$$\Pr(\boldsymbol{f}|\boldsymbol{e}) = \sum_{\boldsymbol{a}} \Pr(\boldsymbol{f},\boldsymbol{a}|\boldsymbol{e})$$

(3)

ここで，$\boldsymbol{f} = f_1, f_2, \dots, f_j, \dots f_m$であり，$\boldsymbol{f}$がある値（フランス単語（現言語の単語）列）に定まるとき，$\boldsymbol{m}$は同時に値が定まっている変数（つまり，定数）であるため，すなわち，

$$\Pr(\boldsymbol{f}|\boldsymbol{e}) = \sum_{\boldsymbol{a}} \Pr(\boldsymbol{f},\boldsymbol{a}|\boldsymbol{e}) = \sum_{\boldsymbol{a}} \Pr(\boldsymbol{f},m,\boldsymbol{a}|\boldsymbol{e})$$

と $\boldsymbol{m}$を追加してもイコール関係は保たれる．

ベイズの定理

$$\Pr(\mathrm{A},\mathrm{B}|\mathrm{C}) = \Pr(\mathrm{A}|\mathrm{B},\mathrm{C})\Pr(\mathrm{B}|\mathrm{C})$$

を適用すると，Σ内の$\Pr(\boldsymbol{f},m,\boldsymbol{a}|\boldsymbol{e})$の部分は以下のように書き換えられる．

$$\Pr(\boldsymbol{f},m,\boldsymbol{a}|\boldsymbol{e}) = \Pr(\boldsymbol{f},\boldsymbol{a}|m,\boldsymbol{e})\Pr(m|\boldsymbol{e})$$

(3.1)

ここで，

$\boldsymbol{f} = f_1, f_2, \dots, f_j, \dots f_m$

$\boldsymbol{a} = a_1, a_2, \dots, a_j, \dots a_m$

より，$\Pr(\boldsymbol{f},\boldsymbol{a}|m,\boldsymbol{e})$の **f** と **a** の部分を展開して書くと，

$$\begin{aligned}\Pr(\boldsymbol{f},\boldsymbol{a}|m,\boldsymbol{e}) &= \Pr\left(f_1, f_2, \dots, f_j, \dots f_m, a_1, a_2, \dots, a_j, \dots a_m \middle| m,\boldsymbol{e}\right)\\ &= \Pr\left(f_1, a_1, f_2, a_2 \dots, f_j, a_j, \dots f_m, a_m \middle| m,\boldsymbol{e}\right)\end{aligned}$$

さらに，$X_j = f_j, a_j$とおくと，

$$= \Pr\left(X_1, X_2, \dots, X_j, \dots X_m \middle| m,\boldsymbol{e}\right)$$

これに Chain rule

$$\Pr(X_1, X_2, \dots, X_k, \dots X_n) = \Pr(X_n|X_1^{n-1})\Pr(X_{n-1}|X_1^{n-2})\dots\Pr(X_2|X_1)\Pr(X_1) = \prod_{k=1}\Pr\left(X_k\middle|X_1^{k-1}\right)$$

を適用すると，

$$= \prod_{j=1}\Pr\left(X_j\middle|X_1^{j-1}, m,\boldsymbol{e}\right)$$

X_jをf_j, a_jに戻すと，

$$= \prod_{j=1}\Pr\left(f_j, a_j\middle|f_1^{j-1}, a_1^{j-1}, m,\boldsymbol{e}\right)$$

さらにベイズの定理を使うと，

$$= \prod_{j=1}\Pr\left(f_j\middle|a_j, f_1^{j-1}, a_1^{j-1}, m,\boldsymbol{e}\right)\Pr\left(a_j\middle|f_1^{j-1}, a_1^{j-1}, m,\boldsymbol{e}\right)$$

$$= \prod_{j=1}\Pr\left(f_j\middle|f_1^{j-1}, a_1^{j}, m,\boldsymbol{e}\right)\Pr\left(a_j\middle|f_1^{j-1}, a_1^{j-1}, m,\boldsymbol{e}\right)$$

(3.1)式へ上の式を戻してやれば（ついでに条件部の a と f の順番をさらっと入れ替えて，$\Pr(m|\boldsymbol{e})$ もさらっと前に出してやれば），

$$\Pr(\boldsymbol{f},\boldsymbol{a}|\boldsymbol{e}) = \Pr(m|\boldsymbol{e})\prod_{j=1}^{m}\Pr\left(a_j\middle|a_1^{j-1},f_1^{j-1},m,\boldsymbol{e}\right)\Pr\left(f_j\middle|a_1^{j},f_1^{j-1},m,\boldsymbol{e}\right) \tag{4}$$

が一般性を失うことなく求められた．

また，$\boldsymbol{a}$や$\boldsymbol{f}$といった大まかな塊でなく，その中のa_jやf_jといった細かい部分は直接 m によらないとすれば，上式はさらに以下のように書き換えられる．

$$\Pr(\boldsymbol{f},\boldsymbol{a}|\boldsymbol{e}) = \Pr(m|\boldsymbol{e})\prod_{j=1}^{m}\Pr\left(a_j\middle|a_1^{j-1},f_1^{j-1},\boldsymbol{e}\right)\Pr\left(f_j\middle|a_1^{j},f_1^{j-1},\boldsymbol{e}\right)$$

（コロナ社の「機械翻訳」の(4.6)式に対応）