Stat 710: Mathematical Statistics
Lecture 31

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Lecture 31: Kolmogorov-Smirnov tests and asymptotic tests

Kolmogorov-Smirnov tests

Let $X_1,...,X_n$ be i.i.d. random variables from a continuous c.d.f. F.
Consider

$$H_0 : F = F_0 \quad \text{versus} \quad H_1 : F \neq F_0$$

with a fixed F_0.
Let F_n be the empirical c.d.f. and

$$D_n(F) = \sup_{x \in \mathscr{R}} |F_n(x) - F(x)|,$$

which is in fact the distance $\rho_\infty(F_n, F)$.
Intuitively, $D_n(F_0)$ should be small if H_0 is true.
From the results in §5.1.1, we know that $D_n(F_0) \to_{a.s.} 0$ iff H_0 is true.
The statistic $D_n(F_0)$ is called the Kolmogorov-Smirnov statistic.
Tests with rejection region $D_n(F_0) > c$ are called Kolmogorov-Smirnov tests.

Kolmogorov-Smirnov tests

In some cases we would like to test "one-sided" hypotheses

$$H_0 : F = F_0 \quad \text{versus} \quad H_1 : F \geq F_0,\ F \neq F_0,$$

or

$$H_0 : F = F_0 \quad \text{versus} \quad H_1 : F \leq F_0,\ F \neq F_0.$$

The corresponding Kolmogorov-Smirnov statistic is

$$D_n^+(F) = \sup_{x \in \mathscr{R}} [F_n(x) - F(x)]$$

or

$$D_n^-(F) = \sup_{x \in \mathscr{R}} [F(x) - F_n(x)].$$

The rejection regions of one-sided Kolmogorov-Smirnov tests are, respectively, $D_n^+(F_0) > c$ and $D_n^-(F_0) > c$.

Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics and define $X_{(0)} = -\infty$ and $X_{(n+1)} = \infty$.

Since $F_n(x) = i/n$ when $X_{(i)} \leq x < X_{(i+1)}$, $i = 0, 1, ..., n$,

Kolmogorov-Smirnov tests

In some cases we would like to test "one-sided" hypotheses

$$H_0 : F = F_0 \quad \text{versus} \quad H_1 : F \geq F_0,\ F \neq F_0,$$

or

$$H_0 : F = F_0 \quad \text{versus} \quad H_1 : F \leq F_0,\ F \neq F_0.$$

The corresponding Kolmogorov-Smirnov statistic is

$$D_n^+(F) = \sup_{x \in \mathscr{R}} [F_n(x) - F(x)]$$

or

$$D_n^-(F) = \sup_{x \in \mathscr{R}} [F(x) - F_n(x)].$$

The rejection regions of one-sided Kolmogorov-Smirnov tests are, respectively, $D_n^+(F_0) > c$ and $D_n^-(F_0) > c$.

Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics and define $X_{(0)} = -\infty$ and $X_{(n+1)} = \infty$.
Since $F_n(x) = i/n$ when $X_{(i)} \leq x < X_{(i+1)}$, $i = 0, 1, ..., n$,

Kolmogorov-Smirnov tests

$$
\begin{aligned}
D_n^+(F) &= \max_{0\le i\le n} \sup_{X_{(i)}\le x< X_{(i+1)}} \left[\frac{i}{n} - F(x)\right] \\
&= \max_{0\le i\le n} \left[\frac{i}{n} - \inf_{X_{(i)}\le x< X_{(i+1)}} F(x)\right] \\
&= \max_{0\le i\le n} \left[\frac{i}{n} - F(X_{(i)})\right].
\end{aligned}
$$

When F is continuous, $F(X_{(i)})$ is the ith order statistic of a sample of size n from the uniform distribution $U(0,1)$ irrespective of what F is.
The distribution of $D_n^+(F)$ does not depend on F, if F is continuous.
The distribution of $D_n^-(F)$ is the same as that of $D_n^+(F)$ (exercise).
Since

$$D_n(F) = \max\{D_n^+(F), D_n^-(F)\},$$

the distribution of $D_n(F)$ does not depend on F.
This means that the distributions of Kolmogorov-Smirnov statistics are known under H_0 if F is continuous.

Theorem 6.10 (The distributions of D_n, D_n^+, and D_n^-)

Assume that F is continuous.

(i) For any fixed n,

$$P\left(D_n^+(F) \le t\right) = \begin{cases} 0 & t \le 0 \\ n! \prod_{i=1}^{n} \int_{\max\left\{0, \frac{n-i+1}{n} - t\right\}}^{u_{n-i+2}} du_1 \cdots du_n & 0 < t < 1 \\ 1 & t \ge 1 \end{cases}$$

and

$$P\left(D_n(F) \le t\right) = \begin{cases} 0 & t \le \frac{1}{2n} \\ n! \prod_{i=1}^{n} \int_{\max\left\{0, \frac{n-i+1}{n} - t\right\}}^{\min\left\{u_{n-i+2}, \frac{n-i}{n} + t\right\}} du_1 \cdots du_n & \frac{1}{2n} < t < 1 \\ 1 & t \ge 1, \end{cases}$$

where $u_{n+1} = 1$.

(ii) For $t > 0$,

$$\lim_{n\to\infty} P\left(\sqrt{n} D_n^+(F) \le t\right) = 1 - e^{-2t^2}$$

and

$$\lim_{n\to\infty} P\left(\sqrt{n} D_n(F) \le t\right) = 1 - 2\sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2t^2}.$$

Remarks

- When n is not large, Kolmogorov-Smirnov tests of size α can be obtained using the results in Theorem 6.10(i).
- When n is large, using the results in Theorem 6.10(i) is not convenient.
  We can obtain Kolmogorov-Smirnov tests of limiting size α using the results in Theorem 6.10(ii).
- It is worthwhile to compare the goodness of fit test introduced in Example 6.23 with the Kolmogorov-Smirnov test.
  - The former requires a partition of the range of observations and may lose information through partitioning, whereas the latter requires that F be continuous and univariate.
  - The latter is of size α (or limiting size α), whereas the former is only of asymptotic significance level α.
  - The former can be modified to allow estimation of unknown parameters under H_0, whereas the latter does not have this flexibility.

Asymptotic tests (tests with asymptotic significance level α)

A simple method of constructing asymptotic tests (for almost all problems, parametric or nonparametric) for testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0,$$

where θ is a vector of parameters, when an asymptotically normally distributed estimator of θ can be found.
However, this simple method may not provide the best or even nearly best solution to the problem, especially when there are different asymptotically normally distributed estimators of θ.

Let $\widehat{\theta}_n$ be an estimator of θ based on a sample of size n from P.
Suppose that under H_0,

$$V_n^{-1/2}(\widehat{\theta}_n - \theta) \to_d N_k(0, I_k),$$

where V_n is the asymptotic covariance matrix of $\widehat{\theta}_n$.
If V_n is known when $\theta = \theta_0$, then we define a test with rejection region

$$(\widehat{\theta}_n - \theta_0)^\tau V_n^{-1}(\widehat{\theta}_n - \theta_0) > \chi^2_{k,\alpha}$$

where $\chi^2_{k,\alpha}$ is the $(1-\alpha)$th quantile of the chi-squared distribution χ^2_k.

Asymptotic tests (tests with asymptotic significance level α)

A simple method of constructing asymptotic tests (for almost all problems, parametric or nonparametric) for testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0,$$

where θ is a vector of parameters, when an asymptotically normally distributed estimator of θ can be found.
However, this simple method may not provide the best or even nearly best solution to the problem, especially when there are different asymptotically normally distributed estimators of θ.

Let $\widehat{\theta}_n$ be an estimator of θ based on a sample of size n from P.
Suppose that under H_0,

$$V_n^{-1/2}(\widehat{\theta}_n - \theta) \to_d N_k(0, I_k),$$

where V_n is the asymptotic covariance matrix of $\widehat{\theta}_n$.
If V_n is known when $\theta = \theta_0$, then we define a test with rejection region

$$(\widehat{\theta}_n - \theta_0)^\tau V_n^{-1}(\widehat{\theta}_n - \theta_0) > \chi^2_{k,\alpha}$$

where $\chi^2_{k,\alpha}$ is the $(1-\alpha)$th quantile of the chi-squared distribution χ^2_k.

Asymptotic tests (tests with asymptotic significance level α)

This test has asymptotic significance level α.
If the distribution of $\widehat{\theta}_n$ does not depend on the unknown population P under H_0, then this test has limiting size α.
If V_n depends on the unknown population P even if H_0 is true ($\theta = \theta_0$), then we have to replace V_n by an estimator $\widehat{V}_n$.
If, under H_0, $\widehat{V}_n$ is consistent in the sense $\widehat{V}_n V_n^{-1} \to_p I$ (Definition 5.4) then the test having the rejection region

$$(\widehat{\theta}_n - \theta_0)^\tau \widehat{V}_n^{-1} (\widehat{\theta}_n - \theta_0) > \chi^2_{k,\alpha}$$

has asymptotic significance level α.
Variance estimation methods introduced in §5.5 can be used to construct a consistent estimator $\widehat{V}_n$.
The following result shows that, under some additional conditions, the previously defined test is asymptotically correct (§2.5.3), i.e., it is a consistent asymptotic test (Definition 2.13).

Theorem 6.12

Assume that

$$V_n^{-1/2}(\widehat{\theta}_n - \theta) \to_d N_k(0, I_k),$$

holds for any P.

Assume also that $\lambda_+[V_n] \to 0$, where $\lambda_+[V_n]$ is the largest eigenvalue of V_n.

(i) The test having rejection region

$$(\widehat{\theta}_n - \theta_0)^\tau V_n^{-1}(\widehat{\theta}_n - \theta_0) > \chi^2_{k,\alpha}$$

with a known V_n (or with V_n replaced by a consistent estimator $\widehat{V}_n$) is consistent.

(ii) If we choose $\alpha = \alpha_n \to 0$ as $n \to \infty$ and $\chi^2_{k,1-\alpha_n}\lambda_+[V_n] = o(1)$, then the test in (i) is Chernoff-consistent.

Proof

We only prove (i) for the case where V_n is known.

Let $Z_n = V_n^{-1/2}(\widehat{\theta}_n - \theta)$ and $l_n = V_n^{-1/2}(\theta - \theta_0)$.

Then $\|Z_n\| = O_p(1)$ and $\|l_n\| = \|V_n^{-1/2}(\theta - \theta_0)\| \to \infty$ when $\theta \neq \theta_0$.

Theorem 6.12

Assume that

$$V_n^{-1/2}(\widehat{\theta}_n - \theta) \to_d N_k(0, I_k),$$

holds for any P.

Assume also that $\lambda_+[V_n] \to 0$, where $\lambda_+[V_n]$ is the largest eigenvalue of V_n.

(i) The test having rejection region

$$(\widehat{\theta}_n - \theta_0)^\tau V_n^{-1}(\widehat{\theta}_n - \theta_0) > \chi^2_{k,\alpha}$$

with a known V_n (or with V_n replaced by a consistent estimator $\widehat{V}_n$) is consistent.

(ii) If we choose $\alpha = \alpha_n \to 0$ as $n \to \infty$ and $\chi^2_{k,1-\alpha_n}\lambda_+[V_n] = o(1)$, then the test in (i) is Chernoff-consistent.

Proof

We only prove (i) for the case where V_n is known.

Let $Z_n = V_n^{-1/2}(\widehat{\theta}_n - \theta)$ and $l_n = V_n^{-1/2}(\theta - \theta_0)$.

Then $\|Z_n\| = O_p(1)$ and $\|l_n\| = \|V_n^{-1/2}(\theta - \theta_0)\| \to \infty$ when $\theta \neq \theta_0$.

Proof (continued)

The result follows from the fact that when $\theta \neq \theta_0$,

$$\begin{aligned}(\widehat{\theta}_n - \theta_0)^\tau V_n^{-1}(\widehat{\theta}_n - \theta_0) &= \|Z_n\|^2 + \|I_n\|^2 + 2I_n^\tau Z_n \\ &\geq \|Z_n\|^2 + \|I_n\|^2 - 2\|I_n\|\|Z_n\| \\ &= O_p(1) + \|I_n\|^2[1 - o_p(1)]\end{aligned}$$

and, therefore,

$$P\left((\widehat{\theta}_n - \theta_0)^\tau V_n^{-1}(\widehat{\theta}_n - \theta_0) > \chi^2_{k,\alpha}\right) \to 1.$$

Example 6.27

Let $X_1, ..., X_n$ be i.i.d. random variables from a symmetric c.d.f. F having finite variance and positive F'.

Consider the problem of testing $H_0 : F$ is symmetric about 0 versus $H_1 : F$ is not symmetric about 0.

Under H_0, there are many estimators that are asymptotically normal.

Proof (continued)

The result follows from the fact that when $\theta \neq \theta_0$,

$$\begin{aligned}(\widehat{\theta}_n - \theta_0)^\tau V_n^{-1}(\widehat{\theta}_n - \theta_0) &= \|Z_n\|^2 + \|l_n\|^2 + 2l_n^\tau Z_n \\ &\geq \|Z_n\|^2 + \|l_n\|^2 - 2\|l_n\|\|Z_n\| \\ &= O_p(1) + \|l_n\|^2[1 - o_p(1)]\end{aligned}$$

and, therefore,

$$P\left((\widehat{\theta}_n - \theta_0)^\tau V_n^{-1}(\widehat{\theta}_n - \theta_0) > \chi^2_{k,\alpha}\right) \to 1.$$

Example 6.27

Let $X_1,...,X_n$ be i.i.d. random variables from a symmetric c.d.f. F having finite variance and positive F'.
Consider the problem of testing $H_0 : F$ is symmetric about 0 versus $H_1 : F$ is not symmetric about 0.
Under H_0, there are many estimators that are asymptotically normal.

Example 6.27 (continued)

We consider the following three estimators:
(1) $\widehat{\theta}_n = \bar{X}$ and $\theta = E(X_1)$;
(2) $\widehat{\theta}_n = \widehat{\theta}_{0.5}$ (the sample median) and $\theta = F^{-1}(\frac{1}{2})$ (the median of F);
(3) $\widehat{\theta}_n = \bar{X}_a$ (the a-trimmed sample mean) and $\theta = \int xJ(F(x))dF(x)$ with $J(t) = (1-2a)^{-1}I_{(a,1-a)}(t)$, $a \in (0, \frac{1}{2})$.
Although the θ's in (1)-(3) are different in general, in all cases $\theta = 0$ is equivalent to that H_0 holds.

For $\bar{X}$, it follows from the CLT that

$$V_n^{-1/2}(\bar{X} - \theta) \to_d N(0,1)$$

with $V_n = \sigma^2/n$ for any F, where $\sigma^2 = \mathrm{Var}(X_1)$.
From the SLLN, S^2/n is a consistent estimator of V_n for any F.
Thus, Theorem 6.12 applies with $\widehat{\theta}_n = \bar{X}$ and V_n replaced by S^2/n.
This test is asymptotically equivalent to the one-sample t-test derived in §6.2.3.

Example 6.27 (continued)

We consider the following three estimators:
(1) $\widehat{\theta}_n = \bar{X}$ and $\theta = E(X_1)$;
(2) $\widehat{\theta}_n = \widehat{\theta}_{0.5}$ (the sample median) and $\theta = F^{-1}(\frac{1}{2})$ (the median of F);
(3) $\widehat{\theta}_n = \bar{X}_a$ (the a-trimmed sample mean) and $\theta = \int xJ(F(x))dF(x)$ with $J(t) = (1-2a)^{-1}I_{(a,1-a)}(t)$, $a \in (0, \frac{1}{2})$.
Although the θ's in (1)-(3) are different in general, in all cases $\theta = 0$ is equivalent to that H_0 holds.

For $\bar{X}$, it follows from the CLT that

$$V_n^{-1/2}(\bar{X} - \theta) \to_d N(0,1)$$

with $V_n = \sigma^2/n$ for any F, where $\sigma^2 = \mathrm{Var}(X_1)$.
From the SLLN, S^2/n is a consistent estimator of V_n for any F.
Thus, Theorem 6.12 applies with $\widehat{\theta}_n = \bar{X}$ and V_n replaced by S^2/n.
This test is asymptotically equivalent to the one-sample t-test derived in §6.2.3.

Example 6.27 (continued)

From Theorem 5.10, $\widehat{\theta}_{0.5}$ satisfies

$$V_n^{-1/2}(\widehat{\theta} - \theta) \to_d N(0,1)$$

with $V_n = 4^{-1}[F'(\theta)]^{-2}n^{-1}$ for any F.
A consistent estimator of V_n can be obtained using the bootstrap method considered in §5.5.3.
Another consistent estimator of V_n can be obtained using Woodruff's interval introduced in §7.4 (see Exercise 86 in §7.6).
Thus, Theorem 6.12 applies with $\widehat{\theta}_n = \widehat{\theta}_{0.5}$ and V_n replaced by a consistent estimator.

It follows from the discussion in §5.3.2 that $\bar{X}_a$ satisfies

$$V_n^{-1/2}(\bar{X}_a - \theta) \to_d N(0,1)$$

A consistent estimator of V_n can be obtained using the formula for σ_a^2.
Thus, Theorem 6.12 applies with $\widehat{\theta}_n = \bar{X}_a$ and V_n replaced by a consistent estimator is asymptotically correct.

Example 6.27 (continued)

From Theorem 5.10, $\widehat{\theta}_{0.5}$ satisfies

$$V_n^{-1/2}(\widehat{\theta} - \theta) \to_d N(0,1)$$

with $V_n = 4^{-1}[F'(\theta)]^{-2}n^{-1}$ for any F.
A consistent estimator of V_n can be obtained using the bootstrap method considered in §5.5.3.
Another consistent estimator of V_n can be obtained using Woodruff's interval introduced in §7.4 (see Exercise 86 in §7.6).
Thus, Theorem 6.12 applies with $\widehat{\theta}_n = \widehat{\theta}_{0.5}$ and V_n replaced by a consistent estimator.

It follows from the discussion in §5.3.2 that $\bar{X}_a$ satisfies

$$V_n^{-1/2}(\bar{X}_a - \theta) \to_d N(0,1)$$

A consistent estimator of V_n can be obtained using the formula for σ_a^2.
Thus, Theorem 6.12 applies with $\widehat{\theta}_n = \bar{X}_a$ and V_n replaced by a consistent estimator is asymptotically correct.

Example 6.27 (continued)

It is not clear which one of the tests discussed here is to be preferred in general.

The results for $\widehat{\theta}_n$ in (1)-(3) still hold for testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ without the assumption that F is symmetric.

An example of asymptotic tests for one-sided hypotheses is given in Exercise 123.

Example 6.27 (continued)

It is not clear which one of the tests discussed here is to be preferred in general.

The results for $\widehat{\theta}_n$ in (1)-(3) still hold for testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ without the assumption that F is symmetric.

An example of asymptotic tests for one-sided hypotheses is given in Exercise 123.