	Binomial	Geometric	Hyper Geo	Neg Bino
Random Variable	# of success	# of trails b4 1st Success	# of success	# of trails b4 x Success
Type of Trails	Independent	Independent	dependent	Independent
Type of outcomes	Success / Failure	Success / Failure	Success / Failure	Success / Failure

	Discrete uniform	Hypergeometric	Binomial	Negative binomial	Geometric	Poisson
When to use	- Fixed # of outcomes $\{a, a+1, \dots, b\}$ - Equally likely	- Fixed # trials - Trials are without replacement	- Fixed # trials - Indep. trials - $P(S)$ is constant	- Trials until we have k S's - Trials independent	- Trials until we have 1 S - Trials independent	- Events occur in time/space ∘ Independence ∘ Individuality ∘ Homogeneity.
Parameters	a: lowest outcome. b: largest outcome.	N: # objects r: # S's among N objects n: # trials (# draws)	n: # trials p: $P(S)$	k: # successes to stop p: $P(S)$	p: $P(S)$	$\mu = \lambda t$; λ = rate of events t = lenght of time/space.
r.v.	X = an outcome from $\{a, a+1, \dots, b\}$ $X \sim Unif(a, b)$	X = # of S's in trials $X \sim Hyp(N, r, n)$	X = # of S's in trials $X \sim Bin(n, p)$	X = # F's in trials $X \sim NB(k, p)$	X = # F's in trials $X \sim Geo(p)$	X_t = # events during period of length t $X_t \sim Poi(\mu = \lambda t)$
p.f.	$\frac{1}{b-a+1}$, $x \in \{a, a+1, \dots, b\}$	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$, max/min range.	$\binom{n}{x} p^x (1-p)^{n-x}$, $x \in \{0, 1, \dots, n\}$	$\binom{x+k-1}{k-1} p^k (1-p)^x$, $x \in \{0, 1, \dots\}$	$p(1-p)^x$, $x \in \{0, 1, \dots\}$	$\frac{e^{-\mu}\mu^x}{x!}$, $x \in \{0, 1, \dots\}$
CDF	$\frac{[x]-a+1}{b-a+1}$	No closed form	No closed form.	No closed form	$1-(1-p)^{[x]+1}$	—
Approx	×	N, r large relative to n $X \sim Bin(n, r/N)$	n large & p small $X \sim Poi(\mu = np)$	×	×	×

Name	Probability Function
Discrete Uniform	$f(x) = \frac{1}{b-a+1}$ for $x = a, a+1, a+2, \dots, b;\ b > a$
Hypergeometric	$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ for $x = \max(0, n-(N-r)), \dots, \min(n, r)$
Binomial	$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, 2, \dots, n;\ 0 < p < 1$
Negative Binomial	$f(x) = \binom{x+k-1}{x} p^k (1-p)^x$ for $x = 0, 1, 2, \dots;\ 0 < p < 1$
Geometric	$f(x) = p(1-p)^x$ for $x = 0, 1, 2, \dots;\ 0 < p < 1$
Poisson	$f(x) = \frac{e^{-\mu}\mu^x}{x!}$ for $x = 0, 1, 2, \dots;\ \mu > 0$

Three Possible Definitions of Probability

Definition (Classical Definition of Probability) Classical Definition of Probability

Let S be the set of all possible distinct outcomes of a random experiment. Then the probability of an *event* is

$$\frac{\text{Number of ways the event can occur}}{\text{Total number of outcomes in } S}$$

Example

The probability of rolling an even number in one throw of a fair die is $\frac{3}{6}$.

{2,4,6} {1,2,3,4,5,6}

Photo courtesy of Don McLeish

Definition (Random Experiment) Random Experiment

When we repeat the experiment under *controlled conditions*, (repetitions are called **trials** of the experiment) different outcomes may occur.

Definition (Countable) Countable it could be infinite.

A set, S, is countable if the elements can be put in a 1-1 correspondence with the positive integers.

Definition (Relative Frequency Definition) Relative Frequency Definition

The probability of an event in an experiment is the (limiting) proportion or fraction of times the event occurs in a very long (theoretically infinite) series of (independent) repetitions of the experiment.

Example

If a coin is tossed 10000 times, and heads appears 4992 times. Therefore, the probability of heads equals 0.4992.

Definition (Subjective Probability) Subjective Probability

The probability of an event is a 'best guess' by a person making the statement of the chances that the event will happen. (e.g., a 30% chance of rain)

Definition (Sample Space) Sample space

A sample space, S, is the set of distinct outcomes for an experiment or process, with the property that in a single trial, one and only one of these outcomes occurs.

Discrete Sample Space

Definition (Discrete Sample Spaces) Discrete Sample Spaces

A **discrete sample space**, S, is one with a **finite** number of sample points or **countably many** sample points.

Definition (Probability Distribution on S) Probability Distribution

Let $S = \{a_1, a_2, a_3, \dots\}$ be a discrete sample space.

Let $P(a_1), P(a_2), P(a_3), \dots$ be a set of numbers associated with the sample points $a_1, a_2, a_3, \dots$ such that:

1. $0 \le P(a_i) \le 1$, $i = 1, 2, \dots$
2. $\sum_{i=1} P(a_i) = 1$. Their Sum is 1.

Then $P(a_i)$ is called a probability.

The set $\{P(a_i), i = 1, 2, \dots\}$ is called a probability distribution on S.

Definition (Discrete Probability Model) Discrete probability Model

A discrete sample space $S = \{a_1, a_2, \dots\}$ together with a probability distribution $\{P(a_i), i = 1, 2, \dots\}$ is referred to as a **discrete probability model**.

Definition (Event) Event

An event, A, defined on a discrete sample space, S, is a subset of S.

Definition (Simple Event) Simple Event

If the event $A \subset S$ consists of only one sample point then A is called a **simple event**.

Definition (Compound Event) Compound Event

If the event $A \subset S$ consists of two or more sample points then A is called a **compound event**. A is said to **occur** on a trial of the experiment if one of the simple events in A occurs.

Definition (Probability of an Event) Probability of an Event

Let S be a discrete sample space.

Let A be an event defined on S, i.e. $A \subset S$.

Then $P(A)$, the probability of the event A, is the sum of the probabilities corresponding to all the simple events that are in A, that is $P(A) = \sum_{a \in A} P(a)$.

Equal likely Outcomes

Definition ($P(A)$ for equi-probable outcomes) (P (A) for equi-probable outcomes)

When all simple events have the same probability, for any event $A \subset S$, $P(A) = \dfrac{\text{number of points in A}}{N}$.

Addition and Multiplication Rules

Definition (Addition Rule) Addition Rule

Suppose we can do job 1 in p ways and job 2 in q ways. Then we can do either job 1 **OR** job 2 (but not both), in $p + q$ ways.

□ □ □ □ □

or

◊ ◊ ◊

Definition (Multiplication Rule) Multiplication Rule

Suppose we can do job 1 in p ways and, for each of these ways, we can do job 2 in q ways. Then we can do both job 1 **AND** job 2, in $p \times q$ distinct ways.

□ □ □ □ □

and

◊ ◊ ◊

Another Combinatorial Symbol

Definition n choose k

$\binom{n}{k} = \frac{n^{(k)}}{k!}$ called 'n choose k'

Multinomial Coefficient

Definition

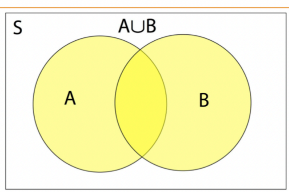
$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! n_2! \dots n_k!}$$

is the number of arrangements of n elements of k different types, there being n_1 of the first type, n_2 of the second type, $\dots$, n_k of the k'th type.

Set and Probability

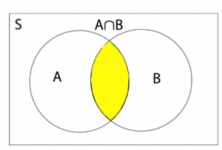
Definition (Union of Two Sets)

The union of A and B (written $A \cup B$) is the set of all points which are in either A or B or both.



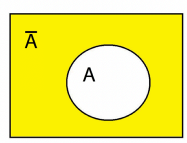
Definition (Intersection of Two Sets) Intersection

The intersection of A and B (written $A \cap B$) is the set of all points which are in both A and B.

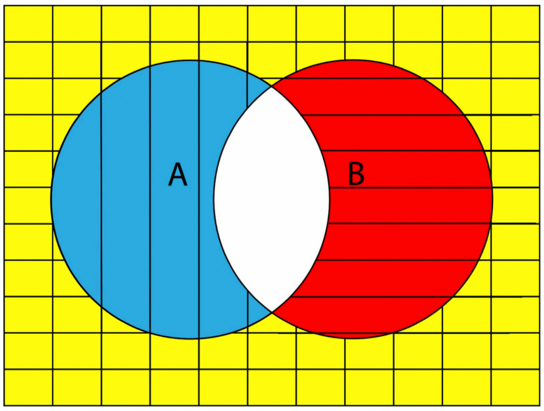


Definition (Complement of a Set) Complement

The complement of A (written $\bar{A}$) is the set of all points in S which are not in A.



De Morgan's Laws



a. The complement of the union is the intersection of the complements: $\overline{A \cup B} = \overline{A} \cap \overline{B}$

b. The complement of the intersection is the union of the complements: $\overline{A \cap B} = \overline{A} \cup \overline{B}$

More generally for k events,

a. $\overline{A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_k} = \overline{A_1} \cap \overline{A_2} \cap \overline{A_3} \cap \ldots \cap \overline{A_k}$

b. $\overline{A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_k} = \overline{A_1} \cup \overline{A_2} \cup \overline{A_3} \cup \ldots \cup \overline{A_k}$

Mutually Exclusive

Definition (Mutually Exclusive)

Two events A and B are **mutually exclusive** if $A \cap B = \emptyset$, where $\emptyset$ is the empty set.

The events $A_1, A_2, \ldots, A_k$ are called **(pairwise) mutually exclusive** if $A_i \cap A_j = \emptyset$ for all i and j with $i \neq j$.

Probability Set Function

Definition (Probability Set Function)

Suppose the function P associates a real value, $P(A)$, with each event A defined on a sample space S such that:

1. $0 \leq P(A)$ for every event A,
2. $P(S) = 1$,
3. If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events, then $P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$,

then P is called a probability set function and $P(A)$ is called the probability of A.

Properties of Probabilities

Theorem

Suppose S is a sample space for an experiment.

Let $A, B, A_1, A_2, \ldots$ be events defined on S.

Then:

1. $0 \leq P(A) \leq 1$.

$$\begin{aligned} P(A) = \sum_{a \in A} P(a) &\leq \sum_{\text{all a}} P(a) \\ &= P(S) \\ &= 1 \end{aligned}$$

2. **Probability of the Complement of an Event:**
$P(\overline{A}) = 1 - P(A)$ which implies $P(\emptyset) = 1 - P(S) = 0$.

$$\sum_{\text{all a}} P(a) = \sum_{a \in A} P(a) + \sum_{a \in \overline{A}} P(a)$$

$$1 = P(A) + P(\overline{A})$$

3. **Probability of a subset of an Event:**
If $A \subset B$, then $P(A) \leq P(B)$.

$$P(A) = \sum_{a \in A} P(a) \leq \sum_{a \in B} P(a) = P(B)$$

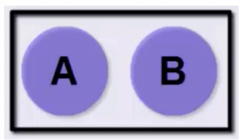
Theorem

Suppose S is a sample space for an experiment.

Let $A, B, A_1, A_2, \ldots$ be events defined on S.

Then:

4. **Probability of the Union of Two Mutually Exclusive Events:**
Let A and B be mutually exclusive events.
Then $P(A \cup B) = P(A) + P(B)$.



5. **Probability of the Union of Mutually Exclusive Events:**
Let $A_1, A_2, \ldots, A_n$ be mutually exclusive events.
Then $P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$.

$A_i \cap A_j = \emptyset,$
$i \neq j$

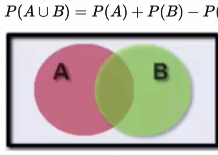
Theorem

Suppose S is a sample space for an experiment.

Let $A, B, A_1, A_2, \ldots$ be events defined on S.

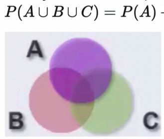
Then:

6. For any two events A and B,
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.



$$P(A) + P(B) = P(A \cup B) + P(A \cap B)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

7. For any three events A, B and C,
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$



Independent Event

Definition (two independent events)

Events A and B are **independent events** if and only if $P(A \cap B) = P(A)P(B)$. If they are not independent, we call the events **dependent**.

Independent events are such that the ratios

$$\frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(S)} = P(A)$$

N Independent Event

Definition (n independent events)

The events $A_1, A_2, \ldots, A_n$ are independent if and only if $P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$ for all sets $(i_1, i_2, \ldots, i_k)$ of distinct subscripts chosen from $(1, 2, \ldots, n)$.

If we really want independence among the n events, it is NOT sufficient to require that the events are pairwise independent, or that $P(A_i \cap A_j) = P(A_i)P(A_j)$ for all $i \neq j$, as we will show shortly with an example.

Independent Theorem

Theorem

A and B are independent events if and only if $\overline{A}$ and B are independent events. Similarly A and $\overline{B}$ or $\overline{A}$ and $\overline{B}$.

Conditional Probability

Definition (Conditional Probability) Definition: Conditional Probability

read as 'given'

The conditional probability of event A, given the event B occurs, is defined by $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ provided $\boxed{P(B) \neq 0}$.

Probability of A given B

Theorem Theorem: Conditional Probability of Independent Events should be the same as its own probability.

Suppose A and B are two events such that $P(A) > 0$ and $P(B) > 0$. A and B are independent events if and only if $P(A|B) = P(A)$ or, equivalently, if $P(B|A) = P(B)$.

Multiplication Rule / Product Rule

Theorem (Multiplication Rule or Product Rule) Theorem: Multiplication Rule / Product Rule

For any two events A and B,

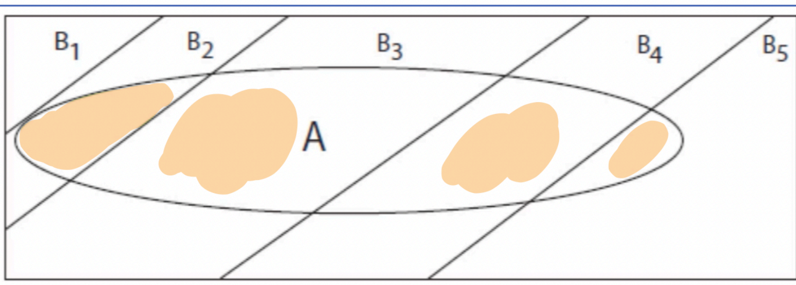
1. $P(A \cap B) = P(A|B)P(B)$, and
2. $P(A \cap B) = P(B|A)P(A)$

Law of Total Probability

Theorem (Law of Total Probability)

Let the sample space, S, be partitioned into k mutually exclusive sets $B_1, B_2, \ldots, B_k$ such that $P(B_i) > 0$, for $i = 1, 2, \ldots, k$ and $S = B_1 \cup B_2 \cup \cdots \cup B_k$.[1]

Then, for any event A, $P(A) = \sum_{i=1}^{k} P(A \cap B_i)$.



Law of Total Probability

Bayes' Theorem

Theorem (Bayes' Theorem)

Let the sample space, S, be partitioned into k mutually exclusive sets $B_1, B_2, \ldots, B_k$ such that $P(B_i) > 0$, for $i = 1, 2, \ldots, k$ and $S = B_1 \cup B_2 \cup \cdots \cup B_k$.
Then, for any event A, and for $j = 1, 2, \ldots, k$, we have

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

Random Variable

Definition Definiation: Random Varibale

A random variable (e.g. X) is a function that assigns a real number to each point in a sample space S.

Definition (Discrete Random Variable X)

The random variable, X, takes a finite or countably infinite number of distinct values (the range of X is countable).

Discrete: 0, 1, 2, ... etc. The positive or non-negative integers, or finite subset of these, it could also be 1/2, 1/3, 1/4, or a subset is countable in the sense that it can be put in 1 to 1 correspondence with positive integers.

Definition (Continuous Random Variable X) Definition: Continuous Random Variable X

X can take any value in a non-degenerate interval (range of X is not countable).

Continuous: Which can take any value in a non-degenerate interval.
Eg: it could be taken any value between 0 and 1, and its range is not a countable set.

Probability Function

Definition (probability function) Definition: Probability function

The **function** $f(x) = P(X = x)$, for all x in the set of possible values of X.

$f(x)$ has two properties:

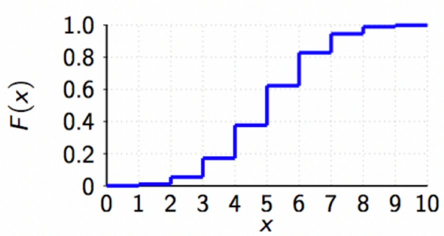
1. $f(x) \geq 0$ for all values of x
2. $\sum_{\text{all } x} f(x) = 1$ (sum of all probabilities is 1)

Cumulative Distribution Function / C.D.F

Definition (c.d.f.)

A cumulative distribution function (c.d.f.) is defined as the function $F(x) = P(X \leq x)$, for all real numbers x.

Properties of a c.d.f. $F(x)$



1. $F(x)$ is a non-decreasing function of x for all $x \in \mathbb{R}$
2. $0 \leq F(x) \leq 1$ for all $x \in \mathbb{R}$
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

Relation between $F(x)$ and $f(x)$

1. If a random variable, X, takes only non-negative integer values, then $F(x)$ is the probability of the values less than or equal to x.
2. $f(x) = F(x) - F(x-1)$ is the size of the **jump in** F at the point x, and
3. $F(x) = \sum_{z \leq x} f(z)$

Discrete Uniform

Theorem Discrete Uniform

Suppose X takes values $a, a+1, a+2, \ldots, b$ with all values being equally likely. Then X has a discrete uniform distribution, on $a, a+1, a+2, \ldots, b$. The probability function of the discrete uniform is

$$f(x) = \begin{cases} \dfrac{1}{b-a+1} & \text{for } x = a, a+1, \ldots, b \\ 0 & \text{otherwise} \end{cases}$$

Hypergeometric Distribution

Theorem

We have a collection of N objects which can be classified into two distinct types, r of type 1 and $N-r$ of type 2.

Pick a sample of $n < N$ objects at random **without replacement**.
Let X be the number of type 1 in the sample.
Then X has a hypergeometric distribution with **probability function**

$$f(x) = \frac{(\text{ways to choose } x \text{ type } 1) \times (\text{ways to choose } n-x \text{ type } 2)}{(\text{ways to choose sample of } n\ )} = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

The range for X is $max(0, n-N+r) \le x \le min(r, n)$.

Theorem

$$\sum_{\text{all } x} \binom{a}{x}\binom{b}{n-x} = \binom{a+b}{n}$$

Acceptance Sampling and the Hypergeometric Distribution

Definition

Acceptance sampling is a statistical method which is used for deciding whether a batch of items produced by a company is acceptable or not for distribution.

Example

Suppose a company sends out items in batches of 200. The company would like to be reasonably sure that the batch of 200 does not contain more than 10 (5%) defective items.

It is too costly to test all items and so the company uses acceptance sampling, i.e. selects n items for testing and decides, based on the observed number of defective items in the sample, whether or not to accept or reject the batch for distribution.

Example (continued)

The company decides that $n = 20$ items will be tested. The batch will be rejected if **more than** 5% of the sample is found to be defective.

1. What is the probability the batch is rejected if there are 10 defective items in the batch of 200?
   $X =$ Number of defective items in sample of 20.
   The batch rejected if $X > 1$.

$$\begin{aligned} P(X>1) &= 1 - P(X=0) - P(X=1) \\ &= 1 - \frac{\binom{190}{20}}{\binom{200}{20}} - \frac{\binom{190}{19}\binom{10}{1}}{\binom{200}{20}} \\ &= 0.263 \end{aligned}$$

2. What is the probability the batch is rejected if there are 20 defective items in the batch of 200? Batch rejected if $X > 1$.

$$\begin{aligned} P(X>1) &= 1 - P(X=0) - P(X=1) \\ &= 1 - \frac{\binom{180}{20}}{\binom{200}{20}} - \frac{\binom{180}{19}\binom{20}{1}}{\binom{200}{20}} \\ &= 0.622 \end{aligned}$$

Example (continued)

What if the company decides to test 40 items (i.e. $n = 40$). What would you conclude about whether one should test more items?

3. If there are 10 defective items in the batch of 200?
   $X = \text{Number of defective items in sample of } 40.$
   Batch rejected if $X > 2$.

$$\begin{aligned} P(X>2) &= 1 - P(X=0) - P(X=1) - P(X=2) \\ &= 1 - \frac{\binom{190}{40}}{\binom{200}{40}} - \frac{\binom{190}{39}\binom{10}{1}}{\binom{200}{40}} - \frac{\binom{190}{38}\binom{10}{2}}{\binom{200}{40}} \\ &= 0.321 \end{aligned}$$

4. If there are 20 defective items in the batch of 200?
   $X = \text{Number of defective items in sample of } 40.$
   Batch rejected if $X > 2$.

$$\begin{aligned} P(X>2) &= 1 - P(X=0) - P(X=1) - P(X=2) \\ &= 1 - \frac{\binom{180}{40}}{\binom{200}{40}} - \frac{\binom{180}{39}\binom{20}{1}}{\binom{200}{40}} - \frac{\binom{180}{38}\binom{20}{2}}{\binom{200}{40}} \\ &= 0.808 \end{aligned}$$

Higher Probability of rejecting batch with 10% defective if n = 40

Binomial Distribution

Definition

Suppose we have an experiment with 2 possible outcomes which, for convenience, we call Success (S) and Failure (F). Suppose also that $P(S) = p$. Repeat the experiment (called a **trial**). Such a sequence of independent trials are called **Bernoulli trials**. Bernoulli trials

Let the random variable X be the number of successes in n Bernoulli trials.

Key Assumptions:

1. The probability of success p must be **constant** over the n trials.
2. The n trials must be **independent**.

Theorem

The **probability function** (p.f.) of X is

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

Check that $\sum f(x) = 1$. If it is a real Binomial Distribution, it should follow.

Theorem (Binomial Theorem)

$$\sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} = (a+b)^n$$

Connection Between Binomial , Hypergeometric Distributions

	Binomial	Geometric	Hypergeometric
Random variables	# of success	# of trails before 1st success	# of success
Type of trails	Independent	Independent	dependent
Type of outcomes	success or failure	success or failure	success or failure

Negative Binomial Distribution

Suppose we have a sequence of Bernoulli trials with $P(S) = p$.
Let the random variable X be the number of failures (F's) observed before obtaining the k'th success (S).
What is the range of X?
Find $f(x) = P(X = x)$, the probability function (p.f.) of X.

Theorem

\# of success \# of failure

The p.f. of X is $f(x) = \binom{x+k-1}{x} p^k (1-p)^x$ for $x = 0, 1, 2, \ldots$

We write $X \sim NB(k, p)$.

Note: Alternatively, the Negative Binomial can be defined to count the total number of **trials*** needed to get the k^{th} success.

Binomial vs. Negative Binomial

Binomial:

1. The total **number of trials is specified** in advance, as n.
2. The **number of successes is unknown** and can only be determined after the experiment.

Negative Binomial:

1. The total **number of trials** is **not specified** in advance because we do not know how many trials will be needed, until after experiment.
2. The number of **successes** is **specified** to be k.

Geometric Distribution: Special Case of Negative Binomial

Geometric distribution is a special case of the Negative Binomial, where $k = 1$. Then the random variable, Y, is the waiting time or the number of failures until the **first** success.

Theorem

1 success   y failure be fore it

The p.f. of Y is $f(y) = P(Y = y) = p(1-p)^y$ for $y = 0, 1, \dots$

We write $Y \sim Geo(p)$.

Poisson Distribution: Approximating Binomial for large n, small p

Definition (Poisson Distribution)

Suppose that X represents the number of events of some type, occurring at a rate of $\mu > 0$. Then a random variable X has a Poisson distibution if the probability function of X is:

$$f(x) = \frac{e^{-\mu}\mu^x}{x!} \text{ for } x = 0, 1, 2, \dots$$

We write $X \sim Poisson(\mu)$. Check that $f(x)$ is a probability function:

Poisson Distribution Models Used for Problems Like . . .

- Number of phone calls to a call center in 1 hour.
- Number of new connections on wireless network in 1 unit time.
- Number of cases of a (non-contagious) rare disease in a region, town, . . .
- Number of vehicle accidents per month in a town, covered by a given insurer.
- Number of cars on a highway passing a given point per hour.

Approximating Binomial distribution with Poission Distribution

Therefore, if n is **large** and p is **small**, we can use the Poisson distribution with $\mu = np$ as an approximation to the Binomial.

Theorem (Poisson approximation to the Binomial)

As $n \to \infty$, μ is constant and $p = \frac{\mu}{n} \to 0$, if $X \sim Bi(n, p)$ then

$$P(X = x) \to \frac{\mu^x}{x!} e^{-\mu}, x = 0, 1, \dots$$

Poisson Process

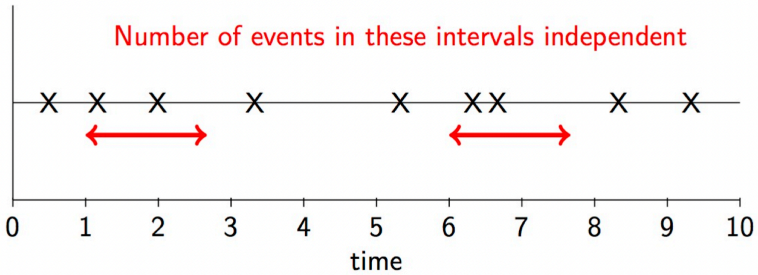
Theorem Distribution with independence, individuality, homogeneity can be described as Poisson distribution.

Suppose a process satisfies the three conditions above (independence, individuality, homogeneity).

Assume events occur at the average rate of λ per unit time.

Let X be the number of events in a time interval of length t units. Then $X \sim Poisson(\mu = \lambda t)$.

Independence: the number of occurrences in non-overlapping time intervals are independent.



Individuality: the probability of 2 or more events in a sufficiently short period of time is approximately zero.
i.e. $P(2 \text{ or more events in}(t, t + \Delta t)) = o(\Delta t)$ as $\Delta t \to 0$.

Note: $o(\Delta t)$ is called the 'order' notation. When a function $g(\Delta t) = o(\Delta t)$ as $\Delta t \to 0$, it means that $\frac{g(\Delta t)}{\Delta t} \to 0$ as $\Delta t \to 0$.

smaller than Δt (vanishingly)

Homogeneity: events occur at a uniform rate, λ, over time.
i.e. $P(\text{one event in } (t, t + \Delta t)) = \lambda \Delta t + o(\Delta t)$ error term

Expected Value

Definition (Expected Value) Definition: Expected Value

If a discrete random variable X has p.f. $f(x)$, then the number $E(X) = \sum_{\text{all x}} x \cdot f(x)$ is the **expected value** of X, denoted $E(X)$ (also referred to as mean or expectation)[2]

[2] if X takes on infinitely many values the series $\sum_{\text{all } x} |x| f(x)$ must be convergent. Otherwise we say the expected value does not exist.

$X \sim \text{Poisson}(N) \Rightarrow E(X) = N$.
$X \sim \text{Bin}(n, p) \Rightarrow E(X) = np$
$X \sim \text{Geo}(p) \Rightarrow E(X) = \frac{1}{p}$
$X \sim \text{NegBin}(r, p) \Rightarrow E(X) = \frac{pr}{1-p}$

HyperGeo (population Size: N; # of trail: n; # of success in population: S; # of success in trails: x) $\Rightarrow E(X) = n\left(\frac{S}{N}\right)$

Definition Definition: Expect Value of a Function of X

The expectation of some function $g(X)$ of a r.v. X with probability function $f(x)$ is

$$\begin{aligned} E(g(X)) &= \sum_{\text{all x}} (\text{value of g(x))} \cdot (\text{probability of } x) \\ &= \sum_{\text{all x}} g(x) f(x) \end{aligned}$$

Law of Expectation / Linearity

(Expected value is linear) Expected value is linear

For any random variable X and constants a, b,

$$E(aX + b) = aE(X) + b$$

Theroem (Linearity) Linearity

For two functions $h(*)$ and $g(*)$,

$$E[g(X) + h(X)] = E[g(X)] + E[(h(X))]$$

$$Var(X) = E(X^2) - \mu^2$$

Variance / Standard Deviation

Definition Definition: Variance

If a random variable X has an expected value of $E(X) = \mu$, then the average squared distance between X and μ is $E[(X - \mu)^2]$ and is called the **variance** of X or $Var(X)$, often denoted σ^2.

$E(X^2) - E(X)^2$

Note: variance is in squared units

Definition Definition: Standard Deviation

The square root of variance is called the **standard deviation**, denoted by σ or $SD(X)$.

$\| \frac{2}{3}$

Note: The units for standard deviation are consistent with those of the original measurements.

Theorem

If X is a random variable and a, b are some constants, then

1. $Var(aX + b) = a^2 Var(X)$ The addition of a constant has no effect on the variance.
2. $SD(aX + b) = a \times SD(X)$, where SD stands for standard deviation

Conclusion: Adding a constant to a random variable **does not** affect the variance.

Theorem

For many standard distributions, with $E(X) = \mu$, we can use $Var(X) = E[X(X-1)] + \mu - \mu^2$.

$X \sim \text{Poisson}(N) \Rightarrow Var(X) = N$
$X \sim \text{Bin}(n, p) \Rightarrow Var(X) = np(1-p)$
$X \sim \text{Geo}(p) \Rightarrow Var(X) = \frac{1-p}{p^2}$
$X \sim \text{NegBin}(r, p) \Rightarrow$

HyperGeo (population Size: N; # of trail: n; # of success in population: S; # of success in trails: x) $\Rightarrow Var(X) = n\left(\frac{S}{N}\right)\left(1 - \frac{S}{N}\right)\left(\frac{N-n}{N-1}\right)$

1	4
2	3
3	2
4	1

$$\frac{4}{64} = \frac{1}{16} = 0.0625$$