Applications of representation theory to statistical problems

Lecture by Persi Diaconis
Notes by Dustan Levenstein

Examples of how statisticians could use representation theory.

(1) Spectral analysis of Time series (babies in New York City)

(2) Ranked data (election data)

(3) Representations of the unitary group U_n (relating to zeroes of the Riemann zeta function)

(4) Monster group

1 Time Series

Suppose we have numbers $f_0, \ldots, f_{N-1} \in \mathbb{R}$ which we think of as a signal. We look for patterns using the discrete Fourier transform

$$\hat{f}(j) = \sum_{k=0}^{N-1} f(k) e^{2\pi ijk/N}.$$

We can recover the signal by the Fourier inversion theorem,

$$f(k) = \frac{1}{N} \sum_{j=0}^{N-1} \hat{f}(j) e^{2\pi ijk/N}.$$

Suppose one $\hat{f}(j)$ is bigger than the rest: then to a good approximation, $f(k)$ is a simple periodic signal.

For example there was a blackout in New York City. The New York Times said there was a big spike in births 9 months after the blackout. There were about 400 kids born every day, and there was a 7 day periodicity to birth rates (fewer kids born on weekends), as well as a 3 year periodicity. Once these factors were removed there was no evidence of a spike.

2 Ranked Data

Suppose we have a study where you taste four chocolates, and rank them from 1 to 4.

Elections: You rank the candidates.

Card shuffling machine: Generates lots of permutations.

In all of these examples, the data is $\sigma_1, \ldots, \sigma_N \in S_n$. We have the associated function

$$f(\sigma) = \#n\{\sigma_n = \sigma\}.$$

Applying the Fourier transform associated to the group S_n gives

$$\hat{f}(\lambda) = \sum_{\sigma} f(\sigma) \rho_\lambda(\sigma),$$

where

$$\rho : S_n \to GL_{d_\lambda}(V)$$

is the corresponding irreducible representation and λ runs over all partitions of n. We can recover

$$f(\sigma) = \frac{1}{|G|} \sum_{t} d_\lambda T_n(\hat{f}(\rho_\lambda)\overline{\rho}_\lambda(\sigma)).$$

Example In the American Psychological Association election data, we have 5 candidates being ranked from 1 to 5, and we can see how many people voted for each permutation of 5.

Ranking	No. of votes cast of this type	Ranking	No. of votes cast of this type	Ranking	No. of votes cast of this type	Ranking	No. of votes cast of this type
54321	29	43521	91	32541	41	21543	36
54312	67	43512	84	32514	64	21534	42
54231	37	43251	30	32451	34	21453	24
54213	24	43215	35	32415	75	21435	26
54132	43	43152	38	32154	82	21354	30
54123	28	43125	35	32145	74	21345	40
53421	57	42531	58	31542	30	15432	40
53412	49	42513	66	31524	34	15423	35
53241	22	42351	24	31452	40	15342	36
53214	22	42315	51	31425	42	15324	17
53142	34	42153	52	31254	30	15243	70
53124	26	42135	40	31245	34	15234	50
52431	54	41532	50	25431	35	14532	52
52413	44	41523	45	25413	34	14523	48
52341	26	41352	31	25341	40	14352	51
52314	24	41325	23	25314	21	14325	24
52143	35	41253	22	25143	106	14253	70
52134	50	41235	16	25134	79	14235	45
51432	50	35421	71	24531	63	13542	35
51423	46	35412	61	24513	53	13524	28
51342	25	35241	41	24351	44	13452	37
51324	19	35214	27	24315	28	13425	35
51243	11	35142	45	24153	162	13254	95
51234	29	35124	36	24135	96	13245	102
45321	31	34521	107	23541	45	12543	34
45312	54	34512	133	23514	52	12534	35
45231	34	34251	62	23451	53	12453	29
45213	24	34215	28	23415	52	12435	27
[illegible]	[illegible]	34152	87	23154	186	12354	28
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	12345	30

We can look at what percentage of people ranked person i in position j — it looks pretty flat. Candidate 3 appears to be the favorite, but there is significant vote against candidate 3. This summarizes the original data by 16 numbers. Is it a good summary?

Candidate	Rank 1	2	3	4	5
1	18	26	23	17	15
2	14	19	25	24	18
3	28	17	14	18	23
4	20	17	19	20	23
5	20	21	20	19	20

We have the isotypic decomposition

$$\mathbb{Q}(S_n) = \bigoplus_{\lambda} V_\lambda.$$

In the case $n = 5$ there are 7 irreducible representations, and we can project the data into these 7 irreducible representations. What we see is the projection on V_3 is pretty large.

Decomposition of the regular representation

$M =$	V_1	$\oplus$	V_2	$\oplus$	V_3	$\oplus$	V_4	$\oplus$	V_5	$\oplus$	V_6	$\oplus$	V_7
Dim 120	1		16		25		36		25		16		1
SS/120	2286		298		459		78		27		7		0

Consider the representation

$$\rho(\sigma)_{i,j} = \sigma_{\sigma(i),j}.$$

The Fourier transform at this subrepresentation is

$$\hat{f}(\ell) = \sum_{\sigma} f(\sigma)\rho(\sigma)_{i,j}.$$

This Fourier transform is exactly the candidate i in position j table. This next table shows the candidate i in position j table with V_1 removed, and we can see it's the same set of numbers normalized so that they add up to zero.

First order effects

Candidate	Rank 1	2	3	4	5
1	−94	371	165	−145	−296
2	−372	−70	267	268	−92
3	461	−187	−354	−97	178
4	24	−175	−58	16	193
5	−18	62	−19	−41	17

Finally we can project the representation on pairs onto V_3, using the fact that V_3 is the representation on un-ordered pairs minus the occurrences of V_1 and V_2.

Second order, unordered effects

Candidate	**Rank 1, 2**	**1, 3**	**1, 4**	**1, 5**	**2, 3**	**2, 4**	**2, 5**	**3, 4**	**3, 5**	**4, 5**
1, 2	−137	−20	18	140	111	22	4	6	−97	−46
1, 3	476	−88	−179	−209	−147	−169	−160	107	128	241
1, 4	−189	51	113	24	−9	98	99	−65	23	−146
1, 5	−150	57	47	45	43	49	56	−48	−53	−48
2, 3	−42	84	19	−61	30	−16	82	−76	−39	72
2, 4	157	−20	−43	−25	−93	−76	−56	8	38	112
2, 5	22	−44	7	15	−117	69	25	62	99	−138
3, 4	−265	−7	72	199	39	140	85	19	−52	−233
3, 5	−169	10	88	70	78	44	47	−51	−36	−80
4, 5	296	−24	−142	−130	−5	−163	−128	38	−9	267

This data can be explained by observing a huge preference for candidates $1, 3$ together among one group of voters, witnessed by the number 476, and a not-as-big preference for candidates 4 and 5 witnessed by the 296. It turns out that 1 and 3 are "clinicians" and 4 and 5 are "academicians", two groups within the American Psychological Association which don't get along. Here Fourier analysis provides a clear picture of the data, where classical statistical analysis failed.

3 Unitary Group

Data: zeroes of the Riemann zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}, \text{ for } \operatorname{re} s > 1$$

It is well known that the zeroes lie within the critical strip $\{0 \leq \operatorname{re} s \leq 1\}$.

If $N(T)$ is defined as the number of zeroes of height T, Riemann showed

$$N(T) = \frac{T}{2\pi} \log \frac{T}{2\pi\ell} + O(\log T).$$

We can rescale the spacings so that they are, on average, spaced 1 apart.

The data in question here is 50000 zeroes from 10^{22} and up.

Hilbert and Polya hypothesized that the zeroes 'look like' the eigenvalues of the unitary matrices.

Pick $M \in U_n$, which has a Haar measure, eigenvalues $\{e^{i\theta_1}, \ldots, e^{i\theta_n}\}$. We can see the distribution of eigenvalues on a circle:

Figure 1:
Unitary Eigenvalues

Here's what uniform randomness looks like:

Figure 2:
Uniform Points

Here's what they would look like equally spaced:

Figure 3:
Picket Fence Model

The eigenvalues are about $\frac{1}{n}$ apart. The zeta zeroes are about $\frac{1}{\log T}$ apart. So we set

$$\frac{1}{n} = \frac{1}{\log T}.$$

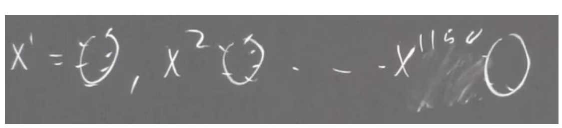
For us $n = 42$ — the zeroes of the zeta functions are supposed to look similar to the eigenvalues of 42×42 unitary matrices. In order to translate the Riemann zeta zeroes from a line into a circle, we take the first 42 zeroes among our 50000 zeta zeroes, and wrap them around the circle, then the next 42, etc. So we have around 1150 circles. (The orientations of each circle were chosen at random.)

The Haar density of of a vector θ of length 42 is given by

$$f(\theta) = \frac{1}{n!2^n} \prod_{j<k} |e^{i\theta_j} - e^{i\theta_k}|^2,$$

a classical formula due to Hurwitz.

So we have $X^1, \ldots, X^{1150}$ circles.



Does this match the Haar measure model from U_n?

One thing we can try is taking the trace: the sum of 42 random eigenvalues is well approximated by a Gaussian bell shaped curve (by the central limit theorem).

Theorem 3.1 *Pick $M \in U_n$ Haar measure.*

$$\sup_A \left| P(T_n(M) \in A) - \int_a \frac{e^{-|z|^2}}{\pi} \, dz \right| \leq \frac{c}{n!}.$$

Here $n = 42$, so $\frac{c}{n!}$ should be a very tiny number.

We can compare the norm-squared "traces" to the expected exponential distribution,

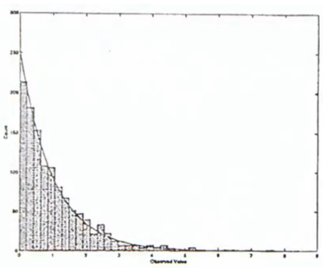


Figure 4: Zeta Function based Norm-Squared-"Traces"

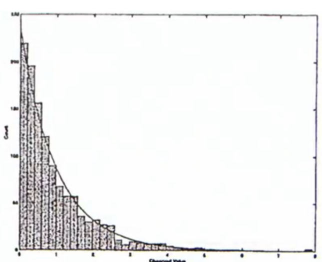


Figure 5: Exponentially Distributed Points

so for this test, the data matches the model.

This is, however, an ad-hoc test. Other tests might reject the model.

For a more systematic test: The characters of U_n are the Schur functions $s_\lambda(x_1, \ldots, x_n)$, where

$$s_0 = 1, s_1 = \sum x_j, s_2 = \sum x_j^2,$$

$$s_{1,1} = \sum_{j<k} x_j x_k,$$

etc. We have orthogonality

$$\langle s_\lambda \mid s_\mu \rangle_{U_n} = \delta_{\lambda,\mu}.$$

$$\mu_N(A) = \frac{1}{N} \#\{i : x^i \in A\}.$$

If $\{x^i\}$ is given by Haar measure, define

$$\hat{\mu}_N(\lambda) = \int s_\lambda(m) \mu_N(ds),$$

and $|\hat{\mu}_N(\lambda)|$, as the approximate inner product of s_λ and the zeta data, should be small, so

$$T_N = \sum_\lambda z^{|\lambda|} |\hat{\mu}_N(\lambda)|$$

gives a statistic.

So V_N converges to Haar $\iff \widehat{V}_N(\lambda) \to 0$.

We can compare these models:

Data Set	Haar on U_{42}	Zeta Zeros	Picket Fence	$U_{84}/2$
W_N	2.17	2.31	7.94	2.82

Table 2: $W_N = N(T_N - 1)$ with $N = 1000$, $n = 42$, $z = \frac{1}{2}$

Here we see that the data fits the model quite closely.

The character theory of μ_n comes in when we rewrite T_n:

$$T_N = \int \int \prod_{j,k} \left(1 - ze^{i(\theta_j - i\theta_k)}\right)^{-1} \mu_N(d\theta)\mu_N(d\theta').$$

Similar considerations arise for L-functions and involve other groups: $\mathcal{O}_n, Sp_{2n}, \ldots$

Similar problems arise for data on homogeneous spaces, e.g., in US lotteries, one picks a subset of $\{1, \ldots, n\}$, giving rise to data on $S_n/(S_k \times S_{n-k})$. This is significant for understanding lottery roll-overs — the lottery wants to understand how often these roll-overs happen, which has to do with how people choose their numbers.

4 Monster

All kinds of computational group theory uses random numbers — question, take a group you're interested in, take some random elements, and use the character theory to test whether the data matches the model.

See [3] for a general introduction to this topic. Section 2 is from [4]. The discrete Fourier transform is computed efficiently in [5].

Remark The Fourier transform computed naively takes $O(N^2)$ times, and the fast Fourier transform takes $O(N \log N)$ times, which is a significant improvement.

Section 3 comes from [1]. And to see an example of how one can do this analysis in a Bayesian manner, see [2].

References

[1] Marc Coram and Persi Diaconis. New tests of the correspondence between unitary eigenvalues and the zeros of riemann's zeta function. Journal of Physics A: Mathematical and General, 36(12):2883, 2003.

[2] P. Diaconis and G. Wang. Bayesian Goodness of Fit Tests: A Conversation for David Mumford. ArXiv e-prints, March 2018.

[3] Persi Diaconis. Group representations in probability and statistics, volume 11 of Institute of Mathematical Statistics Lecture Notes—Monograph Series. Institute of Mathematical Statistics, Hayward, CA, 1988.

[4] Persi Diaconis. A generalization of spectral analysis with application to ranked data. Ann. Statist., 17(3):949–979, 09 1989.

[5] Rockmore Diaconis. Efficient computation of the fourier transform on finite groups. Journal of the American Mathematical Society, 3(2):297–332, 04 1990.