علی قاسم زاده

۴۰۱۱۰۶۳۳۹

۱-۱. داریم که ۱

$$s = \sum_{i=1}^{N-1} \alpha_i v_i \qquad x[n] = A\, s[n] + w[n]$$

$$\underline{x} = A\underline{s} + \underline{w} \longrightarrow \underline{w} = \underline{x} - A\underline{s}$$

با فرض تخمین گر BLUE داریم که ۱

$$\hat{A} = \alpha^T x$$

$$\mathbb{E}\{\hat{A}\} = A \longrightarrow \mathbb{E}\{\alpha^T x\} = \mathbb{E}\{A\alpha^T s\} = A\alpha^T s = A \rightarrow$$

$$\boxed{\alpha^T s = 1}$$

حالا می‌خواهیم واریانس را کمینه کنیم یعنی داریم:

$$var(\hat{A}) = \mathbb{E}\{(\hat{A}-A)^2\} = \mathbb{E}\{(\alpha^T \underline{x} - \alpha^T \mathbb{E}(\underline{x}))^2\} =$$

$$\alpha^T \mathbb{E}\{(\underbrace{\underline{x} - \mathbb{E}(\underline{x})}_{w})(\underbrace{\underline{x} - \mathbb{E}(\underline{x})}_{w})^T\}\alpha = \alpha^T \mathbb{E}\{ww^T\}\alpha =$$

$$\alpha^T C_w \alpha \,, \quad \alpha^T s = 1$$

$$\min_{\alpha} \ \alpha^T C_w \alpha \quad s.t.\ \alpha^T s = 1 \longrightarrow \mathcal{L} = \alpha^T C_w \alpha + \lambda(1 - \alpha^T s) \rightarrow$$

مشتق گیری در خواهیم داشت که

$$\alpha = \frac{\lambda}{2} C_w^{-1} s \Rightarrow \lambda = \frac{2}{s^T C_w^{-1} s} \longrightarrow \alpha = \frac{C_w^{-1} s}{s^T C_w^{-1} s}$$

α بهینه

حالا داریم که:

$$\boxed{\hat{A} = a^T x = \frac{s^T C_w^{-1} x}{s^T C_w^{-1} s}}$$

حالا اگر بردار ویژه متناظر یکه باشد (یکه نبود بیاییم ضریب α_i را عوض می‌کنیم که و بردار را بر عددی تقسیم می‌کنیم که یکه بشود.)

$$\Rightarrow \quad Var(\hat{A}_{BLUE}) = a^T C_w a = \frac{s^T C_w^{-1} \cancel{C_w} C_w^{-1} s}{s^T C_w^{-1} s \cdot s^T C_w^{-1} s} = \frac{1}{s^T C_w^{-1} s}$$

می‌دانیم که اگر $s \in eigen\_vector(C_w)$ باشد $\Leftarrow$

$$C_w s = \lambda s \to s = \lambda C_w^{-1} s \to C_w^{-1} s = \frac{1}{\lambda} s \to$$

$\frac{1}{\lambda}$ مقدار ویژه‌ی C_w^{-1} است $\Leftarrow$ پس اگر s یک مقدار ویژه‌ی C_w باشد،

داریم که:

$$\boxed{Var(\hat{A}_{BLUE}) = \frac{\lambda}{s^T s}}$$

مقدار ویژه‌ی متناظر با s

۲.۱- حالا باز اگر به صورت جمع بردارهای ویژه‌ی یکه‌ی متناظر بنویسیم، داریم خواهیم داشت که:

$$s^T C_w^{-1} s = \sum \frac{\alpha_i^2}{\lambda_i}$$

چراکه داشتیم:

$$\left(\sum \alpha_i v_i\right) C_w^{-1} \left(\sum \alpha_i v_i\right) = \left(\sum \alpha_i v_i\right)\left(\sum \alpha_i C_w^{-1} v_i\right) =$$

$$\left(\sum \alpha_i v_i\right)\left(\sum \alpha_i \frac{1}{\lambda_i} v_i\right) \overset{\text{تعامد}}{=} \sum \frac{\alpha_i^2}{\lambda_i} \underbrace{v_i^2}_{1} = \sum \frac{\alpha_i^2}{\lambda_i}$$

$$\Rightarrow \boxed{\operatorname{Var}(\hat{A}_{BLUE}) = \frac{1}{\sum \frac{\alpha_i^2}{\lambda_i}}}$$

حالا فرض کنیم که $\sum \alpha_i^2 = \varepsilon = s^T s$, $\varepsilon < \varepsilon_0$

می دانیم که بزرگترین مقدار $s^T C_w^{-1} s$ به ازای زمانی حاصل می شود که

s بردار ویژه‌ی متناظر با بزرگترین مقدار ویژه‌ی C_w^{-1} هست $\Leftarrow$ ~~متناظر با~~

بردار ویژه‌ی متناظر با کوچکترین مقدار ویژه‌ی C_w هست.

$\leftarrow$ به ازای بردار ویژه‌ی متناظر با کوچکترین مقدار ویژه‌ی C_w، $s^T C_w^{-1} s$ بیشینه

می شود یعنی $\operatorname{Var}(\hat{A}_{BLUE}) = \frac{1}{s^T C_w^{-1} s}$ کمینه می شود. $\Leftarrow$ خواهیم داشت که

با انتخاب $s = \alpha v_{min}$ بهترین انتخاب هست حالا چه α ای برداریم که

واریانس BLUE کمتر شود؟

$$\boxed{\operatorname{Var}(\hat{A}_{BLUE}) = \frac{1}{\frac{s^T s}{\lambda_{min}}} = \frac{\lambda_{max}}{\alpha^2 \underbrace{v_{min}^T v_{min}}_{1}} = \frac{\lambda_{max}}{\alpha^2}}$$

ماکزیمم α ای که می توانیم برداریم را باید برداریم تا Var کمینه شود، پس خواهیم داشت که

$s^T s = \alpha^2 v_{min}^T v_{min} = \alpha^2$, داریم که $\varepsilon_0 = c^2 = s^T s \leftarrow$

$\leftarrow \alpha^2 = c^2 \leftarrow$ با انتخاب $s = c\, v_{min}$ ،

به کمترین واریانس خواهیم رسید. $\rightarrow$ واریانس تخمین‌ها کمتر داریم

منطقیه چون داریم روی کمترین مقدار ویژه و بردار ویژه‌اش پیش میریم که $\Leftarrow$

همچنین می توانیم با حل مسئله ی

$$\min_{\alpha, s} \quad \alpha^T C_w \alpha$$

$$\text{s.t.} \quad \alpha^T s = 1$$

$$s^T s \leq \varepsilon_0$$

جواب مسئله را بدست بیاوریم. با همان روشی قبلی که گفته شد هم

اولی هست.

۲-۱. داریم که:

$$P(\underline{x};\theta)=\prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{x_i^m e^{-x_i/2\theta}}{(2\theta)^{m+1}\, m!}$$

$$\ln P(x;\theta) = \sum_{i=1}^{n} m \ln x_i + \sum -\frac{x_i}{2\theta} - \sum (m+1)\ln 2\theta - \sum \ln m!$$

$$\rightarrow \frac{\partial \ln P(\underline{x};\theta)}{\partial\theta} = \frac{1}{2\theta^2}\sum x_i - \frac{n(m+1)}{\theta} = 0 \rightarrow \frac{n(m+1)}{\theta} = \frac{1}{2\theta^2}\sum x_i$$

$$\rightarrow \boxed{\hat{\theta}_{ml} = \frac{\sum x_i}{2n(m+1)}}$$

توزیع $Gamma(\alpha = m+1, 2\theta)$ $f(x;\theta)$ هست $\Leftarrow$ $\mathbb{E}\{X_i\} = \alpha 2\theta$

$= 2(m+1)\theta$

$$\mathbb{E}\{\hat{\theta}_{mL}\} = \frac{\sum \mathbb{E}\{x_i\}}{2n(m+1)} = \frac{2n(m+1)\theta}{2n(m+1)} = \theta$$

بدون بایاس است.

$$Var(\hat{\theta}_{mL}) = \frac{1}{(n(m+1))^2} n(m+1)4\theta^2 = \frac{4\theta^2}{n(m+1)}$$

حالا کران کرامر رائو را بدست می‌آوریم:

$$I(\theta) = \mathbb{E}\{(\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta))^2\} = \frac{n(m+1)}{4\theta^2} \rightarrow Var(\hat{\theta}) = \frac{4\theta^2}{n(m+1)}$$

$\longrightarrow$ تخمین گر ml به باند کرامر رائو می‌رسد $\Leftarrow$ بهینه است.

۲_۲. داریم:

$$x_k = \theta^{1/2} N_k \quad k=1,\dots,n$$

$$\Rightarrow \quad \underline{x} = \theta^{1/2} N \longrightarrow \quad x \sim \mathcal{N}(0, \theta\Sigma)$$

$$p(x;\theta) = \frac{1}{\sqrt{(2\pi)^n |\theta\Sigma|}} \exp\left(-\frac{1}{2} x^T (\theta\Sigma)^{-1} x\right) =$$

$$\frac{1}{\sqrt{(2\pi)^n |\theta\Sigma|}} \exp\left(-\frac{1}{2\theta} x^T \Sigma^{-1} x\right) \Longrightarrow$$

$$\frac{\partial}{\partial\theta} \ln p(x;\theta) = \frac{n}{2\theta} - \frac{1}{2\theta^2} x^T \Sigma^{-1} x = 0 \Longrightarrow \cancel{\theta_{ml} = x^T \Sigma^{-1} x}$$

$$\boxed{\hat{\theta}_{mL} = \frac{x^T \Sigma^{-1} x}{n}} \qquad \mathbb{E}\{\hat{\theta}_{mL}\} = \frac{1}{n}\mathbb{E}\{tr(\Sigma^{-1} x x^T)\} =$$

$$\frac{1}{n} tr(\Sigma^{-1}\theta\Sigma) = \frac{1}{n} n\theta = \theta \longrightarrow$$ بدون بایاس است.

$$Var(\hat{\theta}_{mL}) = \frac{1}{n^2} var(x^T \Sigma^{-1} x) = \frac{2n\theta^2}{n^2} = \frac{2\theta^2}{n}$$

حال کران کرامر رائو را می‌نویسیم:

$$I(\theta) = -\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2} \ln p(x;\theta)\right\} = -\mathbb{E}\left\{\frac{n}{2\theta^2} - \frac{1}{\theta^3} x^T \Sigma^{-1} x\right\}$$

$$-\mathbb{E}\left\{\frac{n}{2\theta^2} - \frac{x^T \Sigma^{-1} x}{\theta^3}\right\} = -\frac{n}{2\theta^2} + \frac{n\theta}{\theta^3} = \frac{-n}{2\theta^2} + \frac{n}{\theta^2} = \frac{n}{2\theta^2}$$

$$\longrightarrow \quad Var(\hat{\theta}) \geq \frac{2\theta^2}{2} \longrightarrow$$ ml به کران کرامر رائو می‌رسد

درنتیجه بهینه است.

۳-۲.

$$p(x;\theta)=\theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \rightarrow \ln p(x;\theta)=\left(\sum x_i\right)\ln\theta+\left(n-\sum x_i\right)\ln(1-\theta)$$

$$\Rightarrow \frac{\partial \ln p(x;\theta)}{\partial\theta}=\frac{\sum x_i}{\theta}-\frac{n-\sum x_i}{1-\theta}=0 \rightarrow \hat{\theta}_{ML}=\frac{\sum x_i}{n}$$

حالا واریانس این تخمینگر را باید حساب کنیم،

$$\mathbb{E}\{\hat{\theta}_{ML}\}=\frac{\sum_{i=1}^{n}\mathbb{E}\{x_i\}}{n}=\theta\ ,\ var(\hat{\theta})=\mathbb{E}\{(\hat{\theta}-\theta)^2\}=\mathbb{E}\{\hat{\theta}^2\}-\theta^2=$$

$$\cancel{\mathbb{E}\left\{\frac{\sum x_i x_j}{n^2}-\theta^2\right\}}\quad \cancel{\mathbb{E}\left\{\frac{\sum x_i x_j}{n^2}-\theta^2\right\}}$$

$$=\mathbb{E}\left\{\frac{\sum x_i x_j}{n^2}-\frac{\sum \mathbb{E}\{x_i\}\mathbb{E}\{x_j\}}{n^2}\right\}=$$

$$\mathbb{E}\left\{\frac{\sum_{i=j} x_i x_j-\mathbb{E}\{x_i\}\mathbb{E}\{x_j\}}{n^2}\right\}=\mathbb{E}\left\{\frac{\sum x_i^2-\mathbb{E}\{x_i\}\mathbb{E}\{x_i\}}{n^2}\right\}=$$

$$\frac{1}{n}var(x_1)=\frac{\theta(1-\theta)}{n} \longrightarrow var(\hat{\theta}_{ML})=\frac{\theta(1-\theta)}{n}$$

حالا کران کرامر رائو را می یابیم،

$$\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}=-\frac{\sum x_i}{\theta^2}-\frac{n-\sum x_i}{(1-\theta)^2}\Rightarrow$$

$$I(\theta)=-E\left\{\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\right\}=\mathbb{E}\left\{\frac{\sum x_i}{\theta^2}+\frac{n-\sum x_i}{(1-\theta)^2}\right\}=$$

$$\frac{n\theta}{\theta^2}+\frac{n-n\theta}{(1-\theta)^2}=\frac{n}{\theta}+\frac{n}{1-\theta}=n\left(\frac{1}{\theta}+\frac{1}{1-\theta}\right)=\frac{n}{\theta(1-\theta)}$$

$$\longrightarrow \boxed{CRLB=\frac{\theta(1-\theta)}{n}} \longrightarrow$$ تخمین ML به این باند رسیده در نتیجه بهینه است

Scanned with CamScanner

۳-۱. داریم که:

$$J \sim Cat(P_1, \ldots, P_m)$$

$$K \sim Cat(q_1, \ldots, q_L)$$

$$X \mid J=j \,,\, K=k \sim \mathcal{N}(\mu_j, \sigma_k^2)$$

$$\theta = \{P_1, \ldots, P_m, \mu_1, \ldots, \mu_m, q_1, \ldots, q_L, \sigma_1^2, \ldots, \sigma_L^2\}$$

حالا برای محاسبه $P(x;\theta)$، داریم که ما پارامترها را $(P_1 - P_m)$، $(q_1 - q_L)$، ______ مشخص هستند و فقط باید احتمال $x;\theta$ را بدست آوریم:

$$P(x;\theta) = \sum_{j=1}^{m} \sum_{k=1}^{L} \overbrace{P(J=j)\,P(K=k)\,P(x \mid J=j, K=k)}^{P(x,\, J=j,\, K=k;\,\theta)} = \sum_{j=1}^{m} \sum_{k=1}^{L} P_j q_k \mathcal{N}(x; \mu_j, \sigma_k^2)$$

۳-۲.

$$\gamma_{jk}(x_n) = P(J_n=j, K_n=k \mid x_n; \theta) = \frac{P_j q_k \mathcal{N}(x_n; \mu_j, \sigma_k^2)}{\sum_{j'=1}^{m} \sum_{k'=1}^{L} P_{j'} q_{k'} \mathcal{N}(x_n; \mu_{j'}, \sigma_{k'}^2)}$$

۳-۳.

$$\log P(\{x_n, J_n, K_n\}_{n=1}^{N}; \theta) = \sum_{n=1}^{N} \left[\log P(J_n) + \log P(K_n) + \log P(x_n \mid J_n, K_n)\right]$$

$$= \sum_{n=1}^{N} \left[\log P_{J_n} + \log P_{K_n} - \frac{1}{2}\log(2\pi\sigma_{K_n}^2) - \frac{(x_n - \mu_{J_n})^2}{2\sigma_{K_n}^2}\right]$$

Q) $$Q(\theta \mid \theta^{old}) = \mathbb{E}_{J,K \mid X, \theta^{old}} \left\{\log P(\{x_n, J_n, K_n\}; \theta)\right\}$$

$$= \sum_{n=1}^{N} \sum_{j=1}^{m} \sum_{k=1}^{L} \gamma_{jk,n}^{old} \left[\log P(\{x_n, J_n, K_n\}; \theta)\right]$$

$$\hookrightarrow P(J_n=j, K_n=k \mid x_n; \theta^{old})$$

$$= \sum_{n=1}^{N} \sum_{j=1}^{m} \sum_{k=1}^{\ell} \underbrace{\gamma_{jk,n}^{old}}_{P(J_n=j, K_n=k \mid X_n; \theta^{old})} \left[\log P_j + \log q_k - \frac{1}{2} \log(2\pi \sigma_k^2) - \frac{(x_n - \mu_j)^2}{2\sigma_k^2} \right]$$

۴. حالا می‌خواهیم $Q(\theta|\theta^{old})$ را بیشینه کنیم و داریم که $\sum P_j = 1$ ، $\sum q_k = 1$

نسبت به تک تک پارامترها ماکزیمم می‌کنیم

$$\sum_{n,j,k} \gamma_{jk,n} \log P_j + \lambda\left(\sum P_j - 1\right) \xrightarrow[\text{قرار دادن}]{\substack{\text{مشتق‌گیری} \\ \text{و برابر صفر}}} \boxed{P_j = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{\ell} \gamma_{jk,n}}$$

برای q_k هم خواهیم داشت که:

$$\boxed{q_k = \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{\ell} \gamma_{jk,n}}$$

برای μ_j ها خواهیم داشت که

$$\sum_{n,k} \gamma_{jk,n} \frac{x_n - \mu_j}{\sigma_k^2} = 0 \Rightarrow \mu_j = \frac{\sum_{n=1}^{N} \sum_{k=1}^{\ell} \gamma_{jk,n} x_n / \sigma_k^2}{\sum_{n=1}^{N} \sum_{k=1}^{\ell} \gamma_{jk,n} / \sigma_k^2}$$

برای σ_k^2 نیز خواهیم داشت که:

$$-\frac{1}{2} \sum_{n,j} \gamma_{jk,n} \left[\log \sigma_k^2 + \frac{(x_n - \mu_j)^2}{\sigma_k^2} \right] \longrightarrow \sigma_k^2 = \frac{\sum_{n=1}^{N} \sum_{j=1}^{m} \gamma_{jk,n} (x_n - \mu_j)^2}{\sum_{n=1}^{N} \sum_{j=1}^{m} \gamma_{jk,n}}$$

داریم که، ←

E-step:

$$\gamma_{jk,n} = \frac{p_j\, q_k\, \mathcal{N}(x_n \mid \mu_j, \sigma_k^2)}{\sum_{j',k'} p_{j'}\, q_{k'}\, \mathcal{N}(x_n \mid \mu_{j'}, \sigma_{k'}^2)}$$

M-step،

$$p_j \leftarrow \frac{1}{N} \sum_{n,k} \gamma_{jk,n}, \qquad q_k \leftarrow \frac{1}{N} \sum_{n,j} \gamma_{jk,n}$$

$$\mu_j \leftarrow \frac{\sum_{n,k} \gamma_{jk,n}\, x_n / \sigma_k^2}{\sum_{n,k} \gamma_{jk,n} / \sigma_k^2}, \qquad \sigma_k^2 \leftarrow \frac{\sum_{n,j} \gamma_{jk,n} (x_n - \mu_j)^2}{\sum_{n,j} \gamma_{jk,n}}$$

Scanned with CamScanner

۱-۴. دیتاها $D=\{(x_i,y_i)\}_{i=1}^{N}$ را داریم و از طرفی می‌دانیم که برای $i \le M$ داریم که $y_i = z_i$ و

برای $N \ge i \ge M+1$ داریم که $z_i \ge y_i$. حالا داریم که،

$\mu = w^T x, \quad z = \mu + \varepsilon$

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

$$P(z,x;w,\sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z_i-\mu_i)^2}{2\sigma^2}\right) \longrightarrow$$

$$\log p(z,x;w,\sigma^2) = -\sum_{i=1}^{N} \log\sqrt{2\pi\sigma^2} - \sum_{i=1}^{N} \frac{(z_i-\mu_i)^2}{2\sigma^2} =$$

$$-\frac{N}{2}\log 2\pi\sigma^2 - \sum_{i=1}^{N} \frac{(z_i-\mu_i)^2}{2\sigma^2}$$

$$\mathbb{E}\{z_i\} = z_i \quad \text{for } i \le M$$

$$\mathbb{E}\{z_i\} = \mathbb{E}\{z_i \mid z_i \ge y_i\} \quad \text{for } M+1 \le i \le n$$

$$Q = \mathbb{E}\{\log P(x,z;w,\sigma^2)\} = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(\mathbb{E}\{z_i^2\} - 2\mu_i\mathbb{E}\{z_i\} + \mu_i^2\right)$$

حالا با مشتق گیری و برابر صفر قرار دادن خواهیم داشت که،

$$\frac{\partial}{\partial w}\left(\sum_{i=1}^{N} \mu_i \mathbb{E}\{z_i\} - \frac{1}{2}\mu_i^2\right) = 0 \longrightarrow$$

$$\sum_{i=1}^{N} x_i \mathbb{E}\{z_i\} - \sum_{i=1}^{N} x_i x_i^T w = 0 \longrightarrow$$

$$\left(\sum_{i=1}^{N} x_i x_i^T\right) w = \sum_{i=1}^{N} \mathbb{E}\{z_i\} x_i \longrightarrow \underline{w = \left(\sum_{i=1}^{N} x_i x_i^T\right)^{-1}\left(\sum_{i=1}^{N} \mathbb{E}\{z_i\} x_i\right)}$$

حالا برای بدست آوردن σ^2 به طور مشابه داریم که:

$$\frac{\partial Q}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N} \mathbb{E}\{(z_i^2 - w^T x_i)^2\} = 0 \longrightarrow$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{E}\{z_i^2\} - 2w^T x_i \mathbb{E}\{z_i\} + (w^T x_i)^2\right)$$

← σ^2 و w بهینه بدست می‌آید برحسب $\mathbb{E}\{z_i\}$, $\mathbb{E}\{z_i^2\}$ همچنین در رابطه بالا بجای w مقدار بهینه آنرا باید جایگذاری کنیم تا σ^2 بهینه بدست بیاید.

۲-۴. داریم که:

$$z_i \sim N(\mu_i, \sigma^2) \longrightarrow \text{define: } \tilde{z}_i = \frac{z_i - \mu_i}{\sigma} \sim N(0,1)$$

با شرط $z_i \geq c_i$:

$$\tilde{z}_i \geq u_i, \quad \text{where} \quad u_i = \frac{c_i - \mu_i}{\sigma}$$

$$\mathbb{E}\{z_i | z_i \geq c_i\} = \mu_i + \sigma \mathbb{E}\{\tilde{z}_i | \tilde{z}_i \geq u_i\}$$

$$\mathbb{E}\{\tilde{z}_i | \tilde{z}_i \geq u_i\} = \frac{\int_{u_i}^{\infty} t\,\phi(t)\,dt}{1 - \Phi(u_i)}$$

از آنجایی که $\phi'(t) = -t\,\phi(t)$ پس خواهیم داشت که:

$$\mathbb{E}\{\tilde{z}_i / \tilde{z}_i \geq u_i\} = \frac{\phi(u_i)}{1 - \Phi(u_i)} = H(u_i) \longrightarrow \mathbb{E}\{z_i | z_i \geq c_i\} = \mu_i + \sigma H(u_i)$$

$$z_i^2 = (\mu_i + \sigma\tilde{z}_i)^2 = \mu_i^2 + 2\mu_i\sigma\tilde{z}_i + \sigma^2\tilde{z}_i^2$$

$$\mathbb{E}\{z_i^2 | z_i \geq c_i\} = \mu_i^2 + 2\mu_i\sigma \underbrace{\mathbb{E}\{\tilde{z}_i | \tilde{z}_i \geq u_i\}}_{H(u_i)} + \sigma^2 \mathbb{E}\{\tilde{z}_i^2 | \tilde{z}_i \geq u_i\}$$

$$\mathbb{E}\{\tilde{z}_i^2 \mid \tilde{z}_i \geq u_i\} = ?$$

$$\int t^2 \phi(t)\, dt = -t\phi(t)\Big|_{u_i}^{\infty} + \int \phi(t)\, dt = u_i\, \phi(u_i) + 1 - \Phi(u_i)$$

$$\Rightarrow \mathbb{E}\{\tilde{z}_i^2 \mid \tilde{z}_i \geq u_i\} = \frac{u_i\, \phi(u_i) + 1 - \Phi(u_i)}{1 - \Phi(u_i)} = 1 + u_i\, H(u_i)$$

$$\Rightarrow \quad \mathbb{E}\{z_i \mid z_i \geq c_i\} = \mu_i + \sigma H\left(\frac{c_i - \mu_i}{\sigma}\right)$$

$$\mathbb{E}\{z_i^2 \mid z_i \geq c_i\} = \mu_i^2 + \sigma^2 + \sigma(c_i + \mu_i) H\left(\frac{c_i - \mu_i}{\sigma}\right)$$

۵-۱. داریم که :

$$P_1 : \min_{x} f_0(x) = \sum_{k=1}^{n} \left(\|x - y_k\|_2^2 - d_k^2 \right)^2$$

$$P_2 : \min_{x,t} f_0(x) = \sum_{k=1}^{n} \left(t - 2y_k^T x + \|y_k\|_2^2 - d_k^2 \right)^2$$

$$\text{s.t.} \quad x^T x - t = 0$$

حالا بردار z را در نظر می‌گیریم، $\begin{bmatrix} x \\ t \end{bmatrix}$، خواهیم داشت که،

$$\sum_{k=1}^{n} \left(t - 2y_k^T x + \|y_k\|_2^2 - d_k^2 \right)^2 = \left\| \begin{bmatrix} t - 2y_1^T x + y_1^T y_1 - d_1^2 \\ \vdots \\ t - 2y_n^T x + y_n^T y_n - d_n^2 \end{bmatrix} \right\|_2^2$$

$$\underbrace{\begin{bmatrix} -2y_1^T & 1 \\ \vdots & \vdots \\ -2y_n^T & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ t \end{bmatrix}}_{z}, \qquad \begin{bmatrix} y_1^T y_1 - d_1^2 \\ \vdots \\ y_n^T y_n - d_n^2 \end{bmatrix} = -b$$

$$= \|Az - b\|_2^2$$

همچنین خواهیم داشت که،

$$z^T C z + 2f^T z = 0 \equiv x^T x - t = 0$$

$$\rightsquigarrow \begin{bmatrix} x^T & t \end{bmatrix} C \begin{bmatrix} x \\ t \end{bmatrix} + 2f^T \begin{bmatrix} x \\ t \end{bmatrix} \equiv x^T x - t$$

بردار صفر

$$\longrightarrow C = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \quad f = \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix} \longrightarrow \begin{matrix} z^T C z = x^T x \\ 2f^T z = -t \end{matrix} \quad ✓$$

۵-۲. شرایط KKT را ابتدا می‌نویسیم، خواهیم داشت که:

① Primal const: $z^T C z + 2 f^T z = 0$ *

② Dual const: $\mathcal{L}(z,\nu) = z^T(A^TA)z - 2b^TAz + b^Tb + \nu(z^TCz + 2f^Tz)$

$$\longrightarrow z^T(A^TA + \nu C)z - 2(A^Tb - \nu f)^T z + \|b\|_2^2$$

← برای اینکه عبارت بالا کران دار باشد، لازم است داشته باشیم که، (کران پایین)

$$** \quad A^TA + \nu C \succeq 0, \quad A^Tb - \nu f \in \mathcal{R}(AA^T + \nu C)$$

حالا نسبت به z مشتق می‌گیریم و صفر قرار می‌دهیم به عبارت زیر می‌رسیم،

$$g(\nu) = -(A^Tb - \nu f)^T (A^TA + \nu C)^{+} (A^Tb - \nu f) + \|b\|_2^2$$

در نتیجه مسئله‌ی دوگان به شکل زیر خواهد شد،

$$\max_{\nu, t} \quad -t + \|b\|_2^2$$

$$\text{s.t.} \quad \begin{bmatrix} A^TA + \nu C & A^Tb - \nu f \\ (A^Tb - \nu f)^T & t \end{bmatrix} \succeq 0$$

③ complementory slackness: $\forall_i \; \lambda_i f_i = 0$ ✓

④ gradient of lagrangian with respect to z vanishes:

$$(A^TA + \nu C)z = A^Tb - \nu f$$

براساس راهنمای حالا داریم که:

$$W = Q^T L^T Z, \quad A^TA = LL^T, \quad L^{-1}CL^{-T} = Q\Lambda Q^T$$

(۱) $Z^TCZ + 2f^TZ = 0 \to Z^TL\underbrace{L^{-1}CL^{-T}}L^TZ + 2f^TZ \to$

$Z^TLQ\Lambda Q^TL^TZ + 2f^TL^{-T}L^TZ \to$

$\underbrace{Z^TLQ}_{W^T}\Lambda\underbrace{Q^TL^TZ}_{W} + 2\underbrace{f^TL^{-T}Q}_{g^T}\underbrace{Q^TLZ}_{W} \to W^T\Lambda W + 2g^TW = 0$

$\Rightarrow W^T\Lambda W + 2g^TW = 0 \quad , \quad g = Q^TL^{-1}f, \; W = Q^TLZ$ ★

(۲) $A^TA + \nu C \succeq 0 \to LL^T + \nu LL^{-1}CL^{-T}L^T \succeq 0 \longrightarrow$

$L(I + \nu Q\Lambda Q^T)L^T \succeq 0 \to LQ(I + \nu\Lambda)Q^TL^T \succeq 0$

چون L و Q هر دو معکوس پذیر اند ← فول رنک هستند و همچنین چون $x^TLQ(I+\nu C)Q^TL^Tx$

مثبت باید باشد به ازای هر x ← Q^TL^Tx تمام بردارها را کاور می‌کند ← $I+\nu C$

نیز باید PSD باشد ←

$$I + \nu C \succeq 0 \quad ★★$$

(۳) gradient of lagrangian = 0

$(A^TA + \nu C)Z = A^Tb - \nu f \to (LL^T + \nu C)Z = A^Tb - \nu f$

$\to L(I + \nu\overbrace{L^{-1}CL^{-T}}^{Q\Lambda Q^T})L^TZ = A^Tb - \nu f \longrightarrow$

$LQ(I + \nu\Lambda)\underbrace{Q^TL^TZ}_{W} = A^Tb - \nu f$

$$(I+\nu\Lambda)w = Q^T L^{-1} A^T b - \nu \underbrace{Q^T L^{-1} f}_{g}$$

$$\longrightarrow \quad (I+\nu\Lambda)w = h - \nu g \quad ***$$

حالا خواهیم داشت که:

$$g(\nu) = -(\underbrace{A^T b - \nu f}_{})^T (A^T A + \nu C)^\dagger (A^T b - \nu f) + \|b\|_2^2$$

$$-z^T(A^TA+\nu C)\,(A^TA+\nu C)^\dagger (A^TA+\nu C)\, z + \|b\|_2^2$$

$$= -z^T(A^TA+\nu C)z + \|b\|_2^2$$

$$z^T(A^TA+\nu C)z = z^T(LL^T+\nu C)z = \underbrace{z^T L Q}_{w^T}(I+\nu\Lambda)\underbrace{Q^T L^T z}_{w}$$

$$= w^T(I+\nu\Lambda)w$$

حالا از بالاتر داشتیم که:

$$(I+\nu\Lambda)w = h - \nu g \longrightarrow w = (I+\nu\Lambda)^\dagger (h-\nu g) \longrightarrow$$

$$w^T(I+\nu\Lambda)w = (h-\nu g)^T (I+\nu\Lambda)^\dagger (h-\nu g) =$$

$$g(\nu) = \sum_k \frac{(h_k - \nu g_k)^2}{1+\nu\lambda_k} \Rightarrow \frac{\partial g(\nu)}{\partial \nu} = 0 \longrightarrow$$

$$\sum_k \frac{2g_k(h_k-\nu g_k)}{1+\nu\lambda_k} + \frac{\lambda_k(h_k-\nu g_k)}{(1+\nu\lambda_k)^2} = 0$$

این معادله حل یک چند جمله‌ای بر حسب ν هست.

۳-۵.

دوال P_2 را می‌نویسیم:

$$\mathcal{L}(z, \lambda) = (Az-b)^T(Az-b) + \lambda(z^TCz + 2f^Tz)$$

$$\rightarrow \frac{\partial \mathcal{L}}{\partial z} = 0 \Rightarrow z(\lambda) = (A^TA + \lambda C)^{-1}(A^Tb - \lambda f)$$

$$g(\lambda) = -(A^Tb - \lambda f)^T(A^TA + \lambda C)^{-1}(A^Tb - \lambda f) + b^Tb$$

حالا با فرض $A^TA + \lambda C \succ 0$ باید این عبارت را پیدا کنیم.

حالا مشتق به λ می‌گیریم و λ^* را بدست می‌آوریم و با قرار دادن آن در g به $d^* = g(\lambda^*)$ خواهیم رسید.

چونکه گفته شده که شرایط KKT برقرار است $\Leftarrow$ پاسخ P_2 با دوگانش برابر خواهد بود یعنی $p^* = d^*$.

$$\frac{\partial g(\lambda)}{\partial \lambda} = -2(-f)^T(A^TA + \lambda C)^{-1}(A^Tb - \lambda f) +$$
$$-(A^Tb - \lambda f)(A^TA + \lambda C)^{-1} C (A^TA + \lambda C)^{-1}(A^Tb - \lambda f)$$

$$= 0 \longrightarrow$$

که اینم حل میشه، λ^* بدست میاد و جایگذاری میشه توی معادله اصلی و d^* درمیاد.

۶-۱. داریم که،

$$\hat{\theta} = g(T) \quad , \mathbb{E}\{T\} = \mu$$

$$\mathbb{E}\{\hat{\theta}\} = ?$$

ابتدا بسط تیلور حول μ را می‌نویسیم، خواهیم داشت که،

$$g(T) \approx g(\mu) + \nabla g(\mu)^T (T-\mu) + \frac{1}{2}(T-\mu)^T \nabla^2 g(\mu)(T-\mu)$$

$$\mathbb{E}\{g(T)\} = \mathbb{E}\{g(\mu)\} + \underbrace{\nabla g(\mu)^T \,\mathbb{E}\{T-\mu\}}_{0} + \frac{1}{2}\{((T-\mu)^T \nabla^2 g(\mu)(T-\mu))\}$$

$$= g(\mu) + \frac{1}{2}\mathbb{E}\left(tr\{\nabla^2 g(\mu)(T-\mu)(T-\mu)^T\}\right) =$$

$$g(\mu) + \frac{1}{2} tr\left(\nabla^2 g(\mu)\,\underbrace{\mathbb{E}\{(T-\mu)(T-\mu)^T\}}_{C_T}\right) =$$

$$= g(\mu) + \frac{1}{2} tr(\nabla^2 g(\mu) C_T) \longrightarrow \mathbb{E}\{\hat{\theta}\} \approx g(\mu) + \frac{1}{2} tr[\nabla^2 g(\mu) C_T]$$

۶-۲.

$$Var(\hat{\theta}) = Var(g(T)) \approx Var(\underbrace{g(\mu)}_{\text{ثابت}} + \nabla g(\mu)^T (T-\mu) + \frac{1}{2}(T-\mu)^T \nabla^2 g(\mu)(T-\mu))$$

$$= Var(\nabla g(\mu)^T (T-\mu)) + \frac{1}{4} Var((T-\mu)^T \nabla^2 g(\mu)(T-\mu)) + \underline{Cov(\nabla g(\mu)^T (T-\mu),}$$

$$\underline{\frac{1}{2}(T-\mu)^T \nabla^2 g(\mu)(T-\mu))} = Var(\nabla g(\mu)^T (T-\mu)) + \frac{1}{4} Var((T-\mu)^T \nabla^2 g(\mu)(T-\mu))$$

چون گاوسی است درجه فرد cov اش صفره

$$= \underline{\mathbb{E}\{(T-\mu)^T \nabla g(\mu) \nabla g(\mu)^T (T-\mu)\}} + \frac{1}{2}\mathbb{E}\{((T-\mu)^T \nabla^2 g(\mu)(T-\mu))^2$$

$$= \nabla g(\mu)^T \mathbb{E}\{(T-\mu)(T-\mu)^T\} \nabla g(\mu) + \frac{1}{2}\mathbb{E}\{ \text{———} \}$$

Scanned with CamScanner

$$\frac{1}{4}\mathbb{E}\Big\{ (T-\mu)^T \underbrace{\nabla^2 g(\mu)}_{A} \underbrace{(T-\mu)(T-\mu)^T}_{C_T} \underbrace{\nabla^2 g(\mu)(T-\mu)} - \mathbb{E}\{(T-\mu)^T \nabla^2 g(\mu)(T-\mu)\}^2 \Big\} =$$

$$\underbrace{\frac{1}{4}\Big(\mathbb{E}\{tr(\underbrace{\nabla^2 g(\mu)}_{A}\, (T-\mu)(T-\mu)^T \underbrace{\nabla^2 g(\mu)}_{A} (T-\mu)(T-\mu)^T)\}}_{\frac{1}{4}(2\,tr(A C_T A C_T))} - \mathbb{E}\{ — \quad — \}\Big) =$$

$$\Rightarrow$$

$$Var(\hat\theta) \approx \nabla g(\mu)^T C_T \nabla g(\mu) + \frac{1}{2} tr\big((\nabla^2 g(\mu) C_T)^2\big) \quad ✓$$

۳-۶. سمپل‌ها iid از توزیع نمایی با پارامتر θ در نظر می‌گیریم، خواهیم داشت که،

$$f(x|\theta) = \theta e^{-\theta x}, \qquad x > 0 \qquad \mathbb{E}\{x\} = \frac{1}{\theta}$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{\hat\theta_{MOM}} \rightarrow \hat\theta_{MOM} = \frac{1}{\bar{X}}$$

حالا آمارگان را می‌گذاریم $T = \bar{X}$ خواهیم داشت، $\hat\theta = g(T) = \frac{1}{T}$

$$E\{\hat\theta\} = g(\mu) + \frac{1}{2} tr\left(\nabla^2 g(\mu) \cdot \frac{C}{n}\right) \quad , \mu = \mathbb{E}\{T\} = \frac{1}{\theta}$$

$$C = Var(X_i) = \frac{1}{\theta^2}$$

$$g'(T) = \frac{-1}{T^2}, \quad g''(T) = \frac{2}{T^3} \rightarrow \nabla^2 g(\mu) = g''\left(\frac{1}{\theta}\right) = 2\theta^3$$

$$\rightarrow tr\left(\nabla^2 g(\mu) \cdot \frac{C}{n}\right) = 2\theta^3 \cdot \frac{1}{n\theta^2} = \frac{2\theta}{n}$$

$$\mathbb{E}\{\hat\theta\} \approx \theta + \frac{1}{2} \cdot \frac{2\theta}{n} = \theta + \frac{\theta}{n}$$

1

از بخش ۲.۶ داریم که: $Var(\hat{\theta}) \approx [\nabla g(\mu)]^T \cdot \frac{C}{n} \cdot \nabla g(\mu) + \frac{1}{2} tr\left(\left(\nabla^2 g(\mu) \cdot \frac{C}{n}\right)^2\right)$

$$[\nabla g(\mu)]^T \cdot \frac{C}{n} = 2\theta^3 \cdot \frac{1}{n\theta^2} = \frac{2\theta}{n}$$

$$tr\left(\left(\frac{2\theta}{n}\right)^2\right) = \frac{4\theta^2}{n^2} \longrightarrow Var(\hat{\theta}) \approx \frac{\theta^2}{n} + \frac{1}{2} \cdot \frac{4\theta^2}{n^2} = \frac{\theta^2}{n} + \frac{2\theta^2}{n^2}$$

برای n بزرگ ترم دوم عملاً نادیده گرفته خواهد شد.

حالا با حالت first-order مقایسه می‌کنیم:

$$\hat{\theta} \approx g(\mu) + \overbrace{g'(\mu)(T-\mu)}^{\text{E اش صفره}} \qquad E\{\hat{\theta}\} \approx g(\mu) = \theta$$

$$Var(\hat{\theta}) \approx [g'(\mu)]^2 \cdot Var(T) = \theta^4 \cdot \frac{1}{n\theta^2} = \frac{\theta^2}{n}$$

* برای مرتبه ۲، بایاس $\frac{\theta}{n}$ داریم و واریانس $\frac{\theta^2}{n} + \frac{2\theta^2}{n^2}$ داریم.
* برای مرتبه ۱، بایاس نداریم و واریانس $\frac{\theta^2}{n}$ داریم.

MLE Theory:

$$I(\theta) = -E\left\{\frac{\partial^2}{\partial\theta^2} \log f(X|\theta)\right\} = \frac{1}{\theta^2} \longrightarrow$$

$$\text{Assymptotic } Var(\hat{\theta}_{mle}) = \frac{1}{nI(\theta)} = \frac{\theta^2}{n}$$

برای n بزرگ $T = \bar{X}$ تقریباً توزیع گاوسی خواهد داشت،

for large n: $\bar{X} \sim N\left(\frac{1}{\theta}, \frac{1}{n\theta^2}\right)$

* برای mle در حالت مجانبی بایاس نداریم و واریانس θ^2/n خواهد بود.