Formulario - Prova Intermedia

Alice Giampino

October 2024

Questo formulario si intende come strumento per il ripasso per sostenere la prova intermedia.
Non sarà possibile portare il formulario in sede d'esame.
Il formulario NON sostituisce lo studio del materiale completo!

1 Probabilità

- **Regola moltiplicativa generale**: se una operazione può essere fatta in n_1 modi e se per ciascuno di essi una seconda operazione può essere fatta in n_2 modi, se inoltre per ciascuna di queste prime due, una terza può essere fatta in n_3 modi (e cosi via), allora la sequenza di k operazioni può essere fatta in (modi)

$$n_1 \cdot n_2 \cdot \ldots \cdot n_k$$

1.1 Permutazioni

- **Permutazione**: è una sequenza ordinata di un insieme di oggetti (o di una parte di esso) senza ripetizione

$$n!$$

- **Permutazioni di n oggetti <u>scelti r alla volta</u>**:

$$\frac{n!}{(n-r)!}$$

- **Permutazioni di n elementi <u>in k categorie</u>**: tutti gli elementi scelti senza ripetizione, ma a gruppi. Si parla anche di partizione in sottoinsiemi.

$$\frac{n!}{n_1!n_2!\ldots n_k!}$$

1.2 Combinazioni

Dato un insieme di n elementi, calcolare il numero di sottoinsieme di ampiezza (classe) r che è possibile ottenere, senza ripetizione, e considerando che l'ordine degli elementi non è rilevante.

- **Combinazioni senza ripetizione**:

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Tipo	**Conta l'ordine?**	**Ripetizioni?**	**Formula**
Permutazione di n elementi	SI	NO	$n!$
Permutazione di r elementi presi da n	SI	NO	$\frac{n!}{(n-r)!}$
Permutazione di k gruppi	SI (ma no dentro i gruppi)	NO	$\frac{n!}{(n_1!\cdot n_2!\cdot\ldots\cdot n_k!}$
Combinazioni di r elementi presi da n	NO	NO	$\binom{n}{r}$

1.3 Regole delle probabilità

- **Regola dell'evento complementare**:
$$P(\bar{A}) = 1 - P(A)$$

- **Regola Additiva**: La probabilità dell'unione di due eventi che non sono mutualmente esclusivi è
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
Se gli eventi sono mutualmente esclusivi:
$$P(A \cup B) = P(A) + P(B)$$

- **Probabilità Condizionata**: la probabilità di un evento, dato che un altro evento si è verificato
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- **Regola moltiplicativa per due eventi A e B**:
$$P(A \cap B) = P(A|B)P(B)$$
$$P(A \cap B) = P(B|A)P(A)$$
Qualora A e B siano mutualmente esclusivi, allora $P(A \cap B) = 0$.

- **Indipendenza**:
Due eventi sono statisticamente indipendenti se e solo se:
$$P(A \cap B) = P(A)P(B)$$
Se A e B sono indipendenti, allora
$$P(A|B) = P(A) \text{ per } P(A) > 0$$
$$P(B|A) = P(B) \text{ per } P(B) > 0$$

- **Teorema della probabilità totale**:
$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} p(A|B_i)P(B_i)$$

- **Teorema di Bayes**:
$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{j=1}^{k} P(B_j \cap A)} = \frac{P(A|B_r)P(B_r)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)} \quad r = 1, \ldots, k$$

2 Variabili casuali

Una variabile aleatoria è una funzione che associa un numero reale a ciascun elemento dello spazio campionario.

- **Spazio campionario discreto**: Uno spazio campionario contenente un numero di elementi finito o una sequenza infinita di elementi che possono essere associati a numeri interi è chiamato spazio campionario discreto.

- **Spazio campionario continuo**: Uno spazio campionario contenente un numero infinito di elementi pari al numero di punti di un segmento è chiamato spazio campionario continuo.

- **Distribuzione di Probabilità discrete**: : l'insieme delle coppie ordinate $(x, f(x))$ è una funzione di probabilità, funzione di massa, o distribuzione di probabilità della variabile casuale discreta X se, per ogni possible esito x:
  1. $f(x) \geq 0$;
  2. $\sum_x f(x) = 1$;
  3. $P(X = x) = f(x)$.

- **Funzione di ripartizione (o cumulativa)**: a probabilità che X non superi il valore x
$$F(x) = P(X \leq x)$$
Relazione con funzione di probabilità:
$$F(x) = \sum_{t \leq x} f(t)$$

- **Funzione di Densità di Probabilità**: : l'insieme delle coppie ordinate $(x, f(x))$ è una funzione di densità di probabilità di una variabile casuale continua X se, per ogni possible esito x sull'asse dei numeri reali
  1. $f(x) \geq 0$ per ogni $x \in \mathbb{R}$;
  2. $\int_{-\infty}^{\infty} f(x) = 1$;
  3. $P(a < X < b) = \int_a^b f(x)dx =$ F(b) - F(a)
- **Funzione di ripartizione (o cumulativa)**: a probabilità che X non superi il valore x

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

$$f(x) = \frac{dF(x)}{dx}$$

- **Valor atteso, varianza e covarianza**:
  - Se X è una costante a (variabile casuale degenere), $E(X) = a$ e $V(X) = 0$.
  - X discreta:

$$E(X) = \mu = \sum_x x \cdot f(x)$$

$$V(X) = \sigma^2 = \sum_x (x - \mu)^2 \cdot f(x)$$

  - X continua:

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

$$V(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx$$

    N.B.: gli estremi dell'integrale dipendono dal supporto.
  - Formula alternativa varianza:

$$V(X) = E(X^2) - E(X)^2$$

    Scarto quadratico medio:

$$\sigma = \sqrt{\sigma^2}$$

    N.B.: La varianza NON può mai essere negativa!
  - DISTRIBUZIONI DOPPIE: valgono le formule precedenti MA con doppia sommatoria e doppio integrale!

$$E(XY) = \sum_x \sum_y xyf(x, y)$$

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy$$

  - Coviarianza:

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y$$

    X, Y discrete

$$\sigma_{XY} = Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X)(y - \mu_Y)f(x, y)$$

    X, Y continue

$$\sigma_{XY} = Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)dxdy$$

  - Correlazione:

$$\rho_{XY} = Corr(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

    $\rho_{XY} = 0$ NO relazione lineare X e Y;
    $\rho_{XY} > 0$ relazione lineare positiva X e Y;
    $\rho_{XY} < 0$ relazione lineare negativa X e Y.
  - Funzioni lineari:

$$E(aX \pm bY) = aE(X) \pm bE(Y)$$

$$V(aX \pm bY) = a^2V(X) + b^2V(Y) \pm 2abCov(X, Y)$$

3 Variabili casuali DISCRETE

3.1 Binomiale

Considera solo due risultati (l' esperimento prevede due soli esiti), cioè successo o insuccesso in n prove indipendenti. Probabilità di x successi in n prove, con probabilità di successo p in ogni prova:

$$P(X) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{(x-n)!x!} p^x (1-p)^{n-x}$$

- x = numero di successi nel campione, ($x = 0, 1, 2, \ldots, n$)
- n = dimensione del campione (numero di prove o osservazioni)
- p = probabilità di successo.

$$E(X) = n \cdot p$$
$$V(X) = n \cdot p \cdot (1-p)$$

RStudio:

- **dbinom(x, n, p)** restituisce la probabilità dell'esperimento binomiale per n ripetizioni di un esperimento con probabilità di successo p;
- **pbinom(x, n, p)** restituisce la probabilità cumulativa in x dell'esperimento binomiale per n ripetizioni di un esperimento con probabilità di successo p.

3.2 Binomiale Negativa

Consideriamo un esperimento binomiale dove però le prove dell'esperimento verranno ripetute fino al raggiungimento di un numero fissato di successi. La probabilità di interesse è data dall'ottenimento del $k-$esimo successo alla $x-$esima prova. Questo esperimento è chiamato binomiale negativo.

Se ripetute prove indipendenti risultano in un successo con probabilità p e in un insuccesso con probabilità $q = 1 - p$, allora la distribuzione di probabilità della variabile casuale X, data dal numero di prove necessarie per il verificarsi del $k-$esimo successo, è:

$$P(X) = \binom{x-1}{k-1} p^k q^{x-k}$$
$$E(X) = \frac{k}{p}$$
$$V(X) = \frac{k(1-p)}{p^2}$$

RStudio:

- **dnbinom(x-k, k, p)**, restituisce la probabilità di ottenere il $k-$successo in un esperimento di ripetute prove indipendenti alla x-esima prova, data una probabilità pari a p.
  N.B.: dnbinom(,,) richiede come primo valore il numero di insuccessi, quindi $x - k$.;
- **pnbinom(x-k, k, p)** restituisce la probabilità di ottenere il $k-$successo in un esperimento di ripetute prove indipendenti alla x-esima prova o ad una qualsiasi prova precedente, data una probabilità pari a p.

3.3 Geometrica

Consideriamo un caso speciale di una distribuzione binomiale negativa in cui $k = 1$. Quello che si ottiene è una distribuzione di probabilità per il numero di prove necessarie affinchè avvenga il primo successo.

Data una serie di prove ripetute e indipendenti, la probabilità che si verifichi un successo è p e la probabilità di insuccesso $q = 1 - p$, allora la distribuzione di una variabile casuale X, che rappresenta il numero di prove affinchè si verifichi il primo successo, è:

$$P(X) = pq^{x-1}$$
$$E(X) = \frac{1}{p}$$
$$V(X) = \frac{1-p}{p^2}$$

RStudio:

- **dgeom(x, p)** restituisce la probabilità di ottenere il primo successo in un esperimento di ripetute prove indipendenti alla x-esima prova, data una probabilità pari a p;
- **pgeom(x, p)** restituisce la probabilità di ottenere il primo successo in un esperimento di ripetute prove indipendenti alla x-esima prova o ad una qualsiasi prova precedente, data una probabilità pari a p.

3.4 Ipergeometrica

Probabilità di x successi in n prove in un campione estratto da una popolazione finita di dimensione N (con N piuttosto piccolo). Campione estratto senza reintroduzione. I risultati delle prove sono quindi dipendenti.

$$P(X) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

- N è il numero totale di elementi nella popolazione;
- k è il numero di elementi di successo nella popolazione;
- n è il numero di elementi estratti senza sostituzione;
- x è il numero di elementi di successo nel campione.

$$E(X) = n\frac{k}{N}$$
$$V(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right)$$

RStudio:

- **dhyper(x, m, n, k)** restituisce la probabilità di avere x successi in k estrazioni dipendenti, dato m successi e n insuccessi in popolazione;
- **phyper(x, m, n, k)** restituisce la probabilità di avere x o meno di x successi in k estrazioni dipendenti, dato m successi e n insuccessi in popolazione.

N.B.: con la notazione usata da noi **dhyper(x, k, N-k, n)**

3.5 Poisson

La distribuzione di probabilità della variabile casuale X di tipo Poisson, che rappresenta il numero di esiti che si verificano in uno specifico intervallo di tempo (o in una specifica regione) indicato con t, è:

$$P(X, \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$
$$E(X) = V(X) = \lambda t$$

RStudio:

- **dpois(x, lambda)** restituisce la probabilità di ottenere x eventi in un intervallo unitario di tempo (o di spazio) in un processo di Poisson con parametro λ.
- **ppois(x, lambda)** restituisce la probabilità di ottenere x eventi o un numero inferiore a x, in un intervallo unitario di tempo (o di spazio) in un processo di Poisson con parametro λ.
- **Approssimazione binomiale con Poisson**: Sia X una variabile casuale binomiale con distribuzione di probabilità dbinom(x, n, p). Quando $n \to \infty$, $p \to 0$ e $np \overset{n\to\infty}{\to} \mu$ rimane costante, allora

  $$dbinom(x, n, p) \to dpois(x, np)$$

  Solo quando n è grande e p è piccolo!

4 Variabili casuali CONTINUE

4.1 Uniforme Continua

La distributione uniforme è la distribuzione di probabilità relativa ad una v.a. che assegna la stessa probabilità a tutti i suoi possibili valori. La funzione di densità della distribuzione Uniforme Continua nell'intervallo $[A, B]$, è:

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B \\ \\ 0, & \text{otherwise} \end{cases}$$

- $f(x)$ = valore della funzione di densità a qualunque valore x nel supporto
- A = valore minimo di x con $f(x) > 0$
- B = valore massimo di x con $f(x) > 0$

$$E(X) = \frac{A+B}{2}$$
$$V(X) = \frac{(B-A)^2}{12}$$

RStudio:

- **dunif(x, A, B)** restituisce la densità, valutata in x, di una v.c. uniforme definita nel supporto $[min = A, max = B]$
  NOTA: dunif(x) restituisce la densità per una uniforme in [0,1]
  NOTA 2: La densità sarà la stessa per tutti i valori di x compresi in [A, B]
- **punif(x, A, B)** restituisce la densità dei valori inferiori o pari a x, per una v.c. uniforme definita nel supporto $[min = A, max = B]$

4.2 Normale

- Forma campanulare
- Simmetrica
- Media, mediana e moda coincidono

La tendenza centrale è determinata dalla media, μ. La variabilità è determinata dallo scarto quadratico medio, σ. La variabile aleatoria ha un campo di variazione teoreticamente infinito: $(-\infty, +\infty)$
La funzione di densità di probabilità normale è:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

- e = la costante matematica approssimata da 2.71828
- π = la costante matematica approssimata da 3.14159
- μ = parametro relative alla tendenza centrale
- σ = parametron relative allo scarto quadratico medio
- x = qualunque valore della variabile continua, $-\infty < x < \infty$

RStudio:

- **dnorm(x, μ, σ)** restituisce la densità, valutata in x, di una v.c. Normale con parametri media = μ e scarto = σ.
  NOTA: dnorm(x) restituisce la densità per una normale con media 0 e varianza 1.
- **curve(dnorm(x, μ, σ), from, to)** permette di rappresentare graficamente la curva normale per valori compresi tra from (il minimo) e to (il massimo).

La probabilità relativa ad un intervallo di valori è misurata dall'area sottesa alla curva

$$P(x_1 < X < x_2) = F(x_2) - F(x_1)$$

dove $F(x_2) = P(X < x_2)$ e $F(x_1) = P(X < x_1)$ (potete usare la funzione R: **pnorm(x, μ, σ)**).

4.2.1 La distribuzione Normale Standardardizzata

Un importante caso speciale di trasformazione $g(x)$ di variabili casuali permette di ottenere la variabile aleatoria standardizzata

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

la quale ha media $\mu_Z = 0$ e varianza $\sigma_Z = 1$. Da cui

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

$$P(x_1 < X < x_2) = P\Big(\frac{x_1 - \mu_X}{\sigma_X} < \frac{X - \mu_X}{\sigma_X} < \frac{x_2 - \mu_X}{\sigma_X}\Big) = P(z_1 < Z < z_2)$$

RStudio:

- **pnorm(x, μ, σ)** restituisce la densità dei valori pari o inferiori ad x, di una v.c. Normale con parametri media $= \mu$ e scarto $= \sigma$.
  NOTA: pnorm(x) restituisce la densità cumulata in x per una normale con media 0 e varianza 1 (quindi restituisce il valore di z)
- **qnorm(p)** è la funzione che dato il valore dell'area sotto la curva normale standardizzata (p) a sinistra di un valore z, restituisce quel valore z

4.2.2 Approssimazione della Distribuzione Binomiale con la Distribuzione Normale

Ricorda la distribuzione binomiale:

- n prove indipendenti
- probabilità di successo in ogni prova = p

Variabile aleatoria X:

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

La forma della distribuzione binomiale è "approssimativamente" normale, se:
• se n è grande (e p può assumere qualsiasi valore superiore a 0 e inferiore a 1)
• se n è "moderatamente" piccolo, e p è vicina a 0,5.
• In generale, la normale è una buona approssimazione per la binomiale quando $np(1–p) > 5$ e $np > 5$

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{np(1 - p)}}$$

4.3 Gamma & Esponenziale

La variabile casuale continua X ha una distribuzione gamma con parametri $\alpha > 0$ e $\beta > 0$ se la sua funzione di densità è data da:

$$f(x;\beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$E(X) = \alpha\beta$$
$$V(X) = \alpha\beta^2$$

La variabile casuale continua X ha una distribuzione esponenziale con parametro $\beta > 0$, se la sua funzione di densità è data da:

$$f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$E(X) = \beta$$
$$V(X) = \beta^2$$

RStudio:

- **dgamma(x, shape, rate, scale)** restituisce il valore della funzione di densità di una variabile casuale Gamma, dove shape = α, rate =$1/\beta$, scale = β.
  Notare che si specifica solo uno tra rate e scale
  Notare inoltre che se rate e scale non vengono inseriti, la funzione assume una gamma con = β.
- **exp(x, rate)** restituisce il valore della funzione di densità di una v.c. esponenziale con argomento rate = $1/\beta = \lambda$
  Notare che il parametro λ è il parametro di Poisson.
- **pexp(x, rate)** il valore della relativa densità cumulata