StatMod 2 1/31/15 Mult. Norm. F

MAURICE DIESENDRUCK

MD

Using the result that ① x is mult. var. norm. iff it's an affine transformation of independent univariate normals

x is multivariate normal, $x \sim N_p(\mu, \Sigma)$, $E[e^{t'x}] = \exp\{t'\mu + t'\Sigma t/2\}$

? → $x = Az + \mu$, where $z \sim N_p(\underline{0}, I_p)$, $z = (z_1, \ldots, z_p)'$, $\phi(z_i)$ is standard Gaussian density,

A is $p \times p$ matrix. (other assumptions?) $\mu \in \mathbb{R}^p$

② PDF of standard mult. var. norm.

$$g(z) = \prod_{i=1}^{p} \phi(z_i) = (2\pi)^{-\frac{p}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{p} z_i^2\right\}, \quad z \in \mathbb{R}^p$$

$$= (2\pi)^{-\frac{p}{2}} \exp\left\{-\frac{1}{2} z'z\right\}$$

$$E[x] = E[Az+\mu] = AE[z] + \mu = A\cdot\underline{0} + \mu = \mu$$

$$\mathrm{cov}(x) = E\Big[\big((Az+\mu) - \underbrace{E[Az+\mu]}_{\mu}\big)\big((Az+\mu) - \underbrace{E[Az+\mu]}_{\mu}\big)'\Big] = E\big[(Az)(Az)'\big] = E[Az\,z'A']$$

$$= AE[zz']A' = A\,I\,A' = \boxed{AA' = \Sigma}$$

$\Sigma^{-1} = (A')^{-1}A^{-1} = (A^{-1})'A^{-1}$

If A (and Σ) is invertible, can do transformation that's 1-to-1 from $\mathbb{R}^p \to \mathbb{R}^p$

$z \to x$

$$x = Az + \mu \Rightarrow z = A^{-1}(x-\mu) = g^{-1}(x) \qquad dg^{-1}(z) = A^{-1}$$

$$f(x) = g(z)\cdot|\det(dg^{-1}(z))| = g(z = A^{-1}(x-\mu))\cdot|\det(A^{-1})|$$

$$= (2\pi)^{-\frac{p}{2}} \exp\left\{-\frac{1}{2}(A^{-1}(x-\mu))'(A^{-1}(x-\mu))\right\}\cdot[\det(A)]^{-1}$$

$$= (2\pi)^{-\frac{p}{2}} \exp\left\{-\frac{1}{2}(x-\mu)'A^{-1\prime}A^{-1}(x-\mu)\right\}\cdot|\Sigma|^{-\frac{1}{2}}$$

$$= (2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}, \quad x \in \mathbb{R}^p$$

$\det A = \sqrt{\det(A)\det(A')} = \sqrt{\det(AA')}$
$= \sqrt{\det(\Sigma)}$
$= \det(\Sigma)^{\frac{1}{2}}$
$= |\Sigma|^{\frac{1}{2}}$

? rank(A) = rank(Σ)
not so when A sing. or rect.

StatMod 2 1/31/15 Mult. Norm G

MD

Let $x_1 \sim N(\mu_1, \Sigma_1)$, $x_2 \sim N(\mu_2, \Sigma_2)$ $x_1 \perp x_2$

let $y = Ax_1 + Bx_2$ for A, B full column rank, appropriate dimension

$\dim(x_1) \neq \dim(x_2)$, $\dim(Ax_1) = \dim(Bx_2)$

Use mgf to identify distribution of y: For $t \in \mathbb{R}^p$

$$E[e^{t'y}] = E[e^{t'(Ax_1 + Bx_2)}] = E[e^{t'(Ax_1)} \cdot e^{t'(Bx_2)}] = E[e^{t'Ax_1}] \cdot E[e^{t'Bx_2}]$$

$$= E[e^{s_1'x_1}] \cdot E[e^{s_2'x_2}]$$

$s_1' = t'A$, $s_1 = (t'A)' = A't$ | $s_2' = t'B$, $s_2 = (t'B)' = B't$

$$= \exp\left\{ s_1'\mu_1 + s_1'\Sigma_1 s_1/2 \right\} \cdot \exp\left\{ s_2'\mu_2 + s_2'\Sigma_2 s_2/2 \right\}$$

$$= \exp\left\{ s_1'\mu_1 + s_2'\mu_2 + s_1'\Sigma_1 s_1/2 + s_2'\Sigma_2 s_2/2 \right\}$$

$$= \exp\left\{ t'A\mu_1 + t'B\mu_2 + (t'A)\Sigma_1(A't)/2 + (t'B)\Sigma_2(B't)/2 \right\}$$

$$= \exp\left\{ t'(A\mu_1 + B\mu_2) + t'(A\Sigma_1 A' + B\Sigma_2 B')t/2 \right\}$$

$$y \sim N\left(A\mu_1 + B\mu_2,\ A\Sigma_1 A' + B\Sigma_2 B' \right)$$

Stat Mod 2 2/1/15 Mult. Norm - Cond/Marg A MD

Let x be mult. var. norm. : $x \sim N(\mu, \Sigma)$

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \begin{matrix} x_1 = (x_1, \dots, x_k) \\ x_2 = (x_{k+1}, \dots, p) \end{matrix}, \qquad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \qquad \Sigma_{21} = \Sigma_{12}^T$$

$\mu_1 = (\mu_1, \dots, \mu_k)$

$\mu_2 = (\mu_{k+1}, \dots, \mu_p)$

$\underset{k \times k}{\Sigma_{11}} \quad \underset{k \times (p-k)}{\Sigma_{12}} \quad \underset{(p-k) \times k}{\Sigma_{21}} \quad \underset{(p-k) \times (p-k)}{\Sigma_{22}}$

Derive marginal of x_1 :

1. Result of affine transformation : $x \sim N(\mu, \Sigma)$

$$\Rightarrow Ax + b \sim N(A\mu + b, \; A \Sigma A')$$

Here, consider $A = (\underset{k \times k}{I}, \underset{k \times (p-k)}{0})$, $b = \underset{k \times 1}{\underline{0}}$: $Ax + b = (1, 0)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = x_1$

$$\Rightarrow Ax + b \sim N\left((1,0)\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; \underbrace{(1,0)\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)$$

$$x_1 \sim N\left(\mu_1, \; \Sigma_{11} \right) \qquad (\Sigma_{11} \;\; \Sigma_{12})\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \Sigma_{11}$$

$\blacksquare$

$$f(x_1) = (2\pi)^{-\frac{k}{2}} |\Sigma_{11}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (x_1 - \mu_1)^T \Sigma_{11}^{-1} (x_1 - \mu_1) \right\}$$

$$k\left(\begin{matrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix} \;\; \begin{matrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{matrix} \right) \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_2 \\ \vdots \end{pmatrix} \Big\} p = \begin{pmatrix} \mu_1 \\ \vdots \\ \vdots \end{pmatrix} \text{ then add } \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \mu_1$$

$$\underbrace{\underbrace{\quad}_{k} \; \underbrace{\quad}_{p-k}}_{p}$$

$$\underset{k \times p}{A} \times \underset{p \times 1}{\mu} = \underset{k \times 1}{\mu_1} \text{ then add } \underset{k \times 1}{\underline{0}}$$

$$k\left(\underbrace{\begin{matrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix}}_{k}\ \underbrace{\begin{matrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{matrix}}_{p-k}\right) \begin{pmatrix} \underset{k}{\boxed{\Sigma_{11}}} & \underset{p-k}{\boxed{\Sigma_{12}}} \\ \underset{k}{\boxed{\Sigma_{21}}} & \underset{p-k}{\boxed{\Sigma_{22}}} \end{pmatrix} = \begin{pmatrix} \boxed{\Sigma_{11}} & \boxed{\Sigma_{12}} \end{pmatrix}$$

with the rows of Σ labelled k and $p-k$.

$$\underset{k \times p}{A} \quad \times \quad \underset{p \times p}{\Sigma} \qquad \qquad \underset{k \times p}{A\Sigma}$$

$$k\begin{pmatrix} \underset{k}{\boxed{\Sigma_{11}}} & \underset{p-k}{\boxed{\Sigma_{12}}} \end{pmatrix} \begin{pmatrix} \begin{matrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix} \\ \begin{matrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{matrix} \end{pmatrix} \begin{matrix} \}\,k \\ \\ \}\,p-k \end{matrix} = \begin{pmatrix} \boxed{\Sigma_{11}} \end{pmatrix}$$

$$\underset{k \times p}{A\Sigma} \quad \times \quad \underset{p \times k}{A'} \quad = \quad \underset{k \times k}{A\Sigma A'}$$

Stat Mod 2 2/1/15 Mult. Norm. Cond/Marg B

Let $\Omega = \Sigma^{-1}$ be the inverse covariance matrix
("precision matrix")

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Sigma\Omega = I_p$$

$$\underbrace{\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}}_{\Omega} = \underbrace{\begin{pmatrix} I_K & 0 \\ 0 & I_{p-K} \end{pmatrix}}_{I_p}$$

(a) $\Sigma_{11}\Omega_{11} + \Sigma_{12}\Omega_{21} = I_K$

(b) $\Sigma_{21}\Omega_{11} + \Sigma_{22}\Omega_{21} = 0 \quad \Rightarrow \quad \Omega_{21} = -\Sigma_{22}^{-1}\Sigma_{21}\Omega_{11}$ (e)

(c) $\Sigma_{11}\Omega_{12} + \Sigma_{12}\Omega_{22} = 0 \quad \Rightarrow \quad \Omega_{12} = -\Sigma_{11}^{-1}\Sigma_{12}\Omega_{22}$ (f)

(d) $\Sigma_{21}\Omega_{12} + \Sigma_{22}\Omega_{22} = I_{p-K}$

Plug into (a): $\Sigma_{11}\Omega_{11} + \Sigma_{12}(-\Sigma_{22}^{-1}\Sigma_{21}\Omega_{11}) = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\Omega_{11} = I_K$

into (d): $\Sigma_{21}(-\Sigma_{11}^{-1}\Sigma_{12}\Omega_{22}) + \Sigma_{22}\Omega_{22} = (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\Omega_{22} = I_{p-K}$

Isolating Ω_{11}, Ω_{22}: $\Omega_{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$, $\Omega_{22} = (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}$

Plug into (e), (f): $\Omega_{21} = -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$, $\Omega_{12} = -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}$

Stat Mod 2 2/1/15 Mult. Norm Cond/Marg. C MD

Derive conditional distribution for $f(x_1|x_2)$ in terms of partitioned elements of x, μ, and Σ.

Define conditional: $f(x_1|x_2) = \dfrac{f(x_1, x_2)}{f(x_2)} = \dfrac{\text{joint}}{\text{marginal}}$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ x_p \end{pmatrix} \begin{matrix} \left.\right\} p-k \quad x_1 \\ \left.\right\} k \quad x_2 \end{matrix}$$

Joint: $f(x) = f(x_1, x_2) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$

Marginal: $f(x_2) = (2\pi)^{-\frac{k}{2}} |\Sigma_{22}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2)\right\}$

Look at joint, and unwrap $(x-\mu)'\Sigma^{-1}(x-\mu)$. Immediately below, expansion w/ $\mu = \underline{0}$.

$$x'\Sigma^{-1}x = (x_1\ x_2)\begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22}\end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1'\Omega_{11}x_1 + x_2'\Omega_{21}x_1 + x_1'\Omega_{12}x_2 + x_2'\Omega_{22}x_2 \quad (*)$$

? join terms?

$$= x_1'\Omega_{11}x_1 + 2(x_1'\Omega_{12}x_2) + x_2'\Omega_{22}x_2$$

Complete the square in x_1: Looking for $f(x_1|x_2) = \dots \exp\left\{-\frac{1}{2}(x_1-m)'M^{-1}(x_1-m)\right\}$

$$x'\Sigma^{-1}x = (x_1-m)'M^{-1}(x_1-m) + c$$

$(*)$ $(x-\mu)'\Sigma^{-1}(x-\mu) = (x_1-\mu_1)'\Omega_{11}(x_1-\mu_1)$
$+ 2(x_1-\mu_1)'\Omega_{12}(x_2-\mu_2)$
$+ (x_2-\mu_2)'\Omega_{22}(x_2-\mu_2)$

Want form $x^2 - 2bx + c = (x-b)^2 - b^2 + c$, for x_1.

Quadratic term is $(x_1-\mu_1)'\Omega_{11}(x_1-\mu_1)$, so $M^{-1} = \Omega_{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$

? Linear term is $2(x_1-\mu_1)'\Omega_{12}(x_2-\mu_2)$, want $-2bx_1$, so $b = -$ coefficient of x_1

$$\cancel{2}(x_1-\mu_1)'\Omega_{12}(x_2-\mu_2) = -\cancel{2}b(x_1)$$

$$-(x_1-\mu_1)'\Omega_{12}(x_2-\mu_2) = b(x_1)$$

$$-(x_1-\mu_1)'\left[-\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\right](x_2-\mu_2)$$

Rewrite joint distribution quadratic using Schur complement.

$$x'\Sigma^{-1}x \quad , \qquad \Sigma^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ -\Sigma_{12}^{-1}\Sigma_{21} & I \end{pmatrix} \begin{pmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix}$$

joint $x'\Sigma^{-1}x$ can now be re-written as:

$$f(x_1, x_2) \propto \exp\left\{ -\frac{1}{2} (x_1 \;\; x_2) \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix} \begin{pmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\}$$

$$\propto \exp\left\{ -\frac{1}{2} \left(x_1 - \Sigma_{12}\Sigma_{22}^{-1}x_2 \,,\; x_2 \right) \begin{pmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} x_1 - \Sigma_{12}\Sigma_{22}^{-1}x_2 \\ x_2 \end{pmatrix} \right\}$$

? why does the Σ_{21} move to front and become Σ_{12}?

$$\propto \exp\left\{ -\frac{1}{2} \left(x_1 - \Sigma_{12}\Sigma_{22}^{-1}x_2 \right)' \left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)^{-1} \left(x_1 - \Sigma_{12}\Sigma_{22}^{-1}x_2 \right) + x_2'\Sigma_{22}^{-1}x_2 \right\}$$

$$\propto \exp\left\{ -\frac{1}{2} (x_1 - m)' M^{-1} (x_1 - m) \right\} \quad \text{where } m = \Sigma_{12}\Sigma_{22}^{-1}x_2, \quad M = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Adding back μ_1, μ_2: $\propto \exp\left\{ -\frac{1}{2} \left[\underbrace{(x_1 - \mu_1) - m}_{x_1 - (\mu_1 + m)} \right]' M^{-1} \left[\underbrace{(x_1 - \mu_1) - m}_{x_1 - (\mu_1 + m)} \right] \right\}$

$$\boxed{\text{So } f(x_1 | x_2) \sim N_k \left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \; \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)}$$

* Comment on regression!

Stat Mod 2 2/1/15 Mult. Norm. Cond/Marg. C (in class) MD

$$P(x_1, x_2) = P(x_1 | x_2)\ P(x_2)$$

$$\log P(x_1, x_2) = \log(P(x_1|x_2)) - \log(P(x_2))$$

$$= -\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu) + \frac{1}{2}(x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2) + C_1$$

MD

Stat Mod 2 2/2/15 Mult Regression A.

Show that Least Squares, MLE under Gaussianity, Method of Moments lead to same estimator of β. Given $y_i = x_i^T\beta + \varepsilon_i$, $i = 1, \dots, n$.

<u>Least $\square$s</u>: Take derivative w.r.t. β to find minimum. Trying to minimize $SSE = e'e$ where $e = Y - X\beta$.

$$SSE = e'e = (Y - X\beta)'(Y - X\beta) = (Y' - \beta'X')(Y - X\beta)$$

$$= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta = Y'Y - 2Y'X\beta + \beta'X'X\beta = (*)$$

Take partial derivative $\frac{\partial}{\partial\beta}$: $\frac{\partial}{\partial\beta}\Big((*)\Big) = (-2Y'X)' + 2X'X\beta$ } when $X'X$ is symmetrix

Set to zero: $(-2Y'X)' + 2X'X\beta = 0 \Rightarrow 2X'X\beta = 2X'Y$

$$\Rightarrow X'X\beta = X'Y \Rightarrow (X'X)^{-1}X'X\beta = (X'X)^{-1}X'Y$$

$$\Rightarrow \hat{\beta} = (X'X)^{-1}X'Y$$

<u>MLE</u>: Choose $\hat{\beta}$ to maximize likelihood. $\mathcal{L} = \prod_{i=1}^{n} N(y_i \mid x_i, \beta, \sigma^2) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(Y_i - x_i^T\beta)^2\right\}$

$$\mathcal{L} = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}|Y - X\beta|^2\right\} = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\}$$

$$\log(\mathcal{L}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta) = L^*$$

$$\frac{\partial}{\partial\beta}(L^*) = -\frac{1}{2\sigma^2}\left((-2Y'X)' + 2X'X\beta\right) = 0$$

$$\cancel{2}X'X\beta = \cancel{2}X'Y$$

$$\Rightarrow \hat{\beta} = (X'X)^{-1}X'Y$$

↗

Method of Moments:

Choose $\hat{\beta}$ so that sample covariance btwn errors and each of p predictors $= 0$.

$$\operatorname{cov}(Y_i - X_i'B,\ X_i) = 0 \qquad \varepsilon_i = Y_i - X_i'\beta \qquad X_i'\beta \in C(X)$$

$$E[\varepsilon_i] = 0 \qquad \varepsilon_i = Y_i - X_i'B \in C(X)^{\perp}$$

For the sample: $E[X_i \varepsilon_i] = 0$

$$E[X_i\varepsilon_i] = \frac{1}{n}\sum_{i=1}^{n} X_i(Y_i - X_i'B) = 0$$

$$= \sum_{i=1}^{n} X_i(Y_i - X_i'B) = 0 \Rightarrow X'Y - X'XB = 0$$

$$\Rightarrow X'Y = X'XB \Rightarrow B = (X'X)^{-1}X'Y$$

$$\operatorname{cov}(Y_i - X_i'B, X_i) = E\Big[\big((Y_i - X_i'B) - E[Y_i - X_i'B]\big)\big(X_i - E[X_i]\big)\Big] = E\Big[\big(\varepsilon - E[\varepsilon]\big)\big(X_i - E[X_i]\big)\Big]$$

$$= E\Big[(\varepsilon - 0)(X - E[X_i])\Big] = E\Big[(Y_i - X_i'B)(X - E[X])\Big] = 0$$

equals 0? or $= 0$ wlog?

$$= E\Big[(Y_i - X_i'B)(X_i)\Big] = 0$$

$$= \frac{1}{n}\Big[\sum_{i=1}^{n} X_{ti}(y_t - X_t B)\Big] = 0$$

$$\Rightarrow X^T(y - XB) = 0$$

$$\Rightarrow X^T y = X^T X B$$

$$\Rightarrow \hat{B} = (X^T X)^{-1} X^T y$$

Stat Mod 2 2/3/15 Mult Regression B MD

Derive estimator under weighted sum of squared errors:

$\sigma_i^2 \propto \frac{1}{w_i}$

$\sigma_i^2 = \frac{\sigma^2}{w_i}$

Motivation: ① Want to fit well where σ_i^2 is small and expect to fit poorly where σ_i^2 is large.

② Or, fit well where we designate a priority

$\sigma_i^2 \cdot w_i = \sigma^2$

$w_i = \frac{\sigma^2}{\sigma_i^2}$

$w^{\frac{1}{2}} = \frac{\sigma}{\sigma_i}$

$$\hat{\beta} = \min_{\beta} \left\{ \sum_{i=1}^{n} w_i (y_i - x_i'\beta)^2 \right\}$$

Let $Q = \sum_{i=1}^{n} w_i (y_i - x_i'B)^2 = \sum_{i=1}^{n} \varepsilon_i^2$

$= (Y_i - X_i'B)'(Y_i - X_i'B)\, w$

$$Q = \sum_{i=1}^{n} \left(w^{\frac{1}{2}}(y_i - x_i'\beta)\right)^2$$

$= (Y_i - XB)'(Y_i - XB)\, w$

$= (Y'Y - Y'XB - B'X'Y + B'X'XB)\, w$

$= (Y'Y - 2Y'XB + B'X'XB)\, w \Rightarrow \frac{\partial Q}{\partial B} = 0 = \left[(-2Y'X)' + 2X'XB\right] w$

$Y_i = x_i'B + \varepsilon_i$

$\sigma_i^2 w_i = \sigma^2 \Rightarrow Var(\varepsilon_i)\, w_i = \sigma^2 \Rightarrow var(w^{\frac{1}{2}}\varepsilon_i) = \sigma^2$

where σ^2 is a fixed constant

$\Downarrow$

$w_i^{\frac{1}{2}} \cdot Y_i = w_i^{\frac{1}{2}} x_i'B + w_i^{\frac{1}{2}}\varepsilon_i$

$\Downarrow$

$Y_i^* = X_i^{*\prime}B + \varepsilon_i^*$

Then, minimizing $\sum_{i=1}^{n}(Y_i^* - X_i^{*\prime}B)^2$ is same as minimizing $\sum_{i=1}^{n}\left(w_i(Y_i - X_i'B)^2\right)$

where $Y_i^* = w_i^{\frac{1}{2}}Y_i$, $X_i^* = w_i^{\frac{1}{2}}x_i$, $\varepsilon_i^* = w_i^{\frac{1}{2}}\varepsilon_i$

So: $\hat{\beta} \sim N_{p\times 1}(\beta, \sigma^2(X'X)^{-1})$

$\beta \sim$ M.V.N.

Suppose $w_i = \frac{1}{\sigma_i^2} \Rightarrow w_i \sigma_i^2 = 1 \equiv w_i Var(\varepsilon_i) = 1$

$\Rightarrow Var(w^{\frac{1}{2}}\varepsilon_i) = 1$ which means that $\varepsilon_i^* = w^{\frac{1}{2}}\varepsilon_i$ has fixed variance, i.e. is homoscedastic, and can be minimized by OLS.

To minimize $\sum_{i=1}^{n} w_i(Y_i - X_i'B)^2$ in terms of B, rewrite with transformation

$Y_i^* = w_i^{\frac{1}{2}} Y_i$, $X_i^* = w_i^{\frac{1}{2}} X_i$, $\varepsilon_i^* = w_i^{\frac{1}{2}}\varepsilon_i$

Now, to minimize over B $\sum_{i=1}^{n}(Y_i - X_i'B)^2$, use OLS to get $\hat{\beta} = (X'X)^{-1}X'Y$.

$w_i^{\frac{1}{2}}\left[Y_i = X_i'B + \varepsilon_i\right] \Rightarrow w_i^{\frac{1}{2}}Y_i = w_i^{\frac{1}{2}}X_iB + w_i^{\frac{1}{2}}\varepsilon_i \Rightarrow Y_i^* = X_i^*\beta + \varepsilon^*$,

$w_i(Y_i - X_i'B) = w_i\varepsilon_i$ $\downarrow$ homoscedastic

where B can also be minimized using OLS, and $\hat{\beta} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$

$$\Rightarrow \hat{\beta} = \left[(w_i^{\frac{1}{2}}X_i)'(w_i^{\frac{1}{2}}X_i)\right]^{-1}(w_i^{\frac{1}{2}}X_i)'(w_i^{\frac{1}{2}}Y_i)$$

$$= \left[X_i' w_i^{\frac{1}{2}\prime} w_i^{\frac{1}{2}} X_i\right]^{-1}\left(X_i' w_i^{\frac{1}{2}\prime} w_i^{\frac{1}{2}} Y_i\right) = \left[X_i' w_i X_i\right]^{-1}\left(X_i' w_i Y_i\right)$$

<u>MLE approach:</u> $Y_i \sim N(X_iB, \sigma_i^2)$ Known σ_i's, but $\sigma_i^2 \neq \sigma_j^2$, Heteroscedastic Gaussian error

$$p(Y_i|\beta,\sigma_i^2) = f_i = (2\pi\sigma_i^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\cdot\frac{1}{\sigma_i^2}(Y_i - X_iB)^2\right\}$$

Likelihood: $\prod_{i=1}^{n} f_i = (2\pi)^{-\frac{n}{2}} \cdot \prod_{i=1}^{n}(\sigma_i^2)^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}\cdot\sum_{i=1}^{n}\frac{1}{\sigma_i^2}(Y_i - X_i'B)^2\right\}$

Max Lik w.r.t. B is equiv. to minimizing $\sum_{i=1}^{n}\frac{1}{\sigma_i^2}(Y_i - X_i'B)^2$ and where $w_i = \frac{1}{\sigma_i^2}$,

$$\equiv \min_{\beta}\left\{\sum_{i=1}^{n} w_i(Y_i - X_i'B)^2\right\} \equiv \text{minimize weighted sum of squared errors.}$$

Stat Mod 2 2/3/15 Quantify Uncertainty: Lin Reg. A MD

Observe data from linear regression model w/ Gaussian error. Y is M.V.N. ?yes?

$y = X\beta + \varepsilon \quad , \quad \varepsilon \sim N(0, \sigma^2 I)$

Derive sampling distribution of $\hat{\beta} = (X'X)^{-1}X'Y$

$E[b] = B$

$\sigma^2(b) = \dfrac{\sigma^2}{X'X}$

$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \varepsilon) = (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'\varepsilon$

$= \beta + (X'X)^{-1}X'\varepsilon$

$E[\hat{\beta}|X] = E[\beta + (X'X)^{-1}X'\varepsilon | X] = \beta + E[(X'X)^{-1}X'\varepsilon | X]$

$= \beta + 0 = \beta$

No correlation btwn ε_i and x_i, so $E[\varepsilon_i | x_{1-N}] = 0$

$\mathrm{cov}(\hat{\beta}|X) = E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)' | X]$

$= E[(\beta + (X'X)^{-1}X'\varepsilon - \beta)(\beta + (X'X)^{-1}X'\varepsilon - \beta)' | X]$

$= E[((X'X)^{-1}X'\varepsilon)((X'X)^{-1}X'\varepsilon)' | X]$

$= E[(X'X)^{-1}X'\varepsilon \cdot \varepsilon'X(X'X)^{-1'} | X]$

$= (X'X)^{-1}X' \cdot E[\varepsilon\varepsilon' | X] \cdot X(X'X)^{-1}$

$= (X'X)^{-1}X' \cdot \sigma^2 \cdot I \cdot X(X'X)^{-1}$

$= \sigma^2 \underbrace{(X'X)^{-1}X'X}_{I}(X'X)^{-1}$

$= \sigma^2 (X'X)^{-1}$

why doesn't $(X'X)^{-1}$ get transposed?

$\because (X'X)' = X'X$ symmetric

and

$[(X'X)^{-1}]' = [(X'X)']^{-1}$ inverse + trans. commute

$\Downarrow$

$[(X'X)^{-1}]' = [(X'X)']^{-1} = (X'X)^{-1}$

$E[\varepsilon\varepsilon'|X] = \mathrm{cov}(\varepsilon) = \sigma^2 I$

?

Assume: no correlation of errors + homoscedastic errors

If $Y \sim$ M.V.N., $\beta \sim$ M.V.N.

So: $\hat{\beta} \sim N_{p \times 1}(\beta, \sigma^2(X'X)^{-1})$

Stat Mod 2 Quant Uncertainty : Lin Reg. B

MD

Sampling Distr for $\underset{(p+1)\times 1}{\hat{\beta}}$ is $\hat{\beta} \sim N_{p+1}\left(\beta, \sigma^2(x'x)^{-1}\right)$, depends on σ^2, yet σ^2 is unknown.

Also know: ① $cov(\hat{\beta}) = \sigma^2 (x'x)^{-1} = s.e.^2 (x'x)^{-1}$

② Since $\hat{\beta}$ is multivariate normal, $\hat{\beta}_j$ is univariate normal

Propose strategy for finding s.e. of β_j.

Using ①, $cov(\hat{\beta}) = \sigma^2 (x'x)^{-1}$, since σ^2 is a fixed value,

plucking j-jth value from $(x'x)^{-1}$ $\Rightarrow$ $var(\hat{\beta}_j) = \sigma^2 \left[(x'x)^{-1}\right]_{jj}$

And $\sigma_{\hat{\beta}_j} = \sqrt{\sigma^2 \left[(x'x)^{-1}\right]_{jj}}$. <u>BUT</u> σ^2 is still unknown!

(Remember: $\sigma^2 = \frac{SSE}{n-(p+1)}$)

$$\Rightarrow \sigma_{\hat{\beta}_j} = \sqrt{\frac{SSE}{n-(p+1)} \cdot \left[(x'x)^{-1}\right]_{jj}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i'\hat{\beta})^2}{n-(p+1)} \left[(x'x)^{-1}\right]_{jj}}$$

In the Ozone dataset, $\underset{203\times10}{X}$ $\begin{array}{cc} & 10 \\ 203 & \left[\quad\right] \end{array}$ and $\underset{10\times1}{\hat{\beta}}$ $\begin{array}{cc} & 1 \\ 10 & \left[\quad\right] \end{array}$, so $\underset{203\times1}{X\beta}$:,

$$\text{betacov} = \frac{\sum_{i=1}^{n} (Y_i - X_i B)^2}{n-(p+1)} (X'X)^{-1}$$

To compare, do betacov - betacovlm. See that values are all $< 1\times10^{-9}$

```
# Load the library
# you might have to install this the first time
library(mlbench)

# Load the data
ozone = data(Ozone, package='mlbench')

# Look at the help file for details
?Ozone

# Scrub the missing values
# Extract the relevant columns
ozone = na.omit(Ozone)[,4:13]

y = ozone[,1]
x = as.matrix(ozone[,2:10])

# add an intercept
x = cbind(1,x)

# compute the estimator
betahat = solve(t(x) %*% x) %*% t(x) %*% y

# Fill in the blank
pred <- x %*% betahat
sum.squared.errors <- sum((y-pred)^2)
n <- length(y)
p <- length(betahat) - 1
beta.variance <- sum.squared.errors / (n-(p+1))
xtx <- t(x) %*% x
xtx.inv <- solve(xtx)
betacov = beta.variance * xtx.inv

# Now compare to lm
# the 'minus 1' notation says not to fit an intercept (we've already hard-coded it as an extra
column)
lm1 = lm(y~x-1)

summary(lm1)
betacovlm = vcov(lm1)
sqrt(diag(betacovlm))
```