Changing the Basis of distributions within the exponential family

Marius Hobbhahn
Department of Computer Science
University of Tübingen
Tübingen, Germany
marius.hobbhahn@gmail.com

1 Introduction

normal distributions are the best, lets transform everything to normal

2 General Information

2.1 Change of Variable for Probability Density Function

This is also referred to as change of basis. Let X be a continuous random variable with PDF $f_X(x)$ over $c_1 < x < c_2$. And let $Y = g(x)$ be a monotonic differentiable function with Inverse $X = g^{-1}(Y)$. Then the PDF of Y is

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{dg^{-1}(y)}{dy} \right| = f_X(g^{-1}(y)) \cdot \left| \frac{dg(y)}{dy} \right|^{-1}$$

Proof: TODO

2.2 Laplace Approximation

The Laplace approximation (LPA) is a tool to fit a normal distribution to the PDF of a given other distribution. The only constraints for the other distribution are: one peak (mode/ point of maximum) and twice differentiable. Laplace proposed a simple 2-term Taylor expansion on the log pdf. If $\hat{\theta}$ denotes the mode of a pdf $h(\theta)$, then it is also the mode of the log-pdf $q(\theta) = \log h(\theta)$. The 2-term Taylor expansion of $q(\theta)$ therefore is:

$$\begin{aligned}
q(\theta) &\approx q(\hat{\theta}) + q'(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})q''(\hat{\theta})(\theta - \hat{\theta}) && (1) \\
&= q(\hat{\theta}) + 0 + \frac{1}{2}(\theta - \hat{\theta})q''(\hat{\theta})(\theta - \hat{\theta}) \qquad [\text{since } q'(\theta) = 0] && (2) \\
&= c - \frac{(\theta - \mu)^2}{2\sigma^2} && (3)
\end{aligned}$$

where c is a constant, $\mu = \hat{\theta}$ and $\sigma^2 = \{-q''(\hat{\theta})\}^{-1}$. The right hand side of the last line matches the log-pdf of a normal distribution $N(\mu, \sigma^2)$. Therefore the pdf $h(\theta)$ is approximated by the pdf of the normal distribution $N(\mu, \sigma^2)$ where $\mu = \hat{\theta}$ and $\sigma^2 = \{-q''(\hat{\theta})\}^{-1}$. Note, that even though this derivation is done for the one dimensional case only, it is also true for the multidimensional case. The second derivative just becomes the Hessian of the pdf at the mode.

IS THERE SUPPOSED TO BE A MINUS IN FRONT OF THE q" ?

3 Dirichlet Distribution

see David McKay

TODO: implement LPA

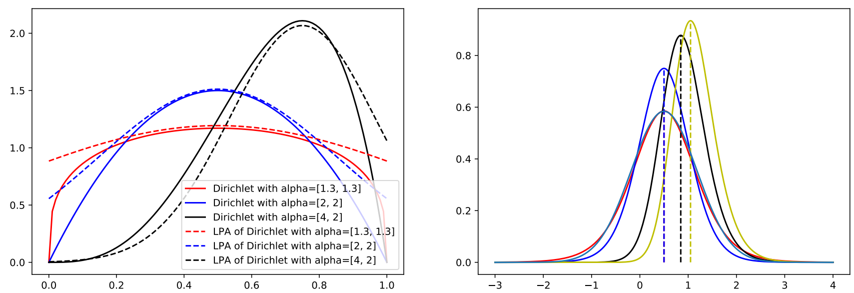

Figure 1: Dirichlet comparison

4 Gamma Distribution

4.1 Standard Gamma Distribution

4.1.1 PDF of the Gamma Distribution

$$f(x, \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot x^{(\alpha-1)} \cdot e^{(-\lambda x)} \tag{4}$$

where $\Gamma(\alpha)$ is the Gamma function. (DO I NEED TO EXPLAIN WHAT THAT IS?)

4.1.2 Laplace Approximation of the Gamma Distribution

To get the LPA of the Gamma function in the standard basis we need its mode and the second derivative of the log-pdf. The mode is already known to be $\hat{\theta} = \frac{\alpha-1}{\lambda}$. For the second derivative of the log-pdf we take the log-pdf and derive it twice and insert the mode for x:

$$\begin{aligned}
\text{log-pdf: } & \log\left(\frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot x^{(\alpha-1)} \cdot e^{(-\lambda x)}\right) \\
& = \alpha \cdot \log(\lambda) - \log(\Gamma(\alpha)) + (\alpha - 1)\log(x) - \lambda x \\
\text{1st derivative: } & \frac{(\alpha-1)}{x} - \lambda \\
\text{2nd derivative: } & \frac{(\alpha-1)}{x^2} \\
\text{insert mode: } & \frac{(\alpha-1)}{(\frac{\alpha-1}{\lambda})^2} = \frac{\lambda^2}{\alpha-1} \\
\text{invert: } & \sigma^2 = \frac{\alpha-1}{\lambda^2}
\end{aligned}$$

The LPA of the Gamma distribution is therefore approximately distributed according to the pdf of $N(\frac{\alpha-1}{\lambda}, \frac{\lambda^2}{\alpha-1})$.

4.2 Log-Transform of the Gamma Distribution

4.2.1 Log-Transformation

We transform the Gamma Distribution with the Log-Transformation, i.e. $Y = \log(X)$. The pdf of the transformed Gamma Distribution is therefore distributed according to

$$f_Y(y) = f_X(\exp(y)) \cdot \left| \frac{d\exp(y)}{dy} \right|$$

as reasoned in Subsection 2.1.

4.2.2 PDF of the log-transformed Gamma Distribution

The new pdf therefore becomes:

$$\begin{aligned} f_t(x, \alpha, \lambda) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \exp(x)^{(\alpha-1)} \cdot e^{(-\lambda\exp(x))} \cdot \exp(x) && (5) \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \exp(x)^{\alpha} \cdot e^{(-\lambda\exp(x))} && (6) \end{aligned}$$

4.2.3 Laplace Approximation of the log-transformed Gamma Distribution

To get the LPA of the Gamma distribution in the (IS IT NOW EXP OR LOG BASIS?) we need to calculate its mode and the second derivative of the log-pdf. To get the mode we take the first derivative and set it to zero.

$$\begin{aligned} \text{log-pdf: } & \log\left(\frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \exp(x)^\alpha \cdot exp(-\lambda\exp(x))\right) \\ & = \alpha\log(\lambda) - \log(\Gamma(\alpha)) + \alpha x - \lambda\exp(x) \\ \text{1st derivative: } & \alpha - \lambda\exp(x) \\ \text{mode: } & \alpha - \lambda\exp(x) = 0 \Leftrightarrow x = \log\left(\frac{\alpha}{\lambda}\right) \\ \text{2nd derivative: } & -\lambda\exp(x) \\ \text{insert mode: } & -\lambda\exp(\log\left(\frac{\alpha}{\lambda}\right)) = -\frac{1}{\alpha} \\ \text{invert and times -1: } & \sigma^2 = \alpha \end{aligned}$$

Therefore the LPA now is $N(\log\left(\frac{\alpha}{\lambda}\right), \alpha)$.

4.3 Comparison

plots and some measure to show similarity of two distributions

5 Beta Distribution

5.1 Standard Beta Distribution

5.1.1 Pdf of the Beta Distribution

$$\begin{aligned} f(x, \alpha, \beta) &= \frac{x^{(\alpha-1)} \cdot (1-x)^{(\beta-1)}}{B(\alpha, \beta)} && (7) \\ \text{with } B(\alpha, \beta) &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} && (8) \end{aligned}$$

where $\Gamma(x)$ is the Gamma function.

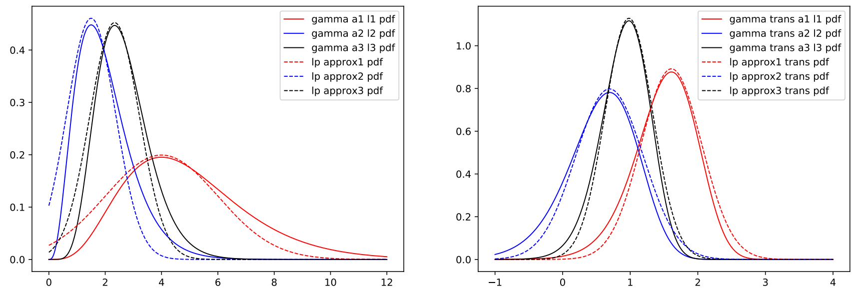

Figure 2: gamma comparison

5.1.2 Laplace approximation of the Beta distribution

TODO: explain stuff

$$
\begin{aligned}
\text{log-pdf: } & \log\left(\frac{x^{(\alpha-1)}\cdot(1-x)^{(\beta-1)}}{B(\alpha,\beta)}\right) \\
& = (\alpha-1)\log(x) + (\beta-1)\log(1-x) - \log(B(\alpha,\beta))) \\
\text{1st derivative: } & \frac{(\alpha-1)}{x} - \frac{(\beta-1)}{1-x} \\
\text{mode: } & \frac{(\alpha-1)}{x} - \frac{(\beta-1)}{1-x} = 0 \Leftrightarrow x = \frac{\alpha-1}{\alpha+\beta-2} \\
\text{2nd derivative: } & \frac{\alpha-1}{x^2} + \frac{\beta-1}{(1-x)^2} \\
\text{insert mode: } & \frac{\alpha-1}{\frac{\alpha-1}{\alpha+\beta-2}^2} + \frac{\beta-1}{(1-\frac{\alpha-1}{\alpha+\beta-2})^2} = \frac{(\alpha+\beta-2)^3}{(\alpha-1)(\beta-1)} \\
\text{invert: } & \frac{(\alpha-1)(\beta-1)}{(\alpha+\beta-2)^3}
\end{aligned}
$$

TODO: full derivations of mode and second derivative in appendix

The Beta distribution in standard basis is therefore approximated by $N(\mu = \frac{\alpha-1}{\alpha+\beta-2}, \sigma^2 = \frac{(\alpha-1)(\beta-1)}{(\alpha+\beta-2)^3})$.

5.2 Logit-Transform of the Beta distribution

5.2.1 Logit-Transform

using $g(x) = \log(\frac{x}{1-x})$. Therefore $g^{-1}(x) = \sigma(x) = \frac{1}{1+\exp(-x)}$.

5.2.2 Pdf after logit transform

$$
\begin{aligned}
f_t(x,\alpha,\beta) &= \frac{\sigma(x)^{(\alpha-1)}\cdot(1-\sigma(x)^{(\beta-1)})}{B(\alpha,\beta)}\cdot(\sigma(x)(1-\sigma(x)) \\
&= \frac{\sigma(x)^{\alpha}\cdot(1-\sigma(x)^{\beta})}{B(\alpha,\beta)}
\end{aligned}
$$

5.2.3 LPA of the logit distribution after transform

mode and variance of log pdf blablabla

$$
\begin{aligned}
\text{log-pdf: } & \log\left(\frac{\sigma(x)^{\alpha}\cdot(1-\sigma(x)^{\beta})}{B(\alpha,\beta)}\right) \\
& = \alpha\log(\sigma(x)) + \beta\log(1-\sigma(x)) - \log(B(\alpha,\beta)) \\
\text{1st derivative: } & \alpha(1-\sigma(x)) - \beta\sigma(x) \\
\text{mode: } & \alpha(1-\sigma(x)) - \beta\sigma(x) = 0 \Leftrightarrow x = -\log(\frac{\beta}{\alpha}) \\
\text{2nd derivative: } & (\alpha+\beta)\sigma(x)(1-\sigma(x)) \\
\text{insert mode: } & (\alpha+\beta)\sigma(-\log(\frac{\beta}{\alpha}))(1-\sigma(-\log(\frac{\beta}{\alpha}))) = \frac{\alpha\beta}{\alpha+\beta} \\
\text{invert: } & \frac{\alpha+\beta}{\alpha\beta}
\end{aligned}
$$

TODO: full derivations of mode and second derivative in appendix

The LPA is therefore $N(\mu = -\log(\frac{\beta}{\alpha}), \sigma^2 = \frac{\alpha+\beta}{\alpha\beta})$.

5.3 Comparison

TODO: KL-Divergence or something similar

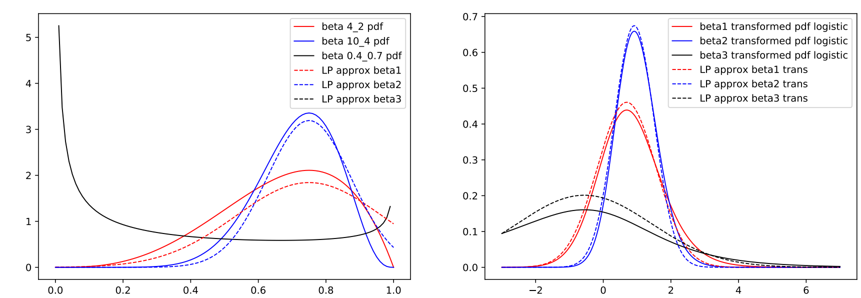

Figure 3: beta comparison

6 Chi-squared Distribution

TODO: write up all

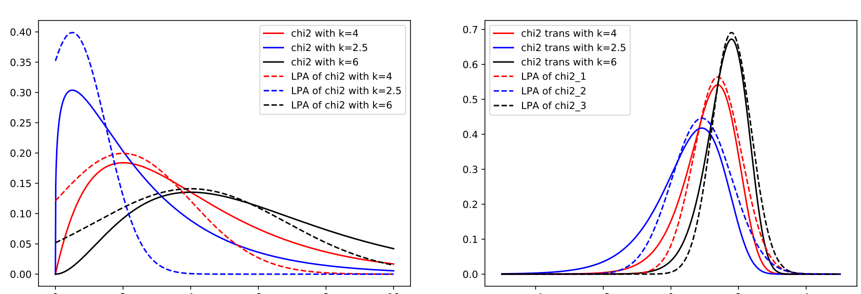

Figure 4: chi2 comparison

7 Wishart Distribution

TODO: get the LPA of standard base and transformed base

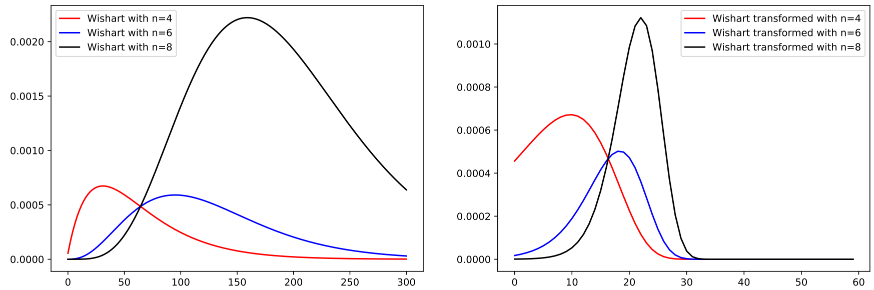

Figure 5: wishart comparison

8 Inverse Wishart Distribution

TODO: see above

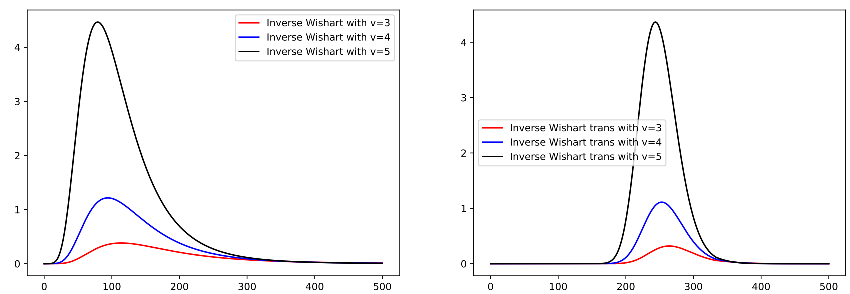

Figure 6: inverse wishart comparison