Frequentist statistics

2018/10/12

Yuhang

Introduce:

1. Parameter estimation

(1)The method of maximum likelihood
(2) The method of least squares

2.Confidence intervals

3.Confidence level upper limit

Fundamental concepts

Consider an experiment whose outcome is characterized by one or more data values, which we can write as a vector $\boldsymbol{x}$. A *hypothesis* H is a statement about the probability for the data, often written $P(\boldsymbol{x}|H)$. (We will usually use a capital letter for a probability and lower case for a probability density. Often the term p.d.f. is used loosely to refer to either a probability or a probability density.) This could, for example, define completely the p.d.f. for the data (a *simple* hypothesis), or it could specify only the functional form of the p.d.f., with the values of one or more parameters not determined (a *composite* hypothesis).

If the probability $P(\boldsymbol{x}|H)$ for data $\boldsymbol{x}$ is regarded as a function of the hypothesis H, then it is called the *likelihood* of H, usually written $L(H)$. Often the hypothesis is characterized by one or more parameters $\boldsymbol{\theta}$, in which case $L(\boldsymbol{\theta}) = P(\boldsymbol{x}|\boldsymbol{\theta})$ is called the likelihood function.

Parameter estimation

(a) consistency, (b) bias, (c)efficiency

(a) An estimator is said to be consistent if the estimate $\hat{\theta}$ converges in probability to the true value θ as the amount of data increases. This property is so important that it is possessed by all commonly used estimators.

(b) The bias, b = E[$\hat{\theta}$] − θ, is the difference between the expectation value of the estimator and the true value of the parameter. When b = 0, the estimator is said to be unbiased. The bias depends on the chosen metric, i.e., if $\hat{\theta}$ is an unbiased estimator of θ, then $\hat{\theta}^2$ is not in general an unbiased estimator for θ^2.

(c).Efficiency is the ratio of the minimum possible variance for any estimator of θ to the variance V [$\hat{\theta}$] of the estimator $\hat{\theta}$.

The method of maximum likelihood :

Suppose we have a set of measured quantities $\boldsymbol{x}$ and the likelihood $L(\boldsymbol{\theta}) = P(\boldsymbol{x}|\boldsymbol{\theta})$ for a set of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$. The *maximum likelihood* (ML) estimators for $\boldsymbol{\theta}$ are defined as the values that give the maximum of L. Because of the properties of the logarithm, it is usually easier to work with $\ln L$, and since both are maximized for the same parameter values $\boldsymbol{\theta}$, the ML estimators can be found by solving the *likelihood equations*,

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \,, \qquad i = 1, \ldots, N \,. \tag{39.9}$$

Often the solution must be found numerically. Maximum likelihood estimators are important because they are unbiased and efficient asymptotically (*i.e.*, for large data samples), under quite general conditions, and the method has a wide range of applicability.

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}) \,. \tag{39.10}$$

In this case the number of events n is regarded as fixed. If however the probability to observe n events itself depends on the parameters $\boldsymbol{\theta}$, then this dependence should be included in the likelihood. For example, if n follows a Poisson distribution with mean μ and the independent x values all follow $f(x; \boldsymbol{\theta})$, then the likelihood becomes

$$L(\boldsymbol{\theta}) = \frac{\mu^n}{n!} e^{-\mu} \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}) \,. \tag{39.11}$$

The method of least squares :

The *method of least squares* (LS) coincides with the method of maximum likelihood in the following special case. Consider a set of N independent measurements y_i at known points x_i. The measurement y_i is assumed to be Gaussian distributed with mean $\mu(x_i;\boldsymbol{\theta})$ and known variance σ_i^2. The goal is to construct estimators for the unknown parameters $\boldsymbol{\theta}$. The log-likelihood function contains the sum of squares

$$\chi^2(\boldsymbol{\theta}) = -2\ln L(\boldsymbol{\theta}) + \text{ constant } = \sum_{i=1}^{N} \frac{(y_i - \mu(x_i;\boldsymbol{\theta}))^2}{\sigma_i^2} \,. \tag{39.19}$$

The parameter values that maximize L are the same as those which minimize χ^2.

The minimum of the chi-square function in Equation (39.19) defines the least-squares estimators $\widehat{\boldsymbol{\theta}}$ for the more general case where the y_i are not Gaussian distributed as long as they are independent. If they are not independent but rather have a covariance matrix $V_{ij} = \mathrm{cov}[y_i, y_j]$, then the LS estimators are determined by the minimum of

$$\chi^2(\boldsymbol{\theta}) = (\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T V^{-1} (\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))\,, \tag{39.20}$$

Frequentist confidence intervals :

Confidence intervals are constructed in order to have a coverage probability greater than or equal to a given confidence level, regardless of the true parameter's value.

confidence intervals are said to have a confidence level (CL) equal to 1 − α.

$$1 - \alpha = P(x_1(\theta) < x < x_2(\theta)) = P(\theta_2(x) < \theta < \theta_1(x)) \, . \quad (39.68)$$

In this probability statement, $\theta_1(x)$ and $\theta_2(x)$, *i.e.*, the endpoints of the interval, are the random variables and θ is an unknown constant. If the experiment were to be repeated a large number of times, the interval $[\theta_1, \theta_2]$ would vary, covering the fixed value θ in a fraction $1 - \alpha$ of the experiments.

Gaussian distributed measurements:

An important example of constructing a confidence interval is when the data consists of a single random variable x that follows a Gaussian distribution; this is often the case when x represents an estimator for a parameter and one has a sufficiently large data sample. If there is more than one parameter being estimated, the multivariate Gaussian is used. For the univariate case with known σ, the probability that the measured value x will fall within $\pm\delta$ of the true value μ is

$$1 - \alpha = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mu-\delta}^{\mu+\delta} e^{-(x-\mu)^2/2\sigma^2}\, dx = \mathrm{erf}\left(\frac{\delta}{\sqrt{2}\,\sigma}\right) = 2\Phi\left(\frac{\delta}{\sigma}\right) - 1\,, \qquad (39.70)$$

Table 39.1: Area of the tails α outside $\pm\delta$ from the mean of a Gaussian distribution.

α	δ	α	δ
0.3173	1σ	0.2	1.28σ
4.55×10^{-2}	2σ	0.1	1.64σ
2.7×10^{-3}	3σ	0.05	1.96σ
6.3×10^{-5}	4σ	0.01	2.58σ
5.7×10^{-7}	5σ	0.001	3.29σ
2.0×10^{-9}	6σ	10^{-4}	3.89σ

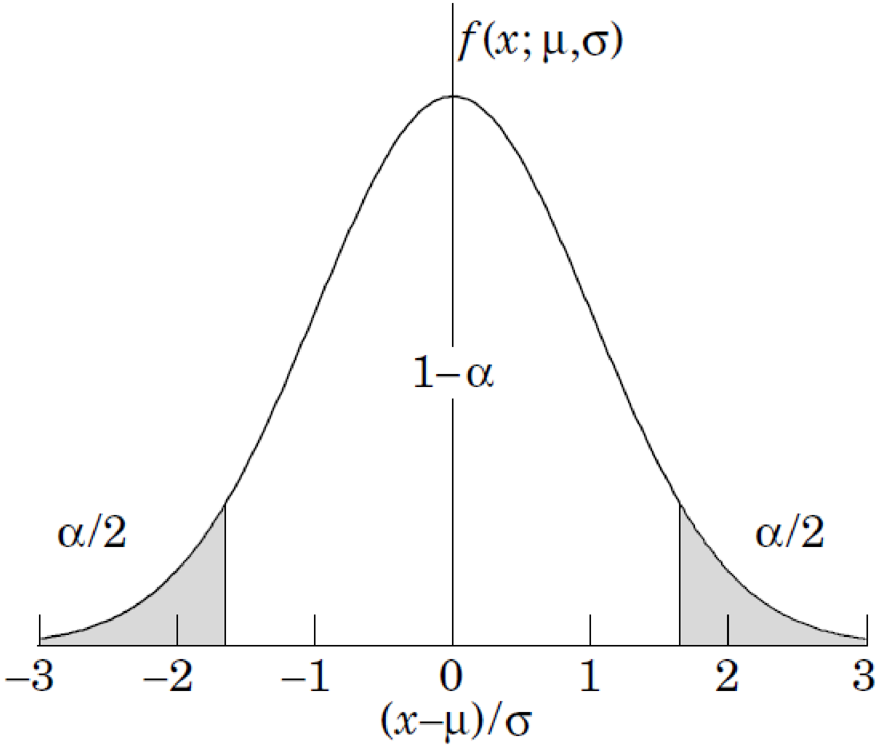


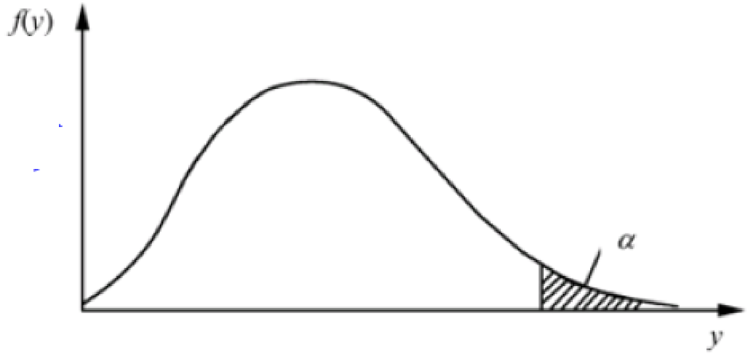
Figure 39.4: Illustration of a symmetric 90% confidence interval (unshaded) for a Gaussian-distributed measurement of a single quantity. Integrated probabilities, defined by $\alpha = 0.1$, are as shown.

The relation (39.70) can be re-expressed using the cumulative distribution function for he χ^2 distribution as

$$\alpha = 1 - F(\chi^2; n) , \tag{39.71}$$

or $\chi^2 = (\delta/\sigma)^2$ and $n = 1$ degree of freedom. This can be seen as the $n = 1$ curve in

The upper α percentage point of $\chi^2(n)$ $(0 < \alpha < 1)$ *approximate* the upper α percentage point of Gaussian distributed N(0,1).



Confidence level upper limit:

In the limit case of subsample capacity n→∞, the likelihood estimation of the parameter $\hat{\theta}$ obeys the Gaussian distribution with the mean value θ and the variance as σ^2, so

$$r = 1 - \alpha = \frac{\int_{-\infty}^{x} L(x|\theta)d\theta}{\int_{-\infty}^{+\infty} L(x|\theta)d\theta}$$

If we want get the 95% confidence level upper limit, we can let r=95%, and we can get the value x. The value x is our upper limit.

Thank you