The Basic Idea of Hypothesis Testing and Confidence Interval

Contents

Hypothesis Testing **1**

Interval estimation: **2**

Hypothesis Testing

Basically, for a parameter θ, you want to know whether it belongs Θ. To realize this, we often create a 'judge criterion':if a statistic $T(x_1,,,x_n) \in W$, then we reject the statement that $\theta \in \Theta$.

So, 'given our statistic' T, we want to find out the W, which we call 'reject area'. But how to find it? To illustrate this, we give a concrete example.

Suppose the sample comes from a normal distribution $N(\mu, \sigma^2)$. Suppose that we already know σ^2 but μ is unknown. suppose we want to test the hypothesis that $\mu \leq \mu_0$

Since $\overline{x}$ is an unbiased estimation of the mean, A natural and straight forward way to set the judge criterion is that: if $\overline{x} > c$ (i.e., the average of the sample is too large, larger than some threshold), then we reject $\mu \leq \mu_0$

but notice! since sample is random, then$\overline{x}$ is random, and it is totally possible that$\overline{x} > c$ EVEN when $\mu \leq \mu_0$ does hold. When such a thing happens, we call that we make a mistake, since we reject the hypothesis when hypothesis actually holds. To be specific, this is called 'Type 1 Error'. In general, we want the probability of such an error to be as small as possible. Therefore, we often want to control such error to be under a level, α, that is:

$$S(\mu, c) \equiv pr(\overline{x} > c|\mu) \leq \alpha, for \quad \forall \mu \leq \mu_0$$

the expression of $pr(\overline{x} > c|\mu)$ is reasonable, since the $\overline{x}$ is a random variable whose distribution is tied to μ. So here, we want to find out how these probability looks like. Again, by creating a (increasing)function $g(\overline{x};\theta)$ such that $g(\overline{x};\theta)$'s distribution function does not rely on θ, we can rewrite this expression.

under the context of the example, we can easily create $g(\overline{x};\mu) = \frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$, and g here follows $N(0,1)$

therefore we have the following:

$$S(\mu, c) \equiv pr(\overline{x} > c|\mu) = pr(g(\overline{x}, \mu) > g(c, \mu)|\mu) = \Phi(g(c, \mu)) = \Phi(\frac{c-\mu}{\frac{\sigma}{\sqrt{n}}})$$

In which Φ is the accumulative distribution function of standard normal distribution.We want $\Phi(\frac{c-\mu}{\frac{\sigma}{\sqrt{n}}}) < \alpha$ for any $\mu \leq \mu_0$. Since this function decreases with μ, to make this condition holds ,we can make $\Phi(\frac{c-\mu_0}{\frac{\sigma}{\sqrt{n}}}) = \alpha$. Through this expression, we get the value of c.

In summary, the basic procedure of doing the the hypothesis test is:

1. Specify the null hypothesis $\theta \in \Theta$ and alternative hypothesis.$\theta \notin \Theta$
2. Find (according to your intuition) a proper statistics $T(x_1,,,,x_n)$ and specify the reject area $T \in W$
3. Now you want $S(\theta, W) \equiv pr(T(x_1,,,,x_n) \in W|\theta) \leq \alpha$ holds for all $\theta \in \Theta$.(i.e., keep the type 1 error small.) the purpose to find the W.

4. To do this, create $g(T, \theta)$ such that the distribution of g does not rely on θ. we then have

$$pr(T(x_1,,,,x_n) \in W|\theta) = pr(g(T(x_1,,,,x_n),\theta) \in g(W,\theta)|\theta)$$

in which $g(W,\theta)$ denotes the domain of $g(w,\theta)$ for any $w \in W$.

5.Find out the W, which is a range of T, s.t.$pr(g(T(x_1,,,,x_n),\theta) \in g(W,\theta)|\theta) \leq \alpha$ holds for any $\theta \in \Theta$.

Interval estimation:

Point estimation gives a specific value, but we cannot tell how accurate this value relative to the real parameter. (notice that in the previous sections we just studied the distribution property of estimators.) Often, we want to the use an interval: to find two estimators value (which are functions of sample)and estimate the parameter to be an interval.

The sample is random, so the estimator, which is a function of the sample, is also random, the interval is also random, and whether this interval can cover the real parameter is also random. Intuitively, you want the probability that the interval can cover the real parameter to be large as possible. But we cannot make the interval too large. Therefore we use the concept of confidence interval.

Each time, you get a sample, and have an interval. Repeat this many many times. Then, if (1-α) ratio of these intervals contain the true parameter, we say that the interval, which is function of the sample, has a confidence level of α. (i.e, if we generate the interval based on this function and a sample and repeat it many many times, then approximately (1-α) of these intervals cover the real value.)

Given a sample and estimator, and α, we want find out 'an interval, which is a function of the sample, a distribution(which is not related to the DGP distribution), and the α', such that among many times of the realization of such an interval, rouhghly $1-\alpha$ contains the true parameters.

Definition 6.6.3 and 6.6.4 gives the motivation for one-sided confidence interval.

How to construct the confidence interval? The idea is also intuitive. For an estimator $\hat{\theta}$, which is a function of sample, (if you know its distribution function), its probability function obviously include θ. But, using some algebra, we can create a another random variable g, which is a function of the sample $\hat{\theta}(x_1, x_2...)$ and θ, such that the pdf of g does not contain any θ:

$$g(\hat{\theta}(x_1, x_2...x_n), \theta)$$

For simplicity we denote this as $g(x_1, x_2...x_n, \theta)$. In practice somtimes, however, we may want to construct g by adding multiple the estimators. The g is good as far as it's distribution does not contain θ. then we want to find out c and d such that:

$$pr(c < g(x_1, x_2,,,x_n, \theta) < d) = 1 - \alpha$$

which means that

$$pr[g^{-1}(x_1, x_2,,,x_n, c) < \theta < g^{-1}(x_1, x_2,,,x_n, d)] = 1 - \alpha$$

of course the candidate c and d is not unique. In Practice we often choose

$$pr(g(x_1, x_2,,,x_n, \theta) > d) = \frac{\alpha}{2}, pr(g(x_1, x_2,,,x_n, \theta) < c) = \frac{\alpha}{2}$$

Of course, different estimators brings different intervals even for the same confidence level. After all these step, we say, if we use $\hat{\theta}(x_1, x_2...x_n)$ as an estimation of θ, we can construct the above interval $[g^{-1}(x_1, x_2,,,x_n, c), g^{-1}(x_1, x_2,,,x_n, d)]$. this interval changes when sample changes, but if we repeat estimation many times, then roughly $1-\alpha$ of these intervals would include the true parameter θ. Always remember that the confidence level is with respect to specific estimator !(i.e., specific function of sample)