ST4231 Cheat-sheet

Basics relationships

$$\begin{aligned}\mathrm{Var}[X] &= E[(X-E[X])^2]\\ &= E[X^2]-E[X]^2\\ \mathrm{Cov}[X,Y] &= E[(X-E[X])(Y-E[Y])]\\ &= E[XY]-E[X]E[Y]\end{aligned}$$

Inversion method

$$T(X) = F^{(-1)}(U) \Rightarrow T(X) \sim f(x)$$

Generalized inverse

$$F^{(-1)}(U) = \inf\{z \in \mathcal{R} : F(z) \geq u\}$$

Change-of-variable formula

Suppose $g(x)$ is one-to-one and $\mathcal{C}^0$, $X \sim f_X(x)$, $Y = g(X)$, then

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right|$$

Central Limit Theorem

$E[X] < \infty$ then the follwing convergence in distribution holds

$$\lim_{n\to\infty} \sqrt{n}(\bar{X}_n - \mu) = \mathcal{N}(0,\sigma^2)$$

Fundamantal theorem of sampling

If X is a random variable with pdf $f(x)$, then simulating X is equivalent to simulating a pair of variables (U, X) jointly from

$$(X,U) \sim \mathrm{Uniform}\{(x,u) : 0 < u < f(x)\}$$

Rejection Sampling Algorithm

Suppose $f(x) = c\tilde{f(x)}$ where $\tilde{f(x)}$ is known and c is not. $\tilde{f}(x) \leq \tilde{M}g(x)\ \forall x$

1. Generate $Y \sim G$
2. Generate $U \sim \mathrm{Uniform}[0,1]$
3. If $U \leq \frac{\tilde{f}(Y)}{\tilde{M}g(Y)}$, then accept: set $X = Y$ Otherwise reject: return to step (1)

Basic distributions

Uniform

$f(x;a,b) = \begin{cases}\frac{1}{b-a}, & \text{if } 0 \leq x \leq b\\ 0 & \text{otherwise}\end{cases}$

$E[X] = \frac{a+b}{2}\ Var[X] = \frac{(b-a)^2}{12}$

Binomial

$f(x;n,p) = \binom{n}{x}p^x(1-p)^{n-x}\ x = 0,1\ldots,n$

$E[X] = np\ Var[X] = np(1-p)$

Geometric

$f(x;p) = p(1-p)^{x-1},\ E[X] = \frac{1}{p}\ Var[X] = \frac{1-p}{p^2}$

Poisson

$f(x) = \frac{\mu^x}{x!}e^{-\mu},\ x = 0,1,2,\ldots$

$E[X] = \mu,\ \mathrm{Var}[X] = \mu$

Negative Binomial

k failures, given r successes

$f(k;r,p) = \binom{k+r-1}{k}(1-p)^k p^r,\ x = 0,1,2\ldots$

$E[X] = \frac{r(1-p)}{p},\ Var[X] = \frac{r(1-p)}{p^2}$

Multinomial

$f(x_1,\ldots,x_k;p_1,\ldots,p_k,n) = \frac{n!}{x_1!\ldots x_k}p_1^{x_1}\ldots p_k^{x_k}$

$E[X_i] = np_i,\ \mathrm{Var}[X_i] = np_i(1-p_i)$

$\mathrm{Cov}(X_i,X_j) = -np_ip_j$

Exponential distribution

$\lambda e^{-\lambda x},\ x \geq 0,\ E[X] = \frac{1}{\lambda},\ \mathrm{Var}[\frac{1}{\lambda^2}]$

$\mathcal{X}^2$ distribution

$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{\nu/2-1}e^{-x/2},\ x \geq 0,\ \nu = 1,2,\ldots$

$E[X] = \nu,\ \mathrm{Var}[X] = 2\nu$

Gamma

$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-x\beta},\ x \geq 0$

$E[X] = \alpha\beta, \mathrm{Var} = \alpha\beta^2$

Weibull

$f(x) = \alpha\beta x^{\beta-1}e^{-\alpha x^\beta},\ x \geq 0$

Beta

$f(x) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1},\ 0 \leq x \leq 1$

$E[X] = \frac{\alpha}{\alpha+\beta},\ \mathrm{Var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Normal

$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ -\infty < x < \infty$

T-distribution

$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}},\ -\infty < x < \infty$

$E[X] = 0,\ \nu \geq 2\ \mathrm{Var}[X] = \frac{\nu}{\nu-2}$

Cauchy

$f(x) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]},\ -\infty < x < \infty$

Rayleigth

$f(x) = \frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}},\ x \geq 0$

$E[X] = \sigma\sqrt{\frac{\pi}{2}},\ \mathrm{Var}[X] = \frac{4-\pi}{2}\sigma^2$

Box-Muller v1

1. $U_1, U_2 \overset{\text{i.i.d.}}{\sim} \mathrm{Uniform}(0,1)$
2. $R = \sqrt{-\ln(U_1)}$, $\theta = 2\pi U_2$
3. $X = R\cos(\theta)$, $Y = R\sin(\theta)$

Box-Muller v2

1. $U_1, U_2 \overset{\text{i.i.d.}}{\sim} \mathrm{Uniform}(0,1)$
2. $V_1 = 2U_1 - 1$,$V_2 = 2U_2 - 1$,$S = V_1^2 + V_2^2$
3. If $S > 1$return to step 1(Rejection sampling)
4. Return $X = \sqrt{-2\ln(S)/S}\cdot V_1$, $Y = \sqrt{-2\ln(S)/S}\cdot V_2$

General Multivariate Normal

To generate d-dimentional normal with mean μ and covariance matrix Σ:

1. Generate $\mathbf{Z} = (Z_1,\ldots,Z_d)^\top$
2. Set $X = \mathbf{LZ} + \boldsymbol{\mu}$

Where $\mathbf{L}$ satisfies $\mathbf{LL}^\top = \mathbf{\Sigma}$(Cholesky)

Simple sampling

Estimate $\theta = E[\phi(X)]$

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^n \phi(X_i),\ \mathrm{Var}(\theta) = \frac{\overbrace{\int_{\mathcal{S}} \phi(x)f(x)dx - \theta^2}^{\text{asymp. variance}}}{n}$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n \phi^2(X_i) - \hat{\theta}^2,\ \hat{\theta} \pm 1.96\sqrt{\frac{\hat{\sigma^2}}{n}}$$

Importance sampling

$$\hat{\theta}_{IS} = \frac{1}{n}\sum_{i=1}^n \frac{\phi(X_i)f(Y_i)}{g(Y_i)}$$

$$\mathrm{Var}(\theta) = \frac{\overbrace{\int_{\mathcal{S}} \frac{\phi^2(x)f^2(x)}{g(x)}dx - \theta^2}^{\text{asymp. variance}}}{n}$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n \phi^2(X_i)\frac{f^2(X_i)}{g^2(X_i)} - \hat{\theta}^2,\ \hat{\theta} \pm 1.96\sqrt{\frac{\hat{\sigma^2}}{n}}$$

Optimal $g(x)$: $g(x) \propto |\phi(x)| \cdot f(x)$

Self-normalized Importance Sampling

$\tilde{w}(x) = \frac{\tilde{f}(x)}{\tilde{g}(x)}\ \forall x \in \mathcal{S}$, $\hat{\theta}_{SIS} = \frac{\sum_{i=1}^n \phi(X_i)\tilde{w}(X_i)}{\sum_i^n \tilde{w}(X_i)}$

$E_f\left[\hat{\theta}_{SIS}\right] \neq \theta$, $\mathrm{bias}(\hat{\theta}_{SIS}) = \mathcal{O}(1/n)$,

$\mathrm{fluctuation}(\hat{\theta}_{SIS}) = \mathcal{O}(1/\sqrt{n})$

$\frac{\hat{\sigma}^2_{SIS}}{n} = \frac{\sum_{i=1}^n\left\{\tilde{w}^2(X_i)\left[\phi(X_i)-\hat{\theta}_{SIS}\right]^2\right\}}{\left\{\sum_{i=1}^n \tilde{w}(X_i)\right\}^2}$

95% asymp. conf. interval: $\hat{\theta}_{SIS} \pm 1.96\sqrt{\frac{\hat{\sigma}^2_{SIS}}{n}}$

Control Variates

Suppose we know:

1. an unbiased estimator $\hat{h}$ of $E[h(X)]$
2. $E_f[h(X)]$ and $\mathrm{Var}[\hat{h}]$
3. the value or sign of $\mathrm{Cov}(th\hat{e}ta, \hat{h})$

Let $\tilde{\theta} = \hat{\theta} + \beta\left\{\hat{h} - E_f[h(X)]\right\}$, then

$\mathrm{Var}(\tilde{\theta}) = \mathrm{Var}(\hat{\theta}) + \beta^2\mathrm{Var}(\hat{h}) + 2\beta\mathrm{Cov}(\hat{\theta},\hat{h})$

which is minimized when $\beta = -\frac{\mathrm{Cov}(\hat{\theta},\hat{h})}{\mathrm{Var}(\hat{h})}$

The corresponding smallest value is

$\mathrm{Var}(\tilde{\theta}) = (1-\rho^2_{\hat{\theta},\hat{h}})\mathrm{Var}(\hat{\theta}),\ \rho^2_{\hat{\theta},\hat{h}} = \mathrm{Cor}(\hat{\theta},\hat{h})$

Antithetic Variates Method

If $g(x)$ is a monotone function then

$[g(u_1) - g(u_2)][g(1-u_1) - g(1-u_2)] \leq 0$

From this we can show that if

$X = F^{-1}(U_1),\ X' = F^{-1}(1-U_1)$

Then $2\mathrm{Cov}(X,X') \leq 0$ which in turn implies that $\mathrm{Var}\left(\frac{X+X'}{2}\right) \leq \frac{1}{2}\mathrm{Var}(X)$

Rao-Blackwellization

$\hat{\theta}_{RB} = \frac{1}{N}\sum_{i=1}^N E[\phi(X_i)|Y = Y_i]$

$\hat{\theta}_{RB}$ is unbiased, and reduces the variance compared to simple sampling, by the law of total variance:

$\mathrm{Var}[\hat{\theta}] = \frac{1}{N}\mathrm{Var}[\phi(X)]$

$= \frac{1}{N}\left(\mathrm{Var}[E[\phi(X)|Y]] + E[\mathrm{Var}[\phi(X)|Y]]\right)$

$\geq \frac{1}{N}\mathrm{Var}[E[\phi(X)|Y]]$

EM-Algorithm For latent variable model:

E-step:

$Q(\theta|\theta^{(k)}) = E_Z[l^C(\theta;Y,Z)|Y,\theta^{(k)}]$

M-step:

$\theta^{(k+1)} = \arg\max_{\theta\in\Theta} Q(\theta|\theta^{(k)})$

Metropolis-Hastings Algorithm

1. Set $\theta^{(0)}$ to some initial value
2. **for** $t = 0$ to $T-1$ **do**:
   (a) Generate θ^* from $q(\theta^*|\theta^{(t)})$
   (b) Compute the acceptance probability: $\alpha(\theta^*, \theta^{(t)}) = \min\left\{1, \frac{p(Y|\theta^*)\cdot\pi(\theta^*)\cdot Q(\theta^*,\theta^{(t)})}{p(Y|\theta^{(t)})\cdot\pi(\theta^{(t)})\cdot Q(\theta^{(t)},\theta^*)}\right\}$
3. Generate $U \sim \text{Uniform}(0,1)$
4. If $U \leq \alpha(\theta^*, \theta^{(t)})$ then $\theta^{(t+1)} = \theta^*$
5. Otherwise $\theta^{(t+1)} = \theta^{(t)}$

Tuning MCMC algorithms:

1. Better to transform all parameters in θ such that they lie in $\mathbb{R}$ unbounded if using Normal transition kernel.
2. We can first maximize the log posterior $\log\pi(\theta|Y)$, if possible, and find the maximizer $\hat{\theta}$. Use $\hat{\theta}$ as $\theta^{(0)}$.
3. Normal kernel $\Rightarrow$ optimal acceptance rate of a random walk Metropolis algorithm is 0.234
4. Achieved by setting $q(\theta|\theta_a) = N(\theta_a, c^2\Sigma)$,where $\Sigma = (-H)^{-1}$ (H is the Hessian matrix), and $c = 2.4/\sqrt{d}$(d is the dimension of θ)
5. Burn-in: discard the first B iterations
6. Thinning: keep every Tth iteration

Gibbs-sampler

1. Initialize $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_d^{(0)})^\top$
2. **for** $t = 0$ to $T-1$ **do**:
   (a) Sample $\theta_1^{(t+1)}$ from $p(\theta_1|\theta_2^{(t)}, \ldots, \theta_d^{(t)}, Y)$
   (b) Sample $\theta_2^{(t+1)}$ from $p(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_d^{(t)}, Y)$
   (c) $\vdots$
   (d) Sample $\theta_d^{(t+1)}$ from $p(\theta_d|\theta_1^{(t+1)}, \ldots, \theta_{d-1}^{(t+1)}, Y)$
3. Set $\theta^{(t+1)} = (\theta_1^{(t+1)}, \ldots, \theta_d^{(t+1)})$

Markov chains

A Markov chain is *irreducible* if all states consist of a single class. Meaning all states are accessible from each other.
An *irreducible* Markov chain X is recurrent if $P[\tau_{ii} < \infty] = 1$ for all states, where $\tau_i i = \min\{t > 0 : X_t = i | X_0 = i\}$
An *irreducible recurrent* Markov chain X is *positive recurrent* if $E[\tau_{ii}] < \infty$ for all states. Otherwise it is *null recurrent.*
If a Markov chain only has a finite number of states, and is *irreducible* then it must be positive recurrent.
positive reccurent $\Leftrightarrow$ there exists stationary pmf $\boldsymbol{\pi}(\cdot)$.
An *irreducible* chain in called *aperiodic* if for some and hence all i, the greatest common divider of $\{t : p_{ii} > 0\} = 1$
The *stationary distribution* satisfies $\boldsymbol{\pi}P = \boldsymbol{\pi}$
If a Markov chain is *irreducible* and *aperiodic*, then it has a unique *stationary distribution*
A *closed class* is one that is impossible to leave, so $p_{ij} = 0$, if $i \in C,\ j \notin C$
A Markov chain is *reversible* if it satisfies Kolmogorov's criterion:
$\pi_i p_{ij} = \pi_j p_{ji},\ \forall i, j$

Convergence Theorem

Let $X = \{X_1, X_2, \ldots\}$ be a *stationary* & *reccurent* Markov chain with the transition matrix P and transition probabilities p_{ij} from any state i to state j. Then

1. the stationary distribution is the unique distribution satisfying $\sum_i \pi_i p_{ij}(t) = \pi_i \forall j \forall t \geq 0$
2. if $E_\pi[|h(X)|] < \infty$ then $\lim_{N\to\infty} N^{-1}\sum_{k=1}^N h(X_K) = E_\pi[h(X)]$
3. if $E_\pi[\phi^2(X)] < \infty$ then $\lim_{T\to\infty}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \phi(X_t) - E_\pi[\phi(X)]\right) \to \mathcal{N}(0,\sigma^2)$
4. if X is *aperiodic* then $p_{ij}(t) \to \pi_j$ as $t \to \infty$ for all i, j

Detailed balance condition

Holds if the Markov chain is *reversible*
$\exists x_j, j = 1, 2, \ldots, K : x_i p_{ij} = x_j p_{ji} \forall i \neq j, \sum_{j=1}^K x_j = 1 \Rightarrow \pi_j \propto x_j$

Common problems

Inversion method

$U \sim \text{Uniform}(0,1)$
$X = -\frac{1}{\lambda}\ln(U) \Rightarrow X \sim \text{Exp}(\lambda)$
$X = (-\frac{1}{\alpha}\ln(U))^{1/p},\ X \sim \text{Weibull}(\alpha,\beta)$
$X = \gamma\tan\left[\pi(U - \frac{1}{2})\right] + x_0,\ X \sim \text{Cauchy}(x_0,\gamma)$
$X = -\sqrt{2\sigma^2\ln(U)},\ X \sim \text{Rayleigth}(\sigma^2)$

Other methods

$Y_1, \ldots Y_\alpha \overset{\text{i.i.d}}{\sim} \text{Exp}(\beta),\ \Rightarrow \sum Y_i \sim \text{Gamma}(\alpha,\beta)$
$U_1, \ldots, U_{\beta+\alpha-1} \overset{\text{i.i.d}}{\sim} \text{Unif}(0,1) \Rightarrow U_{(\alpha)} \sim \text{Beta}(\alpha,\beta)$

EM-algorithm

Mixture of normals:
$y_i \sim p\mathcal{N}(\mu_1,\sigma_1^2) + (1-p)\mathcal{N}(\mu_2,\sigma_2^2)$
Let
$f_i^{(k)}(x) = \frac{1}{\sqrt{2\pi(\sigma_i^2)^2}}\exp\left\{-\frac{(x-\mu_i^{(k)})^2}{2(\sigma_i^2)^{(k)}}\right\},\ i = 1,2$
$\alpha_i^{(k,1)} = \frac{p^{(k)}f_1(y_i)}{p^{(k)}f_1(y_i)+(1-p^{(k)})f_2(y_i)}$
$\alpha_i^{(k,2)} = 1 - \alpha_i^{(k,1)}$
$\mu_1^{(k+1)} = \frac{\sum_{i=1}^N \alpha_i^{(k,1)y_i}}{\sum_{i=1}^N \alpha_i^{(k,1)}},\ \mu_2^{(k+1)} = \frac{\sum_{i=1}^N \alpha_i^{(k,2)y_i}}{\sum_{i=1}^N \alpha_i^{(k,2)}}$
$(\sigma_1^2)^{(k+1)} = \frac{\alpha_i^{(k,1)}(y_i-\mu_1^{(k+1)})^2}{\sum_{i=1}^N \alpha_i^{(k,1)}}$
$(\sigma_2^2)^{(k+1)} = \frac{\alpha_i^{(k,2)}(y_i-\mu_2^{(k+1)})^2}{\sum_{i=1}^N \alpha_i^{(k,2)}}$
$p^{(k+1)} = \frac{\sum_{i=1}^N \alpha_i^{(k,1)}}{N}$
Mixtures of Poissons:
$y_i \sim p\text{Poisson}(\lambda_1) + (1-p)\text{Poisson}(\lambda_2)$
Let:
$f_i^{(k)}(x) = \frac{\lambda_i^{y_i}}{y_i!}\exp\{-\lambda_i\},\ i = 1,2$
$\alpha_i^{(k,1)} = \frac{p^{(k)}f_1^{(k)}}{p^{(k)}f_1^{(k)}+(1-p^{(k)})f_2^{(k)}},\ \alpha_i^{(k,2)} = 1-\alpha_i^{(k,1)}$
$\lambda_1^{(k+1)} = \frac{\sum_{i=1}^N \alpha_i^{(k,1)}y_i}{\sum_{i=1}^N \alpha_i^{(k,1)}},\ \lambda_2^{(k+1)} = \frac{\sum_{i=1}^N \alpha_i^{(k,2)}y_i}{\sum_{i=1}^N \alpha_i^{(k,2)}}$
$p^{(k+1)} = \frac{\sum_{i=1}^N \alpha_i^{(k,1)}}{N}$

Miscellaneous

Divergence/convergence of integrals of rational functions

$\int_1^\infty \frac{1}{x^p} = \begin{cases}\frac{1}{p-1},\ p > 1\\ \to\infty,\ p \leq 1\end{cases}$

$\int_0^a \frac{1}{x^p} = \begin{cases}\frac{1}{1-p},\ p < 1\\ \to\infty,\ p \geq 1\end{cases}$

Convex function

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if, for all $x, y \in \mathbb{R}^n$ and for all $\lambda \in [0,1]$, $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$.
For a function $\mathbb{R} \to \mathbb{R}$ this is equivalent to the second derivative being non-negative.

Inequalities

Jensen's inequality

Assume f is convex and X is a random variable, then $E[f(X)] \geq f(E[X])$

Cauchy-Schwarz inequality

$E[XY]^2 \leq E[X^2]E[Y^2]$

Chebyshev's inequality

Let X be a random variable with mean μ and variance σ^2, then $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

Markov's inequality

Let X be a non-negative random variable, then $P(X \geq k) \leq \frac{E[X]}{k}$

Some integrals

$\int \sin(x)\cos(x)dx = \frac{1}{2}\sin^2(x)$
$\int \sin^2(x)dx = \frac{1}{2}x - \frac{1}{4}\sin(2x)$
$\int \frac{1}{x^2+1}dx = \arctan(x)$
$\int_{-\infty}^\infty e^{-b(x-a)^2}dx = \sqrt{\frac{\pi}{b}}$

Trigonometric identities

$$\sin(\alpha+\beta) = \sin(\alpha)\sin(\beta) + \cos(\alpha)\sin(\beta)$$
$$\cos(\alpha+\beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$$
$$i\sin\theta = \sinh i\theta$$
$$\cos\theta = \cosh i\theta$$
$$\cosh x = (e^x + e^{-x})/2$$
$$\sinh x = (e^x - e^{-x})/2$$
$$\sin^2 x = (1 - \cos 2x)/2$$
$$\cos^2 x = (1 + \cos 2x)/2$$

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$