



Università degli Studi di Padova

Dipartimento di Matematica "Tullio Levi-Civita"

Laurea Triennale in Matematica

**PROCESSI DI GALTON-WATSON E IL TEOREMA
DI KESTEN-STIGUM**

Supervisore:
Prof. Alessandra Bianchi

Studente laureando:
Andrea Pizzolato
Matricola: 1229048

Anno Accademico 2021/2022
23 Settembre 2022

Indice

Introduzione	iii
1 Definizioni e risultati preliminari	1
1.1 Processi stocastici e catene di Markov	1
1.1.1 Martingale	3
1.2 Processi di Galton-Watson e loro rappresentazione tramite alberi	4
1.3 La funzione generatrice delle probabilità	6
1.4 L'albero di Ulam-Harris	7
1.4.1 Definizione e prime proprietà	7
1.4.2 Lo spazio degli alberi planari e gli alberi casuali	9
2 Risultati notevoli sui processi di Galton-Watson	11
2.1 Gli alberi planari e il cycle lemma	11
2.2 Processi di ramificazione e alberi di Bienaymé	15
2.2.1 Il teorema fondamentale dei processi di ramificazione .	17
2.3 Condizionamento per la finitezza degli alberi di Bienaymé . .	21
3 Processi di Galton-Watson con immigrazione e il teorema di Kesten-Stigum	27
3.1 Processi di Galton-Watson con immigrazione	27
3.2 Alberi spinali	32
3.3 Il Teorema di Kesten-Stigum	37
Conclusioni	43

Bibliografia

45

Introduzione

In questo lavoro vogliamo parlare dei *processi di Galton-Watson*, una particolare categoria di processi di ramificazione, di cui studieremo caratteristiche di base e comportamento asintotico. Nel caso che analizzeremo, questi processi puntano a modellizzare l'evoluzione di una determinata popolazione in linea maschile sotto l'ipotesi che il numero di figli di ogni individuo della popolazione sia indipendente da quello degli altri, e che sia modellizzabile tramite una variabile aleatoria applicabile ad ogni individuo. Tale problema, di discreta importanza storica, venne introdotto nella seconda metà dell'Ottocento nell'Inghilterra Vittoriana, dove molte famiglie nobiliari si interrogavano sulla probabilità di sopravvivenza del loro cognome, ragionando quindi solo relativamente alla linea maschile (considerare un solo sesso è assai importante per la modellizzazione del problema).

Nel 1874 Sir Francis Galton e il reverendo Henry William Watson pubblicarono un articolo intitolato “On the probability of the extinction of families” sul “*Journal of the Anthropological Institute of Great Britain and Ireland*” in cui condussero un'analisi su questi argomenti che portò a risultati fondamentali, analoghi a quelli presentati nel Capitolo 2 di questo lavoro. Tale modellizzazione si sviluppò indipendentemente dal lavoro per certi versi simile di Bienaymé, che pochi anni prima aveva affrontato un problema simile studiando i processi di ramificazione utilizzando una struttura come i grafi ad albero. Nel seguito di questa tesi coniugheremo le due idee, utilizzando gli *alberi di Bienaymé* come medium grafico per rappresentare al meglio lo scorrere del tempo, associando a ogni livello dell'albero una generazione.

Lo studio di questo soggetto continuò nel corso del tempo, principalmente con lo scopo di determinare condizioni relative al comportamento asintotico del modello, raffinando la descrizione data dal Teorema fondamentale dei processi di ramificazione e ottenendo svariati risultati che quantificano in modo preciso l'andamento del numero di individui ad ogni generazione all'avanzare del tempo. Un risultato fondamentale in questo senso è il *Teorema di Kesten-Stigum*, che delinea delle condizioni equivalenti per l'andamento asintotico del numero di individui all' n -esima generazione legandolo al valore atteso della variabile aleatoria che descrive la progenie di un singolo individuo ([5]). Tale Teorema è il risultato chiave di questa tesi, l'obiettivo che ci poniamo di dimostrare. Per riuscirci seguiamo un percorso che parte dalle basi e introduce gradualmente tutti i risultati e i costrutti necessari per la dimostrazione: nel Capitolo 1 vengono trattati gli strumenti di base necessari a parlare di processi di ramificazione, quali le definizioni di *processo stocastico* e quindi di filtrazione e *catena di Markov*, che permettono di definire le proprietà necessarie ad analizzare correttamente processi dove il tempo avanza discretamente (come per esempio le generazioni di una popolazione).

Approfondiamo inoltre una famiglia di processi stocastici di importanza fondamentale, le martingale, che sono processi "a media costante", che in questa tesi utilizzeremo principalmente per rappresentare la normalizzazione di processi stocastici: introdurremo inoltre un risultato di convergenza per le (sub)martingale che useremo per determinare il limite della martingala che descrive il rapporto tra il numero di individui nella generazione n , Z_n , e α^n , dove α rappresenta il valore atteso della progenie di un singolo individuo.

Definiremo nella sezione successiva i processi di Galton-Watson e la loro rappresentazione tramite alberi. Risolveremo successivamente il problema dell'enumerazione dei nodi tramite l'*albero di Ulam Harris*, che definisce un modo univoco di etichettatura dei nodi rispettoso delle generazioni. A questo punto introdurremo lo *spazio degli alberi planari*, che permette di rendere i grafi ad albero uno spazio misurabile grazie a un'opportuna scelta di σ -algebra e filtrazione appropriata. In questo capitolo inoltre descriveremo e dimostre-

remo le proprietà analitiche della funzione generatrice delle probabilità, uno strumento di cui ci avvarremo nella dimostrazione di risultati chiave di questa tesi quali il Teorema fondamentale dei processi di ramificazione.

Nel secondo capitolo iniziamo a parlare di risultati di base per la teoria dei processi di Galton-Watson, che costituiscono l'ossatura di tale teoria e/o sono di vitale importanza per dimostrare risultati più complessi: iniziamo introducendo il *processo depth-first queue*, una descrizione dell'albero che descrive quantitativamente (in unione con la nomenclatura dei nomi precedentemente introdotta) la parte del grafo ad albero in esame, e che dimostreremo contenere l'informazione sulla struttura completa dell'albero. Dimostriamo poi il *Cycle Lemma di Dwass*, una versione del Cycle Lemma che afferma che data una n -upla di valori interi a somma negativa $-r$ allora esistono solo r sottostringhe *ordinate* di lunghezza $i < n$ tali che la loro somma è maggiore di $-r$. La dimostrazione è interessante e si basa su una periodizzazione della n -upla e della definizione della somma parziale, e su riscritture intelligenti del problema.

Nella sezione successiva 2.2 introduciamo gli *alberi di Bienaymé*, un tipo di struttura che permette di trattare con approccio probabilistico lo spazio dei grafi ad albero, derivandone la distribuzione a partire dalla distribuzione della progenie di un singolo individuo; a questo punto è quasi naturale introdurre concetti quali *estinzione* e *sopravvivenza* di una popolazione, e il problema in esame è quindi quello di determinare le condizioni per cui tali eventi risultano essere probabilisticamente certi, possibili o impossibili. Il primo risultato che otteniamo è il *Teorema fondamentale dei processi di ramificazione*, che dice che la probabilità che una popolazione sopravviva è strettamente positiva se e solo se ogni individuo ha quasi certamente un figlio oppure se ha in valore atteso più di un figlio. Al fine di dimostrare questo teorema necessitiamo la *proprietà fondamentale dei processi di ramificazione*, che lega la probabilità di estinzione alla funzione generatrice delle probabilità introdotta nel Capitolo 1.

Nella sezione 2.3 affrontiamo il problema di quale sia la distribuzione di

probabilità di un albero di Bienaymé condizionato al fatto di essere finito, risultato che permette di dimostrare la Proprietà 2.3.1 che lega la probabilità che il grafo ad albero sia finito con cardinalità n con la probabilità che la somma di n v.a. distribuite come la progenie di un singolo individuo sia pari a $n - 1$. Tale proprietà può essere riscritta e dimostrata usando il Cycle Lemma di Dwass in modo creativo e a questo punto, note le condizioni necessarie per la sopravvivenza della popolazione e le relazioni che determinano quanto risulti numerosa una popolazione in caso di estinzione, possiamo muovere gli ultimi passi per poi affrontare e dimostrare il Teorema di Kesten-Stigum.

Nel Capitolo 3 iniziamo introducendo un fattore di immigrazione nel processo di Galton-Watson in esame, che tratteremo come un processo stocastico di cui ogni elemento corrisponde al numero di individui immigrati in una certa generazione. Il primo risultato rilevante di questo capitolo è un teorema, dimostrato da Hyde e Seneta nel 1970 ([4]) che lega il valore atteso del logaritmo della v.a. rappresentante il processo di immigrazione al $\limsup_n \frac{Z_n}{\alpha^n}$, determinando la finitezza di tale rapporto: come conseguenza di tale teorema otteniamo la proprietà che dice che $\frac{Z_n}{\alpha^n}$ converge quasi certamente ed è una submartingala (mentre nel caso di un processo di Galton-Watson senza immigrazione tale processo stocastico è una martingala).

Nella sezione successiva introduciamo il concetto di immigrazione negli alberi di Bienaymé tramite gli *alberi spinali*, alberi dove a ogni generazione si ha un "individuo" la cui distribuzione della progenie differisce da quella degli altri individui, ed è sempre positiva, quindi nel caso dello studio di una popolazione è una v.a. a valori nei numeri naturali. Tale successione di v.a. all'interno dell'albero determina una "spina" che corre dentro l'albero, e per descrivere la legge della v.a. introduciamo la *versione size biased* di una v.a. \widehat{X} , cioè una distribuzione tale per cui la probabilità che \widehat{X} assuma un certo valore è il valore atteso $\mathbb{E}(X \mathbf{1}_A(x))$ normalizzato rispetto al valore atteso della variabile X . In questo modo modellizziamo efficacemente la distribuzione del fenomeno di immigrazione. Costruiamo poi uno spazio misurabile per gli alberi spinali come fatto in precedenza per gli alberi di Bienaymé, spazio che

dimostriamo contenere lo spazio degli alberi di Bienaymé.

Dimostriamo ora la relazione che lega la misura di probabilità che si può costruire sullo spazio degli alberi di Bienaymé con processo di immigrazione \widehat{B}_μ (legata a quella che si costruisce sugli alberi spinali) e quella introdotta nel capitolo 1 sugli alberi di Bienaymé standard B_μ , di modo che si possa passare dall'una all'altra. Dopo aver ottenuto un risultato tecnico sulla derivata di Radon-Nykodim di \widehat{B}_μ rispetto a B_μ .

Nella sezione 3.3 trattiamo infine il punto principale di questa tesi, il *Teorema di Kesten-Stigum*, un risultato che descrive delle condizioni equivalenti per determinare la probabilità di estinzione di una popolazione in modo preciso, risultando quindi assai più potente del Teorema fondamentale (che diceva solo in quali casi l'estinzione risultasse certa): la prima condizione lega la probabilità di estinzione della popolazione alla probabilità che la variabile M , definita come il limite delle variabili aleatorie $\frac{Z_n}{\alpha^n}$, sia pari a 0; la seconda condizione equivalente richiede che il valore atteso di M sia pari a 1; la terza condizione, la più interessante, richiede che il valore atteso della variabile aleatoria $X \log^+ X$ sia finito. Questa condizione è la più utile poiché lega la probabilità di estinzione ad una v.a. legata alla progenie del singolo individuo, che è nota differentemente dalla distribuzione di M . La dimostrazione di tale teorema utilizza vari risultati visti precedentemente, oltre a un Teorema che afferma che quando una misura è assolutamente continua rispetto all'altra, dette \mathbb{P} e \mathbb{Q} tali misure, con X la derivata di Radon-Nykodim tra di loro, allora

$$\mathbb{Q} = X\mathbb{P} + \mathbf{1}_{X=\infty}$$

Si può quindi vedere come partendo da poche definizioni di base si riesca a costruire una teoria completa per lo studio del comportamento asintotico del nostro modello di popolazione arrivando a un risultato raffinato quale il T. di Kesten-Stigum, che lega il comportamento di un oggetto complesso come l'intera popolazione a quello di una variabile legata alla progenie del singolo individuo e quindi assai più semplice da calcolare.

Capitolo 1

Definizioni e risultati preliminari

1.1 Processi stocastici e catene di Markov

In questa sezione introduttiva vogliamo parlare di cosa sia un *processo stocastico*, un oggetto che utilizzeremo continuamente nel corso di questa tesi. Un processo stocastico è un oggetto matematico che vuole descrivere l'evoluzione di una quantità al variare di un parametro (solitamente il tempo) in modo stocastico, cioè non deterministico.

Definizione 1.1.1. Un *processo stocastico* è una famiglia di variabili aleatorie $(X_t)_{t \in T}$ con $T \subseteq \mathbb{R}^+$, definite da uno spazio misurabile con spazio campione Ω , σ -algebra \mathcal{F} e filtrazione $(\mathcal{F}_n)_{n \geq 0}$ in uno spazio d'arrivo, per noi sempre un sottoinsieme dei numeri reali (usualmente l'insieme dei numeri interi \mathbb{Z}). La filtrazione è definita sullo spazio Ω per definire gli eventi misurabili al livello n -esimo, cioè la σ -algebra dei possibili esiti del mio processo. Nel caso in analisi dei modelli di studio di popolazioni la filtrazione al livello n -esimo \mathcal{F}_n descrive le possibili popolazioni fino a quella generazione nota la distribuzione della progenie di un singolo individuo.

I valori via via assunti dalla variabile aleatoria sono detti stati del processo

e l'insieme dei possibili stati è detto spazio degli stati. La successione delle osservazioni è detta realizzazione o storia del processo stocastico.

Un processo stocastico si dice *markoviano* se la probabilità che determina il passaggio tra uno stato e il successivo dipende solo dallo stato presente e non dai precedenti (tale proprietà viene anche chiamata *proprietà della perdita di memoria*). Equivalentemente ciò vuol dire che il futuro dipende solo dal presente e non dal passato.

Questa interpretazione è particolarmente utile per determinare se nella modellizzazione del fenomeno che ci interessa dobbiamo utilizzare un processo markoviano o meno: nello studio di una popolazione è chiaro che il numero di figli di un individuo dipenda solo da sè stesso e non dalle generazioni che lo precedono, e pertanto in questa tesi analizzeremo solo processi stocastici markoviani.

Il primo e più importante esempio di processi stocastici markoviani sono le *catene di Markov*. Le catene di Markov sono processi stocastici markoviani (quindi "con perdita di memoria") di cui considereremo il caso a tempo discreto e con un numero di stati fissato (potenzialmente anche infinito) dove ad ogni passaggio la variabile aleatoria che descrive il comportamento del sistema viene aggiornata rispetto al tempo precedente da una matrice di transizione stocastica per righe, costante nel tempo.

Formalmente si definisce catena di Markov una famiglia di variabili aleatorie dotate della proprietà di Markov, cioè tali che $\mathbb{P}(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x | X_n = x_n)$ se entrambe le probabilità condizionate sono ben definite, cioè se $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) > 0$.

Tuttavia le catene di Markov si possono estendere a processi continui nel tempo e dove i possibili stati del sistema appartengano a uno spazio continuo mantenendo la proprietà della perdita di memoria.

Le catene di Markov sono molto utilizzate nella descrizione di situazioni che variano nel tempo; due esempi classici del suo utilizzo sono le passeggiate aleatorie e il "problema della rovina del giocatore", dove a ogni passaggio viene descritto come varia la quantità in esame (nella passeggiata aleato-

ria l'altezza della passeggiata) in funzione di una quantità stocastica, la cui distribuzione dipende dall'ambiente circostante (per esempio, nel problema della rovina del giocatore la probabilità che ad ogni iterazione egli vinca o perda dipende dalle condizioni del gioco).

1.1.1 Martingale

Per poter apprezzare in modo corretto lo studio dei processi di Galton-Watson per la modellizzazione delle popolazioni necessitiamo di introdurre un ulteriore concetto, quello di *martingala*.

Definizione 1.1.2. Si definisce martingala o processo martingala un processo stocastico adattato a una filtrazione di σ -algebre \mathcal{F}_n (quindi tale che X_l è misurabile rispetto a \mathcal{F}_l) tale che il valore atteso condizionato della v.a. al tempo $n + 1$ rispetto all'informazione al tempo n (che è la σ -algebra \mathcal{F}_n) è pari alla v.a. al tempo n . In simboli, detta $(X_n)_{n \geq 0}$ la famiglia di v.a. che compone il processo stocastico e $(\mathcal{F}_n)_{n \geq 0}$ la filtrazione associata, si ha $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$.

Tale proprietà si traduce spesso nella nozione di *gioco equo*: un processo martingala che rappresenti un gioco a scommesse è un gioco equo poiché la mia scommessa sulla puntata successiva ($n + 1$ -esima), con le informazioni note nel presente, il tempo n , che, essendo \mathcal{F}_n una filtrazione comprendono anche le informazioni ottenute nel passato, è in media equivalente alla mia scommessa nel presente, e quindi si può considerare una martingala un processo stocastico a media costante.

In base al rapporto di comparazione tra $\mathbb{E}(X_{n+1}|\mathcal{F}_n)$ e X_n introduciamo i concetti di *supermartingala* se $\mathbb{E}(X_{n+1}|\mathcal{F}_n) < X_n$ e di *submartingala* se $\mathbb{E}(X_{n+1}|\mathcal{F}_n) > X_n$ (sono quindi processi a media strettamente decrescente/crescente, da cui il nome).

Introduciamo ora un importantissimo risultato relativo alle martingale, il *Teorema di convergenza delle martingale*.

Teorema 1.1.1. *Sia $(X_n)_{n \geq 0}$ supermartingala adattata a una filtrazione di σ -algebre, con $\sup_{n \in \mathbb{N}} \mathbb{E}(X_n^+) < +\infty$ allora $X_n \rightarrow X_\infty$ q.c., con $X_\infty \in L^1(\mathcal{F})$, con $\mathcal{F} = \bigcup_n \mathcal{F}_n$*

Dimostrazione. Identifichiamo con $k = \sup_{n \in \mathbb{N}} \mathbb{E}(X_n^+)$. Definiamo un *upcrossing* rispetto a un intervallo $[a, b]$ come una famiglia di variabili aleatorie U_n pari al numero di intervalli disgiunti $[n_{i_1}, n_{i_2}]$, con $n_{i_2} \leq n$, e $X_{n_{i_1}} < a < b < X_{n_{i_2}}$. Allora vale la **disuguaglianza sugli upcrossing di Doob**, che dice che $\mathbb{E}(X_n - a)^+ \geq (b - a)\mathbb{E}(U_n)$.

Questa disuguaglianza è di fondamentale importanza per la dimostrazione del teorema.

$$\begin{aligned} \text{Poniamo ora } \Lambda &= \{\omega \text{ t.c. } X_n(\omega) \text{ non converge}\} \\ &= \{\omega \text{ t.c. } \liminf_n X_n(\omega) < \limsup_n X_n(\omega)\} \\ &= \bigcup_{a, b \in \mathbb{Q}} \{\omega \text{ t.c. } \liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)\} \\ &= \bigcup \Lambda_{a, b}. \end{aligned}$$

Per definizione di upcrossing si ha che, detto U_∞ il limite delle v.a., U_n chiaramente $\Lambda_{a, b} \subset \{\omega \mid U_\infty(\omega) = \infty\}$ perciò, dall'ipotesi sulla finitezza del *sup* dei valori attesi delle parti positive delle v.a. X_n , si ha che deve essere $\mathbb{P}(\Lambda_{a, b}) = 0 \implies \mathbb{P}(\Lambda) = 0$ poichè unione numerabile di eventi a probabilità nulla e dunque q.c. X_n converge a una v.a. X_∞ , che è di valore atteso finito per il Lemma di Fatou. \square

1.2 Processi di Galton-Watson e loro rappresentazione tramite alberi

Definizione 1.2.1.

Un *processo di Galton-Watson* è un processo stocastico $\{Z_n\}_{n \in \mathbb{N}}$, che descrive

1.2 Processi di Galton-Watson e loro rappresentazione tramite alberi

l'evoluzione di una popolazione tramite la ricorsione

$$\begin{cases} Z_0 = 1 \\ Z_{n+1} = \sum_{j=1}^{Z_n} X_j \end{cases}$$

dove X_j è una famiglia di variabili aleatorie a valori naturali indipendenti e identicamente distribuite (d'ora in poi, i.i.d.).

Questa notazione verrà considerata standard d'ora in poi: le variabili Z_i rappresenteranno il numero di individui della popolazione nella generazione i -esima, mentre le variabili X_j rappresenteranno il numero di figli avuti dall'individuo individuato dalla lettera j e saranno scritte anche come X qualora non sia necessario specificare a chi sono riferiti i figli (per esempio qualora si parli del numero medio di figli avuti da un individuo $\mathbb{E}(X)$).

Già dalla definizione risulta chiaro come i processi di Galton-Watson possiedano la proprietà di Markov, e cioè che il futuro del processo stocastico dipende solo dal presente e non dal passato: questo risulta chiaro sia dall'interpretazione pratica della definizione data (il numero dei figli dipende solo dai genitori, non dai nonni o dalle generazioni ancora precedenti) sia dalla struttura ricorsiva definita, poiché gli individui della generazione $(n + 1)$ -esima sono la somma del numero di figli di ciascun individuo della generazione n -esima, e quindi non si ha alcuna dipendenza dal passato.

I processi di Galton-Watson si prestano molto bene ad essere rappresentati tramite dei grafi ad albero, cioè grafi connessi e aciclici, di cui ricordiamo qui alcune proprietà di base che ci saranno utili nel seguito di questo lavoro:

Proprietà 1.2.1 (proprietà degli alberi).

1. In un albero $T(V, E)$ si ha $|E| = |V| - 1$

2. Dati due nodi v_i, v_j di un generico albero T , esiste un unico cammino che li unisce all'interno dell'albero

1.3 La funzione generatrice delle probabilità

Data una variabile aleatoria X a valori in \mathbb{N} definiamo la sua funzione generatrice delle probabilità G_X , definita da $[0, 1]$ in $[0, 1]$ da

$$G_X(s) = \mathbb{E}[s^X]$$

Tale funzione è ben definita poiché $s \in [0, 1]$ e X a valori naturali mi garantisce che $s^X \in [0, 1]$ come variabile aleatoria funzione di X e quindi anche il suo valore atteso sta in tale intervallo. Ne vediamo ora alcune proprietà

Proprietà 1.3.1 (Studio analitico di $G_X(s)$).

- $G_X : [0, 1] \rightarrow [0, 1]$ è continua: infatti, indicando con $\mathbb{P}(X = k) = \mu(k)$, con $k \in \mathbb{N}$, si ha che $G_X(s) = \sum_{k=0}^{+\infty} \mu(k)s^k$. Vogliamo dimostrare che $\forall \varepsilon > 0 \exists \delta > 0$, con $\delta = \frac{\varepsilon}{\mathbb{E}[X]}$ tale che si ha che $|z_1 - z_2| < \delta \Rightarrow |G_X(z_1) - G_X(z_2)| < \varepsilon$.

Infatti abbiamo che

$$\begin{aligned} |G_X(z_1) - G_X(z_2)| &= \sum_{k=0}^{+\infty} \mu(k)(z_1^k - z_2^k) \\ &= \sum_{k=0}^{+\infty} \mu(k)(z_1 - z_2) \left(\sum_{i=0}^{k-1} z_1^i z_2^{k-i-1} \right) \\ &\leq \sum_{k=0}^{+\infty} \mu(k) \frac{\varepsilon}{\mathbb{E}[X]} \sum_{i=0}^{k-1} 1 \\ &\leq \frac{\varepsilon}{\mathbb{E}[X]} \sum_{k=0}^{+\infty} k \mu(k) = \varepsilon \end{aligned}$$

- Dal punto precedente discende banalmente che $G_X(0) = \mu(0)$ e $G_X(1) = \mathbb{E}[1] = 1$

- G_X è crescente in $[0, 1]$: infatti è derivabile e vale

$$G'_X(s) = \sum_{k=1}^{+\infty} k\mu(k)s^k \geq 0$$

poiché ogni termine è non negativo e, poiché $\mu(k)$ non è sempre uguale a 0, la derivata è non negativa (positiva se $z > 0$) e quindi $G_X(s)$ è non decrescente, e con banali sostituzioni si verifica che $G'_X(0) = 0$ e che

$$G'_X(1) = \sum_{k=1}^{+\infty} k\mu(k) = \mathbb{E}[X].$$

L'unico caso escluso è il banale $\mu(0) = 1$ che non è di alcun interesse

- G_X è convessa, infatti in modo analogo al punto precedente si ha che

$$G''_X(s) = \sum_{k=2}^{+\infty} k(k-1)\mu(k)s^k$$

che è non negativa per ogni $s \in [0, 1]$.

L'unico caso in cui si ha derivata seconda nulla si trova quando $\mathbb{P}(X \geq 2) = 0$, nel qual caso $G''_x(s) = 0 \forall s \in [0, 1]$, ma vedremo che questo caso non ci crea nessun disturbo quando utilizzeremo la funzione G_X .

1.4 L'albero di Ulam-Harris

1.4.1 Definizione e prime proprietà

Definizione 1.4.1. L'albero di Ulam-Harris \mathcal{U} è un albero dove i nodi sono etichettati in modo tale da descrivere solo tramite la propria etichettatura il livello dell'albero a cui si trovano una volta determinata la radice e la loro posizione all'interno del livello. Identificando i nodi dell'albero con le loro etichette, si ha che $\mathcal{U} = \bigcup_{n=0}^{+\infty} \mathbb{N}^n$, con la convenzione che $\mathbb{N}^0 = \emptyset$, che rappresenta il nodo *radice* dell'albero (nel nostro modello, il progenitore da cui si inizia a considerare la popolazione).

Un nodo a livello n ha etichetta $v = v_1v_2\dots v_n$, con genitore $par(v) = v_1v_2\dots v_{n-1}$ e figli $v_1v_2\dots v_ni$, $i \geq 1$.

I figli nascono uno alla volta e in ordine, quindi i figli del nodo generico v saranno $v1, v2, \dots$ e così via. Si dice inoltre che vi è *fratello maggiore* di vj se sono entrambi figli di v e vale $i < j$.

Ai fini della nostra discussione ci serve introdurre la definizione di un *sottoalbero* di \mathcal{U} :

Definizione 1.4.2. Un sottoalbero $t \subset \mathcal{U}$ è un sottoinsieme dell'albero di Ulam-Harris tale che

1. $\emptyset \in t$;
2. $v \in t \Rightarrow \text{par}(v) \in t$ e quindi anche i suoi antenati stanno in t ;
3. $v = w_i \in t \Rightarrow w_j \in t \forall j \leq i$, cioè i fratelli maggiori di un nodo presente nel sottoalbero sono anch'essi presenti nel sottoalbero t .

Definiamo ora una notazione standard per il numero di figli di un determinato nodo all'interno di un sottoalbero t

Notazione 1.4.1. Il *numero di figli/out-degree* di $v \in t$ è

$$c(v, t) = \max\{i \in \mathbb{N} \mid v_i \in t\}$$

Tale numero può essere anche infinito purchè siano rispettate le condizioni della definizione di sottoalbero sopra.

Notazione 1.4.2. Indichiamo l' n -esimo livello di un albero t con $t_n = t \cap \mathbb{N}^n$ e l'albero troncato al livello n -esimo con $t_{\leq n} = \bigcup_{m=0}^n t_m$.

Si dice che $t \subset \mathcal{U}$ è *finito* se $|t| < +\infty$ ed è *localmente finito* se $|t_n| < +\infty \forall n \in \mathbb{N}$.

Un esempio di albero localmente finito ma non finito è l'albero definito dalla legge ricorsiva

$$\begin{cases} Z_0 = 1 \\ Z_{n+1} = \sum_{j=1}^{Z_n} X_j \\ \mathbb{P}(X_j = 1) = 1 \quad \forall j \in \mathbb{N} \end{cases}$$

Tale ricorsione dà origine a un albero con esattamente un individuo per ogni livello, che rispetta la definizione di localmente finito ma non quella di

finitezza (infatti ha tanti elementi quanti sono i numeri naturali).

L'altezza di t è $ht(t) = \max_{n \in \mathbb{N}} \{t_n \neq \emptyset\}$

Introduciamo ora un concetto importante, quello di *albero planare*. Tale concetto ci permetterà di lavorare sui sottoalberi di \mathcal{U} come se lavorassimo su uno spazio metrico.

Definizione 1.4.3. Si definisce *albero planare* un sottoalbero di \mathcal{U} localmente finito, cioè un albero con radice tale che i suoi vertici possano essere ordinati per livelli e da sinistra a destra in ogni livello.

Tali vertici possono essere etichettati nel seguente modo:

- Radice dell'albero $t \rightarrow \emptyset$
- Ricorsione: dato un nodo $v = v_1 v_2 \dots v_k$, esso ha figli della forma $v_j = v_1 v_2 \dots v_k j$, con $j \in \{1, 2, \dots, c(v, t)\}$
- Per comodità di notazione la radice viene omessa nell'etichettazione dei nodi, quindi i figli della radice sono indicati come $\{1, 2, \dots, c(\emptyset, t)\}$ invece che da $\{\emptyset 1, \emptyset 2, \dots, \emptyset c(\emptyset, t)\}$. Tale scelta non inficia la correttezza della notazione in quanto la radice è antenato comune di ogni nodo dell'albero.

è importante notare come tale assegnazione dei nomi ad ogni nodo sia univoca e permetta tramite l'unico cammino dalla radice al suddetto nodo di determinarlo in modo preciso.

1.4.2 Lo spazio degli alberi planari e gli alberi casuali

Vogliamo ora trattare gli alberi planari prima introdotti tramite un approccio probabilistico, trattando il numero di figli di un individuo come una variabile aleatoria e di conseguenza trattando gli alberi rappresentanti i processi di Galton-Watson generati come variabili aleatorie su un determinato spazio di probabilità che vogliamo ora definire. Vogliamo in primis che lo spazio $\mathcal{T} = \{\text{spazio degli alberi planari}\}$ sia uno spazio misurabile con una σ -algebra e possibilmente una filtrazione su tale σ -algebra.

Definizione 1.4.4. Dato un albero planare t , dato $n \geq 0$, definiamo $[t]_n = \{t \text{ planare} \mid t'_{\leq n} = t_{\leq n}\}$. La relazione $t \sim_n t' \iff t, t' \in [t]_n$ è di equivalenza in quanto è banalmente riflessiva, simmetrica e transitiva.

Definiamo poi una filtrazione con

$$\begin{aligned}\mathcal{F}_n &= \sigma(\{[t]_n \mid t \in \mathcal{T}\}); \\ \mathcal{F} &= \sigma\left(\bigcup_{n \geq 0} \mathcal{F}_n\right)\end{aligned}$$

La classe di equivalenza $[\cdot]_{n+1}$ contiene la classe $[\cdot]_n$, poiché per poter essere uguali fino al livello $n+1$ gli alberi devono necessariamente essere uguali fino al livello n . Pertanto si ha che \mathcal{F}_n è una filtrazione e la σ -algebra generata dalle \mathcal{F}_n è una σ -algebra definita sullo spazio degli alberi planari quozientato rispetto all'equivalenza detta sopra, e quindi abbiamo uno spazio filtrato per gli alberi planari su cui è ora possibile definire misure di probabilità.

Definizione 1.4.5. Dato t albero planare finito, definiamo il suo *ordinamento lessicografico* come l'ordinamento lessicografico dei suoi vertici etichettati tramite l'albero di Ulam-Harris. Con questa scelta di ordinamento ogni nodo viene dopo i suoi antenati e prima dei suoi discendenti e i figli di ogni nodo appaiono in ordine da sinistra a destra

Capitolo 2

Risultati notevoli sui processi di Galton-Watson

In questo capitolo vogliamo ricavare alcuni importanti risultati, dalla dimostrazione non scontata ma elementare, tra i quali il Teorema Fondamentale dei processi di ramificazione che lega l'estinzione di una popolazione al valore atteso della sua distribuzione di progenie.

2.1 Gli alberi planari e il cycle lemma

In questa sezione ci occuperemo di un importante risultato sulla struttura degli alberi planari, il *cycle lemma di Dwass*. Consideriamo ora un albero planare finito t , con $|t| = n$, i cui vertici siano v_1, v_2, \dots, v_n siano ordinati rispetto all'ordinamento lessicografico definito nel capitolo precedente.

Definizione 2.1.1. Definiamo $d(i) = c(v_i; t)$ e $\forall 1 \leq i \leq n$ sia

$$s(i) = 1 + \sum_{j=1}^i (d(j) - 1) = s(i-1) + d(i) - 1 \quad (2.1)$$

con $s(0) = 1$.

Posta questa definizione si dice che $(s(0), s(1), \dots, s(n))$ è il *processo depth-first queue* di t .

Tale costruzione è particolarmente significativa se si considera iterativamente a cosa corrisponde la grandezza $s(i)$. Analizzo l'albero esplorando esattamente un nodo a ogni passaggio, seguendo l'ordinamento dei vertici. Nel nodo radice ho scoperto un solo nodo (la radice per l'appunto) e devo ancora esplorare tutti i nodi dell'albero: la quantità $s(0)$ rappresenta tutti i nodi che riesco a vedere ma che devo ancora esplorare, che al tempo zero è solo la radice. Considero la formula ricorsiva $s(i) = s(i-1) + d(i) - 1$ che descrive come al passaggio i -esimo, in cui esploro un nodo, aggiungo un numero di nodi da esplorare pari al numero di figli di quel nodo e ne tolgo uno, cioè il nodo che ho appena esplorato. Da tale ragionamento risulta come conseguenza che $s(n) = 0$ poiché dopo aver esplorato n nodi ho esplorato tutto l'albero e quindi non mi rimane nessun nodo da esplorare.

Dimostrazione. Tale asserzione si può dimostrare sfruttando la proprietà 1.2.1 assieme all'aciclicità degli alberi. Difatti poiché $|E_t| = |V_t| - 1 = n - 1$ e ogni arco corrisponde in modo univoco a un rapporto genitore-figlio, si ha che

$$s(n) = 1 + \sum_{j=1}^n (d(j) - 1) = 1 + \sum_{j=1}^n d(j) - n \Rightarrow s(n) = 1 + (n - 1) - n = 0$$

poiché il numero totale di figli è il numero di vertici tolta la radice che non discende da alcun nodo. \square

Inoltre si ha anche che $s(i) \geq s(i-1) - 1$ in quanto a ogni passaggio esploro un nodo e per la formula ricorsiva 2.1 si ha che la differenza $s(i) - (s(i-1) - 1)$ è pari a $d(i) \geq 0$.

Possiamo anche verificare che vale il risultato inverso: nota una $(n+1)$ -upla $(s(0), s(1), \dots, s(n))$ con $s(0) = 1$, $s(n) = 0$ e $s(i) > 0$ per ogni $i \in \{1, \dots, n-1\}$ e $s(i) \geq s(i-1) - 1$, allora posto $d(i) = s(i) - s(i-1) - 1 \forall i \in \{1, \dots, n\} = [n]$ esiste un unico albero planare t con sequenza dei gradi $d(1), \dots, d(n)$. Quindi un albero planare si può ricostruire a partire dal suo processo depth-first queue.

Dimostrazione. Tale fatto si prova in modo costruttivo: consideriamo al tempo zero l'informazione $s(0) = 1$, per l'interpretazione data prima ci dice che vediamo solo la radice, poi al momento successivo vediamo i suoi $d(1) = s(1)$ figli che ordiniamo secondo l'ordinamento lessicografico introdotto prima. Tramite la definizione dei gradi $d(i) = s(i) - (s(i-1) - 1)$ posso trovare il numero di figli di ogni nodo e ricostruire quindi l'albero nella sua interezza, e tale ricostruzione è univoca poiché l'ordine viene assegnato ai nodi in modo univoco. \square

Introduciamo e dimostriamo ora il risultato principale di questa sezione: il *Cycle lemma di Dwass*.

Teorema 2.1.1. *Siano fissati n numeri interi $x_1, x_2, \dots, x_n \in \{-1, 0, 1, 2, \dots\} = \{j \mid j \in \mathbb{Z}, j \geq -1\}$, tali che $\sum_{i=1}^n x_i = -r, \exists r \in \mathbb{N}$. Per $j \in \mathbb{Z}, i \in \{1, \dots, n\} = [n]$, sia $s_j(i) = x_{(j+1) \bmod n} + x_{(j+2) \bmod n} + \dots + x_{(j+i) \bmod n}$, con $y \bmod n$ che rappresenta il resto della divisione euclidea di y per n con $(y, n) \in \mathbb{Z}$. Allora esistono esattamente r valori di $k \in [n]$ per cui $s_k(i) > -r \forall 0 < i < n$.*

Partiamo dall'osservare che si può restringere il problema a soli n valori di j invece che all'intero insieme \mathbb{Z} per proprietà delle classi resto. Tale considerazione ci aiuta poiché nella dimostrazione necessitiamo di estendere la definizione di $s_j(i)$ a tutto \mathbb{Z} . Dimostriamo ora questo lemma.

Dimostrazione. Estendiamo la sequenza degli x_i a \mathbb{Z} con la regola che $x_k = x_{k+tn}, t \in \mathbb{Z}$, ed estendiamo anche la definizione di $s(j)$ a \mathbb{Z} tramite la regola $s(k) - s(k-1) = x_k$, che è evidentemente equivalente alla definizione degli $s_j(i)$ data in precedenza. Senza perdita di generalità posso porre $s(0) = 0$ e ciò mi determina in modo univoco la sequenza degli $s(j)$: difatti vale

$$s(k) = \begin{cases} \sum_{j=1}^k x_j & k \geq 0 \\ -\sum_{j=k+1}^0 x_j & k < 0 \end{cases}$$

Osservazione 2.1.1. Poiché la somma di n termini consecutivi della sequenza degli x_j vale sempre $-r$, si ha che vale la relazione $s(k+n) = s(k) - r \forall k \in \mathbb{Z}$.

Osservazione 2.1.2. Notiamo inoltre che vale $s(i+j) = s(i) + s_i(j)$ poiché $s(i+j) - s(i) = \sum_{t=i+1}^{i+j} x_t s_i(j)$ con la definizione data nel testo del teorema

Definiamo ora $m(k) = \min\{s(j) \mid -\infty \leq j \leq k\}$. Grazie all'*Osservazione 2.1.1* si ha che posso ricercare il minimo solo negli ultimi n valori, quindi $m(k) = \min\{s(j) \mid k-n \leq j \leq k\}$ e sempre per quell'osservazione si ha che $m(k+n) = m(k) - r$.

Osservazione 2.1.3. Dal fatto che $s(k+1) - s(k) = x_{k+1}$ e $x_{k+1} \in \{-1, 0, 1, 2, \dots\}$ si ha che

$$m(k+1) = \begin{cases} m(k) - 1 & \text{se } x_{k+1} = -1 \\ m(k) & \text{altrimenti} \end{cases}$$

Questo ci permette di notare che allo scorrere del contatore k sugli interi la quantità $m(k)$ decresce al più di uno a ogni passaggio, pertanto assume tutti i possibili valori interi (in quanto è illimitata inferiormente grazie all'*Osservazione 2.1.1*).

Da tutto ciò discende quindi che la condizione del nostro teorema si può provare così:

$$s_k(i) > -r \quad \forall i \in [n-1] \iff \quad (2.2)$$

$$s(k+i) - s(k) > -r \quad \forall i \in [n-1] \iff \quad (2.3)$$

$$s(k+i) - s(k) + r > 0 \quad \forall i \in [n-1] \iff \quad (2.4)$$

$$s(k+i-n) - s(k) > 0 \quad \forall i \in [n-1] \iff \quad (2.5)$$

$$s(j) > s(k) \quad \forall j \in \{k-n, \dots, k\} \iff \quad (2.6)$$

$$m(k-1) > s(k) \iff \quad (2.7)$$

$$m(k-1) > m(k) \quad (2.8)$$

Le prime tre doppie implicazioni sono conseguenze immediate delle Osservazioni precedenti, mentre la quarta equivale al cambio di variabile $k+i-n = j$.

La quinta doppia implicazione si dimostra velocemente: l'implicazione diretta discende dal fatto che $m(k-1)$ è il minimo di $s(j)$ per $j \in \{k-n-1, \dots, k-1\}$ che è una richiesta più debole rispetto all'ipotesi $s(j) > s(k) \forall j \in \{k-n, \dots, k\}$ grazie al fatto che $s(k-n-1) = s(k-1) + r > s(k-1) > s(k)$; l'implicazione inversa discende semplicemente dalla definizione di $m(k-1)$. L'ultima doppia implicazione ha una direzione ovvia (l'implicazione diretta è un'applicazione della definizione di $m(k)$), mentre l'implicazione inversa è leggermente più articolata.

Poiché $m(k-1) = \min\{s(j) \mid k-n-1 \leq j \leq k-1\}$ e $m(k) = \min\{s(j) \mid k-n \leq j \leq k\}$ si ha che qualora esistesse $l \in \{k-n, \dots, k-1\} \subset \{k-n-1, \dots, k-1\}$ tale che $s(l) < s(k)$ la disuguaglianza dei minimi non sarebbe realizzata e quindi deve essere $m(k) = s(k)$ da cui l'implicazione voluta.

A questo punto possiamo dimostrare che esistono r valori diversi tali che $m(k-1) > m(k)$ in $\{1, \dots, n-1\}$. Poiché $m(n) = m(0) - r$ e $m(i+1) \geq m(i) - 1$ esistono esattamente r valori per cui $m(k-1) > m(k)$, da cui la tesi. \square

2.2 Processi di ramificazione e alberi di Bienaymé

In questa sezione introduciamo gli alberi di Bienaymé, cioè alberi planari su cui si definisce una misura di probabilità a partire dalla distribuzione della progenie della popolazione in esame.

Definizione 2.2.1. Data una distribuzione di probabilità μ definita a valori naturali, si definisce *albero di Bienaymé con distribuzione μ* o $\text{Bienaymé}(\mu)$ un albero planare casuale che rappresenta un processo di ramificazione di una popolazione con distribuzione della progenie con legge μ . Ciò induce una distribuzione di probabilità sullo spazio degli alberi planari come introdotto nella sezione 1.4.2 che si denota con B_μ , ed è determinata univocamente nel seguente modo: la probabilità che un dato albero di Bienaymé T sia uguale a un albero planare t fino al livello h per ogni intero $h \geq 1$ e per ogni t albero

planare di altezza $\leq h$ è pari a:

$$\mathbb{P}(T^{\leq h} = t) = B_\mu(\{\tau \in T \mid \tau^{\leq h} = t\}) = \prod_{v \in t^{\leq h-1}} \mu(\deg_t(v)) \quad (2.9)$$

dove T è lo spazio degli alberi planari.

Questa definizione non è particolarmente intuitiva, quindi vediamo un approccio equivalente per la costruzione di Bienaymé(μ):

- Considero il nodo \emptyset radice, il suo numero di figli X_\emptyset è una variabile aleatoria con legge μ , pertanto la radice ha figli $1, \dots, X_\emptyset$;
- per ogni $i = \{1, \dots, X_\emptyset\}$ si ha che il numero di figli di i cioè X_i ha legge μ , e quindi il nodo i ha figli $i1, \dots, iX_i$
- procedo in questo modo costruendo l'albero per livelli.

Introduciamo ora un concetto fondamentale, la *sopravvivenza* di una popolazione.

Definizione 2.2.2. Sia $Z_n = Z_n(T)$ il numero di individui all' n -esima generazione di un albero di Bienaymé T . Sia ora $|T| = \sum_{n=0}^{+\infty} Z_n$ la cardinalità totale della popolazione. Si dice che si ha *sopravvivenza* della popolazione se $Z_n > 0 \forall n \in \mathbb{N}$, altrimenti si ha *estinzione* della suddetta popolazione. Tale definizione può essere scritta in termini del tutto equivalenti in base a $|T|$, difatti si ha $Z_n > 0 \forall n \in \mathbb{N} \iff |T| = +\infty$ poiché se la somma è finita ho estinzione e quindi il numero totale di individui è finito, mentre se ad ogni generazione ho un numero di individui strettamente positivi la serie che definisce $|T|$ diverge chiaramente.

A questo punto viene naturale chiedersi sotto quali condizioni una popolazione sopravviva e sotto quali raggiunga l'estinzione. Vedremo ora che mentre è abbastanza immediato trovare delle condizioni per avere la sopravvivenza di una popolazione come evento possibile, è più complicato trovare tali condizioni per il fenomeno dell'estinzione (anche se otterremo come risultato collaterale del prossimo Teorema delle condizioni per cui si ha quasi certamente l'estinzione della popolazione).

2.2.1 Il teorema fondamentale dei processi di ramificazione

Questa sezione presenta il testo e la dimostrazione di un teorema che fornisce condizioni necessarie e sufficienti per avere probabilità di sopravvivenza non nulla, si capisce quindi quanto sia un risultato importante nella teoria dello studio dei processi di ramificazione.

Teorema 2.2.1. *Sia X una variabile aleatoria a valori in \mathbb{N} con distribuzione μ , sia $T = \text{Bienaymé}(\mu)$. Allora vale $\mathbb{P}(|T| = +\infty) > 0$ se e solo se vale una delle due seguenti condizioni:*

- $\mathbb{P}(X = 1) = 1$
- $\mathbb{E}(X) > 1$

Prima di dimostrarlo, notiamo che le due condizioni equivalgono o alla possibilità di avere quasi certamente un figlio o al fatto di avere in media più di un figlio.

Per procedere alla dimostrazione del teorema necessitiamo di alcuni risultati che ora enunceremo e dimostreremo.

Lemma 2.2.2. *Sia X una v.a. μ -distribuita, allora $\forall n \in \mathbb{N}$ vale $\mathbb{E}(Z_n) = (\mathbb{E}(X))^n$*

Dimostrazione. Poiché $Z_0 = 1$ il caso per $n = 0$ è banalmente verificato.

Lavoriamo per induzione su $n \geq 1$. Consideriamo il processo condizionato ad avere i figli della radice, allora si ha che ognuno di questi i figli è radice di un processo con ramificazione con $n - 1$ livelli su cui posso applicare l'ipotesi induttiva, e per ognuno di questi il valor medio del livello $(n - 1)$ -esimo è uguale, quindi sfruttando la probabilità condizionata e il fatto che

$\mathbb{E}(X) = \sum_{i=0}^{+\infty} i\mathbb{P}(X = i)$ si ha che

$$\begin{aligned} \mathbb{E}(Z_{n+1}) &= \sum_{i=0}^{+\infty} \mathbb{E}(Z_n | X_\emptyset = i) \mathbb{P}(X_\emptyset = i) = \\ &= \sum_{i=0}^{+\infty} i\mathbb{E}(Z_n) \mathbb{P}(X = i) = \mathbb{E}(Z_n) \mathbb{E}(X) = (\mathbb{E}(X))^{n+1} \end{aligned} \quad (2.10)$$

Ciò prova il lemma. \square

Questo lemma ha un corollario immediato, che ci dà una condizione utile sull'estinzione di una popolazione.

Corollario 1. *Se il valore atteso di X è minore di 1 allora si ha che il valore atteso $\mathbb{E}(|T|)$ è finito anch'esso, cioè si ha estinzione della popolazione.*

Dimostrazione. Poichè

$$\mathbb{E}(|T|) = \sum_{n=0}^{+\infty} \mathbb{E}(Z_n) = \sum_{n=0}^{+\infty} (\mathbb{E}(X))^n = \frac{1}{1 - \mathbb{E}(X)} < +\infty$$

ho che per la disuguaglianza di Markov la probabilità che la cardinalità della popolazione sia infinita è pari a zero, quindi ho estinzione della popolazione. La convergenza della serie geometrica è assicurata dal fatto che il valore atteso di X è minore di 1 \square

Dal lemma precedente si può ricavare una conseguenza fondamentale che permette di dire che il processo stocastico $(M_n)_{n \geq 0}$, con $M_n = \frac{Z_n}{(\mathbb{E}(X))^n}$, è una $\tilde{\mathcal{F}}_n$ -martingala, con $\tilde{\mathcal{F}}_n = \sigma(Z_0, \dots, Z_n)$ la σ -algebra generata dalle v.a. Z_i .

Considero $Z_{n+1} = \sum_{v \in T_n} X_v$ con T_n l' n -esimo livello dell'albero di Bienaymé sotto esame. Allora si ha che $\forall S \subseteq \mathbb{N}^n$ fissato

$$\mathbb{E}(Z_{n+1} | T_n = S) = \mathbb{E}\left(\sum_{v \in S} X_v | T_n = S\right) = \sum_{v \in S} \mathbb{E}(X) = |S| \cdot \mathbb{E}(X)$$

dove la seconda uguaglianza è vera per indipendenza della v.a. X rispetto a

S e linearità del valore atteso. Poiché $|S| = Z_n$ abbiamo che $\mathbb{E}(Z_{n+1}|\tilde{\mathcal{F}}_n) = Z_n\mathbb{E}(X)$ perché tale condizionamento è indipendente dalla scelta di S . Pertanto il processo stocastico $(Z_n)_{n \geq 0}$ è una martingala se $\mathbb{E}(X) = 1$. Per dimostrare che $(M_n)_{n \geq 0}$ è $\tilde{\mathcal{F}}_n$ -martingala osserviamo che adattabilità e integrabilità sono ovvie poiché Z_n è misurabile rispetto a $\tilde{\mathcal{F}}_n$. Dimostriamo la proprietà di martingalità usando la notazione $\mathbb{E}(X) = \alpha$.

$$\begin{aligned} \mathbb{E}(M_{n+1}|\tilde{\mathcal{F}}_n) &= \mathbb{E}\left(\frac{Z_n}{\alpha^{n+1}}|\tilde{\mathcal{F}}_n\right) = \\ \frac{1}{\alpha^{n+1}}\mathbb{E}(Z_{n+1}|\tilde{\mathcal{F}}_n) &= \frac{Z_n\alpha}{\alpha^{n+1}} = \frac{Z_n}{\alpha^n} = M_n \end{aligned} \quad (2.11)$$

e ciò prova la proprietà di martingalità voluta.

Necessitiamo di un'ulteriore proprietà prima della dimostrazione del teorema fondamentale, anche questa di grande importanza poiché fornisce un metodo di calcolo per la probabilità di estinzione. Consideriamo la funzione generatrice della probabilità introdotta nella *Sezione 1.2*, di cui adesso faremo largo uso: la funzione $G_X(s) = \mathbb{E}(s^X)$.

Proprietà 2.2.1 (*proprietà fondamentale dei processi di ramificazione*). In un processo di ramificazione con distribuzione della progenie con legge μ , detta X una v.a. con la stessa legge si ha che

$$\mathbb{P}(X = 1) < 1 \Rightarrow \mathbb{P}(|T| < +\infty) = \min_{z \geq 0} \{G_X(z) = z\}.$$

Procediamo ora a dimostrare questa importantissima proprietà.

Dimostrazione. Chiamiamo $p = \mathbb{P}(|T| < +\infty)$. L'evento $(|T| < +\infty)$ è l'unione degli eventi $(Z_n = 0)$ su tutti i valori di n naturali. Pertanto $\mathbb{P}(|T| < +\infty) = \mathbb{P}\left(\bigcup_{n=0}^{+\infty} (Z_n = 0)\right)$, e tali eventi sono crescenti (se $Z_n = 0$ allora anche $Z_{n+1} = 0$) e quindi posso sfruttare la continuità dal basso e scrivere che $p = \lim_{n \rightarrow +\infty} \mathbb{P}(Z_n = 0)$.

Sia ora $F_1(z) = G_X(z)$ la funzione generatrice delle probabilità, e definiamo una successione di funzioni come $F_n(z) = G_X(F_{n-1}(z))$, vogliamo ora dimostrare che $\mathbb{P}(Z_n = 0) = F_n(0)$. Il caso $n = 1$ è banale in quanto $F_1(0) = G_X(0) = \mu(0) = \mathbb{P}(X = 0) = \mathbb{P}(Z_1 = 0)$.

$$\begin{aligned} \text{Consideriamo il caso } n > 1. \quad \mathbb{P}(Z_n = 0) &= \sum_{i=0}^{+\infty} \mathbb{P}(Z_n = 0 | Z_1 = i) \mathbb{P}(Z_1 = i) = \\ &= \sum_{i=0}^{+\infty} \mathbb{P}(Z_{n-1} = 0)^i \mathbb{P}(X = i) = \sum_{i=0}^{+\infty} (F_{n-1}(0))^i \mathbb{P}(X = i) = G_X(F_{n-1}(0)) = F_n(0) \end{aligned}$$

dove abbiamo sfruttato l'indipendenza delle i ramificazioni per determinare la probabilità condizionata.

Abbiamo quindi $p = \lim_{n \rightarrow +\infty} F_n(0)$, quindi per continuità di G_X ho che $F_{n+1}(0) = G_X(F_n(0)) \rightarrow F(p)$ e quindi si ha che $p = G_X(p)$ poiché limiti di due successioni che hanno lo stesso limite.

Dimostriamo ora che tale p è il minimo tra i valori che realizzano tale condizione, sia q un'altra soluzione di $G_X(z) = z$. Poiché G_X crescente, ho che $q = G_X(q) \geq G_X(0)$ e se continuo ad applicare G_X da ambo i lati q rimane fisso in quanto punto fisso della successione $F_n(z)$, mentre a destra ottengo $\lim_{n \rightarrow +\infty} F_n(0) = p$, da cui si ha $q \geq p$ (le proprietà utilizzate sono quelle della funzione generatrice delle probabilità vista nella Sezione 1.3). \square

A questo punto possiamo finalmente dimostrare il Teorema fondamentale dei processi di ramificazione grazie agli strumenti che abbiamo acquisito nelle ultime pagine e a quanto visto nella sezione 1.3.

Dimostrazione. Partiamo dal fatto che grazie al Corollario 1 sappiamo che se il valore atteso della distribuzione della progenie è minore di 1 si ha estinzione della popolazione. Consideriamo le due condizioni date dal Teorema.

- qualora si abbia $\mathbb{P}(X = 1) = 1$, risulta chiaro che $Z_n = 1 \forall n \geq 1$ e che quindi $|T| = \sum_{n \in \mathbb{N}} Z_n = +\infty$. Passiamo quindi al secondo caso.
- Se $\mathbb{E}(X) > 1$ per continuità si deve avere che $\exists x \in [0, 1]$ tale che $G_X(x) < x$, poiché ricordiamo che $\mathbb{E}(X)$ è la derivata di G_X in 1, quindi ho che in un intorno sinistro di 1 la funzione $G_X(z) - z$ è negativa. Tuttavia poiché $G_X(0) = \mu(0) \geq 0$ ho dal teorema di esistenza degli zeri che la funzione $G_X(z) - z$ ha almeno uno zero nell'intervallo $[0, 1)$, e dalla Proposizione fondamentale ho che la minima tra le soluzioni (che

chiameremo y^* è la probabilità di estinzione. Pertanto la probabilità di sopravvivenza è pari a $1 - y^* > 0$.

- Rimane da analizzare il caso in cui il valore atteso della progenie sia pari a 1, con $\mathbb{P}(X = 1) < 1$. In tale casistica si ha che deve esistere $k \in \mathbb{N} \setminus \{0, 1\}$ con $\mathbb{P}(X = k) > 0$. Questo fatto mi garantisce che la funzione G_X è strettamente convessa $\forall z \geq 0$, e pertanto ho che G'_X è strettamente crescente e vale $G'_X(1) = \mathbb{E}(X) = 1$. Pertanto la funzione $G_X(z) - z$ ha derivata ≤ 0 in ogni suo punto quindi è decrescente e vale 0 in 1 che è dunque il suo unico zero, quindi la probabilità di estinzione è 1, come volevasi dimostrare.

□

2.3 Condizionamento per la finitezza degli alberi di Bienaymé

In questa sezione vogliamo capire quale è la distribuzione di un albero di Bienaymé condizionato all'essere finito. Lavoreremo sempre con la distribuzione della progenie di legge μ , con valore atteso della v.a. associata $\mathbb{E}(X) > 1$ e con conseguente distribuzione B_μ per l'albero di Bienaymé associato.

Richiediamo prima di tutto che $\mathbb{P}(X = 0) > 0$, altrimenti non è possibile richiedere la finitezza dell'albero. Per determinare la distribuzione nuova, condizionata alla finitezza dell'albero, ragioniamo pensando al numero di figli della radice. Chiamiamo $q = \mathbb{P}(|T| < +\infty)$, allora abbiamo che usando la formula di Bayes

$$\begin{aligned} \mathbb{P}(c(\emptyset; T) = j \mid |T| < +\infty) &= \frac{\mathbb{P}(c(\emptyset; T) = j)}{\mathbb{P}(|T| < +\infty)} \mathbb{P}(|T| < +\infty \mid c(\emptyset; T) = j) = \\ &= \frac{\mu(j)}{q} q^j = \mu(j) q^{j-1} = \widehat{\mu}(j) \end{aligned} \tag{2.12}$$

Ciò è vero poiché $\mathbb{P}(|T| < +\infty | c(\emptyset; T) = j) = q^j$ poiché ognuno dei j figli della radice è, indipendentemente dagli altri, radice di un albero cui stiamo richiedendo di essere finito, pertanto tale probabilità è effettivamente q^j .

Possiamo effettuare un sanity check e verificare che la misura $\hat{\mu}$ definita sopra descriva effettivamente una probabilità:

$$\sum_{j=0}^{+\infty} \hat{\mu}(j) = \sum_{j=0}^{+\infty} \mu(j)q^{j-1} = q^{-1}G_X(q) = q^{-1}q = 1 \quad (2.13)$$

dove abbiamo usato la scrittura estesa della funzione generatrice delle probabilità e la proposizione fondamentale per dire $G_X(q) = q$, e pertanto concludiamo che $\hat{\mu}$ definisce una misura di probabilità.

Sia inoltre \hat{X} una v.a. $\hat{\mu}$ -distribuita, si ha che $\mathbb{E}(\hat{X}) = \sum_{j=0}^{+\infty} j\mu(j)q^{j-1} = G'_X(q) < 1$ poiché G_X convessa. Questo ci assicura che grazie a questo condizionamento l'albero è finito poiché il valore atteso della v.a. associata alla legge $\hat{\mu}$ è inferiore ad 1, e ho che l'albero di Bienaymé associato \hat{T} ha distribuzione $B_{\hat{\mu}}$.

Consideriamo ora un'importante proprietà che lega la cardinalità degli alberi di Bienaymé finiti alle distribuzioni della progenie della popolazione collegata.

Proprietà 2.3.1. Sia μ una distribuzione di probabilità con supporto i numeri naturali, e T un albero di Bienaymé con distribuzione B_μ , e sia $(D_i)_{i \in \mathbb{N}}$ una famiglia di v.a. i.i.d. μ -distribuite. Allora si ha che per ogni $n \in \mathbb{N} \setminus \{0\}$

$$\mathbb{P}(|T| = n) = \frac{1}{n} \mathbb{P}(D_1 + D_2 + \dots + D_n = n - 1)$$

Dimostrazione. Sia \mathcal{D} l'insieme delle sequenze dei gradi uscenti (così da non avere ripetizioni nel conteggio) degli alberi piani con n vertici. Per $0 \leq i < n$, sia $\mathcal{D}_i = \{(d_{i+1}, \dots, d_n, d_1, \dots, d_i) \mid (d_1, \dots, d_n) \in \mathcal{D}\}$, cioè l'insieme delle sequenze dei gradi traslate di i posizioni in avanti. Per il *Cycle Lemma di Dwass* (Th. 2.1.1), $\mathcal{D}_0, \dots, \mathcal{D}_{n-1}$ sono disgiunti, poiché considero i gradi associati ai vertici con ordine e quindi due sequenze uguali nei gradi ma riferite a

vertici diversi sono distinte poiché rappresentano due alberi diversi e la loro unione è

$\bigcup_{i=0}^{n-1} \mathcal{D}_i = \{(d_1, \dots, d_n) \in \mathbb{N}^n \mid d_1 + \dots + d_n = n - 1\}$ dato che in un albero con n vertici la somma dei gradi è il numero degli archi, cioè $n - 1$. Inoltre tali sequenze sono tutte le sequenze di gradi possibili (a somma $n - 1$) poiché parto dall'insieme di tutti gli alberi e considero ogni sequenza ciclata in tutti i possibili modi.

Sia ora t un qualsiasi albero con n vertici v_1, \dots, v_n con gradi $d(1), \dots, d(n)$, quindi ho che t è un albero planare finito.. Allora ha sicuramente un numero di livelli $ht(t) \leq (n + 1)$, posso usare la definizione della misura di probabilità per l'albero di Bienaymé e avere che $\mathbb{P}(T = t) = \prod_{v \in t} \mu(deg_t(v)) =$

$\prod_{i=1}^n \mu(d(i)) = \mathbb{P}(D_1, \dots, D_n) = (d(1), \dots, d(n))$ poiché ognuna delle D_i è μ -distribuita e pertanto rappresenta il grado di uno degli n nodi: unendo a ciò l'indipendenza delle v.a. in esame si ottiene l'ultima uguaglianza.

A questo punto sommiamo su tutti i possibili alberi planari a n vertici così da ottenere la probabilità che la cardinalità $|T|$ sia esattamente pari a n . Ciò è possibile poiché ognuno degli alberi è distinto dagli altri.

Si ha dunque $\mathbb{P}(|T| = n) = \mathbb{P}((D_1, \dots, D_n) \in \mathcal{D}) = \frac{1}{n} \mathbb{P}((D_1, \dots, D_n) \in \mathcal{D}_i)$, dove l'ultima uguaglianza è vera per disgiunzione dei \mathcal{D}_j . Ora usando l'i-

dentità sopra dimostrata $\bigcup_{i=0}^{n-1} \mathcal{D}_i = \{(d_1, \dots, d_n) \in \mathbb{N}^n \mid d_1 + \dots + d_n = n - 1\}$

ricavo che la probabilità che sto cercando è pari a $\frac{1}{n} \mathbb{P}((D_1, \dots, D_n) \in \mathcal{D}_i) = \frac{1}{n} \mathbb{P}(D_1 + \dots + D_n = n - 1)$ e ciò prova la proprietà. □

Dimostriamo ora un'altra caratterizzazione della probabilità che un albero di Bienaymé abbia cardinalità fissata, che fa uso del Cycle Lemma di Dwass.

Proprietà 2.3.2. Siano $(X_i, i \geq 1)$ v.a. i.i.d., con $X_i \sim D_i - 1$, e per $n \geq 0$ sia $S_n = 1 + X_1 + \dots + X_n$. Definiamo ora τ il minimo tempo affinché la

quantità S_m si annulli, cioè

Definizione 2.3.1. $\tau = \inf_{m \in \mathbb{N}} \{S_m = 0\}$

Si ha che $\mathbb{P}(|T| = n) = \mathbb{P}(\tau = n)$

La dimostrazione si basa sulla lettura della quantità S_m introdotta come visto nel Cycle Lemma di Dwass, e sul fatto che $\sum_i X_i = \sum_i (D_i - 1) = \sum_i D_i - n$

Dimostrazione. Abbiamo visto nella proprietà precedente che $\mathbb{P}(|T| = n) = \frac{1}{n} \mathbb{P}(D_1 + D_2 + \dots + D_n = n - 1)$ che per l'osservazione sulla relazione tra D_i e X_i equivale a $\frac{1}{n} \mathbb{P}(X_1 + \dots + X_n = -1) = \frac{1}{n} \mathbb{P}(S_n = 0)$ con n minimo numero naturale che soddisfa tale condizione.

Consideriamo ora $\mathbb{P}(\tau = n) = \mathbb{P}(S_n = 0, S_m > 0 \forall 0 \leq m < n)$. Tale evento può essere scomposto tramite la definizione di S_m : ho n valori x_1, \dots, x_n tali che la loro somma è pari a -1 e di cui so la sequenza ma non l'elemento da cui inizio (ho pertanto n possibili ordinamenti di questi valori): siccome richiedo che $S_m > 0 \forall m < n$ e che $S_n = 0 \iff \sum_i X_i = -1$ sono esattamente nelle ipotesi del Cycle Lemma e dunque, poiché $r = 1$ esiste un unico ordinamento possibile che rispetta queste ipotesi, e pertanto $\mathbb{P}(S_n = 0, S_m > 0 \forall 0 \leq m < n) = \frac{1}{n} \mathbb{P}(S_n = 0)$. \square

Possiamo analizzare cosa succede in un caso specifico, quando le variabili D_i sono distribuite come variabili aleatorie di Poisson di parametro $\lambda \in [0, 1]$ i.i.d.. Consideriamo l'albero di Bienaymé associato a tale processo T , sappiamo dal Teorema fondamentale dei processi di ramificazione che, poiché la v.a. $X \sim Poi(\lambda)$ ha valore atteso $\mathbb{E}(X) = \lambda \leq 1$ e che $\mathbb{P}(X = 1) < 1$, allora si ha estinzione della popolazione, cioè $\mathbb{P}(|T| < +\infty) = 1$. Inoltre vale che $\forall k \in \mathbb{N} \mathbb{P}(|T| = k) = \frac{1}{k} \mathbb{P}(D_1 + \dots + D_k = k - 1) = \frac{1}{k} \mathbb{P}(Poi(\lambda k) = k - 1) = \frac{1}{k} e^{-\lambda k} \frac{(\lambda k)^{k-1}}{(k-1)!} = \frac{e^{-\lambda k} (\lambda k)^{k-1}}{k!}$.

Dal fatto dimostrato sopra che la cardinalità dell'albero di Bienaymé associato è finita si ha che vale la relazione
$$\sum_{i=0}^{+\infty} \frac{e^{-\lambda k} (\lambda k)^{k-1}}{k!} = 1.$$

Osservazione 2.3.1. La proprietà 2.3.1 si può estendere a foreste di alberi nel seguente modo:

Proprietà 2.3.3. Sia μ distribuzione di probabilità sui numeri naturali, $(T_i, i \geq 1)$ una famiglia di alberi di Bienaymé indipendenti e identicamente B_μ -distribuiti. Vale che $\forall 1 \leq i \leq n \mathbb{P}(|T_1| + \dots + |T_n| = r) = \frac{r}{n} \mathbb{P}(D_1 + \dots + D_n = n - r)$ con $(D_i, i \geq 1)$ v.a. i.i.d. μ -distribuite.

La dimostrazione non è trattata qui, in quanto semplice estensione della dimostrazione già vista.

La proprietà 2.3.1 si può dimostrare in modo alternativo a partire da un lemma più debole del Cycle lemma di Dwass, ma di dimostrazione più diretta e intuitiva.

Lemma 2.3.1. *Siano $(X_i, i \geq 1)$ v.a. con $X_i \sim D_i - 1$, sia $r \in \mathbb{N}$ e sia $S_n = r + X_1 + \dots + X_n$, detto $\tau_0 = \inf_{m \in \mathbb{N}} \{S_m = 0\}$ si ha che $\mathbb{P}(\tau_0 = n) = \mathbb{P}(S_n = 0)$.*

Dimostrazione. Definiamo $\mathbb{P}_k\{\cdot\} = \mathbb{P}(\cdot | S_k = 0)$, e dimostriamo la proprietà per induzione su n, k . Il caso $k = 0$ è banale, in quanto risulta $0 = 0$. Il caso con $n = k = 1$ ci dà $\mathbb{P}(\tau_0 = 1) = \mathbb{P}(X_1 = -1) = \mathbb{P}(S_1 = 0)$ quindi abbiamo i due casi base che ci servono per l'induzione.

Consideriamo ora $n > 1, 0 < k \leq n$ (le altre casistiche rientrano nei casi base) e condizioniamo su X_1 sfruttando la *proprietà di Markov*, cioè che dato $X_1 = i$ il valore parziale di S_n si comporta come una passeggiata aleatoria che inizia al valore $k + i$. Notiamo ora che $k + i \geq 0$ poiché $k \geq 1$ e $i \geq -1$ per definizione delle v.a. X_j e $n - 1 \geq 0$.

Alla luce di quanto appena detto, posso scrivere

$$\begin{aligned}
\mathbb{P}_k(\tau_0 = n) &= \sum_{i=-1}^{+\infty} \mathbb{P}_{k+i}(\tau_0 = n-1) \mathbb{P}(X_1 = i) = \\
&= \frac{1}{n-1} \sum_{i=-1}^{+\infty} (k+i) \mathbb{P}_{k+i}(S_{n-1} = 0) \mathbb{P}(X_1 = i) = \\
&= \frac{1}{n-1} \sum_{i=-1}^{+\infty} (k+i) \mathbb{P}_k(X_1 = i, S_n = 0) = \\
&= \frac{1}{n-1} \sum_{i=-1}^{+\infty} (k+i) \mathbb{P}_k(X_1 = i | S_n = 0) \mathbb{P}_k(S_n = 0) = \\
&= \frac{1}{n-1} \mathbb{P}_k(S_n = 0) \left[k \sum_{i=-1}^{+\infty} \mathbb{P}_k(X_1 = i | S_n = 0) + \right. \\
&\quad \left. + \sum_{i=-1}^{+\infty} i \mathbb{P}_k(X_1 = i | S_n = 0) \right] = \\
&= \frac{1}{n-1} \mathbb{P}_k(S_n = 0) [k + \mathbb{E}_k(X_1 | S_n = 0)]
\end{aligned}$$

Per calcolare il valore atteso nell'ultima riga della catena di uguaglianze qui sopra possiamo notare che $S_n = 0$, $S_0 = k$ e poiché le X_i sono i.i.d. il valore medio è $\mathbb{E}(X_i) = -\frac{k}{n}$, e quindi sostituendo ho che

$$\mathbb{P}_k(\tau_0 = n) = \frac{1}{n-1} \mathbb{P}_k(S_n = 0) \left[k - \frac{k}{n} \right] = \frac{k}{n} P_k(S_n = 0) \quad (2.14)$$

A questo punto è immediato osservare come quanto dimostrato sia equivalente alla proposizione di partenza: difatti è sufficiente applicare questo lemma appena visto ponendo $k = 1$ e si ottiene la proprietà 2.3.1.

□

A questo punto abbiamo ottenuto tutti gli strumenti necessari per avvicinarci a dimostrare il Teorema di Kesten-Stigum.

Capitolo 3

Processi di Galton-Watson con immigrazione e il teorema di Kesten-Stigum

3.1 Processi di Galton-Watson con immigrazione

In questo capitolo vogliamo trattare cosa succede se consideriamo un processo di immigrazione all'interno della nostra popolazione. Ciò equivale a considerare un processo per cui nel nostro albero introduciamo a ogni generazione un certo numero di individui (eventualmente anche 0) che non sono figli di nessun individuo precedentemente presente nella popolazione, ma i cui discendenti apparterranno alla popolazione. Tale numero può essere deterministico oppure una variabile aleatoria con una certa distribuzione. Solitamente considereremo un processo migratorio stocastico, con distribuzione nota.

Costruiamo la nostra popolazione per generazioni con questo metodo: siano $(X_{n,k})_{n,k \geq 1}$ v.a. i.i.d. con distribuzione μ , e $(Y_n)_{n \geq 1}$ con distribuzione ν , con ν a supporto nei numeri naturali (consideriamo quindi solo processi di immi-

grazione e non di emigrazione nella nostra analisi). Fissiamo $Z_0 = Y_0$ come inizio della popolazione in esame, e poi consideriamo le generazioni successive legate alle precedenti dalla legge ricorsiva $Z_{n+1} = Y_{n+1} + X_{n,1} + \dots + X_{n,Z_n}$. Ciò significa che alla generazione $n + 1$ -esima ho un quantitativo Y_{n+1} di membri della popolazione che immigrano dall'esterno, e per ognuno degli individui presenti alla generazione n -esima aggiungiamo il suo numero di figli (che come al solito sono v.a. i.i.d.).

In questo modo rappresentiamo una popolazione dove a ogni generazione si considera un flusso migratorio non negativo. Si può anche considerare il caso in cui il flusso migratorio non sia stocastico ma deterministico, nel qual caso si otterrà un vettore Y di componenti a valori naturali, con la legge ricorsiva che definisce la popolazione non cambia.

Introduciamo ora un importante risultato provato da Seneta nel 1970 ([4]) che ci servirà nella dimostrazione del Th. di Kesten-Stigum.

Teorema 3.1.1. *Siano X, Y, Z famiglie di v.a. con distribuzione come definite sopra, con valore atteso $\mathbb{E}(X) = \alpha$. Se $\mathbb{E}(\max(\log(Y), 0)) = \mathbb{E}(\log^+(Y)) < +\infty$ allora $\lim_{n \rightarrow \infty} \frac{Z_n}{\alpha^n} < +\infty$ quasi certamente, mentre se $\mathbb{E}(\log^+(Y)) = +\infty$ allora $\lim_{n \rightarrow \infty} \frac{Z_n}{c^n} = +\infty$ quasi certamente $\forall c > 0$.*

Per dimostrare questo teorema necessitiamo prima di provare un altro lemma, che ci dà un'utile proprietà collegata a v.a. con valore atteso finito.

Lemma 3.1.2. *Siano $(R_n)_{n \geq 1}$ delle v.a. i.i.d. a valori non negativi. Allora valgono le seguenti:*

1. Se $\mathbb{E}(R_1) < \infty$ allora q.c. $\limsup \frac{R_n}{n} = 0$ e inoltre $\sum_{n \geq 1} e^{R_n} c^n < +\infty$
per ogni $c \in [0, 1]$
2. Se $\mathbb{E}(R_1) = \infty$ allora q.c. $\limsup \frac{R_n}{n} = +\infty$ e inoltre $\sum_{n \geq 1} e^{R_n} c^n = +\infty$
per ogni $c \in [0, 1]$

Dimostrazione. Dimostriamo separatamente i due casi.

1. Fissiamo un numero reale $\varepsilon > 0$. Allora si ha che

$$\sum_{n>0} \mathbb{P}(R_n \geq \varepsilon n) = \sum_{n>0} \mathbb{P}(R_1 \geq \varepsilon n) \leq \frac{1}{\varepsilon} \sum_{n \geq 0} \mathbb{P}(R_1 \geq n) = \frac{\mathbb{E}(R_1)}{\varepsilon} < +\infty \quad (3.1)$$

La prima uguaglianza discende immediatamente dal fatto che le R_i sono i.i.d., mentre la disuguaglianza centrale è una stima che si dimostra nel seguente modo: consideriamo tutti i valori di εn appartenenti all'intervallo $[k, k+1)$, allora per monotonia maggiore ognuna delle probabilità ad essi associati con $\mathbb{P}(X \geq k)$, e ne ho sicuramente meno di $\frac{1}{\varepsilon}$ come quantità. Poiché al crescere di n ho intersezione con tutti i $[k, k+1)$ al variare di $k \in \mathbb{N}$ ho che la serie viene indicizzata non più a partire da 1 bensì da 0. Invece l'ultima uguaglianza si ricava da una riscrittura del valore atteso grazie a una doppia sommatoria: infatti

$$\mathbb{E}(X) = \sum_k k \mathbb{P}(X = k) = \sum_{k=1}^{+\infty} \sum_{n=1}^k \mathbb{P}(X = k) = \sum_{n \in \mathbb{N}} \sum_{k=n}^{+\infty} \mathbb{P}(X = k) = \sum_n \mathbb{P}(X \geq n)$$

dove posso scambiare la serie e la sommatoria poiché entrambe convergono.

A questo punto usiamo il *Lemma di Borel-Cantelli*, che dice che data una successione numerabile di eventi S_n tali che $\sum \mathbb{P}(S_n) < +\infty$ allora $\mathbb{P}(\limsup S_n) = 0$. Nel nostro caso, consideriamo $S_n = (R_1 \geq \varepsilon n)$ e poiché tali eventi sono crescenti in n per continuità dal basso posso dire che $\mathbb{P}(\frac{R_n}{n} \geq \varepsilon) = 0$ da cui, per arbitrarietà di $\varepsilon > 0$ ho la prima tesi del punto 1.

Per quanto riguarda il secondo risultato, notiamo innanzitutto che per $a \in (0, 1)$ vale $\log(1-a) < -a$ (ovvio da un semplice studio di funzione). Definiamo $N_0 = \sup(n \in \mathbb{N} | R_n \geq \varepsilon n) < +\infty$ q.c.; allora posso dire $\forall c \in (0, 1-2\varepsilon)$ valgono le seguenti stime:

$$\begin{aligned} \sum_{n \geq 1} e^{R_n} c^n &= \sum_{n \geq 1} e^{R_n + n \log c} < \sum_{n \geq 1} e^{R_n + n \log(1-2\varepsilon)} \leq \sum_{n \geq 1} e^{R_n - 2n\varepsilon} \leq \\ &\leq \sum_{n \leq N_0} e^{R_n - 2n\varepsilon} + \sum_{n > N_0} e^{-\varepsilon n} < +\infty \end{aligned} \quad (3.2)$$

dove la prima disuguaglianza è dovuta alla proprietà di crescita del logaritmo sui numeri reali, la seconda si ricava dalla proprietà sopra elencata e la terza si basa sulla definizione di N_0 . Le due parti dell'ultimo termine nella catena di disuguaglianze sono entrambe finite, poiché la prima è una somma finita e la seconda serie è convergente per il criterio del confronto. Dall'arbitrarietà di ε si ricava che la proprietà dimostrata vale per ogni $c \in (0, 1)$.

2. Consideriamo ora il caso in cui $\mathbb{E}(R_1) < +\infty$, e sia $C > 1$ un numero reale. Usiamo ora un ragionamento analogo a quello visto nel punto 1, per cui vale

$$\sum_{n \geq 0} \mathbb{P}(R_n \geq Cn) = \sum_{n \geq 0} \mathbb{P}(R_1 \geq Cn) \geq \frac{1}{C} \sum_{n \geq C} \mathbb{P}(R_1 \geq n) \geq \frac{\mathbb{E}(R_1)}{C} = +\infty \quad (3.3)$$

dove la serie finale parte da C poiché la quantità Cn è sempre almeno C .

A questo punto usiamo il *secondo Lemma di Borel-Cantelli*, che dice che data una successione numerabile di eventi indipendenti E_n tali che la somma delle loro probabilità sia infinita, allora $\mathbb{P}(\limsup_n E_n) = 1$. Nel caso in esame gli eventi indipendenti che consideriamo sono $E_n = (R_n \geq Cn)$ che so essere indipendenti poiché le variabili R_n lo sono, e ho quindi

$$\mathbb{P}(\limsup R_n \geq Cn) = 1 \implies \limsup \frac{R_n}{n} \geq C \text{ q.c. } \forall C > 1.$$

Per dimostrare la seconda parte della tesi è sufficiente considerare che, fissato $C > 1$, $\forall c > \frac{1}{C}$ si ha che

$$\sum_{n > 0} e^{R_n} c^n \geq \sup_{n \geq 0} e^{R_n} c^n \geq \sup_{n \geq 0} (Cc)^n = +\infty \quad (3.4)$$

poiché considerando l'estremo superiore $e^{R_n} \geq e^{Cn} \geq C^n$ dato che essendo $C > 1$ vale $E^C > C$. Per arbitrarietà di C concludo che la proprietà vale per ogni $c > 0$.

Abbiamo pertanto completato la dimostrazione del lemma e con questa possiamo procedere a dimostrare il teorema. \square

Procediamo ora con la dimostrazione del teorema 3.1.1:

Dimostrazione. Partiamo dal dimostrare cosa succede nel caso $\mathbb{E}(\log^+(Y)) = +\infty$. Sotto questa ipotesi, ricordando che $Z_n \geq Y_n$, possiamo procedere a dimostrare che $\lim_{n \rightarrow \infty} \frac{Z_n}{c^n} = +\infty$ quasi certamente $\forall c > 0$. Infatti, $\forall M > 0$, dall'ipotesi sulla non finitezza del valore atteso della parte positiva di $\log(Y)$ e dal lemma precedente ho che $\limsup \frac{\log(Y_n)}{n} > \frac{\log M}{n} + \log c \iff \limsup \log(Y_n) > \log(Mc^n) \iff \limsup \frac{Y_n}{c^n} > M$ da cui la tesi poiché $Y_n \leq Z_n$, e ciò vale $\forall M > 0$ da cui la condizione sul limite voluta, e $\forall c > 0$ come volevamo dimostrare.

Per quanto riguarda il caso dove il valore atteso è finito, consideriamo $Z_{n,k}$ il numero di discendenti alla generazione n degli immigrati nella popolazione alla generazione k , con $k < n$. Condizionando questa variabile aleatoria a Y_k , essa si comporta come un processo di Galton-Watson con Y_k radici e $n - k$ generazioni. Inoltre tale v.a. è indipendente dai processi di immigrazione nelle generazioni diversa dalla k -esima, quindi se definiamo $\mathcal{G} = \sigma\{Y_n, n \geq 1\}$ abbiamo che il valore atteso condizionato è pari a $\mathbb{E}(Z_{n,k}|\mathcal{G}) = \mathbb{E}(Z_{n,k}|Y_k) = Y_k \alpha^{n-k}$, dove l'ultima uguaglianza è legata al fatto che a ogni generazione ogni individuo ha in media α figli e quindi moltiplico il numero di individui immigrati (Y_k) per la media di figli per generazione elevato al numero di generazioni. A questo punto abbiamo che

$$Z_n = \sum_{k=0}^n Z_{n,k} \text{ con l'assunzione di chiamare } Y_0 \text{ il numero di individui presente alla nascita della popolazione (tale assunzione è assolutamente lecita in quanto anche una costante può essere vista come una v.a.), e quindi}$$

$$\mathbb{E}\left(\frac{Z_n}{\alpha^n}|\mathcal{G}\right) = \sum_{k \leq n} \mathbb{E}\left(\frac{Z_{n,k}}{\alpha^n}|\mathcal{G}\right) = \sum_{k \leq n} \frac{Y_k}{\alpha^k}.$$

Per il lemma precedente abbiamo che $\lim_{n \rightarrow +\infty} \sum_{k \leq n} \frac{Y_k}{\alpha^k} = \sum_{n \geq 0} \frac{Y_n}{\alpha^n} < +\infty$ q.c. e quindi per il *Lemma di Fatou Condizionale*, che posso applicare poiché ognu-

no dei termini della serie è non negativo, vale $\mathbb{E}(\liminf X_n | \mathcal{G}) \leq \liminf \mathbb{E}(X_n | \mathcal{G})$. Applicandolo al caso in esame, si ha che $\mathbb{E}(\liminf_n \alpha^{-n} Z_n | \mathcal{G}) \leq \liminf_n \mathbb{E}(\alpha^{-n} Z_n | \mathcal{G}) = \sum_n \frac{Y_n}{\alpha^n} < +\infty$, da cui ricaviamo che $\mathbb{P}(\liminf_{n \rightarrow \infty} \alpha^{-n} Z_n = +\infty) = 0$ (poiché il suo valore atteso condizionato è finito), ma allora quasi certamente il $\liminf \frac{Z_n}{\alpha^n}$ è finito (poiché non negativo).

A questo punto dobbiamo dimostrare che il processo stocastico $\frac{Z_n}{\alpha^n}$ converge quasi certamente, e per farlo utilizziamo il teorema di convergenza delle martingale visto nella sezione 1.1.1. Tale teorema richiede, per avere la convergenza, che il $\sup \mathbb{E}(\alpha^{-n} Z_n) < +\infty$ e che tale processo sia una submartingala. Per dimostrare quest'ultima proprietà consideriamo $\mathbb{E}(Z_{n+1} | Z_1, \dots, Z_n, \mathcal{G}) = \alpha Z_n + Y_{n+1}$, vero grazie al lemma 2.2.2 (dove abbiamo usato uno dei risultati intermedi della dimostrazione) e che ci dice che dividendo per α^{n+1} otteniamo una submartingala rispetto alla sua filtrazione naturale \mathcal{G} , poiché $Y_j \geq 0 \forall j \in \mathbb{N}$. La limitatezza del valore atteso è già stata dimostrata e quindi possiamo applicare il suddetto teorema e ottenere che $\frac{Z_n}{\alpha^n}$ converge quasi certamente a una variabile aleatoria M □

3.2 Alberi spinali

Si può affrontare il problema dei processi di Galton-Watson con immigrazione all'interno dell'ambiente degli alberi di Bienaymé.

Definizione 3.2.1. Un albero spinale è una coppia (t, p) tale che $t \in \mathcal{T}$, cioè lo spazio degli alberi planari e $p = p_0, p_1, \dots$ un cammino sui nodi con radice nel nodo p_0 . Lo spazio degli alberi spinali si indica con \mathcal{T}^* . Nel caso in esame lavoriamo sempre con nodo iniziale di p la radice dell'albero. Definiamo con la notazione $p_{\leq n}$ la troncatura al livello n -esimo del cammino p .

Il nostro obiettivo è costruire un albero che rappresenti la popolazione che stiamo studiando in cui il cammino p (che d'ora in avanti sarà chiamato anche *spina* dell'albero) rappresenti il numero di immigrati a ogni generazione. Per

poter fare ciò necessitiamo di introdurre il concetto di versione *size-biased* di una variabile aleatoria.

Definizione 3.2.2. La *versione size-biased* di una variabile aleatoria X con distribuzione μ è una v.a. \widehat{X} tale che $\mathbb{P}(\widehat{X} \in A) = \frac{\mathbb{E}(X \mathbf{1}_A(X))}{\mathbb{E}(X)}$ con $A \subseteq [0, +\infty)$ intervallo, condizione che per note proprietà di approssimazione delle funzioni Boreliane con funzioni caratteristiche e per linearità del valore atteso può essere riformulata con qualsiasi funzione boreliana $f : [0, +\infty) \rightarrow [0, +\infty)$.

Un buon modo per visualizzare questa modifica nel caso di v.a. a valori nei numeri naturali è il seguente: supponiamo di avere un'urna di palline numerate, tale che la probabilità che esca la pallina j sia q_j , con $\sum_{j=0}^{+\infty} jq_j < +\infty$, la versione *size-biased* di questa probabilità consiste della probabilità di uscita delle varie palline etichettate quando nell'urna sono presenti j palline numerate con l'etichetta $j \forall j$. Da ciò possiamo ricavare una legge esplicita per la probabilità *size-biased*: relativamente alla notazione sopra introdotta, $\mathbb{P}(\widehat{X} = j) = \frac{\mathbb{E}(X \mathbf{1}_{(X=j)}(X))}{\mathbb{E}(X)} = \frac{jq_j}{\mathbb{E}(X)}$ che è coerente poiché la somma su tutti i possibili valori di j al numeratore restituisce esattamente il valore atteso della v.a. X , quindi la misura così introdotta è effettivamente una misura di probabilità (il caso in cui X sia a valori reali si affronta in modo analogo, ma non essendo di nostro interesse non lo espandiamo). Definiamo questa nuova misura $\widehat{\mu}$.

Possiamo ora parlare della costruzione di alberi di Bienaymé *size-biased*, che si realizza scegliendo randomicamente un vertice per ogni generazione dell'albero di Bienaymé "normale" associato, e associando a esso una distribuzione della progenie con legge la versione *size-biased* (con argomento opportunamente modificato come vedremo dopo) della progenie degli individui della popolazione.

Costruiamo ora in modo esplicito la misura *size-biased* sugli alberi. Per farlo dobbiamo lavorare sugli alberi spinali, in questo modo: consideriamo una famiglia di v.a. $(X_v, v \in \mathcal{U})$ i.i.d. con distribuzione μ , un vettore

di numeri naturali $y = (y_n, n \geq 1)$ e un altro vettore $i = (i_n, n > 0)$ con $1 \leq i_n \leq y_n \forall n$, definiamo ora $P_n = P_n(i) = i_1 i_2 \dots i_n$ il cammino di nodi determinato dal vettore i troncato all' n -esima generazione, definiamo inoltre $P = P_0, P_1, \dots$ la collezione di questi cammini troncati al crescere delle generazioni che nel limite approssima il vettore i .

Definiamo un albero casuale che dipende da queste grandezze e che contiene il cammino P , lo nominiamo $T = T(X, y, i)$:

- $\emptyset \in T, p_0 = \emptyset$
- per $n \geq 0$, noto l'albero fino al livello n che è $T_{\leq n}$, sia $c(p_n; T) = y_{n+1}$ (ciò mi garantisce che il cammino prosegue all'interno dell'albero poiché $i_{n+1} \leq y_{n+1}$ quindi il nodo successivo di P esiste nell'albero), mentre per gli altri nodi v nel livello n -esimo sia X_v il loro numero di figli.

In pratica, tutti i nodi di un livello hanno distribuzione identica eccettuato il nodo nel cammino, la cui distribuzione chiameremo Y laddove la progenie di un membro della popolazione ha distribuzione X .

Ci interessa a questo punto definire, in modo analogo a quanto fatto per gli alberi planari, delle classi di equivalenza.

Definizione 3.2.3. Dato \mathcal{T}^* , per ogni $n \in \mathbb{N}, t \in \mathcal{T}, v \in t_n$ sia

$$[t, v]_{\leq n} = \{(t', p') \in \mathcal{T}^* \mid t'_{\leq n} = t_{\leq n}, p'_n = v\}.$$

Possiamo a questo punto definire una filtrazione $\mathcal{F}_n^* = \sigma\left\{\bigcup_{i=0}^n \{[(t, p)]_{\leq i} \mid (t, p) \in \mathcal{T}^*\}\right\}$ (dove la notazione (t, p) identifica una classe di equivalenza dove il cammino coincide dalla radice all' n -esimo livello) e sia $\mathcal{F}^* = \sigma\left\{\bigcup_{n \geq 0} \mathcal{F}_n^*\right\}$.

Il fatto che $(\mathcal{F}_n^*)_{n \geq 0}$ sia una filtrazione è ovvio, e si vede facilmente che tale filtrazione raffina la filtrazione $(\mathcal{F}_n)_{n \geq 0}$ definita nella sezione 1.4.2, poiché la contiene e non ne è contenuta, dato che fissati due vertici diversi nello stesso livello dello stesso albero t essi danno luogo a due classi di equivalenza diverse in \mathcal{F}_n^* , ma appartengono alla stessa classe di equivalenza in \mathcal{F}_n .

Mostriamo ora esplicitamente, in termini di variabili aleatorie, come costruire una misura di probabilità sugli alberi spinali.

Proprietà 3.2.1 (*Cambio spinale di misura*). Sia data una distribuzione a valori nei numeri naturali μ , e sia $\hat{\mu}$ la sua versione size-biased. Notiamo che se \hat{X} è una v.a. con distribuzione $\hat{\mu}$, allora $\mathbb{P}(\hat{X} \geq 1) = 1$. Definiamo ora la misura ν a valori nei numeri naturali, tale che $\nu(i) = \hat{\mu}(i - 1)$. Tale misura è la distribuzione di una variabile aleatoria con insieme immagine i numeri naturali positivi e rappresenta quindi il processo di immigrazione che stiamo cercando di modellizzare.

Siano $X = (X_v, v \in \mathcal{U})$ v.a. i.i.d. con legge μ e $\mathbb{E}(X) = \alpha$, $Y = (Y_n, n > 0)$ i.i.d. con legge ν e $Z = (Z_n, n > 0)$ i.i.d. con legge $Unif([0, 1])$, con X, Y, U mutualmente indipendenti tra di loro. Ora $\forall n > 0$ sia $I_n = \lceil (Y_{n+1}Z_n) \rceil$, cioè I_n rappresenta un elemento uniformemente casuale dell'insieme $\{1, \dots, Y_{n+1}\}$ (ciò permette di definire stocasticamente il vettore i che era stato precedentemente definito per stabilire il cammino).

Chiamiamo ora \hat{B}_μ^* la legge della distribuzione dell'albero spinale $(T, P) = (T(X, Y, I), P(I))$ (comprendente quindi sia l'albero sia il cammino), mentre sia B_μ la distribuzione dell'albero di Bienaymé senza spina.

Allora per ogni albero spinale (t, p) vale che $\hat{B}_\mu^*(t_{\leq n}, p_{\leq n}) = \frac{B_\mu(t_{\leq n})}{\alpha^n}$

Dimostrazione. Sia (T, P) definito come sopra, allora abbiamo che $\hat{B}_\mu^*(t_{\leq n}, p_{\leq n}) = \mathbb{P}(T_{\leq n} = t_{\leq n}, P_{\leq n} = p_{\leq n})$. Posso scomporre questo evento nel prodotto di eventi del tipo "il livello i -esimo di T e P coincide con quelli di t e p condizionato al fatto che i livelli da 1 a $i - 1$ coincidano". In questo modo la probabilità di cui sopra diventa il prodotto delle probabilità di questi eventi, poiché essi sono indipendenti tra loro e $A = (T_{\leq n} = t_{\leq n}, P_{\leq n} = p_{\leq n})$ è la loro intersezione come eventi.

Quindi per avere l'evento di cui sopra devono verificarsi per ogni livello le seguenti condizioni:

- $\forall i \geq 0$ il nodo i -esimo del cammino p_i deve avere il giusto numero di figli \rightarrow la probabilità di tale evento è $\hat{\mu}(c(p_i; t))$ poiché il nodo del cammino ha distribuzione size-biased, in quanto rappresenta il processo migratorio;

- devo estendere il cammino a p_{i+1} nel modo giusto \rightarrow la probabilità di tale evento è $\frac{1}{c(p_i; t)}$ poiché la scelta del $(i+1)$ -esimo nodo del cammino è uniforme su tutti i nodi figli dell' i -esimo nodo del cammino;
- tutti gli altri nodi di t_i (i -esimo livello dell'albero) devono avere il giusto numero di figli \rightarrow la probabilità di tale evento è $\prod_{v \in t_i, v \neq p_i} \mu(c(v; t))$.

Unendo queste cose otteniamo che

$$\begin{aligned} \mathbb{P}(T_{i+1} = t_{i+1}, P_{i+1} = p_{i+1} \mid (T_{\leq i}, P_{\leq i}) = (t_{\leq i}, p_{\leq i})) &= \\ \widehat{\mu}(c(p_i; t)) \frac{1}{c(p_i; t)} \prod_{v \in t_i, v \neq p_i} \mu(c(v; t)) &= \\ \frac{c(p_i; t)}{c(p_i; t)} \frac{1}{\alpha} \mu(c(p_i; t)) \prod_{v \in t_i, v \neq p_i} \mu(c(v; t)) &= \\ \frac{1}{\alpha} \prod_{v \in t_i} \mu(c(v; t)) & \end{aligned}$$

A questo punto uniamo questo risultato con la scomposizione dell'evento come intersezione di eventi indipendenti e abbiamo che

$$\begin{aligned} \widehat{B}_\mu^*(t_{\leq n}, p_{\leq n}) &= \mathbb{P}(T_{\leq n} = t_{\leq n}, P_{\leq n} = p_{\leq n}) = \\ \prod_{i=0}^{n-1} \left(\frac{1}{\alpha} \prod_{v \in t_i} \mu(c(v; t)) \right) &= \frac{1}{\alpha^n} B_\mu(t_{\leq n}) \end{aligned} \tag{3.5}$$

dove l'ultima uguaglianza deriva dalla definizione della misura B_μ definita sugli alberi di Bienaymé. Ciò prova la tesi voluta. □

Da questo risultato ne deriva un corollario che risulterà utile nella dimostrazione del Teorema di Kesten-Stigum.

Corollario 2. Per ogni $n \geq 0$ vale la seguente proprietà che lega la misura $B_{\hat{\mu}}$ sugli alberi size biased e B_{μ} sugli alberi planari: la derivata di Radon-Nykodim di $B_{\hat{\mu}}$ su B_{μ} è pari alla quantità M_n , cioè

$$\frac{dB_{\hat{\mu}}|_{\mathcal{F}_n^*}}{dB_{\mu}|_{\mathcal{F}_n}} = M_n = \frac{Z_n}{\alpha^n} \quad (3.6)$$

Dimostrazione. Per la definizione della misura B_{μ} sappiamo che per ogni sottoalbero $t \subset \mathcal{U}$ vale che $B_{\hat{\mu}}(t_{\leq n}) = \sum_p B_{\hat{\mu}}^*(t_{\leq n}, p)$ con p cammino che inizia nella radice e che arriva all' n -esimo livello. Per la struttura degli alberi, il numero di tali cammini è esattamente $|t_n|$, cioè il numero di elementi al livello n -esimo in quanto per ogni nodo esiste un unico cammino che lo connette con la radice. Usando la proprietà precedente e il fatto che $M_n(t) = \frac{Z_n(t)}{\alpha^n} = \frac{|t_n|}{\alpha^n}$, abbiamo che $B_{\hat{\mu}}(t_{\leq n}) = \sum_p B_{\hat{\mu}}^*(t_{\leq n}, p) = \frac{|t_n|}{\alpha^n} B_{\mu}(t_{\leq n}) = M_n(t) B_{\mu}(t_{\leq n})$ e da ciò discende il risultato poiché tale uguaglianza vale $\forall t \subset \mathcal{U}$. \square

Abbiamo a questo punto costruito la misura size-biased sugli alberi $B_{\hat{\mu}}$ che unita alla misura B_{μ} ci dà tutti gli elementi di cui abbiamo bisogno per il Teorema di Kesten-Stigum. Infatti il corollario che abbiamo appena provato ci dà una relazione tra queste due misure, e il cambio spinale di misura precedentemente introdotto sarà di importanza chiave nella dimostrazione.

3.3 Il Teorema di Kesten-Stigum

In questa sezione trattiamo il *Teorema di Kesten-Stigum*, risultato che analizza il comportamento asintotico di una popolazione con distribuzione della progenie con valore atteso superiore a 1, cioè una popolazione con probabilità di sopravvivenza non nulla. Il teorema fondamentale dei processi di ramificazione ci aveva dato condizioni sufficienti per l'estinzione, mentre ora è nostro interesse raffinare questo risultato. Lavoreremo d'ora in poi nell'ipotesi che la progenie di un individuo della popolazione sia distribuita come una v.a. X con legge μ e con $\mathbb{E}(X) = \alpha > 1$. Ricordiamo come ultima

cosa che $M_n = \frac{Z_n}{\alpha^n}$ e che $M = \lim_n M_n$ è il suo limite martingala che abbiamo dimostrato esistere quasi certamente.

Teorema 3.3.1. *Sia T un albero di Bienaymé B_μ -distribuito, siano M_n e M definite come sopra, allora le seguenti condizioni sono equivalenti:*

1. $\mathbb{P}(M = 0) = \mathbb{P}(|T| < +\infty)$
2. $\mathbb{E}(M) = 1$
3. $\mathbb{E}(X \log X) < +\infty$

Prima di procedere alla dimostrazione notiamo due fatti che serviranno poi.

Remark 1. • Se ω è un evento tale che $|T(\omega)| < +\infty$ allora $\exists n_0 \in \mathbb{N}$ tale che $M_n(T(\omega)) = 0 \forall n \geq n_0$ e quindi anche $M(T(\omega))=0$, e quindi si ha che $\mathbb{P}(M = 0) \geq \mathbb{P}(|T| < +\infty)$.

- Per il teorema di convergenza delle Martingale uniformemente integrabili, $\mathbb{E}(M_n) \rightarrow \mathbb{E}(M)$ se e solo se la v.a. M_n è uniformemente integrabile $\forall n \geq 0$ e quindi una quarta condizione equivalente del teorema è che la famiglia di v.a. $(M_n)_{n \geq 0}$ sia uniformemente integrabile (è immediato dimostrare l'equivalenza con la prima condizione del teorema)

Ci serve un lemma prima di procedere con la dimostrazione, che lega la prima condizione del Teorema di Kesten-Stigum con la funzione generatrice delle probabilità studiata nella sezione 1.3.

Lemma 3.3.2. *Dato un albero di Bienaymé T e la v.a. M definita come il limite delle v.a. $M_n = \frac{Z_n}{\alpha^n}$, allora vale una delle seguenti:*

- $\mathbb{P}(M = 0) = \mathbb{P}(|T| < +\infty)$;
- $\mathbb{P}(M = 0) = 1$.

Dimostrazione. Sia $i \in Z_1$ un nodo appartenente alla prima generazione della popolazione (radice esclusa), allora il sottoalbero radicato in i è anch'esso un albero di Bienaymé distribuito come quello di partenza. Se chiamiamo $M_n(i) = \frac{1}{\alpha^{n-1}} \#\{v \in T_n \mid i \text{ è antenato di } v\}$ (dove $\#$ rappresenta la cardinalità di un insieme), allora si può dimostrare che tale processo stocastico è una martingala (rispetto alla sua filtrazione naturale \mathcal{F}_n): ciò è chiaro se si considera che il numero di individui al livello $n + 1$ che discendono da un vertice specifico è distribuito come Z_n e quindi

$$\mathbb{E}(M_{n+1}^{(i)} | \mathcal{F}_n) = \mathbb{E}\left(\frac{Z_{n+1}^{(i)}}{\alpha^n} | \mathcal{F}_n\right) = \mathbb{E}\left(\frac{Z_n}{\alpha^n} | \mathcal{F}_n\right) = M_{n-1} = M_n^{(i)}$$

poiché il ragionamento visto sopra si può pensare anche al contrario e quindi la distribuzione della $n - 1$ -esima generazione è identica alla distribuzione della n -esima generazione condizionata alla discendenza da uno specifico vertice.

Tornando al lemma, possiamo dire che $M^{(i)}$ è il limite q.c. della famiglia di v.a. $(M_n^{(i)})$. Allora possiamo decomporre il limite martingala delle v.a. M_n riferite a tutto l'albero guardando ai sottoalberi figli degli individui della prima generazione, e si ha quindi che

$$M = \frac{1}{\alpha}(M^{(1)} + \dots + M^{(X(\emptyset))}) \quad (3.7)$$

Condizionalmente al fatto che la v.a. $X(\emptyset)$ sia pari a k , le v.a. $M^{(i)}$ sono copie indipendenti della v.a. M , e $M = 0 \iff M^{(1)} = 0, \dots, M^{(X(\emptyset))} = 0$ e pertanto abbiamo che

$$p = \mathbb{P}(M = 0) = \sum_{k \geq 0} \mathbb{P}(X(\emptyset) = k) \mathbb{P}(M = 0)^k = \mathbb{E}(p^X) = F(p)$$

che è la funzione generatrice delle probabilità.

Per quanto visto durante la dimostrazione della Proprietà fondamentale dei processi di ramificazione (sezione 2.2) e usando le proprietà di convessità e crescita della funzione generatrice delle probabilità viste nella sezione 1.3 le soluzioni dell'equazione $p = F(p)$ sono la probabilità di estinzione della popolazione oppure 1 (che è sempre soluzione). Da ciò abbiamo la tesi del lemma. \square

Possiamo quindi approssciare finalmente la dimostrazione del Teorema di Kesten-Stigum.

Dimostrazione. Siano come al solito $X \sim \mu$, $Y \sim \nu$, con $\nu(i) = \widehat{\mu}(i+1)$, sia $L = \log(Y+1)$. Si ha che $\mathbb{L} < +\infty \iff \mathbb{E}(\log^+ Y) < +\infty$, sfruttando la proprietà che il logaritmo di un prodotto è pari alla somma dei logaritmi e la linearità del valore atteso: esplicitando i conti si ha che

$$\begin{aligned} \mathbb{E}(L) &= \sum_{i=0}^{\infty} \nu(i) \log(i+1) = \sum_{i=0}^{\infty} \nu(i) \log(i) + \nu(i) \log\left(1 + \frac{1}{i}\right) \leq \\ &\leq \sum_{i=0}^{\infty} \nu(i) \log(i) + 2\nu(i) = \sum_{i=0}^{\infty} \nu(i) \log(i) + 2 < +\infty \iff \quad (3.8) \\ &\iff \sum_{i=0}^{\infty} \nu(i) \log(i) = \mathbb{E}(\log^+ Y) < +\infty \end{aligned}$$

dove abbiamo usato la parte positiva del logaritmo di Y per non avere problemi qualora Y dovesse valere 0.

Inoltre vale anche la seguente uguaglianza,

$$\mathbb{E}(X \log^+ X) = \sum_{j>0} j \log j \mu(j) = \alpha \sum_{j>0} \log j \frac{j \mu(j)}{\alpha} = \alpha \mathbb{E}(L)$$

che ci dice che la finitezza del valore atteso di $X \log^+ X$ è equivalente a quella di L .

Per il Teorema 3.1.1 abbiamo che $\mathbb{E}(X \log^+ X) < +\infty \iff \mathbb{E}(L) < +\infty$ implica $\lim_n \frac{Z_n}{\alpha^n} < +\infty$ quasi certamente, cioè $\mathbb{P}(\limsup_n M_n < +\infty) = 1$ e tale limite è la v.a. M .

Usiamo ora il corollario precedente (corollario 2, sez. 3.2) che ci dice che $M = \limsup_n M_n = \limsup \frac{d\widehat{B}_\mu|_{\mathcal{F}_n^*}}{dB_\mu|_{\mathcal{F}_n}} = \frac{d\widehat{B}_\mu}{dB_\mu}$. Allora ho che posso scrivere $\mathbb{E}(M) = \int M(t) B_\mu(dt) = B_\mu(M)$, e a questo punto ci serve un ultimo risultato che permette di concludere velocemente la dimostrazione.

Teorema 3.3.3. *Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio con misura, sia inoltre \mathbb{Q} una misura finita su (Ω, \mathcal{F}) . Sia $(\mathcal{F}_n)_{n \geq 0}$ una filtrazione tale che la σ -algebra generata dall'unione delle σ -algebre della filtrazione è \mathcal{F} . Sia $\mathbb{P}_n = \mathbb{P}|_{\mathcal{F}_n}$ e*

$\mathbb{Q}_n = \mathbb{Q}|_{\mathcal{F}_n}$, e sia inoltre $\mathbb{Q}_n \ll \mathbb{P}_n$. Sia $X_n = \frac{d\mathbb{Q}_n}{d\mathbb{P}_n} : \Omega \rightarrow [0, +\infty)$ la corrispondente derivata di Radon-Nykodim. Allora posto $X = \limsup_n X_n$, si ha che $\mathbb{Q} = X\mathbb{P} + \mathbb{Q}\mathbf{1}_{(X=\infty)}$

Dimostrazione. Sia π la media aritmetica di \mathbb{P} e \mathbb{Q} , e sia π_n la sua restrizione a \mathcal{F}_n , che risulta quindi essere $\pi_n = \frac{\mathbb{P}_n + \mathbb{Q}_n}{2}$.

Allora $\mathbb{P}, \mathbb{Q} \ll \pi$ e $\mathbb{P}_n, \mathbb{Q}_n \ll \pi_n$ poiché se $\pi(A) = 0 \implies \mathbb{P}(A) = \mathbb{Q}(A) = 0$.

Siano $U_n = \frac{d\mathbb{Q}_n}{d\pi_n}$, $V_n = \frac{d\mathbb{P}_n}{d\pi_n}$ le derivate di Radon-Nykodim rispetto alla misura π introdotta, $U = \limsup U_n$, $V = \limsup V_n$. Poiché $\mathbb{Q}_n \ll \pi$ dal

Lemma che vedremo dopo si ha che $U_n \rightarrow U$ π -q.c. e $V_n \rightarrow V$ π -q.c., che

$\mathbb{Q} = U\pi$ e $\mathbb{P} = V\pi$.

Ora π -q.c. vale che $U_n + V_n = \frac{d\mathbb{Q}_n}{d\pi_n} + \frac{d\mathbb{P}_n}{d\pi_n} = 2 \cdot \frac{d\pi_n}{d\pi_n} = 2$.

Quindi $\pi(U + V = 0) = \pi(\limsup(U_n + V_n) = 0) = \pi(2 = 0) = 0$, quindi π -q.c. ho $\frac{U}{V}$ ben definita (fa ∞ se $U = \infty$ e/o $V = 0$) e vale $\frac{U}{V} = \frac{\lim_n U_n}{\lim_n V_n} =$

$\lim_n \frac{U_n}{V_n} = \lim X_n = X$.

Posso scrivere $U = XV + U\mathbf{1}_{(V=0)} = XV + U\mathbf{1}_{(X=\infty)}$ e poiché $\mathbb{Q} = U\pi = XV\pi + \mathbf{1}_{(X=\infty)}U\pi = X\mathbb{P} + \mathbb{Q}\mathbf{1}_{(X=\infty)}$.

□

Lemma 3.3.4. *Nel contesto del Teorema precedente, se $\mathbb{Q} \ll \mathbb{P}$ si ha $\mathbb{Q} = X\mathbb{P}$ (con X che è il $\limsup X_n$).*

Dimostrazione. Ricordiamo anzitutto che $\mathbb{Q} = X\mathbb{P}$ vuol dire che $\forall E \in \mathcal{F}$, $\mathbb{Q}(E) = \int_E d\mathbb{Q} = \int_E X d\mathbb{P} = \mathbb{E}_{\mathbb{P}}[X\mathbf{1}_E]$. Sia \tilde{X} la derivata di Radon-Nykodim di \mathbb{Q} rispetto a \mathbb{P} , essa esiste e soddisfa $\mathbb{Q}(\tilde{X} = \infty) = 0$ (per finitezza della misura \mathbb{Q}).

Sappiamo che vale per definizione di derivata di Radon-Nykodim $\mathbb{E}_{\mathbb{Q}}(\mathbf{1}_E) = \int_{\mathbb{R}} \mathbf{1}_E d\mathbb{Q} = \int_E \tilde{X} d\mathbb{P} = \mathbb{E}_{\mathbb{P}}[\tilde{X}\mathbf{1}_E]$, e quindi possiamo concludere che

$\forall E \in \mathcal{F}_n$ abbiamo che $\mathbb{E}_{\mathbb{P}} = \int_E X_n d\mathbb{P} = \int_E X_n d\mathbb{P}_n = \int_E 1 d\mathbb{Q}_n = \int_E 1 d\mathbb{Q} = \mathbb{E}_{\mathbb{Q}}(\mathbf{1}_E)$ quindi $X_n = \mathbb{E}[\tilde{X}|\mathcal{F}_n]$ (ciò risulta evidente poiché la v.a. X_n ha esattamente il comportamento che ci si aspetta da un valore atteso condizionato di \tilde{X} rispetto a \mathcal{F}_n). Nei passaggi sopra esplicitati, la seconda uguaglianza è valida per la definizione di $d\mathbb{P}_n$, la terza per come si definisce la derivata di Radon-Nykodim e la quarta per la definizione di \mathbb{Q}_n).

Poiché $\sigma(\bigcup_n \mathcal{F}_n) = \mathcal{F}$, il Teorema di convergenza delle martingale non negative (introdotto nella sezione 1.1.1) ci dice che \mathbb{P} q.c. $X_n \rightarrow \mathbb{E}[\tilde{X}|\mathcal{F}_{\infty}] = \mathbb{E}[\tilde{X}|\mathcal{F}] = \tilde{X}$ e ciò implica che \mathbb{P} q.c. $\tilde{X} = X$.

Quindi $\mathbb{E}_{\mathbb{P}}[\tilde{X}\mathbf{1}_E] = \mathbb{E}_{\mathbb{P}}[X\mathbf{1}_E]$ e il risultato segue da quanto dimostrato poco sopra, poiché abbiamo che $\mathbb{E}_{\mathbb{Q}}(\mathbf{1}_E) = \mathbb{E}_{\mathbb{P}}[X\mathbf{1}_E]$. □

Torniamo alla dimostrazione di Kesten-Stigum.

Dal Teorema appena visto segue che $\mathbb{E}(M) = 1$ se e solo se $\hat{B}_{\mu}(M < \infty) = 1$ se e solo se $\mathbb{E}[X \log^+ X] < \infty$.

Ora se $\mathbb{E}[M] = 1$ si deve avere $\mathbb{P}(M = 0) < 1 \implies \mathbb{P}(M = 0) = \mathbb{P}(|T| < \infty)$ dal lemma 3.3.2. Per l'altra implicazione, $\mathbb{E}(X \log^+ X) = \infty \implies \hat{B}_{\mu}(M = \infty) = 1 \implies B_{\mu}(M = 0) = 1$ per il teorema 3.3.3 e quindi $\mathbb{P}(M = 0) = 1 \implies \mathbb{E}(M) = 0 < 1$, e ciò conclude la doppia implicazione, quindi abbiamo dimostrato la completa equivalenza delle tre condizioni del teorema di Kesten-Stigum grazie a tale doppia implicazione..

□

Conclusioni

In questo lavoro abbiamo ripercorso brevemente una possibile evoluzione dello studio dei processi di Galton-Watson applicato allo studio delle popolazioni. In un modello semplificato, dove abbiamo assunto che ogni individuo abbia una distribuzione della prole identica e indipendente dagli altri X a valori in \mathbb{N} (assunzione non aderente alla realtà, ma utile per popolazioni ristrette e di specie biologicamente semplici, dove vi siano pochi fattori esterni da tenere in conto nello studio delle popolazioni), abbiamo visto proprietà utili per la costruzione di tali modelli e ci siamo concentrati sullo studio della sopravvivenza di tali popolazioni, giungendo a determinare delle condizioni sufficienti per l'estinzione, legate solo alla distribuzione della prole di un singolo individuo e non all'andamento complessivo della popolazione, assai più difficile da modellizzare e studiare matematicamente: una popolazione si estingue quasi certamente quando il valore atteso del numero di figli di un singolo individuo è minore o uguale ad 1 (nel caso sia uguale, è necessario che $\mathbb{P}(X = 0) > 0$ altrimenti si ricade in un caso banale).

Questa considerazione però non ci permette di quantificare la probabilità di estinzione nel caso essa sia inferiore ad 1, e per ottenere ciò è necessario un ulteriore lavoro di rifinitura, introducendo processi di immigrazione nel nostro modello che permettono di aumentare la profondità del modello: difatti ci consentono di dire che il processo stocastico $\frac{Z_n}{\alpha^n} = M_n$, che rappresenta il numero di elementi all' n -esima generazione diviso per il valore atteso della v.a. che rappresenta la prole di un singolo individuo elevata all' n -esima potenza, è una supermartingala e quindi ha media crescente nel tempo.

Il teorema di Kesten-Stigum (vedi Cap.3.3) è il coronamento di tale lavoro di indagine, in quanto lega la probabilità di estinzione della popolazione alla probabilità che la v.a. M , limite q.c. delle variabili M_n prima introdotte, sia pari a 0; inoltre ci permette di dire che la condizione appena espressa è equivalente a due condizioni molto più semplici da dimostrare: il fatto che il valore atteso della v.a. M sia pari a 1 oppure che $\mathbb{E}(X \log^+ X) < +\infty$, che coinvolgendo solo la distribuzione di X , risulta di assai più semplice trattazione rispetto ad oggetti complessi quali M e il suo valore atteso.

In questo modo si ottiene una struttura chiara e concisa del fenomeno della sopravvivenza (e, per il rovescio della medaglia, dell'estinzione): escludendo casi banali, condizione necessaria per la sopravvivenza è che in media ogni individuo abbia almeno un figlio, mentre sotto le ipotesi del Teorema di Kesten-Stigum la condizione sufficiente per la sopravvivenza è che, nel limite di tempo che tende all'infinito, ad ogni generazione la quantità M_n non si annulli mai q.c.

La potenza di tale Teorema sta soprattutto nel fatto che, a differenza del Teorema fondamentale dei processi di ramificazione, esso permette uno studio quantitativo della probabilità di sopravvivenza di una popolazione, rendendolo *de facto* il completamento dello studio del comportamento asintotico in un modello di popolazione che rispetti le nostre ipotesi di lavoro.

Bibliografia

- [1] Louigi Addario-Berry, 2021, "*Lecture notes: 2021 CRM-PIMS Probability Summer School*".
- [2] Remco van der Hofstad, 2009, "*Random graphs and complex networks- Volume 1*", disponibile su <http://www.win.tue.nl/rhofstad/NotesRGCN>.
- [3] Russell Lyons e Yuval Peres, 2017 "*Probability on trees and networks*", edito da *Cambridge University Press*.
- [4] Eugene Seneta, 1970, "*On the supercritical Galton-Watson process with immigration*", pubblicato su "Mathematical Biosciences" da *Elsevier*.
- [5] Harry Kesten e Bernt P. Stigum, 1966, "*A limit theorem for multi-dimensional Galton-Watson processes*", pubblicato su "The Annals of Mathematical Statistics" da *JSTOR*.

Ringraziamenti

Qui possiamo ringraziare il mondo intero!!!!!!!!!!
Ovviamente solo se uno vuole, non è obbligatorio.