برآورد پارامتر

همانطور که سابقا به تعریف پارامتر در جامعه آماری پرداختیم، تمام تلاش‌های محققین معطوف به معلوم کردن مقدار پارامتر در جامعه می باشد. بدین منظور از مشاهدات استفاده می شود تا حداقل برآورد خوبی از مقدار واقعی پارامتر ارائه دهد. بمنظور انجام این مهم دو رویکرد وجود دارد

1. **بدون هیچ اطلاعی از جامعه آماری**: در این حالت از ابزارهای ارائه شده در مبحث آمار توصیفی و بصورت پیشرفته‌تر آن از روشهای ناپارامتری بهره برده می شود.

2. **دانستن توزیع مشاهدات:** یعنی توزیع حاکم بر مشاهدات مشخص می باشد. روش‌های توسعه داده شده در علم آمار با دانستن توزیع حاکم بر مشاهدات را روشهای پارامتری می نامند. همچنین برآورد حاصل از این روش‌ها تحت عنوان برآورد پارامتری معروف هستند. این روشها دارای انواع مختلف هستند و از معروف‌ترین آنها روش برآورد ماکزیمم درستنمایی[1] (**MLE**) است که در ادامه شرح و با ارائه چند مثال معرفی خواهد شد.

روش برآورد ماکزیمم درستنمایی: بمنظور معرفی این روش برآورد، در ابتدا فرض می کنیم مشاهدات از توزیعی با یک پارامتر بدست آمده باشند. برای تشریح بیشتر گوییم، هرگاه با بهره از روش‌های نمونه‌گیری مناسب یک نمونه n تایی از جامعه‌ای با تابع توزیع $F_X(x,\theta)$ داشته باشیم به منظور برآورد پارامتر θ بهترین ایده آن است که θ را به گونه‌ای انتخاب نماییم که احتمال مشاهده آن نمونه بیشترین مقدار ممکن باشد. به روش مبتنی بر این ایده روش برآورد **ماکزیمم درستنمایی** گویند.

$$let \quad X_1,\dots,X_n \sim F_X(x,\theta) \quad then$$

$$L(\theta)=\begin{cases} p(X_1=x_1,\dots,X_n=x_n\ ;\ \theta) & \text{ها گسسته}X_i \\ \\ f_{X_1,\dots,X_n}(x_1,\dots,x_n\ ;\ \theta) & \text{ها پیوسته}X_i \end{cases}$$

$$\xRightarrow{if\ X_i\ are\ independent} \begin{cases} \prod_{i=1}^{n} p(X_i=x_i\ ;\ \theta) & \text{ها گسسته}X_i \\ \prod_{i=1}^{n} f_{x_i}(x_i\ ;\ \theta) & \text{ها پیوسته}X_i \end{cases}$$

---

[1] Maximum Likelihood Estimator

تابع $L(\theta)$ معروف است به **تابع درستنمایی**. حال می بایست پارامتر θ بگونه‌ای تعیین شود که تابع $L(\theta)$ را بیشینه نماید که بدین منظور از روشهای ماکزییم‌یابی همانند مشتق یا سایر روش‌ها استفاده می نماییم. بمنظور درک بهتر این روش، در ادامه مثال‌های مختلفی را ارائه می نماییم.

توجه: برآورد حاصل از روش برآورد درستنمایی ماکزیمم برای پارامتر θ را بصورت $\hat{\theta}$ نمایش می دهیم و تابع حاصل از مشاهدات (یعنی تابع مرتبط با $\hat{\theta}$)، حتما می بایست بصورت متغیر تصادفی نوشته شود.

Example: $let\ X_1, \dots, X_n \sim Ber(p).\ find\ MLE\ of\ p\ when\ X_i\ are\ independent.$

$$L(p) = p(X_1 = x_1, \dots, X_n = x_n\ ; p) \overset{ind}{\Longrightarrow} L(p) = \prod_{i=1}^{n} p(X_i = x_i\ ; p)$$

$$= \prod_{i=1}^{n} p^{x_i}\,(1-p) \quad = p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i}.$$

به دنبال تعیین مقداری از p هستیم که تابع فوق را ماکزیمم نماید. بسیار پرواضح است که تابع $L(p)$ نسبت به پارامتر p یک تابع پیوسته بر روی بازه $(0,1)$ می باشد، بنابراین به منظور ماکزیمم‌یابی از مشتق عبارت بالا نسبت به p و برابر صفر قرار دادن آن، استفاده می نماییم. بدین منظور تعریف می کنیم $m = \sum_{i=1}^{n} X_i$، آنگاه

$$\frac{d}{dp}L(p) = \frac{d}{dp}\left(p^m(1-p)^{n-m}\right)$$
$$= mp^{m-1}(1-p)^{n-m} + p^m(n-m)(1-p)^{n-m-1}(-1) = 0$$
$$\Rightarrow mp^{m-1}(1-p)^{n-m} = p^m(n-m)(1-p)^{n-m-1} \Rightarrow p = \frac{m}{n}.$$

بنابراین برآورد MLE برابر خواهد بود با

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

Example: $X_1, \dots, X_n \sim pow(\alpha, \beta)\ such\ that\ (s.t.)\ \alpha\ is\ known. Find\ MLE\ of\ \beta.$

$$if\ X \sim pow(\alpha, \beta), then\ f_X(x) = \frac{\alpha x^{\alpha-1}}{\beta^\alpha} \quad 0 < x < \beta.$$

$$L(\beta) = f_{X_1,\dots,X_n}(x_1,\dots,x_n;\ \beta) \overset{ind}{\Longrightarrow} L(\beta) = \prod_{i=1}^{n} f_X(x_i,\beta) = \prod_{i=1}^{n} \frac{\alpha x_i^{\alpha-1}}{\beta^{\alpha}}\ I_{(0,\beta)}(x_i)$$
$$= \frac{\alpha^n\left(\prod_{i=1}^{n} x_i\right)^{\alpha-1}}{\beta^{n\alpha}}\ \prod_{i=1}^{n} I_{(0,\beta)}(x_i).$$

تابع $I_A(x)$ معروف است به تابع نشانگر[2]، به بیان دیگر این تابع بصورت زیر تعریف می شود

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

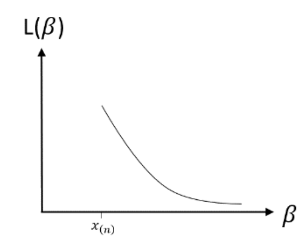
با دانستن این تعریف خواهیم داشت

$$\prod_{i=1}^{n} I_{(0,\beta)}(x_i) = \begin{cases} 1 & \forall i\ 0 < x_i < \beta \\ 0 & o.w. \end{cases} = \begin{cases} 1 & x_{(n)} < \beta \\ 0 & o.w. \end{cases} = I_{(x_{(n)},\infty)}(\beta).$$

با در نظر گرفتن $c = \alpha^n\left(\prod_{i=1}^{n} x_i\right)^{\alpha-1}$، خواهیم داشت

$$L(\beta) = \frac{c}{\beta^{n\alpha}} I_{(x_{(n)},\infty)}(\beta).$$

از عبارت بالا نمیتوان مشتق گرفت اما میدانیم که تابع $L(\beta)$ نسبت به پارامتر β یک تابع نزولی است، بنابراین



این بدان معنا است که ماکزیمم مقدار تابع فوق ابتدای بازه، یعنی $x_{(n)}$، رخ می دهید، در نتیجه برآورد MLE پارامتر β برابر خواهد بود با

$$\hat{\beta} = X_{(n)}.$$

توجه: در ماکزیمم سازی فوق از مشتق‌گیری استفاده نشد، دلیل این موضوع آن است که پارامتر مدنظر در کران تکیه‌گاه متغیر تصادفی قرار داشت.

---

[2] indicator

گاهی مشتق گرفتن از تابع درستنمایی کار ساده‌ای نیست. در این مواقع چون خود نقطه ماکزیمم اهمیت دارد و نه مقدار آن، ازln تابع درستنمایی مشتق می‌گیریم، زیرا این تابع یکنوای صعودی است. بدین منظور خواهیم داشت

$$l(\theta)=\ln\big(L(\theta)\big).$$

Example: $X_1,\dots,X_n\sim pow(\alpha,\beta)\ s.t.\ \beta\ is\ known, find\ MLE\ of\ \alpha.$

$$L(\alpha)=\prod_{i=1}^{n}\frac{\alpha x_i^{\alpha-1}}{\beta^{\alpha}}=\frac{\alpha^n\left(\prod_{i=1}^{n}x_i\right)^{\alpha-1}}{\beta^{n\alpha}}.$$

$$l(\alpha)=\ln\big(L(\alpha)\big)=n\ln(\alpha)+(\alpha-1)\sum_{i=1}^{n}\ln(x_i)-n\alpha\ln(\beta).$$

$$\frac{d}{d\alpha}\left(n\ln(\alpha)+(\alpha-1)\sum_{i=1}^{n}\ln(x_i)-n\alpha\ln(\beta)\right)=0.$$

$$\Rightarrow\frac{n}{\alpha}+\sum_{i=1}^{n}\ln(x_i)-n\ln(\beta)=0\Rightarrow\ \alpha=\frac{n}{n\ln(\beta)-\sum_{i=1}^{n}\ln(x_i)}.$$

بنابراین برآورد ماکزیمم درستنمایی پارامتر α برابر خواهد بود با

$$\hat{\alpha}=\frac{n}{n\ln(\beta)-\sum_{i=1}^{n}\ln(X_i)}.$$

برآورد ماکزیمم درستنمایی برای بیش از یک پارامتر: هرگاه تعداد پارامتر مجهول یک توزیع بیش از یک عدد باشد و این پارامترها در فضاهای نامرتبط مقدار بپذیرند (به هم وابسته نباشند) آنگاه به منظور برآورد MLE پارامترها با فرض معلوم بودن سایر پارامترها، اقدام به برآورد یکی از پارامترها می‌نماییم و در گام دوم با در نظر گرفتن مقدار برآورد به دست آمده برای پارامتر اول، پارامترهای بعدی را از با بهره از همین روش‌ها انجام می دهیم. بمنظور درک مسیر فوق به مثال‌های زیر توجه نمایید.

نکته: روش فوق بمنظور برآورد ماکزیمم درستنمایی برای چند پارامتر یک روش کلی نیست بلکه تنها زمانی کاربرد دارد که فضای پارامترهای توزیع وابسته به هم نباشند (به بیان دیگر فضاهای تعریف پارامترها از یکدیگر مستقل باشند). در حالت کلی می توان از روشهای ماکزیمم‌سازی معمول همانند مشتق‌های توام استفاده نمود.

Example: $X_1, \dots, X_n \sim pow(\alpha, \beta)$, find MLE if α, β both are unknown (we know parameter spaces are $\alpha > 0, \beta > 0$).

$$L(\alpha,\beta) = f_{X_1,\dots,X_n}(x_1,\dots,x_n;\ \alpha,\beta) \overset{ind}{\Longrightarrow} L(\alpha,\beta)$$
$$= \prod_{i=1}^{n} f_{X_i}(x_i;\ \alpha,\beta) = \prod_{i=1}^{n} \frac{\alpha x_i^{\alpha-1}}{\beta^\alpha} I_{(0,\beta)}(x_i)$$

Let α is known, so $\hat{\beta} = X_{(n)}$,then

$$l(\alpha,\hat{\beta}) = \ln\left(L(\alpha,\hat{\beta})\right) = n\ln(\alpha) + (\alpha-1)\sum_{i=1}^{n}\ln(x_i) - n\alpha\ln(\hat{\beta})$$

$$\Rightarrow \hat{\alpha} = \frac{n}{n\ln(\mathrm{X}_{(\mathrm{n})}) - \sum_{i=1}^{n}\ln(x_i)}.$$

Example: Let $X_1, \dots, X_n$ have distributed as $N(\mu, \sigma^2)$ $s.t.$ they are independent and μ and σ^2 are both unknown. Find MLE of (μ, σ^2).

$$L(\mu,\sigma^2) = f_{X_1,\dots,X_n}(x_1,\dots,x_n;\ \mu,\sigma^2) \overset{ind}{\Longrightarrow} L(\mu,\sigma^2)$$
$$= \prod_{i=1}^{n} f_{X_i}(x_i;\ \mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}}.$$

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

$let\ \sigma^2\ be\ known, then$

$$\frac{d}{d\mu}\left(l(\mu,\sigma^2)\right) = \frac{d}{d\mu}\left(\frac{-n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}\right) = 0.$$

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

این بدان معنا است که برآورد MLE میانگین جامعه توزیع نرمال برابر خواهد بود با

$$\hat{\mu} = \bar{X}.$$

The MLE of $\mu\ is\ \bar{X}, then$

$$\frac{d}{d\sigma^2}\left(l(\mu,\sigma^2)\right)=\frac{d}{d\sigma^2}\left(\frac{-n}{2}\ln(2\pi\sigma^2)-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}\right)=0.$$

$$\Longrightarrow \frac{-n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\hat{\mu})^2=0.$$

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

Example: let $X_1,\dots,X_n$ be distributed as $\beta(\alpha,b)\ s.t., b=2$. Find MLE of α.

می‌دانیم برای توزیع بتا با پارامترهای (α,β) داریم

$$f_X(x)=\frac{x^{\alpha-1}(1-x)^{2-1}}{\beta(\alpha,2)}=\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}x^{\alpha-1}(1-x);\quad 0<x<1.$$

$$L(\alpha)=f_{X_1,\dots X_n}(x_1,\dots,x_n;\alpha)\overset{ind}{\Longrightarrow}L(\alpha)=\prod_{i=1}^{n}f_{X_i}(x_i;\alpha).$$

$$=\prod_{i=1}^{n}\alpha(\alpha-1)x_i^{\alpha-1}(1-x_i)=\left(\alpha(\alpha-1)\right)^n\left(\prod_{i=1}^{n}x_i\right)^{\alpha-1}\left(\prod_{i=1}^{n}(1-x_i)\right).$$

$$\Longrightarrow l(\alpha)=\ln\left(L(\alpha)\right)=n\ln(\alpha^2+\alpha)+(\alpha-1)\ln(\prod_{i=1}^{n}x_i)+\ln(\prod_{i=1}^{n}(1-x_i))$$

$$\Longrightarrow \frac{d}{d\alpha}\left(l(\alpha)\right)=\frac{n(2\alpha+1)}{\alpha^2+\alpha}+\ln(\prod_{i=1}^{n}x_i)+0=0.$$

$$\Rightarrow\frac{2\alpha+1}{\alpha^2+\alpha}=\frac{-\sum_{i=1}^{n}\ln(x_i)}{n}-\sum_{i=1}^{n}\ln(x_i)=t.$$

$$\Rightarrow\ t\alpha^2+(t-2n)\alpha-n=0.$$

$$\hat{\alpha}=\frac{2n+\sum_{i=1}^{n}\ln(X_i)+\sqrt{\left(\sum_{i=1}^{n}\ln(X_i)\right)^2+4n^2}}{-2\sum_{i=1}^{n}\ln(X_i)}.$$