



دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته  
دانشکده مهندسی برق و کامپیوتر

# تئوری تخمین

## ESTIMATION THEORY

نیم سال اول سال تحصیلی ۰۴-۰۵

# تخمین بیشترین درست نمایی

2

- مقدمه
- پیدا کردن MLE
- ویژگی های MLE
- تبدیل پارامترها
- یافتن عددی MLE
- تعمیم به حالت برداری

- تخمین بیشترین درست نمایی یک جایگزین برای تخمین  $MVU$  است زمانی که  $MVU$  وجود ندارد، یا وجود دارد ولی نمی توانیم آن را پیدا کنیم.
- این تخمین بیشترین استفاده را در کاربرد دارد و در داده گسترده به تخمینگر  $MVU$  میل می کند.
- در این تخمینگر به دنبال ماکزیمم کردن تابع درست نمایی هستیم.
- نشان داده می شود که این روند در حالت مجانبی (وقتی داده به سمت بی نهایت میل کند) به  $MVU$  منجر می شود.

□ به طور خلاصه تخمین MLE از  $\theta$  مقداری است که تابع  $p(\mathbf{x}; \theta)$  را بیشینه می کند.

The MLE for a scalar parameter is defined to be *the value of  $\theta$  that maximizes  $p(\mathbf{x}; \theta)$*

□ مثال ۱:

$$x[n] = A + w[n] \quad n = 0, 1, \dots, N-1$$

$$p(\mathbf{x}; A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right].$$

$$\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A) \quad \longrightarrow \quad \hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n].$$

□ اگر یک تخمینگر «کارا» وجود داشته باشد، MLE آن را خواهد یافت.

□ مثال ۲

$$x[n] = A + w[n] \quad n = 0, 1, \dots, N-1$$

□ در این مثال،  $w[n]$  نویز WGN با واریانس  $A$  است. بنابراین  $A$  مقدار مثبتی است که باید تخمین زده شود.

$$p(\mathbf{x}; A) = \frac{1}{(2\pi A)^{\frac{N}{2}}} \exp \left[ -\frac{1}{2A} \sum_{n=0}^{N-1} (x[n] - A)^2 \right].$$

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}; A)}{\partial A} &= -\frac{N}{2A} + \frac{1}{A} \sum_{n=0}^{N-1} (x[n] - A) + \frac{1}{2A^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \\ &\stackrel{?}{=} I(A)(\hat{A} - A). \end{aligned}$$

□ در ضمن می توان محاسبه نمود:

$$\text{var}(\hat{A}) \geq \frac{A^2}{N(A + \frac{1}{2})}.$$

□ حال به محاسبه MLE می پردازیم:

$$\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} = -\frac{N}{2A} + \frac{1}{A} \sum_{n=0}^{N-1} (x[n] - A) + \frac{1}{2A^2} \sum_{n=0}^{N-1} (x[n] - A)^2$$

□ با مساوی صفر قرار دادن مشتق دو جواب به دست می آوریم:

$$\hat{A} = -\frac{1}{2} \pm \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] + \frac{1}{4}}.$$

□ چون A مثبت است، جواب زیر انتخاب می شود:

$$\hat{A} = -\frac{1}{2} + \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] + \frac{1}{4}}$$

□ باید توجه کنیم که این تخمینگر بایاس دارد:

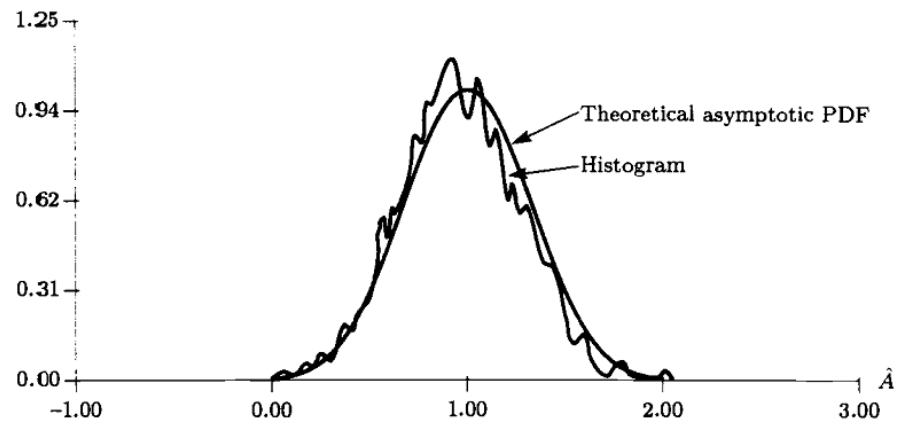
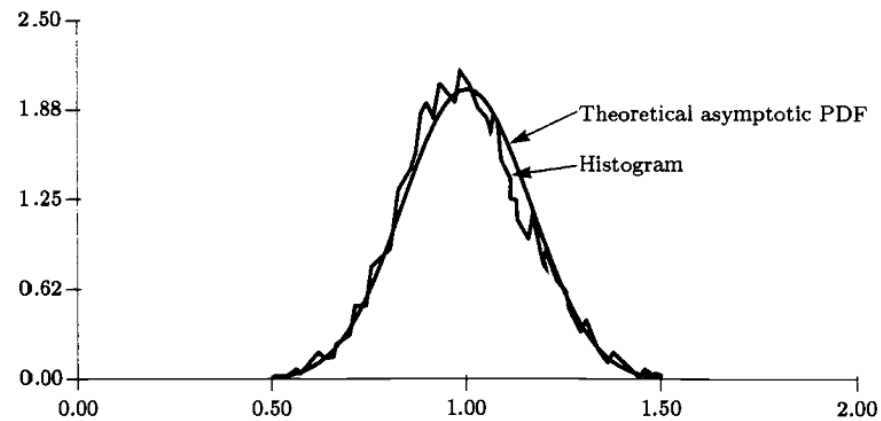
$$\begin{aligned}
 E(\hat{A}) &= E\left(-\frac{1}{2} + \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] + \frac{1}{4}}\right) \\
 &\neq -\frac{1}{2} + \sqrt{E\left(\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]\right) + \frac{1}{4}} \quad \text{for all } A \\
 &= -\frac{1}{2} + \sqrt{A + A^2 + \frac{1}{4}} \\
 &= A.
 \end{aligned}$$

□ اما با توجه به اینکه در  $N$  بزرگ مقدار تخمین به  $A$  میل می کند، این تخمینگر «به صورت مجانبی بدون بایاس» است. در  $N$  بزرگ داریم:

$$\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \rightarrow E(x^2[n]) = A + A^2 \quad \longrightarrow \quad \hat{A} \rightarrow A.$$

□ ضمنا می توان نشان داد مقدار واریانس این تخمینگر به صورت مجانبی (وقتی  $N$  به بی نهایت میل می کند) به حد کرامر-راو میل می کند. یعنی این تخمینگر به صورت مجانبی کارا است. مثلا به ازای  $A=1$ ، جدول زیر مقدار میانگین و واریانس را در  $N$  های مختلف با استفاده از شبیه سازی مونت کارلو محاسبه نموده است:

| Data Record Length, $N$      | Mean, $E(\hat{A})$ | $N \times \text{Variance}, N \text{ var}(\hat{A})$ |
|------------------------------|--------------------|----------------------------------------------------|
| 5                            | 0.954              | 0.624                                              |
| 10                           | 0.976              | 0.648                                              |
| 15                           | 0.991              | 0.696                                              |
| 20                           | 0.996 (0.987)      | 0.707 (0.669)                                      |
| 25                           | 0.994              | 0.656                                              |
| Theoretical asymptotic value | 1                  | 0.667                                              |

(a)  $N = 5$ (b)  $N = 20$

- به طور کلی تخمینگر MLE به صورت مجانبی بدون بایاس است و واریانس آن به حد کرامر-راو میل می کند و در  $N$  بزرگ PDF آن به گوسی میل می کند.
- در عمل معمولاً نمی دانیم مقدار  $N$  چقدر باید بزرگ باشد تا به رابطه بالا دست یابیم.

**Theorem 7.1 (Asymptotic Properties of the MLE)** *If the PDF  $p(\mathbf{x}; \theta)$  of the data  $\mathbf{x}$  satisfies some “regularity” conditions, then the MLE of the unknown parameter  $\theta$  is asymptotically distributed (for large data records) according to*

$$\hat{\theta} \stackrel{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta)) \quad (7.11)$$

where  $I(\theta)$  is the Fisher information evaluated at the true value of the unknown parameter.

□ مثال ۳. تخمین MLE فاز سینوسی

$$x[n] = A \cos(2\pi f_0 n + \phi) + w[n] \quad n = 0, 1, \dots, N-1$$

□ نویز WGN است و مقدار A و  $f_0$  را می دانیم.

□ باید مقدار فازی که تابع زیر را بیشینه می کند پیدا کنیم:

$$p(\mathbf{x}; \phi) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A \cos(2\pi f_0 n + \phi))^2 \right]$$

□ یا به عبارتی باید نقطه کمینه تابع زیر را پیدا کنیم:

$$J(\phi) = \sum_{n=0}^{N-1} (x[n] - A \cos(2\pi f_0 n + \phi))^2.$$

□ مثال ۳. ادامه

□ با مشتق گیری از تابع فوق داریم:

$$\frac{\partial J(\phi)}{\partial \phi} = 2 \sum_{n=0}^{N-1} (x[n] - A \cos(2\pi f_0 n + \phi)) A \sin(2\pi f_0 n + \phi)$$

□ برای یافتن مقدار کمینه، مشتق را مساوی صفر قرار می دهیم:

$$\sum_{n=0}^{N-1} x[n] \sin(2\pi f_0 n + \hat{\phi}) = A \sum_{n=0}^{N-1} \sin(2\pi f_0 n + \hat{\phi}) \cos(2\pi f_0 n + \hat{\phi}).$$

□ با توجه به خاصیت زیر، سمت راست عبارت فوق برای فرکانس غیر صفر و  $1/2$  تقریبا برابر با صفر است:

$$\frac{1}{N} \sum_{n=0}^{N-1} \sin(2\pi f_0 n + \hat{\phi}) \cos(2\pi f_0 n + \hat{\phi}) = \frac{1}{2N} \sum_{n=0}^{N-1} \sin(4\pi f_0 n + 2\hat{\phi}) \approx 0$$

□ مثال ۳. ادامه

□ بنابراین:

$$\sum_{n=0}^{N-1} x[n] \sin(2\pi f_0 n + \hat{\phi}) = 0.$$

$$\sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n \cos \hat{\phi} = - \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \sin \hat{\phi}$$

□ یا به عبارتی:

□ و در نهایت:

$$\hat{\phi} = -\arctan \frac{\sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n}{\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n}.$$

□ مثال ۳. ادامه

□ با توجه به قضیه جانبی MLE داشتیم:  $\hat{\phi} \stackrel{a}{\sim} \mathcal{N}(\phi, I^{-1}(\phi))$ .

□ در اینجا:

$$I(\phi) = \frac{NA^2}{2\sigma^2} \Rightarrow \text{var}(\hat{\phi}) = \frac{1}{N \frac{A^2}{2\sigma^2}} = \frac{1}{N\eta}$$

□ شبیه سازی:

| Data Record Length, $N$      | Mean, $E(\hat{\phi})$ | $N \times \text{Variance}, N \text{var}(\hat{\phi})$ |
|------------------------------|-----------------------|------------------------------------------------------|
| 20                           | 0.732                 | 0.0978                                               |
| 40                           | 0.746                 | 0.108                                                |
| 60                           | 0.774                 | 0.110                                                |
| 80                           | 0.789                 | 0.0990                                               |
| Theoretical asymptotic value | $\phi = 0.785$        | $1/\eta = 0.1$                                       |

□ مثال ۳. ادامه

□ در مورد تخمین حاصل شده داریم:

$$\hat{\phi} = -\arctan \frac{\sum_{n=0}^{N-1} [A \cos(2\pi f_0 n + \phi) + w[n]] \sin 2\pi f_0 n}{\sum_{n=0}^{N-1} [A \cos(2\pi f_0 n + \phi) + w[n]] \cos 2\pi f_0 n} \approx -\arctan \frac{-\frac{NA}{2} \sin \phi + \sum_{n=0}^{N-1} w[n] \sin 2\pi f_0 n}{\frac{NA}{2} \cos \phi + \sum_{n=0}^{N-1} w[n] \cos 2\pi f_0 n}$$

□ اگر مقدار N یا مقدار SNR بزرگ باشد، بخش نویزی کوچک خواهد بود:

$$\hat{\phi} \approx \arctan \frac{\sin \phi - \frac{2}{NA} \sum_{n=0}^{N-1} w[n] \sin 2\pi f_0 n}{\cos \phi + \frac{2}{NA} \sum_{n=0}^{N-1} w[n] \cos 2\pi f_0 n}.$$

# تبدیل پارامترها

16

□ مثال ۴. فرض کنید در مثال ولتاژ DC در نویز WGN می خواهیم مقدار  $\exp(A)$  را تخمین بزنیم.

$$x[n] = A + w[n] \quad n = 0, 1, \dots, N-1$$

$$p(\mathbf{x}; A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] \quad -\infty < A < \infty$$

□ با تبدیل پارامتر:

$$p_T(\mathbf{x}; \alpha) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - \ln \alpha)^2 \right] \quad \alpha > 0$$

□ باید بیشینه تابع فوق پیدا شود.

□ مثال ۴. ادامه

□ مشتق گیری از تابع و مساوی صفر قرار دادن منجر می شود به:

$$\sum_{n=0}^{N-1} (x[n] - \ln \hat{\alpha}) \frac{1}{\hat{\alpha}} = 0 \quad \Rightarrow \quad \hat{\alpha} = \exp(\bar{x}).$$

**Theorem 7.2 (Invariance Property of the MLE)** *The MLE of the parameter  $\alpha = g(\theta)$ , where the PDF  $p(\mathbf{x}; \theta)$  is parameterized by  $\theta$ , is given by*

$$\hat{\alpha} = g(\hat{\theta})$$

where  $\hat{\theta}$  is the MLE of  $\theta$ . The MLE of  $\hat{\theta}$  is obtained by maximizing  $p(\mathbf{x}; \theta)$ . If  $g$  is not a one-to-one function, then  $\hat{\alpha}$  maximizes the modified likelihood function  $\bar{p}_T(\mathbf{x}; \alpha)$ , defined as

$$\bar{p}_T(\mathbf{x}; \alpha) = \max_{\{\theta: \alpha = g(\theta)\}} p(\mathbf{x}; \theta).$$

# تبدیل پارامترها

18

□ مثال ۵. تخمین مقدار توان نویز WGN بر حسب dB:

$$P = 10 \log_{10} \sigma^2.$$

□ تابع PDF نویز:

$$p(\mathbf{x}; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right].$$

□ مشتق گیری از تابع log-likelihood (چرا؟)

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}; \sigma^2)}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2} \left[ -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right] \\ &= -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=0}^{N-1} x^2[n] \end{aligned}$$

□ مثال ۵. ادامه

□ در نهایت تخمین توان و توان بر حسب dB:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n].$$

$$\begin{aligned} \hat{P} &= 10 \log_{10} \hat{\sigma}^2 \\ &= 10 \log_{10} \frac{1}{N} \sum_{n=0}^{N-1} x^2[n]. \end{aligned}$$