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

Lecture 5

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

OPTION 1. Run this notebook **locally in your computer**:

1. Confirm that you have the '3dasm' mamba (or conda) environment (see Lecture 1).
2. Go to the 3dasm_course folder in your computer and pull the last updates of the
   repository:

   ```
   git pull
   ```

   - Note: if you can't pull the repo due to conflicts (and you can't handle these
     conflicts), use this command (with **caution**!) and your repo becomes the same
     as the one online:

     ```
     git reset --hard origin/main
     ```

3. Open command window and load jupyter notebook (it will open in your internet
   browser):

   ```
   jupyter notebook
   ```

4. Open notebook of this Lecture and choose the '3dasm' kernel.

Outline for today

- Bayesian inference for one hidden rv
    - Prior
    - Likelihood
    - Marginal likelihood
    - Posterior
    - Gaussian pdf's product

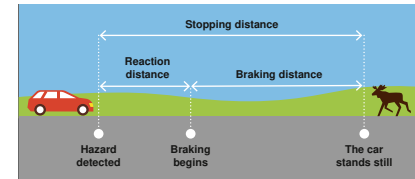**Reading material**: This notebook + Chapter 3

Car distance problem at constant velocity

Our favorite problem at constant velocity $x = 75$ m/s is described by:

$$y = z \cdot 75 + 0.1 \cdot 75^2 = 75z + 562.5$$

- y is the **output**: the car stopping distance (in meters)

- z is a hidden variable: an rv representing the driver's reaction time (in seconds)

where $z \sim \mathcal{N}(\mu_z = 1.5, \sigma_z^2 = 0.5^2)$

```
In [9]: # vvvvvvvvvvv this is just a trick so that we can run this cell multiple times vvvvvvvvvvv
        fig_car_new, ax_car_new = plt.subplots(1,2); plt.close() # create figure and close it
if fig_car_new.get_axes():
    del ax_car_new; del fig_car_new # delete figure and axes if they exist
    fig_car_new, ax_car_new = plt.subplots(1,2) # create them again
# ^^^^^^^^^^^ end of the trick ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
N_samples = 3 # CHANGE THIS NUMBER AND RE-RUN THE CELL
x = 75; empirical_y = samples_y(N_samples, x); # Empirical measurements of N_samples at x=75
empirical_mu_y = np.mean(empirical_y); empirical_sigma_y = np.std(empirical_y); # empirical mean and st
car_fig(ax_car_new[0]) # a function I created to include the background plot of the governing model
for i in range(2): # create two plots (one is zooming in on the error bar)
    ax_car_new[i].errorbar(x , empirical_mu_y,yerr=1.96*empirical_sigma_y, fmt='m*', markersize=15);
    ax_car_new[i].scatter(x*np.ones_like(empirical_y),empirical_y, s=40,
                          facecolors='none', edgecolors='k', linewidths=2.0)
print("Empirical mean[y] is",empirical_mu_y, "(real mean[y]=675)")
print("Empirical std[y] is",empirical_sigma_y,"(real std[y]=37.5)")
fig_car_new.set_size_inches(25, 5) # scale figure to be wider (since there are 2 subplots)
```

```
Empirical mean[y] is 672.5378635093475 (real mean[y]=675)
Empirical std[y] is 35.419420813521 (real std[y]=37.5)
```
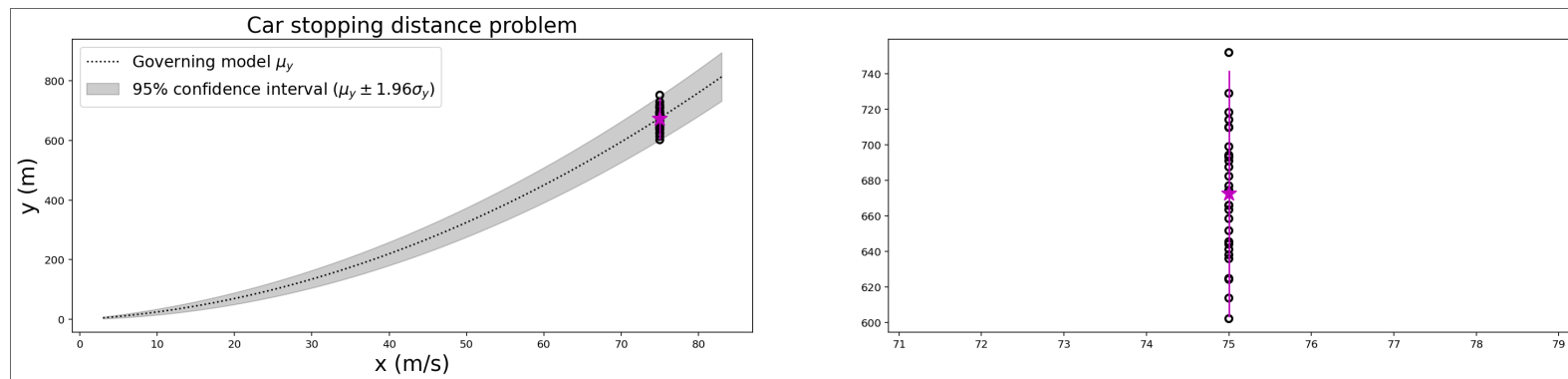
Car stopping distance problem without completely knowing the underlying model

Let's focus (again) on our favorite problem for a constant velocity $x = 75$ m/s, **but this time consider the following (more realistic) scenario**:
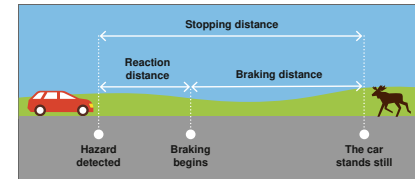
- Assume that you are **not 100% sure** that the governing model for the stopping distance is: $y = 75z + 562.5$. In other words, you **suspect** that the model looks like that but you are not sure that this is the case (it just so happens that you are right!).

- Also, assume that you don't have the slightest idea about the pdf of the reaction time z, i.e. you **do not know** that $z \sim \mathcal{N}(\mu_z = 1.5, \sigma_z^2 = 0.5^2)$.

How can we construct a reasonable joint pdf $p(y, z)$?

Car stopping distance problem without completely knowing the underlying model

Let's construct a reasonable joint pdf $p(y, z) = p(y|z)p(z)$ by making assumptions about $p(z)$ and $p(y|z)$:

1. Since we have no idea about the pdf of z, the simplest assumption we can make is that the prior is the **Uniform distribution** (a.k.a. "noninformative" prior).

1. Since we think that $y = 75z + 562.5$ but we are not sure, we can assume the conditional pdf to be a Gaussian distribution with **mean** $\mu_{y|z} = 75z + 562.5$ and with a standard deviation of our choice $\sigma_{y|z} = 80$ m.
    - Note: (for now) this value for the standard deviation can be viewed as a safety margin that we assign to our prediction (it could be any number! I chose 80 because it's close to the actual standard deviation for y when using the actual distribution for z at $x = 75$ m/s and that you calculated in Homework 2).

Prior: our beliefs about the problem

As we said, when we have no clue about what the distribution of the hidden rv z is, then we can use a **Uniform distribution** (a.k.a. uninformative prior).
This distribution assigns equal probability to any value of z within an interval $z \in (z_{min}, z_{max})$.

$$p(z) = \frac{1}{C_z}$$

where $C_z = z_{max} - z_{min}$ is the **normalization constant** of the Uniform pdf, i.e. the value that guarantees that $p(z)$ integrates to one.

For the time being, we will not assume any particular values for z_{max} and z_{min}. So, we will consider the case when $z_{max} \to +\infty$ and $z_{min} \to -\infty$. Mathematically this leads to some indefinition, but it's commonly ignored. If we had some information, we could consider some values for these bounds (e.g. $z_{min} = 0$ seconds would be the limit of the fastest reaction time that is humanly possible, and $z_{max} = 3$ seconds would be the slowest reaction time of a human being).

Conditional pdf: including a little safety...

Now that we settled on the prior, let's focus on the conditional pdf $p(y|z)$ that we decided to assume to be a Gaussian distribution like this:

$$p(y|z) = \mathcal{N}\left(y|\mu_{y|z} = wz + b, \sigma^2_{y|z} = s^2\right)$$

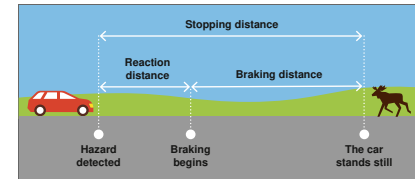where w, b and s are all constants that we defined as:

$w = x = 75$

$b = 0.1x^2 = 562.5$

$s^2 = 80^2$

- Note: we could have assumed any other values, but of course that would have consequences on our predictions... We will discuss this extensively in the next lectures.

$$p(y|z) = \mathcal{N}\left(y|\mu_{y|z} = wz + b, \sigma^2_{y|z} = s^2\right)$$

- Compare the newly assumed Gaussian conditional pdf $p(y|z)$ with the the conditional pdf we defined previously when we were certain about the stopping distance, $y = wz + b$. As you may recall, when we were certain about it then we used:

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

- Using the Dirac delta "distribution" is equivalent to considering the Gaussian distribution when the standard deviation goes to zero: $s \to 0$.

  - It is also equivalent to the Uniform distribution when the interval shrinks to a point.

Summary of our Model

1. The **observation distribution** (our conditional pdf):

$$p(y|z) = \mathcal{N}\left(y|\mu_{y|z} = wz + b, \sigma_{y|z}^2\right) \tag{1}$$

$$= \frac{1}{C_{y|z}} \exp\left[-\frac{1}{2\sigma_{y|z}^2}(y - \mu_{y|z})^2\right] \tag{2}$$

where $C_{y|z} = \sqrt{2\pi\sigma_{y|z}^2}$ is the **normalization constant** of the Gaussian pdf, and where $\mu_{y|z} = wz + b$, with w, b and $\sigma_{y|z}^2$ being constants, as previously mentioned.

1. and the **prior distribution**: $p(z) = \frac{1}{C_z}$

where $C_z = z_{max} - z_{min}$ is the **normalization constant** of the Uniform pdf, i.e. the value that guarantees that $p(z)$ integrates to one.

So, how can we make our prediction for the car stopping distance problem at constant velocity?

Today we will start making our first prediction!
Recall that we want to predict the stopping distance y at constant velocity $x = 75$ m/s without knowing $p(z)$ and without being certain about the stopping distance model.

- So, **how** can we predict y without this information?

Yes!! If we use **Bayes' rule** (and another step... in the next lecture)!
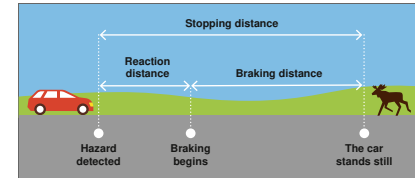
Understanding the Bayes' rule

$$p(z|y) = \frac{p(y|z)p(z)}{p(y)}$$

- $p(z)$ is the **prior distribution**
- $p(y|z)$ is the **observation distribution** (conditional pdf)
- $p(y)$ is the **marginal distribution**
- $p(z|y)$ is the **posterior distribution**

Posterior from Bayes' rule

Since we have defined the **observation distribution** and the **prior distribution**, we can now compute the posterior distribution from Bayes' rule.

But this requires a bit of algebra... Let's do it!

First, in order to apply Bayes' rule $p(z|y) = \frac{p(y|z)p(z)}{p(y)}$ we need to calculate $p(y)$.

$p(y)$ is obtained by marginalizing the joint distribution wrt z:
$p(y) = \int p(y|z)p(z)dz$
which implies an integration over z. So, let's rewrite $p(y|z)$ so that the integration becomes easier.

$$p(y|z) = \mathcal{N}\left(y|\mu_{y|z} = wz + b, \sigma^2_{y|z}\right) \tag{3}$$

$$= \frac{1}{C_{y|z}} \exp\left[-\frac{1}{2\sigma^2_{y|z}}(y - (wz + b))^2\right] \tag{4}$$

$$= \frac{1}{C_{y|z}} \exp\left\{-\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2}\left[z - \left(\frac{y-b}{w}\right)\right]^2\right\} \tag{5}$$

$$= \frac{1}{|w|} \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{-\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2}\left[z - \left(\frac{y-b}{w}\right)\right]^2\right\} \tag{6}$$

Note: This Gaussian pdf $\mathcal{N}\left(z|\frac{y-b}{w}, \left(\frac{\sigma_{y|z}}{w}\right)^2\right)$ is unnormalized when written wrt z (due to $\frac{1}{|w|}$).

We can now calculate the marginal distribution $p(y)$:

$$p(y) = \int p(y|z)p(z)dz \qquad (7)$$

$$= \int \frac{1}{|w|} \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{-\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2}\left[z - \left(\frac{y-b}{w}\right)\right]^2\right\} \frac{1}{C_z}dz \qquad (8)$$

We can rewrite this expression as,

$$p(y) = \frac{1}{|w|\cdot C_z} \int \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{-\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2}\left[z - \left(\frac{y-b}{w}\right)\right]^2\right\}dz \qquad (9)$$

What is the result for the blue term?

From where we conclude that the marginal distribution is:

$$p(y) = \frac{1}{|w|C_z}$$

So, now we can determine the posterior:

$$p(z|y) = \frac{p(y|z)p(z)}{p(y)} \tag{10}$$

$$= |w|C_z \cdot \frac{1}{|w|} \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{-\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2}\left[z - \left(\frac{y-b}{w}\right)\right]^2\right\} \cdot \frac{1}{C_z} \tag{11}$$

$$= \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{-\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2}\left[z - \left(\frac{y-b}{w}\right)\right]^2\right\} \tag{12}$$

which is a **normalized** Gaussian pdf in z: $\mathcal{N}\left(z|\frac{y-b}{w}, \left(\frac{\sigma_{y|z}}{w}\right)^2\right)$

- **This is what the Bayes' rule does!** Computes the posterior $p(z|y)$ from $p(y|z)$ and $p(z)$.

Why should we care about the Bayes' rule?

There are a few reasons:
1. As we will see, models are usually (always?) wrong.
1. But our beliefs may help to get closer to reality! Bayes' rule enables us to get better models if our beliefs are reasonable and we don't have lots of data!
1. We don't observe distributions. We observe **DATA**. Bayes' rule is a very powerful way to predict the distribution of our quantity of interest (here: y) from data!

Bayes' rule applied to observed data

Previously, we already introduced Bayes' rule when applied to observed data $\mathcal{D}_y$.

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

- $p(z)$ 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.

Bayes' rule applied to observed data

But remember that Bayes' rule is just a way to calculate the posterior:

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

Usually, what we really want is to be able to predict the distribution of the quantity of interest (here: y) after observing some data $\mathcal{D}_y$:

$$p(y|y = \mathcal{D}_y) = \int p(y|z)p(z|y = \mathcal{D}_y)dz$$

which is often written in simpler notation: $p(y|\mathcal{D}_y) = \int p(y|z)p(z|\mathcal{D}_y)dz$

Bayesian inference for car stopping distance problem

Now we will solve the first Bayesian ML problem from some given data $y = \mathcal{D}_y$:

y_i (m)
601.5
705.9
693.8
...
711.3

where the data $\mathcal{D}_y$ could be a Pandas dataframe with N data points (N rows).

- **Very Important Question (VIQ)**: Can we calculate the **likelihood** function from this data?

Likelihood for car stopping distance problem

Of course! As we saw a few cells ago, the **likelihood** is obtained by evaluating the **observation distribution** at the data $\mathcal{D}_y$.
Noting that each observation in $\mathcal{D}_y$ is independent of each other and that they have the same distribution (**indepentent and identically distributed**: i.i.d.), then:

$$p(y = \mathcal{D}_y | z) = \prod_{i=1}^{N} p(y = y_i | z) = p(y = y_1 | z)p(y = y_2 | z) \cdots p(y = y_N | z)$$

which gives the **probability density** of observing that data if using our observation distribution (part of our model!).

CALCULATING THE LIKELIHOOD

Let's calculate it:

$$p(y = \mathcal{D}_y | z) = \prod_{i=1}^{N} p(y = y_i | z) \tag{13}$$

$$= \prod_{i=1}^{N} \frac{1}{|w|} \frac{1}{\sqrt{2\pi \left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y_i - b}{w}\right) \right]^2 \right\} \tag{14}$$

$$= \prod_{i=1}^{N} \frac{1}{|w|} \mathcal{N}\left(z | \mu_i, \sigma_i^2\right) \tag{15}$$

where $\mu_i = \frac{y_i - b}{w}$ and $\sigma_i = \frac{\sigma_{y|z}}{w}$

This seems a bit daunting... I know. Do not dispair yet!

Product of Gaussian pdf's of the same rv z

It can be shown that the product of N univariate Gaussian pdf's of the same rv z is:

$$\prod_{i=1}^{N} \mathcal{N}(z|\mu_i, \sigma_i^2) = C \cdot \mathcal{N}(z|\mu, \sigma^2)$$

with mean: $\mu = \sigma^2 \left(\sum_{i=1}^{N} \frac{\mu_i}{\sigma_i^2} \right)$

variance: $\sigma^2 = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$

and normalization constant: $C = \frac{1}{(2\pi)^{(N-1)/2}} \sqrt{\frac{\sigma^2}{\prod_{i=1}^{N} \sigma_i^2}} \exp\left[-\frac{1}{2} \left(\sum_{i=1}^{N} \frac{\mu_i^2}{\sigma_i^2} - \frac{\mu^2}{\sigma^2} \right) \right]$

Curiosity: the normalization constant C is itself a Gaussian! You can see it more clearly if you consider $N = 2$

In Homework 3...

Show that the product of two Gaussian pdf's for the same rv z is:

$$\mathcal{N}(z|\mu_1, \sigma_1^2) \cdot \mathcal{N}(z|\mu_2, \sigma_2^2) = C \cdot \mathcal{N}(z|\mu, \sigma^2)$$

$$\sigma^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \tag{16}$$

$$\mu = \sigma^2 \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2} \right) \tag{17}$$

$$C = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp \left[-\frac{1}{2(\sigma_1^2 + \sigma_2^2)} (\mu_1 - \mu_2)^2 \right] \tag{18}$$

Back to calculating the likelihood

$$p(y = \mathcal{D}_y|z) = \prod_{i=1}^{N} p(y = y_i|z) \tag{21}$$

$$= \prod_{i=1}^{N} \frac{1}{|w|} \frac{1}{\sqrt{2\pi \left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y_i - b}{w}\right) \right]^2 \right\} \tag{22}$$

$$= \frac{1}{|w|^N} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi \left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y_i - b}{w}\right) \right]^2 \right\} \tag{23}$$

So, using the result of a product of N Gaussian pdf's to calculate the likelihood, and noting that $\sigma_i = \frac{\sigma_{y|z}}{w}$ and $\mu_i = \frac{y_i - b}{w}$ we get:

$$p(y = \mathcal{D}_y | z) = \frac{1}{|w|^N} \cdot C \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(z - \mu)^2\right]$$

where

$$\mu = \frac{\sigma^2}{\sigma_i^2} \sum_{i=1}^{N} \mu_i = \frac{w^2\sigma^2}{\sigma_{y|z}^2} \sum_{i=1}^{N} \mu_i$$

$$\sigma^2 = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} = \frac{1}{\sum_{i=1}^{N} \frac{w^2}{\sigma_{y|z}^2}} = \frac{\sigma_{y|z}^2}{w^2 N}$$

$$C = \frac{1}{(2\pi)^{(N-1)/2}} \sqrt{\frac{\sigma^2}{\left(\frac{\sigma_{y|z}^2}{w^2}\right)^N}} \exp\left[-\frac{1}{2}\left(\frac{w^2}{\sigma_{y|z}^2} \sum_{i=1}^{N} \mu_i - \frac{\mu^2}{\sigma^2}\right)\right] = \frac{1}{(2\pi)^{(N-1)/2}} \sqrt{\frac{\sigma^2}{\left(\frac{\sigma_{y|z}^2}{w^2}\right)^N}}$$

Calculating the Marginal Likelihood

$$p(y = \mathcal{D}_y) = \int p(y = \mathcal{D}_y | z)p(z)dz \tag{24}$$

$$= \int \frac{1}{|w|^N} C \cdot \mathcal{N}(z|\mu, \sigma^2) \cdot \frac{1}{C_z} dz \tag{25}$$

$$= \frac{C}{|w|^N C_z} \int \mathcal{N}(z|\mu, \sigma^2) dz = \frac{C}{|w|^N C_z} \tag{26}$$

We can now calculate the posterior:

$$p(z|y = \mathcal{D}_y) = \frac{p(y = \mathcal{D}_y | z)p(z)}{p(y = \mathcal{D}_y)} \tag{27}$$

$$= \frac{1}{p(y = \mathcal{D}_y)} \cdot \frac{1}{|w|^N} C \cdot \mathcal{N}(z|\mu, \sigma^2) \cdot \frac{1}{C_z} \tag{28}$$

$$= \mathcal{N}(z|\mu, \sigma^2) \tag{29}$$

CALCULATING THE POSTERIOR PREDICTIVE DISTRIBUTION (PPD)

Having found the posterior, we can determine the PPD:

$$p(y|\mathcal{D}_y) = \int p(y|z)p(z|\mathcal{D}_y)dz$$

To calculate this, we will have to use the identity for a product of two Gaussians.

$$p(y|\mathcal{D}_y) = \int \frac{1}{|w|}\mathcal{N}\left(z|\frac{y-b}{w},\left(\frac{\sigma_{y|z}}{w}\right)^2\right)\mathcal{N}(z|\mu,\sigma^2)dz \tag{30}$$

$$= \int \frac{1}{|w|}C^*\mathcal{N}\left(z|\mu^*,(\sigma^*)^2\right)dz \tag{31}$$

Next class

In the next class we will finish this example, by solving this integral to determine the PPD $p(y|\mathcal{D}_y)$.

See you next class

Have fun!