Bayesian Estimation

(CIV6540 - Probabilistic Machine Learning for Civil Engineers)

Professor: James-A. Goulet

Département des génies civil, géologique et des mines
Polytechnique Montréal

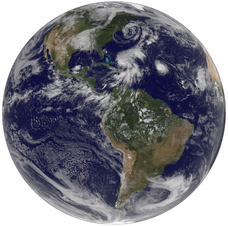
Chapter 6 – Goulet (2020)
Probabilistic Machine Learning for Civil Engineers
MIT Press

Example – Screening test

Given a deadly disease so rare that **only one human on Earth has it**.

We have a screening test so that

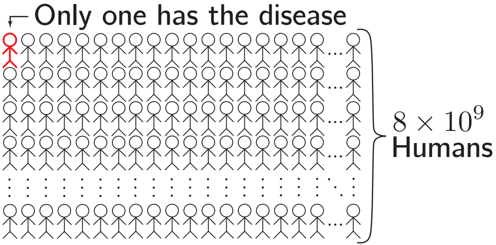
$$\text{test}+ \rightarrow \begin{cases} \Pr(\text{test}+|\text{disease}) = 0.999 \\ \Pr(\text{test}+|\neg\text{disease}) = 0.001 \end{cases}$$

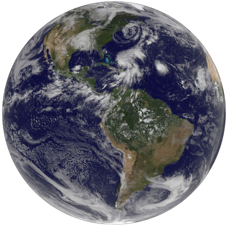


If you test positive, should you be worried?

[NASA]

Example – Screening test





We expect $\approx 0.001 \times (8 \times 10^9) = 8 \times 10^6$ **false diagnoses.**

$$\Pr(\text{disease}|\text{test}+) \approx \frac{1}{8 \times 10^6} \approx 2 \times \Pr\left(\text{649}\right)$$

If you want to **properly extract information from data, you must consider the prior probability** of the phenomenon you are interested in.

[NASA, Google Images]

Module #3 Outline

Introduction
Bayes
Discrete
Continuous
Monte Carlo
PDF parameters
Conjugate Prior

Topics organization

Background	1	Revision probability & linear algebra
	2	Probability distributions
Machine Learning Basics	0	Introduction
	3	Bayesian Estimation $p(A\|B) = \frac{p(B\|A)p(A)}{p(B)}$
	4	MCMC sampling & Newton
Supervised learning	5	Regression
	6	Classification
	7	LSTM networks for time series
Unsupervised learning	7	State-space model for time-series
Decision Making & RL	8	Decision Theory
	9	AI & Sequential decision problems

Section Outline

Why Bayes?

y: Observation

$$p(\text{unknown}|\text{known})$$

x: Constant

X: Random variable

$\boldsymbol{\theta}$: $X \sim f(x;\boldsymbol{\theta})$

Bayes's rule formulation

Given two random variables $X \sim f(x)$ and $Y \sim f(y)$, their joint PDF is obtained by the product of the conditional PDF and the marginal PDF

$$f(x,y) = \begin{cases} f(x|y) \cdot f(y) \\ f(y|x) \cdot f(x) \end{cases}$$

In the case where y **is known**, but x **is not**, we employ Bayes rule

$$f(x|y) \cdot f(y) = f(y|x) \cdot f(x)$$

$$f(x|y) = \frac{f(y|x) \cdot f(x)}{f(y)}$$

Bayes's rule

Given $\mathbf{X} = [X_1, X_2, \cdots, X_{\mathtt{X}}]^\intercal$ a vector of random variables so that $\mathbf{X} \sim f(\mathbf{x})$

and given $\mathcal{D} = \{y_1, y_2, \cdots, y_{\mathtt{D}}\}$ **a set of observations** corresponding to realizations of $\mathbf{Y} = [Y_1, Y_2, \cdots, Y_{\mathtt{D}}]^\intercal$ so that $\mathbf{Y} \sim f(\mathbf{y})$

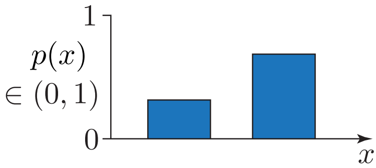
$$f(\mathbf{x}|\mathbf{y} = \mathcal{D}) = \frac{f(\mathbf{y} = \mathcal{D}|\mathbf{x}) \cdot f(\mathbf{x})}{f(\mathbf{y} = \mathcal{D})}$$

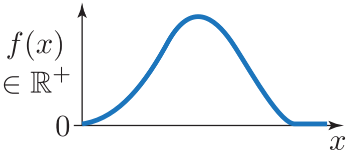
$$\underbrace{f(\mathbf{x}|\mathcal{D})}_{\text{posterior}} = \frac{\overbrace{f(\mathcal{D}|\mathbf{x})}^{\text{likelihod}} \cdot \overbrace{f(\mathbf{x})}^{\text{prior}}}{\underbrace{f(\mathcal{D})}_{\text{normalization cte.}}}$$

Prior knowledge

$f(\mathbf{x})$ describes our prior knowledge for the values that values $\mathbf{x}$ can take. Prior knowledge can be based on

- Engineering heuristics (expert knowledge)
- The posterior PDF obtained from previous data
- Non-informative prior (i.e. absence of prior knowledge)





Likelihood – $f(\mathcal{D}|\mathbf{x})$

$f(\mathbf{y}=\mathcal{D}|\mathbf{x}) \equiv f(\mathcal{D}|\mathbf{x})$ describes the conditional probability of a set of observations $\mathcal{D}$ given the values that $\mathbf{x}$ can take.

Note: you **do not need Bayes in the special case where $\mathbf{y}=\mathbf{x}$**

$$\begin{array}{rcll} p(\text{test}+|\text{disease}) & = & p(\text{test-}|\neg\text{disease}) & =1 \\ p(\text{test-}|\text{disease}) & = & p(\text{test}+|\neg\text{disease}) & =0 \end{array}$$

Because the observation is exact, the prior does not play a role.

Evidence – $f(\mathcal{D})$

$f(\mathbf{y}=\mathcal{D}) \equiv f(\mathcal{D})$ is called the *evidence* or the *normalization constant*.

The posterior integral must be equal to 1

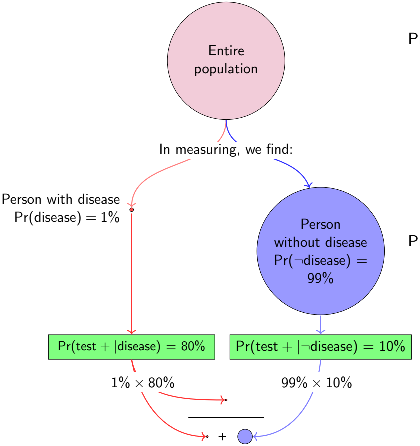
$$\underbrace{\sum_{\mathbf{x}} p(\mathbf{x}|\mathcal{D})}_{\text{discrete case}} \equiv \underbrace{\int f(\mathbf{x}|\mathcal{D})d\mathbf{x}}_{\text{continuous case}} = 1$$

so that

$$p(\mathcal{D}) = \underbrace{\sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) \cdot p(\mathbf{x})}_{\text{discrete case}}, \quad f(\mathcal{D}) = \underbrace{\int f(\mathbf{y}|\mathbf{x}) \cdot f(\mathbf{x})d\mathbf{x}}_{\text{continuous case}}$$

Section Outline

Example – Disease screening



$\Pr(\text{disease} \mid \text{test+})$

$$= \frac{\Pr(\text{test+}|\text{disease})\cdot\Pr(\text{disease})}{\Pr(\text{test+}|\text{disease})\cdot\Pr(\text{disease})+\Pr(\text{test+}|\neg\text{disease})\cdot\Pr(\neg\text{disease})}$$

$$= \frac{1\%\times 80\%}{(1\%\times 80\%)+(99\%\times 10\%)}$$

$$= 7.5\%$$

$\Pr(\neg\text{disease} \mid \text{test+})$

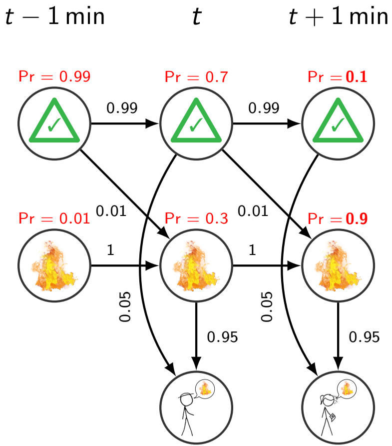
$$= \frac{\Pr(\text{test+}|\neg\text{disease})\cdot\Pr(\neg\text{disease})}{\Pr(\text{test+}|\text{disease})\cdot\Pr(\text{disease})+\Pr(\text{test+}|\neg\text{disease})\cdot\Pr(\neg\text{disease})}$$

$$= \frac{10\%\times 99\%}{(1\%\times 80\%)+(99\%\times 10\%)}$$

$$= 92.5\%$$

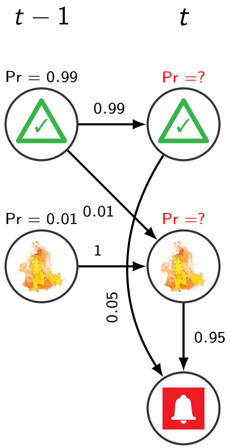
[John Henderson]

Example – Fire alarm (Role of the prior)



States : $x \in \{\text{fire}, \neg\text{fire}\} \equiv \{🔥, △\}$

Observation: $y \in \{\text{alarm}, \neg\text{alarm}\} \equiv \{🔔, \neg🔔\}$



Likelihood

$$p(🔔|x) = \begin{cases} \Pr(🔔|🔥) = 0.95 \\ \Pr(🔔|△) = 0.05 \end{cases}$$

Prior knowledge

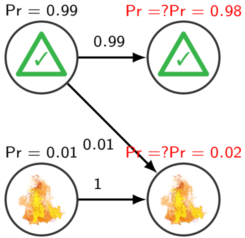
$$p(x_{t-1}) = \begin{cases} \Pr(🔥) = 0.01 \\ \Pr(△) = 0.99 \end{cases}$$

Transition probability

$$p(x_t|x_{t-1}) = \begin{cases} \Pr(△|△) = 0.99 \\ \Pr(🔥|△) = 0.01 \\ \Pr(🔥|🔥) = 1 \\ \Pr(△|🔥) = 0. \end{cases}$$

From $t-1$ to t, there are 4(3) possibilities

$t-1$ $\qquad$ t



$$\begin{aligned}\Pr(\text{🔥}_t, \text{🔥}_{t-1}) &= \Pr(\text{🔥}_t|\text{🔥}_{t-1})\cdot\Pr(\text{🔥}_{t-1})\\ &= 1\cdot 0.01 = 0.01\end{aligned}$$

$$\begin{aligned}\Pr(\text{🔥}_t, \triangle_{t-1}) &= \Pr(\text{🔥}_t|\triangle_{t-1})\cdot\Pr(\triangle_{t-1})\\ &= 0.01\cdot 0.99 = 0.0099\end{aligned}$$

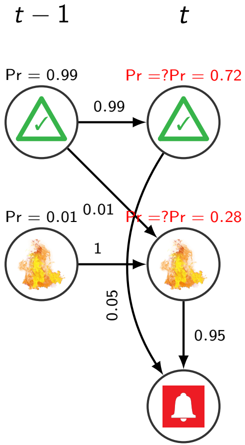
$$\begin{aligned}\Pr(\triangle_t, \triangle_{t-1}) &= \Pr(\triangle_t|\triangle_{t-1})\cdot\Pr(\triangle_{t-1})\\ &= 0.99\cdot 0.99 = 0.9801\end{aligned}$$

$$\begin{aligned}\Pr(\triangle_t, \text{🔥}_{t-1}) &= \Pr(\triangle_t|\text{🔥}_{t-1})\cdot\Pr(\text{🔥}_{t-1})\\ &= 0\cdot 0.01 = 0\end{aligned}$$

$$\begin{aligned}\Pr(\text{🔥}_t) &= \Pr(\text{🔥}_t, \text{🔥}_{t-1}) + \Pr(\text{🔥}_t, \triangle_{t-1})\\ &= 0.01 + 0.0099 = \boxed{0.0199}\end{aligned}$$

$$\begin{aligned}\Pr(\triangle_t) &= \Pr(\triangle_t, \triangle_{t-1}) + \Pr(\triangle_t, \text{🔥}_{t-1})\\ &= \boxed{0.9801}\end{aligned}$$

Example – Role of the prior



$$
\begin{aligned}
\Pr(\text{🔥}_t|\text{🔔}) &= \frac{\overbrace{\Pr(\text{🔔}|\text{🔥}_t)\cdot\Pr(\text{🔥}_t)}^{\Pr(\text{🔔},\text{🔥}_t)}}{\underbrace{(\Pr(\text{🔔}|\text{🔥}_t)\cdot\Pr(\text{🔥}_t)) + (\Pr(\text{🔔}|\triangle_t)\cdot\Pr(\triangle_t))}_{\Pr(\text{🔔})}}\\
&= \frac{0.95\cdot 0.02}{(0.95\cdot 0.02)+(0.05\cdot 0.98)}\\
&= 0.28
\end{aligned}
$$

$$
\Pr(\triangle_t|\text{🔔}) = 1-\Pr(\text{🔥}_t|\text{🔔}) = 0.72
$$

Example – post-earthquake damage assessment

Given an earthquake of intensity:

$$X : x \in \{\text{Light (L)}, \text{Moderate (M)}, \text{Important (I)}\}$$

and a structure that can be in a state:

$$Y : y \in \{\text{damaged } (D), \text{ undamaged } (U)\}$$

We know the likelihood of a damage given an intensity as well as their prior probability

$$\begin{array}{rcl} \Pr(y=D|x=L) & = & 0.01 \\ \Pr(y=D|x=M) & = & 0.10 \\ \Pr(y=D|x=I) & = & 0.60 \end{array} \quad \& \quad \left.\begin{array}{rcl} \Pr(x=L) & = & 0.90 \\ \Pr(x=M) & = & 0.08 \\ \Pr(x=I) & = & 0.02 \end{array}\right\} \sum = 1$$

$$p(x|y=D) = \frac{p(y=D|x)\cdot p(x)}{p(y=D)} = \textbf{?}$$

$$\begin{aligned} p(y=D) &= \overbrace{p(D|L)p(L)}^{p(y=D,x=L)} + \overbrace{p(D|M)p(M)}^{p(y=D,x=M)} + \overbrace{p(D|I)p(I)}^{p(y=D,x=I)} \\ &= 0.01 \times 0.90 + 0.10 \times 0.08 + 0.6 \times 0.02 \\ &= 0.029 \end{aligned}$$

$$p(x=I|y=D) = \frac{p(D|I)p(I)}{p(D)} = \frac{0.60 \times 0.02}{0.029} = 0.41$$

$$p(x=M|y=D) = \frac{p(D|M)p(M)}{p(D)} = \frac{0.10 \times 0.08}{0.029} = 0.28$$

$$p(x=L|y=D) = \frac{p(D|L)p(L)}{p(D)} = \frac{0.01 \times 0.90}{0.029} = 0.31$$

Section Outline

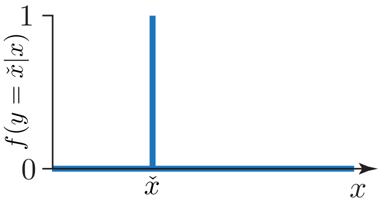
Continuous

Discrete v.s. Continuous

The main difference between the discrete and the continuous case is in the definition of the likelihood function $f(\mathcal{D}|\mathbf{x})$

Like for the discrete case, if $y = x$ there is no need for Bayes because

$$\rightarrow f(x|y) = \delta(y) = \delta(\check{x})$$



Observation model

Given $\check{x}$, **the true, yet unknown value**.
We have access to imprecise observation so that either

$$\left.\begin{array}{lcll} y & = & \check{x}+v & \text{(direct)} \\ y & < & \check{x}+v & \text{(lower-bound censored)} \\ y & > & \check{x}+v & \text{(upper-bound censored)} \end{array}\right\} v : V \sim \mathcal{N}(v;0,\sigma_V^2)$$

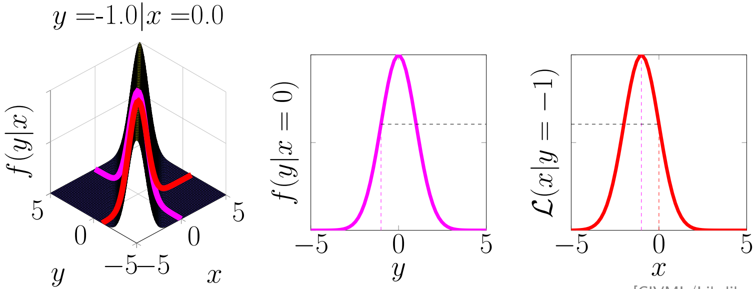
Where V is a zero-mean Gaussian observation error.
Note: $V_i \perp\!\!\!\perp V_j, \forall i \neq j$

Likelihood: $y = \check{x} + v$ [.m]

Given an observation y_1

$$\overbrace{y = \check{x} + v}^{\text{additive}}, \;\; v : V \sim \mathcal{N}(v; 0, \sigma_V^2)$$

$$f(y|x) \equiv \mathcal{L}(x|y) = \mathcal{N}(y; x, \sigma_V^2) = \frac{1}{\sqrt{2\pi} \cdot \sigma_V} \exp\left[-\frac{1}{2}\left(\frac{y-x}{\sigma_V}\right)^2\right]$$



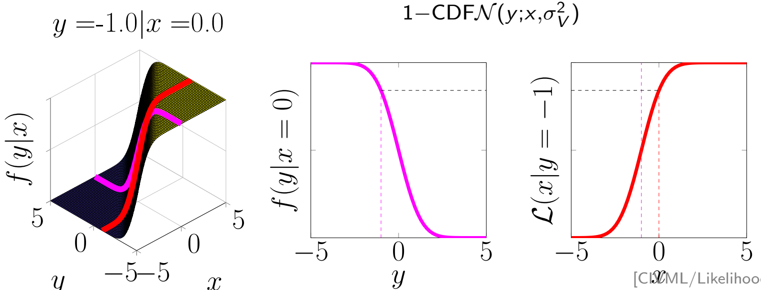
[CIVML/Likelihood_example.m]

Likelihood: $y < \check{x} + v$ [.m]

Given an observation y_1

$$\overbrace{y < \check{x} + v}^{\text{lower bound}},\ \ v : V \sim \mathcal{N}(v; 0, \sigma_V^2)$$

$$f(y < \check{x} + v|x) = \underbrace{\int_y^{\infty} \frac{1}{\sqrt{2\pi}\cdot\sigma_V} \exp\left[-\frac{1}{2}\left(\frac{y-x}{\sigma_V}\right)^2\right] dy}_{1-\text{CDF}\mathcal{N}(y;x,\sigma_V^2)}$$



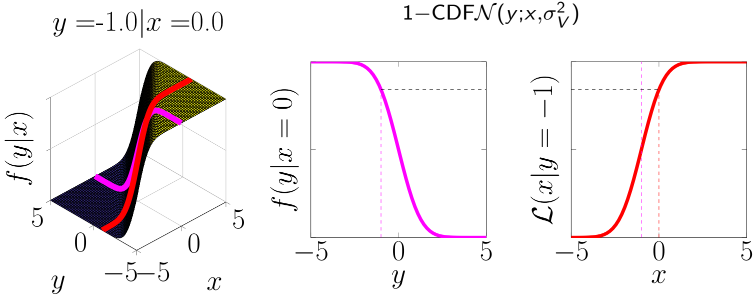
[CIVML/Likelihood_example.m]

Likelihood: $y < \check{x} + v$ [.m]

Given an observation y_1

$$\overbrace{y < \check{x} + v}^{\text{lower bound}}, \quad v : V \sim \mathcal{N}(v; 0, \sigma_V^2)$$

$$f(y < \check{x} + v|x) = \underbrace{1 - \left(\frac{1}{2} + \frac{1}{2}\text{erf}\left[\frac{y - x}{\sqrt{2}\sigma_V}\right]\right)}_{1-\text{CDF}\mathcal{N}(y;x,\sigma_V^2)}$$



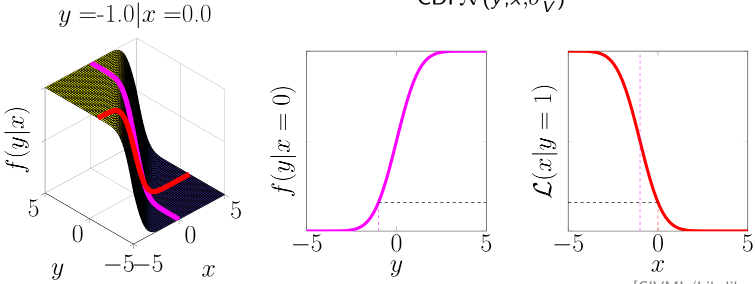
[CIVML/Likelihood_example.m]

Likelihood: $y > \check{x} + v$ [.m]

Given an observation y_1

$$\overbrace{y > \check{x} + v}^{\text{upper bound}}, \quad V \sim \mathcal{N}(v; 0, \sigma_V^2)$$

$$f(y > \check{x} + v|x) = \underbrace{\int_{-\infty}^{y} \frac{1}{\sqrt{2\pi} \cdot \sigma_V} \exp\left[-\frac{1}{2}\left(\frac{y-x}{\sigma_V}\right)^2\right] dy}_{\text{CDF}\mathcal{N}(y;x,\sigma_V^2)}$$



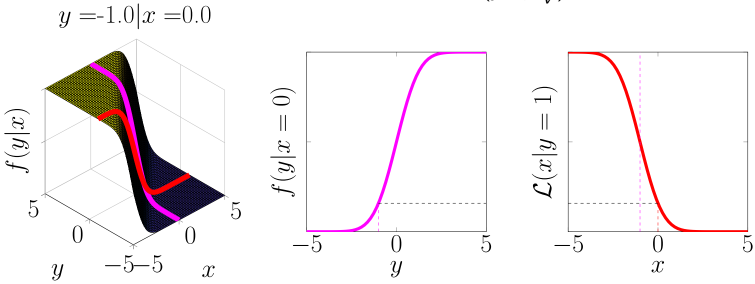
[CIVML/Likelihood_example.m]

Likelihood: $y > \check{x} + v$ [.m]

Given an observation y_1

$$\overbrace{y > \check{x} + v}^{\text{upper bound}}, \;\; V \sim \mathcal{N}(v; 0, \sigma_V^2)$$

$$f(y > \check{x} + v|x) = \underbrace{\frac{1}{2} + \frac{1}{2}\text{erf}\left[\frac{y-x}{\sqrt{2}\sigma_V}\right]}_{\text{CDF}\mathcal{N}(y;x,\sigma_V^2)}$$



[CIVML/Likelihood_example.m]

$f(\mathcal{D}|x)$: Likelihood for multiple observations

If $V_i \perp\!\!\!\perp V_j, \forall i \neq j$
(i.e. observations are independent of each others)

$$\begin{aligned} f(Y_1 = y_1, \cdots, Y_{\mathtt{Y}} = y_{\mathtt{D}}|x) &= \overbrace{\prod_{i=1}^{\mathtt{D}} f(Y_i = y_i|x)}^{=0,\mathtt{D}\gg 1} \\ &= \exp\left(\sum_{i=1}^{\mathtt{D}} \ln(f(Y_i = y_i|x))\right) \end{aligned}$$

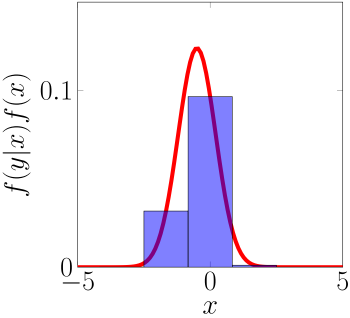
In practice, one should always employ the second formulation which is **more robust towards numerical underflow**.

$f(\mathcal{D})$: Evidence $\equiv$ Normalization constant [.m]

$$f(\mathcal{D}) = \int f(\mathcal{D}|x) \cdot f(x)dx \approx \sum_{i=1}^{N} f(\mathcal{D}|x_i) \cdot f(x_i)\Delta x_i$$

It if often not possible to perform this integration analytically.
A first crude method to perform this integration is the **rectangle rule**

$N = 7,\ \Delta x_i = \frac{10}{7},\ f(\mathcal{D}) \approx 0.2158$



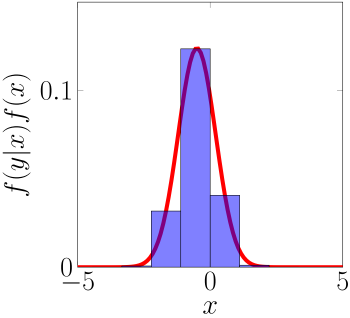
[CIVML/Likelihood_example_integration.m]

$f(\mathcal{D})$: Evidence $\equiv$ Normalization constant [.m]

$$f(\mathcal{D}) = \int f(\mathcal{D}|x) \cdot f(x)dx \approx \sum_{i=1}^{N} f(\mathcal{D}|x_i) \cdot f(x_i)\Delta x_i$$

It if often not possible to perform this integration analytically.
A first crude method to perform this integration is the **rectangle rule**

$N = 10,\ \Delta x_i = \frac{10}{10},\ f(\mathcal{D}) \approx 0.2198$



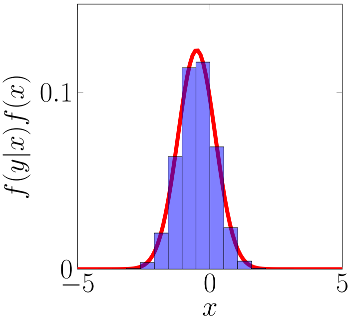
[CIVML/Likelihood_example_integration.m]

$f(\mathcal{D})$: Evidence $\equiv$ Normalization constant [.m]

$$f(\mathcal{D}) = \int f(\mathcal{D}|x) \cdot f(x)dx \approx \sum_{i=1}^{N} f(\mathcal{D}|x_i) \cdot f(x_i)\Delta x_i$$

It if often not possible to perform this integration analytically.
A first crude method to perform this integration is the **rectangle rule**

$N = 20,\ \Delta x_i = \frac{10}{10},\ f(\mathcal{D}) \approx 0.2197$



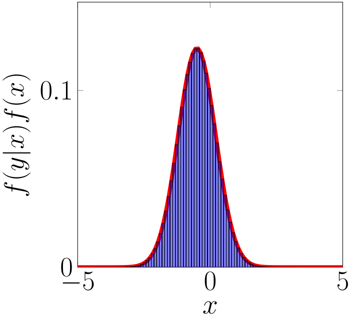
[CIVML/Likelihood_example_integration.m]

$f(\mathcal{D})$: Evidence $\equiv$ Normalization constant [.m]

$$f(\mathcal{D}) = \int f(\mathcal{D}|x) \cdot f(x)dx \approx \sum_{i=1}^{N} f(\mathcal{D}|x_i) \cdot f(x_i)\Delta x_i$$

It if often not possible to perform this integration analytically.
A first crude method to perform this integration is the **rectangle rule**

$N = 100,\ \Delta x_i = \frac{10}{100},\ f(\mathcal{D}) \approx 0.2197$



[CIVML/Likelihood_example_integration.m]

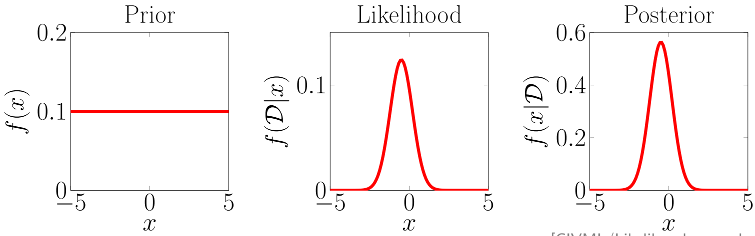
Posterior – $f(\mathbf{x}|\mathcal{D})$ []

Given two observations $\mathcal{D} = \{y_1 = -1, y_2 = 0\}$.

$$y = \check{x} + v, \;\; v : V \sim \mathcal{N}(v; 0, 1^2), \quad f(x) = \mathcal{U}(x; -5, 5)$$

What is $f(\mathbf{x}|\mathcal{D})$?

$$f(x|\mathcal{D}) = \frac{f(\mathcal{D}|x) \cdot f(x)}{f(\mathcal{D})}, \quad f(\mathcal{D}) \approx 0.2197$$



[CIVML/Likelihood_example_integration.m]

Limitation – $f(\mathcal{D})$

$$\begin{aligned} f(\mathcal{D}) &= \int f(\mathcal{D}|x)\cdot f(x)dx \\ &\approx \sum_{i=1}^{N} f(\mathcal{D}|x_i)\cdot f(x_i)\Delta x_i \end{aligned}$$

...the rectangle method is limited to 1D or 2D problems.

(Note: # dimension corresponds to the # of parameters $\mathtt{X}$)

Solution → Monte Carlo sampling

Monte Carlo Sampling – Example

Estimating the area of a circle using Monte-Carlo sampling?

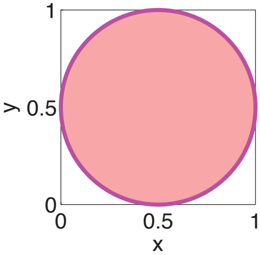
Given a circle of diameter $d = 1$ ($r = 0.5$) so that

$$(x - 0.5)^2 + (y - 0.5)^2 = r^2$$

and given the indicator function

$$I(x, y) = \begin{cases} 1 & \text{if } (x - 0.5)^2 + (y - 0.5)^2 \leq r^2 \\ 0 & \textit{else} \end{cases}$$

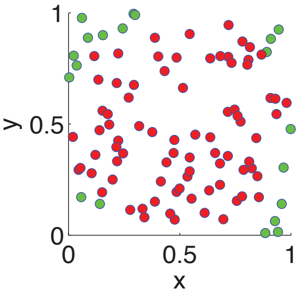
$$\begin{aligned} \underbrace{a}_{\text{area}} &= \int_y \int_x I(x, y) f_{XY}(x, y) dxdy \\ &= \mathbb{E}[I(X, Y)] \end{aligned}$$



Monte-Carlo Sampling (cont.) [.m]

What is the area, a?

If we genreate samples $\{x_s, y_s\}$, $s = 1, 2, \cdots, \mathtt{S}$ following a uniform probability density, i.e. $f_{XY}(x,y) = 1$.



$$a = \mathbb{E}[I(X,Y)] = \lim_{\mathtt{S}\to\infty} \frac{1}{\mathtt{S}} \sum_{s=1}^{\mathtt{S}} I(x_s, y_s)$$

$$a \cong \mathbb{E}[\widehat{I(X,Y)}] = \frac{1}{\mathtt{S}} \sum_{s=1}^{\mathtt{S}} I(x_s, y_s)$$

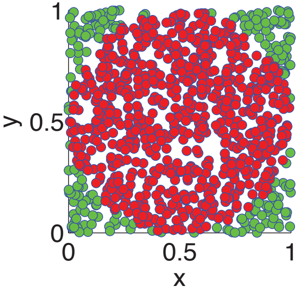
For $\mathtt{S} = 100$, $\mathbb{E}[\widehat{I(X,Y)}] = 0.800$ $(\pi r^2 = 0.785)$

[CIV8530/M4_Numerical_integration_exemple.m]

Monte-Carlo Sampling (cont.) [.m]

What is the area, a?

If we genreate samples $\{x_s, y_s\}$, $s = 1, 2, \cdots, \mathtt{S}$ following a uniform probability density, i.e. $f_{XY}(x, y) = 1$.



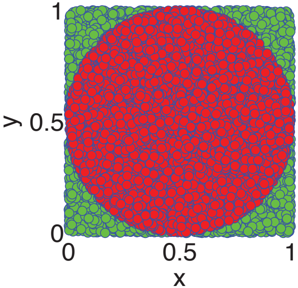
For $\mathtt{S} = 1\,000$, $\mathbb{E}[\widehat{I(X,Y)}] = 0.759$ ($\pi r^2 = 0.785$)

[CIV8530/M4_Numerical_integration_exemple.m]

Monte-Carlo Sampling (cont.) [.m]

What is the area, a?

If we genreate samples $\{x_s, y_s\}$, $s = 1, 2, \cdots, \mathtt{S}$ following a uniform probability density, i.e. $f_{XY}(x, y) = 1$.



For $\mathtt{S} = 10\,000$, $\mathbb{E}[\widehat{I(X, Y)}] = 0.784$ $(\pi r^2 = 0.785)$

[CIV8530/M4_Numerical_integration_exemple.m]

Monte-Carlo Sampling (cont.) [.m]

What is the area, a?

If we genreate samples $\{x_s, y_s\}$, $s = 1, 2, \cdots, \mathtt{S}$ following a uniform probability density, i.e. $f_{XY}(x, y) = 1$.

Estimation quality:

1 - $\mathbb{E}[\widehat{I(X,Y)}]$ depends on the number of samples

2- Independent of the number of dimensions

[CIV8530/M4_Numerical_integration_exemple.m]

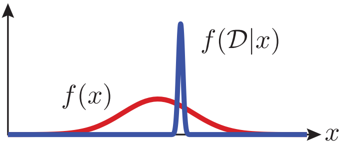
Evidence – $f(\mathcal{D})$

$$f(\mathcal{D}) = \int f(\mathcal{D}|\mathbf{x})f(\mathbf{x})d\mathbf{x} = \mathbb{E}_{\mathbf{x}}[f(\mathcal{D}|\mathbf{x})]$$

$$\widehat{f(\mathcal{D})}_{\mathrm{MC}} = \frac{1}{\mathtt{S}}\sum_{s=1}^{\mathtt{S}} \frac{f(\mathcal{D}|\mathbf{x}_s)\cdot f(\mathbf{x}_s)}{h(\mathbf{x}_s)}, \; \mathbf{x}_s : \mathbf{X}_s \sim h(\mathbf{x}) : \text{Sampling PDF}$$

$$= \frac{1}{\mathtt{S}}\sum_{s=1}^{\mathtt{S}} f(\mathcal{D}|\mathbf{x}_s), \; \mathbf{x}_s : \mathbf{X}_s \sim h(\mathbf{x}) = f(\mathbf{x})$$

Warning: Monte Carlo integrating with $h(\mathbf{x}) = f(\mathbf{x})$ is not an option when the number of observations $\mathtt{D}$ is large.



Posterior – $f(\mathbf{x}|\mathcal{D})$

$$f(\mathbf{x}|\mathcal{D}) = \frac{f(\mathcal{D}|\mathbf{x})f(\mathbf{x})}{f(\mathcal{D})}$$

$f(\mathbf{x}|\mathcal{D}) \rightarrow$ Most often, it does not have a simple analytic solution

Solution: compute summary statistics

$$\mathbb{E}[\mathbf{x}|\mathcal{D}] = \int \mathbf{x} \cdot f(\mathbf{x}|\mathcal{D})d\mathbf{x}$$

$$\text{var}[\mathbf{x}|\mathcal{D}] = \int (\mathbf{x} - \mathbb{E}[\mathbf{x}|\mathcal{D}])^2 \cdot f(\mathbf{x}|\mathcal{D})d\mathbf{x}$$

Summary statistics estimation using MC – $f(\mathbf{x}|\mathcal{D})$

$$
\begin{aligned}
\mathbb{E}[\mathbf{x}|\mathcal{D}] &= \int \mathbf{x}\cdot f(\mathbf{x}|\mathcal{D})d\mathbf{x}\\
&\approx \frac{1}{\mathrm{S}}\sum_{s=1}^{\mathrm{S}} \mathbf{x}_s \cdot \frac{f(\mathcal{D}|\mathbf{x}_s)\cdot f(\mathbf{x}_s)}{f(\mathcal{D})}\cdot\frac{1}{h(\mathbf{x}_s)}\\
&\approx \frac{1}{\mathrm{S}}\sum_{s=1}^{\mathrm{S}} \mathbf{x}_s \cdot \frac{f(\mathcal{D}|\mathbf{x}_s)}{f(\mathcal{D})},\ \mathbf{x}_s : \mathbf{X}_s \sim h(\mathbf{x}) = f(\mathbf{x})
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}[\mathbf{x}|\mathcal{D}] &= \int (\mathbf{x}-\mathbb{E}[\mathbf{x}|\mathcal{D}])^2\cdot f(\mathbf{x}|\mathcal{D})d\mathbf{x}\\
&\approx \frac{1}{\mathrm{S}}\sum_{s=1}^{\mathrm{S}} (\mathbf{x}_s-\mathbb{E}[\mathbf{x}|\mathcal{D}])^2 \cdot \frac{f(\mathcal{D}|\mathbf{x}_s)\cdot f(\mathbf{x}_s)}{f(\mathcal{D})}\cdot\frac{1}{h(\mathbf{x}_s)}\\
&\approx \frac{1}{\mathrm{S}}\sum_{s=1}^{\mathrm{S}} (\mathbf{x}_s-\mathbb{E}[\mathbf{x}|\mathcal{D}])^2 \cdot \frac{f(\mathcal{D}|\mathbf{x}_s)}{f(\mathcal{D})},\ \mathbf{x}_s : \mathbf{X}_s \sim h(\mathbf{x}) = f(\mathbf{x})
\end{aligned}
$$

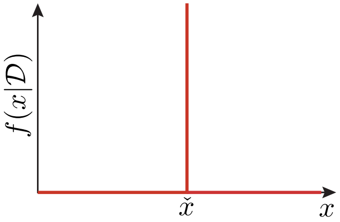
Posterior – $f(\mathbf{x}|\mathcal{D})$: effect of $\nearrow$ D

When the number of independent observations $\mathtt{D} \to \infty$

$$f(\mathbf{x}|\mathcal{D}) = \frac{f(\mathcal{D}|\mathbf{x})f(\mathbf{x})}{f(\mathcal{D})} \to \underbrace{\delta(\overbrace{\check{\mathbf{x}}}^{\text{true value}})}_{\text{Dirac delta PDF}}$$

$$\mathbb{E}[\mathbf{x}|\mathcal{D}] \to \underbrace{\check{\mathbf{x}}}_{\text{true value}}$$

$$\text{var}[\mathbf{x}|\mathcal{D}] \to \underbrace{0}_{\text{true value}}$$



*For **identifiable** problems

Posterior – $f(\mathbf{x}|\mathcal{D})$: $\nearrow$ $\mathtt{D}$ & non-identifiability [.m]

We want to estimate the **contaminant concentration in two streams** $\{x_1, x_2\}$, and where we only have access to contaminant concentration observations

$$y_i = x_1 + x_2 + v,\;\; v : V \sim \mathcal{N}(v; 0, 2^2)$$

at the output of a river after both streams have merged

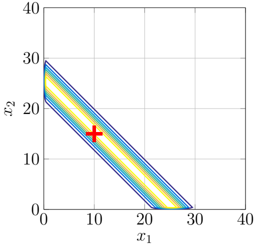
Prior: $f(x_1, x_2) \propto 1$

Posterior: $f(\mathbf{x}|\mathcal{D}) =?$

$\mathtt{D} = 1$

$y_1 = 25.8$

$f(x_1, x_2) \propto 1$



[CIVML/Example_identifiability.m]

Posterior – $f(\mathbf{x}|\mathcal{D})$: $\nearrow$ $\mathtt{D}$ & non-identifiability [.m]

We want to estimate the **contaminant concentration in two streams** $\{x_1, x_2\}$, and where we only have access to contaminant concentration observations

$$y_i = x_1 + x_2 + v, \;\; v : V \sim \mathcal{N}(v; 0, 2^2)$$

at the output of a river after both streams have merged

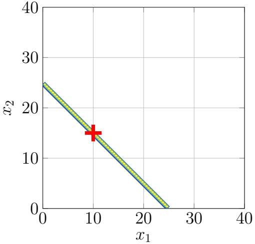
Prior: $f(x_1, x_2) \propto 1$

Posterior: $f(\mathbf{x}|\mathcal{D}) = ?$

$$\mathtt{D} = 100$$

$$y_1 = 25.8, \cdots, y_{100} = 24.6$$

$$f(x_1, x_2) \propto 1$$



[CIVML/Example_identifiability.m]

Posterior – $f(\mathbf{x}|\mathcal{D})$: $\nearrow$ D & non-identifiability [.m]

We want to estimate the **contaminant concentration in two streams** $\{x_1, x_2\}$, and where we only have access to contaminant concentration observations

$$y_i = x_1 + x_2 + v, \;\; v : V \sim \mathcal{N}(v; 0, 2^2)$$

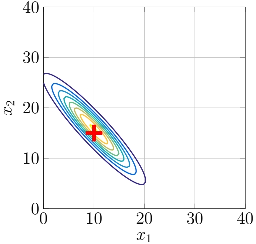
at the output of a river after both streams have merged

Prior: $f(x_1, x_2) \propto \mathcal{N}(x_1; 10, 5^2)$

Posterior: $f(\mathbf{x}|\mathcal{D}) = ?$

$\mathtt{D} = 1$

$f(x_1, x_2) \propto \mathcal{N}(x_1; 10, 5^2)$



[CIVML/Example_identifiability.m]

Posterior – $f(\mathbf{x}|\mathcal{D})$: $\nearrow$ D & non-identifiability [.m]

We want to estimate the **contaminant concentration in two streams** $\{x_1, x_2\}$, and where we only have access to contaminant concentration observations

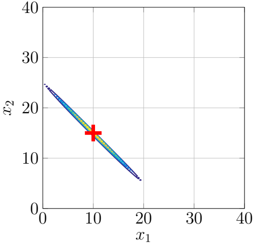
$$y_i = x_1 + x_2 + v, \;\; v : V \sim \mathcal{N}(v; 0, 2^2)$$

at the output of a river after both streams have merged

Prior: $f(x_1, x_2) \propto \mathcal{N}(x_1; 10, 5^2)$

Posterior: $f(\mathbf{x}|\mathcal{D}) = ?$

$\mathtt{D} = 100$
$f(x_1, x_2) \propto$
$\mathcal{N}(x_1; 10, 5^2)$



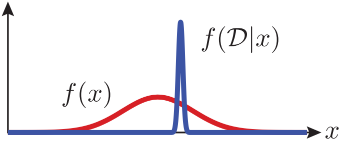
[CIVML/Example_identifiability.m]

Summary – Monte-Carlo Sampling

For D observations and S samples : $\mathbf{x}_s : \mathbf{X}_s \sim h(\mathbf{x}) = f(\mathbf{x})$

$$
\begin{aligned}
f(\mathcal{D}) &\approx \frac{1}{\mathtt{S}}\sum_{s=1}^{\mathtt{S}}\left[\prod_{j=1}^{\mathtt{D}} f(y_j|\mathbf{x}_s)\right]\\
\mathbb{E}[\mathbf{x}|\mathcal{D}] &\approx \frac{1}{\mathtt{S}}\sum_{s=1}^{\mathtt{S}}\left[\mathbf{x}_s\cdot\frac{\prod_{j=1}^{\mathtt{D}} f(y_j|\mathbf{x}_s)}{f(\mathcal{D})}\right]\\
\text{var}[\mathbf{x}|\mathcal{D}] &\approx \frac{1}{\mathtt{S}}\sum_{s=1}^{\mathtt{S}}\left[(\mathbf{x}_s-\mathbb{E}[\mathbf{x}|\mathcal{D}])^2\cdot\frac{\prod_{j=1}^{\mathtt{D}} f(y_j|\mathbf{x}_s)}{f(\mathcal{D})}\right]
\end{aligned}
$$

Warning: Monte Carlo integration using $h(\mathbf{x}) = f(\mathbf{x})$ is not a good choice when D is large.



Example – Lab [] estimation

A laboratory test for measuring [ ] Hg follows

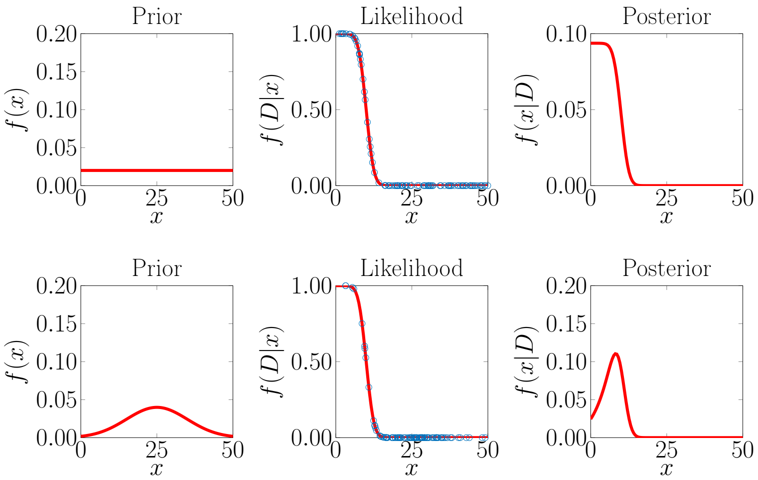
$$y : \begin{cases} y = \check{x} + v, & if\ \check{x} + v \geq 10\mu g/L \\ y = \texttt{error}, & if\ \check{x} + v < 10\mu g/L \end{cases}$$

$$v : V \sim \mathcal{N}(v; 0, 2^2)$$

What is the posterior PDF describing the [ ] if $\mathcal{D} = \{\texttt{error}\}$ if the prior is either

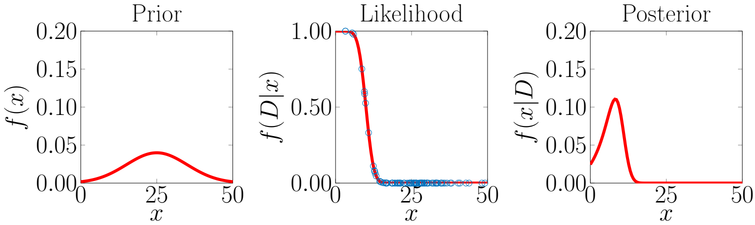
- $f(x) = \mathcal{U}(x; 0, 50)$
- $f(x) = \mathcal{N}(x; 25, 10^2)$

Example – Lab [] estimation [.m]



[CIVML/contamination_lab_test.m]

Example – Lab [] estimation [.m]



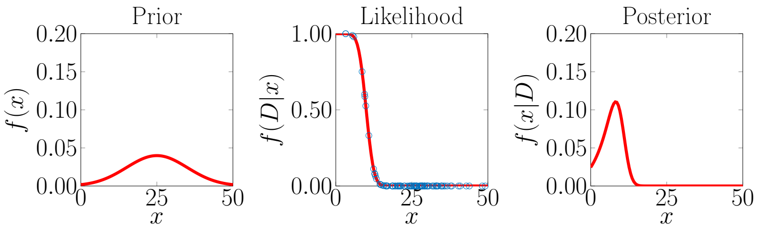
$$f(\mathcal{D}) \approx \frac{1}{\mathtt{S}}\sum_{s=1}^{\mathtt{S}}\left[\prod_{j=1}^{\mathtt{D}} f(y_j|\mathbf{x}_s)\right] = 0.067$$

$$\mathbb{E}[\mathbf{x}|\mathcal{D}] \approx \frac{1}{\mathtt{S}}\sum_{s=1}^{\mathtt{S}}\left[\mathbf{x}_s \cdot \frac{\prod_{j=1}^{\mathtt{D}} f(y_j|\mathbf{x}_s)}{f(\mathcal{D})}\right] = 6.27$$

$$\mathrm{var}[\mathbf{x}|\mathcal{D}] \approx \frac{1}{\mathtt{S}}\sum_{s=1}^{\mathtt{S}}\left[(\mathbf{x}_s - \mathbb{E}[\mathbf{x}|\mathcal{D}])^2 \cdot \frac{\prod_{j=1}^{\mathtt{D}} f(y_j|\mathbf{x}_s)}{f(\mathcal{D})}\right] = 17.96$$

[CIVML/contamination_lab_test.m]

Example – Lab [] estimation



$$f(\mathcal{D}) \approx \frac{1}{\mathrm{S}}\sum_{s=1}^{\mathrm{S}}[f(y|\mathbf{x}_s)] = 0.067$$

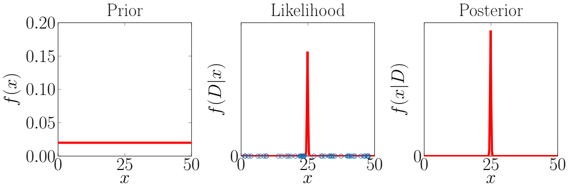
$$\mathbb{E}[\mathbf{x}|\mathcal{D}] \approx \frac{1}{\mathrm{S}}\sum_{s=1}^{\mathrm{S}}\left[\mathbf{x}_s \cdot \frac{f(y|\mathbf{x}_s)}{f(\mathcal{D})}\right] = 6.27$$

$$\mathrm{var}[\mathbf{x}|\mathcal{D}] \approx \frac{1}{\mathrm{S}}\sum_{s=1}^{\mathrm{S}}\left[(\mathbf{x}_s - \mathbb{E}[\mathbf{x}|\mathcal{D}])^2 \cdot \frac{f(y|\mathbf{x}_s)}{f(\mathcal{D})}\right] = 17.96$$

[CIVML/contamination_lab_test.m]

Example – Lab [] estimation [.m]

What is the posterior PDF describing the []
if $\mathcal{D} = \{25.4, 23.6, 24.1, \cdots, 24.3\}_{1\times 100}$
and the prior is $p(x) = \mathcal{U}(x; 0, 50)$?



$$
\begin{aligned}
f(\mathcal{D}) &= 1.19\times 10^{-78} \\
&\neq \tfrac{1}{S}\sum_{s=1}^{50}\left[\prod_{j=1}^{100} f(y_j|\mathbf{x}_s)\right] = 1.95\times 10^{-83}
\end{aligned}
$$

Solution: $\nearrow$ # MC samples or MCMC (Module #4)

[CIVML/contamination_lab_test.m]

Section Outline

PDF parameters

determininistic v.s. random variables

So far, we have employed data $\mathcal{D}$ to estimate our posterior knowledge given the hypothesis that **a true quantity $\check{x}$ exists**.

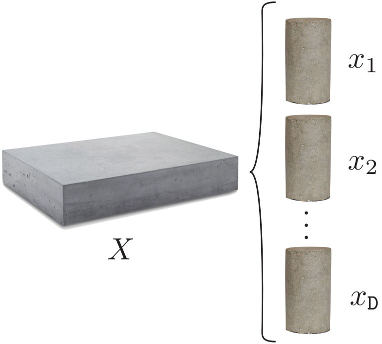
$$f(x|\mathcal{D}) = \frac{f(\mathcal{D}|x)\cdot f(x)}{f(\mathcal{D})}$$

When the number of independent observations $\mathtt{D} \to \infty$

$$\mathbb{E}[x|\mathcal{D}] \to \underbrace{\check{\mathbf{x}}}_{\text{true value}}, \quad \text{var}[x|\mathcal{D}] \to \underbrace{0}_{\text{true value}}$$

What happens if a true quantity $\check{x}$ does not exists because X is a random variable?

Bayes for random variables



- X: random variable, $f_X(x;\boldsymbol{\theta})$
- $\boldsymbol{\theta}$: PDF's parameters
- $\{x_1, x_2, \cdots, x_{\mathtt{D}}\}$: realizations of X

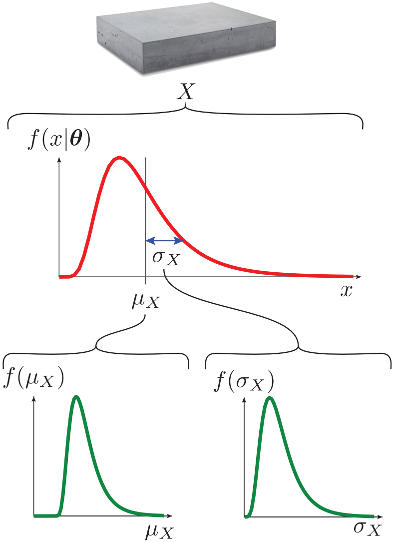
$$f(\boldsymbol{\theta}|\mathcal{D}) = \frac{f(\mathcal{D}|\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta})}{f(\mathcal{D})}$$

$$X \sim f(x|\boldsymbol{\theta}) : \text{Random variable}$$

$$\underbrace{\mu_X, \sigma_X}_{\boldsymbol{\theta}} : \text{Parameters of } f(x|\boldsymbol{\theta})$$

$$\left.\begin{aligned} \mu_X &\sim f(\mu_x) \\ \sigma_X &\sim f(\sigma_x) \end{aligned}\right\} \text{Prior}$$

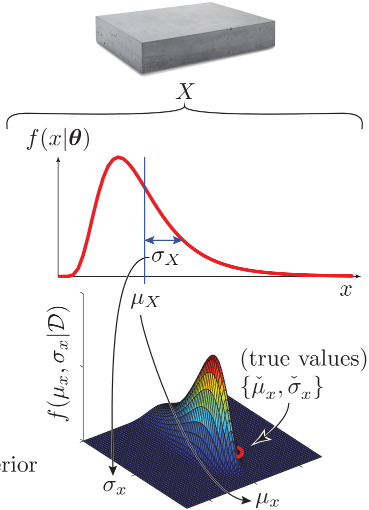
$$\left.\begin{aligned} \mu_X|\mathcal{D} &\sim f(\mu_x|\mathcal{D}) \\ \sigma_X|\mathcal{D} &\sim f(\sigma_x|\mathcal{D}) \end{aligned}\right\} \text{Posterior}$$



$X \sim f(x|\boldsymbol{\theta})$: Random variable

$\underbrace{\mu_X, \sigma_X}_{\boldsymbol{\theta}}$: Parameters of $f(x|\boldsymbol{\theta})$

$\{\mu_X, \sigma_X\} \sim f(\mu_x, \sigma_X)\}$ Prior
$\{\mu_X, \sigma_X\}|\mathcal{D} \sim f(\mu_x, \sigma_X|\mathcal{D})\}$ Posterior



Notation

$x : X \sim f(x) \equiv f(x;\boldsymbol{\theta})$: PDF of X

$\boldsymbol{\theta} : \boldsymbol{\theta} \sim f(\boldsymbol{\theta})$: Prior PDF of unknown parameters

$f(\mathcal{D}|\boldsymbol{\theta})$: Likelihood of observations $\mathcal{D} = [y_1, \cdots, y_{\mathtt{D}}]^\intercal$

$f(\boldsymbol{\theta}|\mathcal{D})$: Posterior PDF for unknown parameters

$f(\mathbf{x}|\mathcal{D}) = \int f(x;\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta}$: Posterior predictive PDF for the random variable X

Prior – $f(\theta)$

$f(\boldsymbol{\theta})$ describes our prior knowledge for the values that values $\boldsymbol{\theta}$ can take. Prior knowledge can be based on

- Engineering heuristics (expert knowledge)
- The posterior PDF obtained from previous data
- Non-informative prior (i.e. absence of prior knowledge)
  - For the mean: $f(\mu) \propto 1$
  - For the standard deviation: $f(\sigma) \propto \frac{1}{\sigma}$
  - ⚠ You cannot sample form a non-informative prior because it is not a proper probability distribution, $\int f(\mu)d\mu = \infty \neq 1$. (Solution: MCMC methods → Module 4)

Likelihood – $f(\mathcal{D}|\boldsymbol{\theta})$

$f(\mathcal{D}|\boldsymbol{\theta}) \equiv f(\mathbf{Y}=\mathbf{y}|\boldsymbol{\theta}) \equiv f(\mathbf{y}|\boldsymbol{\theta}) \equiv \mathcal{L}(\boldsymbol{\theta}|\mathbf{y})$: likelihood of observing $\mathcal{D}$ given $\boldsymbol{\theta}$.

Given an observation $\mathcal{D} = \{y_1\}$ being a realization of $\underbrace{y = x + v}_{\text{additive}}$,

where $v : V \sim \mathcal{N}(v; 0, \sigma_V^2)$ and $x : X \sim \mathcal{N}(x; \underbrace{\mu_X, \sigma_X^2}_{\boldsymbol{\theta}})$

$$f(y|\boldsymbol{\theta}) = \mathcal{N}(y; \mu_X, \sigma_X^2 + \sigma_V^2)$$

Given an observation $\mathcal{D} = \{y_1\}$ being a realization of $\overbrace{y = x \cdot v}^{\text{multiplicative}}$,
where $v : V \sim \ln\mathcal{N}(v; 0, \sigma_{\ln V}^2)$ and $x : X \sim \ln\mathcal{N}(x; \underbrace{\mu_{\ln X}, \sigma_{\ln X}^2}_{\boldsymbol{\theta}})$

$$f(y|\boldsymbol{\theta}) = \ln\mathcal{N}(y; \mu_{\ln X}, \sigma_{\ln X}^2 + \sigma_{\ln V}^2)$$

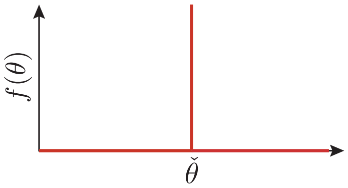
$f(\boldsymbol{\theta}|\mathcal{D})$: Posterior PDF – Effect of D

When the number of independent observations $\texttt{D} \rightarrow \infty$

$$f(\boldsymbol{\theta}|\mathcal{D}) = \frac{f(\mathcal{D}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathcal{D})} \rightarrow \underbrace{\delta(\overbrace{\ \check{\boldsymbol{\theta}}\ }^{\text{true value}})}_{\text{Dirac delta PDF}}$$

$$\mathbb{E}[\boldsymbol{\theta}|\mathcal{D}] \rightarrow \underbrace{\check{\boldsymbol{\theta}}}_{\text{true value}}$$

$$\text{var}[\boldsymbol{\theta}|\mathcal{D}] \rightarrow \underbrace{0}_{\text{true value}}$$



Posterior predictive – $f(x|\mathcal{D})$

Using Bayes:

$$f(\boldsymbol{\theta}|\mathcal{D}) = \frac{f(\mathcal{D}|\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta})}{f(\mathcal{D})}$$

From there, we seek to obtain the **posterior predictive PDF**,

$$\boxed{f(x|\mathcal{D}) = \int f(x;\boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta}}$$

Monte-Carlo sampling – Posterior PDF

For $\mathtt{D}$ observations and $\mathtt{S}$ samples : $\boldsymbol{\theta}_s \sim h(\boldsymbol{\theta}) = f(\boldsymbol{\theta})$

$$f(\mathcal{D}) \approx \frac{1}{\mathtt{S}} \sum_{s=1}^{\mathtt{S}} \left[\prod_{j=1}^{\mathtt{D}} f(y_j|\boldsymbol{\theta}_s) \right]$$

$$\mathbb{E}[\boldsymbol{\theta}|\mathcal{D}] \approx \frac{1}{\mathtt{S}} \sum_{s=1}^{\mathtt{S}} \left[\boldsymbol{\theta}_s \cdot \frac{\prod_{j=1}^{\mathtt{D}} f(y_j|\boldsymbol{\theta}_s)}{f(\mathcal{D})} \right]$$

$$\mathrm{var}[\boldsymbol{\theta}|\mathcal{D}] \approx \frac{1}{\mathtt{S}} \sum_{s=1}^{\mathtt{S}} \left[(\boldsymbol{\theta}_s - \mathbb{E}[\boldsymbol{\theta}|\mathcal{D}])^2 \cdot \frac{\prod_{j=1}^{\mathtt{D}} f(y_j|\boldsymbol{\theta}_s)}{f(\mathcal{D})} \right]$$

Monte-Carlo sampling – Posterior predictive PDF

For D observations and S samples : $\boldsymbol{\theta}_s \sim h(\boldsymbol{\theta}) = f(\boldsymbol{\theta})$

$$\mathbf{x}_s : \mathbf{X} \quad \sim \quad f(\mathbf{x};\boldsymbol{\theta}_s)$$

$$\mathbb{E}[\mathbf{X}|\mathcal{D}] \quad \approx \quad \frac{1}{\mathtt{S}}\sum_{s=1}^{\mathtt{S}}\left[\mathbf{x}_s \cdot \frac{\prod_{j=1}^{\mathtt{D}} f(y_j|\boldsymbol{\theta}_s)}{f(\mathcal{D})}\right]$$

$$\mathrm{var}[\mathbf{X}|\mathcal{D}] \quad \approx \quad \frac{1}{\mathtt{S}}\sum_{s=1}^{\mathtt{S}}\left[(\mathbf{x}_s - \mathbb{E}[\mathbf{x}|\mathcal{D}])^2 \cdot \frac{\prod_{j=1}^{\mathtt{D}} f(y_j|\boldsymbol{\theta}_s)}{f(\mathcal{D})}\right]$$

Example - Concrete strength

Reality (unknown!, $\check{\mu}_R = 42\,MPa$, $\check{\sigma}_R = 2\,MPa$)

$$R \sim \ln\mathcal{N}(r; \underbrace{\mu_{\ln R}}_{\mu_R=?}, \underbrace{\sigma_{\ln R}}_{\sigma_R=?})\,\text{MPa}$$

(Note: Concrete strength (R) has an intrinsic variability)

We seek:

$$\boxed{\overbrace{f(\mu_R, \sigma_R|\mathcal{D})}^{\text{Posterior PDF}} = \frac{\overbrace{f(Y_1 = y_1, \cdots, Y_{\mathtt{D}} = y_{\mathtt{D}}|\mu_R, \sigma_R)}^{\text{Likelihood}} \cdot \overbrace{f(\mu_R, \sigma_R)}^{\text{Prior PDF}}}{\underbrace{f(Y_1 = y_1, \cdots, Y_{\mathtt{D}} = y_{\mathtt{D}})}_{\text{Normalization constant}}}}$$

$$\overbrace{f(\mu_R,\sigma_R|\mathcal{D})}^{\text{Posterior PDF}} = \frac{\overbrace{f(Y_1=y_1,\cdots,Y_\mathtt{D}=y_\mathtt{D}|\mu_R,\sigma_R)}^{\text{Likelihood}}\cdot\overbrace{f(\mu_R,\sigma_R)}^{\text{Prior PDF}}}{\underbrace{f(Y_1=y_1,\cdots,Y_\mathtt{D}=y_\mathtt{D})}_{\text{Normalization constant}}}$$

Prior knowledge: $f(\mu_R,\sigma_R)=f(\mu_R)\cdot f(\sigma_R),\ (\mu_R \perp\!\!\!\perp \sigma_R)$

$$\underbrace{f(\mu_R)=\ln\mathcal{N}(\mu_R;\mu_{\ln\mu_R},\sigma_{\ln\mu_R})}_{\substack{\mu_{\mu_R}=45\,\text{MPa}\\ \sigma_{\mu_R}=7\,\text{MPa}}},\quad \underbrace{f(\sigma_R)=\ln\mathcal{N}(\sigma_R;\mu_{\ln\sigma_R},\sigma_{\ln\sigma_R})}_{\substack{\mu_{\sigma_R}=5\,\text{MPa}\\ \sigma_{\sigma_R}=2\,\text{MPa}}}$$

$$\overbrace{f(\mu_R,\sigma_R|\mathcal{D})}^{\text{Posterior PDF}} = \frac{\overbrace{f(Y_1=y_1,\cdots,Y_{\mathtt{D}}=y_{\mathtt{D}}|\mu_R,\sigma_R)}^{\text{Likelihood}}\cdot\overbrace{f(\mu_R,\sigma_R)}^{\text{Prior PDF}}}{\underbrace{f(Y_1=y_1,\cdots,Y_{\mathtt{D}}=y_{\mathtt{D}})}_{\text{Normalization constant}}}$$

Press capacity $> R$

Observation errors: $\ln y = \ln r + v,\ \ v : V \sim \mathcal{N}(v; 0, 0.01^2)\,\text{MPa}$

Likelihood: $f(y|\boldsymbol{\theta}) = \ln\mathcal{N}(y; \mu_{\ln R}, \sigma^2_{\ln R} + \underbrace{\sigma^2_V}_{=0.01^2})$

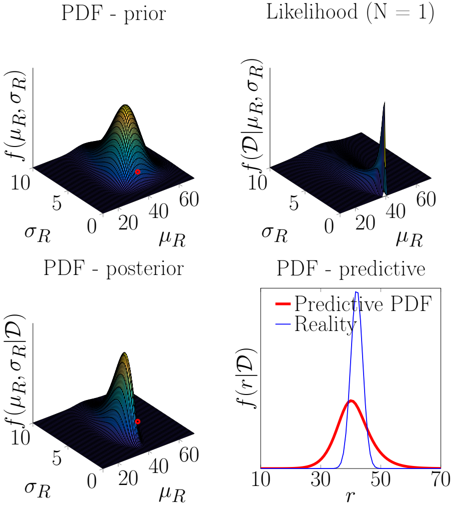
$$f(Y_1=y_1,\cdots,Y_{\mathtt{D}}=y_{\mathtt{D}}|\mu_R,\sigma_R) = \prod_{i=1}^{\mathtt{D}} \underbrace{\frac{1}{y_i\sqrt{2\pi}\sqrt{\sigma^2_{\ln R}+0.01^2}}\exp\left[-\frac{1}{2}\left(\frac{\ln y_i - \mu_{\ln R}}{\sqrt{\sigma^2_{\ln R}+0.01^2}}\right)^2\right]}_{\text{Lognormal PDF}}$$

Concrete strength – Results [.m]

$y_1 = 39.7$ MPa

Posterior predictive

$$f(r|\mathcal{D}) = \int f(r|\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}$$



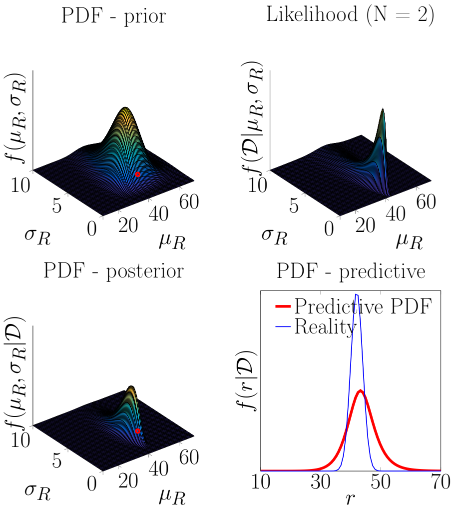
[CIVML/Example_Bayes_concrete.m]

Concrete strength – Results [ ]

$y_1 = 39.7$ MPa
$y_2 = 42.0$ MPa

Posterior predictive

$$f(r|\mathcal{D}) = \int f(r|\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}$$



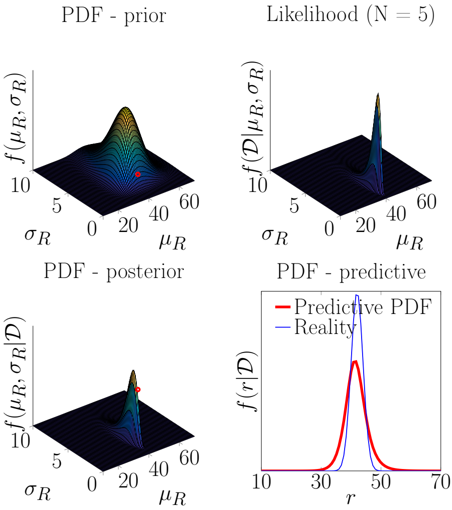
[CIVML/Example_Bayes_concrete.m]

Concrete strength – Results [.m]

D = 5

Posterior predictive

$$f(r|\mathcal{D}) = \int f(r|\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}$$



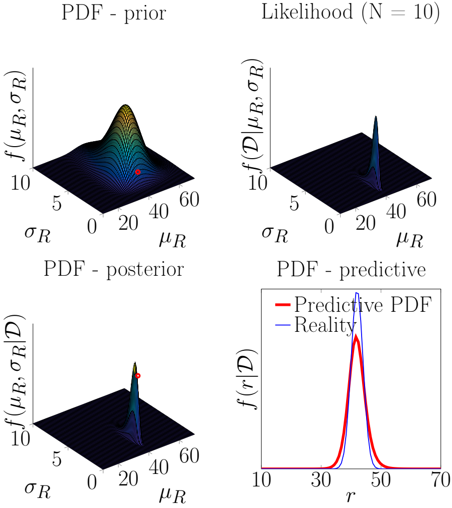
[CIVML/Example_Bayes_concrete.m]

Concrete strength – Results [.m]

$\mathtt{D} = 10$

Posterior predictive

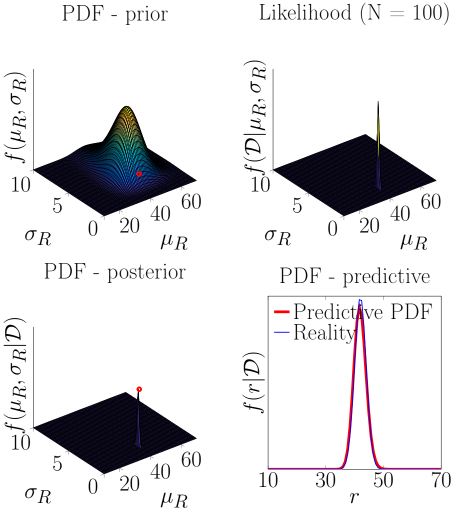
$$f(r|\mathcal{D}) = \int f(r|\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}$$



[CIVML/Example_Bayes_concrete.m]

Concrete strength – Results [.m]

$\mathtt{D} = 100$

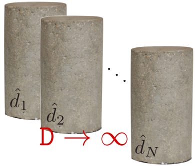


[CIVML/Example_Bayes_concrete.m]

Posterior predictive

$$f(r|\mathcal{D}) = \int f(r|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta}$$

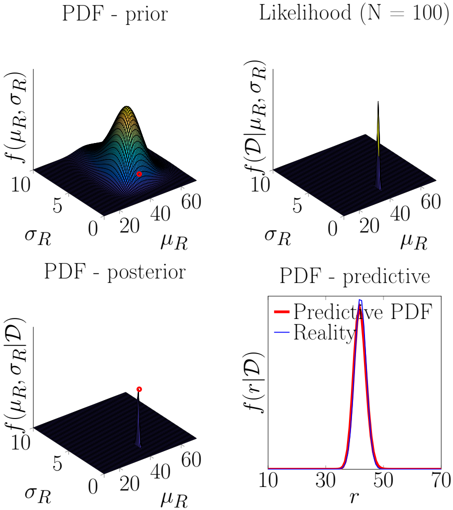
Concrete strength – Results [.m]



$$f(\boldsymbol{\theta}|\mathcal{D}) \rightarrow \underbrace{\delta(\check{\boldsymbol{\theta}})}_{\text{Dirac delta PDF}}$$

$$\mathbb{E}[\boldsymbol{\theta}|\mathcal{D}] \rightarrow \underbrace{\check{\boldsymbol{\theta}}}_{\text{true value}}$$

$$f(r|\mathcal{D}) \rightarrow \underbrace{f(r)}_{\text{true PDF}}$$



[CIVML/Example_Bayes_concrete.m]

Section Outline

Conjugate Prior

Definition – Conjugate prior

For specific combinaisons of **prior** distribution and **likelihood** function, the **posterior PDF follows the same type of distribution than the prior PDF**.

Moreover, **analytic formulation for the parameters of the posterior** PDF exists.

Example – Conjugate prior [.m]

$$\mathcal{S} = \{\text{Head, Tail}\}, \quad \Pr(\{\text{Head}\}) = X$$

What is the likelihood $f(\underbrace{\{\text{Head}, \text{Tail}, \text{Head}\}}_{\mathcal{D}}|x)$**?**

[CIVML/Beta_example.m]

Beta-Binomial – Conjugate prior

If the **prior follows a Beta** PDF,

$$f(x) = \mathcal{B}(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathbf{B}(\alpha, \beta)}$$

the **likelihood** of binary observations $\mathcal{D}_0 = \{0, 0, \cdots, 0\}_{1\times \mathtt{D}_0}$ $\mathcal{D}_1 = \{1, 1, \cdots, 1\}_{1\times \mathtt{D}_1}$ **follows a Binomial** distribution

$$f(\mathcal{D}|x) \propto x^{\mathtt{D}_1}(1-x)^{\mathtt{D}_0}$$

then the **posterior follows a Beta** PDF

$$\begin{aligned} f(x|\mathcal{D}) &= x^{\mathtt{D}_1}(1-x)^{\mathtt{D}_0} \cdot \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathbf{B}(\alpha, \beta)} \\ &= \frac{x^{\alpha+\mathtt{D}_1-1}(1-x)^{\beta+\mathtt{D}_0-1}}{\mathbf{B}(\alpha+\mathtt{D}_1, \beta+\mathtt{D}_0)} \\ &= \mathcal{B}(x; \alpha+\mathtt{D}_1, \beta+\mathtt{D}_0) \end{aligned}$$

Example – traffic safety

Given an intersection where the number of users is $u = 3 \times 10^5/\textbf{yr}$ and for which $a = 20$ accidents occurred **over the last five years**.

Following **modifications to decrease the probability** of accident π_{old}, we observe $a = 1$ accident over a period of one year.

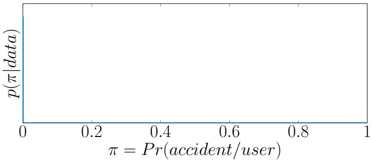
What is the predictive probability of accident π_{new} for this intersection after modification?

Example – traffic safety [.m]

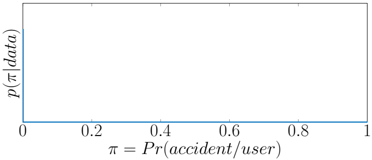
$$f(\pi|\mathcal{D}_{\text{old}}) = \mathcal{B}(\pi; 20, 5\times 3\times 10^5) \quad \text{(Conjugate)}$$

$$f(\pi|\mathcal{D}_{\text{new}}) \propto \underbrace{\pi^1\cdot(1-\pi)^{3\times 10^5}}_{\text{likelihood}}\cdot\underbrace{(1-F(\pi|\mathcal{D}_{\text{old}}))}_{\text{prior(upper bound)}} \quad (\neg\,\text{Conjugate})$$

Before interventions | # users $= 1.4999\times 10^6$, # accidents $= 20$



After interventions | # users $= 3.0903\times 10^5$, #a ccidents $= 1$



We cannot see anything if we plot the Beta PDF in the original space. **Solution** $\rightarrow$ transform in $\log_{10}$-space

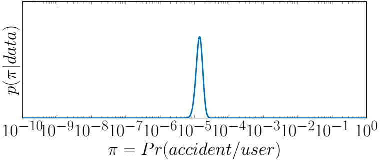
$$f(\log_{10}(\pi)) = f(\pi)\cdot\left|\frac{d\log_{10}(\pi)}{d\pi}\right|^{-1}$$

[CIVML/congugate_prior_intersection_safety.m]

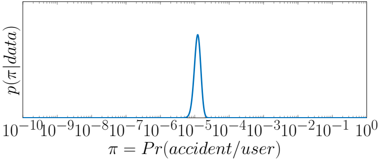
Example – traffic safety [.m]

$$
\begin{aligned}
f(\pi|\mathcal{D}_{\text{old}}) &= \mathcal{B}(\pi; 20, 5\times 3\times 10^5) && \text{(Conjugate)}\\
f(\pi|\mathcal{D}_{\text{new}}) &\propto \underbrace{\pi^1\cdot(1-\pi)^{3\times 10^5}}_{\text{likelihood}}\cdot\underbrace{(1-F(\pi|\mathcal{D}_{\text{old}}))}_{\text{prior(upper bound)}} && (\neg\,\text{Conjugate})
\end{aligned}
$$

Before interventions | # users $= 1.4999\times 10^6$, # accidents $= 20$



After interventions | # users $= 3.0903\times 10^5$, #a ccidents $= 1$



We cannot see anything if we plot the Beta PDF in the original space. **Solution** $\rightarrow$ transform in $\log_{10}$-space

$$f(\log_{10}(\pi)) = f(\pi)\cdot\pi\log(10)$$

[CIVML/congugate_prior_intersection_safety.m]

List of conjugate priors

Parameter	Likelihood	Prior/Posterior	Posterior predictive
$\Pr(E_i)$: probability	Binomial	Beta	$\Pr(\cdot)$
N: nb. of occurence	Poisson	Gamma	Negative Binomial
μ: mean	Normal	Normal	Normal
σ: std.	$\mathcal{IG}$	$\mathcal{IG}$	$\mathcal{T}$
μ, σ: mean and std.	...	$\mathcal{NIG}$	$\mathcal{T}$
$\boldsymbol{\mu}, \boldsymbol{\Sigma}$: mean and covariance	Normal	$\mathcal{NIG}$	$\mathcal{T}$

Complete list:

http://en.wikipedia.org/wiki/Conjugate_prior

Bayes's rule: $$\underbrace{f(\mathbf{x}|\mathcal{D})}_{\text{posterior}} = \frac{\overbrace{f(\mathcal{D}|\mathbf{x})}^{\text{likelihod}} \cdot \overbrace{f(\mathbf{x})}^{\text{prior}}}{\underbrace{f(\mathcal{D})}_{\text{normalization cte.}}}$$

Prior – $f(x), f(\theta)$: based on: Engineering heuristics, Previous posterior PDF, Non-informative prior

Likelihood – $f(\mathcal{D}|x)$ or $f(\mathcal{D}|\theta)$: Conditional probability of a set of observations $\mathcal{D}$ given the values that $\mathbf{x}$ or $\boldsymbol{\theta}$ can take

Evidence – $f(\mathcal{D})$:

$$f(\mathcal{D}) = \underbrace{\int f(\mathbf{y}|\mathbf{x}) \cdot f(\mathbf{x}) d\mathbf{x}}_{\text{continuous case}} = 1$$

MC sampling from the prior

$$\begin{aligned} f(\mathcal{D}) &\approx \frac{1}{\mathtt{S}} \sum_{s=1}^{\mathtt{S}} \left[\prod_{j=1}^{\mathtt{D}} f(y_j|\mathbf{x}_s)\right] \\ \mathbb{E}[\mathbf{x}|\mathcal{D}] &\approx \frac{1}{\mathtt{S}} \sum_{s=1}^{\mathtt{S}} \left[\mathbf{x}_s \cdot \frac{\prod_{j=1}^{\mathtt{D}} f(y_j|\mathbf{x}_s)}{f(\mathcal{D})}\right] \\ \text{var}[\mathbf{x}|\mathcal{D}] &\approx \frac{1}{\mathtt{S}} \sum_{s=1}^{\mathtt{S}} \left[(\mathbf{x}_s - \mathbb{E}[\mathbf{x}|\mathcal{D}])^2 \cdot \frac{\prod_{j=1}^{\mathtt{D}} f(y_j|\mathbf{x}_s)}{f(\mathcal{D})}\right] \end{aligned}$$

Limited to simple cases $\rightarrow$ MCMC Module 4

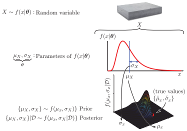
Posterior – $f(\mathbf{x}|\mathcal{D})$: When the number of independent observations $\mathtt{D} \rightarrow \infty$

$$f(\mathbf{x}|\mathcal{D}) = \frac{f(\mathcal{D}|\mathbf{x})f(\mathbf{x})}{f(\mathcal{D})} \rightarrow \underbrace{\delta(\overbrace{\check{\mathbf{x}}}^{\text{true value}})}_{\text{Dirac delta PDF}}$$

$$f(\boldsymbol{\theta}|\mathcal{D}) = \frac{f(\mathcal{D}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathcal{D})} \rightarrow \underbrace{\delta(\overbrace{\check{\boldsymbol{\theta}}}^{\text{true value}})}_{\text{Dirac delta PDF}}$$

Posterior Predictive – $f(\mathbf{x}|\mathcal{D})$

$$f(x|\mathcal{D}) = \int f(x;\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}$$



Conjugate priors: For specific combinaisons of prior distribution and likelihood function, the posterior PDF follows the same type of distribution than the prior PDF