Niech n_i wskazuje liczbę razy, gdy *i*-ty typ bloku wystąpi w ciągu wśród $n + 1 - k$ kolejnych *k*-gramów. Wtedy prawdopodobieństwo p_i dla *i*-tego bloku jest szacowane jako częstotliwość n_i/N, dla której godzimy się napisać $(n + 1 - k)$ po prostu jako N. Na przykład ciąg

CAAABBCBABBCABAACBACC

ma długość $n = 21$, a m jest równe 3. Istnieje $m^2 = 9$ możliwych digramów, a ponieważ $k = 2$, można je znaleźć wśród $N = 20$ kolejnych bloków o długości 2 w ciągu. Jednym z 9 możliwych do uzyskania digramów jest AB i występuje on trzy razy, a więc prawdopodobieństwo tego konkretnego bloku wynosi 3/20.

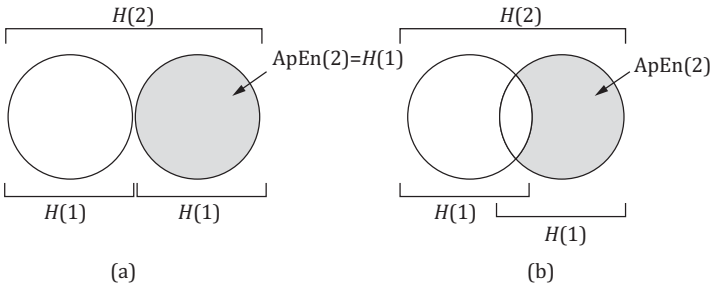
Nietrudno zobaczyć, że n_i muszą się sumować do N, gdy *i* zmienia się od 1 do m^k. Poniższe przybliżone wyrażenie jest często otrzymywane dla „entropii" $H(k)$ na blok o wielkości *k*:

$$H(k) = -\text{suma z } (n_i/N) \log(n_i/N),$$

gdzie indeks *i* zmienia się od 1 do m^k. $H(1)$ jest identyczne ze zwykłym wyrażeniem na H. Musicie być jednak świadomi, że „entropia" jest tu nadużyciem językowym, gdyż entropia zgodnie z wcześniejszą definicją dla H odnosi się do alfabetu symboli wybieranych niezależnie, podczas gdy bloki o wielkości *k* mogą być skorelowane ze względu na sekwencyjne zależności. Co więcej, entropia zakłada, że istnieje jakieś konkretne źródło generujące, podczas gdy teraz mamy tylko jeden ciąg z nieznanego źródła i możemy jedynie oszacować źródło ze z grubsza szacowanych prawdopodobieństw.

Mogę teraz wprowadzić kluczowe pojęcie *szacowanej entropii* ApEn(*k*) jako różnicę między $H(k) - H(k - 1)$, gdzie ApEn(1) jest po prostu równa $H(1)$. Idea polega na tym, że chcemy oszacować „entropię" bloku o długości *k* pod warunkiem, że znamy prefiks o długości $k - 1$. To daje *nowe informacje wniesione przez ostatni element bloku, pod warunkiem*,

że znamy jego poprzedniki w bloku. Jeśli ciąg ma dużą nadmiarowość, spodziewacie się, że wiedza danego bloku w dużym stopniu określa już następny symbol i bardzo niewiele informacji zostanie pozyskane. W tej sytuacji ApEn będzie mała. Można to pokazać schematycznie dla digramów z rysunku 2.4, na którym w jednym przypadku owale reprezentują średnią entropię pary nakładających się (sekwencyjnie zależnych) symboli oraz rozłączne (niezależne) symbole w drugim. Pojedyncze owale mają „entropie" $H(1)$, a ich związek ma „entropię" $H(2)$.



Rysunek 2.4. Schematyczna reprezentacja dwóch kolejnych symboli, które są sekwencyjnie niezależne (a) i zależne (b). Stopień ich nakładania się wskazuje na zakres sekwencyjnej zależności. Brak nałożenia oznacza niezależność. Średnia „entropia" na jeden dysk (czytaj: symbol) to $H(1)$, a średnia „entropia" digramu to $H(2)$, podczas gdy zacieniony obszar reprezentuje informacje, jakie wnosi drugi element digramu, który nie jest jeszcze zawarty w swoim poprzedniku, pod warunkiem znajomości pierwszego elementu pary, a konkretnie $H(2) - H(1) = \text{ApEn}(2)$

Zacieniony obszar na diagramie (a) reprezentuje nowe informacje wnoszone przez drugi element digramu, który nie jest jeszcze zawarty w swoim poprzedniku. Jest to średnia entropia ApEn(2) digramu, pod warunkiem znajomości pierwszego elementu pary. Zacieniony obszar na diagramie (b), a stąd i niepewność, jest maksymalny, gdy owale się nie nakładają.

Inny sposób, żeby zobaczyć, dlaczego ApEn(2) jest bliska zeru, gdy druga cyfra z pary w digramie jest w pełni określona przez poprzednią,