Supplemental Information

Here we give the closed form derivation for the information gain metric we use in our simulations. We are interested in computing

$$IG(e) = D_{KL}(B_T||B_S) - D_{KL}(B_T||B_{S+e}) \tag{1}$$

where the divergence measure is computed in closed form for e.g., B_T and B_S, as

$$\begin{aligned} D_{KL}(B_T||B_S) =& \log(\frac{B(\alpha_S, \beta_S)}{B(\alpha_T, \beta_T)}) + (\alpha_T - \alpha_S)\psi(\alpha_T) \\ &+ (\beta_T - \beta_S)\psi(\beta_T) + (\alpha_T - \alpha_S + \beta_T - \beta_S)\psi(\alpha_T + \beta_T). \end{aligned} \tag{2}$$

where ψ denotes the digamma function and $B(a, b)$ denotes the beta function. We can substitute Equation 2 into Equation 1 to get

$$\begin{aligned} IG(e) =& \log(\frac{B(\alpha_S, \beta_S)}{B(\alpha_T, \beta_T)}) + (\alpha_T - \alpha_S)\psi(\alpha_T) \\ &+ (\beta_T - \beta_S)\psi(\beta_T) + (\alpha_T - \alpha_S + \beta_T - \beta_S)\psi(\alpha_T + \beta_T) \\ &- \log(\frac{B(\alpha_{S+e}, \beta_{S+e})}{B(\alpha_T, \beta_T)}) - (\alpha_T - \alpha_{S+e})\psi(\alpha_T) \\ &- (\beta_T - \beta_{S+e})\psi(\beta_T) - (\alpha_T - \alpha_{S+e} + \beta_T - \beta_{S+e})\psi(\alpha_T + \beta_T) \end{aligned} \tag{3}$$

Consider the case where e is a series of h 1's (heads) and t 0's (tails). Then $\alpha_{S+e} = \alpha_S + h$ and $\beta_{S+e} = \beta_S + t$, so we can simplify Equation 3 to

$$\begin{aligned} IG(e) =& \log(\frac{B(\alpha_S, \beta_S)}{B(\alpha_T, \beta_T)}) - \log(\frac{B(\alpha_S + h, \beta_S + t)}{B(\alpha_T, \beta_T)}) \\ &+ (\alpha_T - \alpha_S)\psi(\alpha_T) + (\beta_T - \beta_S)\psi(\beta_T) \\ &+ (\alpha_T - \alpha_S + \beta_T - \beta_S)\psi(\alpha_T + \beta_T) \\ &- (\alpha_T - \alpha_S - h)\psi(\alpha_T) - (\beta_T - \beta_S - t)\psi(\beta_T) \\ &- (\alpha_T - \alpha_S - h + \beta_T - \beta_S - t)\psi(\alpha_T + \beta_T) \\ =& \log(\frac{B(\alpha_S, \beta_S)}{B(\alpha_T, \beta_T)} \cdot \frac{B(\alpha_T, \beta_T)}{B(\alpha_S + h, \beta_S + t)}) \\ &+ (h)\psi(\alpha_T) + (t)\psi(\beta_T) + (h + t)\psi(\alpha_T + \beta_T). \\ =& \log(\frac{B(\alpha_S, \beta_S)}{B(\alpha_S + h, \beta_S + t)}) \\ &+ (h)\psi(\alpha_T) + (t)\psi(\beta_T) + (h + t)\psi(\alpha_T + \beta_T). \end{aligned} \tag{4}$$

And, since

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}, \tag{5}$$

we can rewrite the first term and reduce:

$$
\begin{aligned}
IG(e) &= \log(\frac{\frac{\Gamma(\alpha_S)\Gamma(\beta_S)}{\Gamma(\alpha_S+\beta_S)}}{\frac{\Gamma(\alpha_S+h)\Gamma(\beta_S+t)}{\Gamma(\alpha_S+\beta_S+h+t)}}) \\
&\quad + (h)\psi(\alpha_T) + (t)\psi(\beta_T) + (h+t)\psi(\alpha_T+\beta_T). \\
&= \log(\frac{\Gamma(\alpha_S)\Gamma(\beta_S)\Gamma(\alpha_S+\beta_S+h+t)}{\Gamma(\alpha_S+h)\Gamma(\beta_S+t)\Gamma(\alpha_S+\beta_S)}) \\
&\quad + (h)\psi(\alpha_T) + (t)\psi(\beta_T) + (h+t)\psi(\alpha_T+\beta_T). \\
&= \log(\frac{\Gamma(\alpha_S)}{\Gamma(\alpha_S+h)}) + \log(\frac{\Gamma(\beta_S)}{\Gamma(\beta_S+t)}) + \log(\frac{\Gamma(\alpha_S+\beta_S+h+t)}{\Gamma(\alpha_S+\beta_S)}) \\
&\quad + (h)\psi(\alpha_T) + (t)\psi(\beta_T) + (h+t)\psi(\alpha_T+\beta_T). \\
&= -\log(\frac{\Gamma(\alpha_S+h)}{\Gamma(\alpha_S)}) - \log(\frac{\Gamma(\beta_S+t)}{\Gamma(\beta_S)}) + \log(\frac{\Gamma(\alpha_S+\beta_S+h+t)}{\Gamma(\alpha_S+\beta_S)}) \\
&\quad + (h)\psi(\alpha_T) + (t)\psi(\beta_T) + (h+t)\psi(\alpha_T+\beta_T).
\end{aligned} \tag{6}
$$

When $h = 0$, the first log term reduces to 0. When $t = 0$, the second log term reduces to 0. Otherwise, since for positive integer n,

$$
\frac{\Gamma(x+n)}{\Gamma(x)} = \prod_{k=0}^{n-1}(x+k), \tag{7}
$$

we can reduce the previous formulation a bit further when $h \neq 0$ and $t \neq 0$, to

$$
\begin{aligned}
IG(e) = &-\log(\prod_{i=0}^{h-1}(\alpha_S+i)) - \log(\prod_{j=0}^{t-1}(\beta_S+j)) + \log(\prod_{k=0}^{h+t-1}(\alpha_S+\beta_S+k)) \\
&+ (h)\psi(\alpha_T) + (t)\psi(\beta_T) + (h+t)\psi(\alpha_T+\beta_T).
\end{aligned} \tag{8}
$$

For computation, to prevent overflow we calculate the series equivalently as follows:

$$
\begin{aligned}
IG(e) = &\sum_{k=0}^{h+t-1}\log(\alpha_S+\beta_S+k) - \sum_{i=0}^{h-1}\log(\alpha_S+i) - \sum_{j=0}^{t-1}\log(\beta_S+j) \\
&+ (h)\psi(\alpha_T) + (t)\psi(\beta_T) + (h+t)\psi(\alpha_T+\beta_T).
\end{aligned} \tag{9}
$$