この 1 年間弱、買い物で生じた小銭を貯金箱に貯めていたが、最近、その貯金箱が一杯になった。その総額を知りたいが硬貨を数えるのは手間がかかるので、硬貨の総重量を測り、その重さから総額を推定する方法を考える。

1　コインの総重量の生成モデル

帰宅した際に財布に小銭があれば、それを全て貯金箱に投入する。この行為を**投入**と呼ぶ。投入の回数を N とし、N は $\{N_{\min}, N_{\min}+1, \ldots, N_{\max}\}$ に値をとるカテゴリカル分布に従うと仮定する。$\theta_i := \Pr(N = i), \boldsymbol{\theta} := (\theta_{N_{\min}}, \theta_{N_{\min}+1}, \ldots, \theta_{N_{\max}})$ とし、$\boldsymbol{\theta}$ の事前分布として $\sum_i \theta_i = 1$ なる単体上の一様分布を仮定する。

i 回目の投入における金額を A_i とする。A_i は $\{1, 2, \ldots, 999\}$ の一様分布に従うとする。金額の発生分布は一様でもない気がするが他にアイデアもないのでこうした。i 回目の投入における p 円玉 ($p \in \{1, 5, 10, 50, 100, 500\}$) の枚数を $n_i^{[p]}$ とする。$n_i^{[p]}$ は、A_i から次のように計算する。

$$
\begin{aligned}
n_i^{[500]} &= A_i/500 \\
n_i^{[100]} &= (A_i - 500n_i^{[500]})/100 \\
n_i^{[50]} &= (A_i - 500n_i^{[500]} - 100n_i^{[100]})/50 \\
n_i^{[10]} &= (A_i - 500n_i^{[500]} - 100n_i^{[100]} - 50n_i^{[50]})/10 \\
n_i^{[5]} &= (A_i - 500n_i^{[500]} - 100n_i^{[100]} - 50n_i^{[50]} - 10n_i^{[10]})/5 \\
n_i^{[1]} &= A_i - 500n_i^{[500]} - 100n_i^{[100]} - 50n_i^{[50]} - 10n_i^{[10]} - 5n^{[5]}
\end{aligned}
$$

ここで、割り算 / は、整数商の計算である。これは、合計金額が A_i となる小銭の組み合わせで最も枚数が少なくなる組み合わせを計算している。これも、必ずしも現実を反映しているとは限らないが他にアイデアもないのでこうした。

p 円玉の重さを $w^{[p]}$ とする。$w^{[p]}, n_i^{[p]}$ を使って、コインの総重量 W の計算式が書ける。

$$
W = \sum_{i=1}^{N} \sum_{p} w^{[p]} n_i^{[p]}
$$

重さの測定に誤差があると想定し、硬貨の総重量 X は W を期待値とするガンマ分布に従うとする。

2　尤度の計算式

記録されていない N, A_i を含めると、データの尤度 L, 対数尤度 l は

$$
\begin{aligned}
L(X, N, A_1, \ldots, A_N) &= \Gamma(X; \beta W, \beta) \left(\frac{1}{999}\right)^N \theta_N \\
l(X, N, A_1, \ldots, A_N) &= \log \Gamma(X; \beta W, \beta) - N \log 999 + \log \theta_N
\end{aligned}
$$

と書ける。β はガンマ分布のパラメータである。パラメータの取り方は、Stan での定義に合わせた。記録がない N, A_i について integrate out するとデータの尤度

$$
L(X) = \sum_{N=N_{\min}}^{N_{\max}} \sum_{i=1}^{N} \sum_{A_i=1}^{999} L(X, N, A_1, \ldots, A_N)
$$

が得られる。この尤度を計算する式を Stan で書けば、サンプリングできる。

3 logsumexp

上の尤度の対数を取ると

$$\begin{aligned}\log(L(X)) &= \log\left(\sum_{N=N_{\min}}^{N_{\max}}\sum_{i=1}^{N}\sum_{A_i=1}^{999} L(X,N,A_1,\ldots,A_N\right)\\ &= \log\left(\sum_{N=N_{\min}}^{N_{\max}}\sum_{i=1}^{N}\sum_{A_i=1}^{999} \exp(l(X,N,A_1,\ldots,A_N))\right)\end{aligned}$$

という `logsumexp` の形になる。2 つの配列 $\mathtt{a},\mathtt{b}$ に対して、

$$\begin{aligned}\mathtt{logsumexp}(\mathtt{a}{+}{+}\mathtt{b}) &= \log(\exp(\mathtt{a}[1])+\exp(\mathtt{a}[2])+\cdots+\exp(\mathtt{a}[m])+\exp(\mathtt{b}[1])+\exp(\mathtt{b}[2])+\cdots+\exp(\mathtt{b}[n]))\\ &\neq \log(\exp(\mathtt{a}[1])+\exp(\mathtt{a}[2])+\cdots+\exp(\mathtt{a}[m]))+\log(\exp(\mathtt{b}[1])+\exp(\mathtt{b}[2])+\cdots+\exp(\mathtt{b}[n]))\\ &= \mathtt{logsumexp}(\mathtt{a})+\mathtt{logsumexp}(\mathtt{b})\end{aligned}$$

つまり、logsumex を分割して計算することはできないことに注意。

4 和の個数

データの尤度を計算するために $\sum_{N=N_{\min}}^{N_{\max}}\sum_{i=1}^{N}\sum_{p}$ の計算が必要になる。この式が、幾つの項から成っているかを計算する。

$\sum_p$ で 6 個出てきて、$\sum_{i=1}^{N}\sum_p$ で $6N$ 出てくるので、全体では、

$$\sum_{N=N_{\min}}^{N_{\max}} 6N = 6\frac{N_{\min}+N_{\max}}{2}(N_{\max}-N_{\min}+1)$$