**Proposizione**

**Dimostrazione**

Per definizione di probabilità

Dato che ⇔

Dato che **abbiamo supposto che le parole sono indipendenti (Naive Bayes)** tra di loro e dato che se A e B sono indipendenti allora allora possiamo concludere che

Applicando gli stessi passaggi per n volte otterremo

Possiamo quindi concludere che

E la costante (una volta definite le parole ) per cui differiscono è

**Perché è necessario usare lo smoothing parameter?**

Supponiamo di dover classificare il messaggio:

***“Dear friend, you won money money money”***

E supponiamo di aver calcolato le seguenti probabilità:

**Probabilità per ham**

**P(friend|ham) = 0.8**

**…**

**P(money|ham) = 0.2**

**Probabilità per spam**

**P(friend|Spam) = 0**

**…**

**P(money|Spam) = 0.6**

E ragionevole pensare che:

***“Dear friend, you won money money money”***

Verra classificato come spam, ma questo non è vero.

**Se la parola friend non è presente in nessun messaggio spam**

**Per via della parola friend**, il messaggio avrà probabilità **P(Spam| *Dear friend … money*) = 0**

P(Spam| *Dear friend … money*) = P(Spam) P(Dear|Spam) P(friend|Spam)… P(money|Spam) = 0

