ENM531: Data-driven modeling and probabilistic scientific computing

Lecture #15: Sampling Methods

Paris Perdikaris
March 26, 2019

Penn
University of Pennsylvania

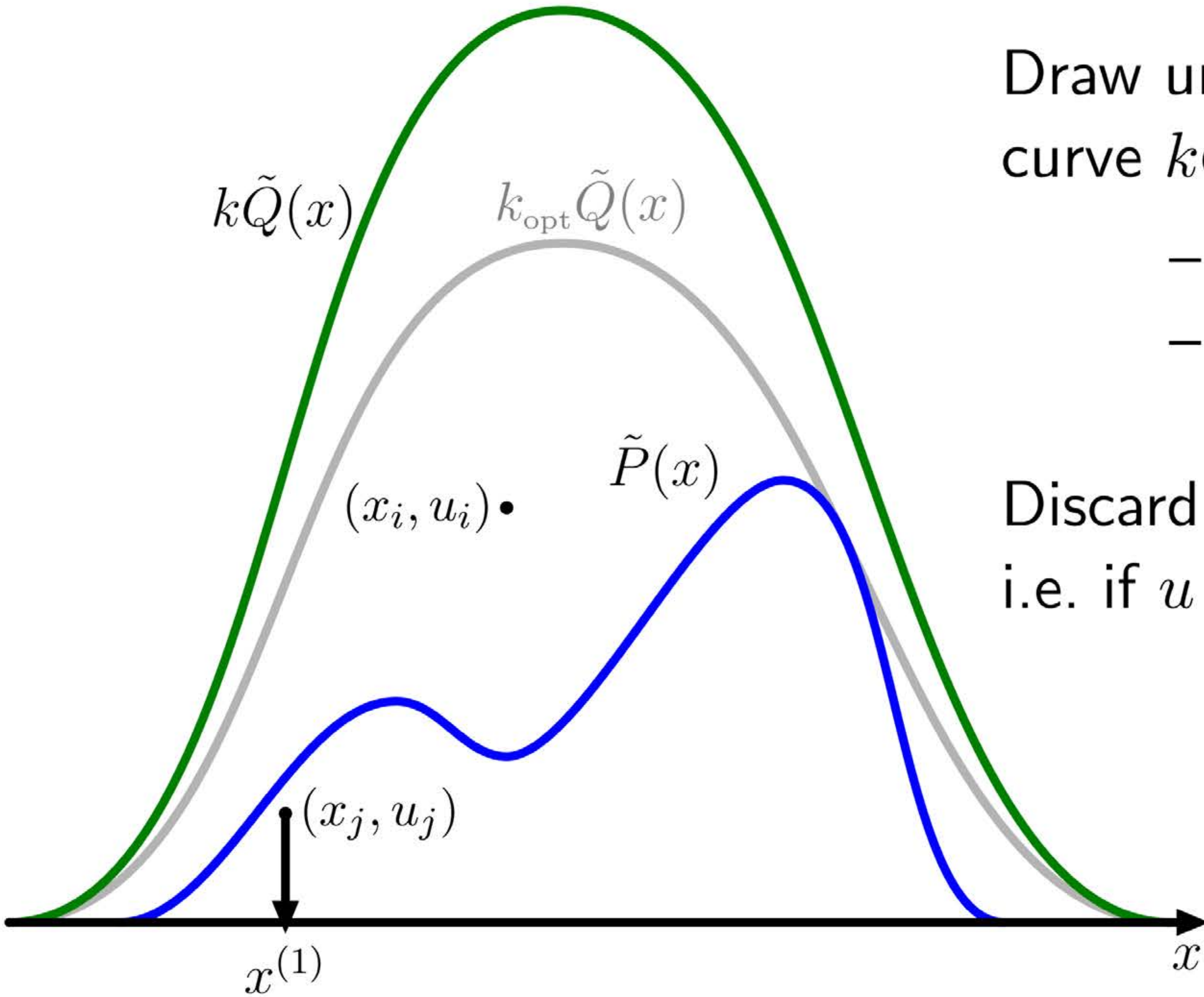
Monte Carlo approximation

$$\mathbb{E}_{x\sim p(x)}[f(x)] = \int f(x)p(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i),$$

where x_i are drawn iid from $p(x)$

Rejection sampling

Sampling underneath a $\tilde{P}(x) \propto P(x)$ curve is also valid



Draw underneath a simple curve $k\tilde{Q}(x) \geq \tilde{P}(x)$:

- Draw $x \sim Q(x)$
- height $u \sim \text{Uniform}[0, k\tilde{Q}(x)]$

Discard the point if above $\tilde{P}$, i.e. if $u > \tilde{P}(x)$

Bayesian linear regression

$$y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, 1/\tau)$$

or equivalently

$$y_i = \beta_0 + \beta_1 x_i + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 1/\tau)$$

The likelihood for this model may be written as the product over N iid observations

$$L(y_1, \ldots, y_N, x_1, \ldots, x_N | \beta_0, \beta_1, \tau) = \prod_{i=1}^{N} \mathcal{N}(\beta_0 + \beta_1 x_i, 1/\tau)$$

Priors on the model parameters:

$$\beta_0 \sim \mathcal{N}(\mu_0, 1/\tau_0)$$

$$\beta_1 \sim \mathcal{N}(\mu_1, 1/\tau_1)$$

$$\tau \sim \mathrm{Gamma}(\alpha, \beta)$$

Gibbs sampling

Gibbs sampling works as follows: suppose we have two parameters θ_1 and θ_2 and some data x.

Our goal is to find the posterior distribution of $p(\theta_1, \theta_2 || x)$.

Gibbs sampling algorithm:

1. Pick some initial $\theta_2^{(i)}$.
2. Sample $\theta_1^{(i+1)} \sim p(\theta_1 || \theta_2^{(i)}, x)$
3. Sample $\theta_2^{(i+1)} \sim p(\theta_2 || \theta_1^{(i+1)}, x)$

Then increment i and repeat K times to draw K samples.

This is equivalent to sampling new values for a given variable *while holding all others constant*.

The general approach to deriving an update for a variable is

1. Write down the posterior conditional density in log-form
2. Throw away all terms that don't depend on the current sampling variable
3. Pretend this is the density for your variable of interest and all other variables are fixed. What distribution does the log-density remind you of?
4. That's your conditional sampling density!

Pros: No parameters need to be tuned (e.g. vs MCMC that needs a proposal distribution)

Cons: It might be hard to analytically derive the conditional distributions.