Prediction

Table of contents

Prediction 1
Bivariate Normal Distribution . . . 1

Prediction via Minimum MSE 3
Best Linear Predictor . . . 4
Specialization to the Bivariate Normal . . . 4
R Example . . . 5

Prediction

Bivariate Normal Distribution

Let (X, Y) be bivariate normal.

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\}.$$

$-\infty < \mu_X < \infty$ and $-\infty < \mu_Y < \infty$.

$0 < \sigma_X, \sigma_Y$ and $-1 < \rho < 1$.

Marginals

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$$

$$Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2).$$

Exercise: Show

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\Big\{-\frac{1}{2}\Big(\frac{x-\mu_X}{\sigma_X}\Big)^2\Big\}$$

Exercise: Show

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\Big\{-\frac{1}{2}\Big(\frac{y-\mu_Y}{\sigma_Y}\Big)^2\Big\}.$$

Conditionals

$$Y \mid X = x \sim \mathcal{N}\Big(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X),\ \sigma_Y^2(1-\rho^2)\Big).$$

$$f_{Y|X}(y \mid x) = \frac{1}{\sigma_Y\sqrt{1-\rho^2}} \exp\Big\{-\frac{1}{2}\frac{\big(y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\big)^2}{\sigma_Y^2(1-\rho^2)}\Big\}.$$

$$\mathrm{E}[Y \mid X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X).$$

$$\mathrm{Var}(Y \mid X = x) = \sigma_Y^2(1-\rho^2).$$

Covariance and Correlation

$$\mathrm{Cov}(X,Y) = \rho\sigma_X\sigma_Y.$$

$$\mathrm{Corr}(X,Y) = \rho.$$

Note: For a bivariate normal, X and Y are independent if and only if $\rho = 0$. $$

Prediction via Minimum MSE

A random variable is observed and then based on the observed value, we predict a second random variable.

Let $g(X)$ denote the predictor function.

Choose a function $g(X)$ so it tends to be close to Y

Criteria:

Goal: minimizing MSE, i.e., minimze

$$\mathrm{E}[(Y - g(X))^2].$$

Claim: The best predictor is

$$g(x) = \mathrm{E}[Y \mid X = x].$$

Proof:

Condition on X and then take $E[\ \ ]$.

$$\mathrm{E}[(Y - g(X))^2 \mid X] = \mathrm{E}[(Y - \mathrm{E}[Y \mid X] + \mathrm{E}[Y \mid X] - g(X))^2 \mid X]$$

$$= \mathrm{E}[(Y - \mathrm{E}[Y \mid X])^2 \mid X] + \mathrm{E}[(\mathrm{E}[Y \mid X] - g(X))^2 | X] + 2\,\mathrm{E}[(Y - \mathrm{E}[Y \mid X])(\mathrm{E}[Y \mid X] - g(X)) \mid X].$$

Note: $\mathrm{E}[Y \mid X] - g(X)$ given X can be treated as a contant.

$$\mathrm{E}[Y - \mathrm{E}[Y \mid X] \mid X] = 0 \Rightarrow \text{cross term} = 0.$$

$$\Rightarrow \mathrm{E}[(Y - g(X))^2 \mid X] \geq \mathrm{E}[(Y - \mathrm{E}[Y \mid X])^2 \mid X].$$

$$\Rightarrow \mathrm{E}[(Y - g(X))^2] \geq \mathrm{E}[(Y - \mathrm{E}[Y \mid X])^2].$$

Exercise: Show

$$\mathrm{E}[(X - a)^2] \text{ is minimized at } a = \mathrm{E}[X].$$

Best Linear Predictor

It sometimes happens that the joint density of X and Y is not completely known, or $E[Y|X]$ is difficult to calculate. However, $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \sigma_{XY}$ are known, then we can at least determine the **Best Linear Predictor** of Y on X..

Assume

$$g(X) = a + bX \text{ and minimize } \mathrm{E}[(Y - (a + bX))^2].$$

Answer:

$$\frac{\partial}{\partial a}\,\mathrm{E}[(Y - (a + bX))^2] = -2\,\mathrm{E}[Y] + 2a + 2b\,\mathrm{E}[X] = 0.$$

$$\frac{\partial}{\partial b}\,\mathrm{E}[(Y - (a + bX))^2] = -2\,\mathrm{E}[XY] + 2a\,\mathrm{E}[X] + 2b\,\mathrm{E}[X^2] = 0.$$

$$b = \frac{\mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]}{\mathrm{E}[X^2] - (\mathrm{E}[X])^2} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} = \rho\frac{\sigma_Y}{\sigma_X}.$$

$$a = \mathrm{E}[Y] - b\,\mathrm{E}[X] = \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X.$$

$$g(x) = a + bx = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

$$(\text{intercept } = \mu_Y - \rho\tfrac{\sigma_Y}{\sigma_X}\mu_X,\ \text{slope } = \rho\tfrac{\sigma_Y}{\sigma_X}).$$

Specialization to the Bivariate Normal

$$g(x) = \mathrm{E}[Y \mid X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

$$\mathrm{MSE}(g) = \mathrm{E}[(Y - \mu_Y - \rho\tfrac{\sigma_Y}{\sigma_X}(X - \mu_X))^2] = \sigma_Y^2(1 - \rho^2).$$

Note: For (X, Y) Binvariate Notemal, the Best Predictor is the Best Linear Predictor $E[Y|X]$.

Note: For (X, Y) Binvariate Notemal, the Best Linear Predictor has MSE $V(Y|X)$.

Note: Thus, higher ρ, then lower prediction error.

R Example

```r
set.seed(1)
n <- 2000
rho <- 0.6
muX <- 0; muY <- 1
sigX <- 2; sigY <- 3
Sigma <- matrix(c(sigX^2, rho*sigX*sigY,
                  rho*sigX*sigY, sigY^2), 2, 2)
Z <- MASS::mvrnorm(n, mu = c(muX, muY), Sigma = Sigma)
X <- Z[,1]; Y <- Z[,2]

b <- rho * sigY / sigX
a <- muY - b * muX
gX <- a + b * X

c(mean_mse = mean((Y - gX)^2),
  theory = sigY^2 * (1 - rho^2))
```

```
mean_mse   theory
6.372515 5.760000
```