Domande dell'orale di Calcolo delle Probabilità e Statistica Matematica

Filippo $\mathcal{L}$ Troncana (ovviamente su domande di Luigi Amedeo Bianchi)

A.A. 2023/2024, appello di Giugno

Disclaimer: Gli interrogati sono raccolti nell'ordine in cui mi sono state riferite le domande, non necessariamente nell'ordine di interrogazione. Aggiungerò le risposte, o almeno un tentativo.

1

Domande

1. Sia $X \sim \exp(\lambda)$ con $\lambda > 0$ e sia $Y = g(X)$ con $g(x) = (1-x)^2$. Quali sono legge e media di Y?
2. Intervalli di confidenza.
3. Introduzione alla Statistica.
4. Media campionaria e dimostrazione della correttezza.

Risposte

1. Piuttosto che usare la formula, si può usare la definizione di legge, notando innanzitutto che $P(Y < 0) = 0$, quindi esaminiamo $y \geq 0$:

   $$F_Y(y) = P(Y \leq y) = P((1-X)^2 \leq y) = P(|1-X| \leq \sqrt{y}) = P(X \geq 1-\sqrt{y} \wedge X \leq 1) + P(X \leq 1+\sqrt{y} \wedge X > 1)$$

   Dunque otteniamo

   $$F_Y(y) = \mathbf{1}_{y\geq 0} \int\limits_{1-\sqrt{y}}^{1+\sqrt{y}} f_X(x)\ \mathrm{d}x = \begin{cases} 0 & y < 0 \\ \int\limits_{0}^{1+\sqrt{y}} \lambda e^{-\lambda x}\ \mathrm{d}x & y > 1 \\ \int\limits_{1-\sqrt{y}}^{1+\sqrt{y}} \lambda e^{-\lambda x}\ \mathrm{d}x & y \in [0,1] \end{cases} = \begin{cases} 0 & y < 0 \\ F_X(1+\sqrt{y}) - F_X(1-\sqrt{y}) & y \geq 0 \end{cases}$$

   Per calcolare la media usiamo

   $$E[g(X)] = \int\limits_{\mathbb{R}} g(x) f_X(x)\ \mathrm{d}x = \int\limits_0^{+\infty} (1-x)^2 \lambda e^{-\lambda x}\ \mathrm{d}x = \int\limits_0^{+\infty} x^2 \lambda e^{-\lambda x}\ \mathrm{d}x - \int\limits_0^{+\infty} 2x\lambda e^{-\lambda x}\ \mathrm{d}x + \int\limits_0^{+\infty} \lambda e^{-\lambda x}\ \mathrm{d}x$$

   Che farà qualche numero reale che non ho voglia di calcolare, immagino.

2. Immaginiamo di avere una popolazione Ω di cui vogliamo trovare un qualche parametro $\theta \in \mathbb{R}$. Estraiamo un campione $X = \{X_i\}_1^n$ e otteniamo una stima $\Theta[X]$. Un intervallo di confidenza bilaterale di livello $1-\alpha$ con $\alpha \in]0,1[$ è un intervallo della forma $I_\alpha = [\Theta[X] - A_\alpha, \Theta[X] + B_\alpha]$ tale che $P(\theta \in I_\alpha) = 1-\alpha$. Se A_α o B_α sono $+\infty$, parliamo di intervallo unilaterale rispettivamente sinistro e destro.

3. TODO

4. Sia $X := \{X_i\}_1^n$ un campione di variabili aleatorie (che assumeremo indipendenti e identicamente distribuite) da una popolazione di media μ sconosciuta. La ***media campionaria*** di X indicata con $\hat{\mu}_n$ è lo stimatore

   $$\hat{\mu}_n = E\left[\frac{1}{n}\sum_{x=1}^n X_i\right] \quad \text{che per la linearità della speranza è uguale a} \quad \frac{1}{n}\sum_{i=1}^n E[X_i]$$

   Ed è corretto come conseguenza della legge debole dei grandi numeri, che ci dice che

   $$\hat{\mu}_n \xrightarrow{n\to+\infty} \mu$$

2

Domande

1. Legge debole dei grandi numeri: enunciato, dimostrazione, significato e applicazione.
2. Variabile aleatoria binomiale: definizione, legge e media (con dimostrazione della media).

Risposte

1. **Teorema 2.0.1: Legge debole dei grandi numeri**

   Sia $\{X_i\}_{\mathbb{N}}$ una successione di variabili aleatorie identicamente distribuite e indipendenti, ciascuna di media μ e varianza σ^2 e sia $S_n := X_1 + ... + X_n$. Allora abbiamo che per ogni $\varepsilon > 0$ si ha

   $$\lim_{n\to+\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0$$

   Dimostrazione

   Sfruttando il fatto che $E[S_n/n] = \mu$ e $\mathrm{Var}[S_n/n] = \sigma^2/n$ e la disuguaglianza di Chebychev, fissiamo $\varepsilon > 0$ e otteniamo

   $$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = P\left(\left|\frac{S_n}{n} - E\left[\frac{S_n}{n}\right]\right| > \varepsilon\right) \leq \frac{\mathrm{Var}\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n\to+\infty} 0$$

   □

   L'idea è che in una successione di variabili aleatorie, $S_n - n\mu = o(n)$, ma non è detto che $S_n - n\mu \to 0$, attenzione! Ci dà un'idea della frequenza dei risultati, non del risultato che dobbiamo aspettarci al prossimo tentativo: se finora sono uscite $42'000'000$ teste e 42 croci, non è per nulla detto che esca croce (anche se potremmo iniziare a farci qualche domanda sulla qualità della moneta).

2. **Definizione 2.0.1: Variabile aleatoria binomiale**

   Sia $\{X_i\}_1^n$ una successione di variabili aleatorie bernoulliane di parametro $p \in [0, 1]$.
   Una variabile aleatoria X si dice ***binomiale*** e si indica $X \sim \mathrm{bin}(n, p)$ se

   $$X = \sum_{i=1}^{n} X_i$$

   La legge di X è

   $$\varphi_X(k) = \mathbf{1}_{\{0\to n\}}(k) \binom{n}{k} p^k (1-p)^{n-k}$$

   Per la linearità della speranza, abbiamo che

   $$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = np$$

3

Domande

1. Intervalli di confidenza: costruzione, interpretazione, esempi, proprietà.
2. In quanti modi si possono suddividere nove ubriachi in tre taxi con tre persone ciascuno?

Risposte

1. TODO

2. Innanzitutto possiamo immaginare 9! al numeratore, in quanto possiamo immaginare che i nostri passeggeri siano rappresentati da $\{1 \to 9\}$ e disporli in tre taxi sarebbe come immaginare una parola fatta di ciascuno di questi simboli, ma c'è un problema: la disposizione 123|456|789 è equivalente a quella 123|789|456 e a quella 132|456|789 ad esempio, dunque dobbiamo anche dividere per i possibili ordini dei taxi e i possibili ordini dei passeggeri, ottenendo $9!/(3!)^2$.

4

Domande

1. Test statistici: definizione, obiettivi, proprietà ed esempi.
2. In un roster di quaranta giocatori numerati, qual è la probabilità che in una squadra da undici giocatori non ci siano due giocatori con numeri consecutivi?

Risposte

5

Domande

1. Successioni di variabili aleatorie.
2. Teorema centrale del limite.

Risposte

1. TODO
2. Allora, prepariamo gli ingredienti che ci servono per questo Teorema

Definizione 5.0.1: Funzione generatrice dei momenti

Sia X una variabile aleatoria con densità f_X e supporto R_X. Si dice ***funzione generatrice dei momenti*** di X la funzione $M_X(t) = E[e^{tX}]$. In particolare vale:

$$M_X(t) = \sum_{x \in R_X} e^{tx} f_X(x) \quad \text{se } X \text{ discreta}, \qquad M_X(t) = \int_{R_X} e^{tx} f_X(x)\, \mathrm{d}x \quad \text{se } X \text{ assolutamente continua.}$$

Lemma 5.0.1

Siano due variabili aleatorie indipendenti X e Y. Valgono i seguenti:

- Se $Z = aX + bY + c$ si ha $M_Z(t) = M_X(at)M_Y(bt)e^{ct}$.
- Se esiste un aperto non vuoto $U \subset \mathbb{R}$ tale che $M_X(t) = M_Y(t)$ per ogni $t \in U$, allora $X \sim Y$
- Vale $\mathrm{d}^n M_X(0) = E[X^n]$, quindi in particolare $M_X'(0) = E[X]$ e se $E[X] = 0$, si ha $M_X''(0) = \mathrm{Var}(X)$.
- Se $X \sim \mathcal{N}(0,1)$ allora $M_X(t) = e^{\frac{x^2}{2}}$

Ora possiamo enunciare il nostro

Teorema 5.0.1: Teorema centrale del limite

Sia $\{X_i\}_\mathbb{N}$ una successione di variabili aleatorie indipendenti ognuna con media μ e varianza σ^2. Si ha

$$S_n^* = \frac{S_n - n\mu}{\sqrt{\sigma^2 n}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1), \quad \text{ovvero,} \quad P\left(\frac{S_n - n\mu}{\sqrt{\sigma^2 n}} \le x\right) \xrightarrow[n\to+\infty]{\text{pt. per } x\in\mathbb{R}} \Phi(x)$$

Dimostrazione

Consideriamo com'è fatta $M_{S_n^*}(t)$. Possiamo riscrivere

$$S_n^* = \frac{\sum_{i=0}^n (X_i - \mu)}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=0}^n \frac{X_i - \mu}{\sigma} =: \frac{1}{\sqrt{n}} \sum_{i=0}^n X_i^*$$

Dunque per il lemma di cui sopra

$$M_{S_n^*}(t) = M_{X^*}^n \left(\frac{t}{\sqrt{n}} \right) = \left[M_X \left(\frac{t}{\sigma\sqrt{n}} \right) e^{-\frac{\mu t}{\sigma}} \right]^n$$

Centrando in $t = 0$ la serie di Taylor di $M_X(t)$ otteniamo che

$$M_X(t) \sim M_X(0) + M_X'(0)t + M_X''(0)\frac{t^2}{2} = 1 + \mu t + \sigma^2 \frac{t^2}{2}$$

Possiamo assumere che $\mu = 0$ senza perdita di generalità e otteniamo

$$M_{S_n^*}(t) = M_X^n \left(\frac{t}{\sigma\sqrt{n}} \right) \sim \left(1 + \frac{\sigma^2}{2} \frac{t^2}{\sigma^2 n} \right)^n \xrightarrow[n \to +\infty]{\text{in un intorno di } 0} e^{\frac{t^2}{2}}$$

Dunque hanno la stessa MGF in un intorno di 0 e pertanto sono identicamente distribuite.

□

6

Domanda

Variabili aleatorie poissoniane: definizioni, proprietà, esempi, intervallo di confidenza per una popolazione di poissoniane.

Risposta

Definizione 6.0.1: Variabile aleatoria poissoniana

Sia X una variabile aleatoria tale che per un certo $\mu \in \mathbb{R}^+$ abbia funzione di ripartizione

$$\varphi_X(x) = \mathbf{1}_{\mathbb{N}} \frac{\mu^x}{x!} e^{-\mu}$$

Allora X si dice ***poissoniana*** di media μ e si scrive $X \sim \text{pois}(\mu)$

Una poissoniana di media μ ha speranza μ (wow!), le poissoniane sono riproducibili e si ha $\text{pois}(\mu) + \text{pois}(\lambda) = \text{pois}(\mu + \lambda)$ e si ha

Proposizione 6.0.1

Data una successione $\{p_n\}_{\mathbb{N}}$ tale che $np_n \to \mu \in \mathbb{R}^+$, la successione $\{\text{bin}(n, p_n)\}_{\mathbb{N}}$ converge in legge a $\text{pois}(\mu)$, ovvero:

$$\binom{n}{k} p_n^k (1 - p_n)^k \xrightarrow[n \to +\infty]{\text{pt. per } k \in \mathbb{N}} \frac{\mu^k}{k!} e^{-\mu}$$

Un esempio di poissoniana potrebbe essere "quante botte mi darà LAB se uso Fubini per il volume di un cubo sapendo che in media ne dà 42?"

7

Domanda

Variabili aleatorie geometriche: definizione, proprietà, varianza e media con dimostrazione.

Risposta