QF5210A Financial Time Series

Chao ZHOU

Department of Mathematics, NUS
matzc@nus.edu.sg

2nd Lecture, 12 May 2019

Population and samples

The population is the total collection of all objects or people to be studied, which may be either finite (very large) or infinite.

Samples are finite observations from the given population.

For example, suppose we have observations, $X_1, ..., X_n$ from the same population, then the sample mean is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ while the population mean $E[X] = \mu$, which is typically an unknown parameter.

Statistical Inference

To estimate unknown population parameters, we use various statistics.

A statistic is a function computed from the data in a sample. In particular, a (point) estimator $\hat{\theta}$ is a statistic computed from a sample that gives a single value for the unknown population parameter θ.

A statistic is a **random variable** while a unknown parameter is a **constant**.

To do hypothesis testing, we use test statistics. A test statistic is a statistic computed from a sample that is used to conduct hypothesis testing.

Outline

Confidence interval : example

Suppose we want to estimate population mean μ from X_1, ..., X_n, which are i.i.d. samples from the population.
A natural estimator (though not necessarily the best one) will be the sample mean $\bar{X}$. In other words, $\bar{X}$ is an estimator for μ by LLN.
By the central limit theorem

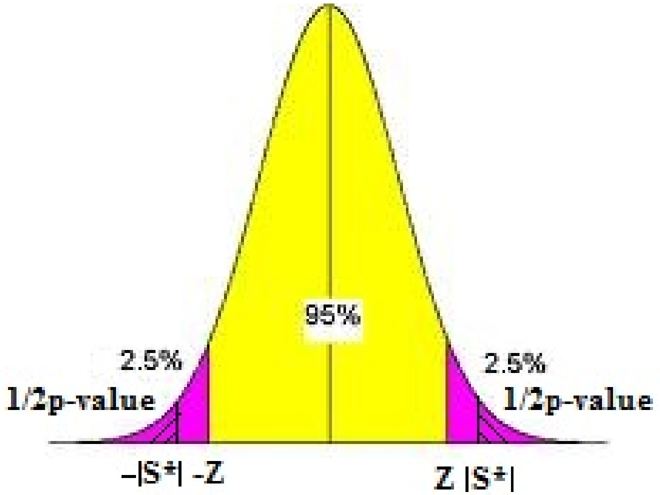
$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \rightarrow N(0,1),$$

in distribution as $n \rightarrow \infty$. In general σ is unknown, we can replace σ by s_n in the central limit theorem and get

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{s_n\sqrt{n}} \rightarrow N(0,1),$$

in distribution.

Normal distribution



Therefore, for any $\alpha \in [0, 1]$,

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i - n\mu}{s_n\sqrt{n}}\right| \leq z_{1-\alpha/2}\right) \approx 1 - \alpha.$$

Thus,

$$P\left(\left|\sqrt{n}\frac{\bar{X} - \mu}{s_n}\right| \leq z_{1-\alpha/2}\right) \approx 1 - \alpha,$$

and

$$P\left(\bar{X} - z_{\alpha/2}\frac{s_n}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\frac{s_n}{\sqrt{n}}\right) \approx 1 - \alpha.$$

The interval

$$(\bar{X} - z_{1-\alpha/2}\frac{s_n}{\sqrt{n}},\ \bar{X} + z_{1-\alpha/2}\frac{s_n}{\sqrt{n}})$$

is called the $(1-\alpha)$ asymptotic confidence interval (c.i.) for μ.
For example, the asymptotic 95% c.i. for μ is

$$(\bar{X} - 1.96\frac{s_n}{\sqrt{n}},\ \bar{X} + 1.96\frac{s_n}{\sqrt{n}}).$$

Confidence interval

In summary, a $(1-\alpha)$ confidence interval for the unknown parameter θ is a random interval, $(\hat{\theta}_L, \hat{\theta}_H)$, such that

$$P(\hat{\theta}_L < \theta < \hat{\theta}_H) \text{ is about } 1-\alpha$$

If the calculation is exact, it is called exact confidence interval ; if the coverage probabilities is asymptotically equal to $1-\alpha$, then it is called asymptotic confidence interval.

Confidence interval

In general, if the estimator $\hat{\theta}$ has an asymptotic normal distribution

$$\hat{\theta} \approx N(\theta, v(\theta)),$$

where $v(\theta)$ is the asymptotic variance of $\hat{\theta}$, then $(1-\alpha)$ asymptotic confidence interval for the unknown parameter θ is

$$(\hat{\theta} - z_{1-\alpha/2}\sqrt{v(\hat{\theta})},\ \hat{\theta} + z_{1-\alpha/2}\sqrt{v(\hat{\theta})}),$$

in which we estimate the variance $v(\theta)$ by $v(\hat{\theta})$.

Desirable Properties of Estimators 1

1. Unbiased Estimator
An estimator $\hat{\theta}$ of an unknown parameter θ is unbiased if

$$E[\hat{\theta}] = \theta.$$

The bias of an estimator $\hat{\theta}$ is defined to be

$$\text{Bias}(\hat{\theta}) := E[\hat{\theta} - \theta]$$

Many estimators are unbiased. For example, suppose we have four i.i.d. observations X_1, X_2, X_3, X_4 and we want to estimate the unknown population mean μ. There are many possible choices, e.g. $(X_1 + X_2 + X_4)/3$, $(X_2 + X_3)/2$, $(X_1 + X_2 + X_3 + X_4)/4$, all of which are unbiased. Which one is the best ?

Suppose we have two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$. We say $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2).$$

Coming back to the example, clearly, among the three estimators $(X_1 + X_2 + X_4)/3$, $(X_2 + X_3)/2$, $(X_1 + X_2 + X_3 + X_4)/4$, the last one (the sample average) has the smallest variance (Why ?), and therefore is the most efficient.

Desirable Properties of Estimators 2

2. Small Mean Squared Error

The mean squared error of an estimator $\hat{\theta}$ is

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

Note that if θ is a scalar, then it can be shown (Why ?)that

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + (\text{Bias}(\hat{\theta}))^2.$$

The definition of the mean squared error attempts to provide a comparison between various estimators, while considering the trade-off between unbiasedness and variance.

Remark : Sometimes an unbiased estimator may have a larger mean squared error than a biased estimator.

Mean Squared Error : Example

Example. Assume that we have i.i.d. samples from the normal distribution with mean μ and variance σ^2. A commonly used estimator of σ^2 is

$$s^2 := \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

which is an unbiased estimator of σ^2. However, a less commonly used estimator is

$$\hat{\sigma}^2 := \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n},$$

which has a smaller mean squared error even it is a biased estimator.

In fact, by using the Chi-squared distribution, we know that

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$$

has a χ^2-distribution with d.f. $n-1$, which has **mean** $n-1$ and **variance** $2(n-1)$. Thus,

$$\begin{aligned} E\left(\frac{s^2}{\sigma^2}\right) &= \frac{1}{n-1}E\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}\right) = 1, \\ Var\left(\frac{s^2}{\sigma^2}\right) &= \frac{1}{(n-1)^2}Var\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}\right) = \frac{2}{n-1}. \end{aligned}$$

Hence, we have

$$E(s^2) = \sigma^2, \quad MSE(s^2) = Var(s^2) = \frac{2\sigma^4}{n-1}.$$

On the other hand,

$$E(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}$$

$$Var(\hat{\sigma}^2) = Var(s^2)\left(\frac{n-1}{n}\right)^2 = \frac{2\sigma^4}{n-1}\left(\frac{n-1}{n}\right)^2$$

Therefore,

$$\begin{aligned} MSE(\hat{\sigma}^2) - MSE(s^2) &= (\frac{\sigma^2}{n})^2 + \frac{2\sigma^4}{n-1}\left(\frac{n-1}{n}\right)^2 - \frac{2\sigma^4}{n-1} \\ &= \sigma^4\left(\frac{2n-1}{n^2} - \frac{2}{n-1}\right) < 0. \end{aligned}$$

In summary, $\hat{\sigma}^2$ has a smaller MSE than that of the standard variance estimator s^2, under the normal distribution assumption.

Desirable Properties of Estimators 3

3. Consistency
An estimator $\hat{\theta}$ is called a consistent estimator of an unknown parameter θ, if $\hat{\theta} \to \theta$ in probability, as the sample size $n \to \infty$.

Consistency is a desirable property for estimators.

Convergence in probability : We say that X_n converges to X in probability, if for any $\varepsilon > 0$ we have

$$P(|X_n - X| > \varepsilon) \to 0,$$

as $n \to \infty$.

Consistency : Example

Example. The sample average $\bar{X}$ is a consistent estimator of the population mean μ, because

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \to E[X] = \mu,$$

in probability, by the law of large numbers.

The sample variance

$$s^2 := \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

is a consistent estimator of σ^2.

$$
\begin{aligned}
\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} &= \frac{\sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i \bar{X} + n(\bar{X})^2}{n-1} \\
&= \frac{\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n(\bar{X})^2}{n-1} \\
&= \frac{\sum_{i=1}^{n} X_i^2 - 2n(\bar{X})^2 + n(\bar{X})^2}{n-1} \\
&= \frac{\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2}{n-1} \\
&= \frac{n}{n-1}\frac{\sum_{i=1}^{n} X_i^2}{n} - \frac{n}{n-1}(\bar{X})^2 \\
&\to E\left(X^2\right) - (E[X])^2 = \sigma^2,
\end{aligned}
$$

in probability, again by the law of large numbers.

Estimation Methods

1. Method of Moments

An easy way (though not most efficient) way for estimation is to use the moments.

More precisely, to match the sample moments with population moments, or more generally to match functions of sample moments with the same functions of the population moments.

This is called the method of the moments, with the number of moments needed is equal to the number of unknown parameters.

An example will be given for the estimation of the parameters in AR(p) model which leads to the so-called Yule-Walker equations.

Estimation Methods

2. Maximum Likelihood Estimator (MLE)
Likelihood function $L(\theta) = f(X_1, ..., X_n; \theta)$ is simply the joint density function for $X_1, ..., X_n$ with the unknown parameter θ.

If $X_1, ..., X_n$ are independent, then clearly

$$L(\theta) = f(X_1, \theta) f(X_2, \theta) \cdots f(X_n, \theta).$$

In general the observation $X_1, ..., X_n$ may be dependent, e.g. in time series analysis. Then it may be more complicated to write down the likelihood.

The maximum likelihood estimator (MLE) $\hat{\theta}$ is given by maximizing $L(\theta)$.

Example. Consider I.I.D. $N(\mu, \sigma^2)$. We want to estimate μ and σ^2. The likelihood is given by

$$L(Y;\theta) = \prod_{i=1}^{n} \frac{1}{\sigma}\phi\left(\frac{Y_i - \mu}{\sigma}\right) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)^2\}.$$

Therefore,

$$\log L(\theta) = -\log(\sigma^n (2\pi)^{n/2}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)^2.$$

Taking derivatives with respect to μ and σ^2 and setting them to zero

$$\sum_{i=1}^{n}(Y_i - \mu) = 0, \; -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i - \mu)^2 = 0.$$

Therefore, the estimators for μ and σ are

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}, \;\; \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Note that $\hat{\sigma}^2$ is exactly the one that we just mentioned before, which is (slightly) biased.

In general, MLE will not give unbiased estimators.

But as we will see that the bias goes to zero asymptotically, as the sample size goes to infinity.

Properties of MLE

MLE is perhaps the most widely used estimation procedure.
(1) Consistency : $\hat{\theta} \to \theta$ in probability.
(2) Asymptotic normality :

$$\hat{\theta} \approx N(\theta, \mathcal{I}(\theta)^{-1}),$$

in distribution, where the Fisher information matrix $\mathcal{I}(\theta)$ is given by

$$\mathcal{I}(\theta) = -E\left\{\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta^T}\right\}.$$

We can estimate the asymptotic variance as $V_E(\hat{\theta})$ where

$$V_E(\theta) = -\left(\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta^T}\right)^{-1}.$$

In other words, we simply ignore the expectation.

In the last Example, we have

$$\log L(\theta) = -\log(\sigma^n (2\pi)^{n/2}) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \mu)^2.$$

Therefore,

$$\frac{\partial}{\partial \mu} \log L(\theta) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \mu).$$

and

$$\frac{\partial^2}{\partial \mu^2} \log L(\theta) = -\frac{\sum_{i=1}^{n} 1}{\sigma^2} = -\frac{n}{\sigma^2}$$

It follows that $\mathcal{I}(\hat{\mu}) = n/\hat{\sigma}^2$ and the asymptotic variance of $\hat{\mu}$ is $\hat{\sigma}^2/n$ (the standard error of $\hat{\mu}$ is $s_{\hat{\mu}} = \hat{\sigma}/\sqrt{n}$).

(3) It can shown the variance of the MLE is asymptotically smallest possible. More precisely, no unbiased estimators can have smaller variance than that of the MLE, at least asymptotically. For instance, $Var(\hat{\sigma}^2) = \frac{2\sigma^4}{n-1}\left(\frac{n-1}{n}\right)^2$ is smaller than $Var(s^2) = \frac{2\sigma^4}{n-1}$ for normal distribution.

(4) Invariance. The MLE of $C(\theta)$ is given by $C(\hat{\theta})$, if the function $C(\cdot)$ is continuously differentiable. For example, if we want to estimate the Sharpe ratio $(\mu - r)/\sigma$, where r is the risk-free rate, we can simply use $(\hat{\mu} - r)/\hat{\sigma}$.

Outline

Confidence Intervals and Hypothesis Testing

How to test the hypothesis $H_0 : \mu = 3$ versus the alternative hypothesis $H_a : \mu \neq 3$?

Assume we have large sample size. One possible solution is simply by using the idea of confidence interval.

More precisely, first we construct a $(1-\alpha)$ confidence interval (c.i. for short) for μ :

$$\bar{X} \pm z_{1-\alpha/2} \frac{s_n}{\sqrt{n}}.$$

Then we reject H_0 if 3 is not within the $(1-\alpha)$ c.i. ; otherwise, we cannot reject it as a plausible value for μ.

In other words, we reject H_0 if

$$3 < \bar{X} - z_{1-\alpha/2}\frac{s_n}{\sqrt{n}} \text{ or } 3 > \bar{X} + z_{1-\alpha/2}\frac{s_n}{\sqrt{n}}.$$

The above formula can be rewritten as

$$\frac{\bar{X} - 3}{s_n/\sqrt{n}} < -z_{1-\alpha/2} \text{ or } \frac{\bar{X} - 3}{s_n/\sqrt{n}} > z_{1-\alpha/2}.$$

By using such a test, we may make a wrong decision with probability α, if we reject the null hypothesis when it is indeed true. This is called type I error.

There is another possible error, which is the error we fail to reject the null hypothesis when it is false. This is called type II error.

Basic Ideas of Hypothesis Testing

Consider μ_0 as a fixed constant. (It is equal to 3 in the previous example.) The basic terminologies of hypothesis testing can be summarized in one table.

Terminology	Example
Null hypothesis H_0	$H_0 : \mu = \mu_0$
Alternative hypo. H_a	$H_a : \mu \neq \mu_0$
Test statistic	$\frac{\bar{X}-\mu_0}{s_n/\sqrt{n}}$
Rejection region	$\left\lvert \frac{\bar{X}-\mu_0}{s_n/\sqrt{n}} \right\rvert > z_{1-\alpha/2}$
Type I error	Reject H_0 when H_0 is true
Type II error	Fail to reject H_0 but H_0 is false

Remarks : (1) Type I error is also called *p*-value. (2) Statistical significance is not practical significance.

In terms of conducting hypothesis testing, we try to control the type I error first, and then minimize the type II error. In other words, we do the following

$$\begin{aligned} \text{min} : &\ \text{type II error} \\ \text{subject to} : &\ p\text{-value} \le \alpha \end{aligned},$$

where α is a fixed number called significance level. Typically $\alpha = 0.05$ and sometimes $\alpha = 0.01$.

In general, the type I error is more important.

Some Examples of Hypothesis Testing 1

1. Large Sample z-Test

The problem of z-test has three versions.

$$\begin{aligned} (1) \quad & H_0: \mu = \mu_0, \ H_a: \mu \neq \mu_0; \\ (2) \quad & H_0: \mu \leq \mu_0, \ H_a: \mu > \mu_0; \\ (3) \quad & H_0: \mu \geq \mu_0, \ H_a: \mu < \mu_0. \end{aligned}$$

Two assumptions that we need to make are independent samples and that the sample size n must be large. The test statistic is given by

$$S^* = \frac{\bar{X} - \mu_0}{s_n/\sqrt{n}} \approx N(0, 1),$$

under H_0.

To make sure that

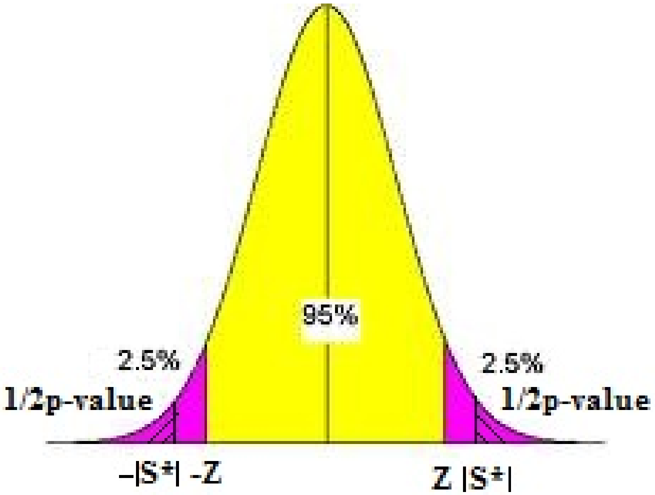
$$\text{type I error} = p - \text{value} \leq \alpha$$

we must have the rejection regions

$$(1)\ |S^*| > z_{1-\alpha/2}, \quad (2)\ S^* > z_{1-\alpha}, \quad (3)\ \ S^* < -z_{1-\alpha}$$

Remarks : (1) The sign in the rejection region is the same as that in H_a. (2) The test statistic and rejection are the same for the test $H_0 : \mu \leq \mu_0$ vs $H_a : \mu > \mu_0$ or for the test $H_0 : \mu = \mu_0$ vs $H_a : \mu > \mu_0$.

Normal distribution



Example : $n = 50$, $\bar{X} = 16$ and $s_n^2 = 9$, $H_0 : \mu \leq 15$, $H_a : \mu > 15$.
The test statistic is

$$S^* = \frac{16 - 15}{\sqrt{9}/\sqrt{50}} = 2.357.$$

Since $S^* > z_{0.95} = 1.69$, we reject H_0 at 5% level. The p-value in this example is approximately

$$P(N(0,1) > 2.357) = 0.0091.$$

If the problem changes to

$$H_0 : \mu = 15, \ H_a : \mu \neq 15,$$

then

$$p\text{-value} \approx P(N(0,1) > 2.357) + P(N(0,1) < -2.357) = 2(0.0091).$$

Some Examples of Hypothesis Testing 2

2. Small Sample t-Test

The problem has three versions

$$(1) \quad H_0 : \mu = \mu_0, \; H_a : \mu \neq \mu_0$$

$$(2) \quad H_0 : \mu \leq \mu_0, \; H_a : \mu > \mu_0$$

$$(3) \quad H_0 : \mu \geq \mu_0, \; H_a : \mu < \mu_0$$

Assume that we have independent samples and normal distribution. Then the test statistic is given by

$$T = \frac{\bar{X} - \mu_0}{s_n / \sqrt{n}} \approx t\text{-distribution}$$

with degree of freedom (d.f. for short) $n - 1$ under H_0.

To make sure that

$$p - \text{value} \le \alpha$$

we must have the rejection regions

$$(1) \quad |T| > t_{1-\alpha/2}, \quad (2) \quad T > t_{1-\alpha}, \quad (3) \quad T < -t_{1-\alpha}.$$

Example : $n = 10$, $\bar{X} = 1290$, $s_n = 110$. Test $H_0 : \mu = 1200$ versus $H_a : \mu \neq 1200$. Use $\alpha = 0.01$ and assume normality.

$$T = \frac{1290 - 1200}{110/\sqrt{10}} = 2.587$$

with d.f. 9. Since $T < t_{0.995}(9) = 3.25$, we fail to reject H_0. Notice that we use $t_{0.995}$ rather than $t_{0.99}$. If we use $\alpha = 0.05$, then $t_{0.975}(9) = 2.262$, and we reject H_0.

Question. What can we say about the p-value in this example?

Hypothesis Testing : Likelihood Ratio Method

The celebrated Neyman-Pearson lemma says that the under some mild assumptions the likelihood ratio method will give the best test statistics.
In other words, the likelihood ratio method will solve the following optimization problem

$$\begin{aligned} \text{min} :\ & \text{type II error} \\ \text{subject to} :\ & p\text{-value} \leq \alpha \end{aligned},$$

where α is a fixed number called significance level.

Basic Principle

Consider a simple setting in which the unknown parameters is given by $\theta = (\theta_1, \theta_2)$.

Hypothesis testing $H_0 : \theta_1 = \theta_{1,0}$ vs. $H_a : \theta_1 \neq \theta_{1,0}$.

The basic principle of the likelihood ratio method

(1) compute the two MLE's, the constrained MLE under H_0 and the unconstrained MLE
(2) compare the ratio of the two likelihood functions corresponding to both constrained and unconstrained MLE's.
(3) reject the null hypothesis if the likelihood from the unconstrained MLE is much larger than the likelihood from the constrained MLE.

Test statistic

More precisely, let $\hat{\theta}$ be the unconstrained MLE, and $(\theta_{1,0}, \hat{\theta}_2)$ be constrained MLE with θ_1 being fixed to be $\theta_{1,0}$. Then the likelihood ratio method rejects H_0 hypothesis if

$$2\log\left\{\frac{L(\hat{\theta})}{L(\theta_{1,0}, \hat{\theta}_2)}\right\} \geq \chi^2_{1-\alpha,k},$$

or equivalently $2\log L(\hat{\theta}) - 2\log L(\theta_{1,0}, \hat{\theta}_2) \geq \chi^2_{1-\alpha,k}$,,where $k = \dim(H_a) - \dim(H_0)$ is the dimension of θ_1 and $\chi^2_{1-\alpha,k}$ is the $(1-\alpha)$ percentile of the χ^2_k distribution.

This is because it can be shown that the test statistic (which is the left side of the above equation) has an asymptotic χ^2_k distribution.

Example. I.I.D. samples from $N(\mu, \sigma^2)$. We want to test $\mu = 0$. We can show that the constrained MLE is given by $(0, \hat{\sigma}_0^2)$, where

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - 0)^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2.$$

For the likelihood $L(\mu, \sigma^2)$, we have

$$2\log L(\mu, \theta) = -2\log((\sigma^2)^{n/2}(2\pi)^{n/2}) - \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)^2.$$

Therefore, the test statistic is

$$-2\log((\hat{\sigma}^2)^{n/2}(2\pi)^{n/2}) - \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2 + 2\log((\hat{\sigma}_0^2)^{n/2}(2\pi)^{n/2}) + \frac{1}{\hat{\sigma}_0^2}\sum_{i=1}^{n} Y_i^2$$

Since

$$\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2 = n, \quad \frac{1}{\hat{\sigma}_0^2}\sum_{i=1}^{n} Y_i^2 = n,$$

the test statistic becomes

$$\begin{aligned}
& -2\log((\hat{\sigma}^2)^{n/2}) + 2\log((\hat{\sigma}_0^2)^{n/2}) \\
= \;& n\log\left(\frac{\sum_{i=1}^n Y_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}\right) \\
= \;& n\log\left(\frac{\sum_{i=1}^n Y_i^2 - 2\sum_{i=1}^n Y_i\bar{Y} + 2\sum_{i=1}^n Y_i\bar{Y} + n(\bar{Y})^2 - n(\bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}\right) \\
= \;& n\log\left(\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2 + n(\bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}\right).
\end{aligned}$$

Thus, we reject the null hypothesis if

$$n \log \left(\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2 + n(\bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \right) \geq \chi^2_{1-\alpha,1},$$

i.e.

$$n \log \left(1 + \frac{(\bar{Y})^2}{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2} \right) = n \log \left(1 + \frac{(\bar{Y})^2}{\hat{\sigma}^2} \right) \geq \chi^2_{1-\alpha,1},$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

is the MLE of σ^2.

Remark : The above test statistic is asymptotically equivalent to the z-test. Indeed using the Taylor expansion $\log(1+x) \approx x$, for small x, we have the test statistic

$$n \log\left(1 + \frac{(\bar{Y})^2}{\hat{\sigma}^2}\right) \approx n \frac{(\bar{Y})^2}{\hat{\sigma}^2} = \left\{ \frac{\bar{Y}}{\hat{\sigma}/\sqrt{n}} \right\}^2,$$

where

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}.$$

This is exactly the z-test, as $\chi^2_{1-\alpha,1} = \left(z_{1-\alpha/2}\right)^2$. In our case, the approximation via the Taylor expansion is valid because under the null hypothesis $\mu = 0$, and we must have $\frac{(\bar{Y})^2}{\hat{\sigma}^2} \to 0$, as the sample size $n \to \infty$.

likelihood ratio test statistic

In general, suppose we want to test $H_0 : c(\theta) = q$. The likelihood ratio test statistic is given by

$$2 \log \frac{L_{H_a}}{L_{H_0}},$$

where L_{H_0} is the computed maximum likelihood with the restriction under H_0, and L_{H_a} is the maximum likelihood with the restriction under H_a.

We shall reject the null hypothesis if the likelihood ratio is too large compared to $\chi^2_{1-\alpha,k}$, where $k = \dim(H_a) - \dim(H_0)$.

Goodness-of-Fit Tests 1

Objective of goodness-of-fit tests : test whether a model fits data.

1. Empirical distribution One idea of testing distribution is to use the empirical distribution. Suppose we observe $X_1, X_2, \dots, X_n$ from a distribution F.
How to estimate the c.d.f. $F(x) = P(X \leq x)$ for a given number x?

An obvious solution is simply to count how many observed data points are below x.
More precisely, the empirical distribution $F_n(x)$ is defined to be

$$F_n(x) = \frac{1}{n}\#\{i : X_i \leq x\} = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$$

Suppose we order the data such that $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Then we can also write

$$F_n(x) = \begin{cases} 0 & x < X_{(1)} \\ \frac{k}{n} & X_{(k)} \leq x < X_{(k+1)} \\ 1 & x \geq X_{(n)} \end{cases} .$$

It can be shown that

$$E(F_n(x)) = F(x),\ \mathrm{Var}(F_n(x)) = \frac{1}{n}F(x)(1 - F(x)),$$

$$F_n(x) \to F(x),$$

in probability as $n \to \infty$. Therefore $F_n(x)$ is an unbiased estimator of $F(x)$ and it is also a consistent estimator of $F(x)$. A central limit theorem for $F_n(x)$ is also easy to derive so that we can construct asymptotic confidence intervals for $F(x)$ for every fixed x.

Goodness-of-Fit Tests 2

2. Kolmogorov-Smirnov Test
We can also construct a test that is valid for all x simultaneously.

In fact, a measure of closeness of the observed data to a given distribution can be found by using the Kolmogorov-Smirnov test statistic

$$K_n = \max_{-\infty < x < \infty} |F_n(x) - F(x)|.$$

Kolmogorov and Smirnov showed that the limiting distribution of K_n, as $n \to \infty$, exists and the limiting distribution does not depend on F ; it has a special distribution.

Therefore, we can use the limiting distribution to construct a test that is valid for all x.

Goodness-of-Fit Tests 3

3. QQ Plot

Quantile-Quantile Plot provides a qualitative graphical assessment of the goodness of fitness of a given distribution.

Example. We want to test

$$H_0 : X_1, X_2, \ldots, X_n \sim U(0,1),$$

where X_1, ..., X_n are i.i.d. samples. The basic idea is consider the order statistics $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$. If H_0 is true, then under H_0

$$E(X_{(k)}) = \frac{k}{n+1}.$$

This suggests if plotting $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ vs. $\frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1}$, we would expect a linear line. This is the key idea of the QQ plot.

More generally, suppose we want to test

$$H_0 : X_1, X_2, \ldots, X_n \text{ i.i.d} \sim F.$$

Recall the fact : if X has c.d.f. F, then $F(X) \sim U(0,1)$. Then

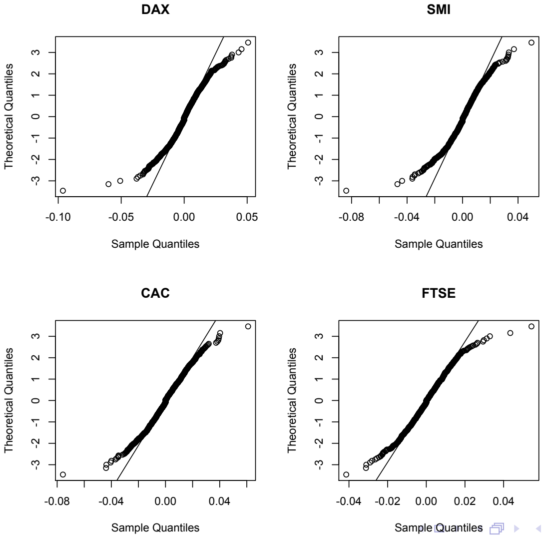
$$E\{F(X_{(k)})\} = \frac{k}{n+1}.$$

Therefore, a plot of $F(X_{(k)})$ vs. $\frac{k}{n+1}$ would display a linear pattern under H_0. Equivalently, a plot of

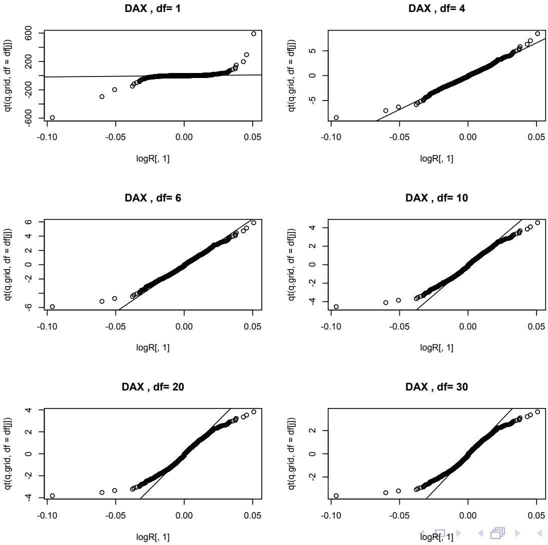
$$X_{(k)} \text{ vs. } F^{-1}(\frac{k}{n+1})$$

would display a linear pattern under H_0.
In R, *qqnorm* for normal distribution and *qqplot* more generally.

QQ plots : EU stock markets Log returns (daily)



QQ plots : DAX Log returns (daily)



Goodness-of-Fit Tests 4

4. Tests for normality Data : $\{x_1, \cdots, x_T\}$

(1) Sample mean : $\widehat{\mu}_x = \frac{1}{T}\sum_{t=1}^{T} x_t$,

(2) Sample variance : $\widehat{\sigma}_x^2 = \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \widehat{\mu}_x)^2$,

(3) Sample skewness : $\widehat{S}(x) = \frac{1}{(T-1)\widehat{\sigma}_x^3}\sum_{t=1}^{T}(x_t - \widehat{\mu}_x)^3$,

(4) Sample kurtosis : $\widehat{K}(x) = \frac{1}{(T-1)\widehat{\sigma}_x^4}\sum_{t=1}^{T}(x_t - \widehat{\mu}_x)^4$,

Under **normality** assumption, for large T,

$$\widehat{S}(x) \sim N(0, \frac{6}{T}), \quad \widehat{K}(x) - 3 \sim N(0, \frac{24}{T})$$

Tests for normality

(1) Test for symmetry : $S^* = \frac{\hat{S}(x)}{\sqrt{6/T}} \sim N(0,1)$ if normality holds.

• Decision rule : Reject H_0 of a symmetric distribution if $|S^*| > z_{1-\alpha/2}$ or p-value is less than α (p-value$=P(|X| \geq |S^*|)$ with $X \sim N(0,1)$).

(2) Test for tail thickness : $K^* = \frac{\hat{K}(x)-3}{\sqrt{24/T}} \sim N(0,1)$ if normality holds.

• Decision rule : Reject H_0 of normal tails if $|K^*| > z_{1-\alpha/2}$ or p-value is less than α.

(3) A joint test (Jarque-Bera test) : $JB = (K^*)^2 + (S^*)^2 \sim \chi^2_2$ if normality holds, where χ^2_2 denotes a chi-squared distribution with 2 degrees of freedom.

• Decision rule : Reject H_0 of normality if $JB > \chi^2_{1-\alpha,2}$ or p-value is less than α (p-value$=P(X \geq JB)$ with $X \sim \chi^2_2$).

Thank you for your attention.