STA

2001

STA 2001 Probability and Statistics I

Dr. CHEN, Yilun.

I. Probability

1.1 Properties of Probability.

- **Fundamental Concepts.**

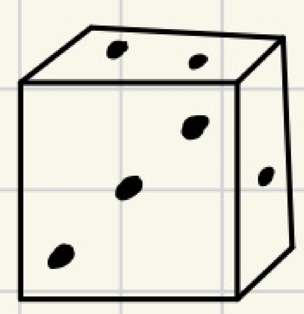
Experiment: Any procedure that can be infinitely repeated and has a well-defined set of possible outcomes.

Random Experiment: An experiment is said to be random if it has more than one possible outcomes

Sample Space: The collections of all possible outcomes of a random experiment. Denoted by S

Event: Given S, an event A is a set that satisfies $A \subseteq S$

e.g. Throwing a fair 6-sided die (A random experiment).
Sample space $S = \{1,2,3,4,5,6\}$.
Event $A = \{1,2\}$.



- **An Intuitive Def. of Prob.**

① Repeat the experiment n times
② Count the number of time that event A occurs, $N(A)$.
$\Rightarrow \frac{N(A)}{n}$ is called the relative frequency of event A in n repetitions of the experiment.

e.g. $S = \{1,2,3,4,5,6\}$. $A = \{1,2\}$.
$\frac{N(A)}{n} \to \frac{1}{3}$, as $n \to \infty$

We have $P(A) = \lim_{n\to\infty} \frac{N(A)}{n}$ ☆

- **Set Theory** (Algebra of Sets)

Set: A collection of distinct elements.

$\emptyset$: The null or empty set.

$A \subseteq B$: A is a subset of B

$A \cup B$: The union of A and B.

$A \cap B$: The intersection of A and B.

A': The complement of A in S is the set of all elements in S that are not in A.

$A_1, A_2, \dots, A_k$ are said to be:

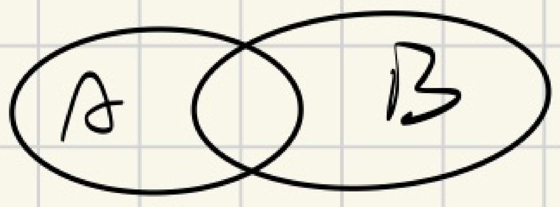
① mutually exclusive if $A_i \cap A_j = \emptyset$, $i \neq j$.

② exhaustive if $A_1 \cup A_2 \cup \cdots \cup A_k = S$.

③ mutually exclusive and exhaustive if ① & ② holds.

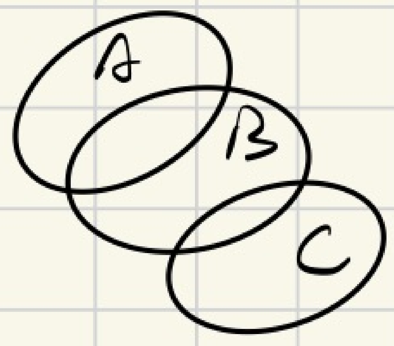
Commutative laws:

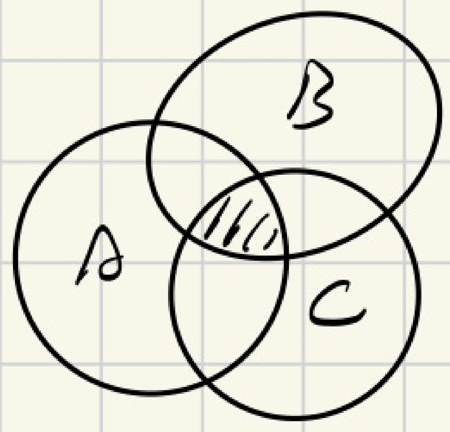
$A \cup B = B \cup A$. $A \cap B = B \cap A$.



Associative law:

$(A \cup B) \cup C = A \cup (B \cup C)$. $(A \cap B) \cap C = A \cap (B \cap C)$





Distributive law:

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

De Morgan's law:

$(A \cup B)' = A' \cap B'$, $(A \cap B)' = A' \cup B'$.

· Definition of Probability (Probability Axioms).

Def. A real-valued set function P that assigns to each event A in the sample space S, a number $P(A)$, called the probability of the event A such that the following:

① $P(A) \geq 0$.

② $P(S) = 1$.

③ If $A_1, A_2, A_3 \dots$ are countable and mutually exclusive events,

$P(A_1 \cup A_2 \cup A_3 \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$

or $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

· Properties of Probability.

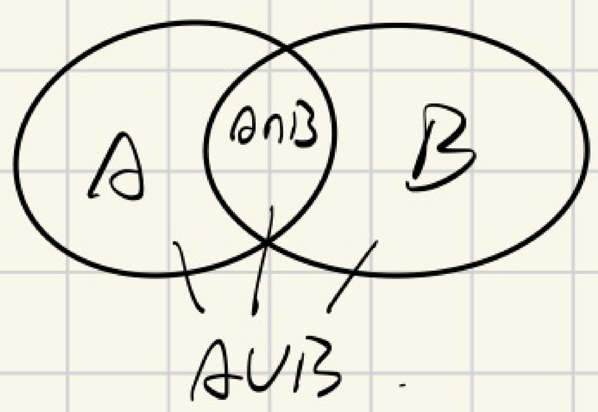
① For each event A, $P(A) = 1 - P(A')$.

② $P(\emptyset) = 0$.

③ If $A \subseteq B$, then $P(A) \leq P(B)$.

④ For each event A, $0 \leq P(A) \leq 1$.

⑤ For any A & B. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.



· Probability Space

1.2 Method of Enumeration

- **Motivation**

For some cases, to define and calculate $P(A)$ can be converted to count the number of outcomes in A → counting techniques.

e.g. $S = \{e_1, e_2, e_3, \dots, e_m\}$.
$P(\{e_k\}) = \frac{1}{m}$, $k = 1, 2, \dots, m$ (equally likely).
Then $P(A) = \frac{N(A)}{N(S)}$, where $N(X)$ is the number of outcomes in $X \subseteq S$.
We need to calculate $N(A)$ & $N(S)$, enumberation is one of the methods.

- **Multiplication Principle**

Consider that an experiment E can be done by a sequential implementation of 2 sub-experiments E_1 & E_2.
→ E_1 → n_1 outcomes
→ E_2 → n_2 outcomes
→ E_1 → E_2 → $n_1 \cdot n_2$ possible outcomes.

- **Permutation of n objects**

Consider that n positions are to be filled with n different objects.
→ pos. 1 → pos. 2 → ⋯ → pos. n
$n \times n-1 \times \cdots \times 1$
in total $n! = n(n-1)\cdots 2 \cdot 1$ arrangements.

Def. Each of the $n!$ arrangement of n different object is called a permutation of n objects

Consider that only r positions are to be filled with objects selected from n different objects.
→ pos. 1 → pos. 2 → ⋯ → pos. r
$n \times n-1 \times \cdots \times n-r+1$
in total $_nP_r = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$ arrangements.

Def. Each of the $_nP_r$ arrangements is called a permutation of n objects taken r a time

e.g. The num of possible 4-English letter words with different letters:
$_{26}P_4 = 26 \times 25 \times 24 \times 23 = \frac{26!}{22!}$

- **Ordered Sample and Ordered Sampling**

Def. If r objects are selected from a set of n different objects and if the order of selection is noted, then the selected set of r objects is called ordered sample of size r.

Ordered sampling without replacement occurs when an object is not replaced after it has been selected ($_nP_r$).

Ordered sampling with replacement occurs when an object is selected and then replaced before the next object is selected (n^r).

- Combination of n objects

Combination is a problem of unordered sampling without replacement

$\to$ pos.1 $\to$ pos.2 $\to \cdots \to$ pos.r $\to$ nPr $\to$ unordered subset of r $\times$ permutation of r objects

$X \qquad\qquad r!$

$\Rightarrow X \cdot r! = nPr \Rightarrow X = \frac{nPr}{r!} = \frac{n!}{r!(n-r)!} \triangleq nCr$

$nCr = \binom{n}{r} = \binom{n}{n-r} = nC_{n-r}$

Def. Each of the nCr unordered subsets is called a combination of n objects taken r at a time.

- Distinguishable Permutation of objects

① Two types:

Consider permutation of n objects of two types: r of one type and $(n-r)$ of the other type.

$\to$ pos.1 $\to$ pos.2 $\to \cdots \to$ pos.n $\to n!$

$\to$ permutation of n objects of two types $\times$ permutation of r objects of one type ($r!$) $\times$ permutation of $(n-r)$ objects of the other type. ($(n-r)!$)

X

$\Rightarrow n! = X \cdot r! \cdot (n-r)! \Rightarrow X = nCr = \binom{n}{r}$

Def. Each of the nCr permutations of n objects of two types with r of one type and $(n-r)$ of the other type

Remark: The number $\binom{n}{r}$ is often called binomial coefficients, because in binomial expansion:

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

② m types

Consider a set of n objects of m types: n_1 of one type, n_2 of one type ... n_m of one type where $n_1 + n_2 + \cdots + n_m = n$.

permutation of n objects of m types $X = \frac{n!}{n_1! \cdot n_2! \cdots n_m!}$

which is sometimes called the multinomial coefficient.

- Unordered Sampling with Replacement

The number of unordered samples of size r selected from a set of n different objects when sampling with replacement is the distinguishable permutation of $n+r-1$ objects of two types with r objects of one type and $n-1$ objects of the other type.

which is $\binom{n+r-1}{r} = \binom{n+r-1}{n-1}$

REMARK Although not needed as often in the study of probability, it is interesting to count the number of possible samples of size r that can be selected out of n objects when the order is irrelevant and when sampling with replacement. For example, if a six-sided die is rolled 10 times (or 10 six-sided dice are rolled once), how many possible unordered outcomes are there? To count the number of possible outcomes, think of listing r 0's for the r objects that are to be selected. Then insert $(n-1)$ |'s to partition the r objects into n sets, the first set giving objects of the first kind, and so on. So if $n = 6$ and $r = 10$ in the die illustration, a possible outcome is

0 0 | | 0 0 0 | 0 | 0 0 0 | 0, two types {0, |}

which says there are two 1's, zero 2's, three 3's, one 4, three 5's, and one 6. In general, each outcome is a permutation of r 0's and $(n-1)$ |'s. Each distinguishable permutation is equivalent to an unordered sample. The number of distinguishable permutations, and hence the number of unordered samples of size r that can be selected out of n objects when sampling with replacement, is

$${}_{n-1+r}C_r = \frac{(n-1+r)!}{r!\,(n-1)!}.$$

1.3 Conditional Probability

Definition

The conditional probability of an event A, given that the event B has occurred, is defined by

$$\triangle\ P(A|B) = \frac{P(A\cap B)}{P(B)} \qquad (P(B)>0)$$

B is the sample space for $P(A|B)$.

Properties:

① $P(A|B) \geq 0$

② $P(B|B) = 1$

③ If $A_1, A_2, A_3 \dots$ are countable and mutually exclusive events, then

$P(A_1 \cup A_2 \cup \dots A_n | B) = P(A_1|B) + P(A_2|B) + \dots + P(A_n)$.

Multiplication Rule

Def. The probability that two events, A and B both occur is given by the multiplication rule:

$P(A\cap B) = P(A)\cdot P(B|A)$. $\quad (P(A)>0)$.

or $P(B\cap A) = P(B)\cdot P(A|B)$. $\quad (P(B)>0)$.

For three events:

Def. The probability that three events, A, B and C all occur is given by the multiplication rule

$P(A\cap B\cap C) = P((A\cap B)\cap C) = P(A\cap B)\cdot P(C|A\cap B)$.

where $P(A\cap B) = P(A)\cdot P(B|A)$.

$\Rightarrow P(A\cap B\cap C) = P(A)\cdot P(B|A)\cdot P(C|A\cap B)$.

Q: More than three?

A: Use the same method!

1.4 Independent Events

Motivation

For certain pairs of events, the occurrence of one of them does not change the probability of the occurrence of the other.

e.g. Flip a coin twice $\{HH, HT, TH, TT\}$.

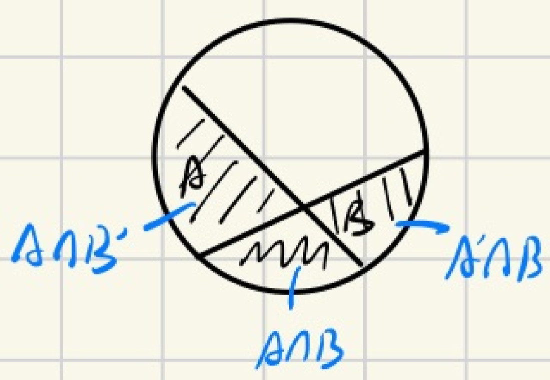
$A = \{$heads on the first flip$\} = \{HH, HT\}$.

$B = \{$tails on the second flip$\} = \{HT, TT\}$.

$P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$

We have $P(B|A) = P(B)$. $P(A|B) = P(A)$. which means they don't affect each other.

Definition



$P(A\cap B) = P(A)\cdot P(B)$.

when $P(A)\cdot P(B) > 0$ & $P(A|B) = P(A)$.

Event A and B are independent if:

$\triangle\ P(A\cap B) = P(A)\cdot P(B)$

Otherwise, events A and B are called dependent events.

when $P(A) \neq 0$, we have $P(B|A) = P(B)$.

when $P(B) \neq 0$, we have $P(A|B) = P(A)$

- **Properties**

Theorem: A and B are independent, if and only if any pair of the following are indep.

① A and B'
② A' and B
③ A' and B'

Three events

Def. Events A, B and C are mutually independent if

1. A, B, C are pairwise independent i.e.,

&

$$\begin{cases} P(A\cap B) = P(A)\cdot P(B) \\ P(A\cap C) = P(A)\cdot P(C) \\ P(B\cap C) = P(B)\cdot P(C). \end{cases}$$

2. $P(A\cap B\cap C) = P(A)\cdot P(B)\cdot P(C)$.

Warning: pairwise independent $\overset{\times}{\underset{\checkmark}{\rightleftarrows}}$ mutually independent.

More events

Mutual independence can be extended to four or more events:

△ Each pair, triple, quartet of events are independent and moreover:

$$P(A_1\cap A_2\cap\cdots\cap A_n) = P(A_1)\cdot P(A_2)\cdots P(A_n)$$

That is, any combination should be independent.

1.5 Bayes' Theorem

- **Definition**

Assume that

1. S is a sample space, and $B_1, B_2, \cdots, B_m$ are mutually exclusive and exhaustive w.r.t S.
2. The prior probabilities of B_i is positive.

Then we have

① For any event A,

△ $P(A) = \sum_{i=1}^{m} P(A\cap B_i) = \sum_{i=1}^{m} P(B_i)\cdot P(A|B_i)$ => Law of total probability.

② If $P(A)>0$, then $P(B_k|A) = \dfrac{P(B_k\cap A)}{P(A)}$,

△ $P(B_k|A) = \dfrac{P(B_k)|P(A|B_k)}{P(A) = \sum_{i=1}^{m} P(B_i)\cdot P(A|B_i)}$ => Bayes' Rule

$P(B_k)$ → prior probability
$P(B_k|A)$ → posterior probability
$P(A|B_k)$ → likelihood of B_k, A is called a data.

II. Discrete Distribution

2.1 Random Variable of the Discrete Type

· **Random Variable (RV)**

Def. Given a random experiment with sample space S, a function $X: S \to \bar{S} \subseteq R$ that assign one real number $X(s) = x$ to each $s \in S$ is called a Random Variable (RV).
$\bar{S}$ denot the range of X: $\bar{S} = \{x \mid X(s) = x, s \in S\}$.

Conventions:
Uppercase letters, e.g. X, Y, Z → RVs.
Lowercase letters, e.g. x, y, z → the numeric values that RV X, Y, Z can take, respectively.
For $X: S \to \bar{S}$, two probability involved:
$P_S(\cdot)$ is the probability function associated with S.
$P(\cdot)$ is the probability function associated with $\bar{S}$.
For any $x \in \bar{S}$,
$P(X = x) \triangleq P(\{X = x\}) = P_S(\{s \mid X(s) = x, s \in S\})$
$P(X \in A) \triangleq P(\{X \in A\}) = P_S(\{s \mid X(s) \in A, s \in S\})$.

e.g. Given a sample space $S = \{a, b, c, d, e, f\}$.
Define a RV: $X(a) = 1$, $X(b) = 2$, ... $X(f) = 6$.
$X: S = \{a, b, c, d, e, f\} \to \bar{S} = \{1, 2, 3, 4, 5, 6\}$.
Let $x = 1$ and $A = \{1, 2\}$.
$P(X = 1) \triangleq P(\{X = 1\}) = P_S(\{s \mid X(s) = 1, s \in S\}) = P_S(\{a\})$.
$P(X \in A) \triangleq P(\{X \in A\}) = P_S(\{s \mid X(s) \in A, s \in S\}) = P_S(\{a, b\})$.

· **Discrete Random Variable**

Def. Recall that $\bar{S}$ denote the range of X: $\bar{S} = \{x \mid X(s) = x, s \in S\}$.
A RV X is said to be discrete if its range $\bar{S}$ is finite or countably infinite.

e.g. $X: S = \{a, b, c\} \to \bar{S} = \{1, 2, 3\}$. RV X is discrete because $\bar{S}$ is finite.

· **Probability Mass Function (pmf)**

Def. Suppose that X is a RV with range $\bar{S}$. Then a function $f(x): \bar{S} \to (0, 1]$ is called pmf. if:
1. $f(x) > 0$, $x \in \bar{S}$.
2. $\sum_{x \in \bar{S}} f(x) = 1$.
3. $P(X \in A) = \sum_{x \in A} f(x)$. $A \subseteq \bar{S}$. Which defines the probability function for an event A.

We often extend the domain of $f(x)$ from $\bar{S}$ to R and let $f(x) = 0$. $x \notin \bar{S}$, $\bar{S}$ is called the support of $f(x)$.

A function $f(x): R \to [0, 1]$ is called pmf, if
1. $f(x) \geq 0$, $x \in R$
2. $\sum_{x \in \bar{S}} f(x) = 1$.
3. $P(X \in A) = \sum_{x \in A} f(x)$. $A \subseteq \bar{S}$.

e.g. $X: S = \{a, b, c\} \to \bar{S} = \{1, 2, 3\}$.
pmf $f(x) = \frac{1}{3}$, $x \in \bar{S}$, and $f(x) = 0$, $x \notin \bar{S}$.

· **Uniform Distribution**

Def. A RV X is said to have a uniform distribution if $f(x) = $ constant for $x \in \bar{S}$.

- Line Graph and Probability Histogram

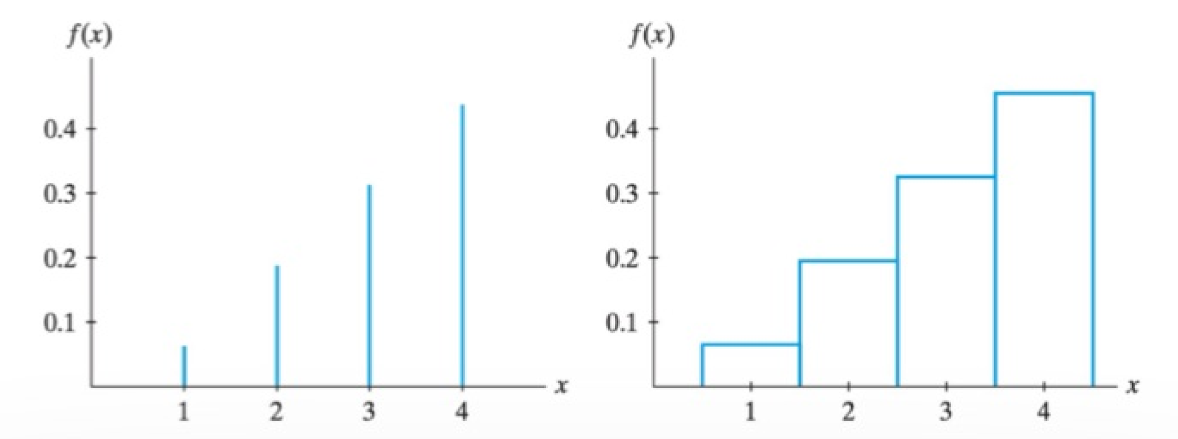


Figure 2.1-1 Line graph and probability histogram

- Cumulative Distribution Function (cdf)

Def. The function $F(x): \mathbb{R} \to [0,1]$:

$$F(x) = P(X \le x)$$

is called the cumulative distribution function (cdf).

1. $F(x)$ is noncreasing and moreover,
$$P(X \le x) = \sum_{x' \le x, x \in \bar{S}} f(x').$$

2. Relation between the probability function and the cdf
$$P(a < X \le b) = F(b) - F(a).$$

e.g. $X: S = \{a, b, c\} \to \bar{S} = \{1, 2, 3\}$.

cdf $F(x) = P(X \le x) = \sum_{x \le x, x \in \bar{S}} f(x')$

$$= \begin{cases} 0, & x < 1 \\ \frac{k}{3}, & k \le x < k+1,\ k = 1, 2 \\ 1, & x \ge 3. \end{cases}$$

2.2 Mathematical Expectation

- Definition

Assume X is a discrete RV with range $\bar{S}$ and $f(x)$ is its pmf. If $\sum_{x \in \bar{S}} g(x) f(x)$ exists, then it's called the mathematical expectation of $g(X)$ and is denoted by

$$\triangle\ E[g(X)] = \sum_{x \in \bar{S}} g(x) f(x).$$

- Properties

Assume that X is a discrete RV with range $\bar{S}$ and $f(x)$ is its pmf. We have:

1. If c is a constant, $E[c] = c$.
2. If c is a constant and $g(x)$ is a function, $E[cg(x)] = cE[g(x)]$.
3. If c_1 and c_2 are constants, $g_1(X)$ and $g_2(X)$ are functions,
$$E[c_1 g_1(X) + c_2 g_2(X)] = c_1 E[g_1(X)] + c_2 E[g_2(X)].$$

Mathematical expectation is a linear operator.

2.3 Special Mathematical Expectations [special $g(x)$]

- Mean and Variance

Def. Mean of a RV [$g(x) = X$]:

$$\circ\ E[X] = \sum_{x \in \bar{S}} x f(x) \xlongequal{\bar{S} = \{x_1, \cdots, x_k\}} \sum_{i=1}^{k} x_i f(x_i).$$

Interpretation of $E[X]$: the average value of x.

Def. Variance of a RV [$g(X) = (X-E[X])^2$]:

$\triangleright$ $Var(X) = E[(X-E[X])^2] = \sum_{x\in\bar{S}} (x-E[X])^2 f(x) = E[X^2] - (E[X])^2$

Standard deviation of a RV: $\sigma_X = \sqrt{Var(X)}$

Properties: let c be a constant.

$Var(c) = 0$, $\quad Var(cX) = c^2 Var(X)$.

· The r^{th} Moment

Def. r^{th} moment of X [$g(X) = X^r$ with r a positive integer]:

If $E[X^r] = \sum_{x\in\bar{S}} x^r f(x)$ exists, then it's called the r^{th} moment.

e.g. $E[X]$ and $E[X^2]$ are the first and second moments, respectively.

In addition,

If $E[(X-b)^r] = \sum_{x\in\bar{S}} (x-b)^r f(x)$ exists, then it's called the r^{th} moment of X about b.

If $E[(X)_r] = E[X(X-1)\cdots(X-r+1)]$ exists, then it's called the r^{th} factorial moment.

· Moment Generating Function (mgf)

Def. Let X be a discrete RV with range space $\bar{S}$ and $f(x)$ be its pmf.

If there exists a $h>0$ such that:

$\triangleright$ $E[e^{tX}] = \sum_{x\in\bar{S}} e^{tx} f(x)$ exists, for $-h<t<h$.

then the function defined by $M(t) = E[e^{tX}]$ is called the moment generating function of X.

The mgf can be used to generate the moments of X.

Properties:

1. $M(0) = 1$
2. If 2 RVs have the same mgf, they have the same probability distribution, i.e., the same pmf.
3. Derivatives

$$M'(t) = \sum_{x\in\bar{S}} x e^{tx} f(x) \qquad\qquad M'(0) = E[X]$$
$$M''(t) = \sum_{x\in\bar{S}} x^2 e^{tx} f(x) \quad \xrightarrow{\text{set } t=0} \quad M''(0) = E[X^2]$$
$$\vdots \qquad\qquad\qquad \vdots$$
$$M^{(r)}(t) = \sum_{x\in\bar{S}} x^r e^{tx} f(x). \qquad\qquad M^{(r)}(0) = E[X^r]$$

Observation: the moments can be computed by differentiating $M(t)$ and evaluating the derivatives at $t=0$.

2.4 Binomial Distribution

· Bernouli Distribution

Bernouli Experiment:

The outcomes can be classified in one of two mutually exclusive and exhaustive ways.

Def. Let X be a RV associated with Bernouli experiment with the probability of success p.

RV: $X: S \to \bar{S}$, where $S = \{success, failure\}$.

Define $X(success) = 1$, $X(failure) = 0$, and thus $\bar{S} = \{0,1\}$.

pmf of X: $f(x): \bar{S} \to (0,1]$.

$f(x) = p^x(1-p)^{1-x}$, $x \in \bar{S}$.

Then we say X has a Bernouli distribution with probability of success p.

- Mathematical expectations

1. $E[X] = \sum_{x\in S} x f(x) = 0\cdot(1-p) + 1\cdot p = p$.

2. $Var[X] = E[X-E[X]^2] = \sum_{x\in S}(x-p)^2 f(x) = p^2(1-p) + (1-p)^2 p = (1-p)p$.

3. Mgf: $M(t) = E[e^{tx}] = [e^t\cdot p + (1-p)]^n$. $t\in(-\infty,\infty)$.

- Bernouli Trials

If a bernouli experiment is performed n times.
1. Independently, i.e., all trials are independent. $(P(A_1\cap A_2\cap\cdots\cap A_n) = \prod_{i=1}^{n} P(A_i))$.
2. The probability of success, say p, remains the same from trial to trial.

then these n repetitions of the Bernouli experiment is called n Bernouli trials.

- Binomial Distribution

Def. A RV X is said to have a binomial distribution, if the range $S = \{0,1,\cdots,n\}$, and the pmf $f(x) = \binom{n}{x}p^x(1-p)^{n-x}$, $x = 0,1,\cdots,n$.
Denoted by $X\sim b(n,p)$, where the constants n,p are parameters expansion.

P.S. It is called the binomial distribution because of its connection with binomial expansion $(a+b)^n = \sum_{x=0}^{n}\binom{n}{x}a^x b^{n-x}$ with $a=p$, $b=1-p$.

- mgf of Binomial Distribution

Let $X\sim b(n,p)$. Then by definition,
$M(t) = E[e^{tX}] = \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x(1-p)^{n-x} = [(1-p)+pe^t]^n$ $(t\in R)$

From the expansion of $(a+b)^n = \sum_{x=0}^{n}\binom{n}{x}a^x b^{n-x}$ with $a = pe^t$, $b = 1-p$.

Use:
$M'(t) = n[(1-p)+pe^t]^{n-1}pe^t \Rightarrow M'(0) = E[X] = np$.
$M''(t) = n(n-1)[(1-p)+pe^t]^{n-2}p^2e^{2t} + n[(1-p)+pe^t]^{n-1}pe^t$
$M''(0) = E[X^2] = n(n-1)p^2 + np$.
$Var[X] = E[X^2] - (E[X])^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p)$.
Btw, when $n=1$ in $b(n,p)$, the binomial distribution reduces to Bernouli distribution by $b(1,p)$.

- cdf of Binomial Distribution

$F(x) = P(X\le x) = \sum_{y\in\{X\le x\}} f(y) = \sum_{y=0}^{[x]}\binom{n}{y}p^y(1-p)^{n-y}$, where $x\in(-\infty,\infty)$ and $[x]$ is the largest integer $\le x$.

2.5 Negative Binomial Distribution

- Motivative

We are interested in the number of Bernouli trials until exactly r success occur, where r is a fixed number.

pmf $f(x) = P(\{\text{at the } x \text{ trial, the } r^{th} \text{ success is observed}\})$
$= P(\{r-1 \text{ success in the first } x-1 \text{ trials}\}) \cap P(\{\text{success at the } x^{th} \text{ trial}\})$.
$= P(A)\cap P(B) = P(A)P(B)$. (independent)

- Negative Binomial Distribution

Def. A RV is said to have a negative bionomial distribution with the probability of success p and the number of success r we are interested in, if the range $\bar{S}=\{r, r+1, \cdots\}$ and the pmf $f(x)$ is in the form of

$$\triangleright\ f(x)=\binom{x-1}{r-1}p^r(1-p)^{x-r}, \quad x=r, r+1, \cdots.$$

P.S. This distribution gets its name due to the negative binomial series:

$$(1-w)^{-r}=\sum_{x=r}^{\infty}\binom{x-1}{r-1}w^{x-r}.$$

- Mathematical Expectations

1. Mean: $E[X]=\frac{r}{p}$.

2. Variance: $Var[X]=E[X^2]-(E[X])^2=\frac{r(1-p)}{p^2}$

3. mgf: $M(t)=E[e^{tX}]=\frac{(pe^t)^r}{[1-(1-p)e^t]^r}$, for $(1-p)e^t<1$.

- Geometric Distribution

Def. A RV is said to have a geometric distribution with the probability of success p, if the range $\bar{S}=\{1,2,\cdots\}$ and the pmf $f(x)$ is in the form of

$$f(x)=p(1-p)^{x-1}, \quad x=1,2,\cdots$$

For a positive integer k,

$$P(X>k)=\sum_{x=k+1}^{\infty}p(1-p)^{x-1}=\frac{(1-p)^kp}{1-(1-p)}=(1-p)^k.$$

$$P(X\le k)=\sum_{x=1}^{k}p(1-p)^{x-1}=1-P(X>k)=1-(1-p)^k.$$

2.6 Poisson Distribution

- Approximate Poisson Process (APP)

Def. Let the number of occurrences of some event in a given continuous interval be counted. Then we have an APP with parameter $\lambda>0$ if

① The number of occurrence is non-overlapping subintervals are independent.

② The probability of exactly one occurrence in a sufficiently short interval of length h is approximately λh.

③ The probability of two or more occurrence in a sufficiently short interval is essentially 0.

- Poisson Distribution

Def. A RV X is said to have a Poisson distribution with the parameter λ, if the range $\bar{S}=\{0,1,\cdots\}$ and the pmf $f(x)$ is in the form of

$$\triangleright\ f(x)=\frac{\lambda^x e^{-\lambda}}{x!}, \quad x=0,1,\cdots.$$

We can simply denoted it by $X\sim \text{Poisson}(\lambda)$

$$P(X=x)=\lim_{n\to\infty}\frac{n!}{(n-x)!\,n^x}\cdot\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}=\frac{\lambda^x e^{-\lambda}}{x!}$$

- Mathematical Expectations

1. $E[X] = \lambda$

2. $Var[X] = E[X^2] - (E[X])^2 = \lambda$

3. mgf $M[t] = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{\lambda(e^t - 1)}$

△ λ is the mean and variance of $X \sim Poisson(\lambda)$: the average number of occurrences in the unit interval.

P.S. Sometimes λ will become θ or else, but it's the same.

Ⅲ. Continuous Distribution

3.1 Random Variable of Continuous Type

· Continuous RV

Def. A RV X with $\bar{S}$ that is an interval or unions of intervals is said to be continuous RV, if there exists a function $f(x): \bar{S} \to (0, \infty)$ such that

1. $f(x) > 0$, $x \in \bar{S}$.
2. $\int_{\bar{S}} f(x)\,dx = 1$.
3. If $[a,b] \subseteq \bar{S}$, $P(a \le x \le b) \triangleq \int_a^b f(x)\,dx$

P.S. f is the so called probability density function (pdf).

Remark: We often extend the domain of $f(x)$ from $\bar{S}$ to R and let $f(x) = 0$, $x \notin \bar{S}$.
In this case, $f(x): R \to [0, \infty)$ and $\bar{S}$ is called the support of X. Then we have:

1. $f(x) \ge 0$, $x \in R$
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$.
3. $P(a \le x \le b) = \int_a^b f(x)\,dx$.

· Cumulative Distribution Function (cdf)

Def. Cumulative distribution function (cdf) $F(x): R \to [0,1]$.

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt.$$

1. $F(x)$ is nondecreasing.
2. $P(a \le X \le b) = F(b) - F(a)$.
3. pdf & cdf: $f(x) = F'(x)$. for those x at which $F(x)$ is differentiable.

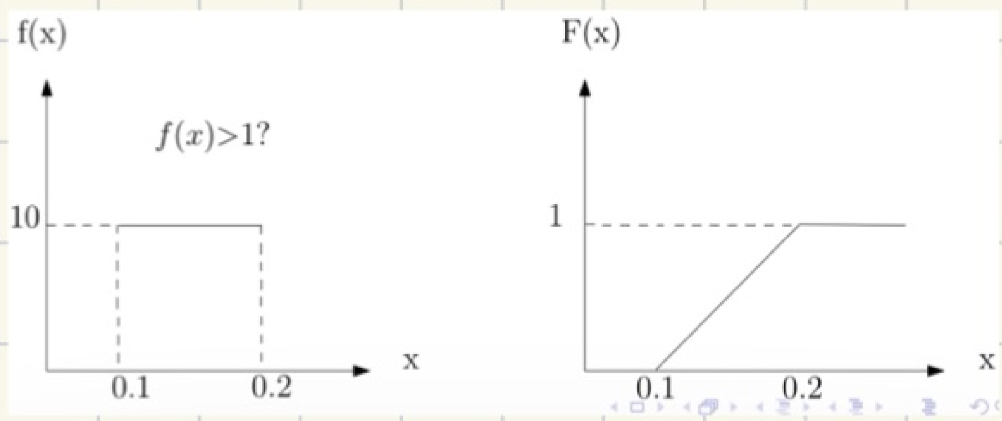
· Uniform Distribution

Def. Let the RV X denote the outcome when a point is selected randomly from $[a,b]$ with $-\infty < a < b < \infty$

pdf $f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b. \\ 0, & \text{otherwise} \end{cases}$

cdf $F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b. \end{cases}$

e.g.



For any $x \in [a,b]$, $P(X \le x) = P(a \le X \le b) = \frac{x-a}{b-a}$. implies the probability of selecting a point from the interval $[a,x]$ is proportional to the length of $[a,x]$. Such distribution is called uniform distribution and denoted by $X \sim U(a,b)$.

$$\begin{cases} E(X) = \frac{a+b}{2} \\ Var(X) = \frac{(b-a)^2}{12} \\ M(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)}, & t \ne 0 \\ 1, & t = 0 \end{cases} \end{cases}$$

· Mathematical Expectation

Def. Let X be a continuous RV with pdf $f(x): \bar{S} \to (0, \infty)$. If $\int_{\bar{S}} g(x) f(x)\,dx$ exists, it is called the mathematical expectation for $g(x)$ and denoted by

$$E[g(x)] = \int_{\bar{S}} g(x) f(x)\,dx.$$

If the range of X is extended from $\bar{S}$ to R with $f(x) = 0$ for $x \notin \bar{S}$, then

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

· Special Mathematical Expectations

1. Mean : $E[X] = \int_S x f(x) dx$

2. Variance: $Var(X) = E[(X-E[X])^2] = \int_S (x-E[x])^2 f(x) dx$

3. Moments: $E[X^r] = \int_S x^r f(x) dx$

4. mgf : $M(t) = E[e^{tX}] = \int_S e^{tx} f(x) dx$ $\quad -h < t < h$ for some $h > 0$.

Similar to discrete distribution!

· $(100p)$th percentile

Def. Given $p \in (0,1)$, π_p is a number such that the area under $f(x)$ to the left of π_p is p. That is
$$p = \int_{-\infty}^{\pi_p} f(x) dx = F(\pi_p)$$

The 50th percentile is called the median. The 25th and 75th percentiles are called the first and 75th percentiles are called the first and third quantiles, respectively.

3.2 Exponential, Gamma and Chi-square Distribution

· Exponential Distribution

Def. A RV X has an exponential distribution if its pdf is
$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad x \geq 0, \ \theta > 0 \qquad \text{Notation}: X \sim Exp(\lambda).$$

Accordingly, the waiting time until the first occurrence for an approximate Poisson process (APP) has an exponential distribution with $\theta = \frac{1}{\lambda}$ (λ: the average number of occurrence per unit time).

· Mathematical Expectations

1. mgf : $M(t) = E[e^{tX}] = \int_0^\infty e^{tx} \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = \frac{1}{1-t\theta}$, $t < \frac{1}{\theta}$.
$M'(t) = \frac{\theta}{(1-\theta t)^2}$, $M''(t) = \frac{2\theta^2}{(1-\theta t)^3}$

2. Mean: $E[X] = M'(0) = \theta$.

3. Variance : $Var[X] = E[X^2] - E[X]^2 = M''(0) - (M'(0))^2 = 2\theta^2 - \theta^2 = \theta^2$

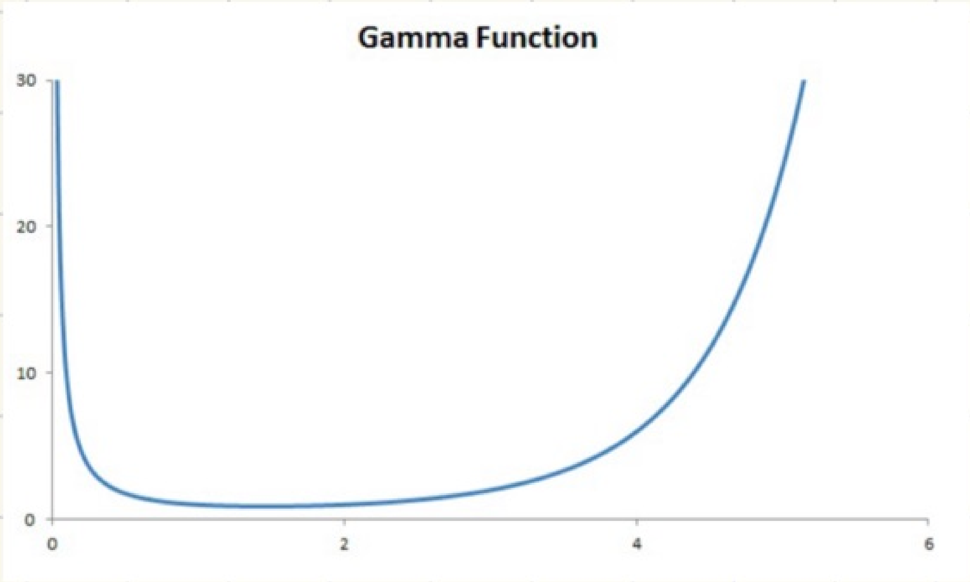
· Poisson Distribution

Def. Let X describe the number of occurrences of some events in a unit interval with
$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, \cdots, \qquad E[X] = Var[X] = \lambda.$$

For an interval with length T, which should be treated as a new "unit interval", the number of occurrences Y has $E[Y] = \lambda T$ and thus its pmf is
$$f(y) = \frac{(\lambda T)^y e^{-\lambda T}}{y!}, \quad y = 0, 1, \cdots$$

Then for $\alpha = 1, 2, \cdots$,
$$P(Y < \alpha) = \sum_{k=0}^{\alpha-1} \frac{(\lambda T)^k e^{-\lambda T}}{k!} = P(\text{the number of occurrence smaller than } \alpha \text{ in the interval with length } T)$$

- Gamma Function



Def. $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$, $t > 0$

$$\Gamma(t) = \left[-y^{t-1} e^{-y}\right]_0^\infty + \int_0^\infty (t-1) y^{t-2} e^{-y} dy = (t-1)\cdot\Gamma(t-1).$$

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) = \cdots = (n-1)!\cdot\Gamma(1) \quad (\text{where } \Gamma(1) = 1).$$

- Gamma Distribution

Def. A RV X has a Gamma distribution if its pdf is

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}}, \quad x \geq 0,\ \alpha > 0,\ \theta > 0.$$

Where θ and α are the two parameters.

mgf: $M(t) = \frac{1}{(1-\theta t)^\alpha}$, $t < \frac{1}{\theta}$.

Mean: $E[X] = \alpha\theta$. Variance: $Var[X] = \alpha\theta^2$.

Special Case: When $\alpha = 1$, Gamma distribution reduces to exponential distribution.

- Chi-square Distribution

Def. Let X have a Gamma distribution with $\theta = 2$, $\alpha = \frac{r}{2}$, r is an integer. The pdf of X is

$$f(x) = \frac{1}{\Gamma(\frac{r}{2}) 2^{\frac{r}{2}}} \cdot x^{\frac{r}{2}-1} e^{-\frac{x}{2}}, \quad x > 0.$$

Then X has a chi-square distribution with degrees of freedom r, and denoted by $X \sim \chi^2(r)$

$E[X] = \alpha\theta = \frac{r}{2}\cdot 2 = r$.

$Var[X] = \alpha\theta^2 = \frac{r}{2}\cdot 2^2 = 2r$.

mgf: $M(t) = (1-2t)^{-\frac{r}{2}}$, $t < \frac{1}{2}$.

3.3 Normal Distribution

- Normal Distribution

Def. A continuous RV X is said to be normal or Gaussian if it has a pdf of the form

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\cdot\frac{(x-\mu)^2}{\sigma^2}\right). \quad -\infty < x < \infty.$$

where μ and σ^2 are two parameters characterizing the normal distribution. Briefly, $X \sim N(\mu, \sigma^2)$.

- Mathematical Expectations

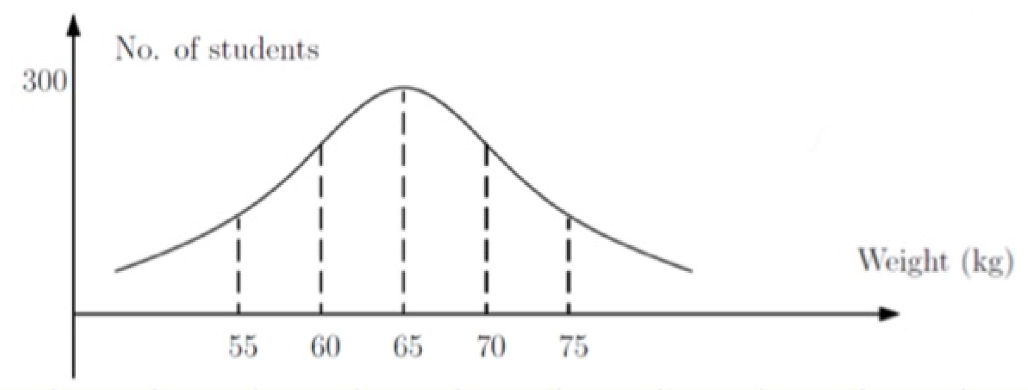
mgf: $M(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$

Mean: $E[X] = \mu$

Variance: $Var[X] = \sigma^2$

When observed over a large population, many things of interests have a "bell-shaped" relative frequency distribution.

- Weight of male students in CUHKsz
- Height
- TOFEL,IELTS test score



· **The Upper 100α Percent Point**

Def. The number z_α, such that $P(Z \geq z_\alpha) = \alpha$.

Note: $P(Z < z_\alpha) = 1 - P(Z \geq z_\alpha) = 1-\alpha$. So z_α is the $100(1-\alpha)^{th}$ percentile.

· **Theorems of Normal Distribution**

Theorem: If Y is $N(\mu, \sigma^2)$, then $X = \frac{Y-\mu}{\sigma}$ is $N(0,1)$. $N(0,1)$ can be calculated by given table.

Theorem: If X is $N(\mu, \sigma^2)$ with $\sigma^2 > 0$, then

$$\frac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1). \quad (\text{Chi-square distribution}).$$

Appendix:

RV X is a function $X: S \to \bar{S} \in R$.

Discrete RV

pmf $f(x): R \to [0,1]$.

1. $f(x) \geq 0$
2. $\sum_{x \in \bar{S}} f(x) = 1$
3. $P(X \in A) = \sum_{x \in A} f(x)$.

Continuous RV:

pdf $f(x): R \to [0, \infty)$

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(X \in A) = \int_A f(x)dx$

with $f(x) = 0$ for $x \notin \bar{S}$. $\bar{S}$ is called the support set of $f(x)$.

	pmf/pdf	mgf	mean	variance	
Binomial ($n=1$ Bernoulli)	$\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0,1,2,\dots,n$.	$[(1-p)+pe^t]^n$, $-\infty < t < \infty$	np	$np(1-p)$.	The total number of success in n Bernoulli trials (no order)
Negative Binomial	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$, $x = r, r+1, r+2, \dots$	$\frac{(pe^t)^r}{[1-(1-p)e^t]^r}$, $(1-p)e^t < 1$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	For a given natural number r, the number of Bernoulli trials on which the r^{th} success is observed.
Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}$, $x = 0,1,2\dots$	$\exp(\lambda(e^t - 1))$	λ	λ	The number of occurrences in a particular event that can be described as an APP.
Gamma ($\alpha=1$, Exponential) ($\theta=2, \alpha=\frac{r}{2}$. χ^2)	$\frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}}$, $0 < x < \infty$	$\frac{1}{(1-\theta t)^\alpha}$, $t < \frac{1}{\theta}$	$\alpha\theta$	$\alpha\theta^2$	The waiting time until the α^{th} occurrences of a particular event for an APP.
Normal ($\mu=0$, $\sigma^2=1$, standard)	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)$, $-\infty < x < \infty$	$\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$, $-\infty < t < \infty$	μ	σ^2	When a large number of outcomes are observed

IV. Bivariate Distribution

4.1 Bivariate Distribution of Discrete Type

· Bivariate RV

Def. Let (X, Y) be a pair of RVs with their range denoted by $\bar{S} \subseteq \mathbb{R}^2$. Then (X, Y) or X and Y is said to be a bivariate RV. If $\bar{S}$ is finite or countably infinite, then (X, Y) or X and Y is said to be a discrete bivariate RV.

Moreover, let $\bar{S}_X \subseteq \mathbb{R}$ and $\bar{S}_Y \subseteq \mathbb{R}$ denote the range of X and Y, respectively.

$\bar{S} = \{\text{all possible values of } (X, Y)\}$.

$\bar{S}_X = \{\text{all possible values of } X\} = \{x \mid (x, y) \in \bar{S}\}$

$\bar{S}_Y = \{\text{all possible values of } Y\} = \{y \mid (x, y) \in \bar{S}\}$

Then, it holds that

$\bar{S} \subseteq \bar{S}_X \times \bar{S}_Y = \{(x, y), x \in \bar{S}_X, y \in \bar{S}_Y\}$.

· Joint pmf

Def. The function $f(x, y): \bar{S} \to [0, 1]$ is called the joint probability mass function (joint pmf) of X and Y or (X, Y), if

1. $f(x, y) > 0$ for $(x, y) \in \bar{S}$
2. $\sum_{(x,y) \in \bar{S}} f(x, y) = 1$
3. For $A \subseteq \bar{S}$, $P[(x, y) \in A] \triangleq P(\{x, y\} \in A) = \sum_{(x,y) \in A} f(x, y)$

which defines the probability function for a set A. In particular, taking $A = \{(x, y)\}$ yields the probability of $X = x$ and $Y = y$, i.e., $P(X = x, Y = y) = f(x, y)$.

Remark:

For $A \subseteq \bar{S}$, $P[(X, Y) \in A] \triangleq P(\{X, Y\} \in A) = \sum_{(x,y) \in A} f(x, y)$

Let $A_X = \{x \mid (x, y) \in A\}$, $A_Y(x) = \{y \mid (x, y) \in A\}$, for $x \in A_X$

Then $P((X, Y) \in A) = \sum_{x \in A_X} \sum_{y \in A_Y(x)} f(x, y)$

Let $A_Y = \{y \mid (x, y) \in A\}$, $A_X(y) = \{x \mid (x, y) \in A\}$, for $y \in A_Y$

Then $P((X, Y) \in A) = \sum_{y \in A_Y} \sum_{x \in A_X(y)} f(x, y)$

· Marginal pmf

Def. Let (X, Y) or X and Y be a bivariate RV and have the joint pmf $f(x, y): \bar{S} \to [0, 1]$. Sometimes, we are interested in the pmf of X or Y alone, which is called the marginal pmf of X and Y.

For $x \in \bar{S}_X$, $f(x) = P_X(X = x) \triangleq P(X = x, Y \in \bar{S}_Y(x)) = \sum_{y \in \bar{S}_Y(x)} f(x, y)$

where $\bar{S}_Y(x) = \{y \mid (x, y) \in \bar{S}\}$ for the given $x \in \bar{S}_X$.

Y is similar.

· Trinomial Distribution

For the trinomial experiment, we let
X be number of "perfect".
Y be number of "seconds".
$n-X-Y$ be number of "defectives".
Then we have
$\bar{S} = \{(x,y) \mid x+y \le n,\ x,y = 0,1,\dots,n\}$.
∘ Joint pmf $f(x,y) = P(X=x, Y=y)$
$$= \frac{n!}{x!\,y!\,(n-x-y)!} P_X^x P_Y^y (1-P_X-P_Y)^{n-x-y},\ (x,y) \in \bar{S}$$
▹ Margin pmf $f_X(x) = \sum_{Y \in \bar{S}_Y(x)} f(x,y) = \sum_{y=0}^{n-x} \binom{n}{x}\binom{n-x}{y} P_X^x P_Y^y (1-P_X-P_Y)^{n-x-y}$
$$= \binom{n}{x} P_X^x (1-P_X)^{n-x}$$

P.S. The n^{th} power of a trinomial
$$(a+b+c)^n = \sum_{x=0}^{n} \binom{n}{x} a^x (b+c)^{n-x}$$
$$= \sum_{x=0}^{n} \binom{n}{x} a^x \sum_{y=0}^{n-x} \binom{n-x}{y} b^y c^{n-x-y}$$
$$= \sum_{x=0}^{n} \sum_{y=0}^{n-x} \frac{n!}{x!\,y!\,(n-x-y)!} a^x b^y c^{n-x-y}$$

We know $X \sim b(n, p_X)$, $Y \sim b(n, P_Y)$.

· Independent Random Variables

Def. The random variables X and Y are said to be independent if for every $x \in \bar{S}_X$ and $y \in \bar{S}_Y$
∘ $f(x,y) = f_X(x) f_Y(y)$.
or equivalently,
$P(X=x, Y=y) = P_X(X=x) P_Y(Y=y)$.
X and Y are said to be dependent if otherwise.
And when X and Y are independent,
$\bar{S} = \bar{S}_X \times \bar{S}_Y$, $\bar{S}$ is said to be rectangular, which is a neccessary condition.

⇒ For any $A \subset \bar{S}_X$ and $B \subset \bar{S}_Y$, the two events $X \in A$ and $Y \in B$ are independent.

· Mathematical Expectation

Def. Let X and Y be discrete RVs with their joint pmf $f(x,y): \bar{S} \to [0,1]$.
Consider a function $g(X,Y)$ of X and Y.
Then the expectation of $g(X,Y)$ is
$$E[g(X,Y)] = \sum_{(x,y) \in \bar{S}} g(x,y) \cdot f(x,y).$$

Two ways to calculate $E[X]$:

$$E[X] = \underbrace{\sum_{x \in \bar{S}_X} x f_X(x)}_{\text{Marginal pmf}}.$$

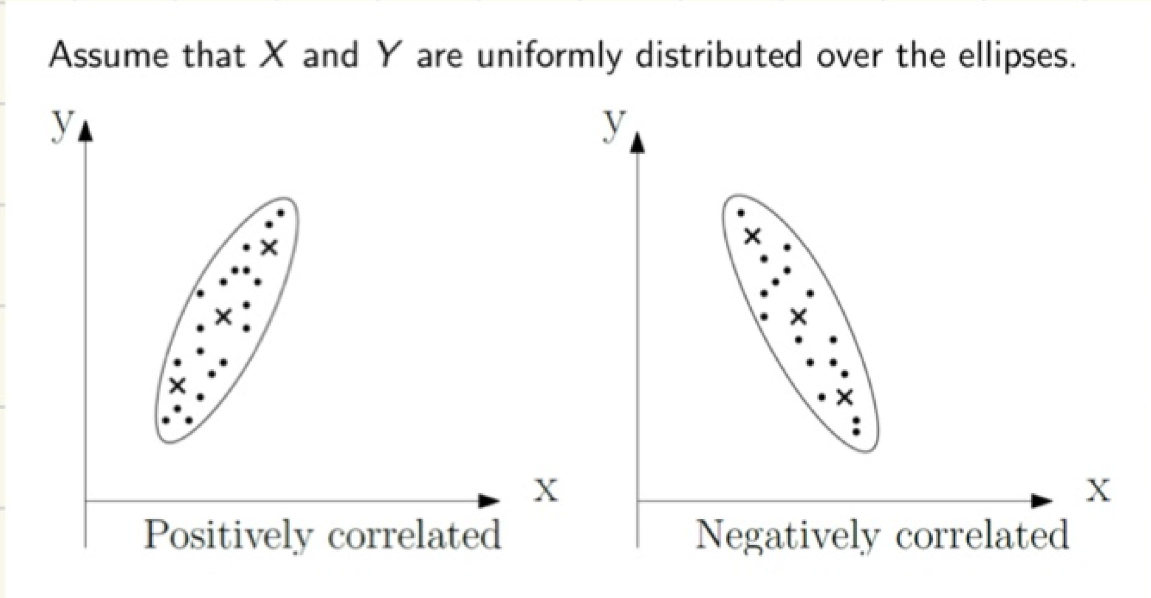
$$E[X] = \underbrace{\sum_{(x,y) \in \bar{S}} x f(x,y)}_{\text{Joint pmf}} = \sum_{x \in \bar{S}_X} x \cdot \underbrace{\sum_{y \in \bar{S}_Y(x)} f(x,y)}_{= f_X(x)}.$$

Variance of X: $E[(X-E[X])^2]$.

4.2 The Correlation Coefficient

· Covariance of X and Y

Def. Let X and Y be RVs with joint pmf $f(x,y): \bar{S} \to [0,1]$.
Take $g(X,Y) = (X-E(X))(Y-E(Y))$.
△ $Cov(X,Y) \triangleq E[(X-E(X))(Y-E(Y))] = E(XY) - E(X)E(Y)$.
$= \sum_{(x,y)\in\bar{S}} (x-E(X))(y-E(Y)) f(x,y)$.



△ 1. When $Cov(X,Y) = 0$, X and Y are uncorrelated
2. When $Cov(X,Y) > 0$, X and Y are positively correlated.
3. When $Cov(X,Y) < 0$, X and Y are negatively correlated.

Or using Indicate Variables to calculate $Cov(X,Y)$.

· Independence and Uncorrelation

Independence $\Rightarrow$ Uncorrelation

If X and Y are independent, we have
$f(x,y) = f_X(x) \cdot f_Y(y) \Rightarrow \bar{S} = \bar{S}_X \times \bar{S}_Y$.
$E(XY) = \sum_{(x,y)\in\bar{S}} xy f(x,y) = \sum_{x\in\bar{S}_X}\sum_{y\in\bar{S}_Y} xy f_X(x) f_Y(y) = E(X)E(Y)$.

Therefore
$Cov(X,Y) = E(XY) - E(X)E(Y) = 0$.

Uncorrelation $\not\Rightarrow$ Independence
$Cov(X,Y) = 0 \not\Rightarrow f(x,y) = f_X(x) f_Y(y)$.

· Correlation Coefficient

Def. The correlation coefficient of X and Y that have nonzero variance is defined as

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}}$$

Interpretation: $\rho > 0$ (or $\rho < 0$) indicate the values of $X - E[X]$ and $Y - E[Y]$ "tend" to have the same (or negative, respectively) sign.

· Properties of the Correlation Coefficient

1. It is a normalized version of $Cov(X,Y)$ and in fact $-1 \le \rho(X,Y) \le 1$.

2. $\rho = 1$ (resp. $\rho = -1$) if and only if there exists a positive (resp. negative) constant c s.t.
$Y - E(Y) = c(X - E(X))$.
and the size of $|\rho|$ provides a normalized measure of the extent to which this is true.

4.3 Conditional Distribution

· Conditional Distribution

Def. Conditional pmf of X given $Y=y$ is defined by

$\triangle$ $g(x|y) = \dfrac{f(x,y)}{f_Y(y)}$, $x \in \bar{S}_X(y)$. $\quad (f_Y(y) > 0)$.

Similarly, the conditional pmf of Y given that $X=x$ is defined by

$\triangle$ $h(y|x) = \dfrac{f(x,y)}{f_X(x)}$, $y \in \bar{S}_Y(x)$. $\quad (f_X(x) > 0)$.

And we have

1. $h(y|x) > 0$
2. $\displaystyle\sum_{y \in \bar{S}_Y(x)} h(y|x) = \sum_{y \in \bar{S}_Y(x)} \frac{f(x,y)}{f_X(x)} = \frac{\sum_{y \in \bar{S}_Y(x)} f(x,y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1.$
3. for $A \subseteq \bar{S}_Y(x)$,

$$P(Y \in A | X=x) = \frac{P(X=x, Y \in A)}{P(X=x)} = \sum_{y \in A} h(y|x).$$

Therefore, $h(y|x)$ determines the distribution of probability of events of Y given $X=x$.
P.S. $g(x|y)$ is similar!

If X and Y are independent, then $f(x,y) = f_X(x) f_Y(y)$ and thus
$g(x|y) = f_X(x)$, and $h(y|x) = f_Y(y)$.

· Conditional Mathematical Expectation

Def. Let $g(Y)$ be a function of Y.
Then the conditional expectation of $g(Y)$ given $X=x$:

$\triangle$ $E(g(Y)|X=x) = \displaystyle\sum_{y \in \bar{S}_Y(x)} g(y) h(y|x)$.

When $g(Y) = y$.

$E(Y|X=x) = \displaystyle\sum_{y \in \bar{S}_Y(x)} y\, h(y|x)$. conditional mean

When $g(Y) = [Y - E(Y|X=x)]^2$,

$Var(Y|X=x) \triangleq E([Y - E(Y|X=x)]^2 | X=x)$.
$= E(Y^2|X=x) - [E(Y|X=x)]^2$ conditional variance

4.4 Bivariate Distribution of Continuous Type

· Bivariate Continuous RV

Def. Let X and Y be two continuous random variables and (X,Y) be a pair of RVs with their range denoted by $\bar{S} \subseteq R^2$. Then (X,Y) or X and Y is said to be a bivariate continuous RV. Moreover, let $\bar{S}_X \subseteq R$ and $\bar{S}_Y \subseteq R$ denote the range of X and Y, respectively.

$\bar{S} = \{$all possible values of $(X,Y)\}$.
$\bar{S}_X = \{$all possible values of $X\} = \{x | (x,y) \in \bar{S}\}$.
$\bar{S}_Y = \{$all possible values of $Y\} = \{y | (x,y) \in \bar{S}\}$.

And we have $\bar{S} \subseteq \bar{S}_X \times \bar{S}_Y = \{(x,y) | x \in \bar{S}_X, y \in \bar{S}_Y\}$.

Joint pdf

Def. The joint pdf of two continuous RVs X and Y is a function $f(x,y): \bar{S} \to (0,\infty)$ with the following properties.

1. $f(x,y) > 0$, $(x,y) \in \bar{S}$
2. $\iint_{\bar{S}} f(x,y)\,dx\,dy = 1$
3. $P((X,Y)\in A) \triangleq P(\{X,Y\}\in A) = \iint_A f(x,y)\,dx\,dy$, $A \subseteq \bar{S}$

Marginal pdf

Def. The marginal pdf of X, $f_X(x): \bar{S}_X \to (0,\infty)$

$$f_X(x) = \int_{\bar{S}_Y(x)} f(x,y)\,dy$$

where $\bar{S}_Y(x) = \{y \mid (x,y)\in \bar{S}\}$ for $x \in \bar{S}_X$.

The marginal pdf of Y is similar.

Mathematical Expectation

Def. Let $g(X,Y)$ be a function of X and Y, whose joint pdf $f(x,y): \bar{S} \to (0,\infty)$. Then

$$E[g(X,Y)] = \iint_{\bar{S}} g(x,y)\,f(x,y)\,dx\,dy$$

Mean of X: $g(X,Y) = X$

$$E[X] = \iint_{\bar{S}} x\,f(x,y)\,dx\,dy = \int_{\bar{S}_X} x \int_{\bar{S}_Y(x)} f(x,y)\,dy\,dx = \int_{\bar{S}_X} x f(x)\,dx$$

Variance of X: $g(X,Y) = g(X-E[X])^2$

$$Var[X] = \iint_{\bar{S}} (x-E[X])^2 f(x,y)\,dx\,dy = \int_{\bar{S}_X} (x-E[X])^2 \int_{\bar{S}_Y(x)} f(x,y)\,dy\,dx = \int_{\bar{S}_X} (x-E[X])^2 f_X(x)\,dx.$$

Independent Continuous RVs

Def. Two continuous RVs X and Y are independent if

$$f(x,y) = f_X(x)\cdot f_Y(y), \quad x\in\bar{S}_X, \quad y \in \bar{S}_Y.$$

Otherwise, they are dependent.

And $\bar{S} = \bar{S}_X \times \bar{S}_Y$ is a necessary condition for independent of X and Y.

Covariance and Correlation Coefficient

Def. $Cov(X,Y) = E[(X-E[X])(Y-E[Y])] = E(XY) - E(X)E(Y)$

$$E(X,Y) = \iint_{\bar{S}} xy\,f(x,y)\,dx\,dy$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\cdot\sqrt{Var(Y)}}, \quad Var(X) > 0,\ Var(Y) > 0.$$

Conditional pdf

Def. Let X and Y have a joint pdf $f(x,y): \bar{S} \to (0,\infty)$ and marginal pdf $f_X(x): \bar{S}_X \to (0,\infty)$ and $f_Y(y): \bar{S}_Y \to (0,\infty)$.

The conditional pdf of Y, given that $X=x$ are

$$h(y|x) = \frac{f(x,y)}{f_X(x)} \quad \text{for } f_X(x) > 0,\ y \in \bar{S}_Y(x).$$

For $A \subseteq \bar{S}_Y(x)$, $P(Y \in A | X=x) = \int_A h(y|x)\,dy$.

The conditional pdf of X is similar.

Conditional Mathematical Expectation

Def. The conditional mathematical expectation of a function of Y, $g(Y)$, given that $X=x$ is

$$E(g(Y)|X=x) = \int_{\bar{S}_Y(x)} g(y)\,h(y|x)\,dy.$$

Mean: $E(Y|X=x) = \int_{\bar{S}_Y(x)} y\,h(y|x)\,dy$

Variance: $Var(Y|X=x) = E\left([Y - E(Y|X=x)]^2 \,\middle|\, X=x\right) = E(Y^2|X=x) - [E(Y|X=x)]^2$.

4.5 Bivariate Normal Distribution

pdf of Bivariate Normal Distribution

Def. Let X and Y be 2 continuous RVs and have the joint pdf

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\tfrac{1}{2}q(x,y)\right), \quad x, y \in \mathbb{R}.$$

where $q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right] \geq 0$. $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X, \sigma_Y > 0$, $|\rho| < 1$.

Then X and Y are said to be bivariate normally distributed.

Key components: Scaled exponential function with a quadratic and negative function as its exponent.

Properties

1. Marginal pdf of X and Y are normal with
   $X \sim N(\mu_X, \sigma_X^2)$ $\quad Y \sim N(\mu_Y, \sigma_Y^2)$.

2. Conditional pdf of X given that $Y=y$ is normal with mean
   $E(X|y) = \mu_X + \frac{\sigma_X}{\sigma_Y}\rho(y-\mu_Y)$
   and variance
   $Var(X|y) = (1-\rho^2)\sigma_X^2$
   i.e. $X|Y=y \sim N\left(\mu_X + \frac{\sigma_X}{\sigma_Y}\rho(Y-\mu_Y),\ (1-\rho^2)\sigma_X^2\right)$.
   $Y|X=x \sim N\left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\rho(X-\mu_X),\ (1-\rho^2)\sigma_Y^2\right)$.

3. Independence $\Leftrightarrow$ Uncorrelation

$f(x,y) = f_X(x) f_Y(y) \Leftrightarrow \rho = 0$. $(Cov(X,Y) = \rho\sigma_X\sigma_Y = 0)$.

Calculation of Covariance through Conditional pdf and Marginal pdf

Recall that

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

First, if the marginal pdf $f_X(x)$ and $f_Y(y)$ are available, then $E(X)$ and $E(Y)$ can be computed.

Then, we only need to consider how to calculate $E(XY)$ through the conditional pdf and the marginal pdf.

$$E(XY) = \int_{\bar{S}} xyf(x,y)dxdy = \int_{\bar{S}_X} x \int_{\bar{S}_Y(x)} yf(x,y)dydx$$
$$= \underbrace{\int_{\bar{S}_X} x \underbrace{\int_{\bar{S}_Y(x)} yh(y|x)dy}_{\text{expectation of function of } Y|X=x} f_X(x)dx}_{\text{expectation of function of } X}$$

Calculation of Covariance through Conditional pmf and Marginal pmf

$$E(XY) = \int_{\bar{S}} xyf(x,y)dxdy = \int_{\bar{S}_X} x \int_{\bar{S}_Y(x)} yf(x,y)dydx$$
$$= \underbrace{\int_{\bar{S}_X} x \underbrace{\int_{\bar{S}_Y(x)} yh(y|x)dy}_{\text{expectation of function of } Y|X=x} f_X(x)dx}_{\text{expectation of function of } X}$$
$$= \sum_{\bar{S}_X} xE(Y|X=x)f_X(x)dx$$

Clearly, if the conditional pdf of Y given that $X=x$, i.e., $h(y|x)$, then the conditional mean $E(Y|X=x)$ can be computed.

Then with $E(Y|X=x)$, one can compute the mathematical expectation of $E(XE(Y|X=x))$ through the marginal pdf $f_X(x)$.

V. Distribution of Functions of Random Variables

5.1 Function of One Random Variable

· Function of One Random Variable

Def. Let X be a RV of either discrete or continuous type with its pmf or pdf denoted by $f(x)$. Consider a function of X, say $Y=u(X)$. Then Y is also a RV and has its pmf or pdf.

1. For discrete RV, when $Y=u(X)$ be a one-to-one mapping with inverse $X=v(Y)$.
   Then the pmf of Y is
   △ $g(y)=f[v(y)]$ for $y \in \bar{S}_Y$.

2. For continuous RV, when $Y=u(X)$ is continuous, strictly decreasing or increasing and has inverse function $X=v(Y)$, whose derivative $\frac{dv(Y)}{dy}$ exists. Then the pdf of Y is
   △ $g(y)=f[v(y)]\left|\frac{dv(y)}{dy}\right|$.

· Random Number Generator

Theorem:
Let $Y \sim U(0,1)$ and $F(x)$ have the properties of a cdf of a continuous RV with $F(a)=0$, $F(b)=1$. Moreover, $F(x)$ is strictly increasing such that $F(x): (a,b) \to [0,1]$, where a could be $-\infty$, b could be ∞. Then $X=F^{-1}(Y)$ is continuous RV with cdf $F(x)$.
1. Generate a random number y from $U(0,1)$.
2. Take $x=F^{-1}(y)$ using where inverse function is easy to find

Then x is a random number generated from the distribution or RV with cdf $F(x)$.

Theorem:
Suppose that X is a continuous RV with $\bar{S}_X=(a,b)$, and moreover, its cdf $F(x)$ is strictly increasing. Then the RV Y, defined by $Y=F(X)$, has a uniform distribution, that is, $Y \sim U(0,1)$.

· Not One-to-one Case

Let X be a RV of either discrete or continuous type with its pmf or pdf denoted by $f(x)$. Consider a function of X, say $Y=u(X)$. Then Y is also a RV and has its pmf or pdf.

5.3 Several Random Variables (Multivariate RVs)

· Joint pmf or pdf

Def. Same as bivariate RVs.

Discrete type: $X_1, X_2, \cdots, X_n$ are all discrete
Joint pmf $f(x_1,\cdots,x_n): \bar{S} \to [0,1]$.
1. $f(x_1,x_2,\cdots,x_n)>0$, $(x_1,\cdots,x_n)\in\bar{S}$
2. $\sum_{x_1,\cdots,x_n\in\bar{S}} f(x_1,\cdots,x_n)=1$
3. $P[(X_1,\cdots,X_n)\in A]=\sum_{(x_1,\cdots,x_n)\in A} f(x_1,\cdots,x_n)$

Continuous type: $X_1, X_2, \cdots, X_n$ are all continuous
Joint pdf $f(x_1,\cdots,x_n): \bar{S} \to (0,\infty)$
1. $f(x_1,\cdots,x_n)>0$, $(x_1,\cdots,x_n)\in\bar{S}$
2. $\int_{\bar{S}} f(x_1,\cdots,x_n)\,dx_1\cdots dx_n=1$.
3. $P[(X_1,\cdots,X_n)\in A]=\int_A f(x_1,\cdots,x_n)\,dx_1\cdots dx_n$

n Independent RVs

Def. The n RVs $X_1, \cdots, X_n$ are said to be (mutually) independent if

$$f(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

A necessary condition for the independence of the n RVs $X_1, \cdots, X_n$ is

$$\bar{S} = \bar{S}_{X_1} \times \cdots \times \bar{S}_{X_n}.$$

Remark: If $X_1, \cdots, X_n$ are independent, then any pair of them, any triple of them, $\cdots$, any $(n-1)$ of them are also independent.

Def. Independently and identically distributed (i.i.d.) RVs $X_1, X_2, \cdots, X_n$, are also called random sample of size n from a common distribution

Mathematical Expectation

Def. Let $X_1, \cdots, X_n$ be multivariate RVs and have the joint pmf or pdf given by $f(x_1, \cdots, x_n)$, $(x_1, \cdots, x_n) \in \bar{S}$.

For a function $u(X_1, X_2, \cdots, X_n)$, its mathematical expectation is

$$E[u(X_1, X_2, \cdots, X_n)] = \begin{cases} \sum\limits_{(x_1, \cdots, x_n) \in \bar{S}} u(x_1, \cdots, x_n) \cdot f(x_1, \cdots, x_n) & \text{discrete RVs} \\ \int_{\bar{S}} u(x_1, \cdots, x_n) \cdot f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n & \text{continuous RVs} \end{cases}$$

In the case where $X_1, \cdots, X_n$ are independent, $f(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$.

$$E[u(X_1, X_2, \cdots, X_n)] = \begin{cases} \sum\limits_{x_1 \in \bar{S}_{X_1}} \cdots \sum\limits_{x_n \in \bar{S}_{X_n}} u(x_1, \cdots, x_n) \cdot f_{X_1}(x_1) \cdots f_{X_n}(x_n) & \text{discrete} \\ \int_{\bar{S}_{X_1}} \cdots \int_{\bar{S}_{X_n}} u(x_1, \cdots, x_n) \cdot f_{X_1}(x_1) \cdots f_{X_n}(x_n)\, dx_1 \cdots dx_n & \text{continuous} \end{cases}$$

Theorem:

Assume that $X_1, X_2, \cdots, X_n$ are independent RVs and $Y = u_1(X_1)\, u_2(X_2) \cdots u_n(X_n)$

If $E[u_i(X_i)]$, $i = 1, \cdots, n$ exist, then

$$\begin{aligned} E[Y] &= E[u_1(X_1)\, u_2(X_2) \cdots u_n(X_n)] \\ &= E[u_1(X_1)] \cdot E[u_2(X_2)] \cdots E[u_n(X_n)]. \end{aligned}$$

This is an extension of $E(XY) = E(X)E(Y)$ when X and Y are independent.

Theorem:

Assume that $X_1, X_2, \cdots, X_n$ are independent RVs with respective mean $\mu_1, \mu_2, \cdots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$, respectively. Consider $Y = \sum_{i=1}^{n} a_i X_i$, where $a_1, a_2, \cdots, a_n$ are real constants. Then

$$E(Y) = \sum_{i=1}^{n} a_i \mu_i \quad \text{and} \quad Var(Y) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

Sample Mean

Def. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed with mean μ. The sample mean:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and a statistic and also an estimator of mean μ.

5.4 Moment Generating Function Technique

· Product Property of mgf

Theorem:

If $X_1, X_2, \cdots, X_n$ are independent RVs with respective mgfs $M_{X_i}(t)$, where $|t| < h_i$ for $h_i > 0$, $i = 1, \cdots, n$.
Then the mgf of $Y = \sum_{i=1}^{n} a_i X_i$ is: → linear combination

$$\triangle \; M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t)$$

where $|a_i t| < h_i$, $i = 1, \cdots, n$.

Corollary:

If $X_1, X_2, \cdots, X_n$ is a random sample of size n from a distribution with mgf $M(t)$ (all the same), where $|t| < h$, then

(a) The mgf of $Y = \sum_{i=1}^{n} X_i$ is

$$M_Y(t) = \prod_{i=1}^{n} M(t) = (M(t))^n, \quad |t| < h$$

(b) The mgf of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is

$$M_{\bar{X}}(t) = \prod_{i=1}^{n} M\left(\frac{1}{n}t\right) = \left[M\left(\frac{t}{n}\right)\right]^n, \quad \left|\frac{t}{n}\right| < h.$$

· Additive Property of Chi-square Distribution

Theorem:

Let $X_1, X_2, \cdots, X_n$ be independent chi-square RVs with $r_1, r_2, \cdots, r_n$ degrees of freedom, i.e. $X_i \sim \chi^2(r_i)$, $i = 1, \cdots, n$. Then:

$$Y = X_1 + X_2 + \cdots + X_n \text{ is } \chi^2(r_1 + r_2 + \cdots + r_n)$$

Corollary:

Let $Z_1, Z_2, \cdots, Z_n$ have standard normal distributions, $N(0,1)$. If these random variables are independent, then:

$$W = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi^2(n).$$

Corollary:

If $X_1, X_2, \cdots, X_n$ are independent and have normal distributions $N(\mu_i, \sigma_i^2)$, $i = 1, \cdots, n$, respectively.

$$W = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2(n).$$