The effect of non-tightness on Bayesian estimation of PCFGs
Shay B. Cohen and Mark Johnson (Proceedings of ACL, 2013)

This text describes step-by-step procedures to indicate the difference that the three approaches for handling tightness of PCFGs in Bayesian estimation have on the posterior over syntactic structures with a uniform prior. See paper for more discussion of this Mathematica procedure.

General problem setting

In this note, we describe a step-by-step procedure in Mathematica that follows from section 8 in the paper.

We assume we have the following grammar:

\[
\begin{align*}
S &\rightarrow S \ S \ & (\text{rule probability } p) \\
S &\rightarrow S \ S \ & (\text{rule probability } q) \\
S &\rightarrow x \ & (\text{rule probability } 1-p-q)
\end{align*}
\]

We are going to calculate the posterior over the three possible trees for the string \(w = "x \ x \ x"\).

The three possible trees (t1, t2 and t3) are:

1. \((S \ (S \ x) \ (S \ x) \ (S \ x))\)
2. \((S \ (S \ x) \ (S \ (S \ x) \ (S \ x)))\)
3. \((S \ (S \ (S \ x) \ (S \ x)) \ (S \ x))\)

Note that the second and third trees have identical probabilities.

We assume a uniform prior over all grammar rules, so that the prior is proportional to a constant.

The measure of t1, t2 and t3 are:

\[
\begin{align*}
\text{measureT1} &:= p \ (1 - p - q)^3 \\
\text{measureT2} &:= q^2 \ (1 - p - q)^3 \\
\text{measureT3} &:= q^2 \ (1 - p - q)^3
\end{align*}
\]
We begin by calculating a function that computes the partition function for this grammar, following Chi (1999):

\[
\text{In[12]:= } \text{sol} = z / . \text{Solve}[z = p z^3 + q z^2 + 1 - p - q, z]
\]

\[
\text{Out[12]= } \left\{1, \frac{-p - q - \sqrt{4 p - 3 p^2 - 2 p q + q^2}}{2 p}, \frac{-p - q + \sqrt{4 p - 3 p^2 - 2 p q + q^2}}{2 p} \right\}
\]

\[
\text{In[2]:= } Z = \text{Min}[1, \text{If}[\text{sol[[2]]} < 0, 10000, \text{sol[[2]]}], \text{If}[\text{sol[[3]]} < 0, 10000, \text{sol[[3]]}]]
\]

\[
\text{Out[2]= Min}[1, \text{If}\left[\frac{-p - q - \sqrt{4 p - 3 p^2 - 2 p q + q^2}}{2 p} < 0, 10000, \text{sol[[2]]}\right],
\text{If}\left[\frac{-p - q + \sqrt{4 p - 3 p^2 - 2 p q + q^2}}{2 p} < 0, 10000, \text{sol[[3]]}\right]]
\]

### The sink-element approach

For the sink-element approach, to compute \(p(t_1 \mid w)\), \(p(t_2 \mid w)\) and \(p(t_3 \mid w)\), we need to integrate the probability of each over the probability simplex:

\[
\text{In[24]:= } \text{sinkElementT1} = \text{Integrate}[\text{measureT1}, \{p, 0, 1\}, \{q, 0, 1-p\}]
\]

\[
\text{Out[24]= } \frac{1}{120}
\]

\[
\text{In[16]:= } \text{sinkElementT2} = \text{Integrate}[\text{measureT2}, \{p, 0, 1\}, \{q, 0, 1-p\}]
\]

\[
\text{Out[16]= } \frac{1}{420}
\]

\[
\text{In[19]:= } \text{sinkElementT3} = \text{Integrate}[\text{measureT3}, \{p, 0, 1\}, \{q, 0, 1-p\}]
\]

\[
\text{Out[19]= } \frac{1}{420}
\]

\[
\text{In[21]:= } \text{sinkElementProbT1} = \text{sinkElementT1} / (\text{sinkElementT1} + \text{sinkElementT2} + \text{sinkElementT3})
\]

\[
\text{Out[21]= } \frac{7}{11}
\]

\[
\text{In[22]:= } \text{N}[\text{sinkElementProbT1}]
\]

\[
\text{Out[22]= } 0.636364
\]
The tight-only approach

For the tight-only approach, we need to create an indicator function that is 1 only if the grammar is tight.

\[
\text{indTight} = \begin{cases} 
1 & \text{if } Z < 1, \\
0 & \text{otherwise}.
\end{cases}
\]

\[\text{Out[26]} = \min\left[1, \min\left[\frac{-p - q - \sqrt{4 p - 3 p^2 - 2 p q + q^2}}{2 p} < 0, 10000, \text{sol[2]}\right], \right.\]

\[\left.\frac{-p - q + \sqrt{4 p - 3 p^2 - 2 p q + q^2}}{2 p} < 0, 10000, \text{sol[3]}\right]\right]

\( < 1, 0, 1\)

(By the way, it can be analytically shown that this grammar is tight when \( (p+1)/2 < q < 1-p \).)

Repeat the same integration, only factoring in the tightness:

\[\text{tightOnlyT1} = \int_{p=0}^{1} \int_{q=0}^{1-p} \text{measureT1 indTight} \, dp \, dq\]

\[\text{Out[27]} = \frac{1597}{466560}\]

\[\text{tightOnlyT2} = \int_{p=0}^{1} \int_{q=0}^{1-p} \text{measureT2 indTight} \, dp \, dq\]

\[\text{Out[28]} = \frac{1007}{1088640}\]

\[\text{tightOnlyT3} = \int_{p=0}^{1} \int_{q=0}^{1-p} \text{measureT2 indTight} \, dp \, dq\]

\[\text{Out[29]} = \frac{1007}{1088640}\]

\[\text{tightOnlyT1Prob} = \frac{\text{tightOnlyT1}}{\text{tightOnlyT1} + \text{tightOnlyT2} + \text{tightOnlyT3}}\]

\[\text{Out[31]} = \frac{11179}{17221} = 0.649149\]
The renormalization approach

The last thing we need to compute is the probability of $t_1$ with the renormalization approach. For that, we repeat the integration of the measure of $t_1$, $t_2$ and $t_3$, only this time factoring in the partition function:

```
In[33]:= 
renormalizationT1 = Integrate[measureT1 / Z, {p, 0, 1}, {q, 0, 1 - p}]

Out[33]= 
\frac{8109 + 160 \sqrt{3} \pi + 1280 \text{ArcCoth}[3] - 640 \text{Log}[2]}{699 840}

In[34]:= 
renormalizationT2 = Integrate[measureT2 / Z, {p, 0, 1}, {q, 0, 1 - p}]

Out[34]= 
\frac{13 509 + 3984 \sqrt{3} \pi + 73 792 \text{ArcCoth}[3] - 13 568 \text{Log}[2] - 11 664 \text{Log}[3]}{9 797 760}

In[35]:= 
renormalizationT3 = Integrate[measureT3 / Z, {p, 0, 1}, {q, 0, 1 - p}]

Out[35]= 
\frac{13 509 + 3984 \sqrt{3} \pi + 73 792 \text{ArcCoth}[3] - 13 568 \text{Log}[2] - 11 664 \text{Log}[3]}{9 797 760}

In[38]:= 
renormalizationT1Prob = 
renormalizationT1 / (renormalizationT1 + renormalizationT2 + renormalizationT3)

Out[38]= 
\frac{8109 + 160 \sqrt{3} \pi + 1280 \text{ArcCoth}[3] - 640 \text{Log}[2]}{699 840} \left(\frac{8109 + 160 \sqrt{3} \pi + 1280 \text{ArcCoth}[3] - 640 \text{Log}[2]}{699 840} + \frac{13 509 + 3984 \sqrt{3} \pi + 73 792 \text{ArcCoth}[3] - 13 568 \text{Log}[2] - 11 664 \text{Log}[3]}{4 898 880}\right)

In[39]:= 
N[renormalizationT1Prob]

Out[39]= 0.619893
```