CHAPTER 1

Sampling from Special Distributions

1. Preliminary Facts

- **Definition:** If $\sum = (\sigma_{ij})$ is a $k \times k$ real-valued matrix, we say $\sum$ is a positive definite if
  1. $\sigma_{ij} = \sigma_{ji}$ i.e. $\sum$ is symmetric
  2. $x'\sum x \geq 0$, for all $x \in \mathbb{R}^k$ with $x'\sum x = 0$ if and only if $x = 0$.

  **Note:** $x'\sum x = \sum_{i,j=1}^{k} \sigma_{ij}x_ix_j$

- **Definition:** If $P$ is a $k \times k$ matrix we say $P$ is orthogonal if $P'P = PP' = I_k$.

- **Spectral Theorem** Let $\sum$ be $k \times k$ positive definite, then there exists an orthogonal $k \times k$ matrix $P$ and a diagonal matrix $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_k \end{bmatrix}$ such that $P\sum P' = D$ with $\lambda_i > 0$.

  **Remark** Let the rows of $P$ be denoted by $p_1, \cdots, p_k$, so that

  $p = \begin{bmatrix} p_1 \\ \vdots \\ p_k \end{bmatrix}$

  $p' = \begin{bmatrix} p_1' \\ \cdots \\ p_k' \end{bmatrix}$

  Then we have

  $\sum p' = p'D$

  i.e. $\sum p_i' = \lambda_ip_i'$

  That is $p_i'$ is an eigenvector(normalized) with eigenvalue $\lambda_i$.

- **Corollary (Cholesky Decomposition)** If $\sum$ is $k \times k$ positive definite, then there exists a $k \times k$ matrix $U$ such that $\sum =UU'$

  **Proof:**

1
Write $\Sigma = PD P'$ and $D = SS'$ where $S = \begin{pmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_k} \end{pmatrix}$

Then take $U = PS, U' = S'P'$.

**Corollary** If $\Sigma$ is $k \times k$ positive definite then there exists a $k \times k$ positive definite matrix $V$ such that $V \Sigma V' = I_k$.

**Proof:**
Write $\Sigma = PD P'$ as above and take $V = S_1 P_0$.

- **Definition** If $X = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix}$ is a random column $k$-vector we define

  $$E(X) = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_k) \end{bmatrix}$$

  $$\sum_x = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq k}$$

- **Proposition** If $A$ is an $n \times k$ matrix, $b$ is an column $n$-vector and $X$ is a random column $k$-vector then

  $$E(AX + b) = A(E(X)) + b$$

  $$\sum_{AX+b} = A \sum A'$$

  **Proof:** Exercise.

- **Definition** Given a $k$-vector $\mu$ and a $k \times k$ positive definite matrix $\Sigma$ we say $x \sim N_k(\mu, \Sigma)$ if $X$ has density function

  $$f_X(x) = (2\pi)^{-\frac{k}{2}}(\det(\Sigma))^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(x - \mu)' \Sigma^{-1} (x - \mu)\right\}$$

- **Properties**

  1. If $X \sim N_k(\mu, \Sigma)$ then for an invertible $k \times k$ matrix $A$ and a $k$-vector $b$ we have

     $$AX + b \sim N(A\mu + b, A \Sigma A')$$

     **Proof:** Transformation.

  2. If $X \sim N_k(0, I_k)$ then $X_i$ are iid $N(0, 1)$.

     **Proof:** $f_X(x) = \prod_{i=1}^k \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2}$

  3. $X \sim N_k(0, I_k)$ \Rightarrow $E(X) = 0, \sum X = I_k$

     **Proof:** $E(X) = (E(X_1), \cdots, E(X_k))$, $\sum X = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq k}$

  4. $X \sim N_k(\mu, \Sigma) \Rightarrow E(X) = \mu, \sum X = \Sigma$

     **Proof:**

     Let $Y = A(X - \mu) = AX - A\mu$ where $A \sum A' = I_k$, then
\[ E(Y) = E(AX - A\mu) = AEX - A\mu = 0 \]
\[ \sum_Y = A \sum A' = I_k. \]
Since \( X = A^{-1}Y + \mu \) we have
\[ E(X) = \mu, \sum_X = A^{-1} \sum_Y A^{-1}' = A^{-1}I_kA^{-1}' = \sum, \]
Note that
\[ X = A^{-1}Y + \mu \]
\[ Y \sim N_k(0, I_k) \]
so we can generate \( X \sim N_k(0, I_k) \) as follows:
Generate \( Y_1, \ldots, Y_k \) iid \( N(0, 1) \)
Take \( X = A^{-1}Y + \mu \)
where \( A^{-1} \) is obtained from Cholesky-Decomposite \( \sum(= A^{-1}A^{-1}') \).

2. \( N(0, 1) \) Deviates via Polar Transformation

If \( X, Y \) indep. \( N(0, 1) \) define \( R = \sqrt{X^2 + Y^2}, \Theta = arg(X, Y) \in (-\frac{\pi}{2}, \frac{\pi}{2}) \).
Then
\[ \Theta \sim \text{Uniform in } (-\pi, \pi) \]
\[ R^2 \sim \chi^2_2 \] i.e. \( f_{R^2}(x) = \frac{1}{2}xe^{-\frac{x}{2}} \)
This implies we can generate \( (X, Y) \sim \text{indep } N(0,1) \) by taking Bux-Muller method:
\[ X = \sqrt{2E}\cos(2\pi U) \]
\[ Y = \sqrt{2E}\sin(2\pi U) \]
where \( E \sim \xi(1), U \sim \text{ Unif } (0,1), E, U \) indep
Note that Box Muller requires:
(1) square root cal.
(2) trig functions.

**Alternative Methods** Instead of generating \( \Theta \sim \text{ Unif } [0, 2\pi] \) and applying a trig transform we can take
\[ (X, Y) \sim \text{ Uniform in Unit Dist } \{(x, y) : x^2 + y^2 \leq 1\} \]
and then use
\[ \sqrt{2E(\frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}})} \]
i.e.
Generate \( X, Y \) indep. \( U(0,1) \)
Until \( X^2 + Y^2 \leq 1 \)
Take \( E = -\log(U) \quad U \sim U(0,1) \)
Return \( \sqrt{2E(\frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}})} \).
3. Stationary Gaussian Processes

3.1. Definition. We use $Z$ to denote the set of integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$. A collection of random variables $\{X_t, t \in Z\}$ is said to form a strongly stationary stochastic process if for every $s$, the distribution of $(X_t, X_{t+1}, \cdots, X_{t+s})$ does not depend on $t$.

The sequence $\{X_t, t \in Z\}$ is said to be stationary Gaussian if it is stationary and for every $k$, $(X_t, \cdots, X_{t+k-1})$ has a $k$-variate normal distribution.

Remark: If $\{X_t, t \in Z\}$ is stationary Gaussian then $E(X_t)$ is constant $\mu$ and the distribution of any $(X_t, X_{t+1}, \cdots, X_{t+k-1})'$ is determined completely by this constant its covariance matrix

$$\sum = (\sigma_{ij})_{1 \leq i, j \leq k}$$

where

$$\sigma_{ij} = Cov(X_{t+i-1}, X_{t+j-1}) = Cov(X_i, X_j) = \gamma_{j-i} = \gamma_{i-j}$$

We call $\gamma_k$ the auto covariance at lag $k$.

Observe that

$$\gamma_{-k} = \gamma_k$$

If $X_t$ is a sequence of jointly Gaussian r.v.'s then it is stationary Gaussian if and only if $Cov(X_t, X_{t+s})$ does not depend on $t$, only on $s$.

Definition: A sequence $\{\gamma_k, k \in Z\}$ is said to be positive semidefinite if for any choice of constants $c_k$, we have

$$\sum c_i c_j \gamma_{i-j} \geq 0$$

and positive definite if equality holds only when $c_k = 0$ for all $k$.

Proposition: If $\{X_t, t = \cdots, -2, -1, 0, 1, 2, \cdots\}$ is stationary Gaussian then its auto covariance function $\gamma_k = Cov(X_t, X_{t+k})$ is positive semidefinite.

Proof:

$$0 \leq Var(\sum_t c_t X_t) = \sum_{s,t} c_s c_t Cov(X_s, X_t) = \sum_{s,t} c_s c_t \gamma_{s-t}$$

3.2. Examples of Stationary Gaussian Processes.

- Gaussian white noise (GWN)
  $$\epsilon_t \text{ iid } N(0, 1)$$
  $$(\epsilon_t, \epsilon_{t+1}, \cdots, \epsilon_{t+k-1})' \sim N(0, I_k)$$
• Filtered Gaussian white noise
If \( \epsilon_t \) is GWN and we define
\[
X_t = \sum_{j=-\infty}^{\infty} c_{t-j} \epsilon_j \text{ i.e. } X = c \ast \epsilon
\]
where \( c_i \) are constants such that \( \sum c_i \leq +\infty \)
then \( X_t \) are jointly Gaussian, and

\[
Cov(X_t, X_{t+s}) = Cov\left( \sum_j c_{t-j} \epsilon_j, \sum_j c_{t+s-l} \epsilon_l \right)
\]
\[
= \sum_j \sum_l c_{t-j} c_{t+s-l} Cov(\epsilon_j, \epsilon_l)
\]
\[
= \sum_j c_{t-j} c_{t+s-j} = \sum_u c_u c_{u+s}
\]

Suppose \( c_i \) are given, and \( c_i = 0 \) for \( |i| > k \), so that
\[
X_t = \sum_i c_{t-i} \epsilon_i
\]
\[
= \sum_u c_u \epsilon_{t-u}
\]
\[
= \sum_{|u| \leq k} c_u \epsilon_{t-u}
\]
and we want to generate a sample \( X_1, X_2, \cdots, X_N \) having the prescribed distribution. We could proceed as follows:

Generate \( \epsilon_i \) iid \( N(0,1) \) for \( i = 1 - K, \cdots, N + k \)
Compute each \( X_t = \sum_{|u| \leq k} c_u \epsilon_{t-u}, t = 1, \cdots, N \)

How many operations Requied?
Each iteration
\( 2K + 1 \) multiplications+additions
\( N \) iterations
\( O(N \times (2K + 1)) \)
Can we do better?
Discrete Fourier Transforms
Definition: Given constant \( x_j, j = 0, \cdots, N - 1 \) we define
\[
\hat{x}_k = \sum_{j=0}^{N-1} x_j \exp\left(\frac{2\pi i j k}{N}\right) \quad k = 0, \cdots, N - 1
\]
\[
\overline{x}_k = \frac{1}{N} \sum_{j=0}^{N-1} x_j \exp\left(-\frac{2\pi i j k}{N}\right)
\]
Facts:
(1) \( \overline{\hat{x}} = x \)
(2) \((x*y)_u = \hat{x}_u \cdot \hat{y}_u\)

(3) FFT and BFT can compute \(\hat{x}, \hat{y}\) for a sequence of size \(N\) in \(O(N \log N)\) time.

Strategy for generating sequence \(X = c * \epsilon\)
- Compute \(\hat{c}, O(N \log N)\)
- Generate \(\epsilon, O(N \log N)\)
- Transform to get \(\hat{\epsilon}, O(N \log N)\)
- Pointwise multiply \(\hat{X} = \hat{\epsilon} \cdot \hat{c}, O(N)\)
- Inverse transform \(X = \hat{X}, O(N \log N)\).

Note: work is no longer \(O(K \cdot N)\) if \(K = N\) would get quadratic.

Question: What if we are given \(\gamma_s = \text{Cov}(X_t, X_{t+s})\) and we want to find \(c\) such that
\(X_t = \sum c_{t-j} \epsilon_j\) where \(\epsilon_j\) iid \(N(0, 1)\).
has this \(\gamma_s\) as its autocovariances function.

It will be advantageous to start with a different problem.

Definition: A sequence \(\{\gamma_i, i = 0, \ldots, N - 1\}\) of real numbers is positive semidefinite if for every choice of complex constants \(c_i\) we have

\[
\sum_{i,j=0}^{N-1} c_i c_j \gamma_{i-j} \geq 0
\]

Fact: If \(\{X_t, t = 0, \ldots, N - 1\}\) is stationary and we define
\(\gamma_i = \text{Cov}(X_t, X_{t+i})\)
then \(\{\gamma_i, i = 0, \ldots, N - 1\}\) is psd.

Remarks:

(1) If \(Z, W\) are complex-valued rv's, i.e. \(Z = X + iY, W = U + iV\), we define
\[E(Z) = E(X) + iE(Y)\]
\[\text{Cov}(Z, W) = E(Z\overline{W}) - E(Z)E(\overline{W})\]
\[\text{Var}(Z) = \text{Cov}(Z, Z)\]
\[= E[(X+iY)(X-iY)] - E[X+iY]E[X-iY]\]
\[= E(X^2+Y^2) - [E^2(X)+E^2(Y)] = \text{Var}(X) + \text{Var}(Y)\]

All of the usual properties of \(E\) & \(\text{Cov}\) go through.

(2) If \(\{X_t, t = 0, \ldots, N - 1\}\) is a real-valued stationary process and we define \(\gamma_k = \text{Cov}(X_t, X_{t+k})\), then for complex-constants \(c_j\) we have

\[
0 \leq \text{Var} \left( \sum_{j=0}^{N-1} c_j X_j \right) = \sum_{j,k=0}^{N-1} c_j c_k \gamma_{j-k}
\]
3. STATIONARY GAUSSIAN PROCESSES

So \( \gamma_k \) is a p.s.d. sequence.

**Theorem** If \( \{ \gamma_i, i = 0, \cdots, N-1 \} \) is a psd sequence then its DFT is nonnegative, and satisfies \( \tilde{\gamma}_p = \tilde{\gamma}_{N-p} \).

**Proof:**
Define \( z_j^{(p)} = exp \left( \frac{2\pi ijp}{N} \right) \), then

\[
0 \leq \sum_{j,k=0}^{N-1} z_j^{(p)} \tilde{z}_k \gamma_{j-k} = \sum_{j,k=0}^{N-1} exp \left( \frac{2\pi ijp}{N} \right) exp \left( -\frac{2\pi ikp}{N} \right) \gamma_{j-k} = \sum_{j,k=0}^{N-1} exp \left( \frac{2\pi p(j-k)}{N} \right) \gamma_{j-k} = N \sum_{u=0}^{N-1} exp \left( \frac{2\pi pu}{N} \right) \gamma_u = N \tilde{\gamma}_p
\]

**Theorem** Let \( \{ \gamma, t = 0, \cdots, N-1 \} \) be p.s.d. with \( \gamma_s = \gamma_{-s} \)
If we define \( c_p = \sqrt{\tilde{\gamma}_p} \) i.e. \( \tilde{c} = \sqrt{\tilde{\gamma}} \)
i.e. \( \sum_{j=0}^{N-1} c_j X_j \) has \( \gamma \) as its autocovariance function.

**Proof:** Let \( h_s = \sum_{p=0}^{N-1} c_p c_{p+s} = \sum_{p=0}^{N-1} c_p c_{-(p+s)} = (c \ast c)_{-s} \), then to check that \( h_s = \gamma \) we check that the DFT’s match up.

\[
\hat{h}_u = (c \ast c)_{-u} = \tilde{c}_{-u} \tilde{c}_{-u} = \tilde{c}_u \tilde{c}_u = \sqrt{\tilde{\gamma}_u} \sqrt{\tilde{\gamma}_u}
\]

Now that we’ve seen that all is well in the circles, we ask what happens when \( X_t \) lives in \( \mathbb{D} \).

**Idea:** Start with \( \{ X_t, t = 0, \cdots, k-1 \} \) having as its autocovariance function \( \gamma_0, \cdots, \gamma_{k-1} \) and “embed” in a larger sequence that lives in \( \mathbb{D}_N \) for some \( N \).

**Minimal embedding:**
Define \( \Sigma = \begin{bmatrix} \gamma_0 & \cdots & \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_1 \\ \vdots \\ \gamma_{k-1} & \gamma_{k-1} \\ \gamma_{k-2} \\ \vdots \\ \gamma_1 \end{bmatrix} \).

If this is positive definite then we proceed as before, sample the larger process.

Suppose \( f \) is a monotonically decreasing and convex function and we define
\[
\gamma_i = f(i \Delta), \ i = 0, \cdots, k - 1,
\]
then the matrix \( \Sigma \) defined by the above minimal embedding is p.s.d.

Procedure:

Start with \( \gamma_0, \gamma_1, \cdots, \gamma_{k-1} \) for sampling \( X_0, \cdots, X_{k-1} \).

\[
\begin{bmatrix} \gamma_0 & \cdots & \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_1 \\ \vdots \\ \gamma_{k-1} & \gamma_{k-1} \\ \gamma_{k-2} \\ \vdots \\ \gamma_1 \end{bmatrix}
\]

Process is stationary Gaussian in the circle, so use circle method.
Restrict to \( X_0, \cdots, X_{k-1} \). Done.