Interlude: avoiding rounding problems

- $L(a_1, a_2, \ldots, a_n|p) = \prod_{i=1}^{n} p^{a_i}(1-p)^{1-a_i} = p^{n_1}(1-p)^{n_0}$
- $$TPr(1, j, h) = \sum_{l=2}^{j-1} TPr(1, l, h-1) L(D(l+1..j), 1).$$
- $L(D(i..j))$ is very small; rounding errors abound
- Compute with logarithms
- Products are easy
- How to sum?

Sums for logs

We know $\log x$ and $\log y$, want to compute $log(x+y)$.

Assume $y < x$.

$$\begin{aligned}
\log(x+y) &= \log(e^{\log x} + e^{\log y}) \\
&= \log(e^{\log x}(1 + e^{\log y - \log x})) \\
&= \log x + \log(1 + e^{\log y - \log x})
\end{aligned}$$

Example

```
for i=1:n
    for j=i:n
        n0 = sum(A(i:j));
        cumllh(i,j,1) =  gamma(n0+1)*gamma(j-i+1-n0+1)/gamma(j-i+1+2);
        scaled(i,j,1) = gammaln(n0+1)+gammaln(j-i+1-n0+1)-gammaln(j-i+1+2);
    end
end

for b=2:k
    for i=1:n
        for j=i+1:n
            cumllh(i,j,b)=0;
            for a=i:j-1
                cumllh(i,j,b) = cumllh(i,j,b)+cumllh(i,a,b-1)*cumllh(a+1,j,1);
            end
            scaled(i,j,b)=scaled(i,i,b-1)+scaled(i+1,j,1);
            for a=i+1:j-1
                y = scaled(i,a,b-1)+scaled(a+1,j,1);
                if y<scaled(i,j,b)
                   scaled(i,j,b)=scaled(i,j,b)+log(1+exp(y-scaled(i,j,b)));
                else
                    scaled(i,j,b)=y+log(1+exp(scaled(i,j,b)-y));
                end
            end
        end
    end
end
```

6. Full Bayesian treatment of segmentation

- The same as in the previous section
- More carefully
- Data: 0-1 points $D = \{t_1, \ldots, t_n\}$
- Model: segments (ending points) and levels: $M = 8(e_1, \ldots, e_{k-1}), (p_1, \ldots, p_k))$
- Note: the model M contains both segments and levels
- Can consider models which have suboptimal levels for the segments

- For each point i let $r(i)$ be the unique h such that i belongs to segment h
- Likelihood of the data given the model

$$L(D|M) = \prod_{i=1}^{n} p_{r(i)}^{t_i} (1 - p_{r(i)})^{1-t_i}$$

- Log likelihood

$$LL(D|M) = \sum_{i=1}^{n} t_i \log p_{r(i)} (1 - t_i) \log(1 - p_{r(i)})$$

Model class

- A collection of models
- $\mathcal{M}(k)$: the set of all models with k segments and levels between 0 and 1
- $\mathcal{M}(k) = \{((e_1, \ldots, e_{k-1}), (p_1, \ldots, p_k)) \mid 1 \leq e_i < e_j \leq n, p_i \in [0, 1]\}$
- $\mathcal{M}$: all models from any $\mathcal{M}(k)$

Bayesian analysis

- Goal: look at all possible models (in $\mathcal{M}(k)$ or $\mathcal{M}$) and see what they tell about the data
- How likely is a model? $P(M|D)$
- Not just the single best model, but probability of a model given the data
- Bayes' rule:

$$Pr(M|D) = Pr(M \wedge D)/Pr(D)$$

$$= Pr(M)(Pr(M \wedge D)/Pr(M))/Pr(D)$$

$$= Pr(M|D) = Pr(M)Pr(D|M)/Pr(D)$$

- $Pr(M)$: prior of the model
- $Pr(D|M)$: likelihood of the data

Prior of the model

- $M = (S, \bar{p}) = ((e_1, \ldots, e_{k-1}), (p_1, \ldots, p_k))$
- Assume all segmentations are equally likely
  Want to have

$$\sum_S \int_{\bar{p}} Pr(S, \bar{p}) = 1$$

-

$$Pr(M) = \binom{n-1}{k-1}^{-1}$$

Probability of the data

- $P(D)$: the data comes from some model in the model class
- $$Pr(D) = \sum_{M \in \mathcal{M}(k)} Pr(M)Pr(D|M)$$
- Thus

$$Pr(M|D) = \frac{Pr(M)(Pr(D|M)}{Pr(D)} = \frac{Pr(M)(Pr(D|M)}{\sum_{M \in \mathcal{M}(k)} Pr(M)Pr(D|M)} = 1$$

- If we only want to compare two models, we don't have to know $Pr(D)$:

$$Pr(M_1|D) = Pr(M_1)Pr(D|M_1)/Pr(D)$$

$$Pr(M_2|D) = Pr(M_2)Pr(D|M_2)/Pr(D)$$

$Pr(D)$ for class $\mathcal{M}(k)$

$$
\begin{aligned}
Pr(D) &= \sum_{M \in \mathcal{M}(k)} Pr(M) Pr(D|M) \\
&= \sum_{S} \int_{pbar} Pr((s, \bar{p})) Pr(D|M) \\
&= \sum_{S} \binom{n-1}{k-1}^{-1} \int_{\bar{p}} Pr(D|S, \bar{p}) \\
&= \binom{n-1}{k-1}^{-1} \sum_{S} \int_{\bar{p}} Pr(D|S, \bar{p})
\end{aligned}
$$

$Pr(D)$ for class $\mathcal{M}(k)$, cont.

$$
\begin{aligned}
\int_{\bar{p}} Pr(D|S,\bar{p}) &= \int_{\bar{p}} \prod_{j=1}^{k} Pr(D_j|p_j) \\
&= \prod_{j=1}^{k} \int_{p_j} Pr(D_j|p_j)
\end{aligned}
$$

D_j: the data in segment j of S
$a(D_j)$: the number of 1s in D_j
$b(D_j)$: the number of 0s in D_j

$$
\begin{aligned}
\int_{p_j} Pr(D_j|p_j) &= \int_{p_j} p_j^{a(D_j)}(1-p_j)^{b(D_j)} \\
&= \frac{\Gamma(a(D_j)+1)\Gamma(b(D_j)+1)}{\Gamma(a(D_j)+b(D_j)+2)}
\end{aligned}
$$

$Pr(D)$ for class $\mathcal{M}(k)$, cont.

Denote

$$f(a(D_j), b(D_j)) = \frac{\Gamma(a(D_j)+1)\Gamma(b(D_j)+1)}{\Gamma(a(D_j)+b(D_j)+2)}$$

Then

$$Pr(D) = \binom{n-1}{k-1}^{-1} \sum_S \int_{\bar{p}} Pr(D|S, \bar{p})$$

and

$$Pr(D) = \binom{n-1}{k-1}^{-1} \sum_S \prod_{j=1}^{k} f(a(D_j), b(D_j))$$

Dynamic programming

- Define

$$Q(k, D) = \sum_{S,|S|=k} \prod_j f(a(D_j), b(D_j)) = Pr(D) \binom{n-1}{k-1}$$

- Thus $Q(1, D) = f(a(D), b(D))$
- Can we compute $Pr(D)$ efficiently?
- $Q(k, D) =$
$\sum_{m=1}^{n} Q(k-1, D(1..m))Q(1, D(m+1..n))$
- The same recurrence as before
- The probability of the 1 segment case is different

Application: the probability of a segment boundary

Model class $\mathcal{M}(k)$
$BP(i|D)$: the probability that i is the last point in a segment boundary
$[i \in M]$: 1, if i is a last point of a segment in M

$\mathcal{M}(k,i)$: k-segment models for D with i a last point
$\mathcal{M}(h,c,d)$: h-segment models for $D(c..d)$

$$
\begin{aligned}
BP(i|D) &= \sum_{M \in \mathcal{M}(k)} [i \in M] Pr(M|D) \\
&= \sum_{M \in \mathcal{M}(k,i)} Pr(M) Pr(D|M) / Pr(D)
\end{aligned}
$$

$M \in \mathcal{M}(k,i)$ if and only if M has h segments for $1..i$ and $k-h$ segments for $i+1..n$, for some h

$$
\begin{aligned}
& BP(i|D) \\
&= \sum_{M_1 \in \mathcal{M}(h,1,i)} \; \sum_{M_2 \in \mathcal{M}(k-h,i+1,n)} Pr(M_1 M_2|D)/P(D) \\
&= \sum_{M_1} \sum_{M_2} Pr(M_1 M_2) Pr(D_1|M_1) Pr(D_2|M_2)/P(D) \\
&= \binom{n-1}{k-1}^{-1} \sum_h \sum_{M_1} Pr(D(1..i)|M_1) \sum_{M_2} Pr(D(i+1..n)|M_2)/Pr(D) \\
&= \binom{n-1}{k-1}^{-1} \sum_h Q(h, D(1..i)) Q(k-h, D(i+1..n))/Pr(D)
\end{aligned}
$$

Probability of a breakpoint

$\mathcal{M} = \cup_k \mathcal{M}(k)$

The model class changes!

$\mathcal{S}(k,i)$: segmentations with k segments, one ending at i

$$
\begin{aligned}
BP(i|D) &= \sum_{M \in \mathcal{M}} [i \in M] Pr(M) Pr(D|M)/Pr(D) \\
&= \sum_k \sum_{M \in \mathcal{M}(k,i)} Pr(M) Pr(D|M)/Pr(D) \\
&= \sum_k \sum_{S \in \mathcal{S}(k,i)} \int_{\bar{p}} Pr(S,\bar{p}) Pr(D|S,\bar{p})/Pr(D)
\end{aligned}
$$

$Pr(S,\bar{p}) = \binom{n-1}{k-1}^{-1} 2^{-k}$, for example

$\sum_k \sum_S \int_{\bar{p}} Pr(S,\bar{p}) = 1$ (well, almost)

Model class $\mathcal{M}$

$$
\begin{aligned}
BP(i|D) &= \sum_k \sum_{M \in \mathcal{M}(k,i)} Pr(M)Pr(D|M)/Pr(D) \\
&= \sum_k \sum_{S \in \mathcal{S}(k,i)} \int_{\bar{p}} Pr(S,\bar{p})Pr(D|S,\bar{p})/Pr(D) \\
&= \sum_k 2^{-k} \binom{n-1}{k-1}^{-1} \sum_{S \in \mathcal{S}(k,i)} \int_{\bar{p}} Pr(D|S,\bar{p})/Pr(D) \\
&= \sum_k 2^{-k} \binom{n-1}{k-1}^{-1} \sum_h Q(h, D(1..i))Q(k-h, i+1..n)/Pr(D
\end{aligned}
$$

Model class $\mathcal{M}$

$$BP(i|D) = \frac{\sum_k 2^{-k} \binom{n-1}{k-1}^{-1} \sum_h Q(h, D(1..i))Q(k-h, i+1..n)}{\sum_k 2^{-k} \binom{n-1}{k-1}^{-1} Q(k, D)}$$