
Changing the Basis of distributions within the exponential family

Marius Hobbhahn
Department of Computer Science
University of Tübingen
Tübingen, Germany
maris.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

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

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:

$$q(\theta) \approx q(\hat{\theta}) + q'(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})q''(\hat{\theta})(\theta - \hat{\theta}) \quad (1)$$

$$= q(\hat{\theta}) + 0 + \frac{1}{2}(\theta - \hat{\theta})q''(\hat{\theta})(\theta - \hat{\theta}) \quad [\text{since } q'(\hat{\theta}) = 0] \quad (2)$$

$$= c - \frac{(\theta - \mu)^2}{2\sigma^2} \quad (3)$$

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

$$f(x|\theta) = h(x) \exp[\eta(\theta)t(x) - A(\theta)] \quad (4)$$

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

3 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

4 Exponential Distribution

4.1 Standard Exponential Distribution

The PDF of the exponential distribution is

$$p(x|\lambda) = \lambda \exp(-\lambda x) \quad (5)$$

which can be written as

$$p(x|\lambda) = \exp[-\lambda x + \log \lambda] \quad (6)$$

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

4.1.1 Laplace Approximation of the Exponential Distribution

log-pdf: $\log(\log \lambda - \lambda x)$

1st derivative: $-\lambda$

2nd derivative: 0

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

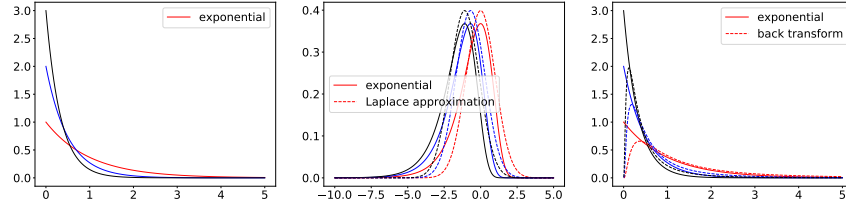


Figure 1: exponential comparison

4.2 Log-transformed Exponential Distribution

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

$$\begin{aligned} f(x|\lambda) &= \lambda \exp(-\lambda \exp(x)) \cdot \exp(x) \\ &= \lambda \exp(-\lambda \exp(x) + x) \end{aligned} \quad (7)$$

which can be written as

$$p(x|\lambda) = \exp[-\lambda \exp(x) + x + \log \lambda] \quad (8)$$

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

4.2.1 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 $\mathcal{N}(x, \log(1/\lambda), 1)$.

4.3 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:

$$\mu = \log(1/\lambda) \quad (9)$$

$$\sigma = 1 \quad (10)$$

$$\lambda = 1/\exp(x) \quad (11)$$

5 Gamma Distribution

5.1 Standard Gamma Distribution

$$f(x, \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot x^{(\alpha-1)} \cdot e^{(-\lambda x)} \quad (12)$$

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

$$f(x, \alpha, \lambda) = \exp[(\alpha - 1) \log(x) - \lambda x + \alpha \log(\lambda) - \log(\Gamma(\alpha))] \quad (13)$$

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

5.1.1 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 :

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

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

5.2 Sqrt-Transform of the Gamma Distribution

5.2.1 Sqrt-Transformation

We transform the Gamma Distribution with the Log-Transformation, i.e. $Y = \sqrt{(X)}$, $g(x) = \sqrt{(x)}$, $g^{-1}(x) = x^2$. The transformed pdf is

$$\begin{aligned}
f_t(x, \alpha, \lambda) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot x^{2(\alpha-1)} \cdot e^{(-\lambda x^2)} \cdot 2x \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot x^{2\alpha} \cdot e^{(-\lambda x^2)}
\end{aligned} \tag{14}$$

which can be rewritten as

$$f_t(x, \alpha, \lambda) = \exp [2\alpha \log(x) - \lambda 2x + \alpha \log(\lambda) - \log(\Gamma(\alpha))] \tag{15}$$

with $T = (\log(x), x^2)$, $\eta = (2\alpha, -\lambda)$ and $A(\alpha, \lambda) = \log(\Gamma(\alpha)) - \alpha \log(\lambda)$.

5.2.2 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 $\mathcal{N}\left(\sqrt{\frac{\alpha}{\lambda}}, \frac{1}{4\lambda}\right)$.

5.2.3 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

$$\mu = \sqrt{\frac{\alpha}{\lambda}} \quad (16)$$

$$\sigma^2 = \frac{1}{4\lambda} \quad (17)$$

$$\lambda = \frac{4}{\sigma^2} \quad (18)$$

$$\alpha = \frac{(\sigma\mu)^2}{4} \quad (19)$$

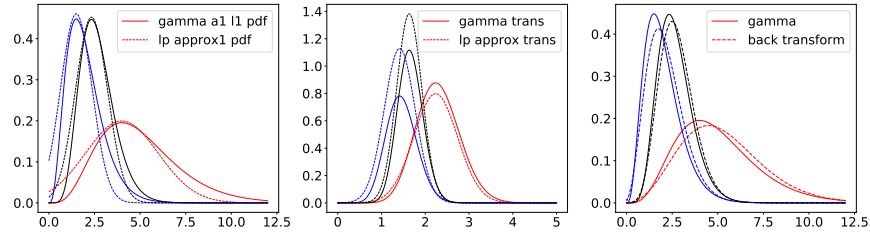


Figure 2: gamma comparison square

5.3 Log-Transform of the Gamma Distribution

5.3.1 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

$$\begin{aligned}
f_t(x, \alpha, \lambda) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \exp(x)^{(\alpha-1)} \cdot e^{(-\lambda \exp(x))} \cdot \exp(x) \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \exp(x)^\alpha \cdot e^{(-\lambda \exp(x))}
\end{aligned} \quad (20)$$

which can be rewritten as

$$f_t(x, \alpha, \lambda) = \exp[\alpha x - \lambda \exp(x) \alpha \log(\lambda) - \log(\Gamma(\alpha))] \quad (21)$$

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

5.3.2 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.

$$\begin{aligned} \text{log-pdf: } & \log\left(\frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \exp(x)^\alpha \cdot \exp(-\lambda \exp(x))\right) \\ & = \alpha \log(\lambda) - \log(\Gamma(\alpha)) + \alpha x - \lambda \exp(x) \end{aligned}$$

$$\text{1st derivative: } \alpha - \lambda \exp(x)$$

$$\text{mode: } \alpha - \lambda \exp(x) = 0 \Leftrightarrow x = \log\left(\frac{\alpha}{\lambda}\right)$$

$$\text{2nd derivative: } -\lambda \exp(x)$$

$$\text{insert mode: } -\lambda \exp(\log\left(\frac{\alpha}{\lambda}\right)) = -\frac{1}{\alpha}$$

$$\text{invert and times -1: } \sigma^2 = \alpha$$

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

5.3.3 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

$$\mu = \log(\alpha/\lambda) \quad (22)$$

$$\sigma^2 = \alpha \quad (23)$$

$$\lambda = \alpha/\exp(\mu) \quad (24)$$

$$\alpha = \sigma^2 \quad (25)$$

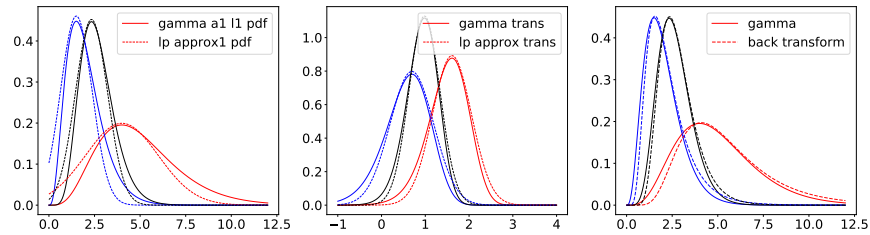


Figure 3: gamma comparison

6 Inverse Gamma Distribution

6.1 Standard Inverse Gamma Distribution

The pdf of the inverse gamma is

$$f(x, \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right) \quad (26)$$

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

$$f(x, \alpha, \lambda) = \exp [(-\alpha - 1) \log(x) - \lambda/x + \alpha \log(\lambda) - \log \Gamma(\alpha)] \quad (27)$$

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

6.1.1 Laplace Approximation of the standard inverse gamma distribution

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

6.2 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

$$f_t(x, \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \exp(x)^{-\alpha} \exp(-\lambda / \exp(x)) \quad (28)$$

which can be written as

$$f_t(x, \alpha, \lambda) = \exp \left[-\alpha x - \frac{\lambda}{\exp(x)} + \alpha \log \lambda - \log \Lambda(\alpha) \right] \quad (29)$$

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

6.2.1 Laplace Approximation of the log-transformed Inverse Gamma Distribution

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

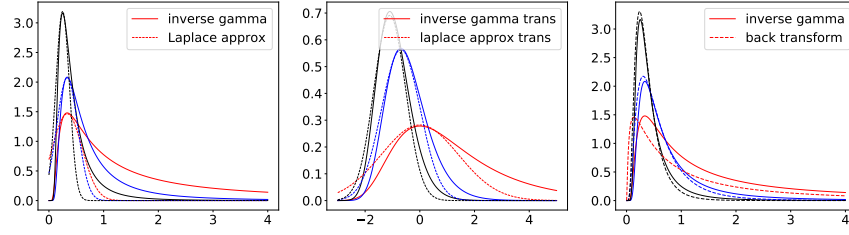


Figure 4: inverse gamma comparison

6.2.2 The Bridge for the log-transformed Inverse Gamma Distribution

$$\mu = \log(\lambda/\alpha) \quad (30)$$

$$\sigma^2 = 1/\alpha \quad (31)$$

$$\lambda = \exp(\mu)\sigma^2 \quad (32)$$

$$\alpha = 1/\sigma^2 \quad (33)$$

7 Chi-squared Distribution

7.1 Standard Chi-squared distribution

The pdf is

$$f(x, k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} \exp(-x/2) \quad (34)$$

which can be written as

$$f(x, k) = \exp \left[(k/2 - 1) \log(x) - x/2 - \log(2^{k/2}\Gamma(k/2)) \right] \quad (35)$$

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

7.1.1 Laplace approximation of the standard Chi-squared distribution

$$\text{log-pdf: } (k/2 - 1) \log(x) - x/2 - \log(2^{k/2}\Gamma(k/2))$$

$$\text{1st derivative: } \frac{k/2 - 1}{x} - \frac{1}{2}$$

$$\text{mode: } \frac{k/2 - 1}{x} - \frac{1}{2} = 0 \Leftrightarrow x = k - 2$$

$$\text{2nd derivative: } -\frac{k/2 - 1}{x^2}$$

$$\text{insert mode: } -\frac{k/2 - 1}{(k - 2)^2}$$

$$\text{invert and times -1: } \sigma^2 = \frac{(k - 2)^2}{k/2 - 1}$$

7.2 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

$$f_t(x, k) = \frac{1}{2^{k/2}\Gamma(k/2)} \exp(x)^{k/2-1} \exp(-\exp(x)/2) \quad (36)$$

which can be written as

$$f_t(x, k) = \exp \left[(k/2)x - \frac{\exp(x)}{2} - \log(2^{k/2}\Gamma(k/2)) \right] \quad (37)$$

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

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

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

7.2.2 The Bridge for log-transform

$$\mu = \log(k) \quad (38)$$

$$\sigma^2 = 2/k \quad (39)$$

$$k = \exp(\mu) \quad (40)$$

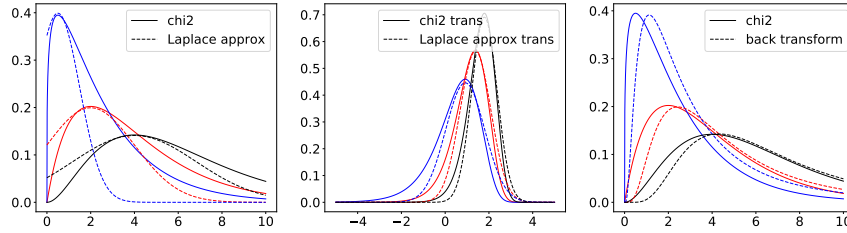


Figure 5: chi2 comparison log transform

7.3 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

$$\begin{aligned} f_t(x, k) &= \frac{1}{2^{k/2}\Gamma(k/2)} x^{2(k/2-1)} \exp(-x^2/2) 2x \\ &= \frac{1}{2^{k/2}\Gamma(k/2)} x^k \exp(-x^2/2) \end{aligned} \quad (41)$$

which can be written as

$$f_t(x, k) = \exp \left[(k \log(x) - \frac{x^2}{2} - \log(2^{k/2} \Gamma(k/2))) \right] \quad (42)$$

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

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

$$\text{log-pdf: } (k \log(x) - \frac{x^2}{2} - \log(2^{k/2} \Gamma(k/2)))$$

$$\text{1st derivative: } \frac{k}{x} - x$$

$$\text{mode: } \frac{k}{x} - x = 0 \Leftrightarrow x = \sqrt{k}$$

$$\text{2nd derivative: } -\frac{k}{x^2} - 1$$

$$\text{insert mode: } -\frac{k}{k} - 1$$

$$\text{invert and times -1: } \sigma^2 = 1/2$$

7.3.2 The Bridge for sqrt-transform

$$\mu = \sqrt{k} \quad (43)$$

$$\sigma^2 = 1/2 \quad (44)$$

$$k = \mu^2 \quad (45)$$

TODO:THE BRIDGE BACK LOOKS A BIT WEIRD

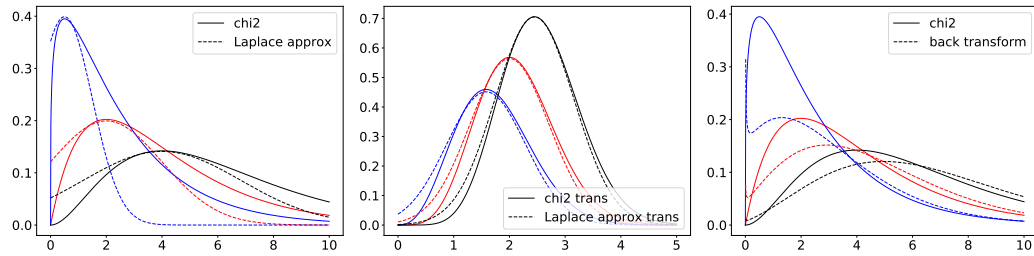


Figure 6: chi2 sqrt comparison

8 Beta Distribution

8.1 Standard Beta Distribution

The pdf of the Beta distribution in the standard basis is

$$f(x, \alpha, \beta) = \frac{x^{(\alpha-1)} \cdot (1-x)^{(\beta-1)}}{B(\alpha, \beta)} \quad (46)$$

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

$$f(x, \alpha, \beta) = \exp [(\alpha - 1) \log(x) + (\beta - 1) \log(1 - x) - \log(B(\alpha, \beta)))] \quad (47)$$

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

8.1.1 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.

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

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

8.2 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

$$\begin{aligned}
f_t(x, \alpha, \beta) &= \frac{\sigma(x)^{(\alpha-1)} \cdot (1-\sigma(x))^{(\beta-1)}}{B(\alpha, \beta)} \cdot (\sigma(x)(1-\sigma(x))) \\
&= \frac{\sigma(x)^\alpha \cdot (1-\sigma(x))^\beta}{B(\alpha, \beta)}
\end{aligned} \tag{48}$$

which can be rewritten to get

$$f_t(x, \alpha, \beta) = \exp [\alpha \log(\sigma(x)) + \beta \log(1-\sigma(x)) - \log B(\alpha, \beta)] \tag{49}$$

$$= \exp [\alpha \log(\sigma(x)) + \beta \log(\sigma(-x)) - \log B(\alpha, \beta)] \tag{50}$$

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

$$\tag{52}$$

TODO: THIS FEELS LIKE IT COULD BE SIMPLIFIED BUT I CANNOT FIND THE SOLUTION

8.2.1 Laplace approximation of the logit transformed Beta distribution

mode and variance of log pdf blablabla

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

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

8.2.2 The Bridge for the logit transformation

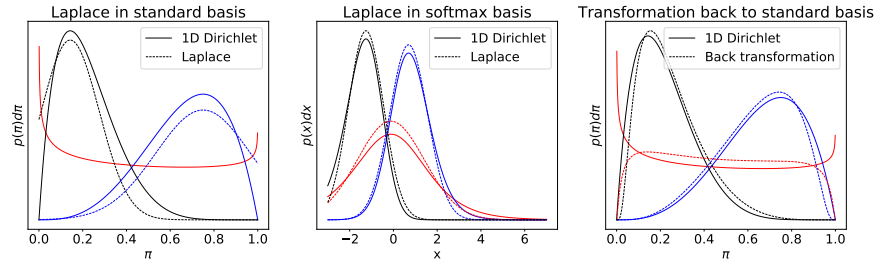


Figure 7: beta comparison

9 Dirichlet Distribution

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

10 Wishart Distribution

10.1 Standard Wishart distribution

the pdf of the Wishart is

$$f(x; n, p, V) = \frac{1}{2^{np/2} |\mathbf{V}|^{n/2} \Gamma_p\left(\frac{n}{2}\right)} |\mathbf{x}|^{(n-p-1)/2} e^{-(1/2) \text{tr}(\mathbf{V}^{-1}\mathbf{x})} \quad (53)$$

which can be written as

$$f(x; n, p, V) = \exp \left[(n - p - 1)/2 \log(|x|) - (1/2) \text{tr}(\mathbf{V}^{-1}\mathbf{x}) - \log \left(2^{np/2} |\mathbf{V}|^{n/2} \Gamma_p\left(\frac{n}{2}\right) \right) \right] \quad (54)$$

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

10.1.1 Laplace Approximation of the standard Wishart distribution

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

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

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 $\mathcal{N}(X; (n-p-1)V, 2(n-p-1)(V \otimes V))$, where the representation of the symmetric positive definite matrices has been changed from $\mathbb{R}^{n \times n}$ to \mathbb{R}^{n^2} .

10.2 Sqrtm-Transformed Wishart distribution

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

$$\begin{aligned} f_t(x; n, p, V) &= \frac{1}{2^{np/2} |\mathbf{V}|^{n/2} \Gamma_p\left(\frac{n}{2}\right)} |\mathbf{x}^2|^{(n-p-1)/2} e^{-(1/2) \text{tr}(\mathbf{V}^{-1} \mathbf{x}^2)} \cdot |2x| \quad (55) \\ &= \frac{1}{2^{np/2} |\mathbf{V}|^{n/2} \Gamma_p\left(\frac{n}{2}\right)} |\mathbf{x}|^{2(n-p-1)/2} e^{-(1/2) \text{tr}(\mathbf{V}^{-1} \mathbf{x}^2)} \cdot 2^k |x| \\ &= \frac{2^k}{2^{np/2} |\mathbf{V}|^{n/2} \Gamma_p\left(\frac{n}{2}\right)} |\mathbf{x}|^{(n-p)} e^{-(1/2) \text{tr}(\mathbf{V}^{-1} \mathbf{x}^2)} \end{aligned}$$

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

$$f_t(x; n, p, V) = \exp \left[(n-p) \log(|x|) - (1/2) \text{tr}(\mathbf{V}^{-1} \mathbf{x}^2) - \log \left(2^{np/2} |\mathbf{V}|^{n/2} \Gamma_p\left(\frac{n}{2}\right) \right) + k \log(2) \right] \quad (56)$$

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

10.2.1 Laplace Approximation of the sqrtm-transformed Wishart distribution

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

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

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

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

TODO: SOMETHING IS WRONG HERE: THE DIMENSIONS DO NOT ADD UP.

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:

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

TODO: SOMETHING IS WRONG HERE: THE DIMENSIONS DO NOT ADD UP.

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

10.2.2 The Bridge for sqrtm-tranform

TODO: COMPUTE THE ACTUAL BRIDGE

TODO:SOMETHING LOOKS A BIT FISHY HERE!!

11 Inverse Wishart Distribution

TODO: see above

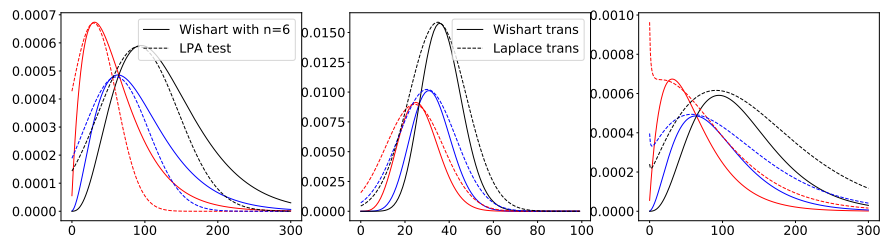


Figure 8: wishart comparison for sqrtm