

实验数据的拟合

袁略真

3130103964

生物信息学

浙江大学

2016 年 3 月 20 日

1 使用线性回归进行数据拟合

物理量 x, y 遵循某种未知的函数关系 $y = f(x)$. 为了确定这个目标函数关系,通常需要实验测量,获得 n 组数据. 对这些数据应用恰当的统计模型,可以描述尽可能真实的规律($y = y(x)$),也就是接近真实的函数关系($y(x) \approx f(x)$).

本节课使用的统计模型是线性回归模型 $y(x) = A_0 + A_1 x$, 选定的拟合标准为最小二乘法,也就是拟合曲线与测量数据有偏差平方和最小. 对于多选线性拟合的情形,模型做适当拓展即可.

1.1 非线性关系的线性拟合

当物理量 x, y 遵循非线性函数关系时,如果仍要使用线性回归模型的方法拟合非线性模型(如 $y = Ae^{Bx}$),需要将数据进行适当转换($Y' = \ln y, x' = x, Y' = \ln A + Bx'$).

2 程序流程

1. Generate standard deviation for average N

2. Transform the data

```
1 ind = 0;
2 if (y(N) == a * e^{bN})
3     Y' = ln y; N' = N; ind = 1; // ln Y = ln a + bx
4 else if (y(N) == a * N^b)
5     Y' = ln y; N' = ln N; ind = 2; // ln Y = ln a + b ln N
```

3. Several average calculation: $\bar{x}, \bar{y}, \bar{xy}, \bar{x^2}, n = \text{length}(Y')$;

4. Linear correlation coefficient:

```
1 ab=0; a2=0; b2=0;
2 for i=1:n:
3     a = (x_i - \bar{x});
4     b = (y_i - \bar{y});
5     ab += a*b;
6     a2 += a*a; b2 += b*b;
7 end for
8 r = a*b / (a2 * b2)^0.5
```

5. Linear fitting:

```

1 A1 = (\bar{xy} - \bar{x} * \bar{y})/(\bar{x^2} - \bar{x}^2);
2 A0 = \bar{y} - A1 * \bar{x};
3
4 tmp = 0;
5 for i = 1:n;
6     tmp += (yi - A0 - A1 * xi)^2;
7 sigma = (1/(n-2) * tmp)^0.5; //standard deviation
8 sigmaA0 = sigma * (\bar{x^2}/(n*\bar{x^2} - \bar{x}^2))^0.5;
9 sigmaA1 = sigma * (1/(n*\bar{x^2} - \bar{x}^2))^0.5

```

6. Transform back

```

1 a = e^A0;b=A1;
2
3 if(ind == 1 or ind == 2)
4     y'pred = e^(A0+A1*N'); //predicted y
5     ypred = e^{y'pred};
6     cout<<N<<y<<ypred;

```

3 拟合结果

3.1 拟合参数和标准偏差

Final Model	$y(N) = 0.0042296e^{-0.0281388N}$
cc*	-0.909336
A_0	-5.46565
A_1	-0.0281388
a	0.0042296
b	-0.0281388
σ	0.189625
σ_{A_0}	0.0270225
σ_{A_1}	0.000922264

表 1: $Y(N) = ae^{bx}$. *: correlation coefficient

Final Model	$y(N) = 0.00911777N^{-0.499482}$
cc*	-0.998902
A_0	-4.69753
A_1	-0.499482
a	0.0042296
b	-0.0281388
σ	0.0210493
σ_{A_0}	0.00300457
σ_{A_1}	0.000970073

表 2: $Y(N) = ae^{bx}$. *: correlation coefficient

3.2 图形结果

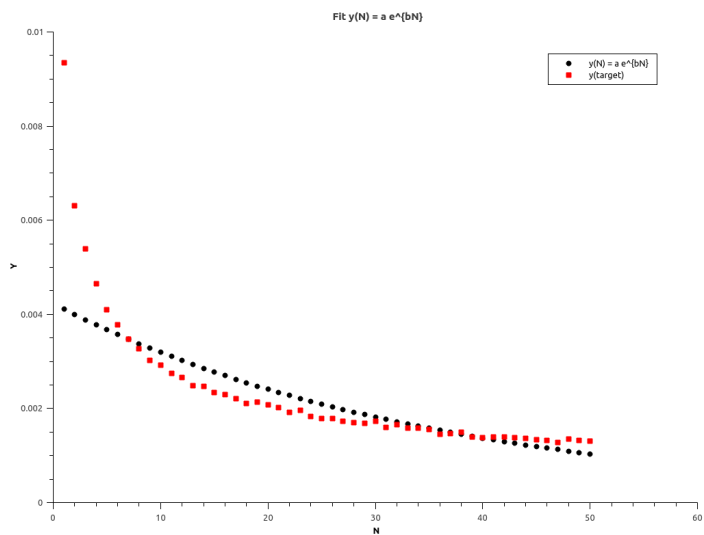


图 1: 使用 $Y(N) = ae^{bx}$ 对 $\sigma_{\bar{x}} \sim N$ 做线性拟合. 数据先调整为 $y' = \ln y, x' = x$ (线性相关系数-0.909336), 然后使用线性回归模型 $y(x) = A_0 + A_1x$ 进行拟合. 拟合直线标准偏差 0.189625.

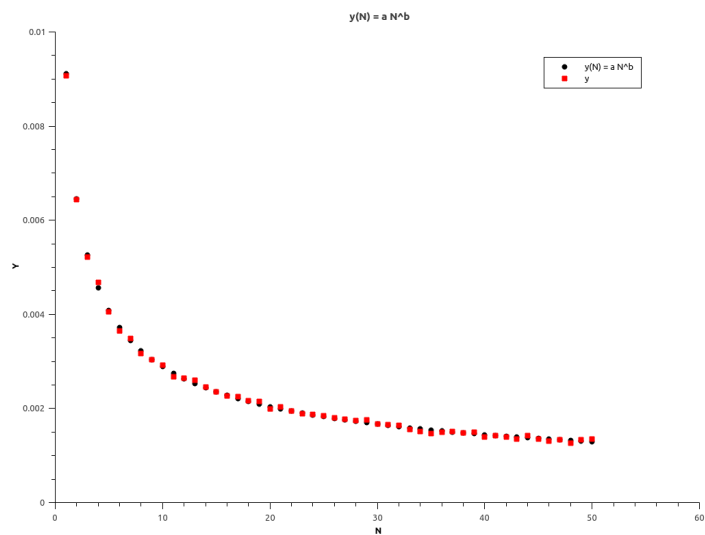


图 2: 使用 $Y(N) = aN^b$ 对 $\sigma_{\bar{x}} \sim N$ 做线性拟合. 数据先调整为 $y' = \ln y, x' = \ln x$ (线性相关系数-0.998902), 然后使用线性回归模型 $y(x) = A_0 + A_1x$ 进行拟合. 拟合直线标准偏差 0.0210493.