CPS 5310: Parameter Estimation

Natasha Sharma, Ph.D.

Example

Suppose our task is to determine the net income for year 2019 based on the net incomes given below

Year	Net Income
2016	48.3 million
2017	90.4 million
2018	249.9 million

Last lecture we tried to answer this with

- the use of a polynomial interpolating three points,
- least squares method
- using MATLAB's *polyfit* function

Example

Suppose our task is to determine the net income for year 2019 based on the net incomes given below

Year	Net Income
2016	48.3 million
2017	90.4 million
2018	249.9 million

Last lecture we tried to answer this with

- the use of a polynomial interpolating three points,
- least squares method
- using MATLAB's *polyfit* function

Example

Suppose our task is to determine the net income for year 2019 based on the net incomes given below

Year	Net Income
2016	48.3 million
2017	90.4 million
2018	249.9 million

Last lecture we tried to answer this with

- the use of a polynomial interpolating three points,
- least squares method
- using `MATLAB`'s *polyfit* function

Summary of MATLAB commands


```
clear all; clf;
year=[0,1,2];
income=[48.3, 90.4, 249.9];
plot(year, income, ‘ro’)
```

For regression, we use

```
polyfit(year, income, N)
```

N is the degree of the polynomial used for curve fitting. The output represents the decreasing coefficients of the polynomial.
In the Figure 1, choose **Tools**, **Basic Fitting**.

Year	Crime Rate	Unemployment
1994	5373.5	6.1 %
1995	5277.6	5.6 %
1996	5086.6	5.4 %
1997	4922.7	4.9 %
1998	4619.3	4.5%
1999	4266.8	4.2 %
2000	4124.8	4%
2001	4160.5	4.8%

Summary of MATLAB commands

Natasha Sharma, Ph.D.

```
clear all; clf;
scatterplot(X);
```

Generate the best fit curves for crime-unemployment data

Regression with more general functions

Consider the following unconstrained optimization problem

$$\min_p E(p) := \sum_{j=1}^{k} |T_j - \sin(m_j - p)|^2.$$

Computing the optimality condition yields:

$$2\sum_{j=1}^{k} \Big(T_j - \sin(m_j - p)\Big)\cos(m_j - p) = 0.$$

Summary of MATLAB commands

```
function value = bigbend(a);
scaledtemp=[-1,-0.71,...-0.93];
value=0;
for j=1:12
m=2*pi*j/12;
value= value +
(scaledtemp(j)-sin(m-p))*cos(m-p);
end
```

Solving for p

```
months=1:12;
temps=[60.9 ...62.2];
p=fzero(@bigbend,1)
modeltemps=18*sin(2*pi*months/12-p)+78.9;
plot(months,temps,'o',months,modeltemps)
```

Lotka-Volterra Prey-Predator Model

Example

Consider the following model capturing the interaction of the species $x(t)$ and $y(t)$:

$$\frac{dx}{dt} = ax - bxy$$
$$\frac{dy}{dt} = -ry + cxy,$$

where a, b, c and r are non-negative parameters.

Can you identify which species is the prey and which is the predator?

Interaction between Lynx and Hare

Year	Lynx	Hare	Year	Lynx	Hare	Year	Lynx	Hare
1900	4	30	1907	13	21.4	1914	45.7	52.3
1901	6.1	47.2	1908	8.3	22	1915	51.1	19.5
1902	9.8	70.2	1909	9.1	25.4	1916	29.7	11.2
1903	35.2	77.4	1910	7.4	27.1	1917	15.8	7.6
1904	59.4	36.3	1911	8	40.3	1918	9.7	14.6
1905	41.7	20.6	1912	12.3	57	1919	10.1	16.2
1906	19	18.1	1913	19.5	76.6	1920	8.6	24.7

Table : Number of pelts collected by the Hudson Bay Company (in 1000s).

Goal: Based on the data given, determine the parameters a, b, r and c without finding an exact solution to the model.

Derivative Approximation Method

The predator equation under the assumption that $y(t)$ is not zero can be expressed as

$$\frac{1}{y}\frac{dy}{dt} = cx - r.$$

Treat "$\frac{1}{y}\frac{dy}{dt}$" as a single variable then c and r can be thought of as the slope and intercept of a line.
We would like to plot $\frac{1}{y}\frac{dy}{dt}$ as a function of x and fit a line through this data.

Method

Replace $\frac{dy}{dt}$ with its numerical derivative either forward or backward or central difference approximation.

$$\frac{1}{y(t^*)}\frac{y(t+h)-y(t-h)}{2h} = cx(t^*) - r,$$

where t^* could either be t or $t+h$ for explicit or implicit Euler Method respectively.
We would like to plot $\frac{1}{y}\frac{dy}{dt}$ as a function of x and fit a line through this data.

Derivative Approximation Method

Returning to the data,

Year	Lynx	Hare	Year	Lynx	Hare	Year	Lynx	Hare
1900	4	30	1907	13	21.4	1914	45.7	52.3
1901	6.1	47.2	1908	8.3	22	1915	51.1	19.5
1902	9.8	70.2	1909	9.1	25.4	1916	29.7	11.2

Assuming $t = 1$, the explicit Euler method reads

$$\frac{1}{y(t)} \frac{y(t+h) - y(t-h)}{2h} = c\ x(t) - r,$$

$$\frac{1}{6.1} \frac{9.8 - 4}{2} = 47.2c - r$$

Repeating for each year up to 1919 we obtain the system of equations that we will solve by regression.

Direct Method

Let us revisit the predator-prey model in a more general notation:

$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y}; \mathbf{p}), \quad \mathbf{y}, \ \mathbf{f} \in \mathbb{R}^n,$$

where $\mathbf{p} = (p_1, p_2, \cdots, p_m)^T \in \mathbb{R}^m$ denote the parameters. Also, assume that we have been given data at k points:

$$\mathbf{y}_i = \mathbf{y}(t_i) \quad i = 1, 2, \cdots k.$$

For the Lynx-Hare model we introduced, we have: $k = 21$, $n = 2$, $m = 4$.

$$\begin{aligned} y_1' &= p_1 y_1 - p_2 y_1 y_2 \\ y_2' &= -p_3 y_2 + p_4 y_1 y_2, \end{aligned}$$

where $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$; $\mathbf{p} = (p_1, p_2, p_3, p_4)^T$.

Direct Method

Consider the following unconstrained optimization problem

$$\min_{\mathbf{p}} E(\mathbf{p}) := \sum_{j=1}^{k} |\mathbf{y}(t_j; \mathbf{p}) - \mathbf{y}_j|^2,$$

where $\mathbf{y} = \left| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right|^2 = y_1^2 + y_2^2.$