STA 114: Statistics

Notes 7. ML Confidence Intervals for Normal Model

ML interval for the normal model

We saw that for the model $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, σ fixed, the ML interval $B_c(x) = \bar{x} \mp c\sigma/\sqrt{n}$ for μ has confidence coefficient $2\phi(c) - 1$. For the more general model, where $\sigma^2 \in (0, \infty)$ is also included as a model parameter, an (approximate) ML interval for μ is $B_c(x) = \bar{x} \mp cs_x/\sqrt{n}$. In this lecture, we shall calculate the confidence coefficients of these intervals. By simple rearrangements (as we did for the known σ case), the coverage of B_c at any (μ_0, σ_0^2) can be expressed as:

$$\gamma(B_c; (\mu_0, \sigma_0^2)) = P_{[X|(\mu_0,\sigma_0^2)]}\left(-c \le \frac{\bar{X} - \mu_0}{s_X/\sqrt{n}} \le c\right)$$

where s_X denotes the random variable $\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Evaluating the probability on the right would require knowing the distribution of the random variable $T = \frac{\bar{X}-\mu_0}{s_X/\sqrt{n}}$ when $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_0, \sigma_0^2)$. To get there, we first need to describe the joint distribution of $\bar{X}$ and s_X^2. We will do this in several steps.

Orthogonal transformation of Normal variables

An $n \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

is called an orthogonal matrix if each row and each column of A has norm one, and the inner product between any two rows or any two columns of A is zero. This implies that both $A'A$ and AA' equal the n-dimensional identity matrix, where A' denotes the transpose of A. In other words $A^{-1} = A'$.

Consider the system of linear equations

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n \\ \vdots \quad & \quad \vdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n. \end{aligned}$$

For an input $x = (x_1, \cdots, x_n)$, write the output $y = (y_1, \cdots, y_n)$ given by the above system as $y = Ax$ (this can be correctly interpreted as A times x if you think of x and y as n-dimensional column vectors). For any $y^* = (y_1^*, \cdots, y_n^*) \in (-\infty, \infty)^n$ there is a unique solution to $y^* = Ax$ in x, given by $x^* = A'y^*$, i.e., $x^* = (x_1^*, \cdots, x_n^*)$ with $x_i^* = a_{1i}y_1^* + \cdots + a_{ni}y_n^*$. Also note that if (x, y) is an input-output pair, i.e., $y = Ax$, then $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n x_i^2$. The orthogonal system A essentially rotates the input vector by a certain angle, without altering its norm.

RESULT 1. Let $X = (X_1, \cdots, X_n)$ with $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(0, 1)$ and define $Y = (Y_1, \cdots, Y_n)$ as $Y = AX$, the output of the above equations when the input is X, i.e., $Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n$, etc. Then $Y_i \overset{\text{IID}}{\sim} \mathsf{Normal}(0, 1)$.

Proof. Let $f(x)$ and $g(y)$ denote the pdfs of X and Y. We know that

$$f(x) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n x_i^2\right), \quad x = (x_1, \cdots, x_n) \in (-\infty, \infty)^n.$$

For any $a = (a_1, \cdots, a_n) \in (-\infty, \infty)^n$ and for any $r > 0$ let $B_r(a)$ denote the sphere of radius r with center at a, i.e., $B_r(a)$ contains all points $z = (z_1, \cdots, z_n)$ such that $\sum_{i=1}^n (z_i - a_i)^2 \le r^2$. Then, for any $x = (x_1, \cdots, x_n)$ and any $y = (y_1, \cdots, y_n)$,

$$f(x) = \lim_{r \to 0} \frac{P(X \in B_r(x))}{\mathrm{vol}(B_r(x))}, \quad g(y) = \lim_{r \to 0} \frac{P(Y \in B_r(y))}{\mathrm{vol}(B_r(y))}$$

where $\mathrm{vol}(B_r(a))$ denotes the volume of $B_r(a)$.

Fix a $y^* = (y_1^*, \cdots, y_n^*)$ and let $x^* = A'y^*$ be the unique solution of $y^* = Ax$. Observe that $X \in B_r(x^*)$ if and only if $Y \in B_r(y^*)$. To see this, let $\hat{X} = (\hat{X}_1, \cdots, \hat{X}_n)$ with $\hat{X}_i = X_i - x_i^*$, and $\hat{Y} = (Y_1, \cdots, Y_n)$ with $\hat{Y}_i = Y_i - y_i^*$. Then $\hat{Y} = A\hat{X}$ and therefore,

$$X \in B_r(x^*) \iff \sum_{i=1}^n \hat{X}_i^2 \le r^2 \iff \sum_{i=1}^n \hat{Y}_i^2 \le r^2 \iff Y \in B_r(y^*).$$

Also $\mathrm{vol}(B_r(x^*)) = \mathrm{vol}(B_r(y^*))$ because the two spheres have the same radius. Therefore

$$\begin{aligned} g(y^*) &= \lim_{r \to 0} \frac{P(Y \in B_r(y^*))}{\mathrm{vol}(B_r(y^*))} = \lim_{r \to 0} \frac{P(X \in B_r(x^*))}{\mathrm{vol}(B_r(x^*))} = f(x^*) \\ &= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n x_i^{*2}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n y_i^{*2}\right), \end{aligned}$$

because $\sum_{i=1}^n x_i^{*2} = \sum_{i=1}^n y_i^{*2}$ since $y^* = Ax^*$. But since y^* is arbitrary, the pdf of Y is,

$$g(y) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n y_i^2\right), y = (y_1, \cdots, y_n) \in (-\infty, \infty)^n,$$

i.e., $Y_i \overset{\text{IID}}{\sim} \mathsf{Normal}(0, 1)$. □

RESULT 2. Let $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(0,1)$. Then,

1. $\bar{X} \sim \mathsf{Normal}(0, \frac{1}{n})$
2. $\sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$
3. $\bar{X}$ and $\sum_{i=1}^n (X_i - \bar{X})^2$ are independent.

Proof. It is possible to construct an $n \times n$ orthogonal matrix A whose first row is $(\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}})$ [in fact, given any n numbers $b_1, \cdots, b_n$ so that $b_1^2 + \cdots + b_n^2 = 1$, it is possible to construct an orthogonal matrix A with first row $= (b_1, \cdots, b_n)$]. Take $X = (X_1, \cdots, X_n)$ and $Y = AX$ in the sense of Result 1 above. Then $Y = (Y_1, \cdots, Y_n)$ with $Y_i \overset{\text{IID}}{\sim} \mathsf{Normal}(0,1)$ and

$$Y_1^2 + \cdots + Y_n^2 = X_1^2 + \cdots + X_n^2.$$

Now, $Y_1 = X_1/\sqrt{n} + \cdots + X_n/\sqrt{n} = \sqrt{n}\bar{X}$, and so

$$\bar{X} = Y_1/\sqrt{n} \sim \mathsf{Normal}(0, 1/n)$$

because $Y_1 \sim \mathsf{Normal}(0,1)$. Also,

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 = \sum_{i=1}^n Y_i^2 - Y_1^2 = Y_2^2 + \cdots + Y_n^2.$$

But $Y_2, \cdots, Y_n \overset{\text{IID}}{\sim} \mathsf{Normal}(0,1)$, therefore $Y_2^2 + \cdots + Y_n^2 \sim \chi^2(n-1)$ and so $\sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$. Also, Y_i's are independent of each other, therefore, $\bar{X}$, which is a function of Y_1 is independent of $\sum_{i=1}^n (X_i - \bar{X})^2$ which is a function of only $Y_2, \cdots, Y_n$. □

RESULT 3. Let $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$. Then

1. $\bar{X} \sim \mathsf{Normal}(\mu, \frac{\sigma^2}{n})$
2. $\frac{(n-1)s_X^2}{\sigma^2} \sim \chi^2(n-1)$
3. $\bar{X}$ and s_X^2 are independent.

Proof. Define $Z_i = (X_i - \mu)/\sigma$, then $Z_1, \cdots, Z_n \overset{\text{IID}}{\sim} \mathsf{Normal}(0,1)$ and

$$\bar{X} = \mu + \sigma\bar{Z}, \quad \frac{(n-1)s_X^2}{\sigma^2} = \sum_{i=1}^n (Z_i - \bar{Z})^2$$

and therefore Result 2 implies Result 3. □

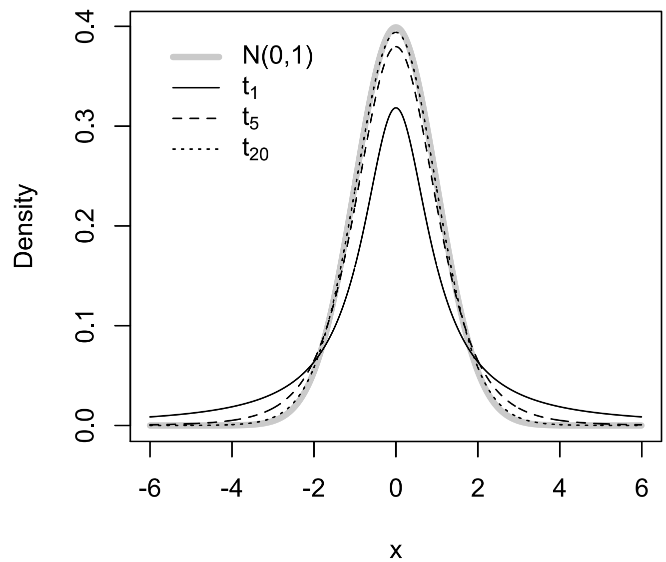

Figure 1: Some $t(k)$-densities and the $\mathsf{Normal}(0,1)$ density. As the degrees of freedom $k \to \infty$, a $t(k)$ density morphs into the $\mathsf{Normal}(0,1)$ density.

The t-distributions

Let W and V be two independent random variables, with $W \sim \mathsf{Normal}(0,1)$ and $V \sim \chi^2(k)$. Then $T = \frac{W}{\sqrt{V/k}}$ is said to have a t-distribution with k degrees of freedom, denoted $T \sim t(k)$. The pdf of $t(k)$ is given by

$$f(x) = \frac{1}{\sqrt{k}B(\frac{1}{2}, \frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}, \quad x \in (-\infty, \infty).$$

The $t(k)$ pdf looks like a normal bell curve, but its tails decay at a slower rate. However, the resemblance improves as k increases. In fact at any x, the $t(k)$ pdf $f(x)$ approaches the $\mathsf{Normal}(0,1)$ pdf $(2\pi)^{-1/2}\exp(-x^2/2)$ when $k \to \infty$. See Figure 1. We will denote the $t(k)$ CDF by $\Phi_k(x)$. Because $t(k)$ is symmetric around 0, for any $T \sim t(k)$,

$$P(-c \le T \le c) = \Phi_k(c) - \Phi_k(-c) = \Phi_k(c) - (1 - \Phi_k(c)) = 2\Phi_k(c) - 1.$$

Note that the normal CDF is a limiting case of Φ_k with $k \to \infty$. I might sometimes write Φ_∞ for Φ. As there are z-tables for values of Φ and its inverse, there are t-tables for Φ_k and their inverses. In `R`, $\Phi_k(x)$ is calculated as `pt(x, df = k)` and its inverse $\Phi_k^{-1}(u)$ is calculated as `qnorm(u, df = k)`.

RESULT 4. Suppose $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, then $T = \frac{\bar{X} - \mu}{s_X/\sqrt{n}} \sim t(n-1)$

Proof. Indeed, $T = \frac{W}{\sqrt{V/(n-1)}}$ where $W = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathsf{Normal}(0,1)$ and $V = \frac{(n-1)s_X^2}{\sigma^2} \sim \chi^2(n-1)$ with W and V independent (Result 3). □

Confidence coefficient of B_c

From the above result, the coverage $\gamma(B_0; (\mu_0, \sigma_0^2))$, at any $(\mu_0, \sigma_0^2) \in (-\infty, \infty) \times (0, \infty)$, can now be calculated as

$$\gamma(B_c; (\mu_0, \sigma_0^2)) = P_{[X|(\mu_0,\sigma_0^2)]}\left(-c \leq \frac{\bar{X} - \mu_0}{s_X/\sqrt{n}} \leq c\right) = 2\Phi_{n-1}(c) - 1,$$

and therefore,

$$\gamma(B_c) = 2\Phi_{n-1}(c) - 1.$$

Confidence intervals

For a given $\alpha \in (0,1)$, a $100(1-\alpha)\%$-CI B_c is obtained by matching

$$2\Phi_{n-1}(c) - 1 = 1 - \alpha \implies c = \Phi_{n-1}^{-1}(1 - \alpha/2).$$

Note that unlike the known σ^2 case, the choice of c now depends on n. Because the $t(k)$ distributions have tails that decay slower than the $\mathsf{Normal}(0,1)$ tails, $\Phi_{n-1}^{-1}(1-\alpha/2)$ is larger than the corresponding $\Phi^{-1}(1-\alpha/2)$ for the known variance model. In other words, the ML 95%-CI for μ in the unknown variance model is wider than the ML 95%-CI for the known variance model.

Notation

For any $\alpha \in (0,1)$, we will use the symbol $z_k(\alpha)$ to denote the quantity $\Phi_k^{-1}(1 - \alpha/2)$. The limiting case, $\Phi^{-1}(1-\alpha/2)$ will be denoted by $z(\alpha)$. Therefore,

1. For the model $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, $\sigma^2 \in (0, \infty)$, a $100(1-\alpha)\%$-CI for μ is $\bar{x} \mp z_{n-1}(\alpha) s_x/\sqrt{n}$,

2. For the model $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, σ^2 fixed, a $100(1-\alpha)\%$-CI for μ is $\bar{x} \mp z(\alpha)\sigma/\sqrt{n}$.

Keep in mind that $z_n(\alpha)$ takes as argument α the reciprocal of the desired confidence coefficient, i.e., for a 95%-CI, $\alpha = 0.05$ and we use $z_{n-1}(0.05)$. Table 1 gives values of $z_k(\alpha)$ for some choices of k, including the limiting case of $k = \infty$, and for confidence coefficients 90% ($\alpha = 0.1$), 95% ($\alpha = 0.05$) and 99% ($\alpha = 0.01$).

Confidence	α	$z_5(\alpha)$	$z_6(\alpha)$	$z_7(\alpha)$	$z_8(\alpha)$	$z_9(\alpha)$	$z_{10}(\alpha)$	$z_{50}(\alpha)$	$z_{100}(\alpha)$	$z(\alpha)$
90%	0.10	2.02	1.94	1.89	1.86	1.83	1.81	1.68	1.66	1.64
95%	0.05	2.57	2.45	2.36	2.31	2.26	2.23	2.01	1.98	1.96
99%	0.01	4.03	3.71	3.50	3.36	3.25	3.17	2.68	2.63	2.58

Table 1: $z_k(\alpha)$ values needed to construct $100(1-\alpha)\%$-CI.