N1 Доказать тождество Вудбери: $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$

Домножим слева на $(A+UCV)$ слева: $I = (A+UCV)(A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1})$

$$I = I + UCVA^{-1} - (A+UCV)A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$$

$$I = I + UCVA^{-1} - (U+UCVA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$$

$$I = I + UCVA^{-1} - UC\underbrace{(C^{-1}+VA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}}_{I}VA^{-1}$$

$$\cancel{I} = \cancel{I} + \cancel{UCVA^{-1}} - \cancel{UCVA^{-1}}$$

чтд.

N2 $p(x) = \mathcal{N}(x|\mu,\Sigma)$ $\quad A\in\mathbb{R}^{m\times n} \Rightarrow x\in\mathbb{R}^n,\ y\in\mathbb{R}^m,\ \Gamma\in\mathbb{R}^{m\times m}$

$p(y|x) = \mathcal{N}(y|Ax,\Gamma)$ $\quad$ Найти $p(x|y)$ $\quad \mu\in\mathbb{R}^{n\times n},\ \Sigma\in\mathbb{R}^{n\times n}$

По т. Байеса: $p(x|y) = \dfrac{p(y|x)p(x)}{\int p(y|x)p(x)dx}$

$$\begin{cases} p(x) = \dfrac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\} \\ p(y|x) = \dfrac{1}{(2\pi)^{m/2}|\Gamma|^{1/2}}\exp\{-\frac{1}{2}(y-Ax)^T\Gamma^{-1}(y-Ax)\} \end{cases}$$

Значит, $p(y|x)p(x) = \dfrac{1}{(2\pi)^{\frac{n+m}{2}}(|\Sigma||\Gamma|)^{\frac{1}{2}}}\exp\{-\frac{1}{2}((x-\mu)^T\Sigma^{-1}(x-\mu) + (y-Ax)^T\Gamma^{-1}(y-Ax))\}$

Так как $p(x) = e^{-x^TFx\ldots}\ldots$ и $p(y|x)$ как ф-я по $x = e^{-x^TFx\ldots}\ldots \Rightarrow$ и их произведение по x будет $e^{-x^TFx\ldots}\ldots$ (перевернутая парабола под экспонентой инвариантна относительно операции умножения).

Значит, $p(x|y)$ – тоже нормальное распределение $\mathcal{N}(x|\mu^*,\Sigma^*)$

Как найти μ^* и Σ^*? Заметим, что для норм. распр-я мат. ожидание равно моде x_{MP}.

1. Найдем $x_{MP} = \arg\max\limits_x p(x|y) = \arg\max\limits_x \dfrac{p(y|x)p(x)}{\int\ldots dx = const\ по\ x} =$

$= \arg\max\limits_x p(y|x)p(x) =$ логарифмируем $= \arg\max\limits_x \ln p(y|x) + \ln p(x) =$

$= \arg\max\limits_x -\cancel{\frac{n}{2}\ln 2\pi} - \cancel{\frac{1}{2}\ln|\Sigma|} - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) - \cancel{\frac{m}{2}\ln 2\pi} - \cancel{\frac{1}{2}\ln|\Gamma|} - \frac{1}{2}(y-Ax)^T\Gamma^{-1}(y-Ax)$

const по x

$= \arg\max\limits_x -(x-\mu)^T\Sigma^{-1}(x-\mu) - (y-Ax)^T\Gamma^{-1}(y-Ax) =$

$= \arg\max\limits_x -\frac{1}{2}x^T\Sigma^{-1}x + \mu^T\Sigma^{-1}x - \frac{1}{2}\mu^T\Sigma^{-1}\mu - \frac{1}{2}y^T\Gamma^{-1}y + y^T\Gamma^{-1}Ax - \frac{1}{2}x^TA^T\Gamma^{-1}Ax$

$= \arg\max\limits_x -\frac{1}{2}x^T(\Sigma^{-1} + A^T\Gamma^{-1}A)x + (\mu^T\Sigma^{-1} + y^T\Gamma^{-1}A)x - \frac{1}{2}(\mu^T\Sigma^{-1}\mu + y^T\Gamma^{-1}y)$ $\quad(*)$

$=$ не зависит от x $= \arg\max\limits_x -\frac{1}{2}x^T\_\_x + \_\_x$

Продифф-м по x: $-\_\_x + \_\_ = 0$

$$(\Sigma^{-1}+A^T\Gamma^{-1}A)x = \mu^T\Sigma^{-1} + y^T\Gamma^{-1}A$$

$$x_{MP} = (\Sigma^{-1}+A^T\Gamma^{-1}A)^{-1}(\mu^T\Sigma^{-1}+y^T\Gamma^{-1}A)$$

т.к $\overbrace{\Sigma^{-1}}^{n\times n} + \underbrace{A^T\Gamma^{-1}A}_{\substack{(n\times m)(m\times m)(m\times n)\\ \| \\ (n\times n)}} \in \mathbb{R}^{n\times n}$

Значит, $\boxed{\mu^* = (\Sigma^{-1}+A^T\Gamma^{-1}A)^{-1}(\mu^T\Sigma^{-1}+y^T\Gamma^{-1}A)}$

2. Найдем Σ^*. Заметим, что для нормального распр-я Σ^{-1} стоит при $x^T\ldots x$ - т.е. при "квадрате" аргумента в экспоненте.
Обратим внимание, что степень экспоненты записана в (*) (т.к. все остальное не зависит от x)
Значит, $\boxed{\Sigma^{*-1} = \Sigma^{-1} + A^T\Gamma^{-1}A}$

Итого $\boxed{p(x|y) = \mathcal{N}(x|\mu^*, \Sigma^*) \quad \mu^* = \Sigma^*(\mu^T\Sigma^{-1}+y^T\Gamma^{-1}A) \quad \Sigma^* = (\Sigma^{-1}+A^T\Gamma^{-1}A)^{-1}}$

N3 Доказать, что в задаче 2 $p(y) = \mathcal{N}(y|A\mu, \Gamma + A^T\Sigma^{-1}A)$

$p(y)$ = так же как и $p(y|x)$, $p(x)$ будет нормальным
(перевернутая парабола под экспонентой у $p(y|x)$, $p(x)$, $p(x|y)$ $\Rightarrow$ $p(y)$ - тоже перевернутая парабола под экспонентой (и не перевернутой она быть не может в силу того, что $p(y)$ - распределение $\Rightarrow \int p(y)dy = 0 \Rightarrow$ обычной параболы у е там быть не может)

Знаем, что $p(y|x) = \mathcal{N}(y|Ax, \Gamma)$ - то есть $y = Ax + \varepsilon$, где $\varepsilon \sim \mathcal{N}(0, \Gamma)$
Тогда $\mathbb{E}y = \mathbb{E}(Ax+\varepsilon) = A\mathbb{E}x + \mathbb{E}\varepsilon = \left\{\begin{matrix}\mathbb{E}x = \mu \\ \mathbb{E}\varepsilon = 0\end{matrix}\right\} = A\mu$

$$\mathbb{D}y = \mathbb{D}(Ax+\varepsilon) = \mathbb{D}Ax + \mathbb{D}\varepsilon = A^T\mathbb{D}xA + \Gamma = A^T\Sigma^{-1}A + \Gamma$$

чтд.

N4 $\dfrac{\partial|x^{-1}+A|}{\partial x} = \dfrac{\partial f(Y(A,x))}{\partial x}$

По правилу дифференцир-ния сложной функции $\dfrac{\partial f}{\partial x} = \dfrac{df(Y)}{dY}\cdot\dfrac{\partial Y(A,x)}{\partial x}$

$= \underbrace{|x^{-1}+A|(x^{-1}+A)^{-1}}_{\frac{d\det(Y)}{dY}} \times \underbrace{(-1)x^{-2}}_{\frac{\partial(x^{-1}+A)}{\partial x} = \frac{dx^{-1}}{dx}} = -|x^{-1}+A|(x^{-1}+A)^{-1}x^{-2}$

$\boxed{\text{Ответ: } -|x^{-1}+A|(x^{-1}+A)^{-1}x^{-2}}$

N5 $\dfrac{\partial\, tr(AX^{-T}BXC)}{\partial X}$ $\quad X \in \mathbb{R}^{n\times n} \Rightarrow A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{n\times k}$,
$m \neq n$, $k \neq n$ - т.к. неквадратные матрица

Значит $AX^{-T}BXC$: $(m\times n)(n\times n)(n\times n)(n\times n)(n\times k) = (m\times k) \Rightarrow$

Считаем A, B, C за const, найдем $d\,tr(AX^{-T}BXC)$.

$tr(AX^{-T}BXC) = \langle I_{mk}, AX^{-T}BXC \rangle$. Поэтому $d\,tr(AX^{-T}BXC) = d\langle I_{mk}, AX^{-T}BXC \rangle =$

$= \langle I_{mk}, dAX^{-T}BXC \rangle = \langle I_{mk}, dAX^{-T}B \cdot XC + AX^{-T}B\,dXC \rangle = -\langle I_{mk}, A(dX^{-T})BXC \rangle$

$+ \langle I_{mk}, AX^{-T}B(dX)C \rangle =$ перекосим по принципу $\langle A, BC \rangle = \langle AC^T, B \rangle =$

$= \langle B^T A, C \rangle$

$= -\langle X^{-T}BXCAX^{-T}, dX \rangle + \langle B^T X^{-T} A^T C^T, dX \rangle =$

$= \langle B^T X^{-T} A^T C^T - X^{-T}BXCAX^{-T}, dX \rangle$

Значит, $\dfrac{d\,tr(\ldots)}{dX} = B^T X^{-T} A^T C^T - X^{-T}BXCAX^{-T}$