Interpolation vs. Backoff

- **Interpolation:** view p(w|REF) as a prior and the actual counts as observed evidence

$$p(w|d) = (1-\lambda)\frac{c(w,d)}{|d|} + \lambda\, p(w|REF)$$

- **Backoff** (Katz-Backoff): if the count is sufficiently high (sufficient evidence), we'd trust the ML estimate, otherwise, we simply ignore the ML estimate and go for p(w|REF)

$$p(w|d) = \begin{cases} \beta\dfrac{c(w,d)}{|d|} & if\ \ c(w,d) > k \\ \lambda p(w|REF) & otherwise \end{cases}$$