Handout #3

Title: FAE — Spring/2015
Course: Econ 368 — Instructor: Dr. I-Ming Chiu

Sets, Functions, Counting Rules, Random Variables, and Probability Distributions

1. Sets and Subsets

Important definitions and theorems

Set: A collection of objects. Each object is an element. How to express a set: a. enumeration method, b. property method.
e.g. (a): Econ368 = {David, Minji, Mitch, Nhung, Nick, Ron, Sebastian, Theresa}
e.g. (b): L (labor force) = {x: x is a person who belongs to the civilian non-institutional population at age 16 and above who has a job or search for a job}

1. Subset: $X \subseteq Y$ (all the elements of a set X are also elements of set Y)
Proper subset: $X \subset Y$ (all the elements of a set X are also elements of set Y, but not all the elements of Y are in X; excludes the situation where X = Y)

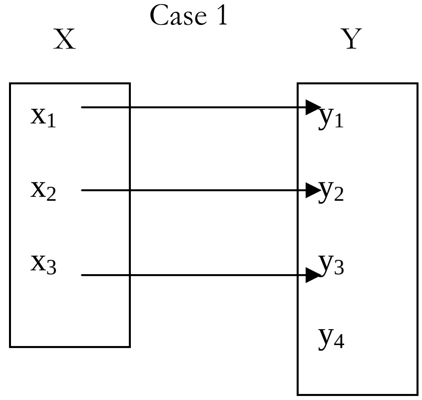
2. Set Operations (Venn diagram can be used to illustrate these operations)
Ω: Universal set (the largest set from the collection of an experiment outcomes)
a. Intersection: $C = A \cap B = \{x: x \in A \text{ and } x \in B\}$
b. Union: $D = A \cup B = \{x: x \in A \text{ or } x \in B\}$
c. Complement: $\overline{A} = \Omega - A = \{x \in \Omega: x \notin A\}$
d. Relative difference: $E = A - B = \{x: x \in A \text{ and } x \notin B\}$
e. Partition: $X_i \subset \Omega$ (i = 1…n), $X_i \cap X_j = \phi$ ($i \neq j$), $\cup_{i = 1 \ldots n} X_i = \Omega$.
e.g. Ω = labor force, X_1 = employed (has a job), X_2 = unemployed (look for a job)
f. Power set of a set X: $P(X) = \{A: A \subseteq X\}$. The total number of $P(X) = 2^N$; N; the number of elements in set X.

e.g.: Throw a dice once
$\Omega = \{1, 2, 3, 4, 5, 6\}$
$A = \{1, 3, 5\}$
$B = \{4, 5, 6\}$
1. $A \cup B$ =? Answer: {1, 3, 4, 5, 6}
2. $A \cap B$ =? Answer: {5}
3. Ω - A =? Answer: {2, 4, 6}
4. A - B =? Answer: {1, 3} Notice that A - B is equivalent to A - ($A \cap B$)
5. ($A \cup B$) - ($A \cap B$) =? Answer: {1, 3, 4, 6}

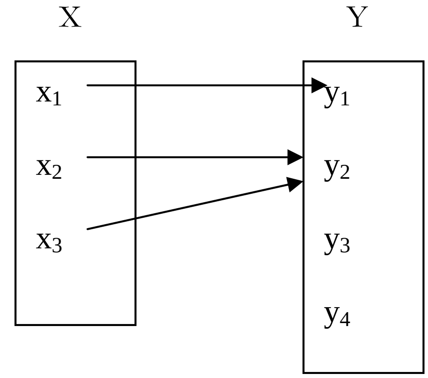
Notice: Three set operation commands in R: “union”, “intersect”, and “setdiff”.

2. Functions

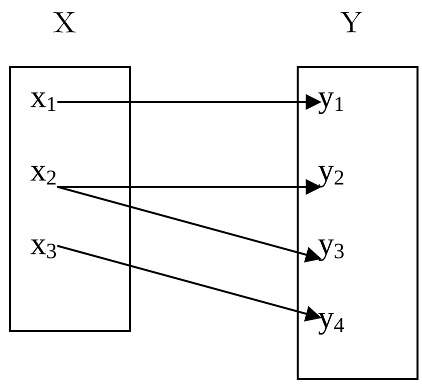
Case 1



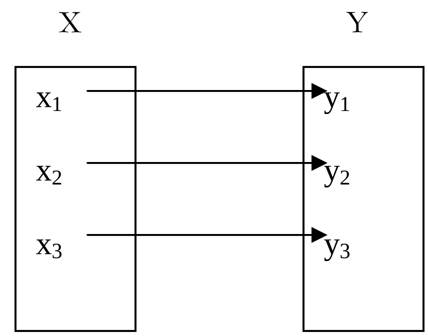
Case 2



Case 3



Case 4



Function

Def: A function from set X to set Y is a rule that associates with <u>each element</u> of X, <u>one and only one</u> element of Y.

Function is an association.

$$f: X \to Y$$

or $$y = f(x),\ x \in X$$

X: Domain
Y: Codomain
Range: the set of elements in Y; that is $\{y_1, y_2, y_3\}$ in case 1 from the above graphical

e.g. X = {contractionary policy, expansionary policy}, Y = {lower GDP, higher GDP}
X = {x: x = study time in # of hours}, Y = {y: y = exam scores}

Function is a mapping.
The range of a function can be written as the Image set:

$$f(X) = \{y \in Y: y = f(x),\ x \in X\}$$

e.g. for case 1 & 4, the image set is $\{y_1, y_2, y_3\}$.
$\{y_1, y_2\}$ is the image set for case 2.

Function: a) one-to-one: Case 1
b) Several-to-one: Case 2

Function: a) Maps onto: $f(X) = Y$, Case 4
b) Maps into: $f(X) \subset Y$, Case 1, 2.

The inverse of a function exists iff the function is one-to-one and maps onto.

3. Counting

Probability theory is a very important subject. I will write up a supplement to briefly introduce probability theory and post it on sakai soon.

Why do we have to count?

Multiplication Principle
e.g. There are three high ways from Boston to NYC and there are two high ways from NYC to Philadelphia, how may combination of routes someone can take traveling between Boston and Philadelphia?

Factorial (n!): $n! = n*(n-1)*(n-2)*...*2*1$
Note: $0! = 1$, $n! = n*(n-1)!$

Permutation: The number of permutations (ordered choices) of r objects from n objects is

$$P(n, r) = n*(n-1)*\ldots*(n-r+1) = \frac{n!}{(n-r)!}$$

Combination: The number of combinations (unordered choices) of r objects from n objects is

$$C(n, r) = \frac{n!}{r!(n-r)!}$$ (the R command for doing combination is "choose(n, r)".

Note: $C(n, r) = \frac{P(n, r)}{r!}$

e.g. A committee that consists of president, vice president, and treasury will be formed from a class of ten students, how many different committees can be formed?
Answer: P(10, 3)
e.g. Three students are needed to form a committee from a class of ten students, how many different committees can be formed?
Answer: C(10, 3)
e.g. What's the odd (probability) to win the grand prize of Powerball lottery?
Answer: 1/C(59, 5)*C(35, 1)

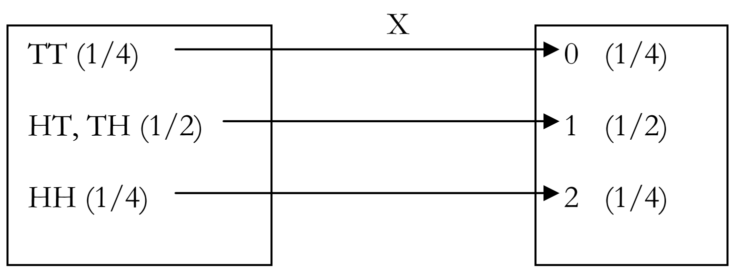
4. Random Variable (RV)

Consider the following experiment: Toss two fair coins simultaneously.
Universal set = {HH, HT, TH, TT}

Suppose we are interested in the number of heads in this experiment, the collection of events (event: a subset) would be {{TT}, {HT, TH}, {HH}}.

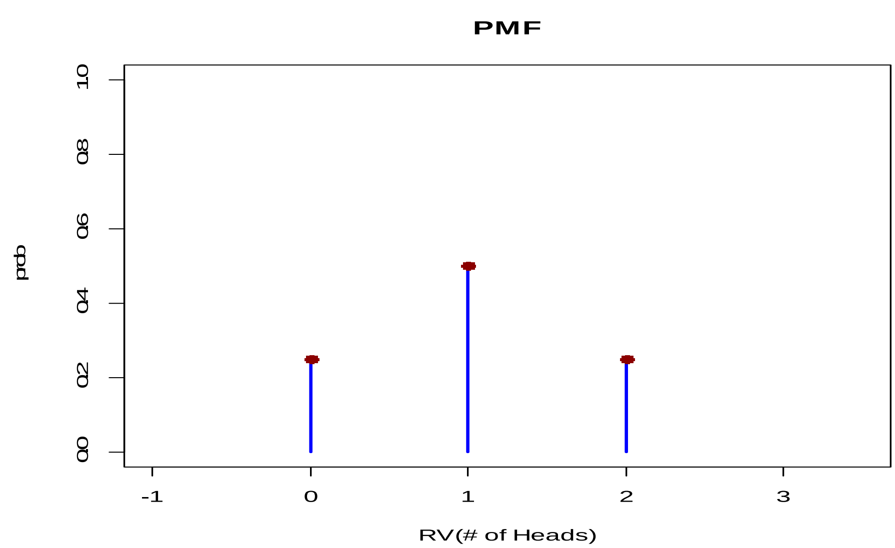
Let's define a function that maps this event to real numbers (non-negative integers 0, 1, 2).
The above statement defines a random variable, X (# of Heads).

The probability distribution function will be



Other e.g. X = return of buying a share of stock; it's a random variable because it's unlikely to know the outcome when it's purchased (i.e., selling price > = < buying price)

Probability Mass Function (PMF)
Def: Let X be a discrete random variable. The probability mass function (pmf) of X is the function $f : R \rightarrow R$ defined by $f(x) = P(X = x)$



Property of Probability Distribution:

a) $0 \leq P(X_i) \leq 1$

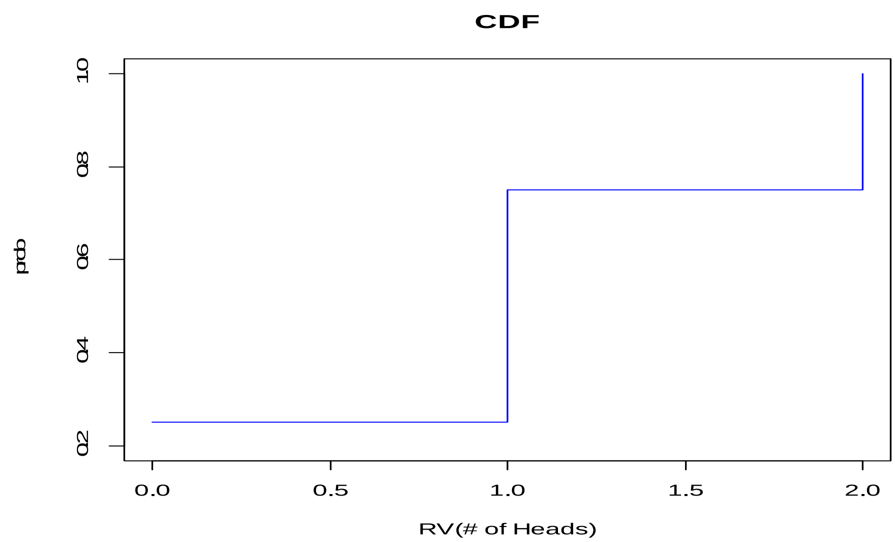
e.g. P (X = 0) = P(X = 2) = 1/4, P(X = 1) = 1/2

b) $\sum_{i=1}^{n} P(Xi) = 1$

e.g. ¼ + ½ + ¼ = 1

Cumulative Distribution Function (CDF)

Def: For each y ∈ R, let L(y) = { x ∈ X(S): x ≤ y} denote the value of X that are less than or equal to y. Then $F(y) = P(X \le y) = P(X \in L(y)) = \sum_{x \in L(y)} P(X = x) = \sum_{x \in L(y)} f(x)$



In order to describe the properties of a probability distribution, a measure called "the moments of a distribution" is commonly used. They include mean, variance, skewness, kurtosis, and higher moments. Here we only focus on the first two moments.

Mean: The expected value of a random variable X

$$\text{Mean } (\mu) = E(X) = \sum_{i=1}^{n} X_i * f(X_i) = 0*\frac{1}{4} + 1*\frac{1}{2} + 2*\frac{1}{4} = 1$$

Variance: The dispersion of a probability distribution

$$\text{Variance } (\sigma^2) = E(X - \mu)^2 = \sum_{i=1}^{n} (X_i - \mu)^2 * f(X_i) = (0-1)^2*\frac{1}{4} + (1-1)^2*\frac{1}{2} + (2-1)^2*\frac{1}{4} = \frac{1}{2}$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\sum_{i=1}^{n} (X_i - \mu)^2 * f(X_i)} = \sqrt{\frac{1}{2}}$$

e.g.: X is a random variable that denotes the number of heads appeared in an experiment of tossing three fair coins.
What is the mean, variance and standard deviation of random variable X?

Theorem: $E(a + b*X) = a + b*E(X)$, $Var(a + b*X) = b^2*Var(X)$

5. Probability Distributions

Discrete RV and PMF

(a) Bernoulli Distribution

Binary random variables (e.g., healthy/diseased) are abundant in scientific studies. There are also numerous binary random variables exist in economic studies; e.g., whether a labor is employed or unemployed, whether a worker is a female or male, whether the economy is expanding or contracting (i.e., business cycle), etc.

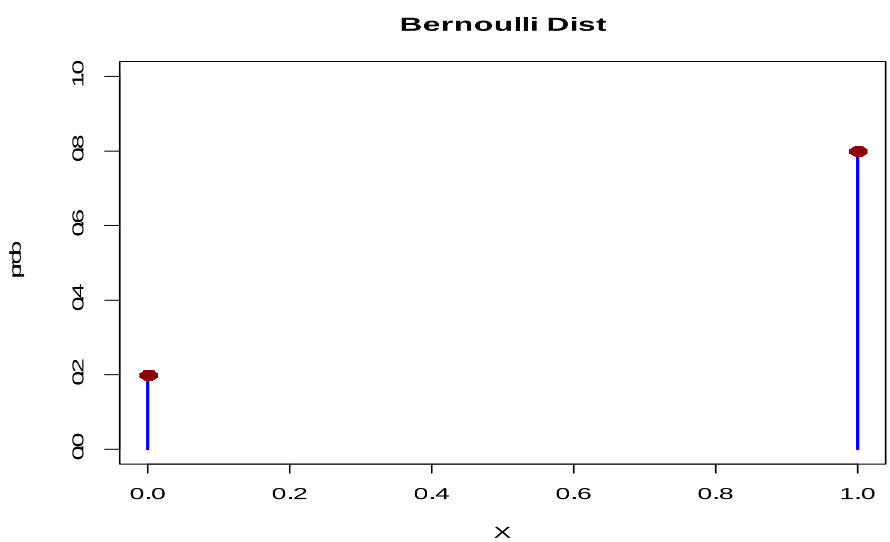
The binary random variable X with possible values 0 and 1 has a Bernoulli distribution with parameter θ.

Here, $P(X = 1) = \theta$ and $P(X = 0) = 1 - \theta$

e.g.
$P(X = x) = 0.2$ for $x = 0$
$P(X = x) = 0.8$ for $x = 1$

We denote this as $X \sim \text{Bernoulli}(\theta)$, where $0 \leq \theta \leq 1$.

$E(X) = \theta$
$Var(X) = \theta(1 - \theta)$



(b) Binomial Distribution

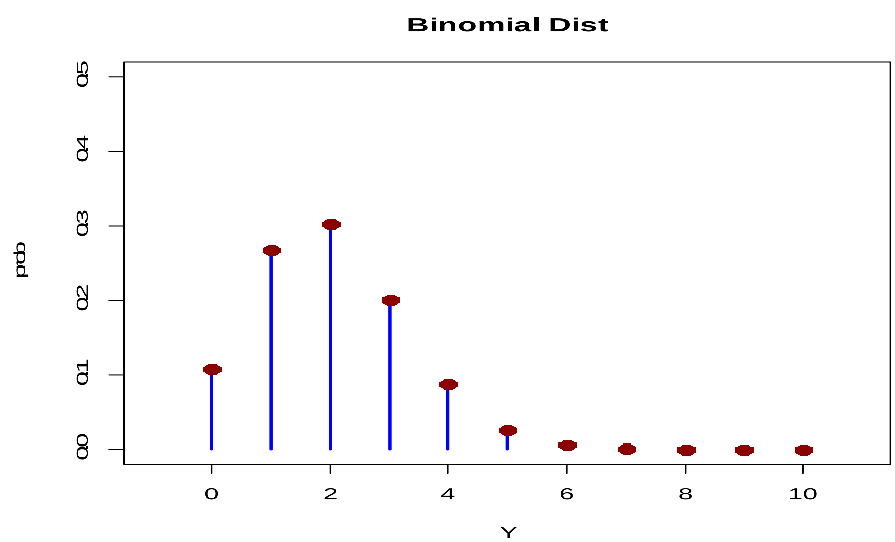
A sequence of binary random variables $X_1, X_2, \ldots, X_n$ is called Bernoulli trials if they all have the same Bernoulli distribution and are independent.

The random variable Y representing the number of times the outcome of interest occurs in n Bernoulli trials (i.e., the sum of Bernoulli trials) has a Binomial(n, θ) distribution. In other words, $Y = X_1 + X_2 + \ldots + X_n$

The pmf of a Binomial(n, θ) specifies the probability of each possible value (integers from 0 through n) of the random variable.

$E(Y) = n\theta$
$Var(Y) = n\theta(1 - \theta)$

In the above, the theoretical (population) mean of a random variable Y with Binomial(n, θ) distribution is $\mu = n\theta$. The theoretical (population) variance of Y is $\sigma^2 = n\theta(1 - \theta)$.



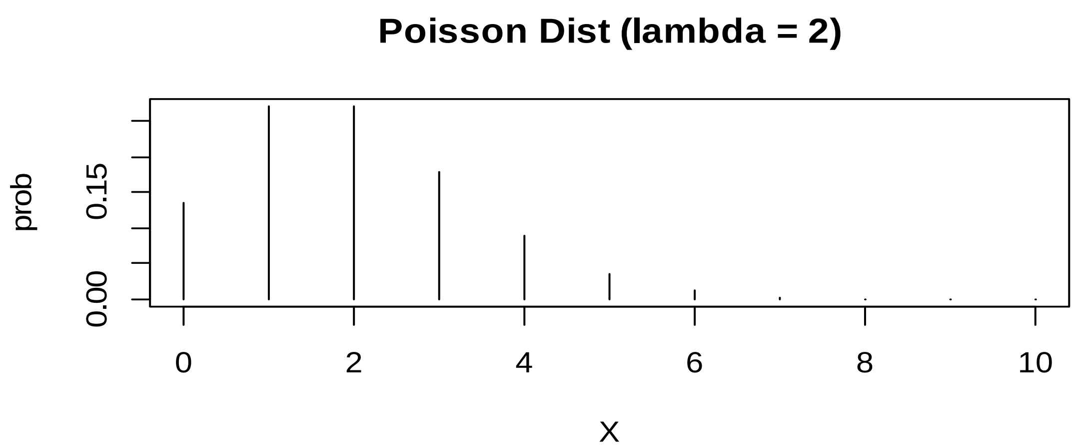
(c) Poisson distribution

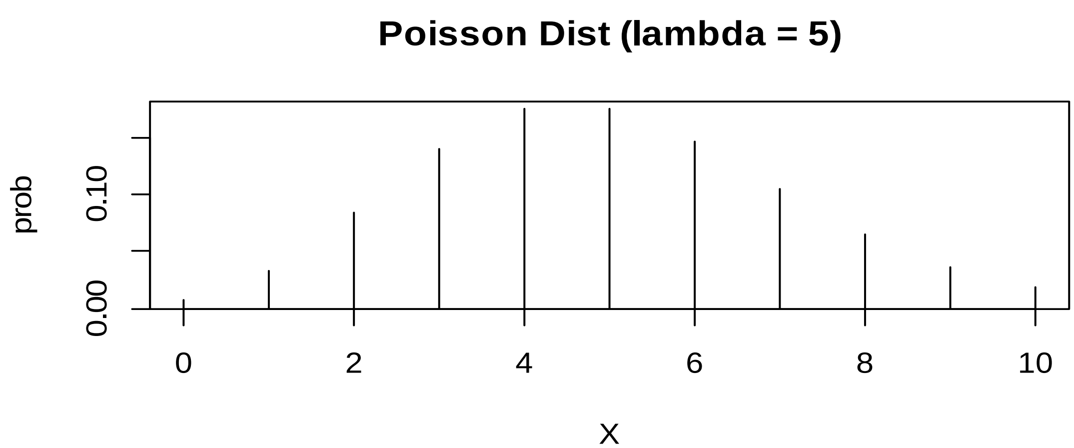
A discrete random variable X is said to have a Poisson distribution with parameter $\lambda > 0$, if for x = 0, 1, 2, ... the probability mass function of X is given by:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{(e: exponential function, x!: x factorial)}$$

E(X) = Var(X) = λ (both the mean and variance of the Poisson random variable equal λ)

Notice that $P(X = x)/P(X = x\text{-}1) = \frac{\lambda}{x}$, so we can obtain the P(X) recursively starting from $P(X = 0) = e^{-\lambda}$

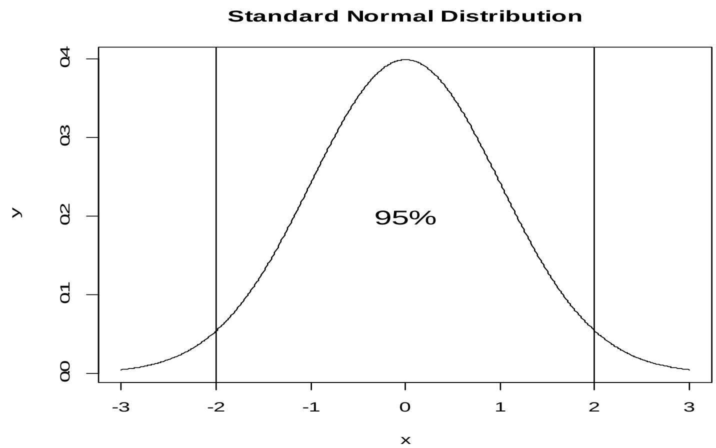




Continuous RV and PDF: The Family of Sampling Distribution

(a) Normal distribution (two parameter distribution)
Suppose $X \sim N(\mu, \sigma^2)$; that is random variable X is normally distributed with mean μ and variance σ^2 and has the probability density function f(X)[1] = $\frac{1}{\sqrt{2\pi\sigma^2}} \exp[\frac{-(X-\mu)^2}{2\sigma^2}]$

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



Property of Normal Distribution:
1. $f(\mu + x) = f(\mu - x)$
2. 68-95-99 rule
3. f(x) decreases as x is moving away from μ.

e.g. If X~N(1, 4), then what is the probability that X assumes a value no more than 3?

e.g. If X~N(2, 16), then what is the probability that X assumes a value between 0 and 10?

e.g. If X~N(-3, 25), then what is the probability that $|X|$ assumes a value greater than 10?

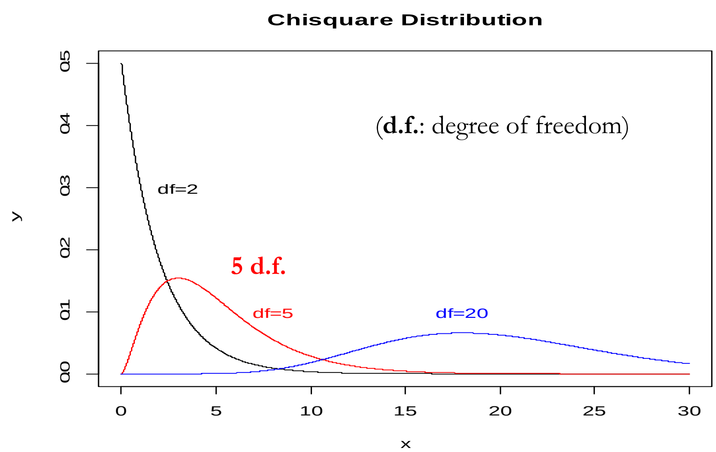
e.g. If X~N(4, 16), then what is the probability that X^2 assumes a value less than 36?

Theorem: If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, given that X_1 and X_2 are independent.

[1] We will use Φ to denote CDF of X.

(b) χ^2 Distribution (one parameter distribution; degree of freedom)

If Z_i (i=1…n) are all independently distributed standard normal distribution (i.e., $Z_i \sim N(0, 1)$), then $\sum_{i=1}^{n} Z_i^2$ is said to have a chi-squared distribution with degree of freedom n, χ^2_n.



Theorem: If $Z_1, Z_2, \ldots, Z_n \sim N(0, 1)$ and $Y = Z_1^2 + Z_2^2 + \ldots + Z_n^2$, then $Y \sim \chi^2(n)$

Show that E(Y) = n

Proof:

$Z_i \sim N(0, 1) \Rightarrow E(Z_i) = 0 \; Var(Z_i) = 1$

$Var(Z_i) = E(Z_i^2) - (E(Z_i))^2 = 1 \Rightarrow E(Z_i^2) = 1$

$E(Y) = E(\sum_{i=1}^{n} Z_i^2) = 1 + 1 + ... + 1 = n$

*Note: Var(Y) = 2n; it's easier to show it using Moment Generating function.

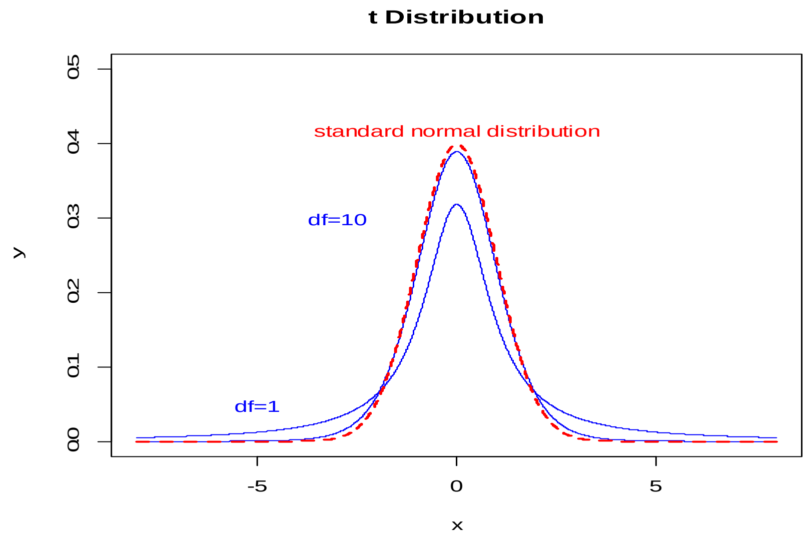
Since Y (Chi-squared r.v.) is non-negative, you can define it in $(0, \infty)$.

e.g. Suppose $Y \sim \chi^2(3)$, please find $\Phi(0.95)$ using R.

e.g. Suppose $Y \sim \chi^2(5)$, please find $f(2 < Y < 8)$ using R.

(c) t Distribution (one parameter distribution; degree of freedom)

If Z_i (i=0…n) are all independently distributed standard normal distribution, then $\frac{Z_0}{\sqrt{\sum_{i=1}^{n} Z^2{}_i / n}}$ is said to have a t distribution.



Theorem: $Z \sim N(0, 1)$ and $Y \sim \chi^2(n)$, then $T = \frac{Z}{\sqrt{Y/n}}$ has a t distribution with degree of freedom equals n [using symbol t(n)], given that Z and Y are independent.

Theorem: Let F_n denote the CDF of t(n) and let Φ denote the CDF of N(0, 1). Then

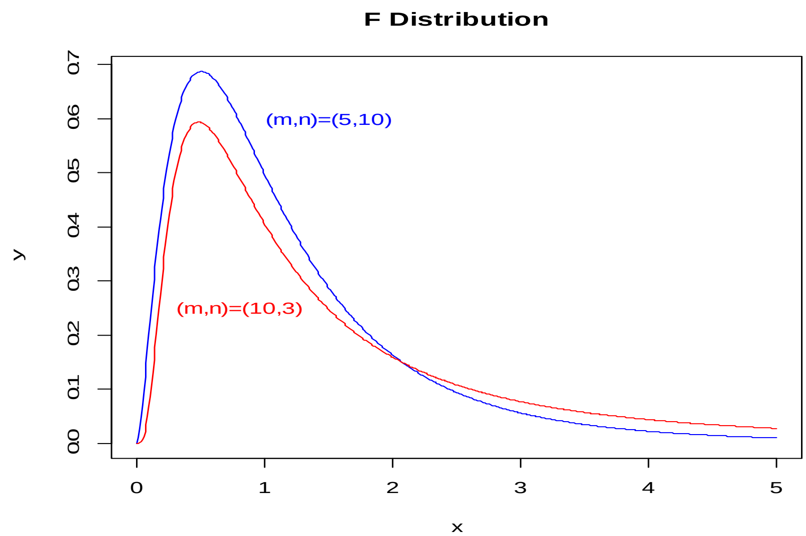
$$\lim_{n\to\infty} F_n(y) = \Phi(y) \text{ for } \forall\, y \in (-\infty, \infty)$$

The above theorem indicates that when degree of freedom in t distribution becomes large, then t distribution can be approximately represented by a standard normal distribution.

e.g. If T ~ t(14) and T assumes a values no greater than -1.5, please find the probability of T using R.

Following the above example, let's increase the degree of freedom of n using 100, 500 and 1000.

(d) F Distribution (two parameter distribution; a pair of degree of freedom)
If V_1 and V_2 are two independent random variables having the Chi-Squared distribution with m_1 and m_2 degrees of freedom respectively, then the following quantity follows an F distribution with m numerator degrees of freedom and n denominator degrees of freedom, i.e., (m, n) degrees of freedom.



Theorem: Let $Y_1 \sim \chi^2(m)$ and $Y_2 \sim \chi^2(n)$ be independent variable, then $F = \dfrac{Y_1/m}{Y_2/n}$ is an F distribution with degree of freedom m and n, respectively. It is denoted by F(m, n).

Theorem: If $T \sim t(n)$, then $T^2 \sim F(1, n)$

Proof:

$$T = \frac{Z}{\sqrt{\chi^2/n}} \Rightarrow T^2 = \frac{Z^2}{\chi^2/n} = \frac{\chi^2/1}{\chi^2/n}, \text{ therefore, } T^2 \sim F(1, n)$$

e.g. If $F \sim F(2, 27)$, please find $P(F > 2.5)$ using R.

Appendix:
Statistical distributions and commands in R

- p for "probability", the cumulative distribution function (c. d. f.)
- q for "quantile", the inverse c. d. f.
- d for "density", the density function (p. f. or p. d. f.)
- r for "random", a random variable having the specified distribution

Normal distribution: A two parameter distribution.

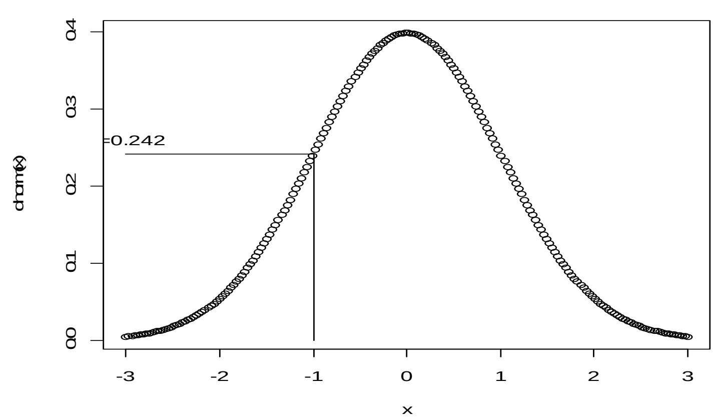
Suppose X~N(μ, σ²); that is $f(X) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp[\frac{-(X-\mu)^2}{2\sigma^2}]$

Standardization ⇒ $\frac{X-\mu}{\sigma}$ ~N(0, 1)

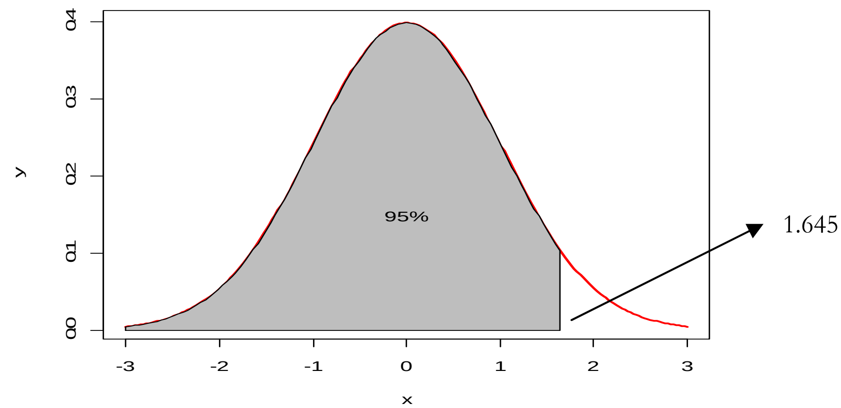
How to use these four commands, (p, q, d, r)?

We will use normal distribution to illustrate the use of (p, q, d, r).

a) **dnorm(x, mean, sd)**; default: mean=0, sd=1.
After defining the domain x, "dnorm" function can be used to find the probability of standard normal distribution given the mean and standard deviation (sd). "d" is for density.
e.g. dnorm(-1), dnorm(0), dnorm(1)
e.g. graph of dnorm(-1)



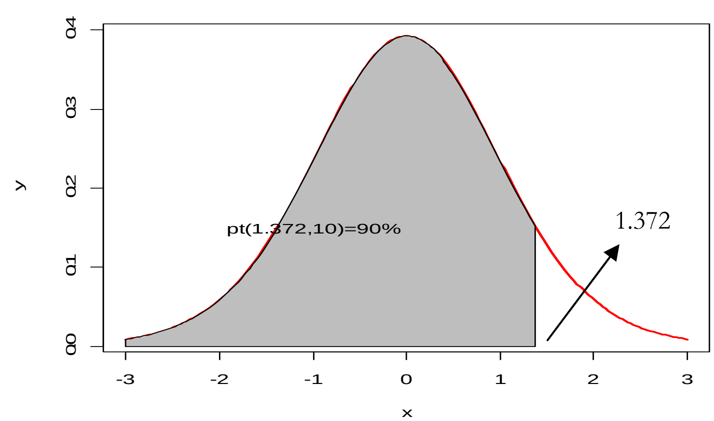
b) **pnorm(x, mean, sd)**
The "pnorm" function provides the cumulative probability from the -∞ till the x value given that mean and sd. "p" is for "probability".
e.g. pnorm(1.645), pnorm(1.96)
e.g. graph of pnorm(1.645)



c) **qnorm(percent, mean, sd)**
The function of "qnorm" is the anti-function of "pnorm". Here "q" is for "quantile". Suppose the value of x is unknown, but we know the cumulative probability at x is 95%. The "qnorm" function can help us find x.
e.g. qnorm(0.95), qnorm(0.975)

d) **rnorm(#, mean, sd)**
The function "rnorm" is used to generate # random numbers that has a normal distribution with given mean and sd. "r" is for "random".
e.g. rnorm(10)

These four commands can be applied to other distributions. For example, "dt", "pt", "qt" and "rt". But notice that t distribution depends on the degree of freedom instead of mean and standard deviation. e.g. pt(1.372, 10) in the diagram below, where 10 is the degree of freedom.



Check the following web site about statistical distributions using R:

http://www.stat.umn.edu/geyer/old/5101/rlook.html