TOPIC:

DATE:

FILE UNDER:

PAGE:

Integration

We often need to compute:

① $\mathbb{E}_f\, g(X)$ — expectations / moments

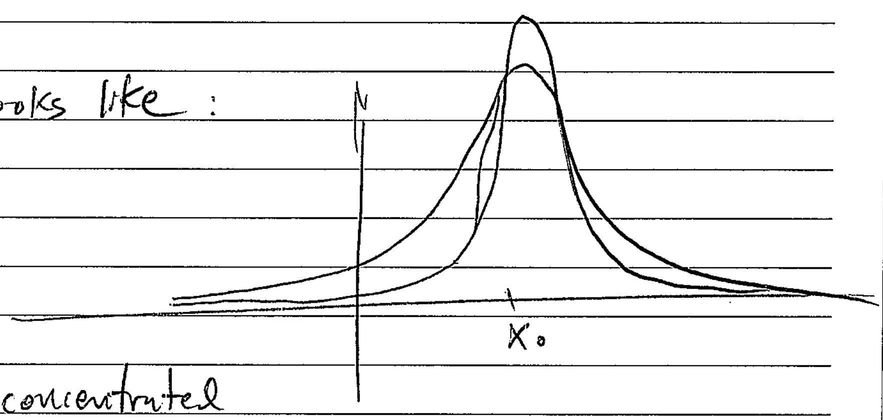
② $\int p(x|\theta)\,dx$ — normalizing constants for densities.

$\Rightarrow$ In general this can be very hard, but there are many special cases.

Suppose we want to calculate:

$$\int g(x)\,dx \qquad \text{and} \qquad g \in \mathcal{L}^2$$

Suppose g looks like:



So g is highly concentrated about x_0

Then we could say (?)

$$\cancel{\int g(x)\,dx \doteq g(x_0)\,\varepsilon} \quad ?$$

$$\int_{x-a}^{x+a} g(x) =$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Suppose we want

$$\int_a^b g(x)\,dx = \int_a^b e^{h(x)}\,dx \qquad \left[h(x) = \log g(x)\right]$$

Suppose h achieves its max at X_0

$$\doteq \int_a^b \exp\Big(h(x_0) + \underbrace{h'(x_0)(x-x_0)}_{=0} + \frac{h''(x_0)(x-x_0)^2}{2}\Big)dx$$

$$= \int_a^b \exp\Big(h(x_0) + \frac{h''(x_0)(x-x_0)^2}{2}\Big)dx$$

$$= \exp(h(x_0))\int_a^b \exp\Big(\frac{h''(x_0)}{2}(x-x_0)^2\Big)dx$$

$$= \exp(h(x_0))\int_a^b \underbrace{\exp\Big(-\frac{1}{2}\,\frac{(x-x_0)^2}{-h''(x_0)^{-1}}\Big)}_{N(x_0,\,-h''(x_0)^{-1})\ \text{density}}dx$$

$$= \underbrace{\exp(h(x_0))}_{=g(x_0)}\left[\frac{\sqrt{2\pi}}{-h''(x_0)^{1/2}}\right]\underbrace{\left[\Phi(b\,|\,x_0, -h''(x_0)^{-1}) - \Phi(a\,|\,x_0, -h''(x_0)^{-1})\right]}_{=1 \text{ if } b=\infty,\ a=-\infty}$$

$$h''(x) = \left(\frac{g'}{g}\right)'$$

$$\frac{gg'' - (g')^2}{g^2}$$

$$\frac{g''}{g} - \frac{g'}{g}\frac{g'}{g}$$

$$g \qquad \frac{\quad}{g'' - \frac{g'g'}{g}}$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

$$\cancel{\mathbb{E}[\theta] = \frac{\int \theta f(y|\theta)\pi(\theta)}{\int}}$$

$$\mathbb{E}\theta = \int \theta \, p(\theta|y)\, d\theta$$

$$= \frac{\int \theta \, f(y|\theta)\pi(\theta)\, d\theta}{\int f(y|\theta)\pi(\theta)\, d\theta}$$

$$= \frac{\int \theta \exp(\log f(y|\theta)\pi(\theta))}{\int \exp(\underbrace{\log f(y|\theta)\pi(\theta)}_{h(\theta|y)})}$$

Need $\hat{\theta}$

$$\doteq \frac{\int \theta \exp\left(h(\hat\theta|y) + h''(\hat\theta|y)(\theta-\hat\theta)^2/2\right) d\theta}{\int \exp\left(h(\hat\theta|y) + h''(\hat\theta|y)(\theta-\hat\theta)^2/2\right) d\theta}$$

$$= \frac{\int \theta \exp\left(\frac{h''(\hat\theta|y)}{2}(\theta-\hat\theta)^2\right) d\theta}{\int \exp\left(\frac{h''(\hat\theta)}{2}(\theta-\hat\theta)^2\right) d\theta}$$

Laplace approx to posterior mean is the posterior mode!

$$= \frac{\int \theta \frac{\sqrt{2\pi}}{-h''(\hat\theta)^{1/2}} \varphi(\theta|\hat\theta, -h''(\hat\theta)^{-1})\, d\theta}{\int \frac{\sqrt{2\pi}}{-h''(\hat\theta)^{1/2}} \varphi(\theta|\hat\theta, -h''(\hat\theta)^{-1})\, d\theta}$$

$$= \hat\theta$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

See Lange Chap 4.6 for Laplace Approx.

Monte Carlo

Suppose we want to compute ~~$\mathbb{E}_f$~~ for $h : \mathbb{R}^k \to \mathbb{R}$

$$\mathbb{E}_f[h(X)] = \int h(x) f(x)\, dx$$

If we can simulate $X_1, \ldots, X_n \overset{iid}{\sim} f$, then by the LLN,

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} h(X_i)}_{\text{sum}} \longrightarrow \underbrace{\mathbb{E}_f[h(X)]}_{\text{integral}}$$

Furthermore, $Var\{ = \frac{1}{n^2} \sum Var(h(X_i))$

$$\cancel{SD\left(\frac{1}{n}\sum h(X_i)\right) = \frac{1}{n}\sqrt{\sum_{i=1}^{n} Var(h(X_i))}}$$

Notice that the variance doesn't depend on dimension of X. That is, $Var \sim \frac{1}{n}$. This is very important, both good and bad.

Applications

1. Monte Carlo / Simulation studies. We have a method that estimates a parameter $h(X) = \hat{\beta}$ and we want to explore performance of $\hat{\beta}$.

2. $\mathbb{E}_f h(X)$ might be a posterior mean (f is post. dens)

3. Might want $\int h(x)\, dx = \int \frac{h(x)}{f(x)} f(x)\, dx$.

$$\approx \frac{1}{n} \sum \frac{h(X_i)}{f(X_i)}, \quad X_1, \ldots, X_n \sim f$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

We need to be able to simulate numbers, i.e. from some dist. f

Random Number Generation

Most popular (and simplest) are linear congruential generators.

Let X_0 be some starting value, called the "seed". Then generate a sequence (for $n = 0, 1, 2, \ldots$)

$$X_{n+1} = (a X_n + b) \bmod m$$

a = multiplier
b = increment
m = modulus

For uniform RN, just let $U_{n+1} = \frac{X_{n+1}}{m}$

Ideally, $X_1, X_2, \ldots$ will hit every number from 0 to $m-1$ before repeating

# of steps until repeat is "period".
A "maximal period generator" has period m.

Setting a, b, and m is a very tricky business.

For example, this is bad

$$X_{n+1} = (2 X_n + 0) \bmod 2^{32}$$

One good set is

$$a = 106, \quad b = 1283, \quad m = 6075$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

There are quadratic and cubic ~~ge~~ RNGs, but they require more work and not much better, i.e.

$$X_{n+1} = (aX_n^2 + bX_n + c) \mod m$$

LCG are NOT useful for things like cryptography. Any polynomial generator can be broken.

RNG are critical for stream ciphers which ~~[illegible]~~ use linear feedback shift registers (LFSR) "bread + butter"

All PRNGs produce deterministic sequences that "look" random

One can check "randomness" of sequence with a test for uniformity

- Kolmogorov-Smirnov test
- Chi-square test

Marsaglia has "die hard" tests

Also, many PRNGs generate good 1-D sequences but do not look random in > 1 dimension

Its may be stuck on a hyperplane.

TOPIC: | DATE:

FILE UNDER: | PAGE:

Uniforms are good but we need values from some density f. First, assume you can simulate Unif(0,1)

(*) Integral transform / rescaling

(1) $U \sim \text{Unif}(0,1)$

(2) $X \sim F^{-1}(U)$, where $F(x) = \int_{-\infty}^{x} f(t)\,dt$ cdf.

(*) Transformation

(1) $Z_1 = \sqrt{-2\log U_1}\,\cos(2\pi U_2)$ iid $\sim N(0,1)$

$Z_2 = \sqrt{-2\log U_1}\,\sin(2\pi U_2)$

(2) ~~Beta(α, β)~~ $\frac{X}{X+Y} \sim \text{Beta}(\alpha,\beta)$

where $X \sim \text{gamma}(\alpha,1)$
$Y \sim \text{gamma}(\beta,1)$

(3) $\mu \sim \text{gamma}(\alpha,\beta)$
$Y|\mu \sim \text{Poisson}(\mu)$

$Y \sim$ neg. binomial

Most standard dist. are implemented in R.

$$p(y) = \int_0^\infty \underbrace{p(y|\mu)}_{\text{Poisson}}\ \underbrace{p(\mu)}_{\text{gamma}}\,d\mu$$

exponential(1)

$f(x) = e^{-x/\lambda}$

$F(x) = \frac{1}{\lambda}\lambda(e^{-x/\lambda} + 1)$

$F(x) = 1 - e^{-x/\lambda}$

$u = 1 - e^{-x/\lambda}$

$-\log(u-1)$

$F^{-1}(u) =$

$F^{-1}(u) = -\log(1-u)$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Multivariate Normal

Want $X \sim N(\mu, \Sigma)$

Let $\Sigma = L^T L$ (cholesky decomp)

(1) Simulate $Z \sim N(0, I)$
(2) $X = \mu + WZ \sim N(\mu, \Sigma)$

~~Sequential sampling~~

Rejection sampling

A way of generating samples/obs. from f by thinning out obs/samples from a candidate density g. ("Random thinning")

Suppose f is our target density and we can evaluate it.

Let g be our candidate density that we can simulate from.

Let $\mathcal{X}_f$ be the support of f and $\mathcal{X}_g$ be support of g and assume $\mathcal{X}_f \subset \mathcal{X}_g$

Assume $c = \sup_{x \in \mathcal{X}_f} \frac{f(x)}{g(x)} < \infty$

and that we can calculate c

TOPIC:

DATE:

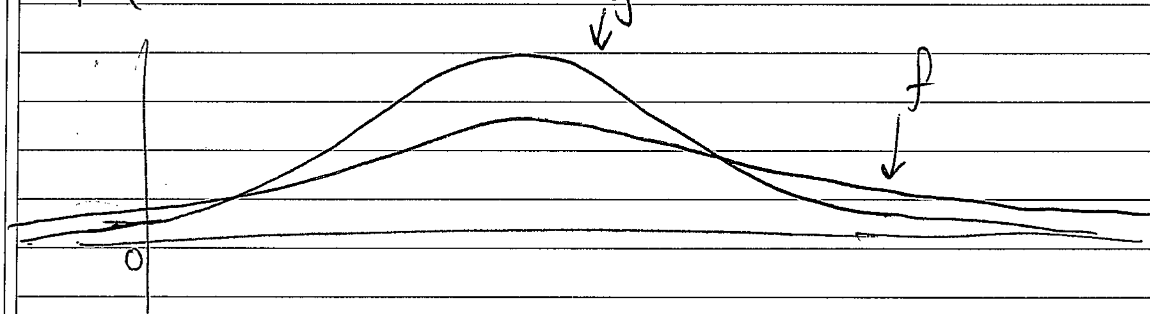
FILE UNDER:

PAGE:

(4) The higher the dimension of f, g the more inefficient rejection sampling will be.

(5) Whether $c = \infty$ depends (usually) on the tails of the candidate, which must be heavier than the target

Bad:



As $x \to \infty$, $g(x) \downarrow 0$ faster than $f(x) \downarrow 0$.

$\Rightarrow f(x)/g(x) \uparrow \infty$.

$\Rightarrow$ What if we cannot calculate $c = \sup \frac{f(x)}{g(x)}$?

Can we estimate c? Yes!

Empirical sup rejection sampling (Caffo '02)

$h(x) = \frac{f(x)}{g(x)}$

(1) Guess $\hat{c} > 1$

(2) Generate $U \sim \text{uniform}(0,1)$, $X \sim g$

$h(x_n), h(x_{n+1})$

$x_{n+1} = x_n - \alpha h'(x_n)$

(3) Accept X if $U \le \frac{f(x)}{\hat{c}\, g(x)}$

(4) Update $\hat{c}^* = \max\left\{\hat{c}, \frac{f(x)}{g(x)}\right\}$ $\Longleftrightarrow$ sup estimator

(5) Goto step (2), set $\hat{c} = \hat{c}^*$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Rejection Sampling

1. Simulate $u \sim \text{Uniform}(0,1)$
2. Simulate $X \sim g(x)$
3. If $u \le \frac{f(X)}{cg(X)}$, "accept" X

   otherwise, ~~go to~~ discard X

   go back to 1.

Let $X_1, X_2, \ldots \overset{iid}{\sim} g$

$$\underbrace{\underbrace{\underset{\text{reject}}{\overset{X_1}{0}} \quad \underset{\text{reject}}{\overset{X_2}{0}} \quad \underset{\text{accept}}{\overset{X_3}{1}}}_{\text{geometric}(p)} \quad \overset{X_4}{0} \quad \underset{\text{accept}}{\overset{X_5}{1}}}_{\text{iid coin flips}} \quad \overset{X_6}{0} \quad \overset{X_7}{0} \quad \cdots$$

sample size of 2.

# of flips until acceptance is geometric w/ success probability $1/c$.

Pf: $\mathbb{P}(X \text{ accepted}) = \mathbb{P}\left(U \le \frac{f(X)}{cg(X)}\right)$

$$= \int \mathbb{P}\left(U \le \frac{f(x)}{cg(x)} \,\Big|\, X = x\right) g(x)\, dx$$

$$= \int \frac{f(x)}{cg(x)}\, g(x)\, dx$$

$$= 1/c$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

The dist. of accepted values is f.

$$\mathbb{P}(X \le t \mid X \text{ accepted})$$

$$= \frac{\mathbb{P}(X \le t,\ X \text{ accepted})}{\mathbb{P}(X \text{ accepted})}$$

$$= \frac{\mathbb{E}_g\left[\mathbb{1}\{X \le t\}\,\mathbb{1}\{X \text{ accepted}\}\right]}{1/c}$$

$$= c\,\mathbb{E}_g \mathbb{E}\left[\mathbb{1}\{X \le t\}\,\mathbb{1}\left\{u \le \frac{f(x)}{c g(x)}\right\} \Big| X\right]$$

$$= c\,\mathbb{E}_g\left(\mathbb{1}\{X \le t\}\frac{f(x)}{c g(x)}\right)$$

$$= c \int \mathbb{1}\{x \le t\} \frac{f(x)}{c g(x)}\, g(x)\, dx$$

$$= \int_{-\infty}^{t} f(x)\,dx = F(t)$$

Notes:

1. Only need to ~~know~~ know f or g up to a constant of proportionality
2. Any number $c' > c$ will work, but will be less efficient.
3. Operations can (and should) be performed on log scale

TOPIC: | DATE:

FILE UNDER: | PAGE:

ESUP

Extra assumption is need for ~~ESUS~~:

$c = \frac{f(x_e)}{g(x_e)}$ for some $x_e \in \mathcal{X}_f$

& (sup is achievable)

This is satisfied if g has heavier tails than f.

~~Caffo proved that If g is discrete~~

let $X_1, X_2, X_3, \_\_\_ \overset{iid}{\sim} g$

	X_1	X_2	X_3	X_4	X_5	X_6 ......
True c $\tilde{Y}_i$	0	0	1 accept	0	1 accept	0
$\hat{c}$ Y_i	1 accept error	0	1 accept	1 accept error	1 accept	0

Caffo showed that

① $\mathbb{P}(Y_i \neq \tilde{Y}_i \text{ i.o.}) = 0$ if f is discrete

Pf: By assumption 3, $\exists\, x_e \in \mathcal{X}_f$ s.t. $c = \frac{f(x_e)}{g(x_e)}$.

Let $\gamma = \min_i \{X_i = x_e\}$ where $X_i \sim g$.

Then $\gamma \sim \text{geometric}(g(x_c))$. Once $\hat{c} = c$, algorithms are the same.

Coupling lemma $\Longrightarrow$ $\mathbb{P}(Y_i \neq \tilde{Y}_i) \leq \mathbb{P}(\gamma \geq i) = (1 - g(x_c))^{i-1}$

$\Rightarrow \sum_{i=1}^{\infty} \mathbb{P}(Y_i \neq \tilde{Y}_i) < \infty$

TOPIC:

DATE:

FILE UNDER:

PAGE:

$\sum_{n=1}^{\infty} \frac{1}{n^2}$

$\int_1^{\infty} x^{-2}$

$= -\frac{1}{x}\Big|_1^{\infty}$

$= 1 - 0 = \ldots$

(2) For continuous f, it's trickier. In general,

$$\mathbb{P}(Y_i \neq \tilde{Y}_i) = O(i^{-1})$$

But if $\log(f/g)$ is smooth (twice differentiable) and unimodal (w/mode at x_c), then

$$\mathbb{P}(Y_i \neq \tilde{Y}_i) = O(i^{-2})$$

Suggests a "burn-in", ~~of~~ discard first several variates.

More on burn-in with MCMC

~~Ex~~ Often we want to simulate from

$$p(\theta|y) \propto \underbrace{f(y|\theta)}_{\text{likelihood}} \underbrace{\pi(\theta)}_{\text{prior}}$$

$\Rightarrow$ use π as the candidate dist.

$$C = \sup_{\theta} \frac{p(\theta|y)}{\pi(\theta)} = \sup_{\theta} f(y|\theta) = f(y|\underbrace{\hat{\theta}}_{\text{MLE}})$$

~~posterior mode~~

Rejection sampling is

(1) Simulate $U \sim \text{uniform}(0,1)$

(2) Simulate $\theta \sim \pi$

(3) Accept θ if $U \leq \frac{p(\theta|y)}{C\pi(\theta)} = \frac{f(y|\theta)}{f(y|\hat{\theta})}$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Ex: $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$

Prior: $\pi(\mu) = N(0, \underbrace{B}_{\to \text{BIG}})$

$$\pi(\sigma^2) = IG(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} e^{-\beta/z} = f(z)$$

MLE: $\hat{\mu} = \bar{Y}$, $\hat{\sigma}^2 = S^2 = \frac{1}{n}\sum(Y_i - \bar{Y})^2$

$$C = f(y|\hat{\mu}, \hat{\sigma}^2) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}^2} e^{-\frac{1}{2\hat{\sigma}^2}\sum(Y_i - \hat{\mu})^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{\sum(Y_i - \bar{Y})^2}{n}\right)^{-n/2} e^{-\frac{n}{2}}$$

(1) Simulate $U \sim \text{Uniform}(0,1)$

(2) Simulate $\mu \sim N(0, B)$, $\frac{1}{\sigma^2} \sim \text{gamma}(\alpha, \beta)$

(3) Accept (μ, σ^2) if

$$U \le \frac{f(y|\mu, \sigma^2)}{f(y|\hat{\mu}, \hat{\sigma}^2)} = \left(\frac{\sum(Y_i - \bar{Y})^2}{n\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum(Y_i - \mu)^2 + \frac{n}{2}\right)$$

Given a sample from $p(\theta|y)$ we can compute confidence/posterior intervals, mean, variance, etc.

What does ESUP do? It searches for MLE by random search.