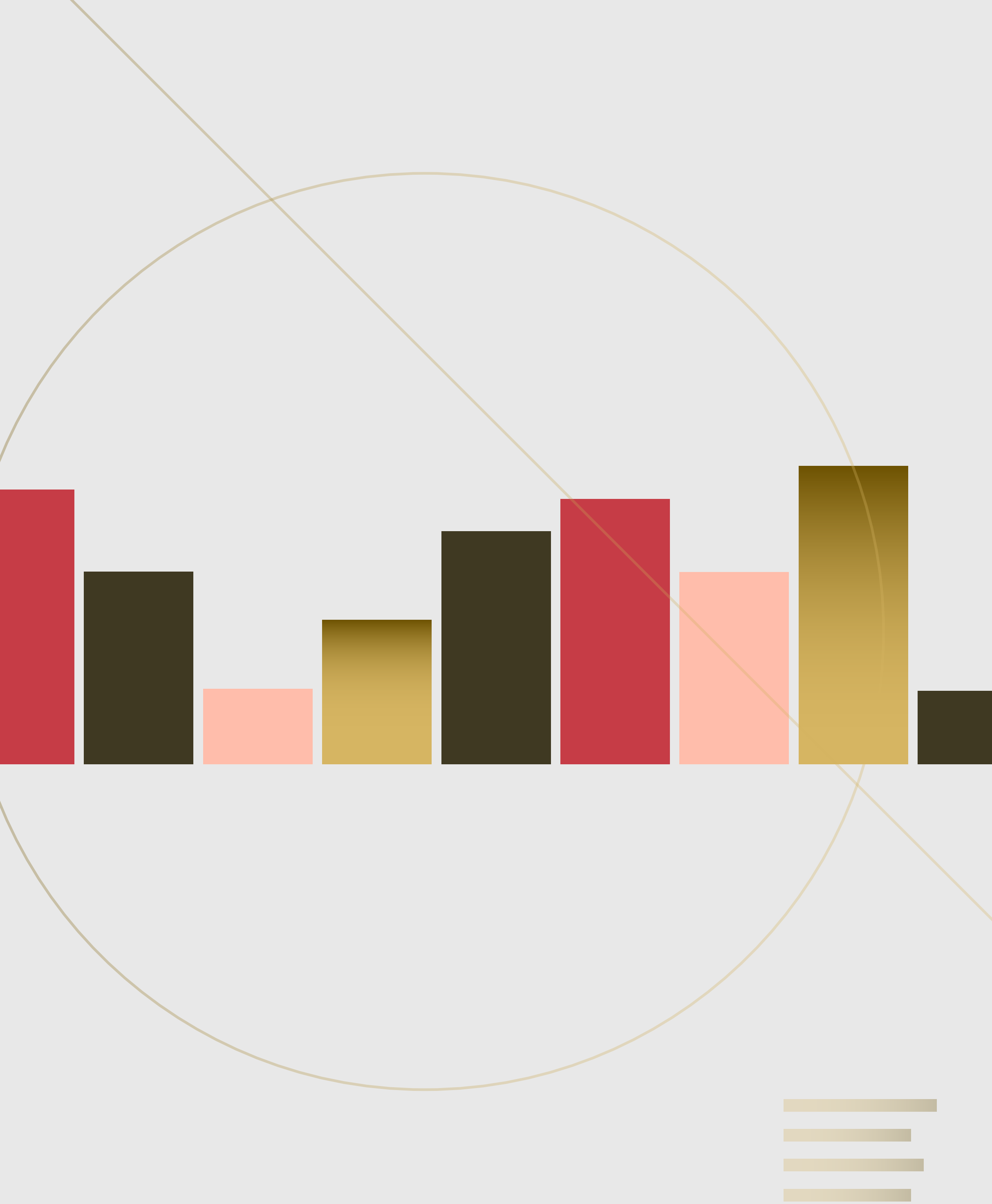

Understanding the Metropolis - Hastings Algorithm
- Chib & Greenberg (1995)
Journal of American Statistical Association

- Correlation can be introduced in sampling making the observations dependent rather than independent as a tool to reduce variance.
Eg- Markov samples

## Acceptance Rejection Sampling

target density - $\pi(x) = \frac{f(x)}{K}$ , $x \in \mathbb{R}^d$

where $f(x)$ = unnormalized density
$K$ = normalizing constant

Density from which we want to sample the observations.

$h(x)$ - some other density [Researcher has to chose it] which can be simulated using some method &
$\exists\, c$ s.t $f(x) \le c\,h(x)$ $\forall x$

Then to obtain one random variate from $\pi(\cdot)$

Step 1: Generate a candidate $Z$ from $h(x)$
Generate a value $u$ from $U(0,1)$

Step 2: If $\frac{f(x)}{c\,h(x)} \le u$ - reject $Z$ and go to step 1

If $\frac{f(x)}{c\,h(x)} \ge u$ - accept $Z \Rightarrow$ sampled from $\pi(\cdot)$

Objective - Chose min $c$ that satisfy condition $f(x) \le c\,h(x)$ to decrease number of rejections

Drawback - Too many rejections, hence inefficient

## Markov Chain Monte Carlo Simulation

State space is continuous
$P(x, A)$ is transition kernal for $x \in \mathbb{R}^d$
$A \in \mathcal{B}$, $\mathcal{B}$ - borel $\sigma$-algebra
- smallest $\sigma$ algebra on open sets in $\mathbb{R}^d$

$P(x, A)$ - conditional probability of moving in set $A$ from $x$.

Invariant distribution

$$\pi^*(dy) = \int_{\mathbb{R}^d} P(x, dy)\, \pi(x)\, dx$$

$$\pi^*(dy) = \pi(y)\, dy \qquad [\pi^*(dy) \text{ is Lebesgue measure}]$$

$$\pi(y)\, dy = \int_{\mathbb{R}^d} P(x, dy)\, \pi(x)\, dx$$

$n^{th}$ iterate converges to the invariant distribution under certain conditions

Idea - To generate samples from $\pi(\cdot)$, the algorithm utilize a transition kernel $P(x, dy)$ whose $n^{th}$ iterate converges to $\pi(\cdot)$ for large $n$.

Now, we have to find appropriate $P(x, dy)$ which converges to $\pi(\cdot)$

Suppose for some $p(x, y)$

The term account for the fact that $p(x,x) = 0$

$$P(x, dy) = p(x, y)\, dy + r(x)\, \delta_x(dy)$$

$$p(x, x) = 0$$

$$\delta_x(dy) = 1 \text{ if } x \in dy$$

$$r(x) = 1 - \int_{\mathbb{R}^d} p(x, y)\, dy$$ → probability that process stays in $x$.

If $p(x,y)$ satisfies reversibility condition

$$\pi(x)\, p(x,y) = \pi(y)\, p(y,x)$$

Then $\pi(\cdot)$ is invariant distribution of $P(x\ )$

Prove: $$\int_A P(x,A)\,\pi(x)\,dx \qquad\qquad x,y \in A$$

$$= \left[\int_A \left[\int_A p(x,y)\,dy + r(x)\,\delta_x(A)\right]\pi(x)\,dx\right.$$

$$= \int_A \left[\int_A p(x,y)\,dy\ \pi(x)\,dx + \int_A r(x)\,\delta_x(A)\,\pi(x)\,dx\right.$$

$$= \int_A \left[\int_A p(y,x)\,\pi(y)\,dy\,dx + \int_A r(x)\,\pi(x)\,dx\right.$$

$$= \int_A \int_A p(y,x)\,dx\ \pi(y)\,dy + \int_A r(x)\ \pi(x)\,dx$$

$$= \int_A [1 - r(y)]\,\pi(y)\,dy + \int_A r(x)\,\pi(x)\,dx$$

$$= \int_A \pi(y)\,dy$$

## Metropolis - Hastings Algorithm

Candidate generating density $= \int q(x,y)\,dy = 1$

If the process is at $x$, this density generate $y$ from $q(x,y)$.

Objective $\Rightarrow$ $q(x,y)$ satisfy reversibility condition

If not then

$$\pi(x)\ q(x,y) \underset{\text{or} \ <}{>} \pi(y)\,q(y,x)$$

prob of going from $x \to y$ is higher than $y \to x$

To correct it, introduce a weight $\alpha(x,y) < 1$ which is multiplied on LHS to reduce probability of going from $x$ to $y$.

In A·R → if rejection, a new pair was drawn independently of the previous value as draws are independent

In MH → the current value is not independendent of previous value. Thus we have $q(x,y)$ interpreted as conditional probability rather than $h(y)$ (independent of previous value $x$) which was the case in A-R.

$$\pi(x)\, q(x,y)\, \alpha(x,y) = \pi(y)\, q(y,x)\, \alpha(y,x)$$

Thus $P_{MH}(x,y) = q(x,y)\,\alpha(x,y)$ [$P(x, A)$ in derivation] is a new density that we created using $q(x,y)$ and $\alpha(x,y)$ which satisfy reversibility

If $\pi(x)\, q(x,y) > \pi(y)\, q(y,x)$

Then put $\alpha(y,x) = 1$ [maximum probability] & solve for $\alpha(x,y)$.

$$\alpha(x,y) = \frac{\pi(y)\, q(y,x)}{\pi(x)\, q(x,y)}$$

Thus generally for both > & < case.

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

Conditions required for convergence to invarient distri-bution is aperiodicity and irreducibility

MH algorithm

Step 1 Generate $y$ from $q(x^t, \cdot)$
Generate $u$ from $U(0,1)$

Step 2 If $u \leq \alpha(x^t, y) \Rightarrow$ set $x^{t+1} = y$
Else $x^{t+1} = x^t$

$\Rightarrow$ Return the values $\{x^{(1)}, x^{(2)}, \dots x^{(T)}\}$

Draws are regarded as sample only after the distribution has passed the transient stage.