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
Klugman S.A., Panjer H.H. and Willmot, G.E. (1998). Wiley

Course objective: 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- Classifications of distributions

5- Creation of new distributions

6- Estimation of parameters

7- Quality of estimators

1- Introduction

- Main differences between P/C insurance + 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 – Prem adjusted based on experience of policyholder (ex: automobile ins)

  - P/c ins: amount paid by ins. may be smaller than loss incurred by policyholder.

- Definitions

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

  - loss: amount of damages incurred by ph following the accident.

  - claim: amount paid to ph 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 per claim.
Investigate a) elimination of deductible to encourage more frequent visits to dentist by ees
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 question like elimination of deductible; imposing policy limit
  - calculation of confidence intervals
  - simplicity (1 distribution + 2 parameters).
  - smoothness

  Estimation of parameters
  - point 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 (ph)
- Total number of claims by all ph of company
- Total claim amounts by all ph in aut. ins.

2- Functions describing a r.v. X

a) Cumulative distribution fct (cdf)

$$F_X(x) = Pr[X \le x]$$

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)

$f_X(x) = \frac{d}{dx} F_X(x)$ for continuous r.v.

Properties: - positive
- $Pr[a < X \le b] = \int_a^b f_X(x)\,dx$.

(if X discrete, $F_X(x) = \sum_{y \le x} Pr[X = y]$.

c) Survival function $s(x) = 1 - F_X(x)$

d) Hazard rate: $h_X(x) = f_X(x)/s(x) = \frac{-d}{dx} \ln s(x)$

Properties: $h_X(x) \ge 0 \quad \forall x$.

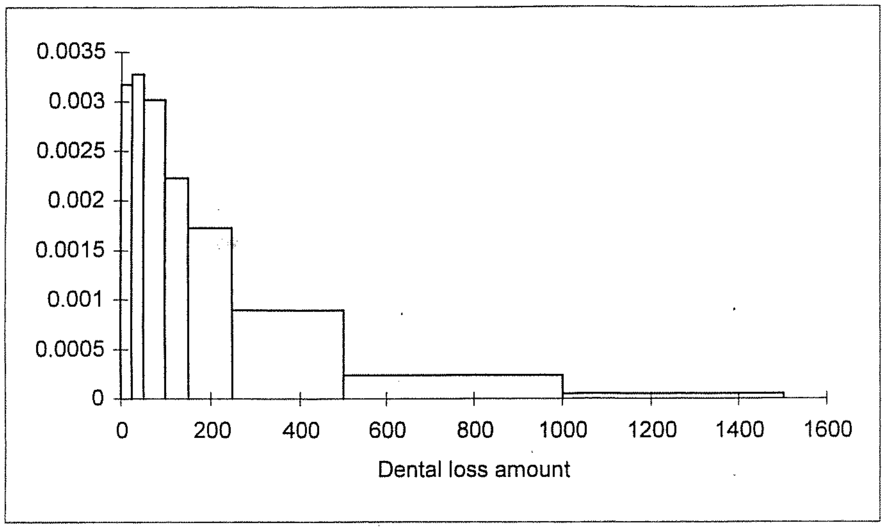


Fig. 2.3 Histogram of grouped dental loss amounts

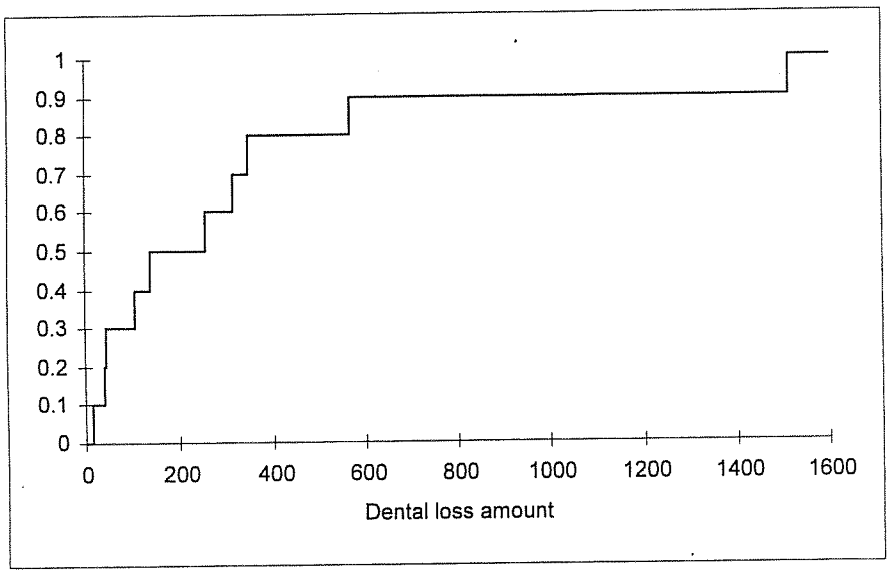
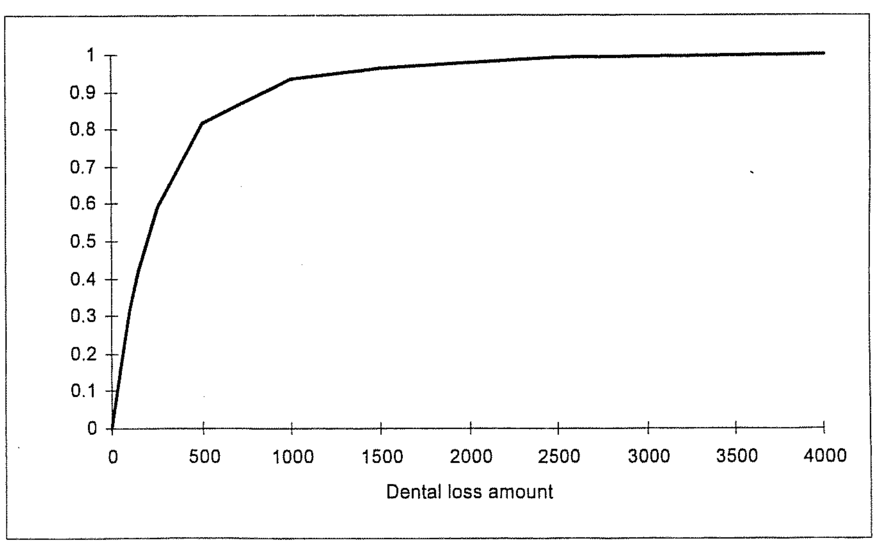


Fig. 2.1 Empirical distribution function of individual dental loss amounts



ref. Loss distributions

Fig. 2.2 Ogive of grouped dental loss amounts

3- Measures characterizing a r.v.

- Mode: value maximizing the pdf $f_X(x)$ or $Pr[X=x]$.
  ? Mode at 0 or positive

- Median: measure of central tendency (symmetric dist. ?)
  value $m \ni Pr[X \leq m] = 1/2$ (if X continuous, m unique)
- Mean: measure of central tendency (moment matching)
  $E(X) = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$ if X continuous r.v.

  $E(X) = \sum_x x \cdot Pr[X=x]$ if X discrete r.v.

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

- Variance $Var(X) = E[X-\mu]^2 = E(X^2) - E^2(X)$.
  measure of variability

N.B. $E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)$ if X continuous

$E[g(X)] = \sum_x g(x) \cdot Pr[X=x]$ 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: $\dfrac{\sqrt{Var(X)}}{E(X)}$

- Coefficient of asymmetry $\dfrac{E[X-E(X)]^3}{(Var(X))^{3/2}}$
  if equal to 0, symmetric dist; if >0, asymmetric to the right.
  verify: $E[X-E(X)]^3 = E[X^3] - 3E(X^2)E(X) + 2E^3(X)$

- kurtosis: measure thickness of tails (compared to normal dist.)
  $\dfrac{E[(X-E(X))^4]}{Var(X)^2}$.

- 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 $E(\ )$ exists.

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

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

- Find $M_X(t)$ if $X \sim Exp(\theta)$.

If $X_1, \ldots, X_n$ are indep. r.v. s.t. $M_{X_i}(t)$ exists for all i, then for $Y = \sum_{i=1}^{n} X_i$

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

if X_i are iid, then $M_Y(t) = \left(M_{X_i}(t)\right)^n$.

- Proof: ...
- mgf uniquely characterizes a r.v.
  Use this to find dist. of $\sum_{i=1}^{n} X_i$, $X_i \overset{iid}{\sim} E(\theta)$.

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.

G. 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- $X \sim N(\mu, \sigma^2)$ $\quad Y \sim N(a\mu, a^2\sigma^2)$
  2- $X \sim Exp(\theta)$ $\quad Y \sim Exp(a\theta)$.

  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 + Exp. have a scale parameter?
  - $X \sim \Gamma(\alpha, \beta)$ Scale parameter? It depends
  - The lognormal represents a scale family, w/o scale param.

  Change of monetary unit: Can \$ → US \$
  $Y = 1.289\, X$

- Mixing of distributions

R.V. Y is a mixing of r.v. $X_1, \dots, X_k$ ($k \in \mathbb{N}$) if its cdf is

$$F_Y(y) = a_1 F_{X_1}(y) + a_2 F_{X_2}(y) + \dots + a_k F_{X_k}(y)$$

where $0 < a_i < 1$ & $\sum_{i=1}^{k} a_i = 1$

To model Y when there are two or more subpopulations behaving differently
(ex: medical insurance: $k=2$: M + F.
car insurance: new drivers, experienced, old)
- it may be difficult to estimate parameters $a_1, \dots, a_k$
- k could be a parameter itself (compare models with $k=2$ or $k=3$)
- also called semi-parametric models.
- if all r.v. X_i 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_Θ
- after realization of θ, r.v. X has conditional density $f_{X|\Theta}$.
- the uncondition distribution of X

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

<u>Ex</u>: If $X \sim Exp(1/\alpha = \text{mean})$ and

$X|\Theta \sim Exp(1/\theta)$, then $X \sim \text{Pareto}(1, \alpha)$.

Proof:

- Length of tails
  interesting classification system, because
  - tails contain information on extreme events.
  - important for financial health of ins. cies.
  - 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 $\forall k \le N$, heavy tail (student)
  - if $E(X^k)$ does not exist $\forall k$, extremely heavy t. (Cauchy)

- existence of mgf
  if mgf ~~does~~ not exist, heavier tail

- limit of ratio of pdf

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \begin{cases} \infty : & f \text{ has lighter right tail} \\ 0 : & g \text{ has lighter R. tail} \\ cst : & \text{similar behavior} \end{cases}$$

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

$$\lim_{x\to\infty} \frac{S_F(x)}{S_G(x)} = 1$$

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

2. With ratio of survival functions,
compare right tail of Burr $(\alpha=1, \theta, \tau>0)$
with Pareto $(\alpha=1, \theta)$.

$$s_B(x) = \frac{\theta^\tau}{\theta^\tau + x^\tau} \qquad s_P(x) = \frac{\theta}{\theta+x}.$$

– hazard rate : increasing vs decreasing
non-short tail   short tail

5- Creation of new distributions

a) Multiplication by a positive constant

$Y = aX$, $a > 0$ Change of scale.

$F_Y(y) = F_X(y/a)$

b) Raising to a power

$$Y = X^{\tau}, \quad \tau \in \mathbb{R}, \quad X > 0.$$

$$F_Y(y) = P_r(X^{\tau} \le y) = \begin{cases} 1 & \text{if } \tau = 0. \\ P_r[X \le y^{1/\tau}] & \text{if } \tau > 0 \\ P_r[X \ge y^{1/\tau}] & \text{if } \tau < 0. \end{cases}$$

$\tau > 0$: transformed dist.

$\tau = -1$: inverse dist.

$\tau < 0$: transformed inverse dist.

Ex. $X \sim Exp(1)$

Distribution of $Y = X^{\theta}$.

c) Exponentiation $Y = e^X$

cdf of Y ?...

1- $X \sim$ Normal $\Rightarrow Y \sim LN$

2- $X \sim Exp(1) \xRightarrow{Y = e^X - 1} Y \sim$ Pareto $(\alpha = 1, \theta = 1)$.

d) Splicing : join 2 or more pdf.

$$f_X(x) = \begin{cases} a_1 f_1(x) & \text{if } c_0 < x \le c_1 \\ \vdots & \vdots \\ a_k f_k(x) & \text{if } c_{k-1} < x < c_k, \end{cases}$$

where $f_i(x)$ is a pdf , $\sum_{i=1}^{k} a_i = 1$

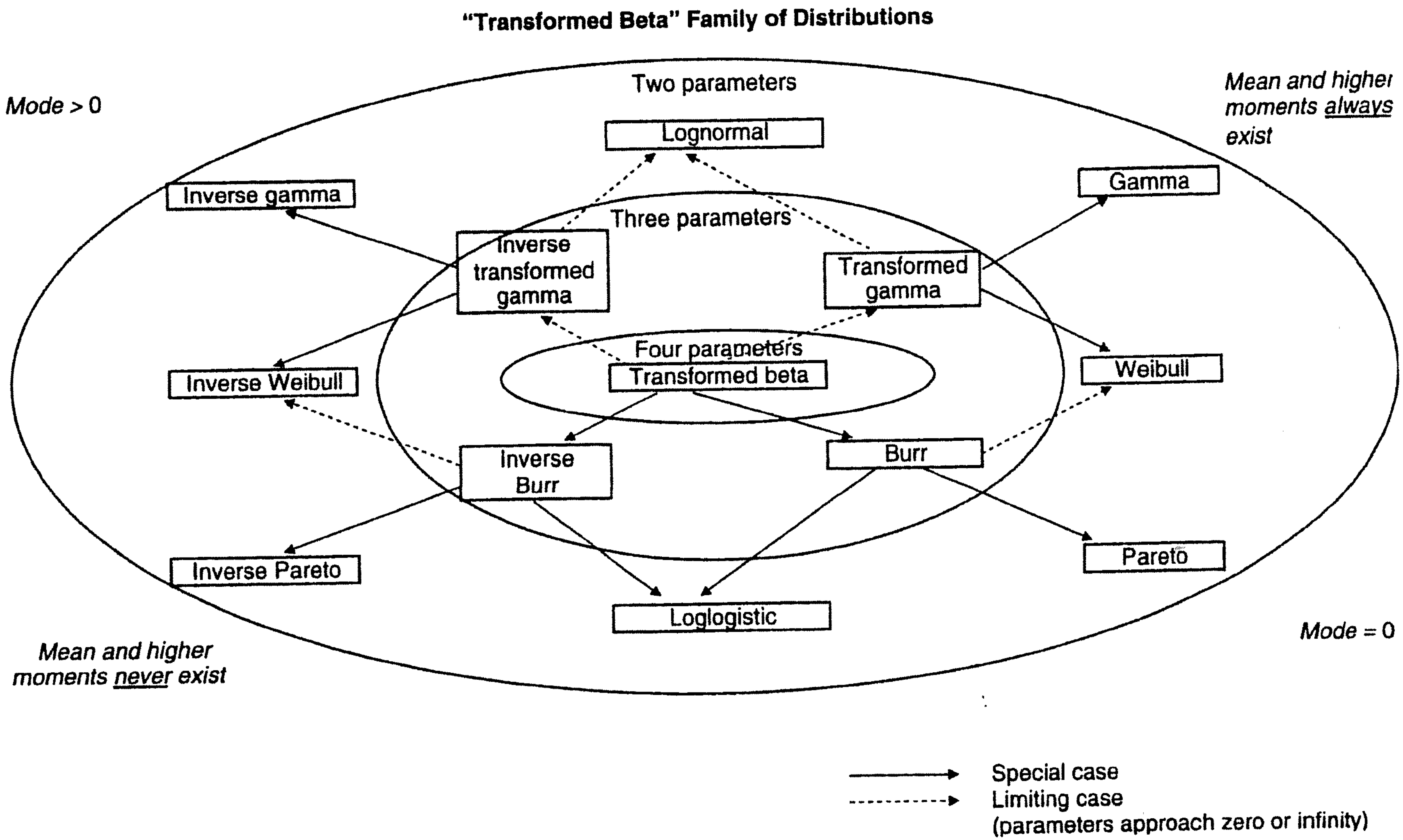


Figure 5.4 Distributional relationships and characteristics.

Ref: Klugman, Panjer, Willmot (2008), 3rd ed. Loss Models: From Data to Decisions

N.B. 1- $f_X(x)$ not necessarily continuous.

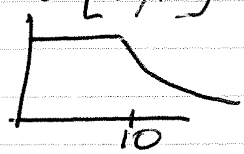
2- $k, c_0, \dots, c_k$ are usually known.

3- Interpretation similar to mixing

Ex: • positive return vs negative return of a stock.

• $f_X(x) = \begin{cases} 0.01\, x & \text{if } 0 < x \leq 10 \\ .05 & \text{if } 20 \geq x > 10. \end{cases}$

• 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 p empirical moments to those of theoretical distribution

$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} F$ with parameter $\theta \in \mathbb{R}^p$.

Solve system of equations

$$E(X_1^k) = \frac{1}{n} \sum_{i=1}^{n} x_i^k \quad , \; k = 1, \dots, p.$$

Difficulty with large value of p (extreme values!)
you could use negative or fractionary moments.

Ex: 1- $X_1, \dots, X_n \overset{iid}{\sim} Exp(\theta)$
... Moment estimator?

2- $X_i \overset{iid}{\sim} LN(\mu, \sigma^2)$.

3- $X_i \overset{iid}{\sim} Gamma(\alpha, \beta)$.

B- Percentile matching

$X_1, \ldots, X$ r.v. with same dist. F.

F depends on $\theta \in \mathbb{R}^p$.

Objective: estimate parameters using p percentiles of distribution F set equal to percentiles of empirical dist. $F_n(x)$

Method: Find p percentiles representing well the dist. (ex: for $p=2$, use 25^{th} and 75^{th} percentile).

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(x_i \le x)}.$$

The empirical cdf could be smoothed.

Smoothing: order sample: $x_{(1)}, \ldots x_{(n)}$

Interpolation between 2 observations.

$$\hat{\pi}_g = (1-h)\, x_{(j)} + h\, x_{(j+1)}$$

$$j = \lfloor (n+1)g \rfloor \quad \text{et} \quad h = (n+1)g - j$$

Find the 25^{th} percentile with obs. $(1.1, 1.75, 2.3, 3.7, 4.2)$

<u>Ex</u>: $X \sim \text{Pareto}(\alpha = 1, \theta)$. $\quad F_X(x) = 1 - \dfrac{\theta}{x+\theta}$

Find $\tilde{\theta}$ with median matching.

C- Maximum likelihood est (MLE)

advantage of A and B: easy to find estimators.

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 $g(\hat{\theta})$.

$X_1, \ldots, X_n$: iid r.v. $\theta \in \mathbb{R}^d$.
we observe $x_1, \ldots, x_n$.

Likelihood function

$$L(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i; \theta) \qquad \text{Find } \theta \text{ maximizing } L(\theta)$$

Log likelihood function

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f_{X_i}(x_i; \theta).$$

Ex: $X_1, \ldots, X_n$ i.i.d. $Exp(\theta)$. (Q ? d.8)

Find $L(\theta)$; $\ell(\theta)$; $\hat{\theta}$ 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 value of θ.

Def. : An estimator is asymptotically unbiased if $\lim_{n \to \infty} E(\hat{\theta}_n) = 0 \quad \forall \theta$.

ii) Consistency

An estimator is consistent if, for $\forall \delta > 0$, $\forall \theta$

$$\lim_{n \to \infty} P_r\left[|\hat{\theta}_n - \theta| > \delta \right] = 0.$$

If $\hat{\theta}_n$ asympt unbiased and its variance tends to 0, it is convergent.

iii) Mean quadratic error : avrg dist. between estimator & parameter

$$MQE = E\left[(\hat{\theta}_n - \theta)^2\right] = Var(\hat{\theta}_n) + \left(E[\hat{\theta}_n] - \theta\right)^2.$$

Ex $X_1, \dots, X_n \overset{i.i.d}{\sim} U[0, \theta]$.

Estimate $\hat{\theta}_n$ by ML. (Q? p. 15-16).
Study properties of $\hat{\theta}_n$.

PROPERTIES OF MLE

Under certain regularity conditions,

1- The probability that $l(\theta) = 0$ has a solution tends to 1 as $n \to \infty$.

2- $\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, I^{-1}(\theta))$ Min. variance estim

where $I(\theta) = -E\left(\frac{\partial^2}{\partial \theta^2} \ln f_{X}(x)\right)$ is the information on θ.

The expected information on θ often estimated with the observed information calculated from the sample.

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

X_n : random vector of dimension p.
$g : \mathbb{R}^p \to \mathbb{R}$ s.t. g' exists at θ.

If $\sqrt{n}\,(X_n - \theta) \xrightarrow{\mathcal{L}} N(0, \Sigma)$, then

$$\sqrt{n}\left(g(X_n) - g(\theta)\right) \xrightarrow{\mathcal{L}} N(0, g'(\theta) \Sigma g(\theta)).$$

Particular case : $p = 1$.

Examples.

Central limit theorem

let $X_1, \ldots, X_n$ be iid. r.v. with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, then

$$\frac{1/n \sum X_i - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{L}} N(0,1).$$