$$p(y, w, X|A) = \prod_j \mathcal{N}(x_j|0, \sigma^2 I_d) \mathcal{N}(w|0, A^{-1}) \prod_j p(y_j|x_j, w)$$

$$\mathbb{P}(y_j = 1) = \frac{1}{1+\exp(-w^T x_j)}$$

a) $p(w|X_{train}, y_{train}, A)$ -?

$p(w|X, y, A) \cdot p(X, y|A) = p(y, w, X|A)$

$p(w|X, y, A) \cdot p(y|X, A) \cdot p(X|A) = p(y, w, X|A)$

$\Downarrow$

$$p(w|X, y, A) = \frac{p(y, w|X, A)}{p(y|X, A)} = \frac{\mathcal{N}(w|0, A^{-1}) \prod_j p(y_j|x_j, w)}{p(y|X, A)}$$

$q(w) = -\log p(y; w|X, A) = -\log p(w|A) - \log p(y|X, w) =$

$= q(w_{MAP}) + \frac{1}{2}(w - w_{MAP})^T H^{-1}(w - w_{MAP}) + O(\|w - w_{MAP}\|^3)$

$H^{-1} = A + X^T R X, \; R = \operatorname{diag}(\sigma(w_{MAP}^T x_j)\sigma(-w_{MAP}^T x_j))$

w_{MAP}: $q(w) = wAw^T + \sum_j \log \sigma(y_j w^T x_j)$

$$\nabla q(w) = Aw + \sum_j \nabla \log \sigma(y_j w^T x_j)$$

$$\nabla \sigma(x) = \nabla \frac{1}{1+e^{-x}} = -\frac{1}{(1+e^{-x})^2}(e^{-x}(-1)) = \frac{e^{-x}}{(1+e^{-x})^2} \Rightarrow$$

$\sigma(x) = \frac{1}{1+e^{-x}}$ $\Rightarrow$ $e^{-x} = \sigma^{-1}(x) - 1$

$\Rightarrow \nabla \sigma(x) = \sigma(x)^2(\sigma^{-1}(x) - 1) = \sigma(x) - \sigma^2(x)$

$$\Rightarrow \nabla \log \sigma(y_j w^T x_j) = \frac{\nabla \sigma(\ldots)}{\sigma(\ldots)} = \frac{(\sigma(\ldots) - \sigma^2(\ldots)) \cdot y_j x_j}{\sigma(\ldots)} =$$

$= (1 - \sigma(x)) \cdot x_j y_j$

$\Downarrow$

$\nabla q(w) = Aw + \sum_j (1 - \sigma(x)) x_j y_j \Rightarrow$

$\Rightarrow w_{MAP}$: $Aw_{MAP} + \sum_j (1 - \sigma(y_j w_{MAP}^T x_j)) x_j y_j = 0$

$y = Xw + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$, σ^2 – известно

$p(w) = \mathcal{N}(w \mid m, \mathrm{diag}(s))$ – $m, \mathrm{diag}(s)$ – неизвестно

а) $p(y, w \mid X, m, s) = \mathcal{N}(y \mid Xw, \sigma^2 I) \cdot \mathcal{N}(w \mid m, \mathrm{diag}(s))$

б) $$p(w \mid X, m, s) = \frac{p(y, w \mid X, m, s)}{p(y \mid X, m, s)} = \frac{\mathcal{N}(y \mid Xw, \sigma^2 I)\, \mathcal{N}(w \mid m, \mathrm{diag}(s))}{p(y \mid X, m, s)}$$

в) $$p(y \mid X, m, s) = \int \underbrace{p(y \mid X, w)}_{\text{правдоподобие}} \cdot \underbrace{p(w \mid m, s)}_{\text{априорное}}\, dw =$$

$$= \frac{p(y, w \mid X, m, s)}{p(w \mid X, m, s)}$$ ← апостериорное

$p(y \mid X, w)$ – нормальное
$p(w \mid m, s)$ – нормальное } ⇒ сопряженные ⇒

⇒ $p(w \mid X, m, s)$ – тоже нормальная.

$$\boxed{\begin{aligned} p(w \mid X, m, s) &= \mathcal{N}(w \mid \mu', \Sigma_1') = \mathcal{N}(\theta \mid \mu', \Sigma_1') \\ p(y \mid X, w) &= \mathcal{N}(y \mid \cancel{\ldots}\, w, \Sigma_1) = \mathcal{N}(x \mid \theta, \Sigma) \\ p(w \mid \mu, \Sigma_1) &= \mathcal{N}(w \mid \mu_0, \Sigma_0) = \mathcal{N}(\theta \mid \mu_0, \Sigma_0) \end{aligned}}$$

$$p(\theta \mid X, \mu_0, \Sigma_0) = \frac{p(\theta, X \mid \mu_0, \Sigma_0)}{p(X \mid \mu_0, \Sigma_0)}$$

$$p(X \mid \mu_0, \Sigma_0) = \frac{p(\theta, X \mid \mu_0, \Sigma_0)}{p(\theta \mid \mu_0, X, \Sigma_0)}$$

$$p(X \mid \mu_0, \Sigma_0) = \frac{\mathcal{N}(X \mid \theta, \Sigma_1)\, \mathcal{N}(\theta \mid \mu_0, \Sigma_0)}{\mathcal{N}(\theta \mid (\Sigma_0^{-1} + \Sigma_1^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + \Sigma_1^{-1}X), (\Sigma_0^{-1} + \Sigma_1^{-1})^{-1})} =$$

$$\sim \frac{\exp(-\frac{1}{2}(x-\theta)^T \Sigma_1^{-1}(x-\theta))\, \exp(-\frac{1}{2}(\theta-\mu_0)^T \Sigma_0^{-1}(\theta-\mu_0))}{\exp(-\frac{1}{2}(\theta - (\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + \Sigma_1^{-1}X))^T (\Sigma_0^{-1}+\Sigma_1^{-1}) \times (\theta - (\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + \Sigma_1^{-1}X)))} =$$

$$= \exp\Big(-\frac{1}{2}\{x^T \Sigma_1^{-1} x - 2\theta^T \Sigma_1^{-1} x + \theta^T \Sigma_1^{-1} \theta + \theta^T \Sigma_0^{-1} \theta - 2\theta^T \Sigma_0^{-1} \mu_0 + \mu_0^T \Sigma_0^{-1} \mu_0 - \theta^T (\Sigma_0^{-1} + \Sigma_1^{-1}) \theta - [(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + \Sigma_1^{-1}x)]^T \cdot (\Sigma_0^{-1}+\Sigma_1^{-1}) \cdot [(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + \Sigma_1^{-1}x)] + 2\theta^T (\Sigma_0^{-1}+\Sigma_1^{-1}) \cdot (\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + \Sigma_1^{-1}x)$$ =

$= \exp\left(-\frac{1}{2}\{x^T\Sigma_1^{-1}x^T + \mu_0^T\Sigma_0^{-1}\mu^T - [(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}\cdot\right.$

$\left.\cdot(\Sigma_0^{-1}\mu_0+\Sigma_1^{-1}x)]^T\cdot(\Sigma_0^{-1}\mu_0+\Sigma_1^{-1}x)]\right) =$

$= \exp\left(-\frac{1}{2}\{x^T\Sigma_1^{-1}x + \mu_0^T\Sigma_0^{-1}\mu^T - (\Sigma_0^{-1}\mu_0+\Sigma_1^{-1}x)^T\cdot\right.$

$\left.\cdot(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}\cdot(\Sigma_0^{-1}\mu_0+\Sigma_1^{-1}x)\}\right) =$

$(\Sigma_0^{-1}\mu_0)^T(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}(\Sigma_0^{-1}\mu_0)^T +$

$+2(\Sigma_0^{-1}\mu_0)^T(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}\Sigma_1^{-1}x + \underbrace{(\Sigma_1^{-1}x)^T(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}(\Sigma_1^{-1}x)}_{x^T[\Sigma_1^{-1}(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}\Sigma_1^{-1}]x}$

$\cancel{= \exp(-\frac{1}{2}\{x^T}$

$= \exp\{-\frac{1}{2}\{x^T\Sigma_1^{-1}(I-(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}\Sigma_1^{-1})x -$

$-2(\Sigma_0^{-1}\mu_0)^T(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}\Sigma_1^{-1}x + \dots\})$

$\mathcal{N}(x|\xi,H) = (x-\xi)^T H^{-1}(x-\xi) = x^T H^{-1}x^T - 2\xi^T H^{-1}x + \dots\}$

под [illegible] экспоненты

$H^{-1} = \Sigma_1^{-1}(I - (\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}\Sigma_1^{-1})$

$H = (I-(\Sigma_0^{-1}+\Sigma_1^{-1})\Sigma_1^{-1})^{-1}\Sigma_1$

$\xi = H\,\Sigma_1^{-1}(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}\Sigma_0\mu_0$

$\xi = (I - (\Sigma_0^{-1}+\Sigma_1^{-1})\Sigma_1^{-1})^{-1}\underbrace{\Sigma_1\Sigma_1^{-1}}_{I}(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}\Sigma_0^{-1}\mu_0 =$

$= (\Sigma_0^{-1}+\Sigma_1^{-1} - \underbrace{(\Sigma_0^{-1}+\Sigma_1^{-1})(\Sigma_0+\Sigma_1^{-1})}_{I}\Sigma_1^{-1})^{-1}\Sigma_0^{-1}\mu_0 =$

$= (\Sigma_0^{-1})^{-1}\Sigma_0^{-1}\mu_0 = \mu_0$

$\Downarrow$

$\cancel{\mathcal{N}}\exp p(x|\mu_0,\Sigma_0) = \mathcal{N}(x|\cancel{\mu_0}, H)$

$H = (I-(\Sigma_0^{-1}+\Sigma_1^{-1})^{-1}\Sigma_1^{-1})^{-1}\Sigma_1$

$\mu_0 = m,\ \Sigma_0 = diag(s)$

$\Sigma_1 = \sigma^2 I$

$p(y|m, dia\,s) = \mathcal{N}(y|\cancel{\mu_0}, H)\ \mathcal{N}(y|m, H)$

$\Sigma_0 = X \operatorname{diag}(s) X^T$
$\Sigma_1 = \sigma^2 I$
$\mu_0 = Xm$

$\forall A \quad [\Sigma_1, A] = 0$

$H = (I - (\Sigma_0^{-1} + \Sigma_1^{-1})^{-1} \Sigma_1^{-1})^{-1} \Sigma_1 =$

$= ((\Sigma_0^{-1} + \Sigma_1^{-1})^{-1} (\Sigma_0^{-1} + \Sigma_1^{-1}) - (\Sigma_0^{-1} + \Sigma_1^{-1})^{-1} \Sigma_1^{-1})^{-1} \Sigma_1 =$

$= ((\Sigma_0^{-1} + \Sigma_1^{-1})^{-1} \Sigma_0^{-1})^{-1} \Sigma_1 = \Sigma_0 (\Sigma_0^{-1} + \Sigma_1^{-1}) \cdot \Sigma_1 =$

$= (I + \Sigma_0 \Sigma_1^{-1}) \Sigma_1 = \Sigma_0 + \Sigma_1$

$$\boxed{p(y \mid X, m, s) = \mathcal{N}(y \mid \underbrace{Xm}_{\mu_0}, \underbrace{\sigma^2 I + X \operatorname{diag}(s) X^T}_{H})}$$

$p(y \mid X, m, s) \sim \frac{1}{\sqrt{\det H}} \exp\{-\frac{1}{2}(y - Xm)^T H^{-1} (y - Xm)\}$

~~a) Будто~~

Если ~~$\exists m, \infty$~~

Если $\exists m_0 \mid Xm_0 = y$, то $\max p(y \mid X, m, s) = \mathcal{N}(y \mid Xm_0,$

a) Если $\exists m_0 \mid Xm_0 = y$ ~~и~~ $\hat{m} = m_0$, $s = 0$ и т.д.

$= \mathcal{N}(y \mid Xm_0, \sigma^2 I\}$ => $\operatorname{rank}(X) = \operatorname{rank}(X \mid y)$ <=>

Если $\exists m_0 \mid Xm_0 = y$ <=> $\operatorname{rank}(X)$ меньше чем признаков

<=> независимых данных меньше чем признаков

1) Практический вывод: Если кол-во независ. точек должно быть как кол-во независимых точек должно быть как минимум не меньше чем кол-во признаков, иначе модель переобучается, т.к. идеально подстраивается под данные

б) Если $\nexists m_0 \mid Xm_0 = y$ и $\operatorname{rank} X < n$, то

Если ~~$\operatorname{rank} X m_0 \le$~~

$\log(y \mid X, m, s) = -\frac{1}{2}\log\{\det(\sigma^2 I + X\,\mathrm{diag}(s)\,X^T)\} -$

$-\frac{1}{2}\log[(y - Xm)^T(\sigma^2 I + X\,\mathrm{diag}(s)\,X)^{-1}(y - Xm)] =$ / H - симметричная / $=$

$(y - Xm)^T H (y - Xm) =$

$q_m(y) = -\frac{1}{2}(y - Xm)^T H (y - Xm) \approx$

$\cancel{H^{-1}(Xm - y)} = -\frac{1}{2}(Xm - y)^T H^{-2}(Xm - y) =$

$= y^T H^{-2} y + (Xm)^T H^{-2}(Xm) - 2(Xm)^T H^{-2} y =$

$= y^T H^{-2} y + m^T X^T H^{-2} X m - 2y^T H^{-2} X m$

$\partial q_m(y) = 2X^T H^{-2} X m - 2(y^T H^{-2} X)^T = 0 \Rightarrow$

$\Rightarrow X^T H^{-2} X m = X^T H^{-2} b \Rightarrow$

$\Rightarrow \boxed{\hat{m} = (X^T H^{-2} X)^{-1} X^T H^{-2} y}$ ~~[illegible]~~

$\hat{m} = m(s)$ (оптимальный m для фиксированного s)

$\Downarrow$

Теперь можно найти s

$\hat{s} = \arg\max \log\det H(s) - \frac{1}{2}(y - X\hat{m})^T H^{-1}(s)(y - X\hat{m})$

~~[illegible]~~ $H = \sigma^2 I + \sum s_k X_k X_k^T$

~~$XX^T = P^T \Sigma P$, где Σ - диаг.~~

$\Downarrow$ ~~$H' = P^{-1} H P$~~, H' - диагональн.

~~[illegible]~~

~~$P^T P = I$~~ ~~$XX^T = P^T \Sigma_1 P$, где~~

$XX^T = P^T X' P$, где $P^T P = I$, X' - диагональн.

~~$y' = Py$~~

~~$m' = Pm$~~

~~$H' = P^T H P$~~

$H = \sigma^2 P^T P + X\,\mathrm{diag}(s)\,X^T = \sigma^2 P^T P + P^T \mathrm{diag}(s \cdot \lambda)$

$\cdot P^T = P^T \mathrm{diag}(s\lambda + \sigma^2) P$

$\Rightarrow y' = Py$

$X' = PX$

$H' = \mathrm{diag}(s\lambda + \sigma^2)$

$\Rightarrow$ находим s_k^* как усредн. отдельных компонент

$s = \max\left(\frac{\hat{D}_k - \sigma^2}{\lambda}, 0\right)$

~~[illegible]~~

№3

$n_1 = 10$

$n_2 = 204$

$k_1 = 3$

$k_2 = 5100$

1) ~~$p(k$~~ правдоподобие: $p(k|n, p_1, p_2) = \binom{n_1}{k_1} p_1^{k_1}(1-p_1)^{n_1-k_1} \cdot \binom{n_2}{k_2} p_2^{k_2}(1-p_2)^{n_2-k_2}$

Т.к. у монеты два события орёл и решка, то оно задаётся одной вероятностью - это вероятность выпадения орла.

Априорное распределение: $p_1 \sim U[0,1]$ - т.к. мы $p_2 \sim U[0,1]$ не имеем информации о том как выбрались монеты, то будем считать что равномерно.

~~$p(k, p_1, p_2|n) = \binom{n_1}{k_1} p_1^{k_1}(1-p_1)^{n_1-k_1}$~~

$$p(k, p_1, p_2|n) = \prod_{i=1}^{2} \binom{n_i}{k_i} p_i^{k_i}(1-p_i)^{n_i-k_i} \cdot U(p_1|[0,1]) \cdot U(p_2|[0,1])$$

б) $U[0,1] \equiv B(1,1) \Rightarrow p(k|n, p_1, p_2)$ и $\Rightarrow U[0,1]$ - сопряжено с

и $q(p_1|k_1, n_1)$

~~$q(p_1|k_1, n_1) = \frac{\int p(k, p_1, p_2|n)}{\int p(k, p_1, p_2|n)}$~~

$$q(p_1, p_2|k, n) = \frac{p(k, p_1, p_2|n)}{\int p(k, p_1, p_2|n)\, dp_1 dp_2}$$

или

$$q(p_1|k_1, n_1) = \frac{p(k, p_1|n_1)}{\int p(k, p_1|n_1)\, dp_1}$$

$$\int_0^1 \binom{n_1}{k_1} p_1^{k_1}(1-p_1)^{n_1-k_1} \cdot \frac{dp_1}{1} =$$

$$= \binom{n_1}{k_1} \int_0^1 p_1^{k_1}(1-p_1)^{n_1-k_1} dp_1 = \binom{n_1}{k_1} \frac{(n_1-k_1)!\,k_1!}{(n_1+1)!} = \frac{n_1!\,(n_1-k_1)!\,k_1!}{k_1!\,(n_1-k_1)!\,(n_1+1)!} =$$

$$= \frac{1}{(n_1+1)} \Rightarrow \boxed{q(p_1|k_1, n_1) = \frac{\binom{n_1}{k_1} p^{k_1}(1-p_1)^{n_1-k_1} U(p_1|[0,1])}{\frac{1}{n_1+1}}}$$

6) ~~$M_2 \sum q(p_1 \cdot p_2)$~~

M_1: $p(k_1, k_2 | p, n_1, n_2) = \int p(k, p_1 | n_2)\, dp_1 =$

$= \int_0^1 \binom{n_1}{k_1}\binom{n_2}{k_2} p^{k_1} \cdot p^{k_2} (1-p)^{n_1-k_1} \cdot (1-p)^{n_2-k_2} dp =$

$= \binom{n_1}{k_1}\binom{n_2}{k_2} \int p^{k_1} \cdot p^{k_2} (1-p)^{n_1+n_2-(k_1+k_2)} dp =$

$= \binom{n_1}{k_1}\binom{n_2}{k_2} \cdot \frac{(n_1-k_1)!(k_1)!}{(n_1+1)!} = \binom{n_1}{k_1}\binom{n_2}{k_2} \left(\frac{(n_1+n_2-(k_1+k_2))!(k_1+k_2)!}{(n_1+n_2+1)!}\right)$

$p(m | k, n) = \frac{p(m, k | n)}{\underset{const}{p(k|n)} = \frac{1}{2}\sum p(k|n, m_1) + p(k|n, m_2)}$

$p(m|k,n) \sim p(k|n,m)$

~~$p(\ldots)$~~ $\sim \frac{1}{n_1+1} \cdot \frac{1}{n_2+1} \Rightarrow$ вероятн равномерн по k

$\cdot\ p(M=2|k,n)$

$p(M=1|k,n)$ - ?

пусть $k = qk$

$\frac{(qk_1)!(qk_2)!}{q(k_1+k_2)+1)!} \cdot \frac{((q-1)(k_1+k_2))!\,(k_1+k_2)!}{k_1!\,((q-1)k_1)!\,k_2!\,((q-1)k_2)!} \sim$

$\sim \frac{\sqrt{2\pi q k_1}}{\sqrt{2\pi k_1}} \cdot \frac{\left(\frac{qk_1}{e}\right)^{qk_1}}{\sqrt{2\pi(q-1)k_1} \cdot \left(\frac{k_1}{e}\right)^{k_1}} \cdot \frac{\sqrt{2\pi q k_2}\left(\frac{qk_2}{e}\right)^{qk_2}}{\left(\frac{(q-1)k_2}{e}\right)^{(q-1)k_2}} \cdot \frac{1}{\sqrt{2\pi k_2}\sqrt{2\pi(q-1)k_2}} \cdot$

$\cdot \frac{\sqrt{2\pi(q-1)(k_1+k_2)}\sqrt{2\pi(k_1+k_2)} \cdot \left(\frac{(q-1)(k_1+k_2)}{e}\right)^{(q-1)(k_1+k_2)} \cdot \left(\frac{k_1+k_2}{e}\right)^{k_1+k_2}}{\left(\frac{k_2}{e}\right)^{k_2} \cdot \left(\frac{(q-1)k_2}{e}\right)^{(q-1)k_2} \cdot \sqrt{2\pi q(k_1+k_2)}\left(\frac{q(k_1+k_2)}{e}\right)^{q(k_1+k_2)}}$

exp! $\frac{\left(\frac{qk_1}{e}\right)^{qk_1} \cdot \left(\frac{qk_2}{e}\right)^{qk_2} \cdot (q-1)^{(q-1)(k_1+k_2)} \cdot \left(\frac{k_1+k_2}{e}\right)^{(q-1)(k_1+k_2)}}{\left(\frac{k_2}{e}\right)^{k_2} \cdot (q-1)^{(q-1)k_2}\left(\frac{k_1}{e}\right)^{(q-1)k_1} \cdots \left(\frac{q(k_1+k_2)}{e}\right)^{q(k_1+k_2)}} =$

$= \frac{\left(\frac{qk_1}{e}\right)^{qk_1} \cdot \left(\frac{qk_2}{e}\right)^{qk_2} \cdot \left(\frac{k_1+k_2}{e}\right)^{q(k_1+k_2)}}{\left(\frac{k_1}{e}\right)^{qk_1} \cdot \left(\frac{k_2}{e}\right)^{qk_2} \cdot \left(\frac{q(k_1+k_2)}{e}\right)^{q(k_1+k_2)}} = \frac{q^{qk_1} \cdot q^{qk_2}}{q^{q(k_1+k_2)}} = 1$

$= 1$

$C(n)$: $\frac{\sqrt{2\pi q h_1}\sqrt{2\pi q h_2}}{\sqrt{2\pi h_1}\sqrt{2\pi(q-1)h_1}\sqrt{2\pi h_2}\sqrt{2\pi(q-1)h_2}}$ ·

$$\cdot\frac{\sqrt{2\pi(q-1)(h_1+h_2)}\sqrt{2\pi(h_1+h_2)}}{\sqrt{2\pi q(h_1+h_2)}} =$$

$$= \frac{\sqrt{q}\sqrt{q}\sqrt{q-1}}{\sqrt{2\pi}\sqrt{q-1}\sqrt{q-1}\sqrt{q}} = \frac{\sqrt{q}}{\sqrt{2\pi}\sqrt{q-1}}\frac{\sqrt{h_1+h_2}}{\sqrt{h_1}\sqrt{h_2}} =$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\frac{1}{q}}}\sqrt{\frac{k_1+k_2}{k_1k_2}} = \frac{1}{\sqrt{2\pi}\sqrt{1-p}}\sqrt{\frac{k_1+k_2}{k_1k_2}} =$$

$\frac{1}{q} = p \quad k = pn$

$$\Rightarrow \cancel{\frac{1}{\sqrt{2\pi}\sqrt{p}\sqrt{1-p}}}\ \frac{1}{\sqrt{2\pi}\sqrt{1-p}}\cdot\sqrt{\frac{p}{p^2}\frac{n_1+n_2}{n_1n_2}} = \frac{1}{\sqrt{2\pi}\sqrt{p}\sqrt{1-p}}\sqrt{\frac{n_1+n_2}{n_1n_2}}$$

$$\min\frac{1}{\sqrt{1-p}\sqrt{p}} = 2. \quad \frac{2}{\sqrt{2\pi}} \approx 0{,}8$$

$$\Rightarrow P(k=1|k,n) \approx \frac{1}{\sqrt{2\pi}\sqrt{p}\sqrt{1-p}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \geq 0{,}79\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

при $k \approx pn$.

$$\frac{P_{M_1}}{P_{M_2}} \approx \frac{0{,}79\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}{\frac{1}{n_1+1}\cdot\frac{1}{n_2+1}} \approx 0{,}79\, n_1 n_2\sqrt{\frac{n_1+n_2}{n_1n_2}} =$$

$$= 0{,}79\sqrt{(n_1n_2)(n_1+n_2)} \geq \sqrt{2}\cdot 0{,}79 \geq 1 \Rightarrow$$

$\Rightarrow P_{M_1} > P_{M_2}$

Причём при $n_1 >> 1$ $P_{M_1} >> P_{M_2}$

Практический вывод:

Более простая модель может ~~лучше~~ для подходить лучше, если она объясняет природу данных.

В данном случае, если $\frac{k_1}{n_1} \approx \frac{k_2}{n_2} \approx \alpha$, то обоснованность у M_1 ~~растёт пропорционально~~ относительно M_2 растёт пропорционально размеру выборки. ~~Следовательно, если~~

Следовательно это позволяет проверять гипотезы, как, например, в данном случае $H_0: (p_1 = p_2 = p)$

№4

$D_{KL}(P\|Q) = \int_X p \log \frac{p}{q} \, d\mu(dx)$, где $P(dx) = p(x)\mu(dx)$
$Q(dx) = q(x)\mu(dx)$

где $\mu(x)$ - любая мера на X не обязательно вероятностная.
Если X-счётно, то $\int_X = \sum_i$, а если $\mu(dx) = \mathbb{P}(dx)$, то
P, Q - абсолютно непрерывны и p, q - плотности P и Q

$D_{KL}(P\|Q)$ показывает прирост информации, достигнутый, если бы вместо Q, принятой сейчас, использовали P

Тогда $D_{KL}(P\|Q)$ определен тем, что $Q(dx) = 0 \Rightarrow P(dx) = 0$,
иначе $D_{KL}(P\|Q) = +\infty$

б) $\int_X p \log \frac{p}{q} \, d\mu(dx) = -\int_X p \log \frac{q}{p} \, d\mu - \int_X p\left(\frac{q}{p} - 1\right) d\mu =$
$= -\int (q - p) \, d\mu = -\int_X q \, d\mu + \int_X p \, d\mu = -\int_X q \, d\mu + 1 \geq 0$

$\log \frac{q}{p} \leq \frac{q}{p} - 1$