Hypothesis testing

Timo Tiihonen

2016

Estimates

Assume we have a random variable x and let $F(x)$ be some property of interest of the variable x.
Now, given a sample $X_1, \ldots, X_n$ we need form two types of estimates for $F(x)$.

- Point estimate: an estimate $A = A(X_1, \ldots, X_n)$ that should estimate $E(F(x))$.
- Interval estimate: two values for which it holds $A_1(X_1, \ldots, X_n) < E(F(x)) < A_2(X_1, \ldots, X_n)$ with given high probability.

We say that point estimate A is unbiased if $E(A) = E(F(x))$.
The point estimate A is consistent if for any $\epsilon > 0$ and $\delta > 0$ we can find n such that $P(|A - E(x)| > \delta) < \epsilon$.

Hypothesis testing

Assume we have a sample $X_1, \ldots, X_n$ and we want to study if this sample represents a random variate x which has some property of interest $E(F(x)) = 0$.
Example: if the sample should be from distribution for which $E(x) = a$ we can study the property $F(x) = x - a$.
Now, given a sample $X_1, \ldots, X_n$ can we infer if it behaves as the conjectured random sequence or can we/must we argue from our sample that $E(F(X)) <> 0$.
Each sample is random. How can we avoid making wrong consequences?

Hypothesis testing

In hypothesis testing we make two hypotheses

- H_0, zero hypothesis: The sampled system behaves as expected and only random fluctuations are observed. (here: the sample X is drawn from x and $E(F(X)) = 0$).
- H_1, hypothesis to be proved: The sampled system has the non trivial property to be shown. ($E(F(X)) <> 0$).

H_0 is accepted always when it is a possible interpretation of the observed simulation results.
H_1 is accepted only in the case, when H_0 would be very improbable given the observed results.

Hypothesis testing - confidence interval

Let x be a random variate, take sample of n values $(X_1, \ldots, X_n)$ with sample average $\bar{X}$. Using this sample we want to make statements of the expectation of x.
For hypothesis testing we have to define two values $a_1(X) < a_2(X)$ such that

$$P(a_1(X) < E(x) < a_2(X)) > 1 - \beta$$

for given confidence level β. This interval is called the confidence interval and its length depends on β, on the probability distribution of x and on n.

Hypothesis testing - confidence interval

Consider the normalized error of the sample average

$$\hat{z}(X) = \frac{\bar{X} - E(x)}{\sigma(x)} n^{1/2}$$

where $\sigma(x)$ is the standard deviation of x. If the distribution of $\bar{X}$ is known, we can compute values z_1 and z_2 such that $P(z_1 < \hat{z} < z_2) = 1 - \beta$ for chosen β. In practice $\sigma(x)$ is often not known and must be approximated.

Hypothesis testing - confidence interval

If X_i:s are independent $\sigma(x)$ can be approximated by sample standard deviation.

$$\sigma^2 \approx s^2(X) = \sum (X_i - \bar{X})^2/(n-1).$$

This leads us to test variable $z = \frac{\bar{X}-E(x)}{s(X)} n^{1/2}$. If x obeys the normal distribution, z obeys t-distribution.
For given β we can define z_1 ja z_2 such that

$$P(\bar{X} - (z_1 s/n^{1/2}) < E(x) < \bar{X} + (z_2 s/n^{1/2})) = 1 - \beta$$

This gives us an interval estimate for $E(x)$ (with confidence level $1-\beta$). The interval gets shorter when n increases and longer if β decreases.

Hypothesis testing - confidence interval

If X_i:s are dependent, the autocorrelation has to be accounted for.

$$\sigma^2 \approx s^2 = \sum (X_i - \bar{X})^2/(n-1) + 2/(n-1) \sum_i \sum_k cov(X_i, X_{i+k})$$

or

$$\sigma^2 \approx s^2 = \sum (X_i - \bar{X})^2/(n-1) + 2 \sum_{k=1}^{\infty} \rho_k$$

where $\rho_k = E(cov(X_i, X_{i+k}))$ are the autocorrelations (at equilibrium). For positively autocorrelated samples the sample standard deviation alone predicts too small variability and confidence interval.

Hypothesis testing

There are two possible types for wrong conclusions

- Type I: we accept H_1 even if it is not true (probability $< \beta$).
- Type II: we accept H_0, but H_1 would be the right conclusion (very probable if we have done only few samples, require high confidence or if the true value is close to treshold).

Type II error means that we can not make the right conclusion because the simulation result is not reliable enough.

χ^2 test

Many hypotheses to be tested can be formulated as: H_0 - the observation $O = O(X)$ is a sample from distribution f. To test this we may use the Pearson χ^2-test:
Divide the range of O to N classes, compute the expected frequences (E_i) to each class (for n observations) and compute the statistics

$$\chi^2 = \sum_{i=1}^{n} (O_i - E_i)^2/(E_i)$$

where O_i is the number of observations for class i.
One should have $E_i > 5$ for all classes for reliable test. H_0 is rejected if the test statistics is too small or too big compared to tresholds for χ^2-distribution with $N - 1$ degrees of freedom. (Low value - lack of randomness, high value - different distribution).