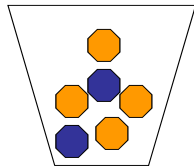


Elementi di calcolo delle probabilità e loro applicazione in medicina

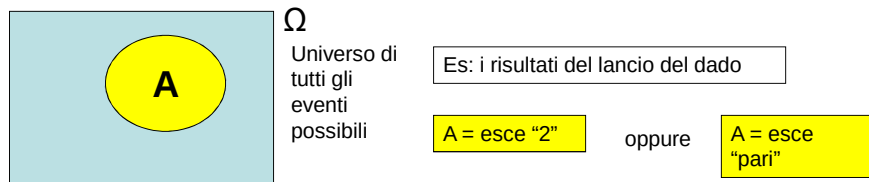


- Gli eventi e la Probabilità: le regole basilari
 - Il concetto di dipendenza probabilistica
- La regola di Bayes e sue implicazioni
 - I test diagnostici
 - Curve ROC
- Le distribuzioni di probabilità
 - Binomiale, Poisson, Normale

Eventi e Probabilità

- Le nozioni di “evento” e “probabilità” sono intuitive e usate in vari contesti comuni, nei giochi e nelle scommesse, ma anche nella vita quotidiana.
- In ambito scientifico vi sono diverse impostazioni “filosofiche” per definire eventi e probabilità. Gli elementi comuni e intuitivi sono:
 - **Un evento** è un oggetto (fenomeno, circostanza o caratteristica) che può verificarsi o no. Dunque **può essere considerato come il risultato di una prova determinata dal caso, assimilabile all'estrazione di una pallina da un'urna.**
 - La definizione non è limitata a oggetti che si determinano nel futuro, è sufficiente che vi sia incertezza sul loro essersi verificati nel passato ovvero sul loro essere veri nel presente.
 - **La probabilità di un evento rappresenta l'aspettativa che l'evento si verifichi, ed è espressa su una scala da 0% a 100% (o da 0 a 1).**

Eventi - Insiemi - e Probabilità



La probabilità dell'evento A è un numero:

$$0 < p(A) < 1$$

Ω è l'evento certo: $p(\Omega)=1$

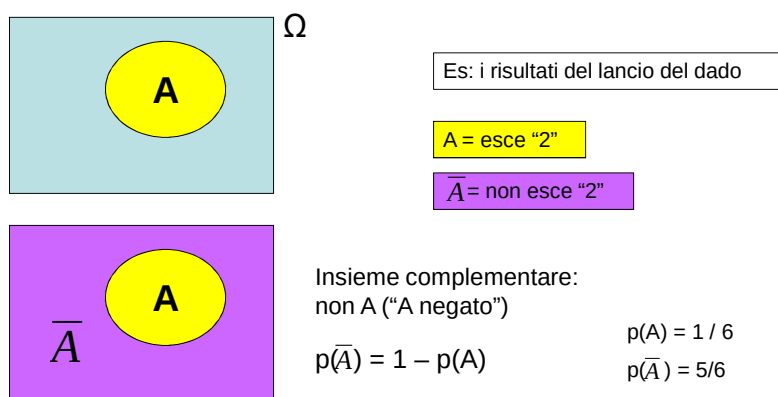
Nei casi più semplici, dove la prova ha un numero finito di possibili esiti, e tutti sono ugualmente probabili, $p(A)$ = numero casi favorevoli / numero casi possibili.

Rispettivamente nei due esempi:

$$p(A) = 1 / 6$$

$$p(A) = 3 / 6 = 1 / 2$$

Evento complementare



L'evento complementare di A è semplicemente l'evento che comprende tutti i casi in cui A non si verifica

$$p(\bar{A}) = 1 - p(A)$$

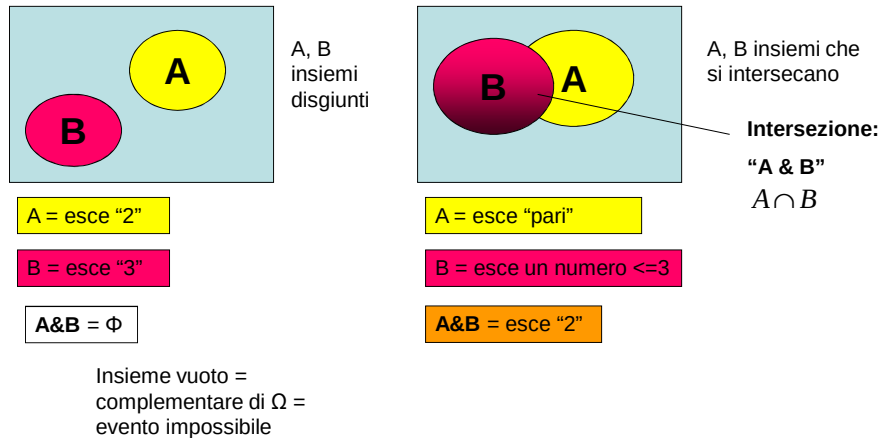
$$p(A) = 1 / 6$$

$$p(\bar{A}) = 5/6$$

$$p(\text{infezione}) = 0.7$$

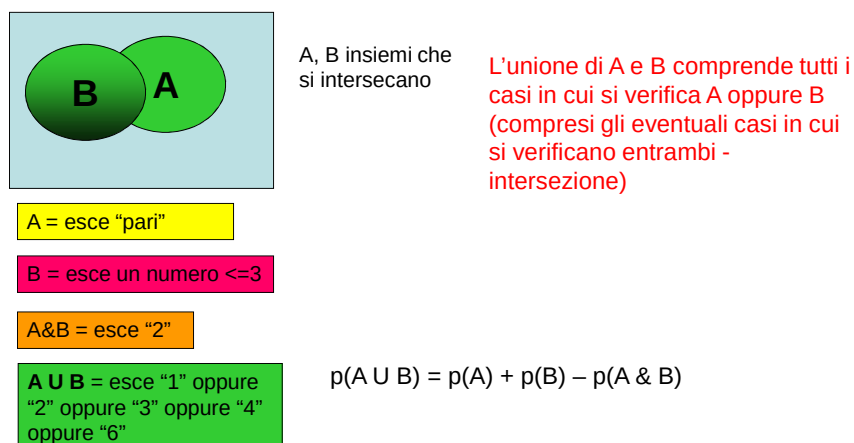
$$p(\text{no infezione}) = 0.3$$

Intersezione $\rightarrow$ "e" (And)

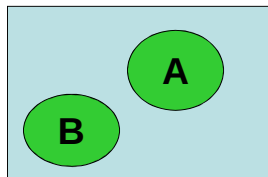


L'intersezione di due eventi A e B comprende tutti i casi in cui si verificano sia A che B: può essere vuota, ossia "impossibile"

Unione $\rightarrow$ "oppure" (Or) (1) "o anche"



Unione → “oppure” (Or) (2) “o anche”



A, B
insiemi
disgiunti

Es: i risultati del lancio del dado

“pari” = “2 o 4 o 6”

$$p(\text{pari}) = p(2) + p(4) + p(6) \\ = 1/6 + 1/6 + 1/6 = 3/6 = 1/2$$

A = esce “2”

B = esce “3”

A ∪ B = esce
“2” oppure “3”

$$p(A \cup B) = p(A) + p(B)$$

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

Nota: il caso generale comprende quello particolare con eventi disgiunti.

La formula è in realtà la stessa, poiché se gli eventi sono disgiunti $p(A \cap B) = 0$

Probabilità condizionata

Spesso, la probabilità di un evento cambia a seconda dell'informazione che abbiamo

Es: $p(\text{esce } 2) = 1/6$

Ma: se so che “esce pari” → la prob. che esca 2 sale a $1/3$

Introduciamo quindi il concetto di probabilità condizionata:

$p(A|B)$ = prob. di A condizionata a B

- “dato B”
- “se si verifica B”
- “sapendo che si verifica B”
- “restringendosi ai casi in cui si verifica B”

Es: Nella popolazione generale, la prob. di decesso per infarto è 5%; fra gli obesi, è 10%.

$$p(\text{decesso per infarto}) = 0.05$$

$$p(\text{decesso per infarto} | \text{obeso}) = 0.10$$

Eventi dipendenti e indipendenti

Quando la probabilità di un evento NON cambia in presenza di condizionamento ad un altro evento, essi si dicono **indipendenti**

$$p(A|B) = p(A)$$

→ Il condizionamento non agisce!

→ L'aspettativa di A non si modifica sapendo che si verifica B

Nota: non è una indipendenza materiale, logica, causale delle prove. E' una indipendenza "della probabilità".

Analogamente, A e B si dicono **dipendenti** se:

$$p(A|B) \neq p(A)$$

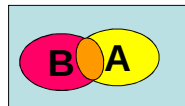
L'evento B non modifica l'evento A in modo materiale, concreto; quello che si modifica è la probabilità

Formule per prob. condizionata e intersezioni

$$p(A | B) = \frac{p(A \cap B)}{p(B)}$$

Prob. di A condizionata a B

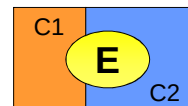
$$\begin{aligned} p(A \cap B) &= p(A | B) \cdot p(B) \\ &= p(B | A) \cdot p(A) \end{aligned}$$



$$p(A \cap B) = p(A) \cdot p(B)$$

Caso particolare per eventi A e B indipendenti

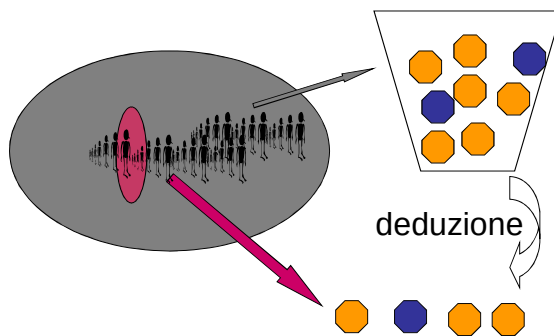
$$p(C1 | E) = \frac{p(E | C1) \cdot p(C1)}{p(E | C1) \cdot p(C1) + p(E | C2) \cdot p(C2)}$$



Formula di Bayes: per calcolare la probabilità a posteriori di C1 dato E ("problema della diagnosi")

Utilizzare le probabilità (in medicina)

Ci interessa un fenomeno "in generale", in una Popolazione obiettivo. Usiamo un modello matematico per rappresentare il suo "andamento teorico". **Se conosciamo i parametri che descrivono come è composta l'urna, possiamo elaborare ulteriormente le nostre informazioni.**



Es: X Risposta al trattamento

Supponiamo di conoscere la composizione dell'urna = sappiamo che nella popolazione / in generale, il trattamento è efficace nel 25% dei casi:

$P(\text{Risposta})=0.25$

Allora in un campione di 4 pazienti, mi aspetto di osservare una risposta.

Dal quesito al problema di probabilità: esempio 1

In una certa popolazione:

- Il 20% dei parti va incontro a complicazioni;
- La metà dei parti con complicazioni richiede un taglio cesareo;
- In generale, il 30% dei parti è cesareo.

Qual è la prob. di avere un parto cesareo o con complicazioni?

Eventi: A = complicazioni; B = cesareo

Informazioni: $p(A)=0.2$ $p(B) = 0.3$ $P(B|A) = 0.5$

Quesito: $P(A \cup B)$

$P(A \cup B) = p(A) + p(B) - p(A \cap B)$

$p(A \cap B) = p(B|A) \cdot p(A) = 0.5 \cdot 0.2 = 0.1$ (prob. cesareo con complicazioni)

$P(A \cup B) = 0.2 + 0.3 - 0.1 = 0.4$

Dal quesito al problema di probabilità: esempio 2

Una certa tecnica chirurgica d'avanguardia viene utilizzata solo in due ospedali. Nell'ospedale A, l'intervento riesce nel 45% dei casi, nell'ospedale B nel 30%. Esprimere la differenza comparando la prob di riuscita in B rispetto a quella di A. Calcolare poi la probabilità di riuscita per un paziente che sa di avere il 70% di probabilità di subire l'intervento in A e il 30% di farlo in B.

Eventi: Notazione: A, B → ospedale. Riuscita: R

Informazioni: $p(R|A)=0.45$ $p(R|B) = 0.3$ $p(A)=0.7$ $p(B)=0.3$

Quesito 1: confronto: $P(R|B)$ confrontata con $P(R|A)$:

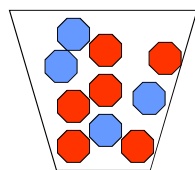
$RR=0.3/0.45=0.67$ [risk ratio]

Quesito 2: $p(R)$

$= p(R \& (A \cup B)) = p(R \& A) + p(R \& B) =$

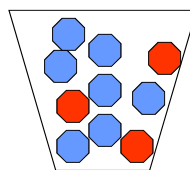
$= p(R|A) \cdot p(A) + p(R|B) \cdot p(B) = 0.45 \cdot 0.7 + 0.3 \cdot 0.3 = 0.315 + 0.09 = 0.405$

Confronto di probabilità: il Risk Ratio



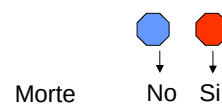
Popolazione (urna) dei
soggetti ESPOSTI ad un
fattore di interesse, ad es.
sottoposti a tecnica chirurgica
A

$\Pr(\text{Morte} | A) = 0.60$



Popolazione (urna)
dei soggetti esposti
NON ESPOSTI, ad
es. sottoposti a
tecnica chirurgica B

$\Pr(\text{Morte} | B) = 0.30$



Molto spesso in medicina ed epidemiologia si vogliono confrontare le probabilità che si verifichi un evento fra due gruppi. Quasi mai il confronto si fa calcolando la differenza fra le due probabilità.

Confronto di probabilità: il Risk Ratio

$$\Pr(\text{Morte} | A) = 0.60$$

$$\Pr(\text{Morte} | B) = 0.30$$

RISK RATIO $RR = \frac{\pi_E}{\pi_{NE}} = \frac{0.60}{0.30} = 2$

= 1 : non c'è relazione tra Esposizione e verificarsi dell'evento

> 1 : Esposizione fattore di rischio per l'evento

Tra 0 e 1 : Esposizione fattore protettivo per l'evento

= 1.6 : Esposizione aumenta il rischio di evento del 60%

= 2: aumento del 100%

= 0.7: Esposizione diminuisce il rischio di evento del 30% (lo riduce al 70% del rischio dei Non Esposti)

Formula di Bayes: contesto ≈ diagnosi

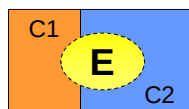
Un problema tipico in medicina è formulare una diagnosi sulla base di una osservazione:

- di sintomi o altre possibili conseguenze di una condizione («malattia»)
- del risultato di una procedura diagnostica

Il contesto più generale è: c'è una EVIDENZA osservata, e ci sono 2 (o più) possibili CAUSE o ipotesi che possono spiegare quell'evidenza.

Per la causa C1 vogliamo conoscere la sua probabilità DATO CHE l'evidenza E è stata osservata:

$$P(C1 | E)$$



Conosciamo:

$P(C1)$: la prob della causa C1 prima di / senza aver osservato alcuna evidenza

$P(E|C1)$: la prob dell'evento E se causa=C1

$P(E|C2)$: la prob dell'evento E se causa= non C1

Aggiornamento delle probabilità a priori a seguito di evidenza

CON CHE PROBABILITA' E' INFARTO??

Un paziente si presenta dal medico per un dolore al braccio, temendo di avere un infarto in corso.

Il medico fa il seguente ragionamento:

- Se c'è un infarto, la probabilità di avere questo tipo di dolore è del 80%;
- D'altra parte, un'inflammation provocherebbe questo dolore nel 30% dei casi;

Il medico prosegue il ragionamento:

- Quest'uomo è giovane, magro, fa attività fisica ... la prob. di infarto in questi casi è bassa, 5%
- Invece, con lo sport che pratica, la prob. di infiammazione è 95%

E = dolore

C1 = infarto

C2 = infiammazione

$$p(E|C1) = 0.8$$

$$p(E|C2) = 0.3$$

VEROSIMIGLIANZE
delle ipotesi C1 e C2
dato E

$$p(C1) = 0.05$$

$$p(C2) = 0.95$$

Prob. a priori delle
ipotesi C1 e C2

La formula di Bayes

CON CHE PROBABILITA' E' INFARTO??

Per fare una diagnosi, il medico deve valutare tutti questi elementi, e valutare la probabilità che stia agendo la causa infarto avendo l'evidenza di un suo sintomo.

Quesito: $P(C1|E)$. Formula di Bayes:

$$= \frac{p(E|C1) \cdot p(C1)}{p(E|C1) \cdot p(C1) + p(E|C2) \cdot p(C2)}$$

$$= \frac{0.8 \cdot 0.05}{0.8 \cdot 0.05 + 0.3 \cdot 0.95} = 0.12$$

$p(C1|E)$: Prob. a posteriori
dell'ipotesi C1

E = dolore

C1 = infarto

C2 = infiammazione

$$p(E|C1) = 0.8$$

$$p(E|C2) = 0.3$$

VEROSIMIGLIANZE
delle ipotesi C1 e C2
dato E

$$p(C1) = 0.05$$

$$p(C2) = 0.95$$

Prob. a priori delle
ipotesi C1 e C2

I test diagnostici

- Il test diagnostico è uno strumento per la diagnosi della presenza di una certa condizione, ad es. un'anomalia genetica, o più semplicemente una malattia, utilizzabile in clinica e negli screening. Test "positivo" indica presenza di quella caratteristica (es malattia).
- Il test diagnostico solitamente non dà risultati sicuri: non tutti i soggetti malati vengono individuati, e viceversa alcuni soggetti sani vengono erroneamente classificati come malati.
 - Si hanno cioè, rispettivamente, i cosiddetti **FALSI NEGATIVI** e **FALSI POSITIVI**
 - Questi test trovano la loro utilità quando effettuare una diagnosi più accurata sia troppo costoso – invasivo – pericoloso – etc

Le caratteristiche di un test diagnostico vengono sintetizzate da due parametri:

SENSITIVITA': la capacità di individuare i soggetti malati

SPECIFICITA': la capacità di riconoscere i soggetti sani

Capire i test diagnostici per la pratica clinica

Esempio: si stima che il 10% delle persone appartenenti ad una certa categoria di rischio sia affetta dal virus dell'HIV (per semplicità, diciamo "malata").

Supponiamo di dover sottoporre a test diagnostico un individuo di quella categoria; il test utilizzato ha sensibilità = 90% e specificità = 80%.

- Le domande che si può porre l'operatore sono:
 - Per quanti soggetti malati mancheremo la diagnosi?
 - Quanti soggetti non malati sottoporremo inutilmente a ulteriori accertamenti?
 - Quanti errori diagnostici commetteremo in tutto?
- Le domande che il soggetto sottoposto al test può porre sono ad esempio:
 - Il test dà un risultato sicuro?
 - Se sono malato uscirà test positivo?
 - Se il test viene positivo, vuol dire che sono malato?

Le probabilità nel test diagnostico (1)

		Risultato del test diagnostico	
		Test +	Test –
Situazione (incognita) del soggetto	Malato	ok	Falso negativo
	Non Malato	Falso positivo	ok

Caratteristiche del test – sensibilità e specificità:

SENSITIVITA': $p(\text{Test} + | \text{Malato})$

SPECIFICITA': $p(\text{Test} - | \text{Non Malato})$

Le caratteristiche di un test diagnostico vengono sintetizzate da due parametri:

SENSITIVITA': la capacità di individuare i soggetti malati (fornendo risultato positivo)

SPECIFICITA': la capacità di riconoscere i soggetti sani (fornendo risultato negativo)

Le probabilità nel test diagnostico (2)

		Risultato del test diagnostico	
		Test +	Test –
Situazione (incognita) del soggetto	Malato	ok	Falso negativo
	Non Malato	Falso positivo	ok

Caratteristiche del test – sensibilità e specificità:

SENSITIVITA': $p(\text{Test} + | \text{Malato})$

SPECIFICITA': $p(\text{Test} - | \text{Non Malato})$

Errori:

Falso negativo $\rightarrow p(\text{Test} - | \text{Malato})$

$= 1 - \text{SENSITIVITA'}$

Falso positivo $\rightarrow p(\text{Test} + | \text{Non Malato})$

$= 1 - \text{SPECIFICITA'}$

Le probabilità nel test diagnostico (3)

		Risultato del test diagnostico	
		Test +	Test –
Situazione (incognita) del soggetto	Malato	ok	Falso negativo
	Non Malato	Falso positivo	ok

Se il test viene positivo, l'individuo è malato?

SENSITIVITA': $p(\text{Test} + | \text{Malato})$
 SPECIFICITA': $p(\text{Test} - | \text{Non Malato})$
 FN: $p(\text{Test} - | \text{Malato}) = 1 - \text{SENS}$
 FP: $p(\text{Test} + | \text{Non Malato}) = 1 - \text{SPEC}$

Valori predittivi del test:

$p(\text{Malato} | \text{Test} +)$

$p(\text{Sano} | \text{Test} -)$

Si tratta di prob. "a posteriori" → dobbiamo la conoscere la prob. "a priori", non condizionata, di avere la malattia

Dobbiamo avere il dato sulla PREVALENZA della malattia

$P(\text{Malato})$

Le probabilità nel test diagnostico (4)

		Risultato del test diagnostico	
		Test +	Test –
Situazione (incognita) del soggetto	Malato	ok	Falso negativo
	Non Malato	Falso positivo	ok

Se il test viene positivo, l'individuo è malato?

$p(M | T+)$

$$= \frac{p(T+ | M) \cdot p(M)}{p(T+ | M) \cdot p(M) + \underbrace{p(T+ | \text{non } M)}_{1 - p(T- | \text{non } M)} \cdot \underbrace{p(\text{non } M)}_{1 - p(M)}}$$

$$= \frac{\text{sens} \cdot \text{prev}}{\text{sens} \cdot \text{prev} + (1 - \text{spec}) \cdot (1 - \text{prev})}$$

SENSITIVITA': $p(\text{Test} + | \text{Malato})$
 SPECIFICITA': $p(\text{Test} - | \text{Non Malato})$
 FN: $p(\text{Test} - | \text{Malato}) = 1 - \text{SENS}$
 FP: $p(\text{Test} + | \text{Non Malato}) = 1 - \text{SPEC}$

Prevalenza = $P(\text{Malato})$

Esempio: test diagnostico

Si stima che una patologia colpisca 1 individuo su 50. L'accertamento della presenza di questa patologia è invasivo. Un test basato su un prelievo di sangue permette di identificare i soggetti affetti. Il test ha sensibilità = 70% e specificità = 90%. Si vuole calcolare la probabilità che un soggetto con Test positivo sia malato.

Eventi: TP= test positivo; M = malattia

Informazioni: $p(\text{TP}|\text{M})=0.7$ $p(\text{non TP}|\text{non M}) = 0.9$ $P(\text{M}) = 1/50=0.02$

Quesito: $P(\text{M} | \text{TP})$

Si applica la formula di Bayes:

$$\begin{aligned} &= \frac{p(\text{TP} | \text{M}) \cdot p(\text{M})}{p(\text{TP} | \text{M}) \cdot p(\text{M}) + p(\text{TP} | \text{non M}) \cdot p(\text{non M})} \\ &= \frac{0.7 \cdot 0.02}{0.7 \cdot 0.02 + 0.1 \cdot 0.98} = 0.125 \end{aligned}$$

Esercizio: la probabilità di errore complessiva in un test diagnostico

Calcolare la prob. complessiva di errore e il numero totale di errori diagnostici nell'applicazione a 40 individui, fra cui il 10% malati, di un test diagnostico con sensibilità = 70% e specificità = 90%.

Suggerimento: Errore cioè: Falso Negativo per i Malati, Falso Positivo per i Non-malati

$$\begin{aligned} \text{Pr(Errore)} &= \text{Pr(Errore \& malato)} + \text{Pr(Errore \& Non malato)} = \\ &= \text{Pr(Errore} | \text{malato)} P(\text{malato}) + \text{Pr(Error} | \text{Non malato)} \text{Pr(Non malato)} = \\ &= \text{Pr(Test -} | \text{malato)} P(\text{malato}) + \text{Pr(Test +} | \text{Non malato)} (1 - \text{Pr(malato)}) = \\ &= (1 - \text{sensibilità}) \text{Pr(malato)} + (1 - \text{specificità}) (1 - \text{Pr(malato)}) \end{aligned}$$

Applicazione:

$$\text{Pr(Errore)} = (1 - 0.7) \cdot 0.1 + (1 - 0.9) \cdot (1 - 0.1) = 0.093$$

$$\text{Numero atteso di Errori} = 0.093 \cdot 40 = 3.72$$

Stima di Sensitività e Specificità

Si vuole stimare la sensitività e la specificità di una nuova tecnica diagnostica per immagini, alternativa ad una con risultato certo, ma meno invasiva / costosa. Si prende quindi un campione di n soggetti che, sottoposti alla "vecchia" tecnica, vengono classificati in "malati" e "non malati"; li si sottopongono poi alla "nuova" diagnostica, ottenendo i seguenti risultati:

	Test +	Test -	
Malati	vp	fn	m
Non Malati	fp	vn	n-m
	vp+fp	fn+vn	n

n soggetti di cui m malati, gli altri $(n-m)$ non malati

Dei malati, vp hanno Test+ e fn hanno Test- (vp sono i "veri positivi", fn sono i "falsi negativi")

etc

$$\text{sensitività} \approx \frac{vp}{m} = \frac{vp}{vp + fn}$$

$$\text{specificità} \approx \frac{vn}{n-m} = \frac{vn}{vn + fp}$$

Attenzione! Il valore predittivo si può calcolare solo se conosciamo la **prevalenza** della malattia. **Solo se possiamo pensare di stimarla dal campione, ($=m/n$)** allora si ha:

$$p(M | T+) \approx \frac{vp}{vp + fp}$$

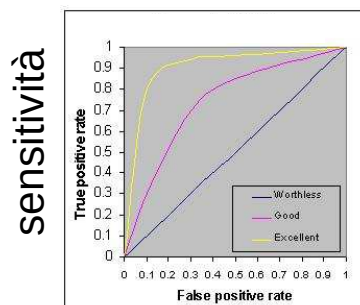
(stesso risultato con la formula di Bayes)

Curva ROC (cenni)

Receiver Operating Characteristic

Contesto: biomarker continuo X . Valori alti indicano «malattia». Si deve fissare un cut-point x tale che se $X > x \rightarrow$ diagnosi di malattia. Come scegliere x ?

Ad ogni x corrisponde una diversa coppia (sensitività, specificità) $\rightarrow$ una curva ROC



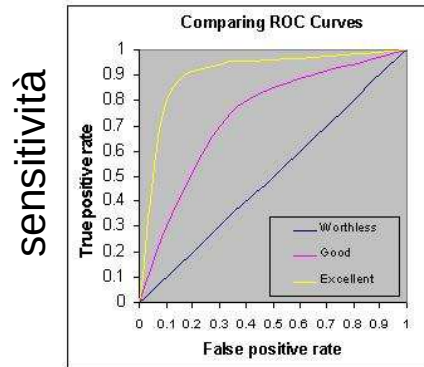
1-specificità

Confronto di curve ROC (cenni)

Receiver Operating Characteristic

Contesto: biomarker continuo X. Per ogni $x \rightarrow$ una curva ROC

Usando un altro biomarker Y, si ha un'altra curva ROC.



AUC = Area sotto la curva:
misura l'accuratezza

= 1 massima accuratezza

1-specificità

Gioco: lancio una moneta 3 volte, quante volte esce Testa?



3 lanci consecutivi di una moneta con
 $\text{prob}(\text{Testa}) = \text{prob}(\text{Croce}) = 50\%$

X = numero di "Testa"

Risultati
possibili:

C	C	C	C	C	T	C	T	T	T	T	T	T
				C	T	C	T	C	T			
				T	C	C	T	T	C			

$X = 0$

$X = 1$

$X = 2$

$X = 3$

$$p(X=0) = 1/8$$

$$p(X=1) = 3/8$$

$$p(X=2) = 3/8$$

$$p(X=3) = 1/8$$

Gioco 2: moneta non bilanciata



3 lanci consecutivi di una moneta con
 $\text{prob}(\text{Testa}) = 30\%$ $\text{prob}(\text{Croce}) = 70\%$

X = numero di "Testa"

Nota ad es: la
terna CCT
stavolta non ha la
stessa probabilità
della terna CTT!!

Risultati
possibili:

C	C	C	C	C	T	C	T	T	T	T	T	T
				C	T	C	T	C	T			
				T	C	C	T	T	C			

$X = 0$

$X = 1$

$X = 2$

$X = 3$

$$\begin{aligned} p(X=0) &= \\ p(C) \cdot p(C) \cdot p(C) &= 0.7^3 \\ &= 0.343 \end{aligned}$$

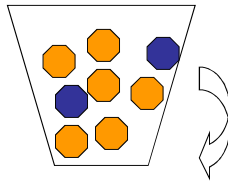
$$\begin{aligned} p(X=1) &= \\ 3 \cdot p(C) \cdot p(C) \cdot p(T) &= 3 \cdot 0.7^2 \cdot 0.3 \\ &= 0.441 \end{aligned}$$

$$\begin{aligned} p(X=2) &= \\ 3 \cdot p(C) \cdot p(T) \cdot p(T) &= 3 \cdot 0.7 \cdot 0.3^2 \\ &= 0.189 \end{aligned}$$

$$\begin{aligned} p(X=3) &= \\ p(T) \cdot p(T) \cdot p(T) &= 0.3^3 \\ &= 0.027 \end{aligned}$$

Variabili aleatorie e distribuzioni di probabilità

Associamo un numero al risultato di una prova. Tale numero non è determinato, ma varia a seconda del caso: è una variabile casuale, o aleatoria (v.a.) (*random variable*).



Es: X Recidiva dopo trattamento
= 0 (no) oppure 1 (sì, yes)
 $\Pr(\text{Recidiva})=25\%$

Es 2: Trattiamo 50 pazienti, quanti recidiveranno?

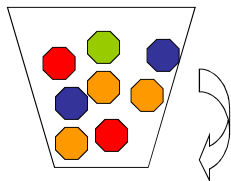
Il numero di recidivati X è una somma di variabili $Y=0,1$:

$X = 0, 1, 2, \dots, 49, 50$

	$X = 1$	$\Pr(X = 1) = 0.25$
	$X = 0$	$\Pr(X = 0) = 0.75$

Variabili aleatorie discrete

Nell'esempio della variabile Recidiva, la X è «dicotomica» (0/1). In altri casi, può essere di interesse un numero aleatorio X con più modalità, in generale k.

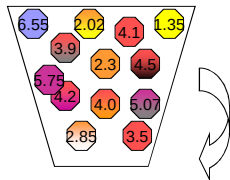


Es: X = numero di ricoveri necessari per paziente sottoposto a chemioterapia
= 0, 1, 2, 3

	$X = 0$	$\Pr(X=0)$
	$X = 1$	$\Pr(X=1)$
	$X = 2$	$\Pr(X=2)$
	$X = 3$	$\Pr(X=3)$

Variabili aleatorie continue

In altre situazioni, la variabile che interessa è «continua» cioè assume infiniti valori: tutti i valori compresi in un certo intervallo.



Es: X = conta dei globuli bianchi (WBC)
= qualunque valore tra (?) 4500 e 8500

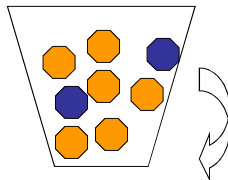
Es 2: X = età all'insorgenza della malattia
= qualunque valore >0

Infinite palline di infiniti colori!! Ovvero, palline marcate da un numero, infiniti diversi numeri

$\Pr(X=x)$ (?)

Variabile aleatoria → distribuzione di probabilità

La variabile aleatoria è descritta attraverso la sua distribuzione di probabilità: l'«elenco» di tutti i valori che X può assumere, con la probabilità di ciascun valore



Es: X Risposta al trattamento
= 0 (no) oppure 1 (sì, yes)

 $X = 1$

$\Pr(X = 1) = p = 0.25$

← p

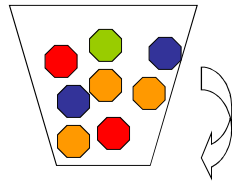
 $X = 0$

$\Pr(X = 0) = 1-p = 0.75$

Nel caso di una v.a. dicotomica, è sufficiente 1 parametro, $p = \Pr(X=1)$, per descrivere la distribuzione

Distribuzione di probabilità per v.a. discreta

In generale, nel caso di una v.a. discreta con $k \geq 2$ possibili valori, sono necessari $k-1$ parametri (tutte le $\Pr(X=j)$ meno una), per descrivere la distribuzione



	$X = 0$	$\Pr(X=0)=p_0$
	$X = 1$	$\Pr(X=1)=p_1$
	$X = 2$	$\Pr(X=2)=p_2$
	$X = 3$	$\Pr(X=3)= 1-p_0-p_1-p_2$

} $\leftarrow p_0, p_1, p_2$

Gioco → Esempio di distribuzione di probabilità discreta



Es: 3 lanci consecutivi di una moneta, X = numero di "Testa"
 X variabile aleatoria

Risultati possibili:

C	C	C	C	C	T	C	T	T	T	T	T
			C	T	C	T	C	T			
			T	C	C	T	T	C			

Distribuzione di probabilità di X

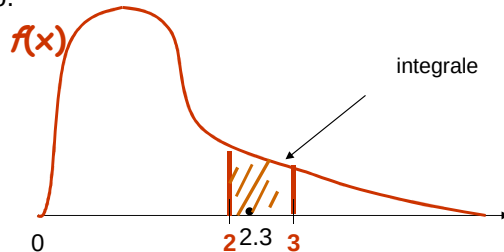
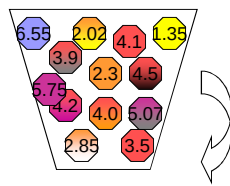
$X = 0$	$X = 1$	$X = 2$	$X = 3$
$p(X=0) = 1/8$	$p(X=1) = 3/8$	$p(X=2) = 3/8$	$p(X=3) = 1/8$

Nota: la somma delle probabilità di tutti i risultati possibili per X vale 1

Distribuzioni di probabilità continue

Quando la variabile aleatoria X è continua (es. età all'insorgenza della malattia; conta dei WBC dopo il trattamento) è come se nell'urna ci fossero palline di infiniti colori: come descriverne la composizione?

I valori possibili di X sono tutti i valori di un intervallo (o su un asse o semiasse cartesiano). Usiamo una **funzione di densità** tale che preso un intervallo di valori la corrispondente **area sotto la curva indica la probabilità** che X appartenga a quell'intervallo.



2.3

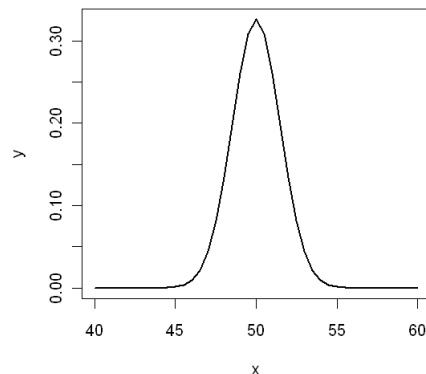
$$\text{Prob}(X=2.3) = 0$$

$$\text{Prob}(X \text{ in } (2,3)) = \int_2^3 f(x)dx = \text{Area}$$

La curva Normale (i)

Un modello per la variabilità biologica / per gli "errori"

La principale **curva di densità teorica** è la Normale (o Gaussiana), che descrive l'andamento di quei fenomeni misurabili come caratteri continui che dipendono "dal caso", come gli errori di misurazione. E' infatti simmetrica e ha una forma a campana.



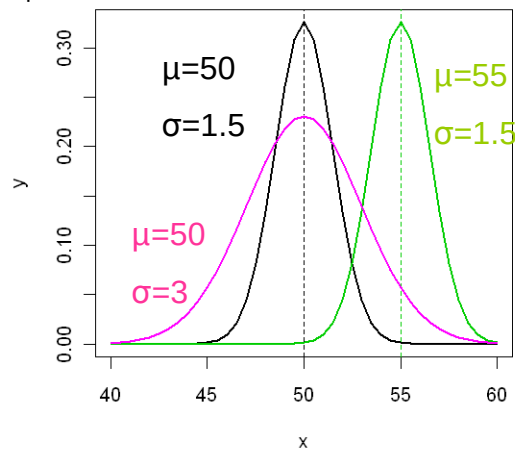
$$y = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Es: distribuzione dei risultati della misurazione ripetuta del peso di un paziente di 50 kg

La curva Normale (ii)

Un modello per la variabilità biologica / per gli "errori"

La formula che descrive la curva contiene 2 parametri μ e σ , che determinano rispettivamente dove si posiziona la curva rispetto all'asse x e quanto è ampia la campana

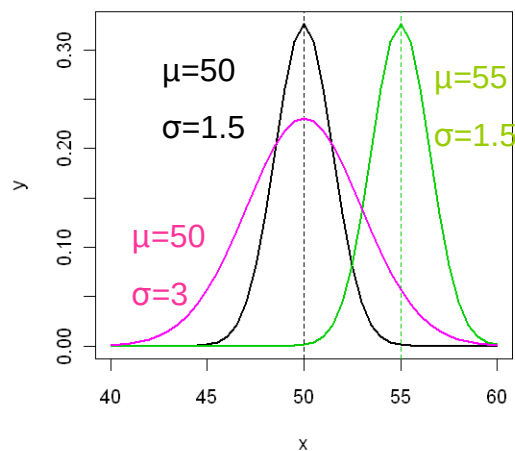


$$y = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

La curva Normale (iii)

I parametri μ e σ

- μ , che posiziona l'asse di simmetria, ed è interpretabile come **valore medio***
- σ , che determina l'ampiezza della campana, ossia la **dispersione** di X, e coincide con la **deviazione standard**



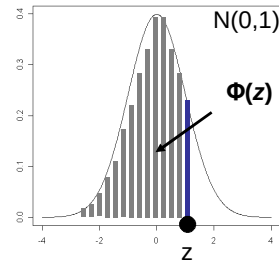
*valore atteso

Calcolare probabilità per la Normale

Per la Normale(0,1) (detta Standard) calcolatori o tavole forniscono i valori dell'area sotto la curva, fino a z : indichiamola con $\Phi(z)$, per ogni z .

Per qualsiasi altra Normale(μ, σ), per avere l'area fino a x , basta calcolare Φ sul valore trasformato:

$$z = \frac{x - \mu}{\sigma} \quad (\text{Standardizzazione})$$



Per calcolare aree con altra forma, basta comporla o scomporla in pezzi del tipo di $\Phi(z)$, ricordando che vale la simmetria attorno all'asse μ , per cui:

$$\text{Area}(Z < -z) = \text{Area}(Z > z)$$

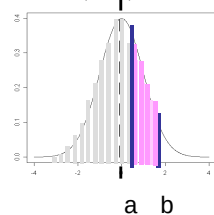
$$\Phi(-z) = 1 - \Phi(z)$$

$$\Phi(0) = 0.5 \quad \Phi(+\infty) = 1$$

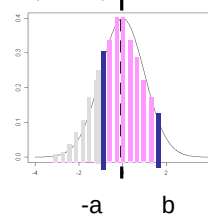
Calcolare probabilità per la Normale

Utilizzando tavole che forniscono $\Phi(z) = \text{Area}(-\infty, z)$ per $z > 0$:

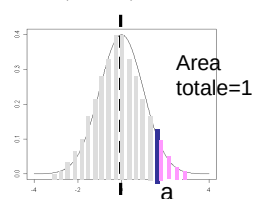
$$\text{Area}(a, b) = \Phi(b) - \Phi(a)$$



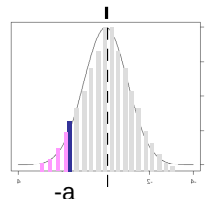
$$\text{Area}(-a, b) = \Phi(b) - (1 - \Phi(a))$$



$$\text{Area}(a, +\infty) = 1 - \Phi(a)$$



$$= \Phi(-a)$$



Esempio: Normale

In una popolazione di ragazze adolescenti, il Body Mass Index (BMI) si distribuisce secondo una Normale con media 23 e varianza 7. Se definiamo "sottopeso" le ragazze con BMI inferiore a 18, qual è la probabilità di essere sottopeso? Quante ragazze risulteranno sottopeso in un gruppo di 60?

Variabile aleatoria: X = valore del BMI

Informazioni: $\mu=23$ $\sigma^2=7$

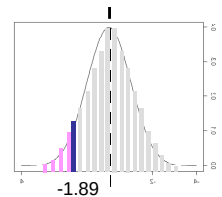
Quesito: $P(X < 18)$

Standardizziamo il valore $x=18$: $z = \frac{18-23}{\sqrt{7}} = -1.89$

(è negativo!) $\Phi(-1.89) = 1 - \Phi(1.89)$

$= 1 - 0.971 = 0.029 \approx 3\%$

Su 60 ragazze, circa il 3%, pari a $0.029 \cdot 60 = 1.74$, dunque circa 2 risulteranno in sovrappeso



Il valore atteso

Data una variabile aleatoria X , definiamo valore atteso la media dei valori che essa può assumere (secondo le probabilità di ciascun valore)

$$E(X) = \sum_i x_i \cdot p_i \quad \text{Variabile aleatoria discreta}$$

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx \quad \text{Variabile aleatoria continua}$$

Descrive in sostanza un valore rappresentativo di tutti i valori che si possono ottenere

Esempio di calcolo del valore atteso per v.a. discreta: Gioco 2



3 lanci consecutivi di una moneta con
 $\text{prob}(\text{Testa})=30\%$ $\text{prob}(\text{Croce})=70\%$

X = numero di "Testa"

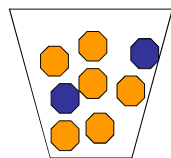
$X = 0$	$X = 1$	$X = 2$	$X = 3$
$p(X=0)$ $= 0.343$	$p(X=1)$ $= 0.441$	$p(X=2)$ $= 0.189$	$p(X=3)$ $= 0.027$

$$E(X) = \sum_i x_i \cdot p_i = 0 \cdot 0.343 + 1 \cdot 0.441 + 2 \cdot 0.189 + 3 \cdot 0.027 = 0.9$$

La legge Binomiale

In ogni contesto assimilabile all'osservazione di un evento ("successo") che ha probabilità π di verificarsi, in N casi, o soggetti, o "prove", in cui interessa il numero totale (X) di successi, si possono usare le seguenti formule

$Y=0,1$ $\Pr(Y=1)=\pi$
 N prove $X=\text{tot}(Y=1)$



- Il numero medio "atteso" di eventi è $N \cdot \pi$
- La probabilità di osservare esattamente x eventi è data da:

$$p(X = x) = \binom{N}{x} \pi^x (1 - \pi)^{N-x}$$

Dove:

$$\binom{N}{k} = \frac{N!}{(N-k)! \cdot k!}$$

$$k! = k \cdot (k-1) \cdot (k-2) \cdot \dots \cdot 2 \cdot 1$$

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$$

$$0! = 1$$

Esempio: Binomiale

In una popolazione di pazienti sottoposti a terapia, la probabilità di risposta è pari a 25%. A un medico vengono assegnati 20 pazienti. Quanti pazienti in media risponderanno al trattamento? Qual è la probabilità che 10 pazienti rispondano al trattamento?

Variabile aleatoria: X = numero di rispondenti fra $N=20$ pazienti

Informazioni: $\pi=0.25$

Quesiti: $E(X)$ $P(X=10)$

Usiamo la Binomiale ($N=20$ $\pi=0.25$):

$$p(X = x) = \binom{N}{x} \pi^x (1 - \pi)^{N-x}$$

Valore atteso: $N \cdot \pi = 20 \cdot 0.25 = 5$

$$p(X = 10) = \binom{20}{10} 0.25^{10} (1 - 0.25)^{20-10} = 0.0099$$

Gioco 2: calcolare le probabilità se $\Pr(\text{Testa})=0.3$ usando la Binomiale



Es: 3 lanci consecutivi di una moneta, X = numero di "Testa"

X variabile aleatoria

Risultati
possibili:

C	C	C	C	C	T	C	T	T	T	T	T
			C	T	C	T	C	T			
			T	C	C	T	T	C			

Distribuzione di
probabilità di X :

**Binomiale con
parametri $N=3$
e $p=0.3$**

$X = 0$	$X = 1$	$X = 2$	$X = 3$
$p(X=0)$ $= 0.343$	$p(X=1)$ $= 0.441$	$p(X=2)$ $= 0.189$	$p(X=3)$ $= 0.027$

$$p(X = x) = \binom{N}{x} \pi^x (1 - \pi)^{N-x}$$

Gioco 2: calcolare le probabilità se $\Pr(\text{Testa})=0.3$ usando la Binomiale



Es: 3 lanci consecutivi di una moneta, X = numero di "Testa"
 X variabile aleatoria

Risultati
possibili:

C	C	C	C	C	T	C	T	T	T	T	T
			C	T	C	T	C	T			
			T	C	C	T	T	C			

$X = 0$

$X = 1$

$X = 2$

$X = 3$

Distribuzione di
probabilità di X :

**Binomiale con
parametri $N=3$
e $p=0.3$**

$$p(X=0) = 0.343$$

$$= \binom{3}{0} 0.3^0 (1-0.3)^{3-0} = \frac{3!}{(3-0)!0!} \cdot 1 \cdot 0.7^3 = \frac{3!}{3! \cdot 1} \cdot 0.7^3$$

$$p(X = x) = \binom{N}{x} \pi^x (1 - \pi)^{N-x}$$

Gioco 2: calcolare il valore atteso usando la Binomiale



Es: 3 lanci consecutivi di una moneta, X = numero di "Testa"
 X variabile aleatoria

Risultati
possibili:

C	C	C	C	C	T	C	T	T	T	T	T
			C	T	C	T	C	T			
			T	C	C	T	T	C			

$X = 0$

$X = 1$

$X = 2$

$X = 3$

Distribuzione di
probabilità di X :

**Binomiale con
parametri $N=3$
e $p=0.3$**

$$p(X=0) = 0.343$$

$$p(X=1) = 0.441$$

$$p(X=2) = 0.189$$

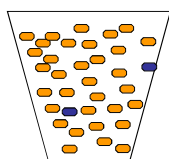
$$p(X=3) = 0.027$$

$$E(X) = N\pi = 3 \cdot 0.3 = 0.9$$

La legge di Poisson

Nel contesto di un fenomeno assimilabile all'osservazione di un evento RARO (π molto piccola) in una serie INFINITA di prove o lungo un INTERVALLO di tempo (o su una superficie, e simili), per il numero totale (X) di eventi osservati si possono usare le seguenti formule

Il parametro λ (detto TASSO) rappresenta il numero di eventi atteso in infinite prove ovvero da la prob. di evento nell'intervallo di osservaz.



- Il numero medio "atteso" di eventi è λ
- La probabilità di osservare esattamente x eventi è data da:

$$p(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Dove: $e^{-\lambda} = \frac{1}{e^{\lambda}}$

$$k! = k \cdot (k-1) \cdot (k-2) \cdot \dots \cdot 2 \cdot 1$$

$$k^0 = 1$$

Se dobbiamo contare il numero X di eventi in N prove con $\pi = P(\text{Evento}) \approx 0$ e $N \gg 0$ usiamo la Poisson con $\lambda = N \cdot \pi$

Esempio: Poisson

Una certa malattia colpisce appena 1 neonato su 1000. Qual è la probabilità che in un campione di 50 neonati si osservi almeno 1 bimbo affetto dalla malattia?

Variabile aleatoria: X = numero di bambini affetti

Informazioni: $p(\text{affetto}) = \pi = 0.001$ numero di prove: $N = 50$

Quesito: $P(X=1 \text{ o } X=2 \text{ o } \dots) = P(X>0)$

Usiamo la Poisson, con $\lambda = N \cdot \pi = 50 \cdot 0.001 = 0.05$

$$P(X>0) = 1 - P(X=0)$$

$$p(X=0) = \frac{e^{-0.05} 0.05^0}{0!} = e^{-0.05} = 0.95$$

$$P(X>0) = 1 - 0.95 = 0.05 = 5\%$$