Parameter Estimation and Hypothesis Testing

03/20/2025

As part of inferential statistics, parameter estimation is to quantitatively determine the few parameters that completely describes the probability distribution for the observed outcomes. Hypothesis testing, on the other hand, is a decision making procedure designed to minimize the chance of making a bad decision as a result of the “Type I Error” when a valid null hypothesis is rejected instead of being accepted. The probability of making a Type I error is also known as the significance of the hypothesis test, denoted as α.

The first step of making parameter estimates is to select the probability distribution. Many factors may influence the choice of probability distributions. For example, Poisson distribution is best for random events occurring at a constant rate, which is also equivalent to the binomial distribution when the probability of success or failure is a constant. Gamma distribution is best for non-negative random variables, *e.g.*, occurrences of floods, wildfires, earthquakes, etc. The normal distribution is often chosen when the sample size is so large that the central limit theorem takes effect for the sample means.

Parameter Estimation

There are two main approaches to parameter estimation. One is the method of moments and the other the method of maximum likelihood.

The Method of Moments:

The moment-generating function of a random variable, X, is $M(t) = E(exp(tX))$. Importantly, $M^{(r)}(0) = E(X^r) = \mu_r$, where $M^{(r)}(t)$ is the r-th differentiation of $M(t)$ and μ_r the r-th moment of a probability distribution. The method of moments states that if $X_1, X_2, \ldots, X_n$ are i.i.d. random variables, then the r-th sample moment as defined as:

$$\hat{\mu}_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

is an estimate of μ_r, the r-th moment of the population probability distribution. Take the normal distribution as an example and use the alternative formula for $\sigma^2 = E(X^2) - (E(X))^2$,

$$\mu_1 = E(X) = \mu$$
$$\mu_2 = E(X^2) = \mu^2 + \sigma^2$$

Therefore,

Ⓒ all rights reserved, Orange County Science Institute, 2025

$\mu = \mu_1$
$\sigma^2 = \mu_2 - \mu_1^2$

The method of moments' estimates of μ and σ^2 are:

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

The Method of Maximum Likelihood:

Let θ be the parameter of the probability distribution of i.i.d. random variables $X_1, X_2, \ldots, X_n$. Assume a joint distribution function exists where X_i take sample values x_i, the likelihood of θ as a function of $x_i, x_2, \ldots, x_n$ is defined as

$$lik(\theta) = f(x_1, x_2 \ldots, x_n|\theta)) = \prod_{\imath=1}^{n} f(X_i|\theta)$$

The log likelihood is

$$l(\theta) = \sum_{i=1}^{n} log(f(X_i|\theta))$$

The maximum likelihood estimate of θ is the value of θ that maximizes the likelihood, or the equivalent log likelihood. Operationally, it is the critical point of $l(\theta)$, *i.e.*, the solution for setting the partial derivative of the function with respect to θ. For example, the log likelihood function of normal distribution is:

$$l(\mu, \sigma) = -nlog\sigma - \frac{n}{2}log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

At the maximum of the likelihood function, the partial derivatives with respect to μ and σ are zero:

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0$$
$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2$$

© all rights reserved, Orange County Science Institute, 2025

The solutions to the above equations or the maximum likelihood estimates are:

$$\hat{\mu} = \overline{X}$$
$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

These are identical to those using the method of moments.

Confidence Intervals:

A confidence interval for μ is based on the fact that the random variable below has a t distribution with $n-1$ degrees of freedom (please see the article “central limit theorem”):

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

and

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Let $t_{n-1}(\alpha/2)$ denote the t-score at which the cumulative t_{n-1} distribution has probability of $\alpha/2$. By definition the following probability holds:

$$P\left(-t_{n-1}(\alpha/2) \leq \frac{\sqrt(n)(\overline{X} - \mu)}{S} \leq t_{n-1}(\alpha/2)\right) = 1 - \alpha$$

Rearrangement of the inequality inside the parentheses leads to:

$$P\left(\overline{X} - \frac{S}{\sqrt{n}}t_{n-1}(\alpha/2) \leq \mu \leq \overline{X} + \frac{S}{\sqrt{n}}t_{n-1}(\alpha/2)\right) = 1 - \alpha$$

Therefore, the 100*(1-α)% confidence interval for μ is:

$$\left(\overline{X} - \frac{S}{\sqrt{n}}t_{n-1}(\alpha/2), \overline{X} + \frac{S}{\sqrt{n}}t_{n-1}(\alpha/2)\right),\ or$$
$$\overline{X} \pm t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}}$$

© all rights reserved, Orange County Science Institute, 2025

This latter format is frequently used in engineering reports. For a large test sample size, $t_{n-1}(\alpha/2)$ is replaced by $z_{\alpha/2}$ of the normal distribution.

<u>Hypothesis Testing</u>

A statistical hypothesis involves a claim about the numerical value of a population parameter, such as the population proportion or mean. The null hypothesis (H_o) is based on the prior knowledge of the population and assumes that the population parameter has not changed. The alternate hypothesis (H_A), however, claims that the population parameter is either more than, less than the prior knowledge value (one tailed), or merely different from the prior knowledge value (two tailed).

A level of significance (α) is selected for hypothesis testing. This value is largely based on the custom of statistical investigation. Often, $\alpha = 5\%$ is assumed. Using α, a critical value defining the rejection region is computed with the correct cumulative probability distribution. If the sample statistic falls into the rejection region, the null hypothesis is rejected. Otherwise, rejection to the null hypothesis is not adopted (called “fail to reject”).

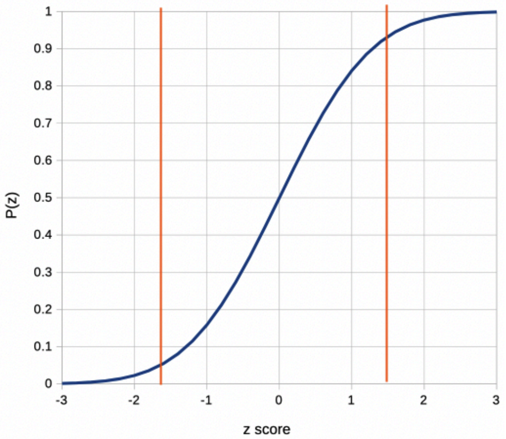
Population Proportion

When the sample size is large enough, the correct probability distribution is the standard normal distribution $N(0,1)$. Therefore, the rejection region defined by the critical value (CV, orange lines, see below) is computed from the following inequality for an one-tailed H_A:

$$P(z \leq CV) \leq \alpha$$
$$P(z \geq CV) \geq 1-\alpha$$

For a two-tailed H_A, the inequality becomes:

$$P(z \leq CV_1) \leq \frac{\alpha}{2}$$
$$P(z \geq CV_2) \geq 1-\frac{\alpha}{2}$$



where CV_1, CV_2 have the same value but opposite signs delineating the boundaries of the two rejection regions on the tails.

The sample statistic, z, is computed as follows:

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

© all rights reserved, Orange County Science Institute, 2025

where $\hat{p}$ is the sample proportion, p the population proportion from the prior knowledge, and n the sample size. If z locates inside the rejection region defined by the critical value (CV), the null hypothesis is rejected.

Alternatively, a p-value (probability) is computed from the cumulative standard normal distribution for the z score. If the p-value is less than α, the null hypothesis is rejected.

For hypothesis testing of two sample proportions, the z score calculation is changed to the following:

$$z = \frac{(\hat{p_1}) - (\hat{p_2}) - (p_1 - p_2)}{\sqrt{\overline{p}(1-\overline{p})(1/n_1 + 1/n_2)}}$$

where

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Population Mean

The correct probability distribution in this case is the t distribution with the degrees of freedom $n-1$. The critical value defining the rejection region is computed from the t distribution and the level of significance (α).

The sample statistic, t, is computed below:

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

where $\overline{X}$ is the sample mean, μ the population mean based on prior knowledge, s sample standard deviation, and n sample size. If t is found inside the rejection region, the null hypothesis is rejected.

Likewise, a p-value is computed from the cumulative t distribution from the t score. If the p-value is less than α, the null hypothesis is rejected.

For hypothesis testing of two sample means, the t score calculation is modified as below:

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

© all rights reserved, Orange County Science Institute, 2025

If the prior knowledge indicates that the two population variances are equal, the t score becomes:

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

For hypothesis testing of paired two samples, the difference between the two values of each pair is $d_i = x_i^{2nd} - x_i^{1st}$ and the t score is computed as follows:

$$t = \frac{\overline{d} - \mu_d}{S_d/\sqrt{n}}$$

where

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$\mu_d = \mu_2 - \mu_1$$

$$S_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1}}$$

Concluding Remarks

The population parameters can be estimated using either the method of moments or the method of maximum likelihood. Both methods yield the same estimates. Importantly, a confidence interval of a point estimate is mandatory for scientific and engineering reporting. The confidence interval is composed of the point estimate plus (or minus) the margin of error. The latter is a product of the sample standard error ($S/\sqrt{n}$) and the t score of the significance ($1-\alpha$). Statistical hypothesis testing is a procedure for decisions that minimizes the bias. Two approaches are commonly used. One is to define a rejection region in the $z \ or \ t$ space. If the test $z \ or \ t$ score falls into the rejection region, the null hypothesis is rejected. The other equivalent approach is to compute the probability of the $z \ or \ t$ score as the p-value. If the p-value is less than the level of significance of the test (α), the null hypothesis is rejected. These two approaches can also be connected to the confidence interval of the point estimate.

© all rights reserved, Orange County Science Institute, 2025