Statistical modelling in insurance and finance

Sichuan University

July 2016

LOUIS G. DORAY, Ph.D., ASA

UNIVERSITÉ DE MONTRÉAL

Prerequisite course: 1 course in statistics

Reference: Loss Models: from data to decisions

Course objectives: the student will learn the various steps in modelling for problems in P/C insurance
- hypothesize an appropriate model for a data set
- estimate the parameters of the model and variance
- test goodness-of-fit of model to data.

The model can then be applied to calculate premiums and measure impact of policy modifications (deductible, limit, coinsurance)

Preparation to exam C of SOA and CAS
Construction and evaluation of actuarial models.

Do problems (many)!
Plan of course

1. Introduction

2. Functions describing a random variable (r.v.)

3. Measures characterizing a r.v.

4. Classification of distributions

5. Creation of new distributions

6. Estimation of parameters

7. Quality of estimators
1 - Introduction

- Main difference between P/C insurance and life insurance.
  P/C: Property/Casualty ins. (fire, home, automobile...) also called general ins. in UK.
  - Life ins: amount (death benefit) is fixed at policy issue
  P/C ins: claim amount is random.

- Life ins: time until death is random (long-term discount)
  P/C ins: also random but short-term, renewable contracts
  No discounting - Premiums adjusted based on experience of policyholder (e.g., automobile ins).

- P/C ins: amount paid by ins. may be smaller than losses incurred by policyholder.

- Definitions

  - Accident: event leading to a loss by th (policyholder). The th incurs damages potentially covered by his policy.

  - Loss: amount of damages incurred by th following the accident.

  - Claim: amount paid to th following following the accident (may be smaller than loss).
• Why amount paid may be smaller than loss?
  - Loss adjustment expenses
    (amount incurred to determine amount paid, legal fees)
  - Policy modifications
    • Deductible: if loss smaller than deductible, no amount paid; otherwise, amount paid equals loss minus deductible.
    • Limit: if loss exceeds limit, amount paid = limit.
    • Coinsurance: percentage of loss (after deductible and limit) paid by insurer.

Ex: Group dental policy
Currently policy has a deductible of 50 for claim. Investigate:
a) Elimination of deductible to encourage more frequent visits to dentist by users
b) Raising deductible to 100 to reduce premiums.

10 claims chosen at random:
141 16 46 40 351 259 317 1511 107 567
• Parametric models

  Advantages:
  - we can answer questions like elimination of deductables; imposing policy limits
  - calculation of confidence intervals
  - simplicity (1 distribution + 2 parameters).
  - smoothness

  Estimation of parameters
  - joint estimation
  - by interval

  Hypothesis testing.

• Hypothesis

  - loss and amount to be paid are known as accident occurs.
  - in practice, there may be a long delay between time of accident and time of final payment by insurer. The claim could also be paid in many small instalments.
Random variable
- Loss in automobile insurance
- Claim in automobile ins.
- Number of claims in a year by a policyholder (k)
- Total number of claims by all jk in company
- Total claim amounts by all jk in auto. ins.

2- Functions describing a r.v. X

a) Cumulative distribution function (cdf)
\[ F_X(x) = \Pr [X \leq x] \]

Properties:
- Non-decreasing
- Continuous to the right
- \( \lim_{x \to -\infty} F_X(x) = 0 \) \( \lim_{x \to \infty} F_X(x) = 1 \)

b) Probability density function (pdf)
\[ f_X(x) = \frac{df}{dx} F_X(x) \text{ for continuous r.v.} \]

Properties:
- Positive
- \( \Pr [a < X \leq b] = \int_a^b f_X(x) \, dx \)
  (if X discrete, \( F_X(x) = \sum_{y \leq x} \Pr [X = y] \))

c) Survival function
\[ \delta(x) = 1 - F_X(x) \]

d) Hazard rate
\[ h_X(x) = \frac{f_X(x)}{\delta(x)} = -\frac{d}{dx} \ln \delta(x) \]

Properties:
- \( h_X(x) \geq 0 \) \( \forall x \)
Fig. 2.1 Empirical distribution function of individual dental loss amounts

Fig. 2.2 Ogive of grouped dental loss amounts

Fig. 2.3 Histogram of grouped dental loss amounts

- **Mode**: value maximizing the pdf \( f_X(x) \) or \( P_X[X=x] \).
  - Mode at 0 or positive

- **Median**: measure of central tendency (symmetric dist.?)
  - Value \( m \) satisfying \( P_X[X \leq m] = \frac{1}{2} \) (if \( X \) continuous, \( m \) unique)

- **Mean**: measure of central tendency (moment matching)
  - \( E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) \, dx \) if \( X \) continuous r.v.
  - \( E(X) = \sum_{x} x \cdot P_X[X=x] \) if \( X \) discrete r.v.

  *What is \( E(X) \) if \( X \) is a mixed r.v.*

- **Variance**: \( \text{Var}(X) = E[X - \mu]^2 = E(X^2) - E^2(X) \).
  - Measure of variability
  - N.B. \( E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \, dx \) if \( X \) continuous
  - \( E[g(X)] = \sum_{x} g(x) \cdot P_X[X=x] \) if \( X \) discrete.

  *If integral or sum does not converge, \( E[g(X)] \) does not exist.

- **\( k^{th} \) moment** (\( k \in \mathbb{N} \)) \( E[X^k] \).

- **Coefficient of variation**: measure of standardized variability, \( \sqrt{\text{Var}(X)} / E(X) \).

- **Coefficient of asymmetry**: \( E[(X-E(X))^3] / (\text{Var}(X))^{3/2} \)

  *If equal to 0, symmetric dist.; if > 0, asymmetric to the right.*

  *Verify:* \( E[X \cdot E(X)^2] = E[X^3] - 3E(X^2)E(X) + 2E^3(X) \)

- **Kurtosis**: measure thickness of tails (compared to normal dist.)
  - \( E[(X-E(X))^4] / \text{Var}(X)^2 \).

- **Percentile**: \( f_X(x) > 0 \); for \( 0 < p < 1 \), unique \( x_p \) s.t. \( F_X(x_p) = p \).
• Moment generating function $M_x(t)$

$$M_x(t) = E(e^{tx})$$ for all $t$ for which $E(\cdot)$ exists.

$$E[X^k] = \frac{d^k}{dt^k} M_x(t)\bigg|_{t=0}, \quad k = 1, 2, \ldots$$

Since

$$M_x(t) = E\left[\sum_{m=0}^{\infty} \frac{t^m X^m}{m!}\right] = \sum_{m=0}^{\infty} \frac{t^m E(X)^m}{m!}$$

- Find $M_x(t)$ if $X \sim E(\lambda)$.

If $X_1, \ldots, X_n$ are independent r.v.s, s.t. $M_{X_i}(t)$ exists for all $i$, then for $Y = \sum_i X_i$

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

if $X_i$ are i.i.d., then $M_Y(t) = (M_{X_i}(t))^n$.

- Proof: ...

- mgf uniquely characterizes a r.v.

  Use this to find dist. of $\sum_i X_i$, $X_i \sim E(\lambda)$.

4- Classification of distributions:

• Complexity of model (number of parameters)

• Shape of distribution (asymmetry, tails, mode).

Complexity of models

Arguments for a simple model

- Few elements to specify
- Model more stable in time

Arguments for a complex model: better fit to data

Parsimony: the simplest model reflecting well the reality should be used.

C. Box: "All models are wrong, but certain are useful".
Class of parametric distributions
- set of distributions where each member is specified by 1 or more parameters.
- Number of parameters is fixed and finite.
- If the values of all parameters are specified, the dist. is completely known.

Scale family:
- let \( a \) be a positive constant.
- A family is closed under a scale transformation if \( Y = aX \) belongs to the same family of dist. as \( X \).
- ex: 1. \( X \sim N(\mu, \sigma^2) \Rightarrow Y \sim N(a\mu, a^2\sigma^2) \)
- 2. \( X \sim \text{Exp}(\theta) \Rightarrow Y \sim \text{Exp}(a\theta) \).

If \( X \) has pdf \( f_X(x) \), then \( f_Y(y) = f_X(y/a) \cdot \frac{1}{a} \).

A scale parameter is such that:
- the parameter is multiplied by \( a \).
- the other parameters are unchanged.

ex: Do the \( N + \text{Exp} \) have a scale parameter?
- \( X \sim \Gamma(\alpha, \beta) \) Scale parameter: \( \alpha \) depends
- The lognormal represents a scale family, w/o scale param.
Change of monetary unit: Can \( \# \) \( \Rightarrow \) US \( \# \)
\[ Y = 10.289 \times X \]
• Mixing of distributions

R.V. Y is a mixing of r.v. X₁, ..., Xₖ (k ∈ ℕ) if its cdf is

\[ F_Y(y) = a_1 F_{X_1}(y) + a_2 F_{X_2}(y) + \ldots + a_k F_{X_k}(y) \]

where \( 0 < a_i < 1 \) and \( \sum_{i=1}^{k} a_i = 1 \)

To model Y when there are two or more subpopulations behaving differently (ex: medical insurance: \( k = 20 \): M + F. car insurance: new drivers, experienced, old).
- it may be difficult to estimate parameters \( a_1, \ldots, a_k \)
- \( k \) could be a parameter itself (compare models with \( k = 20, k = 3 \))
- also called semi-parametric models.
- if all r.v. have the same dist. (e.g. Exp.),
  it is a mixing of Exponential distributions

Continuous mixing
- the limit of the mixing of dist. as \( k \to \infty \)
  is a continuous mixing of distributions.
- let \( \Theta \) \( \in \mathbb{R} \) a r.v. with density \( f_{\Theta} \)
- after realization of \( \Theta \), r.v. \( X \) has conditional density \( f_{\Theta}(x) \).
- the unconditional distribution of \( X \)

\[ f_X(x) = \int_{-\infty}^{\infty} f_{X|\Theta}(x|\theta) f_{\Theta}(\theta) d\theta. \]

Ex: If \( X \sim \text{Exp}() \) (\( \lambda = \text{mean} \)) and

\( X|\Theta \sim \text{Exp}(\lambda \Theta) \), then \( X \sim \text{Pareto}(1, \lambda) \).

Proof:
• Length of tails
  - Interesting classification systems, because tails contain information on extreme events.
  - Important for financial health of insurers.
  - Can order distributions according to length of tails: light, heavy, extremely heavy.

Measures of tail length:
  - Existence of moments
    - If $E[X^k]$ exists for all $k$, light tail (Normal).
    - If $E[X^k]$ exists $\forall k < N$, heavy tail (Student).
    - If $E[X^k]$ does not exist $\forall k$, extremely heavy (Cauchy).
  - Existence of mgf
    - If mgf does not exist, heavier tail.
  - Limit of ratio of pdf

\[
\lim_{n \to \infty} \frac{f(x)}{g(x)} = \begin{cases} 
\infty & \text{if } f \text{ has lighter tail} \\
0 & \text{if } g \text{ has lighter tail} \\
\text{const: similar behavior} & \text{if } g \text{ has heavier tail}
\end{cases}
\]

Proposition: If $g(x) > 0$ and $\lim_{n \to \infty} \frac{f(x)}{g(x)} = \infty \in [0, \infty)$,

\[
\lim_{n \to \infty} \frac{S_n(x)}{S_n(x)} = 1
\]

Ex: 1 $\Gamma(2, \theta)$ lighter right tail than $Exp(\theta)$.

Ratio of pdf.

2. With ratio of survival functions, compare right tail of $\text{Bern}(\alpha=1, \theta)$ with Pareto $(\alpha=1, \theta)$.

\[
P_B(n) = \frac{\theta^n}{\theta^n + n} \quad P_P(n) = \frac{\theta}{\theta + n}.
\]
- hazard rate: increasing vs decreasing
  non-shortfall, shortfall

5. Creation of new distributions
a) Multiplication by a positive constant
   \[ Y = ax \quad , \quad a > 0 \] Change of scale.
   \[ F_Y(y) = F_X(y/a) \]

b) Raising to a power
   \[ Y = x^\tau \quad , \quad \tau \in \mathbb{R} \quad , \quad x > 0. \]
   \[ F_Y(y) = P_X(x \leq y) = \begin{cases} 1 & \text{if } \tau = 0 \\ P_X \left( \frac{y}{x} \right) & \text{if } \tau > 0 \\ 1 - P_X \left( \frac{y}{x} \right) & \text{if } \tau < 0. \end{cases} \]
   \[ \tau > 0 : \text{transformed dist.} \]
   \[ \tau = -1 : \text{inverse dist.} \]
   \[ \tau < 0 : \text{transformed inverse dist.} \]

Ex. \( X \sim \text{Exp}(1) \)
Distribution of \( Y = X^\tau \).

c) Exponentiation \( Y = e^X \)
   cdf of \( Y \)?
   1. \( X \sim \text{Normal} \Rightarrow Y \sim \text{LN} \)
   2. \( X \sim \text{Exp}(1) \Rightarrow Y \sim \text{Pareto} (\alpha = 1, \theta = 1) \).

d) Slicing: join 2 or more jdf.
   \[ f_X(x) = \begin{cases} a_i f_i(x) & \text{if } c_i \leq x < c_{i+1} \\ a_k f_k(x) & \text{if } c_{k-1} < x < c_k, \end{cases} \]
   where \( f_i(x) \) is a jdf, \( \sum_{i=1}^{k} a_i = 1 \)
"Transformed Beta" Family of Distributions

Two parameters
- Lognormal

Three parameters
- Inverse transformed gamma
- Transformed gamma

Four parameters
- Transformed beta

Inverse gamma
- Inverse Weibull
- Inverse Burr
- Inverse Pareto
- Loglogistic

Gamma
- Weibull
- Burr
- Pareto

Mean and higher moments always exist

Mode > 0

Mean and higher moments never exist

Mode = 0

Special case
Limiting case (parameters approach zero or infinity)

Figure 5.4 Distributional relationships and characteristics.

N.B. 1- f might not necessarily be continuous.
2- k, c₀, ..., cₖ are usually known.
3- Interpretation similar to mixing

Ex: Positive return vs negative return of a stock.

\[ f_X(x) = \begin{cases} 
0.01 & \text{if } 0 < x \leq 10 \\
0.05 & \text{if } 10 < x \\
2x & \text{if } x > 10.
\end{cases} \]

Join \( U[0,10] \) with \( g_X(x) = e^{-(x-10)}, x > 10 \)

(multiply by \( \frac{10}{11} \))

6- Estimation of parameters

A- Method of moments

Equate the first \( k \) empirical moments to those of the theoretical distribution \( X_1, ..., X_n \) \( i.i.d. \) from parameter \( \theta \in \mathbb{R}^d \).

Solve system of equations

\[ E(X_i^k) = \frac{1}{m} \sum_{i=1}^{m} X_i^k, \quad k = 1, ..., d. \]

Difficulty with large values of \( k \) (extreme values!)
you could use negative or fractional moments.

Ex: 1- \( X_1, ..., X_n \) \( i.i.d. \) \( \text{Exp}(\theta) \) moment estimator?

... of moment estimator?

2- \( X_i \) \( i.i.d. \) \( \text{LN}(M, \sigma^2) \).

3- \( X_i \) \( i.i.d. \) \( \text{Gamma}(\alpha, \beta) \).
B- Percentile matching

\[ X_1, \ldots, X_n \text{ i.i.d. with same dist. } F. \]

\[ F \text{ depends on } \theta \in \Theta. \]

Objective: estimate parameters using a percentile of distribution \( F \) equal to percentiles of empirical dist. \( F_n(x) \)

Method: find \( j \) percentiles representing well the dist. \( F \).

例: for \( j = 2 \), use 25th and 75th percentiles.

\[ F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x). \]

The empirical cdf could be smoothed.

Smoothing: order sample: \( X_1, \ldots, X_n \)

Interpolation between 2 observations:

\[ \hat{X}_{ij} = (1-h)X_{(i)} + hX_{(i+1)} \]

Find the 25th percentile with 5 obs. \( (1.1, 1.75, 2.3, 3.7, 4.2) \)

例: \( X \sim \text{Pareto} (\alpha = 1, \theta). \)

\[ F_x(x) = 1 - \frac{x}{\theta} \]

Find \( \hat{\theta} \) with median matching.

C- Maximum likelihood est (MLE)

Advantages of A and B: easy to find estimates.

Problems with A and B:

- does not use all the information available
- arbitrary decision for choice of percentiles
- moments may not exist

MLE will correct these problems.

- give dist. of \( g(\theta). \)
\[ X_1, \ldots, X_n \text{i.i.d.} \quad \theta \in \mathbb{R}^d. \]

we observe \( \gamma_1, \ldots, \gamma_n \).

Likelihood function
\[ L(\theta) = \prod_{i=1}^n f_{X_i}(\gamma_i; \theta) \]

Find \( \hat{\theta} \) maximizing \( L(\theta) \)

Log likelihood function
\[ l(\theta) = \ln L(\theta) = \sum_{i=1}^n \ln f_{X_i}(\gamma_i; \theta). \]

Ex: \( X_1, \ldots, X_n \) i.i.d. \( \text{Exp}(\theta). \) \( \hat{\theta} \) MLE

7. Quality of estimators:
- Performance of estimators
- Can we compare them?
- Measures permitting this

i) Bias: on average, does the estimator give the good value?

Definition: The bias of an estimator is equal to \( E(\hat{\theta}) - \theta \)

Def.: An estimator is unbiased if its bias equals 0 for all values of \( \theta \).
Def.: An estimator is asymptotically unbiased if \( \lim_{m \to \infty} E(\hat{\theta}_m) = \theta \) for all \( \theta \).

ii) Consistency
An estimator is consistent if, for \( \forall \delta > 0 \) and \( \forall \theta \)
\[ \lim_{m \to \infty} P_n[|\hat{\theta}_m - \theta| > \delta] = 0. \]

If \( \hat{\theta}_m \) is asymptotically unbiased and its variance tends to 0, it is consistent.
Mean quadratic error. avg dist. between estimator of parameter
\[ \text{MAE} = E[(\hat{\theta}_m - \theta)^2] = \text{Var}(\hat{\theta}_m) + (E[\hat{\theta}_m] - \theta)^2. \]

Ex. \( X_1, \ldots, X_n \sim \text{i.i.d. } \text{U}[0, \theta]. \)
Estimate \( \hat{\theta}_m \) by MLE.
Study properties of \( \hat{\theta}_m \).

PROPERTIES of MLE

Under certain regularity conditions
1. The probability that \( \ell(\theta) = 0 \) has a solution tends to 1 as \( n \to \infty \).
2. \( \sqrt{n} (\hat{\theta}_m - \theta) \xrightarrow{d} N(0, I^{-1}(\theta)) \)  Min. variance estim.

where \( I(\theta) = -E(\frac{d^2}{d\theta^2} \ln f(x|\theta)) \) is the information m \( \theta \).

The expected information on \( \theta \) often estimated with the observed information calculated from the sample.

Method of statistical differentials: dist. of \( g(X) \).

\( X_\infty \): random vector of dimension \( g \).
\( g: \mathbb{R}^\theta \to \mathbb{R} \) s.t. \( g' \) exists at \( \theta \).

If \( \sqrt{n} (X_m - \theta) \xrightarrow{d} N(0, \Sigma) \), then
\[ \sqrt{n} \left( g(X_m) - g(\theta) \right) \xrightarrow{d} N(0, g'(\theta)\Sigma g(\theta)). \]

Particular case: \( \theta = 1 \).

Examples.
Central limit theorem

Let \( X_1, \ldots, X_n \) be iid r.v. with \( E(X_i) = \mu \) and \( \text{Var}(X_i) = \sigma^2 \), then

\[
\frac{n \sum X_i - n \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0,1).
\]