REJECTION SAMPLING, METROPOLIS-HASTINGS AND MCMC

REJECTION SAMPLING

ASSUME YOU WANT TO SAMPLE FROM A DISTRIBUTION p BUT YOU DON'T KNOW HOW TO. THERE ARE MANY REASONS WHY YOU WOULD WANT TO SAMPLE. FOR INSTANCE, MAYBE YOU WANT TO COMPUTE $E[X]$

$$E[X] = \int x\, p(x)\, dx$$

BUT YOU CAN'T INTEGRATE. IF YOU HAVE ENOUGH SAMPLES $x_1, x_2, \dots x_n$ FROM $p(x)$, THEN

$$\sum \frac{x_i}{n} \longrightarrow E[X]$$

INSTEAD OF p, ASSUME YOU KNOW HOW TO SAMPLE FROM ANOTHER DISTRIBUTION q SUCH THAT:

$$c\, q(x) \geq p(x)$$

FOR SOME CONSTANT c.

REJECTION SAMPLING:

SAMPLE $X \sim q(x)$

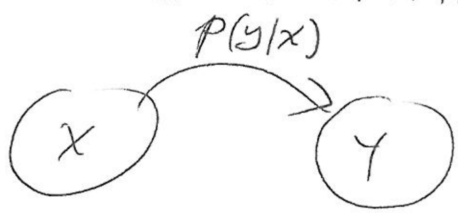
ACCEPT SAMPLE WITH PROB. $\dfrac{p(x)}{c\, q(x)}$

$$P(x \leq X \leq x+\delta) \propto \delta\, q(x) \cdot \frac{p(x)}{c\, q(x)} \propto \delta\, p(x)$$

WITH REJECTION SAMPLING, WE KNOW $p(x)$ BUT WE DON'T KNOW HOW TO SAMPLE FROM p. BUT WHAT IF WE DON'T EVEN KNOW p ITSELF? WHY IS THAT A CONCERN? THIS IS TYPICAL SCENARIO IN BAYESIAN ANALYSIS. RECALL

$$f(\theta|x) = \frac{f(x|\theta)\, f(\theta)}{\int f(x|\theta) f(\theta)\, d\theta}$$

SO $f(\theta)$ posterior IS REALLY NOT KNOWN UNLESS WE KNOW HOW TO INTEGRATE $\int f(x|\theta) f(\theta)\, d\theta$. THE METROPOLIS-HASTINGS ALGORITHM IS A KIND OF REJECTION SAMPLING THAT CAN OVERCOME THIS PROBLEM. CONSIDER A MARKOV CHAIN WITH UNDERLYING TRANSITIONS $p(y|x)$.
ASSUME OUR DESIRED DISTRIBUTION (WHICH WE DON'T KNOW) IS $\pi(x)$. p and π HAVE THE SAME STATE SPACE. IF π IS REVERSIBLE FOR OUR MARKOV CHAIN, THEN π WILL BE THE STATIONARY DISTRIBUTION.



ASSUME ON THE OTHER HAND THAT

$$\pi(x)\, p(y|x) > \pi(y)\, p(x|y)$$

THEN MULTIPLYING THE LEFT HAND SIDE BY $\dfrac{\pi(y)\, p(x|y)}{\pi(x)\, p(y|x)}$ MAKES BOTH SIDES EQUAL (REVERSIBILITY CONDITION).
EFFECTIVELY, WE CHANGE $p(y|x)$ to $\dfrac{\pi(y)}{\pi(x)}\, p(x|y)$

AT TIME n,

SAMPLE $y \sim p(y|x_n)$

WITH PROB $\alpha = \min\left[\frac{\pi(y)\, p(x_n|y)}{\pi(x_n)\, p(y|x_n)}, 1\right]$

MAKE $x_{n+1} = y$ (MOVE)

ELSE $x_{n+1} = x_n$ (STAY)

THIS IS THE METROPOLIS-HASTINGS ALGORITHM, WHICH IS THE BASIS FOR MARKOV CHAIN MONTE CARLO (MCMC).

SYMMETRIC MCMC

$p(y|x) = p(x|y)$, THEN

$$\alpha = \min\left[\frac{\pi(y)}{\pi(x_n)}, 1\right]$$

INDEPENDENT MCMC

$p(y|x) = p(y)$, THEN

$$\alpha = \min\left[\frac{\pi(y)\, p(x_n)}{\pi(x_n)\, p(y)}\right]$$

THIS FORM CAN BE USED IN BAYESIAN ANALYSIS TO USE THE LIKELIHOOD RATIO AS A BASIS FOR MOVING IN THE MARKOV CHAIN.

EXAMPLE:

$$x_i | \mu \sim N(\mu, \sigma^2)$$

WE OBSERVE $\bar{x}$ AND WE HAVE A PRIOR $f(\mu) \propto \frac{1}{1+\mu^2}$

$$f(\mu|\bar{x}) \propto f(\bar{x}|\mu) f(\mu)$$

$$\propto e^{-\frac{(\bar{x}-\mu)^2}{2\sigma^2/n}} \cdot \frac{1}{1+\mu^2}$$

THIS IS OUR (UNKNOWN) DESIRED DISTRIBUTION.

WITH AN INDEPENDENT MCMC WITH UNDERLYING ~~[illegible]~~ TRANSITIONS $\propto \frac{1}{1+\mu^2}$, WE COMPUTE α AS

$$\alpha = \min\left[\frac{e^{-\frac{(\mu_2-\bar{x})^2}{2\sigma^2/n}} \cdot \frac{1}{1+\mu_2^2} \cdot \frac{1}{1+\mu_1^2}}{e^{-\frac{(\mu_1-\bar{x})^2}{2\sigma^2/n}} \cdot \frac{1}{1+\mu_1^2} \cdot \frac{1}{1+\mu_2^2}}, 1 \right]$$

$$\alpha = \min\left[\frac{e^{-\frac{(\mu_2-\bar{x})^2}{2\sigma^2/n}}}{e^{-\frac{(\mu_1-\bar{x})^2}{2\sigma^2/n}}}, 1 \right]$$

WHERE μ_1 IS CURRENT SAMPLE AND μ_2 IS THE NEXT SAMPLE TO BE ACCEPTED WITH PROB. α

SO THE BASIC IDEA HERE IS TO USE AN INDEPENDENT MCMC WITH UNDERLYING TRANSITIONS BASED ON THE PRIOR.

GIBBS SAMPLER

GIBBS SAMPLER IS A SPECIAL CASE MCMC WHERE $\alpha = 1$ (ALWAYS ACCEPT). SO THE CHAIN IS ALREADY REVERSIBLE.

IT IS USEFUL WHEN SAMPLING MULTIPLE VALUES.

ASSUME $(x_1^t, \dots x_n^t)$ IS OUR SAMPLE AT TIME t. TO OBTAIN A SAMPLE AT TIME $t+1$

SAMPLE $x_1^{t+1} \sim p(x_1 / x_2^t, \dots, x_n^t)$

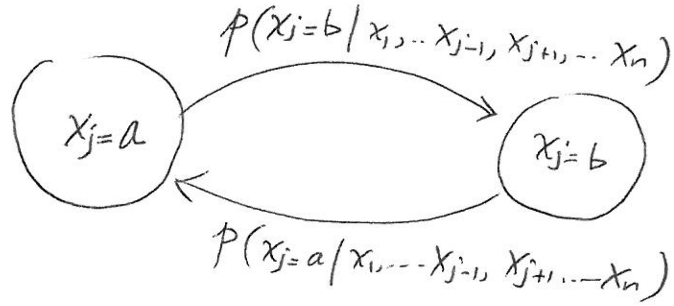
SAMPLE $x_2^{t+1} \sim p(x_2 / x_1^{t+1}, x_3^t, \dots, x_n^t)$

SAMPLE $x_3^{t+1} \sim p(x_3 / x_1^{t+1}, x_2^{t+1}, x_4^t, \dots x_n^t)$

$\vdots$

SAMPLE $x_n^{t+1} \sim p(x_n / x_1^{t+1}, \dots, x_{n-1}^{t+1})$

WE CAN SHOW THAT p, THE JOINT DISTRIBUTION IS REVERSIBLE FOR THE UNDERLYING MARKOV CHAIN.



$$p(x_1, \dots x_j = a, \dots x_n) \cdot p(x_j = b / x_1 \dots x_{j-1}, x_{j+1}, \dots x_n)$$

$$= p(x_1 \dots x_j = a, \dots x_n) \cdot \frac{p(x_1 \dots x_{j-1}, x_j = b, x_{j+1}, \dots x_n)}{p(x_1, \dots x_{j-1}, x_{j+1}, \dots x_n)}$$

$$= p(x_j = a / x_1, \dots x_{j-1}, x_{j+1}, \dots x_n)\, p(x_1, \dots x_j = b, \dots x_n)$$

THE GIBBS SAMPLER IS USEFUL WHEN THE JOINT PROBABILITY IS HARD TO DEAL WITH, BUT MARJINAL PROB. ARE EASY

EXMPLE:

$$x_i \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$$

ASSUME $\frac{S_0}{\sigma^2} \sim \chi^2_K$

AND $\mu \sim N(\beta, \tau^2)$

WE KNOW THAT $\mu \mid \bar{x}, \sigma^2 \sim N\left(\frac{\sigma^2 \beta / n + \tau^2 \bar{x}}{\sigma^2/n + \tau^2}, \frac{\sigma^2 \tau^2 / n}{\sigma^2/n + \tau^2} \right)$

AND $\frac{S + S_0}{\sigma^2} \mid \mu, \bar{x} \sim \chi^2_{K+n}$ WHERE $S = \sum (x_i - \mu)^2$

SO WE CAN SAMPLE (μ, σ) FROM POSTERIOR DISTRIBUTION BY USING A GIBBS SAMPLER.

- START WITH ARBITRARY (μ^0, σ^0)
- REPEAT FOR $t = 0, 1, 2, 3, \ldots$
  SAMPLE μ^{t+1} GIVEN σ^t
  SAMPLE σ^{t+1} GIVEN μ^{t+1}