RECAP

► We looked at n-length random vectors $\mathbf{X}$ which are essentially multivariate random variables.

► Its CDF/pdf is simply joint CDF/pdf of components

► $E[\mathbf{X}] = [E[X_1], \ldots E[X_n]]$. $C_{\mathbf{X}}$ is the covariance matrix.

> Let $\mathbf{Y} = G(\mathbf{X})$ where $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$, continuous invertible with continuous partial derivatives. Let H denote its inverse. Then $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(H(\mathbf{y}))|J|$ where J is the determinant of the Jacobian matrix.

►

► If $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, then $C_{\mathbf{Y}} = AC_{\mathbf{X}}A^T$ and

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|det(A)|} f_{\mathbf{X}}(A^{-1}(\mathbf{y} - \mathbf{b}))$$

Standard Normal Vector

▶ An n length random vector $\mathbf{Z}$ is called as a standard normal vector if its components Z_i are independent and standard normal.

▶ What is $E[\mathbf{Z}]$ and $C_{\mathbf{Z}}$?

▶ Show that the pdf is given by

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}} e^{\left\{-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right\}}$$

Standard Normal Vector

▶ Now suppose $\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}$. What is $E[\mathbf{X}]$ and $C_\mathbf{X}$?

▶ $E[\mathbf{X}] = \boldsymbol{\mu}$ and $C_\mathbf{X} = AA^T$.

▶ Note that A can have dimension $n \times l$ in which case $\mathbf{Z}$ is an l length standard normal.

▶ What is $f_\mathbf{X}(\mathbf{x})$?

Multivariate Gaussian Vector

▶ What is $f_{\mathbf{X}}(\mathbf{x})$?

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{|det(A)|} f_{\mathbf{Z}}(A^{-1}(\mathbf{x} - \boldsymbol{\mu})) \\
&= \frac{1}{(2\pi)^{n/2}\sqrt{det(C_{\mathbf{X}})}} e^{\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T C_X^{-1}(\mathbf{x}-\mu)\right\}}
\end{aligned}
$$

▶ Henceforth we will use the notation Σ to represent $C_{\mathbf{X}}$.

Definition 1: $\mathbf{X}$ is multivariate Gaussian with mean vector μ and covariance matrix Σ (denoted by $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$) if for some A and $\boldsymbol{\mu}$, it can be written as $\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}$ where $\mathbf{Z}$ is a standard normal vector and $\Sigma = AA^T$.

Equivalent definitions of a Gaussian vector

The following are equivalent definitions (without proof)

$\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$ if for some A and $\boldsymbol{\mu}$, it can be written as
$\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}$

$\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$ if it has the pdf

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{det(\Sigma)}} e^{\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}}$$

$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ iff for all vectors $\mathbf{a} \in \mathbb{R}^n$, it turns out that $\mathbf{a}^\mathsf{T}\mathbf{X}$ is univariate Gaussian $\mathcal{N}(a^T\boldsymbol{\mu}, a^T\Sigma a)$.

For equivalent definitions see `https://en.wikipedia.org/wiki/Multivariate_normal_distribution`

Affine transformations preserve Gaussianity

▶ Suppose $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, then what can we say about $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$? Is it a Gaussian vector ?

▶ Easy to see that $E[\mathbf{Y}] = AE[\mathbf{X}] + b$ and $C_{\mathbf{Y}} = A\Sigma A^T$.

▶ Like in the univariate case, we can use MGF (for multivariate MGF see Bertsekas) to show the following

> Suppose $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Now consider $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$. Then we have $\mathbf{Y} \sim \mathcal{N}(AE[\mathbf{X}] + b, A\Sigma A^T)$.

Marginalization and Conditioning (without proof)

▶ Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and partition $\mathbf{X}$ as $[\mathbf{X_1}, \mathbf{X_2}]^T$ where $\mathbf{X_1}$ is $m \times 1$ and $\mathbf{X_2}$ is $(n - m) \times 1$.

▶ We similarly have $\boldsymbol{\mu} = [\boldsymbol{\mu_1}, \boldsymbol{\mu_2}]^T$ and $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ where Σ_{11} is $m \times m$ matrix and so on ..

Marginalization property: The $m-$dimensional marginal distribution of $\mathbf{X_1}$ is $\mathcal{N}(\boldsymbol{\mu_1}, \Sigma_{11})$ and $\mathbf{X_2}$ is $\mathcal{N}(\boldsymbol{\mu_2}, \Sigma_{22})$

Conditioning property: The $m-$dimensional conditional distribution of $\mathbf{X_1}$ given $\mathbf{X_2} = \mathbf{x_2}$ is

$$\mathcal{N}(\boldsymbol{\mu_1} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x_2} - \boldsymbol{\mu_2}), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

Note the decrease in variance which does not depend on $\mathbf{x_2}$.

Towards Bivariate Gaussians

▶ Suppose $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Also suppose X_1 and X_2 are independent.

▶ For $\mathbf{x} = [x_2, x_2]^T$, we have

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= f_{X_1}(x_1) f_{X_2}(x_2) \\
&= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\left(\frac{(x_1-\mu_1)^2}{2\sigma_1^2} + \frac{(x_2-\mu_2)^2}{2\sigma_2^2} \right)} \\
&= \frac{1}{(2\pi)\sqrt{det(\Sigma)}} e^{\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\}}
\end{aligned}
$$

where $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$ and hence $\mathbf{X}$ is bivariate Gaussian.

▶ In general, a vector composed of independent Gaussians is a Gaussian vector. The converse is not true: a vector of dependent Gaussian components need not be Gaussian vector (EX 5.35 in probabilitycourse.com).
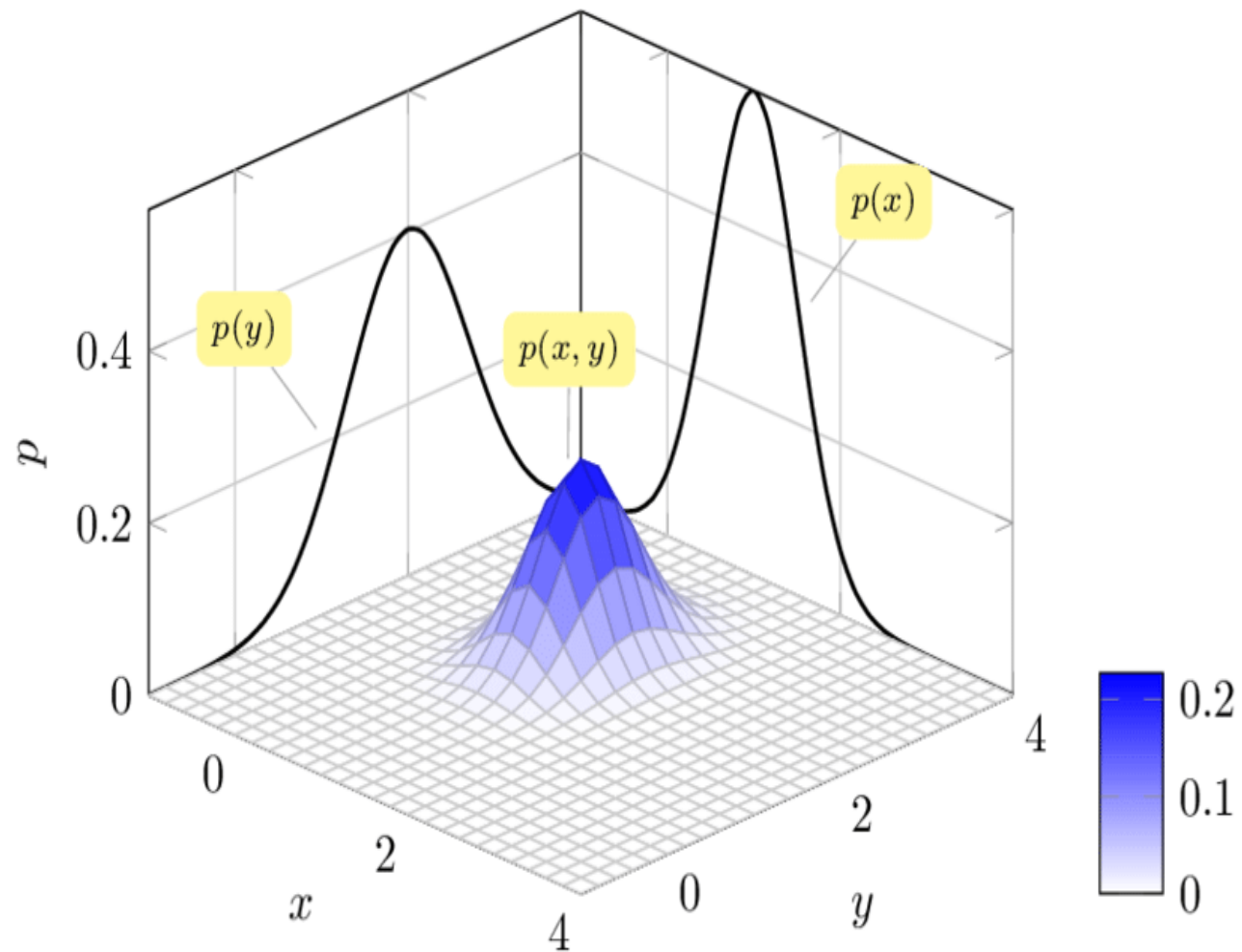
Bivariate Gaussians

▶ In general, X_1 and X_2 need not be independent in which case we have a general bivariate Gaussian
$\mathbf{X} = [X_1, X_2]^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\mu = [\mu_1, \mu_2]^T$ and
$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

▶ Show that Bivariate Gaussian is closed under marginalization,i.e., $X_1 \sim \mathcal{N}(\mu_1, \sigma_{11})$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_{22})$.

▶ Show that Bivariate Gaussian is closed under conditioning. (For proof see Theorem 5.4, probabilitycourse.com).

▶ This means that Given $X_2 = x_2$, one can show that $f_{X_1|X_2}(x_1|x_2)$ is Gaussian.

▶ These two properties make multivariate Gaussians as efficient modelling tools and the handy in Gaussian processes and Bayesian optimization.

Some Bivariate gaussian pdfs

Some Bivariate gaussian pdfs