Calculating Probabilities

September-12-11 12:43 PM

Sample Space

S is called the Sample Space or the Universal Set in set theory.

Probability

$P(x)$ will denote probability

The Classical Approach to Probability

Let S be the set of all possible outcomes (points) to an experiment.

Then the probability of some event is:

$$\frac{Number\ of\ ways\ event\ can\ occur}{Number\ of\ points\ in\ S}$$

We will use this definition to formalize the mathematical concept of probability.

S has few points, one of the easiest methods of finding probabilities is to list all the outcomes.

Example:

What is the probability of rolling an odd number on a single throw of a die?

$S = \{1, 2, 3, 4, 5, 6\}$

$P(odd\ number) = \frac{3}{6} = \frac{1}{2}$

While a sample space is not necessarily unique, an important point to keep in mind about sample spaces is that they should have equally likely points in order to apply our formula for probability.

Example:

Toss a coin 3 times. What is the probability of getting a total of 2 tails?

Solution #1: $S_1 = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

$\therefore P(two\ tails) = \frac{3}{8}$

Solution #2: $S_2 = \{0\ Tails, 1\ Tail, 2\ Tails, 3\ Tails\}$

$\therefore P(two\ tails) = \frac{1}{4}$

False. The sample space consists of points which are not equally likely. It is harder to get 3 tails than either 1 or 2 tails.

What would be the probability of getting a head on the first toss?

Let $A \equiv$ getting a head on the first toss.

$A = \{HHH, HHT, HTH, HTT\} \subseteq S$

$\therefore P(A) = \frac{4}{8} = \frac{1}{2}$

Biggest disadvantage to using this listing method

Not practical when there are a lot of points to write out.

We need more efficient ways of keeping track of outcomes without having to list them all.

Basic Principles of Counting

September-14-11 12:34 PM

The Addition Rule
Suppose experiment #1 can be performed in p ways and an unrelated experiment #2 can be perform in q ways. Then, either experiment #1 or experiment #2 can be performed in p+q ways.

The Multiplication Rule
suppose experiment #1 can be performed in p ways, and an unrelated experiment #2 can be performed in q ways. Then both experiments 1 and 2 can be performed in p x q ways.

OR <-- Addition
AND <-- Multiplication

Permutation
Order is important; objects are drawn without replacement.

$$n^{(r)} = \frac{n!}{(n-r)!}$$

Combination
Order is not important; object are drawn without replacement.

$$\binom{n}{r} = \frac{n!}{(n-r)!\,r!} = \frac{n^{(r)}}{r!} = \binom{n}{r, n-r}$$

Example of Addition Rule
Suppose a class has 30 men and 25 women. Then, there are 30+25=55 ways the professor can select one student to answer a question.

Example of Multiplication Rule
A small community consists of 10 men, each of whom has 3 sons. If one man and one of his sons are to be chosen as farther and son of the year, there are 10 x 3 = 30 possible different choices.

Important Point
When objects are selected and replaced after each draw, the Addition and Multiplication Rules are generally sufficient to calculate probabilities.

Permutations
1. The number of ways to order n distinct objects is $n!$
2. The number of ordered arrangements of r objects from a set of n distinct objects. $(n \geq r)$ is $n^{(r)} = \frac{n!}{(n-r)!}$, "n taken to r terms"
3. the number of distinct arrangements of n objects when n_1 are alike of one type, n_2 are alike of one type, …, n_k are alike of one type is $\frac{n!}{n_1!n_2!\dots n_k!} = \binom{n}{n_1, n_2, \dots\ n_k,}$

Notes
$n^{(0)} = \frac{n!}{(n-0)!} = \frac{n!}{n!} = 1$
$n^{(n)} = n!$
$n^{(r)}$ still has mathematical meaning even when n is not a positive integer.

$(-2)^{(3)} = (-2)(-3)(-4) = -24$

Combinations
The number of distinct subsets, or combinations, that can be selected from n distinct object (n≥ r) is given by
$\binom{n}{r}$ − "n choose r"

Note
$\binom{n}{r}$ still has mathematical meaning even when n is not a non-negative integer $\geq r$ so long as $r \in \mathbb{N}$

Ex.
$$\binom{\frac{1}{2}}{3} = \frac{\left(\frac{1}{2}\right)^{(3)}}{3!} = \frac{\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)}{6} = \frac{1}{16}$$

Rules of Probability

12:30 PM

Universal Set / Sample Space
S, the set containing all elements

Null / Empty Set
$\emptyset$, the set containing no elements

Union
$A \cup B = \{e: e \in A \text{ or } e \in B\}$

Intersection
$A \cap B = AB = \{e: e \in A \text{ and } e \in B\}$

Complement
$\bar{A} = A^c = \{\, e : e \in S \text{ and } e \notin A\}$

Disjoint / Mutually Exclusive
A and B are disjoint when $A \cap B = \emptyset$

Note
The sets $A_1, A_2, \ldots$ are said to be mutually exclusive or pairwise disjoint if no to sets have any common elements.
$i.e. A_i \cap B_i = \emptyset\ \forall i \neq j$

Important Set Relationships
Associative Laws
$A \cup (B \cup C) = (A \cup B) \cup C$
$A \cap (B \cap C) = (A \cap B) \cap C$
Distributive Laws
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
DeMorgan's Laws
$\overline{A \cup B} = \bar{A} \cap \bar{B}$
$\overline{A \cap B} = \bar{A} \cup \bar{B}$

Partition
$A \cap B = \emptyset, A \cup B = S$

Axioms of Probability
Let $P(x)$ be a probability rule or measure.

i. $0 \leq P(A) \leq 1$
ii. P(S) = 1
iii. If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events, then
$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

General inequality
$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

Independent Events
Two events A and B are independent iff $P(A \cap B) = P(A)P(B)$

Remarks
i) Suppose that A and B are independent. Then $\bar{A}$ and B are also independent.
In general: A,B ind. $\Leftrightarrow \bar{A}B$ ind. $\Leftrightarrow A\bar{B}$ ind. $\Leftrightarrow \bar{A}\bar{B}$ ind.
ii) The concept of independence generalizes to more than 2 events: $A_1, A_2, \ldots, A_n$ are **independent events** iff
$$P\left(\bigcap_{j=1}^{k\leq n} A_{i_j}\right) = \prod_{j=1}^{k\leq n} A_{i_j}, \forall \{i_1, i_2, \ldots, i_k\} \subseteq \{1, \ldots, n\}$$

In this chapter, we will learn important rules which will provide the means of taking complicated probability problems and breaking them down into simpler, smaller pieces that are easier to work with.

A pictorial representation of sets is obtained by Venn diagrams. To construct a Venn diagram draw a rectangle whose interior will represent S. Any set A is represented as the interior of a closed curve contained in S.

Example
In a college class of 500 students, it was found that 210 smoke, 258 drink alcohol, 216 eat between meals, 122 smoke and drink alcohol, 83 eat between meals and drink alcohol, 97 smoke and eat between meals, and 52 engage in all three. If a student is selected at random, find the probability that:

a) The student smokes but does not drink alcohol
b) The student eats between meals and drinks alcohols but does not smoke
c) The student neither smokes nor eats between meals.

Axioms of Probability
Suppose a random selection from a specified population has associated with it a sample space S. a probability rule (or measure) is a numerically-valued function that assigns a number (call it P(A)) so that the following axioms hold.

i. $0 \leq P(A) \leq 1$
ii. P(S) = 1
iii. If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events, then
$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

Important Rules Governing Unions
Note that $S = A \cup \bar{A}$ and $A, \bar{A}$ are mutually exclusive
$\Rightarrow 1 = P(X) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$
$\Rightarrow P(A) = 1 - P(\bar{A})$

Extension
$P(S) = 1 - P(\bar{S}) \Rightarrow P(\emptyset) = 0$

Suppose we want to know $P(A \cup B)$
$A = (A \cap \bar{B}) \cup (A \cap B), B = (B \cap \bar{A}) \cup (A \cap B)$
$A \cup B = (A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B)$, all disjoint
$P(A \cup B) = P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B) = P(A) + P(B) - P(A \cap B)$

Remarks (Addition rule)
i) If A and B are mutually exclusive then $P(A \cup B) = P(A) + P(B) - P(\emptyset) = P(A) + P(B)$
ii) Since $P(A)$ is contained in [0,1] for any event A note that $P(A \cup B) = P(A) + P(B) - P(A \cup B) \leq P(A) + P(B)$.
Note that this inequality becomes an equality when A and B are exclusive.
iii) In the case of 3 events A, B, and C, look at
$P(A \cup B \cup C) = P(A) + P(B \cup C) - P\big(A \cap (B \cup C)\big)$
$= P(A) + P(B) + P(C) - P(B \cap C) - P\big((A \cap B) \cup (A \cap C)\big)$
$= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$

This rule generalizes to the case of an arbitrary number of events, say $A_1, A_2, \ldots, A_n$

iv) In the case of 4 events: $P(A \cup B \cup C) \leq P(A) + P(B \cup C) \leq P(A) + P(B) + P(C)$
v) Suppose that $A \subseteq B$. Note that $B = A \cup (\bar{A} \cap B) \Rightarrow P(B) = P(A) + P(\bar{A} \cap B) \geq P(A)$
In fact: $P(\bar{A} \cap B) = P(B) - P(A)$

Proof of Remark i)
$B = (\bar{A} \cap B) \cup (A \cap B)$
$P(B) = P(\bar{A} \cap B) + P(A \cap B) \Rightarrow P(\bar{A} \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B)$
$= P(B)\big(1 - P(A)\big) = P(B)P(\bar{A}) \Rightarrow \bar{A} \text{ and } B \text{ are indep.}$

Conditional Probability

September-23-11 12:45 PM

Conditional Probability

The conditional probability of an event A given that an event B has occurred is denoted $P(A \mid B)$ "Probability of A given B"

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

Note: $P(B) > 0$

Multiplication Rule

$P(A \cap B) = P(B)P(A|B)$

We can generalize the multiplication rule beyond the case of 2 events.

If $A_1, A_2, \dots, A_n$ is some sequence of events, then

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P\left(A_i \middle| \cap_{j=1}^{i-1} A_j\right)$$

Moreover, the notion of conditional probability satisfies the axioms of probability defined earlier. Specifically, let A be some event in S such that P(A)>0 Then we have

1) $0 \le P(B|A) \le 1$ for some event B
2) $P(S|A) = 1$
3) If $E_1, E_2, E_3, \dots$ is a sequence of mutually exclusive events, then $P(\cup_{i=1}^{\infty} E_i \,|A) = \sum_{i=1}^{\infty} P(E_i|A)$.

To see this consider

1) $P(B|A) = \dfrac{P(A \cap B)}{P(A)} \ge 0$

   Note that $A \cap B \subseteq B$

   Therefore $P(A \cap B) \le P(A)$, which implies

   $P(B|A) \le \dfrac{P(A)}{P(A)} = 1$
2) $P(S|A) = \frac{P(S \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$ since $S \cap A = A$
3) $P(\cup_{i=1}^{\infty} E_i \,|A) = \dfrac{P\left((\cup_{i=1}^{\infty} E_i) \cap A\right)}{P(A)} = \dfrac{P(\cup_{i=1}^{\infty} AE_i)}{P(A)}$

   $= \dfrac{\sum_{i=1}^{\infty} P(AE_i)}{P(A)} = \displaystyle\sum_{i=1}^{\infty} \frac{P(AE_i)}{P(A)} = \sum_{i=1}^{\infty} P(E_i|A)$

Example

From five computer chips, of which one is defective, two chips are randomly chosen for use. Find the probability that the 2nd chip selected is non-defective, given that the 1st chip chosen is also non-defective.

Intuition

Desired probability should be 3/4

Let $A \equiv$ event that $2nd$ chip selected is a non-defective.

Let B $\equiv$ event that the 1st chip selected is non-defective.

We want to get $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Clearly, $P(B) = \frac{4}{5}$

$$P(A \cap B) = P(both\ chips\ non-defective) = \frac{\binom{4}{2}\binom{4}{0}}{\binom{5}{2}} = \frac{6}{10} = \frac{3}{5}$$

$$P(A|B) = \frac{\frac{3}{5}}{\frac{4}{5}} = \frac{3}{4}$$

Note

What happens when A and B are independent events?

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

An equivalent way of definite independence is to say A and B are independent iff $P(A|B) = P(A)$ and $P(B|A) = P(B)$

Note

$$P(A) = P(A|S) = \frac{P(A \cap S)}{P(S)} = \frac{P(A)}{1} = P(A)$$

Multiplication Rule

3 events

$P(A \cap B \cap C) = P(A \cap B)P(C|AB) = P(A)P(B|A)P(C|AB)$

Total Probability

September-26-11 12:56 PM

Law of Total Probability

$\{B_i\}$ partition S

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

Bayes' Theorem

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

Consider

We have two events, A and B.
Note that $S = B \cap B^c$, mutually exclusive
$A = (A \cap B) \cup (A \cap B^c)$, mutually exclusive
$P(A) = P(A \cap B) + P(A \cap B^c) = P(B)P(A|B) + P(B^c)P(A|B^c)$

Assume now that $S = B_1 \cup B_2 \cup \cdots \cup B_k, where\ P(B_i) >= \forall 1 \le i \le k$ and $B_i \cap B_j\ \forall 1 \le i, j \le k\ , i \ne k$
Note that

$$A = \bigcup_{i=1}^{k} AB_i\ , P(A) = \sum_{i=1}^{k} P(AB_i)$$

If $B_1, \dots, B_k$ is a partition of S and A is some event in S then

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

Example

A cab was involved in a hit and run accident at night. Three cab companies (Green, Blue, and Purple) operate in the city. You are given the following information:

- 70% of the cabs in the city are Green, 20% are Blue and 10% are Purple
- A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the three colours 80% of the time but failed 20% of the time (split equally between the other two colour choices)

Knowing all this, what is the probability that the cab involved in the accident was indeed Blue?

$P(B) = 0.2, \qquad P(W_B|B) = 0.8$
$P(G) = 0.7, \qquad P(W_B|G) = 0.1$
$P(P) = 0.1, \qquad P(W_B|P) = 0.1$

$$P(B|W_B) = \frac{P(B \cap W_B)}{P(W_B)} = \frac{P(B)P(W_B|B)}{P(W_B)} = \frac{P(B)P(W_B|B)}{P(B)P(W_B|B) + P(G)P(W_B|G) + P(P)P(W_B|P)}$$
$$= \frac{(0.2)(0.8)}{(0.2)(0.8) + (0.7)(0.1) + (0.1)(0.1)} = \frac{0.16}{0.24} = \frac{2}{3}$$

Discrete Random Variables

September-30-11 12:31 PM

Random Variable (rv)

A random variable is a function which assigns a real number to each point in a sample space.

If S is a sample space then a rv X is a function $X: S \to \mathbb{R}$

In practice we use capital letters to denote random variables and their associated lower-case letters to represent the values they can take.

Random variables typically fall into one of two categories:

1) Continuous random variables
   - Can take on values in some interval or continuum of real numbers .
2) Discrete random variables
   - Can take on only a finite or a countable number of values.

Probability Mass Function (pmf)

If X is a discrete random variable, then we can define its probability mass function by the following:
$p(x) = P(X = x)$

Clearly $0 \leq p(x) \leq 1$ for any $x \in \mathbb{R}$, also

$$\sum_x p(x) = 1$$

The pmf is an important function for characterizing the **probability distribution** of a random variable.

Cumulative Distribution Function (cdf)

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$$

Properties

1. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$
2. $F(\infty) = \lim_{x \to \infty} F(x) = 1$
3. $F(x)$ is a non-decreasing function of x
4. If X takes on values in the set $\{x_1, x_2, \dots\}$ where $x_i < x_{i+1}$ for $i \in \mathbb{Z}^+$then $p(x_1) = F(x_1)$ and $p(x_i) = F(x_i) - F(x_{x-1})$ for $i \geq 2$

Example

Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls.

Let X be the rv indicating the number of red balls chosen.
Write out a sample space S corresponding to this experiment.

One such sample space might be $S = \{(B,B), (B,R), (R,B), (R,R)\}$

Samples space outcomes	X=x
(B,B)	0
(B,R)	1
(R,B)	1
(R,R)	2

There are 3 possible values X can take on, namely {0, 1, 2} (discrete).

Example

Toss a coin until it comes up heads.
Define the rv X to be the total number of tosses required to obtain the first head.
The, $X = x, where\ x \in \mathbb{N}^+$
So X is a discrete random variable.

PMF, earlier example

$X = x$ where $x \in \{0, 1, 2\}$

$$p(0) = P(X = 0) = P(no\ red\ balls) = \frac{\binom{4}{0}\binom{3}{2}}{\binom{7}{2}} = \frac{1}{7}$$

$$p(1) = P(X = 1) = \frac{\binom{4}{1}\binom{3}{1}}{\binom{7}{2}} = \frac{4}{7}$$

$$p(2) = p(X = 2) = \frac{2}{7}$$

Thus the pmf of X is given by

x	0	1	2	Other
p(x)	1/7	4/7	2/7	0

OR

$$p(x) = \begin{cases} \dfrac{\binom{4}{x}\binom{3}{2-x}}{\binom{7}{2}}, & x \in \{0, 1, 2\} \\ 0, & x \notin \{0, 1, 2\} \end{cases}$$

Common Probability Distributions

There are some common or special discrete probability distributions that have particular meaning and arise frequently in practice. We will introduce and study these distributions on at a time.

Discrete Uniform Distribution

October-03-11 12:44 PM

Discrete Uniform Distribution

If X is a random variable taking on values in the set $\{a, a+1, a+2, \dots, b\}$ $b = a+k, k \in \mathbb{N}$ with all values being equally likely. Then X is said to have a discrete uniform distribution.

Note

There are b-a+1 values that X can possible take on and we want each value to have the same probability. Thus, the pmf of X is given by

$$p(x) = \begin{cases} \frac{1}{b-a+1}, x \in \{a, a+1, \dots, b\} \\ 0, otherwise \end{cases}$$

For $x \in \{a, a+1, \dots, b\}$

$$F(x) = P(X \le x) = \sum_{y=a}^{x} p(y) = \frac{x-a+1}{b-a+1}$$

In general

$$F(x) = \begin{cases} 0, & x < a \\ \frac{\lfloor x-a \rfloor + 1}{b-a+1}, & a \le x \le b \\ 1, & x \ge b \end{cases}$$

Example

Computer simulation of 3 digits
Generate 3 digits independently from $\{0, \dots, 9\}$
Interest: The rv Y, where Y represents the smallest digit among our 3-digit number.
We want to determine $p(y) = P(Y = y)$

Steps:

1. Find $P(Y > y)$
2. Find F(y)
3. Find $p(y)$

Define:
X_i to be the digit generated in the i^{th} selection $i = 1, 2, 3$
We clearly see that X_i is a discrete uniform rv with $a = 0, b = 9, i = 1,2,3$

$$\Rightarrow P(X_i > x) = \frac{9-x}{10} \Rightarrow P(Y > y) = P(\min\{X_1, X_2, X_3\} > y) = P(each\ of\ X_i > y)$$

$$= P(\{X_1 > y\} \cap \{X_2 > y\} \cap \{X_3 > y\}) = P(X_1 > y)P(X_2 > y)P(X_3 > y) = \left(\frac{9-y}{10}\right)^3$$

$for\ y = 0, \dots 9$, then $F(y) = 1 - P(Y > y) = 1 - \left(\frac{9-y}{10}\right)^{10}$

$$p(y) = F(y) - F(y-1) = 1 - \left(\frac{9-y}{10}\right)^3 - \left(1 - \left(\frac{9-y+1}{10}\right)^3\right) = \left(\frac{10-y}{10}\right)^3 - \left(\frac{9-y}{10}\right)^3$$

(This happens to work out for $y = 0$ because $F(-1) = 0$

Binomial Distribution

October-05-11 12:52 PM

Binomial Distribution

Setup

Suppose we repeat the same experiment a fixed number of times. Let this number be n. If:

1. The repetitions are independent of one another.
2. There are only 2 possible outcomes in a single experiment: 'success' and 'failure' or 1/0
3. Each repetition has the same probability p of being a success.

Remark:

Such an experiment is referred to as a Bernoulli experiment or trial. The two outcomes are often labeled 0 for failure and 1 for success.

Let X count the total number of successes obtained in n repetitions of the same experiment.
X is said to have a **binomial distribution**.

Note:

There are 2 parameters which need to be specified, namely n and p.

Shorthand Notation

$X \sim Bi(n, p)$
X is a binomial rv with parameters n and p

We wish to determine the pmf of a binomial rv.
First of all, X can take on values in the set $\{0, 1, \ldots, n\}$
Next: For each $x = 0, 1, 2, \ldots, n$ what does $p(x) = P(X = x)$ look like?

How many possible arrangements exist in which there are x Ss and n-x Fs?

$$\Rightarrow \frac{n!}{x!\,(n-x)!} = \binom{n}{x}$$

Each such arrangement has probability $p^x(1-p)^{n-x}$

$$\Rightarrow p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \ for\ x = 0, 1, 2, .., n$$

Remarks

1) Note that for large n and small p, the function
   $\binom{n}{x} p^x (1-p)^{n-x}$ could be potentially subject to numerical instability.
   It may be more practical to calculate such probabilities in a recursive manner

$$P(x+1) = \binom{n}{x+1} p^{x+1}(1-p)^{n-x-1} = \frac{n-x}{x+1} \times \frac{p}{1-p} P(X = x)$$

2) The binomial distribution derives its name from the famous sum formula called the Binomial Sum or Binomial Theorem

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

   Look at $\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = \big(p + (1-p)\big)^n = 1$ from above formula

3) Recall that in the case of discrete uniform distribution, we derived explicit nice formulas for both the pmf and cdf. Unfortunately, in the case of the $Bi(n, p)$ distribution, no simplified expression exists for the cdf.

   $$P(X \le x) = \sum_{y=0}^{x} \binom{n}{y} p^y (1-p)^{n-y} \ for\ x = 0, 1, \ldots, n$$

Hypergeometric Distribution

October-07-11 12:42 PM

Hypergeometric Distribution

Setup

Consider a collection of N objects which can be classified into 2 distinct types: "Success" or "Failure"

Out of this group of N total objects, suppose r of them are classified as "success" and the remaining N-r or them are classified as "failure".
Pick n objects are random from this collection without replacement.
If X represents the number of successes obtained then X is said to have a hypergeometric probability distribution with parameters N, r and n.

The next question: What is the form of the pmf and what values can X take on?
Using combinatorial techniques from earlier:

$$p(x) = P(X = x) = P(x \text{ successes obtained in sample of size } n) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Restrictions: $x \leq \min\{n, r\}$ and $x \geq \max\{0, n - N + r\}$
Oftentimes it is easier to simply write

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, for\ x \in \mathbb{N}$$

Like in the case of the binomial distribution this distribution derives its name from the famous "Hypergeometric Identity" given by

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

We see

$$\sum_{x=0}^{\infty} p(x) = \frac{\left(\sum_{x=0}^{\infty}\binom{r}{x}\binom{N-r}{n-x}\right)}{\binom{N}{n}} = \frac{\binom{r+N-r}{n}}{\binom{N}{n}} = 1$$

Also like the binomial case, the cdf looks like:

$$F(x) = P(X \leq x) = \sum_{y=0}^{x} \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}\ for\ x \in \mathbb{N}$$

Key distinction between the binomial and hypergeometric models:

- Binomial distribution requires independent repetitions
- Hypergeometric distribution uses selections without replacement $\Rightarrow$ dependent

However, in the event that we are drawing a fairly small proportion of a large collection of objects (i.e. N is large and n is relatively small), then it would seem unlikely to get the same object more than once even if it was replaced.
In this situation, the binomial and hypergeometric models approximate each other.

Rule of Thumb

If $\frac{n}{N} \leq 0.05$
Then a reasonable (i.e 2-3 decimal places of accuracy) approximation is obtained.

Geometric Distribution

October-12-11 12:47 PM

The Geometric Distribution

Recall that a single **Bernoulli trial** can result in two outcomes: 'success' and 'failure'. Suppose our experiment consists of repeating a Bernoulli trial until the first success is obtained.

If X counts the number of failures obtained prior to the first success occurring, the X is a geometric random variable with parameter p (which represents the probability of success on a single Bernoulli trial)

What sis the form of $p(x) = P(X = x) = P(x\ failures\ occur\ before\ the\ 1st\ success)$
What are the possible values of x?
$x \in \mathbb{N}$, It has a countably infinite number of values.

$p(x) = (1-p)^x p$

Remarks

1. The geometric distribution derives its name from the geometric series. ($\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}, |r| < 1$)
   If $0 < p < 1$ then
   $$\sum_x p(x) = \sum_{x=0}^{\infty} (1-p)^x p = p \sum_x (1-p)^x = p \cdot \frac{1}{1-(1-p)} = 1$$
2. Recall that the cdf of X is given by $F(x) = P(X \leq x)$
   For $x \in \mathbb{N}$ then
   $$F(x) = P(X \leq x) = \sum_{y=0}^{x} p(y) = p \sum_{y=0}^{x} (1-p)^y = p \cdot \frac{(1-(1-p)^{x+1})}{1-(1-p)} = 1-(1-p)^{x+1}$$
3. In some texts, the definition of the geometric distribution is slightly different. Instead of modelling the number of failures prior to the 1st success, we could model the number of trials run to obtain the 1st success.
   Let Y be this rv. Clearly Y = X+1
   Thus the pmf of Y is given by $P(Y = y) = P(X + 1 = y) = P(X = y - 1) = p(y-1) = (1-p)^{y-1}p$ for $y \in \mathbb{N}^+$

   Thus: For $y \in \mathbb{N}^+, P(Y \leq y) = P(X + 1 \leq y) = F(y-1) = 1-(1-p)^y$

Negative Binomial Distribution

October-14-11 12:41 PM

Negative Binomial Distribution

The setup for this distribution is almost the same as for the geometric distribution.

- An experiment has 2 distinct types of outcomes
- Repetitions are independent with same probability p of success each time.

However, the repetitions are continued until a fixed (pre-determined) number r of successes have been obtained.

If X counts the number of failures obtained before the r^{th} success occurs, then X is said to have a **negative binomial distribution**. It has two parameters, r and p.

Form

There are $x+r$ trials. The last is S. The remaining $x+r-1$ trials have xF's and (r-1)S's in any order.

PMF

$p(x) = \binom{x+r-1}{x} p^r (1-p)^x \; for \; x \in \mathbb{N}$

If r = 1 then
$p(x) = (1-p)^x p$, which is the geometric PMF

Alternate

Just as in the case of a geometric distribution, one can define an alternative version of the negative binomial distribution. Specifically, $Y \equiv rv$ indicating the total number of trials run to obrtain the r^{th} success. Clearly, $Y = X + r$

The pmf of Y is given by $P(Y=y) = P(X+r=y) = P(X=y-r) = p(y-r) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$ for $y \geq r$

Recall

Consider the function $f(x) = (1+x)^n$ where $n \in \mathbb{R}$. The Taylor series of $f(x)$ about the point $x=0$ is given by

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=0}^{\infty} \binom{n}{k} x^k \text{ where } |x| < 1 \; and \; \binom{n}{r} = \frac{n^{(k)}}{k!}$$

Let's verify that $\sum p(x) = 1$

$$p(x) = \binom{x+r-1}{x} p^r (1-p)^x$$

$$\binom{x+r-1}{x} = \frac{(x+r-1)^{(x)}}{x!}$$
$$= \frac{\big((x+r-1)(x+r-2)\dots\big(x+r-(x-1)\big)(x+r-x)\big)}{x!}$$
$$= \frac{(x+r-1)(x+r-2)\dots(r+1)(r)}{x!}$$
$$= \frac{(-1)^x(-r)(-r-1)\dots\big(-r-(x-1)\big)}{x!} = (-1)^x \binom{-r}{x}$$

Now we have

$$\sum_{(x=0)}^{\infty} p(x) = \sum_{x=0}^{\infty} \binom{x+r-1}{x} p^r (1-p)^x$$
$$= \sum_{x=0}^{\infty} (-1)^x \binom{-r}{x} p^r (1-p)^x = p^r \sum_{x=0}^{\infty} \binom{-r}{x} (p-1)^x$$
$$= p^r (1+p-1)^{-r} = p^r p^{-r} = 1$$

The Poisson Distribution

October-17-11 10:21 PM

Poisson Distribution

PMF

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Recall the form of the binomial distribution.
i.e. $X \sim B(n,p)$ if its pmf looks like $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, \dots, n$

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

What happens when n gets large $(n \to \infty)$, p gets small $(p \to 0)$ and the product np remains fixed at some constant value $np = \lambda$.

Note that $\lambda = np \Rightarrow p = \frac{\lambda}{n}$

$$p(x) = \frac{n(n-1)(n-2)\dots(n-x+1)}{x!}\left(1-\frac{\lambda}{n}\right)^{n-x}\left(\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!}\frac{n(n-1)(n-2)\dots(n-x+1)}{n^x}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}$$

$$= \frac{\lambda}{x!} \times 1 \times \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{x+1}{n}\right)\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}$$

Letting $n \to \infty$ not that $p = \frac{\lambda}{n} \to 0$

$$\lim_{n\to\infty} p(x) = \frac{\lambda^x}{n}(1)(1)\dots(1) \times \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n \times 1 = \frac{\lambda^x e^{-\lambda}}{x!}, for\ x = 0, 1, 2, \dots$$

The Poisson Distribution arises as the limiting form of the binomial distribution.

Check that $\sum_{x=0}^{\infty}\frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$

Rule of Thumb

The Poisson distribution will provide a "good" approximation (i.e. at least 2 decimal places) when $n \geq 20$ and $p \leq 0.05$.) When $n \geq 100$ and $p < 0.1$ the approximation is generally excellent (i.e. 3-4 places of accuracy)

Note

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \lim_{n\to\infty} e^{n\ln\left(1-\frac{\lambda}{n}\right)} = e^{\lim_{n\to\infty}\frac{\ln\left(1-\frac{\lambda}{n}\right)}{n^{-1}}}$$

$\lim_{n\to\infty}\frac{\ln\left(1-\frac{\lambda}{n}\right)}{n^{-1}} = \frac{0}{0}$ so apply L'Hopital's Rule

$$e^{\lim_{n\to\infty}\frac{\ln\left(1-\frac{\lambda}{n}\right)}{n^{-1}}} = e^{\lim_{n\to\infty}\frac{\left(\frac{1}{1-\frac{\lambda}{n}}\right)(\lambda n^{-2})}{-n^{-2}}} = e^{\lim_{n\to\infty} -\frac{\lambda}{1-\frac{\lambda}{n}}} = e^{-\lambda}$$

Example

The probability that any given page of the first five chapters of the stat 240 textbook contains at least one typographical error is 0.005 and errors are independent from page to page.
What is the probability that out of the 210 total pagers, there are at most 22 pages with errors?

Let X count the number of pages with errors. Then $X \sim B_i(n = 210, p = 0.005)$
Thus $p(x) = \binom{210}{x}(0.005)^x(0.995)^{210-x}\ for\ x \in \mathbb{N}$
We want to calculate $P(X \leq 2)$
$F(2) = p(0) + p(1) + p(2) = 0.9107$

However, since n is "large" and p is "small", we can use the Poisson distribution to get an approximation to the exact answer. To do so, let $\lambda = np = (210)(0.005) = 1.05$
Now, let Y be a Poisson rv with parameter $\lambda = 1.05$ so that

$$p(Y = y) = \frac{(1.05)^y e^{-1.05}}{y!} \cong 0.9103$$

Poisson Process

October-19-11 12:46 PM

Poisson Process

Consider the following assumptions about the way in which events occur:

1) The number of occurrences in non-overlapping (i.e. disjoint) intervals are independent.
2) Let Δt represent a very short period of time. Then the probability of 2 or more occurrences in this short period of time is approximately 0.
3) Events occur at a uniform rate over time

   What this means is that there exists a parameter or constant $\lambda > 0$ such that the probability of a single occurrence is proportional to the length of time.

   $p_{\Delta t}(1) = \lambda$

This 3 conditions define what's known as a **Poisson Process.**

It's called a Poisson process because if you let X count the number of events to occur in the time interval [0, t] then X_t actually has a Poisson distribution.

Let $p_t(x)$ be the pmf of X_t.

$$\boxed{\frac{d}{dt}p_t(x) = \lambda p_t(x-1) - \lambda p_t(x)}$$

and

$$p_t(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}\ \forall x \in \mathbb{N}$$

Lambda is the constant of proportionality and is often referred to (or interpreted as) the average rate of occurrence per unit of time.

λt represents the average number of occurrences in the interval $[0, t]$

Limiting form of the binomial distribution, but it also arises in connection with the random occurrence of events of a particular type over time (or space).

Derivation of Poisson Process

To see this, let $p_t(x)$ be the pmf of X_t $p_t(x) = P(X_t = x)$

Let $t \geq 0$ and consider the time interval $[0, t + \Delta t]$

Look at $p_{t+\Delta t}(x) = P(x \text{ events occur in the interval } [0, t + \Delta t])$

$$= \sum_{k=0}^{x} P(k \text{ events occur in the interval } [0,t])P(x\ in\ [0, t+\Delta t] | k\ in\ [0,t])$$

$$= p_t(x)P(p_{t+\Delta t}(x)|p_t(x)) + p_t(x-1)P(p_{t+\Delta t}(x)|p_t(x-1)$$

$$+ \sum_{k=0}^{x-2} p_t(k)P(p_{t+\Delta t}(x)|p_t(k))$$

$$= p_t(x)p_{\Delta t}(0) + p_t(x-1)p_{\Delta t}(1) + \sum_{k=0}^{x-2} p_t(k)p_{\Delta t}(x-k)$$

$$\cong p_t(x)(1-\lambda\Delta t) + p_t(x-1)\lambda\Delta t = p_t(x) - \lambda\Delta t p_t(x) + p_t(x-1)\lambda\Delta t = p_{t+\Delta t}(x)$$

$$\Rightarrow \frac{p_{t+\Delta t}(x) - p_t(x)}{\Delta t} = \lambda p_t(x-1) - \lambda p_t(x)$$

$$\lim_{\Delta t \to 0} \frac{p_{t+\Delta t}(x) - p_t(x)}{\Delta t} = \lambda p_t(x-1) - \lambda p_t(x)$$

$$\boxed{\frac{d}{dt}p_t(x) = \lambda p_t(x-1) - \lambda p_t(x)}$$

If one specifies $p_0(0) = 1$ and $p_0(x) = 0\ \forall x \in \mathbb{Z}^+$ as the set of initial conditions it is possible to solve for $p_t(x)$ and obtain the solution

$$p_t(x) = \frac{(\lambda x)^x e^{-\lambda t}}{x!}\ \forall x \in \mathbb{N}$$

Distribution Example

October-21-11 12:32 PM

Example

A person working in telephone sales has a 20% chance of making a sale on each call, with calls being independent. Assume calls are made at a constant rate, with the numbers in non-overlapping periods being independent. On average, there are 20 calls made per hours.

A) Find the probability there are 2 sales in 5 calls
B) Find the probability exactly 8 calls are needed to make 2 sales.
C) Find the probability of 3 calls being made in a 15-minute period.

A

Let X represent the number of calls resulting in sales.

Here $X \sim B_i(n = 5, p = 0.2)$. We want to calculate $P(X = 2) = p(2) = \binom{5}{2} 0.2^2 0.8^3$

B

Let Y be the number of calls which result in non-sales prior to obtaining 2 calls which result in sales.
Here, Y is negative binomial with parameters r=2 and p=0.2

$P(Y = y) = \binom{y + 2 - 1}{y} 0.2^2 0.8^y \; for \; y \in \mathbb{N}$

Thus P(exactly 8 calls needed to make 2 sales)$= P(Y = 6) = \binom{7}{6} 0.2^2 0.8^6$

C

Let Z represent the # of calls made in a 15-minute time period.

Then, $Z \sim$Poisson with parameter $\lambda t = (20)\left(\frac{1}{4}\right) = 5$

$P(Z = 3) = \frac{e^{-5} 5^3}{3!}$

Expectation

October-21-11 12:45 PM

The Concept of Expectation

For a discrete rv X with pmf $p(x) = P(X = x)$ the **expected value** of X is given by

$$\mu = \mu_X = E(X) = \sum_x xp(x)$$

This is the weighted average of the values X can assume.

This $E(X)$ is also referred to as the mean of X

Functions of Random Variables

In many problems of statistics, one is interested not only in the expected value of a random variable X but also in the expected values of random variables related to X by means of the equation $Y = g(X)$.

The expected value of $Y = g(X)$ is given by:

$$E(Y) = E[g(X)] = \sum_x g(x)p\big(g(x)\big) = \sum_x g(x)p(x)$$

Remark

the expected value operator "E" is a linear operator.
$E(ag(x) + b) = aE[g(X)] + b$

In general

$$E\left(\sum_{i=1}^{n} a_i g_i(X)\right) = \sum_{i=1}^{n} a_i E\big(g_i(X)\big)$$

Probability Distribution Means

What are the means of the 6 special discrete probability distributions?

1) **Discrete Uniform:**

$$p(x) = \frac{1}{b-a+1}, for\ x = a, a+1, \dots, b$$
$$E(X) = \frac{a+b}{2}$$

2) **Binomial:**

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x \in \mathbb{N}$$
$$E(X) = np$$

3) **Hypergeometric**

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}\ for\ \max\{0, n-N+r\} \le x \le \min\{r, n\}$$

But also works for $x \in \mathbb{N}$ since the chooses will be 0

$$E(X) = n\left(\frac{r}{N}\right)$$

5) **Negative Binomial (also geometric)**

$$p(x) = \binom{x+r-1}{x} p^r (1-p)^x, \qquad for\ x \in \mathbb{N}$$
$$E(x) = \frac{r(1-p)}{p}$$

For Geometric: $r = 1$

$$E(X) = \frac{1-p}{p}$$

Also: $Y = X + r$

$$\Rightarrow E(Y) = E(X) + r = \frac{r(1-p)}{p} + r = \frac{r}{p}$$

6) **Poisson**

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Example

A lot of 12 television sets includes 2 that are defective. IF three of the sets are chosen at random from the shipment to a hotel, how many defective sets can they expect to receive?

Let X represent the # of defective TV sets received by the hotel.
X is hypergeometric with parameters $N = 12, r = 2, n = 3$

$$p(x) = \frac{\binom{2}{x}\binom{10}{3-x}}{\binom{12}{3}}\ for\ x \in 0,1,2$$

x	0	1	2
$p(x)$	$\frac{6}{11}$	$\frac{9}{22}$	$\frac{1}{22}$

$$E(X) = \frac{9}{22} + \frac{2}{22} = \frac{1}{2}$$

Example

In the last example, suppose the shipping company charges at $100 flat rate for the actual shipping of the TV sets. However, for any defective set received the company must pay out $50 in additional costs.
Might be interested in the rv $Y = 100 - 50X$ representing the company's revenue.

Proof of Remark

$$E(Y) = E(ag(x) + b) = \sum_x (ag(x) + b)p(x) = \sum_x ag(x)p(x) + bp(x)$$
$$= a\sum_x g(x)p(x) + b\sum_x p(x) = aE[g(x)] + b$$

Back to Example

$$E(X) = \frac{1}{2}, \qquad Y = 100 - 50X$$
$$E(Y) = -50 \times E(X) + 100 = -50\left(\frac{1}{2}\right) + 100 = \$75$$

Derivation of Probability Distribution Means

Discrete Uniform

$$E(X) = \sum_{x=a}^{b} \frac{x}{b-a+1} = \frac{1}{b-a+1}\left(\sum_{y=1}^{b-a+1} y + (a-1)\right)$$
$$= \frac{1}{b-a+1}\left(\frac{(b-a+1)(b-a+2)}{2} + (a-1)(b-a+1)\right) = \frac{a+b}{2}$$

Binomial

$$E(X) = \sum_{x=1}^{n} x\left(\frac{n!}{x!\,(n-x)!}\right)p^x(1-p)^{n-x}$$
$$= np\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,\big((n-1)-(x-1)\big)!} p^{x-1}(1-p)^{(n-1)-(x-1)}$$
$$= np\sum_{x=1}^{n}\binom{n-1}{x-1}p^{x-1}(1-p)^{(n-1)-(x-1)} = np\sum_{y=0}^{n-1}\binom{n-1}{y}p^y(1-p)^{(n-1)-y}$$
$$= np$$

Hypergeometric

$$E(X) = \sum_{x=1}^{\infty} x\frac{\left(\frac{r!}{x!\,(r-x)!}\right)\binom{N-r}{n-x}}{\binom{N}{n}}$$
$$= \frac{r}{\binom{N}{n}}\sum_{x=1}^{\infty}\frac{(r-1)!}{(x-1)!\,\big((r-1)-(x-1)\big)!}\binom{N-r}{n-x}$$
$$= \frac{r}{\binom{N}{n}}\sum_{y=0}^{\infty}\binom{r-1}{y}\binom{N-r}{(n-1)-y}$$

Recall:

$$\sum_{x=0}^{\infty}\binom{a}{x}\binom{b}{m-x} = \binom{a+b}{m}$$

$$E(x) = \frac{r}{\binom{N}{n}}\sum_{y=0}^{\infty}\binom{r-1}{y}\binom{N-r}{(n-1)-y} = \frac{r}{\binom{N}{n}}\cdot\binom{N-1}{n-1}$$
$$= r\cdot\frac{n!\,(N-n)!}{N!}\cdot\frac{(N-1)!}{(n-1)!\,(N-n)!} = \frac{rn}{N}$$

Negative Binomial

$$E(x) = \sum_{x=1}^{\infty} x \cdot \frac{(-1)^x(-r)^{(x)}}{x!} p^r(1-p)^x$$

$$= p^r \sum_{x=1}^{\infty} \frac{(-1)^x(-r)(-r-1)^{(x-1)}}{(x-1)!}(1-p)^x$$

$$= rp^r(1-p)\sum_{x=1}^{\infty}\binom{-r-1}{x-1}(p-1)^{x-1} = rp^r(1-p)\sum_{y=0}^{\infty}\binom{-r-1}{y}(p-1)^y$$

Recall:

For n $\notin \mathbb{N}, |a| < 1, (1+a)^n = \sum_{x=0}^{\infty}\binom{n}{x}a^x$

$$E(x) = rp^r(1-p)p^{-r-1} = \frac{r(1-p)}{p}$$

Poisson

$$E(x) = \sum_{x=1}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x!} = \lambda\sum_{x=1}^{\infty}\frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda\sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^y}{y!} = \lambda$$

Measures of Variability

October-26-11 12:38 PM

Variance

The variance of a random variable X is defined by
$\sigma^2 = Var(X) = E[(X-\mu)^2]$
The average squared distance x is away from its mean.

If X happens to be a discrete rv with pmf $p(x)$ then :

$$Var(X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x)$$

$$Var(X) = E(X^2) - \mu^2 = E(X^2) - E(X)^2$$
$$Var(X) = E[X(X-1)] + \mu - \mu^2$$

Standard Deviation

The standard deviation of X is defined by
$\sigma_X = \sigma = SD(X) = \sqrt{Var(X)}$

Note

Variance (and Standard Deviation) can never be negative.
When does $Var(X) = 0$? When X is an rv that is equal to a single constant with probability 1.

Important Property

Let $Y = aX + b$
We know that
$\mu_Y = E(U) = a\mu_X + b$
What is σ_Y^2, σ_Y?
To determine this, $\sigma_Y^2 = Var(U) = E[(Y-\mu_Y)^2] = E[aX + b - a\mu_X - b] = E[a^2(X-\mu_X)] = a^2 E[(X-\mu_X)] = a^2 Var(X) = a^2\sigma_X^2$

$\therefore \sigma_Y^2 = a^2\sigma_X^2$
$\therefore \sigma_Y = |a|\sigma_X$

Variances of the Distributions

1) **Discrete Uniform**

$$Var(X) = \frac{(b-a)(b-a+2)}{12}$$

2) **Binomial**

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

3) **Hypergeometric**

$$Var(X) = \frac{nr}{N} \cdot \frac{(N-r)(N-n)}{N(N-1)} = n\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Variability (or dispersion) is a term used in statistics to characterize individual differences. The greater the variability between observations, the more they will be spread out. Variability is an important consideration in statistics as measures of central tendency by themselves are often not enough information to make sound decisions regarding the distribution.

What qualifies as a "good" or "appropriate" measure of spread?

Calculating Variation

$\sum_x (x-\mu)^2 p(x)$ can be awkward to use
Instead a more computationally-friendly procedure to calculate Var(X) can be used:
$Var(X) = E[(X-\mu)^2] = E(X^2 - 2\mu X + \mu^2)$
$= E(X^2) - 2\mu E(X) + \mu^2$
$E(X) = \mu$ so
$Var(X) = E(X^2) - \mu^2 = E(X^2) - E(X^2)^2$

Generally speaking, this is the most popular method of computing variances. Another often useful method for calculating variances (particularly if p(X) contains $\frac{1}{x!}$ in its formula) is to use:

Note that $E[X(X-1)] = E(X^2 - X) = E(X^2) - E(X)$
$\Rightarrow E(X^2) = E[X(X-1)] + E(X)$

$Var(X) = E[X(X-1)] + \mu - \mu^2$

Continuous Random Variables

October-28-11 12:43 PM

Continuous Random Variable

A rv X is called continuous if X can take on the uncountably infinite number of possible values associated with intervals of real numbers .

Probability Density Function

Associated with each continuous rv X, there is a function $f(x)$ (called the probability density function, or pdf) such that

i) $f(x) \geq 0 \; \forall x \in \mathbb{R}$

ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

iii) $P(a \leq X \leq b) = \int_a^b f(x)dx$

Note

One obtains probabilities for continuous random variables by determining the appropriate areas under the pdf.

Note also for a continuous rv X,

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x)dx = 0$$

For continuous random variables, on instead talk about the probability of being in a certain ranges like $P(a \leq X \leq b)$ or $P(X \leq b)$

Cumulative Distribution Function

The cdf of X where X is a continuous rv is defined as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy \; \forall x \in \mathbb{R}$$

From this definition, we have a relationship between the pdf and the cdf of X, namely:

$$\frac{d}{dx}F(x) = \frac{d}{dx}\left(\int_{-\infty}^{x} f(y)dy\right) = f(x)$$

Important Remarks:

1) The cdf of a continuous rv is always a continuous function of x
2) F(x) is a non-decreasing function of x
3) $F(-\infty) = \lim_{x\to-\infty} F(x) = 0$ and $F(+\infty) = \lim_{x\to\infty} F(x) = 1$
4) $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < x < b)$ $= F(b) - F(a)$

Relations

$$F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(y)dy$$

$$f(x) = F'(x)$$

A moment of reflection on statistical problems encountered in the real world should convince you that not all random variables take on only a finite number or countable infinity number of values. Many random variables seen in practice have more than a countable collection of possible values.

Examples

- The daily rainfall at a certain geographical point
- The lifetime of a battery
- The change in the depth of a river from one day to the next
- The diameters of machined rods produced at a plant

Random variables of this type have a continuum of possible values. Random variables that take on any value in an interval are call continuous random variables.

Example

Consider the following pdf:

$$f(x) = \begin{cases} k(2x - x^2), & 0 < x < 2 \\ 0, & elsewhere \end{cases}$$

We require $\int_0^2 f(x)dx = 1$

$$k\int_0^2 (2x - x^2)dx = 1 \Rightarrow k\left(x^2 - \frac{x^3}{3}\right)\Big|_0^2 = 1 \Rightarrow k\left(4 - \frac{8}{3}\right) = 1 \Rightarrow k = \frac{3}{4}$$

So we really have

$$f(x) = \begin{cases} \frac{3}{4}(2x - x^2), & 0 < x < 2 \\ 0, & elsewhere \end{cases}$$

Find the cdf of X

For $x < 0$, then $F(x) = P(X \leq x) = 0$

For $0 \leq x \leq 2, F(x) = \int_{-\infty}^{x} f(y)dy = \int_0^x f(y)dy = \frac{3}{4}\int_0^x (2y - y^2)dy$

$$= \frac{3}{4}\left(x^2 - \frac{x^3}{3}\right) = \frac{1}{4}x^2(3 - x)$$

$For\ x > 2, F(x) = P(X \leq x) = \int_{\infty}^{x} f(y)dy = \int_0^2 f(y)dy = 1$

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}x^3(3 - x), & 0 \leq x \leq 2 \\ 1, & x > 2 \end{cases}$$

What is $P(X \geq 1.5)$?

$$P(X \geq 1.5) = 1 - P(X \leq 1.5) = 1 - F(1.5) = 1 - \frac{1}{4}(1.5)^2(3 - 1.5)$$

$$= \frac{5}{32}$$

What about $P(1 \leq X \leq 1.5)$

$$P(1 \leq X \leq 1.5) = P(1 < X \leq 1.5) = P(X \leq 1.5) - P(X \leq 1)$$

$$= F(1.5) - F(1) = \frac{1}{4}(1.5)^2(3 - 1.5) - \frac{1}{4}(1)^2(3 - 1) = \frac{11}{32}$$

Expected Values of Continuous Random Variables

October-31-11 12:57 PM

Expected Value

If X is a continuous rv with pdf $f(x)$ and if $Y = g(X)$ is any real-values function of X, then

$$E(Y) = E[G(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

Expected Value (Mean)

The expected value of X is given by

$$\mu = E(x) = \int_{-\infty}^{\infty} xf(x)dx$$

Linearity

Just as in the discrete case, 'E' is a linear operator

$$E\left(\sum_{i=1}^{n} a_i g_i(X)\right) = \sum_{i=1}^{n} a_i E[g_i(X)]$$

Variance

$$Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = E(X^2) - \mu^2$$

For constants a, b define $Y = aX + b$

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

$Var(Y) = Var(aX + b) = a^2 Var(X)$

$\sigma(Y) = |a|\sigma(X)$

Example

Calculate the mean and variance of the rv X when X has pdf

$$f(x) = \begin{cases} \frac{3}{4}(2x - x^2), & 0 < x < 2 \\ 0, & elsewhere \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^2 x\frac{3}{4}(2x - x^2)dx = \frac{3}{4}\left(\frac{2x^3}{3} - \frac{x^4}{4}\right)\Big|_0^2 = \frac{3}{4}\left(\frac{16}{3} - \frac{16}{4}\right)$$

$$= 4 - 3 = 1$$

$$Var(X) = E(X^2) - \mu^2 = E(X^2) - 1 = \int_0^2 x^2 \frac{3}{4}(2x - x^2)dx - 1$$

$$= \frac{3}{4}\left(\frac{x^4}{2} - \frac{x^5}{5}\right)\Big|_0^2 - 1 = \frac{3}{4}\left(8 - \frac{32}{5}\right) - 1 = \frac{6}{5} - 1 = \frac{1}{5}$$

$$\sigma(X) = \frac{1}{\sqrt{5}}$$

What is $E(Y)$ where $Y = \frac{1}{X}$

$$E(Y) = \int_0^2 \frac{1}{x}\cdot\frac{3}{4}(2x - x^2)dx = \frac{3}{4}\int_0^2 (2-x)dx = \frac{3}{4}\left(2x - \frac{x^2}{2}\right)\Big|_0^2 = \frac{3}{4}(4-2)$$

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

Calculation Methods

November-02-11 12:33 PM

Important Result

Suppose that X is a continuous rv (having pdf $f_X(x)$).
Suppose that $Y = g(X)$ where the function g is either strictly increasing or strictly decreasing.
Then the rv Y has pdf given by

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|, \qquad \text{where } x = g^{-1}(y)$$

Interesting Property

Suppose X is a continuous random variable with cdf $F(x) = P(X \le x)$ and $f(x) = F'(x)$

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx$$

If X happens to be defined only on $(0, \infty)$ (i.e. X is a non-negative rv) then

$$E(X) = \int_0^{\infty} x f(x)dx = \int_0^{\infty} \left(1 - F(x)\right)dx$$

Proof

Suppose that g is a strictly increasing function. Thus $y = g(x)$ implies that $x = g^{-1}(y)$
Now, look at the cdf of $Y = g(X)$

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P\left(g^{-1}(g(X)) \le g^{-1}(y)\right) = P\left(X \le g^{-1}(y)\right)$$
$$= F_X\left(g^{-1}(y)\right)$$

To get pdf of Y

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}\left(F_Y(y)\right) = \frac{d}{dy}\left(F_X\left(g^{-1}(y)\right)\right) = \frac{d}{d\left(g^{-1}(y)\right)}\left(F_X\left(g^{-1}(y)\right)\right)\cdot\frac{d\left(g^{-1}(y)\right)}{dy}$$
$$= f_X\left(g^{-1}(y)\right)\cdot\frac{d\left(g^{-1}(y)\right)}{dy} = f_X(x)\frac{dx}{dy} \text{ Since } x = g^{-1}(y)$$

Similarly, if g is strictly decreasing then

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P\left(X \ge g^{-1}(y)\right) = P\left(X > g^{-1}(y)\right) = 1 - F_X\left(g^{-1}(y)\right)$$
$$f_Y(y) = \frac{d}{dy}F_Y(y) = -f_x(x)\frac{dx}{dy} = f_x(x)\left|\frac{dx}{dy}\right|$$

Example

Recall the earlier example with pdf

$$f(x) = \begin{cases} \frac{3}{4}(2x - x^2), & 0 < x < 2 \\ 0, & elsewhere \end{cases}$$

Find the pdf of $Y = \frac{1}{X}$
Applying the result, the pdf of Y is given by

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X\left(\frac{1}{y}\right)\left|-\frac{1}{y^2}\right| = \frac{3}{4}\left(\frac{2}{y} - \frac{1}{y^2}\right)\left(\frac{1}{y^2}\right) = \frac{3}{4y^4}(2y - 1) \; for \; y \in \left[\frac{1}{2}, \infty\right)$$

Exercise

Verify $\int_{\frac{1}{2}}^{\infty} f_Y(y)dy = 1$

$$E(Y) = E\left(\frac{1}{X}\right) = \frac{3}{2} = \int_{\frac{1}{2}}^{\infty} yF_Y(y)dy$$

Example

Suppose that X is a continuous rv on the interval $(-\infty, \infty)$
Define $Y = X^2$ and find the pdf of Y. Note that Y is not monotonic relative to X
To get the pdf of $Y = X^2$ look first ad the cdf $F_Y(y)$

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

Taking derivative with respect to y we obtain

$$f_Y(y) = \frac{d}{dy}[F_x(\sqrt{y}) - F_x(-\sqrt{y})] = \frac{d}{dy}\left(F_X(\sqrt{y})\right) - \frac{d}{dy}\left(F_X(-\sqrt{y})\right)$$
$$= f_X(\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) - f_X(-\sqrt{y})\right)$$

Look at a slight change to the problem: Suppose that X is a continuous rv on $(0, \infty)$
From $y = x^2, x = \sqrt{y} = g^{-1}(y)$

$$\frac{dx}{dy} = \frac{1}{2\sqrt{y}}$$

Thus $f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} \; for \; y > 0$

Proof of Interesting Property

Consider

$$\int_0^{\infty}(1 - F(x))dx = \int_0^{\infty}(1 - P(X \le x))dx = \int_0^{\infty} P(X > x)dx = \int_0^{\infty}\left(\int_x^{\infty} f(y)dy\right)dx$$
$$= \int_0^{\infty}\left(\int_0^y f(y)dx\right)dy = \int_0^{\infty} f(y)\left(\int_0^y dx\right)dy = \int_0^{\infty} yf(y)dy$$

Exercise

$$f(x) = \begin{cases} \frac{3}{4}(2x - x^2), & 0 < x < 2 \\ 0, & elsewhere \end{cases}$$

$$E(X) = 1 = \int_0^2 x \times \frac{3}{4}(2x - x^2)dx$$

Calculate using this alternate method

Example

Suppose X is a rv with pdf $f(x) = 2e^{-2x}$ for $x > 0$

Veryify $\int_0^{\infty} 2e^{-2x}dx = 1$

Calculate the cdf:

$$F(y) = \int_0^x 2e^{-2y}dy = (-e^{-2y})\Big|_0^x = 1 - e^{-2x}$$
$$E(X) = \int_0^{\infty} x2e^{-2x}dx \text{ OR}$$
$$E(X) = \int_0^{\infty} e^{-2x}dx = \left(-\frac{e^{-2x}}{x}\right)\Big|_0^{\infty} = \frac{1}{2}$$

Continuous Uniform Distribution

November-04-11 12:31 PM

Suppose X takes on values in the interval $[a, b]$. We assume that sub-intervals of $[a, b]$ that have the same length are assigned the equal probability.
We say that X is a continuous uniform r.v. on $[a, b]$

PDF

The pdf of X is given by $f(x) = \begin{cases} \dfrac{1}{b-1}\ for\ a < x < b \\ 0\ elsewhere \end{cases}$

CDF

For $a \leq x \leq b$

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{a} f(y)dy + \int_{a}^{x} f(y)dy = \int_{a}^{x} \frac{1}{b-a}dy = \frac{x-a}{b-a}$$

$$F(x) = \begin{cases} 0\ if\ x < 0 \\ \dfrac{x-a}{b-a}\ if\ a \leq x \leq b \\ 1\ if\ x > b \end{cases}$$

Mean and Variance

$$\mu = E(X) = \int_{a}^{b} \frac{x}{b-a}dx = \left(\frac{1}{b-a}\right)\left(\frac{x^2}{2}\right)\Big|_{a}^{b} = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

$$E(X^2) = \int_{a}^{b} \frac{x^2}{b-a}dx = \left(\frac{1}{b-a}\right)\left(\frac{x^3}{3}\right)\Big|_{a}^{b} = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}$$

$$SD(X) = \frac{b-a}{\sqrt{12}}$$

Example

Suppose a value x is chosen at random in the interval $[0, b]$ where $b > 0$. The value x divides the interval $[0, b]$ into 2 subintervals. Find the pdf of the length of the shortest subinterval.

Let X be the rv indicating the random position chosen on $[0, b]$. Bo construction, X is continuous uniform on $[0, b]$.

Let Y represent the length of the shortest subinterval. So $Y = \min\{X, b-X\}$
To get the pdf of Y, let us first consider the function $1 - F_Y(y) = P(Y > y)$
For $y \in \left[0, \frac{b}{2}\right]$, $P(Y > y) = P(\min\{X, b-X\} > y) = P(X > y \cap b - X > y\} = P(X > y \cap X < b - y) = P(y < X \leq b - y) = F_X(b-y) - F_x(y) = \frac{b-y}{b} - \frac{y}{b} = \frac{b-2y}{b} = 1 - \frac{2}{b}y$

$$F_Y(y) = \frac{2}{b}y\ for\ 0 \leq y < \frac{b}{2}$$

$$f_Y(y) = F_Y'(y) = \frac{2}{b} = \frac{1}{\frac{b}{2} - 0}$$

So Y is a uniform r.v. on $\left[0, \frac{b}{2}\right]$

Exponential Distribution

12:59 PM

PDF

X is said to have an exponential distribution (with parameter $\lambda > 0$) if its pdf is of the form

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Many random variables occuring in engineering and the sciences can be modelled as having an exponential distribution.

CDF

For $x \geq 0$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{0} f(y)dy + \int_{0}^{x} f(y)dy$$

$$= \int_{0}^{x} \lambda e^{-\lambda y} dy = \left(-e^{-\lambda y}\right)\Big|_{0}^{x} = 1 - e^{-\lambda x}$$

$$P(X > x) = 1 - F(x) = 1 - \left(1 - e^{-\lambda x}\right) = e^{-\lambda x}$$

Aside: Gamma Function

$$\Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1} e^{-x} dx, \qquad \alpha > 0$$

$$\Gamma(1) = \int_{0}^{\infty} x^{0} e^{-x} dx = \int_{0}^{\infty} e^{-x} dx = (-e^{-x})\Big|_{0}^{\infty} = 1$$

Properties of the Gamma Function

1) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ use integration by parts once.
2) For $\alpha \in \mathbb{Z}^{+}$, then
   $$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) = (\alpha - 1)(\alpha - 2) \ldots \Gamma(1) = (\alpha - 1)!$$
3) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Mean and Variance

$$E(X) = \int_{0}^{\infty} x \lambda e^{-\lambda x} dx = \int_{0}^{\infty} [1 - F(x)]dx = \int_{0}^{\infty} e^{-\lambda x} dx$$

$$= \frac{-e^{-\lambda x}}{\lambda}\Big|_{0}^{\infty} = \frac{1}{\lambda}$$

$$E(X^2) = \int_{0}^{\infty} x^2 \lambda e^{-\lambda x} dx$$

Let $y = \lambda x \Rightarrow x = \frac{y}{\lambda} \Rightarrow \frac{dx}{dy} = \frac{1}{\lambda}$

$$E(X^2) = \int_{0}^{\infty} \left(\frac{y}{\lambda}\right)^2 \lambda e^{-y} \frac{dy}{\lambda} = \frac{1}{\lambda^2} \int_{0}^{\infty} y^{3-1} e^{-\lambda} = \frac{1}{\lambda^2}\Gamma(3) = \frac{2}{\lambda^2}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

$$\sigma(X) = \frac{1}{\lambda}$$

Memoryless Property

Suppose X is a rv which has an exponential distribution with parameter $\lambda > 0$

Let a, b be two positive real constants.

Consider $P(X > a + b | X > a) = \dfrac{P\big((X > a + b) \cap (X > a)\big)}{P(X > a)}$

$$= \frac{P(X > a + b)}{P(X > a)} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b} = P(X > b)$$

So the exponential distribution has the **memoryless property**, which is

$P(X > a + b | X > a) = P(X > b)$

It turns out that the exponential distribution is the unique continuous distribution possessing this property.

What about discrete distributions?

Verify that the geometric distribution has this property.

Example

A sugar refinery has 3 processing plants which operate independently of one another. The amount of sugar that one plant can process in a day can be modelled by an exponential distribution with a mean of 4 tonnes.

a) Find the probability that exactly 2 of the 3 plants process more than 4 tonnes on a given day.
b) For a particular plant, how much raw sugar should be stocked for that plant each day so that the chance of running out of the product is only 0.05.

a) Let X be the rv indicating the amount of sugar processed in any one plant on any given day.

We know that X is exponential distributed with parameter $\lambda = \frac{1}{4}$

The pdf of X is given by $f(x) = \lambda e^{-\lambda x} = \frac{1}{4} e^{-\frac{x}{4}}$ for $x > 0$

$P(X > 4) = e^{-\lambda(4)} = e^{-1} \cong 0.37$

Let Y be the rv indicating the number of plants producing more than 4 tonnes on a given day. Thus, $Y \sim B_i(3, p = e^{-1})$

Hence $P(Y = 2) = \binom{3}{2}(e^{-1})^2(1 - e^{-1})^{3-2} = 3e^{-2}(1 - e^{-1}) \cong 0.2566$

b) Let a denote the amount of raw sugar to be stocked on a given day.

We want to choose a so that $P(X > a) = 0.05$

$\Rightarrow e^{-\frac{a}{x}} = 0.05 \Rightarrow -\frac{a}{4} = \ln(0.05) \Rightarrow a = -4 \times \ln(0.04) \cong 11.98$

Relationship to the Poisson Process

Suppose that the number of events Y_t occurring in the interval $[0, t]$ is governed by a Poisson process at rate λ per unit time.

i.e. $P(Y_t = y) = \dfrac{e^{-\lambda t}(\lambda t)^y}{y!}$ $for\ y \in \mathbb{N}$

How long do I have to wait to see the very first event occur?

Let X denote the length of time until the 1st event occurs.

$P(X > x) = P(\text{time until the 1st event occurs} > x) = P(Y_x = 0)$

But Y_x is Poisson distributed with mean λx

$P(X > x) = P(Y_x = 0) = \dfrac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}$ $for\ x \geq 0$

$F_{X(x)} = 1 - P(X > x) = 1 - e^{-\lambda x}$ $for\ x \geq 0$

$\Rightarrow f_X(x) = F'_X(x) = \lambda e^{-\lambda x}$ $for\ x > 0$

Thus X is exponentially distributed with mean $\frac{1}{\lambda}$

More generally

The length of time until the nth event occurs after the (n-1)th event occurs is an exponential distribution with mean $\frac{1}{\lambda}$

Gamma Distribution

November-07-11 1:20 PM

PDF

$f(x) = \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}\ for\ x > 0$

$\lambda, \alpha > 0$

Mean and Variance

$E(X) = \frac{\alpha}{\lambda}$

$E(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2}$

$Var(X) = E(X^2) - E(X)^2 = \frac{\alpha}{\lambda^2}$

Property

When $\alpha = 1$, the Gamma distribution becomes the exponential distribution.

Gamma Function

Recall $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx\ for\ \alpha > 0$

Consider $\int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx\ where\ \lambda > 0$

Let $y = \lambda x \Rightarrow x = \frac{y}{\lambda} \Rightarrow dx = \frac{1}{\lambda} dy$

$\int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx = \int_0^{\infty} \left(\frac{y}{\lambda}\right)^{\alpha-1} e^{-y} \frac{dy}{\alpha} = \frac{1}{\lambda^{\alpha}} \int_0^{\infty} y^{(\alpha-1)} e^{-y} dy = \frac{\Gamma(\alpha)}{\lambda^{\alpha}}$

$\frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx = 1 \Rightarrow \int_0^{\infty} \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda}}{\Gamma(\alpha)} dx = 1 = \int_0^{\infty} f(x) dx$

We say that X has a Gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$ if its pdf is of the form

$f(x) = \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}\ for\ x > 0$

Not that if $\alpha = 1$ then we get the exponential distribution (with mean $\frac{1}{\lambda}$) as a special case

For a Gamma distribution, what are $E(X)$ and $Var(X)$?

$E(X) = \int_0^{\infty} x f(x) dx = \int_0^{\infty} x \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx = \int_0^{\infty} \frac{\lambda^{\alpha} x^{\alpha+1-1} e^{-\lambda x}}{\Gamma(\alpha)} dx$

$= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \int_0^{\infty} \frac{\lambda^{\alpha+1} x^{(\alpha+1)-1} e^{-\lambda x}}{\Gamma(\alpha+1)} dx = \frac{\alpha}{\lambda}$

Normal Distribution

November-09-11 12:37 PM

The normal distribution is considered to be the most important one in all of probability and statistics. It turns out that many numerical populations have distributions that can be fit very closely by the "bell-shaped" curve of the normal distribution.

PDF

X is said to have a normal distribution with parameters μ and σ^2 $(-\infty < \mu < \infty,\ \sigma^2 > 0)$ if its pdf is of the form:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty$$

Must have $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = 1$

Derivation is in the textbook

Normal pdf's are symmetric and bell-shaped.

Remark

The parameters μ and σ^2 are the mean and variance, respectively, and completely characterize the shape of the normal distribution.

Notation

$X \sim N(\mu, \sigma^2)$

Standard Normal Distribution

$Z \sim N(0,1)$

The transformation

$$Z = \frac{(X-\mu)}{\sigma} \sim N(0,1)$$

Transforms any normal X into a standard normal random variable.

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Mean and Variance

$$E(Z) = \int_{-\infty}^{\infty} z \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{\infty} g(z) dz, where\ g(-z) = g(z)$$

Since g is odd,

$E(Z) = 0$

$$E(Z) = 0 = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma}$$

$\therefore E(X) = \mu$

Similarly, one can prove that $Var(Z) = 1$ and so $Var(X) = \sigma^2$

CDF and Standard Normal Distribution

Because $f(x)$ is complicated to work with and

$$F(x) = \int_{-\infty}^{\infty} f(t) dt$$

isn't known explicitly we have to rely on a table to compute probabilities for the standard normal distribution.

Properties

What kind of nice properties does the normal distribution possess?

Suppose that X is a random variable and $Y = aX + b,\ a, b \in \mathbb{R}, a \neq 0$

Based on what we already know:

$\mu_Y = E(Y) = a\mu_X$

$\sigma_Y^2 = Var(Y) = a^2\sigma_X^2$

Now, assume $X \sim N(\mu, \sigma^2)$

$\Rightarrow \mu_Y = a\mu + b, \qquad \sigma_Y^2 = a^2\sigma^2$

It turns out that $Y \sim N(\mu_Y, \sigma_Y^2) = N(a\mu + b, a^2\sigma^2)$

Proof

Note that $Y = aX + b = g(X)$ is strictly monotone.

By our earlier result, the pdf of Y is given by

$$f_Y(y) = f_X\big(g^{-1}(y)\big)\left|\frac{d\,g^{-1}(y)}{dy}\right|$$

So $y = ax + b \Rightarrow x = \frac{y-b}{a} = g^{-1}(y)$

$$\frac{dx}{dy} = \frac{1}{a}$$

Therefore, the pdf of Y looks like

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right)\frac{1}{|a|} = \frac{1}{|a|\sigma 2\pi} e^{-\frac{\left(\left(\frac{y-b}{a}\right)-\mu\right)^2}{2\sigma^2}} = \frac{1}{|a|\sigma 2\pi} e^{-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}}$$

$$= \frac{1}{\sigma_Y 2\pi} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}$$

$\therefore Y$ is normal with mean $\mu_Y = a\mu + b$ and $\sigma_Y^2 = a^2\sigma^2$ ∎

Standard Normal Distribution

Suppose that $X \sim N(\mu, \sigma^2)$. Consider the random variable $Z = \frac{X-\mu}{\sigma}$ which we can write as $Z = \frac{1}{\sigma}X + \left(-\frac{\mu}{\sigma}\right)$

According to the previous property, Z has a normal distribution with mean:

$$\mu_Z = a\mu + b = \frac{1}{\sigma}\mu - \frac{1}{\sigma}\mu = 0$$

and variance:

$$\sigma_Z = a^2\sigma^2 = \left(\frac{1}{\sigma}\right)^2 \times \sigma^2 = 1$$

In other words $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

Example

The time it takes to complete a large file transfer across a particular communications link varies according to a normal distribution with mean 266 minutes and standard deviation 16 minutes.

i) What percent of file transfers last between 240 and 270 minutes?
ii) How long do the top 80% of file transfers last?

Let X be the rv indicating the file transfer length. We a told that $X \sim N(266, 16^2)$

i) We want to calculate

$P(240 < X < 270)$ for

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1) \Rightarrow P\left(\frac{240-266}{16} < Z < \frac{270-266}{16}\right)$$

$= P(Z < 0.25) - P(X < -1.63) = 0.54663$

Discrete Multivariate Distributions

November-14-11 12:34 PM

Joint Probability Function

Let X_1, X_2 be discrete random variables. The joint probability function of X_1, X_2 is given by
$p(x_1,x_2) = P(X_1 = x_1, X_2 = x_2)$
$= P(\{X_1 = x_1\} \cap \{X_2 = x_2\})$

Properties

1. $p(x_1,x_2) \geq 0\ \forall x_1, x_2$
2. $\displaystyle\sum_{x_1}\sum_{x_2} p(x_1,x_2) = 1$

Marginal Probability Distribution

The marginal probability distribution of X_1 is defined by its marginal pmf given by

$$p_{X_1}(x_1) = p_1(x_1) = P(X_1 = x_1) = \sum_{x_2} p(x_1,x_2)$$

Similarly

$$p_{X_2} = \sum_{x_1} p(x_1,x_2)$$

Remarks

For n discrete random variables, the joint pmf is given by
$p(x_1,x_2,\dots,x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$

The marginal pmf of X_i is given by

$$p_{X_i}(x_i) = p_i(x_i) = \sum_{x_1}\sum_{x_2}\dots\sum_{x_{i-1}}\sum_{x_{i+1}}\dots\sum_{x_n} p(x_1,x_2,\dots,x_n)$$

Conditional Probability Distributions

If X_1 and X_2 are both discrete random variables with pmf $p(x_1,x_2)$ and marginal pmfs $p_{X_1}(x_1),\ p_{X_2}(x_2)$ respectively then the conditional probability distribution of X_1 given $X_2 = x_2$ is given by

$$p_{X_1|X_2} = P(X_1 = x_1 | X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)}$$
$$= \frac{p(x_1,x_2)}{p_{X_2}(x_2)}$$

Similarly, we could define

$$p_{X_2|X_1}(x_2|x_1) = \frac{p(x_1,x_2)}{p_{X_1}(x_1)}$$

provided that $p_{X_1}(x_1) \neq 0$

Recall

Two events A and B are independent if $P(A \cap B) = P(A)P(B)$ or $P(A|B) = P(A)$

Independent Random Variables

Discrete random variables X_1 and X_2 are independent if
$p(x_1,x_2) = p_{X_1}(x_1)p_{X_2}(x_2)\ \forall x_1, x_2$

Om general, $X_1, \dots, X_n$ discrete random variables are said to be independent iff

$$p(x_1,\dots,x_n) = \prod_{i=1}^{n} p_{X_i}(x_i), \qquad \forall x_1,\dots,x_n$$

In Chapter 4, we concerned ourselves with random variables that produced a single numerical response. However, we are often interested in studying the joint behaviour of 2 or more random variables. To proceed with such a study we must learn how to handle multivariate probability distributions. When only2 random variables are considered, these joint distributions are called **bivariate**. We will discuss the bivariate case in some detail - extensions to more than 2 random variables follow a similar approach.

Example

From a group of 3 seniors, 2 juniors, and 1 freshman, a committee of 2 people is to be randomly formed. Let Y_1 denote the number of seniors and Y_2 the number of juniors on the committee. Find the joint probability mass function of Y_1 and Y_2

We note that $Y_1 = 0, 1\ or\ 2$ and $Y_2 = 0, 1, or\ 2$
For example, to find

$$P(1,1) = \frac{\binom{3}{1}\binom{2}{1}\binom{1}{0}}{\binom{6}{2}} = \frac{2}{5}$$

$$P(2,0) = \frac{\binom{3}{2}\binom{2}{0}\binom{1}{0}}{\binom{6}{2}} = \frac{1}{5}$$

Calculating for all combinations

$Y_2 \setminus Y_1$	0	1	2	ΣY_2
0	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
1	$\frac{2}{15}$	$\frac{2}{5}$	0	$\frac{8}{15}$
2	$\frac{1}{15}$	0	0	$\frac{1}{15}$
ΣY_1	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$	1

The summation columns are the marginal probabilities

Marginal pmf of Y_1:

$$p_1(y_1) = \begin{cases} \frac{1}{5}, & y_1 = 0 \\ \frac{3}{5}, & y_1 = 1 \\ \frac{1}{5}, & y_1 = 2 \end{cases} = \frac{\binom{3}{y_1}\binom{3}{2-y_1}}{\binom{6}{2}}$$

Given that one of the two members on the committee is a junior, find the conditional distribution of the number of seniors on the committee.

$$p_{Y_1|Y_2}(y_1|1) = \frac{p(y_1,1)}{p_{Y_2}(1)} = \frac{8}{15} * \begin{cases} \frac{2}{15}, & y_1 = 0 \\ \frac{2}{5}, & y_1 = 1 \\ 0, & y_1 = 2 \end{cases} = \begin{cases} \frac{1}{4}, & y_1 = 0 \\ \frac{3}{4}, & y_1 = 2 \end{cases}$$

Are Y_1 and Y_2 independent random variables? No.

Example

In an automobile parts company, an average of μ defective parts are produced per shift. The number, X, of defective parts produced (per shift) has a Poisson distribution. An inspector checks all parts prior to shipping them, but there is a $100 \times p\%$ changes that a defective part will ship by undetected. Let Y be the number of defective parts the inspector finds on a shift. Find the conditional distribution of X given $Y = y$

We are given the distribution of the rv X, namely:

$$p_X(x) = P(X = x) = \frac{e^{-\mu}\mu^x}{x!}, \qquad x \in \mathbb{N}$$

We want to find $p_{X|Y}(x|y) = P(X = x|Y = y) = \dfrac{p(x,y)}{p_Y(y)}$

Now, given $X = x$ then $P_{Y|X}(y|x) = P(Y = y|X = x) = \binom{x}{y}(1-p)^y p^{x-y}$

Also, $p_{Y|X}(y|x) = \dfrac{p(x,y)}{p_X(x)} \Rightarrow p(x,y) = p_{Y|X}(y|x)p_X(x) = \binom{x}{y}(1-p)^y p^{x-y} \times \dfrac{e^{-\mu}\mu^x}{x!}$

$$= \left(\frac{x!}{y!\,(x-y)!}\right)(1-p)^y p^{x-y}\left(\frac{e^{-\mu}\mu^x}{x!}\right) = \frac{(1-p)^y p^{x-y} e^{-\mu}\mu^x}{y!\,(x-y)!}, \qquad x \in \mathbb{N}, y \in \{0,1,\dots,x\}$$

What is another way to represent the range of x and y?
$y \in \mathbb{N}, x \in \{y, y+1, y+2, \dots\}$

Next we need to find

$$p_Y(y) = P(Y = y) = \sum_{x=y}^{\infty} p(x,y) = \sum_{x=y}^{\infty} \frac{(1-p)^y p^{x-y} e^{-\mu}\mu^x}{y!\,(x-y)!} = \frac{(1-p)^y e^{-\mu}\mu^y}{y!}\sum_{x=y}^{\infty}\frac{p^{x-y}\mu^{x-y}}{(x-y)!}$$

$$= \sum_{l=0}^{\infty}\frac{(p\mu)^l}{l!} = e^{p\mu} = \frac{\left((1-p)\mu\right)^y e^{-(1-p)\mu}}{y!}$$

We recognize that the marginal distribution of Y is Poisson but with adjusted mean $\mu(1-p)$
Therefore,

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} = \left(\frac{(1-p)^y p^{x-y} e^{-\mu} \mu^x}{y!\,(x-y)!}\right)\left(\frac{y!}{(1-p)^y \mu^y e^{-(1-p)\mu}}\right) = \frac{(\mu p)^{x-y} e^{-\mu p}}{(x-y)!}$$

For $x \in \{y, y+1, \ldots\}$

The conditional rv X given $Y = y$ has the same distribution as $W = V + y$ where $V \sim Poisson(\mu p)$

Recall

Two rvs are independent iff $p(x_1, x_2) = p_{X_1}(x_1) p_{X_2}(x_2)$ for all values of x_1 and x_2

Sum of Random Variables

November-18-11 12:34 PM

Result for Sum of Poisson

If X_i is a Poisson distribution with mean λ_i for $i = 1, 2, \dots, n$ and $X_1, X_2, \dots, X_n$ are independently distributed then

$$T = \sum_{i=1}^{n} X_i \text{ is Poisson distributed with mean } \sum_{i=1}^{n} \lambda_i$$

Result for Sum of Binomial

If X_i is a binomial distribution with parameters n_i and p and $X_1, X_2, \dots, X_m$ are independently distributed then

$$T = \sum_{i=1}^{m} X_i \text{ is a Binomial distribution with parameters } \sum_{i=1}^{n} n_i \text{ and } p$$

Result for Sum of Negative Binomial

If X_i is a negative binomial with parameters r_i and p for $i = 1, 2, \dots, n$ and $X_1, X_2, \dots, X_n$ are independent then

$$T = \sum_{i=1}^{n} X_i \text{ is negative binomial with parameters } \sum_{i=1}^{n} r_i \text{ and } p$$

Conversely

If X is negative binomial with parameters r and p then $X = X_1 + X_2 + \cdots + X_r$ where each $X_i \sim \text{Geometric}(p)$ and independent.

$$E(X) = \sum_{i=1}^{r} E(X_i) = \frac{r}{p}$$

Sum of Poisson

Suppose $X_1 \sim \text{Poisson}(\lambda_i)$, $X_2 \sim \text{Poisson}(\lambda_2)$
Also, X_1, and X_2 are assumed to be independent.
Define $T = X_1 + X_2$
Find the distribution of T.

We know that

$$p(x_1, x_2) = p_{X_1}(x_1) p_{X_2}(x_2) = \frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!}, \qquad for\ x_1, x_2 \in \mathbb{N}$$

So T takes on values in $\mathbb{N}$. For $t = 0, 1, 2, \dots$

$$P(T = t) = P(X_1 + X_2 = t) = \sum_{\substack{all\ x_1, x_2 \\ x_1 + x_2 = t}} p(x_1, x_2) = \sum_{x_1=0}^{t} p(x_1, t - x_1) = \sum_{x_1=0}^{t} \frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2}\lambda_2^{t-x_1}}{(t-x_1)!}$$

$$= e^{-\lambda_1-\lambda_2} \sum_{x_1=0}^{t} \frac{1}{x_1!\,(t-x_1)!} \lambda_1^{x_1}\lambda_2^{t-x_1} = \frac{e^{-\lambda_1-\lambda_2}}{t!} \sum_{x_1=0}^{t} \binom{t}{x_1} \lambda_1^{x_1}\lambda_2^{t-x_1} = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^t}{t!}$$

$$\Rightarrow T = X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$$

In general, suppose $X_1, \dots, X_n$ are independent and $X_i \sim \text{Poisson}(\lambda_i)$, $i = 1, 2, \dots, n$

$$\text{Then } T = \sum_{i=1}^{n} X_i \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right)$$

Sum of Binomial

Suppose $X_1 \sim Bi(n_1, p)$ and $X_2 \sim Bi(n_2, p)$
Find the distribution of $T = X_1 + X_2$

Since X_1 and X_2 are independent,

$$p(x_1, x_2) = p_{X_1}(x_1) p_{X_2}(x_2) = \binom{n_1}{x_1} p^{x_1}(1-p)^{n_1-x_1} \times \binom{n_2}{x_2} p^{x_2}(1-p)^{n_2-x_2}$$

for $x_1 = 0, 1, 2, \dots, n$ and $x_2 = 0, 1, 2, \dots, n$

Thus $T = X_1 + X_2$ takes on values in the set $\{0, 1, 2, \dots, n_1 + n_2\}$
For $t \in \{0, 1, 2, \dots, n_1 + n_2\}$ consider

$$p_T(t) = P(T = t) = \sum_{\substack{all\ x_1, x_2 \\ x_1 + x_2 = t}} p(x_1, x_2) = \sum_{x_1=0}^{t} p(x_1, t - x_1)$$

$$= \sum_{x_1=0}^{t} \binom{n_1}{x_1} p^{x_1}(1-p)^{n_1-x_1} \binom{n_2}{t-x_1} p^{t-x_1}(1-p)^{n_2-t+x_1}$$

$$= p^t(1-p)^{n_1+n_2-t} \sum_{x_1=0}^{t} \binom{n_1}{x_1}\binom{n_2}{t-x_1}$$

$$= \binom{n_1+n_2}{t} p^t (1-p)^{n_1+n_2-t}\ [by\ Hypergeometric\ Identity\ ]$$

Therefore

$$T = X_1 + X_2 \sim Bi(n_1 + n_2, p)$$

Example

Suppose $X_i \sim B_i(n_i, p)$ for $i = 1,2$ independent.
Find the conditional distribution of X_1 given $X_1 + X_2 = m$
The conditional pmf of X_1 given $X_1 + X_2 = m$ is given by

$$P(X_1 = x_1 \mid X_1 + X_2 = m) = \frac{P(X_1 = x \cap X_1 + X_2 = m)}{P(X_1 + X_2 = m)} = \frac{p(x_1, m - x_1)}{P(X_1 + X_2 = m)}$$

$$= \frac{\binom{n_1}{x_1} p^{x_1}(1-p)^{n_1-x_1} \binom{n_2}{m-x_1} p^{m-x_1}(1-p)^{n_2-m+x_1}}{\binom{n_1+n_2}{m} p^m (1-p)^{n_1+n_2-m}} = \frac{\binom{n_1}{x}\binom{n_2}{m-x_1}}{\binom{n_1+n_2}{m}}$$

Hypergeometric with

$$N = n_1 + n_2, \qquad n = m, \qquad r = n_1$$

Exercise

Assume that X_1 is geometric with success probability p and similarly, X_2 is geometric with the same success probability. Assume X_1 and X_2 are independent.
Show that $T = X_1 + X_2$ is negative binomial with parameters $r = 2$ and the same success probability.

The Multinomial Distribution

November-18-11 1:13 PM

PMF

The form of the joint pmf of $(X_1, X_2, \dots, X_n)$ is

$$p(x_1, x_2, \dots, x_k) = \frac{n!}{\prod_{i=1}^{k} x_i!} \prod_{i=1}^{k} p_i^{x_i}$$

$$= \binom{n}{x_1, x_2, \dots, x_n} \prod p_i^{x_i}$$

For $x_i \geq 0, \sum_{i=1}^{k} p_i = 1$ and $\sum_{i=1}^{k} x_i = n$

Note

$k = 2$ gives the binomial distribution.

Multinomial Distribution

Suppose an experiment is repeated a fixed number of times (n). If

1. the repetitions involve the same experiment and are independent of one another;
2. instead of only 2 possible outcomes in a single trial, suppose there are now k possible outcomes;
3. for each $i = 1, 2, \dots, k$ each repetition has some probability p_i of being outcome i;

X_i counts the number of times outcome i occurs (for $i = 1, 2, \dots, k$) then $(X_1, X_2, \dots, X_k)$ has a **multinomial distribution.**

Important Results

1. What is the marginal distribution of X_i?
   $X_i \sim Bi(n, p_i)$
2. What is the distribution of $T = X_{i_1} + X_{i_2} + \cdots + X_{i_r}$ where $(i_1, i_2, \dots, i_k)$ are indices from $(1, 2, \dots, k)$?
   $T \sim Bi\left(n, p_{i_1} + p_{i_2} + \cdots + p_{i_r}\right)$

Example

According to the 2010 census figures, the proportion of adults in Canada associated with 5 age categories were as follows:

Age	Proportion
18-24	0.18
25-34	0.23
35-44	0.16
45-64	0.27
65+	0.16

Given that 2 out of 10 randomly selected adults in a sample are both over 65, what is the probability that the 10 chosen adults will include 2 from 15-24, 3 from 35-44, and 3 from 45-64?

Solution

Let X_i represent the number of adults from group i that are chosen, $i = 1, 2, 3, 4, 5$
By construction, $(X_1, X_2, X_3, X_4, X_5)$ has multinomial distribution with $n = 10$ and
$p_1 = 0.18, p_2 = 0.23, p_3 = 0.16, p_4 = 0.27, p_5 = 0.16$
Want to calculate:

$$P(X_1 = 2, X_2 = 0, X_3 = 3, X_4 = 4 \mid X_5 = 2) = \frac{p(2, 0, 3, 3, 2)}{p_{X_5}(2)}$$

$$= \frac{\binom{10}{2, 0, 3, 3, 2} (0.18)^2 (0.23)^0 (0.16)^3 (0.27)^3 (0.16)^2}{\binom{10}{2} (0.16)^2 (0.84)^8} = 0.0059$$

Alternate Solution

Given $X_5 = 2$ there are 8 other individuals (out of 10 in total) to be accounted for.
For these 8 other individuals, they are either
18-24 group, or 25-34 group, 35-44 group, or 45-64 group. Normalize probabilities.

$$p_1^* = \frac{0.18}{1 - 0.16}, \quad p_2^* = \frac{0.23}{0.84}, \quad p_3^* = \frac{0.16}{0.84}, \quad p_4^* = \frac{0.27}{0.84}$$

$$P(X_1 = 2, X_2 = 0, X_3 = 3, X_4 = 3 \mid X_5 = 2) = P(X_1^* = 2, X_2^* = 0, X_3^* = 3, X_4^* = 3)$$

$$= \binom{8}{2, 0, 3, 3} {p_1^*}^2 {p_2^*}^0 {p_3^*}^3 {p_4^*}^3$$

Continuous Multivariate Distributions

November-21-11 12:57 PM

Motivation Behind this Case

$$f(x) = \lim_{dx \to 0} \frac{F(x+dx) - F(x)}{dx}$$

Similarly, for small values dx and dy

$P(x \le X \le x + dx, y \le Y \le y + dy) \cong f(x,y)dx\, dy$

$$f(x,y) \cong \frac{P(X \le X \le x + dx, y \le Y \le y + dy)}{dx\, dy}$$

More formally,

$$f(x,y) = \lim_{\substack{dx \to 0 \\ dy \to 0}} \frac{P(x \le X \le x + dx, y \le Y \le xy)}{dx\, dy}$$

The joint cumulative distribution function of X and Y is defined as

$F(x,y) = P(X \le x, Y \le y)$

Properties

1. $F(+\infty, +\infty) = \lim_{\substack{x \to \infty \\ y \to \infty}} F(x,y) = 1$
2. $F(x, +\infty) = P(X \le x) = F_X(x)$
3. $F(+\infty, y) = P(Y \le y) = F_Y(y)$

Also

$$F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(y,v)du\, dv \Leftrightarrow f(x,y) = \frac{\partial^2}{\partial x\, \partial y} F(x,y)$$

Conditional Distributions

Consider $P(x \le X \le x + dx\, |y \le Y \le y + dy)$

$$= \frac{P(x \le X \le x + dx, y \le Y \le y + dy)}{P(y \le Y \le y + dy)} \cong \frac{f(x,y)dx\, dy}{f_Y(y)dy} = \frac{f(x,y)}{f_Y(y)} dx$$

$$\therefore \frac{f(x,y)}{f_Y(y)} \cong \frac{P(x \le X \le x + dx \mid y \le Y \le y + dy)}{dx}$$

More formally

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \lim_{\substack{dx \to 0 \\ dy \to 0}} \frac{P(x \le X \le x + dy \mid y \le Y \le dy)}{dx}$$

This is the **conditional pdf** of X given $Y = y$.

Look at relationship between $f(x,y)$ and $f_Y(y)$

Discrete case

$$p_Y(y) = P(Y = y) = \sum_{all\, x} p(x,y)$$

Continuous case

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Independence

Independence in the continuous case is defined as follows.

We say X_1 and X_2 are independent (continuous) random variables iff

$f(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)\ \forall x_1, x_2$

Example

Suppose X and Y are jointly continuous random variables having joint pdf of the form

$$f(x,y) = \begin{cases} \frac{12}{5} x(2 - x - y) \text{ for } 0 < x < 1, 0 < y < 1 \\ 0 \text{ elsewhere} \end{cases}$$

a) Calculate $P(X < Y)$

$$P(X < Y) = \int_0^1 \left(\int_0^y \frac{12}{5} x(2 - x - y)dx \right) dy$$

$$= \frac{12}{5} \int_0^1 \left(\int_0^y (2x - x^2 - xy)dx \right) dy = \frac{12}{5} \int_0^1 \left(x^2 - \frac{x^3}{3} - \frac{x^2 y}{2} \right) \Big|_o^y dy$$

$$= \frac{12}{5} \int_0^1 \left(y^2 - \frac{y^3}{3} - \frac{y^3}{2} \right) dy = \frac{12}{5} \int_0^1 \left(y^2 - \frac{5y^3}{6} \right) dy$$

$$= \frac{12}{5} \left(\frac{y^3}{3} - \frac{5y^4}{24} \right) \Big|_0^1 = \frac{12}{5} \left(\frac{1}{3} - \frac{5}{24} \right) = \frac{12}{5} \left(\frac{1}{8} \right) = \frac{12}{40} = \frac{3}{10}$$

b) Find the marginal pdf of Y.

From our theory, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$

$$f_Y(y) = \int_0^1 \frac{12}{5} (2 - x - y)dx = \frac{12}{5} \left(x^2 - \frac{x^3}{3} - \frac{x^{2y}}{2} \right) \Big|_0^1$$

$$= \frac{12}{5} \left(1 - \frac{1}{3} - \frac{y}{2} \right) = \frac{12}{5} \left(\frac{2}{3} - \frac{y}{2} \right) = \frac{2}{5} (4 - 3y)$$

Verify:

$$\int_0^1 \frac{2}{5} (4 - 3y) = 1$$

c) Find the conditional pdf of X given $Y = y$ where $y \in (0,1)$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{12}{5} x(2 - x - y)}{\frac{2}{5} (4 - 3x)} = \frac{6x(2 - x - y)}{4 - 3y}, \qquad 0 < x < 1$$

Example

Suppose the joint pdf of X and Y is given by

$$f(x,y) = \begin{cases} 5e^{-3x-y}, & 0 < 2x < y < \infty \\ 0 \text{ elsewhere} \end{cases}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = \int_0^{\infty} \left(\int_{2x}^{\infty} 5e^{-3x-y} dy \right) dx = \int_0^{\infty} \left(\int_0^{\frac{y}{2}} 5e^{-3x-y} dx \right) dy$$

$= 1$

Find the conditional pdf of Y given $X = x$ where $x \in (0, \infty)$

We know $f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{5e^{-3x-y}}{f_X(x)}$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{2x}^{\infty} 5e^{-3x-y} dy = 5e^{-3x} \int_{2x}^{\infty} e^{-y} dy$$

$$= 5e^{-3x} (-e^y) \Big|_{2x}^{\infty} = 5e^{-3x} e^{-2x} = 5e^{-5x}$$

Note that the marginal distribution of x is exponential with parameter $\lambda = 5$

Let's also find $f_Y(y), y \in (0, \infty)$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^{\frac{y}{2}} 5e^{-3x-y} dx = 5e^{-y} \int_0^{\frac{y}{2}} e^{-3x} dx$$

$$= 5e^{-y} \left(-\frac{e^{-3x}}{3} \right) \Big|_0^{\frac{y}{2}} = 5e^{-y} \left(-\frac{1}{3} e^{-\frac{3}{2} y} + \frac{1}{3} \right) = \frac{5}{3} e^{-y} \left(1 - e^{-\frac{3y}{2}}\right)$$

Note that

$$f_X(x) f_Y(y) = (5e^{-5x}) \left(\frac{5}{3} e^{-y} \left(1 - e^{-\frac{3y}{2}}\right) \right) \ne f(x,y) = 5e^{-3x-y}$$

Sum of Continuous RVs

November-25-11 12:56 PM

Sum of Gamma/Exponential Distributions

If $\{X_i\}_{i=1}^{n}$ is a sequence of independent and identically distributed Gamma random variables with parameters λ and α_i then

$$T=\sum_{i=1}^{n} X_i \sim \Gamma\left(\sum_{i=1}^{n}\alpha_i, \lambda\right)$$

Sum of Normal Distributions

If $X\sim N(0,1)$ and $Y\sim N(0,\sigma^2)$ then $T=X+Y\sim N(0,\sigma^2+1)$

In General

Suppose $X_1\sim N(\mu_1,\sigma_1^2)$, $X_2\sim(\mu_2,\sigma_2^2)$ independent

Then $T=X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$

More Generally

$\{X_i\}_{i=1}^{n}$ are independent $X_i\sim N(\mu_i,\sigma_i^2)$, $a_i\in\mathbb{R}$ then

$$T=\sum_{i=1}^{n} a_iX_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

Implication

$X_1\sim N(\mu_1,\sigma_1^2), X_2\sim N(\mu_2,\sigma_2^2)$ independent

$X_1-X_2\sim N(\mu_1-\mu_2,\sigma_1^2+\sigma_2^2)$

Example

Suppose X and Y are independent random variables where X and Y are exponentially distributed with parameter λ. Find the distribution of $T=X+Y$

By the independence assumption, the joint pdf of X and Y is $f(x,y)=f_X(x)f_Y(y)=(\lambda e^{-\lambda x})(\lambda e^{-\lambda y})=\lambda^2 e^{-\lambda(x+y)}$ for $x,y>0$

Define $F_T(t)=P(T\le t)$

Note: T can take on values in the interval $(0,\infty)$

Let $t\in(0,\infty)$ and consider

$$F_T(t)=P(T\le t)=P(X+Y\le t)=\int_0^t\left(\int_0^{x-t}\lambda^2 e^{-\lambda(x+y)}dy\right)dx$$

$$=\lambda^2\int_0^t e^{-\lambda x}\left(\int_0^{x-t}e^{-\lambda y}dy\right)dx=\lambda^2\int_0^t e^{-\lambda x}\left[-\frac{e^{-\lambda y}}{\lambda}\right]_0^{x-t}dx$$

$$=\lambda\int_0^t e^{-\lambda x}(1-e^{-\lambda(t-x)}dx=\lambda\left(\int_0^t e^{-\lambda x}dx-\int_0^t e^{-\lambda t}dx\right)=\lambda\left(\left[-\frac{e^{-\lambda x}}{\lambda}\right]_0^t-te^{-\lambda t}\right)$$

$$=1-e^{-\lambda t}-\lambda te^{-\lambda t}, for\ t>0$$

$$f_T(t)=F_T'(t)=\frac{d}{dt}\left(1-e^{-\lambda t}-\lambda te^{-\lambda t}\right)=\lambda e^{-\lambda t}+\lambda^2te^{-\lambda t}-\lambda e^{-\lambda t}$$

$$=\frac{\lambda^2t^{2-1}e^{-\lambda t}}{\Gamma(2)}\quad for\ t>0$$

Alternate PDF method

$$\int_0^t f(x,t-x)dx=\int_0^t\lambda^2e^{-\lambda(x+t-x)}dx=\int_0^t\lambda^2e^{-\lambda t}dx=\lambda^2te^{-\lambda t}$$

We recognize that

$T=X+Y$ is Gamma distributed with parameters $\alpha=2$ and λ

Example

Suppose X and Y are independent random variables such that $X\sim N(0,1)$ and $Y=N(0,\sigma^2)$ where $\sigma^2>0$

By independence we know that

$$f(x,y)=f_X(x)f_Y(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y^2}{2\sigma^2}}=\frac{1}{2\pi\sigma}e^{-\frac{x^2}{2}}e^{-\frac{y^2}{2\sigma^2}}\ for\ -\infty<x,y<\infty$$

$$F_T(t)=P(X+Y\le t)=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{t-x}\frac{1}{2\pi\sigma}e^{-\frac{x^2}{2}}e^{-\frac{y^2}{2\sigma^2}}dy\right)dx$$

$$=\frac{1}{2\pi\sigma}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}\left(\int_{-\infty}^{t-x}e^{-\frac{y^2}{2\sigma^2}}dy\right)dx$$

Let $u=\frac{y}{\sigma}, y=\sigma u, \frac{dy}{du}=\sigma$

$$F_T(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}\cdot\left(\int_{-\infty}^{\frac{t-x}{\sigma}}e^{-\frac{u^2}{2}}du\right)dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}\Phi\left(\frac{t-x}{\sigma}\right)dx,$$

where $\Phi(z)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{z}}du$ is the cdf of $N(0,1)$

$$F_T(t)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\Phi\left(\frac{t-x}{\sigma}\right)dx, \text{where } \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

Let $w=\Phi\left(\frac{t-x}{\sigma}\right), \frac{dw}{dx}=\Phi'\left(\frac{t-x}{\sigma}\right)\times\left(-\frac{1}{\sigma}\right)$

$$dw=-\Phi'\left(\frac{t-x}{\sigma}\right)\frac{dx}{\sigma}=-f_X\left(\frac{t-x}{\sigma}\right)\frac{dx}{\sigma}$$

Alternate

$$f(t)=\int_{-\infty}^{\infty}f(x,t-x)dx=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma}e^{-\frac{1}{2\sigma^2}(\sigma^2x^2+(t-x)^2)}dx$$

Reset

$$F_T(t)=P(T\le t)=P(X+Y\le t)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_X(x)f_Y(t)dy\,dx$$

$$=\int_{-\infty}^{\infty}f_X(x)\left(\int_{-\infty}^{t-x}f_Y(y)dy\right)dx=\int_{-\infty}^{\infty}F_Y(t-x)f_X(x)dx$$

We know that the pdf of T is given by $f_T(t)=F_T'(t)$

$$\Rightarrow f_T(t)=\frac{d}{dt}\left(\int_{-\infty}^{\infty}F_Y(t-x)f_X(x)dx\right)=\int_{-\infty}^{\infty}\frac{d}{dx}(F_Y(t-x))f_X(x)dx$$

$$=\int_{-\infty}^{\infty}f_Y(t-x)f_X(x)dx$$

Claim

$$f_T(t)=\int_{-\infty}^{\infty}f_Y(t-x)f_X(x)dx$$

Proof

$$F_T(t)=P(T\le t)=\int_{-\infty}^{t}f_T(w)dw=\int_{-\infty}^{t}\int_{-\infty}^{\infty}f_Y(w-x)f_X(x)dx\,dw$$

$$=\int_{-\infty}^{\infty}f_X(x)\left[\int_{-\infty}^{\infty}f_Y(w-x)dw\right]dx=\int_{-\infty}^{\infty}f_X(x)\left[\int_{-\infty}^{t-x}f_Y(u)du\right]dx$$

$$=\int_{-\infty}^{\infty}F_Y(t-x)f_X(x)dx$$

Recall

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad -\infty < x < \infty$$

$$f_Y(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}}$$

$$f_Y(t-x)f_X(x) = \frac{1}{2\pi\sigma} e^{-\frac{(t-x)^2}{2\sigma^2}} \cdot e^{-\frac{x^2}{2}} = \frac{1}{2\pi\sigma}\exp\left(-\frac{1}{2}\left(x^2 + \frac{(t-x)^2}{\sigma^2}\right)\right)$$

$$= \frac{1}{2\pi\sigma}\exp\left(-\frac{1}{2}\left(x^2 + \frac{t^2}{\sigma^2} - \frac{2tx}{\sigma^2} + \frac{x^2}{\sigma^2}\right)\right) = \frac{1}{2\pi\sigma}\exp\left(-\frac{1}{2}\left(\left(1+\frac{1}{\sigma^2}\right)x^2 - \frac{2tx}{\sigma^2} + \frac{t^2}{\sigma^2}\right)\right)$$

$$= \frac{1}{2\pi\sigma}\exp\left(-\frac{1}{2\left(\frac{1}{c}\right)}\left(x^2 - \frac{2t}{c\sigma^2}x + \frac{t^2}{c\sigma^2}\right)\right), \qquad c = \left(1 + \frac{1}{\sigma}\right)$$

$$= \frac{1}{2\pi\sigma}\exp\left(-\frac{1}{2\left(\frac{1}{c}\right)}\left[x^2 - \frac{2t}{c\sigma^2}x + \frac{t^2}{c^2\sigma^4} - \frac{t^2}{c^2\sigma^4} + \frac{t^2}{c\sigma^2}\right]\right)$$

$$= \frac{1}{2\pi\sigma}\exp\left(-\frac{1}{2\left(\frac{1}{c}\right)}\left(x - \frac{t}{c\sigma^2}\right)^2\right)\exp\left(-\frac{1}{2\left(\frac{1}{c}\right)}\left(\frac{t^2}{c\sigma^2} - \frac{t^2}{c^2\sigma^4}\right)\right)$$

$$= \frac{1}{\sqrt{2\pi}\left(\frac{1}{\sqrt{c}}\right)}\exp\left(-\frac{\left(x - \frac{t}{c\sigma}\right)^2}{2\left(\frac{1}{c}\right)}\right) \cdot \Bigg(\frac{\left(\frac{1}{\sqrt{c}}\right)}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\left(\frac{1}{c}\right)}\left(\frac{t^2}{c\sigma^2} - \frac{t^2}{c^2\sigma^4}\right)\right)$$

$$f_T(t) = \int_{-\infty}^{\infty} f_Y(t-x)f_X(x)dx$$

$$= \frac{\frac{1}{\sqrt{c}}}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\left(\frac{1}{c}\right)}\left(\frac{t^2}{c\sigma^2} - \frac{t^2}{c^2\sigma^4}\right)\right)\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\left(\frac{1}{\sqrt{c}}\right)}\exp\left(-\frac{\left(x - \frac{t}{c\sigma^2}\right)^2}{2\left(\frac{1}{c}\right)}\right)dx$$

$$\frac{1}{\sqrt{2\pi}\left(\frac{1}{\sqrt{c}}\right)}\exp\left(-\frac{\left(x - \frac{t}{c\sigma^2}\right)^2}{2\left(\frac{1}{c}\right)}\right) = \text{pdf of } N\left(\frac{t}{c\sigma^2}, \frac{1}{c}\right)$$

$$f_T(t) = \frac{1}{\sqrt{2\pi}\sqrt{c}\sigma}\exp\left(-\frac{t^2}{2\sigma^2}\left(1 - \frac{1}{c\sigma^2}\right)\right)$$

But $c = 1 + \frac{1}{\sigma^2},\ c\sigma^2 = \sigma^2 + 1$

$$1 - \frac{1}{c\sigma^2} = 1 - \frac{1}{\sigma^2+1} = \frac{\sigma^2+1-1}{\sigma^2+1} = \frac{\sigma^2}{\sigma^2+1}$$

$$f_T(t) = \frac{1}{\sqrt{2\pi c}\sigma}\exp\left(-\frac{t^2}{2(\sigma^2+1)}\right)$$

$$\sqrt{c}\sigma = \sigma\sqrt{1 + \frac{1}{\sigma^2}} = \sqrt{\sigma^2+1}$$

$$f_T(t) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2+1}} e^{-\frac{t^2}{2(\sigma^2+1)}} \text{ pdf of } N(0, \sigma^2+1), \qquad t \in (-\infty, \infty)$$

General Case

Suppose $X_1 \sim N(\mu_1, \sigma_1^2),\ X_2 \sim (\mu_2, \sigma_2^2)$ independent

Consider $T = X_1 + X_2$

$$T = X_1 + X_2 = \sigma_2\left(\frac{X_1 - \mu_1}{\sigma_2} + \frac{X_2 - \mu_2}{\sigma_2}\right) + \mu_1 + \mu_2 = \sigma_2(Y + X) + \mu_1 + \mu_2$$

Where $X = \dfrac{X_2 - \mu_2}{\sigma_2} \sim N(0,1), \qquad Y = \dfrac{X_1 - \mu_1}{\sigma_2} \sim N\left(0, \dfrac{\sigma_1^2}{\sigma_2^2}\right)$ independent

By the above proof, $X + Y \sim N\left(0, \dfrac{\sigma_1^2}{\sigma_2^2} + 1\right)$

$$\therefore T \sim N\left(0 + \mu_1 + \mu_2, \left(\frac{\sigma_1^2}{\sigma_2^2} + 1\right)\sigma^2\right) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Extending Expectation to Joint Distributions

November-30-11 12:42 PM

Expected Value for Joint Distributions

Let X and Y be random variables having either a joint pmf $p(x,y)$ or a joint pdf $f(x,y)$. The expected value of any real-valued function of (X,Y), say $g(X,Y)$, is

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)p(x,y) \text{ if } X,Y \text{ are jointly discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dx\,dy \text{ if } X,Y \text{ are jointly continuous} \end{cases}$$

Obviously, this idea of expected value can naturally be extended to more than 2 random variables.

Linear Combinations

Consider $g(X,Y) = ah_1(X,Y) + bh_2(X,Y)$ where $h_{1,2}$ are real-valued functions.
$E[g(X,Y)] = aE[h_1(X,Y)] + bE[h_2(X,Y)]$

General Property

If $g(X_1, X_2, \dots, X_n) = \sum_{i=1}^{m} a_i h_i(X_1, X_2, \dots, X_n)$ then we have

$E[g(X_1, X_2, \dots, X_n)] = \sum_{i=1}^{m} a_i E[h_i(X_1, X_2, \dots, X_n)]$

If $g(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n} X_i$, $\quad E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$

Covariance

$g(X,Y) = (X-\mu_x)(Y-\mu_Y)$
$E[g(X,Y)] = E\left((X-\mu_x)(Y-\mu_y)\right) = Cov(X,Y)$

Note

$Cov(X,X) = E[(X-\mu_x)(X-\mu_x)] = E[(X-\mu_x)^2] = Var\,(X)$

Recall

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

Similarly

$Cov(X,Y) = E\big[(X-\mu_x)(Y-\mu_y)\big] = E[XY - \mu_Y X - \mu_X Y + \mu_X\mu_Y]$
$= E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y = E(XY) - \mu_X\mu_Y$

$\boxed{Cov(X,Y) = E(XY) - \mu_X\mu_Y}$

Recall

$Var\,(X) \geq 0$
What about $Cov(X,Y)$? $Cov(X,Y)$ can be negative, positive, and even equal to 0.

What can we say about the relationship between X and Y based on the value $Cov(X,Y)$?
$Cov(X,Y) > 0 \Rightarrow$X and Y have a positive association
$Cov(X,Y) < 0 \Rightarrow$X and Y have a negative association
$Cov(X,Y) = 0 \Rightarrow$X and Y are uncorrelated

Proof of Linear Combinations

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)p(x,y) \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)p(x,y)dx\,dy \end{cases}$$

$$= \begin{cases} \sum_x \sum_y ah_1(x,y)p(x,y) + bh_2(x,y)p(x,y) \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ah_1(x,y)p(x,y) + bh_2(x,y)p(x,y)dx\,dy \end{cases}$$

$$= \begin{cases} a\sum_x \sum_y h_1(x,y)p(x,y) + b\sum_x \sum_y h_2(x,y)p(x,y) \\ a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(x,y)p(x,y)dx\,dy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(x,y)p(x,y)dx\,dy \end{cases}$$

$= aE[h_1(X,Y)] + bE[h_2(X,Y)]$

Example

If $X \sim Bi(n,p)$ then $p(x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, \dots, n$

and $E(X) = np = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x}$

Instead of this approach, consider

$Y_i = \begin{cases} 0 \text{ if } i^{th} \text{ trial is FAILURE} \\ 1 \text{ if } i^{th} \text{ trial is SUCCESS} \end{cases}$

$P(Y_i = 0) = 1-p$ and $P(Y_i = 1) = p$
$E(Y_i) = 0 \times (1-p) + 1 \times p = p$

Note

If $X \sim Bi(n,p), X = \sum_{i=1}^{n} Y_i$

$\therefore E(X) = E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} p = np$

Example

If $Y \sim Gamma(x,\lambda)$ then $Y = \sum_{i=1}^{n} X_i$ where $X_i \sim Exponential(\lambda)$

$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{1}{\lambda} = \frac{n}{\lambda}$

Correlation

December-02-11 12:36 PM

Problem: 7.4, 7.6, 7.9, 7.11, 7.20, 7.31, 7.32, 7.36, 7.38, 7.29, 7.40, 7.45
Theoretical: 7.22, 7.23
Self-test: 7.2, 7.4, 7.22, 7.23, 7.29

Covariance

$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X\mu_Y$

$Cov(X,Y) > 0 \Rightarrow$ X and Y have positive association
$Cov(X,Y) < 0 \Rightarrow$ X and Y have negative association
$Cov(X,Y) = 0 \Rightarrow$ X and Y are uncorrelated.

X, Y independent $\Rightarrow Cov(X,Y) = 0$

Correlation Coefficient

$\rho = \dfrac{Cov(X,Y)}{\sigma_x\sigma_y}$

The correlation coefficient measures the direction of the association, but it also tells us the strength of the association.
First of all, ρ is a unitless quantity.
Also $-1 \le \rho \le 1$

Clearly, if X and Y are independent, then $Cov(X,Y) = 0$ and so $\rho = 0$ in this case.

Also, ρ measures the strength of the **linear association** between X and Y.
i.e. if $Y = aX + b$ then $\rho = \pm 1$
If $a,0$ then $a = -1$ and if $a > 0$ then $p = 1$

Independent Random Variables

Let X and Y be independent random variables. Then, we know that $p(x,y) = p_X(x)p_Y(y)$ (if X and Y are both discrete) or $f(x,y) = f_X(x)f_Y(y)$ (if X and Y are both continuous)

WLOG, suppose X and Y are both continuous and consider
$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy = \left(\int_{-\infty}^{\infty} xf_X(x)dx\right)\left(\int_{-\infty}^{\infty} yf_Y(y)dy\right)$
$= E(X)E(Y)$
$Cov(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$
∴Independent ⇒ Uncorrelated
But note that Uncorrelated ⇏ Independent

Example

Suppose the joint pdf of X and Y is

$$f(x,y) = \begin{cases} \frac{1}{2}x(2-x-y) \; for \; 0 < x,y < 1 \\ 0 \; elsewhere \end{cases}$$

What is the value of the correlation coefficient?

$E(X) = \int_{-\infty}^{\infty}\int_{\infty}^{\infty} xf dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \times \frac{12}{5}x(2-x-y)dx\,dy = \frac{3}{5}$

$E(X^2) = \frac{21}{50}, \sigma_X^2 = E(X^2) - E(X)^2 = \frac{3}{50}$

$E(Y) = \frac{2}{5}, E(Y^2) = \frac{7}{30}, Var(Y) = \frac{7}{30} - \left(\frac{2}{5}\right)^2 = \frac{11}{150}$

$E(XY) = \int_0^1\int_0^1 xy\frac{12}{5}x(2-x-y)dx\,dy = \frac{7}{30}$

$Cov(X,Y) = \frac{7}{30} - \left(\frac{3}{5}\right)\left(\frac{2}{5}\right) = -\frac{1}{150}$

$\therefore \rho = \dfrac{Cov(X,Y)}{\sigma_X\sigma_Y} = \dfrac{-\frac{1}{150}}{\sqrt{\frac{3}{50}}\sqrt{\frac{11}{150}}} \cong -0.100504$

It seems that there is very little evidence to suggest a linear relationship exists between X and Y.

Linear Combinations

December-05-11 12:49 PM

Variance/Expectation Linear Combinations

$T = aX + bY$

$E(T) = aE(X) + bE(Y)$
$Var(T) = a^2Var(X) + b^2Var(Y) + 2ab \cdot Cov(X,Y)$

If X and Y are independent
$Var(T) = a^2Var(X) + b^2Var(Y)$

Covariance, Variance Linear Combinations

$$Cov\left(\sum_{i=1}^{m} a_iX_i, \sum_{j=1}^{n} b_jY_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_ib_j \cdot Cov(X_iY_j)$$

$$Var\left(\sum_{i=1}^{n} a_iX_i\right) = Cov\left(\sum_{i=1}^{n} a_iX_i, \sum_{j=1}^{n} a_jX_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_jCov(X_i,X_j)$$
$$= \sum_{i=1}^{n} a_i^2\, Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_ia_j \cdot Cov(X_i,X_j)$$

If X_i are independent then
$$Var\left(\sum_{i=1}^{n} a_iX_i\right) = \sum_{i=1}^{n} a_i^2Var(X_i)$$

Consider

$T = aX + bY$ where a, b are constants
$E(T) = aE(X) + bE(Y)$
What about
$$Var(T) = E[(T-\mu_T)^2] = E\left[\left(aX + bY - aE(X) - bE(Y)\right)^2\right]$$
$$= E\left[\left(a(X-E(X)) + b(Y-E(X))\right)^2\right]$$
$$= E\left[a^2(X-E(X))^2 + b^2(Y-E(Y))^2 + 2ab(X-E(X))(Y-E(Y))\right]$$
$$= a^2E\left((X-E(X))^2\right) + b^2E\left((Y-E(Y))^2\right) + 2abE\left((X-E(X))(Y-E(Y))\right)$$
$$= a^2Var(X) + b^2Var(Y) + 2ab \cdot Cov(X,Y)$$

If X, Y are independent then
$Var(T) = a^2Var(X) + b^2Var(Y)$

Covariance

$X = aX_1 + bX_2, Y = cY_1 + dY_2$
$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= E\left[(aX_1 + bX_2 - -a\mu_{X_1} - b\mu_{X_2})(cY_1 + dY_2 - c\mu_{Y_1} - d\mu_{Y_2})\right]$$
$$= E\left[\left(a(X_1-\mu_{X_1}) + b(X_2-\mu_{X_2})\right)\left(c(Y_1-\mu_1) + d(Y_2-\mu_2)\right)\right]$$
$$= E\left[ac(X_1-\mu_{X_1})(Y_1-\mu_{Y_1}) + ad(X_1-\mu_{X_1})(Y_2-\mu_{Y_2}) + bc(X_2-\mu_{X_2})(Y_1-\mu_{Y_1}) + bd(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})\right]$$
$$= ac \cdot Cov(X_1,Y_1) + ad \cdot \mathrm{Cov}(X_1,Y_2) + bc \cdot Cov(X_2,Y_1) + bd \cdot Cov(X_2,Y_2)$$

Example: Indicator Random Variables

Recall the hypergeometric distribution with pmf
$$p(X) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$
Earlier $E(X) = n\left(\frac{r}{N}\right), Var(X) = n\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$
Instead, let's define the following indicator random variables
$$X_i = \begin{cases} 1 \text{ if the } i^{th} \text{ selection is of success type} \\ 0 \text{ if the } i^{th} \text{ selection is of failure type}\end{cases}$$
Then, our hypergeometric rv is expressed as
$$X = \sum_{i=1}^{n} X_i, \qquad (X_i \text{ are dependent})$$
$$P(X_i = 1) = \frac{(N-1)^{(n-1)}r}{N^{(n)}} = \frac{(N-1)^{(n-1)}r}{N(N-1)^{(n-1)}} = \frac{r}{N}$$
$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n}\left[1 \cdot \frac{r}{N} + 0 \cdot \left(1-\frac{r}{N}\right)\right] = n\left(\frac{r}{N}\right)$$

$$Var(X_i) = E(X_i^2) - E(X_i)^2 = \left[1^2 \cdot \frac{r}{N} + 0^2\left(1-\frac{r}{N}\right)\right] - \left(\frac{r}{N}\right)^2 = \frac{r}{N} - \left(\frac{r}{N}\right)^2 = \frac{r}{N}\left(1-\frac{r}{N}\right)$$
$$= \frac{r}{N}\left(\frac{N-r}{N}\right)$$
$$Cov(X_i,X_j) = E(X_iX_j) - E(X_i)E(X_j) = \sum_{x_i=0}^{1}\sum_{x_j=0}^{1} x_ix_jp(x_i,x_j) - \left(\frac{r}{N}\right)^2$$
$$= P(X_i = 1, X_j = 1) - \left(\frac{r}{N}\right)^2$$

Assuming that $i \neq j$
$$P(X_i = 1, X_j = 1) = (N-2)^{(n-2)}r(r-1) = \frac{r(r-1)}{N(N-1)}$$
$$\Rightarrow Cov(X_i,X_j) = \frac{r(r-1)}{N(N-1)} - \left(\frac{r}{N}\right)^2 = \frac{r}{N}\left(\frac{r-1}{N-1} - \frac{r}{N}\right) = \frac{r(Nr - N - rN + r)}{N^2(N-1)}$$
$$= -\frac{r(N-r)}{N^2(N-1)}$$

$$Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(X_i,X_j)$$
$$= \frac{nr}{N} - 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{r(N-r)}{N^2(N-1)} = \frac{nr}{N} - 2\binom{n}{2}\frac{r(N-r)}{N^2(N-1)}$$