Convergence of Random Processes

DS GA 1002 Probability and Statistics for Data Science
http://www.cims.nyu.edu/~cfgranda/pages/DSGA1002_fall17

Carlos Fernandez-Granda
Define convergence for random processes

Describe two convergence phenomena: the law of large numbers and the central limit theorem
Types of convergence

Law of Large Numbers

Central Limit Theorem

Monte Carlo simulation
Convergence of deterministic sequences

A deterministic sequence of real numbers $x_1, x_2, \ldots$ converges to $x \in \mathbb{R}$,

$$\lim_{i \to \infty} x_i = x$$

if $x_i$ is arbitrarily close to $x$ as $i$ grows

For any $\epsilon > 0$ there is an $i_0$ such that for all $i > i_0$ $|x_i - x| < \epsilon$

Problem: Random sequences do not have fixed values
Convergence with probability one

Consider a discrete random process $\tilde{X}$ and a random variable $X$ defined on the same probability space.

If we fix the outcome $\omega$, $\tilde{X}(i, \omega)$ is a \textbf{deterministic} sequence and $X(\omega)$ is a \textbf{constant}.

We can determine whether

$$\lim_{i \to \infty} \tilde{X}(i, \omega) = X(\omega)$$

for \textit{that particular} $\omega$. 
Convergence with probability one

\( \tilde{X} \) converges with probability one to \( X \) if

\[
\mathbb{P} \left( \left\{ \omega \mid \omega \in \Omega, \lim_{i \to \infty} \tilde{X}(\omega, i) = X(\omega) \right\} \right) = 1
\]

Deterministic convergence occurs with probability one.
Puddle

Initial amount of water is uniform between 0 and 1 gallon

After a time interval $i$ there is $i$ times less water

$$\tilde{D}(\omega, i) := \frac{\omega}{i}, \quad i = 1, 2, \ldots$$
Puddle

\[
\tilde{D}(\omega, i) = \begin{cases} 
0.31 & \text{for } \omega = 0.31 \\
0.89 & \text{for } \omega = 0.89 \\
0.52 & \text{for } \omega = 0.52
\end{cases}
\]
If we fix $\omega \in (0, 1)$

$$\lim_{i \to \infty} \tilde{D}(\omega, i) = \lim_{i \to \infty} \frac{\omega}{i} = 0$$

$\tilde{D}$ converges to zero with probability one
Alternative idea

Idea: Instead of fixing $\omega$ and checking deterministic convergence:

1. Measure how close $\tilde{X}(i)$ and $X$ are for a fixed $i$ using a deterministic quantity

2. Check whether the quantity tends to zero
Convergence in mean square

The mean square of $Y - X$ measures how close $X$ and $Y$ are.

If $E\left((X - Y)^2\right) = 0$ then $X = Y$ with probability one.

Proof: By Markov’s inequality for any $\epsilon > 0$

$$P\left((Y - X)^2 > \epsilon\right) \leq \frac{E\left((X - Y)^2\right)}{\epsilon} = 0$$
Convergence in mean square

\( \tilde{X} \) converges to \( X \) in mean square if

\[
\lim_{i \to \infty} E \left( (X - \tilde{X}(i))^2 \right) = 0
\]
Convergence in probability

Alternative measure: Probability that \( |Y - X| > \epsilon \) for small \( \epsilon \)

\( \tilde{X} \) converges to \( X \) in probability if for any \( \epsilon > 0 \)

\[
\lim_{i \to \infty} P \left( \left| X - \tilde{X}(i) \right| > \epsilon \right) = 0
\]
Conv. in mean square implies conv. in probability

\[ \lim_{i \to \infty} P \left( \left| X - \tilde{X}(i) \right| > \epsilon \right) \leq \lim_{i \to \infty} E \left( \left( X - \tilde{X}(i) \right)^2 \right) \leq 0 \]

Convergence with probability one also implies convergence in probability
Conv. in mean square implies conv. in probability

$$\lim_{i \to \infty} P \left( \left| X - \tilde{X}(i) \right| > \epsilon \right) = \lim_{i \to \infty} P \left( \left( X - \tilde{X}(i) \right)^2 > \epsilon^2 \right)$$
Conv. in mean square implies conv. in probability

\[
\lim_{i \to \infty} P \left( \left| X - \tilde{X}(i) \right| > \epsilon \right) = \lim_{i \to \infty} P \left( \left( X - \tilde{X}(i) \right)^2 > \epsilon^2 \right) \\
\leq \lim_{i \to \infty} \frac{E \left( \left( X - \tilde{X}(i) \right)^2 \right)}{\epsilon^2}
\]
Conv. in mean square implies conv. in probability

\[
\lim_{i \to \infty} P\left(\left|X - \tilde{X}(i)\right| > \epsilon\right) = \lim_{i \to \infty} P\left(\left(X - \tilde{X}(i)\right)^2 > \epsilon^2\right)
\]

\[
\leq \lim_{i \to \infty} \frac{E\left(\left(X - \tilde{X}(i)\right)^2\right)}{\epsilon^2}
\]

\[
= 0
\]
Conv. in mean square implies conv. in probability

\[
\lim_{i \to \infty} P \left( \left| X - \tilde{X}(i) \right| > \epsilon \right) = \lim_{i \to \infty} P \left( \left( X - \tilde{X}(i) \right)^2 > \epsilon^2 \right)
\leq \lim_{i \to \infty} \frac{E \left( \left( X - \tilde{X}(i) \right)^2 \right)}{\epsilon^2}
= 0
\]

Convergence with probability one also implies convergence in probability
Convergence in distribution

The distribution of $\tilde{X}(i)$ converges to the distribution of $X$ if

$$\lim_{i \to \infty} F_{\tilde{X}(i)}(x) = F_X(x)$$

for all $x$ at which $F_X$ is continuous.
Convergence in distribution

Convergence in distribution does not imply that $\tilde{X}(i)$ and $X$ are close as $i \to \infty$!

Convergence in probability does imply convergence in distribution
Binomial tends to Poisson

- \( \tilde{X}(i) \) is binomial with parameters \( i \) and \( p := \lambda/i \)

- \( X \) is a Poisson random variable with parameter \( \lambda \)

- \( \tilde{X}(i) \) converges to \( X \) in distribution

\[
\lim_{i \to \infty} p_{\tilde{X}(i)}(x) = \lim_{i \to \infty} \binom{i}{x} p^x (1 - p)^{(i-x)}
\]

\[
= \frac{\lambda^x e^{-\lambda}}{x!}
\]

\[
= p_X(x)
\]
Probability mass function of $\tilde{X}(40)$
Probability mass function of $\tilde{X}(80)$
Probability mass function of $\tilde{X} (400)$
Probability mass function of $X$

![Bar chart showing the probability mass function of $X$. The x-axis represents $k$ ranging from 0 to 40, and the y-axis represents probabilities ranging from $5 \times 10^{-2}$ to 0.15. The chart peaks around $k = 20$.](image-url)
Types of convergence

Law of Large Numbers

Central Limit Theorem

Monte Carlo simulation
The moving average $\tilde{A}$ of a discrete random process $\tilde{X}$ is

$$\tilde{A}(i) := \frac{1}{i} \sum_{j=1}^{i} \tilde{X}(j)$$
Weak law of large numbers

Let $\tilde{X}$ be an iid discrete random process with mean $\mu_{\tilde{X}} := \mu$ and bounded variance $\sigma^2$

The average $\tilde{A}$ of $\tilde{X}$ converges in mean square to $\mu$
Proof

\[ \mathbb{E}\left( \tilde{A}(i) \right) \]
Proof

\[ E\left( \tilde{A}(i) \right) = E\left( \frac{1}{i} \sum_{j=1}^{i} \tilde{x}(j) \right) \]
Proof

\[ E\left(\tilde{A}(i)\right) = E\left(\frac{1}{i} \sum_{j=1}^{i} \tilde{X}(j)\right) = \frac{1}{i} \sum_{j=1}^{i} E\left(\tilde{X}(j)\right) \]
Proof

\[
E(\tilde{A}(i)) = E \left( \frac{1}{i} \sum_{j=1}^{i} \tilde{X}(j) \right) \\
= \frac{1}{i} \sum_{j=1}^{i} E(\tilde{X}(j)) \\
= \mu
\]
Proof

\[ \text{Var} \left( \tilde{A}(i) \right) \]
Proof

\[ \text{Var} \left( \tilde{A}(i) \right) = \text{Var} \left( \frac{1}{i} \sum_{j=1}^{i} \tilde{X}(j) \right) \]
Proof

$$\text{Var}\left(\tilde{A}(i)\right) = \text{Var}\left(\frac{1}{i} \sum_{j=1}^{i} \tilde{X}(j)\right)$$

$$= \frac{1}{i^2} \sum_{j=1}^{i} \text{Var}\left(\tilde{X}(j)\right)$$
Proof

\[ \text{Var} \left( \tilde{A}(i) \right) = \text{Var} \left( \frac{1}{i} \sum_{j=1}^{i} \tilde{X}(j) \right) \]

\[ = \frac{1}{i^2} \sum_{j=1}^{i} \text{Var} \left( \tilde{X}(j) \right) \]

\[ = \frac{\sigma^2}{i} \]
Proof

\[
\lim_{i \to \infty} \mathbb{E} \left( \left( \tilde{A}(i) - \mu \right)^2 \right) = \lim_{i \to \infty} \text{Var} \left( \tilde{A}(i) \right) = \lim_{i \to \infty} \sigma^2_i = 0
\]
Proof

\[ \lim_{i \to \infty} \mathbb{E} \left( \left( \tilde{A}(i) - \mu \right)^2 \right) = \lim_{i \to \infty} \mathbb{E} \left( \left( \tilde{A}(i) - \mathbb{E} \left( \tilde{A}(i) \right) \right)^2 \right) \]
Proof

\[
\lim_{i \to \infty} E \left( \left( \tilde{A}(i) - \mu \right)^2 \right) = \lim_{i \to \infty} E \left( \left( \tilde{A}(i) - E(\tilde{A}(i)) \right)^2 \right) = \lim_{i \to \infty} \text{Var}(\tilde{A}(i))
\]
Proof

\[
\lim_{i \to \infty} \mathbb{E} \left( (\tilde{A}(i) - \mu)^2 \right) = \lim_{i \to \infty} \mathbb{E} \left( (\tilde{A}(i) - \mathbb{E}(\tilde{A}(i)))^2 \right) \\
= \lim_{i \to \infty} \text{Var}(\tilde{A}(i)) \\
= \lim_{i \to \infty} \frac{\sigma^2}{i}
\]
Proof

\[
\lim_{i \to \infty} \mathbb{E} \left( \left( \tilde{A}(i) - \mu \right)^2 \right) = \lim_{i \to \infty} \mathbb{E} \left( \left( \tilde{A}(i) - \mathbb{E} \left( \tilde{A}(i) \right) \right)^2 \right) \\
= \lim_{i \to \infty} \text{Var} \left( \tilde{A}(i) \right) \\
= \lim_{i \to \infty} \frac{\sigma^2}{i} \\
= 0
\]
Let $\tilde{X}$ be an iid discrete random process with mean $\mu_{\tilde{X}} := \mu$ and bounded variance $\sigma^2$

The average $\tilde{A}$ of $\tilde{X}$ converges with probability one to $\mu$
iid standard Gaussian

Moving average
Mean of iid seq.
iid standard Gaussian

Moving average
Mean of iid seq.
Moving average
Mean of iid seq.
iid geometric with $p = 0.4$
iid geometric with $p = 0.4$
iid geometric with $p = 0.4$
iid Cauchy

Moving average
Median of iid seq.

0 10 20 30 40 50
i

Moving average
Median of iid seq.
iid Cauchy

Moving average
Median of iid seq.
 iid Cauchy

Moving average
Median of iid seq.
Types of convergence

Law of Large Numbers

Central Limit Theorem

Monte Carlo simulation
Central Limit Theorem

Let $\tilde{X}$ be an iid discrete random process with mean $\mu_{\tilde{X}} := \mu$ and bounded variance $\sigma^2$

$$\sqrt{n} \left( \tilde{A} - \mu \right)$$ converges in distribution to a Gaussian random variable with mean 0 and variance $\sigma^2$

The average $\tilde{A}$ is approximately Gaussian with mean $\mu$ and variance $\sigma^2/i$
Height data

- Example: Data from a population of 25,000 people
- We compare the histogram of the heights and the pdf of a Gaussian random variable fitted to the data
Height data
Sketch of proof

Pdf of sum of two independent random variables is the convolution of their pdfs

\[ f_{X+Y}(z) = \int_{y=-\infty}^{\infty} f_X(z-y) f_Y(y) \, dy \]

Repeated convolutions of any pdf with bounded variance result in a Gaussian!
Repeated convolutions
Repeated convolutions
iid exponential $\lambda = 2, \ i = 10^2$
iid exponential $\lambda = 2$, $i = 10^3$
iid exponential $\lambda = 2$, $i = 10^4$
iid geometric $p = 0.4$, $i = 10^2$
iid geometric $p = 0.4, \ i = 10^3$
iid geometric $p = 0.4$, $i = 10^4$
iid Cauchy, $i = 10^2$
iid Cauchy, $i = 10^3$
iid Cauchy, \( i = 10^4 \)
Gaussian approximation to the binomial

$X$ is binomial with parameters $n$ and $p$

Computing the probability that $X$ is in a certain interval requires summing its pmf over the interval

Central limit theorem provides a quick approximation

$$X = \sum_{i=1}^{n} B_i, \quad \mathbb{E}(B_i) = p, \quad \text{Var}(B_i) = p(1-p)$$

$\frac{1}{n}X$ is approximately Gaussian with mean $p$ and variance $p(1-p)/n$

$X$ is approximately Gaussian with mean $np$ and variance $np(1-p)$
Gaussian approximation to the binomial

Basketball player makes shot with probability $p = 0.4$ (shots are iid)

Probability that she makes more than 420 shots out of 1000?

Exact answer:

$$P(X \geq 420) = \sum_{x=420}^{1000} p_X(x)$$

$$= \sum_{x=420}^{1000} \binom{1000}{x} 0.4^x 0.6^{(n-x)} = 10.4 \times 10^{-2}$$

Approximation:

$$P(X \geq 420)$$
Gaussian approximation to the binominal

Basketball player makes shot with probability $p = 0.4$ (shots are iid)

Probability that she makes more than 420 shots out of 1000?

Exact answer:

$$P(X \geq 420) = \sum_{x=420}^{1000} p_X(x)$$

$$= \sum_{x=420}^{1000} \binom{1000}{x} 0.4^x 0.6^{(n-x)} = 10.4 \times 10^{-2}$$

Approximation:

$$P(X \geq 420) \approx P\left(\sqrt{np(1-p)}U + np \geq 420\right)$$
Gaussian approximation to the binomial

Basketball player makes shot with probability $p = 0.4$ (shots are iid)

Probability that she makes more than 420 shots out of 1000?

Exact answer:

$$P(X \geq 420) = \sum_{x=420}^{1000} p_X(x)$$

$$= \sum_{x=420}^{1000} \binom{1000}{x} 0.4^x 0.6^{(1000-x)} = 10.4 \times 10^{-2}$$

Approximation:

$$P(X \geq 420) \approx P \left( \sqrt{np(1-p)}U + np \geq 420 \right)$$

$$= P(U \geq 1.29)$$
Gaussian approximation to the binomial

Basketball player makes shot with probability $p = 0.4$ (shots are iid)

Probability that she makes more than 420 shots out of 1000?

Exact answer:

$$P(X \geq 420) = \sum_{x=420}^{1000} p_X(x)$$

$$= \sum_{x=420}^{1000} \binom{1000}{x} 0.4^x 0.6^{(n-x)} = 10.4 \times 10^{-2}$$

Approximation:

$$P(X \geq 420) \approx P\left(\sqrt{np(1-p)}U + np \geq 420\right)$$

$$= P(U \geq 1.29)$$

$$= 1 - \Phi(1.29) = 9.85 \times 10^{-2}$$
Types of convergence

Law of Large Numbers

Central Limit Theorem

Monte Carlo simulation
Monte Carlo simulation

Simulation is a powerful tool in probability and statistics.

Models are too complex to derive closed-form solutions (life is not a homework problem!)

**Example:** Game of solitaire
Game of solitaire

Aim: Compute the probability that you win at solitaire

If every permutation of the cards has the same probability

\[ P(\text{Win}) = \frac{\text{Number of permutations that lead to a win}}{\text{Total number}} \]

Problem: Characterizing permutations that lead to a win is very difficult without playing out the game

We can’t just check because there are \(52! \approx 8 \times 10^{67}\) permutations!

Solution: Sample many permutations and compute the fraction of wins
The first thoughts and attempts I made to practice (the Monte Carlo Method) were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaires. The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers.
Monte Carlo approximation

**Main principle:** Use simulation to approximate quantities that are challenging to compute exactly

To approximate the probability of an event $\mathcal{E}$

1. Generate $n$ independent samples from $1_\mathcal{E}$: $l_1, l_2, \ldots, l_n$
2. Compute the average of the $n$ samples

$$\tilde{A}(n) := \frac{1}{n} \sum_{i=1}^{n} l_i$$

By the law of large numbers $\tilde{A}$ converges to $P(\mathcal{E})$ as $n \to \infty$ since

$$E(1_\mathcal{E}) = P(\mathcal{E})$$
Basketball league

Basketball league with \( m \) teams

In a season every pair of teams plays once

Teams are ordered: team 1 is best, team \( m \) is worst

Model: For \( 1 \leq i < j \leq m \)

\[
P(\text{team } j \text{ beats team } i) := \frac{1}{j - i + 1}
\]

Games are independent
Basketball league

**Aim:** Compute probability of team ranks at the end of the season

The rank of team $i$ is modeled as a random variable $R_i$.

Pmf of $R_1$, $R_2$, $\ldots$, $R_m$?
\[ m = 3 \]

<table>
<thead>
<tr>
<th>Game outcomes</th>
<th>Rank</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2 1-3 2-3</td>
<td>( R_1 ) 1 ( R_2 ) 2 ( R_3 ) 3</td>
<td>1/6</td>
</tr>
<tr>
<td>1 1 2</td>
<td>1 2 3</td>
<td>1/6</td>
</tr>
<tr>
<td>1 1 3</td>
<td>1 3 2</td>
<td>1/6</td>
</tr>
<tr>
<td>1 3 2</td>
<td>1 1 1</td>
<td>1/12</td>
</tr>
<tr>
<td>1 3 3</td>
<td>2 3 1</td>
<td>1/12</td>
</tr>
<tr>
<td>2 1 2</td>
<td>2 1 3</td>
<td>1/6</td>
</tr>
<tr>
<td>2 1 3</td>
<td>1 1 1</td>
<td>1/6</td>
</tr>
<tr>
<td>2 3 2</td>
<td>3 1 2</td>
<td>1/12</td>
</tr>
<tr>
<td>2 3 3</td>
<td>3 2 1</td>
<td>1/12</td>
</tr>
</tbody>
</table>
$m = 3$

**Probability mass function**

<table>
<thead>
<tr>
<th></th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7/12</td>
<td>1/2</td>
<td>5/12</td>
</tr>
<tr>
<td>2</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>3</td>
<td>1/6</td>
<td>1/4</td>
<td>1/3</td>
</tr>
</tbody>
</table>
Problem: Number of possible outcomes is $2^{m(m-1)/2!}$

For $m = 10$ this is larger than $10^{13}$

Solution: Apply Monte Carlo approximation
\( m = 3 \)

<table>
<thead>
<tr>
<th>Game outcomes</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>1-3</td>
</tr>
<tr>
<td>\hline 1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>\hline 1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
$m = 3$

Estimated pmf ($n = 10$)

<table>
<thead>
<tr>
<th></th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6 (0.583)</td>
<td>0.7 (0.5)</td>
<td>0.3 (0.417)</td>
</tr>
<tr>
<td>2</td>
<td>0.1 (0.25)</td>
<td>0.2 (0.25)</td>
<td>0.4 (0.25)</td>
</tr>
<tr>
<td>3</td>
<td>0.3 (0.167)</td>
<td>0.1 (0.25)</td>
<td>0.3 (0.333)</td>
</tr>
</tbody>
</table>
$m = 3$

Estimated pmf ($n = 2,000$)

<table>
<thead>
<tr>
<th></th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.582 (0.583)</td>
<td>0.496 (0.5)</td>
<td>0.417 (0.417)</td>
</tr>
<tr>
<td>2</td>
<td>0.248 (0.25)</td>
<td>0.261 (0.25)</td>
<td>0.244 (0.25)</td>
</tr>
<tr>
<td>3</td>
<td>0.171 (0.167)</td>
<td>0.245 (0.25)</td>
<td>0.339 (0.333)</td>
</tr>
</tbody>
</table>
Running times

![Graph showing running times for different numbers of teams. The x-axis represents the number of teams, and the y-axis represents running time in seconds. Two lines are plotted: one for exact computation and another for Monte Carlo approximation.]

- Exact computation
- Monte Carlo approximation
<table>
<thead>
<tr>
<th>$m$</th>
<th>Average error</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>$9.28 \times 10^{-3}$</td>
</tr>
<tr>
<td>4</td>
<td>$12.7 \times 10^{-3}$</td>
</tr>
<tr>
<td>5</td>
<td>$7.95 \times 10^{-3}$</td>
</tr>
<tr>
<td>6</td>
<td>$7.12 \times 10^{-3}$</td>
</tr>
<tr>
<td>7</td>
<td>$7.12 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
$m = 5$
$m = 20$
$m = 100$