Slot machine game:

(1) Pull arm of slot machine

(2) Machine draws payout Y from Bernoulli distribution

(3) Receive random payout Y (either \$0 or \$1)

Bernoulli distribution: $\Omega = \{0, 1\}$ success probability

$$P(Y=1 \mid \theta) = \theta \qquad P(Y=0 \mid \theta) = 1 - \theta$$

cost to play : \$0.50

Expected payout: \$0.50 or \$0.60

$Y_1, Y_2, \dots, Y_n \overset{i.i.d.}{\sim} P(y \mid \theta)$ — independently & identically distributed

Bernoulli r.v.

$\theta \in \{\theta_1, \theta_2\}$

$\theta_1 = 0.5$, $\theta_2 = 0.6$

$D = (Y_1, Y_2, \dots, Y_n)$

Likelihood function

$P(D \mid \theta)$

$D = (1 \quad 0 \quad 0 \quad 1)$

$$P(Y_1 = 1, Y_2 = 0, Y_3 = 0, Y_4 = 1 \mid \theta)$$

$$= P(Y_1 = 1 \mid \theta)\, P(Y_2 = 0 \mid \theta)\, P(Y_3 = 0 \mid \theta)\, P(Y_4 = 1 \mid \theta)$$

$$= \theta \cdot (1-\theta) \cdot (1-\theta) \cdot \theta$$

Likelihood Function

$$P(D \mid \theta) = P(Y_1 = y_1, \ldots, Y_n = y_n \mid \theta)$$

$$= \prod_{j=1}^{n} P(Y_j = y_j \mid \theta)$$

$$= \prod_{\substack{j=1 \\ y_j = 1}}^{n} \theta \cdot \prod_{\substack{j=1 \\ y_j = 0}}^{n} (1-\theta)$$

$n_1 = \#1\text{'s}$

$n_0 = \#0\text{'s}$

$$= \theta^{n_1} \cdot (1-\theta)^{n_0}$$

Maximum Likelihood Estimation (MLE)

$$\hat{\theta}_{MLE} = \underset{\theta \in \{\theta_1, \theta_2\}}{\operatorname{argmax}} P(D \mid \theta)$$

$$P(D \mid \theta) = \theta^{n_1}(1-\theta)^{n_0}$$

Maximum Likelihood Estimation (MLE)

$$\hat{\theta}_{MLE} = \underset{\theta \in \{\theta_1, \theta_2\}}{\operatorname{argmax}} \; P(D \mid \theta)$$

$$D = (1 \ 0 \ 0 \ 1)$$

$$P(D \mid \overset{0.5}{\theta_1}) = 0.5^2 \cdot 0.5^2 = 0.0625 \leftarrow \hat{\theta}_{MLE}$$

$$P(D \mid \overset{0.6}{\theta_2}) = 0.6^2 \cdot 0.4^2 = 0.0566$$

Estimator $\hat{\theta}$ is an estimate for unknown parameter θ

$\uparrow$ depends on data $= D =$ training set

An estimator $\hat{\theta}$ is unbiased if $\mathbb{E}[\hat{\theta}] = \theta$

Example: sample mean: $\frac{1}{n}\sum_{j=1}^{n} Y_j =: \hat{\theta}$ is unbiased estimator of θ

$$\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n} Y_j\right] = \frac{1}{n}\sum_{j=1}^{n} \overbrace{\mathbb{E}[Y_j]}^{\theta} = \theta$$

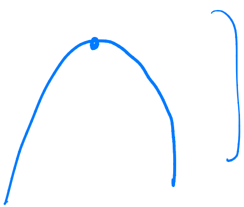
suppose all we know is that $\theta \in [0,1]$ $\quad 0 \le \theta \le 1$

What is MLE $\hat{\theta}_{MLE}$?

$$\underset{\theta \in [0,1]}{\arg\max}\ P(D \mid \theta) = \underset{\theta \in [0,1]}{\arg\max}\ \theta^{n_1}(1-\theta)^{n_0}$$

$$= \underset{\theta \in [0,1]}{\arg\max}\ \log\left(\theta^{n_1}(1-\theta)^{n_0}\right)$$

$$= \underset{\theta \in [0,1]}{\arg\max}\ \underbrace{n_1 \log \theta + n_0 \log(1-\theta)}_{g(\theta)}$$



$g(\theta)$

$$g'(\theta) = 0 \xrightarrow{\text{solve}} \theta = \frac{n_1}{n_0 + n_1} = \frac{n_1}{n}$$

Log likelihood

$$\log P(D \mid \theta) = \log \left(\prod_{j=1}^{n} P(Y_j = y_j \mid \theta) \right)$$

$$= \sum_{j=1}^{n} \log P(Y_j = y_j \mid \theta)$$

Loss functions

0-1 loss : $\mathcal{Y} = \{0,1\}$ or $\mathcal{Y} = \{-1,+1\}$

$$\ell_{0-1}(y, \hat{y}) = 1[y \neq \hat{y}]$$

squared loss: $\mathcal{Y} = \mathbb{R}$ $\quad \ell_{sq.}(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$

hinge loss: $\mathcal{Y} = \{-1,+1\}$ $\quad \ell_{hinge}(y, \underset{\in \mathbb{R}}{\underset{\uparrow}{\hat{y}}}) = \max\{0, 1 - y\hat{y}\}$

cross-entropy loss: $\ell : \{0,1\} \times [0,1] \rightarrow \mathbb{R}$

$$\ell(y, \hat{\theta}) = -y \log \hat{\theta} - (1-y) \log (1-\hat{\theta})$$

cross-entropy loss: $\ell : \{0,1\} \times [0,1] \to \mathbb{R}$

$$\ell(y, \hat{\theta}) = -y \log \hat{\theta} - (1-y) \log (1-\hat{\theta})$$

suppose $y = 1$

suppose $\hat{\theta} = 1$ $\quad \ell(1,1) = -1 \underbrace{\log(1)}_{0} - \underbrace{0 \log 0}_{0} = 0$

suppose $\hat{\theta} = 0$ $\quad \ell(1,0) = \underbrace{-1 \cdot \log 0}_{+\infty} - \underbrace{0 \log (1-0)}_{0} = +\infty$

$$\max_{\theta} \log p(y_1, \dots, y_n \mid \theta) \overset{\text{(equiv.)}}{\equiv} \min_{\theta} -\log p(y_1, \dots, y_n \mid \theta)$$

$$
\begin{aligned}
-\log \prod_{i=1}^{n} p(y_i \mid \theta) &= \sum_{i=1}^{n} -\log p(y_i \mid \theta) \\
&= \sum_{i=1}^{n} -\log\left(\theta^{y_i} (1-\theta)^{(1-y_i)}\right) \\
&= \sum_{i=1}^{n} \left(-y_i \log \theta - (1-y_i) \log(1-\theta)\right) \\
&= \sum_{i=1}^{n} \ell(y_i, \theta)
\end{aligned}
$$