GLOSSARIO SULLA PARTE FINALE DEL CORSO DI MATEMATICA, PARTE 2

DALLA PROBABILITA' DISCRETA ALLA PROBABILITA' CONTINUA

Popolazione statistica: Insieme di individui detti unita statistiche, di qualsiasi natura, su cui e' possibile rilevare una serie di caratteri. Possiamo distinguere tra popolazioni reali, ad esempio una popolazione di neonati in una fissata citta' durante un fissato intervallo di tempo, e popolazioni virtuali, come l'insieme delle n-ple di neonati (con $n > 1$) scelti tra tutti i neonati della popolazione (campione di dimensione n).

Variabile statistica: La funzione che associa ad ogni individuo di una popolazione statistica, la modalita' che assume un certo carattere oggetto di studio (supponendo le possibili modalita' incompatibili).

Distribuzione di frequenza (relativa) di una variabile statistica: La funzione che associa ad ogni modalita' di un carattere oggetto di studio in una popolazione statistica, la proporzione di individui che possiedono tale modalita'. Se n_i e' il numero di individudi che possiedono la modalita' i del carattere, la frequenza (relativa) della modalita i e' $\frac{n_i}{n}$, dove n e' il numero totale di individui. Tali proporzioni sono evidentemente numeri postivi la cui somma, su tutte le modalita' i possibili per il carattere, e' uguale a 1.

Funzione di ripartizione di una variabile statistica: Se le modalita' che assume un carattere sono numeri (reali) la funzione $F(x)$ che associa ad ogni numero reale x la proporzione di individui che possiedono una modalita' del carattere minore o uguale a x. Se le modalita' del carattere sono i numeri $x_{(1)} < x_{(2)} < \ldots < x_{(N)}$, allora $F(x) = \sum_{i: x_{(i)} \le x} n_i/n$. Si tratta evidentemente di una funzione non decrescente, costante tranne che in corrispondenza delle modalita' $x_{(i)}$ effettivamente presenti nella popolazione, in cui ha dei salti di ampiezza uguale alla frequenza (relativa) $\frac{n_i}{n}$ della modalita'. La funzione vale 0 per valori minori di $x_{(1)}$, 1 per valori maggiori o uguali a $x_{(N)}$.

Media di una variabile statistica: Definita per variabili aventi per modalita' numeriche, si tratta della media aritmetica dei valori della variabile, individuo per individuo, nella popolazione. Se gli individui sono numerati da un indice $k = 1, \ldots, n$ e x_k e' la modalita' rilevata sull'individuo k, allora la media e' uguale a $\bar{x} = \sum_{k=1}^{n} x_k$. Essa puo' anche essere calcolata moltiplicando ciascuna modalita' osservata per la sua frequenza (sul totale delle unita') e sommando su tutte le modalita', vale a dire $\bar{x} = \sum_{k=1}^{n} x_k = \sum_{i=1}^{N} x_{(i)} \frac{n_i}{n}$. La media e' quella modalita' del carattere che, se posseduta in quantita' uguale da tutte le unita', fornisce lo stessa quantita' totale del carattere posseduta dall'intera popolazione. La media da' un'idea piu' o meno rozza della regione della retta reale dove sono situate le modalita' piu' frequenti della variabile statistica. Data una variabile statistica, prendendo delle funzoni di questa possono essere ottente altre variabili di cui puo' ugualmente essere calcolata la media.

Varianza di una variabile statistica: Rilevando su ciascuna unita' statistica lo scarto del valore della variabile osservato sull'individuo dalla media, elevato al quadrato, si ottiene quindi una nuova variabile. La sua media si sice varianza della variabile e misura la dispersione della variabile intorno alla sua media. Se tutti gli individui hanno lo stesso valore della variabile, che in tal caso e' ovviamente la media, la varianza e' nulla. Precisamente la varianza e' uguale a

$$s^2 = n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{N} (x_{(i)} - \bar{x})^2 \frac{n_i}{n} = \sum_{i=1}^{N} x_{(i)}^2 \frac{n_i}{n} - (\bar{x})^2.$$

L'ultima uguaglianza fornisce uno strumento per il calcolo della varianza senza calcolare gli scarti $x_{(i)} - \bar{x}$: la varianza e' uguale alla media del quadrato della variabile meno il quadrato della media. La formula e' particolarmente utile quando le modalita' $x_{(i)}$ sono numeri interi,

mentre la media $\bar{x}$ non lo e'.
Variabili aleatorie e distribuzioni di massa, caso finito: In queste definizioni assumiamo che la popolazione statistica sia finita, nel qual caso anche le modalita' di qualsiasi carattere (rilevabili su tale popolazione) saranno anch'esse finite e i valori della distribuzione di frequenza del carattere saranno numeri razionali. Una prima generalizzazione si ottiene se assumiamo che ciascuna unita' statistica abbia un peso positivo, non necessariamente uguale a $\frac{1}{n}$ (probabilita'), con la somma dei pesi di tutte le unita' uguale a 1. Piuttosto che di variabile statistica e della sua distribuzione di frequenza parliamo allora di variabile aleatoria e della sua distribuzione di massa, la funzione che ad ogni modalita' fa corrispondere la somma dei pesi degli individui che la possiedono. Analogamente definiamo la funzione di ripartizione (che potremo chiamare anche **funzione di distribuzione**, e la media e la varianza di una variabile aleatoria.
Distribuzione di probabilita' binomiale: A partire da una popolazione statistica i cui individui possono essre distinti in due categorie mutuamente escludentisi (che convenzionalmente possiamo codificare con i numeri 1 e 0) in proporzione p e $1-p$ (popolazione binaria), formiamo la popolazione virtuale delle n-ple dei campioni di individui. La probabilita' di scegliere un campione con esattamente i individui (tra gli n scelti) che presentano la modalita' 1 e' allora, se la popolazione e' molto grande (per cui le estrazioni possono essere pensate indipendenti), pari a

$$\binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, 2, \dots, n$$

Il primo fattore (**coefficiente binomiale**) esprime il numero dei modi in cui si puo' scegliere in quali estrazioni vengono estratte le i unita' 1, mentre il secondo e' la probabilita' dell'uscita di i volte un'unita' 1 e $n-i$ volte un'unita' 0, in un ordine prefissato. Inserendo i coefficienti binomiali in una tabella con n quale indice di riga e i quale indice di colonna si ottiene un triangolo (detto di Tartaglia o di Pascal). Questo triangolo da' anche un algoritmo di calcolo dei coefficienti binomiali dato che possiamo ottenere tutte le scelte possibili di i oggetti tra n oggetti distinti andando a distinguere se queste scelte includono o meno un oggetto fissato. Ce ne sono $\binom{n-1}{i}$ che non lo fanno e $\binom{n-1}{i-1}$ che al contrario includono l'elemento. Sommando questi due numeri (che troviamo sulla riga $n-1$ del triangolo, nelle posizioni al di sopra e immediatamente a sinistra di quella desiderata) si ottiene $\binom{n}{i}$. Si noti che la somma su i delle probabilita' binomiali, per un qualunque intero n ed un qualunque $p \in (0, 1)$ e' uguale a 1 a questo si puo' verificare attraverso la celeberrima formula del binomio di Newton

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}$$

valida per a e b qualunque, prendendo $a = p$ e $b = 1-p$. Nel caso particolare $p = \frac{1}{2} = 1-p$ (distribuzione binomiale simmetrica), si ha $a^i b^{n-i} = 2^{-n}$ e quindi la somma della n-esima riga di coefficienti binomiali e' 2^n. I coefficienti binomiali sono simmetrici rispetto al valore centrale $\frac{n}{2}$, nel senso che $\binom{n}{i} = \binom{n}{n-i}$, mentre un valore di $p < 1/2$ ($p > \frac{1}{2}$) sposta le probabilita' verso la sinistra (destra) della distribuzione (valori piu' piccoli (grandi) di $\frac{n}{2}$). Una distribuzione binomiale ha media np e varianza $np(1-p)$, quindi e' tanto piu' grande quanto piu' p e' vicino ad 1/2.
Distribuzione di probabilita' geometrica: Una volta associati dei pesi agli individui di una popolazione possiamo anche trattare consentire popolazioni costituite da un numero infinito di unita' statistiche e quindi variabili aleatorie con un numero infinito di modalita' (che

devono tuttavia poter essere distinte da un indice a valori interi). Consideriamo ad esempio una popolazione formata mediante una procedura di campionamento di una popolazione binaria che si arresta una volta estratta per la prima volta un'unita' di tipo 1 (successo). La variabile che conta le estrazioni effettuate fino al conseguimento del primo successo ha una distribuzione che associa alla modalita' $i = 1, 2, \ldots$ la probabilita' $(1-p)^{i-1}p$. Di nuovo la somma di queste probabilita' su k e' uguale a 1, nel senso che

$$\sum_{i=1}^{+\infty}(1-p)^{k-1}p = \lim_{N\to\infty}\sum_{j=0}^{N-1}(1-p)^j p = \lim_{N\to\infty} p\frac{1-(1-p)^N}{1-(1-p)} = p\frac{1}{p} = 1$$

dove abbiamo utilizzato la formula $\sum_{k=0}^{N-1} a^N = \frac{1-a^N}{1-a}$, valida per $a \neq 1$. Utilizzando la stessa formula abbiamo la funzione di ripartizione

$$F(m) = \sum_{k=1}^{m}(1-p)^{k-1}p = 1-(1-p)^m, m = 1, 2, \ldots.$$

La media di una geometrica e' $1/p$ (piu' la probabilta' di successo e' grande meno si deve aspettare vedere un successo) e la varianza $\frac{1-p}{p}$.

Istogramma: Mentre una funzione di massa sugli interi puo' essere visualizzata con un grafico sul piano cartesiano in cui in ascissa, alla modalita' intera i, e' associata come ordinata la probabilita' di questa modalita', quando il carattere di interesse assume un continuo di valori (come altezza e peso), e' necessario utilizzare (a causa del carattere discreto di un qualsiasi strumento di misura) istogrammi. In corrispondenza ad un intervallo di valori della variabile rilevabile dallo strumento si riporta un rettangolo che ha questo intervallo come base ed area proporzionale alla frequenza (probabilita') della classe nella popolazione di interesse. Nel caso delle popolazioni finite, se ammettono strumenti di misura di precisione via via crescenti e quindi istogrammi con basi di lunghezza sempre piu' piccola, gli stogrammi divengono necessariamente piu' irregolari.

Funzioni di densita': Ricordando che l' integrale definito di una funzione reale di variabile reale e' ottenuto come limite dell'area totale di un istogramma, sembra naturale considerare una funzione positiva f su di un intervallo $[a, b]$ con la proprieta' che $\int_a^b f(x)dx = 1$ (detta funzione di densita') come distribuzione (virtuale) di una variabile aleatoria che assume un continuo di valori nell'intervallo $[a, b]$. La probabilita' che la variabile aleatoria assuma valori in un intervallo $[c, d] \subset [a, b]$ va calcolata come l'integrale definito come $\int_c^d f(x)$. Diciamo in questo caso che la variabile aleatoria (di solito indicata con lettere maiuscole X, Y, ecc.) ha densita' di probabilita' f in $[a, b]$. Dato che sembra naturale consentire che questo intervallo possa essere illimitato siamo condotti in questo caso a interpretare gi integrali come integrali impropri.

Funzioni di distribuzione: Riferita ad una densita' f, la funzione F che associa ad ogni x reale associa l'integrale improprio $F(x) = \int_{-\infty}^x f(t)dt$, che rappresenta la probablita' che una variabile aleatoria con densita' f assuma valori minori di (o uguali a) x. Se la densita' e' nulla per valori minori di a (ad esempio se la variabile misura delle quantita' positive, es. un tempo, sara' necessariamente $a = 0$), allora fino al punto a non c'e' alcuna densita' da integrare e quindi $F(x) = \int_a^x f(t)dt$. Per il teorema fondamentale del calcolo F e' l'unica primitiva di f che tende a 0 a $-\infty$ (ad a, nella situazione considerata sopra). Inoltre e' non decrescente e tende a 1 per $x \to +\infty$ (o per $x \to b < +\infty$ se la densita' e' nulla per valori superiori a b). Se F e' crescente (su tutta la retta reale o su di un intervallo $[a, b]$ che raccoglie tutta la densita') la funzione di distribuzione e' invertibile e la sua inversa si dice **funzione quantile**. Alcuni quantili hanno un interesse particolare, come la mediana, che e' il punto in cui la funzione di distribuzione vale $\frac{1}{2}$, e pertanto le probabilita' di valori superiori o inferiori alla mediana sono uguali. Di particolare importanza sono le funzioni di ripartizioni simmetriche che corrispondono alle densita' pari, per le quali per ogni $x > 0$

$$F(x) - F(0) = \int_0^x f(t)dt = \int_0^x f(-t)dt = \int_{-x}^0 f(s)ds = F(0) - F(-x)$$

Mandando x a $+\infty$ si ottiene $F(0) = 1/2$ e quindi $F(x) - \frac{1}{2}$ e' una funzione dispari, cioe' $F(x) + F(-x) = 1$.

Osservazione sulle funzioni di distribuzione: La funzione di distribuzione di una variabile aleatoria e' definita sia quando questa ha una funzione di massa sia quando ha una funzione di densita'. Nel primo caso la funzione e' costante tranne in corrispondenza delle modalita' della variabile, nel secondo cresce con continuita' dal valore 0 al valore 1. Se quescita crescita e' sufficientemente regolare da ammettere una derivata continua, allora tale derivata e' la funzione di densita'. Ma esistono delle funzioni di distribuzione continue che non derivabili che, pur avendo un valore essenzialmente teorico, in particolari applicazioni possono risultare importanti (fenomeni dotati di similarita' a tutte le scale, frattali).

Le probabilita' degli intervalli con la funzione di distribuzione: Se X e' una variabile con densita' f e funzione di ripartizione F abbiamo

$$P(a < X < b) = P(X < b) - P(X < a) = F(b) - F(a) = \int_a^b f(x)dx.$$

Tale formula vale anche se $a = -\infty$ e/o $b = +\infty$, nel qual caso si intende che $F(-\infty) = 0$ e $F(+\infty) = 1$. La funzione $\bar{F}(x) = 1 - F(x) = P(X > x)$ si dice **funzione di sopravvivenza**.

Densita' di una funzione di variabile aleatoria: Consideriamo una funzione monotona (crescente) e derivabile h definita sull'intervallo (a, b) dove e' concentrata la densita' f della variabile X. La funzione h ha allora un'inversa definita sull'intervallo (c, d) immagine di (a, b) sotto h. Grazie alla formula di sostituzione negli integrali definiti, detta Y la variabile aleatoria $h(X)$, e presi $c < \alpha < \beta < d$, si ha

$$P(\alpha < Y < \beta) = P(h^{-1}(\alpha) < X < h^{-1}(\beta)) = \int_{h^{-1}(\alpha)}^{h^{-1}(\beta)} f(x)dx = \int_\alpha^\beta \frac{f(h^{-1}(y)}{h'(h^{-1}(y))}dy$$

con la sostituzione $x = h^{-1}(y)$ nell'integrale. Abbiamo quindi che Y ha densita' $g(y) = \frac{f(h^{-1}(y)}{h'(h^{-1}(y))}$ per $y \in (c, d)$. Allo stesso risultato si poteva pervenire calcolando la funzione di distribuzione G di Y a partire dalla funzione di distribuzione F di X e poi derivando utilizzando la formula di derivazione delle funzioni composte e quella della funzione inversa

$$G(y) = P(Y < y) = P(X < h^{-1}(y)) = F(h^{-1}(y)), \quad G'(y) = F'(h^{-1}(y))\frac{1}{h'(h^{-1}(y))}.$$

Densita' di una variabile aleatoria in un nuovo sistema di riferimento: Di particolare interesse e semplicita' sono le funzioni affini $h(x) = cx + d$ (con $c > 0$), che corrispondono ad un cambiamento dell'origine e dell'unita' di misura in cui viene misurata la variabile aleatoria X. In questo caso la densita' e la funzione di distribuzione divengono

$$g(y) = \frac{1}{c}f(\frac{y-d}{c}), \quad G(y) = F(\frac{y-d}{c})$$

dove f e F sono rispettivamente la funzioni di densita' e quella di distribuzione della variabile X. Ogni funzione di densita' genera quindi una famiglia a due parametri di funzioni di densita' al variare di $c > 0$ e d reale qualsiasi. Il grafico di queste funzioni puo' essere ottenuto da quello della densita' originaria semplicemente riportando sull'asse delle ascisse e delle ordinate le nuove coordinate dei punti. Ad esempio se f e' pari, sull'origine si riporta la nuova ascissa d e il punto 1 viene avere distanza c da essa, mentre il punto 1 sull'asse delle ordinate viene ad avere la nuova ordinata $\frac{1}{c}$, in modo da mantenere invariate le aree.

Media di una variabile aleatoria con densita': La nozione di media per variabili aleatorie con funzione di massa si generalizza al caso di variabili aleatorie X con densita' f attraverso l'integrale $\int_a^b xf(x)dx$, che viene indicato con $E(X)$ (dall'inglese expectation). Al solito, $[a, b]$ e' l'intervallo dove e' concentrata tutta la densita'. Se tale intervallo e' limitato e f e' continua su questo intervallo, allora l'integrale esiste. Altrimenti, se l'intervallo e' illimitato e la funzione f ha degli asintoti, la media va interpretata come integrale improprio. La presenza di asintoti, tuttavia, non provoca divergenza dell'integrale perche' l'integrale di f e' convergente se tutta la retta reale e la funzione x e' continua in qualunque punto l'asintoto possa essere situato. L'eventuale divergenza della media e' quindi legata all'andamento della densita' a $-\infty$ e a $+\infty$. Dato che una funzione di densita' e' non negativa, il contributo positivo alla media viene dall'integrale sul semiasse reale positivo e quello negativo viene dal semiasse reale negativo. Possiamo quindi scrivere la media come differenza tra quantita' positive

$$E(X) = \int_0^{+\infty} xf(x)dx - \int_{-\infty}^0 |x|f(x)dx = \int_0^{+\infty} xf(x)dx - \int_0^{+\infty} xf(-x)dx$$

con l'avvertenza che se i due integrali sono uguali a $+\infty$, la media non esiste (se e' $+\infty$ solo una delle due, la media vale $+\infty$ o $-\infty$). In particolare se f e' pari, i due integrali sono uguali e quindi la media, se esiste, e' uguale a 0. Si noti che la somma dei due integralii non e' altro che $E(|X|) = \int_{-\infty}^{\infty} |x|f(x)dx$, la finitezza del quale e' condizione necessaria e sufficiente per l'esistenza della media. Entrambi gli addendi possono essere espressi mediante la funzione di ripartizione come

$$\int_0^{+\infty} xf(x)dx = \int_0^{+\infty} \bar{F}(x)dx, \quad \int_{-\infty}^0 |x|f(x)dx = \int_{-\infty}^0 F(x)dx$$

Interpretabili graficamente come l'area compresa tra l'asintoto orizzontae $y = 1$ e il grafico della funzione di ripartizione, per ascisse positive, e l'area tra lo stesso grafico e l'asintoto orizzontale $x = 0$, per ascisse negative. Quindi l'esistenza della media e' legata all'ordine di infinitesimo di $f(x)$ (o $\bar{F}(x)$), per $x \to +\infty$ e all'ordine di infinitesimo di $f(x)$ (o $F(x)$) per $x \to -\infty$. La finitezza della media, che sembra una questione piuttosto distante dalle applicazioni, viene ad avere risvolti concreti di grande importanza attraverso la **legge dei grandi numeri**, cui si daranno dei cenni nella parte finale dedicata alla statistica.

Varianza di una variabile aleatoria con densita': La media di una funzione $h(X)$ puo' essere calcolata senza passare per la densita' di X come $E(h(X)) = \int_{-\infty}^{+\infty} h(x)f(x)dx$. Supponiamo ora che la $E(X) = m$ sia finito e definiamo la varianza $var(X) = E(X - m)^2)$; dato che si tratta della media di una variabile positiva questa esiste finita o diverge a $+\infty$. Dato che $(x - m)^2 = x^2 - 2mx + m^2$, la linearita' degli integrali ci permette di arrivare alla comoda formula

$$var(X) = E(X^2) - (E(X))^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx - (\int_{-\infty}^{+\infty} xf(x)dx)^2$$

che chiarisce che l'esistenza della varianza pone vincoli piu' stringenti ad una densita' rispetto all'esistenza della media (l'integrabilita' di $x^2 f(x)$ piuttosto che quella di $|x|f(x)$ per $|x|$ grande. La radice quadrata della varianza prende il nome di **deviazione standard**. Ha il pregio di essere espressa nella stessa unita' di misura di X, mentre la varianza e' espressa nel quadrato dell'unita' di misura di X, visto che $var(aX) = a^2 var(X)$. Anche la finitezza della varianza e' legata ad un risultato di grande importanza applicativa, il **teorema del limite centrale**.

Standardizzazione di una variabile aleatoria: Se X e' una variabile aleatoria di media m e deviazione standard $\sigma > 0$, la variabile aleatoria che si ottiene con la trasformazione affine $Z = \frac{X-m}{\sigma}$ e' la **standardizzata di X**. In qualche modo la standardizzazione esprime la variabile in un sistema di riferimento intrinseco alla variabile, che dipende soltanto dalla sua distribuzione. Viceversa, data una variabile di media nulla e deviazione standard unitaria, la sua trasformata affine $X = m + \sigma Z$ ha media m e deviazione standard σ. Dato che $P(\alpha < X < \beta) = P(\frac{\alpha-m}{\sigma} < Z < \frac{\beta-m}{\sigma})$, nota la funzione di distribuzione di Z possiamo facilmente calcolare la probabilita' che X prenda valori in un qualsiasi intervallo. Il confronto tra le distribuzioni di due variabili aleatorie va effettuato sulla funzioni di distribuzione delle loro standardizzate, in modo che non dipenda dall'unita' di misura in cui le variabili sono espresse. Se la variabile non ha media e varianza finita, la standardizzazione non e' possibile e bisogna ricorrere ad indici alternativi. Ad esempio, per densita' simmetriche, la media puo' essere sostituita dalla mediana (che e' in questo caso il centro di simmetria) e la deviazione standard dalla meta' della **differenza interquartile**, vale a dire la differenza tra la funzione quantile in 3/4 (**terzo quartile**) e in 1/4 (**primo quartile**). In questo modo $P(Z < -1) = P(-1 < Z < 0) = P(0 < Z < 1) = P(Z > 1) = \frac{1}{4}$.

BESTIARIO DI FUNZIONI DI DENSITA' (controllare su Wikipedia in caso di dubbi)

Densita' beta: Le densita' beta sono proporzionali a $x^{a-1}x^{b-1}$ per $0 < x < 1$. I parametri a e b devono essere positivi. La costante di normalizzazione $B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ per cui la funzione va divisa per ottenere una densita' $f(x)$ non e' calcolabile esattamente a meno che a (o $2a$) e/o b (o $2b$) siano interi (grazie alla formula di ricorsione per la funzione Gamma e il fatto che $\Gamma(1) = 1$ e $\Gamma(1/2) = \sqrt{\pi}$). Solo in questi casi anche la funzione di ripartizione e' accessibile. La media e' uguale ad $\frac{a}{a+b}$ e la varianza a $\frac{ab}{(a+b)^2(a+b+1)}$. Abbiamo gia' visto che quando $a = b = 1/2$, $B(a, b) = \pi$, la funzione di distribuzione e' $\frac{1}{\pi}\arcsin(2x-1) + \frac{1}{2}$, per $0 < x < 1$ (**legge dell'arcoseno**) e la funzione quantile $\frac{1}{2}(1 + \sin[\pi(y - \frac{1}{2})])$. Quando $a = b = 1$ la funzione e' costantemente uguale a 1, la funzione di distribuzione e' uguale a x per $0 < x < 1$ (identica alla funzione quantile), la media uguale a $\frac{1}{2}$ (come sempre quando $a = b$) e la deviazione standard a $\frac{1}{2\sqrt{3}}$. In questo caso parliamo di densita' **uniforme** nell'intervallo $(0, 1)$. Con la trasformazione affine $Y = a + (b-a)X$ si ottiene una variabile uniformemente distribuita in un qualunque intervallo limitato (a, b). Con la stessa trasformazione si porta evidentemente una qualunque densita' beta ad essere positiva soltanto nell'intervallo (a, b).

Densita' Pareto: Abbiamo gia' osservato che $\frac{a}{x^{a+1}}$ e' una densita' sulla semiretta $[1, +\infty)$ per tutti gli $a > 0$. La funzione di ripartizione e' uguale a $1 - \frac{1}{x^a}$, per $x > 1$, e la funzione quantile a $(1-y)^{-1/a}$ per $0 < y < 1$. Dato che

$$\int_1^{+\infty} \frac{x}{x^{a+1}}dx = \frac{1}{a-1}, \int_1^{+\infty} \frac{x^2}{x^{a+1}}dx = \frac{1}{a-2},$$

rispettivamente se $a > 1$ e se $a > 2$, si ottiene che (nel primo caso) la media e' uguale a $\frac{a}{a-1}$ e (nel secondo caso) la varianza e' uguale a $\frac{1}{(a-2)(a-1)^2}$. Considerando la trasformazione lineare

$Y = cX$ si porta la densita' di Pareto ad essere concentrata su $[c, +\infty)$ (c e' il reddito minimo, nelle applicazioni economiche).

Densita' Gamma: Data la costante $a > 0$, la funzione $x^{a-1}e^{-x}$ ha un integrale convergente a $\Gamma(a)$ sulla semiretta positiva. Quindi dividendo la funzione per $\Gamma(a)$ si ottiene una funzione di densita'. La media e la varianza sono finite qualunque sia $a > 0$, e si ottengono facilmente, visto che $E(X) = \frac{\Gamma(a+1)}{\Gamma(a)} = a$ e $E(X^2) = \frac{\Gamma(a+2)}{\Gamma(a)} = (a+1)a$, da cui $var(X) = E(X^2) - (EX)^2 = a$. La funzione di distribuzione e' calcolabile esattamente solo quando a e' intero (si parla in questo caso di densita' di **Erlang**). Per $a = 1$ si ha la **densita' esponenziale negativa** e^{-x} sulla semiretta reale positiva che ha la funzione di distribuzione $\int_0^x e^{-t}dt = 1 - e^{-x}$ per $x > 0$ (e la funzione quantile $-\log(1-y)$ per $0 < y < 1$). Si noti che la media e la varianza dell'esponenziale negativa sono uguali a 1. Le funzioni di ripartizione con a intero si ottengono dall'esponenziale negativa con integrazioni per parti. Ad esempio, per $x > 0$

$$\int_0^x te^{-t}dt = [-te^{-t}]_0^x + \int_0^x e^{-t}dt = 1 - e^{-x}, \quad \int_0^x \frac{t^2}{2}e^{-t}dt = [-\frac{t^2}{2}e^{-t}]_0^x + \int_0^x te^{-t}dt = 1 - e^{-x} - xe^{-x}$$

e cosi' via. La densita' esponenziale e' un modello teorico per tempi di vita che non sono affetti da invecchiamento. Infatti, se e' noto che un apparecchio con tempo di vita X con densita' esponenziale negativa e' in funzione al tempo s, la probabilita' che si guasti prima del tempo $t + s$ e'

$$\frac{P(s < X < t+s)}{P(X > s)} = \frac{e^{-s} - e^{-(t+s)}}{e^{-s}} = 1 - e^{-t} = P(X < t)$$

la probabilita' che si rompa prima al tempo t da nuovo (proprieta' di mancanza di memoria). Oltre alle densita' gamma con parametro a intero hanno particolare interesse quelle con $2a$ intero. Prima di darne un cenno, osserviamo che possiamo cambiare l'unita' di misura di una variabile X con distribuzione Gamma moltiplicando per la costante λ^{-1} in modo da ottenere la densita' di $Y = \frac{X}{\lambda}$ della forma piu' generale

$$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1}e^{-x}, x > 0.$$

Quando $a = 1$ si ottengono ancora delle densita' esponenziali parametrizzate da λ (che si vede facilmente essere il reciproco della media, mentre la deviazione standard e' $\frac{1}{\sqrt{\lambda}}$). Per $\lambda = 1/2$ e $2a = m$ intero si ottengono invece le densita' chi quadrato con m gradi di liberta', particolarmente importanti in varie applicazioni statistiche. Il chi quadrato con m gradi di liberta' e' la densita' della somma dei quadrati di m variabili aleatorie gaussiane standard indipendenti. Siamo in grado di verificare questo solo per $m = 1$. Dato che $X^2 = |X|^2$, se X e' gaussiana standard ha densita' pari e quindi possiamo utilizzare la forma delle densita' di funzioni di variabili aleatorie con $h(x) = x^2$ e $f(x) = \sqrt{\frac{2}{\pi}}$ per $x > 0$ per ottenere la densita' g di X^2. Dato che $h^{-1}(y) = \sqrt{y}$, $h'(x) = 2x$, si ha

$$g(y) = \frac{f(h^{-1}(y)}{h'(h^{-1}(y))} = \frac{1}{\sqrt{2\pi y}} e^{-y/2}, y > 0,$$

che e' la densita' gamma con $a = \lambda = 1/2$.

Densita' Gaussiana o normale: La funzione $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ e'universalmente nota come densita' gaussiana standard. Mentre il fatto che l'integrale sull'intera retta reale e' unitario richiede strumenti piu' sofisticati per la verifica, la standardizzazione e' di verifica immediata. Infatti la funzione $xe^{-\frac{x^2}{2}}$ ha come primitiva $-e^{-\frac{x^2}{2}}$ e quindi

$$\int_0^{+\infty} xe^{-\frac{x^2}{2}}dx = [-e^{-\frac{x^2}{2}}]_0^{+\infty} = 1$$

da cui, per la parita' di ϕ, la media della densita' gaussiana standard e' 0. Di nuovo per la parita' di ϕ la varianza e' allora

$$2\int_0^{+\infty} x^2\phi(x)dx = \frac{2}{\sqrt{2\pi}}[xe^{-\frac{x^2}{2}}]_0^{+\infty} + \frac{2}{\sqrt{2\pi}}\int_0^{+\infty} e^{-\frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}dx = 1$$

come volevasi dimostrare. La funzione di distribuzione corrispondente $\Phi(x) = \int_{-\infty}^{x}\phi(t)dt$ non si puo' calcolare mediante funzioni elementari ed e' riportata in forma tabellare in tutti i libri di probabilita' e statistica (in virtu' della sua simmetria, e' riportata solo per $x > 0$ o talvolta viene riportata la funzione dispari $\Phi(x) - \frac{1}{2}$, sempre per $x > 0$). Naturalmente anche la funzione quantile viene calcolata attraverso le tavole. Molto spesso sono i quantili di $|X|$ di interesse, quando X e' Gaussiana standard. Dato che la probabilita' che $|X|$ sia inferiore o uguale ad un numero positivo x e' $\Phi(x) - \Phi(-x) = 2\Phi(x) - 1$, la soluzione in x di $2\phi(x) - 1 = p$ (cioe' la funzione quantile di $|X|$ calcolata in p e' data da $\Phi^{-1}(1 + \frac{p}{2})$, utile quando e' Φ ad essere riportata nella tabella. Ad esempio e' noto che $\Phi(1) = 0.84$, e quindi $2\Phi(1) - 1 = 0.68$ (la probabilita' che una Gaussiana standard si allontani dalla media non piu' di una deviazione standard). Ancor piu' utile e' sapere che $\Phi(1.96) = 0.975$ e quindi $2\Phi(1.96) - 1 = 0.95$ (19 volte su 20 una Gaussiana standard prende valori che distano non piu' di 1.96 deviazioni standard dalla media). Per m reale qualsiasi e $\sigma > 0$ la densita'

$$\frac{1}{\sigma}\phi(\frac{y-m}{\sigma}) = \frac{1}{\sigma\sqrt{2\pi}}^{-\frac{(y-m)^2}{2\sigma^2}}$$

e' la densita' gaussiana di media m e varianza σ^2. Le tavole della sua funzione di distribuzione $\Phi(\frac{y-m}{\sigma})$ si possono ottenere da quelle di Φ. Se si desidera la probabilita' di un intervallo, si devono standardizzare gli estremi e procedere direttamente con la funzione di distribuzione standard Φ. La densita' gaussiana ha un carattere universale (attrae le altre forme distribuzional quando si sommano variabili indipendenti di varianza finita con identica distribuzione) che viene espresso dal **teorema del limite centrale**, che vedremo in seguito.

Densita' Cauchy e Student: Una funzione di densita' che presenta caratteristiche completamente diverse da quelle della gaussiana e' la densita' di Cauchy $f(x) = \frac{1}{\pi(1+x^2)}$ che e' anch'essa pari e definita sull'intera retta reale. Dato che la densita' e' equivalente a x^{-2} per $|x|$ grande la sua media non esiste perche' $\frac{x}{x^2}$ non ha un integrale convergente su semirette illimitate a destra. La sua funzione di distribuzione e'

$$\int_{-\infty}^{x} \frac{1}{\pi(1+t^2)}dt = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$$

e la funzione quantile e' $\tan[(y - \frac{1}{2})\pi]$, definita al solito per $0 < y < 1$. La funzione di distribuzione cresce ad 1 molto piu' lentamente di una gaussiana, al crescere di x, mentre la funzione quantile tende a $+\infty$ molto piu' velocemente di una gaussiana, al tendere di y a 1. A conferma di questo si noti che la funzione quantile in 3/4 vale $\tan(\pi/4) = 1$ e quindi la densita' di Cauchy da' la stessa probabilita' $\frac{1}{4}$ a valori rispettivamente minori di -1, compresi tra -1

e 0, tra 0 e 1 e maggiori di 1. In particolare l'intervallo $(-1, 1)$ ha probabilita' 0.5 contro lo 0.68 della gaussiana standard, mentre l'intervallo $(-\sqrt{3}, \sqrt{3})$ ($\sqrt{3} = 1.732$), ha probabilita' $2/3$, leggermente inferiore a 0.68 (si verifichi queste assezioni tenendo presente che $\tan(\frac{\pi}{4}) = 1$ e $\tan(\frac{5\pi}{6}) = \sqrt{3}$. Naturalmente questo confronto e' in qualche modo mal posto, perche' se consideriamo la variabile $Y = b+aX$, con X avente densita' di Cauchy, questi numeri vengono cambiati (le ampiezze degli intervalli vanno divise per a). Tuttavia sara' sempre vero che le probabilita' di valori sufficientemente grandi della variabile saranno relativamente piu' grandi che per una gaussiana. Incidentalmente notiamo che la densita' di Y si puo' scrivere, con semplici passeggi, nella forma

$$\frac{1}{a} f(\frac{y-b}{a}) = \frac{a}{\pi(a^2 + (y-b)^2)}.$$

Concludiamo questa presentazione della densita' di Cauchy osservando che, se U e' un angolo scelto uniformemente in $(-\pi/2, \pi/2)$, allora $\tan U$ ha la densita' di Cauchy (si calcoli la funzioni di densita' o la funzione di distribuzione di $\tan U$). Per finire, accenniamo anche al fatto che la densita' di Cauchy e' caso particolare di una famiglia di densita' che dipendono da un parametro intero m (il numero di gradi di liberta'), le densita' di Student. Tali densita' sono proporzionali a

$$(1 + x^2/m)^{-\frac{m+1}{2}}.$$

La densita' di Cauchy si ottiene quindi per $m = 1$. Come si vede al crescere di m queste densita' convergono sempre piu' velocemente a 0 per $|x| \to +\infty$. Per $m \to \infty$ la densita' di Student converge alla Gaussiana standard (nel senso che le probabilita' di un qualsiasi intervallo convergono a quelle calcolate con la Φ). Utilizzando i teoremi di confronto per integrali impropri, e' facile verificare che per l'esistenza della media deve essere $m > 1$ e per l'esistenza della varianza deve essere $m > 2$. Un'applicazione delle densita' di Student verra' data nei cenni sulla statistica che chiuderanno il corso.