Pushing Dynamic Estimation to the Extremes: from the Moon to Imprecise Probability

A. Benavoli
alessio@idsia.ch

July 1, 2013
This talk is about:

1. Hidden Markov Models
   More precisely:
   - estimate of the state \( X_t \) given all observations \( y_1, \ldots, y_t \).

2. Why/How is Imprecise Probability important for this problem?
Example: the Sinking of the Ship

An Engineer, a Bayesian and an Imprecise guy are involved in a sinking

to be rescued, they need to communicate via radio an accurate position of the raft;

they know the position before the sinking but, in the meantime, the raft keeps moving on a wavy sea;

they have a sextant on-board.

A discussion begins about the optimal way to obtain the best accurate position of the raft.
Example: the Sinking of the Ship

An Engineer, a Bayesian and an Imprecise guy are involved in a sinking to be rescued, they need to communicate via radio an accurate position of the raft;

- they know the position before the sinking but, in the meantime, the raft keeps moving on a wavy sea;
- they have a sextant on-board.

A discussion begins about the optimal way to obtain the best accurate position of the raft.
Example: the Sinking of the Ship

An Engineer, a Bayesian and an Imprecise guy are involved in a sinking to be rescued, they need to communicate via radio an accurate position of the raft;

- they know the position before the sinking but, in the meantime, the raft keeps moving on a wavy sea;
- they have a sextant on-board.

A discussion begins about the optimal way to obtain the best accurate position of the raft.
The engineer’s solution

We know the GPS position of the ship and its variance:

\[ x_0 \leftarrow \hat{x}_0, \sigma_0^2 \]

before the sinking (time 0)

Assume that the raft is moving along the wind direction, so its position is:

\[ x_{k+1} = x_k + d + w_k \quad \text{for} \quad k = 0, 1, 2, \ldots, t \]

where \( d \) is a drift term (due to the wind) and

\[ w_k \leftarrow 0, \sigma_w^2 \]

is a disturbance (due to the waves).

Then at time \( k = 1 \),

\[ x_1 \leftarrow \hat{x}_{1|0}, \sigma_{1|0}^2 \]

with \( \hat{x}_{1|0} = \hat{x}_0 + d \), \( \sigma_{1|0}^2 = \sigma_0^2 + \sigma_w^2 \)
The engineer’s solution

We know the GPS position of the ship and its variance:

\[ x_0 \leftarrow \hat{x}_0, \sigma_0^2 \]

before the sinking (time 0)

Assume that the raft is moving along the wind direction, so its position is:

\[ x_{k+1} = x_k + d + w_k \quad \text{for} \quad k = 0, 1, 2, \ldots, t \]

where \( d \) is a drift term (due to the wind) and \( w_k \leftarrow 0, \sigma_w^2 \)

is a disturbance (due to the waves).

Then at time \( k = 1 \),

\[ x_1 \leftarrow \hat{x}_{1|0}, \sigma_{1|0}^2 \]

with \( \hat{x}_{1|0} = \hat{x}_0 + d \), \( \sigma_{1|0}^2 = \sigma_0^2 + \sigma_w^2 \)
The engineer’s solution

We know the GPS position of the ship and its variance:

\[ x_0 \leftarrow \hat{x}_0, \sigma_0^2 \]

before the sinking (time 0)

Assume that the raft is moving along the wind direction, so its position is:

\[ x_{k+1} = x_k + d + w_k \quad \text{for} \quad k = 0, 1, 2, \ldots, t \]

where \( d \) is a drift term (due to the wind) and

\[ w_k \leftarrow 0, \sigma_w^2 \]

is a disturbance (due to the waves).

Then at time \( k = 1 \),

\[ x_1 \leftarrow \hat{x}_{1|0}, \sigma_{1|0}^2 \]

with \( \hat{x}_{1|0} = \hat{x}_0 + d \), \( \sigma_{1|0}^2 = \sigma_0^2 + \sigma_w^2 \)
Without measurements, the uncertainty in the position will keep growing until we have no clue anymore about where we are.

With the sextant, we can measure our position:

\[ y_1 = x_1 + \nu_1 \]

where \( \nu \) is the measurement error. Assuming that

\[ \nu_1 \sim 0, \sigma_\nu^2, \]

if we use the measurements, we can improve our previous estimate:

\[ x_1 \leftarrow \hat{x}_{1|0}, \sigma_{1|0}^2 \]

as follows:
The engineer’s solution

Without measurements, the uncertainty in the position will keep growing until we have no clue anymore about where we are.

With the sextant, we can measure our position:

\[ y_1 = x_1 + v_1 \]

where \( v \) is the measurement error. Assuming that

\[ v_1 \leftarrow 0, \sigma_v^2, \]

if we use the measurements, we can improve our previous estimate:

\[ x_1 \leftarrow \hat{x}_{1|0}, \sigma_{1|0}^2 \]

as follows:
Without measurements, the uncertainty in the position will keep growing until we have no clue anymore about where we are.

With the sextant, we can measure our position:

\[ y_1 = x_1 + \nu_1 \]

where \( \nu \) is the measurement error. Assuming that

\[ \nu_1 \leftarrow 0, \sigma^2_{\nu}, \]

if we use the measurements, we can improve our previous estimate:

\[ x_1 \leftarrow \hat{x}_{1|0}, \sigma^2_{1|0} \]

as follows:
The engineer’s solution

\[ x_1 \leftarrow \hat{x}_1, \sigma_1^2 \]

where

\[
\hat{x}_1 = \frac{1}{\sigma_{1|0}^2} \hat{x}_{1|0} + \frac{1}{\sigma_v^2} y_1 \]

the estimate is a weighted average the measurement \( y_1 \) and \( \hat{x}_{10} \), and

\[
\frac{1}{\sigma_1^2} = \frac{1}{\sigma_{1|0}^2} + \frac{1}{\sigma_v^2}
\]

For instance, if \( \sigma_{1|0}^2 = \sigma_v^2 = \sigma^2 \) then \( \hat{x}_1 = \frac{\hat{x}_{1|0} + y_1}{2} \) and \( \sigma_1^2 = \frac{\sigma^2}{2} \).
The engineer’s solution

\[ x_1 \leftarrow \hat{x}_1, \sigma^2_1 \]

where

\[
\hat{x}_1 = \frac{\frac{1}{\sigma^2_{1|0}}}{\frac{1}{\sigma^2_{1|0}} + \frac{1}{\sigma^2_v}} \hat{x}_{1|0} + \frac{\frac{1}{\sigma^2_v}}{\frac{1}{\sigma^2_{1|0}} + \frac{1}{\sigma^2_v}} y_1
\]

the estimate is a weighted average the measurement \( y_1 \) and \( \hat{x}_{1|0} \), and

\[
\frac{1}{\sigma^2_1} = \frac{1}{\sigma^2_{1|0}} + \frac{1}{\sigma^2_v}
\]

For instance, if \( \sigma^2_{1|0} = \sigma^2_v = \sigma^2 \) then

\[
\hat{x}_1 = \frac{\hat{x}_{1|0} + y_1}{2} \quad \text{and} \quad \sigma^2_1 = \frac{\sigma^2}{2}
\]
The engineer’s solution

Summing up: dynamic and measurement model

\[ x_{k+1} = x_k + d + w_k, \quad y_k = x_k + v_k \]

assumptions:

\[ w_k \leftarrow 0, \sigma_w^2, \quad v_k \leftarrow 0, \sigma_v^2, \]

Initialization:

\[ x_0 \leftarrow \hat{x}_0, \sigma_0^2, \]

Repeat for \( k = 1, \ldots, t \):

\[
\begin{align*}
\hat{x}_{k|k-1} &= \hat{x}_{k-1} + d \\
\sigma_{k|k-1}^2 &= \sigma_{k-1}^2 + \sigma_w^2 \\
\hat{x}_k &= \hat{x}_{k|k-1} + K_k (y_k - \hat{x}_{k|k-1}) \\
\sigma_k^2 &= (1 - K_k) \sigma_{k|k-1}^2 \\
K_k &= \frac{\sigma_{k|k-1}^2}{\sigma_{k|k-1}^2 + \sigma_v^2}
\end{align*}
\]

Solution:

\[ x_t \leftarrow \hat{x}_t, \sigma_t^2 \]
The engineer’s solution

Summing up: dynamic and measurement model

\[ x_{k+1} = x_k + d + w_k, \quad y_k = x_k + v_k \]

assumptions:

\[ w_k \leftarrow 0, \sigma^2_w, \quad v_k \leftarrow 0, \sigma^2_v, \]

Initialization:

\[ x_0 \leftarrow \hat{x}_0, \sigma^2_0, \]

Repeat for \( k = 1, \ldots, t \):

\[
\begin{align*}
\hat{x}_{k|k-1} &= \hat{x}_{k-1} + d \\
\sigma^2_{k|k-1} &= \sigma^2_{k-1} + \sigma^2_w \\
\hat{x}_k &= \hat{x}_{k|k-1} + K_k (y_k - \hat{x}_{k|k-1}) \\
\sigma^2_k &= (1 - K_k) \sigma^2_{k|k-1} \\
K_k &= \frac{\sigma^2_{k|k-1}}{\sigma^2_{k|k-1} + \sigma^2_v}
\end{align*}
\]

Solution:

\[ x_t \leftarrow \hat{x}_t, \sigma^2_t \]
The engineer’s solution

Summing up: dynamic and measurement model

\[ x_{k+1} = x_k + d + w_k, \quad y_k = x_k + v_k \]

assumptions:

\[ w_k \leftarrow 0, \sigma_w^2, \quad v_k \leftarrow 0, \sigma_v^2, \]

Initialization:

\[ x_0 \leftarrow \hat{x}_0, \sigma_0^2, \]

Repeat for \( k = 1, \ldots, t \):

\[
\begin{align*}
\hat{x}_{k|k-1} &= \hat{x}_{k-1} + d \\
\sigma_{k|k-1}^2 &= \sigma_{k-1}^2 + \sigma_w^2 \\
\hat{x}_k &= \hat{x}_{k|k-1} + K_k (y_k - \hat{x}_{k|k-1}) \\
\sigma_k^2 &= (1 - K_k) \sigma_{k|k-1}^2 \\
K_k &= \frac{\sigma_{k|k-1}^2}{\sigma_{k|k-1}^2 + \sigma_v^2}
\end{align*}
\]

Solution:

\[ x_t \leftarrow \hat{x}_t, \sigma_t^2 \]
The Bayesian solution

Model:
\[
\begin{align*}
x_{k+1} &= x_k + d + w_k \\
y_k &= x_k + v_k
\end{align*}
\]

Goal: to estimate \(x_t\) given the measurements \(y_1, y_2, \ldots, y_t\) and the prior knowledge:
\[
x_0 \leftarrow \hat{x}_0, \sigma_0^2, \quad w_k \leftarrow 0, \sigma_w^2, \quad v_k \leftarrow 0, \sigma_v^2.
\]

Since \(v_k \in \mathbb{R}\) is a measurement error

a natural choice is:
\[
p(v_k) = N(v_k; 0, \sigma_v^2)
\]

and, thus, for mathematical convenience:
\[
p(x_0) = N(x_0; \hat{x}_0, \sigma_0^2), \quad p(w_k) = N(w_k; 0, \sigma_w^2)
\]
The Bayesian solution

Model:

\[
\begin{align*}
    x_{k+1} &= x_k + d + w_k \\
    y_k &= x_k + v_k
\end{align*}
\]

Goal: to estimate \( x_t \) given the measurements \( y_1, y_2, \ldots, y_t \) and the prior knowledge:

\[
x_0 \leftarrow \hat{x}_0, \sigma_0^2, \quad w_k \leftarrow 0, \sigma_w^2, \quad v_k \leftarrow 0, \sigma_v^2.
\]

Since

\[ v_k \in \mathbb{R} \] is a measurement error

a natural choice is:

\[ p(v_k) = N(v_k; 0, \sigma_v^2) \]

and, thus, for mathematical convenience:

\[ p(x_0) = N(x_0; \hat{x}_0, \sigma_0^2), \quad p(w_k) = N(w_k; 0, \sigma_w^2) \]
The Bayesian solution

Model:

\[
\begin{aligned}
  x_{k+1} &= x_k + d + w_k \\
  y_k &= x_k + v_k
\end{aligned}
\]

Goal: to estimate \( x_t \) given the measurements \( y_1, y_2, \ldots, y_t \) and the prior knowledge:

\[
x_0 \leftarrow \hat{x}_0, \sigma_0^2, \quad w_k \leftarrow 0, \sigma_w^2, \quad v_k \leftarrow 0, \sigma_v^2.
\]

Since

\( v_k \in \mathbb{R} \) is a measurement error

a natural choice is:

\[
p(v_k) = N(v_k; 0, \sigma_v^2)
\]

and, thus, for mathematical convenience:

\[
p(x_0) = N(x_0; \hat{x}_0, \sigma_0^2), \quad p(w_k) = N(w_k; 0, \sigma_w^2)
\]
The Bayesian solution

From the assumption

\[ p(v_k) = N(v_k; 0, \sigma_v^2), \quad p(w_k) = N(w_k; 0, \sigma_w^2) \]

and

\[
\begin{align*}
    x_{k+1} &= x_k + d + w_k \\
    y_k &= x_k + v_k
\end{align*}
\]

it follows that:

\[ p(x_{k+1} | x_k) = N(x_{k+1}; x_k + d, \sigma_w^2), \quad p(y_k | x_k) = N(y_k; x_k, \sigma_v^2) \]

Goal

to estimate \( x_t \) given the measurements \( y_1, y_2, \ldots, y_t \), which means to compute

\[ p(x_t | y^t) \]

here \( y^t = \{y_1, y_2, \ldots, y_t\} \).
The Bayesian solution

From the assumption

\[ p(v_k) = N(v_k; 0, \sigma_v^2), \quad p(w_k) = N(w_k; 0, \sigma_w^2) \]

and

\[
\begin{cases}
  x_{k+1} &= x_k + d + w_k \\
y_k &= x_k + v_k
\end{cases}
\]

it follows that:

\[ p(x_{k+1}|x_k) = N(x_{k+1}; x_k + d, \sigma_w^2), \quad p(y_k|x_k) = N(y_k; x_k, \sigma_v^2) \]

Goal

to estimate \( x_t \) given the measurements \( y_1, y_2, \ldots, y_t \), which means to compute

\[ p(x_t|y^t) \]

here \( y^t = \{y_1, y_2, \ldots, y_t\} \).
The Bayesian solution

Solution: assuming that we have already computed

\[ p(x_{t-1}|y^{t-1}) \]

then

- **Prediction:**
  \[ p(x_t|y^{t-1}) = \int_{x_{t-1}} p(x_t|x_{t-1}) p(x_{t-1}|y^{t-1}) \, dx_{t-1} \]

- **Correction:**
  \[ p(x_t|y^t) = \frac{\int_{x_t} p(y_t|x_t)p(x_t|y^{t-1}) \, dx_t}{\int_{x_t} p(y_t|x_t)p(x_t|y^{t-1}) \, dx_t} \]

We can then compute recursively the previous steps starting from the initialization:

\[ p(x_0|y^{-1}) = p(x_0). \]
The Bayesian solution

Solution: assuming that we have already computed

\[ p(x_{t-1}|y^{t-1}) \]

then

- Prediction:

  \[ p(x_t|y^{t-1}) = \int_{x_{t-1}} p(x_t|x_{t-1})p(x_{t-1}|y^{t-1}) \, dx_{t-1} \]

- Correction:

  \[ p(x_t|y^t) = \frac{p(y_t|x_t)p(x_t|y^{t-1})}{\int_{x_t} p(y_t|x_t)p(x_t|y^{t-1}) \, dx_t} \]

We can then compute recursively the previous steps starting from the initialization:

\[ p(x_0|y^{-1}) = p(x_0). \]
The Bayesian solution

Solution: assuming that we have already computed

\[ p(x_{t-1}|y^{t-1}) \]

then

- **Prediction:**
  \[ p(x_t|y^{t-1}) = \int_{x_{t-1}} p(x_t|x_{t-1})p(x_{t-1}|y^{t-1}) \, dx_{t-1} \]

- **Correction:**
  \[ p(x_t|y^t) = \frac{p(y_t|x_t)p(x_t|y^{t-1})}{\int_{x_t} p(y_t|x_t)p(x_t|y^{t-1}) \, dx_t} \]

We can then compute recursively the previous steps starting from the initialization:

\[ p(x_0|y^{-1}) = p(x_0). \]
The Bayesian solution

Solution: assuming that we have already computed

\[ p(x_{t-1}|y^{t-1}) \]

then

- **Prediction:**
  \[
  p(x_t|y^{t-1}) = \int_{x_{t-1}} p(x_t|x_{t-1})p(x_{t-1}|y^{t-1}) \, dx_{t-1}
  \]

- **Correction:**
  \[
  p(x_t|y^t) = \frac{p(y_t|x_t)p(x_t|y^{t-1})}{\int_{x_t} p(y_t|x_t)p(x_t|y^{t-1}) \, dx_t}
  \]

We can then compute recursively the previous steps starting from the initialization:

\[ p(x_0|y^{-1}) = p(x_0). \]
The Bayesian solution

In the Gaussian case, solution:

\[ p(x_t|y^t) = N(x_t; \hat{x}_t, \sigma_t^2) \]

with

\[
\begin{align*}
\hat{x}_{t|t-1} &= \hat{x}_{t-1} + d \\
\sigma_{t|t-1}^2 &= \sigma_{t-1}^2 + \sigma_w^2 \\
\hat{x}_t &= \hat{x}_{t|t-1} + K_t (y_t - \hat{x}_{t|t-1}) \\
\sigma_t^2 &= (1 - K_t) \sigma_{t|t-1}^2 \\
K_t &= \frac{\sigma_{t|t-1}^2}{\sigma_{t|t-1}^2 + \sigma_v^2}
\end{align*}
\]
The Bayesian solution

In the Gaussian case, solution:

\[ p(x_t | y^t) = N(x_t; \hat{x}_t, \sigma_t^2) \]

with

\[
\begin{align*}
\hat{x}_{t|t-1} &= \hat{x}_{t-1} + d \\
\sigma_{t|t-1}^2 &= \sigma_{t-1}^2 + \sigma_w^2 \\
\hat{x}_t &= \hat{x}_{t|t-1} + K_t(y_t - \hat{x}_{t|t-1}) \\
\sigma_t^2 &= (1 - K_t)\sigma_{t|t-1}^2 \\
K_t &= \frac{\sigma_{t|t-1}^2}{\sigma_{t|t-1}^2 + \sigma_v^2}
\end{align*}
\]

Same recursive equations as the Engineer one!
The Kalman filter is named after Rudolf E. Kalman. Kalman first published his ideas on:


First important application, Apollo missions:
The Bayesian solution

In the Gaussian case, solution:

\[ p(x_t|y^t) = N(x_t; \hat{x}_t, \sigma_t^2) \]

with

\[
\begin{align*}
\hat{x}_{t|t-1} &= \hat{x}_{t-1} + d \\
\sigma_{t|t-1}^2 &= \sigma_{t-1}^2 + \sigma_w^2 \\
\hat{x}_t &= \hat{x}_{t|t-1} + K_t(y_t - \hat{x}_{t|t-1}) \\
\sigma_t^2 &= (1 - K_t)\sigma_{t|t-1}^2 \\
K_t &= \frac{\sigma_{t|t-1}^2}{\sigma_{t|t-1}^2 + \sigma_v^2}
\end{align*}
\]

Same recursive equations as the Engineer one!
Optimal estimate in the MMSE:

\[ \hat{x}_t = E[x_t|y^t] = \arg \min_\mu E_{X_t, Y_t}[(x_t - \mu)^2] \]
Bayesian Engineer

Optimal estimate in the MMSE:

\[ \hat{x}_t = E[x_t|y^t] \]
\[ = \arg \min_{\mu} E_{X^t, Y^t}[(x_t - \mu)^2] \]

From the posterior:

\[ p(x_t|y^t) = N(x_t; \hat{x}_t, \sigma^2_t) \]

my \approx 99\% \text{ credible interval}

\[ [\hat{x}_t - 3\sigma_t, \hat{x}_t + 3\sigma_t] \]
Bayesian Engineer

Optimal estimate in the MMSE:

\[ \hat{x}_t = E[x_t|y^t] \]
\[ = \arg \min_{\mu} E_{X^t, Y^t}[(x_t - \mu)^2] \]

From the posterior:

\[ p(x_t|y^t) = N(x_t; \hat{x}_t, \sigma_t^2) \]

my \approx 99\% \text{ credible interval}

\[ [\hat{x}_t - 3\sigma_t, \hat{x}_t + 3\sigma_t] \]

Engineer

Linear Minimum Variance estimator:

\[ \min_{\gamma_k} E_{X^t, Y^t}[(x_t - \hat{x}_t)^2] \]

with \( \hat{x}_t = \sum_{k=1}^t \gamma_k y_k + \gamma_0 \)
Optimal estimate in the MMSE:

\[ \hat{x}_t = E[x_t|y^t] \]
\[ = \arg \min_{\mu} E_{X^t, Y^t}[(x_t - \mu)^2] \]

From the posterior:

\[ p(x_t|y^t) = N(x_t; \hat{x}_t, \sigma^2_t) \]

my ≈ 99% credible interval

\[ [\hat{x}_t - 3\sigma_t, \hat{x}_t + 3\sigma_t] \]

Linear Minimum Variance estimator:

\[ \min_{\gamma_k} E_{X^t, Y^t}[(x_t - \hat{x}_t)^2] \]

with \( \hat{x}_t = \sum_{k=1}^{t} \gamma_k y_k + \gamma_0 \)

From \( \hat{x}_t, \sigma^2_t \) and Chebyshev inequality:

\[ P( |X - \mu| \leq \gamma \sigma ) \geq 1 - \frac{1}{\gamma^2} , \]

my ≈ 99% credible interval:

\[ [\hat{x}_t - 19\sigma_t, \hat{x}_t + 19\sigma_t] \]

no need of the Gaussian assumption.
Imprecise Probability guy

Model:

\[
\begin{align*}
    x_{k+1} &= x_k + d + w_k \\
    y_k &= x_k + v_k
\end{align*}
\]

Prior knowledge:

\[
    x_0 \leftarrow \hat{x}_0, \sigma^2_0, \quad w_k \leftarrow 0, \sigma^2_w, \quad p(v_k) = N(v_k; 0, \sigma^2_v)
\]

Goal: to estimate \( x_t \) given the measurements \( y_1, y_2, \ldots, y_t \) and to determine a \( \approx 99\% \) credible interval.
Idea: consider all probability distributions that are compatible with

\[ x_0 \leftarrow \hat{x}_0, \sigma^2_0, \quad w_k \leftarrow 0, \sigma^2_w, \quad p(v_k) = N(v_k; 0, \sigma^2_v) \]

and compute

\[ \mathcal{P} = \{ p(x_t|y^t) : \text{for each of such distributions} \} \]

Solution, assuming that

\[ \hat{x}_{1|0} = 0, \quad \sigma^2_{1|0} = \sigma^2_v = 1 \quad \rightarrow \quad \hat{x}_1 = \frac{y_1}{2}, \sigma^2_1 = \frac{1}{2}, \]
Imprecise Probability guy

Idea: consider all probability distributions that are compatible with
\[ x_0 \leftarrow \hat{x}_0, \sigma^2_0, \quad w_k \leftarrow 0, \sigma^2_w, \quad p(v_k) = N(v_k; 0, \sigma^2_v) \]
and compute
\[ \mathcal{P} = \{ p(x_t|y^t) : \text{for each of such distributions} \} \]

Solution, assuming that
\[ \hat{x}_{1|0} = 0, \quad \sigma^2_{1|0} = \sigma^2_v = 1 \quad \rightarrow \quad \hat{x}_1 = \frac{y_1}{2}, \sigma^2_1 = \frac{1}{2}, \]
Imprecise Probability guy

Idea: consider all probability distributions that are compatible with

\[
x_0 \leftarrow \hat{x}_0, \sigma^2_0, \quad w_k \leftarrow 0, \sigma^2_w, \quad p(v_k) = N(v_k; 0, \sigma^2_v)
\]

and compute

\[
\mathcal{P} = \{ p(x_t|y^t) : \text{for each of such distributions} \}
\]

Solution, assuming that

\[
\hat{x}_{1|0} = 0, \quad \sigma^2_{1|0} = \sigma^2_v = 1 \quad \rightarrow \quad \hat{x}_1 = \frac{y_1}{2}, \sigma^2_1 = \frac{1}{2},
\]
Imprecise information, only moments known

Consider a real-valued $X$ and assume that the only probabilistic information about $X$ is:

$$E[l\{X\}] = 1, \quad E[X] = \mu_1, \ldots, E[X^m] = \mu_m.$$ 

where $l\{X\}$ is the indicator function of the set $\mathcal{X}$ (the space of possibilities of $X$) and $\mu_i$ are the first $m$ non-central moments of $X$.

$m$ moments are not enough to uniquely specify the distribution of $X$, so we can consider the set of all Probability Density Functions (PDF) which are compatible with this information:

$$\mathcal{P} = \left\{ p(\cdot) \geq 0 : \begin{array}{l} \int l\{X\} p(x) \, dx = 1 \\ \int xp(x) \, dx = \mu_1 \\ \vdots \\ \int x^m p(x) \, dx = \mu_m \end{array} \right\}$$

The constraints $p(\cdot) > 0$ and $\int l\{X\} p(x) \, dx = 1$ ensure that $p(\cdot)$ is a well-defined PDF.
Imprecise information, only moments known

Consider a real-valued $X$ and assume that the only probabilistic information about $X$ is:

$$E[I\{X\}] = 1, \ E[X] = \mu_1, \ldots, E[X^m] = \mu_m.$$ 

where $I\{X\}$ is the indicator function of the set $\mathcal{X}$ (the space of possibilities of $X$) and $\mu_i$ are the first $m$ non-central moments of $X$.

$m$ moments are not enough to uniquely specify the distribution of $X$, so we can consider the set of all Probability Density Functions (PDF) which are compatible with this information:

$$\mathcal{P} = \left\{ p(\cdot) \geq 0: \begin{array}{l} \int l\{X\}p(x)dx = 1 \\ \int xp(x)dx = \mu_1 \\ \vdots \\ \int x^m p(x)dx = \mu_m \end{array} \right\}$$

The constraints $p(\cdot) > 0$ and $\int l\{X\}p(x)dx = 1$ ensure that $p(\cdot)$ is a well-defined PDF.
The optimization problem

\[ E[g] = \min_{p \in \mathcal{P}} \int g(x)p(x)dx, \quad \bar{E}[g] = \max_{p \in \mathcal{P}} \int g(x)p(x)dx \]

with

\[ \mathcal{P} = \left\{ p(\cdot) \geq 0 : \begin{array}{l} \int l_{\{X\}} p(x)dx = 1 \\ \int xp(x)dx = \mu_1 \\ \vdots \\ \int x^m p(x)dx = \mu_m \end{array} \right\} \]

is a linear programming problem (linear on the unknown \( p \)).

Hereafter, since \( E[g] = -\bar{E}[-g] \), we focus on maximization problems (upper bound).
The optimization problem

\[
\begin{align*}
\mathbb{E}[g] &= \min_{p \in \mathcal{P}} \int g(x)p(x)\,dx, \quad \overline{\mathbb{E}}[g] = \max_{p \in \mathcal{P}} \int g(x)p(x)\,dx
\end{align*}
\]

with

\[
\mathcal{P} = \left\{ p(\cdot) \geq 0 : \begin{array}{l}
\int \mathbb{1}_\mathcal{X} p(x)\,dx = 1 \\
\int xp(x)\,dx = \mu_1 \\
\vdots \\
\int x^m p(x)\,dx = \mu_m
\end{array} \right\}
\]

is a linear programming problem (linear on the unknown \( p \)).

Hereafter, since \( \mathbb{E}[g] = -\overline{\mathbb{E}}[-g] \), we focus on maximization problems (upper bound).
Primal

Theorem

**Primal**: from the fundamental theorem of linear program

\[
E[g] = \max_{p \in \mathcal{P}} \int g(x)p(x)dx = \max_{P \in \text{ext} \mathcal{P}} \int g(x)p(x)dx
\]

where \( \text{ext} \mathcal{P} = \left\{ \sum_{i=1}^{m+1} w_i \delta(x - m_i) \right\} \) with

\[
\sum_{i=1}^{m+1} w_i = 1 \\
\sum_{i=1}^{m+1} w_i m_i = \mu_1 \\
\vdots \\
\sum_{i=1}^{m+1} w_i m_i^m = \mu_m
\]

The extreme probabilities that give the lower and upper expectations of \( g \) are **mixtures of Dirac’s deltas with at most \( m + 1 \) components!**
Primal: from the fundamental theorem of linear program

\[ \overline{E}[g] = \max_{p \in P} \int g(x)p(x)dx = \max_{P \in \text{ext}P} \int g(x)p(x)dx \]

where \( \text{ext}P = \left\{ \sum_{i=1}^{m+1} w_i \delta(x - m_i) \right\} \) with

\[
\begin{align*}
\sum_{i=1}^{m+1} w_i &= 1 \\
\sum_{i=1}^{m+1} w_i m_i &= \mu_1 \\
&\quad \ldots \\
\sum_{i=1}^{m+1} w_i m_i^m &= \mu_m
\end{align*}
\]

The extreme probabilities that give the lower and upper expectations of \( g \) are mixtures of Dirac’s deltas with at most \( m + 1 \) components!
Example: $m = 0$ moment problem

Assume we only know that:

$$X \in \mathcal{X} = [a, b], \quad \rightarrow \quad E[l_{[a,b]}] = 1,$$

with $[a, b] \subset \mathbb{R}$ and consider the function

$$g = X$$

Since $m + 1 = 1$, the set of extreme points are (unimodal mixtures of) Dirac’s deltas on $[a, b]$.

Then, it follows that

$$E[X] = a, \quad \bar{E}[X] = b.$$

where the lower is obtained by $\delta(x - a)$ and the upper by $\delta(x - b)$
Example: $m = 0$ moment problem

Assume we only know that:

$$X \in \mathcal{X} = [a, b], \quad \rightarrow \quad E[l_{[a,b]}] = 1,$$

with $[a, b] \subset \mathbb{R}$ and consider the function

$$g = X$$

Since $m + 1 = 1$, the set of extreme points are (unimodal mixtures of) Dirac’s deltas on $[a, b]$.

Then, it follows that

$$E[X] = a, \quad \bar{E}[X] = b.$$

where the lower is obtained by $\delta(x - a)$ and the upper by $\delta(x - b)$.
Example: \( m = 0 \) moment problem

Assume we only know that:

\[ X \in \mathcal{X} = [a, b], \quad \rightarrow \quad E[I_{[a,b]}] = 1, \]

with \([a, b] \subset \mathbb{R}\) and consider the function

\[ g = X \]

Since \( m + 1 = 1 \), the set of extreme points are (unimodal mixtures of) Dirac’s deltas on \([a, b]\).

Then, it follows that

\[ E[X] = a, \quad \overline{E}[X] = b. \]

where the lower is obtained by \( \delta(x - a) \) and the upper by \( \delta(x - b) \)
Example: \( m = 2 \) moment problem

Assume that

\[ X \in \mathcal{X} = \mathbb{R}, \quad E[X] = \mu_1, \quad E[X^2] = \mu_1^2 + \sigma^2 \]

This means that:

\[ \text{ext} \mathcal{P} = \{ w_1 \delta(x - m_1) + w_2 \delta(x - m_2) + w_3 \delta(x - m_3) \} \]

\[ \text{s.t.} \]

\[ w_1 + w_2 + w_3 = 1 \]

\[ w_1 m_1 + w_2 m_2 + w_3 m_3 = \mu_1 \]

\[ w_1 m_1^2 + w_2 m_2^2 + w_3 m_3^2 = \mu_1^2 + \sigma^2 \]

the set of extreme points are at most trimodal \((m + 1 = 3)\) mixtures of Dirac’s deltas.
Example: $m = 2$ moment problem

Assume that

$$X \in \mathcal{X} = \mathbb{R}, \quad E[X] = \mu_1, \quad E[X^2] = \mu_1^2 + \sigma^2$$

This means that:

$$\text{ext}\mathcal{P} = \{w_1 \delta(x - m_1) + w_2 \delta(x - m_2) + w_3 \delta(x - m_3)\}$$

s.t.

$$w_1 + w_2 + w_3 = 1$$

$$w_1 m_1 + w_2 m_2 + w_3 m_3 = \mu_1$$

$$w_1 m_1^2 + w_2 m_2^2 + w_3 m_3^2 = \mu_1^2 + \sigma^2$$

the set of extreme points are at most trimodal ($m + 1 = 3$) mixtures of Dirac’s deltas.
Example: $m = 2$ moment problem

In this case, if

$$g = X$$

we have that:

$$E[X] = E[X] = E[X] = \mu_1.$$  

So, consider the function

$$g = I_{\{|X-\mu_1| \leq \gamma \sigma\}}$$

for some $\gamma > 0$.

Then, it can be shown that:

$$E[I_{\{|X-\mu_1| \leq \gamma \sigma\}}] = 1 - \frac{1}{\gamma^2}$$

Notice that the lower expectation of $I_{\{|X-\mu_1| \leq \gamma \sigma\}}$ corresponds to the worst-case (equality) in the Chebyshev inequality

$$P(|X - \mu_1| \leq \gamma \sigma) \geq 1 - \frac{1}{\gamma^2},$$

i.e., the probability of the set $|X - \mu_1| \leq \gamma \sigma$ is exactly equal to $1 - \frac{1}{\gamma^2}$.
Example: $m = 2$ moment problem

In this case, if

$$g = X$$

we have that:

$$E[X] = E[X] = E[X] = \mu_1.$$ 

So, consider the function

$$g = I_{\{|x - \mu_1| \leq \gamma \sigma\}}$$

for some $\gamma > 0$.

Then, it can be shown that:

$$E[I_{\{|x - \mu_1| \leq \gamma \sigma\}}] = 1 - \frac{1}{\gamma^2}$$

Notice that the lower expectation of $I_{\{|x - \mu_1| \leq \gamma \sigma\}}$ corresponds to the worst-case (equality) in the Chebyshev inequality

$$P(|X - \mu_1| \leq \gamma \sigma) \geq 1 - \frac{1}{\gamma^2},$$

i.e., the probability of the set $|X - \mu_1| \leq \gamma \sigma$ is exactly equal to $1 - \frac{1}{\gamma^2}$. 
Example: $m = 2$ moment problem

In this case, if

$$g = X$$

we have that:

$$E[X] = \overline{E}[X] = E[X] = \mu_1.$$

So, consider the function

$$g = I_{\{|X - \mu_1| \leq \gamma \sigma\}}$$

for some $\gamma > 0$.

Then, it can be shown that:

$$E[I_{\{|X - \mu_1| \leq \gamma \sigma\}}] = 1 - \frac{1}{\gamma^2}$$

Notice that the lower expectation of $I_{\{|X - \mu_1| \leq \gamma \sigma\}}$ corresponds to the worst-case (equality) in the Chebyshev inequality

$$P(|X - \mu_1| \leq \gamma \sigma) \geq 1 - \frac{1}{\gamma^2},$$

i.e., the probability of the set $|X - \mu_1| \leq \gamma \sigma$ is exactly equal to $1 - \frac{1}{\gamma^2}$.
Example: \( m = 2 \) moment problem

In this case, if

\[ g = X \]

we have that:

\[ \mathbb{E}[X] = \mathbb{E}[X] = \mathbb{E}[X] = \mu_1. \]

So, consider the function

\[ g = I\{|X - \mu_1| \leq \gamma \sigma\} \]

for some \( \gamma > 0 \).

Then, it can be shown that:

\[ \mathbb{E}[I\{|X - \mu_1| \leq \gamma \sigma\}] = 1 - \frac{1}{\gamma^2} \]

Notice that the lower expectation of \( I\{|X - \mu_1| \leq \gamma \sigma\} \) corresponds to the worst-case (equality) in the Chebyshev inequality

\[ P(|X - \mu_1| \leq \gamma \sigma) \geq 1 - \frac{1}{\gamma^2}, \]

i.e., the probability of the set \( |X - \mu_1| \leq \gamma \sigma \) is exactly equal to \( 1 - \frac{1}{\gamma^2} \).
Primal and dual

**Primal:** from the fundamental theorem of linear program

\[
\bar{E}[g] = \max_{p \in \mathcal{P}} \int g(x) p(x) dx = \max_{p \in \text{ext}\mathcal{P}} \int g(x) p(x) dx
\]

where \(\text{ext}\mathcal{P} = \{\sum_{i=1}^{m+1} w_i \delta(x - m_i)\}\) with

\[
\sum_{i=1}^{m+1} w_i = 1 \\
\sum_{i=1}^{m+1} w_i m_i = \mu_1 \\
\vdots \\
\sum_{i=1}^{m+1} w_i m_i^m = \mu_m
\]

The PDFs which give the lower and upper expectations of \(g\) are mixtures of Dirac’s deltas with at most \(m + 1\) components!

**Dual:** Given the unknown \(\mathbf{z} = [z_0, z_1, \ldots, z_m]\) with \(z_i \in \mathbb{R}\):

\[
\bar{E}[g] = \min \mathbf{z}^T \mu \\
\text{s.t.} \quad \mathbf{z}^T \mathbf{f}(x) \geq g(x), \quad \forall x \in \mathcal{X}.
\]

where \(\mathbf{\mu} = [1, \mu_1, \mu_2, \ldots, \mu_m]^T\) and \(\mathbf{f} = [1, x, x^2, \ldots, x^m]^T\)

Natural Extension!
Primal and dual

**Primal:** from the fundamental theorem of linear program

\[
\bar{E}[g] = \max_{p \in \mathcal{P}} \int g(x)p(x)dx = \max_{P \in \text{ext}\mathcal{P}} \int g(x)p(x)dx
\]

where \( \text{ext}\mathcal{P} = \{\sum_{i=1}^{m+1} w_i \delta(x - m_i)\} \) with

\[
\sum_{i=1}^{m+1} w_i = 1 \\
\sum_{i=1}^{m+1} w_i m_i = \mu_1 \\
\vdots \\
\sum_{i=1}^{m+1} w_i m_i^m = \mu_m
\]

The PDFs which give the lower and upper expectations of \( g \) are mixtures of Dirac’s deltas with at most \( m + 1 \) components!

**Dual:** Given the unknown \( \mathbf{z} = [z_0, z_1, \ldots, z_m] \) with \( z_i \in \mathbb{R} \):

\[
\bar{E}[g] = \min \mathbf{z}^T \mu \\
\text{s.t. } \mathbf{z}^T \mathbf{f}(x) \geq g(x), \forall x \in \mathcal{X}.
\]

where \( \mu = [1, \mu_1, \mu_2, \ldots, \mu_m]^T \) and \( \mathbf{f} = [1, x, x^2, \ldots, x^m]^T \)

Natural Extension!
How did the Imprecise guy solve the problem?

\[
\begin{cases}
    x_{k+1} = x_k + d + w_k, \\
    y_k = x_k + v_k,
\end{cases}
\]

\[E[X_0] = \mu_1 = \hat{x}_0, \quad \mu_2 = \hat{x}_0^2 + \sigma_0^2, \quad E[X_0] = \hat{x}_0, \quad E[X_0^2] = \hat{x}_0^2 + \sigma_0^2,\]

\[E[W_k] = \mu_1 = 0, \quad \mu_2 = \sigma_w^2, \quad E[X_{k+1}|x_k] = x_k + d, \quad E[X_{k+1}^2|x_k] = (x_k + d)^2 + \sigma_w^2,\]

\[E[h] = \int h(v_k)N(v_k; 0, \sigma_v^2)dv_k, \quad E[h|x_k] = \int h(y_k)N(y_k; x_k, \sigma_v^2)dy_k.\]

\[
E_{x_{k+1}}[g|x_k] = \min_{z}^2 z^T \mu
\]

s.t. \( z^T f(x_{k+1}) - g(x_{k+1}) \geq 0, \quad \forall x_{k+1} \in X_{k+1}, \)

where \( \mu = [1, x_k + d, (x_k + d)^2 + \sigma_w^2]^T \) and \( f = [1, x_{k+1}, x_{k+1}^2]^T \).
The rescue problem

How did the Imprecise guy solve the problem?

\[
\begin{align*}
x_{k+1} &= x_k + d + w_k, \\
y_k &= x_k + v_k,
\end{align*}
\]

\[
E[X_0] = \mu_1 = \hat{x}_0, \quad E[X_0^2] = \mu_2 = \hat{x}_0^2 + \sigma_0^2, \quad E[X_0] = \hat{x}_0, \quad E[X_0^2] = \hat{x}_0^2 + \sigma_0^2, \\
E[W_k] = \mu_1 = 0, \quad E[W_k^2] = \mu_2 = \sigma_w^2, \quad E[X_{k+1}|x_k] = x_k + d, \quad E[X_{k+1}^2|x_k] = (x_k + d)^2 + \sigma_w^2, \\
E[h] = \int h(v_k)N(v_k; 0, \sigma_v^2)dv_k, \quad E[h|x_k] = \int h(y_k)N(y_k; x_k, \sigma_v^2)dy_k.
\]

\[
\begin{align*}
E_{X_{k+1}|g|x_k} &= \min_z z^T \mu \\
s.t. \quad z^T f(x_{k+1}) - g(x_{k+1}) &\geq 0, \quad \forall x_{k+1} \in \mathcal{X}_{k+1}, \\
\end{align*}
\]

where \( \mu = [1, x_k + d, (x_k + d)^2 + \sigma_w^2]^T \) and \( f = [1, x_{k+1}, x_{k+1}^2]^T \).
The rescue problem

How did the Imprecise guy solve the problem?

\[
\begin{aligned}
x_{k+1} &= x_k + d + w_k, \\
y_k &= x_k + v_k,
\end{aligned}
\]

\[
\begin{aligned}
E[X_0] = \mu_1 = \hat{x}_0, & \quad E[X_0^2] = \mu_2 = \hat{x}_0^2 + \sigma_0^2, \\
E[W_k] = \mu_1 = 0, & \quad E[W_k^2] = \mu_2 = \sigma_w^2, \\
E[h] = \int h(v_k)\mathcal{N}(v_k; 0, \sigma_v^2)dv_k. & \quad E[h|x_k] = \int h(y_k)\mathcal{N}(y_k; x_k, \sigma_v^2)dy_k.
\end{aligned}
\]

\[
\begin{aligned}
E[X_{k+1}|x_k] = x_k + d, & \quad E[X_{k+1}^2|x_k] = (x_k + d)^2 + \sigma_w^2, \\
E[h|x_k] &= \min_z z^T \mu \\
\text{s.t.} \quad z^T f(x_{k+1}) - g(x_{k+1}) \geq 0, \quad \forall x_{k+1} \in \mathcal{X}_{k+1},
\end{aligned}
\]

where \( \mu = [1, x_k + d, (x_k + d)^2 + \sigma_w^2]^T \) and \( f = [1, x_{k+1}, x_{k+1}^2]^T \).
The rescue problem

How did the Imprecise guy solve the problem?

\[
\begin{align*}
    x_{k+1} &= x_k + d + w_k, \\
y_k &= x_k + v_k,
\end{align*}
\]

\[
E[X_0] = \mu_1 = \hat{x}_0, \quad E[X_0^2] = \mu_2 = \hat{x}_0^2 + \sigma_0^2, \quad E[X_0] = \hat{x}_0, \quad E[X_0^2] = \hat{x}_0^2 + \sigma_0^2, \\
E[W_k] = \mu_1 = 0, \quad E[W_k^2] = \mu_2 = \sigma_w^2, \quad E[X_{k+1}|x_k] = x_k + d, \quad E[X_{k+1}^2|x_k] = (x_k + d)^2 + \sigma_w^2, \\
E[h] &= \int h(v_k)N(v_k; 0, \sigma_v^2)dv_k, \quad E[h|x_k] = \int h(y_k)N(y_k; x_k, \sigma_v^2)dy_k.
\]

\[
\bar{E}_{X_{k+1}}[g|x_k] = \min_z z^T\mu \\
s.t. \quad z^Tf(x_{k+1}) - g(x_{k+1}) \geq 0, \quad \forall x_{k+1} \in X_{k+1},
\]

where \( \mu = [1, x_k + d, (x_k + d)^2 + \sigma_w^2]^T \) and \( f = [1, x_{k+1}, x_{k+1}^2]^T \)
The rescue problem

How did the Imprecise guy solve the problem?

\[
\begin{align*}
  x_{k+1} &= x_k + d + w_k, \\
  y_k &= x_k + v_k,
\end{align*}
\]

\[E[X_0] = \mu_1 = \hat{x}_0, \quad E[X_0^2] = \mu_2 = \hat{x}_0^2 + \sigma_0^2, \quad E[X_0] = \hat{x}_0, \quad E[X_0^2] = \hat{x}_0^2 + \sigma_0^2,\]

\[E[W_k] = \mu_1 = 0, \quad E[W_k^2] = \mu_2 = \sigma_w^2, \quad E[X_{k+1}|x_k] = x_k + d, \quad E[X_{k+1}^2|x_k] = (x_k + d)^2 + \sigma_w^2,\]

\[E[h] = \int h(v_k)N(v_k; 0, \sigma_v^2)dv_k. \quad E[h|x_k] = \int h(y_k)N(y_k; x_k, \sigma_v^2)dy_k.\]

\[E_{X_{k+1}}[g|x_k] = \min_z z^T \mu \]

s.t. \[z^T f(x_{k+1}) - g(x_{k+1}) \geq 0, \quad \forall x_{k+1} \in X_{k+1},\]

where \( \mu = [1, x_k + d, (x_k + d)^2 + \sigma_w^2]^T \) and \( f = [1, x_{k+1}, x_{k+1}^2]^T \)
HMM filtering through Coherent Lower Previsions

Input: 
\[ \overline{E}_X(0), \quad \overline{E}_{X_k} (\cdot | x_{k}), \quad E_{Y_k} [\cdot | x_{k}] \]

Goal: derive from them \( \overline{E}_X(t| y^t) \).

This can be done by regular extension on \( \overline{E}_{X_t, Y_t} \):
\[
\overline{E}_X(t| y^t) = \nu \quad \text{s.t.} \quad \inf_{\nu} \overline{E}_{X_t, Y_t} [l_{yt} \cdot (g - \nu)] \geq 0.
\]

The joint \( \overline{E}_{X_t, Y_t} \) is obtained by marginal extension:
\[
\overline{E}_{X_t, Y_t} [\cdot] = \overline{E}_X(0) [\overline{E}_{X_1} [\overline{E}_{Y_1} [\ldots \overline{E}_{X_t} [\overline{E}_{Y_t} [\cdot | X_t] | X_{t-1}] \ldots | X_1] | X_0]]
\]

We also assume that:
\( X^{k-2} \) and \( Y^{k-1} \) are irrelevant to \( X_k \) given \( X_{k-1} \);
\( X^{k-1} \) and \( Y^{k-1} \) are irrelevant to \( Y_k \) given \( X_k \), i.e.:
\[
HMM \rightarrow \begin{cases} 
\overline{E}_{X_k} [\cdot | x_{k-1}] = \overline{E}_{X_k} [\cdot | x^{k-1}, y^{k-1}] \\
\overline{E}_{Y_k} [\cdot | x_k] = \overline{E}_{Y_k} [\cdot | x^{k}, y^{k-1}] 
\end{cases}
\]
HMM filtering through Coherent Lower Previsions

Input:

\[
\mathcal{E}_X^0, \quad \mathcal{E}_{X_{k+1}}(\cdot|X_k), \quad \mathcal{E}_{Y_k}[\cdot|X_k]
\]

Goal: derive from them \( \mathcal{E}_{X_t}[g|y^t] \).

This can be done by regular extension on \( \mathcal{E}_{X_t,Y_t} \):

\[
\mathcal{E}_{X_t}[g|y^t] = \nu \quad \text{s.t.} \quad \inf_{\nu} \mathcal{E}_{X_t,Y_t}[l_{y^t}(g - \nu)] \geq 0.
\]

The joint \( \mathcal{E}_{X_t,Y_t} \) is obtained by marginal extension:

\[
\mathcal{E}_{X_t,Y_t}[\cdot] = \mathcal{E}_X^0[\mathcal{E}_X^1[\mathcal{E}_{Y_1}[\ldots \mathcal{E}_{X_t}[\mathcal{E}_{Y_t}[\cdot|X_t]|X_{t-1}] \ldots |X_1]|X_0]]
\]

We also assume that:

- \( X^{k-2} \) and \( Y^{k-1} \) are irrelevant to \( X_k \) given \( X_{k-1} \);
- \( X^{k-1} \) and \( Y^{k-1} \) are irrelevant to \( Y_k \) given \( X_k \), i.e.:

\[
\text{HMM} \rightarrow \begin{cases} \\
\mathcal{E}_{X_k}[\cdot|X_k] = \mathcal{E}_{X_k}[\cdot|x^{k-1}, y^{k-1}] \\
\mathcal{E}_{Y_k}[\cdot|X_k] = \mathcal{E}_{Y_k}[\cdot|x^{k}, y^{k-1}]
\end{cases}
\]
HMM filtering through Coherent Lower Previsions

Input:
\[ \overline{E}_{X_0}, \quad \overline{E}_{X_{k+1}}(\cdot|X_k), \quad E_{Y_k}[\cdot|X_k] \]

Goal: derive from them \( \overline{E}_{X_t}[g|y^t] \).

This can be done by regular extension on \( \overline{E}_{X_t,Y_t} \):

\[ \overline{E}_{X_t}[g|y^t] = \nu \quad s.t. \quad \inf_{\nu} \overline{E}_{X_t,Y_t}[I_{y^t} \cdot (g - \nu)] \geq 0. \]

The joint \( \overline{E}_{X_t,Y_t} \) is obtained by marginal extension:

\[ \overline{E}_{X_t,Y_t}[] = \overline{E}_{X_0}[\overline{E}_{X_1}[\overline{E}_{Y_1}[\ldots \overline{E}_{X_t}[\overline{E}_{Y_t}[\cdot|X_t]|X_{t-1}] \ldots |X_1]|X_0]] \]

We also assume that:
\( X^{k-2} \) and \( Y^{k-1} \) are irrelevant to \( X_k \) given \( X_{k-1} \);
\( X^{k-1} \) and \( Y^{k-1} \) are irrelevant to \( Y_k \) given \( X_k \), i.e.:

\[
HMM \rightarrow \left\{ \begin{array}{ll}
\overline{E}_{X_k}[\cdot | x_{k-1}] &= \overline{E}_{X_k}[\cdot | x^{k-1}, y^{k-1}] \\
\overline{E}_{Y_k}[\cdot | x_k] &= \overline{E}_{Y_k}[\cdot | x^{k}, y^{k-1}] 
\end{array} \right. 
\]
HMM filtering through Coherent Lower Previsions

Input:

\[ \overline{E}_X^0, \overline{E}_{X_{k+1}}(\cdot|x_k), E_{Y_k}[\cdot|x_k] \]

Goal: derive from them \( \overline{E}_x^t[g|y^t] \).

This can be done by regular extension on \( \overline{E}_{X^t,Y^t} \):

\[ \overline{E}_x^t[g|y^t] = \nu \quad s.t. \quad \inf_{\nu} \overline{E}_{X^t,Y^t}[l_y^t \cdot (g - \nu)] \geq 0. \]

The joint \( \overline{E}_{X^t,Y^t} \) is obtained by marginal extension:

\[ \overline{E}_{X^t,Y^t}[\cdot] = \overline{E}_X^0[\overline{E}_{X_1}[\overline{E}_{Y_1}[\ldots \overline{E}_{X_t}[\overline{E}_{Y_t}[\cdot|X_t]|X_{t-1}] \ldots |X_1]|X_0]] \]

We also assume that:

\( X^{k-2} \) and \( Y^{k-1} \) are irrelevant to \( X_k \) given \( X_{k-1} \);

\( X^{k-1} \) and \( Y^{k-1} \) are irrelevant to \( Y_k \) given \( X_k \), i.e.:

\[
\begin{align*}
HMM \rightarrow & \quad \frac{\overline{E}_{X_k}[\cdot|x_{k-1}]}{\overline{E}_{Y_k}[\cdot|x_k]} = \frac{\overline{E}_{X_k}[\cdot|x^{k-1}, y^{k-1}]}{\overline{E}_{Y_k}[\cdot|x^k, y^{k-1}]}
\end{align*}
\]
Input: \( \overline{E}_X^0, \overline{E}_{X_{k+1}}(.|x_k), \overline{E}_{Y_k}[.|X_k] \)

Goal: derive from them \( \overline{E}_X[|y^t] \).

This can be done by regular extension on \( \overline{E}_{X^t, Y^t} \):

\[
\overline{E}_X[|y^t] = \nu \text{ s.t. } \inf_{\nu} \overline{E}_{X^t, Y^t}[l_y \cdot (g - \nu)] \geq 0.
\]

The joint \( \overline{E}_{X^t, Y^t} \) is obtained by marginal extension:

\[
\overline{E}_{X^t, Y^t}[.] = \overline{E}_X^0[\overline{E}_X^1[\overline{E}_{Y_1}[\cdots \overline{E}_X^t[\overline{E}_{Y_t}[.|X_t]|X_{t-1}] \cdots |X_1]|X_0]]
\]

We also assume that:
- \( X^{k-2} \) and \( Y^{k-1} \) are irrelevant to \( X_k \) given \( X_{k-1} \);
- \( X^{k-1} \) and \( Y^{k-1} \) are irrelevant to \( Y_k \) given \( X_k \), i.e.:

\[
HMM \rightarrow \begin{cases} 
\overline{E}_X[.|x_{k-1}] = \overline{E}_X[.|x^{k-1}, y^{k-1}] \\
\overline{E}_{Y_k}[.|x_k] = \overline{E}_{Y_k}[.|x^k, y^{k-1}]
\end{cases}
\]
So we are able to compute $\overline{E}_{X_t[g|y^t]}$ for any $g$:

1. For $g = X_t$, we can compute
   - posterior lower mean $\underline{E}_{X_t[X_t|y^t]}$
   - posterior upper mean $\overline{E}_{X_t[X_t|y^t]}$

2. For $g = I_S$, we can find the smallest set (interval) $S$ such that
   \[
   E_{X_t[I_S|y^t]} = P_{X_t[S|y^t]} = 0.99
   \]
   which corresponds to the 99% interval computed by the Imprecise Probability guy in the ship problem.
So we are able to compute $\overline{E}_{X_t}[g|y^t]$ for any $g$:

1. For $g = X_t$, we can compute
   - posterior lower mean $E_{X_t}[X_t|y^t]$
   - posterior upper mean $\overline{E}_{X_t}[X_t|y^t]$

2. For $g = I_S$, we can find the smallest set (interval) $S$ such that
   \[
   E_{X_t}[I_S|y^t] = P_{X_t}[S|y^t] = 0.99
   \]
   which corresponds to the 99% interval computed by the Imprecise Probability guy in the ship problem.
Algorithm for the case $X_k \in \mathbb{R}$ (scalar case)

1. For each $k = 0, \ldots, t$ discretize $X_k$ by generating $n$ points equally spaced in $X' = [x_{\text{min}}, x_{\text{max}}]$.

2. Set a value of $\nu$ and set $g(\cdot, t, \nu) = g(\cdot) - \nu$.

3. Do the following backward propagation for $k = t, \ldots, 1$:
   For each discretized value $x_{k-1}^j$ of $X_{k-1}$, solve
   \[
   g(x_{k-1}^j, k - 1, \nu) = \min_{z} z^T q \\
   \text{s.t. } z^T f(x_k^j) - g(x_k^j, k, \nu) \mathcal{N}(y_k; x_k^i, r) \geq 0, \quad \forall x_k^i \in X',
   \]
   where $q = [1, \mu_1, \mu_2]^T$, $f(x) = [1, x, x^2]^T$, $\mu_1 = x_{k-1}^j$ and $\mu_2 = (x_{k-1}^j)^2 + \sigma^2_w$.

4. Solve:
   \[
   \text{res} = \min_{z} z^T q \\
   \text{s.t. } z^T f(x_0^i) - g(x_0^i, 0, \nu) \geq 0, \quad \forall x_0^i \in X',
   \]
   where $q = [1, \mu_1, \mu_2]^T$, $f(x) = [1, x, x^2]^T$, $\mu_1 = \hat{x}_0$ and $\mu_2 = \hat{x}_0^2 + \sigma_0^2$.

5. If $\text{res} = 0$ return $\nu$, otherwise change $\nu$ based on some root-finding algorithm (e.g., bisection) and back to 2).
Simulations: linear case

\[ x_{k+1} = x_k + w_k, \quad y_k = x_k + v_k, \]

Disturbance \( w_k \):

Simulation results:

**Figure:** Trajectory (brown), KF estimate (red), OPF estimate (green), GMBF lower and upper posterior means (blue).
Optimality of the IP solution

In the Bayesian setting,
\[ \hat{x} = E[X|y] = \arg \min_{\hat{x}} E_{X,Y}[(X - \hat{x})^2], \]

In IP, we compute lower and upper posterior expectations of \( X \), i.e., \( \underline{E}[X|y] \) and \( \overline{E}[X|y] \).
Are these values optimal in some sense?
Under maximality, \( \hat{x}_2 \) dominates \( \hat{x}_1 \) under the squared loss if
\[ E_{X,Y}[(X - \hat{x}_1)^2 - (X - \hat{x}_2)^2] > 0. \]

Theorem

A necessary and sufficient condition for \( \hat{x}_2 \) to be undominated under maximality is:
\[ \hat{x}_2 \in \mathcal{X}^* = \left\{ \int x p(x|y) dx : p \in \mathcal{P} \right\}, \]

where \( \mathcal{P} \) is the closed convex set of probabilities associated to \( E_{X,Y} \).
In the Bayesian setting,

\[ \hat{x} = E[X|y] = \arg\min_{\hat{x}} E_{X,Y}[(X - \hat{x})^2], \]

In IP, we compute lower and upper posterior expectations of \( X \), i.e., \( E[X|y] \) and \( \bar{E}[X|y] \).

Are these values optimal in some sense?

Under maximality, \( \hat{x}_2 \) dominates \( \hat{x}_1 \) under the squared loss if

\[ E_{X,Y}[(X - \hat{x}_1)^2 - (X - \hat{x}_2)^2] > 0. \]

**Theorem**

A necessary and sufficient condition for \( \hat{x}_2 \) to be undominated under maximality is:

\[ \hat{x}_2 \in \mathcal{X}^* = \left\{ \int x p(x|y)dx : p \in \mathcal{P} \right\}, \]

where \( \mathcal{P} \) is the closed convex set of probabilities associated to \( E_{X,Y} \).
Optimality of the IP solution

In the Bayesian setting,

\[ \hat{x} = E[X|y] = \arg \min_{\hat{x}} E_{X,Y}[(X - \hat{x})^2], \]

In IP, we compute lower and upper posterior expectations of \( X \), i.e., \( \underline{E}[X|y] \) and \( \overline{E}[X|y] \).

Are these values optimal in some sense?

Under maximality, \( \hat{x}_2 \) dominates \( \hat{x}_1 \) under the squared loss if

\[ \underline{E}_{X,Y}[(X - \hat{x}_1)^2 - (X - \hat{x}_2)^2] > 0. \]

Theorem

A necessary and sufficient condition for \( \hat{x}_2 \) to be undominated under maximality is:

\[ \hat{x}_2 \in X^* = \left\{ \int xp(x|y)dx : p \in \mathcal{P} \right\}, \]

where \( \mathcal{P} \) is the closed convex set of probabilities associated to \( E_{X,Y} \).
In the Bayesian setting,
\[ \hat{x} = E[X|y] = \underset{\hat{x}}{\arg\min} E_{X,Y}[(X - \hat{x})^2], \]

In IP, we compute lower and upper posterior expectations of \( X \), i.e., \( E[\cdot|y] \) and \( E[\cdot|y] \).
Are these values optimal in some sense?
Under maximality, \( \hat{x}_2 \) dominates \( \hat{x}_1 \) under the squared loss if
\[ E_{X,Y}[(X - \hat{x}_1)^2 - (X - \hat{x}_2)^2] > 0. \]

**Theorem**
A necessary and sufficient condition for \( \hat{x}_2 \) to be undominated under maximality is:
\[ \hat{x}_2 \in \mathcal{X}^* = \left\{ \int xp(x|y)dx : p \in \mathcal{P} \right\}, \]
where \( \mathcal{P} \) is the closed convex set of probabilities associated to \( E_{X,Y} \).
Nonlinear systems

\[
\begin{align*}
    x_{k+1} &= u_k(x_k) + w_k, \\
    y_k &= h_k(x_k) + v_k,
\end{align*}
\]

Example:

\[
\begin{align*}
    x_{k+1} &= 0.5x_k + 2 \frac{x_k}{1+x_k^2} + w_k, \\
    y_k &= \frac{x_k^2}{20} + v_k,
\end{align*}
\]

Even if we assume \( w_k, v_k \) are Normal distributed and we apply Bayesian method

\[
p(x_t|y^{t-1}) = \int_{x_{t-1}} p(x_t|x_{t-1})p(x_{t-1}|y^{t-1}) \, dx_{t-1}
\]

\[
p(x_t|y^t) = \frac{p(y_t|x_t)p(x_t|y^{t-1})}{\int_{x_t} p(y_t|x_t)p(x_t|y^{t-1}) \, dx_t}
\]
Nonlinear systems

\[
\begin{align*}
    x_{k+1} &= u_k(x_k) + w_k, \\
y_k &= h_k(x_k) + v_k,
\end{align*}
\]

Example:

\[
\begin{align*}
    x_{k+1} &= 0.5x_k + 2 \frac{x_k}{1+x_k^2} + w_k, \\
y_k &= \frac{x_k^2}{20} + v_k,
\end{align*}
\]

Even if we assume \( w_k, v_k \) are Normal distributed and we apply Bayesian method

\[
p(x_t|y^{t-1}) = \int_{x_{t-1}} p(x_t|x_{t-1}) p(x_{t-1}|y^{t-1}) \, dx_{t-1}
\]

\[
p(x_t|y^t) = \frac{p(y_t|x_t)p(x_t|y^{t-1})}{\int_{x_t} p(y_t|x_t)p(x_t|y^{t-1}) \, dx_t}
\]
Nonlinear systems

\[
\begin{align*}
    x_{k+1} &= u_k(x_k) + w_k, \\
    y_k &= h_k(x_k) + v_k,
\end{align*}
\]

Example:

\[
\begin{align*}
    x_{k+1} &= 0.5x_k + 2 \frac{x_k}{1+x_k^2} + w_k, \\
    y_k &= \frac{x_k^2}{20} + v_k,
\end{align*}
\]

Even if we assume \( w_k, v_k \) are Normal distributed and we apply Bayesian method

\[
p(x_t|y^{t-1}) = \int_{x_{t-1}} p(x_t|x_{t-1})p(x_{t-1}|y^{t-1}) \, dx_{t-1}
\]

\[
p(x_t|y^t) = \frac{p(y_t|x_t)p(x_t|y^{t-1})}{\int_{x_t} p(y_t|x_t)p(x_t|y^{t-1}) \, dx_t}
\]
Nonlinear systems

Kalman filter is not optimal anymore, the optimal filter is a nonlinear function of the observations.

A closed form solution does not exist, but there are approximations:

1. linearise the model about the current best estimate of the state (EKF);

2. if the distributions of $X_0, W_k, V_k$ are known we can solve the Bayesian estimation problem by sequential Monte Carlo methods (particle filter).

Issue: if the distributions of $X_0, W_k, V_k$ are unknown?
Kalman filter is not optimal anymore, the optimal filter is

- a nonlinear function of the observations.

A closed form solution does not exist, but there are approximations:

1. linearise the model about the current best estimate of the state (EKF);

2. if the distributions of $X_0, W_k, V_k$ are known we can solve the Bayesian estimation problem by sequential Monte Carlo methods (particle filter).

Issue: if the distributions of $X_0, W_k, V_k$ are unknown?
Kalman filter is not optimal anymore, the optimal filter is

- a nonlinear function of the observations.

A closed form solution does not exist, but there are approximations:

1. linearise the model about the current best estimate of the state (EKF);

2. if the distributions of $X_0, W_k, V_k$ are known we can solve the Bayesian estimation problem by sequential Monte Carlo methods (particle filter).

Issue: if the distributions of $X_0, W_k, V_k$ are unknown?
Nonlinear systems

\[
\begin{align*}
    x_{k+1} &= u_k(x_k) + w_k, \\
    y_k &= h_k(x_k) + v_k,
\end{align*}
\]

\[
E[X_0] = \mu_1 = \hat{x}_0, \quad E[X_0^2] = \mu_2 = \hat{x}_0^2 + \sigma_0^2,
\]

\[
E[W_k] = \mu_1 = 0, \quad E[W_k^2] = \mu_2 = \sigma_w^2,
\]

\[
E[h] = \int h(v_k) \mathcal{N}(v_k; 0, \sigma_v^2) dv_k.
\]

\[
E[X_0] = \hat{x}_0, \quad E[X_0^2] = \hat{x}_0^2 + p_0,
\]

\[
E[W_k] = \mu_1 = 0, \quad E[W_k^2] = \mu_2 = \sigma_w^2,
\]

\[
E[h] = \int h(y_k) \mathcal{N}(y_k; h_k(x_k), \sigma_v^2) dy_k.
\]

\[
E[X_k|g_{x_{k-1}}] = \min_z z^T \mu
\]

\[
s.t. \ z^T f(x_k) - g(x_k) \geq 0, \quad \forall x_k \in X_k,
\]

where \( \mu = [1, u_{k-1}(x_{k-1}), (u_{k-1}(x_{k-1}))^2 + \sigma_w^2]^T \) and \( f = [1, x_k, x_k^2]^T \).
Nonlinear systems

\[
\begin{align*}
    x_{k+1} &= u_k(x_k) + w_k, \\
    y_k &= h_k(x_k) + v_k,
\end{align*}
\]

- \(E[X_0] = \bar{x}_0, \quad E[X_0^2] = \bar{x}_0^2 + \sigma_0^2,\)
- \(E[W_k] = \mu_1 = 0, \quad E[W_k^2] = \mu_2 = \sigma_w^2,\)
- \(E[h] = \int h(v_k)N(v_k; 0, \sigma_v^2)dv_k.\)

- \(E[X_0] = \bar{x}_0, \quad E[X_0^2] = \bar{x}_0^2 + p_0,\)
- \(E[W_k] = \mu_1 = 0, \quad E[W_k^2] = \mu_2 = \sigma_w^2,\)
- \(E[h] = \int h(v_k)N(v_k; 0, \sigma_v^2)dv_k.\)

\[
\begin{align*}
    E[X_k | x_{k-1}] &= u_{k-1}(x_{k-1}), \\
    E[X_k^2 | x_{k-1}] &= \left(u_{k-1}(x_{k-1})\right)^2 + \sigma_w^2, \\
    E[h | x_k] &= \int h(y_k)N(y_k; h_k(x_k), \sigma_v^2)dy_k. \\
\end{align*}
\]

- \(E_{X_k}[g | x_{k-1}] = \min_z z^T \mu,\)
- \(s.t. \quad z^T f(x_k) - g(x_k) \geq 0, \quad \forall x_k \in X_k,\)

where \(\mu = [1, u_{k-1}(x_{k-1}), (u_{k-1}(x_{k-1}))^2 + \sigma_w^2]^T\) and \(f = [1, x_k, x_k^2]^T.\)
Nonlinear systems

\[
\begin{align*}
    x_{k+1} & = u_k(x_k) + w_k, \\
    y_k & = h_k(x_k) + v_k,
\end{align*}
\]

\[
\begin{align*}
    E[X_0] = \mu_1 = \hat{x}_0, & \quad E[X_0^2] = \mu_2 = \hat{x}_0^2 + \sigma_0^2, \\
    E[W_k] = \mu_1 = 0, & \quad E[W_k^2] = \mu_2 = \sigma_w^2, \\
    E[h] = \int h(v_k)N(v_k; 0, \sigma_v^2)dv_k. & \quad E[h|x_k] = \int h(y_k)N(y_k; h_k(x_k), \sigma_v^2)dy_k.
\end{align*}
\]

\[
\begin{align*}
    \bar{E}_{x_k}[g|x_{k-1}] & = \min_{z} z^T \mu \\
    \text{s.t. } z^T f(x_k) - g(x_k) & \geq 0, \quad \forall x_k \in \mathcal{X}_k,
\end{align*}
\]

where \( \mu = [1, u_{k-1}(x_{k-1}), (u_{k-1}(x_{k-1}))^2 + \sigma_w^2]^T \) and \( f = [1, x_k, x_k^2]^T \)
Simulations: nonlinear case

The following nonlinear model

\[
\begin{align*}
    x_{k+1} &= 0.5x_k + 2 \frac{x_k}{1 + x_k^2} + w_k, \\
    y_k &= \frac{x_k^2}{20} + v_k,
\end{align*}
\]

with $\sigma_w^2 = 1$ and $\sigma_v^2 = 0.01$.

**Figure:** Trajectory (brown), EKF estimate (red), OPF estimate (green), MBF lower and upper limits of the credible interval (blue) and Chebyshev inequality based interval (black).
Consider a multivariate variable $X = [X_1, \ldots, X_n]^T$ on $\mathcal{X} \subseteq \mathbb{R}^n$ and assume that the first $\alpha$ raw moments of $X$ are known, i.e.,
\[ E[X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_n^{\alpha_n}] = \mu_{\alpha_1 \alpha_2 \cdots \alpha_n}, \]
for any non-negative integer $\alpha_i$ such that $\alpha_1 + \alpha_2 + \cdots + \alpha_n \leq \alpha$.

For instance, in case $X = [X_1, X_2]^T$, if we have 6 constraints $E[1] = 1$, $E[X_1] = \mu_{10}$, $E[X_2] = \mu_{01}$, $E[X_1 X_2] = \mu_{11}$, $E[X_1^2] = \mu_{20}$, $E[X_2^2] = \mu_{02}$, this is equivalent to assume the knowledge of support, mean and covariance matrix of $X$.

All previous results continue to hold!

In theory there are not problems!
Consider a multivariate variable $X = [X_1, \ldots, X_n]^T$ on $\mathcal{X} \subseteq \mathbb{R}^n$ and assume that the first $\alpha$ raw moments of $X$ are known, i.e., $E[X_1^{\alpha_1}X_2^{\alpha_2} \cdots X_n^{\alpha_n}] = \mu_{\alpha_1\alpha_2\cdots\alpha_n}$, for any non-negative integer $\alpha_j$ such that $\alpha_1 + \alpha_2 + \cdots + \alpha_n \leq \alpha$.

For instance, in case $X = [X_1, X_2]^T$, if we have 6 constraints $E[1] = 1$, $E[X_1] = \mu_{10}$, $E[X_2] = \mu_{01}$, $E[X_1X_2] = \mu_{11}$, $E[X_1^2] = \mu_{20}$, $E[X_2^2] = \mu_{02}$, this is equivalent to assume the knowledge of support, mean and covariance matrix of $X$.

All previous results continue to hold!

In theory there are not problems!
Consider a multivariate variable $X = [X_1, \ldots, X_n]^T$ on $\mathcal{X} \subseteq \mathbb{R}^n$ and assume that the first $\alpha$ raw moments of $X$ are known, i.e.,

$$E[X_1^{\alpha_1}X_2^{\alpha_2} \cdots X_n^{\alpha_n}] = \mu_{\alpha_1\alpha_2\cdots\alpha_n},$$

for any non-negative integer $\alpha_i$ such that $\alpha_1 + \alpha_2 + \cdots + \alpha_n \leq \alpha$.

For instance, in case $X = [X_1, X_2]^T$, if we have 6 constraints $E[1] = 1$, $E[X_1] = \mu_{10}$, $E[X_2] = \mu_{01}$, $E[X_1X_2] = \mu_{11}$, $E[X_1^2] = \mu_{20}$, $E[X_2^2] = \mu_{02}$, this is equivalent to assume the knowledge of support, mean and covariance matrix of $X$.

All previous results continue to hold!

In theory there are not problems!
Multivariate extension in practice

\[ \bar{E}[g] = \inf_z z^T \mu, \quad s.t. \quad z^T f(x) \geq g(x), \quad \forall x \in \mathcal{X}. \]

Observe that when \( \mathcal{X} \) is infinite, the dual is a semi-infinite linear program, since the number of constraints is infinite.

The dual can also be rewritten in the equivalent minimax formulation (natural extension):

\[ \bar{E}[g] = \inf_{z} \sup_{x \in \mathcal{X}} g(x) - z^T (f(x) - \mu). \]

Other possibilities?
\[ \overline{E}[g] = \inf_{z} z^T \mu, \quad s.t. \quad z^T f(x) \geq g(x), \quad \forall x \in \mathcal{X}. \]

Observe that when \( \mathcal{X} \) is infinite, the dual is a semi-infinite linear program, since the number of constraints is infinite.

The dual can also be rewritten in the equivalent minimax formulation (natural extension):

\[ \overline{E}[g] = \inf_{z} \sup_{x \in \mathcal{X}} g(x) - z^T (f(x) - \mu). \]

Other possibilities?
Multivariate extension in practice

\[ \overline{E}[g] = \inf_z z^T \mu, \quad s.t. \quad z^T f(x) \geq g(x), \quad \forall x \in \mathcal{X}. \]

Observe that when \( \mathcal{X} \) is infinite, the dual is a semi-infinite linear program, since the number of constraints is infinite.

The dual can also be rewritten in the equivalent minimax formulation (natural extension):

\[ \overline{E}[g] = \inf_z \sup_{x \in \mathcal{X}} g(x) - z^T (f(x) - \mu). \]

Other possibilities?
Conclusions

HMM and Imprecise Probability:

- IP can make the difference in practice!
- it is able to model the available information without the need of extra assumptions.
- The Moment Based filter includes Set-membership estimation as particular case.
- The Moment Based filter can be extended to the case:

  \[ E[l_{\{X\}}] = 1, \ E[f_1(X)] = \mu_1, \ldots , \ E[f_m(X)] = \mu_m. \]

- We can also derive a Quantiles based filter.
Conclusions

HMM and Imprecise Probability:

- IP can make the difference in practice!
- it is able to model the available information without the need of extra assumptions.
- The Moment Based filter includes Set-membership estimation as particular case.
- The Moment Based filter can be extended to the case:

\[
E[I_{\{\chi\}}] = 1, \ E[f_1(X)] = \mu_1, \ldots, \ E[f_m(X)] = \mu_m.
\]
- We can also derive a Quantiles based filter.
Conclusions

HMM and Imprecise Probability:

- IP can make the difference in practice!
- it is able to model the available information without the need of extra assumptions.
- The Moment Based filter includes Set-membership estimation as particular case.
- The Moment Based filter can be extended to the case:

\[ E[l_{\{X\}}] = 1, \ E[f_1(X)] = \mu_1, \ldots, E[f_m(X)] = \mu_m. \]

- We can also derive a Quantiles based filter.
HMM and Imprecise Probability:

- IP can make the difference in practice!
- it is able to model the available information without the need of extra assumptions.
- The Moment Based filter includes Set-membership estimation as particular case.
- The Moment Based filter can be extended to the case:

\[
E[I_{\{x\}}] = 1, \ E[f_1(X)] = \mu_1, \cdots, \ E[f_m(X)] = \mu_m.
\]

- We can also derive a Quantiles based filter.
Conclusions

HMM and Imprecise Probability:

- IP can make the difference in practice!
- it is able to model the available information without the need of extra assumptions.
- The Moment Based filter includes Set-membership estimation as particular case.
- The Moment Based filter can be extended to the case:

\[
E[l_{\{X\}}] = 1, \ E[f_1(X)] = \mu_1, \cdots, E[f_m(X)] = \mu_m.
\]

- We can also derive a Quantiles based filter.
Conclusions

Issues and Future work

- Natural extension in theory and in practice
- How can we efficiently solve the problem in the multivariate case?
  - this is important for the practical success of IP solutions in HMM estimation problems.
$X_k$ is a discrete variable:


$X_k \in \mathbb{R}^n$ is a continuous (vector of) variable(s):


THANK YOU FOR THE ATTENTION
Bayesian filters

In other cases, we may only know that

\[ |w_t| \leq c_1, \quad |v_t| \leq c_2, \]

i.e., we know that the disturbances are bounded because of some physical constraint:

- maximum shift;
- maximum speed;
- maximum acceleration etc.

How can we solve the dynamic estimation problem in this case?

- Robust filters!
In case,

\[ |w_t| \leq c_1, \quad |v_t| \leq c_2, \]

we can apply robust filters:

- \( H_\infty, H_2 \);
- set-membership estimation.

\( H_\infty, H_2 \) restrict to linear estimators and derive the minimax solution of the filtering problem based on the knowledge of \( C_1, c_2 \).

Set-membership estimation uses the above information and set estimation to compute \( \mathcal{X}_t \) such that:

\[ x_t \in \mathcal{X}_t \]

to compute a (closed-convex) set that includes with guarantee the state \( x_t \).