10/10/2019

Regression & Classification:

Logistic Regression:

3 species:

<u>sp1</u> <u>SP2</u> <u>SP3</u>

$$\log \frac{P(SP1|x)}{P(SP3|x)} = x^T\beta_1.$$

$$\log \frac{P(SP2|x)}{P(SP3|x)} = x^T\beta_2.$$

$$\boxed{\log \frac{P(SP1|x)}{P(SP2|x)}}$$ $\Rightarrow$ we can get this from above two.

$$\log \frac{P(SP1|x)}{P(SP3|x)} = x^T\beta_1 \Rightarrow P(SP1|x) = P(SP3|x)\exp(x^T\beta_1)$$

$$\log \frac{P(SP2|x)}{P(SP3|x)} = x^T\beta_2$$

$$\Rightarrow P(SP2|x) = P(SP3|x)\exp(x^T\beta_2)$$

We know,

$$\boxed{P(SP1|x) + P(SP2|x) + P(SP3|x) = 1.}$$

$$\therefore P(SP3|x)\exp(x^T\beta_1) + P(SP3|x)\exp(x^T\beta_2) + P(SP3|x) = 1$$

$$\Rightarrow P(SP3|x)\left[\exp(x^T\beta_1) + \exp(x^T\beta_2) + 1\right] = 1.$$

$$\therefore P(SP3|x) = \frac{1}{1 + \sum_{i=1}^{2} \exp(x^T\beta_i)}$$

For multinomial logistic regress-
-ion, use negative log likelihood as optimizer.

In the Irish dataset, likelihood would be a trinomial distribution,

$\underset{\sim}{p} = (p_1, p_2, p_3)$, where each p_i represent probabilities for species i, $i = 1, 2, 3$.

Bootstrap Statistics:

① . $x_1, x_2, \ldots, x_n \overset{i.i.d}{\sim} N(\mu, 1)$

$$\bar{x} \sim N\left(\mu, \frac{1}{\sqrt{n}}\right).$$

② $x_1, \ldots, x_n \overset{i.i.d}{\sim} f(\ )$, such that, $E(x_i) = \mu$, $V(x_i) = 1$, $\forall i$.

What is the sampling distribution of $\bar{x}$?

$H_0: \mu = \mu_0$ vs $H_A: \mu \neq \mu_0$

C.I of μ

→ for both, we need sampling distribution of $\bar{x}$.

Then from CLT,

If $n \to \infty$ (i.e. n is large)

then $\bar{x} \overset{approx}{\sim} N\left(\mu, \frac{1}{\sqrt{n}}\right)$

<u>Note:</u> $x_1, x_2, \ldots, x_n \sim Exp(\lambda)$

$\bar{X} \sim Gamma(n\lambda, n)$

$$E(\bar{x}) = \frac{n\lambda}{n} = \lambda.$$

Let, $X_1, \ldots, X_n \overset{i.i.d}{\sim} F(\,)$, $E(X) = \mu$,

$V(X) = \sigma^2 < \infty$, $\bar{X} \sim G_n(\,)$,

<u>Define</u>: $X_1, \ldots, X_n \sim F(\,)$.

$$F(x) = P[X \leq x].$$

<u>Emperical CDF</u>:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x).$$

$$= (\# X_i \leq x)/n.$$

Let, $Y_i = I(X_i \leq x)$.

$$Y_i \overset{i.i.d}{\sim} \text{Bernoulli}(P).$$

where, $P = P(X_i \leq x)$.

Hence, we can say that,

$F_n(x)$ is MLE of $F(x)$.

Simulate $\underbrace{(x_1^*, x_2^*, \ldots, x_n^*)}_{\text{estimate } \bar{x}^*} \leftarrow F_n(x)$

$\bar{X}_1^*, \bar{X}_2^*, \ldots, \bar{X}_M^*$

Draw histogram of $\bar{x}^*$.

$$\boxed{\hat{G}_n^* \xrightarrow{\text{asymp.}} G_n}$$

$\Downarrow$

Bootstrap Process.

We are just using resampling technique with SRSWR.