# VE401 Recitation Class Note4

## Multivariate Random Variables

Chen Siyi
siyi.chen_chicy@sjtu.edu.cn

# 1 Definition

**Discrete**:

Let S be a sample space and $\Omega$ a countable subset of $\mathbb{R}^n$. A discrete multivariate random variable is a map

$$X : S \to \Omega$$

together with a function $f_X : \Omega \to \mathbb{R}$ with the properties that

(i) $f_X(x) \geqslant 0$ for all $x = (x_1, ..., x_n) \in \Omega$ and

(ii) $\sum_{x\in\Omega} f_X(x) = 1$

**Continuous**: Let S be a sample space. A continuous multivariate random variable is a map

$$X : S \to \mathbb{R}^n$$

together with a function $f_X : \mathbb{R}^n \to \mathbb{R}$ with the properties that

(i) $f_X(x) \geqslant 0$ for all $x = (x_1, ..., x_n) \in \mathbb{R}^n$ and

(ii) $\int_{\mathbb{R}^n} f_X(x)dx = 1$

# 2 Density and Independence

## 2.1 Marginal Density

$$f_{X_k}(x_k) = \sum_{x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n} f_{\mathbf{X}}(x_1, \ldots, x_n)$$

$$f_{X_k}(x_k) = \int_{\mathbb{R}^{n-1}} f_X(x)dx_1 \ldots dx_{k-1}dx_{k+1} \ldots dx_n$$

## 2.2 Conditional Density

$$f_{X_1|x_2}(x_1) := \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)} \quad \text{with } f_{X_2}(x_2) > 0$$

## 2.3 Independence

Two continuous random variables are independent if:

$$f_X(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

A full set of n components of a multivariate random variable is independent if:

$$f_X(x_1, ..., x_n) = f_{X_1}(x_1) ... \cdot f_{X_n}(x_n)$$

**Question: Visualization**

X and Y are continuous random variables. X takes on values between 0 and 2 while Y takes on values between 0 and 1. Their joint PDF is indicated below in the graph.

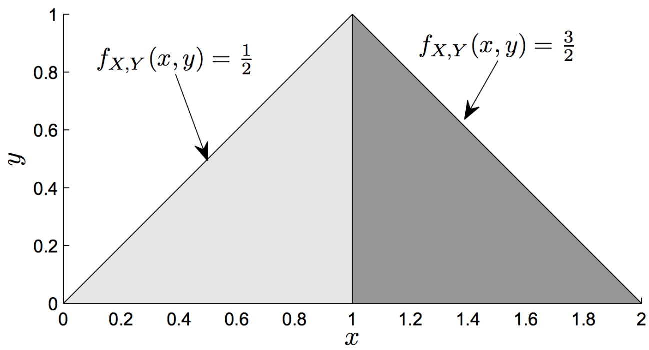


Figure 1: The exponential distribution

1. Are X and Y independent?
2. Find $f_X(x)$ and $f_{Y|X}(y|0.5)$.

**Answer: Visualization**

1. If X and Y to be independent, any observation of X should not give any information on Y. But in the graph, we see X is observed to be equal to 0, then Y must be 0. You can also describe it mathematically:

$$f_{(X,Y)}(0,0) \neq f_X(0) \cdot f_Y(0)$$

Since $f_{(X,Y)}(0,0) = \frac{1}{2}$, $f_X(0) = 0$, $f_Y(0) = 2$.

You can also think in other ways...

2.

$$f_X(x) = \begin{cases} x/2, & \text{if } 0 \leq x \leq 1 \\ -3x/2 + 3, & \text{if } 1 < x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

$$f_{Y|X}(y|0.5) = \begin{cases} 2, & \text{if } 0 \leq y \leq 1/2 \\ 0, & \text{otherwise} \end{cases}$$

# 3 Expectation

**Definition:**

$$\mathrm{E}[X] = \begin{pmatrix} \mathrm{E}[X_1] \\ \vdots \\ \mathrm{E}[X_n] \end{pmatrix}$$

$$\mathrm{E}[X_k] = \sum_{x_k} x_k f_{X_k}(x_k) = \sum_{x \in \Omega} x_k f_X(x)$$

$$\mathrm{E}[X_k] = \int_{\mathbb{R}} x_k f_{X_k}(x_k)\, dx_k = \int_{\mathbb{R}^n} x_k f_X(x) dx$$

**Property:**

For a function $\varphi : \mathbb{R}^n \to \mathbb{R}$:

$$\mathrm{E}[\varphi \circ X] = \sum_{x \in \Omega} \varphi(x) f_X(x), \quad \text{or} \quad \mathrm{E}[\varphi \circ X] = \int_{\mathbb{R}^n} \varphi(x) f_X(x) dx$$

# 4 Variance and Covariance

## 4.1 Definition

**Covariance**

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

which comes from:

$$\begin{aligned}\mathrm{Var}[X+Y] &= \mathrm{E}\left[((X+Y)-\mathrm{E}[X+Y])^2\right] \\ &= \mathrm{E}\left[((X-\mathrm{E}[X])+(Y-\mathrm{E}[Y]))^2\right] \\ &= \mathrm{E}\left[(X-\mathrm{E}[X])^2+(Y-\mathrm{E}[Y])^2+2(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\right] \\ &= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])]\end{aligned}$$

**Variance**

The covariance matrix for a multivariate random variable X is defined as:

$$\mathrm{Var}[X] = \begin{pmatrix} \mathrm{Var}\left[X_1\right] & \mathrm{Cov}\left[X_1, X_2\right] & \dots & \mathrm{Cov}\left[X_1, X_n\right] \\ \mathrm{Cov}\left[X_1, X_2\right] & \mathrm{Var}\left[X_2\right] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathrm{Cov}\left[X_{n-1}, X_n\right] \\ \mathrm{Cov}\left[X_1, X_n\right] & \dots & \mathrm{Cov}\left[X_{n-1}, X_n\right] & \mathrm{Var}\left[X_n\right] \end{pmatrix}$$

## 4.2 Property

1. Cov[X,Y]=Cov[Y,X]
2. Cov[X,X]=Var[X]
3. Var[CX] = CVar[X]$C^T$, C $\in$ Mat(n $\times$ n; $\mathbb{R}$) is a constant matrix with real coefficients.

# 5 *Discussion: Covariance-Linearity-Independence

**Covariance and Independence**:

1. X and Y are independent $\to$ Cov[X, Y] = 0.
2. Cov[X, Y] = 0 $\nRightarrow$ X and Y are independent.

**Therefore, covariance is not a measure of independence.**
Then, what does covariance measure?

**Covariance and Linearity**:
**Covariance measures "linearity".** In other words, it shows how much is the relationship between X and Y is like the form "Y = a + bX".

Let's first have a mathematical taste of the relationship of covariance and linearity. (You do not need to understand exactly clear now.)
We wish to find a best estimation of the linear relationship between variables X and Y, such that the errors between a + bX and Y are minimized. Also notice hopefully we wish $b \neq 0$

to satisfy this is a ”linear relationship”. The method we choose is to find such a pair of (a, b) making $e = E[(Y - (a + bX))^2]$ minimized. Since e is a function of a and b:

$$e = E[(Y - (a + bX))^2] = E(Y^2) + b^2 E(X^2) + a^2 - 2bE(XY) + 2abE(X) - 2aE(Y)$$

To make e minimized, we simply make

$$\begin{cases} \frac{\partial e}{\partial a} = 2a + 2bE(X) - 2E(Y) = 0 \\ \frac{\partial e}{\partial b} = 2bE\left(X^2\right) - 2E(XY) + 2aE(X) = 0 \end{cases}$$

Solve the above equations, we obtain:

$$\begin{cases} b = \frac{Cov[X,Y]}{Var[X]} \\ a = E(Y) - bE[X] \end{cases}$$

Now we notice, when Cov[X, Y] = 0, b = 0; showing X and Y are unlikely to have a linear relationship.

Just relax, let’s see a concrete example.

**Question: Covariance, Linearity, and Independence**

X is a continuous random variable with:

$$f_X(x) = \begin{cases} \frac{1}{2}, & -1 < x < 1 \\ 0, & otherwise \end{cases}$$

Let $Y = X^2$. Then:

1. What is Cov[X, Y]?
2. Are X and Y independent?
3. Are X and Y linearly related?

**Answer: Covariance, Linearity, and Independence**

By definition, $Cov[X, Y] = E[XY] - E[X]E[Y] = E[X^3] - E[X]E[X^2]$
And recall, $\mathrm{E}[X^k] := \int_{-\infty}^{\infty} x^k \cdot f_X(x)$
Then it's clear $E[X] = E[X^3] = 0$. Hence:

1. Cov[X, Y] = 0
2. X and Y are dependent
3. X and Y are not linearly related

# 6 Pearson Correlation Coefficient

**Definition**

$$\rho_{XY} = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}$$

This idea comes from "the covariance of the standardized X and Y". First standardize X and Y to $\widetilde{X}$ and $\widetilde{Y}$, and Then calculate $Cov[\widetilde{X}, \widetilde{Y}]$.

$$\begin{aligned} \mathrm{Cov}[\widetilde{X}, \widetilde{Y}] &= \mathrm{E}[\widetilde{X}\widetilde{Y}] - \mathrm{E}[\widetilde{X}]\mathrm{E}[\widetilde{Y}] \\ &= \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} \end{aligned}$$

So you can interpret $\rho_{XY}$ as the covariance of the standardized variables, and it is a measure of linearity.

**Property**

1. $-1 \leqslant \rho_{XY} \leqslant 1$
2. $|\rho_{XY}| = 1$ if and only if there exist numbers $\beta_0, \beta_1 \in \mathbb{R}$, $\beta_1 \neq 0$, such that almost surely $Y = \beta_0 + \beta_1 X$.
3. Further $\rho_{XY} = 1$ leads to almost surely $\widetilde{X} - \widetilde{Y} = 0$, and $\rho_{XY} = -1$ leads to almost surely $\widetilde{X} + \widetilde{Y} = 0$.

## 6.1 Fisher Transformation

$Var[\widetilde{X} - \widetilde{Y}] = 2 - 2\rho_{XY}$, and $Var[\widetilde{X} + \widetilde{Y}] = 2 + 2\rho_{XY}$. The Fisher transformation is defined as:

$$\ln(\sqrt{\frac{\mathrm{Var}[\widetilde{X} + \widetilde{Y}]}{\mathrm{Var}[\widetilde{X} - \widetilde{Y}]}}) = \frac{1}{2} \ln\left(\frac{1 + \rho_{XY}}{1 - \rho_{XY}}\right) = \mathrm{Artanh}\,(\rho_{XY}) \in \mathbb{R}$$

Then we also have:

$$\rho_{XY} = \tanh\left(\ln\left(\frac{\sigma_{\widetilde{X}+\widetilde{Y}}}{\sigma_{\widetilde{X}-\widetilde{Y}}}\right)\right)$$

## 6.2 $\rho_{XY}$ and Linearity

1. The closer $|\rho_{XY}|$ is to 1, the more likely X and Y have a linear relationship.

2. **Positive correlation:** $\rho_{XY} > 0$, $Var[\widetilde{X} - \widetilde{Y}] < Var[\widetilde{X} + \widetilde{Y}]$. X and Y tend to be more like $\widetilde{X} = \widetilde{Y}$, in other words, more positively related. When X is larger, then Y is likely to be larger.

3. **Negative correlation:** $\rho_{XY} < 0$, X and Y tend to be more like $\widetilde{X} = -\widetilde{Y}$. When X is larger, then Y is likely to be smaller.

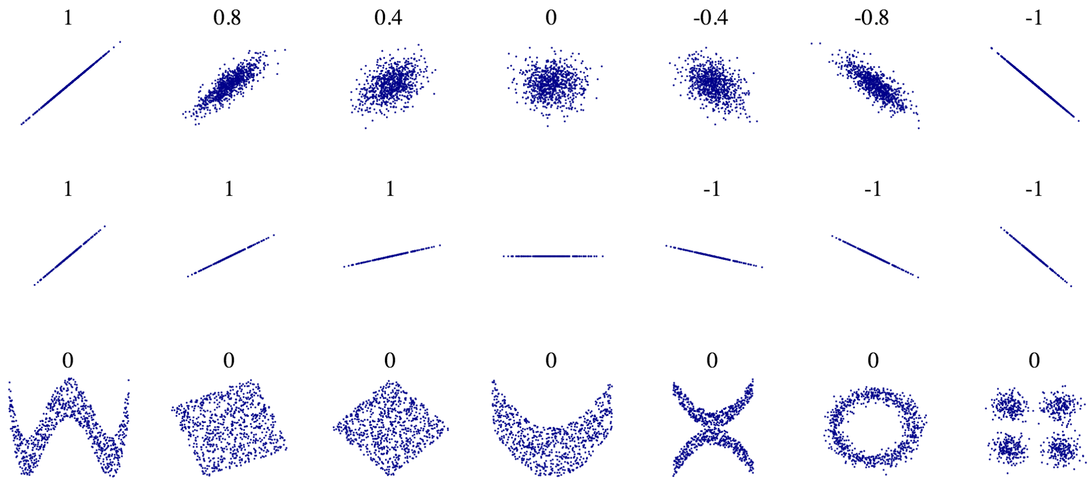


Figure 2: The Pearson Correlation Coefficient

## 6.3 Bivariate Normal Distribution

X and Y should each follow a normal distribution, but may be not independent. Then the joint density function is:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}} e^{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\varrho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

**Property:**

1. $-1 < \rho < 1$, $\rho = \rho_{XY}$

2. $\mu_X = \mathrm{E}[X]$, $\delta_X^2 = Var[X]$. Same for Y.

3. $\rho = 0 \iff$ X and Y are independent. (Notice this is special.)

# 7 Hypergeometric Distribution

**Interpretation:**

A total of N balls, r red and N − r black. Draw n balls out without putting back. Assume r > n and N − r > n. The random variable X describes the number of red balls in the n drawn balls.

**Comment:**
It is a sequence of identical but not independent Bernoulli trials. Each draw is a Bernoulli trial with $p = \frac{r}{N}$.

**Features:**

1. N, r, n are the parameters
2. $f_X(x) = \dfrac{\begin{pmatrix} r \\ x \end{pmatrix}\begin{pmatrix} N-r \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$
3. E[X]=$n\frac{r}{N}$
4. $\text{Var}\, X = n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}$

**Approximate Hypergeometric with Binomial:**
Recall for a Binomial distribution, $E[X] = np$, $\text{Var}\, X = npq$.
If the sampling fraction $\frac{n}{N}$ is small (less than 0.05), we can use $E[X] = n\frac{r}{N} = p$, $n\frac{N-r}{N} = q$ to estimate the Hypergeometric distribution with a Binomial distribution.
We can interpret as: when the sample size n is small, it does not affect the left balls much. So, the trials are approximately independent.

**Question: Covariance and Hypergeometric**

There are totally N balls, r red and N-r black. We draw 2 balls out one by one without putting back. And we define three Bernoulli random variables representing the below events:
$X_1$: Red ball at the first draw.
$X_2$: Red ball at the second draw.
$X_3$: Black ball at the second draw.
Then, what is $\text{Cov}[X_1, X_2]$, and what is $\text{Cov}[X_1, X_3]$?
And, what is $\rho_{X_1X_2}$, and what is $\rho_{X_1X_3}$?

**Answer: Covariance and Hypergeometric**

We have calculated Cov$[X_1, X_2]$ in class.
Cov$[X_1, X_2]$ = E$[X_1X_2]$ − E$[X_1]$E$[X_2]$ = - $\frac{1}{N}\frac{r(N-r)}{N(N-1)} < 0$.
Similarly since $X_1X_3$ is still a Bernoulli trial, hence E$[X_1X_3]$ = P$[X_1$ =1 and $X_3$ =1$]$ = $\frac{r}{N}\frac{N-r}{N-1}$.
Then Cov$[X_1, X_3]$ = E$[X_1X_2]$ − E$[X_1]$E$[X_2]$ = $\frac{1}{N}\frac{r(N-r)}{N(N-1)} > 0$.
Further with $Var[X_1] = Var[X_2] = \frac{r}{N}(1 - \frac{r}{N})$ and $Var[X_3] = \frac{N-r}{N}(1 - \frac{N-r}{N})$. We obtain:
$\rho_{X_1X_2} = -\frac{1}{N-1} < 0$, negative correlated.
$\rho_{X_1X_3} = \frac{1}{N-1} > 0$, positive correlated.
This also help you understand the positive and negative $\rho_{XY}$ better.

1. When $X_1 = 1$, $X_2$ is more likely to be 0, and $X_3$ is more likely to be 1.
2. When N becomes larger, $|\rho_{X_1X_2}|$ and $|\rho_{X_1X_3}|$ become smaller. So the effect of $X_1$ on $X_2$ and $X_3$ becomes less. When N is large enough, $\rho$ is close to 0, and we estimate the two trials as independent.