

**STATISTICS CONCEPTS USED IN THE DOMAIN OF  
DATA SCIENCE AND MACHINE LEARNING**

XAVIER TANG

2020

# Contents

<b>1</b>	<b>Distribution</b>	<b>1</b>
1.1	Bernoulli Trials	1
1.2	Discrete variable: Binomial Distribution	1
1.3	Discrete variable: Poisson Distribution	1
1.4	Continuous variable: Normal/Gaussian Distribution	1
1.5	Continuous variable: Exponential Distribution	1
<b>2</b>	<b>Metrics</b>	<b>2</b>
2.1	ROC	2
2.2	Standard Deviation	2
2.3	Variance	2
2.4	Confusion Matrix	2
2.4.1	Recall	2
2.4.2	Precision	2
2.4.3	F1	2
2.5	P value	2
2.6	T value	2
2.7	Z value	2
2.8	Correlation/Pearson	2
2.9	Covariance	2

---

<b>3</b>	<b>Testing</b>	<b>3</b>
3.1	T test	4
3.2	Chi-square test	4
3.3	Z test	4
3.4	A/B testing	4
3.4.1	Null hypothesis	4
3.4.2	Alternative hypothesis	4
3.4.3	Type I/II error	4
3.4.4	Statistical Power	4
3.5	Test of Significance	4
3.6	Hypothesis testing (one-way and two-way)	4
3.7	ANOVA	4
3.8	ANCOVA	4
3.9	One-sample/Two-sample bootstrap hypothesis test	4
3.10	Time series: p, d, q parameters, unit root and box test	4
<b>4</b>	<b>Thoerem</b>	<b>5</b>
4.1	Central Limit Theorem	5
4.2	Law of the large number	5
4.3	Naive Bayes Algorithm	5
4.4	Bayesian Statistics/Bayes Theorem	5
4.5	Sampling Theory	5

---

<b>5</b>	<b>Distribution</b>	<b>6</b>
5.1	Confidence Interval	6
5.2	Probability Distributed Function	6
5.3	Cumulative Distributed Function	6
5.4	Conditional Probability	6
5.5	Normalisation	6
5.6	Standardisatio	6
5.7	Least-squared error	6
5.8	R-squared error	6
5.9	Mean-squared error	6
5.10	Inferential Statistics	6
5.11	Bias-variance trade off	6

# CHAPTER 1

## Distribution

- 1.1 Bernoulli Trials
- 1.2 Discrete variable: Binomial Distribution
- 1.3 Discrete variable: Poisson Distribution
- 1.4 Continuous variable: Normal/Gaussian Distribution
- 1.5 Continuous variable: Exponential Distribution

# CHAPTER 2

## Metrics

### 2.1 ROC

### 2.2 Standard Deviation

### 2.3 Variance

### 2.4 Confusion Matrix

#### 2.4.1 Recall

#### 2.4.2 Precision

#### 2.4.3 F1

### 2.5 P value

### 2.6 T value

### 2.7 Z value

### 2.8 Correlation/Pearson

### 2.9 Covariance

# CHAPTER 3

## Testing

### 3.1 T test

### 3.2 Chi-square test

### 3.3 Z test

### 3.4 A/B testing

#### 3.4.1 Null hypothesis

#### 3.4.2 Alternative hypothesis

#### 3.4.3 Type I/II error

#### 3.4.4 Statistical Power

### 3.5 Test of Significance

### 3.6 Hypothesis testing (one-way and two-way)

### 3.7 ANOVA

### 3.8 ANCOVA

### 3.9 One-sample/Two-sample bootstrap hypothesis test

### 3.10 Time series: p, d, q parameters, unit root and box test



# CHAPTER 4

## Thorem

4.1 Central Limit Theorem

4.2 Law of the large number

4.3 Naive Bayes Algorithm

4.4 Bayesian Statistics/Bayes Theorem

4.5 Sampling Theory

# CHAPTER 5

## Distribution

5.1 Confidence Interval

5.2 Probability Distributed Function

5.3 Cumulative Distributed Function

5.4 Conditional Probability

5.5 Normalisation

5.6 Standardisatio

5.7 Least-squared error

5.8 R-squared error

5.9 Mean-squared error

5.10 Inferential Statistics

5.11 Bias-variance trade off