8. Probabilistic Analysis Info • November 8 • Cormen 115

Intro Vi bruger probabilistic analysis til at estimere køretider eller andre forbundne omkostninger med en algoritme. Vi antager en fordeling på vores inputs og prøver at approximere en avarage-case running time. Da hele udregningen er afhængig af vores initielle antagelser, er det vigtigt at vi er forsigtige. Hvis ikke vi kan lave fornuftige valg, så er probabilistic analysis ikke brugbart.

Indicator Random Variables Vi kan ofte udregne kørertid på en algoritme ved at udregne dens expected value på en random variable der beskriver kørertiden ved at tælle hvornår et event sker. Vi kan beskrive om et sådan event ved en indicator random variable der kan tage en binære valg og evaluerer til enten én hvis eventet sker eller nul hvis ikke.

$$X = X_1 + X_2 + ... + X_n, X_i \in 0, 1$$

Vi er derved interesseret i at finde den expected value af antallet af events.

$$E[X] = E[\sum_{i=0}^{n} X_i] = \sum_{i=0}^{n} E[X_i] = \sum_{i=0}^{n} X_i * pr(X_i)$$

Et eksempel er the hiring problem, lad os antage at vi har

$$n$$

kandidater hvor hver kandidat er vurderet til at have en score mellem

$$0..n$$

. Hver gang vi får en ny kandidat der er bedre end den foregående, ansætter vi personen. Under antagelse at rækkefølgen er tilfældig, hvor mange personer skal vi da ansætte? Givet overstående, kan vi modellere om en person bliver ansat med en random indikater variable. Vi udregner at en person i rækken bliver ansat med sandsynligheden

$$E[X_I] = pr(X_i) = 1/i$$

$$E[X] = \sum_{i=0}^{n} 1/i = ln(n)$$

Discrete Time Markov Chains En anden måde at udregne køretid på, er ved at bruge discrete markov chains. Det er en måde at beskrive transitioner på med en hvis sandsynlighed, hvor sandsynlighden for at lave en transition ved tiden t ikke tager tidligere states i betragtning. De er "independent". Givet sandynsligheden for at ende i forskellige states beskrevet med en row vector, kan vi med en transition matrix beskrive sandsynlighden for at være i næste step ved

$$p(t + n) = p(t) * P^n$$

Antag at vi har to glas med et lige antal m kugler i alt, hvor alle sammen starter i første glas. Vi ønsker at finde ud af, hvor mange gange vi skal flytte kugler rundt, sådan at der er lige mange kugler i hvert glas. Der er en uniform chance for at hvilken som helst kugle bliver valgt og der er 0.5 chance for at vi faktisk vælge at flytte kuglen efter vi har valgt den. Vi kan udfylde transitionsmatrixen, hvor i er antallet af kugler i første glas før en transition og j er efter.

Chancen for at flytte en bold fra den første kop til den anden

$$p_{i,j} = \frac{i}{2m}, j = i - 1$$

Chancen for at flytte en bold fra den anden kop til første

$$p_{i,j} = \frac{m-i}{2m}, j = i + 1$$

Hvis der er lige mange bolde i hver

$$p_{i,j} = \frac{1}{2}, j = i$$

Alle states der prøve at flytte mere end en bold

$$p_{i,j} = 0, |i - j| > 0$$

Vi vil nu forsøge at udregne det expected antal flytninger man skal lave for at komme til lighed. Lad

$$X_i$$

notere hvor mange skridt vi skal tage givet at der er i bolde i den første kop. Vi vil da gerne finde

$$E[X_m]$$

Vi ved allerede at

$$E[X_{\frac{m}{2}}] = 0$$

For hvis alle bolde er ligeligt fordelt, så behøver vi ikke flere træk. Ellers for en abitrær n

$$E[X_n] = pr_{n,n-1} \cdot (E[X_{n-1}] + 1) + pr_{n,n+1} \cdot (E[X_{n+1}] + 1), n = \frac{m}{n} + 1...m - 1$$

$$E[X_m] = 1 + E[X_{m-1}]$$

Vi kan nu løse expected value som ligninger med flere ubekendte.