Statistics 840 Lecture 2

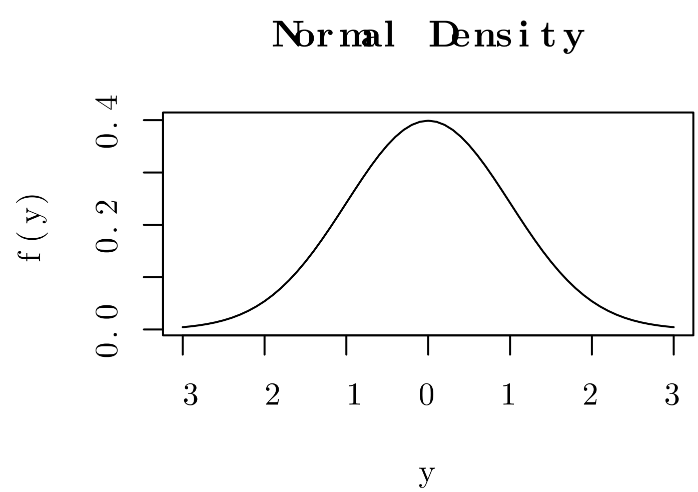
Some Notation:
Y = name of a random variable
y = value of the random variable after it has been observed
Sometimes notation is abused by writing y instead of Y.

$Y \sim F$ means "The random variable Y has the distribution F"
$Y \sim F_Y$ means "The random variable Y has the distribution F_Y"

F the normal distribution: $\mathcal{N}(\mu, \sigma^2)$.

The density of the normal distribution, with mean μ and standard deviation σ is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$



$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ A vector of two random variables, jointly Normally distributed.

$$Y \sim \mathcal{N}(\mu, \Sigma), \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

$$\sigma_{21} = \sigma_{12}, \quad \sigma_{11}\sigma_{22} - \sigma_{21}\sigma_{12} > 0$$

Σ symmetric positive definite, $\Sigma \succ 0$.

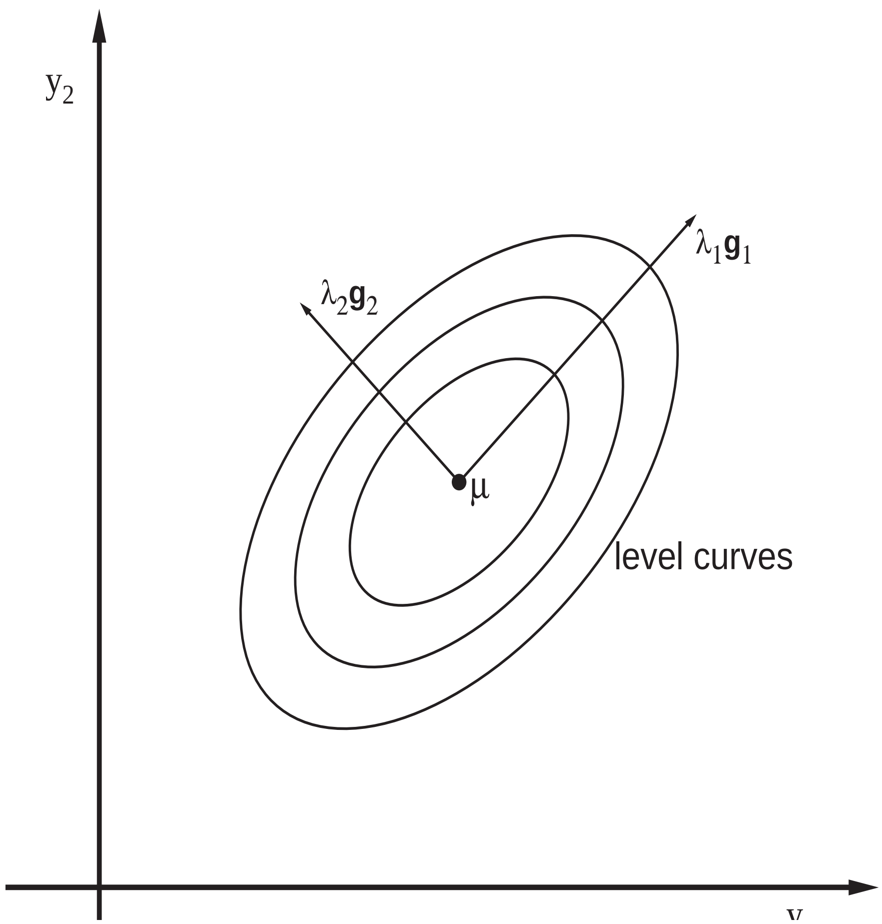
$$F_Y : y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} f_Y(y) = \frac{1}{2\pi \det \Sigma} e^{-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)}$$

The exponent is a quadratic form in y

Eigenvalue-eigenvector decomposition of $\Sigma = \Gamma D \Gamma'$, where $\Gamma\Gamma' = I = \Gamma'\Gamma$ and

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Axes are along eigenvectors $\mathbf{g}_1$ and $\mathbf{g}_2$ stretched according to eigenvalues λ_1 and λ_2.



Multivariate Normal Distribution

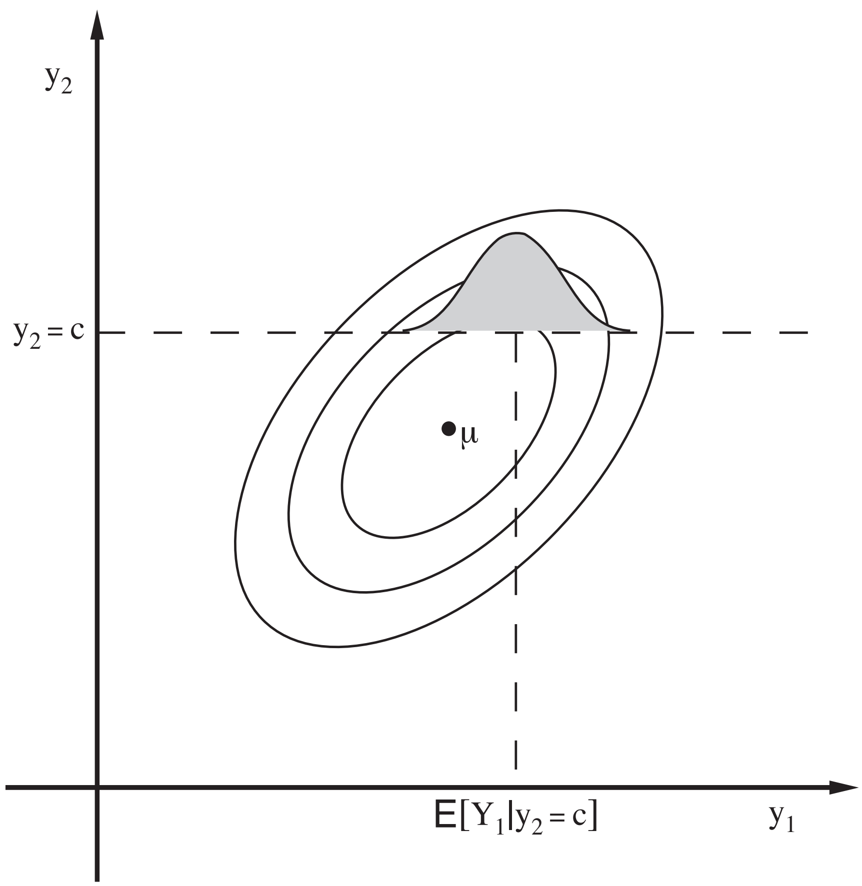
$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma)$$

n jointly normal random variables.

Conditional distribution of $Y_1|y_2$ is

$$\mathcal{N}\left(\mu_1 + \sigma_{12}\sigma_{22}^{-1}(y_2 - \mu_2), \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21}\right).$$

In particular, if $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, then $EY_1|y_2 = \sigma_{12}\sigma_{22}^{-1}y_2$.



More generally, let

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \sim \mathcal{N}(0, \Sigma)$$

Let

$$Y = \begin{pmatrix} U \\ V \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_{n_1} \end{pmatrix}, \quad V = \begin{pmatrix} V_1 \\ \vdots \\ V_{n_2} \end{pmatrix}$$

Conditional expectation of U given $V = v$ is

$$E(U \mid v) = \Sigma_{12}\Sigma_{22}^{-1}v$$

(Standard multivariate Gaussian result, Anderson (1958), Thm 2.5.1, see anderson1958.pdf or mvnormaldist.pdf in the pdf2 directory), Wilks, etc.)

Let U be 1-dimensional. Then Σ_{11} a number, Σ_{12} a row vector

$$E(U \mid v_1, \ldots, v_n) = \sum a_\ell v_\ell$$

Find $a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ to minimize

$$E \left\| U - \sum a_\ell v_\ell \right\|^2 = \Sigma_{11} - 2a'\Sigma'_{12} + a'\Sigma_{22}a$$

$$\frac{d}{da}: \qquad -2\Sigma'_{12} + 2\Sigma_{22}a = 0$$

$$a = \Sigma_{22}^{-1}\Sigma'_{12}$$

$$E(U \mid v_1, \ldots, v_n) = a'v = \Sigma_{12}\Sigma_{22}^{-1}v$$

(This least squares calculation is unique to multivariate normal distributions)

Duality between Bayes estimates and optimization problems:

$$y_i = \int_0^1 K(t(i), s) f(s) ds + \epsilon_i \qquad i = 1, \ldots, n$$

want to estimate f

$$y_n = \underbrace{X_{n \times p}}_{K} \underbrace{\beta_p}_{f} + \epsilon \qquad \epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$$

X has rank $q \leq p$, $\beta \sim \mathcal{N}(0, b\Sigma_{p \times p})$

To look at this as a Bayes estimate:

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

$E(U|v) = \Sigma_{12}\Sigma_{22}^{-1} v$

$$\begin{pmatrix} \beta \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} b\Sigma & b\Sigma X' \\ bX\Sigma & bX\Sigma X' + \sigma^2 I \end{pmatrix} \right)$$

$$\begin{aligned} \widehat{\beta} = E(\beta \mid y) &= b\Sigma X'(bX\Sigma X' + \sigma^2 I)^{-1} y \\ &= \Sigma X'(X\Sigma X' + \tfrac{\sigma^2}{b} I)^{-1} y \end{aligned}$$

Optimization problem: Minimize

$$\frac{1}{n}\underbrace{\|y - X\beta\|^2}_{\text{Euclidean norm}} + \lambda\beta'\Sigma^{-1}\beta \qquad (*)$$

$$= \frac{1}{n}(y'y - 2y'X\beta + \beta'X'X\beta + n\lambda\beta'\Sigma^{-1}\beta)$$

$$\frac{\partial}{\partial\beta}: \quad -2X'y + 2X'X\beta + 2n\lambda\Sigma^{-1}\beta = 0$$

gives

$$\tilde{\beta} = (X'X + \lambda\Sigma^{-1})^{-1}X'y$$

Lemma

$$\underbrace{(X'X + n\lambda\Sigma^{-1})^{-1}}_{p\times p}\underbrace{X'}_{p\times n} = \underbrace{\Sigma}_{p\times p}\underbrace{X'}_{p\times n}\underbrace{(X\Sigma X' + n\lambda I)^{-1}}_{n\times n}$$

provided $X'X + n\lambda\Sigma^{-1} > 0$

Proof

$$\begin{aligned} X' &\overset{?}{=} (X'X + n\lambda\Sigma^{-1})\Sigma X'(X\Sigma X' + n\lambda I)^{-1} \\ &= (X'X\Sigma + n\lambda I)X'(X\Sigma X' + n\lambda I)^{-1} \\ &= X'(X\Sigma X' + n\lambda I)(X\Sigma X' + n\lambda I)^{-1} \end{aligned}$$

OK. So

$$
\begin{aligned}
\tilde{\beta} &= (X'X + n\lambda\Sigma^{-1})X'y = \Sigma X'(X\Sigma X' + n\lambda I)^{-1}y \\
&= \widehat{\beta}
\end{aligned}
$$

THE BOTTOM LINE:

The minimizer of the optimization problem $(*)$ is a Bayes estimate if $n\lambda = \sigma^2/b$, the “noise to signal” ratio

Question for self, to be studied later—improper priors:

$$\Sigma = \begin{pmatrix} \Sigma_0 & 0 \\ 0 & \xi\Sigma_1 \end{pmatrix} \qquad \Sigma^{-1} = \begin{pmatrix} \Sigma_0^{-1} & 0 \\ 0 & \frac{1}{\xi}\Sigma_1^{-1} \end{pmatrix}$$

$\xi \to \infty$ Improper prior $\frac{1}{\xi} \to 0$

$$\lim_{\xi\to\infty} \beta'\Sigma^{-1}\beta = \beta_0'\Sigma_0^{-1}\beta_0 \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Optimization problem OK provided $X'X + n\lambda\Sigma^{-1} \succ 0$. Take limits, will be OK

To solve integral equations: (Tihonov) regularization:

$$\frac{1}{n}\|y - X\beta\|^2 + \beta'\Sigma^{-1}\beta$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \int K(t(i),s)f(s)ds\right)^2 + \underbrace{\lambda J(f)}_{\lambda\beta'\Sigma^{-1}\beta}$$

$J(f) = \int (f''(t))^2 dt$ a quadratic form in a function space—leads to a matrix not of full rank

Geometry of Euclidean space E_n, projections, duality

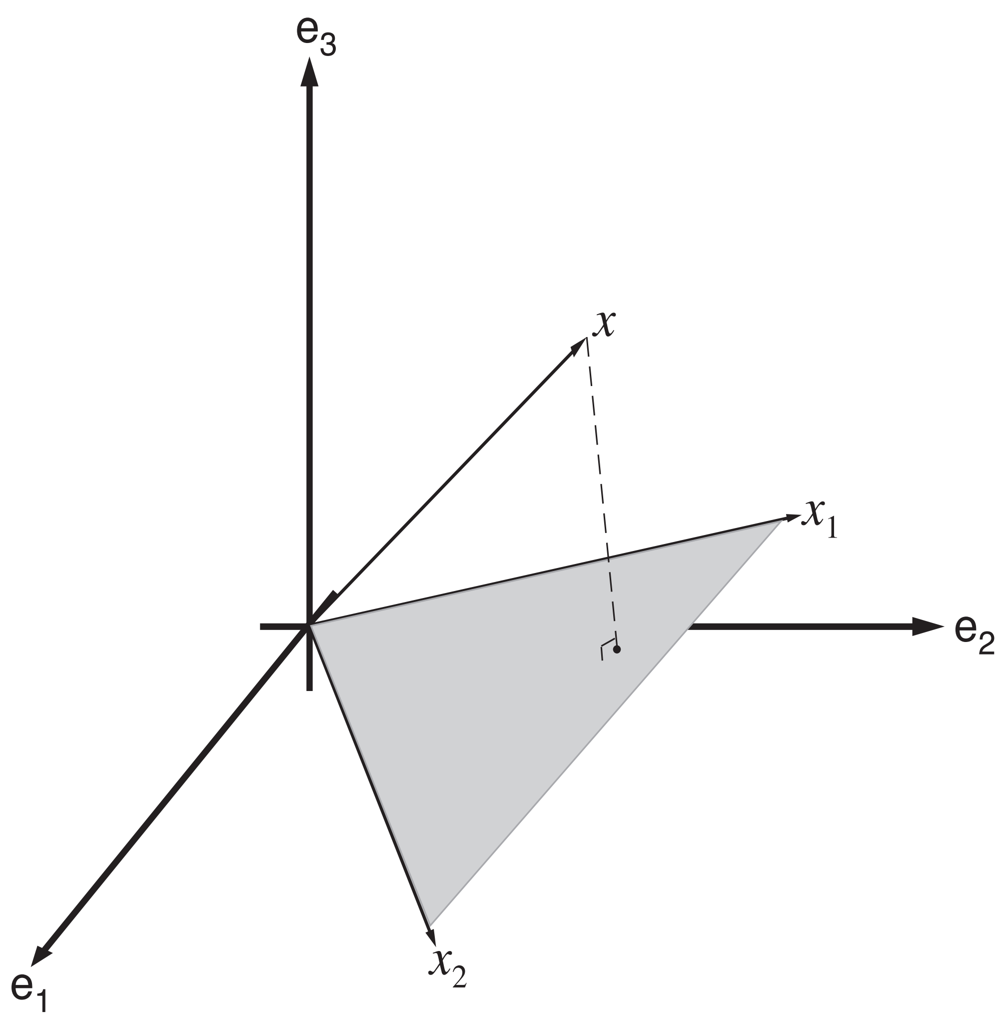
$$(x_1, x_2) = \Sigma_i x_1(i) x_2(i) \qquad \text{Euclidean inner product}$$

$$x_k = (x_k(1), x_k(2), \ldots, x_k(n))'$$

$$\frac{(u, v)}{\|u\|\|v\|} = \cos \text{angle between } u, v$$

$$E_n(\Sigma) \quad \Sigma \succ 0 \quad \langle x_1, x_2 \rangle_\Sigma \stackrel{\text{def}}{=} x_1' \Sigma^{-1} x_2$$

What is the projection of x onto plane spanned by x_1, x_2?



$\widehat{x} =$ orthogonal projection of x onto subspace spanned by x_1, x_2: $\boxed{\widehat{x} = a_1 x_1 + a_2 x_2}$
Find $a = (a_1, a_2)'$ to minimize $\|x - a_1 x_1 - a_2 x_2\|_\Sigma^2$

$$x'\Sigma^{-1}x - 2a_1 x'\Sigma^{-1}x_1 - 2a_2 x'\Sigma^{-1}x_2 + a_1^2 x_1'\Sigma^{-1}x_1 + 2a_1 a_2 x_1'\Sigma^{-1}x_2 + a_2^2 x_2'\Sigma^{-1}x_2$$

$$\frac{1}{2}\frac{d}{da_1} : \quad -x'\Sigma^{-1}x_1 + a_1 x_1'\Sigma^{-1}x_1 + a_2 x_1'\Sigma^{-1}x_2 = 0$$

$$\frac{1}{2}\frac{d}{da_2} : \quad -x'\Sigma^{-1}x_2 + a_1 x_1'\Sigma^{-1}x_2 + a_2 x_2'\Sigma^{-1}x_2 = 0$$

"normal equations" express perpendicularity:

$$\langle x, x_1 \rangle_\Sigma = \langle \widehat{x}, x_1 \rangle_\Sigma$$
$$\langle x, x_2 \rangle_\Sigma = \langle \widehat{x}, x_2 \rangle_\Sigma$$
$$\langle x - \widehat{x}, x_1 \rangle_\Sigma = 0$$
$$\langle x - \widehat{x}, x_2 \rangle_\Sigma = 0$$
$$x - \widehat{x} \perp x_1$$
$$x - \widehat{x} \perp x_2$$

Normal equations:

$$\begin{pmatrix} \langle x_1, x_1 \rangle_\Sigma & \langle x_1, x_2 \rangle_\Sigma \\ \langle x_2, x_1 \rangle_\Sigma & \langle x_2, x_2 \rangle_\Sigma \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \langle x, x_1 \rangle_\Sigma \\ \langle x, x_2 \rangle_\Sigma \end{pmatrix}$$

There is an **isometric isomorphism** between $E_n(\Sigma)$ and $\mathcal{X}$–the n-dimensional space spanned by the random variables

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim \mathcal{N}(0, \Sigma)$$

with the inner product $\langle X_i, X_j \rangle = E(X_i X_j)$

Isometric Isomorphism: `1 : 1` **inner product preserving map.**

$$
\begin{array}{rcll}
 & & E_n(\Sigma) & \mathcal{X} \\
x_1 & = & (\sigma_{11}, \cdots, \sigma_{1n}) & X_1 \\
x_2 & = & (\sigma_{21}, \cdots, \sigma_{2n}) & X_2 \\
\vdots & & & \vdots \\
x_n & = & \underbrace{(\sigma_{n1}, \cdots, \sigma_{nn})}_{rows\ of \Sigma} & X_n
\end{array}
$$

$$
< x_i, x_j >_\Sigma \;=\; \sigma_{ij} \;=\; EX_iX_j
$$

$$
\underbrace{(\sigma_{i1}, \cdots, \sigma_{in})\Sigma^{-1}\begin{pmatrix}\sigma_{j1}\\ \vdots \\ \sigma_{jn}\end{pmatrix}}_{=\sigma_{ij}} : \Sigma^{-1}\begin{pmatrix}\sigma_{j1}\\ \vdots \\ \sigma_{jn}\end{pmatrix} = \begin{pmatrix}0\\ \cdot \\ 1 \\ \cdot \\ 0\end{pmatrix}
$$

Let $x \in E_n(\Sigma)$	Let $X \in \mathcal{X}$
$\widehat{x} =$ projection of x onto $x_1, \ldots, x_k$	$\widehat{X} = E(X \mid X_1, \ldots, X_k)$
$\widehat{x} = a_1 x_1 + \ldots + a_k x_k$	$\widehat{X} = a_1 X_1 + \ldots + a_k X_k$

The normal equations are:

$$\begin{pmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_k \rangle \\ \vdots & \vdots \\ \langle x_1, x_k \rangle & \langle x_k, x_1 \rangle \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} = \begin{pmatrix} \langle x, x_1 \rangle \\ \vdots \\ \langle x, x_k \rangle \end{pmatrix}$$

$$\begin{pmatrix} E\,X_1 X_2 & E\,X_1 X_k \\ \vdots & \vdots \\ E\,X_k X_1 & E\,X_k X_k \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} = \begin{pmatrix} E\,X X_1 \\ \vdots \\ E\,X X_k \end{pmatrix}$$