Parameter Estimation

پارامتر چیست؟

- توزیع‌های احتمال زیر را در نظر بگیرید:

➢ $Ber(p)$	$\theta = p$
➢ $Poi(\lambda)$	$\theta = \lambda$
➢ $U(\alpha, \beta)$	$\theta = (\alpha, \beta)$
➢ $N(\mu, \sigma^2)$	$\theta = (\mu, \sigma^2)$
➢ $Beta(a, b)$	$\theta = (a, b)$

- این توزیع‌ها را مدل‌های پارامتری می‌نامیم.
- با داشتن مدل، این پارامترها هستند که شکل حقیقی توزیع را مشخص می‌کنند.
- پارامتر را غالبا با θ نمایش می‌دهیم.
- توجه کنید که θ می‌تواند اسکالر (مثلا $\theta = \lambda$) و یا برداری (مثلا $\theta = (\mu,\sigma^2)$) باشد.

تخمین نقطه‌ای (Point Estimation)

- با داشتن نمونه‌ای از متغیر تصادفی X با اندازه n:

$$\overbrace{(X_1, X_2, \dots, X_n)}^{\text{Sample}}$$

$$\hat{\theta} = \overbrace{g(X_1, X_2, \dots, X_n)}$$

- تخمین ما بی‌غرض (unbiased) یا ناریب خواهد بود، اگر:

$$\boxed{E(\hat{\theta}) = \theta}$$

- $\underbrace{\left(E(\hat{\theta}) - \theta\right)}$ را بایاس تخمین می‌نامند.

Moment Matching

روش گشتاور (Method of Moments)

- گشتاور نمونه (sample moment) مرتبه k-ام را به صورت زیر تعریف می‌کنیم:

$$\boxed{\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k}$$

$X_1, \ldots, X_n$

$m_1 = E[X]$

$m_2 = E[X^2]$

در روش گشتاور با استفاده از تخمین $\hat{m}_k$ برای گشتاور m_k، پارامتر θ را تخمین می‌زنیم.

مثال ۱: توزیع نمایی

$$f_X(x;\lambda) = \lambda e^{-\lambda x} \ : \quad x > 0$$

$$m_1 = \frac{1}{\lambda}$$

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \longrightarrow \hat{\lambda} = \frac{1}{\hat{m}_1}$$

مثال ۲: توزیع نرمال

$N(\mu, \sigma^2)$

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\boxed{\hat{\mu} = \hat{m}_1}$$

$$\hat{m}_2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

$$\hat{\sigma}^2 = \hat{m}_2 - \hat{m}_1^2 = \frac{1}{n} \sum_{i} x_i^2 - \left(\frac{1}{n} \sum_{i} x_i\right)^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

مثال ۳: توزیع پرتو (Pareto)

$$f_X(x) = \theta\sigma^{\theta}x^{-\theta-1} \quad : \quad x > \sigma$$

$$\hat{m}_1 = E[X] = \boxed{\frac{\theta\sigma}{\theta - 1}} \qquad \theta > 1$$

$$\hat{m}_2 = E[X^2] = \boxed{\frac{\theta\sigma^2}{\theta - 2}} \qquad \theta > 2$$

ادامه مثال ۳: توزیع پرتو

$$\hat{\theta} = \sqrt{\frac{\widehat{m}_2 - \widehat{m}_1^2}{\widehat{m}_2} + 1}$$

$$\hat{\sigma} = \frac{\widehat{m}_1(\hat{\theta} - 1)}{\hat{\theta}}$$

تخمین پایدار (Consistent Estimation)

○ تخمینگر $\hat{\theta}$ برای پارامتر θ را پایدار می‌نامیم اگر برای هر $\epsilon > 0$:

$$\boxed{\lim_{n\to\infty} P\left(|\underline{\hat{\theta}} - \overset{?}{\underline{\theta}}| < \epsilon\right)} = 1$$

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$$

• تعریف بالا در حقیقت تعریف پایداری ضعیف (weak consistency) است.

$$P(|\bar{X} - \mu| > \epsilon) = 0$$

روش بیشینه درست‌نمایی (Maximum Likelihood)

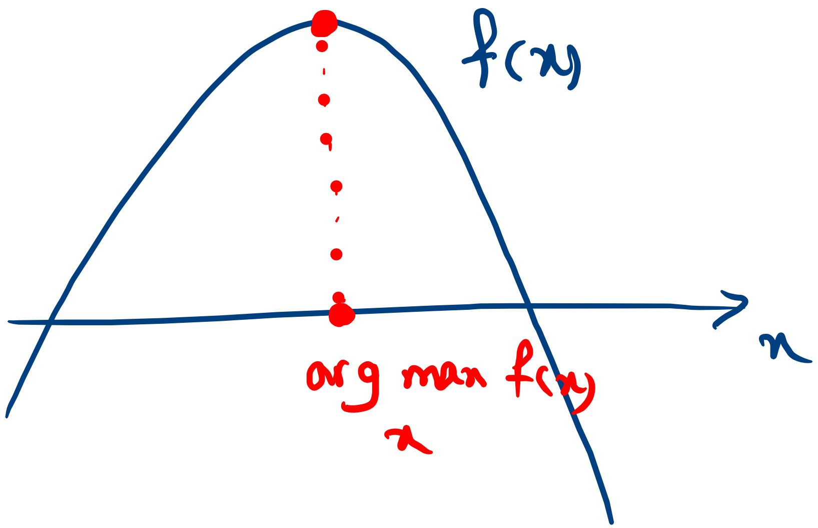
$\boxed{X_1, X_2, \dots, X_n}$ $\quad x_i$ i.i.d.

$\boxed{f_X(x;\theta)}$ $\quad f_X(x|\theta)$ $\quad \boxed{\theta = ?}$

$\mathcal{N}(x;\mu,\sigma^2)$

$D = \{x_1, \dots, x_n\}$

$$\boxed{\hat{\theta}_{ML} = \arg\max_{\theta} P(D)}$$



$$\max_{x} f(x)$$

مثال ۱: توزیع برنولی

$D = \{x_1, \dots, x_n\}$ $\boxed{\theta = ?}$

$P_X(x) = \theta^{x} (1-\theta)^{1-x} \quad x \in \{0, 1\}$

$\hat{\theta}_{ML} = \arg\max_{\theta} \overbrace{\log \underbrace{P(D)}_{\text{log Likelihood}}}$

$LL(\theta) = \log P(D) = \log \underline{P(X_1, \dots, X_n)} = \log \underline{P_X(X_1) P_X(X_2) \cdots P_X(X_n)}$

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

$= \sum_{i=1}^{n} \left(x_i \log\theta + (1-x_i) \log(1-\theta) \right)$

$\frac{d LL(\theta)}{d\theta} = 0 \implies \sum_{i=1}^{n} \frac{x_i}{\theta} - \frac{1-x_i}{1-\theta} = 0$

$\frac{1}{\theta} \sum_{i=1}^{n} x_i - \frac{1}{1-\theta} \left(n - \sum_{i=1}^{n} x_i \right) = 0 \implies \boxed{\hat{\theta}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i}$

مثال ۱: توزیع برنولی

$D = \{x_1, \dots, x_n\}$ $\boxed{\theta = ?}$

$P_X(x) = \theta^{x} (1-\theta)^{1-x} \quad x \in \{0,1\}$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \overbrace{\log \underbrace{P(D)}}$$

log Likelihood

$$LL(\theta) = \log P(D) = \log \underline{P(X_1, \dots, X_n)} = \log \underline{P_X(X_1) P_X(X_2) \cdots P_X(X_n)}$$

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

$$= \sum_{i=1}^{n} \left(x_i \log\theta + (1-x_i) \log(1-\theta) \right)$$

$$\frac{d\,LL(\theta)}{d\theta} = 0 \implies \sum_{i=1}^{n} \frac{x_i}{\theta} - \frac{1-x_i}{1-\theta} = 0$$

$$\frac{1}{\theta} \sum_{i=1}^{n} x_i - \frac{1}{1-\theta} \left(n - \sum_{i=1}^{n} x_i \right) = 0 \implies \boxed{\hat{\theta}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

Estimation

$D = \{\underline{x_1, x_2, \ldots, x_n}\}$ $\quad x_i$ i.i.d.

$x_i \sim f_X(x;\theta)$ $\quad \theta = ?$

$\hat{\theta} = g(\underbrace{x_1, \ldots, x_n})$

R.V.

$E[\hat{\theta}] = \theta \longrightarrow$ $\hat{\theta}$ is an unbiased estimator for θ.

$E[\hat{\theta}] - \theta \longrightarrow$ bias

$\hat{m}_1 \quad \hat{m}_2$

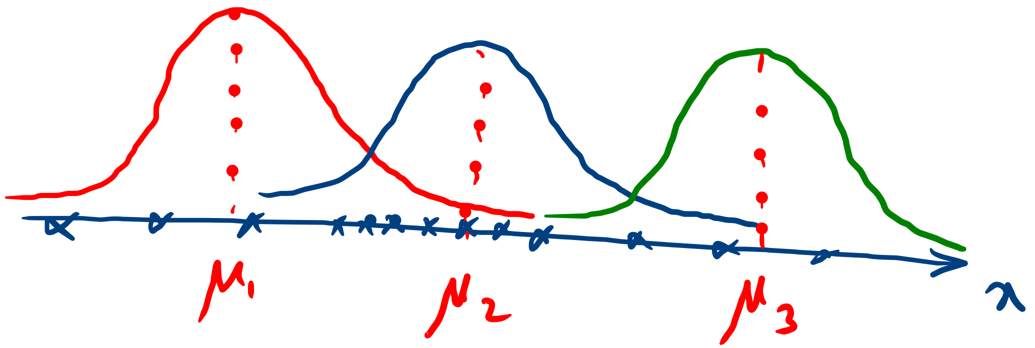
$$\hat{\mu} = \hat{m}_1$$

$$\hat{\sigma}^2 = \hat{m}_2 - \hat{m}_1^2$$

Maximum Likelihood

$$D = \{x_1, \dots, x_n\}$$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \underbrace{\log \underline{P(D)}}_{\text{Log-Likelihood}}$$



مثال ۲: توزیع ویبول (Weibull)

$$f_X(x) = cx\, e^{-c\frac{x^2}{2}}$$

$D = \{x_1, \dots, x_n\} \longrightarrow \hat{c}_{ML} = ?$

$$\hat{c}_{ML} = \arg\max_c \log P(D) = \arg\max_c \log P(x_1, \dots x_n) = \arg\max_c \log\left(f_X(x_1)\dots f_X(x_n)\right)$$

$$= \arg\max_c \sum_{i=1}^{n} \log f_X(x_i) = \arg\max_c \underbrace{\sum_{i=1}^{n}\left(\log c + \log x_i - c\frac{x_i^2}{2}\right)}_{LL(c)}$$

$$\frac{d\,LL(c)}{dc} = \sum_{i=1}^{n}\left(\frac{1}{c} - \frac{x_i^2}{2}\right) = 0$$

$$\frac{n}{c} - \frac{1}{2}\sum_{i=1}^{n} x_i^2 = 0 \Longrightarrow \boxed{\hat{c}_{ML} = \frac{2n}{\sum_{i=1}^{n} x_i^2}}$$

ادامه مثال ۲

○اگر مقادیر نمونه مشاهده شده از این توزیع $(1,1,2,1,3,3)$ باشد، تخمین c به روش ML چیست؟

$$\hat{c}_{ML} = \frac{2 \times 6}{1+1+4+1+9+9} = \frac{12}{25} = 0.48$$

مثال ۳: توزیع پواسون

$$P_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

$D = \{x_1, \dots, x_n\}$ $\quad\quad \hat{\lambda}_{ML} = ?$

$$LL(\lambda) = \sum_{i=1}^{n} \log P_X(x_i) = \sum_{i=1}^{n}\left(-\lambda + x_i \log\lambda - \log x_i!\right)$$

$$\frac{d\,LL(\lambda)}{d\lambda} = \sum_{i=1}^{n} -1 + \frac{1}{\lambda}x_i = 0 \Rightarrow \boxed{\hat{\lambda}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

تخمین ML بیش از یک پارامتر: توزیع نرمال

$$f_X(x_i|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$\hat{\mu}_{ML} = ?$ $\quad$ $\hat{\sigma}^2_{ML} = ?$

$$LL(\mu,\sigma) = \sum_{i=1}^{n} \log f_X(x_i) = \sum_{i=1}^{n} -\log\sqrt{2\pi} - \log\sigma - \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{d\,LL(\mu,\sigma)}{d\mu} = \sum_{i=1}^{n} \frac{(x_i-\mu)}{\sigma^2} = 0 \Rightarrow \sum_{i=1}^{n} x_i - n\mu = 0$$

$$\Rightarrow \boxed{\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

$$\frac{d\,LL(\mu,\sigma)}{d\sigma} = \sum_{i=1}^{n} \frac{-1}{\sigma} + \frac{(x_i-\mu)^2}{\sigma^3} = 0 \Rightarrow \sum_{i=1}^{n}\left(-\sigma^2 + (x_i-\mu)^2\right) = 0$$

$$-n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2 = 0 \Rightarrow \boxed{\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu}_{ML})^2}$$

تخمین ML توزیع یکنواخت

$$f(x_i|\alpha,\beta) = \frac{1}{\beta-\alpha} : \boxed{\underline{\alpha} < x_i < \underline{\beta}}$$

$$LL(\alpha,\beta) = \sum_{i=1}^{n} \log f_X(x_i) = \sum_{i=1}^{n} -\log(\beta-\alpha) = -n\log(\beta-\alpha)$$

$$\frac{d\,LL}{d\alpha} = \boxed{\frac{n}{\beta-\alpha}} > 0 \Longrightarrow \hat{\alpha}_{ML} = \min\{x_1, \ldots, x_n\}$$

$$\frac{d\,LL}{d\beta} = \frac{-n}{\beta-\alpha} < 0 \Longrightarrow \hat{\beta}_{ML} = \max\{x_1, \ldots, x_n\}$$

مثال

- تخمین ML پارامترهای θ_1 و θ_2 را برای توزیع احتمال زیر بیابید:

$$f_X(x;\theta_1,\theta_2) = \theta_1 \theta_2^{\theta_1} x^{-\theta_1-1} \quad , \quad \theta_2 \le x \quad , \theta_1,\theta_2 > 0$$