S2 Text. Mathematical Derivations.

# Derivation of equation (2)

We solve the constrained minimization problem (1) by seeking critical points of the Lagrange function given by

(S2.1.1)

where *X* denotes the set of all length L sequences, and denote the probabilities of sequence under PMFs and , and and are Lagrange multipliers associated with the constraints. Setting the partial derivatives to zero, and solving for gives:

(S2.1.2)

where is the normalization constant for the probability distribution.

As stated in the main text, we choose to be a product of identically distributed single nucleotide distributions:

where is the indicator function for a G or C nucleotide at position , and controls the average GC content of a sequence position. To fix we let denote the genome GC content and require

(S2.1.4)

(S2.1.3)

where denotes expectation with respect to and is an arbitrary sequence positon. Solving for gives

(S2.1.5)

Substituting given by (S2.1.3) with this value of into (S2.1.2) and absorbing the denominator of into , we obtain equation (2).

# Approximating the decay of

To estimate the decay of from its maximum in terms of quantities measurable from our MaxEnt samples of , we take logs of equation (5) to obtain

(S2.2.2)

(S2.2.1)

The first term of (S2.2.2) is the log of the marginal distribution of the input-wide feature when is distributed according to . We denote this marginal distribution as

(S2.2.3)

Similarly, the second term is the log of the marginal distribution of the input-wide feature when is distributed according to . We denote this marginal distribution as

(S2.2.4)

Our choice of *q* as a product of identical single nucleotide distributions (equation (S2.1.3)) implies that when is distributed according to *q*, is a linear combination of the independent random variables with weights (equation (4) main text). Assuming this linear combination involves a sufficient number of weights of roughly the same magnitude, we can apply the Lindeberg version of the Central Limit Theorem to approximate

where and denote the mean and variance of , respective, when is distributed according to *q*. These moments can be calculated directly from the simple form for *q* (equation (S2.1.3)).

(S2.2.5)

Approximation (S2.2.5) treats as a smooth function of . To approximate the decay of , we also treat as a smooth function of and expand in a second order Taylor Series about the value that maximizes

(S2.2.6)

.

The 1st order term is zero since we are expanding about a maximum, and we have substituted for . Truncation of higher order terms is justified if the distribution is approximately normal, a condition which can be checked directly by estimating the distribution of from MCMC samples. Moreover, in this case we can estimate

(S2.2.7)

where is the variance of under the distribution , estimated from MCMC samples. With this estimate and under the above approximations, we obtain:

which is equation (6).