Lecture Notes: Gaussian identities

Marc Toussaint

January 25, 2011

Definitions
We define a Gaussian over x with mean a and covariance matrix A as the function

$$\mathcal{N}(x\,|\,a,A) = \frac{1}{|2\pi A|^{1/2}}\ \exp\{-\frac{1}{2}(x\text{-}a)^\top\ A^{\text{-}1}\ (x\text{-}a)\} \tag{1}$$

with property $N(x\,|\,a,A) = N(a|\ x,A)$. We also define the canonical form with precision matrix A as

$$\mathcal{N}[x\,|\,a,A] = \frac{\exp\{-\frac{1}{2}a^\top A^{\text{-}1} a\}}{|2\pi A^{\text{-}1}|^{1/2}}\ \exp\{-\frac{1}{2}x^\top\ A\ x + x^\top a\} \tag{2}$$

with properties

$$\mathcal{N}[x\,|\,a,A] = \mathcal{N}(x\,|\,A^{\text{-}1}a, A^{\text{-}1}) \tag{3}$$
$$\mathcal{N}(x\,|\,a,A) = \mathcal{N}[x\,|\,A^{\text{-}1}a, A^{\text{-}1}]\ . \tag{4}$$

Non-normalized Gaussian

$$\overline{\mathcal{N}}(x,a,A) = |2\pi A|^{1/2}\ \mathcal{N}(x|a,A) \tag{5}$$
$$= \exp\{-\frac{1}{2}(x\text{-}a)^\top\ A^{\text{-}1}\ (x\text{-}a)\} \tag{6}$$

Matrices [matrix cookbook]

$$(A^{\text{-}1} + B^{\text{-}1})^{\text{-}1} = A\ (A\text{+}B)^{\text{-}1}\ B = B\ (A\text{+}B)^{\text{-}1}\ A \tag{7}$$
$$(A^{\text{-}1} - B^{\text{-}1})^{\text{-}1} = A\ (B\text{-}A)^{\text{-}1}\ B \tag{8}$$
$$\partial_x|A_x| = |A_x|\ \text{tr}(A_x^{\text{-}1}\ \partial_x A_x) \tag{9}$$
$$\partial_x A_x^{\text{-}1} = -A_x^{\text{-}1}\ (\partial_x A_x)\ A_x^{\text{-}1} \tag{10}$$
$$(A + UBV)^{\text{-}1} = A^{\text{-}1} - A^{\text{-}1}U(B^{\text{-}1} + VA^{\text{-}1}U)^{\text{-}1}VA^{\text{-}1} \tag{11}$$
$$(A^{\text{-}1} + B^{\text{-}1})^{\text{-}1} = A - A(B + A)^{\text{-}1}A \tag{12}$$
$$(A + J^\top BJ)^{\text{-}1}J^\top B = A^{\text{-}1}J^\top(B^{\text{-}1} + JA^{\text{-}1}J^\top)^{\text{-}1} \tag{13}$$
$$(A + J^\top BJ)^{\text{-}1}A = \mathbf{I} - (A + J^\top BJ)^{\text{-}1}J^\top BJ \tag{14}$$

(11)=Woodbury; (13,14) holds for pos def A and B

Derivatives

$$\partial_x\mathcal{N}(x|a,A) = \mathcal{N}(x|a,A)\ (-h^\top)\ , \quad h := A^{\text{-}1}(x\text{-}a) \tag{15}$$
$$\partial_\theta\mathcal{N}(x|a,A) = \mathcal{N}(x|a,A)\ \cdot$$
$$\Big[-h^\top(\partial_\theta x) + h^\top(\partial_\theta a) - \frac{1}{2}\text{tr}(A^{\text{-}1}\ \partial_\theta A) + \frac{1}{2}h^\top(\partial_\theta A)h\Big] \tag{16}$$
$$\partial_\theta\mathcal{N}[x|a,A] = \mathcal{N}[x|a,A]\ \Big[-\frac{1}{2}x^\top\partial_\theta Ax + \frac{1}{2}a^\top A^{\text{-}1}\partial_\theta AA^{\text{-}1}a$$
$$+ x^\top\partial_\theta a - a^\top A^{\text{-}1}\partial_\theta a + \frac{1}{2}\text{tr}(\partial_\theta AA^{\text{-}1})\Big] \tag{17}$$
$$\partial_\theta\overline{\mathcal{N}}_x(a,A) = \overline{\mathcal{N}}_x(a,A)\ \cdot$$
$$\Big[h^\top(\partial_\theta x) + h^\top(\partial_\theta a) + \frac{1}{2}h^\top(\partial_\theta A)h\Big] \tag{18}$$

Product
The product of two Gaussians can be expressed as

$$\mathcal{N}(x\,|\,a,A)\ \mathcal{N}(x\,|\,b,B)$$
$$= \mathcal{N}[x\,|\,A^{\text{-}1}a + B^{\text{-}1}b, A^{\text{-}1} + B^{\text{-}1}]\ \mathcal{N}(a\,|\,b, A+B)\ , \tag{19}$$
$$= \mathcal{N}(x\,|\,B(A\text{+}B)^{\text{-}1}a + A(A\text{+}B)^{\text{-}1}b, A(A\text{+}B)^{\text{-}1}B)\ \mathcal{N}(a\,|\,b, A+B)\ , \tag{20}$$

$$\mathcal{N}[x\,|\,a,A]\ \mathcal{N}[x\,|\,b,B]$$
$$= \mathcal{N}[x\,|\,a+b, A+B]\ \mathcal{N}(A^{\text{-}1}a\,|\,B^{\text{-}1}b, A^{\text{-}1} + B^{\text{-}1}) \tag{21}$$
$$= \mathcal{N}[x|\dots]\ \mathcal{N}[A^{\text{-}1}a\,|\,A(A\text{+}B)^{\text{-}1}b, A(A\text{+}B)^{\text{-}1}B] \tag{22}$$
$$= \mathcal{N}[x|\dots]\ \mathcal{N}[A^{\text{-}1}a\,|\,(1\text{-}B(A\text{+}B)^{\text{-}1})\ b,\ (1\text{-}B(A\text{+}B)^{\text{-}1})\ B]\ , \tag{23}$$

$$\mathcal{N}(x\,|\,a,A)\ \mathcal{N}[x\,|\,b,B]$$
$$= \mathcal{N}[x\,|\,A^{\text{-}1}a + b, A^{\text{-}1} + B]\ \mathcal{N}(a\,|\,B^{\text{-}1}b, A + B^{\text{-}1}) \tag{24}$$
$$= \mathcal{N}[x|\dots]\ \mathcal{N}[a\,|\,(1\text{-}B(A^{\text{-}1}\text{+}B)^{\text{-}1})\ b,\ (1\text{-}B(A^{\text{-}1}\text{+}B)^{\text{-}1})\ B] \tag{25}$$

Convolution

$$\int_x \mathcal{N}(x\,|\,a,A)\ \mathcal{N}(y-x\,|\,b,B)\ dx = \mathcal{N}(y\,|\,a+b, A+B) \tag{26}$$

Division

$$\mathcal{N}(x|a,A)\ /\ \mathcal{N}(x|b,B) = \mathcal{N}(x|c,C)\ /\ \mathcal{N}(c|b, C+B)$$
$$C^{\text{-}1}c = A^{\text{-}1}a - B^{\text{-}1}b$$
$$C^{\text{-}1} = A^{\text{-}1} - B^{\text{-}1} \tag{27}$$
$$\mathcal{N}[x|a,A]\ /\ \mathcal{N}[x|b,B] \propto \mathcal{N}[x|a-b, A-B] \tag{28}$$

Expectations
Let $x \sim \mathcal{N}(x\,|\,a,A)$,

$$\text{E}_x\{g(x)\} := \int_x \mathcal{N}(x\,|\,a,A)\ g(x)\ dx \tag{29}$$
$$\text{E}_x\{x\} = a\ , \quad \text{E}_x\{xx^\top\} = A + aa^\top \tag{30}$$
$$\text{E}_x\{f + Fx\} = f + Fa \tag{31}$$
$$\text{E}_x\{x^\top x\} = a^\top a + \text{tr}(A) \tag{32}$$
$$\text{E}_x\{(x\text{-}m)^\top R(x\text{-}m)\} = (a\text{-}m)^\top R(a\text{-}m) + \text{tr}(RA) \tag{33}$$

Transformation Linear transformations imply the following identities,

$$\mathcal{N}(x\,|\,a,A) = \mathcal{N}(x+f\,|\,a+f,\ A)\ ,\quad \mathcal{N}(x\,|\,a,A) = |F|\ \mathcal{N}(Fx\,|\,Fa,\ FAF^\top) \tag{34}$$

$$\mathcal{N}(Fx+f\,|\,a,A) = \frac{1}{|F|}\ \mathcal{N}(x\,|\ F^{\text{-}1}(a-f),\ F^{\text{-}1}AF^{\text{-}\top}) = \frac{1}{|F|}\ \mathcal{N}[x\,|\ F^\top A^{\text{-}1}(a-f),\ F^\top A^{\text{-}1}F]\ , \tag{35}$$

$$\mathcal{N}[Fx+f\,|\,a,A] = \frac{1}{|F|}\ \mathcal{N}[x\,|\ F^\top(a-Af),\ F^\top AF]\ . \tag{36}$$

"Propagation" (propagating a message along a coupling, using eqs (19) and (25), respectively)

$$\int_y \mathcal{N}(x\,|\,a+Fy,A)\ \mathcal{N}(y\,|\,b,B)\ dy = \mathcal{N}(x\,|\,a+Fb, A+FBF^\top) \tag{37}$$

$$\int_y \mathcal{N}(x\,|\,a+Fy,A)\ \mathcal{N}[y\,|\,b,B]\ dy = \mathcal{N}[x\,|\,(F^{\text{-}\top}\text{-}K)(b+BF^{\text{-}1}a),\ (F^{\text{-}\top}\text{-}K)BF^{\text{-}1}]\ ,\quad K = F^{\text{-}\top}B(F^{\text{-}\top}A^{\text{-}1}F^{\text{-}1}\text{+}B)^{\text{-}1} \tag{38}$$

marginal & conditional:

$$\mathcal{N}(x\,|\,a,A)\ \mathcal{N}(y\,|\,b+Fx,B) = \mathcal{N}\left(\begin{matrix}x\\y\end{matrix}\,\middle|\,\begin{matrix}a\\b+Fa\end{matrix},\ \begin{matrix}A & A^\top F^\top\\ FA & B\text{+}FA^\top F^\top\end{matrix}\right) \tag{39}$$

$$\mathcal{N}\left(\begin{matrix}x\\y\end{matrix}\,\middle|\,\begin{matrix}a\\b\end{matrix},\ \begin{matrix}A & C\\ C^\top & B\end{matrix}\right) = \mathcal{N}(x\,|\,a,A)\cdot\mathcal{N}(y\,|\,b+C^\top A^{\text{-}1}(x\text{-}a),\ B-C^\top A^{\text{-}1}C) \tag{40}$$

$$\mathcal{N}[x\,|\,a,A]\ \mathcal{N}(y\,|\,b+Fx,B) = \mathcal{N}\left[\begin{matrix}x\\y\end{matrix}\,\middle|\,\begin{matrix}a+F^\top B^{\text{-}1}b\\ B^{\text{-}1}b\end{matrix},\ \begin{matrix}A+F^\top B^{\text{-}1}F & -F^\top B^{\text{-}1}\\ -B^{\text{-}1}F & B^{\text{-}1}\end{matrix}\right] \tag{41}$$

$$\mathcal{N}[x\,|\,a,A]\ \mathcal{N}[y\,|\,b+Fx,B] = \mathcal{N}\left[\begin{matrix}x\\y\end{matrix}\,\middle|\,\begin{matrix}a+F^\top B^{\text{-}1}b\\ b\end{matrix},\ \begin{matrix}A+F^\top B^{\text{-}1}F & -F^\top\\ -F & B\end{matrix}\right] \tag{42}$$

$$\mathcal{N}\left[\begin{matrix}x\\y\end{matrix}\,\middle|\,\begin{matrix}a\\b\end{matrix},\ \begin{matrix}A & C\\ C^\top & B\end{matrix}\right] = \mathcal{N}[x\,|\,a-CB^{\text{-}1}b,\ A-CB^{\text{-}1}C^\top]\cdot\mathcal{N}[y\,|\,b-C^\top x, B] \tag{43}$$

$$\left|\begin{matrix}A & C\\ D & B\end{matrix}\right| = |A|\ |\widehat{B}| = |\widehat{A}|\ |B|\ ,\text{where}\ \begin{matrix}\widehat{A} = A-CB^{\text{-}1}D\\ \widehat{B} = B-DA^{\text{-}1}C\end{matrix} \tag{44}$$

$$\left[\begin{matrix}A & C\\ D & B\end{matrix}\right]^{\text{-}1} = \left[\begin{matrix}\widehat{A}^{\text{-}1} & -A^{\text{-}1}C\widehat{B}^{\text{-}1}\\ -\widehat{B}^{\text{-}1}DA^{\text{-}1} & \widehat{B}^{\text{-}1}\end{matrix}\right] = \left[\begin{matrix}\widehat{A}^{\text{-}1} & -\widehat{A}^{\text{-}1}CB^{\text{-}1}\\ -B^{\text{-}1}D\widehat{A}^{\text{-}1} & \widehat{B}^{\text{-}1}\end{matrix}\right] \tag{45}$$

pair-wise belief We have a message $\alpha(x) = \mathcal{N}[x|s,S]$, transition $P(y|x) = \mathcal{N}(y|Ax+a,Q)$, and a message $\beta(y) = \mathcal{N}[y|v,V]$, what is the belief $b(y,x) = \alpha(x)P(y|x)\beta(y)$?

$$b(y,x) = \mathcal{N}[x|s,S]\ \mathcal{N}(y|Ax+a,Q^{\text{-}1})\ \mathcal{N}[y|v,V] \tag{46}$$

$$= \mathcal{N}\left[\begin{matrix}x\\y\end{matrix}\,\middle|\,\begin{matrix}s\\0\end{matrix},\ \begin{matrix}S & 0\\ 0 & 0\end{matrix}\right]\ \mathcal{N}\left[\begin{matrix}x\\y\end{matrix}\,\middle|\,\begin{matrix}A^\top Q^{\text{-}1}a\\ Q^{\text{-}1}a\end{matrix},\ \begin{matrix}A^\top Q^{\text{-}1}A & -A^\top Q^{\text{-}1}\\ -Q^{\text{-}1}A & Q^{\text{-}1}\end{matrix}\right]\ \mathcal{N}\left[\begin{matrix}x\\y\end{matrix}\,\middle|\,\begin{matrix}0\\v\end{matrix},\ \begin{matrix}0 & 0\\ 0 & V\end{matrix}\right] \tag{47}$$

$$\propto \mathcal{N}\left[\begin{matrix}x\\y\end{matrix}\,\middle|\,\begin{matrix}s+A^\top Q^{\text{-}1}a\\ v+Q^{\text{-}1}a\end{matrix},\ \begin{matrix}S+A^\top Q^{\text{-}1}A & -A^\top Q^{\text{-}1}\\ -Q^{\text{-}1}A & V+Q^{\text{-}1}\end{matrix}\right] \tag{48}$$

Entropy

$$H(\mathcal{N}(a,A)) = \frac{1}{2}\log|2\pi eA| \tag{49}$$

Kullback-Leibler divergence

$$p = \mathcal{N}(x|a,A)\ ,\quad q = \mathcal{N}(x|b,B)\ ,\quad n = \dim(x)\ ,\quad D\big(p\,\big\|\,q\big) = \sum_x p(x)\log\frac{p(x)}{q(x)} \tag{50}$$

$$2\ D\big(p\,\big\|\,q\big) = \log\frac{|B|}{|A|} + \text{tr}(B^{\text{-}1}A) + (b-a)^\top B^{\text{-}1}(b-a) - n \tag{51}$$

$$4\ D_{\text{sym}}\big(p\,\big\|\,q\big) = \text{tr}(B^{\text{-}1}A) + \text{tr}(A^{\text{-}1}B) + (b-a)^\top(A^{\text{-}1}+B^{\text{-}1})(b-a) - 2n \tag{52}$$

λ-divergence

$$2\ D_\lambda\big(p\,\big\|\,q\big) = \lambda\ D\big(p\,\big\|\,\lambda p + (1\text{-}\lambda)q\big)\ +\ (1\text{-}\lambda)\ D\big(p\,\big\|\,(1\text{-}\lambda)p + \lambda q\big) \tag{53}$$

For $\lambda = .5$: Jensen-Shannon divergence.

Log-likelihoods

$$\log \mathcal{N}(x|a,A) = -\frac{1}{2}\Big[log|2\pi A| + (x\text{-}a)^\top\ A^{\text{-}1}\ (x\text{-}a)\Big] \tag{54}$$

$$\log \mathcal{N}[x|a,A] = -\frac{1}{2}\Big[log|2\pi A^{\text{-}1}| + a^\top A^{\text{-}1} a + x^\top A x - 2x^\top a\Big] \tag{55}$$

$$\sum_x \mathcal{N}(x|b,B) \log \mathcal{N}(x|a,A) = -D\big(\mathcal{N}(b,B)\,\|\,\mathcal{N}(a,A)\big) - H(\mathcal{N}(b,B)) \tag{56}$$

Mixture of Gaussians Collapsing a MoG into a single Gaussian

$$\operatorname*{argmin}_{b,B} D\big(\sum_i p_i\ \mathcal{N}(a_i,A_i)\,\|\,\mathcal{N}(b,B)\big) \quad = \quad \Big(b = \sum_i p_i a_i\ ,\ B = \sum_i p_i (A_i + a_i a_i^\top - b\, b^\top)\Big) \tag{57}$$