Appendix 1

September 24, 2020

Contents

1 Mode 1: Gamma bulk distribution with constant generalized pareto distribution 1
  1.1 Sampling $\xi$ ................................................................. 2
  1.2 Sampling $\sigma$ ............................................................. 2
  1.3 Sampling $u$ ................................................................. 2
  1.4 Sampling $\alpha$ and $\beta$ ................................................. 2

2 Model 2 & 3: Constant bulk regression with time-varying generalized pareto distribution 2
  2.1 Sampling $\beta$ and $\sigma^2$ ............................................... 3
  2.2 Sampling $\{\xi_t\}_{t=1}^T$ ................................................. 3
  2.3 Sampling $\{\sigma_t\}_{t=1}^T$ ............................................ 3
  2.4 Sampling $u$ ................................................................. 4

3 Model 4: Constant bulk regression and generalized pareto distribution regression 4
  3.1 Sampling $\beta$ and $\sigma^2$ ............................................... 4
  3.2 Sampling $\{\xi_t\}_{t=1}^T$ ................................................. 4
  3.3 Sampling $\{\sigma_t\}_{t=1}^T$ ............................................ 5
  3.4 Sampling $u$ ................................................................. 5

4 Model 5: Constant bulk regression and time-varying generalized pareto regression 5
  4.1 Sampling $\beta$ and $\sigma^2$ ............................................... 5
  4.2 Sampling $\{\xi_t\}_{t=1}^T$ ................................................. 5
  4.3 Sampling $\{\sigma_t\}_{t=1}^T$ ............................................ 6
  4.4 Sampling $u$ ................................................................. 6

5 Model Assessment 6

1 Mode 1: Gamma bulk distribution with constant generalized pareto distribution

Below the threshold $u$, the bulk distribution is a gamma distribution, with shape and scale parameters $\alpha$ and $\beta$. Above the threshold $u$, we place the generalized pareto distribution (GPD) with parameters $(\xi, \sigma, u)$:

$$G(y; u, \Theta_u) = \begin{cases} 1 - \left(1 + \frac{\xi(y-u)}{\sigma}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{(y-u)}{\sigma}\right) & \text{if } \xi = 0 \end{cases}$$

(1)

We apply a non-informative prior for $(\xi, \sigma, u)$ following Castellanos and Cabras 2007 and Do Nascimento 2012:

$$p(\xi, \sigma) \propto \sigma^{-1}(1 + \xi)^{-1}(1 + 2\xi)^{-\frac{3}{2}}$$

A truncated normal distribution with parameters $(\mu_u, \sigma_u^2)$ with the following density is placed on $u$:

$$p(u|\mu_u, \sigma_u^2) \propto \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-0.5(u - \mu_u)^2/\sigma_u^2\right)$$

\Phi(\mu_u/\sigma_u)

With the following priors set for parameters below the threshold,

$$\alpha \sim Ga(a, b)$$

$$\mu \sim Ga(c, d)$$

where $\mu = \frac{\alpha}{\beta}$
1.1 Sampling $\xi$

$\xi^*$ is sampled from a $N(\xi(s), V_\xi)I(-\sigma(s)/(M - u(s), \infty))$ distribution, where $M = max(x_1, \ldots, x_n)$. Therefore $\xi(s+1) = \xi^*$ is drawn with acceptance probability $\alpha_\xi$:

$$
\alpha_\xi = \min \left(1, \frac{p(\theta^*|x)\Phi((\xi^* + \sigma(s)/(M - u(s)))/(\sqrt{V_\xi}))}{p(\theta|x)\Phi((\xi^* + \sigma(s)/(M - u(s)))/(\sqrt{V_\xi}))} \right)
$$

1.2 Sampling $\sigma$

If $\xi(s+1) \geq 0$ then $\sigma^*$ is sampled from a $Ga(a_s, b_s)$ distribution with $a_s = (\sigma(s))^2/V_\sigma$, $b_s = \sigma^*/V_\sigma$. If $\xi(s+1) < 0$ then $\sigma^*$ is sampled from a $N(\sigma(s), V_\sigma)I(-\xi(s+1)/(M - u(s), \infty))$ distribution. Therefore, $\sigma(s+1) = \sigma^*$ with acceptance probability $\alpha_\sigma$:

$$
\alpha_\sigma = \min \left(1, \frac{p(\theta^*|x)g(\sigma^*(a_s, b_s))}{p(\theta|x)g(\sigma^*|a^*, b^*)} \right) \quad \text{if} \quad \xi(s+1) \geq 0
$$

$$
\alpha_\sigma = \min \left(1, \frac{p(\theta^*|x)\Phi((\sigma^* + \xi(s+1)/(M - u(s)))/(\sqrt{V_\sigma}))}{p(\theta|x)\Phi((\sigma^* + \xi(s+1)/(M - u(s)))/(\sqrt{V_\sigma}))} \right) \quad \text{if} \quad \xi(s+1) < 0
$$

1.3 Sampling $u$

$u^*$ is sampled from a $N(u(s), V_u)I(a(s+1), M)$ distribution, where

$$
a^{(s+1)} = \begin{cases} 
\min(x_1, \ldots, x_n), & \text{if} \; \xi^{(s+1)} \geq 0 \\
M + \sigma^{(s+1)}/\xi^{(s+1)}, & \text{if} \; \xi^{(s+1)} < 0
\end{cases}
$$

Therefore $u^{(s+1)} = u^*$ with probability $\alpha_u$ where

$$
\alpha_u = \min \left(1, \frac{p(\theta^*|x)\Phi((M - u(s))/(\sqrt{V_u}) - \Phi((a^{(s+1)} - u(s))/(\sqrt{V_u}))}{p(\theta|x)\Phi((M - u^*)/(\sqrt{V_u}) - \Phi((a^{(s+1)} - u^*)/(\sqrt{V_u}))} \right)
$$

1.4 Sampling $\alpha$ and $\beta$

$\alpha^*$ and $\beta^*$ are sampled, respectively, from $N(\alpha(s), V_\alpha)I(0.1, \infty)$, $N(\beta(s), V_\beta)I(0.1, \infty)$. Therefore, $\alpha^{(s+1)}, \beta^{(s+1)} = (\alpha^*, \beta^*)$ with probability

$$
\alpha_\alpha = \min \left(1, \frac{p(\theta^*|x)h(\alpha^*(\alpha^*, V_\alpha))g(\beta^{(s+1)}|a_s, b_s)}{p(\theta|x)h(\alpha^*(\alpha(s), V_\alpha))g(\beta^*|a^*, b^*))} \right)
$$

2 Model 2 & 3: Constant bulk regression with time-varying generalized pareto distribution

Below the threshold $u$, we place the following regression structure on $Y, X$ the dependent and independent variables respectively, $\beta$ the regression parameters of interest and white noise parameterized by $\epsilon$. $n$ denotes the number of observations below the threshold $u$ and $p$ the number of dependent variables in the regression equation:

$$
Y_{n \times 1} = X_{n \times p}^\beta + \epsilon_{n \times 1}
$$

$$
\epsilon \sim N(0, \sigma^2)
$$

We estimate sequentially $\Theta_{-u} = \{\beta, \sigma^2\}$ by placing the following priors on parameters:

$$
\beta \sim N(\beta_0, P_0)
$$

with $\beta_0 = 0_{p \times 1}$ and $P_0 = \text{diag}(100)_{p \times p}$ for our parameters to be centered around 0 and having a wide variance to impose noninformativeness with scalar $T_0 = \theta_0 = 1$ to yield a non-informative inverse-gamma prior distribution for $\sigma$:

$$
\sigma^2 \sim IG\left(\frac{T_0}{2}, \frac{\theta_0}{2}\right)
$$

Above the threshold $u$, we place the generalized pareto distribution (GPD) with time varying parameters:
\[ G(y; u, \Theta_u) = \begin{cases} 1 - \left( 1 + \frac{\xi(y-u)}{\sigma_t} \right)^{-1/\xi} & \text{if } \xi_t \neq 0 \\ 1 - \exp\left(-\frac{(y-u)}{\sigma_t}\right) & \text{if } \xi_t = 0 \end{cases} \] (4)

The GPD parameters \( \xi_t, \sigma_t \) follow random walk state equations with white noise as follows:

\[
\begin{align*}
\xi_t &= \xi_{t-1} + w_{\xi,t} \quad w_{\xi,t} \sim N(0, 1/W_{\xi}) \\
\sigma_t &= \sigma_{t-1} + w_{\sigma,t} \quad w_{\sigma,t} \sim N(0, 1/W_{\sigma})
\end{align*}
\]

We log-transform \( \xi_t, \sigma_t \) to allow parameters to be within the allowable bounds for the GPD \( (\xi < -1) \) with \( l_{\xi_t} = \log(\xi_t + 1), l_{\sigma_t} = \log\sigma_t \):

\[
\begin{align*}
l_{\xi_t} &= \theta_{\xi,t} + v_{\xi,t} \quad v_{\xi,t} \sim N(0, 1/V_{\xi}) \\
l_{\sigma_t} &= \theta_{\sigma,t} + v_{\sigma,t} \quad v_{\sigma,t} \sim N(0, 1/V_{\sigma}) \\
\theta_{\xi,t} &= \theta_{\xi,t-1} + w_{\xi,t} \quad w_{\xi,t} \sim N(0, 1/W_{\xi}) \\
\theta_{\sigma,t} &= \theta_{\sigma,t-1} + w_{\sigma,t} \quad w_{\sigma,t} \sim N(0, 1/W_{\sigma})
\end{align*}
\]

We estimate sequentially \( \{V_{\xi}, V_{\sigma}, W_{\xi}, W_{\sigma}\} \) by placing the following priors on parameters:

\[
\begin{align*}
V_{\xi} &\sim \text{Ga}(l_{\xi}, m_{\xi}) \\
V_{\sigma} &\sim \text{Ga}(l_{\sigma}, m_{\sigma}) \\
W_{\xi} &\sim \text{Ga}(f_{\xi}, o_{\xi}) \\
W_{\sigma} &\sim \text{Ga}(f_{\sigma}, o_{\sigma})
\end{align*}
\]

With the initial information for \( \theta_{\xi,0}, \theta_{\sigma,0} \) given by:

\[
\begin{align*}
\theta_{\sigma,0} &\sim N(\mu_{\sigma,0}, C_{\sigma,0}) \\
\theta_{\xi,0} &\sim N(\mu_{\xi,0}, C_{\xi,0})
\end{align*}
\]

We apply the same threshold prior as Model 1.

### 2.1 Sampling \( \beta \) and \( \sigma^2 \)

\( \beta \) is sampled from:

\[ \beta^* \sim (\Sigma_0^{-1} + \frac{1}{\sigma^2} X'X)^{-1}(\Sigma_0^{-1}\beta_0 + \frac{1}{\sigma^2}X'y) \]

\( \sigma^2 \) is sampled from:

\[ \sigma^2 \sim (\Sigma_0^{-1} + \frac{1}{\sigma^2} X'X)^{-1} \]

### 2.2 Sampling \( \{l_{\xi_t}\}_{t=1}^T \)

If \( x_t < u^{(s)} \), we sample parameters from \( l_{\xi_t}^{(s+1)} \sim N(\theta_{\xi,t}^{(s)}, 1/V_{\xi}^{(s)}) \). If \( x_t \geq u^{(s)} \), \( l_{\xi_t}^* \) is sampled from \( N(l_{\xi_t}^{(s)}, K_{\xi}) \). Therefore, \( l_{\xi_t}^{(s+1)} = l_{\xi_t}^* \) with acceptance probability:

\[ \min\left(1, \frac{p(\theta^{*}\mid y)}{p(\theta\mid y)} \right) \]

### 2.3 Sampling \( \{l_{\sigma_t}\}_{t=1}^T \)

If \( x_t < u^{(s)} \), then the parameter can be sampled by: \( l_{\sigma_t}^{(s+1)} \sim N(\theta_{\sigma,t}^{(s)}, 1/V_{\sigma}^{(s)}) \). If \( x_t \geq u^{(s)} \), \( l_{\sigma_t}^* \) is sampled from \( N(l_{\sigma_t}^{(s)}, K_{\sigma}) \). Therefore, \( l_{\sigma_t}^{(s+1)} = l_{\sigma_t}^* \) with probability:

\[ \min\left(1, \frac{p(\theta^{*}\mid y)}{p(\theta\mid y)} \right) \]
2.4 Sampling $u$

$u^*$ is sampled from $N(u(s), V_u)I(u_L^{(s)}, \infty)$, where $u_L^{(s)} = \max \left( \min(x_{1:t}, \max_{\xi_t^{t+1} < 0, x_t > u^*} (x_{1:t} + \sigma^{s+1}) / (\xi_t^{t+1} + 1)) \right)$. Therefore, $u^{(s+1)}$ is accepted with probability:

$$\min \left( 1, \frac{\pi(\Theta^*|x, y)\Phi((u^* - u_L^*)/\sqrt{V_u})}{\pi(\Theta|x, y)\Phi((u^* - u_L^*)/\sqrt{V_u})} \right)$$

${\{V_\xi, W_\xi, \theta_\xi, V_\sigma, W_\sigma, \theta_\sigma\}}$ are updated via the following Gibbs steps:

$$V_\xi^{(s+1)} \sim G \left( f_\xi + \frac{T}{2}, \theta_\xi + \frac{1}{2} \sum_{t=1}^T (l_{\xi_t}^{(s+1)} - \theta_\xi^{(s+1)})^2 \right)$$

$$W_\xi^{(s+1)} \sim G \left( l_\xi + \frac{T}{2}, m_\xi + \frac{1}{2} \sum_{t=1}^T (l_{\xi_t}^{(s+1)} - \theta_\xi^{(s+1)})^2 \right)$$

$$\theta_\xi^{(s+1)} \sim N \left( \frac{W_\xi^{(s+1)}\theta_\xi^{(s)} + m_\xi, \theta_\xi^{(s)} - \theta_\xi^{(s+1)}}{W_\xi^{(s+1)} + 1} \right)$$

$$\theta_\sigma^{(s+1)} \sim N \left( \frac{V_\sigma^{(s+1)}\xi_t^{(s+1)} + W_\sigma^{(s+1)}\theta_\sigma^{(s)} + \theta_\sigma^{(s+1)}}{V_\sigma^{(s+1)} + W_\sigma^{(s+1)}} \right)$$

Posterior distributions for $V_\sigma, W_\sigma, \theta_\sigma$ follow the same functional form.

3 Model 4: Constant bulk regression and generalized pareto distribution regression

We apply the same bulk distribution and extreme value distribution as Model 2 and 3, but additionally impose regression structure for the extreme parameters for Model 4. We set the following priors for $\beta_\xi$ and $\beta_\sigma$:

$$\beta \sim N(\beta_0, 1/\Sigma)$$

where $\beta_0 = 0_{p \times 1}$, $\Sigma = \text{diag}(\frac{1}{100})_{p \times p}$

$$l_{\xi_t} = \beta_{\xi,t} + v_{\xi,t} \quad v_{\xi,t} \sim N(0, 1/V_\xi)$$

$$l_{\sigma_t} = \beta_{\sigma,t} + v_{\sigma,t} \quad v_{\sigma,t} \sim N(0, 1/V_\sigma)$$

where $\beta_{\xi,t} = \sum_{k=0}^\rho \beta_{\xi,k} X_{t,k}$, $\beta_{\sigma,t} = \sum_{k=0}^\rho \beta_{\sigma,k} X_{t,k}$

3.1 Sampling $\beta$ and $\sigma^2$

$\beta$ is sampled from

$$\beta^* \sim (\Sigma_0^{-1} + \frac{1}{\sigma^2} X'X)^{-1}(\Sigma_0^{-1}\beta_0 + \frac{1}{\sigma^2} X'Y)$$

$\sigma^2$ is sampled from

$$\sigma^2 \sim (\Sigma_0^{-1} + \frac{1}{\sigma^2} X'X)^{-1}$$

3.2 Sampling $\{l_{\xi_t}^{(s)}\}_{t=1}^T$

If $x_t < u(s)$, then the parameter can be sampled by: $l_{\xi_t}^{(s+1)} \sim N(\beta_{\xi,t}^{(s)} 1/V_\xi^{(s)})$. If $x_t \geq u(s)$, $l_{\xi_t}^*$ is sampled from $N(l_{\xi_t}^{(s)}, K_\xi)$. Therefore, $l_{\xi_t}^{(s+1)} = l_{\xi_t}^*$ with probability

$$\min \left( 1, \frac{p(\theta^*|y)}{p(\theta|y)} \right)$$

where $\beta_{\xi,t} = \sum_{k=0}^\rho \beta_{\xi,t,k} X_{t,k}$.
3.3 Sampling \( \{l\sigma_t\}_t^T \)

If \( x_t < u^{(s)} \), then the parameter can be sampled by: \( l\sigma_t^{(s+1)} \sim N(\beta^{(s)}_{\sigma,t}, 1/V^{(s)}_\sigma) \). If \( x_t \geq u^{(s)} \), \( l\sigma_t \) is sampled from \( N(l\sigma_t^{(s)}, K_{\sigma}) \).

Therefore, \( l\sigma_t^{(s+1)} = l\sigma_t^* \) with probability

\[
\min \left( 1, \frac{p(\theta^*|y)}{p(\theta|y)} \right)
\]

where \( \beta_{\sigma,t} = \sum_{k=0}^p \beta_{\xi,t,k}X_{t,k} \).

3.4 Sampling \( u \)

\( u^* \) is sampled from a \( N(u^{(s)}, V_u^*) \), where \( u_L^* = \max \left( \min(x_{1:t}), \frac{\max(\{x_{1:t} + \sigma^{(s)} + 1\})}{(\xi^*_{t+1})} \right) \). Therefore, \( u^{(s+1)} \) is accepted with probability:

\[
\min \left( 1, \frac{p(\theta^*|x,y)\Phi((u^* - u_L^*)/\sqrt{V_u})}{p(\theta|x,y)\Phi((u^* - u_L^*)/\sqrt{V_u})} \right)
\]

\( \{V_\xi, \beta_\xi, V_\sigma, \beta_\sigma\} \) can be updated via Gibbs steps.

\[
V^{(s+1)}_\xi \sim G\left( f_\xi + \frac{T}{2}, a_\xi + \frac{1}{2} \sum_{t=1}^T (l\xi_{t}^{(s+1)} - \beta^{(s)}_{\xi,t})^2 \right)
\]

\[
\beta_\xi \sim N\left( \frac{X'V_\xi l\xi + \Sigma_\beta_0}{X'V_\xi X + \Sigma}, \frac{1}{X'V_\xi X + \Sigma} \right)
\]

where \( l\xi = \{l\xi_t\}_t^T \). Posterior distributions for \( V_\sigma, \beta_\sigma \) also follow the same functional form.

4 Model 5: Constant bulk regression and time-varying generalized pareto regression

In Model 4, the coefficients are constant in the regression structure. We use linear dynamic model which allows time-varying coefficients in Model 5.

\[
l_{t} = \beta_{\xi,t} + v_{\xi,t}, \quad v_{\xi,t} \sim N(0, 1/V_{\xi})
\]

\[
\beta_{\xi,t,k} = \beta_{\xi,t-1,k} + \omega_{\xi,t}, \quad \omega_{\xi,t} \sim N(0, 1/W_{\xi,k})
\]

\[
l\sigma_t = \beta_{\sigma,t} + v_{\sigma,t}, \quad v_{\sigma,t} \sim N(0, 1/V_{\sigma})
\]

\[
\beta_{\sigma,t,k} = \beta_{\sigma,t-1,k} + \omega_{\sigma,t}, \quad \omega_{\sigma,t} \sim N(0, 1/W_{\sigma,k})
\]

where \( \beta_{\xi,t} = \sum_{k=0}^p \beta_{\xi,t,k}X_{t,k} \), \( \beta_{\sigma,t} = \sum_{k=0}^p \beta_{\sigma,t,k}X_{t,k} \).

4.1 Sampling \( \beta \) and \( \sigma^2 \)

\( \beta \) is sampled from

\[
\beta^* \sim (\Sigma^{-1}_0 + \frac{1}{\sigma_2} X'X)^{-1}(\Sigma^{-1}_0 \beta_0 + \frac{1}{\sigma_2} X'Y)
\]

\( \sigma^2 \) is sampled from

\[
\sigma^2 \sim (\Sigma^{-1}_0 + \frac{1}{\sigma_2} X'X)^{-1}
\]

4.2 Sampling \( \{l\xi_t\}_t^T \)

If \( x_t < u^{(s)} \), then the parameter can be sampled by: \( l\xi_t^{(s+1)} \sim N(\beta^{(s)}_{\xi,t}, 1/V^{(s)}_{\xi}) \). If \( x_t \geq u^{(s)} \), \( l\xi_t^* \) is sampled from \( N(l\xi_t^{(s)}, K_{\xi}) \).

Therefore, \( l\xi_t^{(s+1)} = l\xi_t^* \) with probability

\[
\min \left( 1, \frac{p(\theta^*|y)}{p(\theta|y)} \right)
\]

where \( \beta_{\xi,t} = \sum_{k=0}^p \beta_{\xi,t,k}X_{t,k} \).
4.3 Sampling \( \{l\sigma_t\}_{t=1}^T \)

If \( x_t < u^{(s)} \), then the parameter can be sampled by: \( l\sigma_t^{(s+1)} \sim N(\beta_{\sigma,t}^{(s)}, 1/V_{\sigma}^{(s)}) \). If \( x_t \geq u^{(s)} \), \( l\sigma_t \) is sampled from \( N(l\sigma_t^{(s)}, K_\sigma) \). Therefore, \( l\sigma_t^{(s+1)} = l\sigma_t \) with probability

\[
\min \left( 1, \frac{p(\theta^s|x,y)}{p(\theta|x,y)} \right)
\]

where \( \beta_{\sigma,t} = \sum_{k=0}^T \beta_{\xi,t,k} X_{t,k} \).

4.4 Sampling \( u \)

\( u^* \) is sampled from a \( N(u^{(s)}, V_u)I(u^{(s)}, \infty) \), where \( u^{(s)} = \max \left( \min(x_{1:t}), \max_{\xi_{t+1} < 0, x_t > u^*} (x_{1:t} + \sigma^{s+1}) / (\xi_t^{s+1} + 1 + \xi_t^{s+1}) \right) \). Therefore, \( u^{(s+1)} \) is accepted with probability:

\[
\min \left( 1, \frac{p(\theta^s| x,y) \phi((u^* - u_{t}^*)/\sqrt{V_u})}{p(\theta|x,y) \phi((u^* - u_{t}^*)/\sqrt{V_u})} \right)
\]

\( \{V_{\xi}, W_{\xi,k}, \beta_{\xi,t,k}, V_\sigma, W_{\sigma,k}, \beta_{\sigma,t,k} \} \) can be updated via Gibbs steps:

\[
V_{\xi}^{(s+1)} \sim G \left( f_{\xi} + \frac{T}{2}, \alpha_{\xi} + \frac{1}{2} \sum_{t=1}^T (l_{\xi,t}^{(s+1)} - \beta_{\xi,t}^{(s)})^2 \right)
\]

\[
W_{\xi,k}^{(s+1)} \sim G \left( l_{\xi} + \frac{T}{2}, m_{\xi} + \frac{1}{2} \sum_{t=1}^T (\beta_{\xi,t,k}^{(s)} - \beta_{\xi,t-1,k}^{(s)})^2 \right)
\]

\[
\beta_{\xi,0,k}^{(s+1)} \sim N \left( W_{\xi,k}^{(s+1)} \beta_{\xi,1,k}^{(s)} + m_{\xi,0}/C_{\xi,0}, 1/W_{\xi,k}^{(s+1)} + 1/C_{\xi,0} \right)
\]

\[
\beta_{\xi,T,k}^{(s+1)} \sim N \left( V_{\xi}^{(s+1)} S_{t,k} X_{t,k} + W_{\xi,k}^{(s+1)} (\beta_{\xi,t+1,k}^{(s)} + \beta_{\xi,t-1,k}^{(s)}) + V_{\xi}^{(s+1)} X_{T,k}^2 + 2W_{\xi,k}^{(s+1)}, 1/V_{\xi}^{(s+1)} X_{T,k}^2 + W_{\xi,k}^{(s+1)} \right)
\]

where \( S_{t,k} = l_{\xi,t} - \sum_{i=0}^T \beta_{\xi,t,i} X_{t,i} \). Posterior distributions for \( V_{\sigma}, W_{\sigma,k}, \beta_{\sigma,t,k} \) also follow the same functional form.

5 Model Assessment

We use the deviation information criterion (DIC) and log Bayes factor (logBF) to assess the appropriateness of each model for the data generating process.

\[
DIC = 2 \times D(\theta) - D(\bar{\theta})
\]

where \( D(\theta) \) is the average of \( D(\theta) \) over all samples of \( \theta \), \( D(\bar{\theta}) \) is the value of \( D \) evaluated at the average of the samples of \( \theta \).

\[
D(\theta) = -2 \times \log(p(y|\theta)) + C
\]

\( C \) is a constant that can be ignored during the calculation.

\[
BF = \frac{p(D|M_1)}{p(D|M_2)}
\]

\[
\log BF = \log(p(D|M_1)) - \log(p(D|M_2))
\]

where \( p(D|M) \) denotes the likelihood that some data is produced under the assumption of model \( M \).