삼성전기 **AI**전문가 양성과정 - 프로젝트 실습 (비영상)
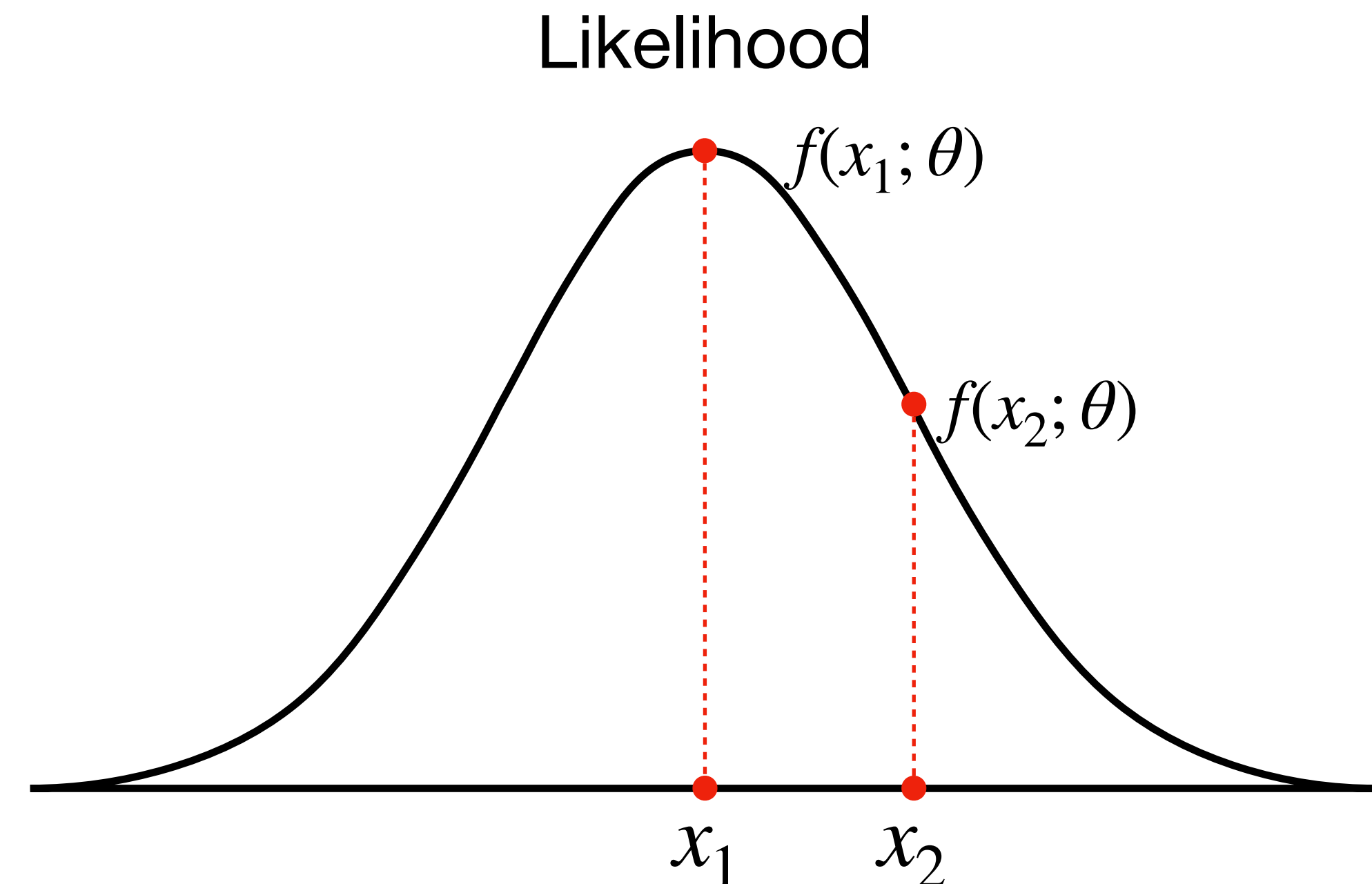
자연어처리를 위한
Negative Log Likelihood

현청천

2022.02.28
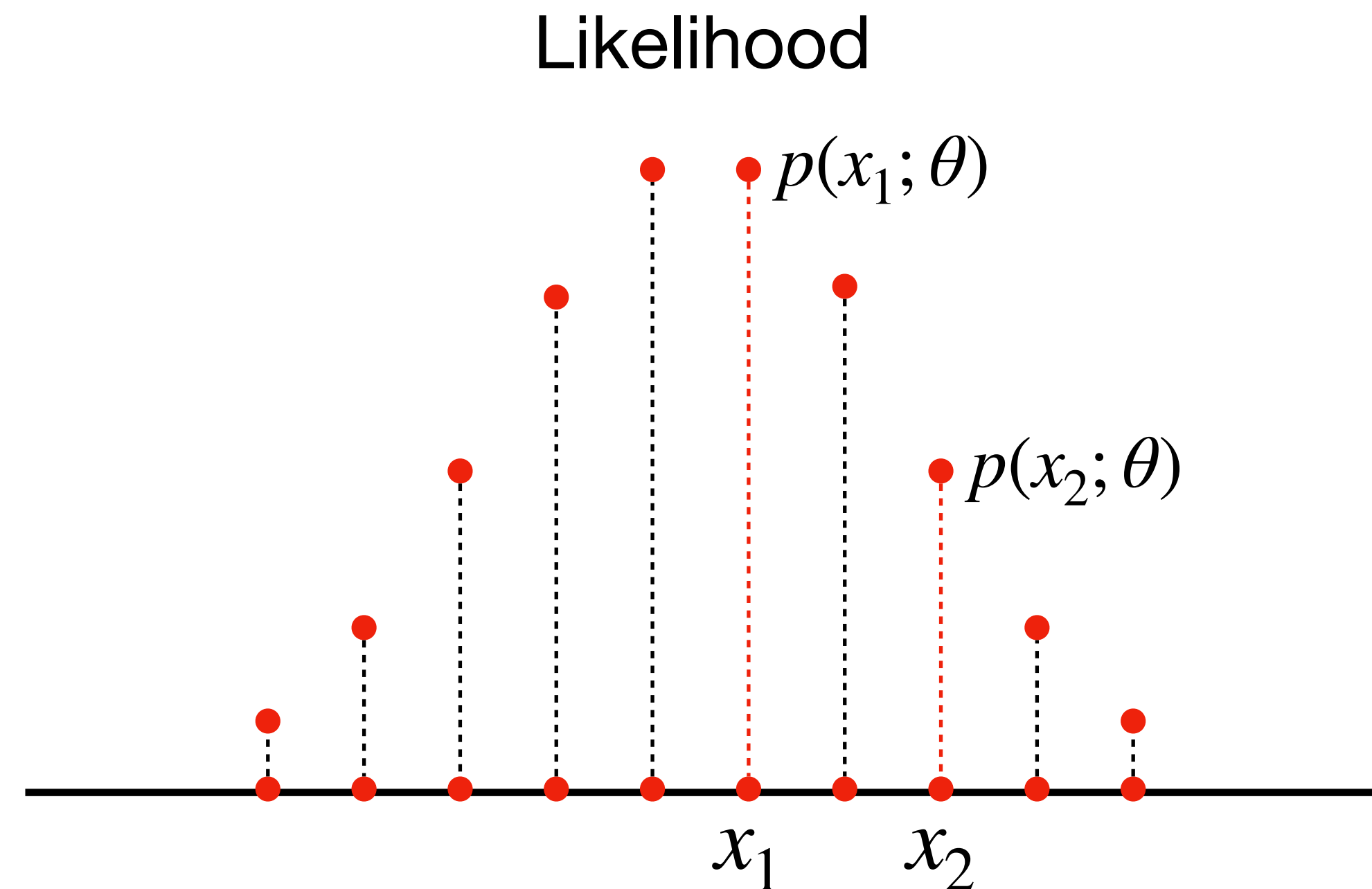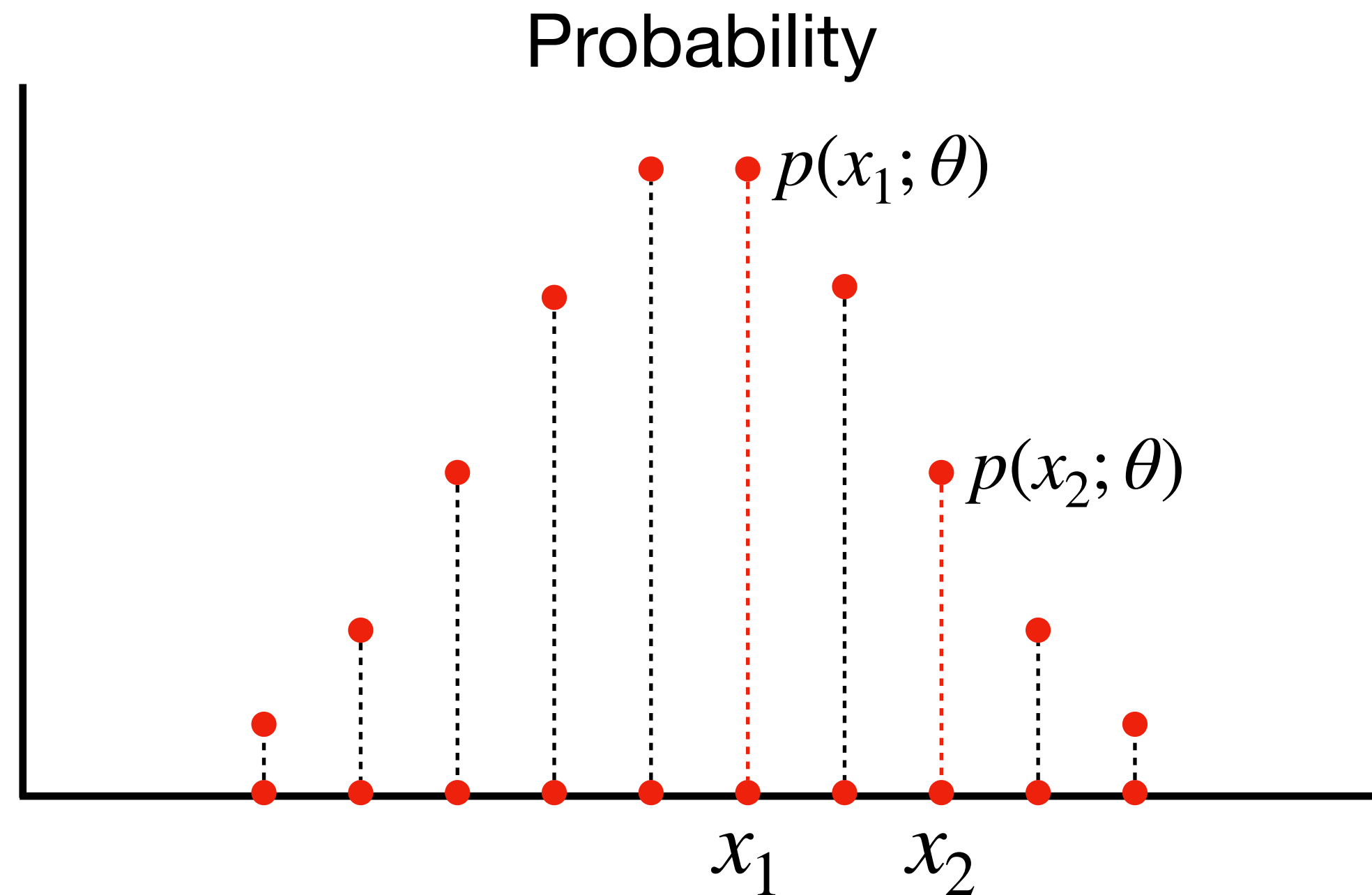
Likelihood (연속확률분포)

Probability

$p(x_1 \leq x \leq x_2; \theta)$

x_1 x_2

Likelihood

$f(x_1; \theta)$

$f(x_2; \theta)$
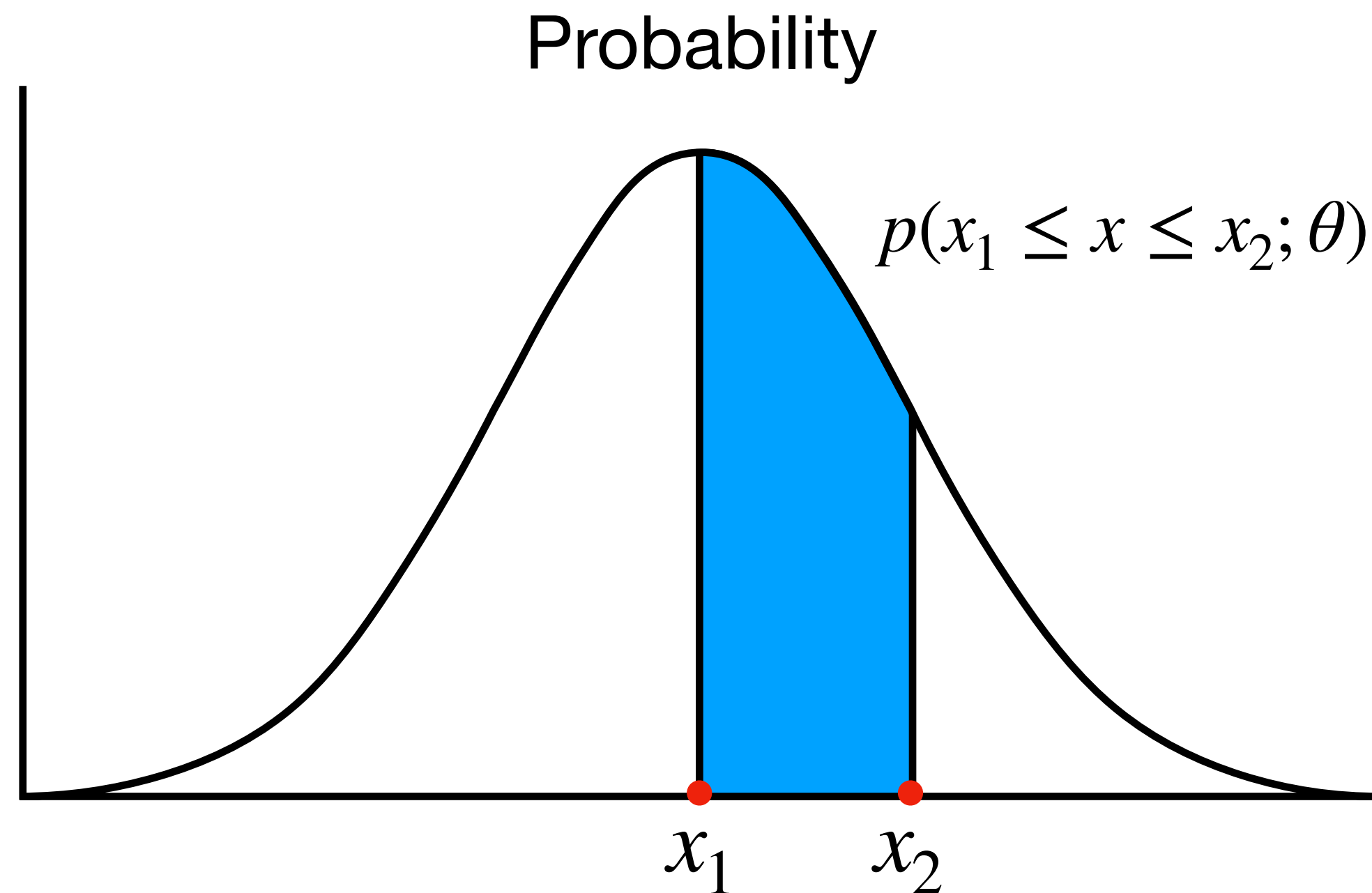
x_1 x_2

$$\mathcal{L}(\theta \,|\, x) = f(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{N} f(x_i; \theta)$$

가능도 (특정 사건들이 일어날 가능성)

Likelihood (이산확률분포)

Probability

$p(x_1; \theta)$

$p(x_2; \theta)$

x_1 x_2

Likelihood

$p(x_1; \theta)$

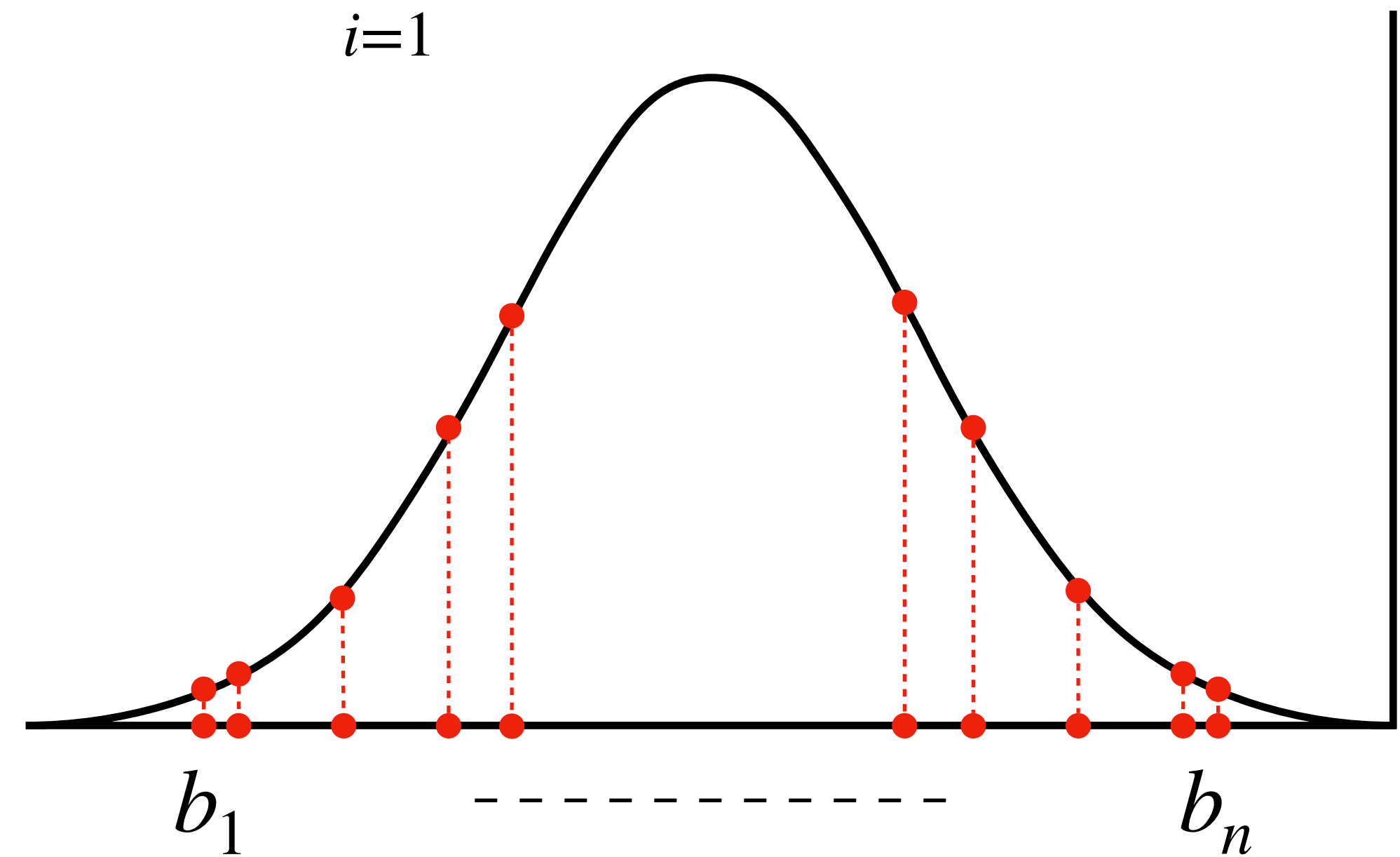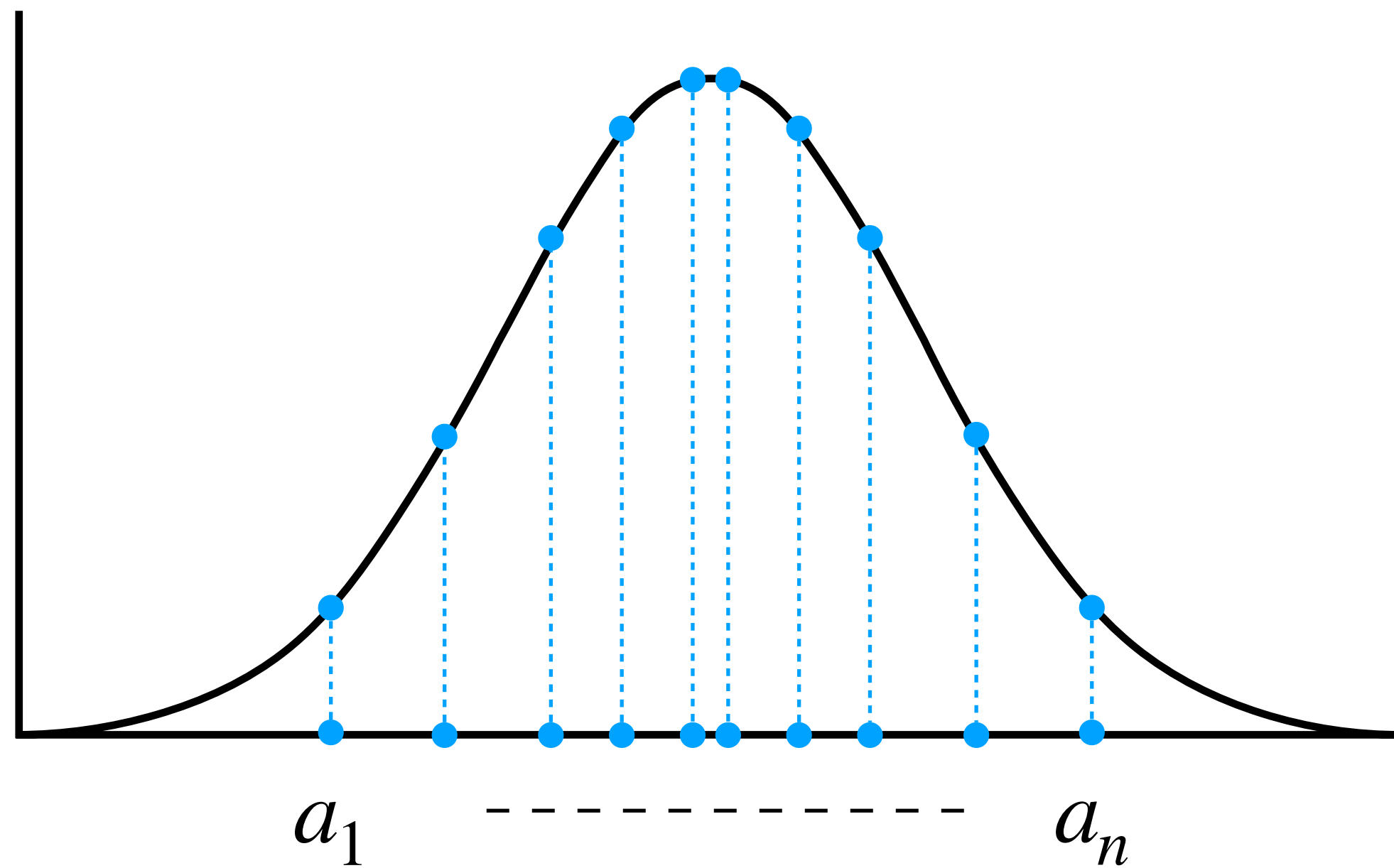$p(x_2; \theta)$

x_1 x_2

$$\mathscr{L}(\theta \mid x) = p(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{N} p(x_i; \theta)$$

가능도 (특정 사건들이 일어날 가능성)

Likelihood

$$\mathscr{L}(\theta \,|\, x) = f(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{N} f(x_i; \theta)$$

$a_1 \quad - - - - - - - - - - \quad a_n$

$b_1 \quad - - - - - - - - - - \quad b_n$

$$\mathscr{L}(\theta \,|\, a) > \mathscr{L}(\theta \,|\, b)$$

Maximum Likelihood Estimation

키
164.95
165.35
165.76
166.16
167.78
168.99
169.80
170.20
171.82
172.63

사건으로부터 확률분포를 예측

Maximum Likelihood Estimation

$$p(X; \theta_1)$$

키
164.95
165.35
165.76
166.16
167.78
168.99
169.80
170.20
171.82
172.63

Maximum Likelihood Estimation

$$p(X : \theta_2)$$
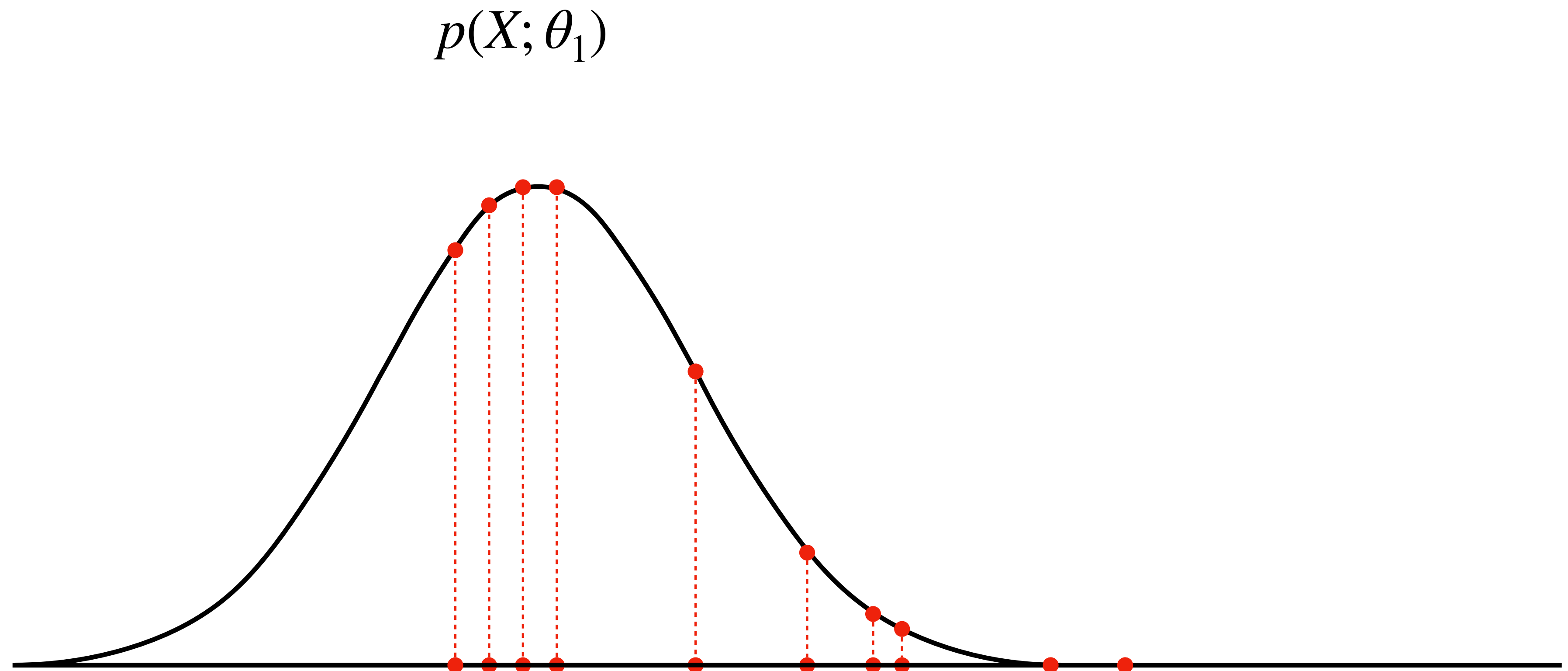
키
164.95
165.35
165.76
166.16
167.78
168.99
169.80
170.20
171.82
172.63
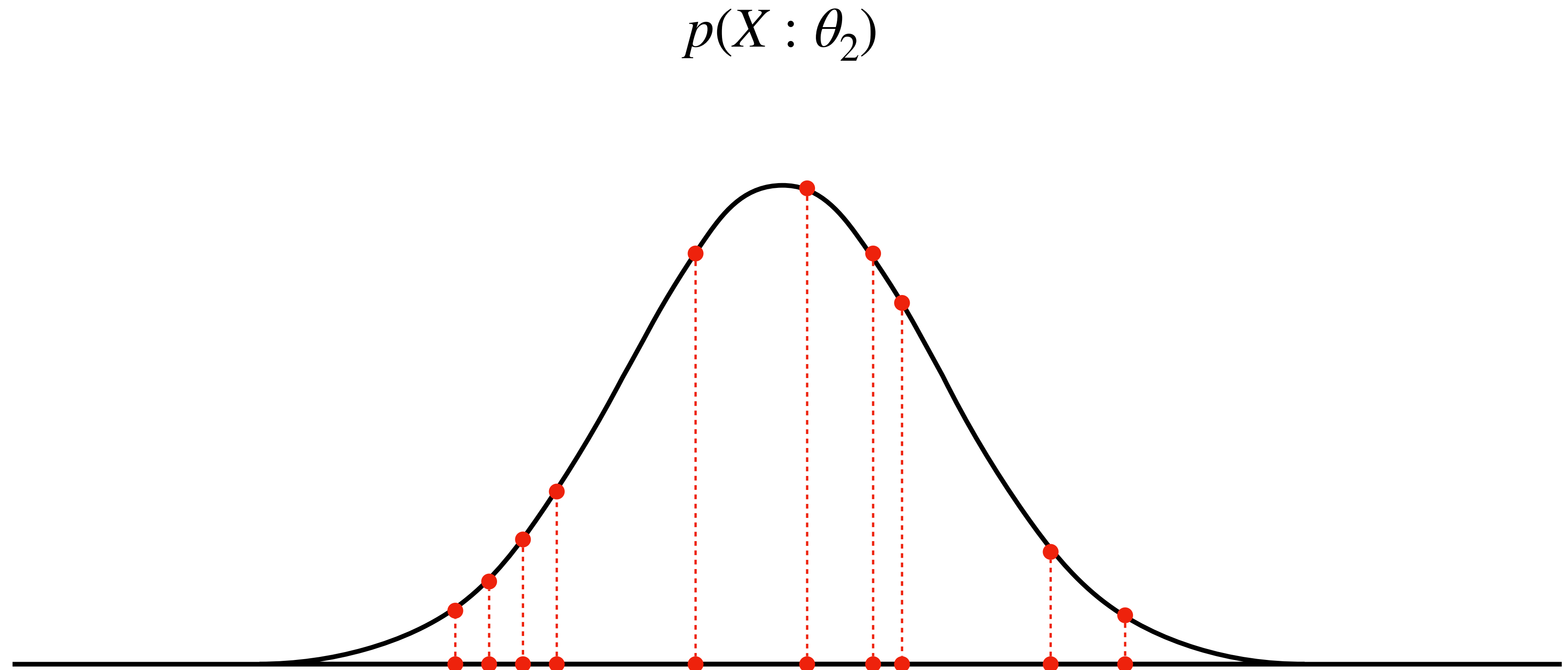
$$\mathscr{L}(\theta_1 \,|\, x) < \mathscr{L}(\theta_2 \,|\, x)$$

Maximum Likelihood Estimation

$$p(X; \theta_3)$$

키
164.95
165.35
165.76
166.16
167.78
168.99
169.80
170.20
171.82
172.63

$$\mathscr{L}(\theta_1 \,|\, x) < \mathscr{L}(\theta_2 \,|\, x) > \mathscr{L}(\theta_3 \,|\, x)$$

8
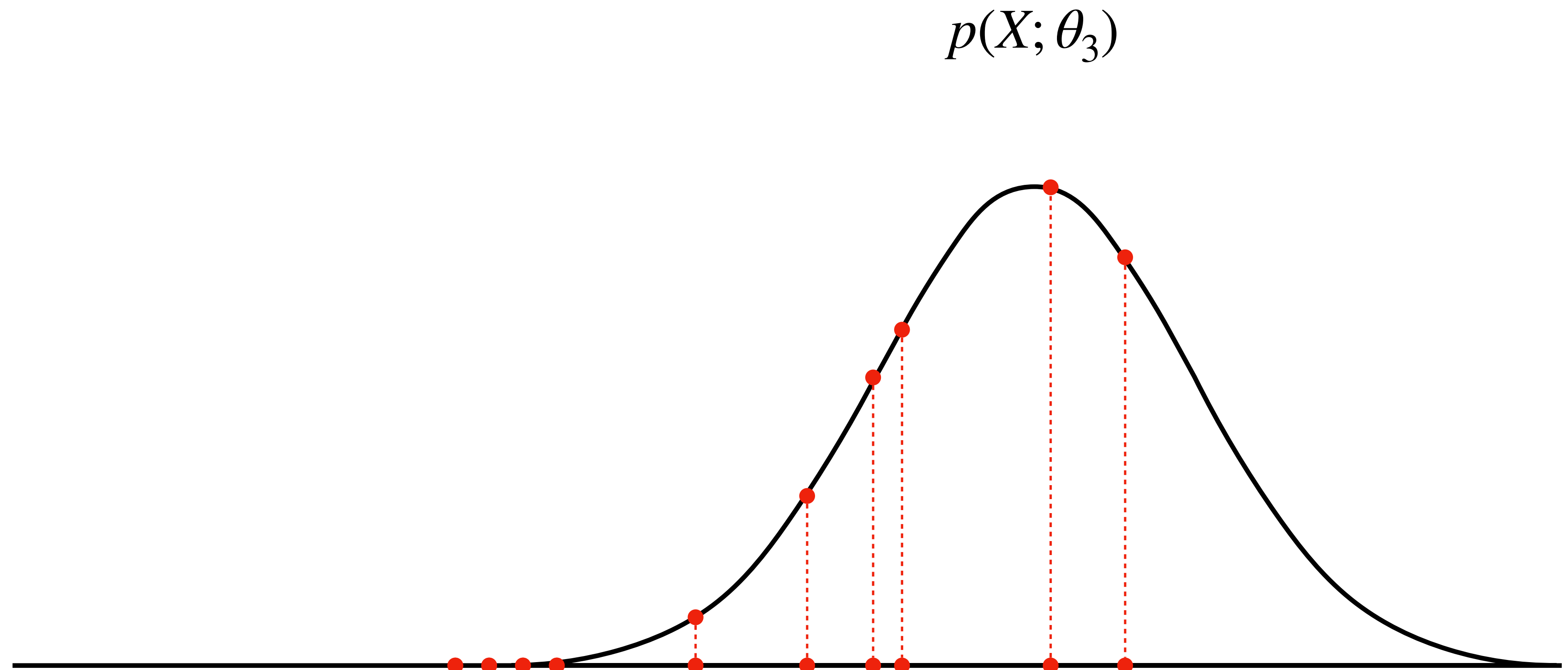
Maximum Likelihood Estimation

Likelihood

$$\mathscr{L}(\theta \,|\, x) = \prod_{i=1}^{N} p(x_i; \theta)$$

$$\hat{\theta} = \underset{\theta}{\mathrm{argmax}} \, \mathscr{L}(\theta \,|\, x)$$

Maximum Likelihood Estimation

키
164.95
165.35
165.76
166.16
167.78
168.99
169.80
170.20
171.82
172.63
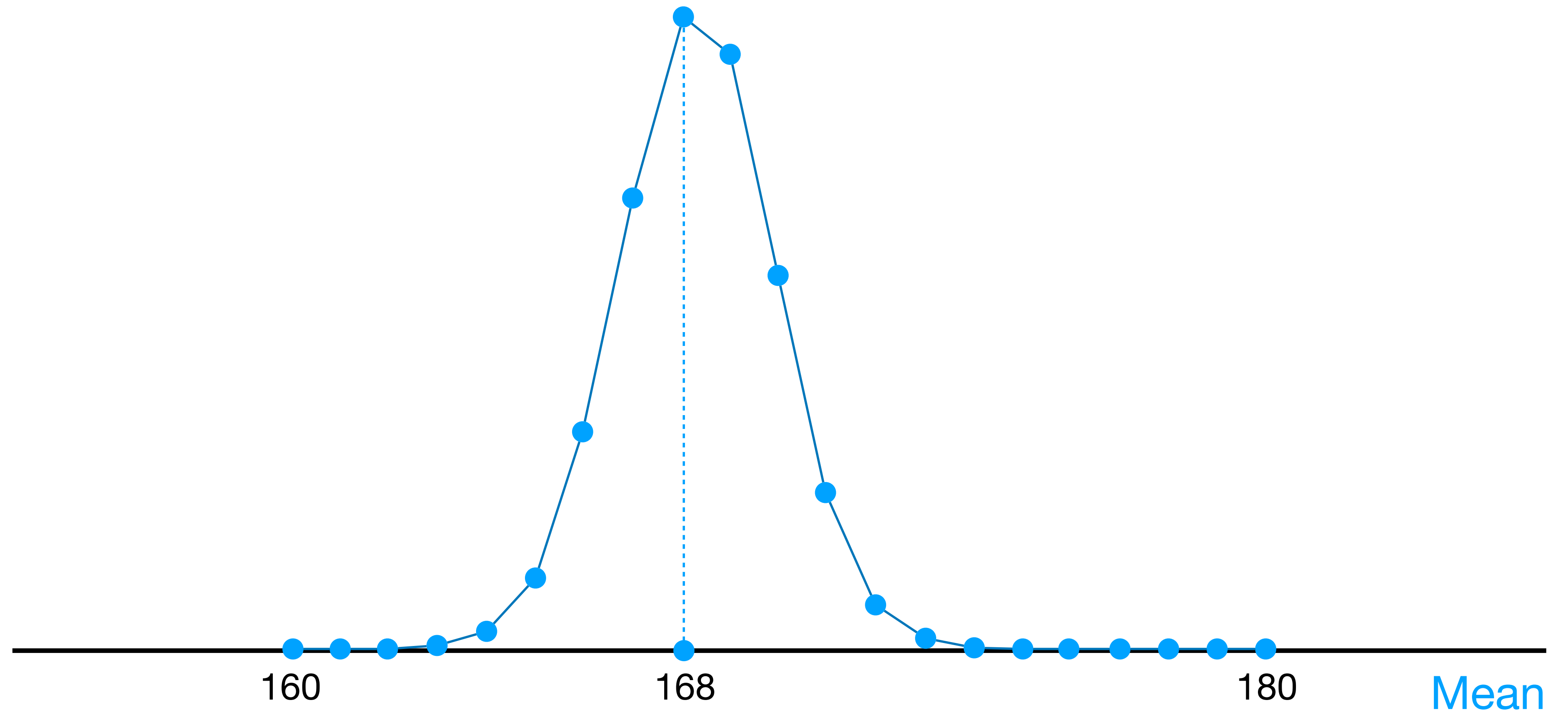
$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \mathscr{L}(\theta \,|\, x)$$

Log Likelihood

Likelihood

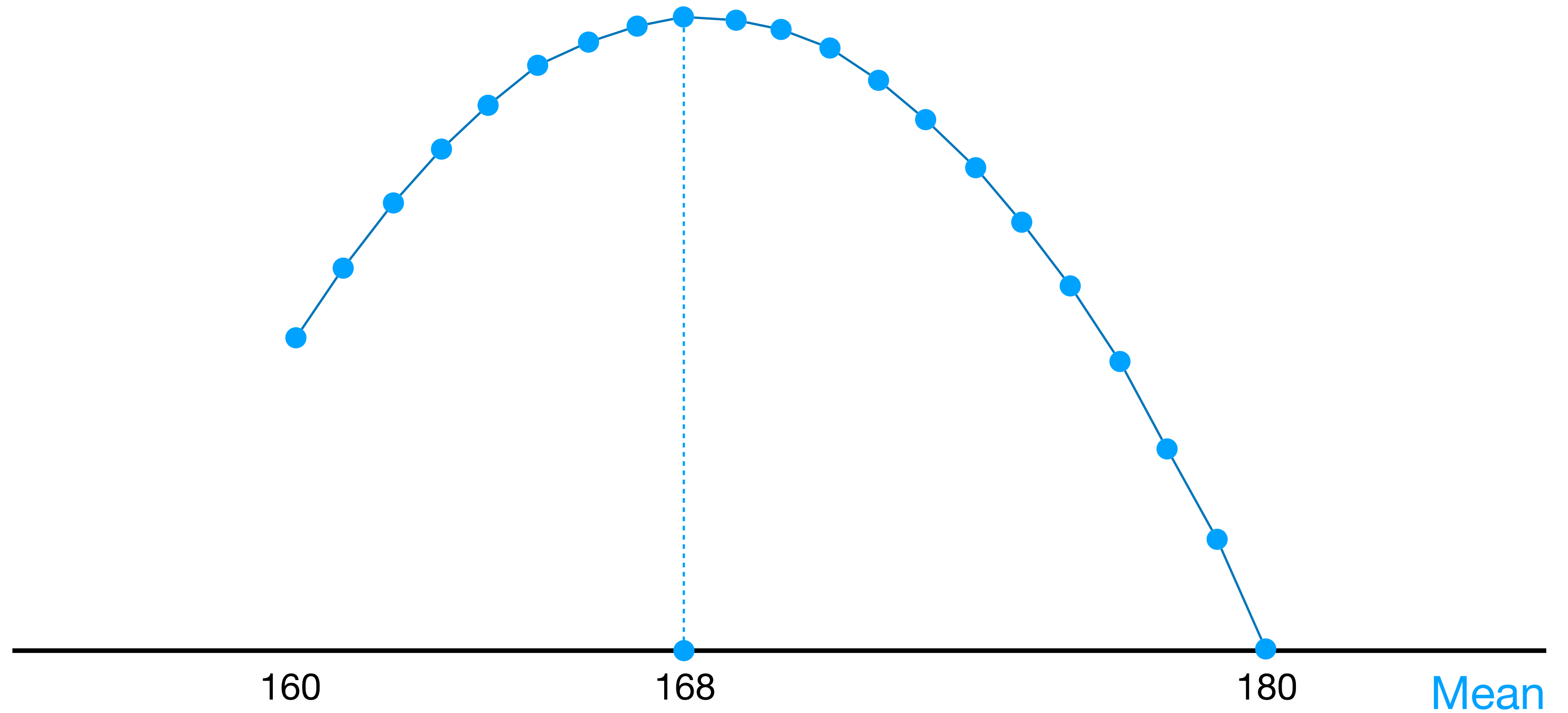$$\mathscr{L}(\theta \,|\, x) = \prod_{i=1}^{N} p(x_i; \theta)$$

Log Likelihood

$$\log \mathscr{L}(\theta \,|\, x) = \sum_{i=1}^{N} \log p(x_i; \theta)$$

$$\hat{\theta} = \underset{\theta}{\mathrm{argmax}} \log \mathscr{L}(\theta \,|\, x)$$

Log Likelihood

키
164.95
165.35
165.76
166.16
167.78
168.99
169.80
170.20
171.82
172.63

160   168   180   Mean

$$\hat{\theta} = \underset{\theta}{\arg\max} \log \mathcal{L}(\theta \,|\, x)$$

Negative Log Likelihood

Likelihood

$$\mathscr{L}(\theta \,|\, x) = \prod_{i=1}^{N} p(x_i; \theta)$$

Log Likelihood

$$\log \mathscr{L}(\theta \,|\, x) = \sum_{i=1}^{N} \log p(x_i; \theta)$$
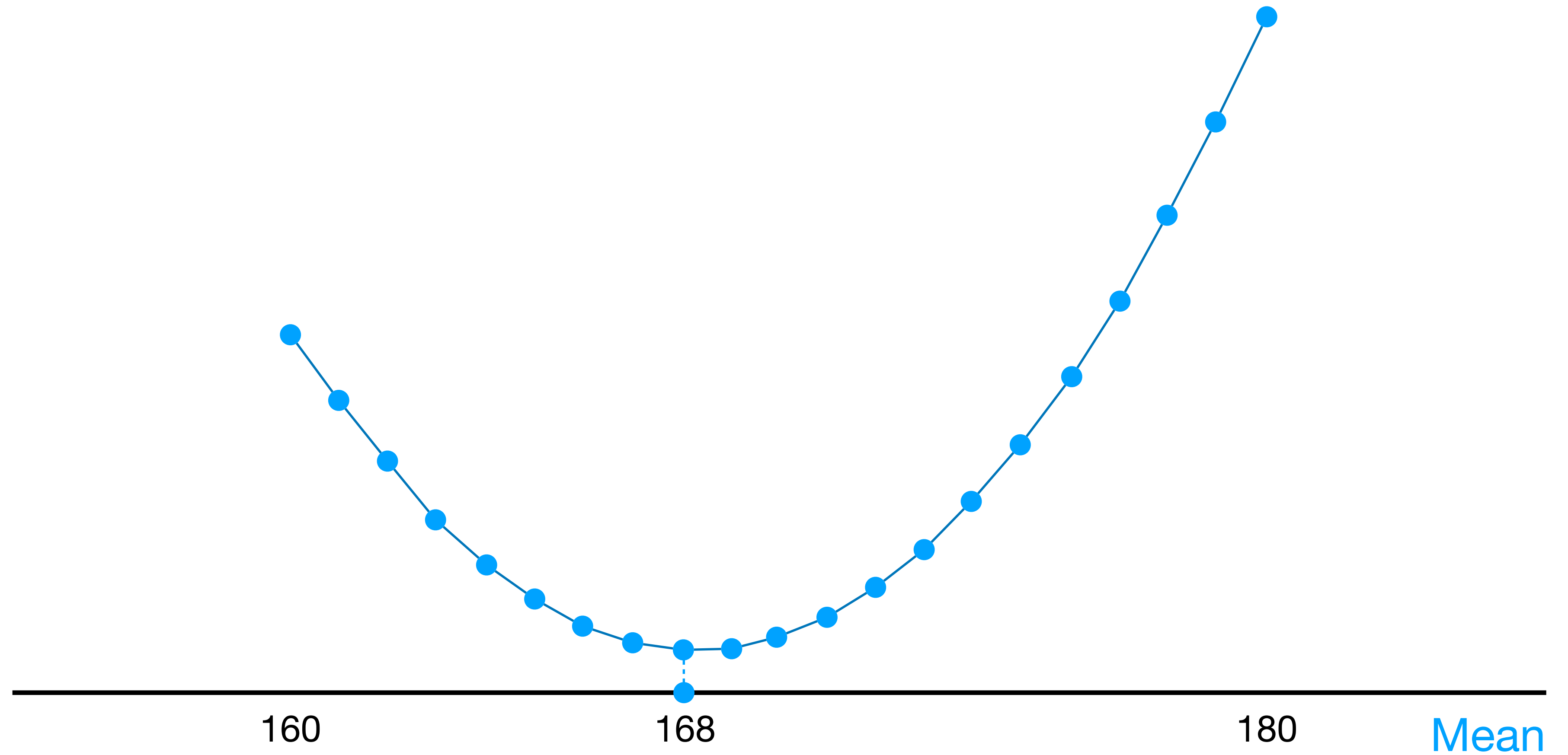
Negative Log Likelihood

$$-\log \mathscr{L}(\theta \,|\, x) = -\sum_{i=1}^{N} \log p(x_i; \theta)$$

$$\hat{\theta} = \operatorname*{argmin}_{\theta} -\log \mathscr{L}(\theta \,|\, x)$$

Negative Log Likelihood

키
164.95
165.35
165.76
166.16
167.78
168.99
169.80
170.20
171.82
172.63

160   168   180   Mean

$$\hat{\theta} = \underset{\theta}{\mathrm{argmin}} - \log \mathscr{L}(\theta \,|\, x)$$

Negative Log Likelihood (Example)

핸든폰 운영체제 점유율 추정

- Simple Example
  - 길가는 사람 10명의 핸드폰 운영체제를 조사했다.
    - Android 7명
    - iOS 3명

Negative Log Likelihood (Example)

Android probability $\qquad\qquad\qquad p$

IOS probability $\qquad\qquad\qquad 1 - p$
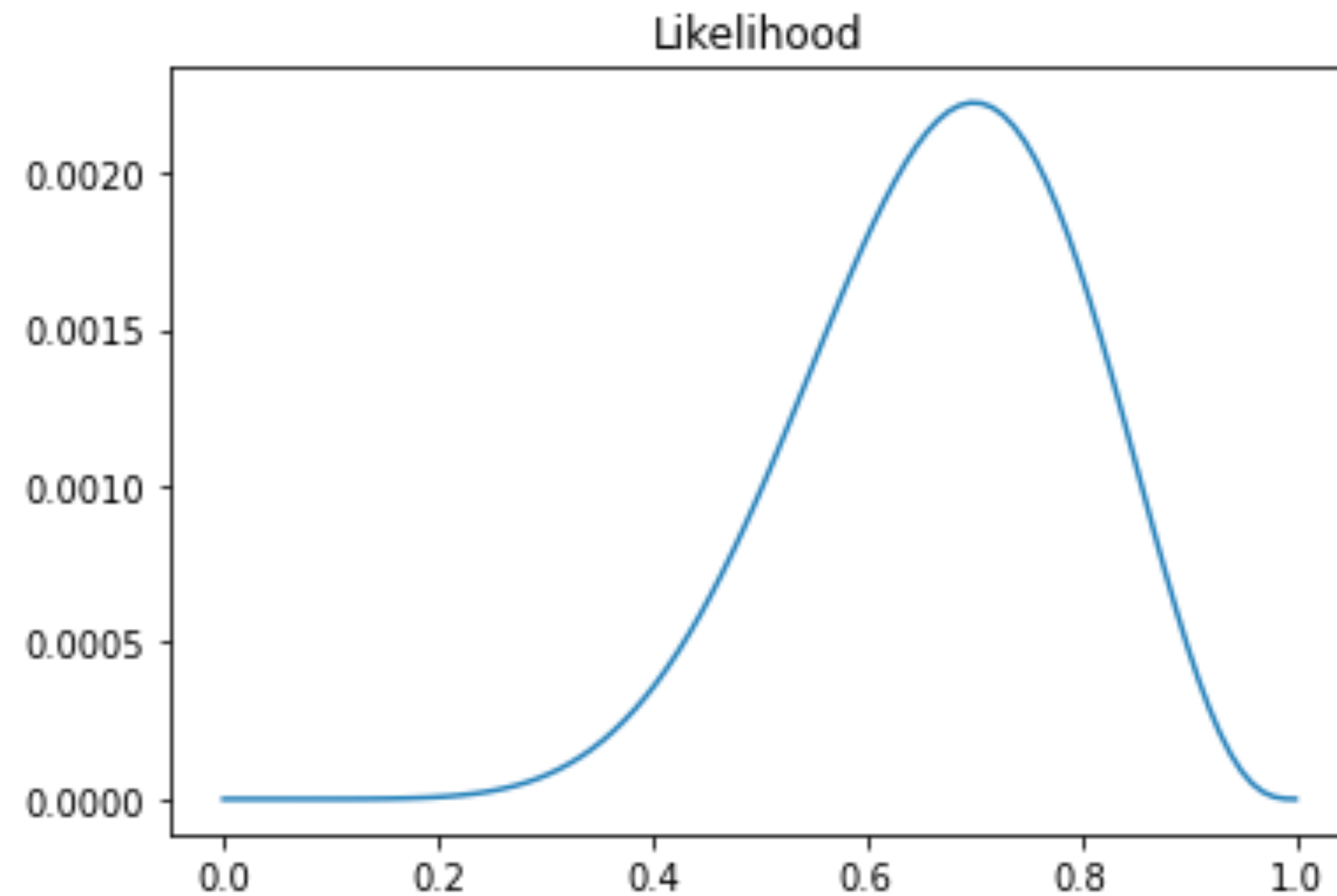
Sampling probability $\qquad\qquad {}_{10}C_7 p^7 (1-p)^3$
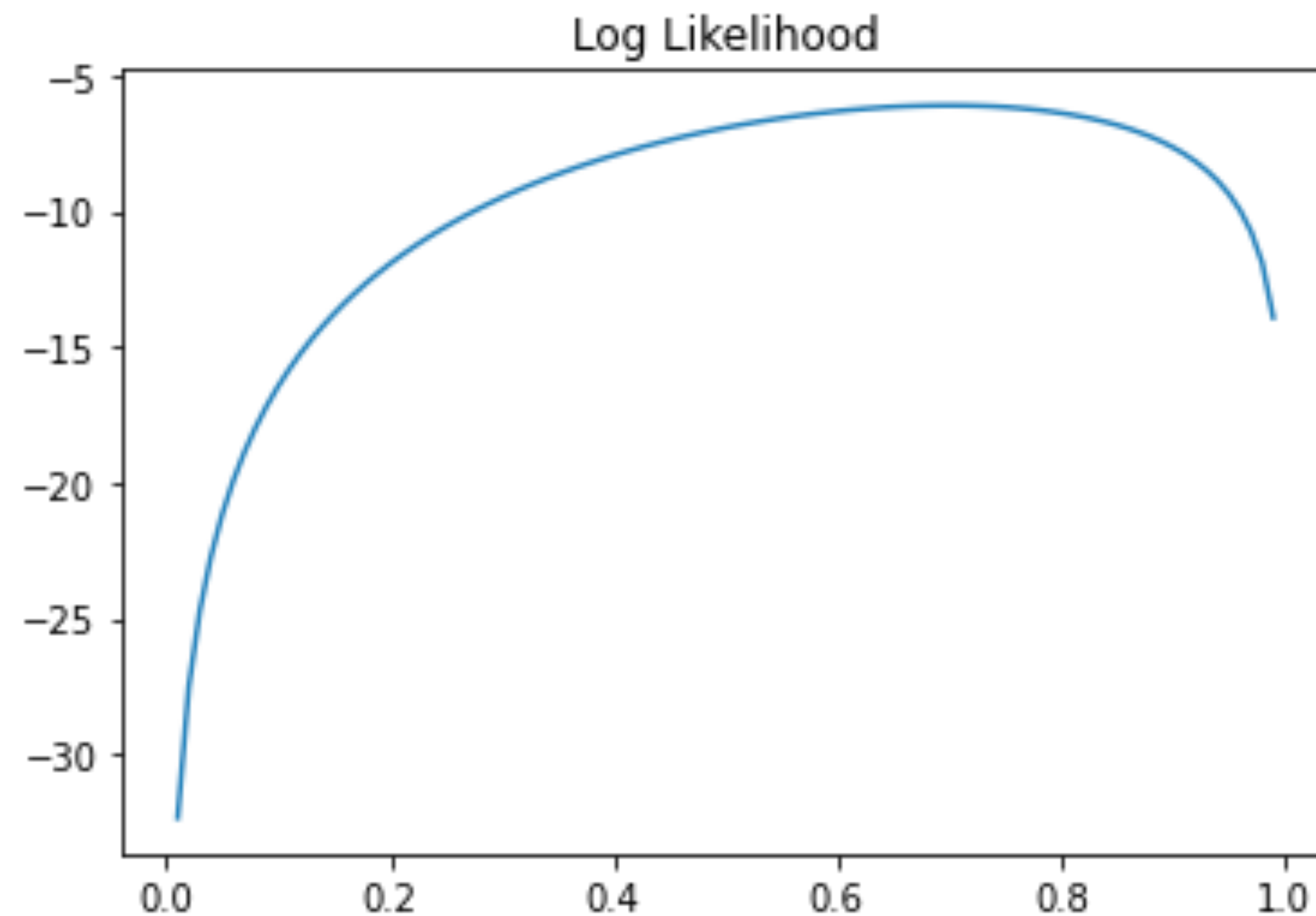
Negative Log Likelihood (Example)

Likelihood

$$\mathcal{L}(p) = p^7(1-p)^3$$

17

Negative Log Likelihood (Example)
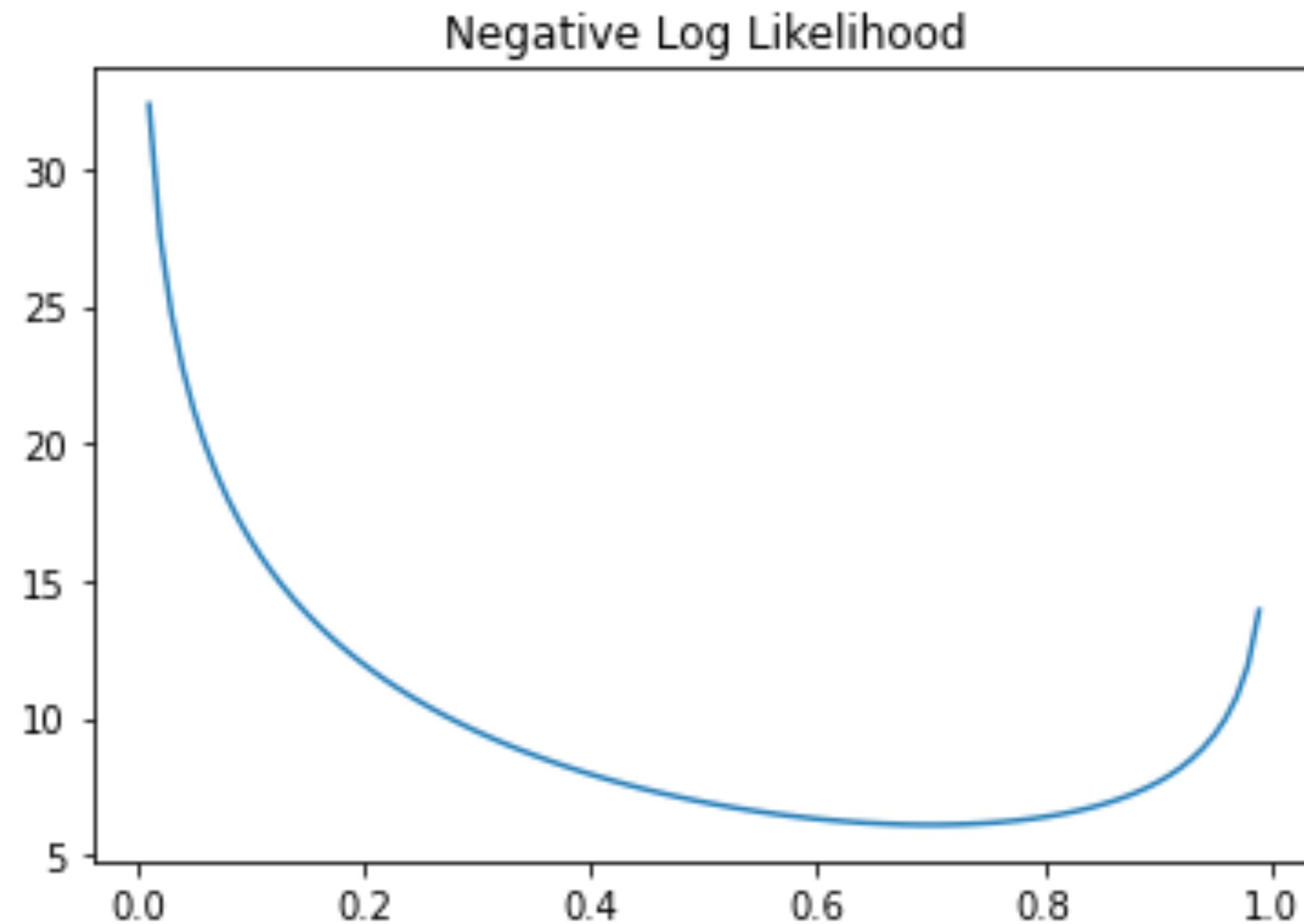
Log Likelihood

$$\log \mathscr{L}(p) = 7 \log p + 3 \log(1 - p)$$
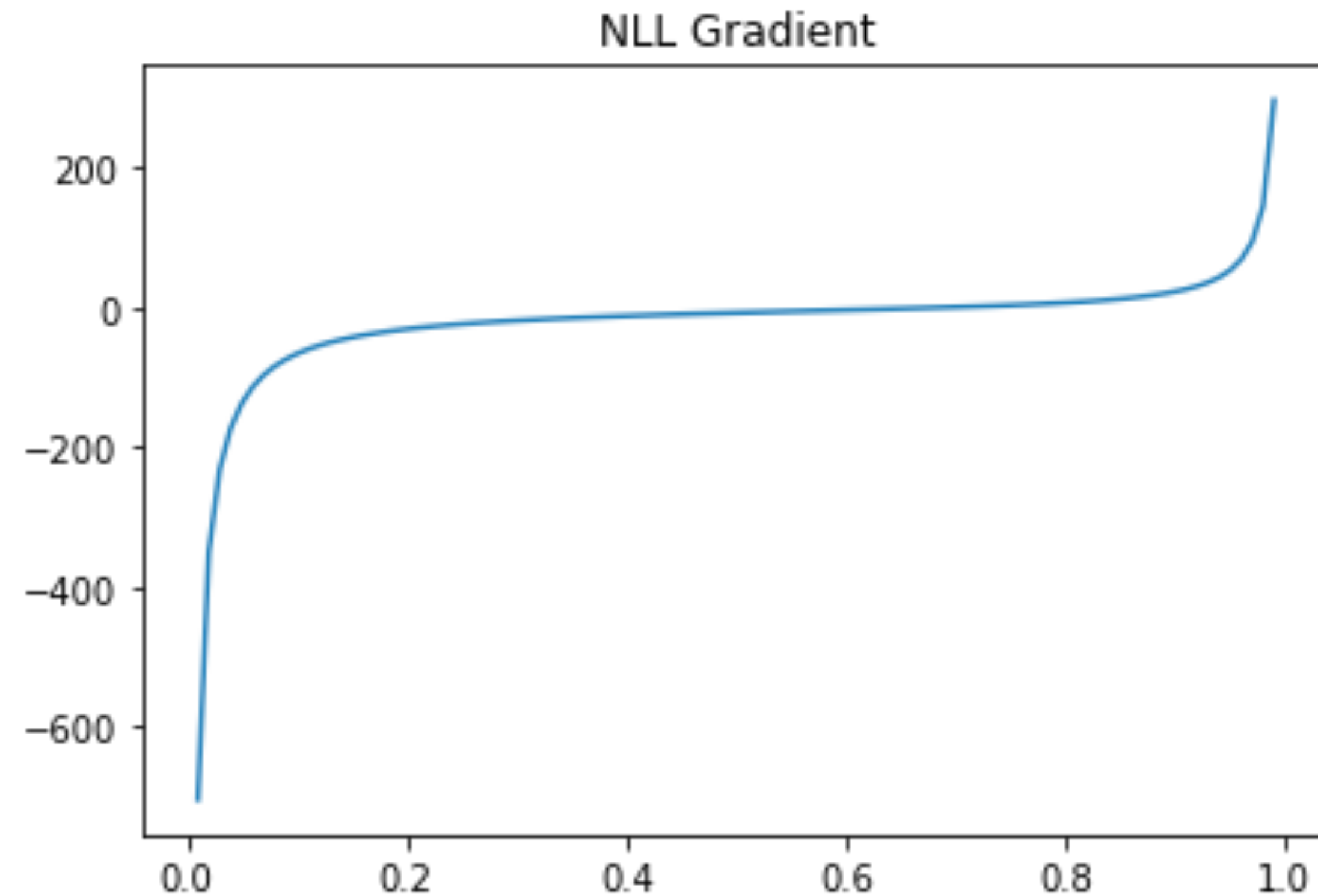
Negative Log Likelihood (Example)

NLL $\qquad -\log \mathscr{L}(p) = -7 \log p - 3 \log(1 - p)$

Negative Log Likelihood (Example)

NLL Gradient

$$-\frac{\log \mathscr{L}(p)}{dp} = -7\frac{1}{p} + 3\frac{1}{1-p}$$

NLL Gradient

감사합니다.