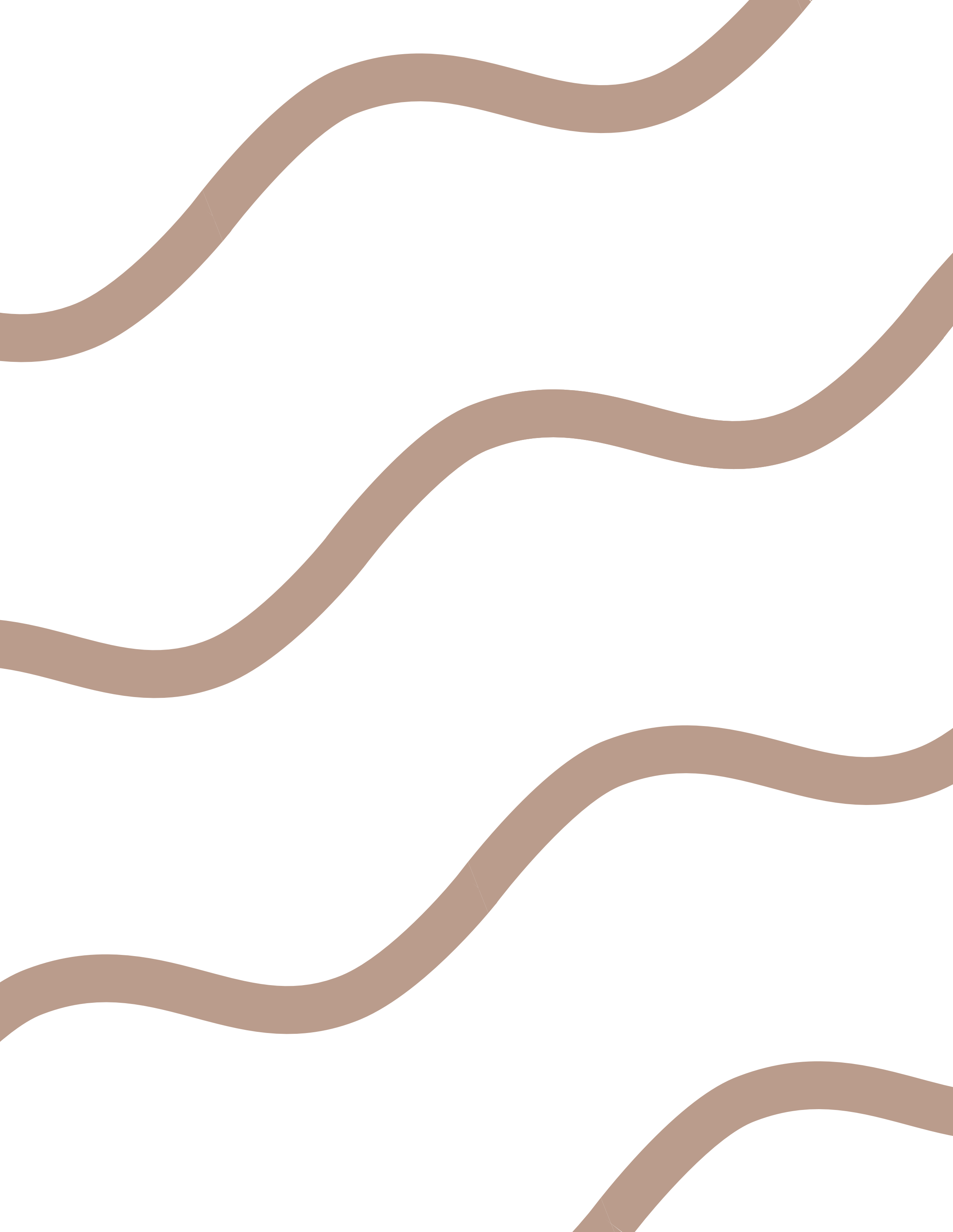

Hierarchical Model

(4) $\mu_k \sim \mathcal{N}(0, \theta^2)$ $\qquad k = 1, \ldots, K$

$$p(\mu_k) = \frac{1}{\sqrt{2\pi\theta^2}} e^{-\frac{\mu_k^2}{2\theta^2}}$$

(5) $c_i \sim \text{Categorical}\left(\frac{1}{K}, \ldots, \frac{1}{K}\right)$ $\qquad i = 1, \ldots, n$

(6) $x_i \mid c_i, \boldsymbol{\mu} \sim \mathcal{N}(\underbrace{c_i^T \boldsymbol{\mu}}_{= \mu_k}, 1)$ $\qquad i = 1, \ldots, n$

$$p(x_i \mid c_i, \boldsymbol{\mu}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \mu_k)^2}{2}}$$

$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$

$c_i = \underbrace{(0, \ldots, 1, \ldots, 0)}_{K\text{-length vector}}$ $\leftarrow$ k^{th} location

if data point x_i belongs to cluster k.

$c_i^T \boldsymbol{\mu} = \mu_k$

Data: $X_1, \dots, X_n$

Sampled in the following way:

(1) Assume μ known
(we've already sampled $\mu_k \sim \mathcal{N}(0, \sigma^2)$)

(2) Step 1: randomly select a cluster
$k = 1, \dots, K$ all equally likely

Assume our selection is cluster k.

(3) Step 2: Sample $X_i \sim \mathcal{N}(\mu_k, 1)$.

Joint Density:

$$p(\mu, c, x) = p(x \mid c, \mu)\, \underbrace{p(c, \mu)}^{\text{indep.}}$$

$$= \overbrace{p(x \mid c, \mu)}^{\text{ind.}}\, \overbrace{p(c)}^{\text{ind.}}\, p(\mu)$$

$$= \prod_{i=1}^{n} p(x_i \mid c_i, \mu)\, p(c_i)\, p(\mu)$$

Marginal is found by integrating a joint.

$$p(x) = \int p(x, y)\, dy$$

(8) $$p(\mathbf{x}) = \int_{\mathbf{c}} \int_{\boldsymbol{\mu}} p(\mathbf{x}, \boldsymbol{\mu}, \mathbf{c})\, d\mathbf{c}\, d\boldsymbol{\mu}$$

(10) $$q^*(\mathbf{z}) = \underset{q(z) \in \mathbb{Q}}{\arg\min}\ KL(q(z) \,\|\, p(z|x))$$

using (12)

$$= \underset{q(z) \in \mathbb{Q}}{\arg\min} \left[\mathbb{E}(\log q(z)) - \mathbb{E}(\log p(z, x)) + \log p(x) \right]$$

$$= \underset{q(z) \in \mathbb{Q}}{\arg\min} \left[\underbrace{\mathbb{E}(\log q(z)) - \mathbb{E}(\log p(z, x))}_{-ELBO(q)} \right]$$

$$= \underset{q(z) \in \mathbb{Q}}{\arg\max}\ ELBO(q)$$

Model:

$$\underbrace{p(z,x)}_{\text{joint}} = \overbrace{p(z)}^{\text{prior}}\,\overbrace{p(x|z)}^{\text{likelihood}}$$

on pg 2.

latent variable

$$(13)\quad ELBO(q) = \mathbb{E}[\log p(z,x)] - \mathbb{E}[\log q(z)]$$

$$= \mathbb{E}[\log p(z)p(x|z)] - \mathbb{E}[\log q(z)]$$

$$= \mathbb{E}[\log p(z)] + \underbrace{\mathbb{E}[\log p(x|z)]}_{\text{expected log likelihood.}} - \mathbb{E}[\log q(z)]$$

$$\mathbb{E}[\log p(\mu_k) ; m_k, s_k^2]$$

$\mu_k \sim \mathcal{N}(m_k, s_k^2)$

$$= \int q(\mu_k ; m_k s_k^2) \log p(\mu_k)\, d\mu_k$$

$$= \int \frac{1}{\sqrt{2\pi s_k^2}} e^{-\frac{(\mu_k - m_k)^2}{2 s_k^2}} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu_k^2}{2\sigma^2}}\right] d\mu_k$$

$$\varphi_{ik} \propto e^{\mathbb{E}(\mu_k) x_i - \mathbb{E}(\mu_k^2)/2}$$

$$\boldsymbol{\varphi}_i = \underbrace{(\varphi_{i1}, \varphi_{i2}, \ldots, \varphi_{iK})}$$

to represent a prob. dist.

need $\sum_{k=1}^{K} \varphi_{ik} = 1$

---

discrete dist:

$$\mathbf{p} = (p_1, p_2, p_3)$$

$$\frac{4}{8}, \frac{3}{8}, \frac{1}{8}$$

$$\tilde{\mathbf{p}} = \underbrace{(4, 3, 1)}_{\tilde{\mathbf{p}}}$$

$$\mathbf{p} = \frac{\tilde{\mathbf{p}}}{\mathrm{sum}(\tilde{\mathbf{p}})} \qquad \mathbf{p} \propto \tilde{\mathbf{p}}$$

$$\mathbb{E}\left[\log p(x_i \mid c_i, \mu) \; ; \; \varphi_i, m, s^2\right]$$

$$\rightarrow \; x_i \mid c_i, \mu \sim \mathcal{N}(c_i^T \mu, 1)$$

$$p(x_i \mid c_i, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - c_i^T\mu)^2}$$

$$= \prod_{k=1}^{K} \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \mu_k)^2}\right]^{c_{ik}}$$

$$c_i = (0, 0, \ldots, 1, \ldots, 0)$$

k^{th} location

$$c_i^T \mu = \mu_k$$

$q(c_i ; \varphi_i)$ = the k^{th} position $= 1$ w.p. φ_{ik}

$$q(\mu ; m, s^2) = \prod_{k=1}^{K} \underbrace{q(\mu_k ; m_k, s_k^2)}_{\text{Gaussian}}$$

$$\int_{c_i}\int_{\mu} q(c_i ; \varphi_i)\, q(\mu ; m, s^2) \sum_{k=1}^{K} c_{ik}\left(\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(x_i - \mu_k)^2\right)$$

$=$

$$\mathbb{E}\left[\log p(x_i \mid c_i, \mu)\right] = \mathbb{E}_{c_i} \mathbb{E}_{\mu}\left[\log p(x_i \mid c_i, \mu)\right]$$

$$= -\mathbb{E}_{\mu} \mathbb{E}_{c_i}\left[\sum_k c_{ik}\left(\tfrac{1}{2}(x_i - \mu_k)^2\right)\right] + \text{const.}$$

$$= -\mathbb{E}_{\mu}\left[\sum_k \varphi_{ik}\left(\tfrac{1}{2}(x_i - \mu_k)^2\right)\right] + \text{const}$$

$$\frac{1}{2}\mathbb{E}_{c_i}\left[\sum_k c_{ik}(x_i - \mu_k)^2\right] =$$

$$= \frac{1}{2}\sum_{k'} \varphi_{ik'} \sum_k \underbrace{\mathbb{1}\{c_{ik'} = 1 \cap c_{ik} = 0\ \text{when}\ k \neq k'\} c_{ik}(x_i - \mu_k)^2}$$

$$= \frac{1}{2}\sum_{k'} \varphi_{ik'}(x_i - \mu_{k'})^2$$

$$\sum_k c_{ik}\, \mathbb{1}\{c_{ik'} = 1 \text{ and } c_{ik} = 0 \text{ for } k \neq k'\}(x_i - \mu_k)^2$$

$= 0$ if $k \neq k'$
$= 1$ if $k = k'$

$$X = \begin{cases} 1 & p_1 \\ 2 & p_2 \\ 3 & p_3 \end{cases} \qquad \mathbb{E}(X) = 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3$$

$$\mathbb{1}\{A\} = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{if } A \text{ is false} \end{cases}$$

$$c_i = \begin{cases} (1,0,0) & \text{w.p. } \varphi_{i1} \\ (0,1,0) & \text{w.p. } \varphi_{i2} \\ (0,0,1) & \text{w.p. } \varphi_{i3} \end{cases} = q(c_i, \varphi_i)$$

$K = 3$

$$\mathbb{E}_{c_i}\left[\sum_{k=1}^{3} c_{ik}\, \Delta_k\right] = \mathbb{E}_{c_i}\left[c_{i1}\Delta_1 + c_{i2}\Delta_2 + c_{i3}\Delta_3\right]$$

$\Delta_k \rightarrow$ const. wrt c_i

$$= \varphi_{i1}\left[\cancel{c_{i1}}^{=1}\Delta_1\right] + \varphi_{i2}\left[\cancel{c_{i2}}^{=1}\Delta_2\right] + \varphi_{i3}\left[\cancel{c_{i3}}^{=1}\Delta_3\right]$$

$$= \sum_{k'=1}^{3} \varphi_{ik'} \underbrace{\Delta_{k'}}_{\Delta_k = (x_i - \mu_k)^2}$$

$$\frac{1}{2}\mathbb{E}_{c_i}\left[\sum_{k=1}^{K} c_{ik}(x_i - \mu_k)^2\right] = \frac{1}{2}\sum_{k'=1}^{K} \varphi_{ik'}(x_i - \mu_{k'})^2$$

$$\|$$

$$\frac{1}{2}\mathbb{E}_{c_i}\left[(x_i - c_i^T \mu)^2\right]$$

$$\boxed{\begin{gathered}\mu = (\mu_1, \ldots, \mu_K) \\ c_i = (c_{i1}, \ldots, c_{iK})\end{gathered}}$$

$$= \frac{1}{2}\left[\varphi_{i1}(x_i - \mu_1)^2 + \varphi_{i2}(x_i - \mu_2)^2 + \cdots + \varphi_{iK}(x_i - \mu_K)^2\right]$$

$$q(c_i; \varphi_i) = \begin{cases} (1,0,\ldots,0) & \varphi_{i1} \\ (0,1,\ldots,0) & \varphi_{i2} \\ \vdots & \\ (0,0,\ldots,1) & \varphi_{iK} \end{cases}$$

$$= \frac{1}{2}\sum_{k} \varphi_{ik}(x_i - \mu_k)^2$$

$$\mathbb{E}\left[\log p(x_i \mid c_i, \mu)\right] = \mathbb{E}_{\mu}\,\mathbb{E}_{c_i}\left[-\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}(x_i - c_i^T\mu)^2\right]$$

$$= -\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}\,\mathbb{E}_{\mu}\,\underbrace{\mathbb{E}_{c_i}\left[(x_i - c_i^T\mu)^2\right]}$$

$$= -\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}\,\mathbb{E}_{\mu}\left[\sum_{k=1}^{K}\varphi_{ik}(x_i - \mu_k)^2\right]$$

$$= -\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}\sum_{k=1}^{K}\varphi_{ik}\,\mathbb{E}_{\mu_k}\left[(x_i - \mu_k)^2\right]$$

$$= -\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}\sum_{k=1}^{K}\varphi_{ik}\int_{\mu_k} q(\mu_k; m_k, s_k^2)\,(x_i - \mu_k)^2\,d\mu_k$$

$$= -\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}\sum_{k=1}^{K}\varphi_{ik}\left[x_i^2 - 2x_i m_k + s_k^2 + m_k^2\right]$$

$$= -\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}x_i^2 + x_i\sum_{k}\varphi_{ik}m_k - \tfrac{1}{2}\sum_{k}\varphi_{ik}(s_k^2 + m_k^2)$$

$$\mathbb{E}_{c_i}\left[(x_i - c_i^T\mu)^2\right] = \varphi_{i1}(x_i - (1,\dots,0)^T\mu)^2$$
$$+ \varphi_{i2}(x_i - (0,1,\dots,0)^T\mu)^2$$
$$+ \dots$$
$$+ \varphi_{iK}(x_i - (0,\dots,0,1)^T\mu)^2$$
$$= \varphi_{i1}(x_i - \mu_1)^2 + \varphi_{i2}(x_i - \mu_2)^2$$
$$+ \dots + \varphi_{iK}(x_i - \mu_K)^2$$

$$c_i^T\mu = \sum_{k} c_{ik}\mu_k$$

$$= (c_{i1}, c_{i2}, \dots, c_{iK})\begin{pmatrix}\mu_1\\ \vdots \\ \mu_K\end{pmatrix}$$

$$\mathbb{E}_{\mu_k}\left[(x_i-\mu_k)^2\right] = \mathbb{E}_{\mu_k}\left[x_i^2-2x_i\mu_k+\mu_k^2\right]$$

$$= x_i^2 - 2x_i\,\mathbb{E}[\mu_k] + \mathbb{E}[\mu_k^2]$$

$$q(\mu_k; m_k, s_k^2) \quad = x_i^2 - 2x_i m_k + s_k^2 + m_k^2$$

$$\sim \mathcal{N}(m_k, s_k^2)$$

$$\mu_k \sim \mathcal{N}(\underset{\text{mean}}{m_k}, \underset{\text{variance}}{s_k^2})$$

$$\mathbb{E}[x] = \mu \qquad x \sim \mathcal{N}(\underset{\text{mean}}{\mu}, \underset{\text{variance}}{\sigma^2})$$

$$\mathbb{E}[x^2] = \sigma^2 + \mu^2$$

$$\mathrm{Var}(x) = \mathbb{E}x^2 - (\mathbb{E}x)^2$$

definition of variance.

$$\sum_{k=1}^{K} \varphi_{ik} x_i^2 = x_i^2 \sum_{k=1}^{K} \varphi_{ik} = x_i^2$$

$$(\varphi_{i1} x_i^2 + \varphi_{i2} x_i^2 + \dots + \varphi_{iK} x_i^2)$$

$$= x_i^2(\varphi_{i1} + \varphi_{i2} + \dots + \varphi_{iK})$$

$$-\frac{1}{2}\sum_{k=1}^{K} \varphi_{ik}\left(x_i^2 - 2x_i m_k + s_k^2 + m_k^2\right)$$

$$= -\frac{1}{2}\sum_k \varphi_{ik} x_i^2 + \sum_k x_i \varphi_{ik} m_k - \frac{1}{2}\sum_k \varphi_{ik}(s_k^2 + m_k^2)$$

$$= -\frac{1}{2} x_i^2 + x_i \sum_k \varphi_{ik} m_k - \frac{1}{2}\sum_k \varphi_{ik}(s_k^2 + m_k^2)$$

Discrete $\quad X = \begin{cases} 5 & \text{w.p } \varphi_1 \\ 10 & \text{w.p } \varphi_2 \\ +2 & \text{w.p } \varphi_3 \end{cases}$

$$\varphi_1 + \varphi_2 + \varphi_3 = 1$$

$$\mathbb{E}[\log X] = \varphi_1 \log(5) + \varphi_2 \log(10) + \varphi_3 \log(+2)$$

$c_i = \left\{ \begin{matrix} (1,0,0) & \text{wp} & \varphi_{i1} \\ (0,1,0) & & \varphi_{i2} \\ (0,0,1) & & \varphi_{i3} \end{matrix} \right\}$ $\quad c_i$ is random vector in $\mathbb{R}^3$

$\mathbb{E}[$

$$\mathbb{E}\left[c_{i1}\Delta_1 + c_{i2}\Delta_2 + c_{i3}\Delta_3\right] = \mathbb{E}\left[c_i^T\Delta\right]$$

$$\Delta = (\Delta_1, \Delta_2, \Delta_3)$$

$$c_i = (c_{i1}, c_{i2}, c_{i3}) = \begin{cases} (1,0,0) & \varphi_{i1} \\ (0,1,0) & \varphi_{i2} \\ (0,0,1) & \varphi_{i3} \end{cases}$$

$$= \varphi_{i1}\left(1\cdot\Delta_1 + \cancel{0\cdot\Delta_2} + \cancel{0\cdot\Delta_3}\right)$$
$$+ \varphi_{i2}\left(\cancel{0\cdot\Delta_1} + 1\cdot\Delta_2 + \cancel{0\cdot\Delta_3}\right)$$
$$+ \varphi_{i3}\left(\cancel{0\cdot\Delta_1} + \cancel{0\cdot\Delta_2} + 1\cdot\Delta_3\right)$$

$$= \varphi_{i1}\Delta_1 + \varphi_{i2}\Delta_2 + \varphi_{i3}\Delta_3$$

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x\,p(x)\,dx$$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\mathbb{E}[\text{const}] = \int_{-\infty}^{\infty} \text{const}\cdot p(x)\,dx = \text{const}\underbrace{\int_{-\infty}^{\infty} p(x)\,dx}_{1}$$

$$\hookrightarrow = \text{const}$$

$$\mathbb{E}_{c_i}\left[\log p(c_i)\right] = \mathbb{E}_{c_i}\left[-\log K\right] = -\log K.$$

$$\mathbb{E}\left((\gamma-\mu)^T \Sigma^{-1} (\gamma-\mu)\right) \quad ; \gamma \sim \mathcal{N}(m_y, S_y)$$

$$\boxed{\text{For a square matrix } A \in \mathbb{R}^{p \times p}, \text{ the quadratic form of } A \text{ is } x^T A x \quad \forall x \in \mathbb{R}^p.}$$

$$x^T A x = \sum_{i=1}^{p} \sum_{j=1}^{p} x_i x_j A_{ij}$$

$$\sum_{i=1}^{p} \sum_{j=1}^{p} \mathbb{E}\left((\Sigma^{-1})_{ij} (\gamma_i - \mu_i)(\gamma_j - \mu_j)\right)$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{p} (\Sigma^{-1})_{ij} \left[\mathbb{E}\gamma_i\gamma_j - \mu_i \mathbb{E}\gamma_j - \mu_j \mathbb{E}\gamma_i + \mu_i\mu_j\right]$$