Data-Driven Design & Analyses of Structures & Materials (3dasm)

Lecture 4

Miguel A. Bessa | **miguel_bessa@brown.edu** | Associate Professor

Outline for today

- Probability: multivariate models
    - The multivariate Gaussian: joint pdf, conditional pdf and marginal pdf
    - Covariance and covariance matrix

Reading material: This notebook (+ Bishop's book Section 2.3)

Summary of Bayes' rule

$$p(z|y=\mathcal{D}_y) = \frac{p(y=\mathcal{D}_y|z)p(z)}{p(y=\mathcal{D}_y)} = \frac{p(y=\mathcal{D}_y,z)}{p(y=\mathcal{D}_y)}$$

- $p(z)$ is the **prior** distribution
- $p(y=\mathcal{D}_y|z)$ is the **likelihood** function
- $p(y=\mathcal{D}_y,z)$ is the **joint likelihood** (product of likelihood function with prior distribution)
- $p(y=\mathcal{D}_y)$ is the **marginal likelihood**
- $p(z|y=\mathcal{D}_y)$ is the **posterior**

We can write Bayes' rule as posterior $\propto$ likelihood $\times$ prior , where we are ignoring the denominator $p(y=\mathcal{D}_y)$ because it is just a **constant** independent of the hidden variable z.

Diving deeper into the joint pdf

Later we will dedicate a lot of effort to using Bayes' rule to update a distribution over unknown values of some quantity of interest, given relevant observed data $\mathcal{D}_y$.
This is what is called *Bayesian inference* (a.k.a. *posterior inference*).

- But before we do that, we need to understand very well multivariate pdfs.
    - In particular, let's focus on the most important one: the **multivariate Gaussian**

Multivariate Gaussian pdf (a.k.a. MVN distribution)

The multivariate Gaussian pdf of a D-dimensional vector $\mathbf{x}$ is given by,

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(x-\mu)}}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}|}} \tag{1}$$

$$= \frac{1}{(2\pi)^{D/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \tag{2}$$

where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}] \in \mathbb{R}^D$ is the mean vector, and $\boldsymbol{\Sigma} = \mathrm{Cov}[\mathbf{x}]$ is the $D \times D$ **covariance matrix**.

Covariance matrix

The covariance matrix is a natural generalization of the variance (Lecture 1) for the multivariate case!

$$\boldsymbol{\Sigma} = \mathrm{Cov}[\mathbf{x}] = \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T\right] \tag{3}$$

$$= \begin{bmatrix} \mathbb{V}[x_1] & \mathrm{Cov}[x_1, x_2] & \cdots & \mathrm{Cov}[x_1, x_D] \\ \mathrm{Cov}[x_2, x_1] & \mathbb{V}[x_2] & \cdots & \mathrm{Cov}[x_2, x_D] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[x_D, x_1] & \mathrm{Cov}[x_D, x_2] & \cdots & \mathbb{V}[x_D] \end{bmatrix} \tag{4}$$

where $\mathrm{Cov}[x_i, x_j] = \mathbb{E}\left[(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])\right] = \mathbb{E}[x_i x_j] - \mathbb{E}[x_i]\mathbb{E}[x_j]$
Also note that $\mathbb{V}[x_i] = \mathrm{Cov}[x_i, x_i]$.

Notes about covariance and normalized covariance (correlation coefficient)

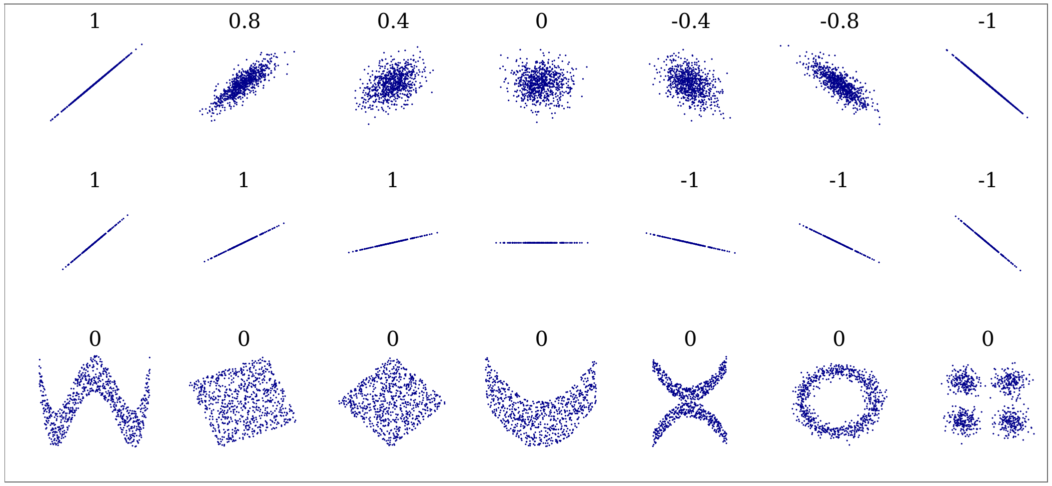
The covariance between two rv's y and z measures the degree to which y and z are **linearly** related.
Covariances can be between negative and positive infinity.
Sometimes it is more convenient to work with a normalized measure, with a finite lower and upper bound. The (Pearson) **correlation coefficient** between y and z is defined as

$$\rho = \mathrm{corr}[y, z] = \frac{\mathrm{Cov}[y, z]}{\sqrt{\mathbb{V}[y]\mathbb{V}[z]}}$$

Covariance and correlation coefficient measure the same relationship.

NOTE ABOUT NORMALIZED COVARIANCE (CORRELATION COEFFICIENT)

Several sets of (y_i, z_i) points, with the correlation coefficient of y and z for each set.



Top row: $\mathrm{corr}[y, z]$ reflects the noisiness and direction of a linear relationship.
Middle row: $\mathrm{corr}[y, z]$ **does not** reflect the slope of that relationship
Bottom row: $\mathrm{corr}[y, z]$ **does not** reflect many aspects of nonlinear relationships.
(Additional note: the figure in the center has a slope of 0 but in that case the correlation coefficient is undefined because the variance of z is zero.)

Understanding the MVN pdf (a common joint pdf)

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu)}}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}} \tag{5}$$

$$= \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \tag{6}$$

where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}] \in \mathbb{R}^D$ is the mean vector, and $\boldsymbol{\Sigma} = \mathrm{Cov}[\mathbf{x}]$ is the $D \times D$ **covariance matrix**.

- Multivariate Gaussian pdf's are very important in ML and Statistics.
- Let's discover their properties by working out some examples.

Homework 2 (Exercise 5): MVN from independent Gaussian rv's

Consider two **independent** rv's x_1 and x_2 where each of them is a univariate Gaussian pdf:

$$x_1 = \mathcal{N}(x_1|\mu_{x_1}, \sigma_{x_1}^2)$$

$$x_2 = \mathcal{N}(x_2|\mu_{x_2}, \sigma_{x_2}^2)$$

where $\mu_{x_1} = 10$, $\sigma_{x_1}^2 = 5^2$, $\mu_{x_2} = 0.5$ and $\sigma_{x_2}^2 = 2^2$.

Answer the following questions:

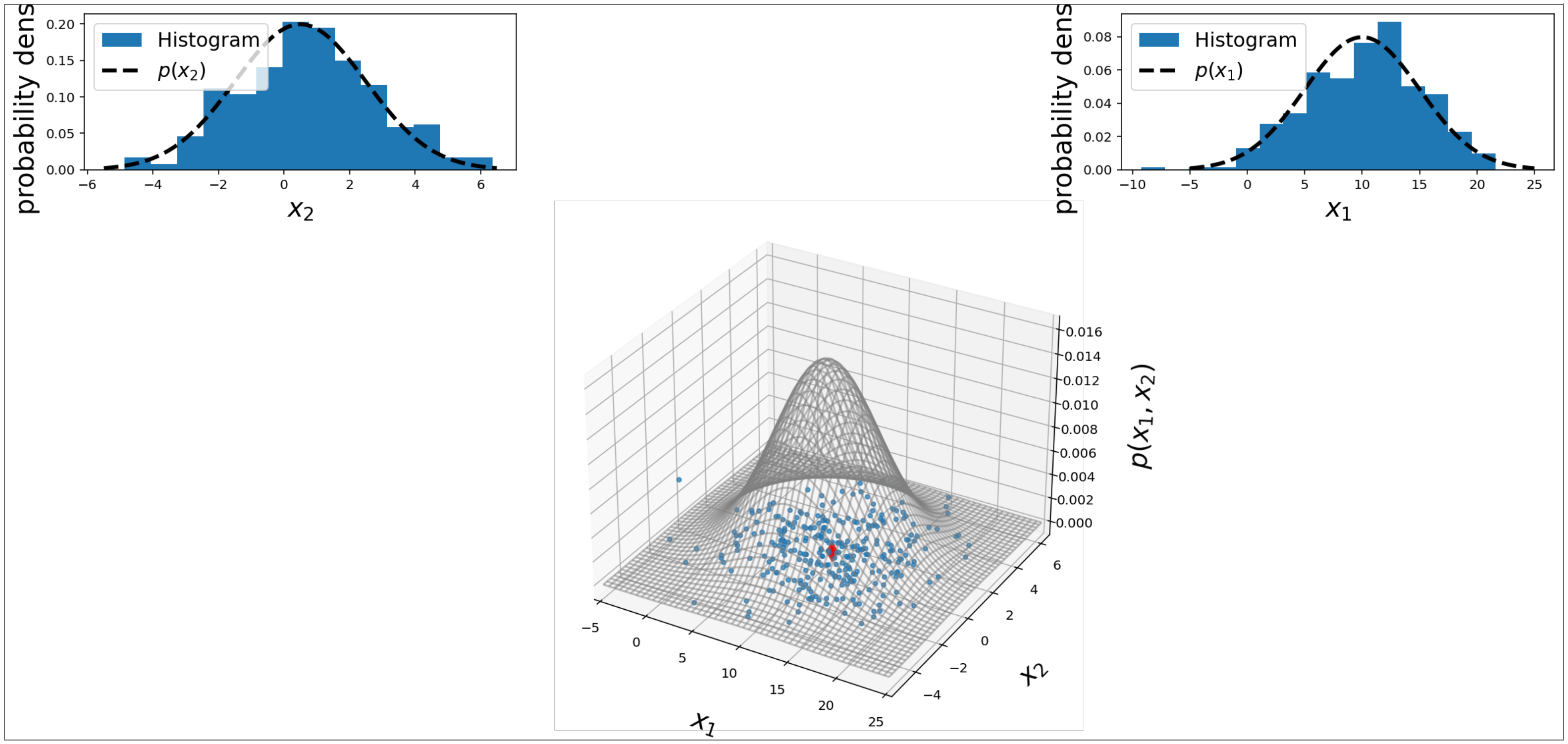
1. What is the joint pdf $p(x_1, x_2)$?
2. Calculate the covariance matrix for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

The next slide plots the solution of the joint pdf... (But do your homework!)

In [3]:

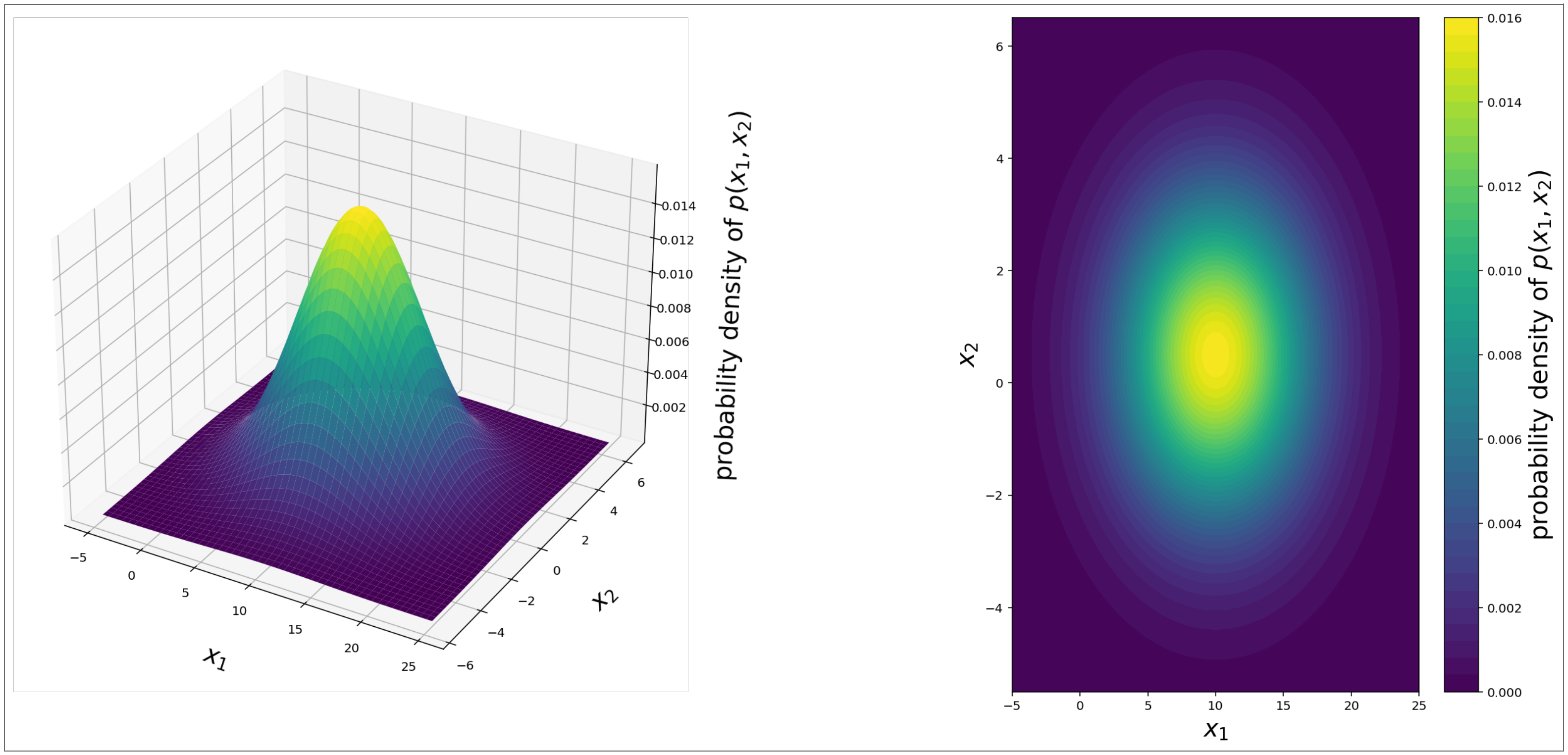
```
# HIGHLIGHT DIFFERENCE IN MAXIMUM PROBABILITY DENSITIES!!
fig_joint_pdf_HW2_ex5 # The joint pdf results from the multiplication...
```

Out[3]:



In [5]:
```
# Same pdf but now as a surface plot and as a contour plot.
fig_joint_pdf_HW2_ex5_color
```

Out[5]:



Car stopping distance problem (I know how much you missed it!)

Back to our simple car stopping distance problem with constant velocity $x = 75$ m/s. We have two rv's for this problem,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y \\ z \end{bmatrix}$$

- Note: this $\mathbf{x}$ has NOTHING to do with our velocity variable x. Be careful!

$$\boldsymbol{\Sigma} = \mathrm{Cov}[\mathbf{x}] = \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T\right] \tag{7}$$

$$= \begin{bmatrix} \mathbb{V}[y] & \mathrm{Cov}[y, z] \\ \mathrm{Cov}[z, y] & \mathbb{V}[z] \end{bmatrix} \tag{8}$$

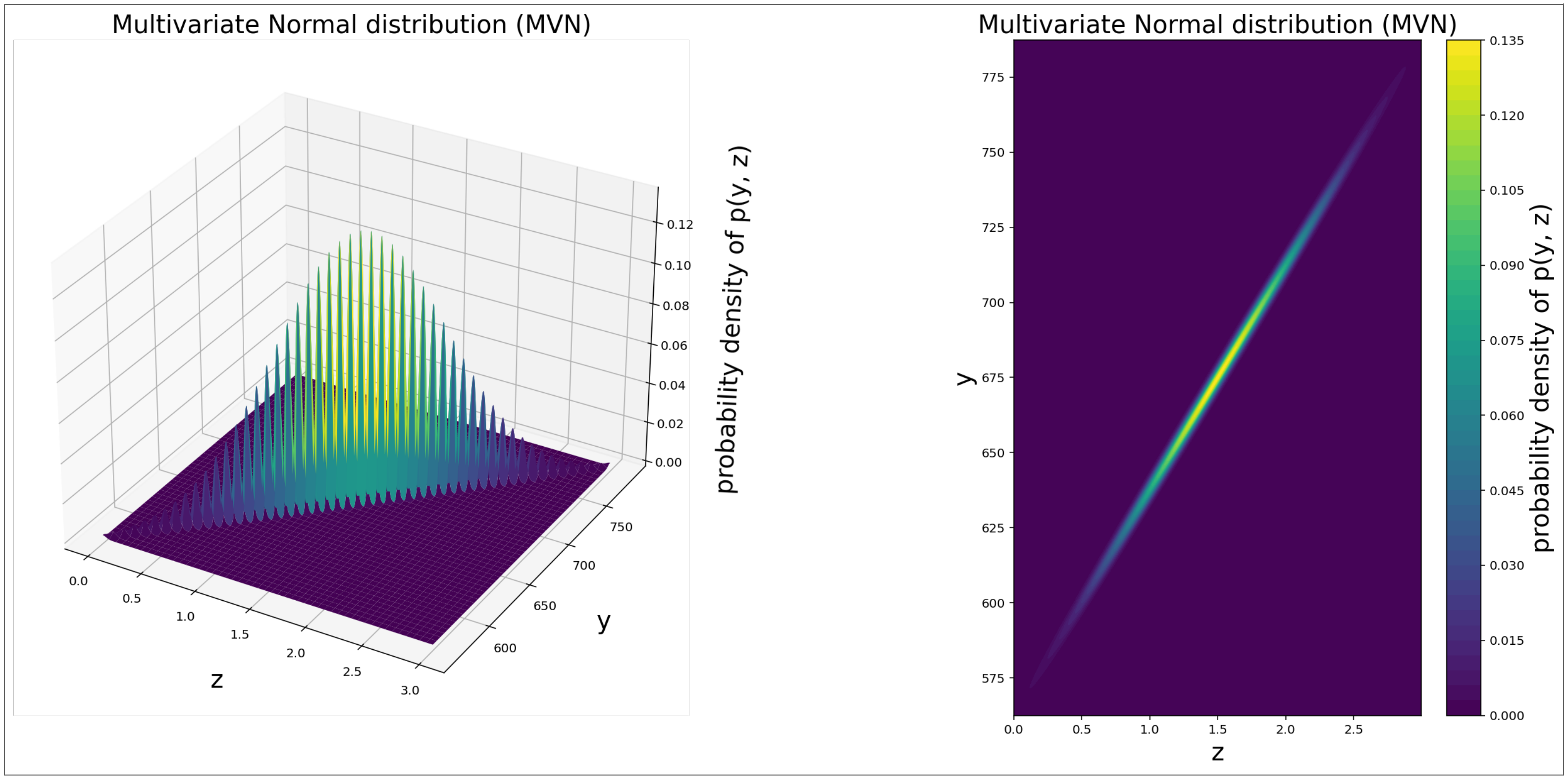
where $\mathrm{Cov}[y, z] = \mathbb{E}\left[(y - \mathbb{E}[y])(z - \mathbb{E}[z])\right] = \mathbb{E}[yz] - \mathbb{E}[y]\mathbb{E}[z]$

Homework 2 (Exercise 6): Covariance matrix for the car problem when $x = 75$ m/s

1. Calculate the mean vector and covariance matrix values for our car stopping distance problem (with $x = 75$ m/s). **Be careful** that y is dependent on z.
2. Calculate the determinant of the covariance matrix.

The next slide plots the multivariate Gaussian $p(y, z)$ obtained from the mean vector and covariance matrix you calculated.

In [7]:
```
# Code to generate this figure is hidden in presentation (shown in notes)
regularizer = 1e-3 # Thikhonov regularization to approximate p(y,z) for car stopping distance probler
plot_car_MVN_regularized(regularizer) # SHOW WHAT HAPPENS IF regularizer is 0, 0.1 and 1e-3
```



Recal the joint pdf $p(y, z)$ we found for this problem in Lecture 3!

We determined in Lecture 3 that the joint pdf $p(y, z)$ for this problem is

$$p(y, z) = \delta\left(y - (75z + 562.5)\right) p(z)$$

where $p(z) = \mathcal{N}(\mu_z = 1.5, \sigma_z^2 = 0.5^2)$, and $p(y|z) = \delta\left(y - (75z + 562.5)\right)$ is the Dirac delta pdf that assigns zero probability everywhere except when $y = 75z + 562.5$.

- Recall that $y = 75z + 562.5$ was obtained from $y = zx + 0.1x^2$ when fixing $x = 75$ m/s.

- Now we see how to approximate this pdf for plotting it:
    - We can consider that the joint pdf $p(y, z)$ is an MVN, and include a small term in the diagonal of the Covariance matrix to plot it! As this term tends to zero, we retrieve the Dirac delta effect.

What if the conditional pdf was different?

Let's consider a slightly different case where, instead of knowing that $p(y|z) = \delta\left(y - (75z + 562.5)\right)$, we consider the conditional pdf as:

$$p(y|z) = \mathcal{N}\left(y|\mu_{y|z} = 75z + 562.5, \sigma_{y|z}^2 = 30^2\right)$$

and that we still consider $p(z) = \mathcal{N}\left(z|\mu_z = 1.5, \sigma_z^2 = 0.5^2\right)$.

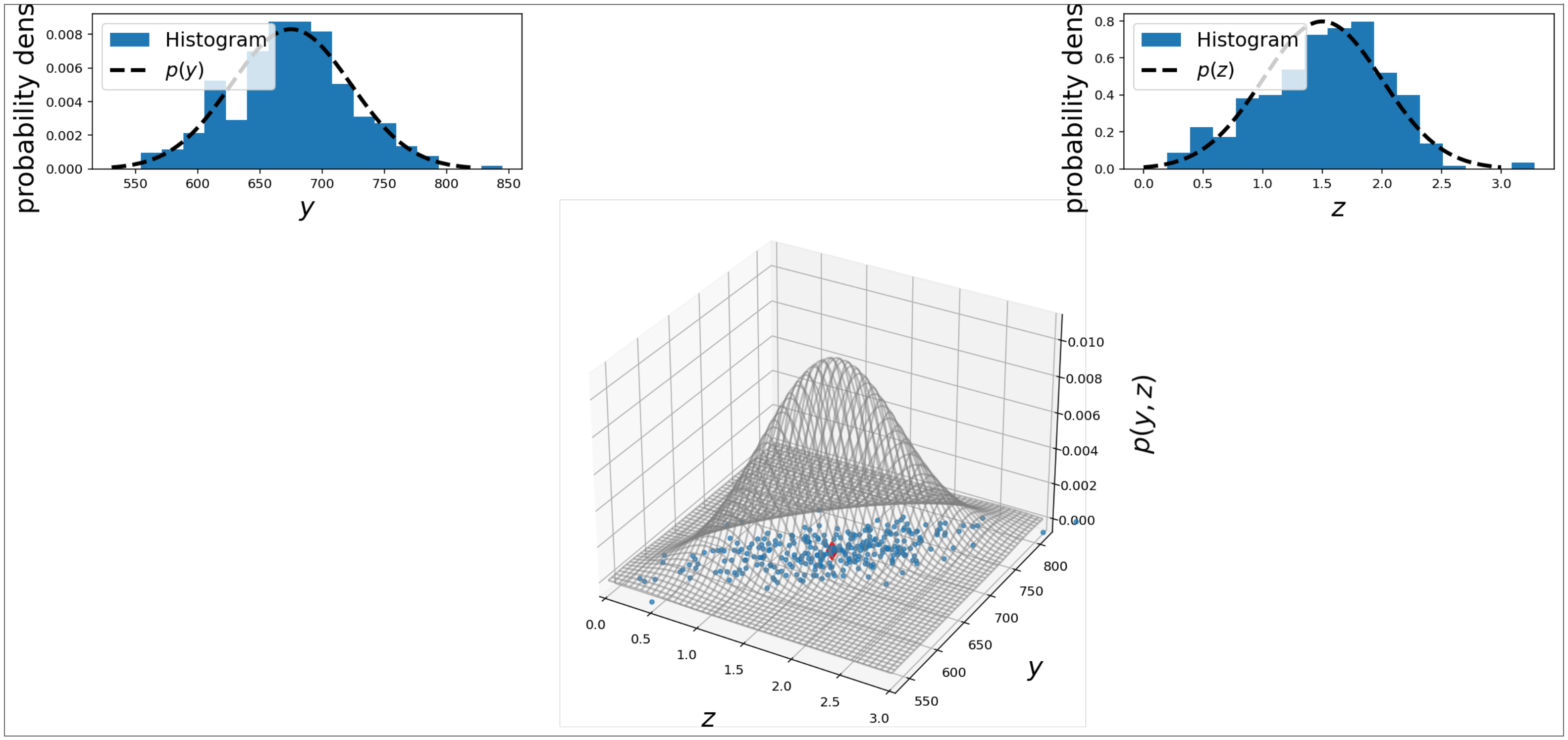
We will see in the next lectures what this choice means in the context of the car stopping distance problem...

Given the above considerations, the joint pdf is now:

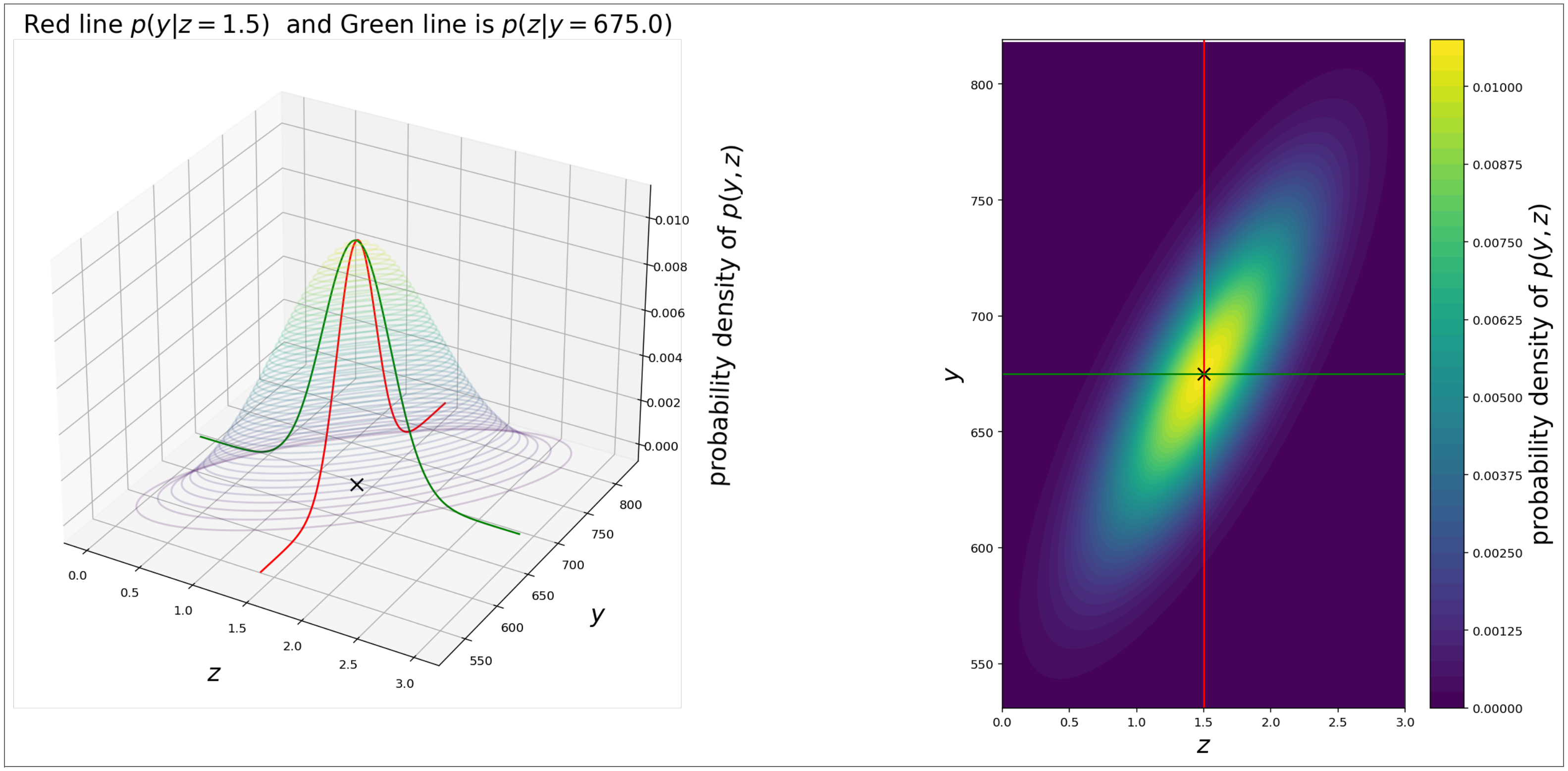
$$p(y,z) = \mathcal{N}\left(y|\mu_{y|z} = 75z + 562.5, \sigma_{y|z}^2 = 30^2\right)\mathcal{N}\left(z|\mu_z = 1.5, \sigma_z^2 = 0.5^2\right)$$

In [9]:
```
# HIGHLIGHT DIFFERENCE IN MAXIMUM PROBABILITY DENSITIES!!
fig_joint_pdf_new # The joint pdf results from the multiplication...
```

Out[9]:



In [11]:
```python
# Static plot (I skip this cell in presentations, but use it when printing slides to PDF)
fig2_joint_pdf_new(y_value=mu_y,z_value=mu_z)
```



Conclusions about Gaussian distributions

Our empirical investigations in this Lecture, have led to some interesting observations! They can be generalized to:

- If two sets of variables are jointly Gaussian, i.e. if their joint pdf is an MVN, then:
    - their conditional pdfs are Gaussian, i.e. the conditional distribution of one set conditioned on the other is again Gaussian!
    - the marginal distribution of either set is also Gaussian!

This is really important because it means that Gaussians are closed under Bayesian conditioning! We will explore this later.

- Note: Bishop's book has a fantastic discussion about the univariate and multivariate Gaussian distribution (Section 2.3). **I recommend reading it.** I also included it in the notes below this cell.

Summary of partitioned Gaussians

Given a joint Gaussian pdf $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Lambda} \equiv \boldsymbol{\Sigma}^{-1}$ and

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{bmatrix}$$

We have the conditional distribution $p(\mathbf{x}_a, \mathbf{x}_b) = \mathcal{N}(\mathbf{x}_a|\boldsymbol{\mu}_{a|b}, \boldsymbol{\Lambda}_{aa}^{-1})$ with the following parameters:

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b)$$

$$\boldsymbol{\Sigma}_{a|b} = \boldsymbol{\Lambda}_{aa}^{-1}$$

where $\boldsymbol{\Lambda}_{aa} = (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}$, and $\boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab} = \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}$.
The marginal distribution is $p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa})$.

See you next class

Have fun!