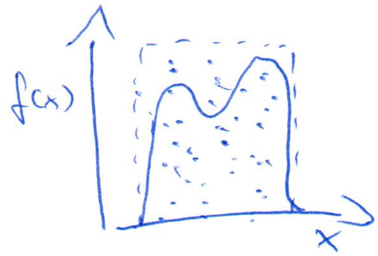
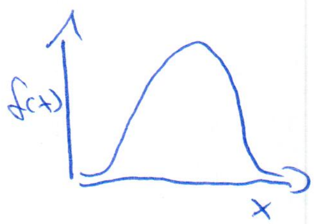
# Lecture 23 - Data Augmentation and slice sampling

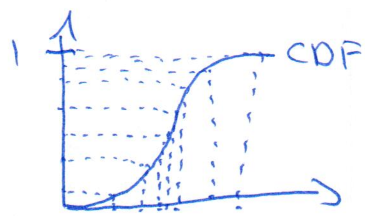
First a recap of our sampling methods:

- rejection sampling




- inverse sampling ⇒ needs CDF and the inverse




- Metropolis hastings
  - current state $x^{(i)}$
  - proposed new state $x^{(i+1)} = x^*$
  - acceptance probability $\frac{p(x^*)}{p(x^{(i)})}$
  - draw $u \sim$ Uniform $[0,1]$
  - if $u <$ acceptance prob $\Rightarrow x^{(i+1)} = x^*$

    otherwise $\Rightarrow x^{(i+1)} = x^{(i)}$

Remember:

- step size of proposal distibution needs to be tuned

Gibbs Sampling:

- Need to know and be able to sample from the conditionals

$$p(x_1 \mid x_2 = x_2^{(i)})$$
$$p(x_2 \mid x_1 = x_1^{(i)})$$

$\Rightarrow$ Sample $x_1^{(i+1)} \sim p(x_1 \mid x_2 = x_2^{(i)})$

Sample $x_2^{(i+1)} \sim p(x_2 / x_1 = x^{(i+1)})$

Remember:
- Gibbs sampling is a form of MH where we always accept the proposal
- If we can sample from some conditionals we can use MH to sample from the other conditionals to still use Gibbs sampling

# Data Augmentation

- Sometimes a problem can be solved more easily when we actually increase the number of dimensions.

- You have seen this in the context of EM where we introduced latent variables to make the optimization easier.

- Remember the point matching problem. Estimating $A$ when knowing $M$ is easy, updating $M$ when knowing $A$ is easy. Estimating $A$ without knowledge of $M$ is hard.

- For sampling with data augmentation we need:

$$p(x) = \int p(x,y)\, \underbrace{dy}_{\text{augmenting variable}}$$

- We win if sampling from $p(x|y)$ and $p(y|x)$ is easier than sampling from $p(x)$.

# Slice Sampling:

Is essentially Gibbs Sampling using data augmentation.

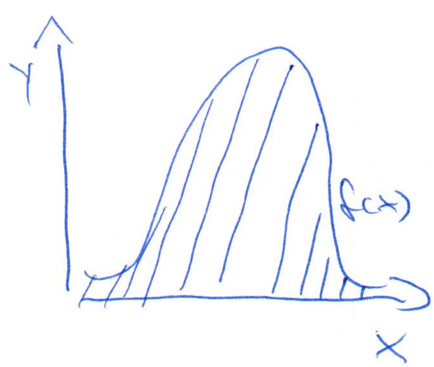
$\Rightarrow$ want to sample from $f(x)$

$\Rightarrow$ sample from $p(x,y) = \begin{cases} \frac{1}{Z} & \text{for } 0 < y < f(x) \\ 0 & \text{otherwise} \end{cases}$

$Z$: normalization: $\int f(x)\,dx$

$$\Rightarrow p(x) = \int_0^{f(x)} \frac{1}{Z}\,dy = \frac{f(x)}{Z}$$

- Our augmented problem makes e.g. a one dimensional function a two dimensional one which is greater than 0 and the curve.



- start with some $x^{(i)}$
- Sample $y \sim \text{Uniform}[0, f(x^{(i)})]$
- Sample $x^{(i+1)}$ from the slice part that is under the curve.
- repeat

⇒ In the end discard all $y$ and only keep the $x$ samples, effectively marginalizing the joint distribution $p(x,y)$ to get $p(x)$

⇒ Again no free lunch! Finding the valid area of the slice to sample $x$ from is not that easy (but easy enough).

⇒ Can use rejection sampling

⇒ Can shrink the borders at rejected samples to make it more efficient.

⇒ can use stepping out

⇒ or doubling (see paper or notebook for details, Neil, "Slice Sampling", 2003)

## Stepping out:

- choose width parameter $w$
- place $w$ interval randomly around $x_0$
- expand left and right border by $w$ until both lie outside of $f(x)$
- Sample from this region. If sample is outside (not under the curve), shrink the valid region accordingly.

Important:
- random positioning of $w$ around $x_0$
- random restriction of maximal slice size.

pseudo code:

$U \sim Uniform(0,1)$
$L \leftarrow x_0 - w \cdot U$
$R \leftarrow L + w$

$V \sim Uniform(0,1)$
$J \leftarrow \lfloor m \cdot V \rfloor$
$K \leftarrow (m-1) - J$

while $J > 0$ and $y < f(L)$:
    $L \leftarrow L - w$
    $J \leftarrow J - 1$

while $K > 0$ and $y < f(R)$:
    $R \leftarrow R + w$
    $K \leftarrow K - 1$