Efficient Operations on Gaussian Distributions and Factors

Ralf Herbrich

August 18, 2023

Abstract

In this short note, we are summarizing some useful identities for working with Gaussian distributions and factors. In particular, we show how to multiply and divide two Gaussian distributions and how to compute the product of a Gaussian distribution and a linear function of a Gaussian distribution. These identities are useful for message passing and Bayesian inference.

1 Gaussian Distributions

In practical probabilistic machine learning, one of the most common distributions that is used is the Gaussian distribution. In this note, we will focus on one-dimensional Gaussian distributions. This distribution can be represented in two different ways:

$$N\left(x;\mu,\sigma^2,\gamma\right) := \exp(\gamma)\cdot\sqrt{\frac{1}{2\pi\sigma^2}}\cdot\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \qquad \textbf{(location-scale parameters)} \tag{1}$$

$$G(x;\tau,\rho,\gamma) := \exp(\gamma)\cdot\sqrt{\frac{\rho}{2\pi}}\cdot\exp\left(-\frac{\tau^2}{2\rho}\right)\cdot\exp\left(\tau\cdot x+\rho\cdot\left(-\frac{x^2}{2}\right)\right), \qquad \textbf{(natural parameters)} \tag{2}$$

where $\int_{-\infty}^{+\infty} N\left(x;\mu,\sigma^2,\gamma\right)\,dx = \int_{-\infty}^{+\infty} G(x;\tau,\rho,\gamma)\,dx = \exp(\gamma)$ for all values of $\mu,\tau,\sigma\geq 0,\rho\geq 0$ and γ Note that the following transformations allow us to easily switch between the two different representations:

$$N\left(x;\mu,\sigma^2,\gamma\right) = G\left(x;\mu\cdot\sigma^{-2},\sigma^{-2},\gamma\right), \tag{3}$$

$$G(x;\tau,\rho,\gamma) = N\left(x;\tau\cdot\rho^{-1},\rho^{-1},\gamma\right). \tag{4}$$

if we drop the third argument, we implicitly assume it to be zero, i.e. $N\left(x;\mu,\sigma^2\right) := N\left(x;\mu,\sigma^2,0\right)$ and $G(x;\tau,\rho) := G(x;\tau,\rho,0)$.

Also note that there are two interesting limit cases: For $\sigma^2\to\infty$, the Gaussian density scaled by $\sqrt{2\pi\sigma^2}$ converges to the constant function $c(\cdot)$ and for the limit case of $\sigma^2\to 0$, the Gaussian density corresponds to a Dirac delta function $\delta(\cdot)$. More formally, we have

$$c(x) := \lim_{\sigma^2\to\infty} N\left(x;0,\sigma^2\right)\cdot\sqrt{2\pi\sigma^2} = \lim_{\sigma^2\to\infty}\exp\left(-\frac{x^2}{2\sigma^2}\right) = 1, \tag{5}$$

$$\delta(x) := \lim_{\sigma^2\to 0} N\left(x;0,\sigma^2,0\right). \tag{6}$$

Note that the speed of convergence of $\lim_{\sigma^2\to\infty}\exp\left(-\frac{x^2}{2\sigma^2}\right)$ is not the same for each x (but the limit value is).

1.1 Multiplication

One of the most frequent operations that we need to perform in message passing and Bayesian inference is multiplying two Gaussian distributions and re-normalizing. The following theorem states an efficient and numerically stable way to achieve this as it relies on additions (mostly) once we switch to natural parameters.

Theorem 1. *Given two non-normalized one-dimensional Gaussian distributions* $G(x;\tau_1,\rho_1,\gamma_1)$ *and* $G(x;\tau_2,\rho_2,\gamma_2)$ *over the same variable* x *we have*

$$G(x;\tau_1,\rho_1,\gamma_1)\cdot G(x;\tau_2,\rho_2,\gamma_2) = G(x;\tau_1+\tau_2,\rho_1+\rho_2,\gamma_1+\gamma_2)\cdot N\left(\mu_1;\mu_2,\sigma_1^2+\sigma_2^2\right), \tag{7}$$

$$= G\left(x;\tau_1+\tau_2,\rho_1+\rho_2,\gamma_1+\gamma_2-\frac{1}{2}\left(\log\left(2\pi\left(\sigma_1^2+\sigma_2^2\right)\right)+\frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\right)\right)$$

where $\sigma_1^2=\rho^{-1}$ *and* $\mu_1=\tau_1\cdot\rho^{-1}$ *(and similarly for* σ_2^2 *and* μ_2*, respectively).*

The proof can be found in Appendix A. By definition (5), we have $G(x;\tau,\rho,\gamma)\cdot c(x)=G(x;\tau,\rho,\gamma)$ for all τ,ρ,γ and x. This can also be shown using the result of Theorem 1: Using the definition of $c(x)$ in (5) we have

$$\begin{aligned} G(x;\tau,\rho,\gamma)\cdot c(x) &= \lim_{\sigma^2\to\infty} G(x;\tau,\rho,\gamma)\cdot N(x;0,\sigma_2,0)\cdot\sqrt{2\pi\sigma^2} \\ &= \lim_{\sigma^2\to\infty} G(x;\tau,\rho+\sigma^{-2},\gamma)\cdot N\left(\rho^{-1}\tau;0,\rho^{-1}+\sigma^2\right)\cdot\sqrt{2\pi\sigma^2} \\ &= G(x;\tau,\rho,\gamma)\cdot\underbrace{\lim_{\sigma^2\to\infty}\sqrt{\frac{\rho}{2\pi\sigma^2\cdot(\rho+\sigma^{-2})}}\cdot\exp\left(-\frac{\rho^{-2}\tau^2}{2\left(\rho^{-1}+\sigma^2\right)}\right)\cdot\sqrt{2\pi\sigma^2}}_{=1} \end{aligned}$$

1.2 Division

An equally frequent operation that we need to perform in message passing is dividing two Gaussian distributions and re-normalizing them. The following theorem states an efficient and numerically stable way to achieve this.

Theorem 2. *Given two non-normalized one-dimensional Gaussian distributions* $G(x;\tau_1,\rho_1,\gamma_1)$ *and* $G(x;\tau_2,\rho_2,\gamma_2)$ *over the same variable* x *we have*

$$\frac{G(x;\tau_1,\rho_1,\gamma_1)}{G(x;\tau_2,\rho_2,\gamma_2)} = G(x;\tau_1-\tau_2,\rho_1-\rho_2,\gamma_1-\gamma_2)\cdot\frac{1}{N\left(\frac{\tau_1-\tau_2}{\rho_1-\rho_2};\frac{\tau_2}{\rho_2},\frac{1}{\rho_1-\rho_2}+\frac{1}{\rho_2}\right)}, \tag{8}$$

$$= G\left(x;\tau_1-\tau_2,\rho_1-\rho_2,\gamma_1-\gamma_2+\log\left(\sigma_2^2\right)+\frac{1}{2}\left(\log\left(\frac{2\pi}{\sigma_2^2-\sigma_1^2}\right)+\frac{(\mu_1-\mu_2)^2}{\sigma_2^2-\sigma_1^2}\right)\right), \tag{9}$$

where $\sigma_1^2=\rho_1^{-1}$ *and* $\mu_1=\tau_1\cdot\rho_1^{-1}$ *(and similarly for* σ_2^2 *and* μ_2*, respectively).*

1.3 Linear Function of Gaussians

When we consider the computation of a Gaussian posterior, we often have a prior $N\left(w;\mu,\sigma^2,\gamma\right)$ over w and a likelihood $N\left(y;aw+b,\beta^2,0\right)$ where the mean is a linear function of the parameter w. If we want to use Theorem 1 or 2, we need to change the likelihood into a Gaussian distribution over w.

Theorem 3. *Given a non-normalized one-dimensional Gaussian distributions* $N\left(y;aw+b,\beta^2,\gamma\right)$ *we have for any* $a\neq 0$, $b\in\mathbb{R}$ *and* $\beta\in\mathbb{R}^+$

$$N\left(y;aw+b,\beta^2,\gamma\right) = N\left(w;a^{-1}(y-b),a^{-2}\beta^2,\gamma-\log(a)\right). \tag{10}$$

Corollary 1. *For any* $a\neq 0$, $b\in\mathbb{R}$ *and* $\beta\in\mathbb{R}^+$*, given a Gaussian prior distribution* $p(w)=N\left(w;\mu,\sigma^2,0\right)$*, and a Gaussian likelihood of a linear function of* w, $p(y|w)=N\left(y;aw+b,\beta^2,0\right)$*, we have the following*

$$p(w|y) = N\left(w,m,s^2,0\right), \tag{11}$$

$$p(y) = N\left(y;a\mu+b,\beta^2+a^2\sigma^2,0\right), \tag{12}$$

where $s^2=\left(\sigma^{-2}+a^2\beta^{-2}\right)^{-1}$ *and* $m=s^2\cdot\left(a\beta^{-2}(y-b)+\sigma^{-2}\mu\right)$.

A Proof for Gaussian Operations

Proof of Theorem 1. Using (2) we see that the left-hand side of (7) equals

$$\exp(\gamma_1+\gamma_2)\cdot\sqrt{\frac{\rho_1\rho_2}{(2\pi)^2}}\cdot\exp\left(-\frac{\tau_1^2}{2\rho_1}-\frac{\tau_2^2}{2\rho_2}\right)\cdot\exp\left((\tau_1+\tau_2)\cdot x+(\rho_1+\rho_2)\cdot\left(-\frac{x^2}{2}\right)\right).$$

Next, we divide this expression by $G(x;\tau_1+\tau_2,\rho_1+\rho_2,\gamma_1+\gamma_2)$ to obtain

$$\sqrt{\frac{\rho_1\rho_2}{2\pi(\rho_1+\rho_2)}}\cdot\exp\left(-\frac{\tau_1^2}{2\rho_1}-\frac{\tau_2^2}{2\rho_2}+\frac{(\tau_1+\tau_2)^2}{2(\rho_1+\rho_2)}\right).$$

It remains to show that this expression equals $N\left(\mu_1;\mu_2,\sigma_1^2+\sigma_2^2,0\right)$. Using (1) this is equivalent to

$$\sqrt{\frac{\rho_1\rho_2}{2\pi(\rho_1+\rho_2)}}=\sqrt{\frac{1}{2\pi\left(\sigma_1^2+\sigma_2^2\right)}}\quad\text{and}\quad-\frac{\tau_1^2}{\rho_1}-\frac{\tau_2^2}{\rho_2}+\frac{(\tau_1+\tau_2)^2}{\rho_1+\rho_2}=-\frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}.$$

Let's start with the first equality. Expanding (3) we see that

$$\rho_1\rho_2\left(\rho_1+\rho_2\right)^{-1}=\rho_1\rho_2\left(\rho_2\left(\rho_1^{-1}+\rho_2^{-1}\right)\rho_1\right)^{-1}=\left(\rho_1^{-1}+\rho_2^{-1}\right)^{-1}=\frac{1}{\sigma_1^2+\sigma_2^2},$$

which proves the first equality. In order to prove the second equality, we use (3) and $\tau=\mu\cdot\rho$ again to obtain

$$\begin{aligned}
-\frac{\tau_1^2}{\rho_1}-\frac{\tau_2^2}{\rho_2}+\frac{(\tau_1+\tau_2)^2}{\rho_1+\rho_2}&=-\mu_1^2\rho_1^2\rho_1^{-1}-\mu_2^2\rho_2^2\rho_2^{-1}+(\mu_1\rho_1+\mu_2\rho_2)^2\left(\rho_1+\rho_2\right)^{-1}\\
&=-\mu_1^2\rho_1-\mu_2^2\rho_2+\left(\rho_2\left(\mu_1\rho_2^{-1}+\mu_2\rho_1^{-1}\right)\rho_1\right)^2\left(\rho_2\left(\rho_1^{-1}+\rho_2^{-1}\right)\rho_1\right)^{-1}\\
&=-\mu_1^2\rho_1-\mu_2^2\rho_2+\left(\mu_1\rho_2^{-1}+\mu_2\rho_1^{-1}\right)^2\cdot\rho_2\rho_1\left(\rho_1^{-1}+\rho_2^{-1}\right)^{-1}\\
&=\frac{\left[-\mu_1^2\rho_2^{-1}\left(\rho_1^{-1}+\rho_2^{-1}\right)-\mu_2^2\rho_1^{-1}\left(\rho_1^{-1}+\rho_2^{-1}\right)+\left(\mu_1\rho_2^{-1}+\mu_2\rho_1^{-1}\right)^2\right]\cdot\rho_2\rho_1}{\rho_1^{-1}+\rho_2^{-1}}\\
&=\frac{\left[-\mu_1^2\rho_2^{-1}\rho_1^{-1}-\mu_1^2\rho_2^{-2}-\mu_2^2\rho_1^{-2}-\mu_2^2\rho_1^{-1}\rho_2^{-1}+\mu_1^2\rho_2^{-2}+2\mu_1\mu_2\rho_1^{-1}\rho_2^{-1}+\mu_2^2\rho_1^{-2}\right]\cdot\rho_2\rho_1}{\rho_1^{-1}+\rho_2^{-1}}\\
&=\frac{\left[-\mu_1^2-\mu_2^2+2\mu_1\mu_2\right]\cdot\rho_1^{-1}\rho_2^{-1}\rho_2\rho_1}{\rho_1^{-1}+\rho_2^{-1}}\\
&=-\frac{(\mu_1-\mu_2)^2}{2\left(\sigma_1^2+\sigma_2^2\right)}.
\end{aligned}$$

The final line follows from using (1) and noticing that

$$\log\left(N\left(\mu_1;\mu_2,\sigma_1^2+\sigma_2^2,0\right)\right)=-\frac{1}{2}\left(\log\left(2\pi\left(\sigma_1^2+\sigma_2^2\right)\right)+\frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\right).$$

□

Proof of Theorem 2. The first equality follows directly from Theorem 1. Rewriting (7) and dividing the expression by $G(x;\tau_2,\rho_2,\gamma_2)$ and $N\left(\mu_3;\mu_2,\sigma_2^2+\sigma_3^2\right)$ and we see that

$$\frac{G(x;\tau_3,\rho_3,\gamma_3)}{N\left(\mu_3;\mu_2,\sigma_2^2+\sigma_3^2,0\right)}=\frac{G(x;\tau_2+\tau_3,\rho_2+\rho_3,\gamma_2+\gamma_3)}{G(x;\tau_2,\rho_2,\gamma_2)}$$

Now setting $\tau_1=\tau_2+\tau_3$, $\rho_1=\rho_2+\rho_3$ and $\gamma_1=\gamma_2+\gamma_3$ and rearranging for τ_3, ρ_3 and γ_3 we have

$$\frac{G(x;\tau_1,\rho_1,\gamma_1)}{G(x;\tau_2,\rho_2,\gamma_2)}=G(x;\tau_1-\tau_2,\rho_1-\rho_2,\gamma_1-\gamma_2)\cdot\frac{1}{N\left(\frac{\tau_1-\tau_2}{\rho_1-\rho_2};\frac{\tau_2}{\rho_2},\frac{1}{\rho_1-\rho_2}+\frac{1}{\rho_2},0\right)},$$

where we used (3) in the $N(\cdot)$ term. It remains to show that

$$-\log\left(N\left(0;\underbrace{\frac{\tau_1-\tau_2}{\rho_1-\rho_2}-\frac{\tau_2}{\rho_2}}_{\mu},\underbrace{\frac{1}{\rho_1-\rho_2}+\frac{1}{\rho_2}}_{\sigma^2},0\right)\right)=\log\left(\sigma_2^2\right)+\frac{1}{2}\left(\log\left(\frac{2\pi}{\sigma_2^2-\sigma_1^2}\right)+\frac{(\mu_1-\mu_2)^2}{\sigma_2^2-\sigma_1^2}\right).$$

Let us start with deriving the expression for σ^2. By virtue of (3) we have

$$\sigma^2=\frac{1}{\rho_1-\rho_2}+\frac{1}{\rho_2}=\frac{\rho_2+(\rho_1-\rho_2)}{(\rho_1-\rho_2)\rho_2}=\frac{\rho_1}{(\rho_1-\rho_2)\rho_2}=\frac{\sigma_2^2}{\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}\right)\sigma_1^2}=\frac{\sigma_2^2}{\left(\frac{\sigma_2^2-\sigma_1^2}{\sigma_1^2\sigma_2^2}\right)\sigma_1^2}=\frac{\sigma_2^2}{\sigma_2^2-\sigma_1^2}\cdot\sigma_2^2.$$

Also, using (3) for μ we get

$$\mu=\frac{\tau_1-\tau_2}{\rho_1-\rho_2}-\frac{\tau_2}{\rho_2}=\frac{\frac{\mu_1}{\sigma_1^2}-\frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}}-\mu_2=\frac{\mu_1\sigma_2^2-\mu_2\sigma_1^2}{\sigma_2^2-\sigma_1^2}-\mu_2=\frac{\mu_1\sigma_2^2-\mu_2\sigma_1^2-\mu_2\left(\sigma_2^2-\sigma_1^2\right)}{\sigma_2^2-\sigma_1^2}=\frac{\mu_1-\mu_2}{\sigma_2^2-\sigma_1^2}\cdot\sigma_2^2.$$

Finally, using (1) we have

$$-\log\left(N\left(0;\mu,\sigma^2,0\right)\right)=-\log\left(\sqrt{\frac{1}{2\pi\sigma^2}}\right)+\frac{\mu^2}{2\sigma^2}=-\frac{1}{2}\log\left(\frac{\sigma_2^2-\sigma_1^2}{2\pi\sigma_2^4}\right)+\frac{1}{2}\frac{(\mu_1-\mu_2)^2}{\sigma_2^2-\sigma_1^2}.$$

□

Proof of Theorem 3. Using the definition (1) we see that

$$\begin{aligned}
N\left(y;aw+b,\beta^2,\gamma\right)&=\exp(\gamma)\cdot\sqrt{\frac{1}{2\pi\beta^2}}\cdot\exp\left(-\frac{1}{2}\frac{(y-a\cdot w-b)^2}{\beta^2}\right)\\
&=\exp(\gamma)\cdot\sqrt{\frac{a^{-2}}{2\pi a^{-2}\beta^2}}\cdot\exp\left(-\frac{1}{2}\frac{\left(a\cdot\left(a^{-1}y-w-a^{-1}b\right)\right)^2}{\beta^2}\right)\\
&=\exp(\gamma-\log(a))\cdot\sqrt{\frac{1}{2\pi a^{-2}\beta^2}}\cdot\exp\left(-\frac{1}{2}\frac{\left(a^{-1}(y-b)-w\right)^2}{a^{-2}\beta^2}\right)\\
&=N\left(w;a^{-1}(y-b),a^{-2}\beta^2,\gamma-\log(a)\right).
\end{aligned}$$

□

Proof of Corollary 1. Using Theorem 3, $p(y|w)$ can be written as $N\left(w;a^{-1}(y-b),a^{-2}\beta^2,-\log(a)\right)$. Using (3) and (7)

$$\begin{aligned}
p(w)\cdot p(y|w)&=G\left(w;\sigma^{-2}\mu,\sigma^{-2},0\right)\cdot G\left(w;a\beta^{-2}(y-b),a^2\beta^{-2},-\log(a)\right)\\
&=G\left(w;a\beta^{-2}(y-b)+\sigma^{-2}\mu,\sigma^{-2}+a^2\beta^{-2},-\log(a)\right)\cdot N\left(\mu;a^{-1}(y-b),a^{-2}\beta^2+\sigma^2,0\right)\\
&=G\left(w;a\beta^{-2}(y-b)+\sigma^{-2}\mu,\sigma^{-2}+a^2\beta^{-2},-\log(a)\right)\cdot N\left(y;a\mu+b,\beta^2+a^2\sigma^2,\log(a)\right)\\
&=N\left(w,m,s^2,0\right)\cdot N\left(y;a\mu+b,\beta^2+a^2\sigma^2,0\right)=p(w|y)\cdot p(y).
\end{aligned}$$

□