Modern Statistics **Lecture 10 - 03/25/2025**

Simple Linear Regression

Lecturer:Xiangyu Chang *Scribe: Rui He,Xuqing Zhao*

Edited by: Zhihong Liu

1 Recall

- MLE (Maximum Likelihood Estimation): Given independent and identically distributed (i.i.d.) samples $X_1, X_2, \ldots, X_n$ from a distribution with parameter θ, the MLE of θ is $\hat{\theta}_n = \arg\max_\theta \prod_{i=1}^n f_\theta(X_i)$. Under regular conditions, $\hat{\theta}_n$ converges in probability to θ^* (WLLN) and converges in distribution to $N(\theta^*, I^{-1}(\theta^*))$ (CLT).
- Score function: $s_\theta(x) = \nabla \log f_\theta(x)$.
- Fisher Information:
  1. $I(\theta) = \mathbb{V}(s_\theta(X))$.
  2. $I_n(\theta) = \mathbb{V}(\nabla \ell_n(\theta))$.
- Properties:
  1. $I_n(\theta) = \sum_{i=1}^n \mathbb{V}(s_\theta(X_i)) = nI(\theta)$.
  2. $\mathbb{E}[s_\theta(X)] = 0$.
  3. $I(\theta) = -\mathbb{E}[\nabla^2 \log f_\theta(X)]$.
  4. Combining 1 and 3, $I_n(\theta) = -\mathbb{E}[\nabla^2 \ell_n(\theta)]$.
- Theorem: $\frac{\hat{\theta}_n - \theta^*}{\hat{se}} \xrightarrow{d} N(0,1)$, where $\hat{se} = \frac{1}{\sqrt{I_n(\hat{\theta}_n)}}$.
- Example: Bernoulli distribution $Ber(p)$.
  - $\hat{p}_n = \bar{X}_n$.
  - $f_p(x) = p^x(1-p)^{1-x}$.
  - $s_p(x) = \nabla \log f_p(x) = \frac{x}{p} - \frac{1-x}{1-p}$.
  - $I(p) = -\mathbb{E}[s_p'(X)] = \frac{1}{p(1-p)}$.
  - $\frac{\hat{p}_n - p}{\hat{se}} \xrightarrow{d} N(0,1)$, where $\hat{se} = \frac{1}{\sqrt{nI(\hat{p}_n)}} = \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$.

2 Hypothesis Testing

- Example: A/B Testing
  - CTR (Click-Through Rate): $R_A = 7.3\%$, $R_B = 7.29\%$.
  - Hypotheses:
    * $H_0 : R_A = R_B$.
    * $H_1 : R_A \neq R_B$.
  - Test Statistic: $T(X) = |R_A - R_B|$.
  - Rejection Region: $T(X) \in C_n = \{T(X) > c\}$.
  - Confidence Interval: $C_n = [\hat{\theta}_n - z_{\frac{\alpha}{2}} \hat{se}, \hat{\theta}_n + z_{\frac{\alpha}{2}} \hat{se}]$ is an approximate $1-\alpha$ confidence interval.
  - $P(\theta^* \in C_n) \to 1 - \alpha$.

3 Simple Linear Regression

Regression is a method for studying the relationship between a response variable Y and a covariate X. One way to summarize the relationship between X and Y is through the regression function.

The simplest version of regression is Simple Linear Regression when X_i is simple (one-dimensional) and $r(x)$ is assumed to be linear.

Data: $\{(x_i, y_i)\}_{i=1}^n$ i.i.d. from $F_{X,Y}$.

Model: $Y = r(X) + \varepsilon$ (prediction).

Variables:

- X: covariance, feature, independent variable.
- Y: response, dependent variable.

MSE (Mean Squared Error): $\min_r \mathbb{E}[(Y - r(X))^2]$.

$$\begin{aligned}
\mathbb{E}[(Y - r(X))^2] &= \mathbb{E}[[Y - \mathbb{E}[Y|X]^2 + \mathbb{E}[Y|X]^2 - r(X)]^2] && (1)\\
&= \mathbb{E}[[Y - \mathbb{E}[Y|X]]^2] + \mathbb{E}[\mathbb{E}[Y|X] - r(X)]^2] + 2\mathbb{E}[[Y - \mathbb{E}[Y|X]]^2 \mathbb{E}[Y|X] - r(X)]^2] && (2)\\
&= \mathbb{E}[[Y - \mathbb{E}[Y|X]]^2] + \mathbb{E}[\mathbb{E}[Y|X] - r(X)]^2], && (3)
\end{aligned}$$

where $2\mathbb{E}[[Y - \mathbb{E}[Y|X]]^2 \mathbb{E}[Y|X] - r(X)]^2] = 0$. (the process of proof is omitted.)

Since the quantity $\mathbb{E}\big[(Y - \mathbb{E}[Y|X])^2\big]$ is independent of the regression function $r(x)$, the minimum value of the expected squared loss $\mathbb{E}\big[(Y - r(X))^2\big]$ is attained *if and only if* $r(X) = \mathbb{E}[Y|X]$.

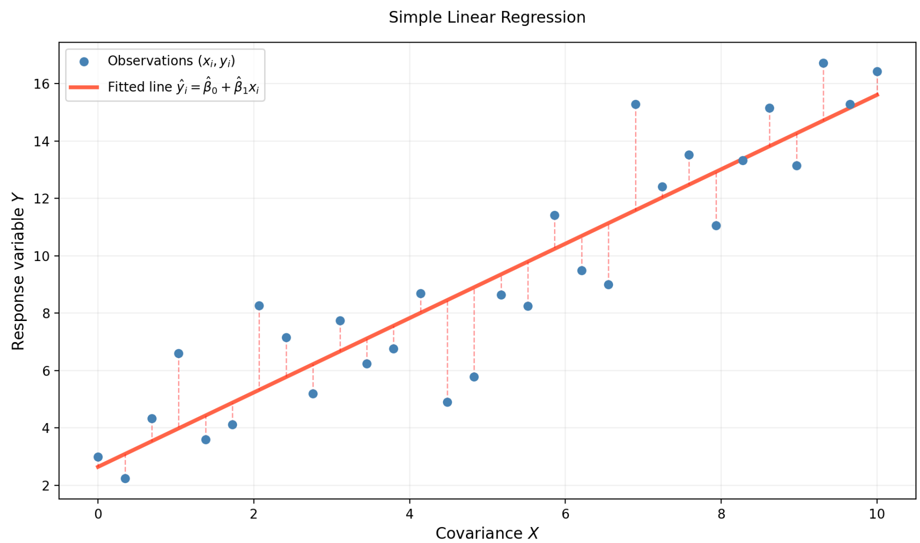

图 1: an Example of SLR

Consequently, the optimal regression function reduces to $r^*(x) = \mathbb{E}[Y|X = x]$, indicating that the functional relationship is fully characterized by the conditional expectation.

Regression Function: $r^*(x) = \mathbb{E}[Y|X = x]$.

Simple Linear Regression:

def: $r(x) \triangleq \beta_0 + \beta_1 x$,

$$Y = r(x) + \epsilon = \beta_0 + \beta_1 X + \epsilon.$$

where $\mathbb{E}[\varepsilon_i|X_i] = 0$, $\mathbb{V}(\varepsilon_i|X_i) = \sigma^2$.

I'd like to show a figure to make it clear:

$$\begin{aligned} \text{Data Generating Process:} \quad & y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2). \\ \text{Estimated Regression Function:} \quad & \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i. \end{aligned}$$

To find the $\hat{\beta}_0$ and $\hat{\beta}_1$,we need the method OLS.

OLS(Ordinary Least Squares):

$$\hat{\beta}_{\text{OLS}} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 .$$

LAD(the Least Absolute Deviations):

$$\hat{\beta}_{\text{LAD}} = \arg\min_{\beta} \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| .$$

we always prefer LS to LAD because LS is more convienient.

Derivation of Ordinary Least Squares Estimators

First-Order Conditions

$$\frac{\partial S}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i,$$
$$\frac{\partial S}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} x_i y_i = \beta_0 \sum x_i + \beta_1 \sum_{i=1}^{n} x_i^2.$$

Solving the Normal Equations:

From the first equation:

$$\bar{y} = \beta_0 + \beta_1 \bar{x}_n \quad \Rightarrow \quad \beta_0 = \bar{y} - \beta_1 \bar{x}_n.$$

Substitute β_0 into the second equation and simplify:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}.$$

The ordinary least squares (OLS) slope estimator is derived from the following sample moments:

$$S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \quad \text{(Sample variance of } X\text{)}.$$
$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n) \quad \text{(Sample covariance)}.$$

Consequently, the OLS estimator for the slope parameter is:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}.$$

Final Estimators:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_n.$$

Theorem 1 *We have:*

$$\mathbb{E}(\hat{\beta}_0) = \beta_0,$$
$$\mathbb{E}(\hat{\beta}_1) = \beta_1,$$
$$\mathbb{V}(\hat{\beta}_0) = \left(\frac{1}{n} + \frac{\bar{x}_n^2}{nS_{xx}}\right)\sigma^2,$$
$$\mathbb{V}(\hat{\beta}_1) = \frac{\sigma^2}{nS_{xx}}.$$

Proof 1 •

$$\mathbb{E}(\hat{\beta}_0) = \mathbb{E}[\frac{S_{xy}}{S_{xx}}] = \frac{\mathbb{E}[S_{xy}]}{S_{xx}}.$$

$$\begin{aligned}
\mathbb{E}[S_{xy}] &= \frac{1}{n}\mathbb{E}[\sum_i (x_i - \bar{x}_n)(y_i - \bar{y}_n)] \\
&= \frac{1}{n}\mathbb{E}[\sum_i (x_i - \bar{x}_n)y_i](where \sum_i (x_i - \bar{x}_n)\bar{y}_n = 0) \\
&= \frac{1}{n}\sum_i (x_i - \bar{x}_n)\mathbb{E}(y_i) \\
&= \frac{1}{n}\sum_i (x_i - \bar{x}_n)(\beta_0 + \beta_1 x_i) \\
&= \frac{\beta_1}{n}\sum_i (x_i - \bar{x}_n)x_i(where \sum_i (x_i - \bar{x}_n)\beta_0 = 0) \\
&= \frac{\beta_1}{n}\sum_i (x_i - \bar{x}_n)(x_i - \bar{x}_n)(where \sum_i (x_i - \bar{x}_n)\bar{x}_n = 0) \\
&= \beta_1 S_{xx}.
\end{aligned}$$

Consequently,

$$\mathbb{E}(\hat{\beta}_1) = \frac{\mathbb{E}[S_{xy}]}{S_{xx}} = \frac{\beta_1 S_{xx}}{S_{xx}} = \beta_1.$$

•

$$\begin{aligned}
\mathbb{V}(\hat{\beta}_1) &= \frac{\mathbb{V}(S_{xy})}{S_{xx}^2} \\
&= \frac{\sum_i (x_i - \bar{x}_n)^2 \mathbb{V}(y_i)}{n^2 S_{xx}^2} \\
&= \frac{\sigma^2 \sum_i (x_i - \bar{x}_n)^2}{n^2 S_{xx}^2} \\
&= \frac{\sigma^2}{n S_{xx}}.
\end{aligned}$$

•

$$\begin{aligned}
\mathbb{E}(\hat{\beta}_0) = \mathbb{E}[\bar{y}_n - \hat{\beta}_1 \bar{x}_n] &= \mathbb{E}[\bar{y}_n] - \bar{x}_n \mathbb{E}(\hat{\beta}_1) \\
&= \beta_1 \bar{x}_n + \beta_0 - \beta_1 \bar{x}_n \\
&= \beta_0.
\end{aligned}$$

•

$$\begin{aligned}
\mathbb{V}(\hat{\beta}_0) &= \mathbb{V}(\bar{y}_n - \beta_1 \bar{x}_n) \\
&= \mathbb{V}(\hat{y}_n) + \bar{x}_n^2 \mathbb{V}(\hat{\beta}_1)(where 2Cov(\bar{y}_n, \beta_1 \bar{x}_n) = 0) \\
&= \frac{\sigma^2}{n} + \bar{x}_n^2 \frac{\sigma^2}{n S_{xx}}.
\end{aligned}$$

The Proof process of $Cov(\bar{y}_n, \beta_1\bar{x}_n) = 0$*:*

$$
\begin{aligned}
Cov(\bar{y}_n, \beta_1\bar{x}_n) &= Cov(\frac{1}{n}\sum_i y_i, \frac{1}{n}\sum_i (x_i - \bar{x}_n)y_i) \\
&= \frac{1}{n^2}Cov(\sum_i y_i, \sum_i (x_i - \bar{x}_n)y_i) \\
&= \sum_i \sum_j (x_i - \bar{x}_n)Cov(y_i, y_j) \\
&= \frac{1}{n^2}\sum_i (x_i - \bar{x}_n)\sigma^2 \\
&= 0.
\end{aligned}
$$

Supplement: Inference of the noise $\hat{\sigma^2}$

$$\hat{\epsilon_i} = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i (i = 1, 2, ..., n).$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\epsilon_i} - \bar{\epsilon}_n)^2.$$

Theorem 2 $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, $\mathbb{E}(\hat{\sigma}^2) = \hat{\sigma}^2$.