1 Gradient Calculations for All Models

For all models, the negative log likelihood function for a vector of parameters θ holding fixed a dataset consisting of N tuples of the form $(x_i^2, t_i^2, x_i^1, t_i^1, y_i)$ is:

$$-\mathcal{L}(\theta) = -\sum_{i=1}^{N} y_i \log(p_i) + (1 - y_i)\log(1 - p_i)$$

We can treat all models as embedded inside of the ϵ-noise formulation since we can always set $\epsilon = 0$ when the additional noise is not appropriate. In this formulation, the probability that $y_i = 1$ is p_i, which is derived from an inner probability without ϵ-noise. We'll call the inner probability $\tilde{p_i}$. The two probabilities are related as follows:

$$p_i = (1 - \epsilon)\tilde{p_i} + \epsilon(\frac{1}{2})$$

Because we always generate $\tilde{p_i}$ using an inverse logit link function of a predictor, we can write $\tilde{p_i}$ as a function of the (potentially non-linear) predictor value, z_i:

$$\tilde{p_i} = L(z_i) = (1 + \exp(-z_i))^{-1}$$

The linear relationship between p_i and $\tilde{p_i}$ implies that their derivatives with respect to model parameters are related as follows:

$$\frac{\partial}{\partial \theta_j} p_i = (1 - \epsilon)\frac{\partial}{\partial \theta_j}\tilde{p_i}$$

A nice property of the inverse link function is that its derivative can be written in terms of a simple function of its value:

$$\frac{\partial}{\partial z_i} L(z_i) = L(z_i)(1 - L(z_i)) = \tilde{p_i}(1 - \tilde{p_i})$$

Given all of this, we can compute gradients with regard to the j-th parameter θ_j as follows using repeated applications of the chain rule to drill down to the places where different models have different functional forms, which generates different gradients for different parameters:

$$
\begin{aligned}
\frac{\partial}{\partial \theta_j} - \mathcal{L}(\theta) &= -\sum_{i=1}^{N} y_i \frac{\partial}{\partial \theta_j} \log(p_i) + (1 - y_i) \frac{\partial}{\partial \theta_j} \log(1 - p_i) \\
&= -\sum_{i=1}^{N} y_i \frac{1}{p_i} \frac{\partial}{\partial \theta_j} p_i + (1 - y_i) \frac{1}{1 - p_i} \frac{\partial}{\partial \theta_j} (1 - p_i) \\
&= -\sum_{i=1}^{N} y_i \frac{1}{p_i} \frac{\partial}{\partial \theta_j} p_i - (1 - y_i) \frac{1}{1 - p_i} \frac{\partial}{\partial \theta_j} p_i \\
&= -\sum_{i=1}^{N} y_i \frac{1}{p_i} (1 - \epsilon) \frac{\partial}{\partial \theta_j} \tilde{p}_i - (1 - y_i) \frac{1}{1 - p_i} (1 - \epsilon) \frac{\partial}{\partial \theta_j} \tilde{p}_i \\
&= -\sum_{i=1}^{N} y_i \frac{1}{p_i} (1 - \epsilon) \tilde{p}_i (1 - \tilde{p}_i) \frac{\partial}{\partial \theta_j} z_i - (1 - y_i) \frac{1}{1 - p_i} (1 - \epsilon) \tilde{p}_i (1 - \tilde{p}_i) \frac{\partial}{\partial \theta_j} z_i
\end{aligned}
$$

In the special case in which $\epsilon = 0$, this last equation simplifies considerably because $p_i = \tilde{p}_i$ when $\epsilon = 0$:

$$
\begin{aligned}
\frac{\partial}{\partial \theta_j} - \mathcal{L}(\theta) &= -\sum_{i=1}^{N} y_i (1 - p_i) \frac{\partial}{\partial \theta_j} z_i - (1 - y_i) p_i \frac{\partial}{\partial \theta_j} z_i \\
&= -\sum_{i=1}^{N} (y_i - p_i) \frac{\partial}{\partial \theta_j} z_i \\
&= \sum_{i=1}^{N} (p_i - y_i) \frac{\partial}{\partial \theta_j} z_i
\end{aligned}
$$

2 Gradient Calculations for Specific Models

2.1 Gradient Calculations for Baseline

$$
\begin{aligned}
\theta &= (\beta_0) \\
z_i &= \beta_0 \\
\frac{\partial}{\partial \beta_0} z_i &= 1
\end{aligned}
$$

2.2 Gradient Calculations for ITCH

$$\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$$

$$z_i = \beta_0 + \beta_1(x_i^2 - x_i^1) + \beta_2(\frac{x_i^2 - x_i^1}{\frac{x_i^2+x_i^1}{2}}) + \beta_3(t_i^2 - t_i^1) + \beta_4(\frac{t_i^2 - t_i^1}{\frac{t_i^2+t_i^1}{2}})$$

$$\frac{\partial}{\partial\beta_0} z_i = 1$$

$$\frac{\partial}{\partial\beta_1} z_i = x_i^2 - x_i^1$$

$$\frac{\partial}{\partial\beta_2} z_i = \frac{x_i^2 - x_i^1}{\frac{x_i^2+x_i^1}{2}}$$

$$\frac{\partial}{\partial\beta_3} z_i = t_i^2 - t_i^1$$

$$\frac{\partial}{\partial\beta_4} z_i = \frac{t_i^2 - t_i^1}{\frac{t_i^2+t_i^1}{2}}$$

2.3 Gradient Calculations for DRIFT

$$\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$$

$$z_i = \beta_0 + \beta_1(x_i^2 - x_i^1) + \beta_2(\frac{x_i^2 - x_i^1}{x_i^1}) + \beta_3(\frac{x_i^2}{x_i^1}^{\frac{1}{t_i^2-t_i^1}} - 1) + \beta_4(t_i^2 - t_i^1)$$

$$\frac{\partial}{\partial\beta_0} z_i = 1$$

$$\frac{\partial}{\partial\beta_1} z_i = x_i^2 - x_i^1$$

$$\frac{\partial}{\partial\beta_2} z_i = \frac{x_i^2 - x_i^1}{x_i^1}$$

$$\frac{\partial}{\partial\beta_3} z_i = \frac{x_i^2}{x_i^1}^{\frac{1}{t_i^2-t_i^1}} - 1$$

$$\frac{\partial}{\partial\beta_4} z_i = t_i^2 - t_i^1$$

2.4 Gradient Calculations for Trade-Off

$$
\begin{aligned}
\theta &= (a, k, \gamma_x, \gamma_t) \\
z_i &= a(\phi(x_i^2, \gamma_x) - \phi(x_i^1, \gamma_x) - k * (\phi(t_i^2, \gamma_t) - \phi(t_i^1, \gamma_t))) \\
\phi(\chi, \gamma) &= \frac{\log(1 + \gamma\chi)}{\gamma} \\
\phi'(\chi, \gamma) &= \frac{\frac{\gamma\chi}{1+\gamma\chi} - \log(1 + \gamma\chi)}{\gamma^2} \\
\frac{\partial}{\partial a} z_i &= \phi(x_i^2, \gamma_x) - \phi(x_i^1, \gamma_x) - k * (\phi(t_i^2, \gamma_t) - \phi(t_i^1, \gamma_t)) \\
\frac{\partial}{\partial k} z_i &= -a(\phi(t_i^2, \gamma_t) - \phi(t_i^1, \gamma_t))) \\
\frac{\partial}{\partial \gamma_x} z_i &= a(\phi'(x_i^2, \gamma_x) - \phi'(x_i^1, \gamma_x)) \\
\frac{\partial}{\partial \gamma_t} z_i &= -ak(\phi'(t_i^2, \gamma_t) - \phi'(t_i^1, \gamma_t))
\end{aligned}
$$

Note constraints:

- $0 < a < \infty$
- $0 < k < \infty$
- $0 < \gamma_x < \infty$
- $0 < \gamma_t < \infty$

2.5 Gradient Calculations for Standard Exponential

$$
\begin{aligned}
\theta &= (a, \delta) \\
z_i &= a(x_i^2 \delta^{t_i^2} - x_i^1 \delta^{t_i^1}) \\
\frac{\partial}{\partial a} z_i &= x_i^2 \delta^{t_i^2} - x_i^1 \delta^{t_i^1} \\
\frac{\partial}{\partial \delta} z_i &= a(x_i^2 t_i^2 \delta^{t_i^2 - 1} - x_i^1 t_i^1 \delta^{t_i^1 - 1})
\end{aligned}
$$

Note constraints:

- $0 < a < \infty$
- $0 < \delta < 1$

2.6 Gradient Calculations for Standard Exponential with Intercept

$$\begin{aligned}
\theta &= (a, \delta, \beta_0) \\
z_i &= \beta_0 + a(x_i^2 \delta^{t_i^2} - x_i^1 \delta^{t_i^1}) \\
\frac{\partial}{\partial a} z_i &= x_i^2 \delta^{t_i^2} - x_i^1 \delta^{t_i^1} \\
\frac{\partial}{\partial \delta} z_i &= a(x_i^2 t_i^2 \delta^{t_i^2 - 1} - x_i^1 t_i^1 \delta^{t_i^1 - 1}) \\
\frac{\partial}{\partial \beta_0} z_i &= 1
\end{aligned}$$

Note constraints:

- $0 < a < \infty$
- $0 < \delta < 1$

2.7 Gradient Calculations for Homothetic Exponential

$$\begin{aligned}
\theta &= (a, \delta) \\
z_i &= a(\log(x_i^2 \delta^{t_i^2}) - \log(x_i^1 \delta^{t_i^1})) \\
\frac{\partial}{\partial a} z_i &= (\log(x_i^2 \delta^{t_i^2}) - \log(x_i^1 \delta^{t_i^1})) \\
\frac{\partial}{\partial \delta} z_i &= a(\frac{1}{x_i^2 \delta^{t_i^2}} x_i^2 t_i^2 \delta^{t_i^2 - 1} - \frac{1}{x_i^1 \delta^{t_i^1}} x_i^1 t_i^1 \delta^{t_i^1 - 1}) \\
&= a(t_i^2 \delta^{-1} - t_i^1 \delta^{-1})
\end{aligned}$$

Note constraints:

- $0 < a < \infty$
- $0 < \delta < 1$

2.8 Gradient Calculations for Standard Hyperbolic

$$
\begin{aligned}
\theta &= (a, \alpha) \\
z_i &= a(x_i^2(1+\alpha t_i^2)^{-1} - x_i^1(1+\alpha t_i^1)^{-1}) \\
\frac{\partial}{\partial a} z_i &= (x_i^2(1+\alpha t_i^2)^{-1} - x_i^1(1+\alpha t_i^1)^{-1}) \\
\frac{\partial}{\partial \alpha} z_i &= a(x_i^2(-1)(1+\alpha t_i^2)^{-2} t_i^2 - x_i^1(-1)(1+\alpha t_i^1)^{-2} t_i^1) \\
&= a(x_i^1(1+\alpha t_i^1)^{-2} t_i^1 - x_i^2(1+\alpha t_i^2)^{-2} t_i^2)
\end{aligned}
$$

Note constraints:

- $0 < a < \infty$
- $0 < \alpha < \infty$

2.9 Gradient Calculations for Homothetic Hyperbolic

$$
\begin{aligned}
z_i &= a(\log(x_i^2(1+\alpha t_i^2)^{-1}) - \log(x_i^1(1+\alpha t_i^1)^{-1})) \\
\theta &= (a, \alpha) \\
\frac{\partial}{\partial a} z_i &= (\log(x_i^2(1+\alpha t_i^2)^{-1}) - \log(x_i^1(1+\alpha t_i^1)^{-1})) \\
\frac{\partial}{\partial \alpha} z_i &= a(\frac{x_i^2(-1)(1+\alpha t_i^2)^{-2} t_i^2}{x_i^2(1+\alpha t_i^2)^{-1}} - \frac{x_i^1(-1)(1+\alpha t_i^1)^{-2} t_i^1}{x_i^1(1+\alpha t_i^1)^{-1}}) \\
&= a(t_i^1(1+\alpha t_i^1)^{-1} - t_i^2(1+\alpha t_i^2)^{-1})
\end{aligned}
$$

Note constraints:

- $0 < a < \infty$
- $0 < \alpha < \infty$

2.10 Gradient Calculations for Standard Hyperbolic with Intercept

$$\begin{aligned}
\theta &= (a, \alpha, \beta_0) \\
z_i &= \beta_0 + a(x_i^2(1+\alpha t_i^2)^{-1} - x_i^1(1+\alpha t_i^1)^{-1}) \\
\frac{\partial}{\partial a} z_i &= (x_i^2(1+\alpha t_i^2)^{-1} - x_i^1(1+\alpha t_i^1)^{-1}) \\
\frac{\partial}{\partial \alpha} z_i &= a(x_i^2(-1)(1+\alpha t_i^2)^{-2} t_i^2 - x_i^1(-1)(1+\alpha t_i^1)^{-2} t_i^1) \\
\frac{\partial}{\partial \beta_0} z_i &= 1
\end{aligned}$$

Note constraints:

- $0 < a < \infty$
- $0 < \alpha < \infty$

2.11 Gradient Calculations for Standard Hyperboloid

$$\begin{aligned}
\theta &= (a, \alpha, \mu) \\
z_i &= a(x_i^2(1+\alpha t_i^2)^{-\mu} - x_i^1(1+\alpha t_i^1)^{-\mu}) \\
\frac{\partial}{\partial a} z_i &= (x_i^2(1+\alpha t_i^2)^{-\mu} - x_i^1(1+\alpha t_i^1)^{-\mu}) \\
\frac{\partial}{\partial \alpha} z_i &= a(x_i^2(-\mu)(1+\alpha t_i^2)^{-\mu-1} t_i^2 - x_i^1(-\mu)(1+\alpha t_i^1)^{-\mu-1} t_i^1) \\
\frac{\partial}{\partial \mu} z_i &= a(-x_i^2(1+\alpha t_i^2)^{-\mu} \log(1+\alpha t_i^2) - (-x_i^1)(1+\alpha t_i^1)^{-\mu} \log(1+\alpha t_i^1))
\end{aligned}$$

Note constraints:

- $0 < a < \infty$
- $0 < \alpha < \infty$
- $0 < \mu < \infty$

2.12 Gradient Calculations for Homothetic Hyperboloid

$$
\begin{aligned}
\theta &= (a, \alpha, \mu) \\
z_i &= a(\log(x_i^2(1+\alpha t_i^2)^{-\mu}) - \log(x_i^1(1+\alpha t_i^1)^{-\mu})) \\
\frac{\partial}{\partial a} z_i &= \log(x_i^2(1+\alpha t_i^2)^{-\mu}) - \log(x_i^1(1+\alpha t_i^1)^{-\mu}) \\
\frac{\partial}{\partial \alpha} z_i &= a(\frac{1}{x_i^2(1+\alpha t_i^2)^{-\mu}} x_i^2(-\mu)(1+\alpha t_i^2)^{-\mu-1} t_i^2 - \frac{1}{x_i^1(1+\alpha t_i^1)^{-\mu}} x_i^1(-\mu)(1+\alpha t_i^1)^{-\mu-1} t_i^1) \\
&= a((-\mu)(1+\alpha t_i^2)^{-1} t_i^2 - (-\mu)(1+\alpha t_i^1)^{-1} t_i^1) \\
&= a\mu((1+\alpha t_i^1)^{-1} t_i^1 - (1+\alpha t_i^2)^{-1} t_i^2) \\
\frac{\partial}{\partial \mu} z_i &= a(\frac{1}{x_i^2(1+\alpha t_i^2)^{-\mu}}(-x_i^2)(1+\alpha t_i^2)^{-\mu}\log(1+\alpha t_i^2) - \frac{1}{x_i^1(1+\alpha t_i^1)^{-\mu}}(-x_i^1)(1+\alpha t_i^1)^{-\mu}\log(1+\alpha t \\
&= a(-\log(1+\alpha t_i^2) - (-\log(1+\alpha t_i^1))) \\
&= a(\log(1+\alpha t_i^1) - \log(1+\alpha t_i^2))
\end{aligned}
$$

Note constraints:

- $0 < a < \infty$
- $0 < \alpha < \infty$
- $0 < \mu < \infty$

2.13 Gradient Calculations for Standard Hyperboloid with Intercept

$$
\begin{aligned}
\theta &= (a, \alpha, \mu, \beta_0) \\
z_i &= a(x_i^2(1+\alpha t_i^2)^{-\mu} - x_i^1(1+\alpha t_i^1)^{-\mu}) \\
\frac{\partial}{\partial a} z_i &= (x_i^2(1+\alpha t_i^2)^{-\mu} - x_i^1(1+\alpha t_i^1)^{-\mu}) \\
\frac{\partial}{\partial \alpha} z_i &= a(x_i^2(-\mu)(1+\alpha t_i^2)^{-\mu-1} t_i^2 - x_i^1(-\mu)(1+\alpha t_i^1)^{-\mu-1} t_i^1) \\
\frac{\partial}{\partial \mu} z_i &= a(-x_i^2(1+\alpha t_i^2)^{-\mu}\log(1+\alpha t_i^2) - (-x_i^1)(1+\alpha t_i^1)^{-\mu}\log(1+\alpha t_i^1)) \\
\frac{\partial}{\partial \beta_0} z_i &= 1
\end{aligned}
$$

Note constraints:

- $0 < a < \infty$
- $0 < \alpha < \infty$
- $0 < \mu < \infty$

2.14 Gradient Calculations for Standard Quasi-Hyperbolic

$$\begin{aligned}
\theta &= (a, \beta, \delta) \\
z_i &= a(x_i^2 \beta^{I(t_2>0)} \delta^{t_i^2} - x_i^1 \beta^{I(t_1>0)} \delta^{t_i^1}) \\
\frac{\partial}{\partial a} z_i &= x_i^2 \beta^{I(t_2>0)} \delta^{t_i^2} - x_i^1 \beta^{I(t_1>0)} \delta^{t_i^1} \\
\frac{\partial}{\partial \beta} z_i &= a(x_i^2 I(t_2 > 0) \delta^{t_i^2} - x_i^1 I(t_1 > 0) \delta^{t_i^1}) \\
\frac{\partial}{\partial \delta} z_i &= a(x_i^2 \beta^{I(t_2>0)} t_i^2 \delta^{t_i^2 - 1} - x_i^1 \beta^{I(t_1>0)} t_i^1 \delta^{t_i^1 - 1})
\end{aligned}$$

Note constraints:

- $0 < a < \infty$
- $0 < \beta < 1$
- $0 < \delta < 1$

2.15 Gradient Calculations for Homothetic Quasi-Hyperboloid

$$
\begin{aligned}
\theta &= (a, \beta, \delta) \\
z_i &= a(\log(x_i^2 \beta^{I(t_2>0)} \delta^{t_i^2}) - \log(x_i^1 \beta^{I(t_1>0)} \delta^{t_i^1})) \\
\frac{\partial}{\partial a} z_i &= \log(x_i^2 \beta^{I(t_2>0)} \delta^{t_i^2}) - \log(x_i^1 \beta^{I(t_1>0)} \delta^{t_i^1}) \\
\frac{\partial}{\partial \beta} z_i &= a(\frac{1}{x_i^2 \beta^{I(t_2>0)} \delta^{t_i^2}} x_i^2 I(t_2 > 0) \delta^{t_i^2} - \frac{1}{x_i^1 \beta^{I(t_1>0)} \delta^{t_i^1}} x_i^1 I(t_1 > 0) \delta^{t_i^1}) \\
&= a(\frac{I(t_2 > 0)}{\beta^{I(t_2>0)}} - \frac{I(t_1 > 0)}{\beta^{I(t_1>0)}}) \\
\frac{\partial}{\partial \delta} z_i &= a(\frac{1}{x_i^2 \beta^{I(t_2>0)} \delta^{t_i^2}} x_i^2 \beta^{I(t_2>0)} t_i^2 \delta^{t_i^2 - 1} - \frac{1}{x_i^1 \beta^{I(t_1>0)} \delta^{t_i^1}} x_i^1 \beta^{I(t_1>0)} t_i^1 \delta^{t_i^1 - 1}) \\
&= a(t_i^2 \delta^{-1} - t_i^1 \delta^{-1})
\end{aligned}
$$

Note constraints:

- $0 < a < \infty$
- $0 < \beta < 1$
- $0 < \delta < 1$

2.16 Gradient Calculations for Standard Quasi-Hyperboloid with Intercept

$$
\begin{aligned}
\theta &= (a, \beta, \delta, \beta_0) \\
z_i &= \beta_0 + a(x_i^2 \beta^{I(t_2>0)} \delta^{t_i^2} - x_i^1 \beta^{I(t_1>0)} \delta^{t_i^1}) \\
\frac{\partial}{\partial a} z_i &= x_i^2 \beta^{I(t_2>0)} \delta^{t_i^2} - x_i^1 \beta^{I(t_1>0)} \delta^{t_i^1} \\
\frac{\partial}{\partial \beta} z_i &= a(x_i^2 I(t_2 > 0) \delta^{t_i^2} - x_i^1 I(t_1 > 0) \delta^{t_i^1}) \\
\frac{\partial}{\partial \delta} z_i &= a(x_i^2 \beta^{I(t_2>0)} t_i^2 \delta^{t_i^2 - 1} - x_i^1 \beta^{I(t_1>0)} t_i^1 \delta^{t_i^1 - 1}) \\
\frac{\partial}{\partial \beta_0} z_i &= 1
\end{aligned}
$$

Note constraints:

- $0 < a < \infty$
- $0 < \beta < 1$
- $0 < \delta < 1$

2.17 Gradient Calculations for Standard System-2

$$\begin{aligned}
\theta &= (a, \omega, \delta_1, \delta_2) \\
z_i &= a(x_i^2(\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2}) - x_i^1(\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1})) \\
\frac{\partial}{\partial a} z_i &= (x_i^2(\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2}) - x_i^1(\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1})) \\
\frac{\partial}{\partial \omega} z_i &= a(x_i^2(\delta_1^{t_i^2} - \delta_2^{t_i^2}) - x_i^1(\delta_1^{t_i^1} - \delta_2^{t_i^1})) \\
\frac{\partial}{\partial \delta_1} z_i &= a(x_i^2 \omega t_i^2 \delta_1^{t_i^2 - 1} - x_i^1 \omega t_i^1 \delta_1^{t_i^1 - 1}) \\
\frac{\partial}{\partial \delta_2} z_i &= a(x_i^2 (1-\omega) t_i^2 \delta_2^{t_i^2 - 1} - x_i^1 (1-\omega) t_i^1 \delta_2^{t_i^1 - 1})
\end{aligned}$$

Note constraints:

- $0 < a < \infty$
- $0 < \delta_1 < 1$
- $0 < \delta_2 < 1$
- $0 < \omega < 1$

2.18 Gradient Calculations for Homothetic System-2

$$
\begin{aligned}
\theta &= (a, \omega, \delta_1, \delta_2) \\
z_i &= a(\log(x_i^2(\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2})) - \log(x_i^1(\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1}))) \\
\frac{\partial}{\partial a} z_i &= (\log(x_i^2(\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2})) - \log(x_i^1(\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1}))) \\
\frac{\partial}{\partial \omega} z_i &= a(\frac{x_i^2(\delta_1^{t_i^2} - \delta_2^{t_i^2})}{x_i^2(\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2})} - \frac{x_i^1(\delta_1^{t_i^1} - \delta_2^{t_i^1})}{x_i^1(\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1})}) \\
&= a(\frac{\delta_1^{t_i^2} - \delta_2^{t_i^2}}{\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2}} - \frac{\delta_1^{t_i^1} - \delta_2^{t_i^1}}{\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1}}) \\
\frac{\partial}{\partial \delta_1} z_i &= a(\frac{x_i^2 \omega t_i^2 \delta_1^{t_i^2 - 1}}{x_i^2(\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2})} - \frac{x_i^1 \omega t_i^1 \delta_1^{t_i^1 - 1}}{x_i^1(\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1})}) \\
&= a(\frac{\omega t_i^2 \delta_1^{t_i^2 - 1}}{\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2}} - \frac{\omega t_i^1 \delta_1^{t_i^1 - 1}}{\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1}}) \\
\frac{\partial}{\partial \delta_2} z_i &= a(\frac{x_i^2 (1-\omega) t_i^2 \delta_2^{t_i^2 - 1}}{x_i^2(\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2})} - \frac{x_i^1 (1-\omega) t_i^1 \delta_2^{t_i^1 - 1}}{x_i^1(\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1})}) \\
&= a(\frac{(1-\omega) t_i^2 \delta_2^{t_i^2 - 1}}{\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2}} - \frac{(1-\omega) t_i^1 \delta_2^{t_i^1 - 1}}{\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1}})
\end{aligned}
$$

Note constraints:

- $0 < a < \infty$
- $0 < \delta_1 < 1$
- $0 < \delta_2 < 1$
- $0 < \omega < 1$

2.19 Gradient Calculations for Standard System-2 with Intercept

$$
\begin{aligned}
\theta &= (a, \omega, \delta_1, \delta_2, \beta_0) \\
z_i &= \beta_0 + a(x_i^2(\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2}) - x_i^1(\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1})) \\
\frac{\partial}{\partial a} z_i &= (x_i^2(\omega\delta_1^{t_i^2} + (1-\omega)\delta_2^{t_i^2}) - x_i^1(\omega\delta_1^{t_i^1} + (1-\omega)\delta_2^{t_i^1})) \\
\frac{\partial}{\partial \omega} z_i &= a(x_i^2(\delta_1^{t_i^2} - \delta_2^{t_i^2}) - x_i^1(\delta_1^{t_i^1} - \delta_2^{t_i^1})) \\
\frac{\partial}{\partial \delta_1} z_i &= a(x_i^2\omega t_i^2\delta_1^{t_i^2-1} - x_i^1\omega t_i^1\delta_1^{t_i^1-1}) \\
\frac{\partial}{\partial \delta_2} z_i &= a(x_i^2(1-\omega)t_i^2\delta_2^{t_i^2-1} - x_i^1(1-\omega)t_i^1\delta_2^{t_i^1-1}) \\
\frac{\partial}{\partial \beta_0} z_i &= 1
\end{aligned}
$$

Note constraints:

- $0 < a < \infty$
- $0 < \delta_1 < 1$
- $0 < \delta_2 < 1$
- $0 < \omega < 1$

3 Handling Constraints

Note that the results shown above all apply to the raw model specification, which involves constraints that will necessitate the use of a constrained optimization routine. It is often convenient to be able to use an unconstrained optimization routine instead by assuming that optima never occur near the boundaries of the open sets used to constrain the model parameters.

The notes above show that only two types of constraints are needed for the models under consideration: constraints to $(0, \infty)$ and to $(0, 1)$. We can map unconstrained parameters to these constrained spaces using $\exp(\theta_j)$ to map

$(-\infty, \infty)$ to $(0, \infty)$ and by using $L(\theta) = (1 + \exp(-\theta_j))^{-1}$ to map $(-\infty, \infty)$ to $(0, 1)$. When calculating the gradients of the negative log likelihood under these unconstrained reparameterizations of the models, we need to introduce an additional term of $\exp(\theta_j)$ to handle the first mapping and an additional term of $L(\theta_j)(1 - L(\theta_j))$ to handle the second mapping.