Perturbative Corrections for

Expectation Propagation

Manfred Opper, Computer Science

joint work with:

Ulrich Paquet (U Cambridge)

Ole Winther (TU Denmark)

Outline

- EP in a nutshell
- EP fixed points and relation to exact inference
- Correction to partition function for models with pairwise couplings
- Perturbation expansion for correction
- Illustration on simple models (GP classification & Ising)
- Outlook

Expectation Propagation (EP) in a nutshell

$x = (x_1, \ldots, x_N)$

Assume factorisation of intractable density

$$p(x) = \frac{1}{Z} f_0(x) \prod_n f_n(x)$$

by a tractable density (in exponential family)

$$q(x) \propto f_0(x) \prod_n g_n(x)$$

Parameters of the g_n's are optimised iteratively by:

- Define *tilted distribution* $q_j(x) \propto f_j(x) \frac{q(x)}{g_j(x)}$
- Optimise q s.t. $q \approx q_j$ by moment matching.

Fixed point equations

Assume exponential family $g_n(x) \propto e^{\Lambda_n^\top \phi(x)} \ \rightarrow$

$$q(x) = \frac{1}{Z_0} f_0(x) e^{\Lambda^\top \phi(x)}$$

with $\Lambda = \sum_n \Lambda_n$. Thus

$$q_j(x) = \frac{1}{Z_j} f_j(x) \ e^{(\Lambda - \Lambda_j)^\top \phi(x)} f_0(x)$$

Moment matching conditions

$$\langle \phi(x) \rangle_q = \langle \phi(x) \rangle_{q_j} \quad \text{for } j = 1, \ldots, N \ .$$

The partition function

Express intractable terms

$$f_n(x) = \frac{Z_n}{Z_0} \frac{q_n(x)}{q(x)} \exp\left(\Lambda_n^T \phi(x)\right) \quad .$$

and insert into the partition function

$$\begin{aligned} Z &= \int dx \; f_0(x) \prod_n f_n(x) \\ &= \underbrace{\prod_n \left[\frac{Z_n}{Z_0}\right]}_{Z_{EP}} \times \underbrace{\int dx \; f_0(x) \prod_n (1 + \varepsilon_n(x)) \; \exp\left(\Lambda^T \phi(x)\right)}_{R} \quad . \end{aligned}$$

where $\varepsilon_n(x) = \frac{q_n(x) - q(x)}{q(x)}$.

EP is optimal to linear order

Expand log partition function into multivariate Taylor expansion with respect to ε_n, i.e.

$$\log Z = \log Z_{EC} + C_1 + C_2 + \ldots$$

C_1 contains corrections linear in ε_n etc.

$$\begin{aligned} C_1 &= \sum_n \frac{\int dx\ f_0(x)\ \varepsilon_n(x) \exp(\Lambda^T \phi(x))}{\int dx\ f_0(x) \exp(\Lambda^T \phi(x))} \\ &= \sum_n \int dx\ q(x) \varepsilon_n(x) = \sum_n \int dx\ q(x) \frac{q_n(x) - q(x)}{q(x)} = 0 \end{aligned}$$

by normalization of q_n and q.

Express joint density via q & q_n

$$Z = Z_{EP} \times R$$

$$p(x) = \frac{q(x) \prod_n \left(\frac{q_n(x)}{q(x)}\right)}{R}$$

where

$$R = \int dx \; q(x) \prod_n \left(\frac{q_n(x)}{q(x)}\right) \; .$$

Corrections for models with pairwise couplings

Models defined by

- the 'Prior'

$$f_0 = \exp\left[\frac{1}{2}x^\top J x\right]$$

- 'likelihood' terms $f_n = f_n(x_n)$ depending on single variable

- 'likelihood approximations'

$$g_n(x) \propto \exp\left[\gamma_n x_n - \frac{1}{2}\lambda_n x_n^2\right]$$

- Multivariate Gaussian approximation

$$q(x) \propto \exp\left[-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right]$$

Correction to marginal likelihood (partition function)

$$R = \int dx\; q(x) \prod_n \frac{q_n^{(n)}(x_n)}{q^{(n)}(x_n)}$$

where $q^{(n)}(x_n)$ and $q_n^{(n)}(x_n)$ are the marginals of q and q_n

We set

$$\prod_n \frac{q_n^{(n)}(x_n)}{q^{(n)}(x_n)} \doteq \frac{p_1(x)}{p_0(x)}$$

- EP accurate if p_1 close to p_0. Measure closeness by cumulants!
- p_0 Gaussian (only 1. & 2. cumulants)
- p_1 and p_0 agree on 1. & 2. cumulants!

Characteristic function & cumulants

- Characteristic function

$$\chi(k) = \int dx\, e^{ik^\top x} q(x) = \langle e^{ik^\top x} \rangle$$
$$q(x) = \int \frac{dk}{(2\pi)^N}\, e^{-ik^\top x} \chi(k)$$

- Cumulants c_l

$$\ln \chi(k) = \sum_l (i)^l \frac{c_l}{l!}\, k^l$$

- Express densities by cumulants

$$p_0(x) = \int \frac{dk}{(2\pi)^N}\, e^{-ik^\top x} \chi_0(k) \qquad \text{with } \chi_0(k) = e^{ik^\top \mu} \prod_n e^{-\frac{1}{2}\Sigma_{nn} k_n^2}$$
$$p_1(x) = \int \frac{dk}{(2\pi)^N}\, e^{-ik^\top x} \chi_0(k)\, e^{\sum_n g_n(k_n)}$$

g_n contains only the higher order cumulants c_{ln} for $l \geq 3$

Express the ratio by cumulants

Use shift of variables $k_j \to \eta_j = k_j + i\frac{(x_j - \mu_j)}{\Sigma_{jj}}$ we get

$$\frac{p_1(x)}{p_0(x)} = \int \prod_n \left(d\eta_n \sqrt{\frac{\Sigma_{nn}}{2\pi}} \right) \; e^{-\sum_n \frac{\Sigma_{nn}\eta_n^2}{2}} \; \exp\left[\sum_n g_n \left(\eta_n - i\frac{(x_n - \mu_n)}{\Sigma_{nn}} \right) \right]$$

Performing the average

Introduce *complex* Gaussian random variable $z_n \doteq \eta_n - i\frac{x_n - m_n}{\Sigma_{nn}}$

$$R = \int dx\; q(x)\, \frac{p_1(x)}{p_0(x)} = \left\langle \exp\left[\sum_n g_n\left(z_n\right)\right]\right\rangle_z$$

z has the covariance

$$\begin{aligned} \langle z_i z_j \rangle_z &= -\frac{\Sigma_{ij}}{\Sigma_{ii}\Sigma_{jj}} \qquad i \neq j \\ \langle z_i^2 \rangle_z &= 0 \end{aligned}$$

The last equation has some nice consequences for the surviving terms of corrections !

Perturbation expansion to Free energy

Assuming that the g_n are small, introduce formal expansion parameter λ, set $\lambda = 1$ at the end.

$$\ln R = \ln \left\langle \exp \left[\lambda \sum_n g_n(z_n) \right] \right\rangle_z =$$

$$\lambda \sum_n \langle g_n \rangle + \frac{\lambda^2}{2} \left\{ \left\langle \left(\sum_n g_n \right)^2 \right\rangle_z - \left(\sum_n \langle g_n \rangle_z \right)^2 \right\} \pm \ldots$$

$$= \frac{\lambda^2}{2} \sum_{m \neq n} \langle g_m g_n \rangle_z \pm \ldots$$

Single marginal terms vanish!

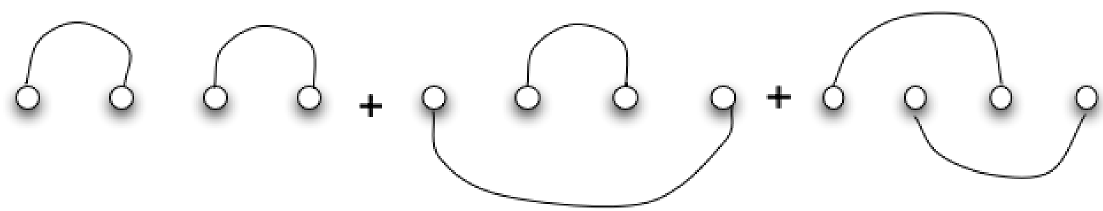
Gaussian averages & Feynman graphs

Let $\langle y_i \rangle = 0$. Then

$$\langle y_1 \cdots y_{2k} \rangle = \sum_{\text{pairings}} \langle y_{i_1} y_{j_1} \rangle \cdots \langle y_{i_k} y_{j_k} \rangle$$

Example:

$$\langle y_1 \cdot y_2 \cdot y_3 \cdot y_4 \rangle =$$



For

$$g_n(k) = \sum_{l \geq 3} i^l \frac{c_{ln}}{l!}\, k^l$$

we get

$$\begin{aligned}
\langle g_n(z_n) g_m(z_m) \rangle_z &= \sum_{l,s \geq 3} i^{l+s} \frac{c_{ln} c_{sm}}{l! s!} \langle z_n^l z_m^s \rangle \\
&= \sum_{l \geq 3} i^{2l} l! \frac{c_{ln} c_{lm}}{(l!)^2} \langle z_n z_m \rangle^l \\
&= \sum_{l \geq 3} \frac{c_{ln} c_{lm}}{l!} \left(\frac{\Sigma_{nm}}{\Sigma_{nn} \Sigma_{mm}} \right)^l
\end{aligned}$$

In practice, truncate after $l = 4$ or $l = 5$.

Conjecture: EP is fairly accurate if:

- the cumulants c_{ln} are small. This holds for GP classification, when posterior variance small compared to the mean.

- if posterior covariances Σ_{ij} small for $i \neq j$.

Gaussian process classification

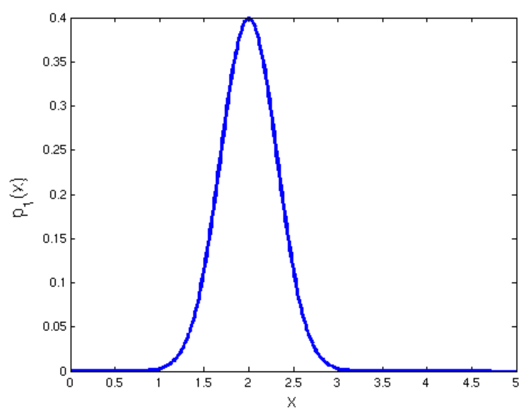
Prior

$$f_0(x) \propto \exp[-\frac{1}{2}x^\top K^{-1}x]$$

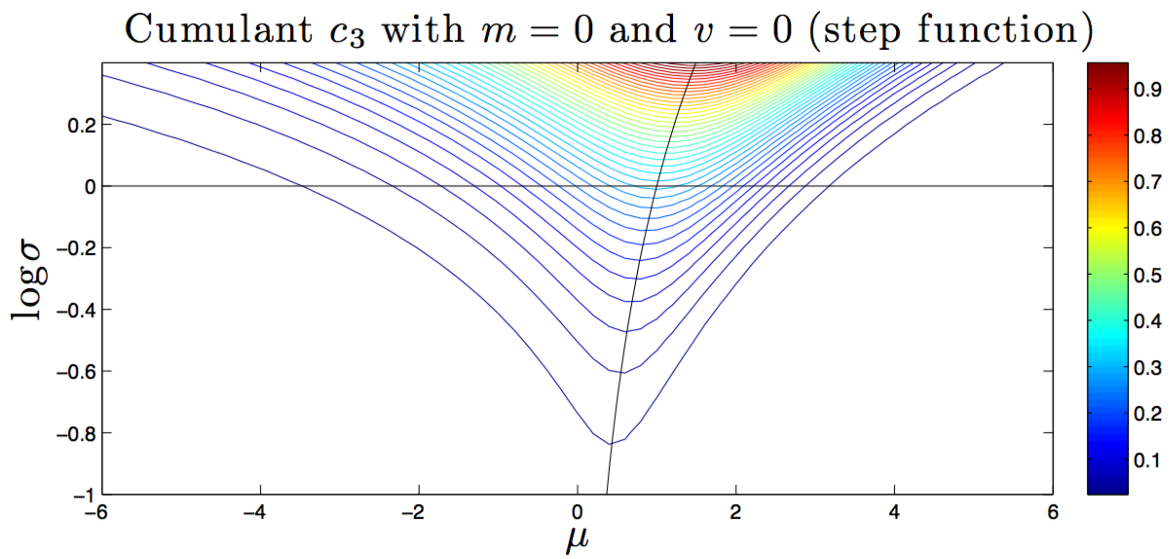
Likelihood

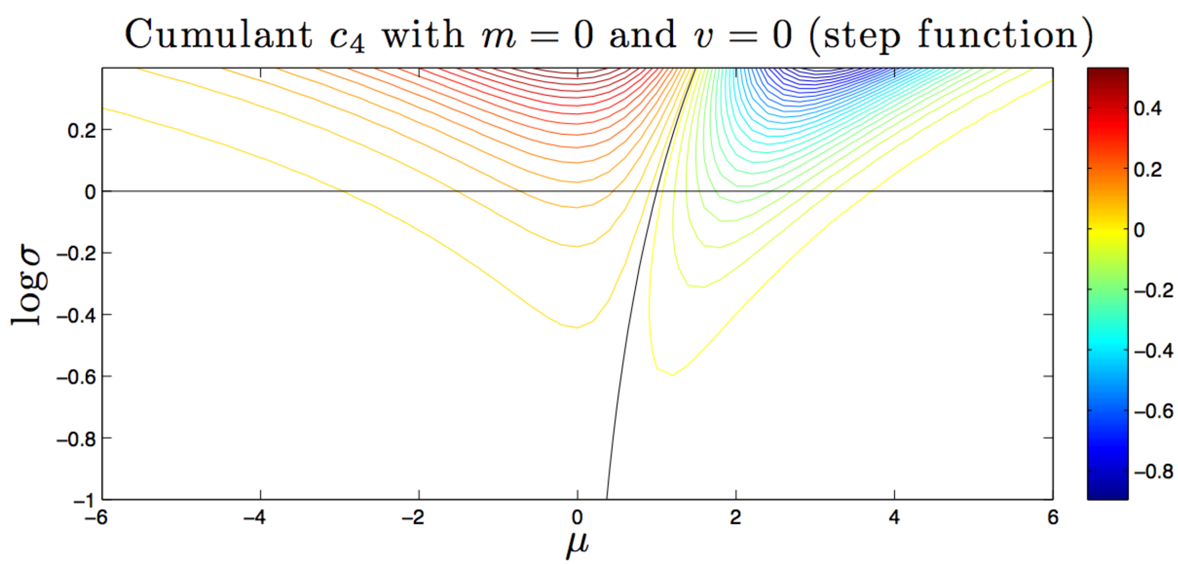
$$f_n(x) = \theta(x_n \cdot \text{label}_n) \qquad \text{(unit step function)}$$

The marginal $q^{(n)}(x)$ might look like this

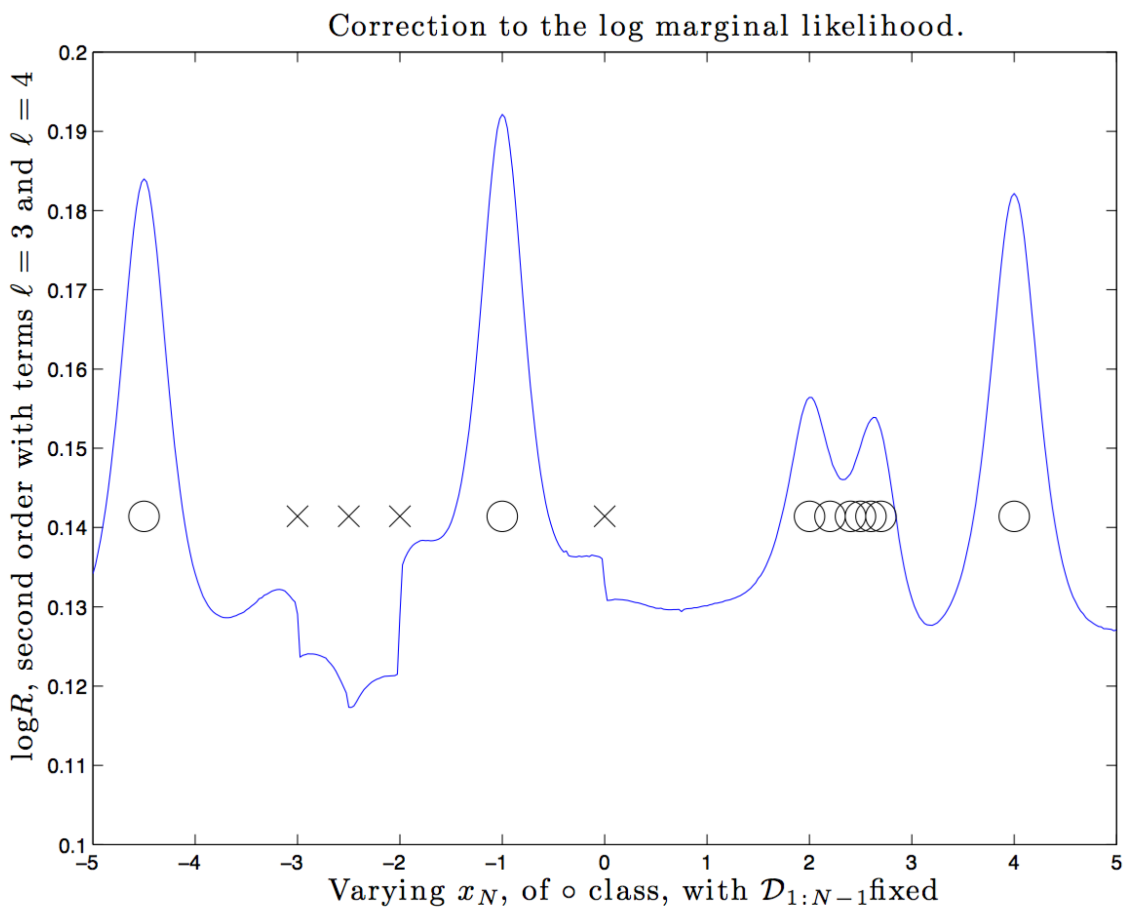


The cumulants

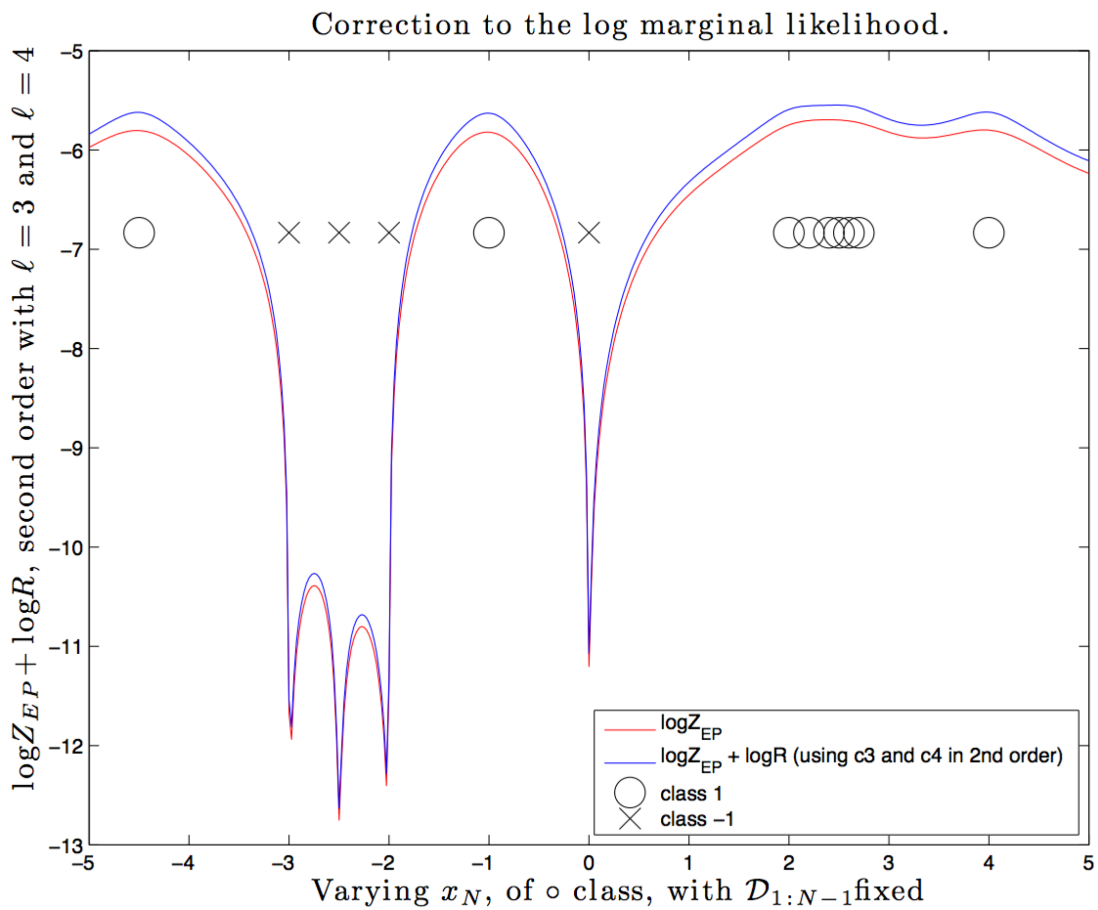




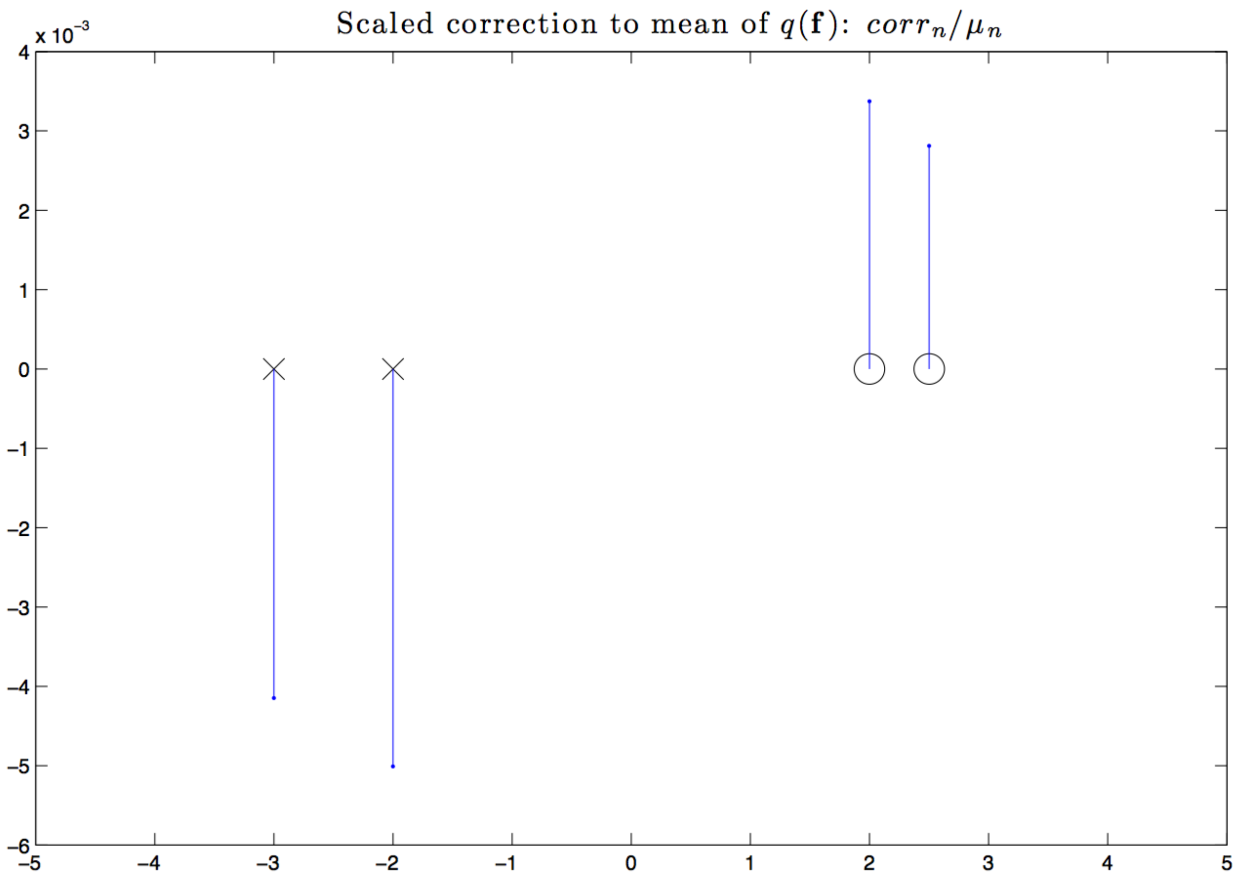
Correction to log partition function



log partition function + correction



Correcting the posterior mean



A toy Ising case

Ising variables $x_{1/2} = \pm 1$ with

$$p(x) = \frac{1}{Z}\, e^{J x_1 x_2} \qquad f_n(x) = \delta(x-1) + \delta(x+1)$$

with

$$\ln Z = \ln 4 + \ln \cosh(J) = \ln 4 + \frac{J^2}{2} - \frac{J^4}{12} \pm \ldots$$

One can show that

$$\ln Z_{EC} = \ln 4 \; + \frac{J^2}{2} - \frac{J^4}{4} \pm \ldots$$

- EC gives correct J^2 coefficient but J^4 comes out wrong.

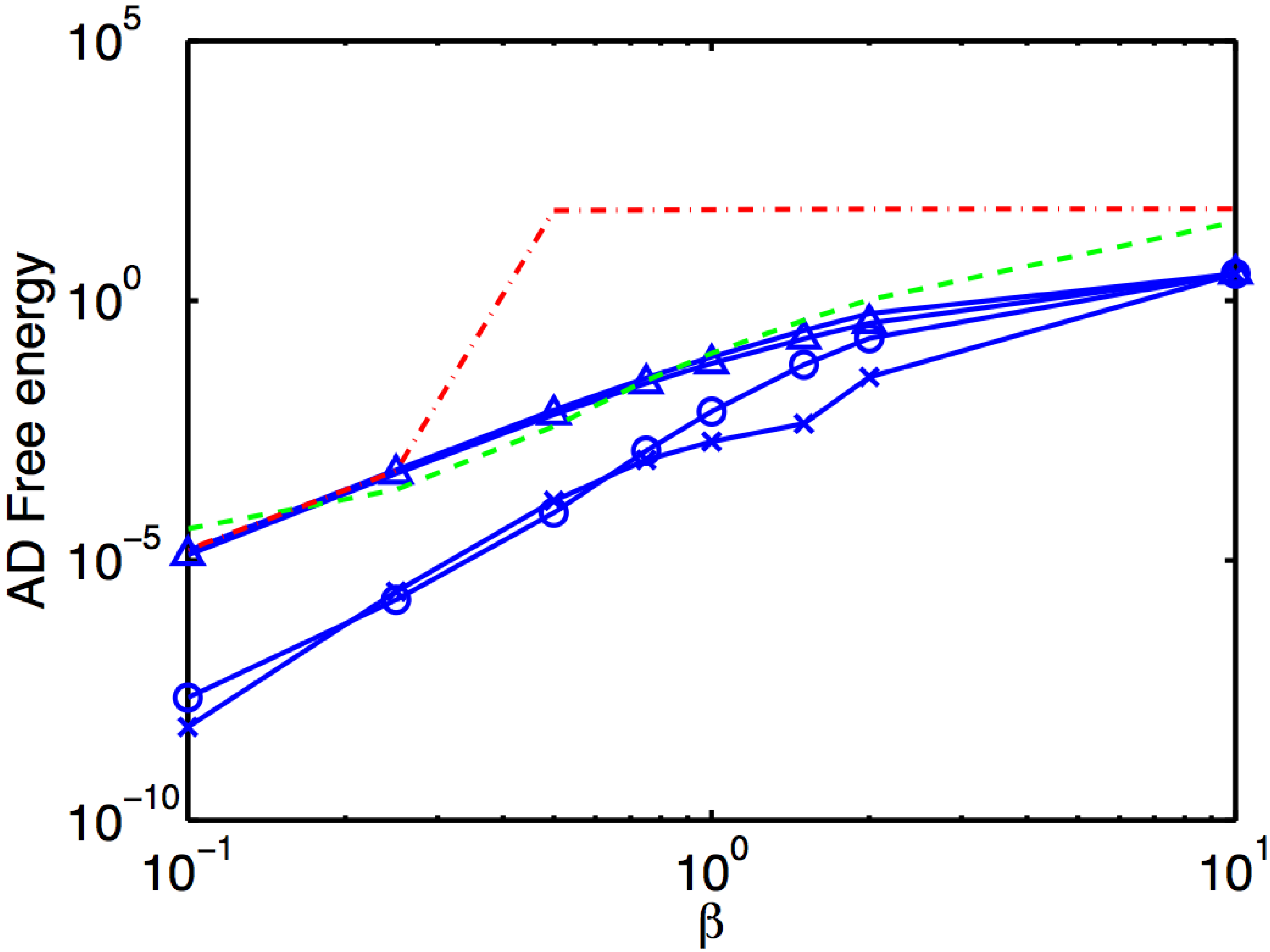
- Adding correction from c_4 makes J^4 exact.

Random Ising networks

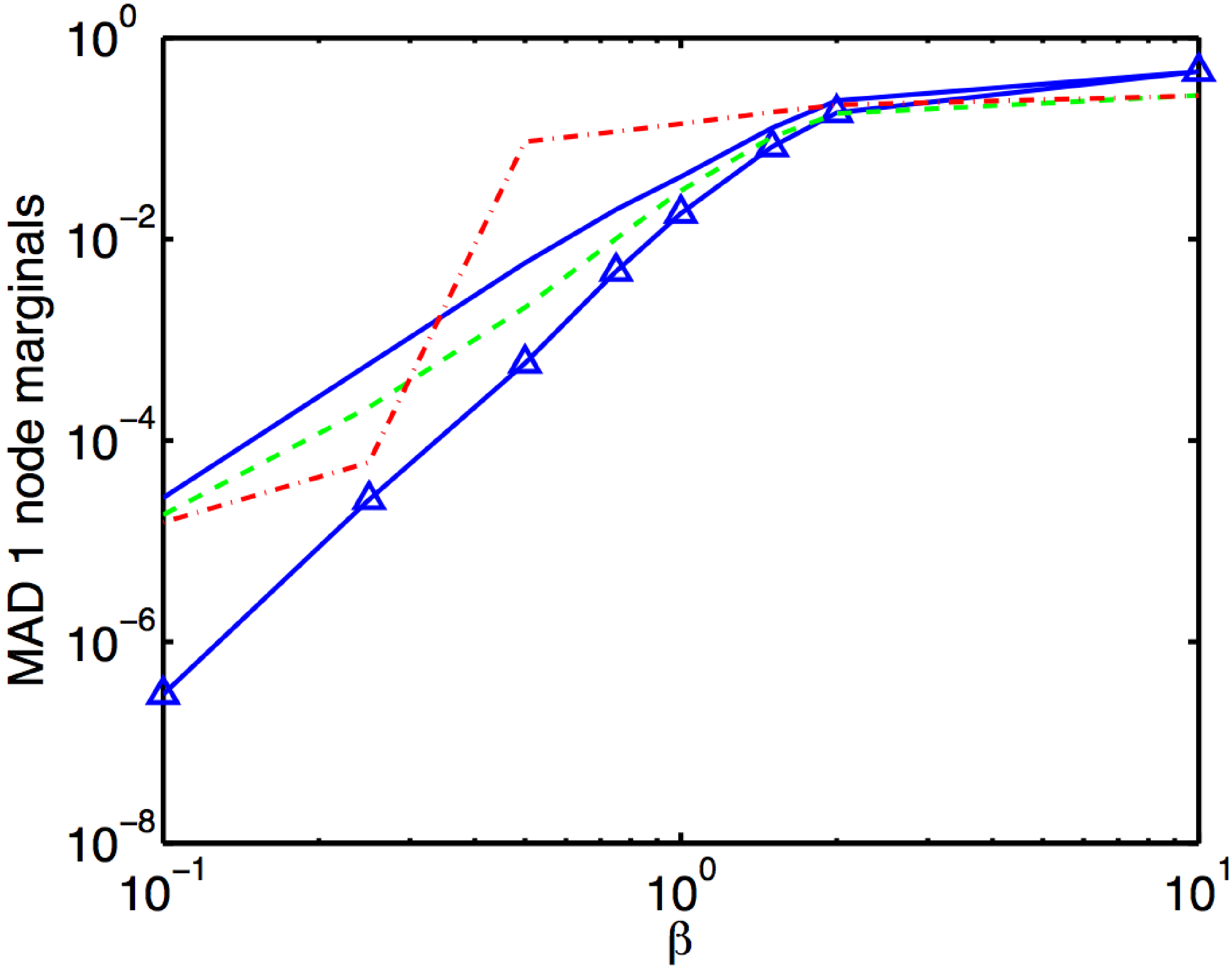
Ising variables $x_{1/2} = \pm 1$ with

$$p(x) = \frac{1}{Z} \exp\left[\frac{1}{2} x^\top J x + \gamma^\top x\right]$$

$J \sim \mathcal{N}(0, \beta^2/N)$ and $N = 10$



AD Free energy
β



MAD 1 node marginals
β
10^0
10^−2
10^−4
10^−6
10^−8
10^−1
10^0
10^1

Outlook

- Systematic expansion w.r.t. Σ_{ij}
- Develop sanity check
- Similar expansions for power EP ?
- Non Gaussian models