Aidan Kardan HW 1 Bayesian

(RS)

1. Last semester, I learned about rejection sampling in my Statistical Computation class. RS is often used in MC methods to generate samples from a complicated distribution when direct sampling is difficult. RS draws candidate values from an easy to sample proposal distribution and accepts them w/ a probability proportional to their likelihood under the target distribution.

My idea: Instead of updating beliefs using Bayes' Rule, I propose an alternative belief update mechanism using rejection sampling based on likelihood acceptance.

Instead of computing: $P(\theta|y) \propto P(y|\theta)P(\theta)$

I propose:

1. Generate a candidate value θ^* from a broad proposal distribution $q(\theta)$. $q(\theta)$ should be broad enough to cover reasonable values of θ.
(Draw a random sample from the proposed distribution)

☆ If we:
- have no prior knowledge, $\theta^* \sim U(a,b)$
- expect values concentrated, $\theta^* \sim N(\mu, \sigma^2)$
- expect $\theta > 0 \ \forall \theta$, $\theta^* \sim Exp(\lambda)$

2. Compute the likelihood of observing the data under θ^*

$$L(\theta^*) = P(y|\theta^*)$$

3. Accept θ^* w/ probability

$$P_{accept}(\theta^*) = \frac{L(\theta^*)}{M} \leq 1 \quad \forall \theta^*$$

where M is a normalizing constant ensuring valid probability,

$M = \max_\theta P(y|\theta) \rightarrow$ maximum likelihood across all candidates

4. Generate a random number from $u \sim U(0,1)$
5. Accept θ^* if $u \leq P_{accept}(\theta^*) \Rightarrow u \leq \frac{L(\theta^*)}{M}$, otherwise reject θ^* and try again w/ new candidate value θ_2^*
6. Repeat w/ new candidate θ_2^* until enough values are accepted!

Enough values: Setting a fixed number of N accepted samples based on Knowledge of data or checking if μ & σ^2 of the accepted samples have stabilized (don't change significantly over last 1,000 samples)

Comparison: Rolling a 6 on a dice

$\theta = \frac{1}{6}$, Roll 10 times , observe 3 6's : information

Bayes' Rule: $P(\theta|y) \propto P(y|\theta)P(\theta) \rightarrow \binom{10}{3}\theta^3(1-\theta)^7 \rightarrow$ follows a Beta distribution,

$\text{Beta}(k+1, (n-k)+1) = \text{Beta}(4,8)$ $\qquad E(\text{Beta}(4,8)) = \frac{4}{12} = \frac{1}{3}$

$k=3, n=10$

My Idea: (Preferably done in R but good enough for this example to compare quickly)

$\hat{\theta}_{MLE} = \frac{k}{n} = \frac{3}{10} = 0.30$ $\qquad \theta$ is between 0 & 1

Generate candidates (done by hand so very few candidates considered, will want to code for better comparison)

$\theta^* = \{0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95\}$, candidates from $U \sim U(0,1)$ ☆

Instead of computing exact values, approximate relative values to M, $\theta^* = 0.35$ is closest to $M = 0.30$,

P-Accept

$\theta^* = 0.05 \rightarrow 0.05M \rightarrow 0.05$
$0.15 \rightarrow 0.20M \rightarrow 0.20$
$0.25 \rightarrow 0.45M \rightarrow 0.45$
$0.35 \rightarrow 0.9M \rightarrow 0.9$
$0.45 \rightarrow 0.7M \rightarrow 0.7$
$0.55 \rightarrow 0.5M \rightarrow 0.5$
$0.65 \rightarrow 0.3M \rightarrow 0.3$
$0.75 \rightarrow 0.15M \rightarrow 0.15$
$0.85 \rightarrow 0.08M \rightarrow 0.08$
$0.95 \rightarrow 0.03M \rightarrow 0.03$

Based on Binomial distribution decay behavior when changing θ^*

$U = \{0.1, 0.25, 0.60, 0.80, 0.35, 0.55, 0.20, 0.50, 0.05, 0.95\}$

Now generate random samples from $U \sim U(0,1)$

Accept/Reject: $0.5 < 0.1$, $0.2 < 0.25$, $0.45 < 0.6$, $0.9 \geq 0.8$ ✓, $0.7 > 0.35$ ✓, $0.5 < 0.55$, $0.3 > 0.20$ ✓, $0.15 < 0.5$, $0.08 > 0.05$ ✓, $0.03 < 0.95$

Accepted $\theta^* = 0.35, 0.45, 0.65, 0.85 \rightarrow \hat{\theta}_{RS} = \frac{1}{N}\sum_{i=1}^{N}\theta_i^* = \frac{0.35+0.45+0.65+0.85}{4} = 0.575$

In this example, my method estimated a 57.5% probability of rolling a 6 after 10 trials & 3 6's. Bayesian inference using a Beta posterior gives a posterior mean of 33% in this case. As more samples are accepted, my rejection sampling method will approximate the Bayesian posterior probability distribution. This happens because my method does not explicitly compute the posterior but instead filters values based on their likelihood, removing unlikely candidates overtime. Since Bayesian updates using the likelihood function, the distribution of accepted samples will approximate the true Bayesian Posterior!

Thus, my idea of rejection sampling using MLE to discard invalid candidates provides an alternative belief-updating mechanism that aligns w/ Bayesian inference while maintaining an element of randomness due to rejection-sampling!

2. Done in Google Docs, submitted as separate pdf

3. $\Pr(X=x|\lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}$, $\lambda > 0$, $X = 0, 1, \ldots\ldots$

a) $L(\lambda) = \prod_{i=1}^{n} \dfrac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \dfrac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} \Longrightarrow \log L(\lambda) = \sum_{i=1}^{n} (x_i \log\lambda - \lambda - \log x_i!)$

do not depend on λ, so just a constant & can be ignored for MLE derivation

$\log L(\lambda) = \sum_{i=1}^{n} x_i \log\lambda - n\lambda$, $\dfrac{\delta}{\delta\lambda}\left(\sum_{i=1}^{n} x_i \log\lambda - n\lambda\right) = \dfrac{\sum_{i=1}^{n} x_i}{\lambda} - n \Rightarrow \dfrac{\sum_{i=1}^{n} x_i}{\lambda} = n \Longrightarrow$

$\dfrac{\lambda \cancel{n}}{n} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \lambda = \sum_{i=1}^{n} x_i / n = \overline{X}$, $\lambda_{MLE} = \overline{X}$

b) $p(\lambda|\alpha, \beta) = \dfrac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)}$, $\lambda, \alpha, \beta > 0$

$\dfrac{\beta^\alpha}{\Gamma(\alpha)}$: normalizing constant

from a, $L(\lambda) = \lambda^{\sum x_i} e^{-n\lambda}$

Using Bayes' Rule, $p(\lambda|X) \propto p(X|\lambda)p(\lambda)$, substituting likelihood & Gamma prior:

$P(\lambda|X) \propto \lambda^{\sum x_i} e^{-n\lambda} \cdot \lambda^{\alpha-1} e^{-\lambda\beta} \Rightarrow \lambda^{\sum x_i + \alpha - 1} e^{-(n+\beta)\lambda} \Rightarrow$

$P(\lambda|X) \propto \lambda^{\sum x_i - \alpha + 1} e^{-(n+\beta)\lambda}$

This is the kernel of a Gamma distribution, so $\lambda|X \sim \text{Gamma}(\alpha + \sum x_i, \beta + n)$ so

updated from $\text{Gamma}(\alpha, \beta) \Rightarrow \text{Gamma}(\alpha + \sum x_i, \beta + n)$

$E(\lambda) = \dfrac{\alpha}{\beta} \Rightarrow \boxed{E(\lambda|X) = \dfrac{\alpha + \sum x_i}{\beta + n}}$

c) Done in R ✓

d) Done in R ✓

4. $f(x|\eta) = \dfrac{\eta}{x^{\eta+1}}$, $x \geq 1$, $\eta > 0$

$f(\eta|\alpha, \beta) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} \eta^{\alpha-1} e^{-\beta\eta}$, $\eta, \alpha, \beta > 0$

a) $f(x|\eta) = \eta x^{-\eta-1}, x \geq 1, \eta > 0$, $L(\eta) = \prod_{i=1}^{n} \eta X_i^{-\eta-1} \rightarrow \eta^n \prod_{i=1}^{n} X_i^{-\eta-1}$: $U = \sum_{i=1}^{n} \log X_i$

$L(\eta) = \eta^n e^{-U\eta}$, multiply by Gamma prior $\Rightarrow$ $p(\eta|X) \propto L(\eta)p(\eta) \Rightarrow$

$p(\eta|x) \propto \eta^n e^{-U\eta} \cdot \eta^{\alpha-1} e^{-B\eta} \Rightarrow \eta^{n+\alpha-1} \cdot e^{-(U+B)\eta}$, this follows the Gamma distribution kernel

so $\eta|X \sim Gamma(\alpha+n, B+U)$

b) $E(\eta) = \frac{\alpha}{B} \Rightarrow \boxed{E(\eta|X) = \frac{\alpha+n}{B+U}}$

$L(\eta) = \eta^n e^{-U\eta} \rightarrow \log L(\eta) = n\log(\eta) - U\eta$, $\frac{\delta}{\delta\eta}(n\log\eta - U\eta) = \frac{n}{\eta} - U = 0 \Rightarrow$

$U = \frac{n}{\eta}$ $\boxed{\hat{\eta}_{MLE} = \frac{n}{U}}$

c) Done in R! ✓

d) Done in R! ✓