1 Derivations in Section 3.2

For $X_c = \{x_i\}_{i \in c} \sim \text{id} p(x|\beta, \theta)$, we have

$$p(\bar{t}_c|\beta, \theta) = \exp \left\{ \beta|c|\bar{t}_c, \theta) - \beta|c|\psi(\theta) - \sum_{i \in c} h_\beta(x_i) \right\}. \quad (1)$$

For notational simplicity, we let $\bar{t}_c = y$ from now. By the normalization property,

$$\beta|c|\psi(\theta) = \log \int \exp \left\{ \beta|c|\langle y, \theta) - \sum_{i \in c} h_\beta(x_i) \right\} dy. \quad (2)$$

Differentiating both sides by $\theta$ yields

$$\beta|c| \frac{d\psi(\theta)}{d\theta} = \int \beta|c| y \cdot p(y|\beta, \theta) dy, \quad \frac{d\psi(\theta)}{d\theta} = \mathbb{E}[y]. \quad (3)$$

Also, we have

$$\beta|c| \frac{\partial^2 \psi(\theta)}{\partial \theta_j \partial \theta_k} = \int \beta|c| y_j p(y|\beta, \theta) \left( \beta|c| y_j - \beta|c| \frac{\partial \psi(\theta)}{\partial \theta_k} \right) dy_j$$

$$= \beta^2 |c|^2 \mathbb{E}[y_j y_k] - \beta^2 |c|^2 \mathbb{E}[y_j] \mathbb{E}[y_k] = \beta^2 |c|^2 \text{cov}(y_j, y_k). \quad (4)$$

Hence,

$$\frac{1}{\beta|c|} \frac{\partial^2 \psi(\theta)}{\partial \theta_j \partial \theta_k} = \text{cov}(y_j, y_k) = \int (y_j - \mathbb{E}[y_j])(y_k - \mathbb{E}[y_k])p(y|\beta, \theta) dy_j dy_k. \quad (5)$$

Differentiating this again yields

$$\frac{1}{\beta|c|} \frac{\partial^3 \psi(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} = \int (y_j - \mathbb{E}[y_j])(y_k - \mathbb{E}[y_k])(y_l - \mathbb{E}[y_l])p(y|\beta, \theta) dy_j dy_k dy_l$$

$$= \mathbb{E}[(y_j - \mathbb{E}[y_j])(y_k - \mathbb{E}[y_k])(y_l - \mathbb{E}[y_l])]. \quad (6)$$

Unfortunately, this relationship does not continue after the third order; the fourth derivative of $\psi(\theta)$ is not exactly match to the fourth order central moment of $y$. However, one can easily maintain the $m$th order central moment by manipulating the $m$th order derivative of $\psi(\theta)$, and $m$th order central moment always have the constant term $(\beta|c|)^{-m}$.

Equation (40) of the paper is a simple consequence of the equation (6). To prove the equation (41) of the paper, we use the following relationship:

$$\mathbb{E} \left[ \frac{\epsilon_\phi(x_c)}{\Delta_\phi(x_c)} \right] \approx \frac{\mathbb{E}[\epsilon_\phi(x_c)]}{\mathbb{E}[\Delta_\phi(x_c)]} - \frac{\text{cov}(\epsilon_\phi(x_c), \Delta_\phi(x_c))}{\mathbb{E}[\Delta_\phi(x_c)]^2} + \frac{\mathbb{E}[\epsilon_\phi(x_c)] \text{var}[\Delta_\phi(x_c)]}{\mathbb{E}[\Delta_\phi(x_c)]^3}. \quad (7)$$

Now it is easy to show that this equation converges to zero when $\beta \to 0$; all the expectations and variances can be obtained by differentiating $\psi(\theta)$ for as many times as needed.