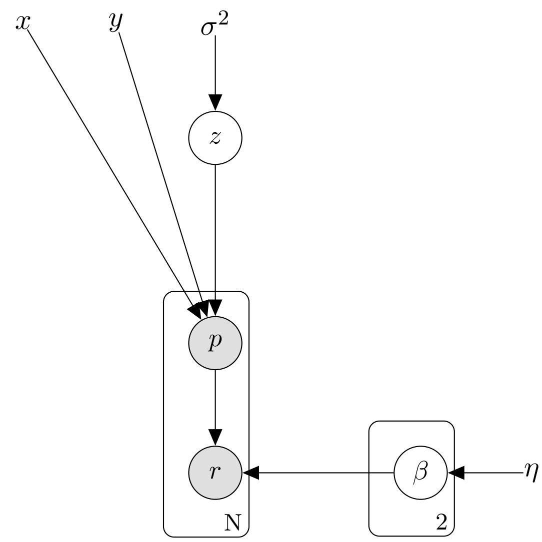

0.1 model

$$
\begin{aligned}
z &\sim \mathcal{N}((0, \sigma^2) \\
p(p=1) &= \frac{1}{1+\exp|XZY^T|} \\
\beta &\sim Beta(\eta) \\
r &\sim Bern(\beta_p)
\end{aligned}
$$

$$
p(R, P, Z, \beta | X, Y, \sigma^2, \eta) = p(R|P, \beta) p(P|X, Y, Z) p(Z|\sigma^2) p(\beta|\eta)
$$

Where:

$$
p(R|P, \beta) = [\beta_1^R (1-\beta_1)^{1-R}]^P [\beta_0^R (1-\beta_0)^{1-R}]^{1-P} \tag{1}
$$

$$
\begin{aligned}
p(P|X, Y, Z) &= [\frac{1}{1+exp(-XZY^T)}]^P [1 - \frac{1}{1+exp(-XZY^T)}]^{1-P} \\
&= e^{XZY^T * P} * \frac{exp(-XZY^T)}{1+exp(-XZY^T)} \\
&\geq e^{XZY^T * P} * \sigma(\xi) * exp\{-\frac{XZY^T + \xi}{2} - \lambda(\xi)((XZY^T)^2 - \xi^2)\}
\end{aligned}
$$

where $\lambda(\xi) = \frac{1}{2\xi}[\sigma(\xi) - \frac{1}{2}]$

$$
\begin{aligned}
p(Z|\sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{Z^2}{2\sigma^2}) \\
p(\beta_1|\eta_1) &= Beta(\beta_1; \eta_{10}, \eta_{11}) \\
p(\beta_0|\eta_0) &= Beta(\beta_0; \eta_{00}, \eta_{01})
\end{aligned}
$$

For the convenience of calculation, the log of them:

$$\ln p(R|P,\beta) = P[R\ln\beta_1 + (1-R)\ln(1-\beta_1)] + (1-P)[R\ln\beta_0 + (1-R)\ln(1-\beta_0)] \tag{2}$$

$$\ln p(P|X,Y,Z) = XZY^T * P + \{-\frac{XZY^T + \xi}{2} - \lambda(\xi)((XZY^T)^2 - \xi^2)\} + \ln\sigma(\xi) \tag{3}$$

$$\ln p(Z|\sigma^2) = -\frac{Z^2}{\sigma^2} - \ln\sigma \tag{4}$$

Assume:

$$q(P, Z, \beta|R, X, Y, \sigma^2, \eta) = q(P|\lambda)q(Z|\mu, v)q(\beta|\rho) \tag{5}$$

Then our goal is to maximize ELBO

$$E_{q(Z,\beta)}[\ln p(R,P,Z,\beta)] - E_{q(Z,\beta)}[\ln q(Z,\beta)] + E_{q(Z,\beta,P)}[\ln P(R,P,Z,\beta)] - E_{q(Z,\beta,P)}[\ln q(Z,\beta,P)] \tag{6}$$

We defining s_1 as a subscript set with 1 in R, and defining s_2 as a subscript set with 0 in R.

0.1.1 Derivation of $\ln q(Z)$

$$\ln q(Z) \quad = \quad \sum_{(R,P)\in s_1} E_\beta[\ln p(R,R,Z,\beta)] + \sum_{(R)\in s_2} E_{\beta,P}[\ln p(R,P,Z,\beta) + const$$

$$E_{\beta,P}[\ln p(R,P,Z,\beta) = E_{\beta,P}[\ln p(P|X,Y,Z)] + E_{\beta,P}[\ln p(Z)] + const \quad (7)$$

$$E_{\beta,P}[\ln p(P|X,Y,Z)] = XZY^T * \lambda - \frac{XZY^T + \xi}{2} - \lambda(\xi)[(XZY^T)^2 - \xi^2] \quad (8)$$

$$E_{\beta,P}[\ln p(R,P,Z,\beta) = XZY^T * \lambda - \frac{XZY^T}{2} - \lambda(\xi)(XZY^T)^2 - \frac{Z^2}{2\sigma^2} + const \quad (9)$$

$$\begin{aligned} E_\beta[\ln p(R,P,Z,\beta)] &= E_\beta[\ln p(p|X,Y,Z)] + E_\beta[\ln p(Z)] + const \\ &= XZY^T - \frac{XZY^T}{2} - \lambda(\xi)(XZY^T)^2 - \frac{Z^2}{2\sigma^2} + const \end{aligned}$$

$$\ln q(Z_{mn}) = [\sum_{(i,j)\in s_2}[(\lambda_{i,j} - \frac{1}{2})x_{im} * (y^T)_{nj}] + \sum_{(i,j)\in s_1} \frac{1}{2} x_{im} * (y^T)_{nj}]Z_{mn} - (\sum_{i,j} \lambda(\xi_{ij}) * x_{im}^2 * {(y^T)_{nj}}^2 + \frac{1}{\sigma^2}) * Z_{mn}^2 \quad (10)$$

$Z_{mn} \sim N(\mu_{mn}, v_{mn})$,so:

$$v_{mn} = \frac{1}{\sqrt{[\sum_{i,j} \lambda(\xi_{ij}) x_{im}^2 {y_{jn}}^2 + \frac{1}{\sigma^2}] * 2}} \quad (11)$$

$$\mu_{mn} = \frac{\sum_{(i,j)\in s_2}[(\lambda_{i,j} - \frac{1}{2})x_{im} * y_{jn}] + \sum_{(i,j)\in s_1} \frac{1}{2} x_{im} * y_{jn}}{2 * (\sum_i \sum_j \lambda(\xi_{ij}) x_{im}^2 {y_{jn}}^2 + \frac{1}{\sigma^2})} \quad (12)$$

0.1.2 Derivation of $\ln q(\beta)$

$$\ln q(\beta) \quad = \quad \sum_{(R,R)\in s_1} E_Z \ln p(R,P,Z,\beta) + \sum_{(R)\in s_2} E_{Z,P} \ln p(R,P,Z,\beta) + const$$

Only $E_Z[\ln p(R|P,\beta)]$, $E_{Z,P}[\ln p(R|P,\beta)]$, $2 * E_{Z,P}[\ln p(\beta|\eta)]$ contain β:

$$\begin{aligned} E_{Z,P}[\ln p(R|P,\beta)] &= \lambda[R \ln \beta_1 + (1-R)\ln(1-\beta_1)] \\ &+ (1-\lambda)[R \ln \beta_0 + (1-R)\ln(1-\beta_0)] \end{aligned}$$

$$
\begin{aligned}
E_{Z,P}[\ln p(\beta|\eta)] &= \ln\Gamma(\eta_{10}+\eta_{11}) - \ln\Gamma(\eta_{10}) - \ln\Gamma(\eta_{11}) \\
&+ (\eta_{10}-1)\ln\beta_1 + (\eta_{11}-1)\ln(1-\beta_1) \\
&+ \ln\Gamma(\eta_{00}+\eta_{01}) - \ln\Gamma(\eta_{00}) - \ln\Gamma(\eta_{01}) \\
&+ (\eta_{00}-1)\ln\beta_0 + (\eta_{01}-1)\ln(1-\beta_0)
\end{aligned}
$$

$$
\begin{aligned}
E_Z[\ln p(R|P,\beta)] &= P[R\ln\beta_1 + (1-R)\ln(1-\beta_1)] \\
&+ (1-P)[R\ln\beta_0 + (1-R)\ln(1-\beta_0)]
\end{aligned}
$$

$$
\begin{aligned}
\ln q(\beta) &= \sum_{(R)\in s_2} \lambda[R\ln\beta_1 + (1-R)\ln(1-\beta_1)] + (1-\lambda)[R\ln\beta_0 + (1-R)\ln(1-\beta_0)] \\
&+ \sum_{(R,P)\in s_1} P[R\ln\beta_1 + (1-R)\ln(1-\beta_1)] + (1-P)[R\ln\beta_0 + (1-R)\ln(1-\beta_0)] \\
&+ [\ln\Gamma(\eta_{10}+\eta_{11}) - \ln\Gamma(\eta_{10}) - \ln\Gamma(\eta_{11}) \\
&+ (\eta_{10}-1)\ln\beta_1 + (\eta_{11}-1)\ln(1-\beta_1) \\
&+ \ln\Gamma(\eta_{00}+\eta_{01}) - \ln\Gamma(\eta_{00}) - \ln\Gamma(\eta_{01}) \\
&+ (\eta_{00}-1)\ln\beta_0 + (\eta_{01}-1)\ln(1-\beta_0)]
\end{aligned}
$$

Remove irrelevant items:

$$
\begin{aligned}
\ln q(\beta) &= (\sum_{(R)\in s_2} \lambda R + \sum_{(R,P)\in s_1} PR + \eta_{10} - 1)\ln\beta_1 \\
&+ [\sum_{(R)\in s_2} \lambda(1-R) + \sum_{(R,P)\in s_1} P(1-R) + \eta_{11} - 1]\ln(1-\beta_1) \\
&+ [\sum_{(R)\in s_2} (1-\lambda)R + \sum_{(R,P)\in s_1} (1-P)R + \eta_{00} - 1]\ln\beta_0 \\
&+ [\sum_{(R)\in s_2} (1-\lambda)(1-R) + \sum_{(R,P)\in s_1} (1-P)(1-R) + \eta_{01} - 1]\ln(1-\beta_0)
\end{aligned}
$$

so we get:

$$
\rho_{00} = \sum_{(R)\in s_2} (1-\lambda)R + \sum_{(R,P)\in s_1} (1-P)R + \eta_{00} \qquad (13)
$$

$$\rho_{01} = \sum_{(R)\in s_2}(1-\lambda)(1-R) + \sum_{(R,P)\in s_1}(1-P)(1-R) + \eta_{01} \tag{14}$$

$$\rho_{10} = \sum_{(R)\in s_2}\lambda R + \sum_{(R,P)\in s_1} PR + \eta_{10} \tag{15}$$

$$\rho_{11} = \sum_{(R)\in s_2}\lambda(1-R) + \sum_{(R,P)\in s_1} P(1-R) + \eta_{11} \tag{16}$$

$$\beta_0 \sim Beta(\rho_{00}, \rho_{01}) \tag{17}$$

$$\beta_1 \sim Beta(\rho_{10}, \rho_{11}) \tag{18}$$

0.1.3 Derivation of $\ln q(P)$

$$\begin{aligned}
\ln q(P) &= E_{Z,\beta}(\ln p(R,P,Z,\beta) \\
&= E_{Z,\beta}[\ln p(R|P,\beta)] + E_{Z,\beta}[\ln p(\beta|\eta)] + E_{Z,\beta}[\ln p(P|X,Y,Z)] + E_{Z,\beta}[\ln p(Z)] \\
&= E_{Z,\beta}[\ln p(R|P,\beta)] + E_{Z,\beta}[\ln p(P|X,Y,Z)]
\end{aligned}$$

Only $E_{Z,\beta}[\ln p(R|P,\beta)], E_{Z,\beta}[\ln p(P|X,Y,Z)]$ contain P:

$$E_{Z,\beta}[\ln p(R|P,\beta)] = P*E_{\beta_1}[R\ln\beta_1+(1-R)\ln(1-\beta_1)]+(1-P)*E_{\beta_0}[R\ln\beta_0+(1-R)\ln(1-\beta_0)] \tag{19}$$

$$E_{Z,\beta}[\ln p(P|X,Y,Z)] = P*E_Z(XZY^T) \tag{20}$$

Because $E(\ln\beta) = \psi(\alpha) - \psi(\alpha+\beta)$,$E(\ln(1-\beta)) = \psi(\beta) - \psi(\alpha+\beta)$,so we get $\ln(q(R))$:

$$\begin{aligned}
\ln q(P) &= P*[R*\psi(\rho_{10}) + (1-R)*\psi(\rho_{11}) - \psi(\rho_{10}+\rho_{11})] \\
&+ (1-P)*[R*\psi(\rho_{00}) + (1-R)*\psi(\rho_{01}) - \psi(\rho_{00}+\rho_{01})] + P*X\mu_z Y \\
&= P*\ln[exp(R*\psi(\rho_{10}))*exp[(1-R)*\psi(\rho_{11})]*exp(-\phi(\rho_{10}+\rho_{11}))*exp(X\mu_z Y)] \\
&+ (1-P)*\ln[exp(R*\psi(\rho_{00}))*exp[(1-R)*\psi(\rho_{01})]*exp(-\phi(\rho_{00}+\rho_{01}))]
\end{aligned}$$

we define:

$$l_1 = exp(R*\psi(\rho_{10}))*exp[(1-R)*\psi(\rho_{11})]*exp(-\psi(\rho_{10}+\rho_{11}))*exp(X_i\mu_z(Y_j)^T)$$
$$l_2 = exp(R*\psi(\rho_{00}))*exp[(1-R)*\psi(\rho_{01})]*exp(-\psi(\rho_{00}+\rho_{01}))$$

So we get:

$$\lambda = \frac{l_1}{l_1+l_2} \tag{21}$$

Where $\frac{1}{exp(\psi(x))} \sim \frac{1}{x} + \frac{1}{2*x^2} + \frac{5}{4*3!*x^3} + \frac{3}{2*4!*x^4}$.

0.1.4 Derivation of ξ

For variational parameters ξ:

$$maxmize \ln \sigma(\xi) - \frac{\xi}{2} - \lambda(\xi)[(XZY^T)^2 - \xi^2] \tag{22}$$

$[\ln \sigma(\xi) - \frac{\xi}{2} - \lambda(\xi)[(XZY^T)^2 - \xi^2]]' = 0$

Then we get $\xi_{ij}^2 = (x_i z (y_j)^T)^2$