1 Basic Formulas

1.1 Probability

— $p(X) = \sum_Y p(X, y)$ (sum rule)
— $p(X, Y) = p(Y|X)p(X)$ (product rule)
— $p(X) = \sum_Y p(X, y)$ (Bayes' rule)
— $\text{Bern}(x|\theta) = \theta^x(1-\theta)^{(1-x)}$ (Bernoulli)
— $\text{Bin}(m|N, \theta) = \binom{N}{m}\theta^m(1-\theta)^{(N-m)}$ (binomial)
— $\text{Mult}(m_1, \ldots., m_K|\theta, N) = \binom{N}{m_1,\ldots.,m_K}\prod_{k=1}^K \theta_k^{m_k}$ (multinomial)
— $\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} exp\Big(\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big)$ (Gaussian)
— Jensen : $f(\mathbb{E}[x]) \leqslant \mathbb{E}[f(X)]$ for f convex.

1.2 Information Theory

— $I(x) = \log \frac{1}{p(x)}$ (information in occurence of x)
— $H(x) = \mathbb{E}_{p(x)}[I(x)] = -\sum_x p(x) \log p(x)$ (entropy of X)
— $D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$ (Kullback-Leibler divergence)
— $I(X, Y) = \sum_{x,y} p_{X,Y}(x, y) \log \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}$ (mutual information of X,Y).

Author : Lilian Besson, (C) December 2015, MIT License.