Statistical Analysis of Simulation Data

(Bolí Spring 2019)

(A) Data Processing
(B) I.I.D. Output
(C) Stationary Output
(D) Asymptotically stationary Output
(E) Empirical CDF
(F) Kernel Density Estimation

(A) Data Processing

Visualization

Histogram: Divide an underlying interval in \( \mathbb{R} \) into small intervals. Count how many data points fall in each small interval. With each small interval, plot a rectangle with height or volume equal to the number or frequency of the data points in that interval.

Scatter plot: Plot data points in \( \mathbb{R}^d \) for \( d = 1, 2, \text{ or } 3 \).

Empirical CDF: (Cumulative Distribution Function) Approximation of a true CDF, represented as a graph of a one-variable function.

Density plot: Approximation of a true PDF for a real random variable.
Data Characterizing Numbers

Let \( X_1, \ldots, X_n \) be i.i.d. random variables in \( \mathbb{R}^d \).
The \textbf{sample mean} is
\[
\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.
\]

If \( d = 1 \) and \( X(1) \leq \cdots \leq X(n) \), then the \textbf{sample median} is
\[
X\left(\frac{n+1}{2}\right) \quad \text{if } N \text{ is odd},
\]
\[
\left\lfloor X\left(\frac{n}{2}\right) + X\left(\frac{n}{2} + 1\right) \right\rfloor / 2 \quad \text{if } N \text{ is even}.
\]
The \textbf{range} is \( x_n - x_1 \).

The \textbf{sample variance} for \( X_1, \ldots, X_n \in \mathbb{R}^d \) is
\[
S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{1}{n-1} \left( \sum_{i=1}^{n} X_i^2 - n \overline{X}^2 \right)
\]
where \( \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \) and \( \|A\|^2 = \|A\|^2 \) (\( A \in \mathbb{R}^d \)).

The \textbf{sample standard deviation} is \( S \) (with \( S^2 \) the sample variance).

The sample \( k \)-th moment is
\[
\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^k
\]
with \( \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \).

For all the \textbf{sample \( r \)-quantile} or \( X \times 100 \) percentile of \( X_1, \ldots, X_n \) is \( X(f \times n) \), where \( X(1) \leq \cdots \leq X(n) \)
and \( \lceil x \rceil \) is the smallest integer \( \geq x \).
Let \((X_1, Y_1), \ldots, (X_n, Y_n) \in \mathbb{R}^m \times \mathbb{R}^m\) be i.i.d. random vectors sampled from a bivariate distribution. The sample covariance is

\[
\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}),
\]

where \(\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i\) and \(\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i\). The sample correlation coefficient is

\[
\frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}.
\]

Let \(X\) be a random variable of finite expectation an \(\mu = \mathbb{E}(X)\). Suppose \(\approx\) approximate \(\hat{\mu} \approx \mu\) and \(\rho > 0\) satisfy

\[
P\{ \mu \in [\hat{\mu} - \rho, \hat{\mu} + \rho] \} = 0.95\ (\sigma \approx)
\]

then we say that the interval \([\hat{\mu} - \rho, \hat{\mu} + \rho]\) is of 95\% (or \(\approx\) percentage) confidence

\[(\text{13})\ I.I.D.\ Output\ .\]

Let \(X_1, \ldots, X_n\) be random variables in \(\mathbb{R}\), i.i.d. according to a density \(f\) with particularly \(\mu = \mathbb{E}(X_i)\) and \(\sigma^2 = \text{Var}(X_i)\), both finite.

Denote \(\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i\). Then

\[
\mathbb{E}(\bar{X}_n) = \mu \quad \text{and} \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.
\]
The central limit theorem implies that for large $N$,
\[
\frac{X_n - \mu}{\sqrt{S_n^2 / N}} \sim N(0, 1)
\] approximately.

One can verify for the sample variance
\[
S_n^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X}_n)^2
\]
that $E(S_n^2) = \sigma^2$ (exactly) and
\[
\lim_{N \to \infty} S_n^2 = \sigma^2
\]
with probability 1. Thus, for $N \gg 1$,
\[
\frac{X_n - \mu}{\sqrt{S_n^2 / N}} \sim N(0, 1)
\] approximately.

Let $\Phi(x)$ be the CDF of the standard normal distribution. Given $\alpha \in (0, 1)$, let $z_\alpha$ be the unique real number such that $1 - \Phi(z_\alpha) = \alpha$, i.e.,
\[
1 - \alpha = \Phi(z_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_\alpha} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{z_\alpha} \phi(x) dx
\]
where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Let $r > 0$ and $\alpha \in (0, 1)$. We have
\[
\mathbb{P}\left( \frac{\bar{X}_n - \mu}{S_n / \sqrt{N}} \in [-r, r] \right) = \alpha
\]
\[
\mathbb{P}\left( \bar{X}_n - \mu \in \left[ -r \frac{S_n}{\sqrt{N}}, r \frac{S_n}{\sqrt{N}} \right] \right) = \alpha
\]
\[
\int_{-r}^{r} \Phi(x) dx = \alpha
\]
\[
\frac{\text{erf}(z_\alpha)}{2} = \frac{\text{erf}(r)}{2}
\]
where $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2} dt$.

Using, $\int_{-\infty}^{\infty} \phi(x) dx = 1$ and $\phi(-x) = \phi(x)$.
Therefore, \( \left[ \overline{x}_n - r \frac{s}{\sqrt{n}}, \overline{x}_n + r \frac{s}{\sqrt{n}} \right] \) is approximately (when \( n \gg 1 \)) a \%\% \% confidence interval for \( \mu \), where \( r \) is determined by \( r = \Phi^{-1} \left( \frac{\alpha + 1}{2} \right) \).

For example, \( \gamma = 0.95 \), \( \alpha = 0.95 \), \( \Phi^{-1} \left( \frac{0.9 + 1}{2} \right) = 1.645 \).

Generalization to vector-valued random variables

Let \( \overline{x}_1, \ldots, \overline{x}_n \) be random variables in \( \mathbb{R}^d \) i.i.d. with the distribution \( f \). Suppose the common expectation and variance are \( \overline{x} = E(\overline{x}_i) \) and \( \sigma^2 = \text{Var}(\overline{x}_i) \), both finite. Then, an approximate \( (1 - \alpha) \) confidence region for \( \overline{x} \) is

\[
\left\{ \overline{x} \in \mathbb{R}^d : \left( \overline{x}_n - \overline{x} \right)^T \left( \frac{1}{\sigma^2} \Sigma^{-1} \left( \overline{x}_n - \overline{x} \right) \right) \leq \frac{\chi_{d, \alpha}^2}{n} \right\}
\]

where \( \Sigma = \frac{1}{n-1} \sum_{i=1}^n (\overline{x}_i - \overline{x}) (\overline{x}_i - \overline{x})^T \) is the sample covariance matrix, and \( \chi_{d, \alpha}^2 \) is the \( \alpha \)-quantile of \( \chi_d^2 \) distribution.
The Delta Method

Suppose \( \bar{x}_1, \ldots, \bar{x}_n \sim \text{iid } f : \mathbb{R}^d \to [0, \infty) \). Then
\[
\frac{1}{\sqrt{n}} \left( \bar{x}_n - \bar{x} \right) \xrightarrow{\text{dist.}} \bar{X} \sim N(0, \Sigma),
\]
where \( \bar{x} = \frac{1}{n} \sum_i x_i \). Then, for any \( C^1 \)-function \( \tilde{g} \), we have
\[
\frac{1}{\sqrt{n}} \left( \tilde{g}(\bar{x}_n) - \tilde{g}(\bar{x}) \right) \xrightarrow{\text{dist.}} \bar{R} \sim N(0, J \Sigma J^T),
\]
where
\[
J = J_{\tilde{g}}(\bar{x}) = \left( \frac{\partial \tilde{g}_i(\bar{x})}{\partial x_j} \right)
\]
matrix of \( \tilde{g} \) at \( \bar{x} \).

Simple proof.
\[
\tilde{g}(\bar{x}_n) = \tilde{g}(\bar{x}) + J_{\tilde{g}}(\bar{x}) (\bar{x}_n - \bar{x}) + o(\|\bar{x}_n - \bar{x}\|^2).
\]

As \( N \to \infty \),
\[
\frac{1}{\sqrt{n}} \left( \tilde{g}(\bar{x}_n) - \tilde{g}(\bar{x}) \right)
\approx \frac{1}{\sqrt{n}} J_{\tilde{g}}(\bar{x}) (\bar{x}_n - \bar{x})
\rightarrow J_{\tilde{g}}(\bar{x}) \bar{R} = \bar{R},
\]
where \( \bar{R} \sim N(0, \Sigma) \). Thus, \( \bar{R} = J_{\tilde{g}}(\bar{x}) \bar{K} \sim N(0, J \Sigma J^T), \) \( J = J_{\tilde{g}}(\bar{x}) \).
(C) Stationary Output

We now consider a stationary stochastic process \( X_1, X_2, \ldots \) where each \( X_j \in \mathbb{R} \) is a random variable. We recall that the stationarity means that, for any integers \( n \geq 1 \) and \( k \geq 1 \), and any \( x_1, \ldots, x_n \in \mathbb{R} \), the joint distributions

\[
P(X_1 < x_1, \ldots, X_n < x_n) = P(X_{n+k} < x_{n+k})
\]

In particular,

\[
P(X_k < x) = P(X_i < x) \quad \forall n \geq 1, \forall x \in \mathbb{R},
\]

i.e., all \( X_k (k \geq 1) \) have the same distribution.

We set \( \mu = E(X_k) \), \( \sigma^2 = \text{Var}(X_k) \), assumed to be finite. Moreover, for any \( i \in \mathbb{N} \) and any \( k \in \mathbb{N} \),

\[
\text{Cov}(X_i, X_{i+k})
\]

is independent of \( i \).

By the Law of Large Numbers and the Central Limit Theorem, we have with \( \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \) that

\[
\lim_{n \to \infty} \frac{\bar{X}_n - \mu}{\sqrt{\text{Var}(\bar{X}_n)}} \sim N(0,1)
\]

approximately for \( N \gg 1 \).

If \( \hat{\sigma}_n \) is a good estimator of \( \text{Var}(\bar{X}_n) \), then, for \( \gamma \in (0,1) \), we can determine \( \mu, \sigma^2 \) with

\[
\gamma = \mathbb{P}(\frac{\bar{X}_n - \mu}{\hat{\sigma}_n} < \frac{1}{\sqrt{n}}), \quad \mathbb{P}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma^2} e^{-x^2/2\sigma^2} dx.
\]
So that 

\[
\left[ \bar{X}_n - r_{\alpha} \sqrt{\frac{V_n}{n}}, \bar{X}_n + r_{\alpha} \sqrt{\frac{V_n}{n}} \right]
\]

is a $\alpha$-confidence interval for $\mu$.

However, unlike the i.i.d. case, the sample variance

\[
S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2
\]

is no longer a good estimator of $\text{Var}(\bar{X}_n)$ in general. It is still true that

\[
\lim_{n \to \infty} S_n^2 = \text{Var}(X_i) = \sigma^2 \quad \text{with prob 1.}
\]

But in general:

\[
\text{Var}(\bar{X}_n) \neq \frac{\text{Var}(X_1)}{n},
\]

since $X_i$'s are not necessarily independent.

Observe that

\[
\text{Var}(\bar{X}_n) = \frac{1}{n^2} \text{Var}(X_1 + \cdots + X_n)
\]

\[
= \frac{1}{n^2} \sum_{i,j=1}^{n} \text{Cov}(X_i, X_j)
\]

Define

\[
C(k) = \text{Cov}(X_i, X_i+k) \quad \forall i \geq 1, \forall k \geq 0.
\]

1. $C(k)$ is independent of $i \geq 1$ by the stationarity.
2. \( C(0) = \text{Cov}(X_i, X_i) = \text{Var}(X_i) = \sigma^2 \quad (\forall i \geq 1). \)
3. If \( k \in \mathbb{Z} \) and \( k < 0 \), we define \( C(k) = C(-k) \).

This definition is consistent with the case
that $k > 0$. If $k \in \mathbb{Z}$, $k < 0$, we can choose $i \in \mathbb{Z}$ s.t. $i + k > 0$. Hence,

\[
C(-k) = \text{Cov}(X_i, X_{i-k}) = \text{Cov}(X_{i+k}, X_i) = \text{Cov}(X_i, X_{i+k}) = C(k).
\]

The function $C(k)$ ($k \in \mathbb{Z}$) is the (auto) covariance function of the process $X(n = 1, 2, \ldots)$.

Now, we have

\[
\text{Var}(\overline{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} C(j-i) = \frac{1}{n^2} \left[ n \sum_{k=0}^{n} C(k) + (n-1)C(1) + (n-1)C(-1) + \cdots \right] = \frac{1}{n^2} \left[ \sum_{k=-n}^{n} C(k) \right] = \frac{1}{n} \sum_{k=-n}^{n} \left( 1 - \frac{|k|}{n} \right) C(k).
\]

**Proposition.** Assume

\[
V := \sum_{k=-\infty}^{\infty} C(k) = \sum_{k=-\infty}^{\infty} \text{Cov}(X_k, X_{k+n})
\]

converges absolutely. Then

\[
\lim_{n \to \infty} n \cdot \text{Var}(\overline{X}_n) = V.
\]

**Proof.** By the above calculations we have

\[
N \cdot \text{Var}(\overline{X}_n) = \sum_{k=-n}^{n} \left( 1 - \frac{|k|}{n} \right) C(k) = \sum_{k=-n}^{n} C(k) - \frac{1}{n} \sum_{k=-n}^{n} |k| C(k).
\]

Since \( \sum_{k=-\infty}^{\infty} C(k) = V \) with absolute convergence it suffices to show that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=-n}^{n} |k| C(k) = 0. \tag{1}
\]

\( \forall \varepsilon > 0, \) since \( \sum_{k=-\infty}^{\infty} |k| C(k) \leq \infty \), there exists \( N \in \mathbb{N} \)
such that
\[ \sum_{|k| > N_0} |c(k)| < \varepsilon. \]

Thus, for \( N > N_0 \),
\[ \sum_{k=-N}^{N} \frac{|k|}{N} |c(k)| = \frac{1}{N} \sum_{|k| \leq N_0} |k| |c(k)| \]
\[ + \sum_{|k| > N_0} \frac{|k|}{N} |c(k)| \]
\[ \leq \frac{1}{N} \sum_{|k| \leq N_0} |k| |c(k)| + \sum_{|k| > N_0} |c(k)| \]
\[ \leq \frac{1}{N} \sum_{|k| \leq N_0} |k| |c(k)| + \varepsilon. \]

Thus,
\[ \lim_{N \to \infty} \sum_{k=-N}^{N} \frac{|k|}{N} |c(k)| = \varepsilon. \]

Hence, since \( \varepsilon > 0 \) is arbitrary, \( (\text{iv}) \) is true.

The Covariance Method

We continue our discussions and now study how to estimate \( \nu = \sigma^2(0) + 2 \sum_{k=1}^{\infty} \sigma(k) \), where \( \sigma(k) = \text{Cov}(X_i, X_{i+k}) \).

Let
\[ \hat{c}_N(k) = \frac{1}{N-k} \sum_{j=1}^{N-k} (X_j - \bar{X}_N)(X_{j+k} - \bar{X}_N). \]

(This is a random variable.) Then for each \( k \),
\[ \lim_{N \to \infty} \hat{c}_N(k) = c(k) \text{ with probability 1}. \]

A natural estimator of \( \nu \) is then
\[ \widehat{\nu}_N = \hat{c}_N(0) + 2 \sum_{k=1}^{N-1} \hat{c}_N(k). \]

But, it turns out this is a bad estimator.
Here is an example. $N = 1,000$, $c(k) \approx e^{-1\cdot k/10}$.

Then $\hat{c}(k) \to 0$ for $k \geq 100$. Hence, $\sum_{k=1}^{1000} \hat{c}(k)$ is mostly "noise." Heuristically, $\text{Var}(\hat{c}(k)) \approx O(1/N)$. So $\text{Var}(\hat{c}^*) \approx N \cdot O(1) = O(1)$. So the variance of $\hat{c}^*$ does not converge to 0.

A better estimator of $V$ is

$$\hat{c}_{N,L} = \hat{c}(0) + 2 \sum_{k=1}^{L} \hat{c}(k),$$

where $L \gg 1$ is fixed! In this case, since $N \text{Var}(\hat{X}_N) \to V$ as $N \to \infty$, $\hat{c}_{N,L}/N$ is a good estimator of $\text{Var}(\hat{X}_N)$. Hence, since

$$\frac{\hat{X}_N - \mu}{\sqrt{\text{Var}(\hat{X}_N)}} \sim N(0,1) \quad \text{approximately for } N \gg 1,$$

the interval $[\hat{X}_N - r \sqrt{\text{Var}(\hat{X}_N)}/N, \hat{X}_N + r \sqrt{\text{Var}(\hat{X}_N)}}$, where $r = \Phi^{-1}(\alpha/2)$ with $\Phi(-\alpha/2)$, is a $\alpha$-confidence interval for $\mu$.

How to choose $L$?

Method 1. Choose $L$ such that $\hat{c}(k)$ is indistinguishable from noise for all $k > L$. 

\[ \hat{c}(k) \quad \downarrow \quad L \]
Method 2: Self-consistent weighting.

Define \( \tau = \frac{V}{\text{Var}(X)} = \frac{V}{\sigma^2} \), where \( V = \sum_{k=-\infty}^{\infty} C(k) = \sum_{k=-\infty}^{\infty} \text{Cov}(X_0, X_k) \). Then

\[
\tau = \sum_{k=1}^{\infty} \text{Cov}(X_0, X_k) = \sum_{k=1}^{\infty} \text{Corr}(X_0, X_k) \text{corr. coeff.}
\]

\[= 1 + 2 \sum_{k=1}^{\infty} \text{Corr}(X_0, X_k) \]

If \( \{X_i\} \) were i.i.d. then \( \tau = 1 \). In general,

\[\text{Var}(\bar{X}_{SN}) \approx \frac{V}{cN} \approx \frac{\sigma^2}{N} \]

Thus, \( TN \) observations from \( \{X_i\} \) gives approximately the same variance (hence the same size of confidence interval) that we would get if we had been able to generate \( N \) i.i.d. samples.

Now, choose some \( L_1 \) for \( L \) as our first guess. (e.g., using Method 1). Set \( T_N = \frac{V}{c_{10}} \).

Then, choose \( L \geq \sqrt{\frac{T_N}{c_{10}}} \) and \( L > L_1 \).

---

The Method of Batch Means for Determining Confidence Intervals

Again, \( X_1, X_2, \ldots \) is a stationary process.

Divide \( X_1, \ldots, X_N \) \((N >> 1)\) into non-overlapping subsections of equal length. Each of such subsections is called a batch. Let \( b \) be the common batch length. So, each batch has \( N/b \) consecutive \( X_i \)'s. (Assume \( N \) is an
integer multiple of $b$.) Let $L = N/b$. So, the $k$th batch is the subsequence of $X_1, \ldots, X_L$. For each $k \in \{1, \ldots, b\}$,

$$Y_k = \frac{1}{L} \sum_{i=(k-1)L+1}^{kL} X_i.$$

If $L$ is sufficiently large, then:

1. $Y_1, \ldots, Y_b \sim N(a, V/L)$ approximately.
2. $Y_1, \ldots, Y_b$ are approximately independent.

Now, using the classical statistics applied to $Y_1, \ldots, Y_b$, we obtain a $\chi^2(\nu, 1)$ confidence interval for $\nu$:

$$\left[ \bar{Y}_b - r_b \sqrt{\frac{S_b^2}{b}}, \bar{Y}_b + r_b \sqrt{\frac{S_b^2}{b}} \right],$$

where $\bar{Y}_b = \frac{1}{b} \sum_{i=1}^{b} Y_i$, $r = \Phi^{-1}(\frac{\nu+1}{2})$, $\Phi(x)$ is the CDF of $N(0, 1)$, i.e., $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx$, and

$$S_b^2 = \frac{1}{b-1} \sum_{i=1}^{b} (Y_i - \bar{Y}_b)^2.$$
$W_k = I$, e.g., this holds if $\exists$ no $J$ s.t. $h(I) = h(J)$. The idea of regeneration is to break $\{X_k\}$ into segments beginning and ending with consecutive visits to $I$. The Markov property implies that these segments are mutually independent, since the process starts fresh ("regenerated") with each visit to $I$.

Formally, assume $W_k = I$. (Otherwise, run the chain until the first visit to $I$, and discard everything before this time.) Let $\omega = 0$. For each $k \geq 1$, let

$$\tau_k = \min \{ t > \tau_{k-1} : W_t = I \}.$$ 

$\tau_k$ is the time of the $k$th visit to $I$.

The $k$th segment is the part of process from $\tau_{k-1} + 1$ to $\tau_k$ (inclusive).

We want to estimate $\mathbb{E}(X_k) = \mathbb{E}_\omega(h(W_k))$.

Set

$$D_k = \tau_k - \tau_{k-1} = \text{duration of } k\text{th segment}.$$ 

The "regenerative" property implies that $D_1, D_2, \ldots$ are i.i.d. Next, set

$$H_k = \sum_{i=\tau_{k-1}+1}^{\tau_k} h(W_i).$$

Then $H_1, H_2, \ldots$ are also i.i.d. For simplicity, assume $N = \sum \tau_m$ for some $m$. Then

$$N = D_1 + \cdots + D_m, \quad \sum_{i=1}^{\omega} X_i = H_1 + \cdots + H_m.$$
By the Strong Law of Large Numbers, we obtain
\[
E(X_1) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i
= \lim_{m \to \infty} \frac{1}{m} \left( \frac{E(H_1) + \cdots + E(H_m)}{E(D_1) + \cdots + E(D_m)} \right)
= \frac{E(H_i)}{E(D_i)} \quad \text{with probability 1.}
\]

Therefore, we can use the estimator
\[
\hat{H}_m = \frac{H_m}{D_m} = \frac{\sum_{i=1}^{m} H_i}{\sum_{i=1}^{m} D_i}
\]
for \( m = E(X_1) (\forall x \geq 0) \).

Note that this is in general biased as
\[
\mathbb{E}(\hat{H}_m) \neq H
\]

Error analysis. By the Central Limit Theorem,
\[
H_m := \frac{1}{m} \sum_{i=1}^{m} H_i = E(H_i) + \varepsilon_m,
\]
\[
D_m := \frac{1}{m} \sum_{i=1}^{m} D_i = E(D_i) + \delta_m,
\]
where \((\varepsilon_m, \delta_m)\) is approximately jointly normally distributed. Taylor’s expansion leads to
\[
\hat{H}_m = \frac{H_m}{D_m} \approx \frac{E(X_1) + \varepsilon_m}{E(D_1) + \delta_m} = \mathbb{E}(H_i) \frac{E(D_i)}{[E(D_i)]^2},
\]
leading to
\[
\text{Var}(\hat{H}_m) \leq \frac{1}{m-1} \sum_{i=1}^{m} \left( \frac{H_i}{D_m} - \frac{H_m}{D_m} \right)^2.
\]

Moreover,
\[
\hat{H}_m - H = O\left(\frac{1}{m}\right).
\]
1) Asymptotically Stationary Output

Let \( \{X_k\}_{k=0}^{\infty} \) be a Markov chain with invariant distribution \( \pi \). We wish to estimate

\[
\mu = \lim_{k \to \infty} E(X_k)
\]

from the simulated samples \( X_0, X_1, \ldots \).

The idea is to choose \( T \) large enough so that \( X_T \) is close to equilibrium, and discard \( X_0, \ldots, X_{T-1} \), and use the previous method for \( X_T, X_{T+1}, \ldots \). The period from \( t=0 \) to \( k=T \) is a period of equilibration, or the "burn-in" period. A short such period may result an "initialization bias".

There are no general rules for selecting \( T \). Also, methods of selecting \( T \) can be case dependent.

A simple and quick method is to divide the output into \( b \) batches (e.g., \( 20 \times 55 \)), and plot the batch means. If the mean of the first batch is significantly larger or smaller than all others, then discard the first batch. Otherwise stop. Repeat the next batch.
(E) Empirical CDF

Suppose Monte Carlo simulations produce random variables \( X_1, X_2, \ldots \) i.i.d. with the (exact) CDF \( F(x) \). (Here, \( X_j \in \mathbb{R} \) \( \forall j \geq 0 \).) \( F(x) \) is not known. But we wish to use \( \{X_k\}_{k=1}^{\infty} \) to find approximations of \( F(x) \).

Define \( \hat{F}_N(x) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{\{X_k \leq x\}} \)

\[ = \frac{1}{N} \left| \{k : X_k \leq x\} \right| \quad \forall x \in \mathbb{R}. \]

Call \( \hat{F}_N(x) \) a random empirical CDF. Let \( U_k = F(X_k) \) \((k = 1, 2, \ldots)\). Then \( U_k \sim U[0,1] \) and \( U_1, \ldots, U_N \) are i.i.d. Here, we assume \( F \) is continuous and strictly increasing. If we denote \( u = F(x) \) and \( x = F^{-1}(u) \), then

\[ \hat{F}_N(x) - F(x) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{\{X_k \leq x\}} - F(x) \]

\[ = \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{\{U_k \leq u\}} - u \]

\[ = \hat{G}_N(u) - u, \]

where

\[ \hat{G}_N(u) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{\{U_k \leq u\}}, \]

is called the reduced empirical CDF.
Define

$$D_n = \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F(x) \right|$$

$$= \sup_{0 \leq u \leq 1} \left| \hat{G}_n(u) - u \right|.$$ 

This is called a **Kolmogorov statistic** of the data.

Note that the distribution of $D_n$ does not depend on $F$ (by the second equality).

Here are some properties:

1. If $X_1 < \cdots < X_N$ then
   $$\hat{F}_n(X_k) = \frac{k}{N}, \quad k = 1, \ldots, N.$$

2. Binomial distribution:
   $$N \hat{F}_n(x) \sim \text{Binomial} \left( N, F(x) \right),$$
   $$N \hat{G}_n(u) \sim \text{Binomial} \left( N, u \right).$$

3. Glivenko-Cantelli:
   $$D_n \xrightarrow{a.s.} 0 \text{ and hence } \hat{F}_n(x) \xrightarrow{a.e.} F(x) \text{ uniformly in } x.$$

4. Central Limit Theorem:
   $$\sqrt{N} \left[ \hat{F}_n(x) - F(x) \right] \xrightarrow{dist.} Z \sim N(0, F(x)(1-F(x))).$$

5. **Conditional Process Poisson**: The probability distribution of the reduced empirical CDF $\{ \hat{G}_n(u) \}$, viewed as a stochastic process on $[0,1]$, is the same as the conditional distribution of a Poisson process $\{ \xi_u : 0 \leq u \leq 1 \}$ with rate $1/N$ given that $\xi_1 = N$. 
(i) Brownian bridge. The stochastic process \( \{ \sqrt{N} [ \hat{G}_n(u) - u], 0 \leq u \leq 1 \} \) converges in distribution to a Brownian bridge process on \([0, 1] \).

(ii) The Kolmogorov distribution.
\[
\lim_{N \to \infty} \mathbb{P}(\sqrt{N} D_N \leq x) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2}, \quad x > 0.
\]

(iii) The confidence interval. An approximate 1-\( \alpha \) confidence interval for \( F(x) \) is
\[
\left( \hat{F}_N(x) - z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{\hat{F}_N(x) (1 - \hat{F}_N(x))}}, \hat{F}_N(x) - z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{\hat{F}_N(x) (1 - \hat{F}_N(x))}} \right),
\]
where \( z_\alpha \) is such that \( 1 - \Phi(z_\alpha) = \gamma \) with \( \Phi(x) \) is CDF of a random variable \( \sim N(0, 1) \).

Equivalently, an approximate 1-\( \alpha \) confidence interval for \( F(X(\xi)) \) is
\[
\left( \frac{\xi}{N} - z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{\xi (1 - \xi N)}}, \frac{\xi}{N} - z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{\xi (1 - \xi N)}} \right).
\]
(F) **Kernel Density Estimation**

This method is for estimating a probability density from simulated data.

Let $X_1, \ldots, X_n$ be independent realizations from an unknown continuous PDF $f$ on some $S \subseteq \mathbb{R}$. Let $K = K(x)$ be a PDF of some random variable on $\mathbb{R}$ and assume it is symmetric, $K(-x) = K(x) \quad \forall x \in \mathbb{R}$ (i.e., $K = K(x)$ is an even function. Let $h > 0$. We define

$$
\hat{f}_{n,h}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K \left( \frac{x - X_i}{h} \right) \quad \forall x \in \mathbb{R}
$$

We call $\hat{f}_{n,h}$ a kernel density estimator of $f$, with the **bandwidth** $h > 0$. Here, $K(x)$ is called a kernel function.

**Example** The Gaussian kernel density estimator,

$$
\hat{f}_{n,h}(x) = \frac{1}{Nh} \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - X_i)^2}{2h^2}} \quad \forall x \in \mathbb{R},
$$

The function $K$ here is given by

$$
K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \forall x \in \mathbb{R}.
$$

This is the PDF of a standard normally distributed random variable.
We define the mean integrated squared error (MISE) of a kernel density estimator \( \hat{f}_{nh} \) by
\[
MISE_N(h) = \frac{1}{|f|} \int_{-\infty}^{\infty} \left| \hat{f}_{nh}(x) - f(x) \right|^2 dx.
\]

We have
\[
MISE_N(h) = \int_{-\infty}^{\infty} \left( \left| \int f(x) \hat{f}_{nh}(x) dx - f(x) \right|^2 + \int \text{Var}(\hat{f}_{nh}(x)) dx \right) dx = \text{pointwise bias of } \hat{f}_{nh} + \text{pointwise variance of } \hat{f}_{nh}.
\]

We can also use an alternative error criterion in the expected \( L^1 \) error:
\[
\int_{-\infty}^{\infty} |\hat{f}_{nh}(x) - f(x)| dx.
\]

How to choose a good bandwidth \( h \)?

**Method 1** The Gaussian rule of thumb.

A first-order asymptotic approximation of the MISE of the Gaussian kernel density estimator is
\[
\frac{1}{4} h^4 \| f'' \|_{L^2(\mathbb{R})}^2 + \frac{1}{2} \frac{h^2}{N^{1/2}}, \quad N \gg 1.
\]
(Here, we assume \( f'' \in L^2(\mathbb{R}) \).) An asymptotic optimal value of \( h \) is then
\[
h^* = \sqrt{\frac{2}{N \sqrt{2\pi}}} \| f'' \|_{L^2(\mathbb{R})}, \quad N \gg 1.
\]
The optimal asymptotic rate of decay of the MISE is
\[ \text{MISE}(h^*) = \frac{5 \, \| f'' \|_{L^2(\mathbb{R})}^2}{4 \sqrt{5 \pi} N^{-\frac{1}{2}}} N^{-\frac{1}{5}} = o(N^{-\frac{4}{5}}) \] as \( N \to \infty. \]

To compute \( h^* \), one needs to estimate \( \| f'' \|_{L^2(\mathbb{R})}^2 \). The Gaussian rule of thumb is to assume that \( f \) is the density of the \( N(\mu, \sigma^2) \) distribution where \( \mu \) and \( \sigma^2 \) are the sample mean and the sample variance of the data, respectively. In this case, \( \| f'' \|_{L^2(\mathbb{R})} = \frac{3}{4 \sqrt{\pi} \sigma} \), and the asymptotic optimal value is \( h^* = h_{\text{asy}} = \left( \frac{12 \sigma}{N} \right)^{1/5} \approx \sqrt{\ln 2 \sigma} N^{-1/5} \) for \( N \gg 1 \).

**Method 2** The Least-Squares Cross Validation

We again work on the selection of an optimal bandwidth for the Gaussian kernel density estimator

\[ \hat{f}_{N, h}(x) = \frac{1}{Nh} \sum_{i=1}^{N} \kappa \left( \frac{x - X_i}{h} \right), \quad \forall x \in \mathbb{R}, \]

where \( \kappa(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-x^2/2} \).

The Least-Squares Cross Validation (LSCV) method defines the optimal bandwidth as the unbiased estimator of the global minimizer of the integrated squared error (ISE), given by...
\[ \text{ISE}_N (\theta) = \int_0^\infty \left| f_{N, \theta} (x) - f(x) \right|^2 \, dx. \]

This is a random variable depending on the particular data. Minimizing \( \text{ISE}_N (\theta) \) is equivalent to minimizing
\[ \int_0^\infty \left| f_{N, \theta} (x) \right|^2 \, dx - 2 \text{IE}_f \left( \hat{f}_{N, \theta} (X) \right). \]

The term \( \text{IE}_f \left( \hat{f}_{N, \theta} (X) \right) \) can be estimated without bias via the cross-validation estimator
\[ \frac{1}{N} \sum_{i=1}^N \hat{f}_{N, \theta}^{(i)} (X_i), \]
where \( \hat{f}_{N, \theta}^{(i)} (x) \) is the Gaussian kernel density estimator based on the data points \( X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N. \) We have
\[ \frac{1}{N} \sum_{i=1}^N \hat{f}_{N, \theta}^{(i)} (X_i) = \frac{1}{hN(N-1)} \sum_{i=1}^N \sum_{j \neq i}^N k \left( \frac{X_i - X_j}{h} \right). \]

Let us introduce
\[ k_\theta (x) = \frac{1}{\sqrt{2\pi h^2}} e^{-\frac{x^2}{2h^2}} = \frac{1}{h} K \left( \frac{x}{h} \right), \quad \forall x \in \mathbb{R}. \]
(\( K \) is the PDF for a random variable \( Z \sim N(0, h^2) \)) The cross-validation estimator of \( \text{IE}_f (\hat{f}_{N, \theta} (X)) \) can be written more as
\[ \frac{1}{N} \sum_{i=1}^N \hat{f}_{N, \theta}^{(i)} (X_i) = \frac{1}{hN(N-1)} \sum_{i=1}^N \sum_{j \neq i}^N k_\theta (X_i - X_j). \]

Note that each \( X_i \) is a random variable. Thus
\[ \mathbb{E}_f \left( \frac{1}{N} \sum_{i=1}^{N} f_{M,h}^{(i)}(x_i) \right) \]
\[ = \frac{1}{N} \sum_{i,j=1}^{N} f(x_i) f_{M,h}(x_j) \] \[ = \frac{1}{N(N-1)} \sum_{i,j=1}^{N} f(x_i) K_h(x-x_j) \] \[ = \frac{1}{N} \sum_{j=1}^{N} f(x_j) K_h(x-x_j) . \]

But \[ \mathbb{E}_f \left( \hat{f}_{M,h}(x) \right) = \frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{N} f(x_i) K_h(x-x_j) \] Thus, the cross-validation estimator is unbiased.

Now, we calculate the integral of \( f_{M,h}^{(i)}(x) \):
\[ \int_{-\infty}^{\infty} |\hat{f}_{M,h}(x)|^2 \, dx = \frac{1}{N^2} \sum_{j=1}^{N} \int_{-\infty}^{\infty} \frac{(x-x_j)^2 + (x-x_j)^2}{2h^2} \, dx \]

For fixed \( x_i, x_j \):
\[ \int_{-\infty}^{\infty} e^{-\frac{1}{2h^2}[(x-x_i) + (x-x_j)]^2} \, dx \]
\[ = \int_{-\infty}^{\infty} e^{-\frac{1}{2h^2}(x-x_i)^2} e^{-\frac{1}{2h^2}((x-x_j)^2} \, dx \]
\[ = e^{-\frac{1}{4h^2}(x_i-x_j)^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2h^2} y^2} \, \frac{h}{\sqrt{2}} \, dy \]
\[ = \frac{\sqrt{2\pi} h}{\sqrt{2}} e^{-\frac{1}{4h^2}(x_i-x_j)^2} \]
(Hence, \[ \int_{-\infty}^{\infty} |\hat{f}_{M,h}(x)|^2 \, dx = \frac{1}{N^2} \sum_{j=1}^{N} K_{2h}(x_i-x_j) \].

Define now the LSCV bandwidth \( h_{LSCV} \) to be
\[ h_{LS} = \arg \min_{h > 0} g(h), \]

where
\[ g(h) = \frac{N}{2} \sum_{i,j=1}^{N} K_{2h}(X_i - X_j) - \frac{2}{N(N-1)} \sum_{i,j=1}^{N} u_i(X_i - X_j). \]

Note that the double-sum can be inefficient in practical implementation.

Back to
\[ f_{N,h}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left( \frac{x - X_i}{h} \right) = \frac{1}{N} \sum_{i=1}^{N} K(x - x_i) \]

with \( K(x) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{x^2}{4t}} \) and \( K_{2h}(x) = \frac{1}{\sqrt{4\pi (2h)^2}} e^{-\frac{x^2}{4(2h)^2}} \).

Let
\[ K(x,t) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{x^2}{4t}} \quad (t > 0, x \in \mathbb{R}). \]

This is the one-dimensional heat kernel. It satisfies
\[ \frac{\partial}{\partial t} K = \Delta K \quad (t > 0, x \in \mathbb{R}) \]

\[ \forall t \in \mathbb{R}, \lim_{t \to 0} \int_{-\infty}^{\infty} K(x,t) \phi(y) dy = \phi(x) \quad \forall x \in \mathbb{R}. \]

Let \( \lambda = h^2 \). Then,
\[ f_{N,h}(x) = f_{N}(x,h) = \frac{N}{h} \sum_{i=1}^{N} K(x - X_i, h) \]

Thus
\[ \frac{\partial}{\partial t} f_{N} = \frac{\partial^2}{\partial x^2} f_{N} \quad (t > 0, x \in \mathbb{R}). \]

\[ f_{N}(x,0) = \frac{N}{h} \sum_{i=1}^{N} \delta_{X_i}(x) \]

where \( \delta_{X_i}(x) \) is the Dirac measure at \( X_i \).

This provides a method of calculating \( f_{N,h} \).
Method 3  Plug-in Bandwidth Selection

This method provides an estimator for \( \| f'' \|_2^2 \) in determining the asymptotically optimal bandwidth \( h^* \) using the Gaussian kernel density estimation.

\[
h^* = \left( 2 \sqrt{\pi} N \| f'' \|_2^2 \right)^{-\frac{1}{5}}.
\]

The estimator of \( \| f'' \|_2^2 \) is \( \| \hat{f}_{N,h_2}'' \|_2^2 \) for some \( h_2 > 0 \), where

\[
\hat{f}_{N,h_2}(x) = \frac{1}{N} \sum_{k=1}^{N} K_{h_2}(x-X_k),
\]

is the Gaussian kernel density estimator.

In general, we have for \( h_2 > 0 \)

\[
\| \hat{f}_{N,h_2}'' \|_2^2 = \int_{-\infty}^{\infty} \left[ \frac{1}{N} \sum_{k=1}^{N} K_{h_2}^{(2)}(x-X_k) \right]^2 dx
\]

\[
= \frac{1}{N^2} \sum_{k,j=1}^{N} \int_{-\infty}^{\infty} K_{h_2}^{(2)}(x-X_k) K_{h_2}^{(2)}(x-X_j) dx
\]

\[
= \frac{(-1)^{a_2} N}{N^2} \sum_{k,j=1}^{N} g^{(2)}(X_k-x_j),
\]

where \( g^{(2)} \) is the second derivative of \( f \).

To use \( \| \hat{f}_{N,h_2}'' \|_2^2 \) as an estimator for \( \| f'' \|_2^2 \), we need to choose \( h_2 \). A good choice is for \( h_2 \) is

\[
h_2 = \left( \frac{1+2^{-3/2}}{3} \frac{(2k-1)!}{N^{1/2} \| f'' \|_2^2 \| \hat{f}_{N,h_2}'' \|_2^2} \right)^{2/3+2k},
\]
To compute $\| \hat{f}^{(k+1)} \|_2$, we need to know $h_{k+1}$, which in turn requires the estimate $h_{k+1}$, and so on.

The idea is now to fix $l$, say $l \geq 3$. First compute $\| \hat{f}^{(l+2)} \|_2$ by assuming $f$ is the normal PDF with the mean and variance estimated from the i.i.d. samples $X_1, \ldots, X_l$. Then compute $h_{l+1}$ using the above formula, which uses $\| \hat{f}^{(l+2)} \|_2$. Then compute $\| \hat{f}^{(l+3)} \|_2$, and then $h_{l+2}$, etc.