Confidence Regions

November 16, 2023

Motivation

- When we perform estimation (or *point estimation* to be specific) as an outcome we get a number that we know that it is not equal to the value one seeks, it only *estimates* it.
- But how well does it do that task? One could provide the mean square error of this estimator.
- However, the MSE is a global entity that is independent of specific data and thus of a specific value of the estimator (often referred to as an estimate).
- One would like to know what one can report in relation to the data driven estimate.
- This is the subject of the confidence interval methodology which often is called interval estimation in the oppose to point estimation.
- We consider the problem of quoting a subset of θ values which are in some sense plausible in the light of the data $\mathbf{x}$.
- We need a procedure which for each possible value $\mathbf{x} \in \mathcal{X}$ specifies a subset $C(\mathbf{x})$ of Θ which we should quote as a set of plausible values for θ.

Definition

Definition

Let $X_1, \ldots, X_n$ be a sample form a distribution that is parameterized by some parameter θ. A *random* set $C(X_1, \ldots, X_n)$ of possible values for θ that is computable from the sample is called a **confidence region** at **confidence level** $1 - \alpha$ if

$$P_\theta(\theta \in C(X_1, \ldots, X_n)) = 1 - \alpha.$$

If the set $C(X_1, \ldots, X_n)$ has the form of an interval (L, U), then we call it a **confidence interval** and we can write $P_\theta(L \leq \theta \leq U) = 1 - \alpha$. If the set is a one-sided unbounded interval (L, ∞) or (∞, U), than U, L is called an upper, lower, respectively, confidence bound.

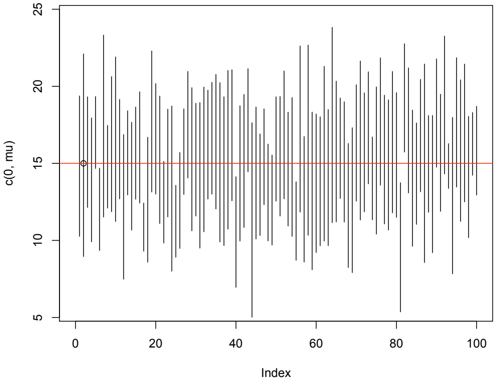
Note: $L = L(\mathbf{x})$, $U = U(\mathbf{x})$ are data dependent and parameter independent, thus are data based statistics.

Interpretation

- In the relation
$$P_\theta(L \le \theta \le U) = 1 - \alpha$$
$L = L(\mathbf{X})$ and $U = U(\mathbf{X})$ are random but θ is not (unless you enter the realm of Baysian statistics which we do not do in this part).
- However, if the specific data $\mathbf{X} = \mathbf{x}$ are given $L = L(\mathbf{x})$, $U = U(\mathbf{x})$ are just numbers and not random at all. One should not interpret the confidence interval as claiming that the parameter θ is between $L(\mathbf{x})$ and $U(\mathbf{x})$ with probability $1 - \alpha$ as all three entities θ, L, and U are just numbers with no probability involved.
- More appropriate is to think that if we repeat our experiment many times and evaluate the confidence intervals each times, in $(1 - \alpha)100\%$ cases the true parameter will be inside the intervals.
- We would not know in which ones, though.
- Our reality 'sits' on only one of these intervals and we do not really repeat this experiment.

Graphical illustration

The following figure illustrate this well:



- The red line represents a true parameter (unknown).
- We are ‘sitting’ on one of these intervals.
- We do not know which but we know its endpoints.

Example

- Suppose we are going to observe data $\mathbf{x}$ where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, and $x_1, x_2, \ldots, x_n$ are the observed values of random variables $X_1, X_2, \ldots, X_n$ which are thought to be iid $N(\theta, 1)$ for some unknown parameter $\theta \in (-\infty, \infty) = \Theta$.
- Consider the subset $C(\mathbf{x}) = [\bar{x} - 1.96/\sqrt{n}, \bar{x} + 1.96/\sqrt{n}]$. If we carry out an infinite sequence of independent repetitions of the experiment, then we will get an infinite sequence of $\mathbf{x}$ values and thereby an infinite sequence of subsets $C(\mathbf{x})$.
- We might ask what proportion of this infinite sequence of subsets actually contain the fixed but unknown value of θ?
- This follows from the fact that $\bar{X}$ has a $N(\theta, \frac{1}{n})$ density and so $Z = \frac{\bar{X} - \theta}{\frac{1}{\sqrt{n}}} = \sqrt{n}(\bar{X} - \theta)$ has a $N(0, 1)$ density. Thus even though θ is unknown, we can calculate that $P[|Z| \leq 1.96] = 0.95$.
- Thus 95% of the time Z will lie between -1.96 and $+1.96$. But

$$
\begin{aligned}
-1.96 \leq Z \leq +1.96 &\Rightarrow -1.96 \leq \sqrt{n}(\bar{X} - \theta) \leq +1.96 \\
&\Rightarrow \bar{X} - 1.96/\sqrt{n} \leq \theta \leq \bar{X} + 1.96/\sqrt{n} \Rightarrow \theta \in C(\mathbf{X})
\end{aligned}
$$

  The proportion of the infinite sequence of subsets $C(\mathbf{X})$ which will actually include the fixed but unknown value of θ is 0.95. For this reason the set $C(\mathbf{X})$ is

The pivotal quantity

The crucial step in the last example was finding the quantity $Z = \sqrt{n}(\bar{X} - \theta)$ whose value depended on the parameter of interest θ but whose distribution was *known* to be that of a standard normal variable. This leads to the following definition.

Definition (Pivotal Quantity)

A pivotal quantity for a parameter θ is a random variable $Q(\mathbf{X}|\theta)$ whose value depends both on (the data) $\mathbf{X}$ and on the value of the unknown parameter θ but whose distribution is explicitly known (parameter free).

The quantity Z in the example above is a pivotal quantity for θ.

Example

Let $X_1, X_2, \ldots, X_n$ be iid observations from a $N(\theta, \sigma^2)$ density where θ is known. Define $Q = \sum_{i=1}^{n}(X_i - \theta)^2/\sigma^2 = \sum_{i=1}^{n} Z_i^2$ where $Z_i = (X_i - \theta)/\sigma \sim \mathcal{N}(0, 1)$ density. Hence, Q has a χ_n^2 density and so is a pivotal quantity for σ. If $n = 20$, then we can be 95% sure that $9.591 \leq Q \leq 34.170$ which is equivalent to a 95% CI for variance σ^2

$$\sum_{i=1}^{n}(X_i - \theta)^2/34.170 \leq \sigma^2 \leq \sum_{i=1}^{n}(X_i - \theta)^2/9.591.$$

Finding a pivotal quantity

The following lemma provides a method of finding pivotal quantities in general.

Lemma

Let X be a random variable with cdf F. Consider the random variable $U = -2\log[F(X)]$. Then U has a χ_2^2 density. Consider the random variable $V = -2\log[1 - F(X)]$. Then V has a χ_2^2 density.

Proof.

Observe that, for $a \geq 0$,

$$\begin{aligned} P[U \leq a] &= P[F(X) \geq \exp(-a/2)] \\ &= 1 - P[F(X) \leq \exp(-a/2)] \\ &= 1 - P[X \leq F^{-1}(\exp(-a/2))] \\ &= 1 - F[F^{-1}(\exp(-a/2))] \\ &= 1 - \exp(-a/2). \end{aligned}$$

Hence, U has density $\frac{1}{2}\exp(-a/2)$ which is the density of a χ_2^2 variable as required. The corresponding proof for V is left as an exercise. □

Application

The above lemma has an immediate, and very important, application.

- Suppose that we have data $X_1, X_2, \ldots, X_n$ which are iid with density $f(x|\theta)$ and the corresponding cdf $F(a|\theta)$. For $i = 1, 2, \ldots, n$, define $U_i = -2\log[F(X_i|\theta)]$.
- Then $U_1, U_2, \ldots, U_n$ are iid each having a χ^2_2 density. Hence $Q_1(\mathbf{X}, \theta) = \sum_{i=1}^n U_i$ has a χ^2_{2n} density and so is a pivotal quantity for θ. Another pivotal quantity (also having a χ^2_{2n} density) is given by $Q_2(\mathbf{X}, \theta) = \sum_{i=1}^n V_i$ where $V_i = -2\log[1 - F(X_i|\theta)]$.

Example

Consider $X_1, X_2, \ldots, X_n$ having the exponential distribution with the intensity θ. Suppose that we want to construct a 95% confidence interval for θ. We need to find a pivotal quantity for θ. Since $F(a|\theta) = 1 - \exp(-\theta a)$,

$$Q_1(\mathbf{X}, \theta) = -2\sum_{i=1}^n \log[1 - \exp(-\theta X_i)], \quad Q_2(\mathbf{X}, \theta) = -2\sum_{i=1}^n \log[\exp(-\theta X_i)] = 2\theta\sum_{i=1}^n X_i$$

are pivotal quantities for θ having a χ^2_{2n} density. Let $A < B$ such that $P[\chi^2_{2n} < A] = P[\chi^2_{2n} > B] = 0.025$. Then

$$0.95 = P[A \le Q_2(\mathbf{X}, \theta) \le B] = P[A \le 2\theta\sum_{i=1}^n X_i \le B] = P[\frac{A}{2\sum_{i=1}^n X_i} \le \theta \le \frac{B}{2\sum_{i=1}^n X_i}]$$

and so the interval $[\frac{A}{2\sum_{i=1}^n X_i}, \frac{B}{2\sum_{i=1}^n X_i}]$ is a 95% confidence interval for θ.

The pivot for the ratio of the two variances

Suppose that we have data $X_1, X_2, \ldots, X_n$ which are iid observations from a $\mathcal{N}(\theta_X, \sigma_X^2)$ density and data $Y_1, Y_2, \ldots, Y_m$ which are iid observations from a $\mathcal{N}(\theta_Y, \sigma_Y^2)$ density where $\theta_X, \theta_Y, \sigma_X$, and σ_Y are all unknown. Let $\lambda = \sigma_X^2/\sigma_Y^2$ and define

$$F^* = \frac{\hat{s}_X^2}{\hat{s}_Y^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)} \frac{(m-1)}{\sum_{j=1}^m (Y_j - \bar{Y})^2}.$$

Let

$$W_X = \sum_{i=1}^n (X_i - \bar{X})^2/\sigma_X^2, \quad W_Y = \sum_{j=1}^m (Y_j - \bar{Y})^2/\sigma_Y^2.$$

Then, W_X has a χ^2_{n-1} density and W_Y has a χ^2_{m-1} density. Hence, by definition of F-distribution (F stands for Fisher),

$$Q = \frac{W_X/(n-1)}{W_Y/(m-1)} \equiv \frac{F^*}{\lambda}$$

has an F density with $n-1$ and $m-1$ degrees of freedom and so is a pivotal quantity for λ. Suppose that $n = 25$ and $m = 13$. Then we can be 95% sure that $0.39 \leq Q \leq 3.02$ which is equivalent to

$$\frac{F^*}{3.02} \leq \lambda \leq \frac{F^*}{0.39}.$$

Approximate confidence intervals

In many situation, we have the limiting distribution of certain functions of the data, where the limit is taken with respect to the sample size n. Many of the results on this will be given later on. To illustrate how this can work for construction confidence intervals let us consider the well-known Central Limit Theorem. It says that under the second moment assumptions, the following holds

$$\lim_{n\to\infty} \sqrt{n}(\bar{X} - \mu)/\sigma \stackrel{d}{=} N(0, 1).$$

Such a limiting result can be used to obtain an approximate pivot for approximate confidence intervals for the probability of success in n Bernoulli trials.

- If $X_1, ..., X_n$ are the indicators of n Bernoulli trials with probability of success θ, then X is the MLE of θ. There is no natural "exact" pivot based on $\bar{X}$ and θ . However, by the De Moivre–Laplace (CLT for Bernoulli) theorem, $\sqrt{n}(\bar{X} - \theta)/\sqrt{\theta(1-\theta)}$ has approximately $N(0, 1)$ distribution. If we use this function as an approximate pivot we obtain that the following occurs with probability approximately equal to $1 - \alpha$.

$$-z_{1-\alpha/2} \leq \sqrt{n}(\bar{X} - \theta)/\sqrt{\theta(1-\theta)} \leq z_{1-\alpha/2}$$

The above inequalities can be solve with respect to θ leading a confidence interval $(\underline{\theta}, \bar{\theta})$ for θ, see the formula 4.4.3 in the text.

Multidimensional parameters

- If one consider multidimensional parameters, then the concept of a confidence region may result in many shapes of the sets.
- These shapes may be more difficult to interpret, and interpretation was the main benefit of the confidence intervals.
- Here we point out on the two shapes that are commonly used in the multivariate case:
  - multidimensional rectangles – most commonly used ones and the easiest to interpret.
  - multidimensional ellipsoides – typically obtained for the confidence regions for the multivariate means with use of the normal approximation.
  - mixture of the above – some parameters maybe easier to estimate by confidence intervals, for example variances, while the means by ellipsoids.
  - bands around functions if the parameters are functions – non-parametric statistics.
- Due to the diversity of possible cases, we limit ourselves to general concepts and very simple examples.
- It should be also mentions that it may be difficult to obtain the confidence regions with exactly specified confidence bands, often we will be satisfied with the *conservative* confidence regions $P_\theta(\theta \in C(\mathbf{X})) > 1 - \alpha$.

Multidimensional parameters

- We can extend the notion of a confidence interval for one-dimensional functions $q(\theta)$ to r-dimensional vectors $q(\theta) = (q_1(\theta), \ldots, q_r(\theta))$.
- Suppose $\underline{q}_j(X) \leq \overline{q}_j(X)$, $j = 1, \ldots, r$. Then the r-dimensional random rectangle $I(X) = [\underline{q}_1(X), \overline{q}_1(X)] \times \cdots \times [\underline{q}_r(X), \overline{q}_r(X)]$ is said to be a level $(1 - \alpha)$ confidence region, if the probability that it covers the unknown but fixed true $(q_1(\theta), \ldots, q_r(\theta))$ is at least $1 - \alpha$.
- We write this as
$$P_\theta(\mathbf{q}(\theta) \in I(X)) \geq 1 - \alpha.$$

From 1D to multiD

- **Coordinatewise independence**: If the pairs $\underline{q}_j(X) \le \overline{q}_j(X)$, $j = 1, \dots, r$ are independent and they yield coordinatewise the confidence level $1 - \alpha_j$, respectively, then

$$P_\theta(\mathbf{q}(\theta) \in I(X)) = \prod_{j=1}^{r}(1 - \alpha_j).$$

  Moreover, if we choose $\alpha_j = 1 - (1 - \alpha)^{1/r}$, $j = 1, \dots, r$, then $I(X)$ has the confidence level $1 - \alpha$.

- **Bonferoni's inequality**: Without assuming independence, we have

$$P_\theta(\mathbf{q}(\theta) \in I(X)) \ge 1 - \sum_{j=1}^{r} P_\theta(q_j \notin I_j(X)) = 1 - \sum_{j=1}^{r} \alpha_j.$$

Proof.

$$P\left(\bigcap_{j=1}^{r} A_j\right) = 1 - P\left(\bigcup_{j=1}^{r} A_j^c\right) \ge 1 - \sum_{j=1}^{r} P(A_j^c)$$

□

Application to the mean/variance estimation of normal parameters

Example

Let $X_1, X_2, \ldots, X_n$ be iid observations from a $N(\theta, \sigma^2)$ density. It is well-known that $\bar{X}$ and S^2 are independent statistics with the distributions $X(\mu, \sigma^2/n)$ and $(n-1)\sigma^2\chi^2_{n-1}$. Moreover, $\sqrt{n}(\bar{X}-\mu)/\sigma$ has Student-t distribution with $n-1$ degrees of freedom. Thus the individual confidence intervals at levels $1-\alpha/2$ are $I_1(X) = \bar{X} \pm St_{n-1,1-\alpha/4}$ and $I_2(X) = [(n-1)S^2/\chi^2_{n-1,1-\alpha/4}, (n-1)S^2/\chi^2_{n-1,\alpha/4}]$ The rectangle $I_1(X) \times I_2(X)$ will be the joint confidence region at level $1-\alpha$ by Bonferroni's inequality.

In this case, one can get also exact confidence region but not in the shape of rectangle. In fact it will be in the shape of the trapezoid if μ and σ (instead of σ^2) are considered to be the parameters of interest. It follows from the independence of $\bar{X}$ and S^2 and the following exact inequalities

$$\bar{X} - z_{1-\alpha_0}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{1-\alpha_0}\frac{\sigma}{\sqrt{n}}, \quad \frac{\sqrt{n-1}S}{\chi_{n-1,1-\alpha_1/4}} \le \sigma \le \frac{\sqrt{n-1}S}{\chi_{n-1,,\alpha_1/4}}$$

which are satisfied with the probability $(1-\alpha_0)(1-\alpha_1)$. One can choose α_0 and α_1 arbitrarily so that $(1-\alpha_0)(1-\alpha_1) = 1-\alpha$, for example $\alpha_0 = \alpha_1 = 1-\sqrt{1-\alpha}$.

The mean for multivariate normal

Finally, let us consider the confidence in the shape of a multidimensional ellipsoid.

Example

Let $X_1, X_2, \ldots, X_n$ be iid observations from a multivariate r dimensional normal density $N(\theta, \Sigma)$, with a known covariance matrix Σ. Then the pivot is

$$\sqrt{n}\Sigma^{-1/2}(\bar{X} - \theta)$$

which has the standard multivariate normal distribution and thus belongs to the k-dimensional ball centered at zero of the radius $\chi_{k,1-\alpha}$ with probability $1 - \alpha$. Denote this ball by $B_d(\chi_{k,1-\alpha})$. Then with the same probability

$$\theta \in \bar{X} + \Sigma^{1/2} B_d(\chi_{k,1-\alpha_0}/\sqrt{n})$$

which makes the right hand side the confidence region at $1 - \alpha$ level. It is clear(?) that this region is an r-dimensional ellipsoid.
This example can be extended to the similar situation as in the previous slide, when the covariance is not known to produce the joint region for θ and Σ. As we see, the increased complexity of the parameters leads to more complex geometry of those regions.

The main idea

- Confidence regions are random subsets of the parameter space that contain the true parameter with probability at least $1 - \alpha$.
- Acceptance regions of statistical tests are, for a given hypothesis *H*, subsets of the sample space with probability of accepting *H* at least $1 - \alpha$ when *H* is true.
- We shall establish a duality between confidence regions and acceptance regions for families of hypotheses.
- The null hypothesis must be about the same parameter for which the confidence regions are built.

Examples

- Let consider the normal distribution case with a known variance and a test $H : \mu = \mu_0$ with an arbitrary version of the alternative.
- Then if we reject this hypothesis when
$$\mu_0 \notin [\bar{x} - \sigma z_{1-\alpha/2}/\sqrt{n}, \bar{x} + \sigma z_{1-\alpha/2}/\sqrt{n}$$
than it has the significance level α.
- But it may be equivalently said that we reject H if the parameter does not belong to the confidence interval at the level $1 - \alpha$.
- The argument is almost obvious. If the parameter does not belong to the confidence region at the level $1 - \alpha$ while it is assumed to be the true parameter than it happens with probability α.
- If we base our rejection based on not belonging to the confidence region we obtain a test with corresponding significance α.
- Since there may be many confidence regions for the same parameters thus there are many tests and different regions can be chosen depending what is in the alternative.

The duality theorem

Theorem

If $C(X)$ a confidence region for a parameter θ at the level $1-\alpha$, then for a given θ_0 a test that rejects H when $\theta_0 \notin C(X)$ is at a significance level α.

Conversely, if R_{θ_0} defines a rejection region for T in testing $H: \theta = \theta_0$ at the significance level α for each $\theta_0 \in \Theta$, then the set $I(X)$ of all θ for which $T \notin R_\theta$ defines the confidence region at level $1-\alpha$.

Proof.

The first part was argued at the previous slide.

The second part follows because for fixed while arbitrary θ_0 $P_{\theta_0}(\theta_0 \notin I(X))$ is simply the probability in the case when H claims θ_0 as truth to have $T \notin R_{\theta_0}$, i.e. not rejecting H which happens with probability $1-\alpha$. $\square$