변분 베이지안 방법은 베이지안 추론과 기계 학습에서 발생하는 다루기 힘든 적분을 근사화하는 기술 계열입니다. 일반적으로 “관측 변수” (일반적으로 "데이터"라고 함)와 “알 수 없는 매개 변수” 및 “잠재 변수”로 구성된 복잡한 통계 모델에 사용되며 그래픽 모델에서 설명하는 것처럼 3 가지 유형의 랜덤 변수 간의 다양한 종류의 관계가 있습니다.

베이지안 추론에서 매개 변수와 잠재 변수는 "관찰되지 않은 변수"와 함께 그룹화됩니다. 변분 베이지안 방법은 주로 두 가지 목적으로 사용됩니다.

1.관찰되지 않은 변수의 사후 확률에 대한 분석적인 근사를 제공하여 이러한 변수에 대한 통계적 추론을 수행합니다.

2.관측 된 데이터.( 즉 관찰되지 않은 변수에 대해 수행 된 marginalization으로 주어진 데이터의 marginal 확률)의 marginal likelihood(때로는 "증거"라고 부름)에 대한 하한을 유도합니다.

이것은 전형적으로 모델을 선택하는데 사용된다

일반적으로 주어진 모델에 대한 더 높은 marginal likelihood은 그 모델에 의한 데이터의 적합(fit)이 더 높다는 것을 의미하고, 문제에서 해당 모델이 데이터를 생성 한 모델 일 확률이 높다는 것이 일반적이다. (베이즈 (Bayes) 기사 참조).

이전의 목적 (사후 확률을 근사화하는 것)에서, 변분 베이즈는 몬테카를로 샘플링 방법에 대한 대안이다. 특히 깁스 샘플링과 같은 마르코프 체인 몬테카를로 방법은 복잡한 분포에 대한 통계적 추론에 베이지안 방식을 완전히 사용하기위한 것이다.

복잡한 분포는 직접 평가하거나 샘플링하기가 어렵습니다. 특히, Monte Carlo 기법은 샘플 세트를 사용하여 정확한 사후(posterior)에 수치 근사를 제공하는 반면, 변분 베이즈는 사후의 근사에 대해 국부적으로 최적의 정확한 분석 솔루션을 제공합니다.

변분 베이즈는 각 매개 변수의 가장 가능성이 높은 값의 최대 사후 추정 (MAP 추정)로부터 매개변수와 잠재 변수의 전체 사후 분포를 계산(근사화)하는 완전 베이지 추정까지의 EM (expectation-maximization) 알고리즘의 확장으로 볼 수 있습니다

EM 에서 처럼 최적의 매개 변수 값 집합을 찾고 분석적으로 해결할 수 없는 상호 연동(interlock) 된 (상호 종속적인) 방정식 집합을 기반으로 EM과 동일한 대체 구조를 갖습니다.

많은 어플리케이션에서 변분 베이즈는 Gibbs 샘플링에 비해 더 빠른 속도의 솔루션을 제공합니다. 그러나 매개 변수를 반복적으로 업데이트하는 데 사용 된 방정식 집합을 파생하려면 종종 유사한 Gibbs 샘플링 방정식을 유도하는 것과 비교하여 많은 양의 작업이 필요합니다. 이는 오직 두 개의 매개 변수와 잠정적 변수가 없는 기본적인 비 계층 적 모델의 경우 아래에 설명 된 것처럼 개념적으로 매우 단순한 많은 모델에 대해서도 마찬가지입니다.

변분 추론에서 관측되지 않은 변수의 집합에 대한 사후분포 와 몇몇 데이터 X 는 변분 분포 Q(Z)으로 근사화한ㄴ다.

분포 Q(Z) 는, 실제 사후 분포 P(Z|X) 와 유사한 Q(Z)을 만들기 위한 의도로 선택된 P(Z|X) 보다 단순한 형태의 분포의 계열에 속하도록 제한된다.

유사성 의 부족은 비유사 함수 d(Q;P) 의 조건으로 측정되고 그러므로 d(Q;P)을 최소화하는 분포 Q(Z)을 선택하여 추론이 된다.

변분 베이즈의 가장 흔한 형태는 비유사 함수의 선택으로 Q로부터 P의 KL발산을 사용한다.

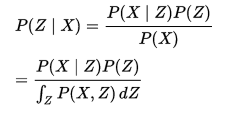
이 선택은 최소화를 다루기 쉽도록 한다.

KL 발산은 아래와 같이 정의된다



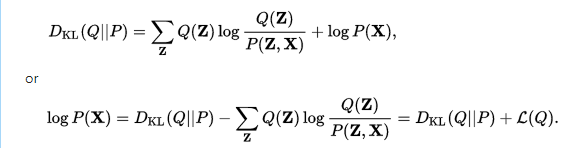
Q와 P는 예상했던 것과는 반대이다. 역 KL- 발산의 사용은 개념적으로 기대 – 최대화(EM) 알고리즘과 유사하다. (다른 방식으로 KL- 발산을 사용하면 기대 전파(Expectation Propagation) 알고리즘이 생성된다

변분 기술은 일반적으로 아래와 같은 근사를 형성하는 데 사용됩니다.

{\displaystyle P(\mathbf {Z} \mid \mathbf {X} )}. 

분모의P (X)를 계산하기 위하여 Z에 대하여 marginalization는 일반적으로 다루기가 어렵다. 예를 들어, Z의 검색 공간이 조합 적으로 크기 때문에. 우리는 Q (Z) ~ P (Z | X) 을 사용하여 근사값을 구합니다.

KL- 발산은 아래와 같이 쓸 수 있습니다.



Q에 대한 로그 증거 ( log P (X))가 고정됨에 따라 마지막 성분  Q을 최대화하여 Q로부터 P의 KL 발산을 최소화한다.

Q의 적절한 선택에 의해, 는 계산되고 최대화 되기 싶다.

따라서 사후분포 P(Z|X) 에 대하여 두개의 분석적인 근사 Q 와 증거 log P(X)에 대하여 하한 경계 를 알고 있다

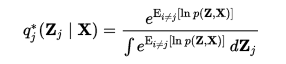
하한 경계  는 (negative) 변분 에너지로 알려져 있다 왜냐하면 Q의 엔트로피 와 에너지 의 합으로 표현되기 떄문이다.

In practice

변분 분포는 일반적으로 잠재 변수의 일부 파티션으로 분리된다고 가정한다. 즉, 잠재 변수 Z 를 ,…,, 으로 분해된다



변분 미적분을 사용하면 "최상의"분포 와 각각의 요소 에 대하여 아래와 같이 표현된다.(위에서 설명한 바와 같이 KL 발산을 최소화하는 분포의 면에서)



여기서  는 파티션에 없는 모든 변수에 대해 선택한 데이터와 잠재 변수의 joint 확률의 로그에 대한 기대치입니다.

실 생활에서 .log를 취하면 아래와 같다



위 식의 상수는 정규화 상수 (에 대해서 위의 표현 식의 분모)와 관련이 있으며 일반적으로 검사에 의해 복원됩니다. 표현의 나머지 부분은 일반적으로 알려진 형태의 분포 (예 : 가우스, 감마 등)로 인식 될 수 있습니다.

기대치의 성질을 사용하여 표현식 는 잠재 변수에 대하여 사전 분포의 고정된 하이퍼 파라미터와 현재 파티션에 없는 잠재 변수.(예, 에 포함되지 않은 잠재 변수)들의 기대치에 대한 함수로 단순화될 수 있다 . 이것은 하나의 파티션에 있는 변수에 대한 분포의 매개 변수와 다른 파티션의 변수에 대한 기대치 사이에 순환 종속성(circular dependencies)을 만듭니다. 이것은 자연스럽게 잠재 변수의 기대치 (아마도 더 높은 모멘트)가 어떤 방식으로 (아마도 무작위로) 초기화 된 EM (기대 최대화 알고리즘)과 같은 반복 알고리즘을 제안하고 각 분포의 매개 변수는 다음과 같습니다. 기대치의 현재 값을 사용하여 차례로 계산되고, 그 후 새롭게 계산된 분포의 기대치는 계산 된 매개편수에 따라 적절히 설정됩니다. 이 종류의 알고리즘은 수렴된다는 것이 보증된다.

즉, 변수의 각 파티션에 대해 파티션의 변수에 대한 분포 식을 단순화하고 문제의 변수에 대한 분포의 함수 종속성을 조사함으로써 일반적으로 분포의 계열을 결정할 수 있습니다.

분포의 매개 변수에 대한 공식은 사전 분포의 하이퍼 파라미터 (상수로 알려져 있음)로 표현 될 뿐만 아니라 다른 파티션의 변수 함수에 대한 기대치로 표현됩니다. 일반적으로 이러한 기대치는 변수 자체 에 대한 기대치(즉, 평균)의 함수로 단순화 될 수 있다. 때로는 제곱 변수 (변수의 분산과 관련 될 수 있음)에 대한 기대치 또는 더 높은 힘 (즉, 더 높은 모멘트)에 대한 기대치가 나타납니다

대부분의 경우, 다른 변수의 분포는 알려진 계열일 것이고 관련 기대치의 공식을 찾을 수 있습니다. 그러나 이러한 수식은 다른 변수에 대한 기대치에 따라 달라지는 분포의 매개 변수에 따라 달라집니다. 결과는 각 변수 분포의 매개 변수에 대한 수식이 변수간에 상호 비선형 종속성을 갖는 일련의 방정식으로 표현 될 수 있다는 것입니다. 보통 방정식 시스템을 직접 풀 수는 없습니다. 그러나 위에서 설명한대로 종속성은 단순한 반복 알고리즘을 제안합니다. 대부분의 경우 수렴이 보장됩니다. 예를 들면 이 과정이 명확해 집니다.

## A basic example

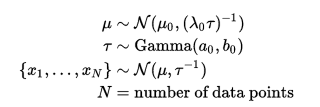
### . 평균 및 분산이 알려지지 않은 가우시안 분포로부터의 i.i.d([**Independent and identically distributed random variables**](https://en.wikipedia.org/wiki/Independent_and_identically_distributed_random_variables)**)의 관측(observation)**의 집합으로 구성된 간단한 비 계층 적 베이지안 모델을 고려해 보자.

### 다음은 이 모델을 통해 변분 베이즈 방법의 작동을 자세히 설명합니다. 수학적 편의를 위해 다음 예제에서 우리는 정밀도 즉 분산의 역수 (또는 다변량 가우스, 공분산 행렬의 역함수)라는 관점에서 작업합니다. (이론적인 관점에서 볼 때 정확도와 분산은 둘 사이에 일대일 대응이 있기 때문에 동일합니다.)

수학적 모델

공분산(conjugate) 사전 분포를 미지의 평균 및 분산에 배치한다. 즉, 평균 또한 가우스 분포를 따르고, 정밀도는 감마 분포를 따른다.

다른 말로 아래와 같다.



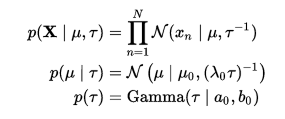
N 개의 데이터 포인트 가 있고 매개변수  와  의 사후 분포  를 추론하는 것이 목표이다

### The joint probability

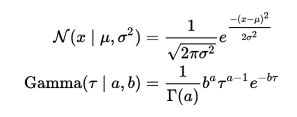
모든 변수의 결합 확률은 아래와 같이 쓸 수 있다.



여기서 각 요소는 아래와 같다.



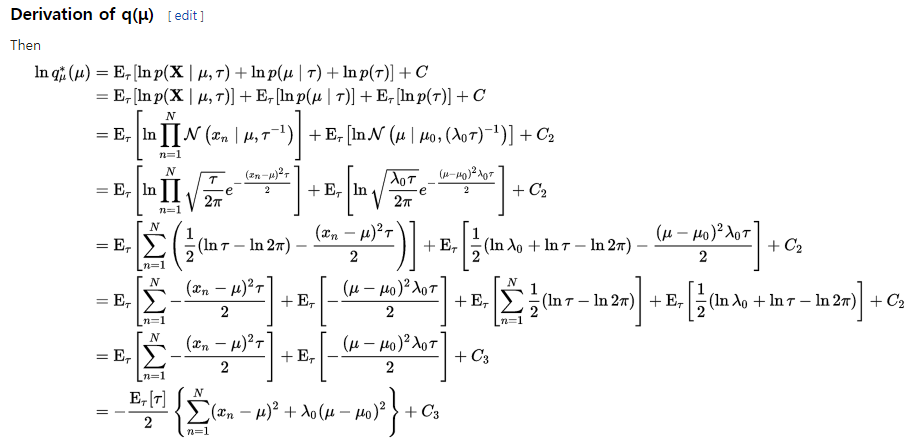
여기서



### Factorized approximation

 사후 분포는 독립 성분 와 로 분해된다고 가정한다

이러한 유형의 가정은 변분 베이지안 방법의 기초가 된다. 실제 사후 분포는 실제로 이 방법으로 분해되지 않는다 (실제로 이 간단한 경우에는 가우스 감마 분포라고 알려져 있음). 따라서 우리가 얻은 결과는 근사값이 됩니다.



위의 유도에서,μ 에 대해서 , 와 는 상수값으로 언급한다.

항목은 μ의 함수가 아니고 μ의 값에 관계없이 같은 값을 가질것이다.

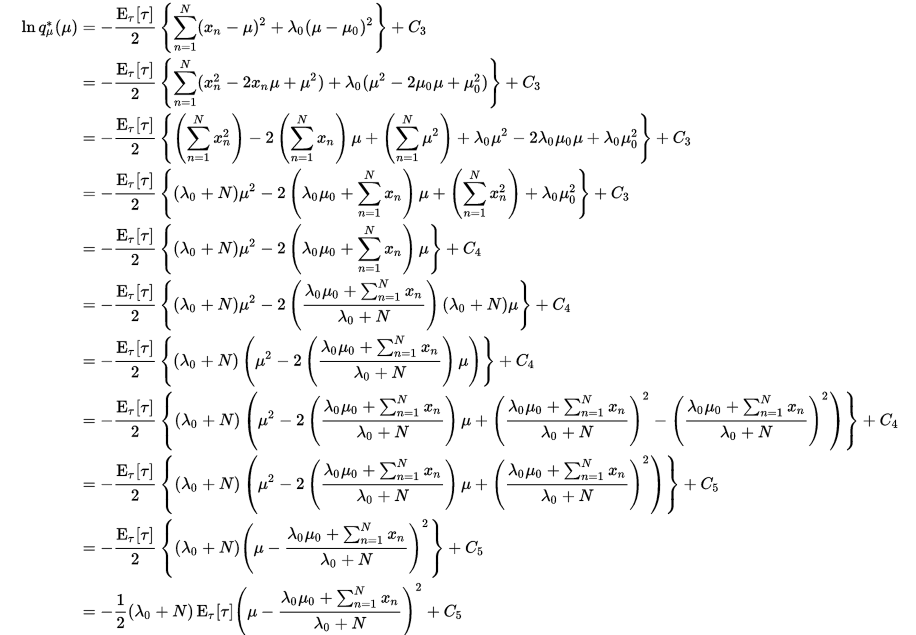
그래서 3번째 라인은 끝에서 상수성분으로 흡수된다.

7번째 라인도 같이 된다

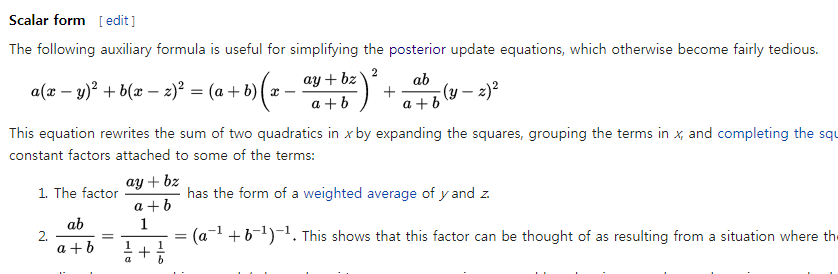
마지막 라인은 μ에서 2차 다항식이다.

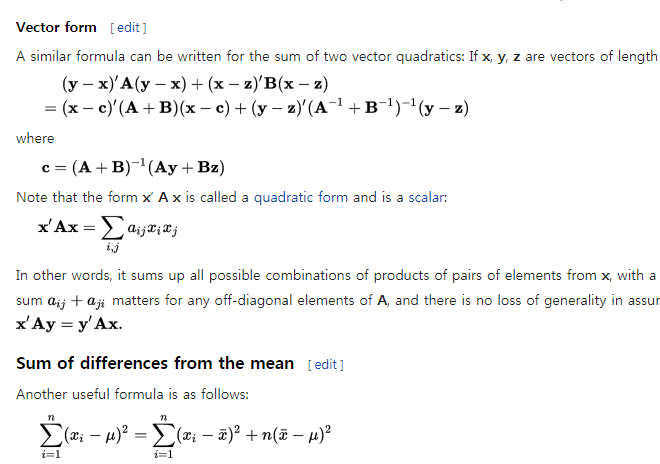
의 로그에 대한 내용이 위의 마지막 식이므로  자체는 가우시안 분포임을 알 수 있다.

{} 내의 제곱을 풀어내고 와 과 관련된 부분을 분리 및그룹화하고 에 대한 완전 제곱식으로 변환)하면 가우시안 분포의 매개 변수를 도출 할 수 있다 :

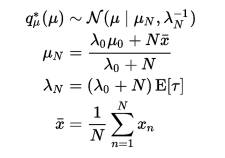


### 위의 모든 단계는 두 개의 2 차 합계(Sum of two quadratics)에 대한 수식을 사용하여 단축 할 수 있습니다.



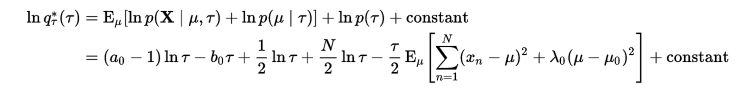


다른 말로

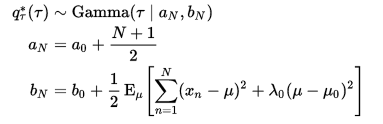




의 유도의 위와 유사하다 , 간결함을 위해 일부 세부 사항은 생략하지만

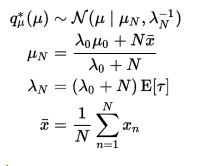


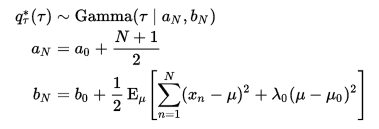
양변에 exponential 을 취하면 는 감마분포이다.



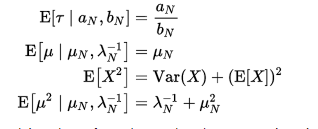
### Algorithm for computing the parameters**[[edit](https://en.wikipedia.org/w/index.php?title=Variational_Bayesian_methods&action=edit&section=9" \o "Edit section: Algorithm for computing the parameters)]**

이전 절에서의 결론을 요약하면

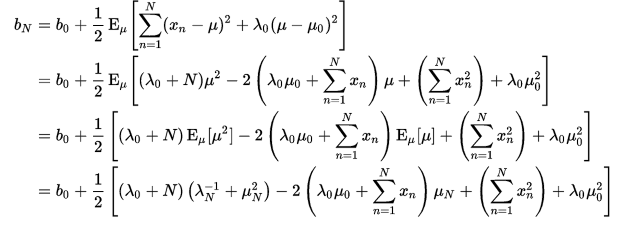




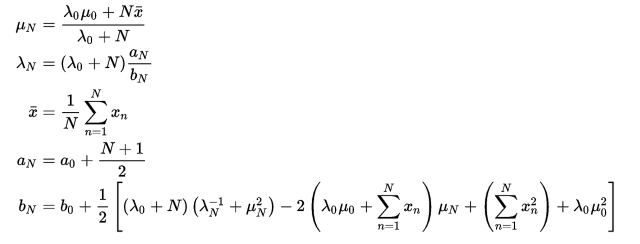
각각의 경우에 변수 중 하나에 대한 분포에 대한 매개 변수는 다른 변수에 대한 기대치에 따라 다릅니다. 가우스 및 감마 분포의 모멘트의 기대값에 대한 표준 공식을 사용하여 기대치를 확장 할 수 있습니다.



대부분의 경우 위의 방정식에 이 수식을 적용하는 것은 간단하지만 의 방정식이 더 많은 작업을 필요로 합니다



그런 다음 매개 변수 방정식을 기대치를 제거하고 다음과 같이 쓸 수 있습니다.



및 에 대한 수식에는 순환 종속성(circular dependency)이 있습니다.

이것은 자연스럽게 EM과 같은 알고리즘을 제안 합니다 :

1. 과 를 계산하고 과 를 계산하는데 이 값을 사용함
2. 임의의 값으로 을 초기화 함



1. 다른 매개변수의 알려진 값을 따라 의 현재 값을 사용하여 을 계산함
2. 다른 매개변수의 알려진 값을 따라 의 현재값을 사용하여 을 계산함
3. 수렴할 때까지 마지막 두 단계를 반복함(어떤 값도 작은 양 이상으로 변화하지 않을 때까지)

사후 매개 변수의 근사 분포의 하이퍼파라미터에 대한 값을 갖습니다. 원하는 특성을 계산하기 위해 사후값을 사용할 수 있습니다. 그것의 평균과 분산, 95 %의 가장 높은 밀도의 영역 (전체 확률의 95 %를 포함하는 가장 작은 간격) 등

이 알고리즘은 로컬 최대 값으로 수렴된다는 것을 알 수 있습니다.

사후 분포는 대응하는 사전 분포와 동일한 형태를 갖는다는 점에도 유의해야 한다. 우리는 이것을 가정하지 않았습니다. 우리가 만든 유일한 가정은 분포가 분해되고, 분포의 형태가 자연스럽게 따른다는(follow) 것입니다. 사후 분포가 사전 분포와 동일한 형태를 갖는다는 사실은 우연이 아니라 사전 분포가 지수족(대부분의 표준 분포의 경우에)에 속하면 일반적인 결과다 .

## Further discussion

### Step-by-step recipe

위의 예는 주어진 베이지안 네트워크에서 사후 확률 밀도에 대한 변분 - 베이지안 근사가 유도되는 방법을 보여줍니다.

1. 그래피컬 모델로 네트워크를 설명함

관찰 된 변수 (데이터) X 와 관측되지 않은 변수 (매개 변수 와 잠재변수 Z)와 그것들의 조건부 확률  분포를 구별하여, 변분 베이즈는 사후 확률 의 근사를 구성한다.

근사는 분할된 분포, 즉 관찰되지 않은 변수의 분리 된(

Disjoint) 부분 집합에 대한 둘 이상의 독립 분포의 곱(내적)라는 기본 특성을 갖는다

1. 관찰되지 않은 변수를 두 개 이상의 하위 집합으로 분할합니다. 이 집합에 대하여 독립적인 요소가 유도됩니다. 이것을 하기 위한 보편적인 절차는 없습니다. 너무 많은 부분 집합을 생성하면 안 좋은 근사값을 산출하는 반면, 너무 적게 생성하면 전체 변분 베이즈 절차를 다루기가 어렵게 만듭니다. 일반적으로 첫 번째 분할은 매개 변수와 잠재 변수를 분리하는 것입니다.

종종 이것은 조잡한 결과를 내기에 충분하다.

파티션은 ,…이라고 가정하자

1. 주어진 파티션 에 대하여 기본 등식 를 이용하여 가장 최적의 근사분포 를 쓴다
2. 그래픽 모델을 사용하여 결합 확률 분포에 대한 공식을 채웁니다.

의 어떤 변수도 포함하지 않는 어떤 성분의 조건 분포는 무시될 수 있다. 그들은 상수항으로 취급된다.

1. 위의 예에 따라 수식을 단순화하고 기대 연산자(operator)를 적용하십시오. 이상적으로, 이것은 에 포함되지 않는 변수의 기본 함수의 기대값으로 단순화 해야한다 (예를 들어 제 1, 제2 기본 모멘트 , 로그의 기대값등). 변분 베이즈 절차가 잘 작동하려면 이러한 기대치가 일반적으로 매개 변수의 함수 및 / 또는 이러한 변수의 분포의 하이퍼 파라미터로 해석 적으로 표현 가능해야 합니다. 모든 경우에 있어 이러한 기대치는 현재 파티션의 변수에 대한 상수입니다.
2. 현재 파티션의 변수와 관련된 공식의 함수 형식은 분포 유형을 나타냅니다. 특히, 수식을 지수화하면 분포의 확률 밀도 함수 (PDF)가 생성됩니다 (또는 최소한 알 수 없는 정규화 상수를 가지고 그것에 비례하는 값). 전체적인 방법이 다루기 쉽도록 하기 위해서, 함수 형태를 알려진 분포에 속하는 것으로 인식 할 수 있어야 한다. 수식을 알려진 분포의 PDF와 일치하는 형식으로 변환하려면 상당한 수학적 조작이 필요할 수 있습니다. 이것이 가능할 때, 정규화 상수는 정의에 의해 복원 될 수 있고, 알려진 분포의 매개 변수에 대한 방정식은 공식의 적절한 부분을 추출함으로써 도출 될 수 있다.
3. 모든 기대치가 현재 파티션에 없는 변수의 함수로 분석적으로 대체되고 PDF가 알려진 분포로 식별 할 수 있는 형식으로 바뀌면 결과는 최적의 매개 변수 값을 다른 파티션의 변수의 매개 변수의 함수로 표현하는 방정식 집합입니다.
4. 이 절차가 모든 파티션에 적용될 수 있으면 결과는 모든 매개 변수의 최적 값을 지정하는 상호 연결된 수식 집합입니다.
5. 기대 최대화 (EM) 형태 절차가 적용되어 각 매개 변수에 대한 초기 값을 선택하고 일련의 단계를 반복합니다. 각 단계에서 방정식을 차례로 순환하여 각 매개 변수를 차례로 업데이트합니다. 이것은 수렴하는 것이 보장됩니다.

### Most important points

관련된 모든 수학적 조작으로 인해 큰 그림을 추적하기가 쉽습니다. 중요한 사항은 다음과 같습니다.

1. 변분 베이즈 (Variational Bayes)의 개념은 데이터가 주어지면서 관측되지 않은 변수 (매개 변수 및 잠재 변수)의 사후 확률에 대한 분석적인 근사를 작성하는 것입니다. 이것은 솔루션의 형태가 깁스 샘플링과 같은 다른 베이지안 추론 방법과 유사하다는 것을 의미합니다. 즉, 변수에 대해 알려진 모든 것을 설명하고자 하는 분포입니다. 다른 베이지안 방법과 마찬가지로 - 그러나 예를 들어. 기대 최대화 (EM) 또는 다른 최대 우도 방법에서 - 관측되지 않은 변수 (즉, 파라미터 및 잠재 변수)의 두 유형은 동일하게, 즉 랜덤 변수로서 취급된다. 그런 다음 변수에 대한 추정치가 표준 베이지안 방식 (예 : 단일 지점 추정치를 얻거나 신뢰할 수 있는 간격, 최고 밀도 지역 등을 얻기 위해 분포의 평균을 계산합니다.

2. "분석적 근사"는 사후 분포에 대해 수식을 작성할 수 있음을 의미합니다. 이 공식은 일반적으로 잘 알려진 확률 분포의 곱으로 구성되며, 각 변수는 관측되지 않은 변수의 집합으로 분해합니다 (즉, 관측 된 데이터가 주어지면 다른 변수와 조건부 독립적입니다). 이 공식은 실제 사후 분포가 아니라 근사입니다. 특히 관측되지 않은 변수의 가장 낮은 모멘트(평균과 분산.)에 상당히 가깝게 일치 할 것이다.

3 모든 수학적 조작의 결과는 (1) 확률 분포가 식별 가능하면 분해 가능 (2) 이러한 분포의 매개 변수에 대한 상호 의존적인 공식입니다. 이러한 매개 변수의 실제 값은 EM과 매우 유사한 반복적 인 절차를 통해 수치로 계산됩니다.

### Compared with expectation maximization (EM)

변분 베이즈 (VB)는 종종 기대 최대화 (EM)와 비교됩니다. 실제 수치 절차는 두 매개 변수가 모두 최적의 매개 변수 값에 연속적으로 수렴되는 반복적 인 절차라는 점에서 매우 유사합니다. 각각의 절차를 유도하는 초기 단계는 둘 다 확률 밀도에 대한 공식으로 시작하고 두 가지 모두 상당한 양의 수학적 조작을 필요로 하는 면에서 모호하게 유사하다

그러나 여러 가지 차이점이 있습니다. 가장 중요한 것은 **무엇**이 계산되는지 이다.

EM은 "매개 변수"로 분류 될 수 있는 랜덥 변수의 사후 분포의 점 추정치를 계산하지만 잠재 변수의 실제 사후 분포만을 추정합니다 (적어도 "소프트 EM"에서는 적어도 잠재 변수가 불연속 인 경우에만 종종 계산 됨) ). 계산 된 점 추정은 이러한 매개 변수의 모드입니다. 다른 정보에 대해서는 계산할 수 없음

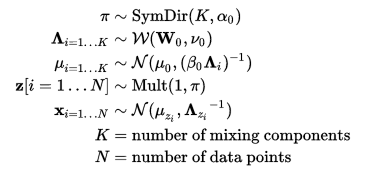
반면에 VB는 모든 변수, 즉 매개 변수와 잠재 변수의 실제 사후 분포의 추정치를 계산합니다. 점 추정치가 도출 될 필요가 있을 때, 일반적으로 베이지안 추론에서 정상적인 것처럼 모드가 아니라 평균이 사용됩니다. 이와 동시에 VB에서 계산 된 매개 변수는 EM의 매개 변수와 동일한 중요성을 갖지 않습니다. EM은 베이즈 네트워크 자체의 매개 변수의 최적 값을 계산합니다.

VB는 베이즈 네트워크의 매개 변수 및 잠재 변수를 근사화하는 데 사용되는 분포 매개 변수의 최적 값을 계산합니다. 예를 들어, 전형적인 가우스 혼합 모델은 각 혼합 성분의 평균 및 분산에 대한 매개 변수를 갖습니다. EM은 이러한 매개 변수에 대한 최적 값을 직접 추정합니다. 그러나 VB는 먼저 이러한 매개 변수에 분포를 맞춥니다. 일반적으로 사전 분포의 형태로 나타납니다. 정규 스케일 역 감마 분포 -를 계산하고, 이 이전의 분포의 파라미터, 즉 본질적으로 하이퍼 파라미터에 대한 값을 계산할 것이다. 이 경우 VB는 구성 요소의 평균 및 분산의 결합 분포를 설명하는 정상 스케일 역 감마 분포의 네 가지 매개 변수에 대한 최적 추정을 계산합니다.

## A more complex example

이 등식은 옳지 않고, 교정될 필요가 있다

베이지안 가우시안 혼합 모델이 아래와 같이 설명된다고 가정하자



주의

SymDir() 는 K 차원의 대칭 Dirichlet 분포이다.

각 성분이 로 설정된 하이퍼 파라미터를 가진

Dirichlet  분포는 categonical 분포 또는 멀티노미얼 분포의 conjugate 사전(prior)이다.

 는 [Wishart](https://en.wikipedia.org/wiki/Wishart_distribution" \o "Wishart distribution) 분포이다, [Wishart](https://en.wikipedia.org/wiki/Wishart_distribution) 분포 는 multivariate가우시안 분포에 대하여 [precision matrix](https://en.wikipedia.org/wiki/Precision_matrix) (역 covariance matrix)의 conjugate prior이다.  .

* Mult() 는 하나의 관측에 대하여  [multinomial 분포이다.](https://en.wikipedia.org/wiki/Multinomial_distribution" \o "Multinomial distribution) ( [categorical 분포와](https://en.wikipedia.org/wiki/Categorical_distribution" \o "Categorical distribution) 등가). 상태공간은 "one-of-K" 표현, 즉. K 차{\displaystyle K} ㄴㅓㅜㅇㄴㅓ우원 벡터, K 차원 벡터에서 요소들 중의 하나만 1(관측의 식별을 규정하기 위한다) 모든 다른 요소는 0이다.

one-hot encoding

* {\displaystyle {\mathcal {N}}()}은 가우시안 분포이다. 이 경우에서 특별히 [multivariate 가우시안 분포](https://en.wikipedia.org/wiki/Multivariate_Gaussian_distribution).

위의 변수에 대한 설명은 아래와 같다.

는 N개의 데이터 점들의 집합

그것의 각각은 multivariate 가우시안 분포에 따라 분포된 K차원의 벡터

는 잠재 변수(해당 데이터 점이 속하는 혼합(Mixture) 성분을 규정하는 데이터 점당 하나)의 집합

위에서 설명한 바와 같이 에 대하여 성분들 로 "one-of-K" 의 벡터 표현을 사용함

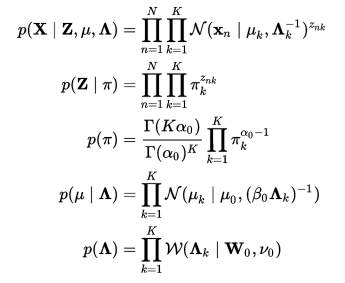
는 K 개의 혼합 성분에 대하여 혼합 비율

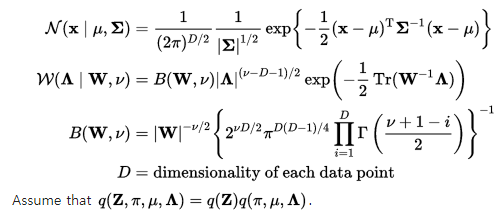
와 는 각 혼합 성분과 연관된 매개변수(평균 과 정확도) 를 규정함

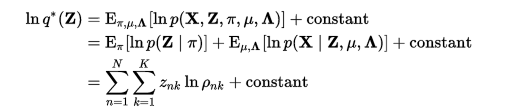
모든 변수의 결합 확률은 아래와 같다.



여기서 각 요소는





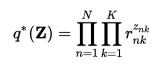




에 대한 공식의 양변에 지수를 취하면 아래와 같다.



이것이 졍규화되면  은 K의 모든값에 대하여 합이 1이다.



여기서



즉 는 하나의 관측 multinomial 분포와 각 에 대한 요소들의 곱이다.

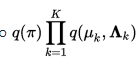
여기서 에 대하여 매개변수 를 가진 하나의 관측 multinomial 분포로서 분포된다.

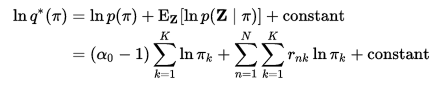
게다가



이것은 categorical 분포에 대한 표준 결과이다.

가우시안 혼합 모델을 정의하는 그래픽 모델의 구조 덕분에 성분 를 고려하면

그 성분들은 로 분해된다.



양변에 exponential을 취하면 를  Dirichlet분포로 인식된다.



여기서



여기서



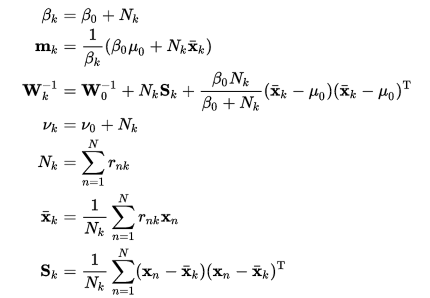
마지막으로



와 에 관련된 성분들을 그룹화하고 읽어내면 결과는 아래와 같은 [Gaussian-Wishart](https://en.wikipedia.org/wiki/Gaussian-Wishart_distribution) 분포이다.

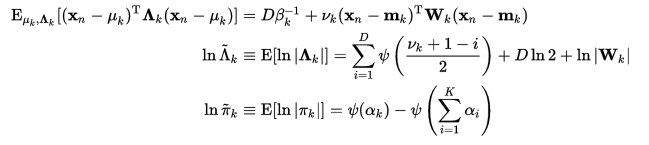


 주어진 정의에 의해



마지막으로 이런 함수는 ,와 에 기반하여 교대로 정의되는 를 사용하는 의 값을 요구한다.

이런 기대치가 취해지는 분포를 결정하고 그것들에 대한 공식을 유도할 수 있다.



이것으로 아래의 결과가 나온다.



 에 대하여 정규화함으로써 절대갑에 비례하도록 변환되고 해당값의 합은 1이다.

1,매개변수 와 변수 와 의 의 갱신 방정식은 통계  에 달려있고 이런 통계는 교대로 에 달려있다

2.변수 의 매개변수 에 대한 갱신 방정식은 교대로 에 의존하는 통계 에 의존한다

3. 에 대한 갱신 방정식은 와 와 직접 원형 의존성을 가진다.

와 를 통해서 와에 대해서는 간접적 원형 의존성을 가진다.

이것은 두 단계를 번갈아 반복하는 절차를 제안합니다.

1. 모든 다른 매개변수의 현재 값을 사용하여 의 값을 계산하는 E 단계
2. 모든 다른 매개변수의 새로운 값을 계산하기 위하여 의 새로운 값을 사용하는 M 단계

이들 단계는 표준 EM 알고리즘과 가깝게 일치하여 가우시안 혼합 모델의 파라미터에 대한 최대 가능도 (maximum likelihood) 또는 최대 사후 확률 (MAP)을 유도한다

E 단계에서의 책임 은 데이터가 주어진 상황에서 잠재 변수 의 사후 확률에 밀접하게 관련되어있다.

통계 와  의 계산은 데이터에 대하여 해당하는 “soft-count” 통계의 연산과 밀접하게 대응합니다

매개변수의 새로운 값을 계산하기 위해서 그런 통계치의 사용는 가우시안 혼합 모델에 대하여 정상적인 EM에서 새로운 값들을 계산하기 위하여 soft count를 이용하는 것과 밀접하게 대응한다.

## Exponential-family distributions

앞의 예에서 관찰되지 않은 변수에 대한 분포가 “매개변수”에 대한 분포와 “잠재 데이터”에 대한 분포로 분할된다고 가정하면, 각 변수에 대하여 유도된 가장 최적의 분포는 변수에 대한 해당 사전 분포로서 같은 족에 있다 이것은 지수족으로부터 파생 된 모든 사전 분포에 적용되는 일반적인 결과입니다.