Università di Firenze - Corso di Laurea in Statistica
Statistica
A.A. 2010/201

Variabili casuali continue & alcuni modelli probabilistici

Cicchitelli: parte del cap 13 e parte del cap 14

A cura di Leonardo Grilli

Distribuzioni di probabilità continue

- Una variabile aleatoria continua è una variabile che può assumere qualunque valore in un intervallo
  - spessore di un oggetto
  - tempo necessario per completare un lavoro
  - temperatura di una soluzione
  - altezza di una persona
- Queste variabili possono potenzialmente assumere qualunque valore
- Strumento di misura di precisione finita → l'insieme dei possibili valori è finito
  - Es. X= lunghezza di un cilindro
    - Se il metro misura fino ai centimetri → (…, 19 cm, 20 cm, …)
    - Se il metro misura fino ai millimetri → (…, 19.8 cm, 19.9 cm, 20.0 cm, 20.1 cm, 20.2 cm,…)

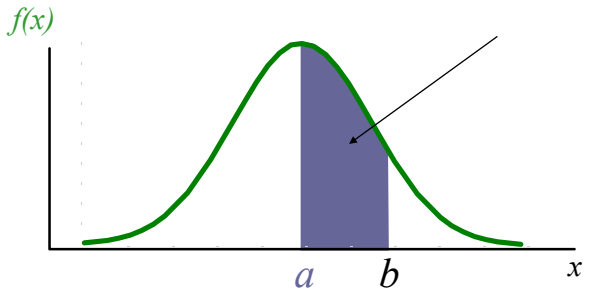
Funzione di densità di probabilità

- Supponiamo di assegnare una probabilità piccola quanto si vuole, ma non nulla, ad ognuno dei punti di un insieme con la cardinalità del continuo (es. l'intervallo [0,1] dei numeri reali)
  - la somma delle probabilità sarebbe infinita e quindi non potrebbe soddisfare il requisito di essere pari a 1 per l'evento certo
- Soluzione: assegnare la probabilità agli intervalli
- Come? Con una **funzione di densità di probabilità** $f(\,)$

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

La probabilità come area

L'area ombreggiata sottesa alla curva è la probabilità che X assuma valori fra a e b



$$P(a \le X \le b)$$
$$= P(a < X \le b)$$
$$= P(a \le X < b)$$
$$= P(a < X < b)$$

La probabilità di un singolo valore è zero: $P(X = x) = 0$

Proprietà della funzione di densità

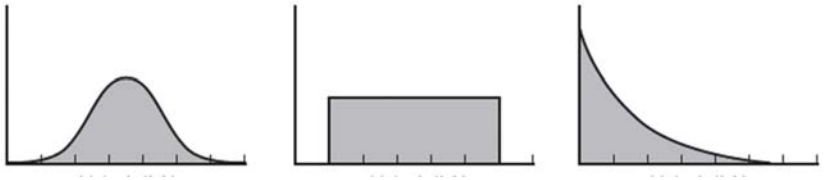
La **funzione di densità di probabilità**, $f(x)$, di una variabile aleatoria X ha le seguenti proprietà:

1. $f(x) \geq 0$ per qualunque numero reale x
2. L'area sottesa alla funzione di densità di probabilità $f(x)$ su tutto l'asse dei reali vale 1:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Si chiama **supporto della v.a. X** il sottoinsieme dei reali per cui la densità è positiva (→ gli integrali possono essere calcolati sul solo supporto senza alterare il risultato)

Funzione di densità: esempi



Distribuzione normale

Supporto: $(-\infty, +\infty)$

Distribuzione uniforme

Supporto: $[a,b]$

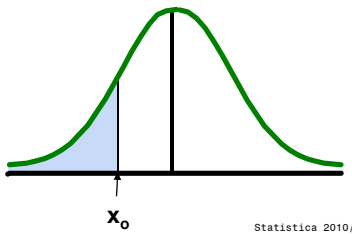
Distribuzione esponenziale

Supporto: $(0, +\infty)$

La funzione di ripartizione

- La **Funzione di ripartizione** (o **Cumulata**), $F(x_0)$, di una variabile aleatoria continua X esprime la probabilità che X non superi il valore x_0

$$F(x_0) = P(X \leq x_0) = \int_{-\infty}^{x_0} f(x)dx$$

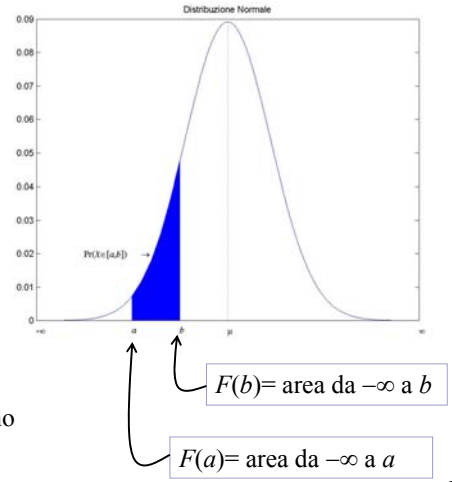


area sottesa alla funzione di densità $f(x)$ fino al valore x_0

Funzione di ripartizione e probabilità

La probabilità corrispondente ad un qualunque intervallo può essere sempre espressa in termini della funzione cumulata $F(\ )$

- $P(a<X\leq b) = F(b)-F(a)$
- $P(X\leq b) = F(b)$
- $P(X>a) = 1- P(X\leq a) = 1-F(a)$



ricordiamo che in una v.a. continua $<$ equivale a $\leq$ per cui con $F(\ )$ si calcolano le probabilità per tutti i tipi di intervallo

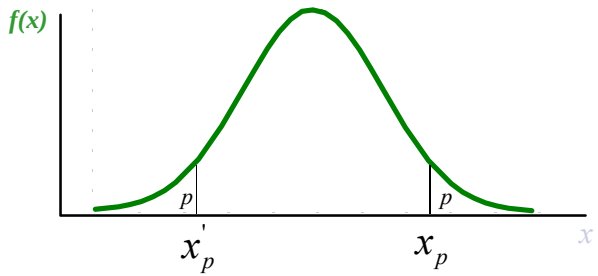
Relazione tra f. densità e f. ripartizione

- La funzione di ripartizione F e la funzione di densità f sono equivalenti poiché si può passare dall'una all'altra in modo univoco:

$$F(x) = \int_{-\infty}^{x} f(u)du \qquad f(x) = \frac{\partial}{\partial u} F(u)\Big|_{u=x}$$

Quantili

- Quantile *p inferiore*: $P(X < x_p') = p$
- Quantile *p superiore*: $P(X > x_p) = p$



Valori attesi di v.a. continue

- Nel continuo le sommatorie diventano integrali
- La media di X, indicata con μ_X, è

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

- La varianza di X, indicata con σ_X^2, è definita come il valore atteso del quadrato degli scarti della variabile dalla sua media, $(X - \mu_X)^2$

$$\sigma_X^2 = E\left((X-\mu_X)^2\right) = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f(x)dx$$

Trasformazione lineare di una v.a.

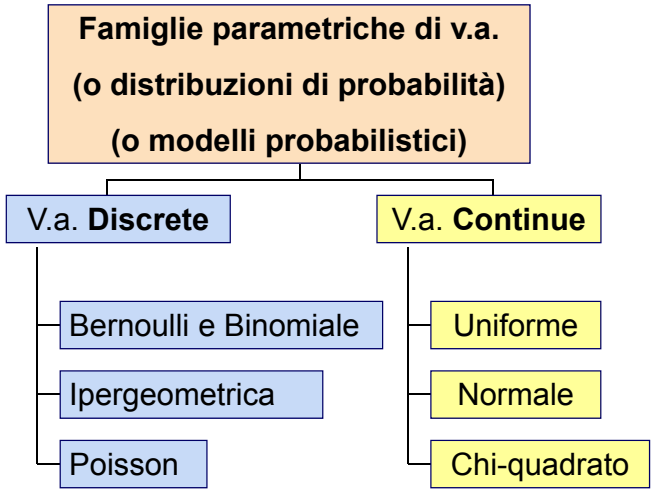
- Il comportamento dell'operatore "valore atteso" per v.a. continue è analogo a quello per v.a. discrete
- Sia $W = a + bX$, dove X ha media μ_X e varianza σ_X^2, e a e b sono costanti
- Allora la media di W è

$$\mu_W = E(a+bX) = a + b\mu_X$$

- la varianza e deviazione std di W sono

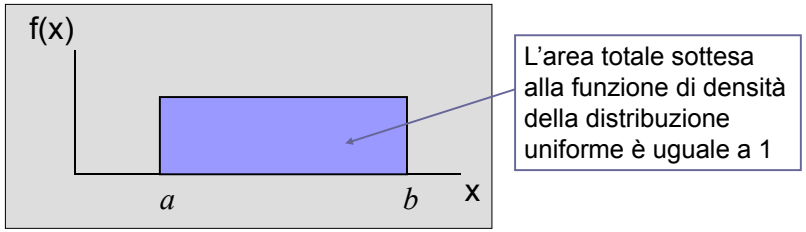
$$\sigma_W^2 = Var(a+bX) = b^2\sigma_X^2 \qquad \sigma_W = |b|\sigma_X$$

Famiglie parametriche che tratteremo



La distribuzione Uniforme /1

- La **distribuzione continua Uniforme** è la distribuzione di probabilità che assegna la ***stessa probabilità*** a tutti gli intervalli



La distribuzione Uniforme /2

Funzione di densità di probabilità

$$f(x) = \begin{cases} \dfrac{1}{b-a} & se\ a \le x \le b \\ 0 & altrove \end{cases}$$

a = valore minimo di x

b = valore massimo di x

Funzione di ripartizione

$$F(x) = \begin{cases} 0 & \text{se } x < a \\ \dfrac{x-a}{b-a} & \text{se } x \in [a,b] \\ 1 & \text{se } x > b \end{cases}$$

La distribuzione Uniforme /3

- La **media** di una distribuzione uniforme è

$$\mu = \frac{a+b}{2}$$

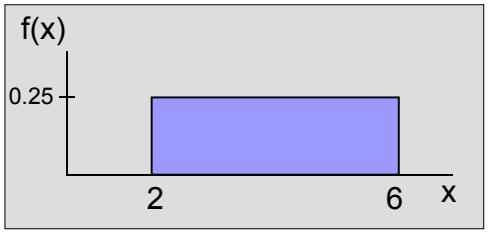
- La **varianza** è

$$\sigma^2 = \frac{(b-a)^2}{12}$$

La distribuzione Uniforme /4

Esempio: Distribuzione di probabilità uniforme nell'intervallo [2, 6]:

$$f(x) = \frac{1}{6-2} = 0.25 \qquad \text{per } 2 \le x \le 6$$



$$\mu = \frac{a+b}{2} = \frac{2+6}{2} = 4$$

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{(6-2)^2}{12} = 1.333$$

La distribuzione Normale /1

- E' la distribuzione più usata perché
  - descrive bene molti fenomeni
  - ha proprietà matematiche convenienti
  - il **teorema limite centrale** afferma che asintoticamente (= al crescere del numero di osservazioni) la distribuzione della media campionaria tende ad una Normale, qualunque sia la distribuzione di probabilità delle osservazioni

È stata proposta da F. Gauss (1809), che la utilizzò per primo nello studio degli errori di misurazione relativi alla traiettoria dei corpi celesti (per questo è chiamata anche *gaussiana*)

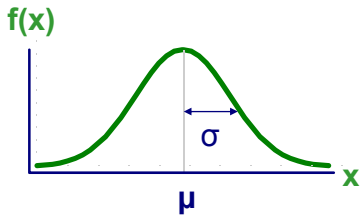
La distribuzione Normale /2

→ **'Forma campanulare'**
→ **Simmetrica**
→ **Media, Mediana e Moda coincidono**

La *tendenza centrale* è determinata dal parametro μ (media)

La *variabilità* è determinata dal parametro σ (deviazione std)

La variabile aleatoria ha un campo di variazione teoricamente infinito: da $-\infty$ a $+\infty$



Media
= Mediana
= Moda

La distribuzione Normale /3

- Famiglia parametrica di distribuzioni continue su supporto $(-\infty, +\infty)$

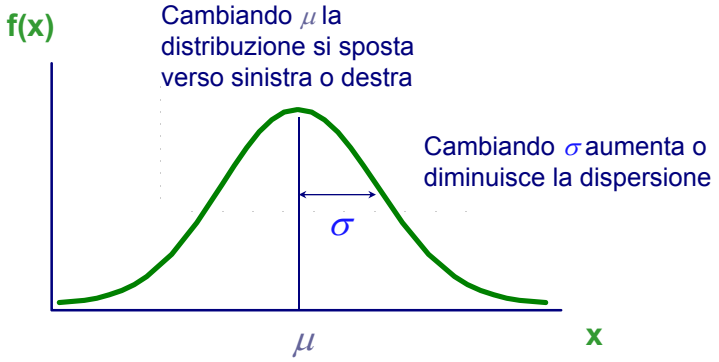
$$X \sim N(\mu, \sigma^2) \qquad \mu \in (-\infty, +\infty) \quad \sigma^2 \in [0, +\infty)$$

- La famiglia Normale è caratterizzata dai due parametri μ e σ^2 → ad es. $N(-8.1, 2.3)$ e $N(-8.1, 2.4)$ sono membri distinti, ma $N(-8.1, -2.3)$ non è un membro
- Per ogni coppia (μ, σ^2) la funzione di densità della Normale è

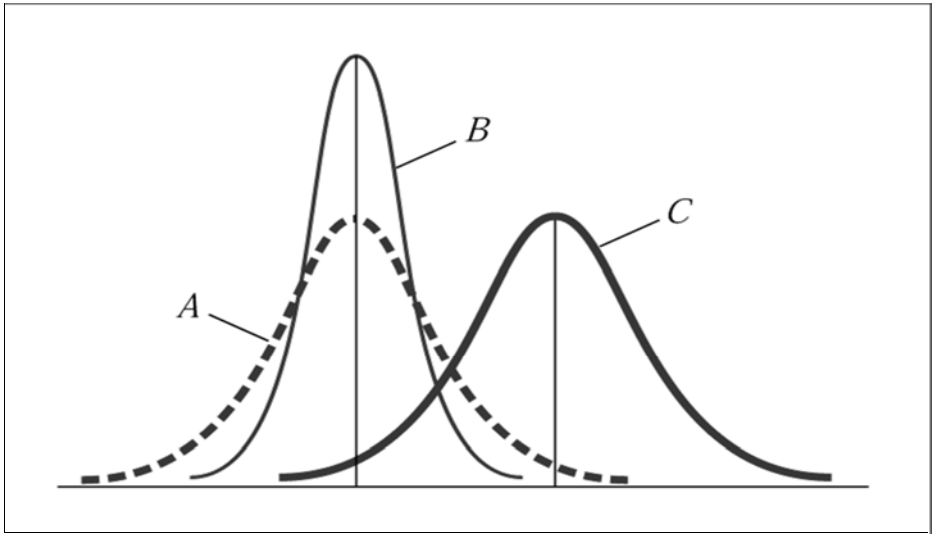
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}\right]}$$

$e \cong 2.71828$
$\pi \cong 3.14159$

La forma della distribuzione Normale



Nella distribuzione Normale la media e la varianza sono due parametri distinti → la varianza non dipende dalla media, come invece accade per molte distribuzioni (es. la binomiale)

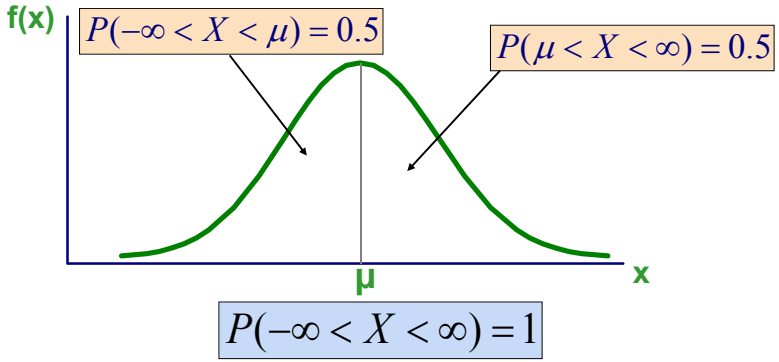


Distribuzioni Normali con valori differenti dei parametri μ e σ

[la distribuzione è individuata indifferentemente usando σ^2 o σ, es. si può dire Normale di media 0 e varianza 9 o Normale di media 0 e deviazione standard 3]

Alcune probabilità notevoli

- L' area totale sottesa alla curva è pari a 1, e la curva è simmetrica, perciò metà è al di sopra della media, e metà è al di sotto



Caratteristiche della Normale

- Per ogni coppia (μ, σ^2) la f. di densità Normale ha le seguenti caratteristiche:
  - È positiva per ogni x reale
  - L'area sottesa alla curva è 1

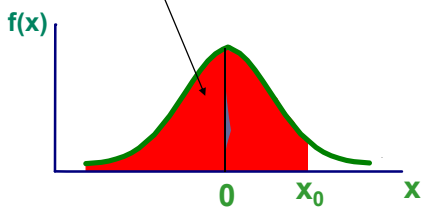
  } Proprietà di ogni densità

  - La media (valore atteso) coincide con il parametro μ (il simbolo del parametro non è stato scelto a caso!)
  - È simmetrica unimodale, per cui μ non è solo la media, ma anche
    - la mediana (μ lascia a sinstra e a destra un'area pari a 0.5)
    - e la moda ($x=\mu$ è il punto in cui la curva ha la massima altezza)
  - La varianza coincide con il parametro σ^2 e quindi la deviazione standard è σ (anche qui il simbolo del parametro non è stato scelto a caso!)
  - La curva ha due punti di flesso (cambia la concavità) in $\mu \pm \sigma$
  - Quando $x \to -\infty$ o $x \to +\infty$ la curva tende a zero (senza mai diventare esattamente zero: l'asse delle ascisse è un asintoto della curva)

Funzione di ripartizione Normale

- Per una v.a. Normale X con media μ e varianza σ^2 , ovvero X~N(μ, σ^2), la funzione di ripartizione è

$$F(x_0) = P(X \le x_0) = \int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}\right]} dx$$



Standardizzazione

- Data una qualunque v.a. X con media μ_X e deviazione standard σ_X , si definisce standardizzata la v.a. Z

$$Z = \frac{X - \mu_X}{\sigma_X}$$

- Per costruzione, si ha $\mu_Z = 0$ e $\sigma_Z = 1$ (si dimostra usando le proprietà delle trasformazioni lineari di v.a.)
- La trasformazione inversa è

$$X = \mu_X + \sigma_X Z$$

- Caso speciale: Se X ~ N(μ_X,σ_X^2), allora Z ~ N(0,1)

Normale standard

- La **Normale standard** Z, il membro con media 0 e varianza 1, funge da "rappresentante" della famiglia

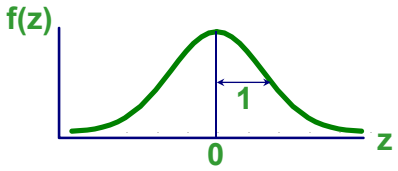
$$Z \sim N(0,1)$$

- funzione di densità:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- funzione di ripartizione:

$$\Phi(x_0) = \int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$



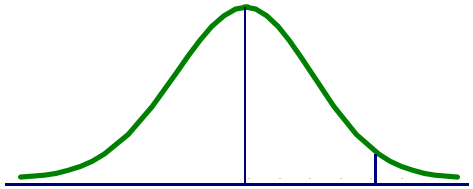
Normale standard: esempio

- Se X ha una distribuzione normale con media 100 e deviazione std 50, il valore di Z corrispondente a X = 200 è

$$Z = \frac{X - \mu}{\sigma} = \frac{200 - 100}{50} = 2.0$$

- Ciò significa che x = 200 è 2.0 deviazioni standard (= 2.0 incrementi di 50 unità) al di sopra del valore medio 100

Confrontando le unità di X e Z



100 200 X (μ = 100, σ = 50)

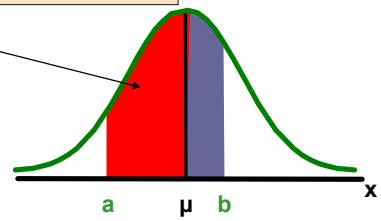
0 2.0 Z (μ = 0, σ = 1)

Notare che la distribuzione è la stessa, è cambiata solo la scala. Possiamo formulare il problema usando le unità originali (X) o le unità standardizzate (Z)

Calcolare le probabilità /1

La probabilità relativa ad un intervallo di valori è misurata dall'area sottesa alla curva e può essere espressa come differenza tra la funzione di ripartizione calcolata negli estremi dell'intervallo

$$P(a < X < b) = F(b) - F(a)$$

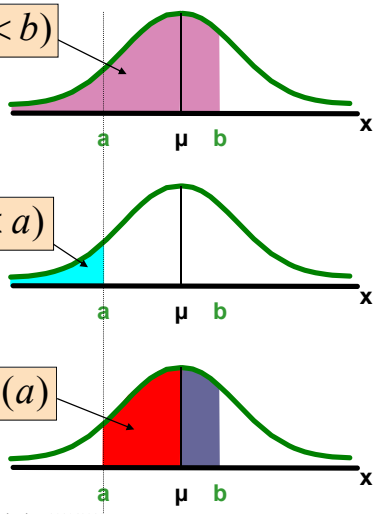


Calcolare le probabilità /2

$$F(b) = P(X < b)$$

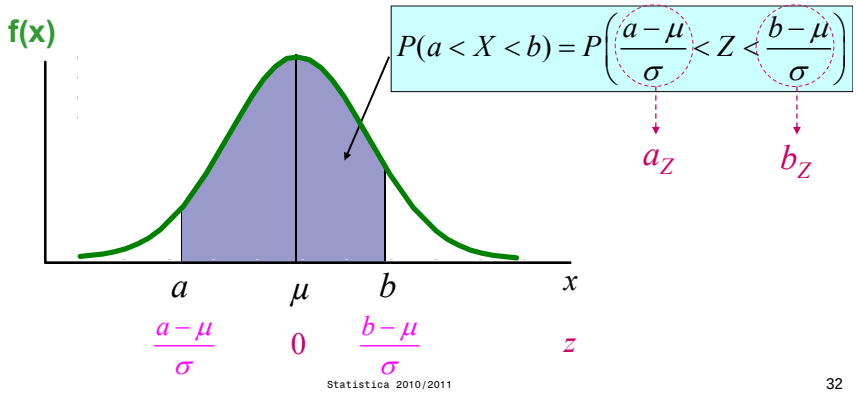
$$F(a) = P(X < a)$$

$$P(a < X < b) = F(b) - F(a)$$



Calcolare le probabilità /3

- Si effettua una ***standardizzazione*** per trasformare $P(a \le X \le b)$ con $X \sim N(\mu, \sigma^2)$ in $P(a_Z \le Z \le b_Z)$ con $Z \sim N(0,1)$



Calcolare le probabilità /4

- Il calcolo delle probabilità per una Normale con media e varianza qualunque si può sempre riportare al calcolo per la Normale standard

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

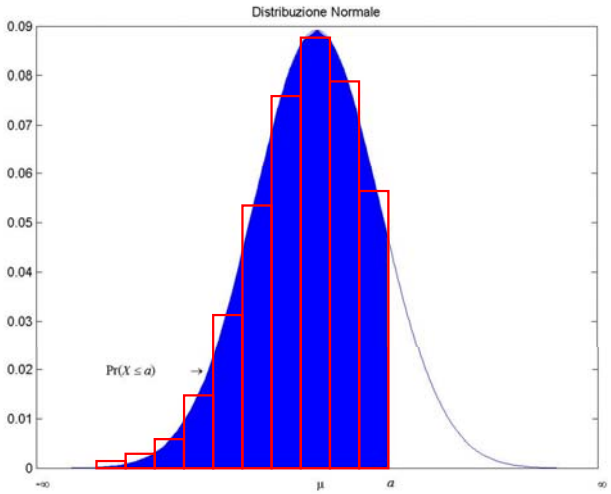
dove

$$\Phi(x_0) = \int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Ma questo integrale non ha soluzione analitica!

- Il valore dell'integrale può essere ben approssimato per via numerica, cioè l'area sottostante alla curva nell'intervallo $(-\infty, x_0]$ può essere calcolata in modo approssimato per mezzo di figure geometriche semplici di cui è facile calcolare l'area, ad es. rettangoli

Approssimazione dell'area

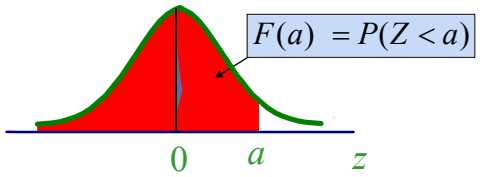


L'area sottesa alla curva in $(-\infty, a]$ è approssimata dall'area totale dei rettangoli (con rettangoli più stretti si ottiene un'approssimazione migliore)

La tavola della Normale standard /1

- La tavola della Normale standard data nel libro (Tavola C.2 dell'Appendice) fornisce i valori della funzione di ripartizione della distribuzione normale
- Per un dato valore a di Z, la tavola fornisce F(a)

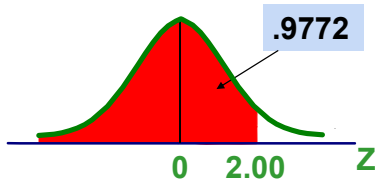
  (l'area sottesa alla curva da meno infinito al valore a)



La tavola della Normale standard /2

- **La tavola C.2 dell'Appendice** fornisce la probabilità F(a) per qualunque valore a tra **0** e **3.49**

Esempio:
P(Z < 2.00) = .9772



P(Z < 3.49) è quasi 1 → la tavola riporta 0.9998

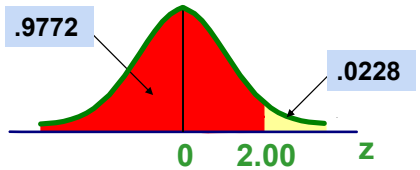
Per un valore più grande di 3.49 la probabilità è ancora più vicina a 1 → la tavola non riporta il valore

Es. P(Z < 5.22) è quasi 1

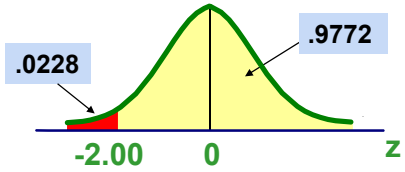
La tavola della Normale standard /3

- Per **valori negativi di Z**, usiamo il fatto che la distribuzione è simmetrica per trovare la probabilità desiderata:

In simboli $\Phi(z)=1-\Phi(-z)$



Esempio:
P(Z < -2.00) = 1 – 0.9772
= 0.0228



Procedura generale

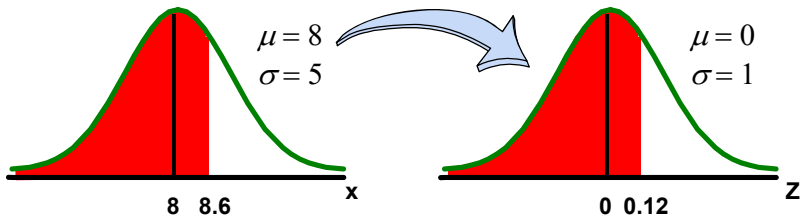
Per calcolare $P(a < X < b)$ quando X ha distribuzione Normale:

- Disegna la curva Normale per il problema in termini di x
- Traduci i valori di x in valori di z
- Usa la Tavola della Funzione di Ripartizione

Esempio coda sinistra: P(X < 8.6)

$$X \sim N(\mu = 8, \sigma^2 = 5^2) \qquad P(X < 8.6) = ?$$

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8.0}{5.0} = 0.12$$



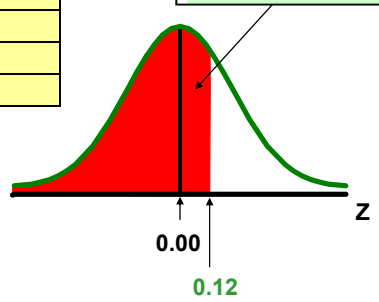
P(X < 8.6) P(Z < 0.12)

Esempio coda sinistra: P(X < 8.6)

Tavola della distribuzione Normale standard

	0.01	0.02	0.03	…
0.1		**0.5478**		
0.2				
0.3				
…				

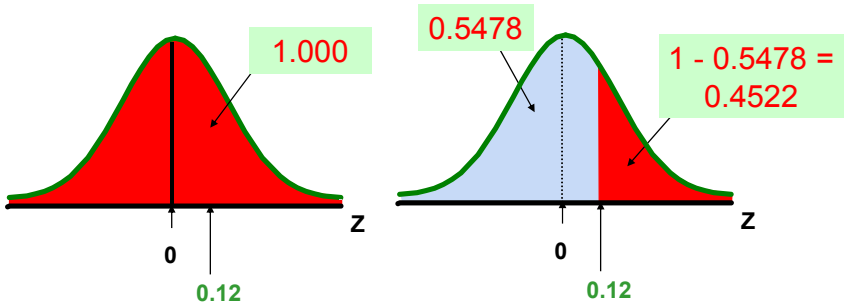
P(X < 8.6)
= P(Z < 0.12)
F(0.12) = 0.5478



Esempio coda destra: P(X > 8.6)

- Adesso calcoliamo P(X > 8.6)…

$$P(X > 8.6) = P(Z > 0.12) = 1 - P(Z \le 0.12)$$
$$= 1 - 0.5478 = 0.4522$$



Problemi diretti e inversi

- Problema diretto: dato un valore di *z* determinare la probabilità cumulata $\Phi(z)$ [in termini geometrici: dato un punto *z* sulle ascisse determinare l'area sottesa alla densità ϕ alla sinistra di *z*]
- Problema inverso: dato un valore *p* della probabilità cumulata, determinare il valore z_p corrispondente, cioè z_p tale che $\Phi(z_p) = p$ [in termini geometrici: determinare il punto z_p per il quale alla sua sinistra l'area sottesa alla densità ϕ è pari ad un certo valore specificato]
- Finora abbiamo visto solo problemi diretti, adesso consideriamo alcuni problemi inversi

Problema inverso

Passi per trovare il valore di x corrispondente ad una data probabilità:

1. Trovare il valore di Z corrispondente alla probabilità data
2. Convertire nelle unità di X usando l'inversa della standardizzazione, cioè:

$$X = \mu + \sigma Z$$

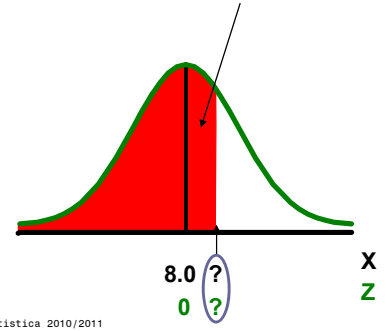
Problema inverso: esempio /1

Esempio:

- Assumiamo che in un certa località la temperatura minima X abbia una distribuzione Normale con media 8 C° e deviazione std 5 C°.
- Adesso troviamo il valore di X tale che l'80% dei valori siano al di sotto

L'80% delle temperature è inferiore a ___ C°

?



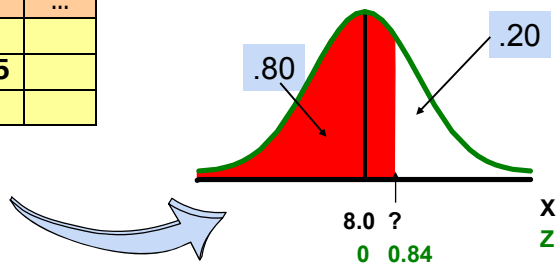
Problema inverso: esempio /2

1. Trova il valore di Z corrispondente alla probabilità data

Tavola della Funzione di Ripartizione Normale

	…	0.04	…
…			
0.8		**0.7995**	
…			

- 80% di area a sinistra corrisponde al valore Z di 0.84



Problema inverso: esempio /3

2. Converti in unità di X:

$$x = \mu + \sigma z = 8 + 5(+0.84) = 12.2$$

Perciò 80% dei valori di una distribuzione Normale con media 8 e deviazione std 5 sono inferiori a 12.2

L'80% delle temperature è inferiore a 12.2 C°

Scala e unità di misura

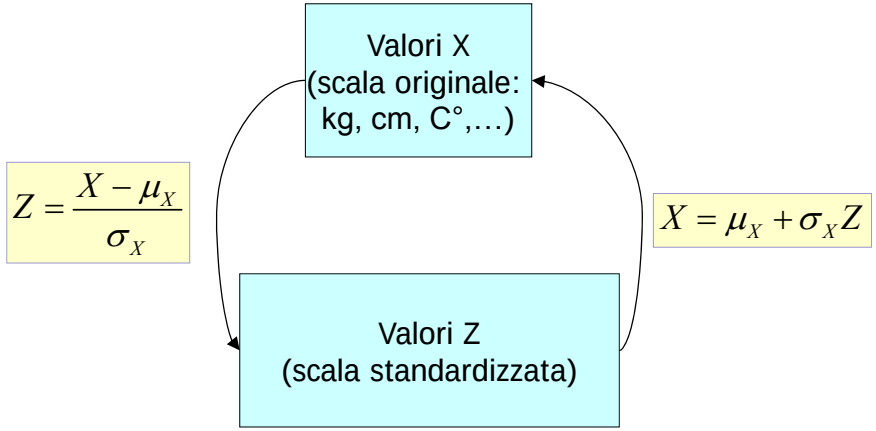
- Quando il problema è di tipo inverso si parte dalla Normale standard per ottenere il valore z desiderato
  - z è in scala standard → non ha unità di misura
- Poi si applica la trasformazione inversa della standardizzazione,

$$x = \mu_X + \sigma_X z$$

che reintroduce la media μ_X e la deviazione standard σ_X originali

  - x è nella scala originale (kg, cm, secondi, C°,…)

Schema dei cambiamenti di scala



Perché la regola empirica

$$a_z \le Z \le b_z \quad \Leftrightarrow \quad \mu_X + \sigma_X a_z \le X \le \mu_X + \sigma_X b_z$$

si possono calcolare le seguenti probabilità:

$\Phi(1) = 0.8413$

$\Rightarrow P(-1 \le Z \le +1) = 0.6826 = P(\mu_X - 1\sigma_X \le X \le \mu_X + 1\sigma_X)$

$\Phi(2) = 0.9772$

$\Rightarrow P(-2 \le Z \le +2) = 0.9544 = P(\mu_X - 2\sigma_X \le X \le \mu_X + 2\sigma_X)$

$\Phi(3) = 0.9987$

$\Rightarrow P(-3 \le Z \le +3) = 0.9974 = P(\mu_X - 3\sigma_X \le X \le \mu_X + 3\sigma_X)$

Ecco spiegata la ***regola empirica***: molti fenomeni sono ben approssimati dalla Normale e quindi la proporzione di osservazioni in un intervallo del tipo $\mu + k\sigma$ è ben approssimata dalla corrispondente probabilità per la Normale

Valori anomali

- In una distribuzione Normale, un valore viene considerato anomalo se è fuori dall'intervallo $\mu_X \pm k\sigma_X$, dove di solito si prende k=2 → le distanze vengono misurate a partire da μ_X e sono in unità di σ_X → un valore non è anomalo in senso assoluto, ma solo relativamente ad una certa distribuzione.
- Esempio: X = lunghezza in mm di un pezzo prodotto, la sua distribuzione è Normale con μ_X = 80 → un pezzo di 85 mm è anomalo se σ_X =1, ma non è anomalo se σ_X =3.
- Attenzione: il criterio dell'intervallo $\mu_X \pm k\sigma_X$ per giudicare l'anomalia non ha senso se la distribuzione è molto diversa dalla Normale (in particolare, se è discreta con poche modalità)

Valutazione dell'ipotesi di normalità /1

- La distribuzione Normale permette di sfruttare una serie di utili proprietà
- Nella maggior parte dei casi, quando la variabile in esame è **continua** la Normale è un modello adeguato, cioè descrive in modo sufficientemente accurato la "vera" distribuzione di probabilità
- Tuttavia vi sono casi in cui la Normale è un modello del tutto inadeguato e quindi usare la Normale porta a risultati inattendibili
- Valutare l'ipotesi di normalità significa confrontare la **distribuzione osservata** (= la distribuzione dei dati da analizzare) con la distribuzione Normale

Valutazione dell'ipotesi di normalità /2

- Alcuni modi per confrontare la distribuzione osservata con la Normale sono:
  - Costruzione di **grafici** per analizzare la forma della distribuzione (boxplot, istogramma)
  - Calcolo delle **misure di sintesi** e confronto fra le caratteristiche dei dati e le proprietà teoriche della distribuzione Normale (la verifica principale consiste nel calcolare media, mediana e moda dei dati e valutare se sono approssimativamente uguali)
  - Verifica della **regola empirica**, calcolando la proporzione di osservazioni che si discostano dalla media per più di 1 volta, 2 volte, 3 volte la deviazione std e confrontando tali proporzioni con le corrispondenti probabilità normali, cioè 68%, 95%, 99%

Valutazione dell'ipotesi di normalità /3

La distribuzione Normale può essere inadeguata per vari motivi. Due motivi frequenti sono:

- **Asimmetria**
  - I dati possono avere una natura fortemente asimmetrica: in tal caso la distribuzione osservata ha media e mediana molto diverse
- **Code pesanti**
  - I dati possono presentare valori estremi (= lontani dalla media) molto più frequentemente di quanto previsto dalla Normale: in tal caso la proporzione di valori al di fuori degli intervalli del tipo $\mu + k\sigma$ è sostanzialmente più elevata delle corrispondenti probabilità normali

Valutazione dell'ipotesi di normalità /4

Due avvertenze finali:

- La distribuzione Normale ha come supporto l'intero asse dei numeri reali e quindi assegna probabilità non nulle anche a intervalli di valori negativi, es. [-3.2, 0]
  - Molti fenomeni analizzati ammettono solo valori positivi, es. tempo, lunghezza, costo. In tali situazioni la Normale può essere adeguata se l'intervallo $\mu + 3\sigma$ (che dovrebbe contenere quasi tutte le osservazioni) è tutto su valori positivi, es. [2.3, 8.2]
- In generale, quando le osservazioni a disposizione sono poche (diciamo meno di 20) è molto difficile stabile se una certa distribuzione di probabilità è adeguata o meno perché eventuali forti discrepanze tra ciò che si osserva e ciò che prescrive il modello potrebbero essere semplicemente frutto del caso

Approssimare la Binomiale con la Normale /1

- Ricorda la distribuzione binomiale:
  - *n* prove indipendenti
  - probabilità di successo in ogni prova = *p*
- Valore atteso e varianza:

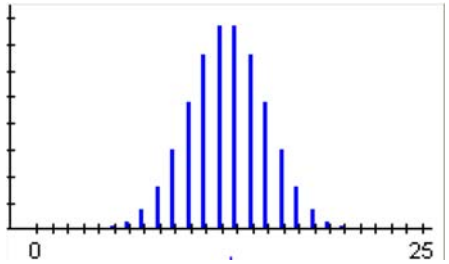
$$E(X) = \mu = np \qquad Var(X) = \sigma^2 = np(1-p)$$

- Quando *n* è grande il calcolo delle probabilità cumulate è complesso: es.

X ~B(*n*=50, *p*) → P(X≤25) = P(X=0) + P(X=1) + … + P(X=25)

Approssimare la Binomiale con la Normale /2

- Quando $np(1-p) > 9$ la Normale è una buona approssimazione per la binomiale



In tal caso la f. di ripartizione della v.a. $X \sim B(n, p)$ è molto simile a quella della v.a. $Y \sim N(np, np(1-p))$

$$P(X \le a) \sim P(Y \le a) = P\left(Z \le \frac{a - np}{\sqrt{np(1-p)}}\right)$$

Approssimare la Binomiale con la Normale /3

- 40% dei cittadini sono favorevoli all'operato del sindaco. Qual è la probabilità che, in un campione di n = 200, il numero di favorevoli sia compreso tra 76 e 80 (ovvero la percentuale di favorevoli sia compresa tra 38% e 40%)?
  - E(X) = $\mu = np$ = 200(0.40) = 80
  - Var(X) = $\sigma^2 = np(1-p)$ = 200(0.40)(1 – 0.40) = 48

    (notare: $np(1-p)$ = 48 > 9)

$$P(76 < X < 80) \sim P\left(\frac{76-80}{\sqrt{48}} \le Z \le \frac{80-80}{\sqrt{48}}\right)$$
$$= P(-0.58 < Z < 0)$$
$$= F(0) - F(-0.58)$$
$$= 0.5000 - 0.2810 = 0.2190$$

Distribuzione Chi-quadrato /1

Famiglia parametrica di v.a. continue:

$$\left\{X \sim \chi^2_r : \quad r = 1, 2, \ldots\right\}$$

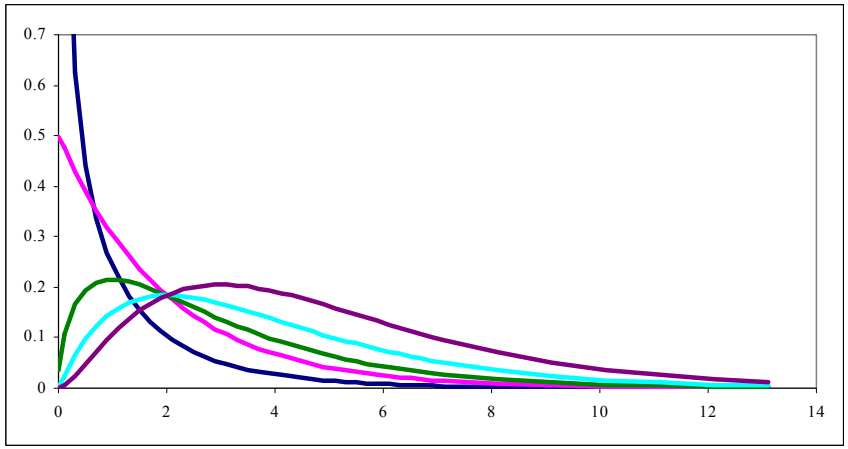
Parametro detto *gradi di libertà (gdl)*

$$f(x) = a x^{\left(\frac{r}{2}-1\right)} e^{-\frac{x}{2}} \qquad x \ge 0$$

(la costante a dipende da r ma non da x)

$$E(X) = r \qquad Var(X) = 2r$$

Distribuzione Chi-quadrato /2



Funzione di densità della v.c. Chi-quadrato con r gradi di libertà

(Blu: r = 1; Rosa: r = 2; Verde: r = 3; Celeste: r = 4; Viola: r = 5)

Distribuzione Chi-quadrato /3

- La Chi-quadrato con 1 gdl si genera elevando al quadrato una v.a. Normale standard:

$$Z \sim N(0,1) \quad \Rightarrow \quad Z^2 \sim \chi^2_1$$

- La Chi-quadrato si riproduce per somma:

$$X_1 \sim \chi^2_{r_1} \qquad X_2 \sim \chi^2_{r_2} \qquad X_1 \text{ e } X_2 \text{ indip.}$$
$$\text{posto } Y = X_1 + X_2 \quad \Rightarrow \quad Y \sim \chi^2_{r_1 + r_2}$$

Distribuzione Chi-quadrato /4

- La funzione di ripartizione non esiste in forma analitica → approssimazione numerica della Tavola C.3
  - Righe: gdl (r = 1,2,…,40,45,50,55,…,100)
  - Colonne: probabilità a destra (p da 0.995 a 0.001)
  - Valori in tabella: quantili superiori

- La Chi-quadrato tende alla Normale per gdl→∞

$$\text{per } r \text{ grande} \quad X \sim \chi^2_r \quad \text{distribuito approx.} \quad N(r, 2r)$$

- I quantili della Chi-quadrato con gdl>100 non si trovano in tavola ma si calcolano con l'approssimazione alla Normale

Distribuzione congiunta di *k* variabili aleatorie (tutte discrete o tutte continue)

Funzione di ripartizione congiunta

- Siano X_1, X_2, . . .X_k variabili aleatorie (discrete o continue)
- La loro **funzione di ripartizione congiunta**,

$$F(x_1, x_2, \ldots, x_k)$$

definisce la probabilità che, simultaneamente, X_1 sia minore di x_1, X_2 sia minore di x_2, …; cioè

$$F(x_1, x_2, \ldots, x_k) = P(X_1 \le x_1 \cap X_2 \le x_2 \cap \cdots X_k \le x_k)$$

Indipendenza

- Le funzioni di ripartizione

$$F(x_1), F(x_2), \ldots, F(x_k)$$

delle singole variabili aleatorie sono chiamate **funzioni di ripartizione marginali**

- Le variabili aleatorie sono **indipendenti** se e solo se

$$F(x_1, x_2, \ldots, x_k) = F(x_1) \times F(x_2) \times \cdots \times F(x_k)$$

Covarianza

- Siano X e Y variabili aleatorie (discrete o continue), con rispettive medie μ_x e μ_y
- Il valore atteso di $(X - \mu_x)(Y - \mu_y)$ viene chiamato **covarianza** tra X e Y

$$Cov(X,Y) = E[(X - \mu_x)(Y - \mu_y)]$$

- Espressione alternativa:

$$Cov(X,Y) = E(XY) - \mu_x \mu_y$$

- Se le variabili X e Y sono indipendenti, allora la covarianza fra loro è 0. In generale, il viceversa non è vero.

Correlazione

- Siano X e Y variabili aleatorie (discrete o continue)
- La **correlazione** tra X e Y è

$$\rho = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Somma di Variabili Aleatorie /1

Siano date k variabili aleatorie $X_1, X_2, \ldots X_k$ (discrete o continue) con medie $\mu_1, \mu_2, \ldots \mu_k$

e varianze $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$.

Allora:

- La media della loro somma è la somma delle loro medie

$$E(X_1 + X_2 + \cdots + X_k) = \mu_1 + \mu_2 + \cdots + \mu_k$$

Somma di Variabili Aleatorie /2

Siano date k variabili aleatorie $X_1, X_2, \ldots X_k$ con medie $\mu_1, \mu_2, \ldots \mu_k$ e varianze $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$. Allora:

- Se la covarianza fra ogni coppia di queste variabili aleatorie è 0, allora la varianza della loro somma è la somma delle loro varianze

$$Var(X_1 + X_2 + \cdots + X_k) = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_k^2$$

- Se le covarianze fra le coppie di variabili non sono 0, la varianza della loro somma è

$$Var(X_1 + X_2 + \cdots + X_k) = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_k^2 + 2\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} Cov(X_i, X_j)$$

Differenza tra due variabili aleatorie

Per due variabili aleatorie X e Y (discrete o continue)

- La media della loro differenza è la differenza fra le loro medie; cioè $E(X-Y)=\mu_X-\mu_Y$
- Se la covarianza tra X e Y è 0, allora la varianza della loro differenza è $Var(X-Y)=\sigma_X^2+\sigma_Y^2$
- Se la covarianza tra X e Y non è 0, allora la varianza della loro differenza è $Var(X-Y)=\sigma_X^2+\sigma_Y^2-2Cov(X,Y)$

Combinazioni lineari di Variabili Aleatorie

- Una combinazione lineare di due variabili aleatorie, X e Y, (dove a e b sono constanti) è $W=aX+bY$
- La media di W è $\mu_W=E[W]=E[aX+bY]=a\mu_X+b\mu_Y$
- La varianza di W è $\sigma_W^2=a^2\sigma_X^2+b^2\sigma_Y^2+2abCov(X,Y)$
- Se entrambe X e Y sono distribuite normalmente allora anche la combinazione lineare, W, è distribuita normalmente

Esempio /1

- Due mansioni devono essere eseguite dallo stesso lavoratore.
  - X = minuti per completare mansione 1; $\mu_X = 20$, $\sigma_X = 5$
  - Y = minuti per completare mansione 2; $\mu_Y = 30$, $\sigma_Y = 8$
  - X e Y sono distribuite normalmente e sono indipendenti
- Quali sono la media e la deviazione std del tempo necessario per completare entrambe le mansioni? Qual è la distribuzione?

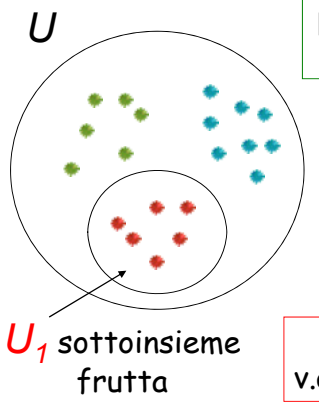
Esempio /2

- X = minuti per completare mansione 1; $\mu_X = 20$, $\sigma_X = 5$
- Y = minuti per completare mansione 2; $\mu_Y = 30$, $\sigma_Y = 8$

- Calcolare media e deviazione std del tempo $W=X+Y$ necessario per completare entrambe le mansioni $\mu_W=\mu_X+\mu_Y=20+30=50$
- Siccome X e Y sono indipendenti, Cov(X,Y) = 0, perciò $\sigma_W^2=\sigma_X^2+\sigma_Y^2+2Cov(X,Y)=(5)^2+(8)^2=89$
- La deviazione std è $\sigma_W=\sqrt{89}=9.434$
- La distribuzione di W è $W\sim N(50,89)$

Misture

Sottopopolazioni



U_1 , U_2 , U_3 partizione della popolazione

$X|U_1$ v.a. condizionata

alimento	*%proteine*
mele	0.2
uva	0.5
limoni	0.6
arance	0.7
pesche	0.8
banane	1.2
pomodori	1
carote	1.1
zucchine	1.3
lattuga	1.8
patate	2.1
spinaci	3.4
pizza	4
biscotti	6.6
riso	7
pane	8.1
crackers	9.4
pasta	10.8
fette bis	11.3
grissini	12.3

Distribuzione marginale come mistura

- Data una popolazione U con partizione $\{U_m : m=1,2,\ldots,M\}$
- Data una v.a. X (discreta o continua) con una distribuzione marginale e M distribuzioni condizionate $X|U_m$
- Sia π_m la probabilità che un'unità appartenga a U_m

Distribuzioni condizionate

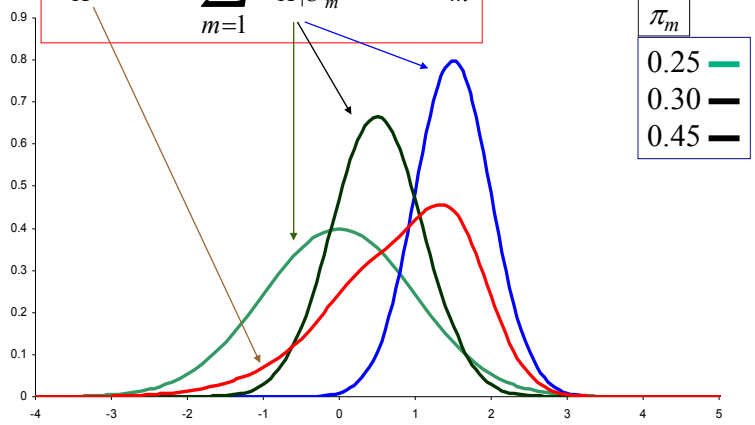
$$f_X(x) = \sum_{m=1}^{M} f_{X|U_m}(x) \cdot \pi_m$$

Distribuzione marginale (mistura)

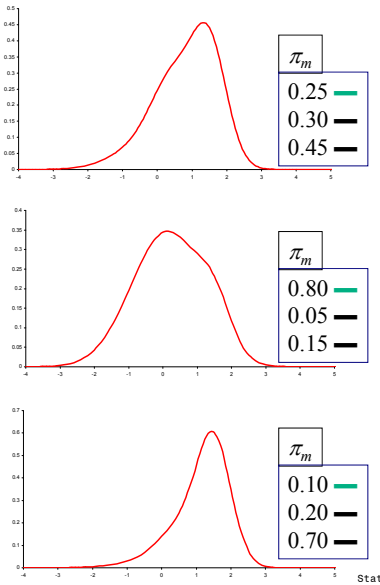
Estensione della formula delle probabilità totali $P(A) = \sum_m P(A|B_m)P(B_m)$

Esempio: mistura di tre densità

$$f_X(x) = \sum_{m=1}^{M} f_{X|U_m}(x) \cdot \pi_m$$

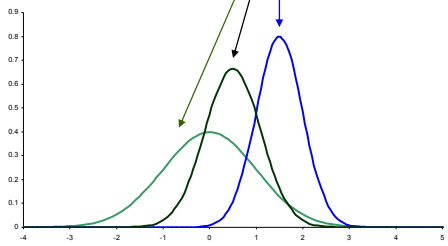


Esempio: mistura di tre densità



Al variare dei pesi cambia la mistura

$$f_X(x) = \sum_{m=1}^{M} f_{X|U_m}(x) \cdot \pi_m$$



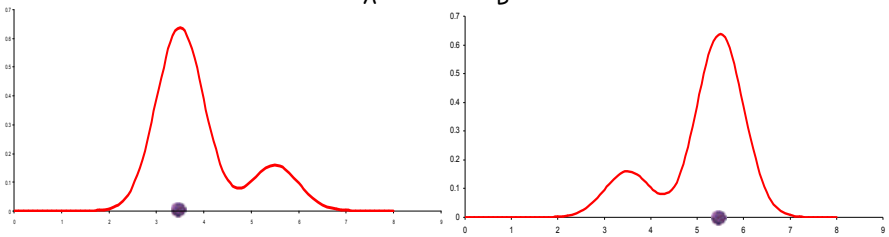
Esempio applicativo

X: tempo necessario per trovare un'occupazione in una popolazione di allievi, maschi e femmine, di due corsi di F.P.

$$f_X(x) = \sum_{m=1}^{M} f_{X|U_m}(x) \cdot \pi_m$$

Qual è il corso più efficace?

$M_A = 3.5 < M_B = 5.5$

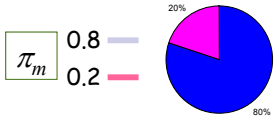


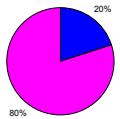
Corso A Corso B

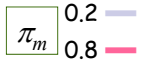
Esempio applicativo: il problema sta nella diversa composizione

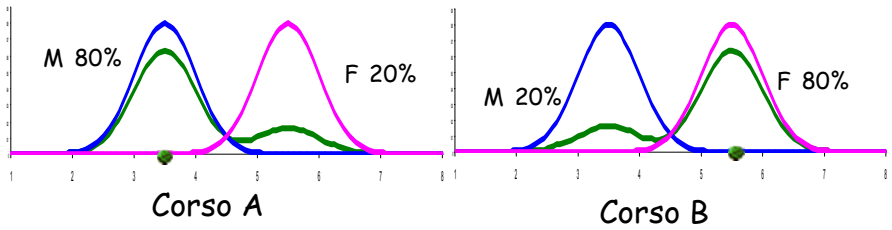
I due corsi hanno una diversa composizione per sesso degli allievi

$$f_X(x) = \sum_{m=1}^{M} f_{X|U_m}(x) \cdot \pi_m$$



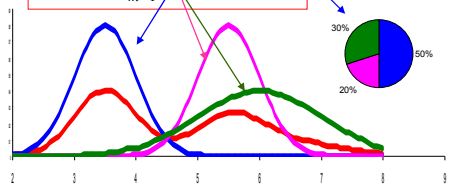


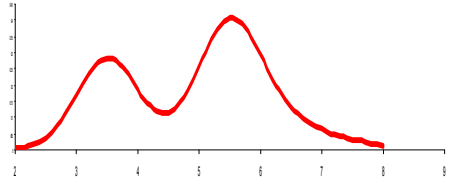




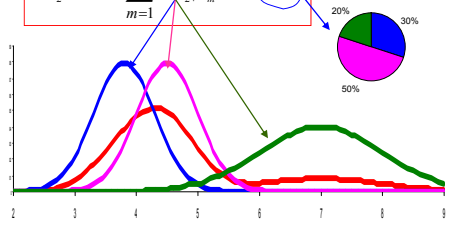
Confronti tra misture "standardizzate"

$$f_{X_1}(x) = \sum_{m=1}^{M} f_{X_1|U_m}(x) \cdot \pi_{1m}$$

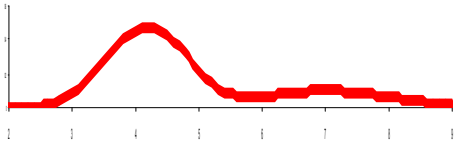




$$f_{X_2}(x) = \sum_{m=1}^{M} f_{X_2|U_m}(x) \cdot \pi_{2m}$$



$\pi_m = (\pi_{1m} + \pi_{2m})/2$



Distribuzione della mistura

- Anche se tutte le distribuzioni condizionate appartengono alla stessa famiglia parametrica (es. Normale), in generale la distribuzione mistura non appartiene alla famiglia
- Se una distribuzione non è Normale nell'intera popolazione, potrebbe essere Normale nelle sottopopolazioni (cioè la non-Normalità potrebbe essere semplicemente una conseguenza della mistura)

Simulare una distribuzione mistura

- Simulare 10 valori da Y = mistura composta da
  - $X_1 \sim N(0,1)$ con peso 0.6
  - $X_2 \sim N(3,1)$ con peso 0.4

simulare 6 valori da una N(0,1)

simulare 4 valori da una N(3,1)

-1.091	3.838
-0.687	3.914
0.695	3.428
0.912	2.978
-2.040	
0.685	

10 valori di Y

Simulare una combinazione lineare

- E' importante capire la differenza tra
  - simulare la distribuzione mistura e
  - simulare la distribuzione combinazione lineare
- Ad es. simulare 10 valori da

$$W = 0.6X_1 + 0.4X_2 \quad \text{con}$$
$$X_1 \sim N(0,1) \quad X_2 \sim N(3,1)$$

simulare 10 valori da una N(0,1)	-1.134	1.904	1.645	-1.576	-1.283	2.064	-1.606	-1.225	1.969	1.106
simulare 10 valori da una N(3,1)	4.714	3.661	2.788	5.049	2.133	3.627	1.775	2.485	3.197	2.696
calcolare media pesata $0.6x_1+0.4x_2$	1.205	2.607	2.102	1.074	0.084	2.689	-0.254	0.259	2.461	1.742