Sequence correction for Gaussian probability

Patrick Ding

2021-05-01

Expectation propagation for Gaussian probability estimation works well for rectangular integration regions, but poorly for polyhedral regions. We provide a way to rewrite the polyhedral region Gaussian probability problem as a rectangular problem.

Cast polyhedral problem as rectangular one

We write the multivariate Gaussian density with covariance Σ truncated to a polytope region $l \leq Ax \leq u$, where $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, and $l, u \in \mathbb{R}^m$, as

$$\begin{aligned} p(x) &= \mathcal{N}_n(x; 0, \Sigma) 1\{l \leq Cx \leq u\} \\ &= \mathcal{N}_n(x; 0, \Sigma) \prod_{i=1}^{m} t_i(x), \end{aligned}$$

where $t_i(x) = 1\{l_i \leq c_i^T x \leq u_i\}$. For random variable X with this density, the normalizing constant of this density is $P(l \leq AX \leq u)$. Let's assume $\Sigma = I$, and call the random variable Z in this case, so we want $P(l \leq AZ \leq u)$.

Let $A_{\cdot,k}$ be the kth column of A. Let $\epsilon_i \sim_{iid} \mathcal{N}(0, 1)$ for $i \in \{1, \dots, n\}$. We construct a sequence of random vectors $\{X_1, X_2, \dots, X_n\}$

$$\begin{aligned} X_1 &= A_{\cdot,1}\epsilon_1 \\ X_2 &= X_1 + A_{\cdot,2}\epsilon_2 \\ &\vdots \\ X_{n-1} &= X_{n-2} + A_{\cdot,n-1}\epsilon_{n-1} \\ X_n &= X_{n-1} + A_{\cdot,n}\epsilon_n. \end{aligned}$$

Substituting X_k for $k \in \{1, \dots, n-1\}$ into the last expression, observe that

$$\begin{aligned} X_n &= A_{\cdot,1}\epsilon_1 + \cdots + A_{\cdot,n}\epsilon_n \\ &= A \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}. \end{aligned}$$

From properties of multivariate normal X_n has the same distribution as AZ, $\mathcal{N}_m(0, AA^T)$, so

$$P(l \leq X_n \leq u) = P(l \leq AZ \leq u).$$

The joint density of the X_i factorizes as

$$p(x_1, \dots, x_n) = p(x_n | x_{n-1}) p(x_{n-1} | x_{n-2}) \cdots p(x_2 | x_1) p(x_1)$$

where

$$p(x_i|x_{i-1}) = \mathcal{N}_m(x_i; x_{i-1}, A_{\cdot,i}A_{\cdot,i}^T) \qquad i \in \{2,\ldots,n\}$$
$$p(x_1) = \mathcal{N}_m(x_1; 0, A_{\cdot,1}A_{\cdot,1}^T).$$

The covariance $\Xi_i = A_{\cdot,i}A_{\cdot,i}^T$ is not invertible in general, it is rank 1. But from Khatri (1968) we can express this singular normal distribution in terms of the Moore-Penrose pseudoinverse of Ξ_i. Let γ_i be the sole eigenvector of Ξ_i, and let λ_i be the sole eigenvalue of Ξ_i. Then

$$\gamma_i = \frac{A_{\cdot,i}}{||A_{\cdot,i}||} \tag{1}$$
$$\lambda_i = ||A_{\cdot,i}||^2 \tag{2}$$
$$\Xi^+ = \frac{1}{\lambda_i}\gamma_i\gamma_i^T, \tag{3}$$

and the density is

$$p(x_i|x_{i-1}) = (2\pi)^{-n/2}|\lambda_i|^{-1/2}\exp\left[-\frac{\lambda_i}{2}(x_i - x_{i-1})^T\gamma_i\gamma_i^T(x_i - x_{i-1})\right]$$
$$p(x_1) = (2\pi)^{-n/2}|\lambda_1|^{-1/2}\exp\left[-\frac{\lambda_1}{2}x_1^T\gamma_1\gamma_1^T x_1\right].$$

Finally we add the indicator for the constraints on X_n and get an augmented truncated multivariate normal density proportional to:

$$\begin{aligned} p(x) &= \frac{1}{S}p(x_1,\ldots,x_n)1\{\ell \le x_n \le u\} \\ &= \frac{1}{S}p(x_1)\left[\prod_{i=2}^{n} p(x_i|x_{i-1})\right]1\{\ell \le x_n \le u\}, \end{aligned} \tag{4}$$

where $x = (x_1,\ldots,x_n)^T \in \mathbb{R}^{mn}$ and S is a normalizing constant. The ratio of this density's normalizing constant and the normalizing constant of $p(x_1,\ldots,x_n)$ will give us $P(\ell \le X_n \le u)$.

EP form

This density is intractable because of the term $1\{\ell \le x_n \le b\}$, but we can approximate the density with a Gaussian using the EP method of Cunningham et al (2011). This will be efficient and accurate since the truncation region is axis aligned, and their method performs well in that case. Equation 4 is exactly of the form that the Cunningham et al method handles:

$$p(x) = p_0(x)\prod_{k=1}^{n} t_k(x) \tag{5}$$

where

$$p_0(x) = p(x_1)\left[\prod_{i=2}^{n} p(x_i|x_{i-1})\right] \qquad \text{prior} \tag{6}$$
$$t_k(x) = 1\{\ell_k \le c_k^T x \le b_k\} = 1\{\ell_k \le x_{n_k} \le b_k\} \qquad \text{likelihood factor} \tag{7}$$
$$c_k = \begin{pmatrix} 0_{1\times m(n-1)} & 0_{1\times k-1} & 1 & 0_{1\times m-k} \end{pmatrix}^T \tag{8}$$

The density of the prior is singular Gaussian,

$$
\begin{aligned}
p_0(x) &= \mathcal{N}_{nm}(x; 0, K) \\
&= (2\pi)^{-n/2} \left(\prod_{i=1}^{n} |\lambda_i|^{-1/2} \right) \exp(x^T K^{-1} x)
\end{aligned}
$$

$$
K^{-1} = \begin{pmatrix}
\Xi_1^+ + \Xi_2^+ & \Xi_2^+ & & & & \\
\Xi_2^+ & \Xi_2^+ + \Xi_3^+ & \Xi_3^+ & & & \\
& \Xi_3^+ & \Xi_3^+ + \Xi_4^+ & & & \\
& & & \ddots & & \\
& & & & \Xi_{n-1}^+ + \Xi_n^+ & \Xi_n^+ \\
& & & & \Xi_n^+ & \Xi_n^+
\end{pmatrix}
$$

where unspecified blocks in K^{-1} are 0.

We also have access to the covariance matrix of x, K. Note that

$$
\begin{aligned}
X_i &= \sum_{k=1}^{i} A_{\cdot,k}\epsilon_k && (9) \\
E[X_i] &= 0, && (10)
\end{aligned}
$$

and K_{ij} is

$$
\begin{aligned}
Cov(X_i, X_j) &= E\left[\left(\sum_{r=1}^{i} A_{\cdot,r}\epsilon_r \right) \left(\sum_{s=1}^{j} A_{\cdot,s}\epsilon_s \right)^T \right] - E[X_i]E[X_j] && (11) \\
&= \sum_{t=1}^{\min(i,j)} A_{\cdot,t} A_{\cdot,t}^T && (12) \\
&= \sum_{t=1}^{\min(i,j)} \Xi_t, && (13)
\end{aligned}
$$

where many terms are 0 because the ϵ_i are independent. The covariance matrix is

$$
K = \begin{pmatrix}
\Xi_1 & \Xi_1 & \cdots & \cdots & \Xi_1 \\
\Xi_1 & \Xi_1 + \Xi_2 & \Xi_1 + \Xi_2 & \cdots & \Xi_1 + \Xi_2 \\
\Xi_1 & \Xi_1 + \Xi_2 & \ddots & & \vdots \\
\vdots & \vdots & & \sum_{t=1}^{n-1} \Xi_t & \sum_{t=1}^{n-1} \Xi_t \\
\Xi_1 & \Xi_1 + \Xi_2 & \cdots & \sum_{t=1}^{n-1} \Xi_t & \sum_{t=1}^{n} \Xi_t
\end{pmatrix} \tag{14}
$$

EP factor updates

We approximate the $t_k(x)$ with unnormalized Gaussian densities with $\tilde{S}_k$, $\tilde{\mu}_k$, and $\tilde{\sigma}_k^2$ the approximating normalizing constant, mean, and variance for the kth factor:

$$
\tilde{t}_k(x) = \tilde{S}_k N(c_k^T x; \tilde{\mu}_k, \tilde{\sigma}_k)
$$

leading to the overall approximating unnormalized Gaussian with normalizing constant S, mean μ, and covariance Σ:

$$
q(x) = p_0(x) \prod_{k=1}^{n} \tilde{t}_k(x) = N(x; \mu, \Sigma)
$$

Initialize $\mu = m$, $\Sigma = K$. The EP cavity distribution is singular Gaussian

$$q^{\backslash k}(x) = Z^{\backslash k} N(x; u^{\backslash k}, V^{\backslash k}) \tag{15}$$

$$V^{\backslash k} = \left(\Sigma^{-1} - \frac{1}{\tilde{\sigma}_k} c_k c_k^T\right)^{-1} \tag{16}$$

$$u^{\backslash k} = V^{\backslash k} \left(\Sigma^{-1}\mu - \frac{\tilde{\mu}_k}{\tilde{\sigma}_k^2} c_k\right) \tag{17}$$

Since we are dealing with rank-one factors and the approximating factors are Gaussians and therefore closed under marginalization, we can moment match the tilted and approximating densities in the one dimension of interest, where the cavity distribution is univariate Gaussian:

$$q_{\backslash k}(c_k^T x) = q_{\backslash k}(x_{k'}) = S^{\backslash k} N(x_{k_k}; \mu_{\backslash k}, \sigma_{\backslash k}) \tag{18}$$

$$\sigma_{\backslash k} = \left(\frac{1}{\Sigma_{k'k'}} - \frac{1}{\tilde{\sigma}_k^2}\right)^{-1} \tag{19}$$

$$\mu_{\backslash k} = \sigma_{\backslash k} \left(\frac{\mu_{k'}}{\Sigma_{k',k'}} - \frac{\tilde{\mu}_k}{\tilde{\sigma}_k^2}\right) \tag{20}$$

$$k' = m(n-1) + k \tag{21}$$

The moments of the tilted distribution $t_k(x)q_{\backslash k}(x)$ are

$$\hat{S}_k = \frac{1}{2}\big(\Phi(\beta) - \Phi(\alpha)\big) \tag{22}$$

$$\hat{\mu}_k = \mu_{\backslash k} + \frac{1}{\hat{S}_k}\frac{\sigma_{\backslash k}}{\sqrt{2\pi}}(\exp(-\alpha^2) - \exp(-\beta^2)) \tag{23}$$

$$\hat{\sigma}_k^2 = \mu_{\backslash k}^2 + \sigma_{\backslash k} + \frac{1}{\hat{S}_k}\frac{\sigma_{\backslash k}}{\sqrt{2\pi}}\big(-(b_k + \mu_{\backslash k})\exp(-\beta^2)\big) - \hat{\mu}_k^2 \tag{24}$$

$$\alpha = \frac{\ell_k - \mu_{\backslash k}}{\sqrt{2\sigma_{\backslash k}^2}} \tag{25}$$

$$\beta = \frac{u_k - \mu_{\backslash k}}{\sqrt{2\sigma_{\backslash k}^2}} \tag{26}$$

We match the moments of $\tilde{t}_k(x)q_{\backslash k}(x)$ to the moments of $t_k(x)q_{\backslash k}(x)$, leading to factor updates:

$$\tilde{\sigma}_k^2 = (\hat{\sigma}_k{}^{-2} - \sigma_{\backslash k}^{-2})^{-1} \tag{27}$$

$$\tilde{\mu}_k = \tilde{\sigma}_k^2(\hat{\sigma}_k^{-2}\hat{\mu}_k - \sigma_{\backslash k}^{-2}\mu_{\backslash k}) \tag{28}$$

$$\tilde{S}_k = \hat{S}_k\sqrt{2\pi}\sqrt{\sigma_{\backslash k}^2 + \tilde{\sigma}_k^2}\exp\left(\frac{1}{2}\frac{(\mu_{\backslash k} - \tilde{\mu}_k)^2}{(\sigma_{\backslash k}^2 + \tilde{\sigma}_k^2)}\right) \tag{29}$$

Overall approximation update

We parameterize the approximate factors in the natural parameterization as in Cunningham et al.:

$$\tilde{\tau}_k = \frac{1}{\tilde{\sigma}_k^2}$$

$$\tilde{\nu}_k = \frac{\tilde{\mu}_k}{\tilde{\sigma}_k^2}$$

We update only the parts of the approximate parameters that are affected by the factor updates. Unfortunately this is all elements of the approximate parameters in general.

$$\Sigma^{new} = \Sigma^{old} - \left(\frac{\Delta\tilde{\tau}_k}{1 + \Delta\tilde{\tau}_k \Sigma^{old}_{k'k'}} \right) (\Sigma^{old}_{\cdot,k'})(\Sigma^{old}_{\cdot,k'})^T \tag{30}$$

$$\mu^{new} = \mu^{old} + \left(\frac{\Delta\tilde{\nu}_k - \Delta\tilde{\tau}_k \mu^{old}_{k'}}{1 + \Delta\tilde{\tau}_k \Sigma^{old}_{k'k'}} \right) \Sigma^{old}_{\cdot,k'} \tag{31}$$

We also need to update the approximate precision Σ^{-1}, which we need to compute the normalizing constant of the approximate Gaussian.

$$\Sigma^{-1} = K^{-1} + \sum_{k=1}^{m} \tilde{\tau}_k c_k c_k^T$$

Next we update the normalizing constant of the approximation:

$$\begin{aligned}
\log S = &-\frac{1}{2}\sum_{k=1}^{n} \log|\lambda_k| \\
&+ \sum_{k=1}^{n} \log\tilde{S}_k - \frac{1}{2}\left(\frac{\tilde{\mu}_k^2}{\tilde{\sigma}_k^2} + \log\tilde{\sigma}_k^2 + \log(2\pi) \right) \\
&+ \frac{1}{2}\left(\mu^T \Sigma^{-1} \mu + \log|\Sigma| \right) \\
= &-\frac{1}{2}\sum_{k=1}^{n} \log|\lambda_k| \\
&+ \sum_{k=1}^{n} \log\hat{S}_k + \frac{1}{2}\log(\tilde{\tau}_k \sigma^2_{\setminus k} + 1) + \frac{1}{2}\frac{\mu^2_{\setminus k}\tilde{\tau}_k - 2\mu_{\setminus k}\tilde{\nu}_k - \tilde{\nu}_k^2 \sigma^2_{\setminus k}}{1 + \tilde{\tau}_k \sigma^2_{\setminus k}} \\
&+ \frac{1}{2}\left(\mu^T \Sigma^{-1} \mu + \log|\Sigma| \right)
\end{aligned} \tag{32}$$

where we rewrite the middle term as Cunningham does. S is the desired probability $P(AZ \le b)$.

Algorithm

Algorithm 1 Sequence corrected EP for Gaussian probability

```
Set σ̃_k = 0
while not converged do
    for k = 1 : n do
        Form cavity distribution q^{\k}(x) = Z^{\k} N(x; u^{\k}, V^{\k})
        Compute moments of tilted distribution t_k(x) q_{\k}(x)
        Update parameters of t̃_k(x)
    end for
    Update overall approximation μ, Σ
end while
Calculate S
return S
```

Appendix

Normalizing constants

$$
\begin{aligned}
p_0(x)\prod_{i=1}^{n}\tilde{t}_i(x) &= \frac{1}{\sqrt{2\pi|K|}}\exp\left(-\frac{1}{2}(x-m)^T K^{-1}(x-m\right)\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\tilde{\sigma}_i^2}}\exp\left(-\frac{1}{2\tilde{\sigma}_i^2}(c_i^T x-\tilde{\mu_i})^2\right) \\
&= \exp\left(-\frac{1}{2}(n\log 2\pi+\log|K|+m^T K^{-1}m)\right)\exp(m^T K^{-1}x-\frac{1}{2}x^T K^{-1}x)\times \\
&\prod_{i=1}^{n}\left(\exp\left(-\frac{1}{2}(\log 2\pi-\log\tilde{\tau}_i+\tilde{\nu}_i^2\tilde{\tau}_i)\right)\exp\left(\tilde{\nu}_i c_i^T x-\frac{1}{2}\tilde{\tau}_i x^T c_i c_i^T x\right)\right) \\
&= \exp\left(-\frac{1}{2}(n\log 2\pi+\log|K|+m^T K^{-1}m)\right)\times \\
&\prod_{i=1}^{n}\left(\exp\left(-\frac{1}{2}(\log 2\pi-\log\tilde{\tau}_i+\tilde{\nu}_i^2\tilde{\tau}_i)\right)\exp\left((m^T K^{-1}+\sum_{i=1}^{n}\tilde{\nu}_i c_i^T)x-\frac{1}{2}x^T(K^{-1}+\sum_{i=1}^{n}\tilde{\tau}_i c_i c_i^T)x\right)\right)
\end{aligned}
$$

We see that this is a multivariate Gaussian distribution in canonical parameterization, with

$$\eta = K^{-1}m+\sum_{i=1}^{n}\tilde{\nu}_i c_i \tag{33}$$

$$\Lambda = K^{-1}+\sum_{i=1}^{n}\tilde{\tau}_i c_i c_i^T \tag{34}$$

and log normalizing constant

$$-\frac{1}{2}(n\log 2\pi+\log|K|+m^T K^{-1}m)+\sum_{i=1}^{n}-\frac{1}{2}(\log 2\pi-\log\tilde{\tau}_i+\tilde{\nu}_i^2\tilde{\tau}_i) \tag{35}$$