Likelihood function

$p(x|\theta)$ ~ probability of x given a model parameterized by θ. θ may be estimate or true.

Example: $P_k(\beta) = Z^{-1} e^{-\beta E_k} = p(k|\beta)$

where $Z = \sum e^{-\beta E_k}$

maximum likelihood: $\frac{\partial}{\partial \theta} p(x|\theta) = 0$

equivalently $\frac{\partial}{\partial \theta} \ln p(x|\theta) = 0$

Example:

$$+\frac{\partial}{\partial \beta} \ln p(k|\beta) = \frac{+1}{p}\left(+Z^{-2} e^{-\beta E_k} \sum E_z e^{-\beta E_z} - p E_k\right)$$

$$= \langle E \rangle - E_k$$

$$\frac{\partial^2}{\partial \beta^2} \ln p(k|\beta) = +\frac{\partial}{\partial \beta}\left(Z^{-1} \sum E_k e^{-\beta E_k}\right)$$

$$= + \left[- Z^{-1} \sum E_k^2 e^{-\beta E_k} - Z^{-2} \left(\sum (-E_k) e^{-\beta E_k} \right) \left(\sum E_k e^{-\beta E_k} \right) \right]$$

$$\partial_\beta^2 \ln p(k|\beta) = - \left[\langle E^2 \rangle - \langle E \rangle^2 \right] < 0$$ around T_c for certain

Note: $\langle E \rangle = f(\beta)$ ~ very non-trivial.

Idea: suppose we have an Ising model ensemble of which we take n energy measurements.

But, we don't know β (say we know $J = 1$).

Interpret $p(\bar{E} | \beta) = p(\beta | \bar{E})$

Then $p(\beta | \bar{E}) = p(\beta ; E_1, \ldots, E_n)$,

where E_i are not neccesarily independent samples.

Why not? If E_i come from MCMC, then there is a finite autocorrelation time.

If E_i come from direct measurement, then E_i are iid. Assume E_i are iid.

Then $p(\beta \mid E_1, \dots, E_n) = \prod_i p(\beta \mid E_i)$

$$= Z^{-n} e^{-\beta \sum E_i}$$

$$= Z^{-n} e^{-\beta n \bar{E}}$$

Note that β is the only variable in $p(\beta \mid E_1, \dots E_n)$.

$$\ln p = -n \ln Z - \beta n \bar{E}$$

$$\frac{\partial}{\partial \beta} \ln p = -\left(\frac{n}{Z} \frac{\partial Z}{\partial \beta} + n \bar{E} \right) = -n \left(\bar{E} - \langle E \rangle \right)$$

$$-\frac{n}{Z} \sum_n E_n e^{-\beta E_n} + n \bar{E} = 0$$

$$\Rightarrow \langle E \rangle = \bar{E} \quad \text{; solve for } \beta \text{!}$$

$\sim$ tune β s.t. $\langle E \rangle_\beta = \bar{E}$!

With degeneracy

$$p(E_k) = Z^{-1}\,\Omega(E_k)\,e^{-\beta E_k}$$

$$p(\{E_i\}_n) = Z^{-n}\prod_{i=1}^{n}\Omega(E_i)\,e^{-\beta E_i} \quad ; \quad \Omega(E_i) = e^{S(E_i)}$$

$$= Z^{-n}\prod_{i=1}^{n} e^{-\beta E_i + S(E_i)}$$

$$= Z^{-n}\prod_{i=1}^{n} e^{-\beta(E_i - TS(E_i))}$$

$$\log(p(\{E_i\}_n)) = -n\log Z + \sum_i(-\beta E_i + \log\Omega(E_i))$$

$$= -n\log Z + \sum_i(-\beta(E_i - TS(E_i))$$

$$= -n(\log Z + \beta\bar{E} + \overline{S(E_i)}_{\{i\}})$$

$$s(x|\beta) = \partial_\beta \log(p(\{E_i\}_n)) = n(\langle E\rangle - \bar{E}) = 0$$

$$\Rightarrow \langle E\rangle = \bar{E} \quad \sim \text{score!}$$

$$\partial_\beta^2 \log p = -(\langle E^2\rangle - \langle E\rangle^2) \sim \text{independent of data!}$$

$$\mathrm{Var}(s(x|\beta)) = \mathrm{Var}(n(\langle E\rangle - \bar{E}))$$

$$= n^2\,\mathrm{Var}(\bar{E}) = n^2\left(\frac{\mathrm{Var}(E)}{n}\right)$$

$$= n\left[\langle E^2\rangle - \langle E\rangle^2\right]$$

Single obs. sample

$$\mathrm{Var}(s(x|\beta)) = -E\left[\partial_\beta^2 \log p\right]$$

$$\mathrm{Var}(s(x|\beta)) = -\left[\langle E^2\rangle - \langle E\rangle^2\right]\Big|_\beta$$

$$\Rightarrow \mathrm{Var}(s(x|\beta))\Big|_{\beta_{MLE}} = -\left[\langle E^2\rangle - \langle E\rangle^2\right]\Big|_{\beta_{MLE}}$$

With autocorrelation

$$\mathrm{Var}(\bar{E}) = 2\tau_{exp}\frac{\mathrm{Var}(E)}{n}$$

$$\Rightarrow \mathrm{Var}(s(x|\beta)) = n^2\,\mathrm{Var}(\bar{E}) = 2\tau_{exp}\, n\,\mathrm{Var}(E)$$

$$= 2n\tau_{exp}\left[\langle E^2\rangle - \langle E\rangle^2\right]$$

$$= -2n\tau_{exp}\, E\left[\partial_\beta^2 \ln p(\{E_i\}|\beta)\right]$$

Questions:

- Confirm relationship in MCMC
  - Compare $Var(E)$ vs. $-\partial_\beta^2 \ln(\underbrace{p(E_i|\beta)}_{\text{function to evaluate}})$ — single sample

- $Var(\hat{\theta})$ vs. $Var(s(E|\beta))$
  - variance of MLE vs. variance of score

- Asymptotic distribution of MLE
  - $\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, I_{\theta_0}^{-1})$

    $n\, Var(\hat{\theta} - \theta_0) \approx I_{\theta_0}^{-1}$

- Fisher information:
  1. expected vs. observed
  2. How to understand $-E_\theta(\partial_\theta^2 \ln p(x|\theta))$

     expectation over θ !?

Properties of MLE

(have lost reference... Close second: Intro Stat Inf... Rohde)

Define $L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log p(x_i|\theta)$ for $\{x_i\}_{i=1}^{n}$.

Take $\ell(x|\theta) = \log p(x|\theta)$, then

$$L(\theta) = E_{\theta_0}\, \ell(x|\theta)$$

$$\sim \int dx \left[\log p(x|\theta)\right] p(x|\theta_0)$$

Note: ① $E_{\theta_0}\, \ell(x|\theta) \neq \ell(E_{\theta_0} x|\theta) = \ell(\mu|\theta)$

② $L(\theta)$ doesn't depend on sample

Law of large numbers (convergence in average)

$$\forall \theta, \quad L_n(\theta) \xrightarrow{n\to\infty} E_{\theta_0}\, \ell(x|\theta) = L(\theta)$$

Lemma: $L(\theta) \leq L(\theta_0)$, moreover

$L(\theta) < L(\theta_0)$ if $P_{\theta_0}\left(p(x|\theta) = p(x|\theta_0)\right) = 1$

Proof: $L(\theta) - L(\theta_0) = E_{\theta_0} \log \frac{p(x|\theta)}{p(x|\theta_0)}$; $\log x \le x - 1$

$$\le E_{\theta_0}\left[\frac{p(x|\theta)}{p(x|\theta_0)} - 1\right]$$

$$= \int dx\, p(x|\theta_0)\left[\frac{p(x|\theta)}{p(x|\theta_0)} - 1\right]$$

$$= \int dx \left(p(x|\theta) - p(x|\theta_0)\right) = 1 - 1$$

$$= 0$$

$$\Rightarrow L(\theta) \le L(\theta_0)$$

Theorem: $\hat{\theta}$ (MLE) is consistent under regularity conditions

Proof: ① $\hat{\theta}$ is the maximizer of $L_n(\theta)$ (by def)

② θ_0 is maximizer of $L(\theta)$ (by Lemma)

③ $\forall \theta$, $L_n(\theta) \to L(\theta)$ by LLN

$[\partial_\theta \ell(x|\theta)]\big|_{\theta=\theta_0}$

Def: $I(\theta_0) = E_{\theta_0}\left(\ell'(x|\theta_0)^2\right)$

Lemma: $I(\theta_0) = -E_{\theta_0}\left[\partial_\theta^2 \ell(x|\theta)\big|_{\theta=\theta_0}\right]$

Theorem: $\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, I(\theta_0)^{-1})$

Proof: $L_n'(\hat{\theta}) = 0$

$$0 = L_n'(\hat{\theta}) = L_n(\theta_0) + L_n''($$