🎲 Monte Carlo Methods & MCMC Cheat Sheet (Lec 9)

🎯 MC Fundamentals & Estimation

Concept	Detail	Formula	Symbol Key
Why MC?	Used to approximate **integrals** or **probabilities** that are analytically hard to compute[1].		
MC Expectation	Estimate the expected value $E[h(Z)]$ using the **sample mean** (relies on LLN)[2].	$E[h(Z)] \approx \hat{h} = \frac{1}{M}\sum^{N} h(z_i)$	M Number of samples/replicates. z_i: Sample drawn from the random variable Z. $h(Z)$: Function of Z.
MC Probability	Estimate $\mathbb{P}(E)$ for event E	$\mathbb{P}(E) = \frac{1}{M}\sum_{i=1}^{M} 1_{z_i \in E}$	1 Indicator function (1 if condition is true, 0 otherwise).
Variance/Error	Variance decreases as $1/M$. Error $e_M(f)$ is approx. $\mathcal{N}$ ..... for large M (CLT)[5].	$Var(\hat{h}) \propto 1/M$ …	

🔢 Random Number Generation (RNG)

Method	Core Idea	Formula	Symbol Key
Inverse CDF Sampling	Draw $U \sim Uniform(0,1)$ and compute $X = F^{-1}(U)$	$X = F_X^{-1}(U)$	F_X^{-1} Inverse Cumulative Distribution Function. U: Uniform random variable.

Method	Core Idea	Formula	Symbol Key
	to get a sample X from $f(x)$		
Box-Muller	Generates two independent standard normal variables (x_1, x_2) from two uniform variables (y_1, y_2)	$x_1 = \sqrt{-2\log y_1}\cos(2\pi y_2)$ $x_2 = \sqrt{-2\log y_1}\sin(2\pi y_2)$	y_1, y_2 Independent $\text{uniform}(0,1)$ variables. π : Pi (≈ 3.14159)

🛠️ Advanced Sampling Methods

Method	Goal	Key Constraint / Requirement	Formula / Acceptance
Importance Sampling (IS)	Estimates $E_f[h(x)]$ when sampling from target $f(x)$ is hard, but proposal $g_1(x)$ is easy[8].	Requires $g_1(x) > 0$ whenever $f(x) \neq 0$ (same support).	$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\frac{f(x_i)}{g_1(x_i)}}_{\omega(x_i)} h(x_i)$
Rejection Sampling	Generates IID samples from an **un-normalized** target $\pi(x) \propto g(x)$	Requires easily sampled proposal $f(x)$ and constant k such that $kf(x) \geq g(x), \forall x$.	Sample $X \sim f$ **Accept** if $u \leq \frac{g(x)}{kf(x)}$ where $u \sim U(0,1)$

Meth od	Goal	Key Constraint / Requirement	Formula / Acceptance
	………… ………… ..		

⚙️ Markov Chains and MCMC

Markov Chain Properties

Property	Definition/Condition
Markov Property	The next state ……X_{n+1}………depends **only** on the current state ……X_n…………… ("memoryless")[13].
Stationary Distribution (π)	A probability distribution such that ……$\pi = \pi P$……….. If the chain starts in ….π.., the distribution remains …π. for all future steps.
Irreducible	Positive probability of eventually reaching any state from any other state.
Aperiodic	No **forced cycles** exist in the state transitions[16]. (Ensures convergence to .π….)

MCMC Algorithms

Goal: Construct a Markov Chain ~~X_n~~ X_n whose stationary distribution is the desired target ~~π~~ π. The samples are taken after a **burn-in** period (~~B~~)[17]. B

Algorithm	Proposal q(x,y)	Acceptance Ratio (An)	Acceptance Probability (α(x,y))
Metropolis (Random Walk)	Symmetric proposal, $q(x,y) = q(y,x)$	$A_n = \frac{\pi(Y_n)}{\pi(X_{n-1})}$	$\min[1, A_n]$ (Ensures reversibility/stationarity if ..q.. is symmetric)[20]. st balanced or mirrored (If u split half → 2 part are the same)
Metropolis -Hastings (MH)	Non-symmetric proposals are allowed.	$A_n = \frac{\pi(Y_n)q(Y_n, X_{n-1})}{\pi(X_{n-1})q(X_{n-1}, Y_n)}$	$\min[1, A_n]$ (Ensures reversibility/stationarit

Algorithm	Proposal q(x,y)	Acceptance Ratio (An)	Acceptance Probability (α(x,y))
			y even if ..q.. is asymmetric).

Not balanced, equal (Two side behave different)

📈 Other Concepts

Moments and MGF

Concept	Detail	Formula	Symbol Key
Moment Generating Function (MGF)	$E(e^{tx})$ .... Derivatives at $t=0$ .. give the raw moments.	$M_x(t) = E[e^{tx}]$	..t..: Variable for which the function is taken. ..X..: Random variable.
MGF Properties	If $Y = \beta_0 + \beta_1 X$ .... , then $M_Y(t) = e^{\beta_0 t} M_X(\beta_1 t)$ .... If X are independent, $M_Y(t) = \prod M_{X_i}(t)$		β_0, β_1 .... : Coefficients/constants M_X : .... MGF of ....

Thompson Sampling

Thompson Sampling is a **Bayesian approach to sequential decision-making** used to manage the **exploration vs. exploitation dilemma** (e.g., in multi-armed bandit problems).

- **Core Idea**: At each step, sample a parameter from each option's posterior, choose the option with the highest sample, observe the result, and update the posterior.
- **Application**: Widely used in digital marketing, pricing, and A/B/n testing[29].

1. What Is a Density?

A probability density describes how likely different values of a random variable are.

For a density f(x):

- The shape tells you where probability mass is.
- The area under the curve must equal 1.

$\int f(x)\,dx = 1$ (Derivative of $f(x) = 1$)

This is a normalized density.

⭐ 2. Normalized Density (Proper Probability Density)

✔ Definition:

A function is a normalized density if it integrates to 1.

$\int p(x)\,dx = 1$

Examples:

- Normal distribution
- Exponential distribution
- Beta distribution

✔ Example: Normal(0,1) PDF

$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

derivative of it = 1

It is normalized because:

$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = 1$

⭐ 3. Unnormalized Density (Not Proper Yet)

✔ Definition:

An unnormalized density is proportional to a probability density but does not integrate to 1.

$\tilde{p}(x) \propto p(x)$

It has the right shape, but not the right height.

It might look like:

$\tilde{p}(x) = e^{-x^2/2}$

$\frac{1}{\sqrt{2\pi}}$

This is missing the normalizing constant ……, so:

$\int \tilde{p}(x)\,dx = \sqrt{2\pi} \neq 1$