Supplements for Loss Data Analytics

An open text authored by the Actuarial Community

2018-11-15

Contents

Chapter 1

Purpose

Chapter 2

Glossary

Chapter Preview.

2.1 Making Changes to Terms and Definitions

- First, open up the issues tab on our repository on GitHub here.
- Click on "create an issue".
- Indicate which chapters you want to make changes to in the title.
- Specify the terms and definitions you wish to change, add or remove.
- Click "Submit new issue".

2.2 Terms and Descriptions by Chapter

Detailed purpose.

2.2.1 Chapter 1 Introduction to Loss Data Analytics

Term	Description
analytics	The process of using data to make decisions. This process involves gathering data, understanding models of uncertainty,making general inferences, and communicating results
business intelligence	May focus on processes of collecting data, often through databases and data warehouses
business analytics	Utilizes tools and methods for statistical analyses of data
data science	Can encompass broader applications in many scientific domains
short-term	Contracts where the insurance coverage is typically provided for six months or a year
property insurance	In the US, policies such as renters and homeowners

Term	Description
casualty insurance	In the US, a policy such as auto that covers medical damages to people
nonlife or general insurance	In the rest of the world, property and casualty insurance are both known as nonlife or general insurance, to distinguish them from life insurance
underwriting	The process of classifying risks into homogenous categoriesand assigning policyholders to these categories, lies at the core of ratemaking. Policyholders within a class have similar risk profiles and so are charged thesame insurance price
ratemaking	Where analysts seek to determine the right price for the right risk
experience rating or merit rating	Modifying premiums with claims history
claims adjustment	The process of determining coverage, legal liability, and settling claims
claims leakage	Dollars lost through claims management inefficiencies
loss reserving	Setting aside money for unpaid claims
claim	At a fundamental level, insurance companies accept premiums in exchange for promises to indemnify a policyholder upon the uncertain occurrence of an insured event. This indemnification is known as aÂ claim
severity	A positive amount is a key financial expenditure for an insurer. So, knowing only the claim amount summarizes the reimbursement to the policyholder
frequency	How often claims arise
pure premium or loss cost	The total severity divided by the number of claims
rating variables	Externally available variables useful in setting insurance rates and premiums

2.2.2 Chapter 2 Frequency Modeling

Term	Description
claim	compensation from insurer to insured upon the occurrence of an insured event
frequency	how often claims arise or how often insured event occurs
severity	amount of each payment for an insured event
expected cost	expected number of claims (frequency) times expected amount per claim (severity)
binomial distribution	discrete frequency distribution and member of (a, b, 0) class; for number of successes in a fixed number of independent identical trials with binary outcomes

Term	Description
negative binomial distribution	discrete frequency distribution and member of (a, b, 0) class; for number of successes until we observe the r-th failure in independent identical trials with binary outcomes
poisson distribution	discrete frequency distribution and member of (a, b, 0) class; for independent events occuring at a constant rate in a certain time period
likelihood	observed value of mass function
maximum likelihood estimator (mle)	to find parameter values that produce the largest likelihood
risk	a unit covered by insurance
parameter	a numerical characteristic of a population
mixture	mixture of subgroups, each with their own distribution
fitted distribution	distribution used for modeling the data
Pearson chi-square statistic	to check for the goodness of fit of the fitted distribution

2.3 Terms and Chapter First Defined

Detailed purpose.

Term	Chapter first defined
analytics	1
binomial distribution	2
business analytics	1
business intelligence	1
casualty insurance	1
claim	1
claims adjustment	1
claims leakage	1
data science	1
expected cost	2
experience rating or merit rating	1
fitted distribution	NA
frequency	1
likelihood	NA
loss reserving	1
maximum likelihood estimator (mle)	2
mixture	NA
negative binomial distribution	NA
nonlife or general insurance	1
parameter	NA
Pearson chi-square statistic	NA
poisson distribution	2
property insurance	1
pure premium or loss cost	1
ratemaking	1
rating variables	1
risk	NA

Term	Chapter first defined
severity	1
short-term	1
underwriting	1

Chapter 3

Table of Distributions

Detailed Purpose.

Name	Probability Density Function f(x)	Mean $\mu = \mathrm{E}\ X$	Variance σ^2 $\mathrm{E}\ (X-\mu)^2$	Moments $\mu'_k = \mathrm{E}\ X^k$ or $\mu_k = \mathrm{E}\ (X-\mu)^k$	$\mathrm{E}\ (X \wedge x)^k$	Momen Function
Uniform	$\frac{1}{\beta-\alpha}$ $-\infty < \alpha, < \beta < \infty$	$\frac{\beta+\alpha}{2}$	$\frac{(\beta-\alpha)^2}{12}$	$\mu_k = 0$ for k odd $\mu_k = \frac{(\beta-\alpha)^k}{2^k(k+1)}$ for k even		
Normal	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $-\infty < \mu < \infty, \sigma > 0$	μ	σ^2	$\mu_k = 0$ for k odd $\mu_k = \frac{k!\sigma^k}{(k/2)!2^{k/2}}$ for k even		exp(μ
Exponential	$\frac{1}{\theta}e^{-x/\theta}$ $\lambda > 0$	θ	θ^2	$\mu'_k = \theta^k\Gamma(k+1)$	$\theta^k\Gamma(k+1)\Gamma(k+1; x/\theta)$ $+x^k e^{-x/\theta}$	
Gamma	$\frac{1}{\theta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\theta}$ $\theta > 0, \alpha > 0$	$\alpha\ \theta$	$\alpha\ \theta^2$	$\mu'_k = \frac{\theta^k\Gamma(k+\alpha)}{\Gamma(\alpha)}$	$\frac{\theta^k\Gamma(k+\alpha)}{\Gamma(\alpha)}\Gamma(k+\alpha; x/\theta)$ $+x^k[1-\Gamma(\alpha; x/\theta)]$	
Beta	$\frac{1}{B(a,b)}u^a(1-u)^{b-1}\frac{\theta}{x}$, $u = x/\theta, a > 0, b > 0$	$\frac{a\theta}{a+b}$	$\frac{ab\theta^2}{(a+b+1)(a+b)^2}$	$\mu'_k = \theta^k \frac{B(k+a,b)}{B(a,b)}$	Not useful	N
Cauchy	$\frac{1}{\pi\beta}[1 + \left(\frac{x-\alpha}{\beta}\right)^2]^{-1}$ $-\infty < \alpha < \infty, \beta > 0$	Does not exist	Does not exist	Does not exist	Does not exist	Doe
Lognormal	$\frac{1}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$ $-\infty < \mu < \infty, \sigma > 0$	$\exp(\mu + \sigma^2/2)$	$\exp(2\mu+2\sigma^2) - \exp(2\mu+\sigma^2)$	$\mu'_k = \exp(k\mu + k\sigma^2)$	$\exp(k\mu + k\sigma^2)$ $\Phi\left(\frac{lnx-\mu-k\sigma^2}{\sigma}\right)$ $+x^k(1-F(x))$	N
Pareto	$\frac{\alpha\theta^\alpha}{x^{\alpha+1}}, \alpha > 0$	$\frac{\alpha\theta}{\alpha-1}$	$\frac{\alpha\theta^2}{\alpha-2} - \left(\frac{\alpha\theta}{\alpha-1}\right)^2$	$\mu'_k = \frac{\alpha\theta^k}{\alpha-k}$	$\frac{\alpha\theta^k}{\alpha-k} - \frac{k\theta^\alpha}{(\alpha-k)x^{\alpha-k}}$	Doe

Chapter 4

Conventions for Notation

Chapter Preview. **Loss Data Analytics** will serve as a bridge between actuarial problems and methods and widely accepted statistical concepts and tools. Thus, the notation should be consistent with standard usage employed in probability and mathematical statistics. See, for example, (Halperin et al., 1965) for a description of one standard.

4.1 General Conventions

- Random variables are denoted by upper-case italicized Roman letters, with X or Y denoting a claim size variable, N a claim count variable, and S an aggregate loss variable. Realizations of random variables are denoted by corresponding lower-case italicized Roman letters, with x or y for claim sizes, n for a claim count, and s for an aggregate loss.
- Probability events are denoted by upper-case Roman letters, such as Pr(A) for the probability that an outcome in the event ''A" occurs.
- Cumulative probability functions are denoted by $F(z)$ and probability density functions by the associated lower-case Roman letter: $f(z)$.
- For distributions, parameters are denoted by lower-case Greek letters. A caret or ''hat" indicates a sample estimate of the corresponding population parameter. For example, $\hat{\beta}$ is an estimate of β .
- The arithmetic mean of a set of numbers, say, $x_1, \ldots, x_n$, is usually denoted by $\bar{x}$; the use of x, of course, is optional.
- Use upper-case boldface Roman letters to denote a matrix other than a vector. Use lower-case boldface Roman letters to denote a (column) vector. Use a superscript prime ''$\prime$" for transpose. For example, $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a quadratic form.
- Acronyms are to be used sparingly, given the international focus of our audience. Introduce acronyms commonly used in statistical nomenclature but limit the number of acronyms introduced. For example, *pdf* for probability density function is useful but *GS* for Gini statistic is not.

4.2 Abbreviations

Here is a list of abbreviations that we adopt. We italicize these acronyms. For example, we can discuss the goodness of fit in terms of the *AIC* criterion.

AIC	Akaike information criterion
BIC	(Schwarz) Bayesian information criterion
cdf	cumulative distribution function
df	degrees of freedom
iid	independent and identically distributed
glm	generalized linear model
mle	maximum likelihood estimate
ols	ordinary least squares
pdf	probability density function
pf	probability function
pmf	probability mass function
rv	random variable

4.3 Common Statistical Symbols and Operators

Here is a list of commonly used statistical symbols and operators, including the latex code that we use to generate them (in the parens).

$I(\cdot)$	binary indicator operator (I). For example, $I(A)$ is one if an outcome in event A occurs and is 0 otherwise.
$\Pr(\cdot)$	probability (`\Pr`)
$\mathrm{E}(\cdot)$	expectation operator (`\mathrm{E}`). For example, $\mathrm{E}(X) = \mathrm{E}\ X$ is the expected value of the random variable X, commonly denoted by μ.
$\mathrm{Var}(\cdot)$	variance operator (`\mathrm{Var}`).For example, $\mathrm{Var}(X) = \mathrm{Var}\ X$ is the variance of the random variable X, commonly denoted by σ^2.
$\mu_k = \mathrm{E}\ X^k$	kth moment of the random variable X. For k=1, use $\mu = \mu_1$.
$\mathrm{Cov}(\cdot, \cdot)$	covariance operator (`\mathrm{Cov}`). For example, $\mathrm{Cov}(X, Y) = \mathrm{E}\left\{(X - \mathrm{E}\ X)(Y - \mathrm{E}\ Y)\right\} = \mathrm{E}(XY) - (\mathrm{E}\ X)(\mathrm{E}\ Y)$ is the covariance between random variables X and Y.
$\mathrm{E}(X\vert\cdot)$	conditional expectation operator. For example, $\mathrm{E}(X\vert Y = y)$ is the conditional expected value of a random variable X given that the random variable Y equals y.
$\Phi(\cdot)$	standard normal cumulative distribution function (`\Phi`)
$\phi(\cdot)$	standard normal probability density function (`\phi`)
$\sim$	means is distributed as (`\sim`). For example, $X \sim F$ means that the random variable x has distribution function F.
$se(\hat{\beta})$	standard error of the parameter estimate $\hat{\beta}$ (`\hat{\beta}`), usually an estimate of the standard deviation of $\hat{\beta}$, which is $\sqrt{Var(\hat{\beta})}$.
H_0	null hypothesis
H_a or H_1	alternative hypothesis

4.4 Common Mathematical Symbols and Functions

Here is a list of commonly used mathematical symbols and functions, including the latex code that we use to generate them (in the parens).

$\equiv$	identity, equivalence (`\equiv`)
$a := b$	defines a in terms of b
$\Longrightarrow$	implies (`\implies`)
$\Longleftrightarrow$	if and only if (`\iff`)
$\to, \longrightarrow$	converges to (`\to`, `\longrightarrow`)
$\mathbb{N}$	natural numbers $1, 2, \ldots$ (`\mathbb{N}`)
$\mathbb{R}$	real numbers (`\mathbb{R}`)
$\in$	belongs to (`\in`)
$\notin$	does not belong to (`\notin`)
$\subseteq$	is a subset of (`\subseteq`)
$\subset$	is a proper subset of (`\subset`)
$\cup$	union (`\cup`)
$\cap$	intersection (`\cap`)
$\emptyset$	empty set (`\emptyset`)
A^c	complement of A
$g * f$	convolution $(g * f)(x) = \int_{-\infty}^{\infty} g(y)f(x-y)dy$
$\exp$	exponential (`\exp`)
$\log$	natural logarithm (`\log`)
$\log_a$	logarithm to the base a
!	factorial
$\mathrm{sgn}(x)$	sign of x(`sgn`)
$\lfloor x \rfloor$	integer part of x, that is, largest integer $\leq x$ (`\lfloor`, `\rfloor`)
$\|x\|$	absolute value of scalar x
$\varGamma(x)$	gamma (generalized factorial) function (`\varGamma`), satisfying $\varGamma(x+1) = x\varGamma(x)(\varGamma)$
$B(x, y)$	beta function, $\varGamma(x)\varGamma(y)/\varGamma(x+y)$

4.5 Further Readings

To make connections to other literatures, see (Abadir and Magnus, 2002) http://www.janmagnus.nl/misc/notation.zip for a summary of notation from the econometrics perspective. This reference has a terrific feature that many latex symbols are defined in the article. Further, there is a long history of discussion and debate surrounding actuarial notation; see (Boehm et al., 1975) for one contribution.

Bibliography

Abadir, K. and Magnus, J. (2002). Notation in econometrics: a proposal for a standard. *The Econometrics Journal*, 5(1):76–90.

Boehm, C., Engelfriet, J., Helbig, M., Kool, A. I., Leepin, P., Neuburger, E., and Wilkie, A. (1975). Thoughts on the harmonization of some proposals for a new international actuarial notation. *Blätter der DGVFM*, 12(2):99–129.

Halperin, M., Hartley, H. O., and Hoel, P. G. (1965). Recommended standards for statistical symbols and notation: Copss committee on symbols and notation. *The American Statistician*, 19(3):12–14.