Prova in Itinere
Calcolo delle probabilità e Statistica (12 crediti)
7 gennaio 2010

Esercizio 1 *Alfa e Beta giocano una serie di partite a freccette; ad ogni turno la probabilità di vittoria di Alfa è $p = 0.45$ mentre la probabilità di pareggio è $r = 0.1$. In ogni turno, se c'è un vincitore, guadagna un euro dell'avversario. Il capitale iniziale di Alfa è 3 euro, quello di Beta è 5 euro.*

a. *Qual è la probabilità che Alfa vinca tutti i soldi di Beta?*

b. *Qual è il capitale medio di Alfa al termine del gioco?*

c. *Quante partite devono disputare, in media, per arrivare al termine del gioco?*

Soluzione Il gioco si può modellizzare con una passeggiata casuale simmetrica, avente $p = q = 0.45$, $r = 0.1$; il capitale iniziale di Alfa è $x = 3$, mentre le barriere sono 0 e $b = 8$. Si applicano le leggi di Wald e si ottiene: **a.** la probabilità di vittoria di Alfa è $\frac{x}{N} = 3/8$; **b.** in media, il capitale finale di Alfa è 3; **c.** il numero medio di partite da giocare risulta $\mathbb{E}[N] = \frac{x(b-x)}{1-r} = 50/3 \simeq 16.67$

Esercizio 2 *Data una catena di Markov su uno spazio degli stati finito E, si definisca il concetto di periodo per uno stato. Si mostri che se una catena è irriducibile allora tutti gli stati hanno lo stesso periodo. Si fornisca infine un esempio di catena avente periodo 3.*

Soluzione Dato uno stato x, definiamo *orbita* di x l'insieme dei tempi in cui è positiva la probabilità di ritorno: $O(x) = \{n \geq 1 : p_{xx}(n) > 0\}$. Il periodo di uno stato è pari al massimo comun denominatore degli elementi dell'orbita: $d_x = MCD\{O(x)\}$.

Siano x e y due stati di una catena irriducibile: allora esiste un percorso $x \to y \to x$; possiamo supporre che il percorso più breve che congiunge $x \to y$ abbia lunghezza m e il più breve che congiunge $y \to x$ abbia lunghezza n. Necessariamente $n + m$ deve essere un multiplo di d_x e di d_y; inoltre $n + m + \alpha d_y$ è un multiplo di d_x per ogni $\alpha \geq 1$, da cui si osserva che deve essere d_y un multiplo di d_x. La costruzione è simmetrica, quindi anche d_x deve essere un multiplo di d_y e quindi devono coincidere.

Un esempio di catena di periodo 3 è: $A \to B \to C \to A$.

Esercizio 3 *Ogni anno Elena sceglie di passare le sue vacanze estive o ad Alassio (A) o a Bonassola (B) o a Corniglia (C), ma ogni anno sceglie di andare in una località diversa da quella dell'anno precedente. Elena ha preparato 12 biglietti riportanti il nome di una località: Alassio è scritto su 2 biglietti, Bonassola su 4 e Corniglia su 6. Ogni anno Elena pone in una scatola tutti i biglietti col nome di una località non visitata l'estate precedente e ne estrae uno, dove andrà a passare le vacanze. Si indichi con X_n la località visitata l'n-esimo anno; si mostri che $\{X_n\}$ forma una catena di Markov omogenea;*

a. *Scrivere la matrice di transizione della catena.*

b. *Disegnare il grafo della catena.*

c. *Classificare gli stati della catena.*

d. *Trovare la distribuzione stazionaria se esiste.*

Soluzione Osserviamo che vale la legge di Chapman-Kolmogorov $\pi_{n+1} = \pi_n T$ quindi il sistema è una catena di Markov omogenea. La matrice di transizione T vale

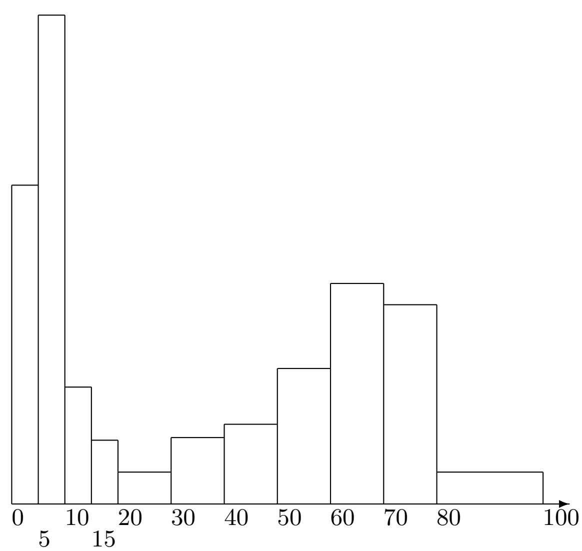
$$T = \begin{pmatrix} 0 & 4/10 & 6/10 \\ 2/8 & 0 & 6/8 \\ 2/6 & 4/6 & 0 \end{pmatrix}$$

La catena è irriducibile; gli stati sono ricorrenti (positivi, dato che la catena è a stati finiti) e aperiodici, quindi la catena è regolare e ergodica. La distribuzione stazionaria (coincide con la distribuzione limite) è $\pi_{inv} = (\frac{5}{22}, \frac{8}{22}, \frac{9}{22})$.

Esercizio 4 *La tabella seguente riporta il numero di pedoni maschi morti in incidenti stradali in Inghilterra nel 1922, divisi per classi di età.* **a.** *si costruisca un istogramma relativo alla tabella,* **b.** *si trovino media e mediana campionaria dell'età al decesso;* **c.** *si calcolino i quartili e IQR.*

Età	*0-5*	*5-10*	*10-15*	*15-20*	*20-30*	*30-40*	*40-50*
Popolazione	*120*	*184*	*44*	*24*	*23*	*50*	*60*
		50-60	*60-70*	*70-80*	*80-100*		
		102	*167*	*150*	*49*		

Soluzione



$N = 973$, media campionaria $\bar{x} = 40.9$, mediana $x_{487} = \frac{487-445}{505-405} * 10 + 40 = 47$; i quartili sono $Q_1 = x_{244} = 8.4$ e $Q_3 = x_{730} = 67.4$ e quindi $IQR = 59$

Esercizio 5 *Come è noto, possiamo supporre che il numero di incidenti che avvengono su un certo percorso, a parità di altri elementi di rischio, sia distribuito con legge di Poisson di parametro* λ. *Si sono osservati i seguenti dati:*

4	*6*	*0*	*5*	*2*	*1*	*2*	*0*	*4*	*3*

Determinare una stima di massima verosimiglianza $\hat{\lambda}$ *per il parametro* λ.

Soluzione Il parametro λ di una distribuzione di Poisson coincide con la media. La stima di massima verosimiglianza per la media di una distribuzione coincide con la media campionaria; risulta allora $MLE(\lambda) = \bar{x} = 2.7$

Esercizio 6 *Sia $\{X_k\}_{k=1,\dots,n}$ un campione ottenuto da una popolazione $\mathcal{N}(\mu,\sigma^2)$. Si chiede di spiegare come si possa ottenere un intervallo di confidenza per la media μ al livello α sapendo che la varianza σ^2 sia nota. Si chiarisca in che senso la conoscenza di σ^2 porti a un intervallo di confidenza migliore di quello che si ha nel caso σ^2 non sia noto.*

Soluzione Dato che $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ si ha

$$\mathbb{P}\left(\left|\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}\right| < \phi_{(1+\alpha)/2}\right) = \alpha)$$

ossia, invertendo rispetto alla media μ,

$$\mathbb{P}\left(\mu \in \left(\bar{X} - \sqrt{\tfrac{\sigma^2}{n}}\,\phi_{(1+\alpha)/2}, \bar{X} + \sqrt{\tfrac{\sigma^2}{n}}\,\phi_{(1+\alpha)/2}\right)\right) = \alpha$$

Se ora sostituiamo alla variabile aleatoria $\bar{X}$ il valore osservato $\bar{x}$ possiamo dire che, con un livello di confidenza α, deve essere

$$\mu \in \left(\bar{x} - \sqrt{\tfrac{\sigma^2}{n}}\,\phi_{(1+\alpha)/2}, \bar{x} + \sqrt{\tfrac{\sigma^2}{n}}\,\phi_{(1+\alpha)/2}\right)$$

Nel caso la varianza non sia nota, possiamo sostituirla con la variabile aleatoria varianza campionaria S^2; in questo caso si ha $\frac{\bar{X}-\mu}{\sqrt{S^2/n}} \sim t(n-1)$. Osserviamo che la legge t di Student approssima dal sopra la legge gaussiana, quindi i suoi intervalli di confidenza (anche a parità di valori numerici) saranno più ampi.

Esercizio 7 *Si investiga la gittata di un certo tipo di mortaio. Vengono osservati i seguenti dati, espressi in metri*

2100	*1950*	*2043*	*2210*	*2018*	*1984*	*1992*
2218	*2152*	*2106*	*2072*	*2096*	*2244*	*1962*
1938	*1898*	*2103*	*2206*	*2007*	*1956*	

Assumendo che la distribuzione delle gittate sia normale, si determini

a. *un intervallo di confidenza al livello $\alpha = 98\%$ per la gittata media dei proiettili;*

b. *il più grande valore v che con il 95% di confidenza è inferiore alla gittata media indagata.*

Soluzione La media campionaria è $\bar{x} = 2062.75$, la varianza campionaria è $s^2 = 10887.57$, quindi l'intervallo di confidenza cercato è

$$\mu \in (\bar{x} - \sqrt{\frac{s^2}{n}}t_{0.99}(19), \bar{x} + \sqrt{\frac{s^2}{n}}t_{0.99}(19)) = (1996, 2129.5)$$

Un intervallo unilatero limitato inferiormente, al livello $\alpha = 95\%$, risulta $(2013.9, +\infty)$ quindi $v = 2013.9$