Statistics Review

① Point estimation → maximum likelihood estimation

② Confidence intervals → normal case → variance known / variance not known

→ bootstrap

③ Hypothesis testing

→ normal case

→ bootstrap

→ basic concepts

Maximum likelihood estimation

$$X_1, \ldots, X_n \sim f(x \mid \theta) \quad \text{contin}$$
$$\sim p(x \mid \theta) \quad \text{discrete}$$

$$L(\theta \mid X_1, \ldots, X_n) \quad \text{discrete}$$

$$= \prod_{i=1}^{n} p(x_i \mid \theta)$$

contin $$= \prod_{i=1}^{n} f(x_i \mid \theta)$$

$i=1$

$\Rightarrow$ maximum likelihood estimate

choose the value of θ

that maximizes $L(\theta)$

in general, we maximize the

log likelihood

$$\ell(\theta) = \log(L(\theta))$$

$X_1, \dots, X_n \sim \text{Poiss}(\lambda)$

Poiss dist'n

$$p(i) = \frac{e^{-\lambda}\lambda^{i}}{i!}$$

$$p(0) = e^{-\lambda}$$

$$p(1) = e^{-\lambda}\lambda$$

$$L(\lambda \mid X_1, \dots, X_n) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{X_i}}{X_i!}$$

$$= e^{-n\lambda}\lambda^{\sum X_i} \cdot \prod^{n} \frac{1}{X_i!}$$

$i=1$

$$\ell(\lambda \mid X_1, \dots, X_n)$$

$$= -n\lambda + \sum x_i \cdot \log(\lambda)$$

$$- \sum_{i=1}^{n} \log(x_i!)$$

$$\frac{\partial \ell}{\partial \lambda} = -n + \frac{\sum x_i}{\lambda} = 0$$

$$n\lambda = \sum x_i$$

$$\hat{\lambda} = \frac{1}{n} \sum x_i = \bar{X}$$

$$X_1, \dots, X_n \sim N(\mu, 1)$$

$$f(x \mid \mu) = \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right)}$$

$$L(\mu \mid X_1, \dots, X_n) =$$

$$\frac{1}{(2\pi)^{n/2}} \cdot \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2\right)$$

$$\ell(\mu \mid x_1, \dots, x_n) =$$

$$-\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial \ell}{\partial \mu} = \sum (x_i - \mu) \overset{\text{set}}{=} 0$$

$$\sum x_i - n\mu = 0$$

$$\hat{\mu} = \frac{1}{n}\sum x_i$$

Estimation

θ, $\hat{\theta}$

$E[\hat{\theta}] = 0$ then it is unbiased

$E[\hat{\theta}] \neq 0$ then it is said to be biased

$$X_1, \dots, X_n \sim N(\mu, \sigma)$$

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

$$\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$E[\hat{\sigma}^2] = \sigma^2$$

$$E[\hat{\sigma}^2_{MLE}] = \frac{n-1}{n} \sigma^2$$

bias of $\hat{\sigma}^2_{MLE}$

is $\frac{n-1}{n} \sigma^2 - \sigma^2$.

$$f(x|\alpha) = \begin{cases} \alpha x^{\alpha-1}, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$$L(\alpha | X_1, \ldots, X_n) = \alpha^n \prod_{i=1}^{n} X_i^{\alpha-1}$$

$$\ell(\alpha | X_1, \ldots, X_n) = n \log(\alpha) + (\alpha - 1) \sum_{i=1}^{n} \log(X_i)$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log(x_i) \overset{\text{set}}{=} 0$$

$$\hat{\alpha} = -\frac{n}{\sum_{i=1}^{n} \log(x_i)}$$

Confidence intervals

↳ statistic you calculate from a sample that tries to quantify a reasonable range of values for the parameter

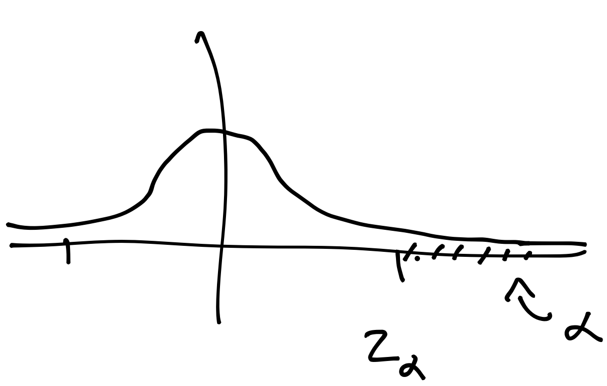
Case 1: $X_1, \dots, X_n \sim N(\mu, \sigma)$

σ known

$(1-\alpha)100\%$ two-sided confidence interval for

μ

$$\bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$



$$Z_{1-\alpha} = -Z_\alpha$$

one-sided upper confidence interval

$$\left(\bar{x} - z_\alpha \cdot \frac{\sigma}{\sqrt{n}},\ \infty\right)$$

$$z_\alpha = \text{scipy.stats.norm.ppf}(1-\alpha)$$
$$= \text{scipy.stats.norm.isf}(\alpha)$$

$X_1, \dots, X_n \sim N(\mu, \sigma)$

σ unknown

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

$$S = \sqrt{S^2}$$

$$\bar{X} \pm t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}$$

$$t_{\alpha, n} = \text{scipy.stats.t.ppf}(1-\alpha, n)$$

Bootstrapping

$\hat{\theta} = f(X$

$X_1, \dots, X_n$ estimate for a parameter θ ...

bootstrap sample is a re-sampling of $X_1, \dots, X_n$ with replacement

m bootstrap samples from $X_1, \dots, X_n$

np.random.choice(X, replace=True, size=(m,n))

calculate theta_hat on each sample

→ m theta_hats

→ sample percentiles to calculate my confidence interval

Hypothesis testing

- see whether you have sufficient evidence to reject H_0
- some critical region C,

test statistic X

reject H_0 if $X \in C$

- the level of a test is the $P_{H_0}(X \in C)$

- generally speaking, we specify α
  choose C s.t. $P_{H_0}(X \in C_\alpha) = \alpha$

- p-value is the smallest level α
  under which we would reject
  H_0

$X_1, \ldots, X_n \sim N(\mu, \sigma)$ σ known

given α, test statistic $\bar{X}$

$H_0: \mu = \mu_0$

$H_A: \mu \neq \mu_0$

critical region is

$$|\bar{X} - \mu_0| \quad z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \}$$

$$|\bar{X} - \mu_0| > z_{\alpha/2} \cdot \sigma\sqrt{n}$$

$|\bar{X}-\mu_0| < -z_{\alpha/2} \cdot \sigma/\sqrt{n}$

p-value: $2\left(1-\Phi\left(\frac{|\bar{X}-\mu_0|}{\sigma/\sqrt{n}}\right)\right)$

If I don't know σ, then need to estimate from data

$$|\bar{X}| > t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}$$

Φ is standard normal

scipy.stats.norm.cdf(...)

→ calculate a $(1-\alpha)100\%$ confidence interval for μ and reject H_0 if confidence interval does not contain 0

→ bootstrap confidence intervals

reject H_0: ~~θ~~ $= k$ if k is not in my bootstrap confidence

$H_0: \theta = k$

$H_A: \theta > k$

If my data

n

$\bar{X}$

s

α

Linear Regression

$(x_1, Y_1), \ldots, (X_n, Y_n)$

$$y_i = \alpha + \beta x_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

Ridge Regression:

add a penalty to the likelihood of α & B of the form $\lambda \sum \beta_i^2$

decision tree - uses series of binary decisions to partition up the observations in your data set