On the Spectrum of Beta and Dirichlet Random Matrices

With Applications to Compressed Sensing

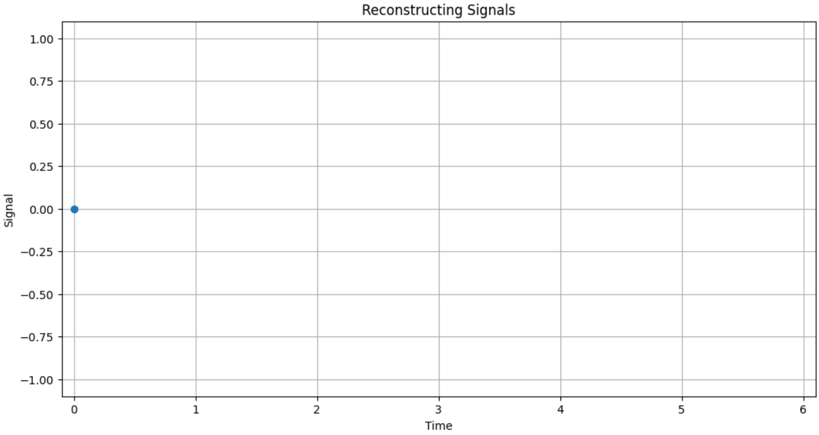
Mitchell Scott

Department of Mathematics, Emory University

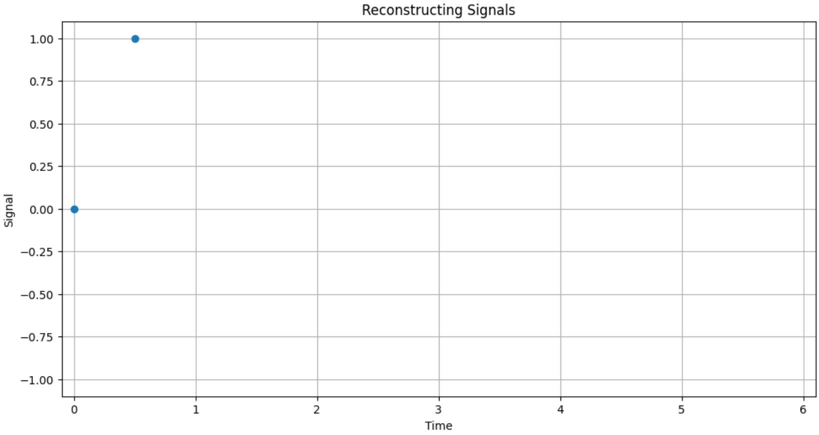
December 3, 2024

Table of Contents

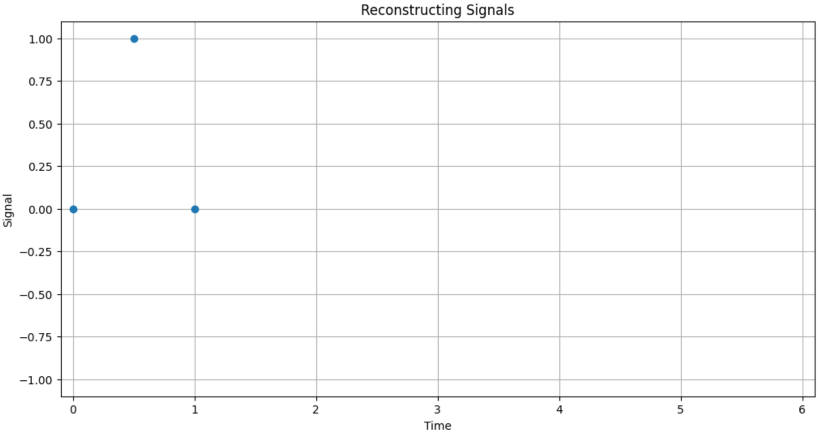
Sampling Problem



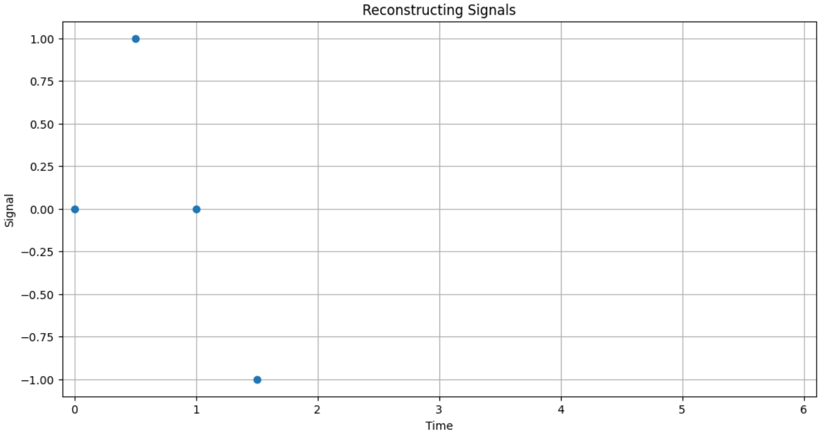
Sampling Problem



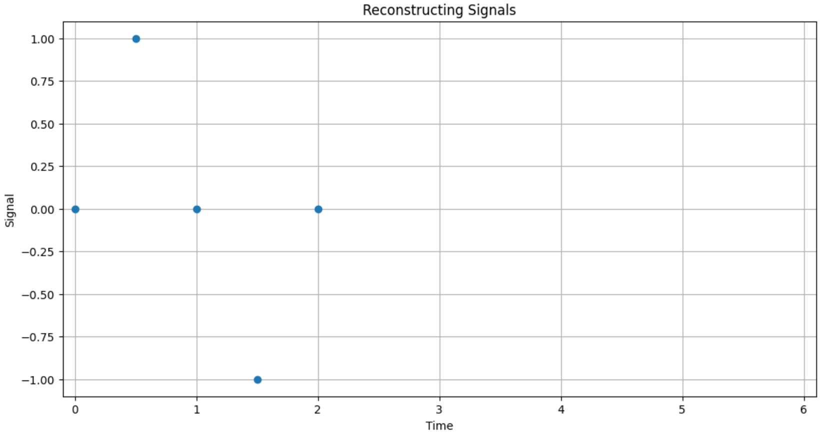
Sampling Problem



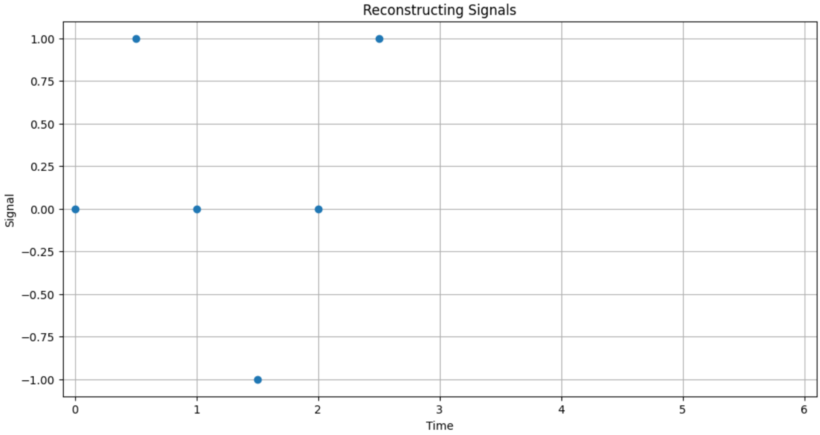
Sampling Problem



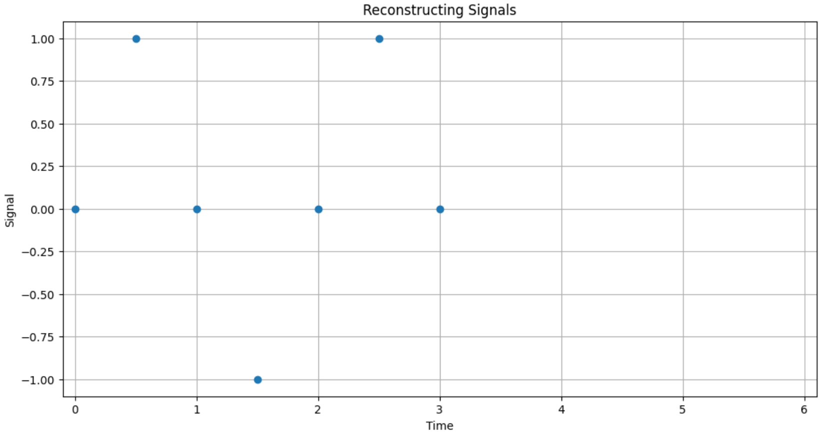
Sampling Problem



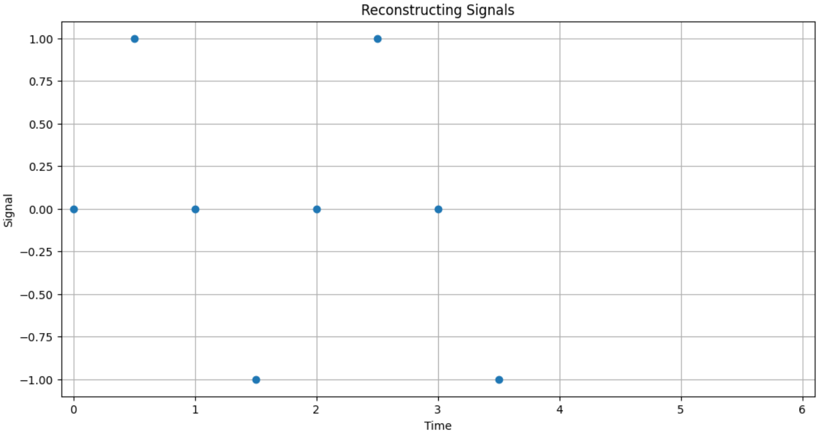
Sampling Problem



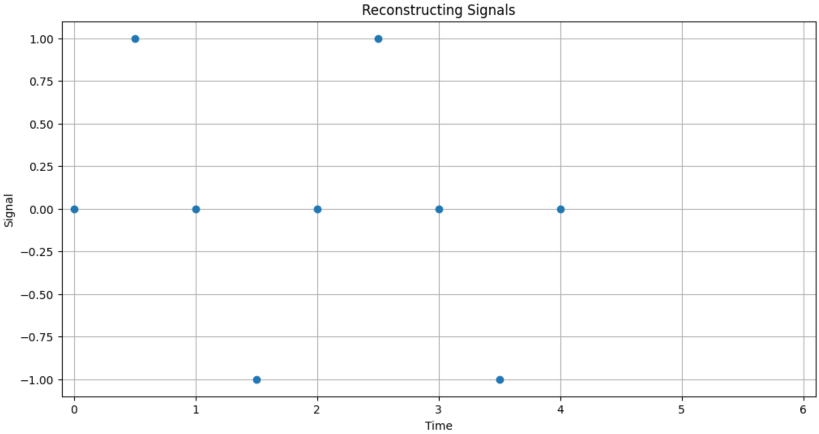
Sampling Problem



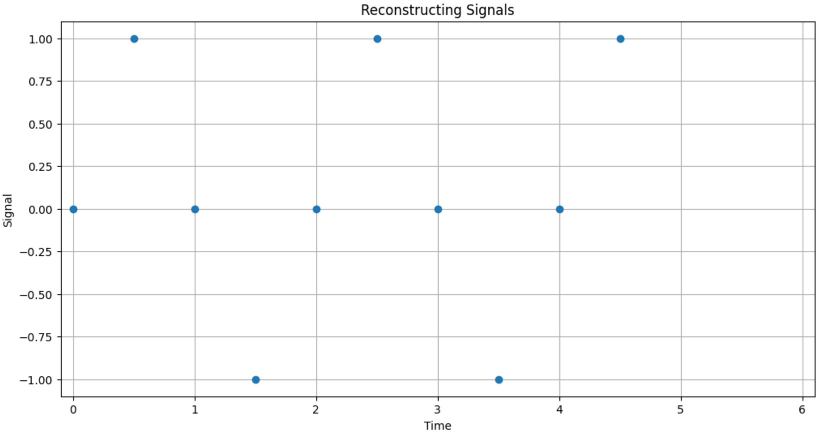
Sampling Problem



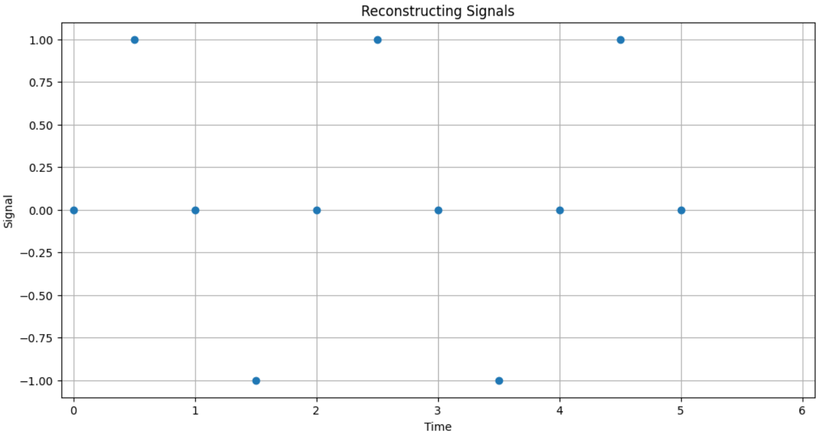
Sampling Problem



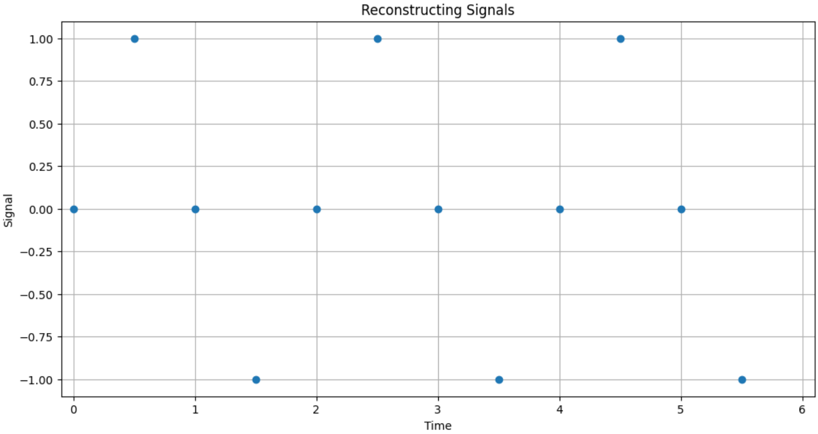
Sampling Problem



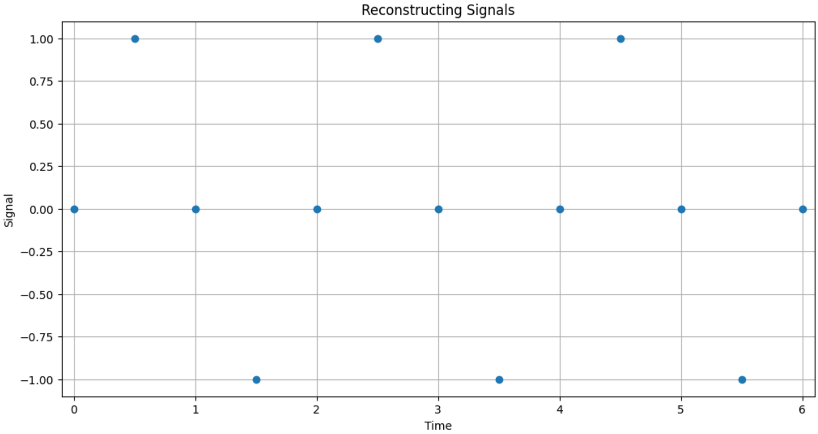
Sampling Problem



Sampling Problem

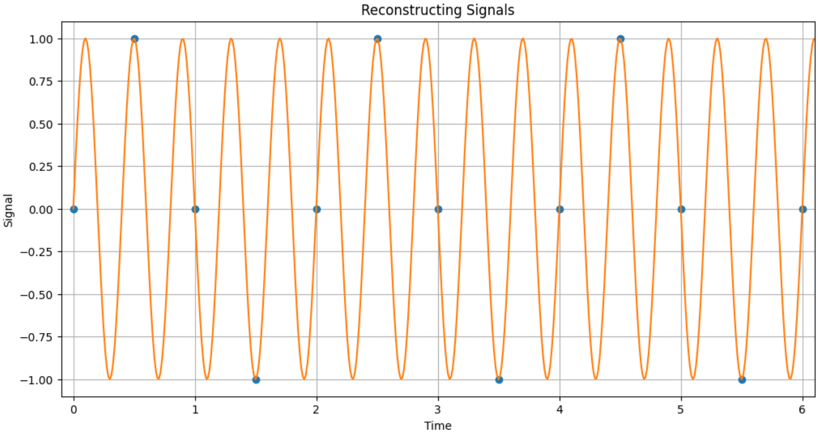


Sampling Problem

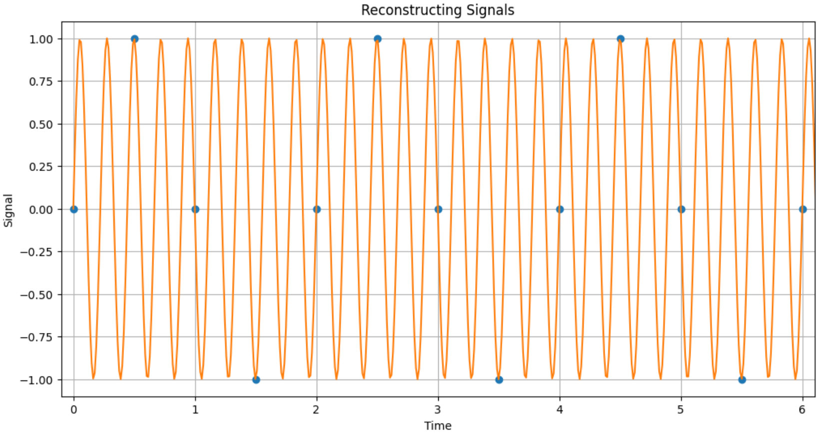


What is the signal?

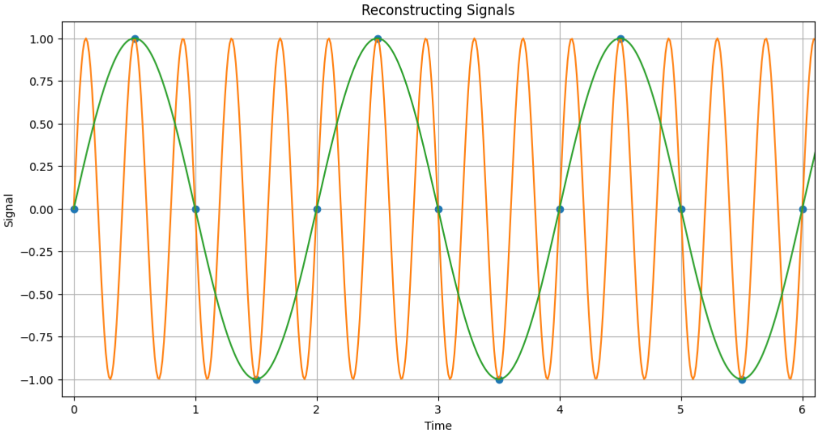
What is the signal?



What is the signal?



What is the signal?



Solution

Theorem (Nyquist - Shannon Sampling Theorem, 1915, [1])

If a function $x(t)$ contains no frequences higher than B hertz $[s^{-1}]$, then it can be completely determined from sampling a sequence of points spaced less than $1/(2B)$ seconds apart.

Solution

Theorem (Nyquist - Shannon Sampling Theorem, 1915, [1])

If a function $x(t)$ contains no frequences higher than B hertz $[s^{-1}]$, then it can be completely determined from sampling a sequence of points spaced less than $1/(2B)$ seconds apart.

Key Take-away

In systems where you want to generate accurate signals from sampling data, you must set the sampling rate high enough to prevent aliasing.

EMORY UNIVERSITY

Solution

Theorem (Nyquist - Shannon Sampling Theorem, 1915, [1])

If a function $x(t)$ contains no frequences higher than B hertz $[s^{-1}]$, then it can be completely determined from sampling a sequence of points spaced less than $1/(2B)$ seconds apart.

Key Take-away

In systems where you want to generate accurate signals from sampling data, you must set the sampling rate high enough to prevent aliasing.

Remark

However, the number of samples needed for high frequency data or long range signals, might be memory intesive to store explicitly. Additionally, how can you set up the sampling rate a priori?

Compressed Sensing

Remark

Perfect reconstruction of a signal can happen even if the N-S criterion isn't satisfied!

Compressed Sensing

Remark

Perfect reconstruction of a signal can happen even if the N-S criterion isn't satisfied!

Definition (Compressed Sensing)

Compressed sensing is a signal processing technique for efficiently acquiring and reconstructing a signal, by finding solutions to underdetermined linear systems.

Compressed Sensing

Remark

Perfect reconstruction of a signal can happen even if the N-S criterion isn't satisfied!

Definition (Compressed Sensing)

Compressed sensing is a signal processing technique for efficiently acquiring and reconstructing a signal, by finding solutions to underdetermined linear systems.

Theorem (Candes, Romberg, Tao (2005)[2])

Given some knowledge about a signal's sparsity, the signal may be reconstructed with even fewer samples than the sampling theorem requires, the basis of compressed sensing.

Table of Contents

Wishart Matrices

Definition (Wishart Ensemble)

The data for a *Wishart Ensemble* is a matrix of $N \times T$ data $\{x_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}$, where we have T observations and each observation contains N variables.

Wishart Matrices

Definition (Wishart Ensemble)

The data for a *Wishart Ensemble* is a matrix of $N \times T$ data $\{x_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}$, where we have T observations and each observation contains N variables.

Example

Wishart matrices can arise in many examples, such as:

- daily returns of N stocks over a certain time period,

Wishart Matrices

Definition (Wishart Ensemble)

The data for a *Wishart Ensemble* is a matrix of $N \times T$ data $\{x_i^t\}_{1 \leq i \leq N, 1 \leq t \leq T}$, where we have T observations and each observation contains N variables.

Example

Wishart matrices can arise in many examples, such as:

- daily returns of N stocks over a certain time period,
- number of spikes fired by N neurons during T consecutive intervals of Δt,

Wishart Matrices

Definition (Wishart Ensemble)

The data for a *Wishart Ensemble* is a matrix of $N \times T$ data $\{x_i^t\}_{1 \le i \le N, 1 \le t \le T}$, where we have T observations and each observation contains N variables.

Example

Wishart matrices can arise in many examples, such as:

- daily returns of N stocks over a certain time period,
- number of spikes fired by N neurons during T consecutive intervals of Δt,
- and so many more.

Sample Covariance Matrices

Definition (Sample Covariance Matrix)

The *sample covariances* of the data are given by

$$\mathbf{E}_{ij} = \frac{1}{T}\sum_{t=1}^{T} x_i^t x_j^t. \tag{1}$$

This results in an $N \times N$ matrix $\mathbf{E}$, called the *sample covariance matrix*, which can be written as

$$\mathbf{E} = \frac{1}{T}\mathbf{H}\mathbf{H}^\top, \tag{2}$$

where $\mathbf{H}$ is the $N \times T$ matrix with $\mathbf{H}_{it} = x_i^t$.

EMORY UNIVERSITY

Convergence of Wishart Matrices' Spectrum

Theorem (Marčenko-Pastur [3])

The full Marčenko-Pastur distribibution can be written as such, let $\mathbf{M} \in \mathbb{R}^{N \times T}$, *where* $N, T \to \infty, N/T \to q \in (0, \infty)$. *Now let* $\lambda_{\pm} = \sigma^2 \left(1 \pm \sqrt{q}\right)^2$. *Then the density of the eigenvalues of* $\mathbf{M}$ *converges weakly to*

$$\rho_{MP}(x) = \frac{1}{2\pi\sigma^2} \frac{\sqrt{\left[(\lambda_+ - x)(x - \lambda_-)\right]_+}}{2\pi q x} \tag{3}$$

where $[a]_+ := \max\{a, 0\}$.

Convergence of Wishart Matrices' Spectrum

Theorem (Marčenko-Pastur [3])

The full Marčenko-Pastur distribribution can be written as such, let $\mathbf{M} \in \mathbb{R}^{N \times T}$, *where* $N, T \to \infty, N/T \to q \in (0, \infty)$. *Now let* $\lambda_{\pm} = \sigma^2 \left(1 \pm \sqrt{q}\right)^2$. *Then the density of the eigenvalues of* $\mathbf{M}$ *converges weakly to*

$$\rho_{MP}(x) = \frac{1}{2\pi\sigma^2} \frac{\sqrt{\left[(\lambda_+ - x)(x - \lambda_-)\right]_+}}{2\pi q x} + \left[\frac{q-1}{q}\right]_+ \delta(x) \tag{3}$$

where $[a]_+ := \max\{a, 0\}$.

Visualizing the Marčenko-Pastur Distribution

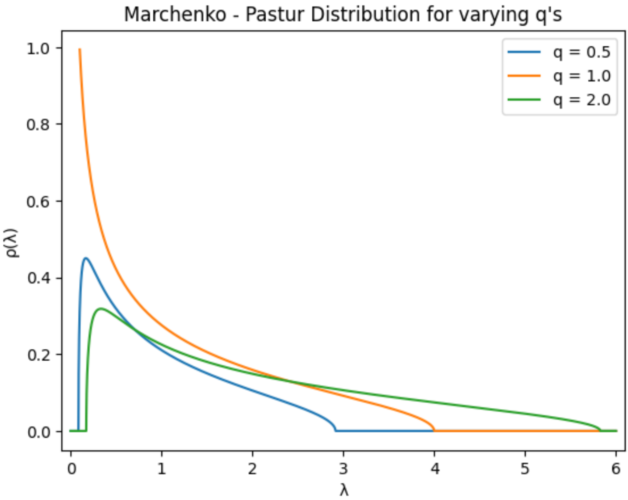


Table of Contents

Beta Distribution

Definition (Beta Distribution PDF)

Let $x \in [0, 1]$. The Beta distribution has two shape parameters $\boldsymbol{\beta} = [\beta_1, \beta_2]^\top$, which control the growth of small and large values of x, respectively. Then the *Beta Distrbution PDF* is given by

$$f(x; \boldsymbol{\beta}) = \frac{1}{B(\boldsymbol{\beta})} x^{\beta_1 - 1} (1 - x)^{\beta_2 - 1}, \quad \text{where} \tag{4}$$

$$B(\boldsymbol{\beta}) = \frac{\Gamma(\beta_1)\Gamma(\beta_2)}{\Gamma(\beta_1 + \beta_2)} \tag{5}$$

Beta Distribution

Definition (Beta Distribution PDF)

Let $x \in [0,1]$. The Beta distribution has two shape parameters $\boldsymbol{\beta} = [\beta_1, \beta_2]^\top$, which control the growth of small and large values of x, respectively. Then the *Beta Distrbution PDF* is given by

$$f(x; \boldsymbol{\beta}) = \frac{1}{B(\boldsymbol{\beta})} x^{\beta_1 - 1} (1-x)^{\beta_2 - 1}, \quad \text{where} \tag{4}$$

$$B(\boldsymbol{\beta}) = \frac{\Gamma(\beta_1)\Gamma(\beta_2)}{\Gamma(\beta_1 + \beta_2)} \tag{5}$$

Example

To sample a vector $\mathbf{v} \in \mathbb{R}^t$ from this distribution, we draw samples from the Beta distribution with parameters $\boldsymbol{\beta}$, and then rescale all samples by $\sum_{i=1}^t \mathbf{v}_i$. This ensures that $\|\mathbf{v}\|_1 = 1$.

Dirichlet Distribution

Definition (Dirichlet Distribution PDF)

Also known as the multivariate Beta distribution, the *Dirichlet Distribution PDF* of order $K \geq 2$ with parameters $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_K]^\top$ is given by

$$f(\mathbf{x}; \boldsymbol{\beta}) = \frac{1}{B(\boldsymbol{\beta})} \prod_{i=1}^{K} x_i^{\beta_i - 1}, \quad \text{where} \tag{6}$$

$$B(\boldsymbol{\beta}) = \frac{\prod_{i=1}^{K} \Gamma(\beta_i)}{\Gamma\left(\sum_{i=1}^{K} \beta_i\right)}, \tag{7}$$

and $x_i \in [0, 1], \forall i \in \{1, 2, \cdots, K\}$, subject to $\sum_{i=1}^{K} x_i = 1$.

Sampling from Dirichlet Distribution

Example

While sampling from the Beta distribution is available in `numpy`, Dirichlet distributions are not. However, we do have access to the Gamma distribution. This means to sample a random vector $\mathbf{x} = [x_1, x_2, \ldots, x_K]^\top$ from the K-dimensional Dirichlet distribution with parameters $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_K]^\top$, we draw K independent samples from Gamma distribution and normalize to sum to 1.

$$\text{Gamma}(\beta_i, 1) = \frac{y_i^{\beta_i - 1} e^{-y_i}}{\Gamma(\beta_i)} \tag{8}$$

$$x_i = \frac{y_i}{\sum_{j=1}^{K} y_j} \tag{9}$$

Visualizing these Vectors

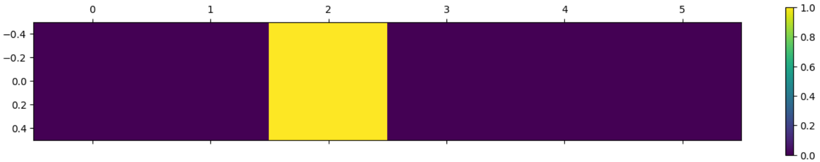


Figure 2: Small values of $\boldsymbol{\beta} = [0.0001, 0.0001]^\top$ promote sparcity.

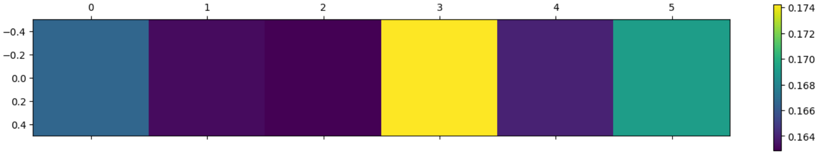


Figure 3: Large values of $\boldsymbol{\beta} = [1000, 1000]^\top$ promote uniformity $(1/N)$.

Making Matrices from these Random Vectors

Definition (Beta/ Dirichlet Random Covariance Matrices[4])

Now that we are able to sample vectors in $\mathbb{R}^t$ from the Beta and Dirichlet distribution, we can stack these vectors on top of each other N times to get an $N \times T$ data matrix $\mathbf{H}$. This makes *Beta Random Matrices* or *Dirichlet Random Matrices*. To observe the spectrum, we construct the covariance matrix to make *Beta Random Covariance Matrices* (BRCM) or *Dirichlet Random Covariance Matrices* (DRCM).

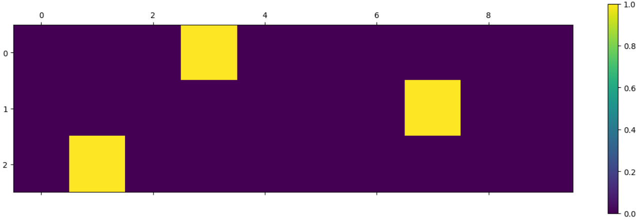


Figure 4: The DRCM is constructed by $\frac{1}{10}\mathbf{A}\mathbf{A}^\top$, where $\mathbf{A}$ is seen above.

Spectrum of the Beta Random Matrices

Conjecture

The spectrum of the BRCM, for arbitrary $\boldsymbol{\beta} = [\beta_1, \beta_2]^\top$, is of Marčenko-Pastur type.

Definition (Sub-Gaussian variables)

A random variable X with $\mu = \mathbb{E}[X] < \infty$ is *sub-Gaussian* if $\exists \sigma > 0$ such that

$$\mathbb{E}[\exp(\lambda(X - \mu))] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right), \quad \forall \lambda \in \mathbb{R}$$

Lemma (Marchal, Arbel, 2017[5])

The Beta(β, β) distribution is strictly sub-Gaussian.

Spectrum of the Beta Random Matrices (cont.)

Proof of Lemma.

First, observe that the j^{th} moment of Beta(β_1, β_2) for a random variable X is given by:

$$\mathbb{E}[X^j] = \frac{(\beta_1)_j}{(\beta_1+\beta_2)_j}, \quad \mathbb{E}[X] = \frac{\beta_1}{\beta_1+\beta_2}, \mathbb{V}[X] = \frac{\beta_1\beta_2}{(\beta_1+\beta_2)^2(\beta_1+\beta_2+1)}$$

Now letting $\beta_1 = \beta_2, \sigma^2(\boldsymbol{\beta}) = \mathbb{V}[\text{Beta}(\beta,\beta)] = 1/(4(2\beta+1))$. Also since X is symmetric around $\frac{1}{2}$, then the even moments are non-zero.

$$\mathbb{E}\left[\exp\left(X-\frac{1}{2}\right)\right] = \sum_{j=0}^{\infty} \mathbb{E}\left[(X-1/2)^{2j}\right]\frac{1}{(2j)!}$$

$$\mathbb{E}\left[\left(X-\frac{1}{2}\right)^{2j}\right]\frac{1}{(2j)!} = \frac{1}{2^{2j}j!}\frac{(\beta)_j}{(2\beta)_{2j}}$$

$$\leq \frac{1}{2^{2j}j!}\frac{1}{(2(2\beta+1))^j} = \frac{\sigma^{2j}(\boldsymbol{\beta})}{2^j j!}$$

□

More Results on the Spectrum of BRCM

Theorem (Vershynin, 2011(Thm 5.39)[6])

Let $\mathbf{A}$ *be an* $N \times n$ *matrix are independent sub-gaussian isotropic random vectors in* $\mathbb{R}^n$. *Then for every* $t \geq 0$ *with probability at least* $1 - 2\exp(-ct^2)$ *one has*

$$\sqrt{N} - C\sqrt{n} - t \leq s_{min}(\mathbf{A}) \leq s_{max}(\mathbf{A}) \leq \sqrt{N} + C\sqrt{n} + t. \tag{10}$$

More Results on the Spectrum of BRCM

Theorem (Vershynin, 2011(Thm 5.39)[6])

Let $\mathbf{A}$ *be an* $N \times n$ *matrix are independent sub-gaussian isotropic random vectors in* $\mathbb{R}^n$*. Then for every* $t \geq 0$ *with probability at least* $1 - 2\exp\left(-ct^2\right)$ *one has*

$$\sqrt{N} - C\sqrt{n} - t \leq s_{min}(\mathbf{A}) \leq s_{max}(\mathbf{A}) \leq \sqrt{N} + C\sqrt{n} + t. \quad (10)$$

Remark

We can even extend this to non-isotropic distributions, which the Beta distribution is since $\mathbb{E}[Beta(\beta_1, \beta_2)] = \frac{1}{N} > 0$.

More Results on the Spectrum of BRCM

Theorem (Vershynin, 2011(Thm 5.39)[6])

Let $\mathbf{A}$ *be an* $N \times n$ *matrix are independent sub-gaussian isotropic random vectors in* $\mathbb{R}^n$*. Then for every* $t \geq 0$ *with probability at least* $1 - 2\exp\left(-ct^2\right)$ *one has*

$$\sqrt{N} - C\sqrt{n} - t \leq s_{min}(\mathbf{A}) \leq s_{max}(\mathbf{A}) \leq \sqrt{N} + C\sqrt{n} + t. \tag{10}$$

Remark

We can even extend this to non-isotropic distributions, which the Beta distribution is since $\mathbb{E}[Beta(\beta_1, \beta_2)] = \frac{1}{N} > 0$.

Possible Proof Directions.

Since the Beta distribution is sub-Gaussian by the previous lemma [5] then by [6] we know that the BRCM has limiting spectral distribution which follows the domain of Marčenko-Pastur density with high probability. □

Spectrum of the Dirichlet Random Matrices

Conjecture

The spectrum of the DRCM is of Marčenko-Pastur type.

Theorem (Yaskov, 2016[7])

If $(\mathbf{x}_p^\top \mathbf{A}_p \mathbf{x}_p - \mathrm{tr}\{\mathbf{A}_p)\})/p \xrightarrow{p} 0$ *as* $p \to \infty$ *for all sequences of* $p \times p$ *complex matrices* $\mathbf{A}_p$ *with uniformly bounded spectral norms* $\|\mathbf{A}_p\|$, *then the spectrum converges weakly to Marčenko-Pastur with probability 1.*

Spectrum of the Dirichlet Random Matrices (cont.)

Proof.

We will use the Cauchy–Stieltjes transform method. By the Stieltjes continuity theorem, it suffices to prove that $s_n(z) \to s(z)$ a.s. $\forall z \in \mathbb{C}$ with $\mathrm{Im}(z) > 0$, where $s_n(z)$ and $s(z)$ are the Stieltjes transforms of μ and μ_{MP}, respectively

$$s_n(z) = \int_{\mathbb{R}} \frac{\mu(\mathrm{d}\lambda)}{\lambda - z} \quad \text{and} \quad s_(z) = \int_{\mathbb{R}} \frac{\mu_{\mathrm{MP}}(\mathrm{d}\lambda)}{\lambda - z}$$

Since μ is isotropic, $s_n(z) = \mathrm{tr}\left(n^{-1}\mathbf{X}\mathbf{X}^\top - z\mathbf{I}_p\right)^{-1}/p$. Now fix $z \in \mathbb{C}$ with $\mathrm{Im}(z) = v > 0$, then through a Martingale type argument, $s_n(z) - \mathbb{E}s_n(z) \to 0$ a.s.. Lastly through a Sherman-Morrison-Woodbury argument we arrive at $\mathbb{E}s_n(z) \to s(z)$. □

Spectrum of the Dirichlet Random Matrices (cont.)

Possible Proof Direction.

By [7], we need to prove that for $\mathbf{A}$, a Dirichlet Random Matrix, that the DRCM $\mathbf{C} = \mathbf{A}^\top \mathbf{A}$ has bounded operator norm. Since $\mathbf{C}_{ij} > 0, \in \mathbb{R}$, then by Perron-Frobenius we know that $\lambda_{\max} \leq \max_i \sum_j \mathbf{C}_{ij} < \infty$, so $\|\mathbf{C}\|_2$ is finite, and then all induced norms are bounded.

Next we need to show that, $\mathbb{V}[\mathbf{x}^\top \mathbf{C}\mathbf{x}/p] \to 0$. Consider $\mathbf{x} = \vec{\mathbf{1}}$, then since each row is rescaled to sum to 1, then we would have $\mathbf{x}^\top \mathbf{x} = p$. Since we are dividing by p, the variance of any constance is 0, so by Theorem, we have DRCM are of Marčenko-Pastur type. □

Numerical Results in the Bulk

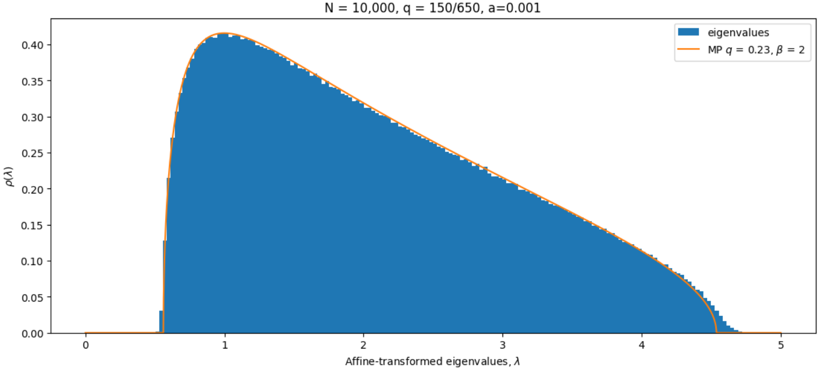


Figure 5: DRCM's spectrum converge to Marčenko-Pastur with hand tuned σ.

Numerical Results on the Edge

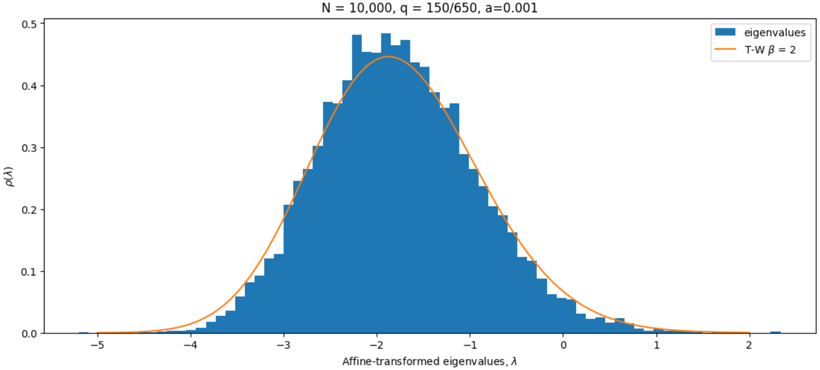


Figure 6: DRCM appears to have a Tracy-Widom like decay on the edge.

Table of Contents

Summary

In this talk, we have

- motivated compressed sensing from the Nyquist Sampling Theorem,

Summary

In this talk, we have

- motivated compressed sensing from the Nyquist Sampling Theorem,
- learned about Wishart covariance matrices and their spectrum (Marčenko-Pastur),

Summary

In this talk, we have

- motivated compressed sensing from the Nyquist Sampling Theorem,
- learned about Wishart covariance matrices and their spectrum (Marčenko-Pastur),
- examimed the Beta and Dirichlet distribution, exploiting them to make RM, and

Summary

In this talk, we have

- motivated compressed sensing from the Nyquist Sampling Theorem,
- learned about Wishart covariance matrices and their spectrum (Marčenko-Pastur),
- examimed the Beta and Dirichlet distribution, exploiting them to make RM, and
- conjectured Marčenko-Pastur spectrum from both classes, numerically verifying it.

Future Directions

Theoretic Considerations:

- The Beta and Dirichlet matrices have $\mathbb{E}(x_{ij}) \neq 0$, so I am trying to find a fix for these, either through a $\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}$ or reworking the proof.

Future Directions

Theoretic Considerations:

- The Beta and Dirichlet matrices have $\mathbb{E}(x_{ij}) \neq 0$, so I am trying to find a fix for these, either through a $\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}$ or reworking the proof.
- Use Markov style arguments and adjust the proof using [8, 9]

Future Directions

Theoretic Considerations:

- The Beta and Dirichlet matrices have $\mathbb{E}(x_{ij}) \neq 0$, so I am trying to find a fix for these, either through a $\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}$ or reworking the proof.
- Use Markov style arguments and adjust the proof using [8, 9]
- Determine the point mass at $x = 0$.

Future Directions

Theoretic Considerations:

- The Beta and Dirichlet matrices have $\mathbb{E}(x_{ij}) \neq 0$, so I am trying to find a fix for these, either through a $\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}$ or reworking the proof.
- Use Markov style arguments and adjust the proof using [8, 9]
- Determine the point mass at $x = 0$.

Numerical Considerations:

- Even with proof that these converge to Marčenko-Pastur distributions, the exact constants in the law aren't known.

Future Directions

Theoretic Considerations:

- The Beta and Dirichlet matrices have $\mathbb{E}(x_{ij}) \neq 0$, so I am trying to find a fix for these, either through a $\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}$ or reworking the proof.
- Use Markov style arguments and adjust the proof using [8, 9]
- Determine the point mass at $x = 0$.

Numerical Considerations:

- Even with proof that these converge to Marčenko-Pastur distributions, the exact constants in the law aren't known.
- I have hand tuned the σ parameter, but I would like to solve $\sigma = \sigma(q, \boldsymbol{\beta}, \dim(\boldsymbol{\beta}), \beta_i)$.

Future Directions

Theoretic Considerations:

- The Beta and Dirichlet matrices have $\mathbb{E}(x_{ij}) \neq 0$, so I am trying to find a fix for these, either through a $\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}$ or reworking the proof.
- Use Markov style arguments and adjust the proof using [8, 9]
- Determine the point mass at $x = 0$.

Numerical Considerations:

- Even with proof that these converge to Marčenko-Pastur distributions, the exact constants in the law aren't known.
- I have hand tuned the σ parameter, but I would like to solve $\sigma = \sigma(\boldsymbol{q}, \boldsymbol{\beta}, \dim(\boldsymbol{\beta}), \beta_i)$.
- Continue to compare different values of $\dim(\boldsymbol{\beta})$, (namely compare Beta and 2-Dirichlet)

EMORY UNIVERSITY

Future Directions

Theoretic Considerations:

- The Beta and Dirichlet matrices have $\mathbb{E}(x_{ij}) \neq 0$, so I am trying to find a fix for these, either through a $\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}$ or reworking the proof.
- Use Markov style arguments and adjust the proof using [8, 9]
- Determine the point mass at $x = 0$.

Numerical Considerations:

- Even with proof that these converge to Marčenko-Pastur distributions, the exact constants in the law aren't known.
- I have hand tuned the σ parameter, but I would like to solve $\sigma = \sigma(q, \boldsymbol{\beta}, \dim(\boldsymbol{\beta}), \beta_i)$.
- Continue to compare different values of $\dim(\boldsymbol{\beta})$, (namely compare Beta and 2-Dirichlet)
- Let $\boldsymbol{\beta}$ change from entry to entry.

EMORY UNIVERSITY

References

[1] C.E. Shannon. “Communication in the Presence of Noise”. In: *Proceedings of the IRE* 37.1 (Jan. 1949). Conference Name: Proceedings of the IRE, pp. 10–21. ISSN: 2162-6634. DOI: 10.1109/JRPROC.1949.232969. URL: https://ieeexplore.ieee.org/document/1697831 (visited on 12/03/2024).

[2] Emmanuel Candes, Justin Romberg, and Terence Tao. *Stable Signal Recovery from Incomplete and Inaccurate Measurements*. Dec. 7, 2005. DOI: 10.48550/arXiv.math/0503066. arXiv: math/0503066. URL: http://arxiv.org/abs/math/0503066 (visited on 12/03/2024).

[3] V A Marčenko and L A Pastur. “DISTRIBUTION OF EIGENVALUES FOR SOME SETS OF RANDOM MATRICES”. In: *Mathematics of the USSR-Sbornik* 1.4 (Apr. 30, 1967), pp. 457–483. ISSN: 0025-5734. DOI: 10.1070/SM1967v001n04ABEH001994. URL: https://www.mathnet.ru/eng/sm4101 (visited on 12/03/2024).

References (cont.)

[4] Philipp Fleig and Ilya Nemenman. *Generative random latent features models and statistics of natural images.* June 13, 2024. DOI: 10.48550/arXiv.2212.02987. arXiv: 2212.02987. URL: http://arxiv.org/abs/2212.02987 (visited on 12/03/2024).

[5] Olivier Marchal and Julyan Arbel. "On the sub-Gaussianity of the Beta and Dirichlet distributions". In: *Electronic Communications in Probability* 22 (none Jan. 1, 2017). ISSN: 1083-589X. DOI: 10.1214/17-ECP92. arXiv: 1705.00048[math]. URL: http://arxiv.org/abs/1705.00048 (visited on 12/03/2024).

[6] Roman Vershynin. *Introduction to the non-asymptotic analysis of random matrices.* Nov. 23, 2011. DOI: 10.48550/arXiv.1011.3027. arXiv: 1011.3027. URL: http://arxiv.org/abs/1011.3027 (visited on 12/03/2024).

References (cont.)

[7] Pavel Yaskov. "A short proof of the Marchenko–Pastur theorem". In: *Comptes Rendus Mathematique* 354.3 (Mar. 1, 2016), pp. 319–322. ISSN: 1631-073X. DOI: 10.1016/j.crma.2015.12.008. URL: https://www.sciencedirect.com/science/article/pii/S1631073X15003362 (visited on 12/03/2024).

[8] Charles Bordenave, Pietro Caputo, and Djalil Chafai. *Circular Law Theorem for Random Markov Matrices*. June 9, 2010. DOI: 10.48550/arXiv.0808.1502. arXiv: 0808.1502. URL: http://arxiv.org/abs/0808.1502 (visited on 12/03/2024).

[9] Djalil Chafai. *The Dirichlet Markov Ensemble*. Nov. 2, 2009. DOI: 10.48550/arXiv.0709.4678. arXiv: 0709.4678. URL: http://arxiv.org/abs/0709.4678 (visited on 12/03/2024).

[10] Olga Friesen and Matthias Löwe. "On the Limiting Spectral Density of Symmetric Random Matrices with Correlated Entries". In: *Random Matrices and Iterated Random Functions*. Ed. by Gerold Alsmeyer and Matthias Löwe. Berlin, Heidelberg: Springer, 2013, pp. 3–29. ISBN: 978-3-642-38806-4. DOI: 10.1007/978-3-642-38806-4_1.