آمار و احتمال مهندسی

تخمین پارامتر

پارامتر چیست؟

- توزیع‌های احتمال زیر را در نظر بگیرید:

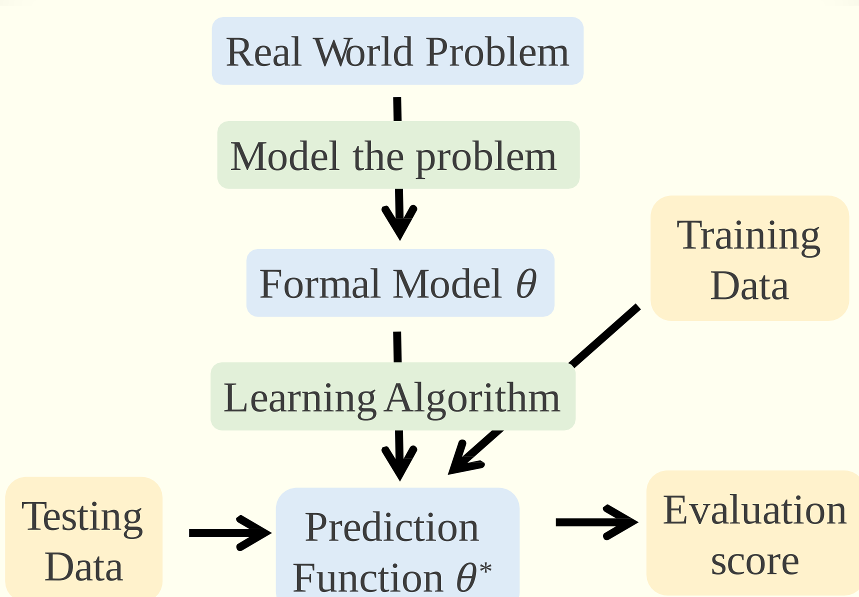
  - $Ber(p)$ $\theta = p$
  - $Poi(\lambda)$ $\theta = \lambda$
  - $U(\alpha, \beta)$ $\theta = (\alpha, \beta)$
  - $N(\mu, \sigma^2)$ $\theta = (\mu, \sigma^2)$
  - $Beta(a, b)$ $\theta = (a, b)$

- 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.

اهمیت تخمین پارامتر

- در دنیای واقعی ما مقدار «صحیح» پارامترهای توزیع احتمالی که برای مدل‌سازی یک پدیده خاص به کار می‌رود را نمی‌دانیم.
  - اما می‌توانیم به مشاهده داده از آن پدیده بپردازیم.
- برای مثال می‌دانیم شیر آمدن یک سکه توزیع $Ber(p)$ دارد ولی p را نمی‌دانیم.
  - اما می‌توانیم سکه را چندین بار پرتاب کنیم و تعداد دفعاتی که شیر می‌آید را مشاهده کنیم.
- باید پارامترهای مدل را از داده مشاهده شده تخمین بزنیم.
- تخمینگر $\hat{\theta}$ یک متغیر تصادفی است که پارامتر θ را تخمین می‌زند.
- تخمین پارامتر به ما اجازه می‌دهد:
  - درک بهتری از فرایند تولید داده داشته باشیم
  - بر مبنای مدل به دست آمده آینده را پیش‌بینی کنیم
  - فرایندهای مختلف را شبیه‌سازی کنیم

Supervised Learning



میانگین و واریانس نمونه

- به دنباله متغیرهای تصادفی *i.i.d.* $(X_1,X_2,\ldots,X_n)$ که از یک جامعه آماری با توزیع F انتخاب شده باشند، یک نمونه از توزیع F می‌گوییم.
- طبق تعریف، میانگین نمونه برابر است با:

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \quad , \qquad E[\bar{X}]=\mu \quad , \qquad \bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)\ \ as\ \ n\to\infty$$

- واریانس نمونه به صورت زیر تعریف می‌شود:

$$S^2=\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n-1} \quad , \qquad E[S^2]=\sigma^2$$

- دیدیم که $\bar{X}$ و S^2 تخمینگرهای ناریب از μ و σ^2 هستند.

تخمین نقطه‌ای (Point Estimation)

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

$$(X_1,X_2,\ldots,X_n)$$

به تخمین پارامتر θ از توزیع این متغیر تصادفی می‌پردازیم:

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

- $g(X_1,X_2,\ldots,X_n)$ تابعی از متغیرهای تصادفی بوده و لذا خود یک متغیر تصادفی است.
  - برای مثال $\hat{\theta}=\bar{X}$ تخمینی از $\theta=\mu$ در توزیع نرمال $X\sim N(\mu,\sigma^2)$ است.
- تخمین ما بی‌غرض (unbiased) یا ناریب خواهد بود، اگر:

$$E\left(\hat{\theta}\right)=\theta$$

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

روش گشتاور (Method of Moments)

- می‌دانیم گشتاور k-ام متغیر تصادفی X برابر است با:

$$m_k = E(X^k)$$

- فرض کنید متغیرهای تصادفی X_i نمونه‌ای از جامعه آماری با توزیع X باشند، بنابراین X_iها i.i.d. هستند و $X_i \sim X$.
- گشتاور نمونه (sample moment) مرتبه k-ام را به صورت زیر تعریف می‌کنیم:

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

- در واقع $\widehat{m}_k$ تخمینی از گشتاور حقیقی m_k است.
- در روش گشتاور با استفاده از تخمین $\widehat{m}_k$ برای گشتاور m_k، پارامتر θ را تخمین می‌زنیم.

مثال ۱

- فرض کنید متغیر تصادفی X دارای توزیع نمایی با پارامتر λ باشد:

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

می‌دانیم که:

$$m_1 = E(X) = \frac{1}{\lambda}$$

یا:

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

پس با رویت نمونه‌های X_i داریم:

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

ادامه مثال ۱

○ فرض کنید نمونه‌هایی که از یک جامعه با توزیع نمایی برداشته‌ایم دارای مقادیر 1, 4, 2, 2, 1, 5, 2, 7 هستند. تخمین پارامتر این توزیع به روش گشتاور چقدر است؟

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

$$= \frac{1}{\frac{1}{8}(1+4+2+2+1+5+2+7)}$$

$$= \frac{1}{3}$$

مثال ۲

○ متغیر تصادفی X دارای توزیع نرمال با میانگین و واریانس نامعلوم است. می‌خواهیم با رویت مقادیر نمونه‌های آن، میانگین و واریانس را تخمین بزنیم. می‌دانیم که:

$$\mu = m_1 \;, \qquad \sigma^2 = m_2 - m_1^2$$

پس:

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

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

مثال ۳

- توزیع پرتو (Pareto) نقش پررنگی در مدل‌سازی ترافیک اینترنت ایفا می‌کند.
- تابع توزیع احتمال پرتو به شکل زیر است:

$$F_X(x) = 1 - \left(\frac{x}{\sigma}\right)^{-\theta} \; : \; x > \sigma$$

در صورت نامعلوم بودن σ و θ، چطور باید آنها را تخمین بزنیم؟

- تا کنون با توزیع پرتو مواجه نشده‌ایم، پس ابتدا نیاز داریم تا گشتاور مرتبه اول و دوم آن را محاسبه کنیم:

$$f(x) = F'(x) = \frac{\theta}{\sigma}\left(\frac{x}{\sigma}\right)^{-\theta-1} = \theta\sigma^{\theta}x^{-\theta-1} \; : \; x > \sigma$$

ادامه مثال ۳

$$m_1 = E(X) = \int_{-\infty}^{+\infty} x\, f(x)dx = \theta\sigma^{\theta}\int_{\sigma}^{+\infty} x^{-\theta}dx$$

$$= \theta\sigma^{\theta}\frac{x^{-\theta+1}}{-\theta+1}\Big|_{\sigma}^{+\infty} = \frac{\theta\sigma}{\theta-1} \; : \; \theta > 1$$

$$m_2 = E\left(X^2\right) = \int_{-\infty}^{+\infty} x^2\, f(x)dx = \theta\sigma^{\theta}\int_{\sigma}^{+\infty} x^{-\theta+1}dx$$

$$= \theta\sigma^{\theta}\frac{x^{-\theta+2}}{-\theta+2}\Big|_{\sigma}^{+\infty} = \frac{\theta\sigma^2}{\theta-2} \; : \; \theta > 2$$

- برای $\theta \leq 1$، متغیر تصادفی پرتو دارای امید ریاضی نامتناهی، و در صورت $\theta \leq 2$ دارای گشتاور دوم نامتناهی خواهد بود.

ادامه مثال ۳

بنابراین داریم:

$$m_1 = \frac{\theta\sigma}{\theta - 1} \quad , \qquad m_2 = \frac{\theta\sigma^2}{\theta - 2}$$

و به دست می‌آوریم:

$$\hat{\theta} = \sqrt{\frac{\widehat{m}_2 - \widehat{m}_1^2}{\widehat{m}_2}} + 1 \quad \text{and} \quad \hat{\sigma} = \frac{\widehat{m}_1(\hat{\theta} - 1)}{\hat{\theta}}$$

- بنابراین با جمع‌آوری نمونه‌هایی از یک متغیر تصادفی پرتو و محاسبه گشتاور اول و دوم این نمونه‌ها، می‌توان پارامترهای σ و θ را با استفاده از روابط بالا تخمین زد.

تخمین پایدار (Consistent Estimation)

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

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

- تعبیر: با داشتن داده بیشتر ($n \to \infty$)، تخمین از مقدار حقیقی فاصله اندکی (ϵ) بگیرد.
- تعریف بالا در حقیقت تعریف پایداری ضعیف (weak consistency) است.
- قانون ضعیف اعداد بزرگ: $\lim_{n\to\infty} P\left(|\bar{X} - \mu| < \epsilon\right) \to 1$
  - $\bar{X}$ همواره تخمین پایداری از μ است.
- تخمینگر گشتاور به طور کلی یک تخمینگر پایدار است.

روش درست‌نمایی بیشینه (Maximum Likelihood)

- فرض کنید می‌خواهیم پارامتر θ از توزیع X را تخمین بزنیم.
- به فرض معلوم بودن θ از X نمونه‌برداری میکنیم: $X_i|\theta$
- می‌خواهیم شانس مشاهده مجموعه داده $(x_1, x_2, \dots, x_n)$ در نمونه $(X_1|\theta, \dots, X_n|\theta)$ را مشخص کنیم.
- واضح است که این شانس به تابع چگالی احتمال مشترک $f_{X_1 \dots X_n}(x_1, \dots, x_n|\theta)$ است و از آنجایی که X_i ها مستقل از هم هستند:

$$f(x_1, \dots, x_n|\theta) = f(x_1|\theta) f(x_2|\theta) \dots f(x_n|\theta)$$

- تابع درست‌نمایی (likelihood function) $L(\theta)$ را به صورت زیر تعریف می‌کنیم:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

تخمین درست‌نمایی بیشینه

- توجه کنید که تابع درست‌نمایی $L(\theta)$ به طور شهودی شانس مشاهده مجموعه داده $(x_1, x_2, \dots, x_n)$ را به ازای یک مقدار θ معلوم نشان می‌دهد.
- تخمینگر درست‌نمایی بیشینه (maximum likelihood estimator) که آن را با $\hat{\theta}_{ML}$ نمایش می‌دهیم، در حقیقت مقداری از پارامتر θ است که تابع $L(\theta)$ را بیشینه می‌کند.

$$\hat{\theta}_{ML} = \arg\max_{\theta} L(\theta)$$

- به عبارت دیگر به ازای چه مقداری از θ احتمال مشاهده $(x_1, \dots, x_n)$ بیشترین است.

تخمین ML

- جهت یافتن $\hat{\theta}_{ML}$ باید از تابع $L(\theta)$ نسبت به θ مشتق بگیریم:

$$\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f(x_i|\theta) = 0 \;\rightarrow\; \hat{\theta}_{ML}$$

- اما از آنجا که $L(\theta)$ یک تابع ضربی است، مشتق آن پیچیده می‌شود.
- به همین دلیل غالباً از لگاریتم آن که یک تابع جمعی می‌شود استفاده می‌کنیم:

$$LL(\theta) = \ln L(\theta) = \ln\prod_{i=1}^{n} f(x_i|\theta) = \sum_{i=1}^{n} \ln f(x_i|\theta)$$

- تابع $LL(\theta)$ را تابع درست‌نمایی لگاریتمی (log-likelihood function) می‌نامیم.

تخمین ML

- با توجه به صعودی اکید بودن تابع لگاریتم $\ln(.)$ ، مقداری از θ که تابع $LL(\theta)$ را بیشینه می‌کند، تابع $L(\theta)$ را هم بیشینه خواهد کرد:

$$\arg\max_{\theta} LL(\theta) = \arg\max_{\theta} L(\theta) = \hat{\theta}_{ML}$$

مثال ۱. برای متغیر تصادفی برنولی $X\sim Ber(p)$ ، تخمین $\hat{p}_{ML}$ را محاسبه کنید.

تابع چگالی احتمال برنولی به شرط معلوم بودن p برابر است با:

$$f(x_i|p) = p^{x_i}(1-p)^{1-x_i}$$

به فرض داشتن n مشاهده $(x_1, \dots, x_n)$ داریم:

مثال ۱

$$L(p) = \prod_{i=1}^{n} f(x_i|p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}\,(1-p)^{\sum_{i=1}^{n}(1-x_i)}$$

$$LL(p) = \left(\sum_{i=1}^{n} x_i\right)\ln p + \left(\sum_{i=1}^{n}(1-x_i)\right)\ln(1-p)$$

$$\frac{\partial LL(p)}{\partial p} = \left(\sum_{i=1}^{n} x_i\right)\left(\frac{1}{p}\right) + \left(\sum_{i=1}^{n}(1-x_i)\right)\left(-\frac{1}{1-p}\right) = 0$$

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

مثال ۲

- توزیع ویبول (Weibull) دارای تابع چگالی احتمال زیر است:

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

تخمین ML برای پارامتر c را محاسبه نمایید.

$$L(c) = \prod_{i=1}^{n}(cx_i\, e^{-c\frac{x_i^2}{2}}) = c^n\left(\prod_{i=1}^{n} x_i\right)e^{-\frac{c}{2}\sum_{i=1}^{n} x_i^2}$$

$$\Rightarrow LL(c) = n\ln c + \sum_{i=1}^{n}\ln x_i - \frac{c}{2}\sum_{i=1}^{n} x_i^2$$

ادامه مثال ۲

- پس خواهیم داشت:

$$\frac{\partial LL(c)}{\partial c}=\frac{n}{c}-\frac{1}{2}\sum_{i=1}^{n}x_i^2=0 \quad \Rightarrow \quad \hat{c}_{ML}=\frac{2n}{\sum_{i=1}^{n}x_i^2}$$

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

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

- بهترین توزیع ویبول که مقادیر فوق را مدل کند عبارت است از:

$$f_X(x)=0.48x\,e^{-0.24x^2}$$

مثال ۳

- تابع چگالی احتمال پواسون به شکل زیر است:

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

می‌خواهیم پارامتر λ را به روش ML تخمین بزنیم.

$$\mathrm{L}(\lambda)=\prod_{i=1}^{n}e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}=\frac{\left(e^{-\lambda}\right)^n\lambda^{\sum_{i=1}^{n}x_i}}{\left(\prod_{i=1}^{n}x_i!\right)}$$

$$\Rightarrow \mathrm{LL}(\lambda)=-n\lambda+\ln\lambda\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\ln(x_i!)$$

ادامه مثال ۳

- پس خواهیم داشت:

$$\frac{\partial LL(\lambda)}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0$$

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

- برای پارامتر λ در متغیر تصادفی پواسون، در صورتی که از روش گشتاور نیز استفاده کنیم به همین تخمین خواهیم رسید.

تخمین ML بیش از یک پارامتر

- فرض کنید می‌خواهیم پارامترهای μ و σ از توزیع نرمال را تخمین بزنیم.
- به این منظور باید از تابع درست‌نمایی یک بار نسبت به μ و بار دیگر نسبت به σ مشتق بگیریم و با حل دستگاه دو معادله و دو مجهول تخمین مناسب را محاسبه کنیم.

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

$$LL(\mu,\sigma) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) = \sum_{i=1}^{n}\left(-\ln\left(\sqrt{2\pi}\sigma\right) - \frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

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

تخمین ML بیش از یک پارامتر

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

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

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

تخمین ML توزیع یکنواخت

- برای تخمین ML پارامترهای توزیع یکنواخت $U(\alpha,\beta)$ داریم:

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

$$LL(\alpha,\beta)=\ln\left(\frac{1}{\beta-\alpha}\right)^n=-n\ln(\beta-\alpha)\ :\ \ \alpha<x_1,x_2,\dots,x_n<\beta$$

$\frac{\partial LL}{\partial \alpha}=\frac{n}{\beta-\alpha}=0\ \Rightarrow$ این معادله جواب ندارد

- اما از آنجا که شرط $\alpha<x_1,x_2,\dots,x_n<\beta$ باید ارضا شود:

$$\hat{\alpha}_{ML}=\min(x_1,x_2,\dots,x_n)\qquad \hat{\beta}_{ML}=\max(x_1,x_2,\dots,x_n)$$

مثال

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

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

- داریم:

$$L(\theta_1,\theta_2) = \prod_{i=1}^{n} f(x_i|\theta_1,\theta_2) = \prod_{i=1}^{n} \theta_1\theta_2^{\theta_1}x_i^{-\theta_1-1}$$

$$= \theta_1^n\theta_2^{n\theta_1}\left(\prod_{i=1}^{n} x_i\right)^{-(\theta_1+1)} \quad : \theta_2 \le x_1, x_2, \dots, x_n$$

$$LL(\theta_1,\theta_2) = n\ln(\theta_1) + n\theta_1\ln(\theta_2) - (\theta_1+1)\sum_{i=1}^{n}\ln(x_i)$$

ادامه مثال

$$\frac{\partial \mathrm{LL}}{\partial\theta_1} = \frac{n}{\theta_1} + n\ln(\theta_2) - \sum_{i=1}^{n}\ln(x_i) = 0 \;\Rightarrow\; \hat{\theta}_{1ML} = \frac{n}{\sum_{i=1}^{n}\ln(x_i) - n\ln\hat{\theta}_{2ML}}$$

$$\frac{\partial \mathrm{LL}}{\partial\theta_2} = \frac{n\theta_1}{\theta_2} > 0$$

- مشتق نسبت به θ_2 همواره مثبت است، بنابراین θ_2 باید تا جای ممکن افزایش یابد، و با توجه به شرط $\theta_2 \le x_1,x_2,\dots,x_n$ خواهیم داشت:

$$\hat{\theta}_{2ML} = \min\{x_1, x_2, \dots, x_n\}$$

و در نتیجه:

$$\hat{\theta}_{1ML} = \frac{n}{\sum_{i=1}^{n}\ln(x_i) - n\ln(\min\{x_1, x_2, \dots, x_n\})}$$