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

Lecture 6

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

Outline for today

- Continuation of previous lecture: Bayesian inference for one hidden rv
    - Prior
    - Likelihood
    - Marginal likelihood
    - Posterior
    - Gaussian pdf's product

Reading material: This notebook + Chapter 3

Recap of Lecture 5: car stopping distance with known x

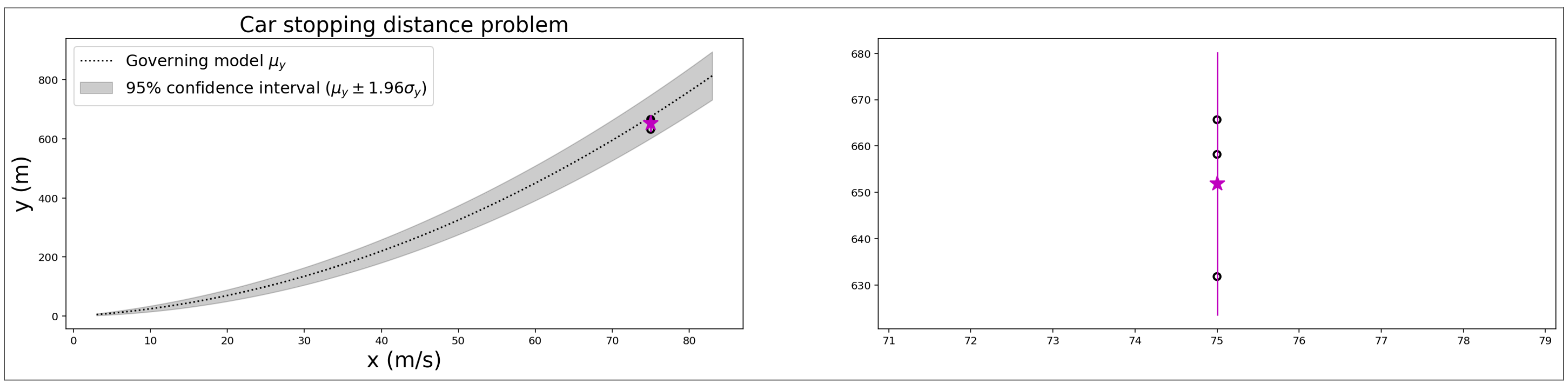
We focused on the car stopping distance problem under the following conditions:

- We kept $x = 75$ m/s.
- But the distribution of the rv z is not known: $p(z) =?$

Under these conditions, recall the "true" model by observing the following plot, including some data observations.

```
In [10]: # 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 std
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 651.907517548372 (real mean[y]=675)
Empirical std[y] is 14.50873859076062 (real std[y]=37.5)
```



Recap of Lecture 5: Summary of our model

1. The **observation distribution**:

$$p(y|z) = \mathcal{N}\left(y|\mu_{y|z} = wz + b, \sigma^2_{y|z}\right) = \frac{1}{C_{y|z}} \exp\left[-\frac{1}{2\sigma^2_{y|z}}(y - \mu_{y|z})^2\right]$$

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

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.

Recap of Lecture 5: Data

- Since we usually don't know the true process, we can only observe/collect data $y = \mathcal{D}_y$:

In [5]:
```
print("Example of N=%1i data points for y at x=%1.1f m/s with :" % (N_samples,x), empirical_y)
```

```
Example of N=3 data points for y at x=75.0 m/s with : [676.34340551 726.27176696 699.9940938
]
```

Recap of Lecture 5: Posterior from Bayes' rule applied to data

Use Bayes' rule applied to data to determine the posterior:

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

That requires calculating the likelihood (here, it results from a product of Gaussian densities):

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

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

$\sigma^2 = \frac{\sigma^2_{y|z}}{w^2 N}$, and

$C = \frac{1}{2\pi^{(N-1)/2}} \sqrt{\frac{\sigma^2}{\left(\frac{\sigma^2_{y|z}}{w^2}\right)^N}}$

After calculating the likelihood, we determined the marginal likelihood:

$$p(y = \mathcal{D}_y) = \frac{C}{|w|^N C_z}$$

From which we got the posterior:

$$
\begin{aligned}
p(z|y=\mathcal{D}_y) &= \frac{p(y=\mathcal{D}_y|z)p(z)}{p(y=\mathcal{D}_y)} && (1) \\
&= \frac{1}{p(y=\mathcal{D}_y)} \cdot \frac{1}{|w|^N} C \cdot \mathcal{N}(z|\mu,\sigma^2) \cdot \frac{1}{C_z} && (2) \\
&= \mathcal{N}(z|\mu,\sigma^2) && (3)
\end{aligned}
$$

which is a **normalized** Gaussian pdf in z with mean and variance as shown in the previous cell.

Determining the Posterior Predictive Distribution (PPD) from the posterior

However, as we mentioned, Bayes' rule is just a way to calculate the posterior:

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

What we really want is the Posterior Predictive Distribution (PPD) . This comes after calculating the posterior given 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: $p(y^*|\mathcal{D}_y) = \int p(y^*|z)p(z|\mathcal{D}_y)dz$

$$p(y^*|\mathcal{D}_y) = \int \underbrace{p(y^*|z)}_{\substack{\text{observation}\\ \text{distribution}}} \overbrace{p(z|y = \mathcal{D}_y)}^{\text{posterior}} dz$$

Considering the terms we found before, we get:

$$p(y^*|\mathcal{D}_y) = \int \underbrace{\frac{1}{|w|} \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{-\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2}\left[z - \left(\frac{y^* - b}{w}\right)\right]^2\right\}}_{\substack{\text{observation}\\ \text{distribution}}} \overbrace{\mathcal{N}(z|\mu, \sigma^2)}^{\text{posterior}} dz \tag{4}$$

$$p(y^*|\mathcal{D}_y) = \frac{1}{|w|}\int \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}}\exp\left\{-\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2}\left[z-\left(\frac{y^*-b}{w}\right)\right]^2\right\}\mathcal{N}(z|\mu,\sigma^2)dz$$

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

This is (again!) the product of two Gaussians!

In Lecture 5 (and the Homework!) you saw (and will demonstrate!) that the product of two or more univariate (and multivariate!) Gaussians is...

- Another Gaussian! Although it needs to be scaled by a constant...

So, we conclude that the PPD is an integral of a Gaussian:

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

where $\mu^* = (\sigma^*)^2 \left(\frac{\mu}{\sigma^2} + \frac{(y^*-b)/w}{\left(\frac{\sigma_{y|z}}{w}\right)^2}\right) = (\sigma^*)^2 \left(\frac{\mu}{\sigma^2} + \frac{(y^*-b)\cdot w}{\sigma^2_{y|z}}\right)$

$(\sigma^*)^2 = \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\left(\frac{\sigma_{y|z}}{w}\right)^2}} = \frac{1}{\frac{1}{\sigma^2}+\frac{w^2}{\sigma^2_{y|z}}}$

$C^* = \frac{1}{\sqrt{2\pi\left(\sigma^2+\frac{\sigma^2_{y|z}}{w^2}\right)}}\exp\left[-\frac{\left(\mu-\frac{y^*-b}{w}\right)^2}{2\left(\sigma^2+\frac{\sigma^2_{y|z}}{w^2}\right)}\right]$

This integral is simple to solve!

$$p(y^*|\mathcal{D}_y) = \frac{1}{|w|}\int C^*\mathcal{N}\left(z|\mu^*, (\sigma^*)^2\right) dz \tag{5}$$

$$= \frac{C^*}{|w|}\int \mathcal{N}\left(z|\mu^*, (\sigma^*)^2\right) dz \tag{6}$$

What's the result of integrating the blue term?

$$p(y^*|\mathcal{D}_y) = \frac{C^*}{|w|}$$

In Homework 3 (exercise 2)

Rewrite the PPD to show that it becomes:

$$p(y^*|\mathcal{D}_y) = \mathcal{N}\left(y^*|b + \mu w, w^2\sigma^2 + \sigma^2_{y|z}\right)$$

a normalized univariate Gaussian!

A long way to show that the PPD is a simple Gaussian...

$$p(y^*|\mathcal{D}_y) = \mathcal{N}\left(y^*|b + \mu w, w^2\sigma^2 + \sigma^2_{y|z}\right)$$

where we recall that each constant is:

$b = 0.1x^2 = 562.5$

$w = x = 75$

$\sigma^2_{y|z} = s^2 = 80^2$

$\sigma^2 = \frac{\sigma^2_{y|z}}{w^2 N} = \frac{s^2}{w^2 N}$

$\mu = \frac{w^2\sigma^2}{\sigma^2_{y|z}} \sum_{i=1}^{N} \mu_i = \cdots = \frac{\sum_{i=1}^{N} y_i}{wN} - \frac{b}{w}$

A long way to show that the PPD is a simple Gaussian...

$$p(y^*|\mathcal{D}_y) = \mathcal{N}\left(y^*|b + \mu w, w^2\sigma^2 + \sigma^2_{y|z}\right) \tag{12}$$

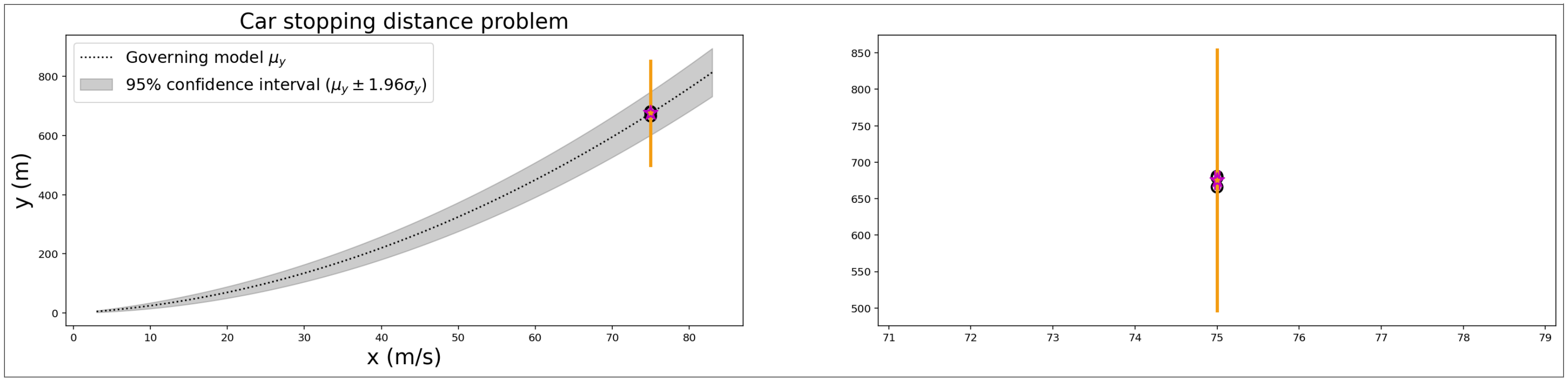
$$= \mathcal{N}\left(y^* \middle| \left(\sum_{i=1}^{N} \frac{y_i}{N}\right), \sigma^2_{y|z}\left(\frac{1}{N} + 1\right)\right) \tag{13}$$

where y_i are each of the N data points of the observed data $\mathcal{D}_y$, and $\sigma^2_{y|z} = s^2 = 80^2$ is the **variance we assumed** because we were not sure about the governing equation for y.

- **Very Important Questions (VIQs)**: What does this result tell us? Did you expect this predicted distribution for y^*?

```
In [15]: fig_car_PPD, ax_car_PPD = plt.subplots(1,2); plt.close() # create figure and close it
         if fig_car_new.get_axes():
    del ax_car_PPD; del fig_car_PPD; fig_car_PPD, ax_car_PPD = plt.subplots(1,2) # delete fig & axes & create t
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 std
# Calculate PPD mean and standard deviation:
PPD_mu_y = np.mean(empirical_y); sigma_yGIVENz = 80; PPD_sigma_y = np.sqrt( sigma_yGIVENz**2*(1/N_samples + 1)
car_fig(ax_car_PPD[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_PPD[i].errorbar(x , empirical_mu_y,yerr=1.96*empirical_sigma_y, fmt='m*', markersize=15, elinewidth=
    ax_car_PPD[i].errorbar(x , PPD_mu_y,yerr=1.96*PPD_sigma_y, color='#F39C12', fmt='*', markersize=5, elinewid
    ax_car_PPD[i].scatter(x*np.ones_like(empirical_y),empirical_y, s=100,facecolors='none', edgecolors='k', lin
print("PPD & empirical mean[y] tend to the same value:",empirical_mu_y, "(real mean[y]=675)")
print("PPD std[y] we predict is",PPD_sigma_y, "& empirical std[y] is",empirical_sigma_y,"(real std[y]=37.5)")
fig_car_PPD.set_size_inches(25, 5) # scale figure to be wider (since there are 2 subplots)
```

```
PPD & empirical mean[y] tend to the same value: 675.1441878349008 (real mean[y]=675)
PPD std[y] we predict is 92.37604307034012 & empirical std[y] is 6.625878282988706 (real std
[y]=37.5)
```



Reflection on what we are observing

1. Generally speaking, our PPD is quite reasonable and the result should be intuitive!
    - For few data points, the variance of the PPD is a bit larger than the variance we assumed for the conditional pdf $p(y|z)$, i.e. $\sigma_{y|z}^2 = s^2 = 80^2$.

1. As the number of data points increases (see PPD as $N \rightarrow \infty$ or play with the figure above by increasing N), then the variance of the PPD $\sigma_y^2 \rightarrow s^2 = 80^2$ tends to the variance we assumed for our model: $\sigma_{y|z}^2 = s^2 = 80^2$.
    - This results from our choice of likelihood and prior... Our model was incorrect in both:
        - The hidden rv z is actually a Gaussian distribution, instead of a noninformative Uniform distribution
        - The real conditional pdf is the Dirac "distribution" around value $y = zx + 0.1x^2$, instead of the Gaussian distribution with a mean of $\mu_{y|z} = zx + 0.1x^2$ and a variance of $\sigma_{y|z} = s^2 = 80^2$.

Please keep this in your head:

- (Bayesian) ML is not magic. Every modeling choice you make affects the predictions you get.
- Of course, there are ways of getting "closer" to the truth! We'll take steps in that direction in the remainder of the course.

Next lecture we will redo everything but for a different prior distribution

Consider the same problem, but now starting from a different model:

1. Same **observation distribution** as before:

$$p(y|z) = \mathcal{N}\left(y|\mu_{y|z} = wz + b, \sigma^2_{y|z}\right) = \frac{1}{C_{y|z}} \exp\left[-\frac{1}{2\sigma^2_{y|z}}(y - \mu_{y|z})^2\right]$$

1. but now assuming a different **prior distribution**: $p(z) = \mathcal{N}\left(z|\overset{<}{\mu}_z = 3, \overset{<}{\sigma}{}^2_z = 2^2\right)$

In my notation, the superscript $\overset{<}{(\cdot)}$ indicates a parameter of the prior distribution.

Notes about the prior distribution we are assuming

- We would have to be very lucky if our "belief" coincided with the "true" distribution of z.
    - Usually, we have beliefs but they are not really true (this is not a comment about religion...).
        - Our hope is that our beliefs are at least reasonable!

- When defining a prior we are making a decision about two things:
    1. The distribution.
        - For example, in this exercise we are assuming that the prior is Gaussian (before we had assumed a "noninformative" Uniform prior). In this case we hit the jackpot! But remember that we are cheating here because we already know the actual distribution of z is a Gaussian!
    2. The parameters of the distribution.
        - For example, in this exercise we are assuming values that are not the true ones! This is normal! As I said, usually we don't know the truth about the "hidden" variable. Most times we don't even know how many hidden variables we have...

See you next class

Have fun!