Technical Note: Computing moments of a truncated Gaussian for EP in high-dimensions

Marc Toussaint
Machine Learning & Robotics group, TU Berlin
Franklinstr. 28/29, FR 6-9, 10587 Berlin, Germany

October 16, 2009

In this technical note we derive an algorithm for computing the moments of a truncated Gaussian in high-dimensions. *In principle, all of this is well known and not novel.* Herbrich has already an (unpublished) technical note on EP with truncated Gaussians at `research.microsoft.com/pubs/74554/EP.pdf`. However, getting an *efficient* algorithm in high-dimension is not so trivial. We derive one in this note. The corresponding source code is available at `user.cs.tu-berlin.de/~mtoussai/source-code/`. Our motivation is the application in the context of Aproximate Inference Control (**?**), where we use approximate inference to compute trajectories under hard constraints: Collision and joint avoidance implis messages of the form of heavyside functions; using Expectation Propagation with truncated Gaussians we can approximate the motion posterior.

1 1D case

Let us first address the simple 1D case. The problem is defined as follows: Let $x \in \mathbb{R}$ and $g(x) = e^{-x^2/2}$ and $\theta(x) = [[x \geq z]]$ (the heavyside function at z). We want to compute a Gaussian approximation of $g(x)\theta(x)$. For this we need to compute the moments of $g(x)\theta(x)$. For the norm (0th moment) we have:

$$\int_0^z e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2}\,\mathrm{erf}(z) \tag{1}$$

$$M_0 := \int_z^\infty e^{-x^2/2}\,dx = \sqrt{2}\int_{z/\sqrt{2}}^\infty e^{-t^2}\,dt$$
$$= \sqrt{\pi/2}\,[1-\mathrm{erf}(z/\sqrt{2})] \tag{2}$$

For the 1st moment:

$$M_1 := \int_z^\infty e^{-x^2/2}\,x\,dx = -\int_{-\frac{1}{2}z^2}^{-\infty} e^t\,dt$$
$$= -\Big[e^t\Big]_{-z^2/2}^{-\infty} = -\Big[0 - e^{-z^2/2}\Big] = e^{-z^2/2} \tag{3}$$

For the n-th moment:

$$I_n(z,a) := \int_z^\infty e^{-ax^2}\,x^n\,dx \tag{4}$$

$$\frac{\partial}{\partial a} I_n(z,a) = \int_z^\infty e^{-ax^2}\,(-x^2)\,x^n\,dx = -I_{n+2}(z,a) \tag{5}$$

$$\frac{\partial}{\partial a}\mathrm{erf}(\sqrt{a}z) = \frac{a^{-1/2}z}{2}\,\frac{2}{\sqrt{\pi}}\,e^{-az^2} \tag{6}$$

And hence for 2nd moment:

$$I_2(z,a) = -\frac{\partial}{\partial a}N(z,a) \tag{7}$$
$$= a^{-3/2}\frac{\sqrt{\pi}}{4}\,[1-\mathrm{erf}(\sqrt{a}z)] + a^{-1/2}\,\frac{\sqrt{\pi}}{2}\,[\frac{a^{-1/2}z}{\sqrt{\pi}}\,e^{-az^2}] \tag{8}$$
$$= a^{-3/2}\frac{\sqrt{\pi}}{4}\,[1-\mathrm{erf}(\sqrt{a}z)] + \frac{z}{2a}\,e^{-az^2} \tag{9}$$

$$M_2 := \int_z^\infty e^{-x^2/2}\,x^2\,dx = I_2(z,\frac{1}{2})$$
$$= \sqrt{\pi/2}\,[1-\mathrm{erf}(z/\sqrt{2})] + z\,e^{-z^2/2} \tag{10}$$
$$= M_0 + zM_1 \tag{11}$$

In summary, we have

$$\text{norm} \quad n := M_0 = \sqrt{\pi/2}\,[1-\mathrm{erf}(z/\sqrt{2})] \tag{12}$$
$$\text{mean} \quad m := M_1/M_0 = e^{-z^2/2}/n \tag{13}$$
$$\text{variance} \quad v := M_2/M_0 - m^2 = 1 + zm - m^2 \tag{14}$$

2 General case

We now have a n-dim Gaussian $f(y)$ and heavyside function $\theta(y)$ along a hyperplane with normal c and offset d,

$$f(y) \propto \exp\{-\frac{1}{2}(y-a)^\top A^{-1}\,(y-a)\} \tag{15}$$
$$\theta(y) = [[c^\top y - d \geq 0]] \tag{16}$$

where $[[\cdot]]$ is the indicator function. We transform this problem such that the Gaussian becomes a standard Gaussian and the constraint is aligned with the x-axis. We need two transformations for this: first a linear transform to standardize the Gaussian, then a rotation to align with the x-axis. Let $A = M^\top M$ be the Cholesky decomposition ($A^{-1} = M^{-1}M^{-\top}$) and we define $x = M^{-\top}(y-a)$. We have

$$f(x) = \exp\{-\frac{1}{2}x^\top x\} \tag{17}$$

Algorithm 1 Truncated Standard Gaussian

Input: z
Output: norm n, mean m, variance v
$n = \sqrt{\pi/2}[1 - \text{erf}(z\sqrt{2})]$
$m = \exp(-z^2/2)/n$
$v = 1 + zm - m^2$

Algorithm 2 Truncate Gaussian

Input: mean a, covariance A, constraint coeffs c, d
Output: mean b, covariance B
$M^T M = A$ *// Cholesky decomposition*
$z = (c^\top a + d)/|Mc|$
$v = Mc/|Mc|$
$R =$ rotation onto v *// as in equation (**??**)*
(m, v) = Truncated Standard Gaussian(z)
$b = M^\top R(m, 0, .., 0) + a$
$B = M^\top R\text{diag}(v, 1, .., 1)R^\top M$

$$\theta(x) = [[c^\top(M^\top x + a - d) \geq 0]] = [[v^\top x + z \geq 0]] \ , \tag{18}$$

$$v := Mc/|Mc| \ , \quad z := [c^\top(a - d)]/|Mc| \tag{19}$$

Note that we defined v to be normalized. (If $|Mc|$ is zero the truncation has no effect or zero likelihood, depending on whether $c^\top a - d > 0$ or $c^\top a - d < 0$, respectively.) We compute a rotation matrix that rotates the unit vector $\boldsymbol{e} = (1, 0, .., 0)$ onto v (implemented in array.cpp). We define $x' = R^{-1}x$. We have $v = R\boldsymbol{e}$ and

$$f(x') = \exp\{-\frac{1}{2}x'^\top x'\} \tag{20}$$

$$\begin{aligned}\theta(x') &= [[v^\top Rx' + z \geq 0]] \\ &= [[(R^{-1}v)^\top x' + z \geq 0]] = [[x'_1 + z \geq 0]]\end{aligned} \tag{21}$$

That is, $\theta(x')$ truncates along the first axis in the x' coordinate system. Given the mean m and variance v of the $(-z)$-truncated standard Gaussian, we have

$$f(x')\ \theta(x') \approx \mathcal{N}(x'|b', B') \tag{22}$$

$$b' = (m, 0, .., 0) \tag{23}$$

$$B' = \text{diag}(v, 1, .., 1) \tag{24}$$

We undo the transformation $x' = R^{-1}M^{-\top}(y - a)$ and get the result

$$f(y)\ \theta(y) \approx \mathcal{N}(y|b, B) \tag{25}$$

$$b = M^\top Rb' + a \tag{26}$$

$$B = M^\top RB'R^\top M \tag{27}$$

which gives the mean and covariance of the truncated Gaussian. The explicit algorithms are given below. The figure illustrates the result of truncating a Gaussian in 2D with the constraint $[[x > 1]]$.

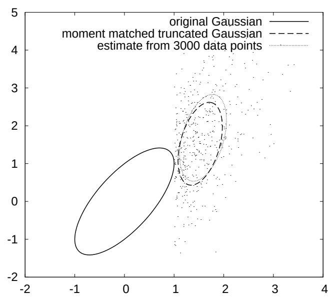

Figure 1: Truncation of a Gaussian at the constraint $[[x > 1]]$ in 2D.