Lecture #3: Primer on Probability and Statistics:

The multi-variate Gaussian distribution:

$$x \in \mathbb{R}^d \quad , \quad x = (x_1, x_2, \ldots, x_d)$$

$$x \sim \mathcal{N}(x \mid \underset{d\times 1}{\mu}, \underset{d\times d}{\Sigma}) = \underbrace{\frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}}}_{\text{normalizing constant}} \exp\left[-\frac{1}{2} \underset{1\times d}{(x-\mu)^T} \underbrace{\underset{d\times d}{\Sigma^{-1}} \underset{d\times 1}{(x-\mu)}}_{d\times 1}\right]$$

$x \sim \mathcal{N}(x|\mu, \Sigma)$

$$\hookrightarrow \quad \log p(x) = -\frac{d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x-\mu^T)\Sigma^{-1}(x-\mu)$$

Key properties:

1.) **Closure under marginalization:**

$$x \in \mathbb{R}^d \quad , \quad y \in \mathbb{R}^q \quad , \quad p(x,y) \sim \mathcal{N}\left(\begin{bmatrix}\mu_x \\ \mu_y\end{bmatrix}, \begin{bmatrix}\Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy}\end{bmatrix}\right).$$

$$p(x) = \int p(x,y)\,dy = \mathcal{N}(\mu_x, \Sigma_{xx})$$

$$p(y) = \int p(x,y)\,dx = \mathcal{N}(\mu_y, \Sigma_{yy})$$

2.) **Closure under conditioning:**

$$p(x|y) \sim \mathcal{N}\Big(\underbrace{\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)}_{\mu_{x|y}}, \underbrace{\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T}_{\Sigma_{x|y}}\Big)$$

Transformations of random variables:

$x \sim p(x)$ is transformed as $x \xrightarrow{g} z$, $z \sim p(z)$?

Affine / Linear transformations:

$$x \xrightarrow{g} z \xrightarrow{g^{-1}} x \quad , \qquad \boxed{x = Az + b}$$

$$z = A^{-1}(x - b)$$

Given $\mu := \mathbb{E}_{z \sim p(z)}[z]$, $\Sigma := Cov[z]$, what is $\mathbb{E}[x]$, $Cov[x]$?

$$\mathbb{E}[x] = \mathbb{E}[Az + b] = A\,\mathbb{E}[z] + b = A\mu + b$$

$$Cov[x] = \mathbb{E}[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^T] = A \Sigma A^T$$

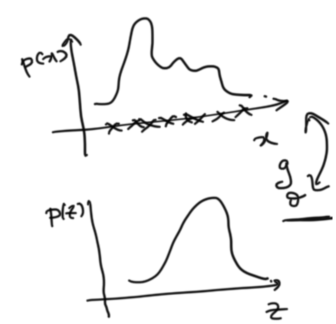
General Transformations:

$x \sim p(x)$, $\in \mathbb{R}^d$ $\qquad x \xrightarrow{g} z \xrightarrow{g^{-1}} x$

$z \sim p(z)$ $\qquad$ Change of variables formula:

$$p(x) = p(z) \left| \det J_{g(x)} \right| , \quad \underset{d \times d}{J_{g(x)}} := \frac{\partial g(x)}{\partial x}$$

Example:



Setup: Given data $\mathcal{D} := \{x_1, \dots, x_n\}$, $x_i \in \mathbb{R}^d$

Goal: Infer $p_\theta(x)$ $\qquad \theta := (A^{-1}, b)$

$$\theta^* = \arg\max_\theta \underbrace{p(x|\theta)}_{\text{likelihood}} \quad (i)$$

$$= \arg\min_\theta - \log p(x|\theta) \quad (ii)$$

$$\log p(x|\theta) = \log p_\theta(x) = \log p(z) + \log \left| \det J_{g_\theta(x)} \right|$$

Normalizing flows