29/03/2021. LECTURE 15: NOTES MADE ON THE VISUALISER.

3.2 POSTERIOR SAMPLING.

If we can draw a sample of values from the posterior, we can use properties of the sample to estimate properties of the posterior.

[Ideas are applicable to sampling from <u>any</u> distribution]

3.2.1 MONTE-CARLO INTEGRATION.

Suppose we want to estimate $E(g(X))$ for some function $g(x)$ with respect to density $f(x)$.

$$E(g(X)) = \int_{X : f(x) > 0} g(x) f(x)\, dx$$

[Typically interested in $E(g(\theta) \mid X) = \int_{\theta} g(\theta) f(\theta \mid x)\, d\theta$]

Monte-Carlo integration does the following:

- Sample N values $x_1, \dots, x_N$ independently from $f(x)$
- Calculate $g(x_i)$ for each x_i
- Estimate $I = E(g(X))$ by the sample mean of $g(x_i)$

$$g_N(x) = \frac{1}{N} \sum_{i=1}^{N} g(x_i)$$

Note: $$E(g_N(X)) = \frac{1}{N} \sum_{i=1}^{N} E(g(X_i)) = \frac{1}{N} \sum_{i=1}^{N} E(g(X))$$
$$= E(g(X))$$

$$Var(g_N(X)) = \frac{1}{N^2} \sum_{i=1}^{N} Var(g(X_i)) = \frac{1}{N} Var(g(X))$$

3.2.2 IMPORTANCE SAMPLING.

Suppose that we can not easily sample from $f(x)$ but we can sample from $q(x)$ which is an approximate to $f(x)$.

We can use importance sampling to use samples from $q(x)$ to estimate $E(g(X))$.

$$E(g(X)) = \int_{X: f(x)>0} g(x) f(x)\, dx \quad \text{(expectation wrt } f(x)\text{)}$$

$$= \int_{X: f(x)>0} \frac{g(x) f(x)}{q(x)} q(x)\, dx \quad \text{assuming } q(x)>0$$

Support of $q(x)$ could be different to $f(x)$.

Enough to have $q(x)>0$ whenever $f(x)g(x) \neq 0$

i.e. support of $q(x)$ includes support of $f(x)g(x)$

i.e. $q(x)=0 \Rightarrow f(x)g(x)=0$.

Then:

$$E(g(X)) = \int_{X: q(x)>0} \frac{g(x)f(x)}{q(x)} q(x)\, dx$$

$$= E\left[\frac{g(X)f(X)}{q(X)} \,\middle|\, X \sim q(x) \right] \quad \text{Expectation wrt } q(x).$$

By Monte-Carlo integration, estimate $E(g(X))$ by:

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} \frac{g(x_i) f(x_i)}{q(x_i)}$$

where $x_1, \dots, x_N$ is an independent sample from $q(x)$.

Note: $E(\hat{I} \mid X \sim q(x)) = E\left(\frac{g(X)f(X)}{q(X)} \mid X \sim q(x)\right)$

$$= E(g(X) \mid X \sim f(x))$$

$$Var(\hat{I} \mid X \sim q(x)) = \frac{1}{N} Var\left(\frac{g(X)f(X)}{q(X)} \mid X \sim q(x)\right)$$

Hence, variance depends upon both N and the choice $q(x)$.

Given a choice of possible $q(x)$s, choose the one that minimizes $Var(\hat{I} \mid X \sim q(x))$. [Includes the case that $f(x) = q(x)$]

3.3 MARKOV-CHAIN MONTE-CARLO (MCMC)

MCMC is a general technique that generates samples from complex multi-dimensional distributions in cases where the distribution is analytically intractable.

A Markov chain is a random process with the property that, conditional upon its current value, future values are independent of the past.

Under certain conditions, such a chain will converge to a stationary distribution. After such point, values can be viewed as observations from the stationary distribution.

Idea behind MCMC techniques for Bayesian statistics:

① Construct a Markov chain with $f(\theta \mid x)$ as its stationary distribution. (How?)

Markov chain

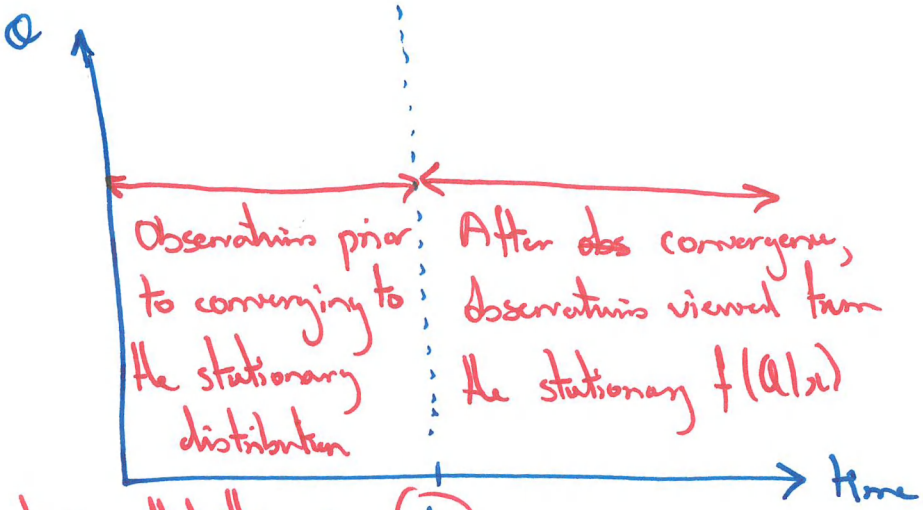
② From some starting point, generate values from the chain until (approximate) convergence (How many?)

③ Continue generating values from the chain which are now a sample from $f(\theta|x)$.

Monte-Carlo

④ Use the sample to estimate properties of the posterior. (How many to take; variability?)

[Non-independent observations].



This is sometimes called the "burn-in" ← b

$$\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(b)} \,\vdots\, \underbrace{\theta^{(b+1)}, \theta^{(b+2)}, \ldots, \theta^{(b+N)}}_{\text{use these to estimate properties of the posterior}}, \ldots$$