Sesion 22: Distribuciones para severidades individuales

Juan Carlos Martínez-Ovando

Primavera 2019

Motivación

- La distribucion del `monto individual de siniestros` juega un papel fundamental en el `calculo de riesgos agregados`.
- Usualmente, no solo un tipo de distribuciones es `util`, sino que se postulan varias alternativas, M, i.e. para el evento X se postulan

$$F_m(x|\theta_m) = \mathbb{P}(X \leq x|\theta_m, \text{modelo } m),$$

  para $m = 1, \ldots, M$.
- Los M tipos de modelos suelen diferir en su `forma estructural` y en el `valor especifico del parametro`.

Objetivo

En la sesion de hoy:

1. Estudiaremos las formas estructurales de varios `tipos de distribuciones` para el monto individual de siniestros.
2. Estudiaremos el problema de `comparacion` y `seleccion` de modelos.

Observacion

En la sesion de hoy trabajaremos con la parte continua de la distribucion de severidades individuales, i.e.

- La distribucion $F_X(x)$ del modelo de `riesgo colectivo`
- La distribucion $F^c_X(x)$ del modelo de `riesgo individual`.

En ambos casos, el **soporte comun** es

$$\mathcal{X} = (0, \infty)$$

.

Tipos de distribuciones

La `incertidumbre` sobre el `riesgo` de ocurrencia de severidades individuales tipicamente se **resume** en dos clases de modelos:

I. La **clase flexible no parametrica**, en la que no se hace ningun `supuesto estructural` sobre $F(x)$ mas alla de que sea una medida de probabilidad.

Tipicamente, si tenemos un conjunto de datos $\{x_1, \ldots, x_n\}$, estos son reexpresados como

$$\{(x_u^*, f_u)\}_{u=1}^{U},$$

donde

- U es el numero maximo de x_is distintas.
- x_u^*s son los `datos unicos ordenados` de los x_is.
- f_us son las fecuencias absolutas de los x_u^*s, i.e. $\sum_{u=1}^{U} f_u = n$.

Observacion

La **clase flexible no parametrica** es util en la practica, pero complica los `calculos matematicos actuariales`.

La distribucion resultante es `escalonada` y `no es absolutamente continua`.

I. La **clase parametrica**, en la que ademas de impoener la `restriccion` sobre $F(x)$ como medida de probabilidad, se imponen `restricciones estructurales adicionales`.

Tipicamente, la distirbucion $F(x|\theta)$ es `absolutamente continua`, por lo que se trabaja con su densidad, $f(x|\theta)$.

En terminos de la `densidad`, observamos que tipicamente es de la forma

$$f(x|\theta) \propto c(x,\theta) \exp\{-h(x,\theta)\} \mathbb{I}(x>0),$$

para algunas funciones $c(x,\theta)$ y $h(x,\theta)$.

Tipos de formas estructurales

Dependiendo de la forma de la funciones $c(x,\theta)$ y $h(x,\theta)$ pensaremos tipicamente en el `parametro` θ como uno de los siguientes tres tipos:

1. **Localizacion**, si $h(x,\theta) = g(x-\theta)$.
2. **Dispersion**, si $h(x,\theta) = g(x/\theta)$.
3. **Forma**, si $h(x,\theta) = g\left(x^{\theta}\right)$.

Para alguna funcion $g(\cdot)$.

Observaciones

- Cada forma estructural hace referencia a un `momento de la distribucion` de severidades individuales.
- La combinacion de las `formas estructurales` brindan `flexibilidad` en el modelo, per una flexibilidad controlada por suavizamiento, monotonicidad, etc.

a. Distribucion Pareto

La funcion de distribucion se define como

$$F_a(x|\alpha,\beta) = \left\{ 1 - \left(\frac{x}{\beta}\right)^{-\alpha} \right\} \mathbb{I}(x \geq \beta),$$

donde

- $\alpha > 0$, es el `parametro de forma`. El parametro de forma mide que tannta probabilidad se le asigna a valores grandes de X.
- $\beta > 0$ es el `parametro de escala`. El parametro de escala mide que tan dispersa es la asignacion de probabilidades para X.

La `funcion de densidad` de la distribucion Pareto tiene funcion de densidad.

$$f_a(x|\alpha, \beta) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}} \mathbb{I}(x \geq \beta).$$

- Como podran anticipar, la estimacoin del parametro α corresponde al aspecto mas sensible para llevar esta distribucion a la practica.

b. Distribucion Lognormal

La distribucion lognormal se origina como la distribucion de $X = \exp(Y)$ con $Y \sim \mathrm{N}(y|\mu, \sigma)$.

Asi, la funcion de distribucion de probabilidades es

$$F_b(x|\mu, \sigma^2) = \Phi\left(\frac{\log(x) - \mu}{\sigma^{1/2}}\right)\mathbb{I}(x > 0),$$

donde

- μ es el `parametro de localizacion`.
- σ es el `parametro de escala`.

La funcion de densidad de la distribucion lognormal es

$$f_b(x|\mu,\sigma) = \frac{1}{x\sqrt{2\pi\sigma}} \exp\left\{-\frac{(\log(x)-\mu)^2}{2\sigma}\right\} \mathbb{I}(x>0).$$

- La estimacion de parametros es relativamente sencilla, empleando la tranformacion inversa que da origen a esta distribucion sobre los datos.

c. Distribucion gamma

La distribucion gamma es definida con base en su `funcion de densidad`, dada por

$$f_c(x|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-x\beta\} \mathbb{I}(x>0),$$

donde

- $\alpha > 0$ es el `parametro de forma`
- β es el `parametro de escala`.

La distribucion gamma generalizada surge modificar la distribucion gamma estandar con una transformacion potencia en γ, i.e. de esta forma la `funcion de densidad modificada` es de la forma

$$f_c(x|\alpha,\beta,\gamma) = \frac{\gamma\beta^{\gamma\alpha}}{\Gamma(\alpha)} x^{\gamma\alpha-1} \exp\{-(x\beta)^{\gamma}\} \mathbb{I}(x>0),$$

donde

- α y $\gamma > 0$ son `parametros de forma`.
- $\beta^{1/\gamma}$ es el `parametro de escala`.

d. Distribuciones tipo beta

La distribucion de severidades individuales tipo beta esta relacionada con la funcion beta incompleta, como su nombre lo indica.

La `funcion de distribucion` de la distribucion beta generalizada del segundo tipo es de la forma,

$$F_d(x|\alpha, \beta, p, q) = I_z(p, q)\mathbb{I}(x > 0),$$

para $z = \left(\frac{x}{\beta}\right)^{\alpha}$, variable auxiliar, donde

$$I_z(p, q) = \frac{1}{B(p, q)} \int_0^z \frac{t^{p-1}}{(1 + t)^{p+q}} \mathrm{d}t.$$

La `funcion de densidad` es de la forma

$$f_d(x|\alpha, \beta, p, q) = \frac{\alpha x^{\alpha p-1}}{\beta^{\alpha p} B(p,q) \left(1 + (x/\beta)^{\alpha}\right)^{p+q}} \mathbb{I}(x > 0),$$

donde

- $\beta > 0$ es el `parametro de escala`
- α, p, q son `parametros de forma`.

e. Dsitribucion Pareto generalizada

La funcion de distribucion Pareto generalizada es de la forma,

$$f_e(x|\alpha,\mu,\sigma) = \left\{1 - \left(1 + \frac{\alpha(x-\mu)}{\sigma}\right)^{-1/\sigma}\right\} \mathbb{I}(x \geq \mu),$$

si $\alpha \geq 0$, donde

- μ es el `parametro de localizacion`
- σ es el `parametro de escala`
- α es el `parametro de forma`.

La `funcion de densidad` de este tipo de distribuciones es

$$f_e(x|\alpha,\mu,\sigma) = \frac{1}{\sigma}\left(1+\frac{\sigma(x-\mu)}{\sigma}\right)^{-\frac{1-\sigma}{\sigma}} \mathbb{I}(x \geq \mu),$$

si $\alpha \geq 0$.

Cuando $\alpha < 0$, el soporte de la distribucion se define como

$$\mathcal{X} = \left\{ x : \mu \leq x \leq \mu - \frac{\sigma}{\alpha} \right\}.$$

- Este sera un caso poco comun en el contexto de los problemas relacionados con nuestro curso.

Comentarios

- Los **cinco** tipos de distribuciones de probabilidad son estructuralmente distintos.
- Cada tipo es tipicamente representado por el `modelo especifico mas representativo` (a.k.a. mas plausible o mas verosimil), i.e.
$$F_m(x|\hat{\theta}_m),$$
donde $\hat{\theta}$ es el EMV o EB del parametro θ_m.
- El problema ahora trasciende a elegir el `modelo mas representativo` entre
$$F_1(x|\hat{\theta}_1), \ldots, F_M(x|\hat{\theta}_M).$$

Comparacion y seleccion de modelos

Los enfoques tradicionales para la `comparacion` y `seleccion` de modelos son de la forma:

A. **Bondad de ajuste**, en el que cada $F_m(x|\hat{\theta}_m)$ se compara con el **modelo flexible no parametrico** como referencia.

- Pruebas como `Kolmogorov`, `Anderson-Darling` o `Xi-cuadrada` sirven a este proposito.

B. **Cociente de verosimilitudes,** en el que cada par $F_m(x|\hat{\theta}_m)$ y $F_l(x|\hat{\theta}_l)$ de modelos contendientes se compara empleando el cociente de sus verosimilitudes respectivas.

Los procedimientos para la `comparacion` y `seleccion` de modelos anteriores se plantean como problemas de pruebas de hipotesis.

Cada procedimiento genera `puntajes` en terminos de sus `estadisticas de pruebas`, dando origen a las siguientes maximas:

- Elegir el modelo con el menor puntaje del estadistico Kolmogorov.
- Elegir el modelo con el menor puntaje del estadistico Anderson-Darling.
- Elegir el modelo con el menor puntaje del estadistico Xi-cuadrado.
- Elegir el modelo con la verosimilitud mas alta.

Observaciones

- Los procedimientos anteriores son del `tipo de ajuste` y pueden incurrir en el riesgo de elegir el modelo que mejor ajusta a un conjunto de datos, pero que no necesariamente predice mejor.
- En los laboratorios exploraremos una forma complementaria de seleccionar modelos basada en el **principio de devianza**, particularmente el **principio de devianza predictiva**.

Lecturas complementarias:

- Kleiber & Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, Caps. 3-6.
- Panjer, *Operational Risk Modeling Analytics*, Cap. 12.