Conditional Monte Carlo Density Estimation

Anton Braun, Huiyu Li and John Stachurski

September 2010

Conditional Monte Carlo (CMC) density estimation:

a way to <u>compute</u> densities

What we do:

1. Define and analyze CMC density estimation in a general setting
2. Provide general asymptotic theory suited to fin/econ/emetrics

But first, some background:

Problem Statement

- We have a given model
- The model defines a density ψ of some r.v. Y
- Want to be able to evaluate $\psi(y)$ for any $y \in \mathbb{X}$
- No analytical solution for ψ available

How to compute an approximation ψ_n to ψ?

Example 1: Discrete Time Markov Models

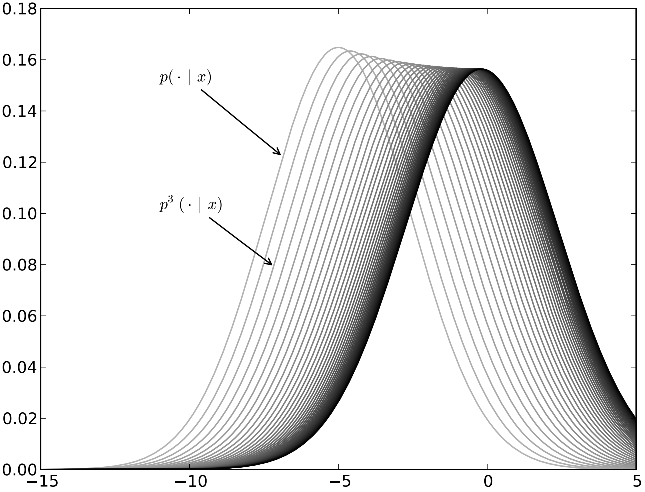
Consider AR(1) model

$$X_{t+1} = \alpha X_t + Z_{t+1} \quad \text{where} \quad \{Z_t\} \overset{\text{IID}}{\sim} N(0,1)$$

Let $p^t(\cdot \,|\, x)$ be the density of X_t given $X_0 = x$

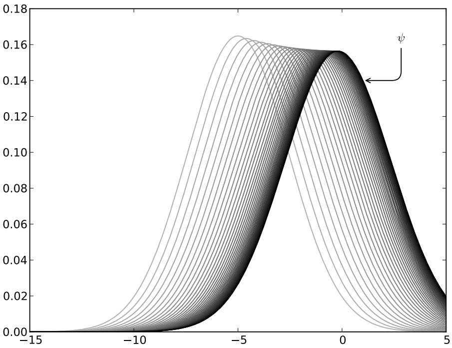
Can be computed analytically:

$$\begin{aligned} p(\cdot \,|\, x) := p^1(\cdot \,|\, x) &= N(\alpha x, 1) \\ p^2(\cdot \,|\, x) &= N(\alpha^2 x, \alpha^2 + 1) \\ p^3(\cdot \,|\, x) &= N(\alpha^3 x, \alpha^4 + \alpha^2 + 1) \\ p^4(\cdot \,|\, x) &= \cdots \end{aligned}$$


0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
−15
−10
−5
0
5
$p(\cdot \mid x)$
$p^3(\cdot \mid x)$

The stationary density is the limit of this sequence

$$\psi := \lim_{t \to \infty} p^t(\cdot \,|\, x) = N\left(0, \frac{1}{1 - \alpha^2}\right)$$



Note that everything here is tractable because of

- linearity
- normality

Without both, $p^t(\cdot \mid x)$ and ψ typically intractable

Let's now consider this case:

Consider general discrete time first order Markov model

$$p(\cdot \,|\, x) = \text{ conditional density of } X_{t+1} \text{ given } X_t = x$$

Given p, the t-step transition density defined by

$$p^t(y \,|\, x) = \int p(y \,|\, z) p^{t-1}(z \,|\, x) dz \quad \text{and} \quad p^1 = p$$

Stationary density ψ defined by

$$\psi(y) = \int p(y \,|\, x) \psi(x) dx, \quad \forall \, y$$

- Typically, ψ and $p^t(\cdot \,|\, x)$ are not tractible
- How to compute approximations to these densities?

Example 2: SDEs

Consider model

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

Let $p(\cdot \mid x, t)$ be the density of X_t given $X_0 = x$

- Sometimes p not tractable
- Sometimes stationary density not tractable

How to compute them?

Example 3: GARCH

Consider a GARCH(1,1) model

$$R_t = \sigma_t Z_t \quad \text{with} \quad Z_t \overset{\text{IID}}{\sim} N(0,1) \tag{1}$$

Process for (σ_t) given by

$$\sigma_{t+1}^2 = \alpha_0 + \beta \sigma_t^2 + \alpha_1 R_t^2$$

Assume:

- All parameters positive
- $\alpha_1 + \beta < 1$

Then R_t has unique stationary density—how to compute it?

Why Compute Densities Anyway?

There are many reasons—here are two

Example 1: To compare

- density implied by model vs
- density constructed from data

Eg: Ait-Sahalia (Rev. Fin. Stud, 1996)

- Let $dr_t = \mu(r_t)dt + \sigma(r_t)dW_t$ be model of interest rates
- Let ψ_m be stationary density of model
- Let ψ_d be est. of same constructed from interest rate data
- Reject model if $\|\psi_m - \psi_d\|$ is large

Example 2: Max likelihood

Let $dr_t = \mu(r_t; \theta)dt + \sigma(r_t; \theta)dW_t$ be model of interest rates

Let $p_\theta(\cdot \mid x, t) =$ the transition density, $\psi_\theta =$ stationary distribution

Let $r_1, \ldots, r_T$ be observations at time interval Δ

Log likelihood given by

$$\ell(\theta) = \ln \psi_\theta(r_1) + \sum_{t=1}^{T-1} \ln p_\theta(r_{t+1} \mid r_t, \Delta)$$

Numerical computation requires lots of computer power:

- Different densities for different parameter values
- Worse in multiple dimensions

How People Approximate Densities

- Discretize the model
- Linearization plus analytical solutions
- Numerical approximation (basis functions, etc.)
- Monte Carlo

Monte Carlo has some advantages:

- Robust to curse of dimensionality
- Good convergence properties
- Simple to program

Direct Monte Carlo Approach

Recall our problem

- Our model defines a density ψ of some r.v. Y
- No analytical solution for ψ available

How to compute approximation ψ_n to ψ?

Direct Monte Carlo method:

1. Generate draws $(Y_1, \ldots, Y_n)$ from ψ (assume possible)
2. Use sample $(Y_1, \ldots, Y_n)$ to construct $\psi_n \cong \psi$

But how should we implement last step???

Intermezzo: Density Estimation Is Hard

Suppose that

- ψ is a density
- $F =$ corresponding cdf ($F' = \psi$)
- neither ψ nor F known
- But have draws $Y_1, \ldots, Y_n$ from ψ (equiv., from F)

In this setting,

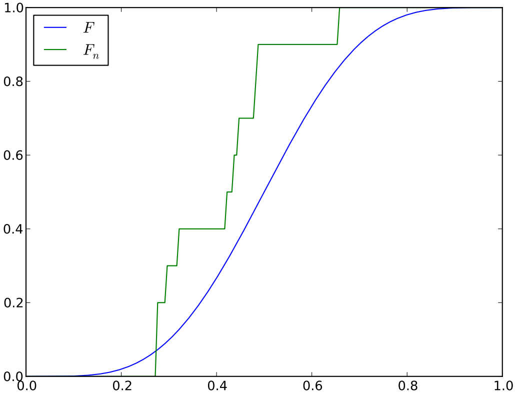
- estimating cdf F from $Y_1, \ldots, Y_n$ is easy, but
- estimating density ψ from $Y_1, \ldots, Y_n$ is hard

Let me explain:

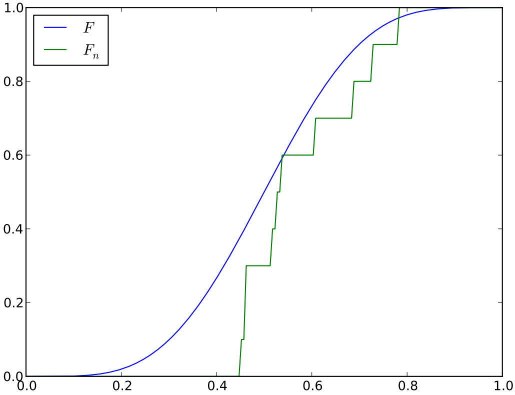
As estimator of F, can take empirical cdf (ecdf) F_n:

$$
\begin{aligned}
F_n(x) &:= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{Y_i \leq x\} \\
&= \text{fraction of the sample less than } x
\end{aligned}
$$

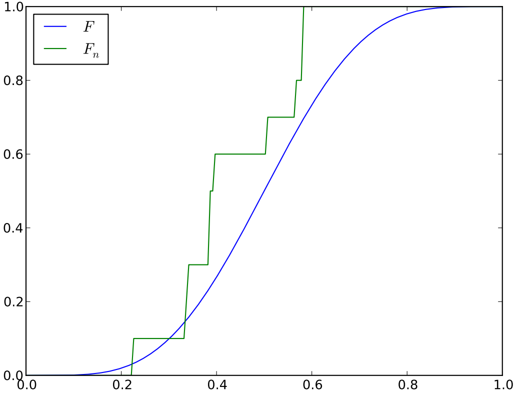
F_n is a random cdf, with jump at each observation:



F_n is a random cdf, with jump at each observation:



F_n is a random cdf, with jump at each observation:



As is well known, F_n converges to F as $n \to \infty$

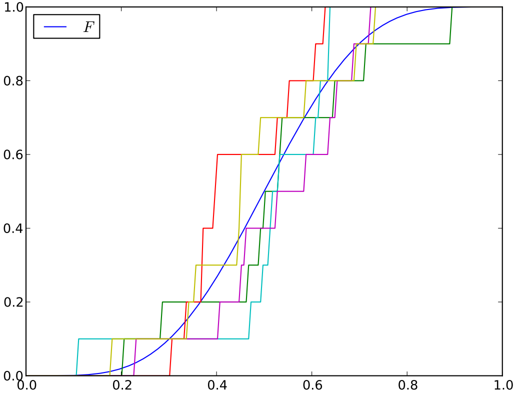


Figure: $n = 10$

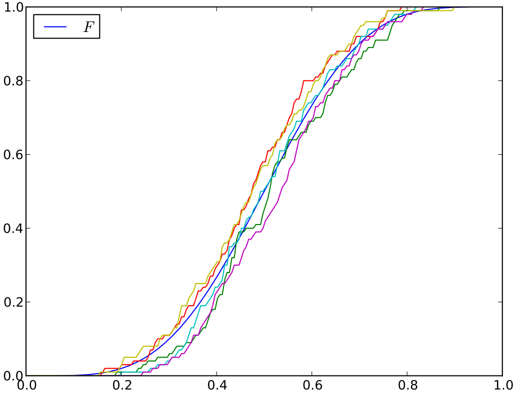


Figure: $n = 100$

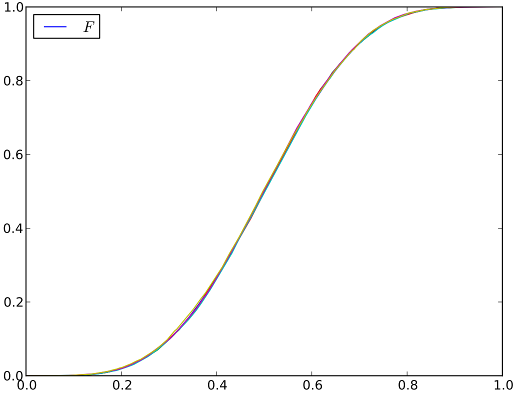
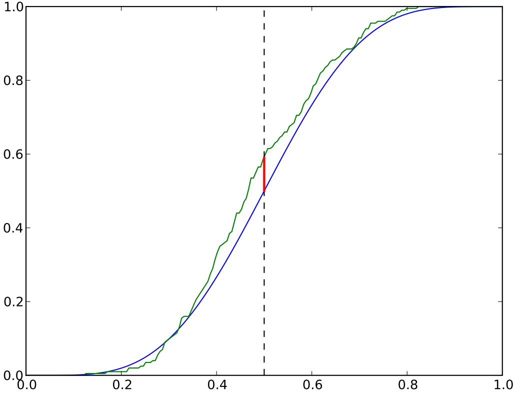


Figure: $n = 1000$

By SLLN, for any x, we have $|F_n(x) - F(x)| \to 0$ with prob one



How about global deviation?

What we really care about is whether $F_n \to F$ as functions

Define

$$\|F_n - F\|_\infty := \sup_x |F_n(x) - F(x)|$$

Does this also converge to zero?

Glivenko-Cantelli:

$$\|F_n - F\|_\infty \to 0 \text{ with prob one}$$

Donsker:

$$\|F_n - F\|_\infty = O_P(n^{-1/2})$$

To paraphrase: F_n is globally $\sqrt{n}$-consistent for F

Density Estimation is Ill-Posed

Consider an inverse problem (functional equation)

$$Tf = g$$

Here

- f and g are functions, T is an operator
- T and g given, goal is to recover f

Assume we can only approximate g with g_n

Ill-posed means:

- g_n is close to g
- but the solution f_n in $Tf_n = g_n$ is not close to f

Density estimation: We know that

$$\int_{-\infty}^{x} \psi(u) du = F(x)$$

We have a great estimator F_n for F

Should be able to estimate ψ by $\psi_n := F_n'$

But F_n is a step function

$\therefore \quad F_n' = 0$ (off a set of measure zero)

In this sense, F_n tells us nothing about the density

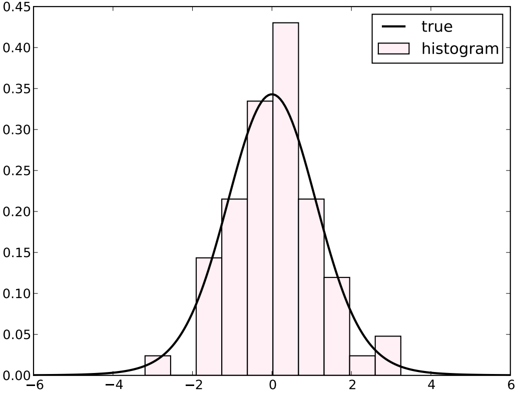
The numerical problem is ill-posed

Moral: Estimating F is easy, estimating ψ is hard

How else can we estimate ψ from the sample $Y_1, \ldots, Y_n$?

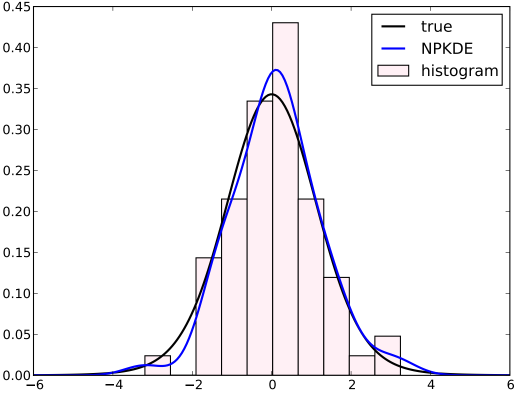
- Histograms?
- Nonparametric kernel density estimation?

Histograms:



true
histogram
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−6
−4
−2
0
2
4
6

Nonparametric kernel density estimation (NPKDE):



Problems with NPKDEs

Theoretical problems:

- Rate of convergence is slower than $O_P(n^{-1/2})$
- And sensitive to dimension of state space

Practical problems with standard implementations:

- Support
- Bias at mode, convex and concave regions

In fact we have no completely general, globally $\sqrt{n}$-consistent density estimator

But with one wee assumption we can get there...

CMC Density Estimation

Same problem: want to compute density ψ of random variable Y

Now suppose $\exists$ second random variable X correlated with Y

- density of Y given X is $q(\cdot \mid X)$
- distribution of X is denoted ϕ

Gives us another way to estimate $\psi(y)$:

- Sample X's from ϕ and then use $q(y \mid X)$

$$Y \sim \psi \xleftarrow{\quad q(\cdot \,|\, x) \quad} X \sim \phi$$

Key eq. for the theory: ψ and ϕ linked by

$$\psi(y) = \int q(y \,|\, x) \phi(dx)$$

Example 1: Transition Densities

Recall general discrete time first order Markov model

$$p(\cdot \,|\, x) = \text{ conditional density of } X_{t+1} \text{ given } X_t = x$$

Suppose we want to compute density $p^t(\cdot \,|\, x)$ of X_t

For conditioning variable, can use X_{t-1}

By definition, we have

$$p^t(y \,|\, x) = \int p(y \,|\, z) p^{t-1}(z \,|\, x) dz$$

This is a version of

$$\psi(y) = \int q(y \,|\, x) \phi(dx)$$

Example 2: GARCH

$R_t = \sigma_t Z_t$ with $Z_t \overset{\text{IID}}{\sim} N(0,1)$

Conditional density of R_t given $\sigma_t = x$ is

$$q(r \,|\, x) = N(0, x^2) = \frac{1}{\sqrt{2\pi x}} \exp\left\{-\frac{y^2}{2x^2}\right\}$$

If

- $\phi :=$ stationary distribution of σ_t
- $\psi :=$ stationary distribution of R_t

Then

$$\psi(r) = \int q(r \,|\, x) \phi(dx)$$

To estimate ψ, can use the CMC density estimator:

$$\text{draw } X_i \overset{\text{IID}}{\sim} \phi \text{ and set } \psi_n(y) := \frac{1}{n} \sum_{i=1}^{n} q(y \,|\, X_i)$$

The key here is that

$$\mathbb{E}\, q(y \,|\, X_i) = \psi(y)$$

$$\left[\text{because } \psi(y) = \int q(y \,|\, x) \phi(dx) \text{ always holds} \right]$$

$$\therefore \quad \psi_n(y) = \text{ sample mean of r.v.s with common mean } \psi(y)$$

Follows that CMC density estimator is:

1. Unbiased

$$\mathbb{E}\psi_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, q(y \,|\, X_i) = \frac{1}{n}\sum_{i=1}^{n} \psi(y) = \psi(y)$$

2. Consistent

$$\psi_n(y) = \frac{1}{n}\sum_{i=1}^{n} q(y \,|\, X_i) \to \mathbb{E}\, q(y \,|\, X_i) = \psi(y)$$

3. $\sqrt{n}$-consistent:

$$\sqrt{n}\{\psi_n(y) - \psi(y)\} = \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n} q(y \,|\, X_i) - \mathbb{E}q(y, \,|\, X_i)\right\}$$

$$\therefore \quad \sqrt{n}\{\psi_n(y) - \psi(y)\} \xrightarrow{\mathscr{D}} N(0, \sigma^2)$$

$$\therefore \quad \psi_n(y) - \psi(y) = O_P(n^{-1/2})$$

Variants of CMC density estimator discovered/rediscovered in many specific settings

- Posterior densities (Gelfand and Smith, 1990; Chib 1995)
- Simulated max likelihood for SDEs (Pedersen, 1995; etc.)
- Look-ahead estimators (Glynn and Henderson, 2001)

Could be used in many more:

- Calculating equilibrium asset/wealth/firm distributions
- Replacing kernel density estimators

What is lacking is a general theory

Wish list for a general theory

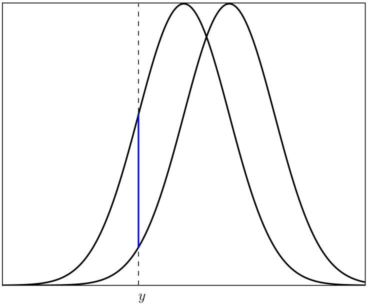
- General X and Y (multivariate, etc.)
- Global convergence
- Sampled X_i's may be correlated, not IID
- Draws from ϕ may not be exact

Finally, want to

- Tailor assumptions to economic/econometric settings
- Accommodate randomness in parameter estimates

Step 1: Global Analysis

So far our arguments are local, at fixed y



For global analysis, can use Hilbert space LLN and CLT

- Let $\{H_i\} = \text{IID}$ sequence of random functions in L_2
- Let $\mathcal{E}H$ be the expectation (def. omitted)

Hilbert space LLN:

$$\left\| \frac{1}{n} \sum_{i=1}^{n} H_i - \mathcal{E}H \right\| \to 0 \quad \text{in } L_2 \text{ with prob one}$$

Hilbert space CLT:

$$n^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^{n} H_i - \mathcal{E}H \right\} \xrightarrow{\mathscr{D}} \text{centered Gaussian on } L_2$$

Can be applied to the CMC density estimator

Key idea: Regard $q(\cdot \mid X_i)$ as a random function in L_2

Recall that $\mathbb{E}q(y \mid X_i) = \psi(y)$ for all y

From this, can prove that $\mathcal{E}q(\cdot \mid X) = \psi$

By Hilbert space LLN:

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} q(\cdot \mid X_i) \to \mathcal{E} q(\cdot \mid X_i) = \psi$$

By Hilbert space CLT:

$$n^{1/2}(\psi_n - \psi) \xrightarrow{\mathscr{D}} N(0, C)$$

Implies global $\sqrt{n}$-consistency: $\|\psi_n - \psi\| = O_P(n^{-1/2})$

Looking back at our wish list for a general theory

- General X and Y (multivariate, etc.) ✓
- Global convergence ✓
- Sampled X_i's may be correlated, not IID **?**
- Draws from ϕ may not be exact **?**

Why do we need the last two?

What sort of situations should we accommodate?

Example 1: Gibbs Sampling, MCMC

Wish to compute a given density ψ (e.g., posterior)

Cannot evaluate ψ, or even simulate draws from ψ directly

Instead, construct transition density p such that

1. can simulate Markov process $(X_t)_{t \geq 0}$ with transition density p
2. p is suitably ergodic (i.e., $X_t \stackrel{\mathscr{D}}{\to}$ stationary density)
3. ψ is the stationary density

Hence, if t large, then X_t an approx. draw from ψ

Can use this approximate sampling to reconstruct ψ

The CMC approach to estimating ψ:

Since ψ stationary, it satisfies

$$\psi(y) = \int p(y \,|\, x)\psi(x)dx$$

Hence, generate process $X_1, \ldots, X_n$ from p and use

$$\psi_n(y) = \frac{1}{n}\sum_{t=1}^{n} p(y \,|\, X_t)$$

Intuitively, should work because X_t almost $\sim \psi$ for t large

But previous global theory does not work, because not IID

Example 2: GARCH

Recall the GARCH(1,1) model

$$R_t = \sigma_t Z_t$$

with

$$\sigma_{t+1}^2 = \alpha_0 + \beta\sigma_t^2 + \alpha_1(\sigma_t Z_t)^2 \tag{2}$$

Conditioning variable is σ_t, with stationary distribution ϕ

Can't sample from ϕ directly, but can generate $\sigma_1, \ldots, \sigma_n$ via (2)

Here $\sigma_1, \ldots, \sigma_n$ is Markov, with ϕ as stationary distribution

Want to accommodate this kind of scenario

Prev. assumption:

- $\{X_t\} \overset{\text{IID}}{\sim} \phi$

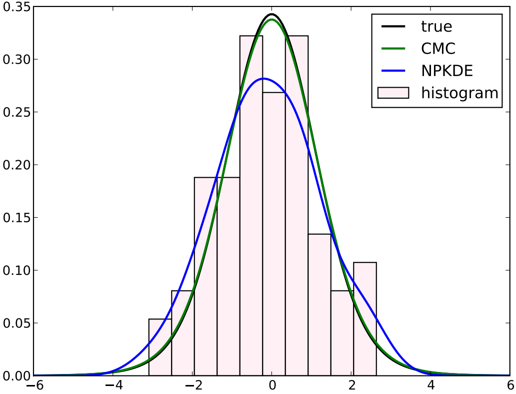
New assumption:

- $\{X_t\} \sim V$-UE Markov with stationary distribution ϕ

Many economic models are V-UE Markov

- various stationary time series models
- stochastic optimal growth models, etc.

Main technical result: Global, $\sqrt{n}$-consistency continues to hold in this Markov setting



Example 1: Stationary ARCH(1) model, sample $X_1, \ldots, X_n$:

Conclusions

CMC density estimation is a neat idea

Could be used much more than it is

Our contributions:

- Provide a general theory
- Global convergence results, in general Markov setting

To do:

- Treat uncertainty in parameters
- More applications?