AMS 597: Statistical Computing

Pei-Fen Kuan (c)

Applied Math and Stats, Stony Brook University

Sums of random variables

- Let $X_1, ..., X_n$ be independent and identically distributed with distribution $X_i \sim F$
- Then $S = X_1 + ... + X_n$ is a convolution of X_i
- Thus, to simulate random variables S, first simulate $X_1, ..., X_n$, then compute the sum

Sums of random variables

- Some useful results:
  - If $Z \sim N(0, 1)$, then $Z^2 \sim \chi^2_1$.
  - If $W_i \sim \chi^2_{k_i}$, then $\sum_{i=1}^n W_i \sim \chi^2_p$, where $p = \sum_{i=1}^n k_i$

Sums of random variables

- Exercise: Simulate 1000 chi-squared r.v. with degrees of freedom 3 using sum of r.v's method. Use only random uniform generator.

Mixtures

- A random variable X is a discrete mixture if the distribution of X is weighted sum $F_X(x) = \sum_i p_i F_{X^i}(x)$ for some sequence of random variables $X_1, X_2, ...$ and $p_i > 0$ such that $\sum_i p_i = 1$
  1. Generate a multinomial r.v $Y \sim \text{Multi}(p_1, ..., p_k)$
  2. Generate X from $F_{X^Y}(x)$

Mixtures

- Exercise: Simulate 10000 r.v from mixture of normals $0.2N(0, 1) + 0.5N(-1, 1) + 0.3N(2, 1)$

Mixtures

- Example of continuous mixture: The negative binomial distribution with mean $pr/(1-p)$ and variance $pr/(1-p)^2$ is a continuous mixture of Poisson distribution where the Poisson rate is a gamma distribution
- That is, we can view the negative binomial as a $Poisson(\lambda)$ distribution, where λ is itself a random variable, distributed as a gamma distribution with shape r and scale $s = p/(1-p)$ (rate $r = 1/s$)
- Note that $X \sim NB(r, p)$ if

$$p(k) = \binom{k+r-1}{k}(1-p)^r p^k, \quad k \in \{0, 1, 2, 3, ...\}$$

Mixtures

- Exercise: Simulate 10000 negative binomial with $p = 0.4$ and $r = 10$ using Gamma-Poisson mixture.

Multivariate Distributions

- Multivariate normal distribution

$$f(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\{-(1/2)(x-\mu)^T\Sigma^{-1}(x-\mu)\}, \quad x \in \mathbb{R}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd} \end{bmatrix}$$

Multivariate Distributions

- To generate a random sample of size n from the $N_d(\mu, \Sigma)$ distribution:
  1. Generate an $n \times d$ matrix Z containing nd random $N(0,1)$ variates (n random vectors in R^d)
  2. Compute a factorization $\Sigma = Q^TQ$
  3. Apply the transformation $X = ZQ + J\mu^T$, where J is a column vector of 1's
  4. Each row of X is a random variate from the $N_d(\mu, \Sigma)$ distribution

Square root matrix

- Related to SVD and spectral decomposition
- $\Sigma = P\Delta P^T$, $Q = P\Delta^{1/2}P^T$, where Δ is the diagonal matrix with the eigenvalues of Σ along the diagonal and P is the matrix whose columns are the eigenvectors of Σ corresponding to the eigenvalues in Δ
- Solve $\det(\Sigma - \lambda I) = 0$ to get eigenvalues
- Solve for eigenvectors e from $(\Sigma - \lambda I)e = 0$
- In R, use `eigen()`

Example

- Simulate multivariate normal with mean $(0, 0, 0)$ and variance

$$\Sigma = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

using spectral decomposition

Cholesky factorization

- Every positive definite matrix A can be factored $A = LL^T$ where L is lower triangular with positive diagonal elements
- L is called the Cholesky factor of A
- Thus, to generate multivariate normal using Cholesky decomposition, set $Q = L^T$ (upper triangular matrix)
- In R, use `chol()`
- Reference: Algorithm for Cholesky decomposition http://www.math.sjsu.edu/~foster/m143m/cholesky.pdf

Example

- Simulate multivariate normal with mean $(0, 0, 0)$ and variance

$$\Sigma = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

using Cholesky decomposition