Ripasso Statistica

Domande aperte

- **Il/la candidato/a definisca la variabile aleatoria distribuita secondo la distribuzione di Poisson. Dopo aver definito tale variabile e la semantica associata, proceda esponendo tutto quanto presentato nel programma del corso su questo argomento:** *Innanzitutto si ha:*

$$X \sim Poi(\lambda)$$

```
_con la probabilità associata così definita:_
```

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

Inoltre si ha che la distribuzione di Poisson rappresenta una binomiale con il n umero di elementi n tendente all'infinito e la probabilità tendente a 0, infatti è detta distribuzione degli eventi rari. Si ha inoltre:

$$p_X(t)=\begin{cases}\frac{\lambda^k}{k!}e^{-\lambda} & \text{se } t \in \{0,\cdots,n\} \\ 0 & \text{altrimenti}\end{cases}$$

$$F_X(t)=\sum_{k\in\mathbb{N}:k\leq t}\frac{\lambda^k}{k!}e^{-\lambda},\ \forall t \in \mathbb{R}$$

$$E[X]=n\cdot p=\lambda$$

$$V[X]=n\cdot(1-p)\cdot p=n\cdot p-n\cdot p^2=\lambda-\frac{\lambda^2}{n}$$

- **Nell'ambito del capitolo sul calcolo della probabilità, il/la candidato/a fornisca definizioni formali di:**

   i. **Probabilità condizionata:** *siano dati uno spazio campione* Ω *ed una misura di probabilità P definita su* $\wp(\Omega)$ *e secondo l'impostazione assiomatica di probabilità, considerati due eventi, A e B con* $P(B)>0$, *viene definita probabilità dell'evento A condizionata dall'evento B la quantità:* $P(A|B)=\frac{P(A\cap B)}{P(B)}$

   ii. **Formula del prodotto:** *sia* $n\in\mathbb{N}_+$ *e data la famiglia* $\{A_i i=1,\cdots,n\}$ *di sottoinsiemi di* Ω, *allora risulta:*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2$$

iii. **Teorema o formula di Bayes:** *sia* $n \in \mathbb{N}_+$ *e data la famiglia* $\{A_i i = 1, \cdots, n\}$ *di sottoinsiemi di* Ω*, allora risulta:* $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=j}^{n} P(B|A_j)P(A_j)}$ *che descrive il modo in cui le opinioni nell'osservare A siano arricchite dall'aver osservato l'evento B*

iv. **Correlazione:** *è una relazione tra due variabili tale che a ciascun valore della prima corrisponda un valore della seconda, seguendo una certa regolarità, non dipende da un rapporto di causa-effetto quanto dalla tendenza di una variabile a cambiare in funzione di un'altra. Si misura con l'indice di covarianza*

v. **Indipendenza stocastica:** *si dice che 2 eventi* $A, B \in \wp(\Omega)$ *sono stocasticamente indipendenti se si ha che:* $P(A) = P(A|B) = P(B)$ *e quindi:* $P(A \cap B) = P(A)P(B)$

vi. **Rapporto correlazione e dipendenza:** *date due serie di dati si ha che se la covarianza è nulla si ha che le due serie sono statisticamente incorrelate. Del resto le due serie sono statisticamente indipendenti se vale che:* $\forall j = 1, \ldots, N \text{ e } k = 1, \ldots, M \text{ si ha } p_{jk} = p_j p_k$ *con* p_{jk} *frequenza relativa doppia mentre le altre sono le frequenze relative marginali. Si ha che la proprietà di indipendenza è più forte dell'incorrelazione, infatti se le due serie di dati sono indipendenti, risultano anche incorrelate mentre il contrario non è detto, infatti si ha che:* $\sum\sum(x_i - \overline{x})(y_i - \overline{y}) = \sum(x_i - \overline{x})\sum(y_i - \overline{y}) = 0$

- **Relativamente alla statistica descrittiva, il/la candidato/a definisca formalmente:**

i. **Frequenza assoluta:** *si dice frequenza assoluta delle modalità* s_j *con* $j = 1, ..., N$ *la quantità* f_j *data dal numero di elementi di* $E = \{x_1, ...x_N\}$ *che hanno valore* s_j*, ovvero il numero di elementi di E con valore* s_j

ii. **Frequenza relativa:** *è il rapporto tra la frequenza assoluta e il numero di elementi:* $p_j = \frac{f_j}{n}$

iii. **Frequenza cumulata assoluta:** *è la somma delle frequenze assolute di tutte le modalità e si indica con:*

$$F_j = \sum_{k: s_k \leq s_j} f_k$$

4. **Frequenza cumulata relativa:** *è la somma delle frequenze relative di tutte le modalità e si indica con:*

$$P_j = \sum_{k: s_k \leq s_j} p_k$$

5. **Frequenza assoluta doppia:** *è la quantita f_{jk} corrispondente al numero di elementi con valore (s_j, u_k)*

6. **Frequenza relativa doppia:** *rapporto tra la frequenza e il numero di elementi:* $p_{jk} = \frac{f_{jk}}{n}$

7. **Frequenza cumulata assoluta doppia:** *somma delle frequenze assolute doppie:*

$$F_{jk} = \sum_{r:s_r \leq s_j; l:u_l \leq u_k} f_{rl}$$

8. **Frequenza cumulata relativa doppia:** *somma delle frequenze relative doppie:*

$$P_{jk} = \sum_{r:s_r \leq s_j; u_l \leq u_k} p_{rl}$$

9. *si può passare dalla frequenze doppie a quelle singole usando le rispettive frequenze marginali (assolute, relative e cumulate) in quanto le distribuzioni marginali sono le distribuzioni dei singoli caratteri presi indipendentementi dagli altri*

R

- **Esercizio 1:**

```
library(prob) # carico il package prob

S <- rolldie(times = 3, makespace = TRUE) # assegno a S lo spazio di probabilità dato dal lancio di un dado 3 volte
S[3:8,] # stampo gli elementi da 3 a 8

Prob(S, X1+X2 > 5) # probabilità che la somma dei primi 2 lanci sia > 5

Prob(S, X1+X2 > 9, given = X1+X2+X3 > 7) # probabilità che la somma dei primi 2 lanci sia > 9 sapendo che la somma dei 3 lanci è > 7

Prob(S, X1>3 & X2!=6 & X3==5) # probabilità che il primo lancio sia > 3 il secondo diverso da 6 e il terzo uguale a 5
```

- **Esercizio 2:**

```
library(prob) # carico il package prob

ball <- urnsamples(1:10,size=1) # assegno a ball lo spazio campione
delle palle numerate da 1 a 10 estratte dall'urna

prob <- rep(1/10,times=10) # assegno a prob la probabilità dei vari
elementi, ovvero ogni estrazione ha probabilità 1/10

S <- probspace(ball, probs=prob) # assegno a S lo spazio di probabilità
dato dallo spazio campione delle 10 palle e dalla loro probabilità

Prob(S, out==10) # stampo la probabilità che esca 10

Prob(S, out==1 | out==5) # stampo la probabilità che esca 1 o 5
```

- **Esercizio 3:**

```
library(prob) # carico il package prob

C <- cards() # assegno a C lo spazio campione dato dal mazzo di carte

MM<-subset(C, !(rank %in% c("J","K","Q"))) # assegno a MM un
sottoinsieme di C senza le figure J, Q, K

MM[8:15,] # stampo gli elementi da 8 a 15
 rank suit
 8 9 Club
 9 10 Club
 13 A Club
 14 2 Diamond
 15 3 Diamond
 16 4 Diamond
 17 5 Diamond
 18 6 Diamond

N <- probspace(MM) # assegno a N lo spazio di probabilità del campione
MM con egual probabilità

Prob(N, rank == 5) # stampo al probabilità che la carta sia di seme
qualsiasi con valore 5

Prob(N, suit == "Spade" , given=(rank == 5)) # stampo la probabilità che
esca il 5 di spade
```

- **Esercizio 4:**

```
library(prob) # carico il package prob

urn <- urnsamples(c("B","B","B","N","N"),size = 3,ordered = TRUE) # assegno a urn lo spazio campione dato da 3 palline bianche e 2 nere da cui si avranno estrazioni da 3 senza reimmissione

space_urn <- probspace(urn) # assegno a space_urn lo spazio di probabilità dato da urn

Prob(space_urn, X1=="N“ | X2=="N“ | X3=="N") # stampo la probabilità che si estragga almeno una biglia nera dalle 3 estrazioni

urn <- urnsamples(c("B","B","B","N","N"),size = 2, ordered = TRUE,replace = TRUE) # assegno a urn lo spazio campione dato da 3 palline bianche e 2 nere da cui si avranno estrazioni da 2 con reimmissione

space_urn <- probspace(urn) # assegno a space_urn lo spazio di probabilità dato da urn

Prob(space_urn, X1=="B“ & X2=="B“) # stampo la probabilità che la prima estrazione sia una biglia bianca e la seconda una biglia bianca
```

Ripasso capitolo 4

- **I valore atteso della variabile che approssima la MEDIA n-esima della successione è dato:** *dal valore atteso delle variabili aleatorie della successione*

- **Data una successione $X_n, n \in \mathbb{N}$ con F_n funzione di ripartizione dell'n-esima variabile aleatoria X_n, cosa deve risultare $\lim_{n\to\inf} F_n(t)$ affinché X_n converga in distribuzione alla variabile aleatoria X, avente distribuzione F ?** *deve risultare* $F(t)$

- **Considerando una successione di variabili aleatorie, la legge dei grandi numeri ci da informazioni circa:** *μ, valore atteso della successione delle medie aritmetiche n-esime*

- **Cosa assicura la Legge dei Grandi Numeri?** *assicura la convergenza della media aritmetica n-esima di una successione*

- **Quali non è una condizione necessaria affinché valga la legge dei grandi numeri per una successione $\{X_n\}$ di variabili aleatorie?** *le variabili aleatorie non devono essere normalmente distribuite*

- **Quale notazione indica che la successione di variabili aleatorie $\{X_n\}$, con F_n generica funzione di distribuzione della n-esima variabile aleatoria della**

successione, converge in distribuzione alla variabile aleatoria X avente funzione di ripartizione F ? *si usa la notazione* $X_n \xrightarrow{d} X$

- **I parametri della distribuzione della variabile che approssima la SOMMA n-esima della successione sono (fatta esclusione per il numero di addendi "n"):** *valore atteso e varianza delle variabili aleatorie* X_i

- **I valore atteso della variabile che approssima la SOMMA n-esima della successione è dato:** *al prodotto della dimensione campionaria per il valore atteso delle variabili aleatorie della successione*

- **Considerando una successione di variabili aleatorie, il teorema del limite centrale:** *permette l'approssimazione della media aritmetica n-esima della succesione con una variabile aleatoria normalmente distribuita*

- **Qual è la condizione di applicazione del teorema del limite centrale?** *il campione di realizzazioni deve soddisfare una numerosità minima, ovvero* $n \geq 30$

Ricapitolando si ha per la variabile che approssima la MEDIA n-esima della successione:

$$E[\overline{X}] = \mu \quad \text{e} \quad V[\overline{X}] = \frac{\sigma^2}{n}$$

e per la variabile che approssima la SOMMA n-sima della successione:

$$E[X_i] = n\mu \quad \text{e} \quad V[X_i] = n\sigma^2$$

Trucchi per gli esercizi

Normale

Vediamo i vari casi, tutti facilmente intuibili guardando il grafico:

- $P(X < a)$, *con* $a > \mu$, *faccio semplicemente* $P(Z < \frac{a-\mu}{\sigma})$

- $P(X > -a)$ *faccio* $P(Z > \frac{-a-\mu}{\sigma})$, *con* $\frac{-a-\mu}{\sigma} = -y$ *che è uguale a* $P(Z < y)$

- $P(X > a)$ *faccio* $1 - P(Z < \frac{a-\mu}{\sigma})$

- $P(X < a)$, *con* $a < \mu$, *comporta un* $P(Z < -y)$ *che è uguale a* *$P(Z > y)$ *che si risolve con* $1 - P(Z < y)$

- $P(a < X < b)$ *si risolve con* $P(X < b) - P(X < a)$

se si chiede di calcolare la P di almeno una su n variabili faccio semplicemente $(P(X < a))^n$

Se devo ragionare sull probabilità che assumono le somme di n variabili devo ricordare che $\mu_{tot} = n\mu$ *e che* $\sigma_{tot} = \sqrt{n\sigma^2}$

Se ho da calcolare la P di due eventi contemporanei uso l'intersezione e quindi il prodotto tra le due P