

INDIVIDUAL CLAIMS SIMULATION MACHINE: DESCRIPTION OF THE R PACKAGE

Andrea Gabrielli Mario V. Wüthrich

Version V1, February 26, 2018

Abstract

The R files `Functions.V1` and `Simulation.Machine.V1` serve at simulating individual claims cash flows of non-life insurance claims for a synthetic insurance portfolio.

Description

The R files `Functions.V1` and `Simulation.Machine.V1` serve at simulating individual claims cash flows of non-life insurance claims for a synthetic insurance portfolio. The construction of the simulation machine is based on a number of different neural networks that have been calibrated to real insurance data coming from accident insurance business. It considers for each individual claim the 6 features `LoB`, `cc`, `AY`, `AQ`, `age` and `inj_part` described below. Based on these 6 features, it simulates a reporting delay for each individual claim and cash flows (thereafter) of a maximal settlement delay of 11 years (including the reporting delay). These cash flows also consider potential recoveries (with a negative sign). Finally, an indicator process is simulated that shows whether the claim is open or closed at the end of each accounting year. This indicator process also considers possible re-openings. For a detailed description of the statistical model used we refer to [1]. Moreover, `Functions.V1` also includes a function that allows for simulating a synthetic claims portfolio. The file `Functions.V1` contains all the necessary R functions and the file `Simulation.Machine.V1` demonstrates their use.

```
Simulation.Machine(features, npb, seed1, std1, std2)
```

Description

This is the main function that generates the individual claims cash flows based on the given portfolio of claims features `features`.

Arguments

- features** This is a matrix that contains the claims portfolio. It consists of one row per individual claim and of $d = 7$ columns having the following structure: (C1Nr, LoB, cc, AY, AQ, age, inj_part). The variable C1Nr is the claim number, this is a unique numerical identifier for each individual claim. LoB is the line of business which should be of factor type (categorical) and may take the labels $1, \dots, 4$. cc is the claims code that should be of factor type (categorical) and may take the labels $1, \dots, 53$ (with possible gaps). AY is the year of claims occurrence (accident year) that should be of integer type and may take the values $1994, \dots, 2005$. AQ is the quarter of claims occurrence (accident quarter) that should be of integer type and may take the values $1, \dots, 4$. age is the age of the injured that should be of integer type and may take the values $15, \dots, 70$. inj_part is the body part injured that should be of factor type (categorical) and may take the labels $1, \dots, 99$ (with gaps).
- npb** For parallel computing the maximal size of the blocks is specified.
- seed1** The seed for the simulation is initialized.
- std1** Standard deviation parameter in the log-normal distribution for the total claim amounts simulation.
- std2** Standard deviation parameter in the log-normal distribution for the total recovery payments simulation.

Output

- features** The output contains the input features (C1Nr, LoB, cc, AY, AQ, age, inj_part).
- RepDel** Reporting delay of the claims (difference between claims occurrence and claims reporting) being in $0, \dots, 11$.
- Pay00, ...** Cash flows paid in development years (delays) $j = 0, \dots, 11$.
- Open00, ...** Indicator showing whether the claim was open (=1) or closed (=0) at the end of development year $j = 0, \dots, 11$.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	
1	C1Nr	LoB	cc	AY	AQ	age	inj_part	RepDel	Pay00	Pay01	Pay02	Pay03	Pay04	Pay05	Pay06	Pay07	Pay08	Pay09	Pay10	Pay11	Open00	Open01	Open02	Open03	Open04	Open05	Open06	Open07	Open08	Open09	Open10	Open11	
2	1	3	17	1994	2	23	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	4	19	1994	3	32	21	0	113	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	4	39	1994	2	47	10	0	180	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	4	3	27	1994	4	35	51	0	0	2582	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	5	2	29	1994	1	15	61	0	0	1656	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	6	2	9	1994	2	63	53	0	3094	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	7	4	42	1994	1	41	21	0	631	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	8	2	43	1994	3	25	52	0	0	2059	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	9	2	34	1994	1	45	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	10	2	1	1994	4	37	34	0	0	1039	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	11	4	6	1994	1	44	51	0	696	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	12	4	7	1994	4	44	30	0	0	307	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	13	4	27	1994	1	48	15	0	208	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	14	4	26	1994	1	37	36	0	192	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	15	1	21	1994	3	52	21	0	332	180	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	16	3	27	1994	4	65	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	17	1	17	1994	3	16	36	0	170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	18	1	15	1994	1	24	36	0	436	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	19	4	49	1994	3	42	30	0	470	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	20	4	42	1994	4	28	30	0	0	330	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Feature.Generation(V, LoB.dist, inflation, seed1)

Description

This function generates a synthetic portfolio of claims features **features**.

Arguments

V Totally expected number of claims (used for Poissonian claims counts).
LoB.dist Probabilities specifying how the claims counts are allocated to the 4 lines of business.
inflation Growth parameter for each line of business acting on the claims counts.
seed1 The seed for the simulation is initialized.

Output

The output contains a portfolio with features (C1Nr, LoB, cc, AY, AQ, age, inj_part).

	A	B	C	D	E	F	G
1	C1Nr	LoB	cc	AY	AQ	age	inj_part
2	1	3	17	1994	2	23	11
3	2	4	19	1994	3	32	21
4	3	4	39	1994	2	47	10
5	4	3	27	1994	4	35	51
6	5	2	29	1994	1	15	61
7	6	2	9	1994	2	63	53
8	7	4	42	1994	1	41	21
9	8	2	43	1994	3	35	52
10	9	2	34	1994	1	45	10
11	10	2	1	1994	4	37	34
12	11	4	6	1994	1	44	51

Acknowledgment

Our greatest thanks go to Suva, Dr. Peter Blum and Dr. Olivier Steiger for providing data and for their immense support.

References

- [1] Gabrielli, A., Wüthrich, M.V. (2018). Individual claims history simulation machine. SSRN Manuscript ID 3130560.