$\epsilon = (\epsilon_1^T, \epsilon_2^T)^T \sim N(0, \Sigma)$

(2)

Assumed

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad \Sigma_{11} = \sigma_1^2 I_{n_1}, \quad \Sigma_{22} = \sigma_2^2 I_{n_2}, \quad \Sigma_{12} = \Sigma_{21}^T = \sigma_{12} I_{n_1 \times n_2}$$

$$\Sigma^{-1} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$$

$$\Lambda_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$\Lambda_{22}^{-1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

(4)

Conditional

$$\Lambda_{12} = -\Lambda_{11}\Sigma_{12}\Sigma_{22}^{-1}$$

$$\Lambda_{21} = -\Sigma_{22}^{-1}\Sigma_{21}\Lambda_{11}$$

Bayes theorem Example:

x: the true value

y: the measurement

$y = x + \epsilon$

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

$p(x|y) = \frac{p(y|x)\,p(x)}{p(y)}$

$p(y|x)$: likelihood, $p(x)$: prior density, $p(y)$: evidence

$N(\mu, \sigma^2)$ → μ: mean, σ: standard deviation

$p(x|\mu, \sigma^2) = N(x; \mu, \sigma^2)$

To get density value: $p(x|y, \mu, \sigma^2)$

$$p(x|y, \mu, \sigma^2) = \frac{p(y|x, \mu, \sigma^2)\,p(x|\mu, \sigma^2)}{p(y|\mu, \sigma^2)}$$

$$p(x|y, \mu, \sigma^2) \propto \exp\left[-\frac{1}{2}(y-x)^2/\sigma^2\right] \cdot \exp\left[-\frac{1}{2}(x-\mu)^2/\sigma^2\right]$$

$$p(x|y, \mu, \sigma^2) \propto \exp\left[-\frac{1}{2}\left(\frac{(y-x)^2}{\sigma^2} + \frac{(x-\mu)^2}{\sigma^2}\right)\right]$$

$p_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}(x-\mu)^2/\sigma^2\right]$

$p_{\mu,\sigma^2}(y) = \frac{1}{\sqrt{2\pi(\sigma^2+\sigma_0^2)}}\exp\left[-\frac{1}{2}(y-\mu)^2/(\sigma^2+\sigma_0^2)\right]$

Marginal distribution for σ

$p(\sigma) = \mathrm{IGAMMA}(\sigma; \alpha, \beta)$

marginal distribution τ

$p(\tau) = \mathrm{GAMMA}(\tau; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1}\exp[-\beta\tau]$ $\quad$ α: shape, β: rate

$E(\tau) = \alpha/\beta$

$V(\tau) = \alpha/\beta^2$

$\sigma_0^2 = \tau^{-1}$

$y = Ax + \epsilon$

I_x $\quad x \sim N(\mu_x, \Sigma_x)$

$p(x|\mu_x, \Sigma_x) \sim N(x|\mu_x, \Sigma_x)$

$$p(x|\mu, \Sigma_x) = \frac{1}{(2\pi)^{n/2}|\Sigma_x|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma_x^{-1}(x-\mu)\right]$$

$\tau = \frac{1}{\sigma_0^2}$ $\quad$ $\mathrm{det}(\tau I) = \tau^n$ for here, τ maximal

likelihood

$$p(y|x, \tau) \propto \left\{(2\pi)^{n/2}(\det\Sigma_y)^{1/2}\right\}\tau^{n/2}\exp\left[-\frac{\tau}{2}(y-Ax)^T\Sigma_y^{-1}(y-Ax)\right]$$

prior $\quad p(x|\mu_x, \tau) = N(x|\mu_x, \tau^{-1}\Sigma_x/\tau/2)$

posterior $\quad p(x, \tau)$ $\quad \sigma_0^2 = \frac{1}{\tau}$ $\quad x \sim N(\mu, \Sigma_x/p, \tau\theta)$

$p(x, \tau)$ $\quad x \sim N(\mu, \tau\Sigma_x/p, \tau\theta)$ $\quad E[x] = \mu$

$D[x] = \left[\frac{p}{p-1}\Sigma_x\right]\frac{1}{\tau\theta}$

marginal distribution for τ Bayes Gamma Distribution

$\tau \sim G(\alpha^2)$ $\quad \tau \sim G(\alpha, \beta)$

$\sigma^2 \sim IG(\alpha, \beta)$

$E[\sigma^2] = \frac{\beta_0}{p_0 - 1} = \sigma_p^2$ → approx $\hat{\sigma}^2$

$V[\sigma^2] = \frac{\beta_0^2}{(p_0-1)^2(p_0-2)}$ $\quad V\sigma^2$ → approx

$p_0 = (\sigma_p^2)^2/V\sigma_p^2 + 2$

$\beta_0 = [(\sigma_p^2)/V\sigma_p^2]\,\sigma_p^2$

$p(x, \tau|y) \sim NG(\mu, \Lambda, b, p)$

$$p(x, \tau|y) \propto \underbrace{\tau^{n/2}\exp\left[-\frac{\tau}{2}(y-Ax)^T\Sigma_y^{-1}(y-Ax)\right]}_{\text{likelihood}}\ \underbrace{\tau^{p_0-1}\exp\left[-\frac{\tau}{2}(2b_0 + (x-\mu_0)^T\Sigma_x^{-1}(x-\mu_0))\right]}_{\text{prior}}$$

$$\propto \tau^{n/2+p_0-1}\exp\left\{-\frac{\tau}{2}\left[2b_0 + (x-\mu_0)^T\Sigma_x^{-1}(x-\mu_0) + (y-Ax)^T\Sigma_y^{-1}(y-Ax)\right]\right\}$$

$$\propto \tau^{n+p_0-1}\exp\left\{\cdot \quad \checkmark\right\}$$

posterior

$$\propto \tau^{n/2+p+1}\exp\left\{-\frac{\tau}{2}\left[2b + (x-\mu)^T V^{-1}(x-\mu)\right]\right\}$$

NG

μ

$V^{-1} = (A^T\Sigma_y^{-1}A + \Sigma_x^{-1})$

$$= \exp\left\{-\frac{\tau}{2}\left[2b + (x-\mu)^T(A^T\Sigma_y^{-1}A + \Sigma_x^{-1})(x-\mu)\right]\right\}$$

$p = p_0 + n/2$

$$p = (\sigma_p^2)^2/V\sigma_p^2 + 2 + \frac{n}{2} = \left[\frac{n}{2} + 2(\sigma_p^2)^2/V\sigma_p^2 + 4\right]/2$$

$$2b_0 + (x-\mu_0)^T\Sigma_x^{-1}(x-\mu_0) + (y-Ax)^T\Sigma_y^{-1}(y-Ax)$$

$$= 2b_0 + y^T\Sigma_y^{-1}y + \mu_0^T\Sigma_x^{-1}\mu_0 - 2x^T(A^T\Sigma_y^{-1}y + \Sigma_x^{-1}\mu_0) + x^T(A^T\Sigma_y^{-1}A + \Sigma_x^{-1})x$$

$$= 2b_0 + y^T\Sigma_y^{-1}y + \mu_0^T\Sigma_x^{-1}\mu_0 - \mu^T(A^T\Sigma_y^{-1}A + \Sigma_x^{-1})\mu + (x-\mu)^T(A^T\Sigma_y^{-1}A + \Sigma_x^{-1})(x-\mu)$$

$$= 2b_0 + y^T\Sigma_y^{-1}y + \mu_0^T\Sigma_x^{-1}\mu_0 - 2\mu^T(A^T\Sigma_y^{-1}A + \Sigma_x^{-1}\mu_0) + \mu^T(A^T\Sigma_y^{-1}A + \Sigma_x^{-1})\mu + (x-\mu)^T(A^T\Sigma_y^{-1}A + \Sigma_x^{-1})(x-\mu)$$

$$= 2b_0 + (\mu_0-\mu)^T\Sigma_x^{-1}(\mu_0-\mu) + (y-A\mu)^T\Sigma_y^{-1}(y-A\mu) + (x-\mu)^T(A^T\Sigma_y^{-1}A + \Sigma_x^{-1})(x-\mu)$$

$$b = \left\{2b_0 + (\mu-\mu_0)^T\Sigma_x^{-1}(\mu-\mu_0) + (y-A\mu)^T\Sigma_y^{-1}(y-A\mu)\right\}/2$$

$$b = \left\{2\left[(\sigma_p^2)^2/V\sigma_p^2 + 1\right]\sigma_p^2 + (\mu-\mu_0)^T\Sigma_x^{-1}(\mu-\mu_0) + (y-A\mu)^T\Sigma_y^{-1}(y-A\mu)\right\}/2$$

$p(x, \tau|y) \sim NG(\mu, \Lambda, b, p)$

$p(x|y) \sim t(\mu, b\Lambda/p, 2p)$

$E[x|y] = \mu$

$p(\tau|y) \sim G(b, p)$

$p(\sigma_b^2|y) \sim IG(b, p)$

$\hat{\sigma}_b^2 = E[\sigma_b^2|y] = \frac{b}{p-1}$

$$\hat{\sigma}_b^2 = (n+2(\sigma_p^2)^2/V\sigma_p^2 + 2)^{-1}\left\{\left[2\left[(\sigma_p^2)^2/V\sigma_p^2 + 1\right]\sigma_p^2\right] + (\mu-\mu_0)^T\Sigma_x^{-1}(\mu-\mu_0) + (y-A\mu)^T\Sigma_y^{-1}(y-A\mu)\right\}$$

$$V(\sigma_b^2|y) = \frac{2(\hat{\sigma}_b^2)^2}{n + 2(\sigma_p^2)^2/V\sigma_p^2}$$

$\sigma_0^2 = 1$

$V\sigma_0^2 = 1$

Adjusted with prior information (Bayes)	Adjusted without prior information
	$\hat{\sigma}^2 = \frac{e^T\Sigma_y^{-1}e}{n-u} = \frac{(y-A\hat{x})^T\Sigma_y^{-1}(y-A\hat{x})}{n-u}$

$\mu = \hat{x} = MAP$

Bayes estimator ← Adjusted with prior information

prior $\quad p(x|\Sigma_x^{-1}) \sim N(\mu, \Sigma_x)$

likelihood $\quad p(y|x, \Sigma_y^{-1}) \sim N(y, \Sigma_y^{-1})$

posterior $\quad p(x|y, \Sigma_x^{-1}, \Sigma_y^{-1}) \sim N(\mu, \Lambda)$

$\mu = \hat{x} = (A^T\Sigma_y^{-1}A + \Sigma_x^{-1})^{-1}(A^T\Sigma_y^{-1}y + \Sigma_x^{-1}\mu)$

$\Lambda = \Sigma_x^{-1}(A^T\Sigma_y^{-1}A + \Sigma_x^{-1})^{-1}$

$\hat{\sigma}_b^2 = (n + 2(\sigma_p^2)^2/V\sigma_p^2)^{-1}\left[2\left[(\sigma_p^2)^2/V\sigma_p^2 + 1\right]\sigma_p^2 + (\mu-\mu_0)^T\Sigma_x^{-1}(\mu-\mu_0) + (y-A\mu)^T\Sigma_y^{-1}(y-A\mu)\right]$