

# **MATH242 Lecture Notes**

Zhangir Azerbayev

Spring 22

# Introduction

Notes from Prof. Zhou Fan's statistics course. These notes are unofficial and all mistakes are almost certainly mine.

# Contents

<b>Introduction</b>	<b>i</b>
<b>1 Probability Review</b>	<b>1</b>
1.1 Moment Generating Function . . . . .	1
1.2 Multivariate Normal Distribution . . . . .	1
<b>2 Hypothesis Testing</b>	<b>2</b>
2.1 Sampling Distributions . . . . .	2
2.2 The Null Hypothesis . . . . .	2
2.3 Simple Hypotheses . . . . .	2
2.4 Composite Hypotheses and Pivotal Statistics . . . . .	3
2.5 One-sample $t$ -test . . . . .	3
2.6 Sign Test . . . . .	3
2.7 Testing Multiple Hypotheses . . . . .	4

# Chapter 1

## Probability Review

### 1.1 Moment Generating Function

A moment-generating function (MGF) is a clothesline on which to hang the moments of a probability distribution. The  $n$ th moment of a random variable  $X$  is  $EX^n$ . the MGF is defined as

$$M_X(t) = Ee^{tX}$$

for  $t \in \mathbb{R}$ . Notice that

$$M_X(t) = \sum_{n=1}^{\infty} \frac{t^n EX^n}{n!}.$$

**Proposition 1.1.1.** If  $X$  and  $Y$  are independent,  $M_{X+Y}(t) = M_X(t)M_Y(t)$ .

So the MGF is useful because it turns convolution into multiplication. Notice that this is like a Laplace transform. We have the following theorem that we won't prove.

**Theorem 1.1.2.** Let  $X$  and  $Y$  be two random variables such that, for some  $h > 0$  and there is a positive measure set on which  $M_X(t)$  and  $M_Y(t)$  are finite and  $M_X(t) = M_Y(t)$ . Then  $X$  and  $Y$  have the same distribution.

If  $Z \sim \mathcal{N}(0, 1)$ , then  $M_X(t) = e^{t^2/2}$ .

### 1.2 Multivariate Normal Distribution

The multivariate normal distribution is defined by a mean vector  $\mu \in \mathbb{R}^k$  and a symmetric covariance matrix  $\Sigma \in \mathbb{R}^{k \times k}$ .

**Definition 1.2.1.**  $(X_1, \dots, X_k)$  has a normal distribution if for every linear combination  $\alpha_1 X_1 + \dots + \alpha_k X_k$  is normal, and  $EX_i = \mu_i$  and  $\text{Cov}(X_i, X_j) = \Sigma_{ij}$ .

## Chapter 2

# Hypothesis Testing

### 2.1 Sampling Distributions

For data  $X_1, \dots, X_n$  a *statistic*  $T(X_1, \dots, X_n)$  is any real-valued function of the data. For example the sample mean, sample variance (remember sample variance is the unbiased estimator for variance), etc. We care about the *sampling distribution* of the statistic  $T$ . A central question of statistics is if we understand the distribution of  $X_1, \dots, X_n$ , how can we understand the sampling distribution of  $T$ ?

The *chi-square distribution with  $n$  degrees of freedom*, abbreviated  $\chi_n^2$ , is the distribution of the statistic  $X_1^2 + \dots + X_n^2$ .

For many simple statistics, the sampling distribution is difficult to describe exactly. For example sample mean of uniform or Bernoulli. In this case, we sampling methods or asymptotic methods. Sampling methods do monte carlo simulation on a computer and asymptotic methods apply the  $n = \infty$  approximation.

The *central limit theorem* is that for large  $n$ , the distribution of  $\sqrt{n} \frac{\bar{X} - \mu}{\sigma}$  approaches  $\mathcal{N}(0, 1)$ .

### 2.2 The Null Hypothesis

A *hypothesis test* is a binary question about the distribution of data. Our goal is to figure out whether the data provides sufficiently strong evidence to reject the null hypothesis  $H_0$  in favor of  $H_1$ . The hypotheses  $H_0$  and  $H_1$  are treated asymmetrically.

A test statistic  $T$  is a statistic computed from the data such that an extreme value of  $T$  provides evidence against  $H_0$  in favor of  $H_1$ . The *null distribution* is the distribution of  $T$  given that  $H_0$  is true. We divide the possible values of  $T$  into a *rejection region* for  $H_0$  and an acceptance region.

*Type I error* is the probability we wrongly reject  $H_0$ , that is  $P_{H_0}(T \text{ belongs to rejection region})$ . We choose the rejection region to ensure that the probability of Type I error is at most  $\alpha$ , we call  $\alpha$  the significance level. We call the smallest significance level at which we wish to reject the test the *p-value*.

To determine the rejection region for a test, we need to know the null distribution.

### 2.3 Simple Hypotheses

A hypothesis  $H_0$  or  $H_1$  is *simple* if it completely specifies the distribution of the data. For simple hypotheses, we can simulate the null distribution on a computer.

*Type II error* is the probability of wrongly accepting the null hypotheses, or  $\beta = P_{H_1}(\text{accept } H_0)$ . Equivalently we can speak of the *power of the test*  $1 - \beta$ , which is the probability of rejecting  $H_0$  in favor of  $H_1$  when  $H_1$  is true.

Suppose we have data  $X = (X_1, \dots, X_n)$ . Suppose  $H_0$  is that  $X$  has distribution  $f_0(x)$  and  $H_1$  is that  $X$  has distribution  $f_1(x)$ . An intuitive statistic for testing two simple hypotheses against each other is

$$L(X) = \frac{f_0(X)}{f_1(X)}.$$

This is called the *likelihood ratio statistic* and the test that rejects  $H_0$  for small values of  $L(X)$  is called the likelihood ratio test.

**Theorem 2.3.1.** Neyman-Pearson Lemma Let  $H_0$  and  $H_1$  be simple hypotheses and fix a significance level  $\alpha$ . Suppose there is a likelihood ratio test which rejects  $H_0$  when  $L(X) < c$  and has Type I error  $\alpha$ . Then for any other test with Type I error  $\alpha$  its power against  $H_1$  is at most the power of the likelihood ratio test.

## 2.4 Composite Hypotheses and Pivotal Statistics

Hypotheses that are not *simple* are *composite*. A test with a composite null hypothesis has to have  $P(\text{reject } H_0) \leq \alpha$  for all distributions specified by  $H_0$ . This makes things very difficult! We have to reason about an infinite family of distributions.

A way we often deal with this is to find a test statistic  $T$  whose sampling distribution is the same for every distribution in  $H_0$ , such a statistic is called *pivotal*. There is a similar consideration for a composite  $H_1$ . We often cannot maximize the power of our test for all possible distributions described by  $H_1$ , so we try to strike a balance.

## 2.5 One-sample $t$ -test

Suppose  $X_1, \dots, X_n$  IID from  $\mathcal{N}(\mu, \sigma^2)$  and we wish to test the hypotheses  $H_0 : \sigma = 0$  and  $H_1$  also something about  $\mu$  where  $\mu$  and  $\sigma^2$  is unknown. The following statistic, called the *one-sample  $t$ -statistic* is pivotal.

$$T = \frac{\sqrt{n}\bar{X}}{S}.$$

If  $Z \sim \mathcal{N}(0, 1)$  and  $U \sim \chi_n^2$  and  $Z$  and  $U$  are independent, then the distribution of  $Z/\sqrt{\frac{1}{n}U}$  is called the  $t$ -distribution with  $n$  degrees of freedom, denoted  $t_n$ . The logic is that the mean and the sample variance are basically independent. This is in fact the distribution of our test statistic  $T$ . So we can do one and two sided hypothesis tests.

## 2.6 Sign Test

We just looked at a parametric test, but what about a non-parametric test? Consider the hypothesis  $H_0 : f$  has median and  $H_1 : f$  has positive median. Then the  $t$ -statistic is no longer pivotal because it relies on a normal assumption. Consider instead the sign statistic

$$S = \sum_{i=1}^n \mathbf{1}\{X_i > 0\}.$$

This has a binomial distribution for the null distribution! We can also use the normal approximation of the binomial distribution.

## 2.7 Testing Multiple Hypotheses

Science is fucked. John P.A Ioannidis “Why Most Published Research Findings Are False”. If testing  $n$  null hypotheses at level  $\alpha$ , all of which are true, then on average I’ll falsely reject  $\alpha n$  of them. Suppose you’re testing if a disease is associated with 1,000,000 different genetic markers.

We will abstract tests merely as returning a  $p$ -value. A  $p$ -value is really a transformed test statistic. Under  $H_0$ , the  $p$ -value  $P$  is uniform on  $[0, 1]$ . Therefore one way to approach testing multiple hypotheses is to check whether the distribution of  $p$ -values is uniform.

The simplest multiple-testing correction is the *Bonferri method*. When testing  $n$  different null hypotheses, perform each at the significance level  $\alpha/n$  instead of  $\alpha$ . This basically sets  $P(\text{rejects any null hypothesis})$  to  $\alpha$ . Another way to think about the Bonferri method is this: suppose we test  $n$  null hypotheses  $n_0$  of which are true nulls and  $n - n_0$  of which are false nulls. The *family-wise error rate (FWER)* is the probability we reject at least one of the  $n_0$  true null hypotheses. A procedure *controls FWER at level  $\alpha$*  if it guarantees that  $\text{FWER} \leq \alpha$ . It’s obvious that the Bonferri method controls FWER at level  $\alpha$ . In fact, the Bonferri method controls FWER at level  $\alpha \frac{n_0}{n}$ , although in practice we do not know  $n_0$ . But if  $n_0$  is small the Bonferri method is quite strict!