При использовании метода максимальной энтропии истинное распределение видов по территории представляется в виде распределения вероятности pi в неком наборе X точек в пределах изучаемой территории. pi принимает неотрицательные значения в каждой точке территории и сумма pi =1 (по сути, pi – вероятность того, что в данной точке будет встречена особь интересующего нас вида). Нам необходима модель этого распределения, которая бы учитывала ограничения, накладываемые имеющимися данными о встречах видов. Эти ограничения выражены в виде простых функций от условий окружающей среды, которые называются features. Среднее значение каждой из таких features должно быть близко к значению, эмпирически полученному на основе фактических данных из точек встреч. Например, для feature «среднегодовое количество осадков», среднее значение, получаемое из модели должно быть близко к таковому, вычисленному на основе фактических данных по среднегодовому количеству осадков на данную территорию. И далее из всего семейства распределения вероятностей, удовлетворяющих имеющемуся набору ограничений мы выбираем модель с максимальной энтропией.

Для объяснения того, как распределение pi описывает реализованное распределение видов по территории, рассмотрим следующий идеализированный вариант. Допустим, наблюдатель выбирает случайный участок **x** из набора **X** и присваивает функции **y** значение 1, если на данном участке есть интересующий вид и 0, если вида нет. В этом случае можно представить **pi(x)** как условную вероятность **P(x|y=1),** т.е. вероятность того, что оказавшись в точке x, наблюдатель обнаружит искомый вид (функция y примет значение 1). Тогда, согласно правилу Байеса, можно записать:

$$\boldsymbol{P(y=1|x) = \frac{P(x|y=1)P(y=1)}{P(X)} = pi(x)P(y=1)|X|} \quad (1)$$

Теорема Байеса (или формула Байеса) — одна из основных теорем теории вероятностей, которая позволяет определить вероятность того, что произошло какое-либо событие (гипотеза) при наличии лишь косвенных тому подтверждений (данных), которые могут быть неточны. Названа в честь её автора, преп. Томаса Байеса (посвящённая ей работа «An Essay towards solving a Problem in the Doctrine of Chances» впервые опубликована в 1763 году, через 2 года после смерти автора). Полученную по формуле вероятность можно далее уточнять, принимая во внимание данные новых наблюдений.

$$\boldsymbol{P(A|B) = \frac{P(B|A)P(A)}{P(B)}} \quad (2)$$

P(A) — априорная вероятность гипотезы A;
P(A|B) — вероятность гипотезы A при наступлении события B;
P(B|A)— вероятность наступления события B при истинности гипотезы A;
P(B)— полная вероятность наступления события B.

Формула Байеса позволяет «переставить причину и следствие»: по известному факту события вычислить вероятность того, что оно было вызвано данной причиной. События, отражающие действие «причин», в данном случае обычно называют гипотезами, так как они — предполагаемые события, повлекшие данное. Безусловную вероятность справедливости гипотезы называют априорной (насколько вероятна причина вообще), а условную — с учетом факта произошедшего события — апостериорной (насколько вероятна причина оказалась с учетом данных о событии).

Вернёмся к формуле (1). P(y=1) – это общая распространённость видов на исследуемой территории. P(y=1|x)—вероятность того, что в точке x функция y примет значение 1. Согласно (1), искомая нами плотность распределения pi(x) пропорциональна вероятности встречи вида. С другой стороны, если у нас есть данные ТОЛЬКО о встрече видов (т.е. изначально мы не знаем, в каких точках y=0), то согласно (Phillips et al., 2006) мы не можем рассчитать общую распространённость видов. С другой стороны, мы будем пытаться определить плотность распределения pi(x).

В общем виде, распределение, используемое в рамках метода максимальной энтропии, относится к семейству распределений Гиббса (Распределение Гиббса представляет наиболее общую и удобную основу для построения равновесной статистической механики. Знание распределения частиц системы позволяет найти средние значения различных характеристик термодинамической системы по формуле математического ожидания. С учетом большого количества частиц в макроскопических системах эти математические ожидания с учетом закона больших чисел совпадают с реально наблюдаемыми значениями термодинамических параметров.**),** и имеет вид:

$$\boldsymbol{q_{\lambda}(x) = \frac{exp(\sum_{j=1}^{n} \lambda_j f_j(x))}{Z_{\lambda}}} \quad \textbf{(3)}$$

где $\boldsymbol{\lambda}$ – некий «вес», присваемые каждому из параметров **f** (те самые features, о которых говорилось выше), $\boldsymbol{Z_{\lambda}}$ **–** константа нормализации, используемая, чтобы сумма q(x) была равна 1. Таким образом, значение q для территории x зависит только от значений features в x, следовательно, от значений переменных окружающей среды в x. Помимо прочего это означает, что модель, полученная для строго определённого набора точек наблюдений может быть использована и для любого другого произвольного набора точек, для которых можно определить тот же список переменных окружающей среды.

Оценивается модель по следующей разнице:

$$\boldsymbol{\frac{1}{m}\sum_{i=1}^{m} \ln(q(x_i) - \sum_{j=1}^{n} \beta_j \lambda_j} \quad \textit{(4)}$$

Здесь β – регуляризующий параметр ширины допустимого интервала ошибки, а x1-xn – точки встречи видов. Первая часть разности, называемая «логарифмическое правдоподобие», увеличивается по мере того, как мы получаем более точное описание (фиттирование) исходных данных. Можно сделать интересное наблюдение: первый параметр больше для моделей, которые назначают более высокую вероятность для точек, где виды были встречены, чем для пустых точек, т.е. моделей, которые лучше позволяют различать точки встречи от фоновых. Вторая часть становится больше, если увеличиваются значения весов отдельных переменных, т.е. если модель усложняется. Идеальным считается ситуация, когда выражение (2) принимает максимальное значение из всех возможных.