02/16/2024 — Terzo appello

Nome e Cognome: ________________________ Matricola: ______________

Indicazioni Importati: Siamo liberi di utilizzare appunti scritti o stampati. **Non possiamo però utilizzare dispositivi elettronici, comunicare tra di noi o con l'esterno, né passarci del materiale tra di noi**.
A meno che non venga esplicitamente detto il contrario, non è necessario calcolare il valore numerico delle soluzioni. A parte questo fatto, **è sempre importante mostrare il ragionamento e le formule usate, ed arrivare ad una risposta esatta**, anche se espressa in funzione di altri operatori (per esempio coefficienti binomiali, radici quadrate, esponenti, ecc). Un esempio: se arriviamo a una espressione del tipo $\binom{10}{5}$, possiamo lasciare questa come risposta senza ulteriori semplificazioni, oppure possiamo fare i calcoli ed arrivare al valore numerico 252. Invece scrivere solo "252" senza che sia chiaro da dove viene quel numero, **sarà considerata una risposta invalida.**

1 Calcolo combinatorio

1. Le targhe automobilistiche sono costituite da 2 lettere, seguite da 3 cifre, e infine 2 lettere. Sapendo che le lettere possono venire scelte tra le 26 lettere dell'alfabeto anglosassone si calcoli quante targhe differenti possono essere ottenute e quindi quante automobili possono essere immatricolate. (2 pt)
2. In quanti modi diversi posso distribuire 12 penne in 5 cassetti? (Si assuma che ogni cassetto possa contenere da 0 a 12 penne e che le 12 penne possano essere considerate indistinguibili, cioè, non ci interessa l'ordine). (4 pt)
3. Supponiamo ora che in ogni cassetto ci debba essere almeno una penna. Quanti modi diversi ci sono di distribuire le penne nei 5 cassetti? (2 pt)
4. Un barman ha a disposizione 4 liquori base, quanti cocktails può ottenere mescolandone 3 alla volta? (2 pt)

1 Soluzione

1. Con due lettere scelte tra 26, **con possibile ripetizione** e **tenendo conto dell'ordine**, posso ottenere 26^2 disposizioni. Con tre cifre, posso ottenere 10^3. Per cui il totale sarà: $26^2 \times 10^3 \times 26^2$
2. Possiamo rappresentare la disposizione delle penne nei cassetti con sequenze di 12 punti (.) separati da 4 barre ($|$). In questo modo, i punti tra una barra e l'altra, considerando anche i punti negli estremi, rappresentano i punti dentro ogni cassetto.

   Per esempio, la sequenza che rappresenta il mettere tre penne in ogni cassetto è:

   ...|...|...|...|...

   Mentre quella che ne mette 0 penne nel primo cassetto, 2 nel secondo, 4 nel terzo, 3 nel quarto e 3 nel quinto si indicherebbe così:

   |..|....|...|...

   Perciò, si può arrivare alla risposta calcolando il numero di permutazioni di questi 12 punti e 4 separatori, dove punti e separatori formano due classi di elementi indistinguibili:

$$\frac{(12+4)!}{12!\cdot 4!} = \frac{(16)!}{12!\cdot 4!}$$

Visto che l'**ordine delle penne non conta**, dato che sono indistinguibili, l'esercizio si può risolvere anche invocando subito la formula che calcola il numero di **combinazioni con ripetizione** di $k = 12$ elementi (le penne) prese da un insieme di $n = 5$ elementi (i cassetti).

$$\binom{n+k-1}{k} = \binom{16}{12} = \frac{(16)!}{12!\cdot 4!}$$

3. Dopo aver distribuito una penna in ognuno dei 5 cassetti, mi rimangono 7 penne da distribuire con le stesse modalità dell'esercizio precedente, per cui il numero di modi possibili sarà:

$$\binom{n+k-1}{k}$$

dove $n = 5$ e $k = 7$, cioè:

$$\binom{n+k-1}{k} = \binom{5+7-1}{7} = \frac{(11)!}{7!\cdot 4!}$$

4. Per fare un qualsiasi cocktail, il barman deve scegliere 3 liquori dall'insieme dei 4 liquori base. Il cocktail ottenuto non dipende dall'ordine in cui si mischiano i liquori, quindi stiamo parlando del numero di combinazioni (senza ripetizione).

$$\binom{4}{3} = \frac{4!}{3!\cdot 1!} = 4$$

2 Caldaie *green*

Ogni anno, una caldaia eco-friendly ha una probabilità fissa del 15% di guastarsi. Sia $\mathcal{X}$ il numero di anni di funzionamento della caldaia (compreso l'anno in cui si guasterà).

1. Che tipo di distribuzione segue la variabile casuale $\mathcal{X}$? Scrivere un'espressione generale per $P(\mathcal{X} = k),\ \forall k \in \mathbb{N}$. (2pt)

2. Calcolare il valore atteso e la varianza di $\mathcal{X}$. (2pt)

3. Per ogni anno di funzionamento (non contando l'anno in cui si guasterà), si stima che la caldaia eco-friendly avrà **ridotto** il *carbon-footprint* (CF) del proprietario di a unità, e, nel momento in cui si guasterà, la riparazione della caldaia **aumenterà** il CF di diecimilla unità. Quanto dev' essere il *minimo* valore di a affinché, in media, l'utilizzo della caldaia porti ad una riduzione del CF? (6pt)

2 Soluzione

1. $\mathcal{X}$ è un tipico esempio di variabile che segue una distribuzione geometrica, con $p = 0.15$; quindi:

   $P(\mathcal{X} = k) = (0.85)^{k-1} \cdot 0.15$

2. Una variabile che segue una distribuzione geometrica con parametro p ha valore atteso $\mathbb{E}[\mathcal{X}] = 1/p$ e varianza $Var(\mathcal{X}) = (1-p)/p^2$. Quindi:

$$\mathbb{E}(\mathcal{X}) = \frac{1}{0.15} = 6.67$$

e

$$Var(\mathcal{X}) = \frac{1-0.15}{0.15^2} = 37.77$$

3. Dato che c'è della stocasticità in $\mathcal{X}$, allora la riduzione del CF dovuta alla caldaia si può modellare con un altra v.a., $\mathcal{Y}$. Se per ogni anno di funzionamento si ottiene una riduzione di a, e, nell'anno in cui si guasta la caldaia, il CF aumenta di diecimila unità, avremo che:

$$\mathcal{Y} = -a \cdot (\mathcal{X} - 1) + 10000$$

e quindi, per le **proprietà del valore atteso**, abbiamo:

$$\mathbb{E}(\mathcal{Y}) = -a(\mathbb{E}(X) - 1) - 10000 = -5.67 \cdot a - 10000.$$

Quindi, affiché $\mathbb{E}(\mathcal{Y})$ sia ≤ 0, dovrà verificarsi:

$$-5.67 \cdot a - 10000 \leq 0 \qquad \text{cioè:} \qquad 5.67 \cdot a + 10000 \geq 0$$

e questo implica:

$$a \geq \frac{10000}{5.67} \approx 1764$$

Quindi, la caldaia porterà ad una riduzione del proprio CF solo se ogni anno che essa è in funzione, aiuta a ridurre il CF di almeno 1764 unità.

3 Container

Ci sono due super-computer sui quali vengono allocati dei container per gli studenti del corso di algoritmi. Li chiameremo computer A e computer B. Quando uno studente fa il login, gli viene assegnato un container in uno dei due super-computer. Solitamente, i container del computer A sono più lenti. Più precisamente, sappiamo che:

- Dato un certo algoritmo, i container istanziati su A eseguono l'algoritmo in un tempo che ha un valore atteso di 200 secondi, con una deviazione standard di 20 secondi; Invece, il tempo impiegato dai container istanziati sul server B per eseguire lo stesso algoritmo ha un valore atteso di 170 secondi, con una deviazione standard di 30 secondi. Quanto vale la varianza del tempo di esecuzione dell'algoritmo su ogni tipo di container? (1pt)

- Supponendo che le distribuzioni dei tempi di esecuzione dell'algoritmo sui container di tipo A e B siano ignote, indicare un limite minimo per le probabilità seguenti: **a)** prob. che il tempo di esecuzione sia compreso tra 160 e 240 secondi per il container di tipo A e **b)** tra 140 e 200 secondi per il container di tipo B. (Dare una soluzione che valga per qualsiasi tipo di distribuzione). (4pt)

- Si stima che i container vengono allocati agli studenti sul computer B la metà delle volte rispetto alle assegnazioni su A. Possiamo inoltre supporre che i tempi di esecuzione dei container seguano una distribuzione normale (con i rispettivi valori attesi e deviazioni standard).

  Uno studente fa login al sistema ed esegue l'algoritmo. Non si sa su quale server gli venga allocato il container; ma l'esecuzione dell'algoritmo dura più di 200 secondi. Con le nuove ipotesi, qual è la probabilità che il container sia sul server B? Trovare una risposta numerica. (6pt)

3 Soluzione

Dati:
Sia $\mathcal{E}$ l'evento "il container viene allocato sul server A" (senza supporre niente riguardo al tempo di esecuzione), e sia $\mathcal{F}$ l'evento "il tempo di esecuzione è maggiore di 200 secondi". Sappiamo dalla traccia dell'esercizio che $P(\mathcal{E}) = 2/3$ e $P(\overline{\mathcal{E}}) = 1/3$.

- La varianza è il quadrato della deviazione standard. Perciò avremmo:

$$\sigma_A^2 = 400\text{sec}^2 \qquad \text{e:} \qquad \sigma_B^2 = 900\text{sec}^2$$

- Visto che veniamo interrogati su intervalli simmetrici rispetto alla media, possiamo usare la disugualianza di Čebyšëv, che ci dice che, per una v.a. $\mathcal{X}$ di media $\mu > 0$ finita e dev. standard $\sigma > 0$ finita, si ha, per ogni $\lambda > 0$:

$$P(\mu - \lambda\sigma < \mathcal{X} < \mu + \lambda\sigma) > 1 - \frac{1}{\lambda^2}$$

  Il caso dei due tipi di container compie le condizioni, quindi avremmo, nel caso dei container di tipo A:

$$P(160 < \mathcal{X}_A < 240) = P(200 - 2 \cdot 20 < \mathcal{X}_A < 200 + 2 \cdot 20) = P(\mu_A - 2\sigma_A < \mathcal{X}_A < \mu_A + 2\sigma_A)$$

  E quindi, nel caso di A, posto $\lambda = 2$ si avrà:

$$P(160 < \mathcal{X}_A < 240) \geq 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = 0.75$$

  Analogamente, nel caso di B avremmo:

$$P(140 < \mathcal{X}_B < 200) = P(170 - 1 \cdot 30 < \mathcal{X}_B < 170 + 1 \cdot 30) = P(\mu_B - \sigma_B < \mathcal{X}_B < \mu_B + \sigma_B)$$

  E, ponendo sempre $\lambda = 1$, qui avremmo:

$$P(110 < \mathcal{X}_B < 230) \geq 0$$

  (L'unica cosa che possiamo dire in questo caso è che la probabilità dell'evento $P(110 < \mathcal{X}_B < 230)$ è non nulla).

- **Ci viene chiesta trovare la probabilità condizionata $P(\overline{\mathcal{E}}|\mathcal{F})$.**

  Iniziamo esprimendo questa probabilità in termini di $P(\mathcal{F}|\overline{\mathcal{E}})$, $P(\mathcal{F}|\mathcal{E})$, visto che il **Teorema di Bayes**, ci può essere di aiuto:

$$P(\overline{\mathcal{E}}|\mathcal{F}) = \frac{P(\mathcal{F}|\overline{\mathcal{E}})P(\overline{\mathcal{E}})}{P(\mathcal{F})} = \frac{P(\mathcal{F}|\overline{\mathcal{E}})P(\overline{\mathcal{E}})}{P(\mathcal{F}|\mathcal{E})P(\mathcal{E}) + P(\mathcal{F}|\overline{\mathcal{E}})P(\overline{\mathcal{E}})}.$$

  Ci manca quindi di calcolare $P(\mathcal{F}|\mathcal{E})$ e $P(\mathcal{F}|\overline{\mathcal{E}})$ e abbiamo finito.

  $P(\mathcal{F}|\mathcal{E})$ e $P(\mathcal{F}|\overline{\mathcal{E}})$ possono essere espresse in termini della funzione Φ utilizzando il **teorema del limite centrale** (in particolare, il fatto che se $\mathcal{X}$ è una distribuzione normale con valore atteso μ e deviazione standard σ, allora $(\mathcal{X} - \mu)/\sigma$ è una distribuzione **normale standard**, e quindi $P((\mathcal{X} - \mu)/\sigma \leq z) = \Phi(z)$.

Ora, il tempo di esecuzione $\mathcal{X}$ sul container di tipo A segue una distribuzione normale con valore atteso $\mu = 200$ e deviazione standard $\sigma = 20$; quindi, la variabile $(\mathcal{X} - 200)/20$ segue una distribuzione normale standard. Dato un container allocato su A, la probabilità che il tempo di esecuzione sia maggiore di 200 secondi è pertanto

$$\begin{aligned} P(\mathcal{F}|\mathcal{E}) &= P(\mathcal{X} > 200) = 1 - P(\mathcal{X} \leq 200) = 1 - P((\mathcal{X} - 200)/20 \leq (200 - 200)/20) = \\ &= 1 - P((\mathcal{X} - 200)/20 \leq 0) = 1 - \mathbf{\Phi}(0) = 1 - 0.5 = 0.5 \end{aligned}$$

un container allocato sul server A ha il 50% di probabilità di eseguire l'algoritmo in più di 200 secondi (questo si sarebbe potuto vedere direttamente osservando che la distribuzione normale è simmetrica rispetto al suo valore atteso, che in questo caso è appunto 200 secondi).

Similmente, il tempo di esecuzione $\mathcal{Y}$ sul container di tipo B segue una distribuzione normale con valore atteso $\mu = 170$ e deviazione standard $\sigma = 30$. Quindi la v.a. $(\mathcal{Y} - 170)/30$ segue una distribuzione normale standard, e

$$\begin{aligned} P(\mathcal{F}|\overline{\mathcal{E}}) &= P(\mathcal{Y} > 200) = 1 - P(\mathcal{Y} \leq 200) = 1 - P((\mathcal{Y} - 170)/30 \leq (200 - 170)/30) = \\ &= 1 - P((\mathcal{Y} - 170)/30 \leq 1) = 1 - \mathbf{\Phi}(1) \approx 1 - 0.8413 = 0.1587 \end{aligned}$$

Cioè, un container allocato su B ha circa il 15.87% di probabilità di consumare più di 200 secondi per eseguire l'algoritmo.

Ora abbiamo tutto il necessario per calcolare la soluzione:

$$P(\overline{\mathcal{E}}|\mathcal{F}) = \frac{P(\mathcal{F}|\overline{\mathcal{E}})P(\overline{\mathcal{E}})}{P(\mathcal{F}|\mathcal{E})P(\mathcal{E}) + P(\mathcal{F}|\overline{\mathcal{E}})P(\overline{\mathcal{E}})} = \frac{(1 - \mathbf{\Phi}(1)) \cdot 1/3}{0.5 \cdot 2/3 + (1 - \mathbf{\Phi}(1)) \cdot 1/3} \approx 0.14$$

c'è all'incirca il 14% di probabilità che il container sia stato allocato sul server B.