

Fox Module 21: Generalized linear models, concepts

(The attached PDF file has better formatting.)

*Homework assignment: maximum likelihood estimation: exponential decay*

[Note: The homework assignment is at the bottom of this posting, after the explanation of the method.]

Health claims occurring in month  $0$  and settled in month  $j$  are a percentage  $P$  of the claims open at the end of month  $j-1$ .

- No claims are settled in the month they occur.
- $P\%$  of claims open at the end of a month are expected to settle in the next month.
- Actual claims settlements are distorted by random fluctuation.

A reserving actuary estimates the percentage  $P$ .

To simplify the mathematics, we assume that if 100 claims occur in December 20X1,

- $100 \times P$  claims close in January,
- $100 \times (1 - P) \times P$  claims close in February, and so forth.

This is not exact, since if more claims close in January, fewer claims are left open, and fewer claims close in later months. But it is roughly correct, and it gives a simple solution.

For claims occurring in December 20X1, the number of claims closed by month in 20X2 are

<i>Jan</i>	<i>Feb</i>	<i>Mar</i>	<i>Apr</i>	<i>May</i>	<i>Jun</i>	<i>Jul</i>	<i>Aug</i>	<i>Sep</i>	<i>Oct</i>	<i>Nov</i>	<i>Dec</i>	<i>&gt;&gt;</i>
29	16	17	4	3	9	4	5	1	1	1	3	7

- 100 claims occur in December 20X1 (the sum of the claims closed and still open).
- Jan, Feb, Mar, ... = January, February, March, ...
- $>>$  = still open at December 31, 20X2.

We compute the maximum likelihood estimate for the percentage  $P$ .

A. Suppose there were no random fluctuation. Using the January figures and the 100 claims in December, what is the percentage  $P$ ?

*Part A:* Were there no random fluctuation, the percentage  $P$  would be constant.

- The claims occurring in December 20X1 is  $100 = 29 + 16 + 17 + \dots + 7$ .
- The percentage of claims closed in January is  $P \times 100 = 29 \Rightarrow P = 29\%$ .

This estimator is not optimal. It ignores the claims closed in later months, so it doesn't use all available information.

This estimator is unbiased, but it is not optimal.

- 71 claims are open at Jan 30, of which 16 close in February  $\Rightarrow P = 16 / 71 = 22.54\%$ .
- 55 claims are open at Feb 28, of which 17 close in March  $\Rightarrow P = 17 / 55 = 30.91\%$ .

We use a maximum likelihood estimator. For claims occurring December 20X1:

- The probability of closing in January 20X8 is  $P$ .
- The probability of closing in February 20X8 is  $(1-P)(P)$ .
- The probability of closing in March 20X8 is  $(1-P)^2(P)$ .
- ....
- The probability of being open at December 31, 20X8 is  $(1-P)^{12}$ .

The probability of observing the actual figures is

$$C \times P^{29} \times [(1-P)(P)]^{16} \times [(1-P)^2(P)]^{17} \times [(1-P)^{12}]^7 \times \dots = C \times P^{93} \times (1-P)^{322}$$

$C$  is the combinatorial constant: the number of ways to arrange 100 claims so that 29 are in January, 16 are in February, 17 are in March, and so forth.

*Illustration:* Of four claims, 2 close in January, 1 closes in February, and 1 closes in March.

We label the claims A, B, C, and D. The possible combinations are

- A and B close in January; C closes in February; and D closes in March.
- A and B close in January; D closes in February; and C closes in March.
- ...

The illustration has 12 possibilities: The claim closing in March has 4 possibilities, for each of which there are three possibilities for the claim closing in February. Computing this constant is more difficult with 100 claims and 12 months. But this constant doesn't affect the maximum likelihood estimation, so we can ignore it.

### *Homework assignment*

To maximize the likelihood, set the derivative of the likelihood (or the loglikelihood) with respect to  $P$  equal to zero. Solve for  $P$ .

[If you have difficulty, ask a question on the discussion forum.]