Multiple Random Variables

In last course notes, we only consider the case of one single random variable. In the note, we consider that case the there are multiple rv's. We aim to understand how to study the relationship among these rv's.

Probability distribution of two random variables

Discrete variables: Assume that there are two random variables, X and Y, and both of them are Discrete.

Joint Probability Mass function of X and Y:

$$p(x,y)=P(X=x \text{ and } Y=y)$$

For any set A which is a subset of R^2, its associated probability is

$$P((X,Y)\in A)=\sum\sum_{(x,y)\in A} p(x,y)$$

Note: joint probability is a legitimate probability mass function defined on 2-dimensional real space. All the probability axioms still hold.

Example: Consider a population (of total number 1000) associated with 2 discrete variables X and Y. The census data is given in the following table:

Number		Variable Y	
		15	25
Variable X	10	200	50
	20	100	150
	30	200	300

The joint pmf table is

p(x,y)		y	
		15	25
x	10	0.2	0.05
	20	0.1	0.15
	30	0.2	0.3

$P(Y=15)=P((X,Y)\in\{(10,15),(20,15),(30,15)\})$ = 0.2+0.1+0.2=0.5;
$P(Y=X+5)=P((X,Y)\in\{(10,15),(20,25)\})$ = 0.2+0.15=0.35;

Marginal Probability Mass function of X and Y:

$$p_X(x)=P(X=x)=P\big((X,Y)\in\{(x,r):r\in R\}\big)=\sum_{\text{all possible y}} p(x,y)$$
$$p_Y(y)=P(Y=y)=P\big((X,Y)\in\{(r,y):r\in R\}\big)=\sum_{\text{all possible x}} p(x,y)$$

The marginal pmf of X is exactly the same pmf of X when only information about X is presented.

<u>Continuous variables</u>: Assume that there are two random variables, X and Y, and both of them are continuous.

<u>Joint Probability density function of X and Y</u>: is the non-negative function f(x,y) satisfies

$$P((X,Y)\in A)=\iint_A f(x,y)dxdy\,.$$

For example,

$$P(a<X<b,c<Y<d)=\int_c^d\int_a^b f(x,y)\,dxdy.$$

<u>Marginal Probability density function of X and Y:</u>

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

Similarly, the marginal pdf of X is exactly the same pdf of X when only information about X is presented.

Example: A bank operates both a drive-up facility and a walk-up window. On a randomly selected day, let *X* = the proportion of time that the drive-up facility is in use (at least one customer is being served or waiting to be served) and *Y* = the proportion of time that the walk-up window is in use. Assume the joint pdf of (X,Y) is

$$f(x,y)=6xy^2 \text{ if } 0\leqslant x,y\leqslant 1;\ f(x,y)=0 \text{ otherwise.}$$

Verify that it is a valid joint pdf:

What is the probability P(X+Y>1):

What is the marginal pdf's of X and Y:

<u>More than two random variables</u>

We can easily generalize two-variable case to the case that there are more than two random variables.

<u>Definition</u>: If $X_1,\ldots, X_n$ are all discrete random variables, the joint pmf of the variables is the function

$$p(x_{1,\ldots},x_n)=P(X_1=x_{1,}\ldots,X_n=x_n)$$

If the variables are continuous, the joint pdf of $X_1,\ldots, X_n$ is the nonnegative function $f(x_1, x_2, \ldots, x_n)$ such that for any n intervals $[a_1, b_1], \ldots, [a_n, b_n]$,

$$P(a_1 \leqslant X_1 \leqslant b_1, \ldots, a_n \leqslant X_n \leqslant b_n) = \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \ldots dx_1 .$$

Example: When a certain method is used to collect a fixed volume of rock samples in a region, there are four resulting rock types. Let X_1, X_2, and X_3 denote the proportion by volume of rock types 1, 2 and 3 in a randomly selected sample (the proportion of rock type 4 is $1-X_1-X_2-X_3$, so a variable X_4 could be redundant). If the joint pdf of X_1, X_2, and X_3 is

$$f(x_1, x_2, x_3) = 144\, x_1 x_2 (1-x_3) \text{ if } 0 \leqslant x_1, x_2, 1-x_1-x_2 \leqslant 1; \text{ and } f(x, y) = 0 \text{ otherwise.}$$

Verify that it is a valid joint pdf:

What is the probability $P(X_1+X_2<0.5)$:

Independence of random variables

Independence among events: Probability of intersection = product of probability of each events.
Independence among rv's: Joint Distribution = product of marginal distribution.
Formal definition:
Random variables X_1 and X_2 are independent if

$$p(x, y) = p(x)p(y) \text{under discrete case; } f(x, y) = f(x) f(y) \text{under continuous case.}$$

Random variables $X_1,\ldots, X_n$ are mutually independent if

$$p(x_{i_1}, \ldots x_{i_k}) = p(x_{i_1}) \ldots p(x_{i_k}) \text{under discrete case; } f(x_{i_1}, \ldots x_{i_k}) = f(x_{i_1}) \ldots f(x_{i_k}) \text{under continuous case,}$$

for any sub indices $i_1,\ldots,i_k$.

Proposition: If X and Y are independent, then

$$P(X \in (a, b), Y \in (c, d)) = P(X \in (a, b))\, P(Y \in (b, c))$$

Example:

(**parallel circuit**) If $X_1,\dots, X_n$ are mutually independent exponential variable, what is the distribution of $\min(X_1,\dots, X_n)$.

X_1 and X_2 have joint pdf:

$$f(x,y)=1/4 \text{ if } 0<x, y<2;\ f(x,y)=0 \text{ otherwise.}$$

Are X_1 and X_2 independent?

X_1 and X_2 have joint pdf:

$$f(x,y)=1/\pi \text{ if } x^2+y^2<1;\ f(x,y)=0 \text{ otherwise.}$$

Are X_1 and X_2 independent?

Expected value, Covariance and Correlation

Definition: Given random variables X and Y, and any function h, the expected mean of h(X,Y) is defined as

$$E(h(X,Y))=\sum_x \sum_y h(x,y)p(x,y) \text{ under discrete case, and}$$
$$E(h(X,Y))=\iint h(x,y)f(x,y)\,dx\,dy \text{ under continuous case.}$$

Note:

By definition of the expected value under one variable case, $EX=\int xf(x)dx$.

Under two variable case, define mapping h(x,y)=x,

$$EX=Eh(X,Y)=\iint xf(x,y)dxdy=\int xf_X(x)dx.$$

There is no contradiction.

Proposition:

$$E(h_1(X,Y)+h_2(X,Y))=Eh_1(X,Y)+Eh_2(X,Y);$$
$$E(X+Y)=EX+EY;$$
$$E(h(X)g(Y))=Eh(X)Eg(Y) \text{ if X and Y are independent.}$$

Example: If X and Y are independent Uniform distribution on (0,1). Compute E(|X-Y|).

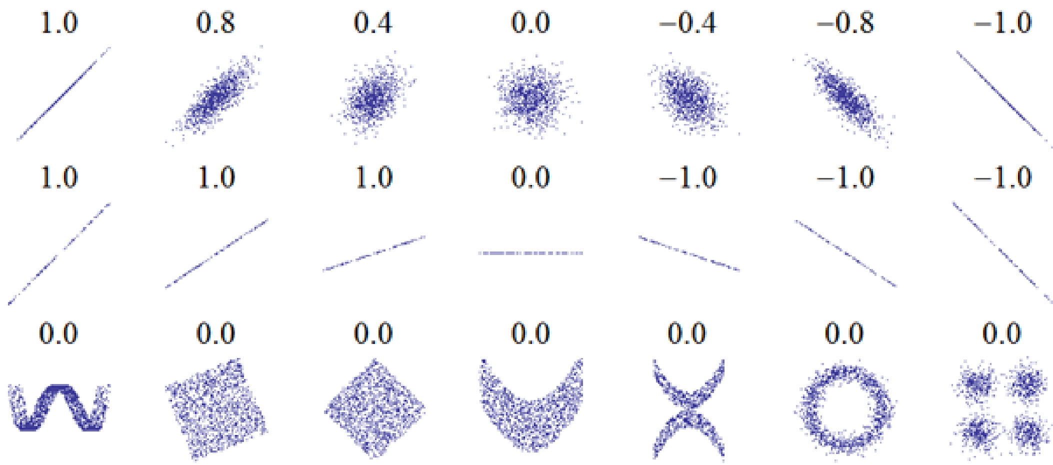
<u>Covariance</u>: $Cov(X,Y)=E(X-\mu_X)(Y-\mu_y)=EXY-\mu_X\mu_Y$.

<u>Correlation (standardized covariance)</u>: $Corr(X,Y)=\rho_{X,Y}=\dfrac{Cov(X,Y)}{\sigma_X\sigma_Y}$.

<u>Proposition</u>:

- $-\sigma_X\sigma_Y\leqslant Cov(X,Y)\leqslant\sigma_X\sigma_Y$, $|\rho_{X,Y}|\leqslant 1$.
- $Cov(X,X)=\sigma_X^2$
- $Cov(X,\text{Constant})=0$
- $Cov(aX+c,bY+d)=a*b*Cov(X,Y)$ and $Corr(aX+c,bY+d)=Corr(X,Y)$;
- $Cov(X,Y)=\rho_{X,Y}=0$ if X and Y are independent.

Covariance and correlation describe the **linear dependence** between X and Y



Corr(X,Y) = 1 or -1 if and only if ____.

Example:
Compute the covariance for joint pdf:
$f(x,y) = 24xy$ if $0 \leqslant x, y, x+y \leqslant 1$, and $f(x,y) = 0$ otherwise.

Sampling Distributions
Measuring how statistics vary from sample to sample

Statistics: A statistic is a numerical summary of the data values measured on subjects in the sample. More specifically, **any function** of the sample data $h(X_1,\ldots,X_n)$ is a statistic.

Example: sample mean, sample median, sample variance, or even the sample data itself.

Statistics are random variables as well, and it has its own distribution. We use an uppercase letter to denote a statistic, and use a lowercase letter to denote the value of statistic given a fixed data.

The distribution of a statistic is call **sampling distribution**, the sampling distribution is completely determined by the joint distribution of $X_1,\ldots,X_n$.

Examples:

1)

Joint pmf of (X_1,X_2)		X_1	
		0	1
X_2	0	0.16	0.24
	1	0.24	0.36

pmf of $Z=X_1+X_2$	0	1	2
p(z)	0.16	0.48	0.32

2)

Joint pmf of (X_1,X_2)		X_1	
		0	1
X_2	0	0.20	0.20
	1	0.20	0.40

pmf of $Z=X_1+X_2$	0	1	2
p(z)	0.2	0.4	0.4

In the above 2 examples, the marginal distribution of X_1 and X_2 are the same. But in 1) there is independence; in 2) X_1 and X_2 are dependent.

In this lecture, we focus on the simplest case that is

1. $X_1,\dots,X_n$ are independent;
2. $X_1,\dots,X_n$ has the same distribution.

In such case, we say that $X_1,\dots,X_n$ are a **simple random sample** of size n.

Examples:

An experiment consists of tossing a dice twice. Let X_1 and X_2 be the 2 values. What is the distribution of $(X_1 + X_2)/2$?

Let X_1 and X_2 be the 2 independent exponential rv's with exponential parameter λ, what is the distribution of $(X_1 + X_2)/2$?

General approach:

1) The cdf of $T= (X_1 + X_2)/2$ is

$$F(t)=P(T\leqslant t)=P(X_1+X_2\leqslant 2t)=\iint_{0\leqslant x_1+x_2\leqslant 2t} f(x_{1,}x_2)\,dx_1dx_2=\int_0^{2t}\int_0^{2t-x_1} f(x_{1,}x_2)\,dx_1dx_2=$$

2) The pdf is the derivative of cdf: $f(t)=F'(t)=$

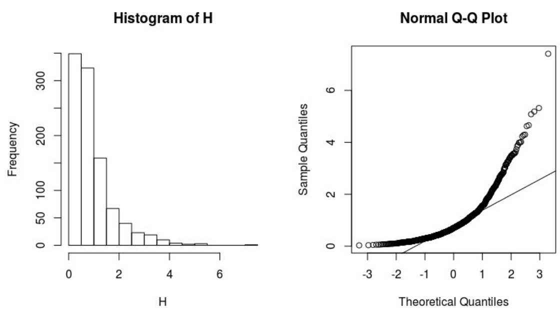
Arithmetically, it may be difficult to derive the distribution of statistic when the distribution of X is complicated or n is large. One can use computer simulation is study the distribution of $h(X_1,\ldots,X_n)$

Compuer simulation study

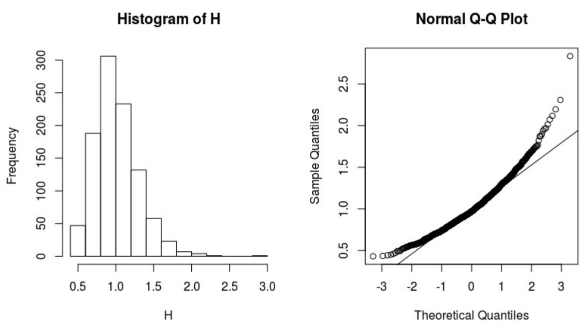
1. Simulation a size n random sample $D_1=(x_1,\ldots,x_n)$ based on the distribution of X.
2. Compute the value of statistic base on D_1, $h_1=h(D_1)=h(x_1,\ldots,x_n)$.
3. Repeat 1 and 2 many many times (say 1000 times, depending your computing power) and obtain $h_1 ,\ldots, h_n$.
4. Make a histogram of $(h_1 ,\ldots, h_n)$.

R code to derive the sampling distribution of sample variance of n=9 exponential observations:

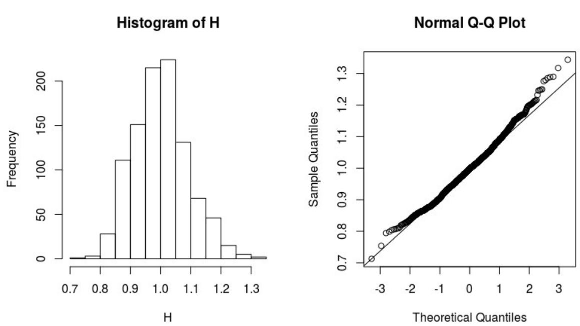
```
n=9; N=1000; H<-numeric(N)

for(i in 1 : N)
{
   D<-rexp(n, 1); H[i]<-var(D);
}

hist(H);qqnorm(H); qqline(H)
```



n=9



n=90



n=900

There are 3 Distinct Distribution for a numerical variable:

- **Population Distribution** (theoretical distribution)
- **Data Distribution** – Observed distribution of data from a single sample.
- **Sampling Distribution** (theoretical distribution) - Probability Distribution of some statistic, it is is the distribution of the set of statistic calculated from the set of all possible samples from the parent population distribution.

Sampling Distribution of the Sample Mean

The previous simulation studies show that when n is sufficient large, the sampling distribution of sample variance and sample mean is more and more normal. Actually, most of the statistics that based on average have this property. Particularly, we study the sampling distribution of sample mean in this section.

An very important statistical result: Central Limit Theory (CLT)

$X_1,\ldots,X_n$ are a simple random sample from a distribution with mean μ and variance σ^2. Then when n is sufficiently large, the sampling distribution of sample mean $\bar{X}$ is approximately normal distribution with mean μ and variance σ^2/n, and the sampling distribution of sample sum $\sum X_i$ is approximately normal distribution with mean $n\mu$ and variance $n\sigma^2$. The larger the value of n, the better this approximation.

Important Comments:

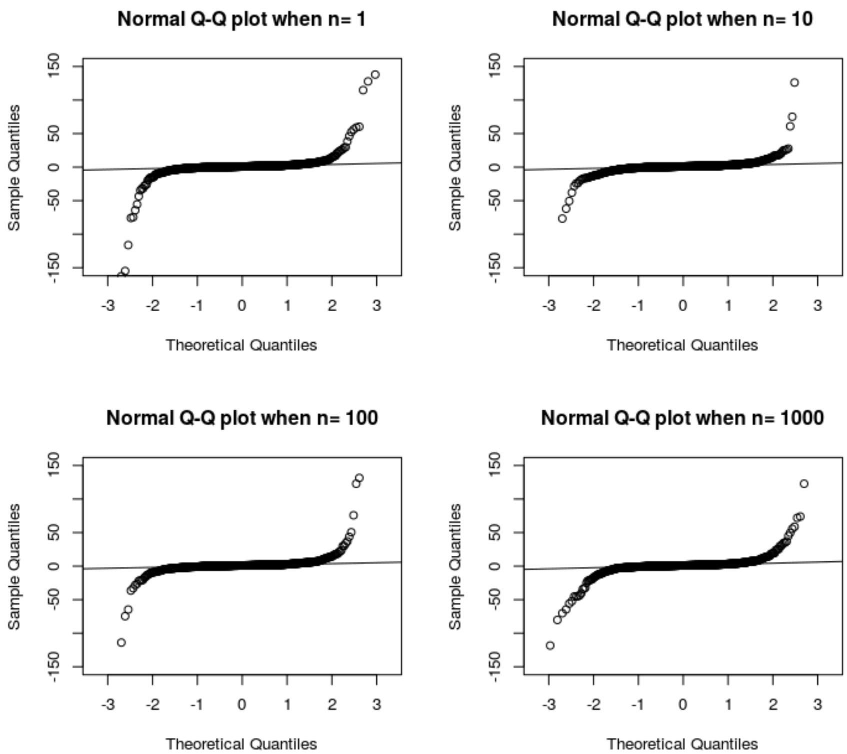
1. What does "good approximation" mean?
   It is good for probability calculation, that is

   $$P(a<\bar{X}<b)=P\left(\text{Normal}(\mu,\sigma^2/n)\in(a,b)\right)=P\left(\frac{a-\mu}{\sigma/\sqrt{n}}<Z<\frac{b-\mu}{\sigma/\sqrt{n}}\right).$$

   A certain consumer organization customarily reports the number of major defects for each new automobile that it tests. Suppose the number of such defects for a certain model is a random variable with mean value 3.2 and standard deviation 2.4. Among 100 randomly selected cars of this model, how likely is it that the sample average number of major defects exceeds 4?

2. What does “large n” mean?It depends on how normal the parent population distribution is. If the parent distribution is exactly normal, the $\bar{X}$ has an **<u>exact</u>** normal distribution for any n, even n=1. If the parent distribution has a very non-normal shape, it requires large n. Textbook gives a Rule of Thumb: n>30.
   Actually, given a fixed n, there always exists a distribution such that its sampling distribution of sample mean doesn’t look like normal.
   Given a fixed parent population distribution, there always exists a large integer N, such that its sampling distribution of sample mean look like normal when n is large than N.

3. CLT holds if and only if the parent population has a variance that is not infinite.
   Counterexample: Cauchy distribution:



Examples:
1) Recall the fact that we can use normal to approximate a Binomial distribution. This is a simple fact following by the CLT.

2) CLT for the sample proportion.

3) Let $X_1,\ldots,X_n$ be a random sample from a distribution for which only positive values are possible [P(X_i> 0) =1]. Then if n is sufficiently large, the product $Y = \prod X_i$ has approximately a lognormal distribution.

Distribution of Linear Combination

In this section, we consider the statistics:

$$Y = a_1 X_1 + \ldots + a_n X_n,$$

given fixed coefficients a_i's.
Note sample mean is a special case where a_i=(1/n).

Proposition:
Let X_i be a random variable with mean μ_i and variance σ_i^2. Then,

1. $EY = E(a_1 X_1 + \ldots + a_n X_n) = a_1 E(X_1) + \ldots + a_n E(X_n) = a_1 \mu_1 + \ldots + a_n \mu_n.$

2. If X_i are mutually independent rv's, then
$$Var(Y) = a_1^2 Var(X_1) + \ldots + a_n^2 Var(X_n) = a_1^2 \sigma_1^2 + \ldots + a_n^2 \sigma_n^2.$$

3. In general
$$Var(a_1 X_1 + \ldots + a_n X_n) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j Cov(X_i, X_j) = \sum_{i=1}^n a_i^2 Var(X_i) + 2 \sum \sum_{i<j} Cov(X_i, X_j).$$

Note: If we let a_i be negative, then we get formula for the different.

The above proposition concerns only mean and variance, but not the distribution. In general, the distribution shape of a linear combination is not easy to figure out.

Usually, a linear combination of normal variable is also a normal random variable. Since a normal distribution is completely determined by its mean and variance, we only need to ___________ _____________________________________ to figure out the distribution of a linear combination of normal rv's.