P.1 Analytical manipulation of Gaussian densities

Let $p(x_a, x_b) = \mathcal{N}(\mathbf{x}\,|\,\mu, \Sigma)$ with

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{aa} & \sigma_{ab} \\ \sigma_{ab} & \sigma_{bb} \end{pmatrix}$$

Marginalization: $p(x_i) = \int p(x_a, x_b)\, dx_i = \mathcal{N}(x_i\,|\,\mu_i, \sigma_{ii})$, $i = a, b$

(does not have to be shown here)

Show that $p(x_a|x_b) = \mathcal{N}(x_a\,|\,\mu_{a|b}, \sigma_{a|b})$ where

$$\mu_{a|b} = \mu_a + \frac{\sigma_{ab}}{\sigma_{bb}}(x_b - \mu_b), \quad \sigma_{a|b} = \sigma_{aa} - \frac{\sigma_{ab}^2}{\sigma_{bb}}$$

We know $\Sigma^{-1} = \frac{1}{\sigma_{aa}\sigma_{bb} - \sigma_{ab}^2} \begin{pmatrix} \sigma_{bb} & -\sigma_{ab} \\ -\sigma_{ab} & \sigma_{aa} \end{pmatrix}$.

Look at log-transformed distributions

$$\log p(x_a|x_b) = \log p(x_a, x_b) - \log p(x_b) =$$

$$= \underbrace{\boxed{-\log\sqrt{(2\pi)^2 \det \Sigma} + \log\sqrt{2\pi\,\sigma_{bb}}}}_{\text{constants, ignore for now}} - \frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) + \frac{1}{2}(x_b - \mu_b)^2 / \sigma_{bb}$$

$$\rightarrow -\frac{1}{2} \frac{1}{\sigma_{aa}\sigma_{bb} - \sigma_{ab}^2} (x_a - \mu_a, x_b - \mu_b) \begin{pmatrix} \sigma_{bb} & -\sigma_{ab} \\ -\sigma_{ab} & \sigma_{aa} \end{pmatrix} \begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix} + \frac{1}{2} \frac{(x_b - \mu_b)^2}{\sigma_{bb}} =$$

$$= -\frac{1}{2} \frac{1}{\sigma_{aa}\sigma_{bb} - \sigma_{ab}^2} \big(\sigma_{bb}(x_a - \mu_a) - \sigma_{ab}(x_b - \mu_b),\; -\sigma_{ab}(x_a - \mu_a) + \sigma_{aa}(x_b - \mu_b)\big) \cdot \begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix} + \frac{1}{2} \frac{(x_b - \mu_b)^2}{\sigma_{bb}} =$$

$$= -\frac{1}{2} \frac{1}{\sigma_{aa}\sigma_{bb} - \sigma_{ab}^2} \Big[\sigma_{bb}(x_a - \mu_a)^2 - \sigma_{ab}(x_b - \mu_b)(x_a - \mu_a) - \sigma_{ab}(x_a - \mu_a)(x_b - \mu_b) + \sigma_{aa}(x_b - \mu_b)^2 - (\sigma_{aa}\sigma_{bb} - \sigma_{ab}^2)/\sigma_{bb}\,(x_b - \mu_b)^2\Big] =$$

$$= -\frac{1}{2} \frac{1}{\sigma_{aa}\sigma_{bb} - \sigma_{ab}^2} \cdot \frac{1}{\sigma_{bb}} \Big[\sigma_{bb}^2 (x_a - \mu_a)^2 - 2\sigma_{ab}\sigma_{bb}(x_a - \mu_a)(x_b - \mu_b) + \sigma_{ab}^2 (x_b - \mu_b)^2\Big] =$$

$$= -\frac{1}{2} \frac{1}{\sigma_{bb}^2} \frac{1}{\sigma_{aa} - \frac{\sigma_{ab}^2}{\sigma_{bb}}} \big(\sigma_{bb}(x_a - \mu_a) - \sigma_{ab}(x_b - \mu_b)\big)^2 =$$

$$= -\frac{1}{2} \frac{\cancel{\sigma_{bb}^2}}{\cancel{\sigma_{bb}^2}} \underbrace{\frac{1}{\sigma_{aa} - \frac{\sigma_{ab}^2}{\sigma_{bb}}}}_{= \sigma_{a|b}} \Big[x_a - \big(\underbrace{\mu_a + \frac{\sigma_{ab}}{\sigma_{bb}}(x_b - \mu_b)}_{= \mu_{a|b}}\big)\Big]^2$$

Exercise II. 2

a) i) $x_{t+1} = 0.4\, x_t + v_t \quad , \; v_t \sim \mathcal{N}(0,1)$

$y_t = -0.5\, x_t + e_t \quad , \; e_t \sim \mathcal{N}(0, 0.01)$

Then
$$\left.\begin{aligned} p(x_t \mid x_{t-1}) &= \mathcal{N}(x_t;\, 0.4\, x_{t-1},\, 1) \\ p(y_t \mid x_t) &= \mathcal{N}(y_t;\, -0.5 x_t,\, 0.01) \end{aligned}\right\} \text{conjugate pdf, FAPF can be computed}$$

Checking a property proposed in the lecture

$$p(y_t \mid x_{t-1}) = \frac{p(y_t, x_{t-1})}{p(x_{t-1})} \, \frac{p(y_t, x_t, x_{t-1})}{p(y_t, x_t, x_{t-1})} =$$
$$= \frac{p(y_t, x_{t-1})}{p(y_t, x_t, x_{t-1})} \cdot p(y_t, x_t \mid x_{t-1}) = \frac{p(y_t, x_t \mid x_{t-1})}{p(x_t \mid y_t, x_{t-1})}$$

and
$$p(y_t, x_t \mid x_{t-1}) = \frac{p(y_t, x_t, x_{t-1})}{p(x_{t-1})} \, \frac{p(x_t, x_{t-1})}{p(x_t, x_{t-1})} = p(y_t \mid x_t, \overset{\text{Markov}}{\cancel{x_{t-1}}})\, p(x_t \mid x_{t-1})$$

$$\Rightarrow p(y_t \mid x_{t-1}) = \frac{p(y_t \mid x_t)\, p(x_t \mid x_{t-1})}{p(x_t \mid y_t, x_{t-1})}$$

ii) $x_{t+1} = \cos(x_t)^2 + v_t \quad , \; v_t \sim \mathcal{N}(0,1)$

$y_t = 2\, x_t + e_t \quad , \; e_t \sim \mathcal{N}(0, 0.01)$

Then
$$\left.\begin{aligned} p(x_t \mid x_{t-1}) &= \mathcal{N}(x_t;\, \cos(x_{t-1})^2;\, 1) \\ p(y_t \mid x_t) &= \mathcal{N}(y_t;\, 2x_t,\, 0.01) \end{aligned}\right\} \text{conjugate pdf, FAPF can be computed}$$

iii) $x_{t+1} = \cos(x_t + v_t)^2 \quad , \; v_t \sim \mathcal{N}(0,1)$

$y_t = 2 x_t + e_t \quad , \; e_t \sim \mathcal{U}([-2, 2])$

Then
$$\left.\begin{aligned} &p(x_t \mid x_{t-1}) \text{ nonlinear transform of normal dist.} \\ &p(y_t \mid x_t) = \mathcal{U}([-2(1+x_t),\, 2(1+x_t)]) \end{aligned}\right\} \text{quite probably not conjugate}$$

b) How to calculate $p(y_t \mid x_{t-1})$?

Markov x_{t-1} not necessary

$$p(y_t \mid x_{t-1}) = \int p(y_t, x_t \mid x_{t-1})\, dx_t = \int p(y_t \mid x_t)\, p(x_t \mid x_{t-1})\, dx_t =$$
$$= \frac{1}{2\pi \cdot 1 \cdot 0.1} \int \exp\!\left(-\frac{1}{2} \frac{(y_t + 0.5 x_t)^2}{0.1^2}\right) \exp\!\left(-\frac{1}{2} \frac{(x_t - 0.4 x_{t-1})^2}{1^2}\right) dx_t =$$

$$= \frac{1}{2\pi \cdot 0.1} \int \exp\left(-\frac{1}{2}\frac{(y_t + 0.5x_t)^2}{0.1^2}\right)\exp\left(-\frac{1}{2}\frac{(x_t - 0.4x_{t-1})^2}{1^2}\right) dx_t =$$

$$= \frac{1}{2\pi\, 0.1} \int \exp\left(-\frac{1}{2}\left(\frac{1}{0.1^2}(y_t - (-0.5x_t))^2 + \frac{1}{1^2}(x_t - 0.4x_{t-1})^2\right)\right) dx_t =$$

$$= \frac{1}{2\pi\, 0.1} \int \exp\left(-\frac{1}{2}\left[\left(\frac{1}{0.1^2}(-0.5)^2 + \frac{1}{1^2}\right)x_t^2 - \frac{1}{0.1^2}2\cdot y_t(-0.5)x_t - \frac{1}{1^2}2\, 0.4x_{t-1}x_t\right]\right)$$

$$\cdot \exp\left(-\frac{1}{2}\left(\frac{1}{0.1^2}y_t^2 + \frac{1}{1^2}0.4^2 x_{t-1}^2\right)\right) dx_t =$$

$$= \frac{1}{2\pi \cdot 0.1} \exp\left(-\frac{1}{2}\left(\frac{1}{0.1^2}y_t^2 + \frac{1}{1^2}0.4^2 x_{t-1}^2\right)\right) \cdot$$

$$\int \exp\left(-\frac{1}{2}\left(\frac{1}{0.1^2}(-0.5)^2 + \frac{1}{1^2}\right)\left(x_t^2 - 2x_t \cdot \left(\frac{0.5^2}{0.1^2} + \frac{1}{1^2}\right)^{-1}\left[\frac{-0.5}{0.1^2}y_t + \frac{0.4}{1^2}x_{t-1}\right] + \right.\right.$$

$$+ \left(\frac{0.5^2}{0.1^2} + \frac{1}{1^2}\right)^{-2}\left(\frac{-0.5}{0.1^2}y_t + \frac{0.4}{1^2}x_{t-1}\right)^2 -$$

$$\left.\left. - \left\{ \quad \sim \quad \right\}^2\right)\right) dx_t =$$

$$= \frac{1}{2\pi \cdot 0.1} \exp\left(-\frac{1}{2}\left(\frac{1}{0.1^2}y_t^2 + \frac{1}{1^2}0.4^2 x_{t-1}^2 - \left(\frac{0.5^2}{0.1^2} + \frac{1}{1}\right)^{-1}\left(\frac{-0.5}{0.1^2}y_t + \frac{0.4}{1^2}x_{t-1}\right)^2\right)\right)$$

$$\cdot \int \exp\left(-\frac{1}{2}(5^2+1)\left[x_t - \frac{1}{5^2+1}\left(\frac{-0.5}{0.1^2}y_t + \frac{0.4}{1^2}x_{t-1}\right)\right]^2\right) dx_t = (*)$$

$$z_t = \left[x_t - \frac{1}{5^2+1}\left(\frac{-0.5}{0.1^2}y_t + \frac{0.4}{1^2}x_{t-1}\right)\right]\sqrt{5^2+1}$$

$$\Rightarrow \frac{dz_t}{dx_t} = \sqrt{5^2+1} \Rightarrow dx_t = \frac{1}{\sqrt{5^2+1}}\, dz_t$$

$$(*) = \frac{1}{2\pi\, 0.1} \cdot \frac{1}{\sqrt{5^2+1}} \exp(\sim) \underbrace{\int \exp\left(-\frac{1}{2}z_t^2\right) dz_t}_{=\sqrt{2\pi}} =$$

$$= \frac{1}{\sqrt{2\pi} \cdot 0.1 \cdot \sqrt{5^2+1}} \cdot \exp\left(-\frac{1}{2}\frac{1}{0.1^2(5^2+1)} \cdot \left(\left(\frac{0.5^2}{0.1^2} + \frac{1}{1^2}\right)y_t^2 + 0.1^2\left(\frac{0.5^2}{0.1^2} - \cancel{\frac{1}{1^2}}\right)\frac{1}{1^2}0.4^2 x_{t-1}^2\right.\right.$$

$$\left.\left. - 0.1^2 \cdot \left[\frac{0.5^2}{0.1^2}y_t^2 + 2\frac{(-0.5)}{0.1^2}y_t\frac{0.4}{1^2}x_{t-1} + \frac{0.4^2}{1^2}x_{t-1}^2\right]\right)\right)$$

$$= \frac{1}{\sqrt{2\pi} \cdot 0.1\sqrt{5^2+1}} \cdot \exp\left(-\frac{1}{2} \cdot \frac{1}{0.1^2(5^2+1)} \cdot \left(y_t^2 + 2 \cdot 0.5 \cdot 0.4\, y_t x_{t-1} + 0.5^2 \cdot 0.4^2 x_{t-1}^2\right.\right.$$

$$= \frac{1}{\sqrt{2\pi} \cdot 0.1\sqrt{5^2+1}} \cdot \exp\left(-\frac{1}{2} \cdot \frac{1}{0.1^2(5^2+1)} \cdot \left(y_t - (-0.5)0.4x_{t-1}\right)^2\right) =$$

$$= \mathcal{N}\left(y_t; -0.5\, 0.4\, x_{t-1},\ 0.1\sqrt{5^2+1}\right)$$

Exercise II 2

General rule

$$p(x_t \mid x_{t-1}) = \mathcal{N}(x_t; f(x_{t-1}), \sigma_1)$$
$$p(y_t \mid x_t) = \mathcal{N}(y_t; c\, x_t, \sigma_2)$$

then $p(y_t \mid x_{t-1}) = \mathcal{N}(y_t; c \cdot f(x_{t-1}), \sigma_1^2 c^2 + \sigma_2^2)$

[see e.g. Bishop (2006), p. 93]

as well as

$$p(x_t \mid x_{t-1}, y_t) = \frac{p(y_t \mid x_t)\, p(x_t \mid x_{t-1})}{p(y_t \mid x_{t-1})},$$

calculate with log-transform, without constants and without pdf in bottom, not relevant, since not depending on x_t in any way, thus constant

$$\log p(y_t \mid x_t) + \log p(x_t \mid x_{t-1}) \hat{=}_{\text{no constants}}$$

$$\hat{=} -\frac{1}{2} \frac{(y_t - c x_t)^2}{\sigma_2^2} - \frac{1}{2} \frac{(x_t - f(x_{t-1}))^2}{\sigma_1^2} =$$

$$= -\frac{1}{2} (\sigma_2^2 \sigma_1^2)^{-1} \left[\sigma_1^2 (y_t^2 - 2 y_t c x_t + c^2 x_t^2) + \sigma_2^2 (x_t^2 - 2 x_t f(x_{t-1}) + f(x_{t-1})^2) \right] =$$

$$= -\frac{1}{2} (\sigma_2^2 \sigma_1^2)^{-1} \Big[\sigma_1^2 \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big) x_t^2 - 2 x_t \sigma_1^2 \Big\{ c y_t + \frac{\sigma_2^2}{\sigma_1^2} f(x_{t-1}) \Big\} +$$
$$+ \sigma_1^2 \Big(y_t^2 + \frac{\sigma_2^2}{\sigma_1^2} f(x_{t-1})^2\Big) \Big] =$$

$$= -\frac{1}{2} \sigma_2^{-2} \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big) \Big[x_t^2 - 2 x_t \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big)^{-1} \Big\{ c y_t + \frac{\sigma_2^2}{\sigma_1^2} f(x_{t-1}) \Big\} +$$
$$+ \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big)^{-1} \Big(y_t^2 + \frac{\sigma_2^2}{\sigma_1^2} f(x_{t-1})^2\Big) \Big] =$$

$$= -\frac{1}{2} \sigma_2^{-2} \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big) \Big[x_t^2 - 2 x_t \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big)^{-1} \Big\{ c y_t + \frac{\sigma_2^2}{\sigma_1^2} f(x_{t-1}) \Big\} +$$
$$+ \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big)^{-2} \Big\{ c y_t + \frac{\sigma_2^2}{\sigma_1^2} f(x_{t-1}) \Big\}^2 -$$
$$\left. \begin{array}{l} - \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big)^{-2} \Big\{ c y_t + \frac{\sigma_2^2}{\sigma_1^2} f(x_{t-1}) \Big\}^2 + \\ + \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big)^{-1} \Big(y_t^2 + \frac{\sigma_2^2}{\sigma_1^2} f(x_{t-1})^2\Big) \Big] \end{array} \right\} \text{Constant}$$

$$\rightsquigarrow -\frac{1}{2} \sigma_2^{-2} \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big) \Big[x_t - \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big)^{-1} \Big(c y_t + \frac{\sigma_2^2}{\sigma_1^2} f(x_{t-1}) \Big) \Big]^2$$

$$\Rightarrow \mathcal{N}\Big(\Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big)^{-1} \Big(c y_t + \frac{\sigma_2^2}{\sigma_1^2} f(x_{t-1}) \Big), \sigma_2^2 \Big/ \Big(c^2 + \frac{\sigma_2^2}{\sigma_1^2}\Big) \Big)$$

Exercise III.3

Let $X^i \sim \pi(x)$ for $i=1,\dots,N$ and $\tilde{\omega}^i \sim \mathcal{U}([0,1])$, $i=1,\dots,N$

Set $\omega^i := \dfrac{\tilde{\omega}^i}{\sum_{i=1}^{N} \tilde{\omega}^i}$ and $\hat{m} := \sum_{i=1}^{N} \omega^i X^i$.

Claim: $\hat{m}$ is an unbiased estimator for the mean m of π.

Proof:

$$\mathbb{E}(\hat{m}) = \iint \sum^{N} \omega^i X^i \, d(\pi \times \mathcal{U}) \overset{\text{Fubini}}{=} \iint \sum_{i=1}^{N} \omega^i X^i \, d\pi \, d\mathcal{U} =$$

$$\overset{\pi,\mathcal{U} \text{ independent}}{\Rightarrow} = \int \underbrace{\sum_{i=1}^{N} \omega^i}_{=1} \underbrace{\int X^i \, d\pi}_{=m} \, d\mathcal{U} = m \qquad \#$$

Assume we have a realisation $\{x^i, \omega^i\}_{i=1}^{N}$ s.th. $\sum_{i=1}^{N} \omega^i = 1$.

Resampling of $\{x^i\}_{i=1}^{N}$ w.r.t. the weights $\{\omega^i\}_{i=1}^{N}$ gives

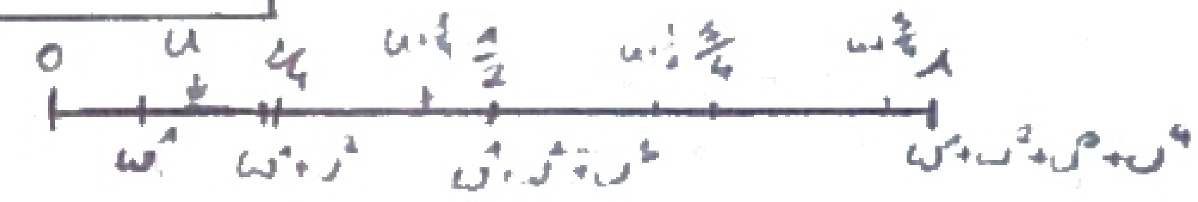
$A^i \sim p(\{\omega^i\}_{i=1}^{N})$ according to some distribution s.th. $\{x^{a^i}\}_{i=1}^{N}$ are all equally weighted.

Define $\underset{\text{multinomial}}{\hat{m}_m} = \underset{\text{systematic}}{\hat{m}_s} = \underset{\text{stratified}}{\hat{m}_t} := \frac{1}{N} \sum_{i=1}^{N} x^{A^i}$.

resampling

Multinomial $p(A^i = k) = \omega^k$

Systematic



$u \sim \mathcal{U}([0, \frac{1}{M}))$, $u^i = \frac{i-1}{M} + u$

$\Rightarrow u^i \sim \mathcal{U}([\frac{i-1}{M}, \frac{i}{M}))$

Assume X is distributed like

$\mathbb{P}(X \in [\frac{i-1}{M}, \frac{i}{M})) = \omega^i$ on $[0,1)$,

with cdf F_X then

$F_X^{-1}(u) = \frac{i-1}{M}$ for $u \in [\sum_{j=1}^{i-1} \omega^j, \sum_{j=1}^{i} \omega^j)$

Set $D^{-1}(u) = N \cdot F_X^{-1}(u) + 1$ then

$D^{-1}(u) = i$ for $u \in [\sum_{j=1}^{i-1} \omega^j, \sum_{j=1}^{i} \omega^j)$

Let $U^i \sim \mathcal{U}([\frac{i-1}{M}, \frac{i}{M}))$ and define $A^i = D^{-1}(U^i)$

Then $\mathbb{P}(A^i = k) = \mathbb{P}(U^i \in [\sum_{j=1}^{k-1} \omega^j, \sum_{j=1}^{k} \omega^j)) =$

$= |[\frac{i-1}{M}, \frac{i}{M}) \cap [\sum_{j=1}^{k-1} \omega^j, \sum_{j=1}^{k} \omega^j)|$

thus $\sum_{i=1}^{N} \mathbb{P}(A^i = k) = \omega_k$

stratified sampling works along the same lines

Claim: $\mathbb{E}(\hat{m}_m) = \mathbb{E}(\hat{m}_s) = \mathbb{E}(\hat{m}_t) = \hat{m}$

Proof:

$$\mathbb{E}\left(\frac{1}{N} \sum_{i=1}^{N} x^{A^i}\right) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}(x^{A^i}) =$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega^j x^j = \sum_{j=1}^{N} \omega^j x^j = \hat{m} \qquad \#$$

holds with different justifications for all resampling methods

Exercise III.3 (continued)

Rao-Blackwell Theorem

Let $\hat{\Theta}$ be an estimator of Θ with $\mathbb{E}(\hat{\Theta}^2) < +\infty$ for all Θ. Suppose that T is sufficient for Θ and let $\Theta^* = \mathbb{E}(\hat{\Theta} \mid T)$. Then for all Θ,

$$\mathbb{E}(\Theta^* - \Theta)^2 \leq \mathbb{E}(\hat{\Theta} - \Theta)^2.$$

Θ^* is called the Rao-Blackwellization of $\hat{\Theta}$.

Here: $\Theta \hat{=} m$, $T = \sum_{i=1}^{N} W^i X^i$ is sufficient for $m = \mathbb{E}[X]$ where $X \sim \pi(x)$

Then $\hat{m}_m = \hat{m}_s = \hat{m}_r = \frac{1}{N} \sum_{i=1}^{N} X^{A^i}$ and

$$\mathbb{E}_{(A^i, W^i, X^i)}\left[\hat{m}_m \,\middle|\, \sum_{i=1}^{N} W^i X^i = t \right] = \mathbb{E}_{(W^i, X^i)}\left[\sum_{i=1}^{N} W^i X^i \,\middle|\, \sum_{i=1}^{N} W^i X^i = t \right] =$$

Not very rigorous...

analogously for $\hat{m}_s$, $\hat{m}_r$. $\qquad = t = \hat{m}$

$\Rightarrow$ $\hat{m}$ is the Rao-Blackwellization of $\hat{m}_m$, $\hat{m}_s$, resp. $\hat{m}_r$ and thus

$$\mathbb{E}(\hat{m} - m)^2 \leq \begin{cases} \mathbb{E}(\hat{m}_m - m)^2 \\ \mathbb{E}(\hat{m}_s - m)^2 \\ \mathbb{E}(\hat{m}_r - m)^2 \end{cases}$$

#