Antithtic Variables

Two identically r.v. Y and Y'

$\frac{Y+Y'}{2}$ $\quad E\left[\frac{Y+Y'}{2}\right] = E[Y] = \mu$

$Var\left(\frac{Y+Y'}{2}\right) = \frac{1}{4}\left[Var(Y) + Var(Y') + 2\,Cov(Y, Y')\right]$

to make $Cov(Y, Y') < 0$

Suppose $X_j = F_x^{-1}(u)$, $j = 1, \dots, m$ $\quad u \sim Unif(0,1)$

$Y_j = g[F_x^{-1}(u)]$, $j = 1, \dots, m$

u and $1-u \sim Unif(0,1)$

$Corr(u, 1-u) = -1$

$Y_j' = g[F_x^{-1}(1-u)]$, $j = 1, \dots, m$

Q: under what conditions are Y_j and Y_j' negatively correlated?

Ans: If $g(\cdot)$ is monotone, Y_j and Y_j' are negatively correlated.

$Y_j = g[F_x^{-1}(u)] = f(u)$ $\quad$ assume $g(\cdot)$ is increasing func.

$-Y_j' = -g[F_x^{-1}(1-u)] = h(u)$ $\quad Cov(Y_j, Y_j') < 0$

$E[f(u)h(u)] \geq E(f(u))\,E(h(u)) = E(Y_j Y_j') - E(Y_j)E(Y_j')$

$E[-Y_j Y_j'] \geq E(Y_j)\,E(-Y_j')$

$E(Y_j Y_j') \leq E(Y_j)\,E(Y_j')$

Algorithm

$u_1, u_2, \dots, u_{m/2} \sim Unif(0,1)$

$Y_j = f(u_j)$

$Y_j' = h(u_j) = f(1-u_j)$

$\hat{\theta} = \sum_{j=1}^{m/2} \frac{Y_j + Y_j'}{m}$

E.g. $x \geq 0$, $F_x(x) + 0.5$ $\quad g(u)$

$\theta = \int_0^1 \frac{\alpha}{\sqrt{2\pi}} e^{-\frac{1}{2}(xu)^2} du$ $\quad u \sim Unif(0,1)$

$Y_j = g(u_j) \Rightarrow f(u)$

$Y_j' = g(1-u_j) \Rightarrow h(u)$

$\hat{\theta} = \sum_{j=1}^{m/2} \frac{Y_j + Y_j'}{m}$

```
MC_phi <- function(x, m = 10000, anti = TRUE){
  u <- runif(m / 2)
  if(!anti)
   v <- runif(m / 2)
  else
   v <- 1 - u

  u <- c(u, v)
  mc <- numeric(length(x))
  for(i in 1:length(x)){
     g <- x[i] * exp(-(u * x[i]) ^ 2 / 2)
     mc[i] <- mean(g) / sqrt(2 * pi) + 0.5
  }
  mc
}
```

The R code to compute the approximate reduction in variance at a $x = 1.3$.

```
n <- 50
MC1 <- MC2 <- numeric(n)
x0 <- 1.3
for(i in 1:n){
  MC1[i] <- MC_phi(x0, m = 1000, anti = FALSE)
  MC2[i] <- MC_phi(x0, m = 1000)
}
(var(MC1) - var(MC2)) / var(MC1)
```

```
## [1] 0.9569375
```

~96% reduction in variance

Control Variates

- $Var(\hat{\theta}_c)$ is minimized at $c = c^*$, where

$$c^* = -\frac{Cov[g(X), f(X)]}{Var[f(X)]}$$

- and minimum variance is

$$Var(\hat{\theta}_{c^*}) = Var[g(X)] - \frac{Cov[g(X), f(X)]^2}{Var[f(X)]}$$

$\hat{\theta}_c = g(x) + c\,[f(x) - \mu]$ $\quad E(f(x)) = \mu$ is known

$E[\hat{\theta}_c] = \theta$ $\quad E[g(x)] = \theta$

$Var(\hat{\theta}_c) \leq Var(\hat{\theta})$

$Var(\hat{\theta}_c) = Var(g(x) + c[f(x) - \mu])$

$= Var(g(x)) + 2c\,Cov[g(x), f(x)] + c^2 Var[f(x)]$

$\min_c Var(\hat{\theta}_c)$

$\frac{\partial}{\partial c} Var(\hat{\theta}_c) = 0$ and solve it for c^*

$c^* = -\frac{Cov[g(x), f(x)]}{Var(f(x))}$

$Var(\hat{\theta}_c) = Var[g(x)] - 2\frac{Cov[g(x), f(x)]^2}{Var(f(x))} + \frac{Cov[g(x), f(x)]^2}{Var[f(x)]}$

$= Var[g(x)] - \frac{Cov[g(x), f(x)]^2}{Var(f(x))} \leq Var[g(x)]$ $\quad$ Control variate

The percent reduction $= 100 \cdot \frac{\frac{Cov[g(x), f(x)]^2}{Var(f(x))}}{Var(g(x))}$

$= 100 \cdot \frac{Cov[g(x), f(x)]^2}{Var(g(x)) \cdot Var(f(x))}$

$= 100 \cdot Corr(g(x), f(x))^2$

$\hat{\theta}_c = c\hat{\theta}_1 + (1-c)\hat{\theta}_2$ $\quad \hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of θ

$E(\hat{\theta}_c) = \theta$

$Var(\hat{\theta}_c) = Cov(c\hat{\theta}_1 + (1-c)\hat{\theta}_2,\ c\hat{\theta}_1 + (1-c)\hat{\theta}_2)$

$= Cov(c(\hat{\theta}_1 - \hat{\theta}_2) + \hat{\theta}_2,\ c(\hat{\theta}_1 - \hat{\theta}_2) + \hat{\theta}_2)$

$= Var(\hat{\theta}_2) + 2c\,Cov(\hat{\theta}_1 - \hat{\theta}_2, \hat{\theta}_2) + c^2 Var(\hat{\theta}_1 - \hat{\theta}_2)$

$Corr(\hat{\theta}_1, \hat{\theta}_2) = -1 \Rightarrow Cov(\hat{\theta}_1, \hat{\theta}_2) = -Var(\hat{\theta}_2)$

$\hat{\theta}_1$ and $\hat{\theta}_2$ are identically distributed $Var(\hat{\theta}_1) = Var(\hat{\theta}_2)$

$\min_c Var(\hat{\theta}_c) \Rightarrow c^* = \frac{1}{2}$

$\hat{\theta}_c = \frac{1}{2}\hat{\theta}_1 + \frac{1}{2}\hat{\theta}_2$

Several control variates

$\hat{\theta}_c = g(x) + \sum_{i=1}^{k} c_i^*(f_i(x) - \mu_i)$ $\quad$ simple linear regression

$E(\hat{\theta}_c) = \theta$

$Var(\hat{\theta}_c) \leq Var(\hat{\theta})$

$g(u) = \frac{e^{-u}}{1+u^2}$ $\quad u \sim Unif(0,1)$

$Corr(g(u), f(u)) \uparrow$

$f(u) = e^{-u}$

$E(f(u)) = \int_0^1 e^{-u}\,du = 1 - e^{-1}$

Algorithm

① find $f(x)$ and $E[f(x)] = \mu$

② $u \sim Unif(0,1)$ $\quad u_1, \dots, u_m$

③ Compute $c^* = -\frac{Cov[g(u), f(u)]}{Var[f(u)]}$

④ $\hat{\theta}_c = \sum_{i=1}^{m} \frac{g(u_i) + c^*\cdot[f(u_i) - \mu]}{m}$

```
f <- function(u){
  exp(u)
}
g <- function(u){
  u*exp(-3*u)
}

u <- runif(10000, 0, 2)
B <-f(u)
A <-g(u)

a <- -cov(A,B)/var(B)
u <- runif(10000, 0, 2)
T1 <- g(u)*2
T2 <- T1 + a * (f(u) - mean(B))
cbind(mean(T1), mean(T2))
```

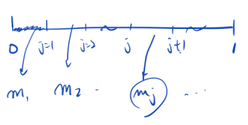
```
##           [,1]     [,2]
## [1,] 0.1082181 0.108405
```

```
con_var <- (var(T1) - var(T2))/var(T1)
con_var
```

```
## [1] 0.529722
```

Stratified Sampling

Stratified sampling $\quad \theta = E[g(x)] = \int_0^1 g(x)\,dx$



① $\hat{\theta}_j = \frac{\sum_{i=1}^{m_j} g(x_i)}{m_j}$

$E[g(x|J=j)] = \theta_j$

$\hat{\theta}_s = \frac{\sum_{j=1}^{k} \hat{\theta}_j}{k}$

$E(\hat{\theta}_s) = \theta$

$E[E[g(x|J)]] \overset{?}{=} E(g(x))$

$E[E[g(x|J)]] = E[g(x|J=1)] \cdot P(J=1) + E[g(x|J=2)] \cdot P(J=2) + \dots + E[g(x|J=k)] \cdot P(J=k)$

$P(A,B) = P(A|B) \cdot P(B)$

$Var(\hat{\theta}_s) = Var\left(\frac{\sum_{j=1}^{k} \hat{\theta}_j}{k}\right) = \frac{1}{k^2}\sum_{j=1}^{k} Var(\hat{\theta}_j) = \frac{1}{k^2}\sum_{j=1}^{k} \frac{\sigma_j^2}{m_j} = \frac{1}{k^2}\sum_{i=1}^{k} \frac{\sigma_j^2}{m}$

$= \sum_{j=1}^{k} E[g(x)] \cdot P(X, J=j) = \sum_{x} g(x) \sum_{j=1}^{k} P(X, J=j)$

$m_1 = m_2 = \dots = m_j = \dots = m_k = m$

$= \frac{1}{km}\sum_{j=1}^{k} \sigma_j^2 = \sum g(x)\,P(x) = E(g(x)) = \theta$

$m = \frac{M}{k}$

$Var(\hat{\theta}_{MC}) = \frac{\sigma^2}{M}$

$E(g(x))^2$

$Var[E(g(x)|J)] = E[E(g(x)|J)^2] - E[E(g(x)|J)]^2$ ①

$Var[g(x|J)] = E[g(x|J)^2] - E[g(x|J)]^2$

$E[Var[g(x)|J]] = E[g(x)^2] - E[E[g(x)|J]^2]$

$Var[E[g(x)|J]] = Var(g(x)) - E[Var(g(x)|J)]$

$Var(g(x)) = Var[E[g(x)|J]] + E[Var(g(x)|J)]$

between-group variance $\quad$ within-group variance

$Var(\hat{\theta}_{MC}) = Var(\overline{g(x)}) = \frac{1}{M} Var\,g(x)$

$Var(\hat{\theta}_{MC}) = \frac{1}{M}\left[Var[E[g(x)|J]] + E[Var(g(x)|J)]\right]$ $\quad \theta_J$

$= \frac{1}{M}\left[Var(\theta_J) + E(\sigma_J^2)\right]$ $\quad P(J=j) = \frac{1}{k}$

$= \frac{1}{M}\left[Var(\theta_J) + \sum_{j=1}^{k} \frac{\sigma_j^2}{k}\right]$

$= \frac{1}{M} Var(\theta_J) + \frac{1}{Mk}\sum_{j=1}^{k} \sigma_j^2$

$Var(\hat{\theta}_{MC}) \geq Var(\hat{\theta}_s)$

1. Initialize Parameters: Set the total number of samples m and the number of strata n_strata.
2. Initialize Storage for Stratified Estimates: Create a numeric vector stratified_estimates to store the mean estimates from each stratum.
3. Stratified Sampling Loop: For each stratum, calculate the lower and upper bounds, generate uniform random samples within these bounds, apply the function fx to these samples, and store the mean of these function values.
4. Calculate the Overall Estimate: Average the stratified means and multiply by the length of the interval (which is 2 in this case).

```
m <- 10000
n_strata <- 100
stratified_estimates <- numeric(n_strata)

for (i in 1:n_strata) {
    lower <- (i - 1) * 2 / n_strata
    upper <- i * 2 / n_strata
    u <- runif(m / n_strata, min = lower, max = upper)
    stratified_estimates[i] <- mean(fx(u))
}

theta_stratified <- mean(stratified_estimates)*2

theta_stratified
```

Suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ. Then for any constant C.

(a) Please show $\hat{\theta} = (1-C)\hat{\theta}_1 + C\hat{\theta}_2$ is an unbiased estimator of θ

(b) Assume the correlation of $\hat{\theta}_1$ and $\hat{\theta}_2$ is -1. Please find C such that $Var(\hat{\theta})$ achieves the minimum.

$$E\left(\hat{\theta}\right) = (1-C)\,\theta + C\theta = \theta$$

b)

$$Var(\theta) = Var((1-C)\hat{\theta}_1 + C\hat{\theta}_2)$$

$$Var(\hat{\theta}) = (1-C)2Var(\hat{\theta}_1) + C2Var(\hat{\theta}_2) - 2C(1-C)Cov(\hat{\theta}_1, \hat{\theta}_2)$$

$$\frac{d}{dC} = 2CVar(\hat{\theta}_2) - 2(1-C)Var(\hat{\theta}_1) + 2(1-2C)(-\sigma\theta_1\sigma\theta_2)$$

$$C = \frac{\sigma(\hat{\theta}_1)\sigma(\hat{\theta}_2)}{2Var(\hat{\theta}_1) - Var(\hat{\theta}_2)}$$

Problem 3: Suppose $X \sim f(x)$, Let $\theta = \int g(x)f(x)dx = E_f[g(X)]$. We draw m iid copies $X_1, \dots X_m$ from $\phi(x)$, which is different from $f(x)$, and define $W(x) = \frac{f(x)}{\phi(x)}$

(a) Prove $E_f[g(X)] = E_\phi[g(X)W(X)]$

$$\phi(x) = \frac{f(x)}{W(x)}$$

$$E_\phi[g(X)W(X)] = \int g(x)W(x)\phi(x)dx$$

$$E_\phi[g(X)W(X)] = \int g(x)\frac{f(x)}{\phi(x)}\phi(x)dx = \int g(x)f(x)dx$$

$$E_f[g(X)] = \int g(x)f(x)dx = E_\phi[g(X)W(X)]$$

(b) Prove $E_\phi[W(X)] = 1$

$$E_\phi[W(X)] = \int W(x)\phi(x)dx$$

$$E_\phi[W(X)] = \int \frac{f(x)}{\phi(x)}\phi(x)dx = \int f(x)dx = 1$$

(c) Let $\hat{\theta} = \Sigma_{i=1}^m g(X)W(X)/m$. Find $E[\hat{\theta}]$ and $Var[\hat{\theta}]$

$$E(\hat{\theta}) = E(\Sigma_{i=1}^m g(X_i)W(X_i)/m) = \frac{1}{m}\Sigma_{i=1}^m E[g(X_i)W(X_i)]$$

Since that X_i are iid samples and $E_f[g(X)] = E_\phi[g(X)W(X)]$

$$E(\hat{\theta}) = E[g(X)W(X)] = E_f[g(X)]$$

$$Var[\hat{\theta}] = Var[\Sigma_{i=1}^m g(X_i)W(X_i)/m] = \frac{1}{m^2}\Sigma_{i=1}^m Var[g(X_i)W(X_i)] = \frac{1}{m}Var[g(X)W(X)]$$

$$Var[\hat{\theta}] = \frac{1}{m}(E[g(X)^2W(X)^2] - E_f[g(X)]^2)$$

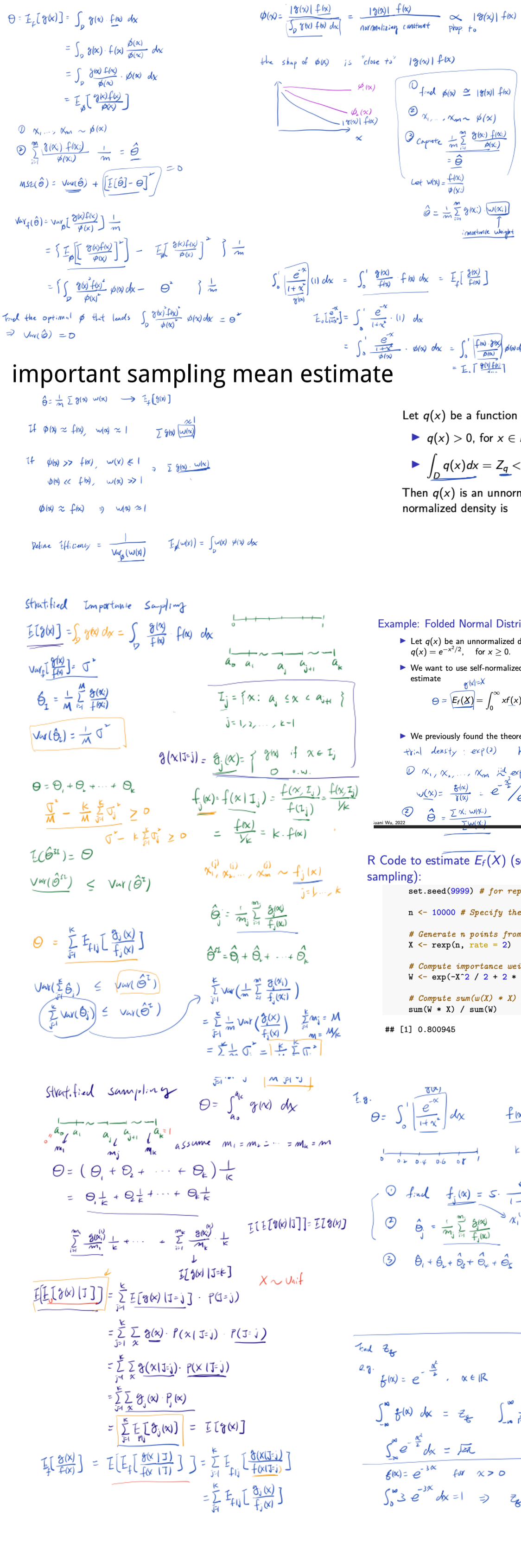
Simple MC

1. Define the Function: You've already defined the function fx(x) = x * exp(-3 * x).
2. Generate Uniform Random Numbers: Generate a large number of uniform random numbers over the interval [0, 2].
3. Apply the Function to Random Numbers: Apply fx to each of these random numbers.
4. Calculate the Average: Compute the average of these function values.
5. Multiply by the Length of the Interval: Multiply this average by the length of the interval (which is 2 in this case) to get the estimate of the integral.

```
fx <- function(x) { x * exp(-3 * x) }

m <- 10000
u <- runif(m, 0, 2)
mc1_estimates <- fx(u)

theta_mc1 <- mean(mc1_estimates)*2
theta_mc1
```



important sampling mean estimate

Example: Folded Normal Distribution

- Suppose we want to estimate $E_f(X)$, where $f(x)$ is the PDF of the folded normal distribution,
$$f(x) = 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \sqrt{\frac{2}{\pi}} e^{-x^2/2}, \quad \text{for } x \geq 0.$$
- The support of $f(x)$ is $D = [0, \infty)$:
  - $f(x) > 0$ for all $x \in D$
  - $\int_0^\infty f(x)dx = 1$
- We want to use importance sampling to estimate $E_f(X)$.

uani Wu, 2022

```r
g <- function(x){
 exp(-x - log(1 + x^2)) * (x > 0) * (x < 1)
}
m <- 10000
#Use Uniform(0, 1) as the candidate function
is1 <- replicate(1000, expr = {
    x <- runif(m)
    phi <- 1
    mean(g(x) / phi)
})
#Use Exponential(1) as the candidate function
is2 <- replicate(1000, expr = {
    u <- runif(m)
    x <- -log(u)
    x <- x[x <= 1]
    phi <- exp(-x)
    sum(g(x) / phi) / m
})
```

```r
#Use exp(-x) / (1 - exp(-1)) as the candidate function
is3 <- replicate(1000, expr = {
    u <- runif(m)
    x <- -log(1 - u * (1 - exp(-1)))
    phi <- exp(-x) / (1 - exp(-1))
    mean(g(x) / phi)
 })
#Use Standard Cauchy as the candidate function
is4 <- replicate(1000, expr = {
    x <- rcauchy(m)
    x <- x[x >= 0 & x <=1]
    phi <- dcauchy(x)
    outcome <- g(x) / phi
    sum(outcome) / m
 })
c(mean(is1), mean(is2), mean(is3), mean(is4))
```

```
## [1] 0.5246914 0.5245130 0.5247545 0.5248252
```

```r
c(var(is1), var(is2), var(is3), var(is4))
```

```
## [1] 6.011627e-06 1.681950e-05 9.240642e-07 9.484520e-05
```

```r
set.seed(9999)
n <- 10000 # Specify the number of points to generate

# Generate n points from Exp(lambda = 2)
x <- rexp(n, rate = 2)

# Compute importance weights
w <- exp(-x^2 / 2 + 2 * x) / sqrt(2 * pi)

# Compute mean(w(x) * g(x)) (g(x) = x here)
mean(w * x)
```

Let $q(x)$ be a function defined on a region D. Suppose that

- $q(x) > 0$, for $x \in D$
- $\int_D q(x)dx = Z_q < \infty$.

Then $q(x)$ is an unnormalized density on D. The corresponding normalized density is

$$f(x) = \frac{q(x)}{Z_q}.$$

Example: Folded Normal Distribution

- Let $q(x)$ be an unnormalized density for $f(x)$, given by $q(x) = e^{-x^2/2}$, for $x \geq 0$.
- We want to use self-normalized importance sampling to estimate
$$\Theta = E_f(X) = \int_0^\infty x f(x)dx = \int_0^\infty x \frac{q(x)}{Z_q} dx.$$
- We previously found the theoretical $E_f(X) = \sqrt{\frac{2}{\pi}}$.

uani Wu, 2022

R Code to estimate $E_f(X)$ (self-normalized importance sampling):

```r
set.seed(9999) # for reproduceability

n <- 10000 # Specify the number of points to generate

# Generate n points from Exp(lambda = 2)
X <- rexp(n, rate = 2)

# Compute importance weights
W <- exp(-X^2 / 2 + 2 * X)

# Compute sum(w(X) * X) / sum(w(X))
sum(W * X) / sum(W)
```

```
## [1] 0.800945
```

Suppose we sample $X_1, \ldots, X_m$ from an importance sampling function $\phi(x)$. Please show that $E(\sum_i^m w_i) = E(f(x_i)/\phi(x_i)) = m$, where f is target density.

$$\mathbf{E}\left[\sum_{i=1}^m w_i\right] = \mathbf{E}\left[\sum_{i=1}^m \frac{f(x_i)}{\phi(x_i)}\right]$$

$$\mathbf{E}\left[\frac{f(x_i)}{\phi(x_i)}\right] = \int \frac{f(x)}{\phi(x)}\phi(x)dx$$

$$\mathbf{E}\left[\frac{f(x_i)}{\phi(x_i)}\right] = \int f(x)dx$$

$$\mathbf{E}\left[\frac{f(x_i)}{\phi(x_i)}\right] = 1$$

$$\sum_{i=1}^m \mathbf{E}\left[\frac{f(x_i)}{\phi(x_i)}\right] = m \times 1 = m$$

We showed that the variance of importance sampling estimator $\mathbf{Var}_f(\hat{\theta}) = \frac{1}{m}\mathbf{Var}_\phi(\frac{g(X)f(X)}{\phi(X)})$, where $\hat{\theta} = \frac{1}{m}\sum_{i=1}^m \frac{g(x_i)f(x_i)}{\phi(x_i)}$ and $x_i \sim \phi(x)$. Please show that $\mathbf{Var}_\phi(\frac{g(X)f(X)}{\phi(X)})$ is equal to $\mathbf{E}_\phi[\frac{(g(X)f(X)-\theta\phi(X))^2}{\phi^2(X)}]$. Here $f(x)$ is importance function, and $\phi(x)$ is importance sampling function.

Your answer:

$\hat{\theta} = \frac{1}{m}\sum_{i=1}^m \frac{g(x_i)f(x_i)}{\phi(x_i)}$

$\mathbf{Var}_\phi(\hat{\theta}) = \frac{1}{m}\mathbf{Var}_\phi\left(\frac{g(X)f(X)}{\phi(X)}\right)$

$\mathbf{E}_\phi\left[\left(\frac{g(X)f(X)}{\phi(X)} - \hat{\theta}\right)^2\right]$

$\mathbf{Var}_\phi(\hat{\theta}) = \mathbf{E}_\phi\left[\left(\frac{g(X)f(X)}{\phi(X)} - \mathbf{E}_\phi\left[\frac{g(X)f(X)}{\phi(X)}\right]\right)^2\right]$

$\mathbf{Var}_\phi(\hat{\theta}) = \mathbf{E}_\phi\left[\left(\frac{g(X)f(X)}{\phi(X)} - \hat{\theta}\right)^2\right]$

Transition Probabilities (2)

If the stat space is finite, the transition probabilities P_{ij} can be represented by a transition matrix

$$\mathbb{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots & P_{0N} \\ P_{10} & P_{11} & P_{12} & \cdots & \\ \vdots & \vdots & \vdots & & \\ P_{i0} & P_{i1} & P_{i2} & \cdots & \\ \vdots & \vdots & \vdots & & \\ P_{N0} & & & & P_{NN} \end{bmatrix}, \quad \sum_{j=1}^{N} P_{ij} = 1$$

where

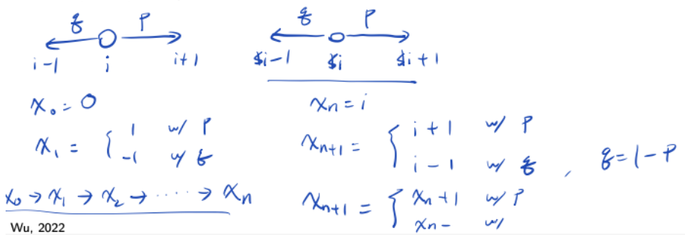
1. All the entries are non-negative:
$$P_{ij} \geq 0, \text{ for all } i, j. \tag{3}$$
2. The sum of each row is 1:
$$\sum_{j=0}^{\infty} P_{ij} = \sum_{j=0}^{\infty} P(X_{n+1} = j | X_n = i) = 1, \text{ for all } i. \tag{4}$$

A **random walk** is a stochastic process:

$$\{X_0, X_1, X_2, \ldots, X_n, \ldots\}, \tag{5}$$

defined on the integers $\mathbb{Z}$, such that:

1. The walk starts at 0: $X_0 = 0$.
2. At each step, the random walk moves to the right 1 unit with probability p and moves to the left 1 unit with probability q (so $p + q = 1$).

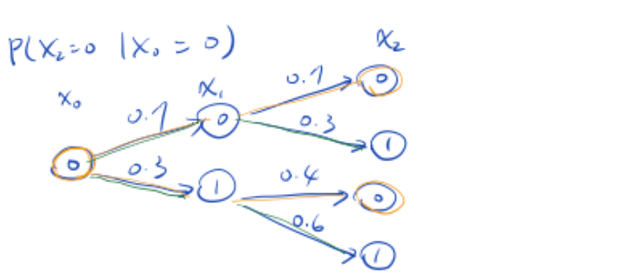


Consider a two-state Markov chain, with state space $\{0, 1\}$ and the transition matrix

$$\mathbb{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}.$$

Suppose we are given that the chain starts at 0 : $X_0 = 0$. Calculate the probability that the state 0 in 2 steps, $P(X_2 = 0) = P_{00}^{(2)}$.

If we generate $X_1, X_2, \ldots, X_n$, for some large n (for example, $n = 10000$). How often is the Markov chain in state 0? How often is the Markov chain in state 1?



$$P_{00}^{(2)} = 0.7 * 0.7 + 0.3 * 0.4$$

$$P_{01}^{(2)} = 0.7 * 0.3 + 0.3 * 0.6$$

$$\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.7*0.7 + 0.3*0.4 & 0.7*0.3 + 0.3*0.6 \\ \bigcirc & \bigcirc \end{bmatrix}$$

$$P^{(n)} = P^n$$

```
n <- 1000
P <- matrix(c(0.7,0.4,0.3,0.6), nrow = 2)
P2 <- P
for(i in 1:(n - 1)){
    P2 <- P2 %*% P
}
```

P2

Problem 2: Two urns A and B contain a total of N balls. Assume that at time t there were exactly k balls in A. At time t + 1, an urn is selected at random in proportion to its contents (i.e., A is chosen with probability k/N and B is chosen with probability (N - k)/N). Then one of the N balls is randomly selected and placed in the chosen urn.

Let X_t denote the number of balls in urn A at time t, so $X_t, t \geq 0$ defines a Markov chain. Determine the transition matrix for this Markov chain.

Since we have that

$$P(X_{t+1} = k - 1 | X_t = k) = \frac{k}{N}$$
$$P(X_{t+1} = k + 1 | X_t = k) = \frac{N - k}{N}$$

Then when $k = 0$, $P(X_{t+1} = k + 1) = 1; P(X_{t-1} = k + 1) = 0$, and when $k = 1$, $P(X_{t+1} = k + 1) = 1/N; P(X_{t-1} = k + 1) = (N - 1)/N$ So on.

Therefore, we will have the following transition matrix:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 1/N & 0 & (N-1)/N & 0 & \ldots & 0 \\ 0 & 2/N & 0 & (N-2)/N & \ldots & 0 \\ | & & & & & | \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

Transition Probabilities (3)

Another way to visualize the transition probabilities of a Markov chain is with a **transition state diagram**:

- Each state in the state space is represented by a node/vertex.
- Each nonzero transition probability P_{ij} is represented by an arrow from vertex i to vertex j.

$$X_1 = X_0 + Y_1 \qquad Y_1 \sim \text{Ber}(P(Y_1 = 1) = p, \; P(Y_1 = -1) = q)$$

$$X_2 = X_1 + Y_2$$

$$X_n = X_{n-1} + Y_n \qquad Y = (-1, 1, \ldots)$$

$$\Rightarrow X_0 + \sum_{i=1}^{n} Y_i$$

$$P(X_{n+1} = j | X_n = i)$$

$$P(X_{n+k} = j | X_n = i)$$

```
set.seed(999) # for reproduceability
n <- 1000 # specify length of random walk
p <- 0.5 # specify P(Y = 1)
Y <- sample(c(1, -1), size = n, replace = TRUE,
            prob = c(p, 1 - p)) # Generate n iid samples from Y
X <- c(0, cumsum(Y)) # Compute the random walk X
plot(X, type = "l", las = 1) # Plot the random walk over time
```

The Limiting Distribution

Let $\pi = (\pi_0, \pi_1, \pi_2, \ldots, \pi_N)$ be a probability distribution on the state space $\{0, 1, 2, \ldots, N\}$. We say π is the **limiting distribution** of a Markov chain $\{X_0, X_1, X_2, \ldots\}$ if

$$\lim_{n \to \infty} P(X_n = j | X_0 = i) = \pi_j, \tag{7}$$

for all $i, j \in \{0, 1, 2, \ldots, N\}$.

Assume the limiting dist. exists, Then for n large enough

$$P(X_n = k | X_0 = i) = \pi_k \quad \text{and} \quad P(X_{n+1} = j | X_0 = i) = \pi_j$$

find $\pi = \{\pi_0, \pi_1, \ldots, \pi_N\}$

$$\pi_j = P(X_{n+1} = j | X_0 = i)$$
$$= \sum_{k=0}^{N} P(X_{n+1} = j, X_n = k | X_0 = i)$$
$$= \sum_{k=0}^{N} P(X_{n+1} = j | X_n = k, X_0 = i) \cdot P(X_n = k | X_0 = i)$$
$$= \sum_{k=0}^{N} P(X_{n+1} = j | X_n = k) \pi_k$$
$$= \sum_{k=0}^{N} P_{kj} \pi_k = \sum_{k=0}^{N} \pi_k P_{kj}$$

$$\begin{cases} \pi_j = \sum_{k=0}^{N} \pi_k P_{kj} & \cdots \; (1) \qquad P\pi = \pi \\ \pi_0 + \pi_1 + \cdots + \pi_N = 1 & \cdots \; (2) \end{cases}$$

Eg $P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$

$$\pi_0 = \pi_0 P_{00} + \pi_1 P_{10}$$
$$\pi_0 = \pi_0 0.7 + \pi_1 0.4$$
$$\pi_1 = \pi_0 P_{01} + \pi_1 P_{11}$$
$$= \pi_0 0.3 + \pi_1 0.6$$
$$\pi_0 + \pi_1 = 1$$
$$0.3\pi_0 = 0.4\pi_1 \Rightarrow \pi_1 = \frac{3}{4}\pi_0$$
$$\pi_0 + \frac{3}{4}\pi_0 = 1 \Rightarrow \pi_0 = \frac{4}{7}$$
$$\pi_1 = 1 - \frac{4}{7} = \frac{3}{7}$$

$$\pi = \{\tfrac{4}{7}, \tfrac{3}{7}\}$$

Suppose $X_0 \sim \pi$ we can show $X_1 \sim \pi$

$$P(X_0 = 0) = \pi_0 \qquad P(X_0 = 1) = \pi_1$$
$$P(X_1 = 0) = \sum_{k=0}^{1} P(X_1 = 0 | X_0 = k) \cdot P(X_0 = k)$$
$$= P_{00}\pi_0 + P_{10}\pi_1$$
$$= 0.7\tfrac{4}{7} + 0.4\tfrac{3}{7} = \tfrac{4}{7} = P(X_0 = 0) = \pi_0$$

If a Markov chain $\{X_0, X_1, X_2, \ldots\}$ is irreducible, aperiodic, and has a stationary distribution $\pi = (\pi_0, \pi_1, \pi_2, \ldots, \pi_N)$, then

$$\lim_{n \to \infty} P(X_n = j | X_0 = i) = \pi_j.$$

That is, π is the limiting distribution of the Markov chain, and π is uniquely determined by the system of equations

$$\begin{cases} \pi = \pi\mathbb{P} \\ \sum_{i=0}^{N} \pi_i = 1 \end{cases}$$

where $\mathbb{P}$ is the transition matrix of the Markov chain.

Chapman-Kolmogorov Equations

We have defined the one-step transition probabilities P_{ij}. We now define the probability that the chain moves from state i to state j in n steps is $P_{ij}^{(n)}$.

$$P_{ij}^{(n)} = P(X_{n+k} = j | X_k = i), \; n \neq 0, \; i, j \neq 0$$

The Chapman-Kolmogorov equations provide a method for computing these n-step transition probabilities:

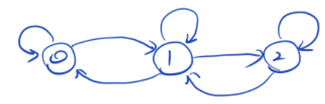
$$P_{ij}^{(n+m)} = \sum_{k=0}^{\infty} P_{ik}^{(n)} P_{kj}^{(m)} \text{ for all, } n, m \neq 0, \text{ all } i, j. \tag{6}$$

Irreducible Markov Chains (1)

Definitation:

1. State j is **accessible** from state i, denoted by $i \to j$, if there is a path from state i to state j.
2. Two states i and j **communicate**, denoted by $i \leftrightarrow j$, if each state is accessible from the other. In other words, if $i \to j$ and $j \to i$, then $i \leftrightarrow j$.
3. A Markov chain is **irreducible** if all states communicate with each other. A Markov chain is **reducible** if it is not irreducible.

$$P = \begin{bmatrix} & 0 & 1 & 2 \\ 0 & 0.6 & 0.4 & 0 \\ 1 & 0.2 & 0.5 & 0.3 \\ 2 & 0 & 0.6 & 0.4 \end{bmatrix}$$



Irreducible Markov Chains (2)

Accessibility can be defined more rigorously using n-step transition probabilities:

Let $\{X_0, X_1, X_2, \ldots\}$ be a Markov chain with state space $\{0, 1, 2, \ldots, N\}$. The **n-step transition probability** from state i to j is defined by

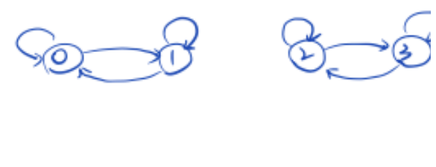
$$P_{ij}^{(n)} := P(X_n = j | X_0 = i). \tag{8}$$

Note that when $n = 1$, $P_{ij}^{(1)} = P_{ij}$.

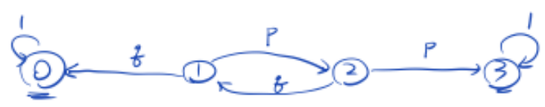
State j is **accessible** from state i, denoted by $i \to j$, if there is positive probability that state j can be reached from state i in a finite number of transitions. In other words, $P_{ij}^{(n)} > 0$ for some n.

Reducible

e.g. $P = \begin{bmatrix} 1/4 & 3/4 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$



e.g



Absorbing states

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad 0 < p, q < 1, \; p + q = 1$$

$d(0) = \gcd\{1, 2, 3, 4, \ldots\} = 1$

$d(3) = 1$

$d(1) = ?$

$1 \to 2 \to 1$ 2 transitions

$1 \to 2 \to 1 \to 2 \to 1$ 4 transitions

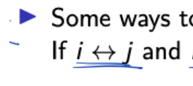
$\gcd(2, 4, 6, 8, \ldots) = 2 \Rightarrow d(1) = 2$

$d(2) = 2$

If $i \leftrightarrow j$, than $d(i) = d(j)$

- The **period** of a state i, denoted by $d(i)$, is the greatest common divisor (gcd) of all integers $n \geq 1$ for which $P_{ii}^{(n)} > 0$.
- To find the period of a state i, we consider the number of steps it takes for any path starting from state i to return to state i.
- A Markov chain is called **aperiodic** if each state has period 1. $d(i) = 1 \; \forall i$
- Some ways to check if a Markov chain is aperiodic: If $i \leftrightarrow j$ and $P_{ii} \geq 0$, then $d(j) = 1$

State i ufte $n = \{k+1, k+2, k+3, \ldots\}$ return i



$d(0) = 1$

$P_{00} > 0, \; 0 \leftrightarrow 1 \Rightarrow d(1) = 1$

$1 \leftrightarrow 1, \; k+1, \; k+2 \cdots \; k+t$

$d(2)$

$2 \leftrightarrow 1$ ✓

$P_{11} > 0$

$2 \leftrightarrow 3$ ✓ $\Rightarrow d(2) = 1$

$P_{33} > 0$

$d(2) = d(1) = d(3) = 1$

Bayesian Perspective

In the Bayesian perspective, we are able to take our prior beliefs into account. We represent our beliefs about θ prior to observing data by a **prior distribution** $P(\theta)$.

- The prior distribution can represent past information, such as past experiments or literature, or subjective beliefs from a knowledgeable person.
- If no prior information is available (or we do not want to take it into account), we can use an **uninformative** (or **flat**) **prior**, which assigns equal density to all possibilities of the parameter.
- When using an uninformative prior, Bayesian estimators tends to be similar (sometimes identical) to frequentist estimators: The data easily outweighs a prior with no information.
- The **posterior distribution** $f(\theta|y)$ represents our updated beliefs about θ *after* observing the data. If the posterior is in the same parametric family as the prior, the prior and posterior are called **conjugate distributions**, and the prior is called a **conjugate prior** for the likelihood.

$\theta|Y \sim f(\ )$ $\quad P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B|A)\,P(A)}{P(B)}$

$f(\theta|Y) = \frac{f(Y|\theta)\cdot f(\theta)}{\int f(Y|\theta)\cdot f(\theta)\,d\theta}$ (posterior)

$\propto f(Y|\theta)\cdot f(\theta|\underbrace{a_1, a_2, \dots, a_k}_{\text{hyper-parameters}})$

likelihood prior

$f(Y|\theta_1, \theta_2, \theta_3, \dots \theta_k)\cdot P(\theta_1)\,P(\theta_2)\cdots P(\theta_k)$

Example: Beta-Binomial Model

For our coin flipping example, suppose our prior is $P(\theta) \sim Beta(\alpha, \beta)$. Find the posterior mean.

$Y \overset{iid}{\sim} Bin(n, \theta)$ $\quad P(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$

$f(\theta|y) \propto \binom{n}{y}\theta^{y}(1-\theta)^{n-y}\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$

$= \binom{n}{y}\frac{1}{B(\alpha,\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$

$\propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$ $\quad$ set $\alpha' = y+\alpha$, $\beta' = n-y+\beta$

$= \theta^{\alpha'-1}(1-\theta)^{\beta'-1}$

$f(\theta|y) = Beta(\alpha', \beta')$

$E[\theta|Y] = \frac{\alpha'}{\alpha'+\beta'} = \frac{y+\alpha}{y+\alpha+n-y+\beta}$

$= \underbrace{\frac{n}{n+\alpha+\beta}}_{w}\frac{y}{n} + \underbrace{\frac{\alpha+\beta}{n+\alpha+\beta}}_{1-w}\frac{\alpha}{\alpha+\beta}$

$= w\hat{\theta}_{MLE} + (1-w)\mathbb{E}(\theta)$

Recall: $\theta \sim Beta(\alpha, \beta)$

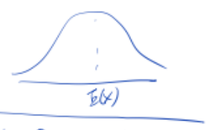
$\mathbb{E}(\theta) = \frac{\alpha}{\alpha+\beta}$

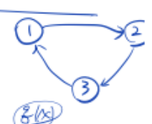
$Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$Mode = \frac{\alpha-1}{\alpha+\beta-2}$

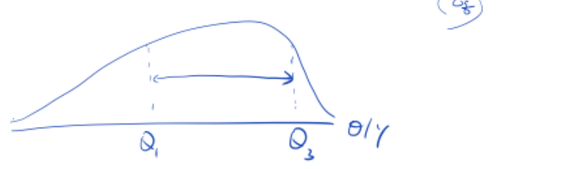
If Y, n are large, $\hat{\theta}_{MLE} \approx \mathbb{E}(\theta|Y)$

If Y, n are small, $\mathbb{E}(\theta) \approx \mathbb{E}(\theta|Y)$





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



The uninformative prior $\theta \sim Unif(0,1)$ $\quad \mathbb{E}(\theta|Y) = \frac{\alpha'}{\alpha'+\beta'} = \frac{y+1}{n+2}$

$f(\theta|y) \propto \binom{n}{y}\theta^{y}(1-\theta)^{n-y}(1)$

$\propto \theta^{y}(1-\theta)^{n-y}$ $\quad$ set $\alpha' = y+1$, $\beta' = n-y+1$

$f(\theta|y) = Beta(\alpha', \beta')$

$Mode = \frac{\alpha'-1}{\alpha'+\beta'-2} = \frac{y}{n}$ MAP

```
alpha <- 6
beta <- 2

sum_y1 <- 20
n1 <- 5
sum_y2 <- 80
n2 <- 20

posterior_alpha1 <- alpha + sum_y1
posterior_beta1 <- beta + n1
posterior_alpha2 <- alpha + sum_y2
posterior_beta2 <- beta + n2

x_seq <- seq(0, 15, length.out = 1000)

plot(x_seq, dgamma(x_seq, shape = alpha, rate = beta), type = 'l',
     col = 'blue', lwd = 2, ylim = c(0, 0.2), ylab = 'Density',
     xlab = expression(lambda), main = 'Gamma Distributions')

# Posterior plot for first scenario
lines(x_seq, dgamma(x_seq, shape = posterior_alpha1, rate = posterior_beta1),
      col = 'red', lwd = 2)

# Posterior plot for second scenario
lines(x_seq, dgamma(x_seq, shape = posterior_alpha2, rate = posterior_beta2),
      col = 'green', lwd = 2)

legend("topright",
       legend = c("Prior", "Posterior with n=5", "Posterior with n=20"),
       col = c("blue", "red", "green"), lwd = 2)
```

Problem 4: We want to model the number of siblings people have in a certain population. We can model the number of siblings a person has as a Poisson random variable Y for some unknown mean parameter λ. The probability mass function of $Y \sim Pois(\lambda)$ is

$$f(y|\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!},\ for\ y = 0, 1, 2, \dots$$

Suppose, before observing any data, we model our prior beliefs about λ by a gamma distribution $Gamma(\alpha, \beta)$ with hyperparameters $\alpha, \beta > 0$, i.e,

$$\pi(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda},\ for\ \lambda > 0$$

where $\Gamma(z) = \int x^{z-1}e^{-x}dx$. Note that + The prior mean is $E(\lambda) = \frac{\alpha}{\beta}$ + The prior variance is $Var(\lambda) = \frac{\alpha}{\beta^2}$ + The prior mode is $mode(\lambda) = \frac{\alpha-1}{\beta}$ Suppose we observe data $y_1, y_2, \dots y_n \sim^{iid} Pois(\lambda)$. Let $y = (y_1, y_2, \dots y_n)$

(a) Show that the gamma distribution is a conjugate prior for the Poisson likelihood. More specifically, show that the posterior distribution $\pi(\lambda|y)$ is of the form

$$\pi(\lambda|y) \sim Gamma(\alpha + \Sigma_{i=1}^{n}y_i, \beta + n)$$

Proof:

$$L(\lambda|y) = \frac{e^{-n\lambda}\lambda^{\Sigma y_i}}{\prod y!}$$

$$\pi(\lambda|y) \propto \frac{e^{-n\lambda}\lambda^{\Sigma y_i}}{\prod y!}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$

$$\pi(\lambda|y) \propto \lambda^{\Sigma y_i+\alpha-1}e^{-(n+\beta)\lambda}$$

$$\pi(\lambda|y) \sim Gamma(\alpha + \Sigma_{i=1}^{n}y_i, \beta + n)$$

(b) Show that the posterior mean $E(\lambda|y)$ is a weighted average of the prior mean and the sample mean $\bar{y} = \frac{1}{n}\Sigma_{i=1}^{n}y_i$. That is, find the weight w so that

$$E(\lambda|y) = w\frac{\alpha}{\beta} + (1-w)\bar{y}$$

Proof:

$$E(\lambda|y) = \frac{\alpha + \Sigma_{i=1}^{n}y_i}{\beta + n}$$

$$w\frac{\alpha}{\beta} + (1-w)\bar{y} = \frac{\alpha + n\bar{y}}{\beta + n}$$

$$w = \frac{\beta}{\beta + n}$$

(c) What is $lim_{n\to\infty}E(\lambda|y)$? What does this limit represent?

$$lim_{n\to\infty}E(\lambda|y) = \bar{y}$$

This limit represents that as we get more data, the posterior mean converges to the sample mean, and the influence of the prior diminishes.

An interval $[\ell(y), u(y)]$, based on the observed data $Y = y$, is a $100(1-\alpha)\%$ credible interval for θ if

$$P[\ell(y) < \theta < u(y)|Y = y] = 1 - \alpha.$$

The probability $1-\alpha$ is called the **(Bayesian) coverage probability**. The interpretation of a credible interval is that it describes the information about the location of the true value of θ *after* you have observed $Y = y$.

This is different from the frequentist interpretation of coverage probability, which describes the probability that the interval will cover the true value *before* the data is observed.

A Bayesian analogue to a frequentist confidence interval is to use posterior quantiles. If $\theta_{\alpha/2}$ and $\theta_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ posterior quantiles of θ, then

$$P(\theta_{\alpha/2} < \theta < \theta_{1-\alpha/2}|Y = y) = 1 - \alpha,$$

so $[\theta_{\alpha/2}, \theta_{1-\alpha/2}]$ is a $100(1-\alpha)\%$ quantile-based credible interval for θ.

Here the quantile function for the Beta distribution in R is `qbeta()`.

The Metropolis-Hastings algorithm induces a Markov chain that converges to the stationary (limiting) distribution $\pi(x)$ if the Markov chain satisfies certain regularity conditions:

- Irreducibility
  (All states communicate with each other.)
- Aperiodicity
  (Every state has a period of 1.)
- Positive recurrence
  (Expected time until the chain returns to any state is finite.)

Positive recurrence is trivially satisfied for irreducible and aperiodic Markov chains with finite state spaces.

Good Mixing Behavior

- If the spread of the proposal is too large:
  - The proposal moves through the state space quickly (good mixing behavior).
  - The proposed state may be far from the current state, and the probability of acceptance will be low.
- If the spread of the proposal is too small:
  - The probability of acceptance will be high.
  - The chain will take a long time to move through the state space, and low density regions will be undersampled.
- Both situations above will exhibit high **autocorrelation**: Correlation between subsequent values in the chain.
- An ideal proposal distribution will balance between good mixing behavior and a high acceptance rate.

The idea of Markov Chain Mote Carlo (MCMC) is to generate a Markov Chain such that its stationary distribution is the target distribution.

Main Goals:

- To generate a sequence of correlated samples from $\pi(x)$.
- To estimate $E_{\pi}[g(X)] = \int g(x)\pi(x)dx$.

Let G be any specified irreducible Markov transition matrix (kernel)

$G = i\begin{pmatrix} & j & \\ & g(j|i) & \end{pmatrix}$ and $g(\cdot|i)$ is a proposal dist. func. given i

when $X_t = i$, generate $Y \sim g(\cdot|i)$

If $Y = j$, then set $X_{t+1} = j$ w/ prob. $\alpha(i,j)$

If $Y \neq j$, then set $X_{t+1} = i$ w/ prob. $1-\alpha(i,j)$

$0 \le \alpha(i,j) \le 1$

$P(X_{t+1} = j | X_t = i) = g(j|i)\cdot\alpha(i,j)$

Recall: $\pi(x)Q(x,y) = \pi(y)Q(y,x)$

$\pi(i)g(j|i)\alpha(i,j) = \pi(j)g(i|j)\alpha(j,i)$

If set $\alpha(i,j) = \frac{\pi(j)g(i|j)}{\pi(i)g(j|i)}$, then $\alpha(j,i) = 1$

If set $\alpha(j,i) = \frac{\pi(i)g(j|i)}{\pi(j)g(i|j)}$, then $\alpha(i,j) = 1$

Set $\alpha(i,j) = \min\left\{\frac{\pi(j)g(i|j)}{\pi(i)g(j|i)}, 1\right\}$

$r = mh = \frac{\pi(j)g(i|j)}{\pi(i)g(j|i)}$

Metropolis-Hastings (M-H) sampling algorithm

There is a candidate point Y generated from a proposal distribution $g(.|X_t)$. If this candidate point is accepted, the chain moves to state Y at time $t+1$ and $X_{t+1} = Y$; otherwise the chain stays in state X_t and $X_{t+1} = X_t$. (Rizzo's book)

$\alpha(i,j) = \min\left\{\frac{\pi(j)g(i|j)}{\pi(i)g(j|i)}, 1\right\}$

If $\frac{\pi(j)g(i|j)}{\pi(i)g(j|i)} \uparrow$, most likely accept $X_{t+1} = j$

$\downarrow$, most likely accept $X_{t+1} = i$

$u \sim Unif(0,1)$

If $(u < mh)$ accept $X_{t+1} = j$

ow set $X_{t+1} = i$

M-H Algorithm $\pi(x)$

1. Choose a proposal distribution $g(.|X_t)$.
2. Generate X_0 from the distribution g.
3. Repeat until the chain has converged to stationary distribution according to some criterion:
4. Generate Y from $g(.|X_t)$
5. Generate U from Uniform(0, 1).
6. Compute $r = \frac{\pi(Y)g(X_t|Y)}{\pi(X_t)g(Y|X_t)}$
7. If (U < r), accept Y and set $X_{t+1} = Y$; otherwise set $X_{t+1} = X_t$
8. Increment t.

$\begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = \begin{pmatrix} - \end{pmatrix}_{1\times|V|}\begin{pmatrix} | \\ | \end{pmatrix}_{|V|\times 1}$

$\pi^T P = \pi^T$

Use the Metropolis-Hastings sampler to generate a sample from Standard Normal Distribution $N(0,1)$.

$\frac{\pi(y)g(X_t|y)}{\pi(X_t)g(y|X_t)} = \frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}g(X_t|y)}{\frac{1}{\sqrt{2\pi}}e^{-\frac{X_t^2}{2}}g(y|X_t)} = e^{-\frac{1}{2}(y^2 - X_t^2)}$

```
set.seed(9999) # for reproduceability
n <- 10000 # specifiy length of chain
X <- 0 # initialize chain
C <- 1
for (t in 2:n) {
     # Generate Y from proposal
     Y <- runif(1, X[t - 1] - C, X[t - 1] + C)
     # Compute MH ratio
     r <- min(1, exp(-0.5 * (Y^2 - X[t - 1]^2)))

     U <- runif(1, 0, 1) # Generate U from Unif(0,1)
     if (U <= r) {
         X[t] <- Y # Move to Y if U <= r
     } else{
         X[t] <- X[t - 1] # Stay at X[t - 1] if U > r
     }
}
```

Consider two urns A and B, where we place n black balls in urn A and n white balls in urn B. At each step, we randomly choose a ball from each urn and interchange the two balls. Let X_t, $t \geq 0$, be the number of black balls in urn A at step t. Please find the transition probabilities.

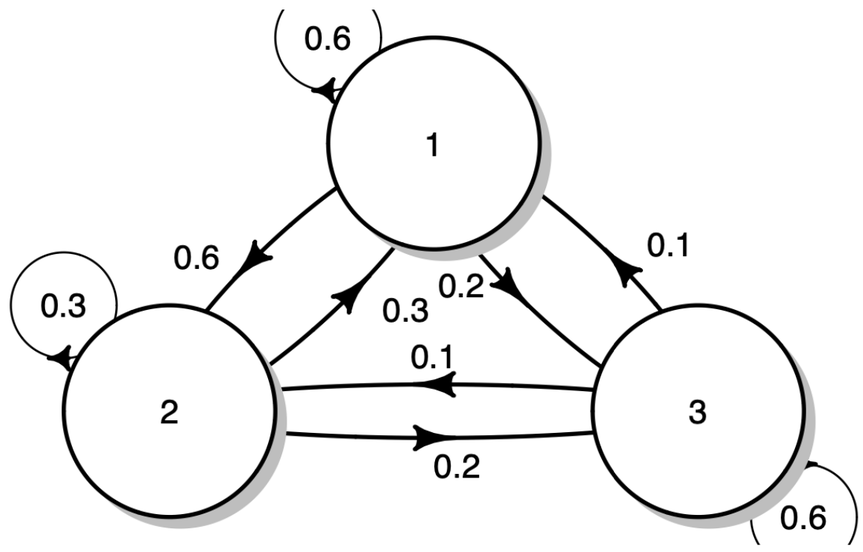
$$P(X_{\{t+1\}} = k+1 \mid Xt = k) = \frac{m-(n-k)}{m} \times \frac{k}{n}$$

$$P(X_{\{t+1\}} = k-1 \mid Xt = k) = \frac{(n-k)}{m} \times \frac{m-(n-k)}{n}$$

$$P(X_{t+1} = k \mid X_t = k) = 1 - P(X_{t+1} = k+1 \mid X_t = k) - P(X_{t+1} = k-1 \mid X_t = k)$$

```
##     1   2   3
## 1 0.6 0.3 0.1
## 2 0.6 0.3 0.1
## 3 0.2 0.2 0.6
```

```
plotmat(M,
        pos = c(1,2),
        self.cex = 0.5,
        self.shiftx = c(-0.1, -0.1, 0.1),
        self.shifty = c(0.1, 0.1, -0.1),
        main = "Transition Diagram")
```



Consider the transition matrix for the state space {0, 1} below.

$\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$. Can we find a stationary distribution $\pi = [\pi_0, \pi_1]$ with this transition matrix? If so, please find the equations for computing π_0 and π_1.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

1. π_0(1−a)+π_1b=π_0
2. $\pi_0 a + \pi_1(1-b) = \pi_1$
3. $\pi_0 + \pi_1 = 1$

transition matrix:

$\pi_0 = \frac{b}{a+b}$

$\pi_1 = \frac{a}{a+b}$

University of California, Los Angeles
Department of Statistics

Statistics 100B **Instructor: Nicolas Christou**

Probability Distributions - Summary

Discrete Distributions				
Distribution	Probability Mass Function	Mean	Variance	Moment-generating Function
Binomial	$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$ $x=0,1,\cdots,n$	np	$np(1-p)$	$[pe^t+(1-p)]^n$
Geometric	$P(X=x)=(1-p)^{x-1}p$ $x=1,2,\cdots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Negative Binomial	$P(X=x)=\binom{x-1}{r-1}p^r(1-p)^{x-r}$ $x=r,r+1,\cdots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$[\frac{pe^t}{1-(1-p)e^t}]^r$
Hypergeometric	$P(X=x)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ $x=0,1,\cdots,n$ if $n\le r$, $x=0,1,\cdots,r$ if $n>r$	$\frac{nr}{N}$	$n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}$	Fairly complicated!
Poisson	$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}$ $x=0,1,\cdots$	λ	λ	$exp[\lambda(e^t-1)]$
Continuous Distributions				
Distribution	Probability Density Function	Mean	Variance	Moment-generating Function
Uniform	$f(x)=\frac{1}{b-a}$ $a\le x\le b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Gamma	$f(x)=\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^\alpha\Gamma(\alpha)}$, $\alpha,\beta>0$, $x\ge 0$	$\alpha\beta$	$\alpha\beta^2$	$(1-\beta t)^{-\alpha}$
Exponential	$f(x)=\lambda e^{-\lambda x}$, $\lambda>0$, $x\ge 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$(1-\frac{1}{\lambda}t)^{-1}$
Beta	$f(x)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ $\alpha>0,\ \beta>0,\ 0\le x\le 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	
Normal	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ $-\infty<x<+\infty$	μ	σ^2	$e^{\mu t+\frac{t^2\sigma^2}{2}}$

Remarks:

Binomial:	X represents the number of successes among n trials.
Geometric:	X represents the number of trials needed until the first success.
Negative Binomial:	X represents the number of trials needed until r successes occur.
Hypergeometric:	X represents the number of items among the n selected that comes from the r group.
Poisson:	X represents the number of events that occur in time, area, etc.