1 Analisi dei dati

1.1 Problema in esame

Test, sottoposto ad un candidato durante un colloquio, composto da *domande a tripla risposta multipla.*

Nel suddetto documento vengono analizzate le relazioni che intercorrono tra due domande, denominate A e B, a seconda se il candidato risulta in grado di rispondervi correttamente o meno.

1.2 Caratteristiche degli eventi di coppia

Tipi di eventi trattati:

- **Eventi indipendenti**;
- **Eventi dipendenti**:
  - A e B sono strettamente dipendenti;
  - A implica B.
- **Evento conosciuto ed evento indovinato**.

Struttura usata per rappresentare la probabilità degli eventi di coppia:

$$AB$$
$$/\backslash$$
$$A\ B$$
$$\backslash/$$
$$Z$$

con:

- AB rappresenta la probabilità complessiva dell'evento che si verifica sempre;
- A rappresenta la probabilità che permette il verificarsi di A, ma non di B;
- B rappresenta la probabilità che permette il verificarsi di B, ma non di A;
- Z rappresenta la probabilità a zero, l'impossibilità del verificarsi dell'evento.

1.2.1 Eventi indipendenti

A e B sono due domande la quali risposte sono completamente scorrelate tra di loro.

$$P(A)P(B)$$
$$/\backslash$$
$$P(A)(1-P(B))\ (1-P(A))P(B)$$
$$\backslash/$$
$$(1-P(A))(1-P(B))$$

Considerazioni generali La probabilità complessiva nel caso di domande indipendenti A e B viene data da $P(A)$ per $P(B)$.
Se è conosciuta dal candidato la risposta alla domanda A ma non alla domanda B la probabilità di ottenere una risposta corretta è $P(A)$, mentre la probabilità di ottenere una risposta non corretta per B vale $1 - P(B)$. Il ragionamento duale è svolto nel calcolo della probabilità per la riposta corretta alla domanda B ma non ad A.
La probabilità di non ottenere alcuna risposta corretta alle due domande viene calcolato prendendo in considerazione gli eventi contrari a quelli coinvolti. Dunque per A la probabilità che il candidato non conosca la soluzione è $1 - P(B)$, dualmente per B la probabilità è $1 - P(A)$.

1.2.2 Eventi dipendenti

A e B sono due domande fortemente correlate tra di loro, se si risponde correttamente ad una delle due domande si risponde correttamente anche all'altra.

$$P(A)^2$$

/\

0 0

\/

$$(1 - P(A))^2$$

Considerazioni generali La probabilità complessiva nel caso di domande dipendenti A e B viene data da $P(A)$ per $P(B)$; ma $P(A) = P(B)$ dunque $P(A)^2 = P(B)^2$.
Conseguentemente se il candidato non conosce la risposta alla domanda A non può conoscere la risposta alla domanda B dunque la probabilità di conoscere uno dei due eventi è pari a 0.
In questo caso la probabilità a 0 è $(1 - P(A))(1 - P(B)) = (1 - P(A))^2$ essendo che A=B.

A implica B Se si sa rispondere alla domanda A di conseguenza si è in grado di rispondere anche alla domanda B.
Tuttavia non vale il ragionamento opposto, se si sa rispondere alla domanda B non significa che si è in grado di rispondere alla domanda A.

$$P(A)$$

/\

$0\ P(B) - P(A)$

\/

$$1 - P(B)$$

Considerazioni generali La probabilità complessiva nel caso di domande dipendenti A e B viene dato esclusivamente da $P(A)$ in quanto la conoscenza di sia di A che di B è possibile solo se si ha piena conoscenza di A.
Dunque la probabilità che si conosca la risposta alla domanda A ma non a B è impossibile (pari a 0); mentre se si ha conoscenza della domanda B ma non di A la probabilità si stanzia a $P(B) - P(A)$
La probabilità a zero è $1 - P(B)$ indicatore dell'impossibilità di avere la risposta corretta per A.

1.3 Evento conosciuto ed evento indovinato

Durante un test il candidato deve saper scegliere la risposta, corretta o meno, alla domanda posta. Le variabili che entrano in gioco durante l'esecuzione dell'atto non riguardano esclusivamente la conoscenza personale del singolo.
La probabilità di un evento A è data dalla formula:

$$P(A) = P(A_C) + P(A_I)$$

Le variabili in uso sono:

- $P(A_C)$: probabilità che il candidato sappia rispondere alla domanda A correttamente per sua conoscenza;
- $P(A_I)$: probabilità che il candidato sappia rispondere alla domanda A correttamente indovinando.

Per quanto appena definito sopra valgono le seguenti proprietà:
$P(B_C|A_C) = 1$
$P(B_C|A_I) = P(B_C)$
$P(B_I|A_C) = 0$
$P(B_I|A_I) = P(B_I)$

1.3.1 Probabilità di rispondere correttamente ad una domanda

Variabili coinvolte:

- $P(A)$: probabilità necessaria perchè si verifichi per la domanda A che il candidato dia la risposta corretta. Per la legge dei grandi numeri la frequenza porta alla probabilità.
- S_0: insieme dei casi in cui in un domanda non viene scartata alcuna risposta dal dominio delle risposte possibili;
- S_1: insieme dei casi in cui in una domanda viene scartata una risposta dal dominio delle risposte possibili;
- S_2: insieme dei casi in cui in una domanda vengono scartate due risposte dal dominio delle risposte possibili.
- $P(I)$: probabilità di dare la risposta corretta alla domanda A indovinando;
- $P(C)$: probabilità di dare la risposta corretta alla domanda A per conoscenza.

Sapendo che $P(I) = P(A) - P(C)$ logicamente vale anche $P(A) = P(I) + P(C)$.

Se un candidato non è in grado scartare alcuna risposta dalla domanda ha 1 possibilità su 3 di, indovinando, dare la risposta corretta. Se un candidato invece risulta in grado di scartare una risposta, sbagliata, alla domanda rimane con 1 possibilità su 2 di poter dare la risposta corretta. Se invece, caso ottimo, il candidato ha piena conoscenza della domanda posta risulta in grado di scartare due risposte sbagliate lasciando un'unica risposta possibile, quella esatta. Il ragionamento sopra espresso può venire espresso con la seguente espressione:

$$P(A) = P(S_0)\tfrac{1}{3} + P(S_1)\tfrac{1}{2} + P(S_2)$$

Ora individuiamo quale è la probabilità effettiva per un candidato di dare la risposta corretta ad una domanda A.

$1 = S_0 + S_1 + S_2$
$S_0 = 1 - S_1 - S_2$

Sostituendo:

P(A)= $(1 - P(S_1) - P(S_2))\frac{1}{3} + P(S_1)\frac{1}{2} + P(S_2)$
$= \frac{1}{3} - \frac{1}{3}P(S_1) - \frac{1}{3}P(S_1) + \frac{1}{2}P(S_1) + P(S_2)$
$=\frac{1}{3} + \frac{1}{6}P(S_1) + \frac{2}{3}P(S_2)$

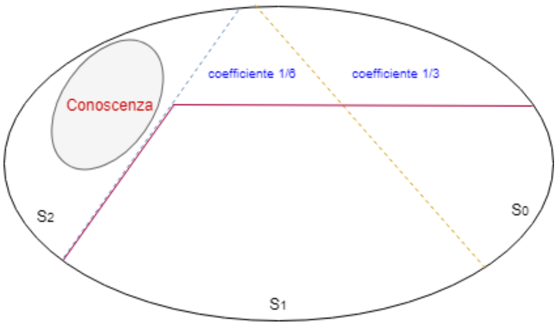

Figure 1: Rappresentazione insiemistica della probabilità di rispondere correttamente ad una domanda: P(A)

Considerazioni importanti

In conclusione P(A)=$\frac{1}{3} + \frac{1}{6}P(S_1) + \frac{2}{3}P(S_2)$. Ovvero la probabilità per un candidato di dare in una domanda A la risposta corretta dipende dai seguenti fattori:

- $\frac{1}{3}$: coefficiente che rappresenta la probabilità effettiva per chi non conosce la risposta alla domanda di dare la risposta corretta;
- $\frac{1}{2}P(S_1)$: coefficiente che rappresenta la probabilità effettiva di dare la risposta corretta quando il candidato è in grado di scartare una risposta sbagliata alla domanda;

- $\frac{2}{3}P(S_2)$: coefficiente che rappresenta la probabilità effettiva di dare la risposta corretta quando il candidato è in grado di scartare due risposte sbagliate alla domanda.

Dall'analisi della tipologia di eventi di coppia e dal calcolo della probabilità necessaria per poter rispondere correttamente ad una domanda, si giunti alla valenza dei seguenti assiomi:

1. Le coppie di domande A e B devono essere fra loro disgiunte, altrimenti si genererebbero situazioni di invalidità dei risultati;
2. Per rispondere correttamente ad una domanda non è necessario che il candidato abbia piena conoscenza di tutti gli argomenti richiesti dall'esame, ma bensì ne risultano sufficienti $n-1$;
3. La probabilità di concedere è contenuta all'interno di S_2, in quanto se un candidato conosce è conseguentemente in grado da una domanda di scartare due risposte sbagliate.

1.3.2 Il piano

La probabilità P(A) che un candidato ha in gioco nel momento in cui si approccia a rispondere ad una domanda può venire rappresentata in un piano.

Di seguito viene mostrata l'immagine di un modellino, rappresentativo di P(A), realizzato durante lo studio.
TODO: foto modello

Ognuno dei tre assi cartesiani rappresenta un insieme dei casi di scarto (S_0, S_1, S_2). L'intersezione tra i punti del piano indica la regione accettabile contenente il range di valori assumibili da P(A). Tale punto proiettato su ognuno dei tre assi permette l'individuazione esatta dei coefficienti delle variabili S_0, S_1, S_2.

Ogni porzione del piano viene individuata con la seguente tecnica:

1. Per individuare ogni retta passante per S_0, S_1 e S_2 è necessario assumere che $S_0 + S_1 + S_2 = 1$;
2. La retta passante per S_0 è rappresentabile per mezzo delle seguenti equazioni:

$$S0 = 0 \text{ e } S_1 + S_2 = 1$$

   In questo modo l'asse S_0 è fissato a 0 e estrapolando S_1 e S_2 da $S_1 = -S_2 + 1$ assumono valori tra (0,1).

3. Il medesimo ragionamento vale per le rette passanti per S_1 e S_2.

$$S_1 = 0 \text{ e } S_0 + S_2 = 1$$

   l'asse S_1 è fissato a 0 e S_0 e S_2 assumono valori tra (0,1)

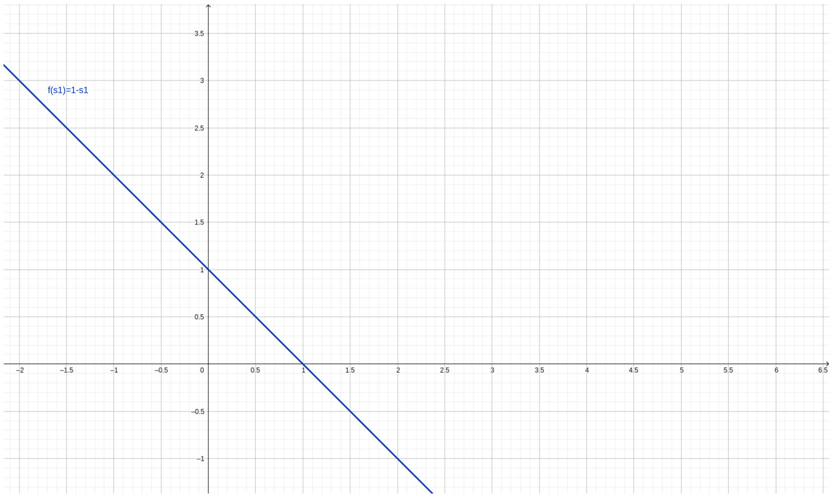

Figure 2: Rappresentazione della retta passante per $S_0 = 0$

$$S_2 = 0 \text{ e } S_1 + S_0 = 1$$

l'asse S_2 è fissato a 0 e S_0 e S_2 assumono valori tra (0,1)

4. In questo modo l'unione di tutte le rette passanti per gli assi creano la regione accettabile dei valori di P(A).

Avendo rappresentato il piano si ottiene nei punti di intersezioni fra le tre rette la regione accettabile per P(A). Inoltre è possibile, ora, individuare il fascio di rette che tangenti il piano permettono di affermare se una specifica domanda è, in base alla sua frequenza, ha difficoltà bassa, media, alta per un candidato.

- Se una domanda ha una difficoltà bassa la retta si situa passante per i punti $0 < S_2 <= 1$ (molto vicino a 1) e $(S_0, S_1) < 0$ (tendenti a 0);
- Se una domanda ha una difficoltà alta la retta si situa passante per i punti $S_2 <= 0$ (molto vicino a 0), $S_1 < 1$ e $S_0 <= 1$ (tendente a non scartare alcuna risposta);
- Se una domanda ha una difficoltà media la retta si situa nella parte centrale della regione accettabile, passante per i punti $0 <= (S_0, S_1, S_2) <= 1$

Rappresentazione di P(A)

Vediamo alcuni casi di come le domande possono venire rappresentate sul piano:

La funzione di partenza è:

$$F = \frac{1}{3} + \frac{1}{6}S_1 + \frac{2}{3}S_2$$

Va esplicitato S_1, i passaggi utili da fare sono i seguenti:

$\frac{-1}{6}S_1 = \frac{1}{3} + \frac{2}{3}S_2 - F \rightarrow S_1 = -4S_2 - 2 + 6F$

Essendo che $0 <= S_2 <= 1$ usando $S_1 = 1$ e $S_2 = 0$ allora si ottiene che $F = \frac{1}{2} = 0.5$

Quanto appena calcolato può venire rappresentato graficamente definendo la retta $S_1 = 1 - S_2$ (responsabile di definisce una porzione del piano in base alle variaibili coinvolte) e mediante la retta $S_1 = -4S_2 - 2 + 6F$ (che permette di calcolare il fascio di rette tangenti alla prima retta).

Figure 3: Rappresentazione di P(A) per una frequenza 0.5 proiettata su assi $S_0 = 0$, S_1 e S_2

Nella figura sopra sono posti i seguenti significati:

- La linea azzurra rappresenta $S_2 = 1 - S_1$;
- La linea rosa rappresenta la retta tangente $S_1 = -4S_2 + 1$;
- punto A (blu):
  $S_1 = 0.5 = \frac{1}{2}$
  $\frac{1}{2} = -4S_2 + 1 \rightarrow 4S_2 = 1 - \frac{1}{2} \rightarrow S_2 = \frac{1}{8}$
  $S_0 = 1 - \frac{1}{2} - \frac{1}{8} = \frac{3}{8}$
  Ovvero metà dei candidati sottoposti alla domanda sanno scartare una delle risposte, lo 0.16% sa dare la risposta corretta e lo 0.36% non sa scartare alcune delle risposte possibili.
- punto B (verde)
  $S_1 = 0$
  $S_2 = \frac{1}{4}$
  $S_0 = 1 - \frac{1}{4} = \frac{3}{4}$
  Ovvero nessun dei candidati sottoposti alla domanda sanno scartare una

delle risposte, lo 0.25% sa dare la risposta corretta e lo 0.75% non sa scartare alcune delle risposte possibili.

- punto C (fucsia):
  $S_1 = \frac{1}{4}$
  $S_2 = \frac{3}{16}$
  $S_0 = 1 - \frac{1}{4} - \frac{3}{16} = \frac{9}{16}$
  Ovvero lo 0.25% dei candidati sottoposti alla domanda sanno scartare una delle risposte, lo 0.19% sa dare la risposta corretta e lo 0.56% non sa scartare alcune delle risposte possibili.

- punto D (arancione):
  $S_1 = \frac{1}{3}$
  $S_2 = \frac{1}{3}$
  $S_0 = 1 - \frac{1}{3} - \frac{1}{3} = \frac{1}{3}$
  Osserviamo che il punto in esame fuoriesce dalla regione delimitata dalla retta tangente di frequenza 0.5 ($S_1 = -4S_2+1$). Conseguenza diretta data dall'impossibilità di ottenere una probabilità del 50% sulla domanda con $\frac{1}{3}$ di candidati che sa scartare 2 risposte, $\frac{1}{3}$ ne sa scartare 1 e $\frac{1}{3}$ nessuna.

Vediamo ulteriori due casi per valutare cosa accade nel piano nel caso di una frequenza quasi in prossimità di 1 e pari alla soglia minima dell'indovinato. Il grafico è il seguente:

Figure 4: Rappresentazione di P(A) per una frequenza 0.33 e 0.8 proiettate su assi $S_0 = 0$, S_1 e S_2

- la linea azzurra mostra la retta tangente con frequenza 0.80%. In questa abbiamo calcolato il punto E (giallo):
  $S_1 = 0$
  $S_2 = \frac{14}{20}$
  $S_0 = 0$

  Ovvero nessuno dei candidati ha la conoscenza per poter scartare nè una nè due risposte, percui l'unica possibilità per un candidato di rispondere alla domanda è indovinare. È evidente come se un candidato non sa la risposta ad una domanda ha una probabilità dello 0.33% di poter indovinare la risposta corretta.

- la linea viola mostra la retta tangente con frequenza 0.33%. In questa abbiamo calcolato il punto F (rosa):
  $S_1 = 0$
  $S_2 = 0$
  $S_0 = 1$

  Ovvero nessuno dei candidati ha la conoscenza per poter scartare nè una nè due risposte, percui l'unica possibilità per un candidato di rispondere alla domanda è indovinare. È evidente come se un candidato non sa la risposta ad una domanda ha una probabilità dello 0.33% di poter indovinare la risposta corretta.