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 Background

2.1 Change of Variable for Probability Density Function

1D

Let X have a continuous density f_X. Let $g : \mathbb{R} \to \mathbb{R}$ be piece-wise strictly monotone and continuously differentiable, i.e. there exists intervals $I_1, I_2, ..., I_n$ which partition $\mathbb{R}$ such that g is strictly monotone and continuously differentiable on the interior of each I_i. For each $i, g : I_i \to \mathbb{R}$ is invertible on $g(I_i)$ and let g_i^{-1} be the inverse function. Let $Y = g(X)$ and $\wedge = \{y|y = g(x), x \in \mathbb{R}\}$ be the range g. Then the density function f_Y of Y exists and is given by

$$f_Y(y) = \sum_{i=1}^{n} f_X(g_i^{-1}(y)) \left| \frac{\partial g_i^{-1}(y)}{\partial y} \right| \mathbf{1}_\wedge(y) \tag{1}$$

Higher dim

TODO: Jacobian. But how do the intervals work?

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:

$$\begin{aligned} q(\theta) &\approx q(\hat{\theta}) + q'(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})q''(\hat{\theta})(\theta - \hat{\theta}) && (2) \\ &= q(\hat{\theta}) + 0 + \frac{1}{2}(\theta - \hat{\theta})q''(\hat{\theta})(\theta - \hat{\theta}) \qquad [\text{since } q'(\theta) = 0] && (3) \\ &= c - \frac{(\theta - \mu)^2}{2\sigma^2} && (4) \end{aligned}$$

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.

2.3 Exponential Family

exponential family form

$$f(x|\theta) = h(x)\exp[\eta(\theta)t(x) - A(\theta)] \tag{5}$$

for sufficient statistics $t : X \to \mathbb{R}$, natural parameters $\eta : \Theta \to \mathbb{R}$, and functions $A : \Theta \to \mathbb{R}$ and $h : X \to \mathbb{R}_+$

2.4 Chi2 <-> Normal

It is already well-known that the Chi-squared distribution describes the sum of independent, standard normal random variables. To introduce a certain 'trick' we show the forward and backward transformation between chi2 and normal.
Let X be normal with $\mu = 0, \sigma^2 = 1$. Let $Y = X^2$ and therefore $g(x) = x^2$, which is neither monotone nor injective. Take $I_1 = (-\infty, 0)$ and $I_2 = [0, +\infty)$. Then g is monotone and injective on I_1 and I_2 and $I_1 \cup I_2 = \mathbb{R}$. $g(I_1) = (0, \infty)$ and $g(I_2) = [0, \infty)$. Then $g_1^{-1} : [0, \infty) \to \mathbb{R}$ by $g_1^{-1}(y) = -\sqrt{y}$ and $g_2^{-1} : [0, \infty) \to \mathbb{R}$ by $g_2^{-1}(y) = \sqrt{y}$. Then

$$\left|\frac{\partial g_i^{-1}(y)}{\partial y}\right| = \left|\frac{1}{2\sqrt{y}}\right| = \frac{1}{2\sqrt{y}}$$

Applying Equation 1 we can transform a normal distribution to a chi-squared.

$$\begin{aligned} f_Y(y) &= f_X(g_1^{-1}(y)) \left|\frac{\partial g_1^{-1}(y)}{\partial y}\right| \mathbf{1}_{\wedge}(y) + f_X(g_2^{-1}(y)) \left|\frac{\partial g_2^{-1}(y)}{\partial y}\right| \mathbf{1}_{\wedge}(y) \\ &= \frac{1}{\sqrt{2\pi}} \exp(-\frac{y}{2}) \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} \exp(-\frac{y}{2}) \frac{1}{2\sqrt{y}} \qquad (y > 0) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} \exp(-\frac{y}{2}) \end{aligned} \tag{6}$$

The 'trick' was to split up the variable transformation in two parts to adjust for the fact that the space of the chi-squared and the Normal are different. We can reverse the same procedure to get from a chi-squared to a normal distribution. We keep the variable names from before. Let $X = \sqrt{Y}$ and therefore $h(x) = \sqrt{x}$. Then $h_1^{-1} : \mathbb{R} \to (-\infty, 0)$ by $h_1^{-1}(x) = -x^2$ and $h_2^{-1} : \mathbb{R} \to [0, \infty)$ by $h_2^{-1}(x) = x^2$. Then

$$\left|\frac{\partial h_i^{-1}(y)}{\partial y}\right| = |2x|$$

and

$$\begin{aligned} f_X(x) &= f_y(h_1^{-1}(x)) \left|\frac{\partial h_1^{-1}(y)}{\partial y}\right| \mathbf{1}_{\wedge}(y) + f_y(h_2^{-1}(x)) \left|\frac{\partial h_2^{-1}(y)}{\partial y}\right| \mathbf{1}_{\wedge}(y) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{x^2}} \exp(-\frac{x^2}{2})|2x|\mathbf{1}_{(-\infty,0)}(x) + \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{x^2}} \exp(-\frac{x^2}{2})|2x|\mathbf{1}_{[0,\infty)}(x) \\ &= \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) \end{aligned} \tag{7}$$

which is defined on the entirety of $\mathbb{R}$.

3 Generalized Laplace Bridge

Define the GLB clearly

- h(x) must be 1
- the distribution must have 2 sufficient statistics
- the sufficient statistics must include an x or x^2 when the original variable is used.

4 General overview

Table 1: Overview

Distribution	$g(x)$	$g^{-1}(x)$	T	T'
Exponential	log	exp	x	$(x, \exp(x))$
Gamma	log	exp	$(\log(x), x)$	$(x, \exp(x))$
Gamma	log	exp	$(\log(x), x)$	$(\log(x), x^2)$
Inverse Gamma	log	exp	$(\log(x), x)$	$(x, \exp(x))$
Inverse Gamma	log	exp	$(\log(x), x)$	$(\log(x), x^2)$
Chi2	log	exp	$(\log(x), x)$	$(x, \exp(x))$
Chi2	log	exp	$(\log(x), x)$	$(\log(x), x^2)$
Beta	logit	logistic	$(\log(x), \log(1-x))$	?
Dirichlet	-	softmax	?	?
Wishart	mlog	mexp	$(\log(X), X)$	$(X, \exp(X))$
Wishart	msqrt	msqr	$(\log(X), X)$	$(\log(X), X^2)$
Inverse Wishart	mlog	mexp	?	?

NOTE:MOST DISTRIBUTIONS ACTUALLY HAVE TWO VALID TRANSFORMS. ONE FOR X AND ONE FOR X^2 AS SUFFICIENT STATISTICS

5 Exponential Distribution

5.1 Standard Exponential Distribution

The PDF of the exponential distribution is

$$p(x|\lambda) = \lambda \exp(-\lambda x) \tag{8}$$

which can be written as

$$p(x|\lambda) = \exp\left[-\lambda x + \log \lambda\right] \tag{9}$$

with $T = x$, $\eta = -\lambda$ and $A(\lambda) = -\log \lambda$

5.1.1 Laplace Approximation of the Exponential Distribution

$$\begin{aligned} \text{log-pdf: } & (\log \lambda - \lambda x) \\ \text{1st derivative: } & -\lambda \\ \text{2nd derivative: } & 0 \end{aligned}$$

The Laplace Approximation is not defined since the second derivative is not positive.

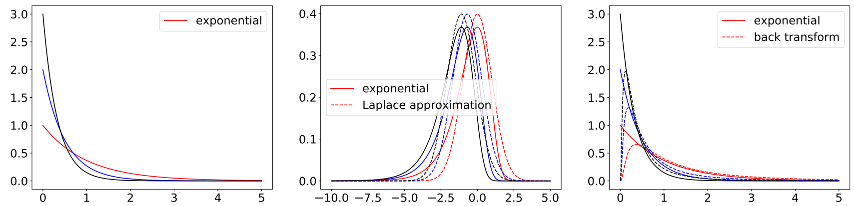

Figure 1: exponential comparison

5.2 Log-transformed Exponential Distribution

We choose $g(x) = \log(x)$, and thereby $g^{-1}(x) = \exp(x)$. It follows that the new pdf is

$$\begin{aligned} f(x|\lambda) &= \lambda \exp(-\lambda \exp(x)) \cdot \exp(x) \\ &= \lambda \exp(-\lambda \exp(x) + x) \end{aligned} \tag{10}$$

which can be written as

$$p(x|\lambda) = \exp\left[-\lambda \exp(x) + x + \log \lambda\right] \tag{11}$$

with $T = x, \exp(x)$, $\eta = 1, -\lambda$ and $A(\lambda) = \log \lambda$

5.2.1 Laplace Approximation of the log-transformed Exponential Distribution

$$\begin{aligned} \text{log-pdf: } & -\lambda \exp(x) + x + \log \lambda \\ \text{1st derivative: } & \lambda - \exp(x) + 1 \\ \text{mode: } & x = \log(1/\lambda) \\ \text{2nd derivative: } & \lambda - \exp(x) \\ \text{insert mode: } & -\lambda \exp(1/\lambda) = -1 \\ \text{invert: } & \sigma^2 = 1 \end{aligned}$$

Therefore the Laplace approximation in the transformed basis is given by $\mathcal{N}(x, \log(1/\lambda), 1)$.

5.3 The Bridge

We have already found μ and σ. The inverse transformation is easily found through the mode $x = \log(1/\lambda) \Leftrightarrow \lambda = 1/\exp(x)$. In summary:

$$\mu = \log(1/\lambda) \tag{12}$$

$$\sigma = 1 \tag{13}$$

$$\lambda = 1/\exp(x) \tag{14}$$

6 Gamma Distribution

6.1 Standard Gamma Distribution

$$f(x, \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot x^{(\alpha-1)} \cdot e^{(-\lambda x)} \tag{15}$$

where $\Gamma(\alpha)$ is the Gamma function. This can be written as

$$f(x,\alpha,\lambda) = \exp\left[(\alpha-1)\log(x) - \lambda x + \alpha\log(\lambda) - \log(\Gamma(\alpha))\right] \tag{16}$$
$$= \frac{1}{x}\exp\left[\alpha\log(x) - \lambda x + \alpha\log(\lambda) - \log(\Gamma(\alpha))\right] \tag{17}$$

with $h(x) = \frac{1}{x}, T = (\log x, x), \eta = (\alpha, -\lambda)$ and $A(\alpha,\lambda) = \log(\Gamma(\alpha)) - \alpha\log(\lambda)$.

6.1.1 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:

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

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

6.2 Sqrt-Transform of the Gamma Distribution

6.2.1 Sqrt-Transformation

We transform the Gamma Distribution with the sqrt-transformation, i.e. $Y = \sqrt{X}, g(x) = \sqrt{x}, g_1^{-1}(y) = -y^2, g_2^{-1}(y) = y^2$ and $\left|\frac{\partial g_i^{-1}(y)}{\partial y}\right| = |2x|$. We use the same 'trick' as in Subsection 2.4 to split up the transformation in two parts.

$$f_t(y,\alpha,\lambda) = \frac{1}{x}\exp\left[\alpha\log(x) - \lambda x + \alpha\log(\lambda) - \log(\Gamma(\alpha))\right] \tag{18}$$

$$\begin{aligned}
f_Y(y) &= f_X(g_1^{-1}(y))\left|\frac{\partial g_1^{-1}(y)}{\partial y}\right|\mathbf{1}_\wedge(y) + f_X(g_2^{-1}(y))\left|\frac{\partial g_2^{-1}(y)}{\partial y}\right|\mathbf{1}_\wedge(y) \\
&= \frac{1}{2y}\exp[\alpha\log(y) - \lambda y - A(\alpha,\lambda)]|2y|\mathbf{1}_{(-\infty,0)}(y) + \frac{1}{2y}\exp[\alpha\log(y) - \lambda y - A(\alpha,\lambda)]|2y|\mathbf{1}_{[0,\infty)}(y) \\
&= \exp[\alpha\log(y) - \lambda y - A(\alpha,\lambda)]\mathbf{1}_{(-\infty,+\infty)}(y) \qquad (19) \\
&= \exp[\alpha\log(y) - \lambda y - A(\alpha,\lambda)]
\end{aligned}$$

which is defined on the entirety of $\mathbb{R}$ and is an exponential family with $h(y) = 1, T = (\log y, y), \eta = (\alpha, -\lambda)$ and $A(\alpha,\lambda) = \log(\Gamma(\alpha)) - \alpha\log(\lambda)$.

6.2.2 Laplace Approximation of the sqrt-transformed Gamma Distribution

To get the LPA of the Gamma distribution in the transformed 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.

$$\begin{aligned}
\text{log-pdf: } & 2\alpha\log(x) - \lambda 2x + \alpha\log(\lambda) - \log(\Gamma(\alpha)) \\
\text{1st derivative: } & \frac{2\alpha}{x} - 2\lambda x \\
\text{mode: } & \frac{2\alpha}{x} - 2\lambda x = 0 \Leftrightarrow x = \sqrt{\frac{\alpha}{\lambda}} \\
\text{2nd derivative: } & -\frac{2\alpha}{x^2} - 2\lambda \\
\text{insert mode: } & -\frac{2\alpha}{\frac{\alpha}{\lambda}} - 2\lambda = -4\lambda \\
\text{invert and times -1: } & \frac{1}{4\lambda}
\end{aligned}$$

Therefore the LPA now is $\mathcal{N}\left(\sqrt{\frac{\alpha}{\lambda}}, \frac{1}{4\lambda}\right)$.

6.2.3 The bridge for the sqrt-transformation

We already know how to get μ and σ from λ and α. To invert we calculate $\mu = \sqrt{\frac{\alpha}{\lambda}} \Leftrightarrow \alpha = \frac{\mu^2}{\lambda}$ and insert $\lambda = \frac{4}{\sigma^2}$. In summary we have

$$\mu = \sqrt{\frac{\alpha}{\lambda}} \tag{20}$$

$$\sigma^2 = \frac{1}{4\lambda} \tag{21}$$

$$\lambda = \frac{4}{\sigma^2} \tag{22}$$

$$\alpha = \frac{(\sigma\mu)^2}{4} \tag{23}$$

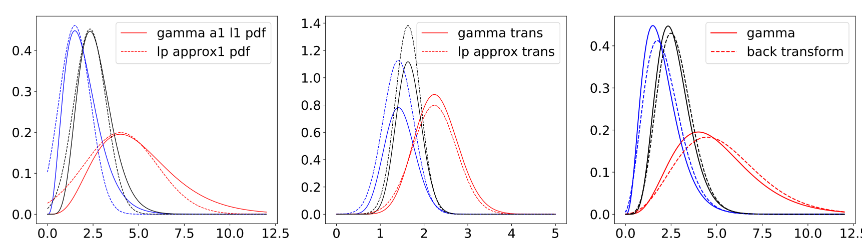

Figure 2: gamma comparison square

6.3 Log-Transform of the Gamma Distribution

6.3.1 Log-Transformation

We transform the Gamma Distribution with the Log-Transformation, i.e. $Y = \log(X), g(x) = \log(x), g^{-1}(x) = \exp(x)$. The transformed pdf is

$$
\begin{aligned}
f_t(x,\alpha,\lambda) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \exp(x)^{(\alpha-1)} \cdot e^{(-\lambda\exp(x))} \cdot \exp(x) \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \exp(x)^{\alpha} \cdot e^{(-\lambda\exp(x))}
\end{aligned} \tag{24}
$$

which can be rewritten as

$$
f_t(x,\alpha,\lambda) = \exp\left[\alpha x - \lambda\exp(x)\alpha\log(\lambda) - \log(\Gamma(\alpha))\right] \tag{25}
$$

with $T = (x, \exp(x)), \eta = (\alpha, -\lambda)$ and $A(\alpha,\lambda) = \log(\Gamma(\alpha)) - \alpha\log(\lambda)$.

6.3.2 Laplace Approximation of the log-transformed Gamma Distribution

To get the LPA of the Gamma distribution in the transformed 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 $N(\log\left(\frac{\alpha}{\lambda}\right), \alpha)$.

6.3.3 The bridge for the log-transformation

We already know how to get μ and σ from λ and α. To invert we calculate $\mu = \log(\alpha/\lambda) \Leftrightarrow \lambda = \alpha/\exp(\mu)$ and insert $\alpha = \sigma^2$. In summary we have

$$
\mu = \log(\alpha/\lambda) \tag{26}
$$

$$
\sigma^2 = \alpha \tag{27}
$$

$$
\lambda = \alpha/\exp(\mu) \tag{28}
$$

$$
\alpha = \sigma^2 \tag{29}
$$

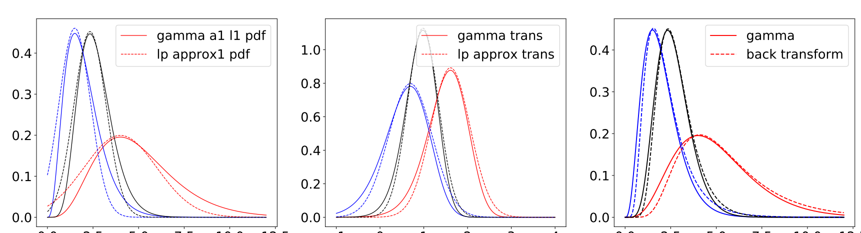

Figure 3: gamma comparison

7 Inverse Gamma Distribution

7.1 Standard Inverse Gamma Distribution

The pdf of the inverse gamma is

$$f(x, \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp(-\frac{\lambda}{\alpha}) \tag{30}$$

where Γ is the Gamma function. This can be rewritten as

$$f(x, \alpha, \lambda) = \exp\left[(-\alpha - 1)\log(x) - \lambda/x + \alpha \log(\lambda) - \log\Gamma(\alpha)\right] \tag{31}$$

where $T = (\log(x), x), \eta = (-\alpha - 1, -\lambda)$ and $A(\alpha, \lambda) = \log\Gamma(\alpha) - \alpha \log \lambda$.

7.1.1 Laplace Approximation of the standard inverse gamma distribution

$$\begin{aligned}
\text{log-pdf: } & (-\alpha - 1)\log(x) - \lambda/x + \alpha \log(\lambda) - \log\Gamma(\alpha) \\
\text{1st derivative: } & \frac{-\alpha - 1}{x} + \frac{\lambda}{x^2} \\
\text{mode: } & \frac{-\alpha - 1}{x} + \frac{\lambda}{x^2} = 0 \Leftrightarrow x = \frac{\lambda}{a+1} \\
\text{2nd derivative: } & \frac{\alpha + 1}{x^2} - 2\frac{\lambda}{x^3} \\
\text{insert mode: } & \frac{\alpha + 1}{\frac{\lambda}{a+1}^2} - 2\frac{\lambda}{\frac{\lambda}{a+1}^3} = -\frac{(\alpha+1)^3}{\lambda^2} \\
\text{invert and times -1: } & \sigma^2 = \frac{\lambda^2}{(\alpha+1)^3}
\end{aligned}$$

7.2 Log-Transform of the inverse gamma distribution

We choose $g(x) = \log(x)$, and thereby $g^{-1}(x) = \exp(x)$. It follows that the new pdf is

$$f_t(x, \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \exp(x)^{-\alpha} \exp(-\lambda/\exp(x)) \tag{32}$$

which can be written as

$$f_t(x, \alpha, \lambda) = \exp\left[-\alpha x - \frac{\lambda}{\exp(x)} + \alpha \log \lambda - \log \Lambda(\alpha)\right] \tag{33}$$

with $T = (x, 1/\exp(x)), \eta(-\alpha, \lambda)$ and $A(\alpha, \lambda) = \log\Gamma(\alpha) - \alpha \log \lambda$.

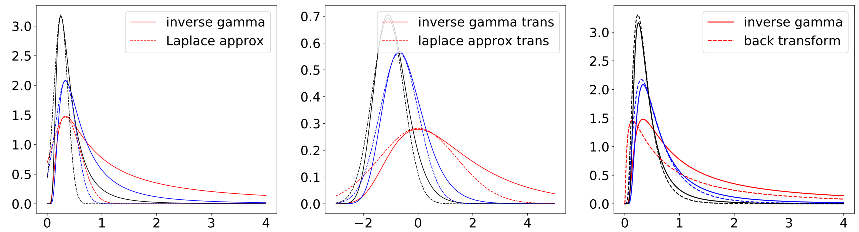

Figure 4: inverse gamma comparison

7.2.1 Laplace Approximation of the log-transformed Inverse Gamma Distribution

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

7.2.2 The Bridge for the log-transformed Inverse Gamma Distribution

$$\mu = \log(\lambda/\alpha) \tag{34}$$

$$\sigma^2 = 1/\alpha \tag{35}$$

$$\lambda = \exp(\mu)\sigma^2 \tag{36}$$

$$\alpha = 1/\sigma^2 \tag{37}$$

8 Chi-squared Distribution

8.1 Standard Chi-squared distribution

The pdf is

$$f(x,k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} \exp(-x/2) \tag{38}$$

which can be written as

$$f(x,k) = \exp\left[(k/2-1)\log(x) - x/2 - \log(2^{k/2}\Gamma(k/2))\right] \tag{39}$$

with $T = (\log(x), x), \eta = (k/2-1)$ and $A(k) = \log(2^{k/2}\Gamma(k/2))$.

8.1.1 Laplace approximation of the standard Chi-squared distribution

$$
\begin{aligned}
\text{log-pdf: } & (k/2-1)\log(x) - x/2 - \log(2^{k/2}\Gamma(k/2)) \\
\text{1st derivative: } & \frac{k/2-1}{x} - \frac{1}{2} \\
\text{mode: } & \frac{k/2-1}{x} - \frac{1}{2} = 0 \Leftrightarrow x = k-2 \\
\text{2nd derivative: } & -\frac{k/2-1}{x^2} \\
\text{insert mode: } & -\frac{k/2-1}{(k-2)^2} \\
\text{invert and times -1: } & \sigma^2 = \frac{(k-2)^2}{k/2-1}
\end{aligned}
$$

8.2 Log-Transformed Chi-squared distribution

we transform the distribution with $g(x) = \log(x)$, i.e. $g^{-1}(x) = \exp(x)$. The new pdf becomes

$$
f_t(x,k) = \frac{1}{2^{k/2}\Gamma(k/2)} \exp(x)^{k/2-1} \exp(-\exp(x)/2) \tag{40}
$$

which can be written as

$$
f_t(x,k) = \exp\left[(k/2)x - \frac{\exp(x)}{2} - \log(2^{k/2}\Gamma(k/2))\right] \tag{41}
$$

meaning $T = (x, \exp(x)), \eta = (k/2)$ and $A(k) = \log(2^{k/2}\Gamma(k/2))$.

8.2.1 Laplace approximation of the log-transformed Chi-squared distribution

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

8.2.2 The Bridge for log-transform

$$
\mu = \log(k) \tag{42}
$$

$$
\sigma^2 = 2/k \tag{43}
$$

$$
k = \exp(\mu) \tag{44}
$$

8.3 Sqrt-Transformed Chi-squared distribution

we transform the distribution with $g(x) = \sqrt{x}$, i.e. $g^{-1}(x) = x^2$. The new pdf becomes

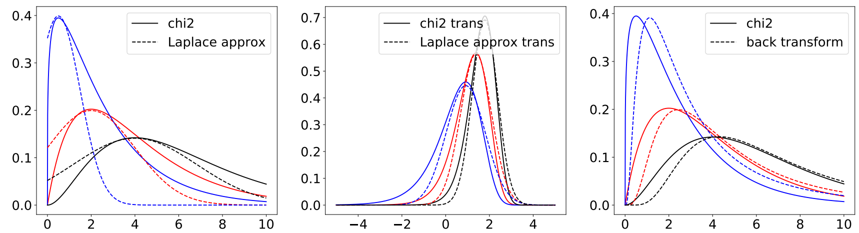

Figure 5: chi2 comparison log transform

$$f_t(x,k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{2(k/2-1)} \exp(-x^2/2)2x \tag{45}$$
$$= \frac{1}{2^{k/2}\Gamma(k/2)} x^{k} \exp(-x^2/2)$$

which can be written as

$$f_t(x,k) = \exp\left[(k\log(x) - \frac{x^2}{2} - \log(2^{k/2}\Gamma(k/2))\right] \tag{46}$$

meaning $T = (\log(x), x^2), \eta = (k, 1/2)$ and $A(k) = \log(2^{k/2}\Gamma(k/2))$.

8.3.1 Laplace approximation of the sqrt-transformed Chi-squared distribution

$$\begin{aligned}
\text{log-pdf: } & (k\log(x) - \frac{x^2}{2} - \log(2^{k/2}\Gamma(k/2)) \\
\text{1st derivative: } & \frac{k}{x} - x \\
\text{mode: } & \frac{k}{x} - x = 0 \Leftrightarrow x = \sqrt{k} \\
\text{2nd derivative: } & -\frac{k}{x^2} - 1 \\
\text{insert mode: } & -\frac{k}{k} - 1 \\
\text{invert and times -1: } & \sigma^2 = 1/2
\end{aligned}$$

8.3.2 The Bridge for sqrt-transform

$$\mu = \sqrt{k} \tag{47}$$
$$\sigma^2 = 1/2 \tag{48}$$
$$k = \mu^2 \tag{49}$$

TODO:THE BRIDGE BACK LOOKS A BIT WEIRD

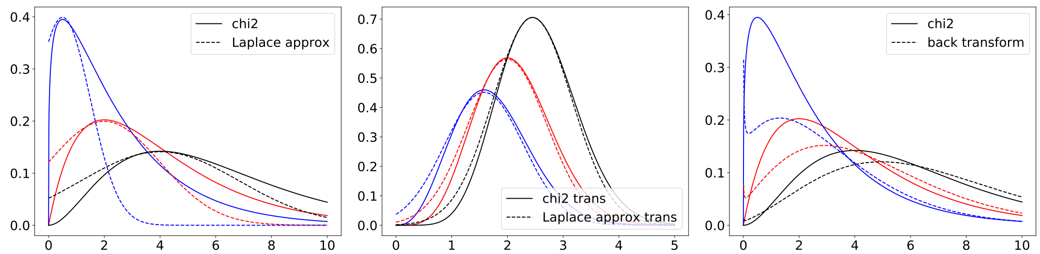

Figure 6: chi2 sqrt comparison

9 Beta Distribution

9.1 Standard Beta Distribution

The pdf of the Beta distribution in the standard basis is

$$f(x, \alpha, \beta) = \frac{x^{(\alpha-1)} \cdot (1-x)^{(\beta-1)}}{B(\alpha, \beta)} \tag{50}$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and $\Gamma(x)$ is the Gamma function. This can be written as

$$f(x, \alpha, \beta) = \exp\left[(\alpha-1)\log(x) + (\beta-1)\log(1-x) - \log(B(\alpha, \beta))\right] \tag{51}$$

$$= \frac{1}{x(1-x)} \exp\left[\alpha \log(x) + \beta \log(1-x) - \log(B(\alpha, \beta))\right] \tag{52}$$

With $h(x) = \frac{1}{x(1-x)}, T = (\log(x), \log(1-x), \eta = (\alpha, \beta)$ and $A(\alpha, \beta) = \log(B(\alpha, \beta)))$.

9.1.1 Laplace approximation of the standard Beta distribution

To get the Laplace approximation we need the mode and Hessian. To get the mode we use the first derivative of the log-pdf and set it to zero. To get the Covariance we use the Hessian at the mode, multiply it with -1 and invert it.

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

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

9.2 Logit-Transform of the Beta distribution

We transform the Beta distribution using $g(x) = \log(\frac{x}{1-x})$. Therefore $g^{-1}(x) = \sigma(x) = \frac{1}{1+\exp(-x)}$. This yields the following pdf

$$f_t(x, \alpha, \beta) = \frac{1}{\sigma(x)(1-\sigma(x))} \exp\left[\alpha \log(\sigma(x)) + \beta \log(1-\sigma(x)) - \log(B(\alpha,\beta))\right] \cdot (\sigma(x)(1-\sigma(x)) \tag{53}$$
$$= \exp\left[\alpha \log(x) + \beta \log(1-x) - \log(B(\alpha,\beta))\right]$$

Which has $h(x) = 1, T = (\log(\sigma(x)), \log(1-\sigma(x)), \eta = (\alpha, \beta)$ and $A(\alpha, \beta) = \log(B(\alpha, \beta))$.

9.2.1 Laplace approximation of the logit transformed Beta distribution

mode and variance of log pdf blablabla

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

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

9.2.2 The Bridge for the logit transformation

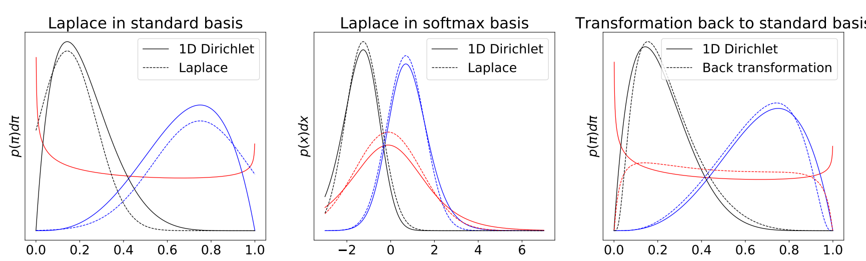

Figure 7: beta comparison

10 Dirichlet Distribution

TODO:REDO THE BRIDGE IN THE EXPONENTIAL FAMILY SETTING. I thought I had already done that but I did not.

11 Wishart Distribution

11.1 Standard Wishart distribution

the pdf of the Wishart is

$$f(x;n,p,V) = \frac{1}{2^{np/2}\left|\mathbf{V}\right|^{n/2}\Gamma_p\left(\frac{n}{2}\right)}|\mathbf{x}|^{(n-p-1)/2}e^{-(1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{x})} \tag{54}$$

which can be written as

$$f(x;n,p,V) = \exp\left[(n-p-1)/2\log(|x|) - (1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{x}) - \log\left(2^{np/2}\left|\mathbf{V}\right|^{n/2}\Gamma_p\left(\frac{n}{2}\right)\right)\right] \tag{55}$$

with $T = (\log(x), x), \eta = ((n-p-1)/2, V^{-1})$ and $A(n,p,V) = \log\left(2^{np/2}\left|\mathbf{V}\right|^{n/2}\Gamma_p\left(\frac{n}{2}\right)\right)$

11.1.1 Laplace Approximation of the standard Wishart distribution

Using $\frac{\partial \det(X)}{\partial X} = \det(X)(X^{-1})^\top$ and $\frac{\partial}{\partial X}Tr(AX^\top) = A$ we can calculate the mode by setting the first derivative of the log-pdf to zero

$$\begin{aligned}
\frac{\partial \log f(X;n,p,V)}{\partial X} &= \frac{(n-p-1)\det(X)(X^{-\top})}{2\det(X)} - \frac{V^{-1}}{2} \\
\Rightarrow 0 &= \frac{(n-p-1)X^{-1}}{2} - \frac{V^{-1}}{2} \\
\Leftrightarrow \frac{(n-p-1)X^{-1}}{2} &= \frac{V^{-1}}{2} \\
\Leftrightarrow X &= (n-p-1)V
\end{aligned}$$

Using the fact that $\frac{\partial A^{-1}}{\partial A} = X^{-1} \otimes X^{-1}$ where $\otimes$ is the Kronecker product we compute the second derivative as

$$\frac{\partial^2 \log f(X;n,p,V)}{\partial^2 X} = -\frac{(n-p-1)}{2}X^{-1}\otimes X^{-1}$$

Using $(\alpha A)^{-1} = \alpha^{-1}A^{-1}$, the linearity of the Kronecker product to pull out scalars and $X^{-1} \otimes X^{-1} = (X \otimes X)^{-1}$ to insert the mode and invert we get:

$$\begin{aligned}
-\frac{(n-p-1)}{2}X^{-1}\otimes X^{-1} &= -\frac{(n-p-1)}{2}\frac{1}{(n-p-1)}V^{-1} \otimes \frac{1}{(n-p-1)}V^{-1} \\
&= -\frac{1}{2(n-p-1)}(V\otimes V)^{-1} \\
\Rightarrow \Sigma &= 2(n-p-1)(V\otimes V)
\end{aligned}$$

In summary, the Laplace approximation of a Wishart distribution in the standard basis is $\mathcal{N}\left(X;(n-p-1)V, 2(n-p-1)(V\otimes V)\right)$, where the representation of the symmetric positive definite matrices has been changed from $\mathbb{R}^{n\times n}$ to $\mathbb{R}^{n^2}$.

11.2 Sqrtm-Transformed Wishart distribution

we transform the distribution with $g(X) = \operatorname{sqrtm}(X) = X^{\frac{1}{2}}$, i.e. $g^{-1}(X) = X^2$, where $\operatorname{sqrtm}(X)$ is the square root of the matrix. The new pdf becomes

$$
\begin{aligned}
f_t(x;n,p,V) &= \frac{1}{2^{np/2}\left|\mathbf{V}\right|^{n/2}\Gamma_p\left(\frac{n}{2}\right)}\left|\mathbf{x^2}\right|^{(n-p-1)/2}e^{-(1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{x^2})}\cdot|2x| \\
&= \frac{1}{2^{np/2}\left|\mathbf{V}\right|^{n/2}\Gamma_p\left(\frac{n}{2}\right)}\left|\mathbf{x}\right|^{2(n-p-1)/2}e^{-(1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{x^2})}\cdot 2^k|x| \\
&= \frac{2^k}{2^{np/2}\left|\mathbf{V}\right|^{n/2}\Gamma_p\left(\frac{n}{2}\right)}\left|\mathbf{x}\right|^{(n-p)}e^{-(1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{x^2})}
\end{aligned}
\tag{56}
$$

where k is the dimensionality of X. This can be rewritten as

$$
f_t(x;n,p,V) = \exp\left[(n-p)\log(|x|) - (1/2)\operatorname{tr}(\mathbf{V}^{-1}\mathbf{x^2}) - \log\left(2^{np/2}\left|\mathbf{V}\right|^{n/2}\Gamma_p\left(\frac{n}{2}\right)\right) + k\log(2)\right]
\tag{57}
$$

with $T = (\log(x), x^2), \eta = ((n-p), V^{-1})$ and $A(n,p,V) = \log\left(2^{np/2}\left|\mathbf{V}\right|^{n/2}\Gamma_p\left(\frac{n}{2}\right)\right) + k\log(2)$

11.2.1 Laplace Approximation of the sqrtm-transformed Wishart distribution

Using $\frac{\partial \det(X)}{\partial X} = \det(X)(X^{-1})^\top$ and $\frac{\partial}{\partial X}Tr(AXX) = 2AX$ we can calculate the mode by setting the first derivative of the log-pdf to zero

$$
\begin{aligned}
\frac{\partial \log f_t(X;n,p,V)}{\partial X} &= \frac{(n-p)\det(X)(X^{-\top})}{\det(X)} - \frac{2V^{-1}X}{2} \\
\Rightarrow 0 &= (n-p)X^{-1} - V^{-1}X \\
\Leftrightarrow (n-p)X^{-1} &= V^{-1}X \\
\Leftrightarrow X^2 &= (n-p)V \\
\Leftrightarrow X &= \sqrt{(n-1)}V^{\frac{1}{2}}
\end{aligned}
$$

Computing the second derivative entry-wise we get

$$
\begin{aligned}
\frac{\partial^2 \log f_t(X;n,p,V)_{kl}}{\partial^2 X_{ij}} &= \frac{\partial}{\partial X_{ij}}(n-p)X^{-1} - V^{-1}X \\
&= -(n-p)\underbrace{(X^{-1})_{ki}(X^{-1})_{jl}}_{I} - \underbrace{\delta_{il}(V^{-1})_{kj}}_{II}
\end{aligned}
$$

We realize that I is the Kronecker product $X^{-1}\otimes X^{-1}$ and name II $\tilde{V^{-1}}$.

Using $(\alpha A)^{-1} = \alpha^{-1}A^{-1}$, the linearity of the Kronecker product to pull out scalars and $X^{-1}\otimes X^{-1} = (X\otimes X)^{-1}$ to insert the mode and invert we get:

$$
\begin{aligned}
-(n-p)X^{-1}\otimes X^{-1} - V^{-1} &= -(n-p)\frac{1}{n-p}V^{-\frac{1}{2}}\frac{1}{n-p}V^{-\frac{1}{2}} - \tilde{V^{-1}} \\
&= -\frac{n-p}{n-p}(V\otimes V)^{-\frac{1}{2}} - \tilde{V^{-1}} \\
\Rightarrow \Sigma &= (V\otimes V)^{\frac{1}{2}} - \tilde{V^{-1}}
\end{aligned}
$$

TODO: SOMETHING IS WRONG HERE: The last inversion is a bit weird but it works exactly as shown here in practice.

In summary, the Laplace approximation of a Wishart distribution in the sqrtm-transformed basis is $\mathcal{N}\left(X; \sqrt{(n-p)}V^{\frac{1}{2}}, (V \otimes V)^{\frac{1}{2}} - V^{\tilde{-1}}\right)$, where the representation of the symmetric positive definite matrices has been changed from $\mathbb{R}^{n \times n}$ to $\mathbb{R}^{n^2}$.

11.2.2 The Bridge for sqrtm-tranform

we use $\mu = ((n-p)V)^{\frac{1}{2}} \Leftrightarrow \mu^2 = (n-p)V \Leftrightarrow V = \frac{\mu^2}{(n-p)}$. Remember that μ is reshaped to be the same size as V even though we usually think of it in vector-form.

$$\mu = ((n-p)V)^{\frac{1}{2}} \tag{58}$$

$$\Sigma = (V \otimes V)^{\frac{1}{2}} - V^{\tilde{-1}} \tag{59}$$

$$V = \frac{\mu^2}{(n-p)} \tag{60}$$

QUESTION: DO WE ASSUME WE KNOW THE n?

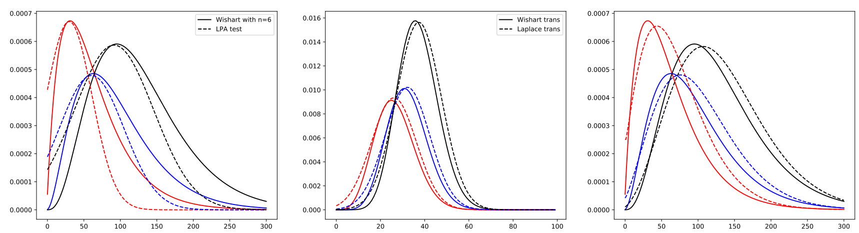

Figure 8: wishart comparison for sqrtm

12 Inverse Wishart Distribution

TODO: see above