Math 152 2-12-09

Method of Maximum Likelihood

With the same set up as for the method of moments examples.

$X_1, \ldots, X_n$ ~~i.i.d.~~ i.i.d

have joint density

$f(x_1, \ldots, x_n \mid \vec{\theta})$

where $\vec{\theta}$ is some parameter (vector).

We define the *likelihood* of θ as a fcn of $x_1, \ldots, x_n$.

$\operatorname{lik}(\theta) = f(x_1, \ldots, x_n \mid \theta)$

In the discrete case this is simply the probability of observing

$X_1 = x_1, \ldots, X_n = x_n$

if the true parameter value is θ.

In the continuous case.

$$\text{Prob}\{x_1 \le X_1 \le x_1+\Delta x_1, \ \ldots \ x_n \le X_n \le x_n + \Delta x_n\}$$

$$\approx f(x_1, \ldots, x_n \mid \theta)\, \Delta x_1 \ldots \Delta x_n.$$

and in either case we see that it is natural to ask which value of θ maximizes the "likelihood" that the observed data $(x_1, \ldots, x_n)$ occur.

When the X_i are i.i.d then

$$lik(\theta) = \prod_{i=1}^{n} f(X_i \mid \theta)$$

and it is more convenient to consider

$$l(\theta) = \log(lik(\theta)) = \sum_{i=1}^{n} \log(f(X_i \mid \theta)).$$

then ↗ (in the one parameter case).

$$l'(\theta) = \sum_{i=1}^{n} \frac{\frac{d}{d\theta}(f(X_i|\theta))}{f(X_i|\theta)}.$$

and we try to solve $l'(\theta)=0$ to find the value of θ which maximizes $l(\theta)$.

Notice that maximizing $l(\theta)$ is the same problem as maximizing $lik(\theta)$ because log is an increasing function

---

e.g. X is Poisson w/ parameter λ

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

$X_1, \ldots, X_n$ are i.i.d. $p(\lambda, k)$.

$$lik(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i}}{(X_i)!} e^{-\lambda}$$

$$l(\lambda) = \sum_{i=1}^{n} \left(X_i \log\lambda - \lambda - \log(X_i!)\right).$$

$$= \log\lambda\left(\sum_{i=1}^{n} X_i\right) - n\lambda - \sum_{i=1}^{n} \log(X_i)!$$

$$\ell'(\lambda) = \frac{1}{\lambda}\left(\sum_{i=1}^{n} X_i\right) - n.$$

$$\ell''(\lambda) = -\frac{1}{\lambda^2}\underbrace{\left(\sum_{i=1}^{n} X_i\right)}_{\geq 0.} \leq 0.$$

So $l(\lambda)$ is maximized at $\lambda = \frac{1}{n}\sum_{i=1}^{n} X_i$.

and $\hat{\lambda} = \bar{X}$

which is the same as the method of moments estimate.

e.g. If $X_1, \ldots X_n$ are i.i.d.

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

$$lik(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{X_i - \mu}{\sigma}\right)^2}$$

$$l(\mu,\sigma) = \sum_{i=1}^{n}\left(-\log\sigma - \frac{1}{2}\log 2\pi - \frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2\right)$$

$$= -n\log\sigma - n\frac{1}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)$$

$$\frac{\partial l}{\partial \sigma} = -n/\sigma + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2$$

Note that: $l(\mu,\sigma) \to -\infty$ if $\sigma \to +\infty$

or if $\sigma \to 0^+$

(since $\frac{1}{\sigma^2}$ beats $\log\sigma$.)

also $l(\mu,\sigma) \to -\infty$ if $\mu \to \pm\infty$.

Solving the system above we get

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

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

since this is the only critical point, it must be a maximum.

The m.l.e. estimates for μ, σ are the same as for the method of moments.

e.g. $X_1, \dots, X_n$ are i.i.d. Gamma(α, λ)

$$f(x \mid \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}$$

$0 \le x < \infty.$

$$\text{lik}(\alpha, \lambda) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} X_i^{\alpha-1} e^{-\lambda X_i}$$

$$\cancel{\ell(\alpha, \lambda) = \sum_{i=1}^{n} \frac{1}{\Gamma(\alpha)} \lambda^{\alpha}}$$

$$\ell(\alpha, \lambda) = \sum_{i=1}^{n} - \log \Gamma(\alpha) + \alpha \log \lambda + (\alpha - 1) \log X_i - \lambda X_i$$

$$= n\alpha \log \lambda + (\alpha - 1)\sum_{i=1}^{n} \log X_i - \lambda\left(\sum_{i=1}^{n} X_i\right).$$
$$- n \log \Gamma(\alpha).$$

$$\frac{\partial l}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i.$$

$$\frac{\partial l}{\partial \alpha} = n \log \lambda + \sum_{i=1}^{n} \log X_i - n\,\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}.$$

Note that $l(\alpha, \lambda) \to -\infty$ if $\lambda \to 0$ or $\lambda \to +\infty$ with fixed α.

also $l(\alpha, \lambda) \to -\infty$ if $\alpha \to +\infty$ or if $\alpha \to 0$.

$\Gamma(\alpha) \sim C \cdot \alpha^{\alpha} e^{-\alpha}$ by Stirling as $\alpha \to +\infty$

Fact $\Gamma(\alpha) \sim \frac{c}{\alpha}$ as $\alpha \to 0^+$.

"Solving" the system gives

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

$$n \log \hat{\alpha} - n \log \bar{X} + \sum_{i=1}^{n} \log X_i - n \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})}$$

$$= 0.$$

So to find $\hat{\alpha}, \hat{\lambda}$ we will need a numerical root finding procedure.

(we get $\hat{\alpha} \approx .441$
$\hat{\lambda} \approx 1.96.$).

In this situation it appears difficult to derive any closed form expression for the ~~variance~~ distributions of $\hat{\alpha}$ and $\hat{\lambda}$

Recall the Bootstrap methodology:

→ if we knew the "true" values of $\alpha = \alpha_0$ and $\lambda = \lambda_0$
we could draw many (say 1000) samples of size $n = 227$ from a gamma(α_0, λ_0)

→ distribution. Treating each such example as a separate instance of our data collection we would form the m.l.e. estimate $\hat{\alpha}_i^*$, $\hat{\lambda}_i^*$ $i = 1, \ldots 1000$.

→ Then a histogram of $\hat{\alpha}_i^*$ would give an approximation to the distribution of $\hat{\alpha}$ and likewise for $\hat{\lambda}$.

→ Since we don't know the true values we use our original $\hat{\alpha}$, $\hat{\lambda}$ instead and proceed the same way.