Distribuzione (e principio di) Pareto.[1]

1 Densità di Pareto

Seguiremo Pareto nell'interpretare la variabile aleatoria X come *reddito* (di un individuo in una data popolazione). La sua osservazione di base è semplice: i ricchi sono pochi ma posseggono la gran parte del reddito totale (della popolazione). In realtà Pareto aveva un'idea precisa: l'80% del reddito totale è in mano al 20% della popolazione. Vedremo qui da dove viene questo cosiddetto "Principio di Pareto".

Sia $x_m > 0$ il reddito minimo. L'assunzione di base è questa: per qualche $\alpha > 1$

$$P(X > x) = \begin{cases} 1 & \text{se } x < x_m \\ (x/x_m)^{-\alpha} & \text{se } x \geq x_m. \end{cases}$$

Per interpretarla facilmente prendi $x_m = 1$: l'idea è che la probabilità di avere un reddito più alto di x scende in misura più che proporzionale ad x (più precisamente secondo una "power law"). Da qui con $F(x) = 1 - P(X > x)$ derivando si ottiene subito la densità di Pareto

$$f(x) = \begin{cases} 0 & x < x_m \\ \dfrac{\alpha x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \end{cases} \equiv f(x; \alpha, x_m)$$

2 Curva di Lorenz e principio di Pareto

Per arrivare al principio 80-20 dobbiamo definire la curva di Lorenz, per capire la quale conviene partire dal discreto. Considera una popolazione di N individui, di cui N_h hanno reddito x_h, per $h = 1, \dots, H$ (quindi $\sum_h N_h = N$). Il reddito totale nella popolazione è $\sum_{h \in H} x_h N_h$, e il reddito pro-capite è

$$\frac{\sum_h x_h N_h}{N} = \sum\nolimits_h x_h \frac{N_h}{N}.$$

Possiamo interpretare la frazione N_h/N di individui con reddito x_h come probabilità di avere reddito x_h. Ponendo allora $N_h/N = P(X = x_h) \equiv p(x_h)$ il reddito pro-capite diventa il *valore atteso del reddito*

$$EX = \sum\nolimits_h x_h p(x_h).$$

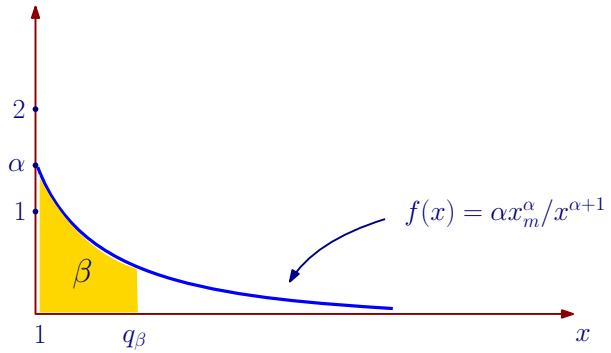
Dato $x > x_m$ considera adesso i redditi non superiori ad x. La frazione di individui con reddito $x_h \leq x$ è allora $\sum_{\{h:\, x_h \leq x\}} N_h/N = \sum_{\{h:\, x_h \leq x\}} p(x_h)$, e il reddito totale da essi pos-

[1] S Modica 2026

seduto è $\sum_{\{h:\, x_h \le x\}} x_h N_h$; quindi *la frazione del reddito totale* che posseggono è (uguaglianza dividendo sopra e sotto per N)

$$\frac{\sum_{\{h:\, x_h \le x\}} x_h N_h}{\sum_{h \in H} x_h N_h} = \frac{\sum_{\{h:\, x_h \le x\}} x_h p(x_h)}{EX}.$$

Ora saltiamo al continuo: la densità discreta p diventa la densità continua f e le somme diventano integrali. La frazione della popolazione con reddito non superiore ad x è $\int_{x_m}^{x} f(x)dx$, e la frazione di reddito da essi posseduta diventa $\int_{x_m}^{x} zf(z)dz/EX$. Se in particolare prendiamo $x = q_\beta$, il quantile di ordine β - definito come sappiamo da $\beta = P(X \le q_\beta) = \int_{x_m}^{q_\beta} f(x)dx$ - l'area sotto la f fino a q_β è il $100 \cdot \beta\%$ più povero della popolazione, vedi figura (dove $x_m = 1$),



e la frazione di reddito da essi posseduto è

$$L(\beta) \equiv \frac{\int_{x_m}^{q_\beta} zf(z)dz}{EX}.$$

Questa è la *Curva di Lorenz*, dove $0 \le \beta \le 1$. Con f Pareto, e assumendo per semplicità $x_m = 1$, otteniamo facilmente

$$L(\beta) = 1 - (1 - \beta)^{1 - 1/\alpha}.$$

Per $\alpha = \log_4 5 = \ln 5 / \ln 4 \approx 1.161$ si ottiene $L(0.8) = 0.2$: l'ottanta per cento della popolazione possiede il venti per cento del reddito; detto altrimenti, l'80% del reddito è in mano al 20% dei più ricchi. **100** esempi di 80-20: cerca "thelangelfirm 80/20" e li trovi.

3 Assicurazioni e franchigie

È prassi comune che le assicurazioni impongano una franchigia (valore del sinistro) al di sotto della quale il sinistro non viene rimborsato. Se la franchigia è d e l'entità del danno è X l'interesse è quindi su $E(X \mid X > d)$, il valore atteso dei sinistri da rimborsare. Ricordiamo

che la densità condizionata ad $\{X > d\}$, che indicheremo con $f_{X|\{X>d\}}$ è definita da

$$f_{X|\{X>d\}}(x) = \frac{f(x)}{P(X > d)} \cdot 1_{(d,\infty)}(x)$$

dove f è la densità di X.

Supponiamo che $EX = 2$ (per esempio in migliaia di Euro per fissare le idee) e che la compagnia imponga una franchigia di 3; dunque vogliamo $E(X \mid X > 3)$. Paragoneremo lo scenario in cui X è Pareto a quello in cui X è esponenziale.

Quindi cominciamo con l'assumere che $X \sim \text{Esp}(\lambda = 1/2)$. Intanto $P(X > 3) = e^{-3/2} \approx 22.3\%$, quindi la compagnia *non* pagherà circa l'80% dei sinistri. Abbiamo $f_{X|\{X>3\}}(x) = e^{3/2} \cdot \frac{1}{2}e^{-x/2} \cdot 1_{(3,\infty)} = \frac{1}{2}e^{-(x-3)/2} \cdot 1_{(3,\infty)}$ quindi (ponendo $x = u + 3$)

$$\begin{aligned} E(X \mid X > 3) &= \frac{1}{2}\int_3^{+\infty} xe^{-\frac{1}{2}(x-3)}dx = \frac{1}{2}\int_0^{+\infty}(u+3)e^{-\frac{1}{2}u}du \\ &= \frac{1}{2}\int_0^{+\infty}(u+3)e^{-\frac{1}{2}u}du = EX + 3 = 5. \end{aligned}$$

La compagnia pagherà in media $5 - 3 = 2$ mila euro per incidente rimborsabile. Nota che anche $EX = 2$ - ma con la franchigia la compagnia paga (in valore atteso) 2 solo sul 22% dei sinistri!

Assumiamo adesso che X sia Pareto con $x_m = 1$ (valore minimo dei sinistri rilevati) e $\alpha = 2$ (di nuovo $EX = 2$). In questo caso $P(X > 3) = \int_3^{+\infty} 2x^{-3}\,dx = \left[-1/x^2\right]_3^{+\infty} = 1/9 \approx 11.1\%$, ed $f_{X|\{X>3\}}(x) = 9 \cdot 2x^{-3} \cdot 1_{(3,\infty)}$ (un'altra Pareto con minimo $x_m = 3$ ed $\alpha = 2$). Perciò

$$E(X \mid X > 3) = \int_3^{+\infty} x \cdot f_{X|A}(x)\,dx = \int_3^{+\infty} x \cdot \frac{18}{x^3}\,dx = 6.$$

Rispetto alla esponenziale la media dei sinistri rimborsabili è più alta del 20% - a causa della "heavy tail" della Pareto - ma di sinistri rimborsabili ce n'è la metà. L'esborso totale della compagnia è in valore atteso più basso.

4 Curiosità: Se $X \sim \Gamma(a, \lambda)$ ed $Y \sim \text{Exp}(X)$ allora $Y \sim \text{Pareto}(\alpha, \lambda)$

L'ipotesi su Y è che la condizionale su X è esponenziale (di parametro X), quindi da $f(x, y) = f_X(x) \cdot f_{Y|X}(y|x)$ marginalizzando otteniamo, per $y \geq 0$,

$$\begin{aligned} f_Y(y) &= \int_0^{\infty} f_X(x) \cdot f_{Y|X}(y|x)dx = \int_0^{\infty} xe^{-xy} \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}dx = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^{\infty} x^\alpha e^{-(\lambda+y)x}dx \\ &= \frac{\lambda^\alpha\Gamma(\alpha+1)}{\Gamma(\alpha)(\lambda+y)^{\alpha+1}}\int_0^{\infty}\frac{(\lambda+y)^{\alpha+1}x^\alpha e^{-(\lambda+y)x}}{\Gamma(\alpha+1)}dx = \frac{\alpha\lambda^\alpha}{(\lambda+y)^{\alpha+1}}. \end{aligned}$$