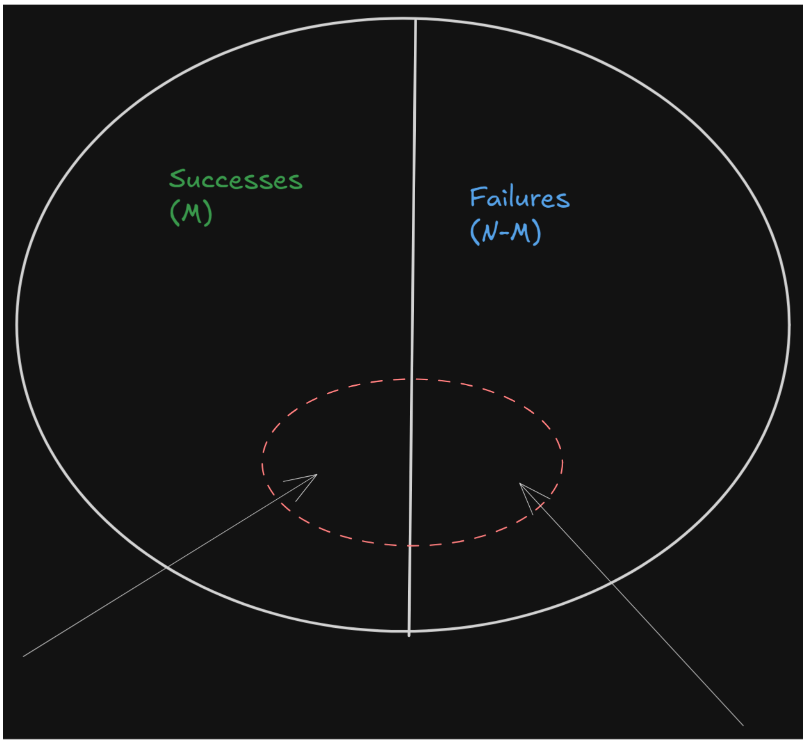
- X=the # of passenger cars, $X\ Bin(n = 25, p = .6)$
- $\$1 * X + \$2.50(25 - X) = X + 62.5 - 2.5X = 62.5 - 1.5X$
- $E(62.5 - 1.5X) = 62.5 - 1.5E(X) = 62.5 - 1.5(25 * .6) = 62.5 - 1.5(15) = 62.5 - 22.5$
- $V(62.5 - 1.5X) = (-1.5)^2 V(X) = 2.25(25 * .6 * .4) = 2.25(6) = \$^2 13.5$

3.5 - Hypergeometric and Negative Binomial Distributions

Hypergeometric Distribution

Assumptions

1. The population or set to be sampled consists of N individuals, objects, or elements (a *finite* population)
2. Each individual can be characterized as a success (S) or a failure (F), and there are M successes in the population
3. A sample of n individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen



Let X be the number of successes in the sample.
Then we say that X has a hypergeometric distribution with parameters
n, sample size
M, the number of S's in a population
N, the size of the population
Notation: $X \ Hyp(n, M, N)$
The pmf of X

- $h(x; n, M, N) = P(X = x) = \frac{\binom{M}{x} * \binom{N-M}{n-x}}{\binom{N}{n}}$
- for x and integer satisfying max $(0, n - N + M) \leq x \leq min(n, M)$
- Note: $n - N + M = n - (N - M)$
- $\binom{M}{x} = C_{x,M}$ is the number of ways to choose x successes for the sample
- …

The cdf of X

- $H(x; n, M, N) = P(X \leq x)$
- has no closed form
- has no table

Example

A deck of 30 cards contain 10 red and 20 black cards
5 cards are randomly chosen and the number of red cards is recorded
The parameters are

- $n = 5$
- $N = 10 + 20 = 30$
- $M = 10$
  The rv X is the number of red cards in the selection
- $P(X = 3) = \frac{C_{3,10} * C_{(5-3),(30-10)}}{C_{5,30}}$
- $= \frac{C_{3,10} * C_{2,20}}{C_{5,30}}$

Example 70

Two sections - first with 20 students, second with 30. Term project assigned, professor randomly ordered them before grading, consider the first 15 graded projects
Given

- $sec1 = F, sec2 = S$
- $N = 50, n = 15, X =$ the number of projects from $sec2$
- $X\ hyp(n = 15, M = 30, N = 50)$
- Probability that exactly 10 projects are from $sec2$
  - $P(X = 10) = h(10, 15, 30, 50)$
  - $\frac{\binom{30}{10} * \binom{20}{5}}{\binom{50}{15}} = .2070$
- Probability that at least 10 projects are from the $sec2$
  - $P(X = 10) = 1 - P(X \leq 9)$
  - $P(X \geq 10) = h(10) + h(11) + \ldots + h(15) = .3798$
- Probability that at least 10 project are from same section
  - P(at least 10 from $sec2$) + P(at least 10 from $sec1$)
  - $P(X \geq 10) + P(15 - X \geq 10)$
  - $= P(X \geq 10) + P(X \leq 5) = .3798 + \sum_{x=0}^{5} h(x; n = 15, M = 30, N = 50)$
  - $= .3798 + .0140 = .3938$
- Mean and std dev of the number among these 15 that are from second section
  - $E(X) = 15 * \frac{30}{50} = 9$
  - $V(X) = \frac{50-15}{50-1} * 15 * \frac{30}{50} * (1 - \frac{30}{50}) = 2.5714$
  - $\sigma = 1.60$
- Mean and std dev of the number of projects not among these first 15 that are from the second section
  - $E(30 - X) = 30 - E(X) = 30 - 9 = 21$

- $V(30 - X) = (-1)^2 V(X) = 2.5714$
- $\sigma = 1.60$

Mean and Variance of X

Proposition

If $X\ hyp(n, M, N)$, then

$$E(X) = n * \frac{M}{N}$$

$$V(X) = (\frac{N-n}{N-1}) * n * \frac{M}{N} * (1 - \frac{M}{N})$$

Rule of Thumb

if $\frac{n}{N} \leq 0.5$ or $N \geq 20n$, then

$$h(x; n, M, N) \approx b(x; n, \frac{M}{N})$$

assuming that $\frac{M}{N}$ is not too close to 0 or 1

Negative Binomial Distribution

Assumption

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 is constant from trial to trial, so $P(S)$ on trial $t) = p$ for $i = 1, 2, 3$
4. The experiment continues (trials are performed) until a total of r successes have been observed, where r is a specified positive integer
   Let X be the number of failures that precede the rth success
   Then X is a negative binomial RV and it has a negative binomial distribution with parameters r and p
   $X \sim nb(r, p)$
   The pmf is $nb(x; r, p)$
   $nb(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x, x = 0, 1, 2, 3, \ldots$

   Example
   Flip an unfair coin until exactly two heads are obtained
   …

Example

10% of engines are defective. Randomly selected one at a time and tested.

- Probability that first non-defective engine will be found on the second trial?
  - $S =$"non-defective engine selected"
  - $p = .9$
  - Let $X =$the number of defective engines before the first non-defective one
  - $r = 1$
  - $X \sim nb(r = 1, p = .9)$
  - $P(X = 1) = (.1)(.9) = .09$
- Probability that the third non-defective engine will be found on the fifth trial?
  $Y =$*the number of defective engines selected before the third non-defective one*
  $r = 3$
  $Y \sim nb(r = 3, p = .9)$
  $P(Y = 2) = \binom{4}{2}(.9)^3(.1)^2$

…

3.6 - Poisson Distribution, Poisson Process, and Approximation of Binomial Distribution

Poisson Distributions

Assumptions

Consider time period $[0, t]$. The interval is divided into subintervals of width Δt.
Suppose Δt is small. Then,

1. The probability that one event occurs during time period of length Δt is approximately directly proportional to Δt. That is $P($one event$= (\alpha\Delta t)$
2. The probability that two or more events occur during the time period of length Δt is approximately 0
3. The number of events observed during any interval of length Δt does not depend on the number of occurrences on the other subintervals
   Let $K =$ the number of events during a time interval of length t.

Proposition

$K \sim Poi(\mu = \alpha t)$,

$$P(K = k) = \frac{e^{-\alpha t} * (\alpha t)^K}{k!}$$

where $k = 0, 1, 2, ...$
α is the *rate of* the process. So, if $t = 1(unit)$, then

$$E(K) = \mu = \alpha * t$$

$\alpha =$ the expected number of events in a unit of time

Example 92

Automobiles arrive at a vehicle equipment inspection station according to a Poisson process with rate $\alpha = 10$ per hour. Suppose that with probability $.5$ an arriving vehicle will have no equipment violations.
$K =$ the number of cars during t hr

1. What is the probability that exactly ten arrive during the hour and all ten have no violations?
   - $t = 1 \Rightarrow K \sim Poi(\mu = \alpha * 1) \Rightarrow \mu = 10$
   - Find $P(K = 10$ and no violations) and $P($no violations | a car arrived at the station$) = .5$
   - $P(K = 10 \cap$ all 10 no violations)
   - $= P(K = 10) * P($no violations | $K = 10)$
   - $= \frac{e^{-10} * 10^{10}}{10!} * (.5)^{10} = .00122$

Example 86

Organisms are present in ballast water discharged from a ship according to a Poisson process with a concentration of $10\ organisms/m^3$
Given: $t =$ the volume of dischage in m^3 concentration = rate = $\alpha = 10\ org./m^3$
$K =$ the number of organisms in tm^3

1. What is the probability that one cubic meter of discharge contains at least 8 organisms?
   - $t = 1\ m^3 \Rightarrow k \sim Poi(\mu = 10)$
   - $P(k \geq 8) = 1 - F(7; \mu = 10) = 1 - .22 = \boxed{.78}$
2. What is the probability that the number of organisms in $1.5m^3$ of discharge exceeds its mean value by more than one standard deviation?
   - $t = 1.5 \Rightarrow k \sim Poi(\mu = 15)$
   - $E(K) = V(K) = \mu = 15 \Rightarrow \sigma_K = \sqrt{15} = 3.87$
   - $P(K > 18.87) = 1 - F(18; 15) = 1 - .819 = \boxed{.181}$