On Constructing Proposal Distributions

Andrew Gelfand
6/16/2011
Importance Sampling

- Want to compute some quantity \( p(x_i = a) = E_p[I(x_i = a)] \)
- Can estimate as \( \hat{p}(x_i = a) = \frac{1}{N} \sum_{n=1}^{N} I(x_i^{(n)} = a) \)
given iid samples \( x^{(1)} \ldots x^{(N)} \) from P
- If cannot sample P, sample surrogate Q instead

\[
p(x_i = a) = E_Q[I(x_i = a) \frac{p(x)}{q(x)}]
\]
\[
p(x_i = a) = E_Q[I(x_i = a) w(x)] \quad w(x) \equiv \frac{p(x)}{q(x)}
\]
\[
\hat{p}(x_i = a) = \frac{1}{N} \sum_{n=1}^{N} I(x_i^{(n)} = a) w(x^{(n)}) \quad \text{(IS estimator)}
\]
\[
\tilde{p}_w(x_i = a) = \sum_{n=1}^{N} I(x_i^{(n)} = a) w(x^{(n)}) / \sum_{n=1}^{N} w(x^{(n)}) \quad \text{(weighted IS estimator)}
\]
Proposal Distributions

What makes Q a good proposal?

1. $p(x) > 0 \Rightarrow q(x) > 0$ (Q includes support of P)
2. Q should be 'close' to P
   - If not, $\tilde{p}(x)$ dominated by few samples
3. Yield a low variance estimate of $\tilde{p}(x)$
4. Easy to sample from
Alpha Divergence

- Often measure 'closeness' in terms of KL
- KL is member of $\alpha$-divergence family
- Can show that [Ali & Silvey 1966]:
  - $D_{KL}(p\|q) = \lim_{\alpha \to 1} D_{\alpha}(p\|q)$ and $D_{KL}(q\|p) = \lim_{\alpha \to 0} D_{\alpha}(p\|q)$
  - For $\alpha \geq 1$, $p(x) > 0 \Rightarrow q(x) > 0$ (zero-avoiding)
  - For $\alpha \leq 0$, $p(x) = 0 \Rightarrow q(x) = 0$ (zero-enforcing)

- So $D_{KL}(p(x|e)\|q(x))$ better than $D_{KL}(q(x)\|p(x|e))$
Is $\alpha$-divergence useful?

- Consider distribution of form: $p(x) = \frac{1}{Z} \prod_k \phi_k(x_k)$

$$D_{KL}(p(x)\|q(x)) = E_p[\ln(p(x)/q(x))] = E_p[\ln p(x)] - E_p[\ln q(x)]$$

$$\min_{q(x)} D_{KL}(p(x)\|q(x)) = \max_{q(x)} E_p[\ln q(x)]$$

- Let $Q$ also be of form: $q(x) = \frac{1}{Z_q} \prod_m \phi_m(x_m)$

$$\max_{q(x)} \sum_{x \in X} p(x) \ln q(x) = \max_{q(x)} \sum_{x \in X} p(x) \sum_m \ln q_m(x_m)$$

$$\frac{\partial}{\partial q_m(x_m)} \left( \sum_{x \in X} p(x) \sum_m \ln q_m(x_m) + \sum_m \lambda_m \sum_{x_m \in X_m} (q_m(x_m) - 1) \right) = 0 \Rightarrow q_m^*(x_m) = p(x_m) \text{ Requires Marginals!}$$
Expectation Propagation [Minka 2001]

- Ideally want to minimize

\[
D_{KL}(p(x)\|q(x)) = D_{KL}\left(\frac{1}{Z_p} \prod_m \phi_m(x_m) \| \frac{1}{Z_Q} \prod_m \tilde{\phi}_m(x_m)\right)
\]

- Minimize KL between \(\phi_m(x_m)\) and \(\tilde{\phi}_m(x_m)\)

- Do so iteratively. For factor \(j\)

\[
\phi_j(x_j) \prod_{m \neq j} \phi_m(x_m) \text{ close to } q^{\text{new}}(x) \propto \phi_j(x_j) \prod_{m \neq j} \tilde{\phi}_m(x_m)
\]

- Idea of Expectation Propagation (EP)

  - If Q in exponential family, find \(q^{\text{new}}\) by matching moments
Minimum Variance Proposal

- Let $\mu = E_p[H(x)]$ be of interest (e.g. $H(x) = I(x_i = a)$)

- $\hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} H(x^{(n)}) w(x^{(n)})$ is unbiased and

\[
q^*(x) = \min_q \text{Var}_Q[\hat{\mu}] = \min_q \frac{1}{N} \text{Var}_Q[H(x) w(x)]
\]

\[
= \min_q E_Q[H^2(x) p^2(x)/q^2(x)] - E_Q[H(x) p(x)/q(x)]^2
\]

\[
\Rightarrow q^*(x) = H(x) p(x)/\mu
\]

- So seek a $Q$ that

\[
\min_{q(x)} D_{KL}(q^*(x) || q(x)) = \min_{q(x)} \sum_{x \in X} q^*(x) \ln q^*(x) - \sum_{x \in X} q^*(x) \ln q(x)
\]

\[
= \max_{q(x)} \sum_{x \in X} H(x) p(x) \ln q(x) = \max_{q(x)} E_p[H(x) \ln q(x)]
\]
Easy to sample

- Desire Q to be a BN
- Choosing structure of Q is important
  - MN to BN requires triangulation (hard)
  - Fully factored Q is too simple (easy)