STAT400: Midterm Review

Fall 2022

Sampling

general sampling.

1. Are there specific characteristics we should look for in pdfs/cdfs that would help us choose which method to go about using? (① Inverse transform, ② transformation, accept reject, ③ ...) Can we maybe go through the most important concepts/rules that need to be met of each method.
   ① ? Write Cdf analytically? Invert it?
   ② ? Is my target dsn the dsn of a transformed rv (can sample from)?
   ③ can always make work
2. For all these different types of simulating, what do real-life examples look like? Sometimes, I am confused about where we are relative to our world and how it relates to what we are doing.

accept-reject

3. Is it possible for us to go over getting the acceptance and rejects and how to envelop exceeds target? Also show us more examples of this?
4. Can you define "g" again for acceptance-reject (is there a way to choose it)?
5. How do you know how to "pick" an envelope without looking at a graph—to ensure that it covers the entire support and the tails?

transform

6. The algorithm for transformation method at 3.1 I was wondering how to compute the G = g(X1,...., XP) drawn from g(X1,..., Xp).
7. For the method of the Transformation Methods(section 3), is it to "add on" something to the distribution of interest to "transform" it?

mixture.

8. Can you give an example of what a distributions of weighted sums would be like (compared to weighted sums of distributions)?
9. Mixture distribution definition how to use the 2 sources of variability?

Transformation method

We are interested in sampling from the dsn of $T = h(X_1,\dots,X_p)$, $X_1,\dots,X_p \sim f$

1. Sample $X_1,\dots,X_p \sim f$
2. $h(X_1,\dots,X_p)$ is one draw from dsn of T.
3. repeat 1.-2. until desired sample size is met.

Accept-Reject

goal: simulate from f with support $\mathcal{X}_f$

Algorithm:

1. find suitable g and $e = c \cdot g$
   - a) • g is a density ① easy to sample from, ② easy to evaluate $g(\,)$, ③ support of g includes $\mathcal{X}_f$.
   - b) • $e(x) = c \cdot g(x)$ where $e(x) \geq f(x)$ $\forall x$.
2. Sample $y \sim g$.
3. Sample $u \sim Unif(0,1)$.
4. If $u < f(y)/e(y)$, accept y.

repeat 2.–4. until we have desired sample size.

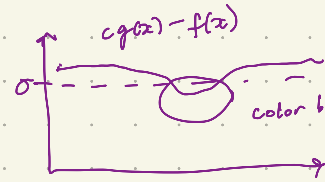
a) If support $\mathcal{X}_f$ is not finite
- Do not use Uniform!
- pick g from table of dsns → pay special attention to support of g relative to support $\mathcal{X}_f$
  → be careful w/ asymptotes.

b) How to find c?

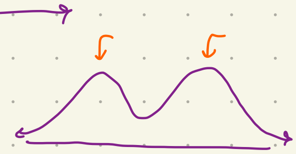
Need c s.t. $c \cdot g(x) \geq f(x)$ $\forall x$.

> try to solve for C: $c \geq \frac{f(x)}{g(x)}$ (this can be difficult).

> make some plots
- try wide values of x to get a sense of potential problem areas
- pick many c's, many x's, for each c:



color by if $cg(x) - f(x) < 0$.



Mixture dsns

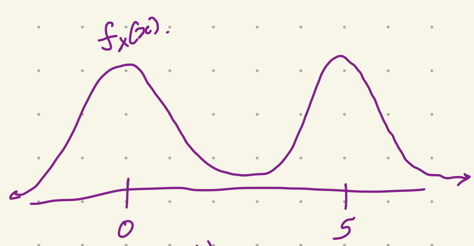
A random variable X has a mixture dsn if $f_X(x) = \sum_{i=1}^{p} \theta_i f_{X_i}(x)$ where $\sum_{i=1}^{p} \theta_i = 1$.

to sample from a two-component mixture dsn $f_X(x) = \theta f_{X_1}(x) + (1-\theta) f_{X_2}(x)$

1. Sample $Z \sim Bern(\theta)$
2. If $Z = 1$, $y \sim f_{X_1}$

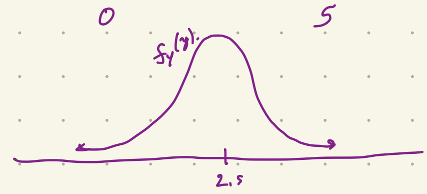
   If $Z = 0$, $y \sim f_{X_2}$

repeat 1–2 until desired sample size met.



let $X_1 \sim N(0,1)$, $X_2 \sim N(5,1)$ independent. If mixture $f_X(x) = \frac{1}{2} f_{X_1}(x) + \frac{1}{2} f_{X_2}(x)$

let $Y = \frac{1}{2}X_1 + \frac{1}{2}X_2$, $Y \sim N\left(2.5, \frac{1}{2}\right)$



Monte Carlo Integration

linear transformation

1. I am a little confused about Chapter 6 page 9. Could you talk about how do you structure the linear transformation? Thanks!

2. For the review on Tuesday, the something I would like practice / clarification on are transformation methods

3. For Monte Carlo Integration for example 1.6 how to get about another approach when we want to use unif(0,1) instead of other method we were given?

indicators

4. How does Method 3 of the Monte Carlo integration example 1.7 (with the indicator function) work?

est. $\theta = \int_A h(x)\,dx \overset{MC}{=} \int_{\mathcal{X}_f} g(x) f(x)\,dx$

Want: $\frac{h(x)}{\phi(x)}$ to be as flat as possible.

importance.

5. Is there a process for the best way to go about choosing a phi distribution or density for importance sampling? What rules or results are we trying to meet with choosing a phi that is the "best"? Is there like an "ideal" or "best" phi that will have the lowest variance/most constant ratio, being technically the "right" phi to use?

6. Going over "Your Turn Importance Sampling Example" since we ran out of time in class.

7. * My question for the review relates to importance sampling. In the class example, we biased the die by making 3 sides have value 1. However, this causes phi to not have positive probability everywhere f does, which violates one of the requirements stated later on. Is it okay to do this if the regions where f has positive probability and phi does not are not of interest to our estimation? Will we face this situation in the future or is it only a concern in our toy discrete example?

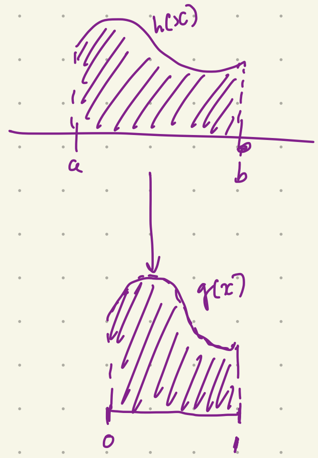
<u>really</u> only need ϕ support to include limits of integration.

MC.

8. I think the one thing I am most confused about is why monte carlo integration still works, even if sampling points does not follow a uniform distribution (for integrating on an interval from a to b). In the notes, it states that the only requirement is that the points are randomly drawn from a distribution with support over the range of integration. It seems like if the distribution we are drawing from is not uniform, then some parts of the curve would be under sampled, and other parts of the curve would be over sampled. I attached a picture to help explain

6-3.

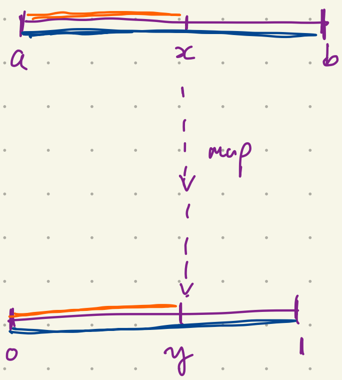
Want to estimate $\theta = \int_{\boxed{a}}^{\boxed{b}} h(x)\,dx$. $\quad a < b < \infty$, $\quad a \neq 0$, $b \neq 1$



Only know how to sample from Unif(0,1).

☆ We have to write θ as $E[g(X)]$, $X \sim f$.

For Unif(0,1), $\theta = \int_{\boxed{0}}^{\boxed{1}} g(x) \cdot 1\, dx$.



$$\frac{x-a}{b-a} = \frac{y-0}{1-0} = y.$$

$$x = y(b-a) + a$$

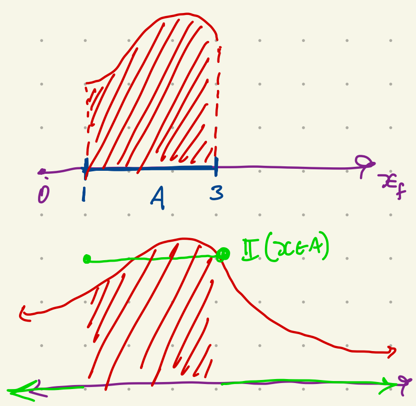
$$dx = (b-a)\,dy$$

$$\theta = \int_a^b h(x)\,dx \overset{\text{change of variable}}{=} \int_0^1 h(y(b-a)+a)\,(b-a)\,dy$$

$$= E\Big[h\big(Y(b-a)+a\big)(b-a)\Big], \quad Y \sim \text{Unif}(0,1).$$

Indicators:

Want $\theta = \int_A h(x)\,dx$, $\quad A \subseteq \mathcal{X}_f$ ← support of f.

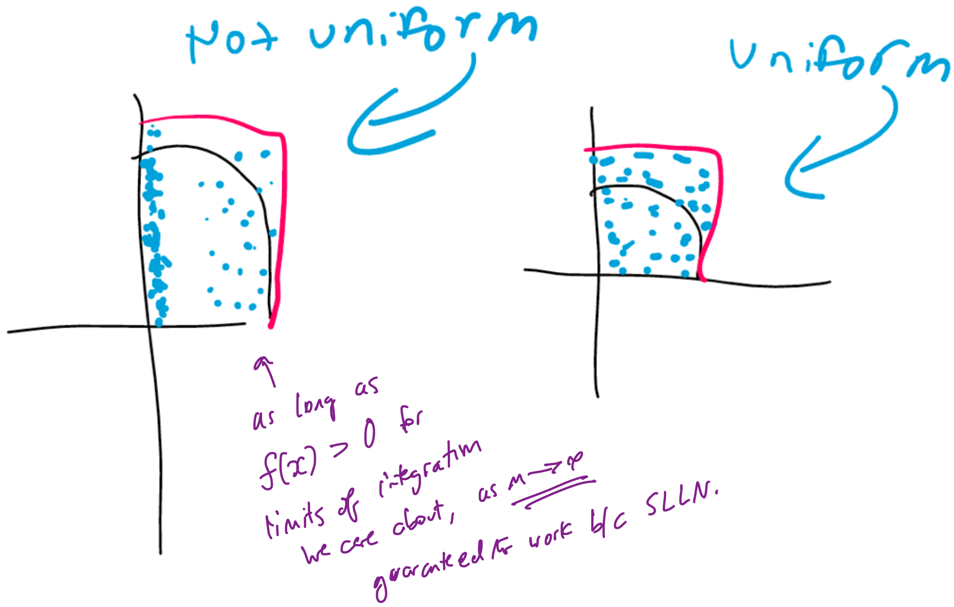


$$= \int_{\mathcal{X}_f} h(x)\,\mathbb{I}(x \in A)\,dx$$

$$= \int_{\mathcal{X}_f} h(x)\,\mathbb{I}(x \in A) \cdot \frac{f(x)}{f(x)}\,dx$$

$$= \int_{\mathcal{X}_f} \underbrace{\frac{h(x)}{f(x)}\,\mathbb{I}(x \in A)}_{g(x)} \cdot f(x)\,dx$$

$$= E\left[\frac{h(X)}{f(X)}\,\mathbb{I}(X \in A)\right], \quad X \sim f$$



In the picture, if we are integrating -x^2 + 4 (from 0 to 2), and the point of monte carlo is to get the ratio of: (dots inside the curve / total dots) * area of distribution (in pink), it seems like this would be affected by the type of distribution we are taking our samples from.

Does this not matter because we are taking enough samples?

as $m \to \infty$, $\frac{1}{m}\sum_{i=1}^{m} g(X_i) \xrightarrow{p} E[g(X)]$. (SLLN).

Other Questions

1. Does convergence almost surely include both convergence in probability and convergence in distribution? Or is there no overlap with those three?

$$\xrightarrow{as} \implies \xrightarrow{p} \implies \xrightarrow{d.}$$

Your Turn

We want to use the following distribution for inference, where we know the shape, but not the full distributional form.

$$f(x) = c\frac{\log(x)}{1+x^2}, \quad x \in [1, \pi]$$

What do we need for this to be a valid pdf?

We need $\int_1^{\pi} c \frac{\log(x)}{1+x^2}\, dx = 1.$

$\Rightarrow \int_1^{\pi} \frac{\log(x)}{1+x^2}\, dx = \frac{1}{c} \Rightarrow c = \frac{1}{\int_1^{\pi} \frac{\log(x)}{1+x^2}\, dx}.$

$f(x) = \frac{1}{\pi - 1}, x \in (1, \pi).$

(1) estimate c w/ MC integration.

a) Choose $f \sim Unif(1, \pi)$. (write integral as a expected value of function wrt $X \sim Unif(1,\pi)$).

$$\int_1^{\pi} \frac{\log(x)}{1+x^2}\, dx = \int_1^{\pi} \frac{\log(x)}{1+x^2} \cdot \frac{\pi - 1}{\pi - 1}\, dx = E\left[\frac{\log(X)}{1+X^2} \cdot (\pi - 1)\right], \quad X \sim Unif(1,\pi).$$

b) Plan:

1. Sample $X_1, \dots, X_m \sim Unif(1, \pi)$
2. $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m}\left[\frac{\log(X_i)}{1+X_i^2} \cdot (\pi - 1)\right]$

c) write in R

↳ easy to estimate $Var(\hat{\theta})$.

(2) could make better? Importance sampling.

(3) How to sample from the density: $f(x)$? Accept-reject!

choose $Unif[1, \pi)$ as proposal.