Lecture 19

- Disadvantages of the grid, G:
  1. Smallest and largest values beyond a computer's storage
  2. dimensionality: Since the number of grid sections/partitions increases exponentialy the partitions of each dimension drops dramatically, so POOR resolution
  3. Region: min., max. of grid can exclude entire regions

- Again: $X_1 \dots X_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$ where θ and σ^2 are both unknown

- $\theta \sim N(\mu_0, \tau^2)$ & $\sigma^2 \sim \text{invGamma}\left(\frac{n_0}{2}, \frac{n_0 \sigma_0^2}{2}\right)$

- so $P(\theta, \sigma^2 \mid x) \neq \text{NormInvGamma}$; $= N(\theta_p, \sigma_p^2)\ K(\sigma^2 \mid x)$
  $\approx P(\theta \mid x, \sigma^2)\ \underbrace{K(\sigma^2 \mid x)}_{\text{an unknown distribution}}$

- Recall: $P(\theta \mid x, \sigma^2) = N(\theta_p, \sigma_p^2)$
  $P(\sigma^2 \mid x, \theta) = \text{InvGamma}\left(\frac{n_0 + n}{2}, \frac{n_0\hat{\sigma}_0^2 + n\hat{\sigma}^2}{2}\right)$

- Imagine Algorithm:
  1. Let $\theta_0 = 0$
  2. Draw σ_0^2 from $P(\sigma^2 \mid x, \theta = \theta_0)$
  3. Now θ_1 from $P(\theta \mid x, \sigma^2 = \sigma_0^2)$
  4. Draw σ_1^2 from $P(\sigma^2 \mid x, \theta = \theta_1)$

  ...

  ...

  ...

  until a stabilization, or "Convergence"

  Gibb's sampling

  ** ideally start with many different θ_0

- Output:

$$< \underbrace{\begin{bmatrix}\theta_0 \\ \sigma_0^2\end{bmatrix}\begin{bmatrix}\theta_1 \\ \sigma_1^2\end{bmatrix}\begin{bmatrix}\theta_2 \\ \sigma_2^2\end{bmatrix}\cdots\begin{bmatrix}\theta_B \\ \sigma_B^2\end{bmatrix}}_{\text{Burn-in period}}\overbrace{\underbrace{\begin{bmatrix}\theta_{B+1} \\ \sigma_{B+1}^2\end{bmatrix}\cdots\begin{bmatrix}\theta_G \\ \sigma_G^2\end{bmatrix}}_{\text{Burned-in Chain}}}^{G \text{ iterations}} >$$

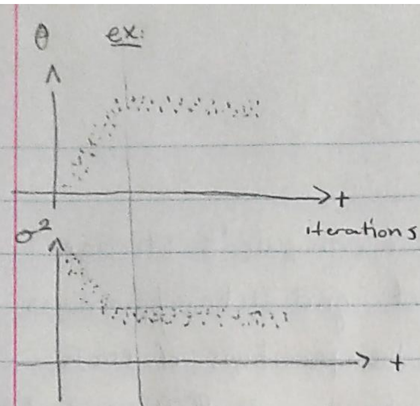
(P dimentional)

Markov Chain

Monte Carlo

: MCMC

ex:



B: Burn in point; point of convergence

Note: Goals were:

$P(\theta, \sigma^2 | X)$
$P(\theta | X)$ } most practical
$P(\sigma^2 | X)$

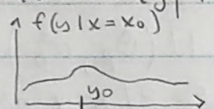
All can be sampled from the MCMC Burned-in Chain

* The Burn in value is gotten visually and is subjective

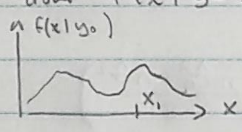
- **Generalization:** Given $f(X|y)$ and $f(y|X)$ are known we want $f(X,y)$

① start at X_0

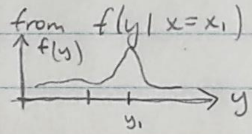
② draw y_0 from $f(y | X = X_0)$



③ draw X_1 from $f(X | y = y_0)$

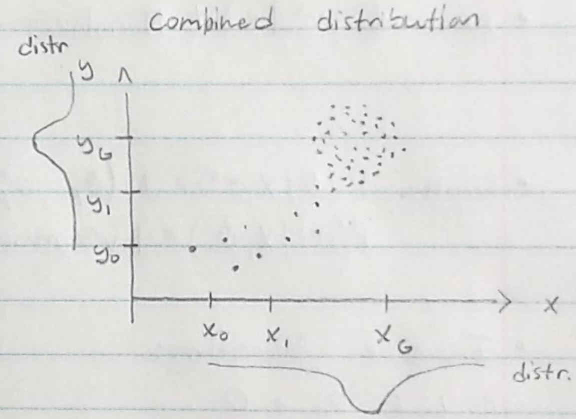


④ draw y_1 from $f(y | X = X_1)$



...
etc.

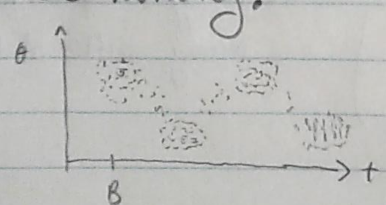
so $f(x,y)$ is the combined distribution



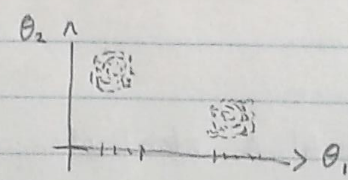
* collapse to a single dimension by ignoring other values *

- Problems of Gibb's sampling:

① Bad mixing:



Such that



- sampling gets stuck if there are various modes especially with large dimensions

** fixable if we start at various points and carry out more chains.

② Not iid samples for $P(\theta, \sigma^2 | X)$
since each of the points are dependent on previous.

Recall: iid samples have a covariance of 0

Note: Covariance $[X,Y] = E[(X-\mu_x)(Y-\mu_y)] = E[X,Y] - \mu_x \mu_y$

$$** \quad \boxed{\rho = \text{cov}[X,Y] = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\underbrace{SE[X] \quad SE[Y]}_{\text{scaling values.}}}}$$

"Condition"

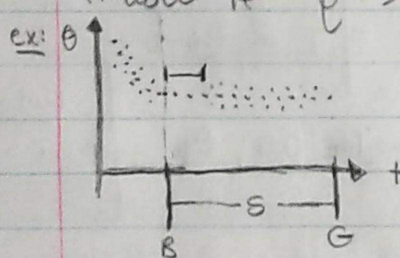
estimated by:

$$\boxed{r := \frac{S_{xy}}{S_x S_y} = \frac{\frac{1}{n-1} \sum (X_i - \bar{x})(Y_i - \bar{y})}{\sqrt{\frac{1}{n-1} \sum (X_i - \bar{x})^2} \sqrt{\frac{1}{n-1} \sum (Y_i - \bar{y})}}} \in [-1, 1]$$

"sample condition"

- fixable if $\rho \to 0$ such that the sample is no longer correlated

ex:



① we want to Average Burned-in Chain: $\bar{\theta}$

$$* \quad \left[\bar{\theta} = \frac{1}{S} \left(\sum_{t=B+1}^{B+S} \theta_t \right) \right]$$

② the we want to test "lags" from $1, 2, \dots, K$ from B to G until a certain lag gives a negligible correlation between points.
we use:

$$** \left[\underset{\substack{\uparrow\uparrow \\ \text{autocorrelation} \\ \text{of lag } K}}{r_{aK}} := \frac{\sum_{t=B+1}^{B+S-K} (\theta_t - \bar{\theta})(\theta_{t+K} - \bar{\theta})}{\sum_{t=B+1}^{B+S} (\theta_t - \bar{\theta})^2 \Big\} \text{ scaling}} \right]$$

Note
"lag" # represents the number of points from one step to the next

ex.

lag 1 3 7

③ when $r_{aK} \to 0$ (≈ 0) we use that K value to determine where in the Burned in chain we have noncorrelated measures of θ and after K points where we have iid samples for θ