Cross Validated

Why does inversion of a covariance matrix yield partial correlations between random variables?

Asked 7 years, 11 months ago Modified 7 months ago Viewed 26k times

51

I heard that partial correlations between random variables can be found by inverting the covariance matrix and taking appropriate cells from such resulting precision matrix (this fact is mentioned in http://en.wikipedia.org/wiki/Partial_correlation, but without a proof).

Why is this the case?

covariance covariance-matrix linear-algebra partial-correlation matrix-inverse

Share Cite Improve this question Follow

edited Mar 17, 2018 at 0:09 kjetil b halvorsen 70.4k 30 160 520

asked Mar 3, 2015 at 6:48 michal 1,238 3 12 14

1 If you mean to get partial correlation in a cell controlled for all the other variables, then the last paragraph here may shed light. – ttnphns Mar 3, 2015 at 8:38

4 Answers

Sorted by: Highest score (default)

46

When a multivariate random variable $(X_1, X_2, \ldots, X_n)$ has a nondegenerate covariance matrix $\mathbb{C} = (\gamma_{ij}) = (\text{Cov}(X_i, X_j))$, the set of all real linear combinations of the X_i forms an n-dimensional real vector space with basis $E = (X_1, X_2, \ldots, X_n)$ and a non-degenerate inner product given by

$$\langle X_i, X_j \rangle = \gamma_{ij} \ .$$

Its *dual basis with respect to this inner product*, $E^* = (X_1^*, X_2^*, \ldots, X_n^*)$, is uniquely defined by the relationships

$$\langle X_i^*, X_j \rangle = \delta_{ij} \ ,$$

the Kronecker delta (equal to 1 when $i = j$ and 0 otherwise).

The dual basis is of interest here because the partial correlation of X_i and X_j is obtained as the correlation between the part of X_i that is left after projecting it into the space spanned by all the other vectors (let's simply call it its "residual", $X_{i\circ}$) and the comparable part of X_j, its residual $X_{j\circ}$. Yet X_i^* is a vector that is orthogonal to all vectors besides X_i and has positive inner product with X_i whence $X_{i\circ}$ must be some non-negative multiple of X_i^*, and likewise for X_j. Let us therefore write

$$X_{i\circ} = \lambda_i X_i^*,\ X_{j\circ} = \lambda_j X_j^*$$

for positive real numbers λ_i and λ_j.

The partial correlation is the normalized dot product of the residuals, which is unchanged by rescaling:

$$\rho_{ij\circ} = \frac{\langle X_{i\circ}, X_{j\circ}\rangle}{\sqrt{\langle X_{i\circ}, X_{i\circ}\rangle\langle X_{j\circ}, X_{j\circ}\rangle}} = \frac{\lambda_i\lambda_j\langle X_i^*, X_j^*\rangle}{\sqrt{\lambda_i^2\langle X_i^*, X_i^*\rangle\lambda_j^2\langle X_j^*, X_j^*\rangle}} = \frac{\langle X_i^*, X_j^*\rangle}{\sqrt{\langle X_i^*, X_i^*\rangle\langle X_j^*, X_j^*\rangle}}\ .$$

(In either case the partial correlation will be zero whenever the residuals are orthogonal, whether or not they are nonzero.)

We need to find the inner products of dual basis elements. To this end, expand the dual basis elements in terms of the original basis E:

$$X_i^* = \sum_{j=1}^n \beta_{ij} X_j\ .$$

Then by definition

$$\delta_{ik} = \langle X_i^*, X_k\rangle = \sum_{j=1}^n \beta_{ij}\langle X_j, X_k\rangle = \sum_{j=1}^n \beta_{ij}\gamma_{jk}\ .$$

In matrix notation with $\mathbb{I} = (\delta_{ij})$ the identity matrix and $\mathbb{B} = (\beta_{ij})$ the change-of-basis matrix, this states

$$\mathbb{I} = \mathbb{B}\mathbb{C}\ .$$

That is, $\mathbb{B} = \mathbb{C}^{-1}$, which is exactly what the Wikipedia article is asserting. The previous formula for the partial correlation gives

$$\rho_{ij\cdot} = \frac{\beta_{ij}}{\sqrt{\beta_{ii}\beta_{jj}}} = \frac{\mathbb{C}_{ij}^{-1}}{\sqrt{\mathbb{C}_{ii}^{-1}\mathbb{C}_{jj}^{-1}}}\ .$$

Share Cite Improve this answer edited Oct 5, 2016 at 13:23 answered Jun 13, 2015 at 16:00

Follow

whuber ♦
304k 56 692
1192

3 +1, great answer. But why do you call this dual basis "dual basis with respect to this inner product" -- what does "with respect to this inner product" exactly mean? It seems that you use the term "dual basis" as defined here mathworld.wolfram.com/DualVectorSpace.html in the second paragraph ("Given a vector space basis $v_1, \ldots, v_n$ for V there exists a dual basis...") or here en.wikipedia.org/wiki/Dual_basis, and it's independent of any scalar product. – amoeba Nov 11, 2015 at 0:57

4 @amoeba There are two kinds of duals. The (natural) dual of any vector space V over a field R is the set of linear functions $\phi : V \to R$, called V^*. There is no canonical way to identify V^* with V, even though they have the same dimension when V is finite-dimensional. Any inner product γ corresponds to such a map $g : V \to V^*$, and *vice versa*, via

$$g(v)(w) = \gamma(v, w).$$

(Nondegeneracy of γ ensures g is a vector space isomorphism.) This gives a way to view elements of V as if they were elements of the dual V^*--but it depends on γ. – whuber ♦ Nov 11, 2015 at 1:22

3 @mpettis Those dots were hard to notice. I have replaced them with small open circles to make the notation easier to read. Thanks for pointing this out. – whuber ♦ Dec 18, 2015 at 18:22

4 @Andy Ron Christensen's *Plane Answers to Complex Questions* might be the sort of thing you are looking for. Unfortunately, his approach makes (IMHO) undue reliance on coordinate arguments and calculations. In the original introduction (see p. xiii), Christensen explains that's for pedagogical reasons. – whuber ♦ Dec 26, 2015 at 15:06

4 This answer is not technically correct and needs to be removed. It tries to show a relationship between partial correlations and entries of the precision matrix, but it uses an incorrect definition of the former (residualizing on all n-1 variables, not n-2). Consequently, it's not surprising that it ends up with a sign error (as the comments have noted). An authoritative reference is (Lauritzen, p130) is noted in the other answer. – user357269 Aug 3, 2020 at 22:24

25

Here is a proof with just matrix calculations.

I appreciate the answer by whuber. It is very insightful on the math behind the scene. However, it is still not so trivial how to use his answer to obtain the **minus sign** in the formula stated in the wikipediaPartial_correlation#Using_matrix_inversion.

$$\rho_{X_i X_j \cdot \mathbf{V} \setminus \{X_i, X_j\}} = -\frac{p_{ij}}{\sqrt{p_{ii} p_{jj}}}$$

To get this minus sign, here is a different proof I found in "Graphical Models Lauriten 1995 Page 130". It is simply done by some matrix calculations.

The key is the following matrix identity:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}G \\ -FE^{-1} & D^{-1} + FE^{-1}G \end{pmatrix}$$

where $E = A - BD^{-1}C$, $F = D^{-1}C$ and $G = BD^{-1}$.

Write down the covariance matrix as

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

where Ω_{11} is covariance matrix of (X_i, X_j) and Ω_{22} is covariance matrix of $\mathbf{V} \setminus \{X_i, X_j\}$.

Let $P = \Omega^{-1}$. Similarly, write down P as

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

By the key matrix identity,

$$P_{11}^{-1} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$$

We also know that $\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$ is the covariance matrix of $(X_i, X_j)|\mathbf{V} \setminus \{X_i, X_j\}$ (from Multivariate_normal_distribution#Conditional_distributions). The partial correlation is therefore

$$\rho_{X_i X_j \cdot \mathbf{V}\setminus\{X_i, X_j\}} = \frac{[P_{11}^{-1}]_{12}}{\sqrt{[P_{11}^{-1}]_{11}[P_{11}^{-1}]_{22}}}.$$

I use the notation that the (k, l)th entry of the matrix M is denoted by $[M]_{kl}$.

Just simple inversion formula of 2-by-2 matrix,

$$\begin{pmatrix} [P_{11}^{-1}]_{11} & [P_{11}^{-1}]_{12} \\ [P_{11}^{-1}]_{21} & [P_{11}^{-1}]_{22} \end{pmatrix} = P_{11}^{-1} = \frac{1}{\det P_{11}} \begin{pmatrix} [P_{11}]_{22} & -[P_{11}]_{12} \\ -[P_{11}]_{21} & [P_{11}]_{11} \end{pmatrix}$$

Therefore,

$$\rho_{X_i X_j \cdot \mathbf{V}\setminus\{X_i, X_j\}} = \frac{[P_{11}^{-1}]_{12}}{\sqrt{[P_{11}^{-1}]_{11}[P_{11}^{-1}]_{22}}} = \frac{-\frac{1}{\det P_{11}}[P_{11}]_{12}}{\sqrt{\frac{1}{\det P_{11}}[P_{11}]_{22}\frac{1}{\det P_{11}}[P_{11}]_{11}}} = \frac{-[P_{11}]_{12}}{\sqrt{[P_{11}]_{22}[P_{11}]_{11}}}$$

which is exactly what the Wikipedia article is asserting.

EDIT: This proof is only valid in the Gaussian case. The proof is actually more simple, and due to the particular definition of partial correlation in terms of residuals of *linear* regression. Note this is not the same as conditional expectation, see reference on wikipedia: Baba, Kunihiro; Ritei Shibata; Masaaki Sibuya (2004). "Partial correlation and conditional correlation as measures of conditional independence". Australian and New Zealand Journal of Statistics. 46 (4): 657–664. doi:10.1111/j.1467-842X.2004.00360.x. S2CID 123130024

I have added a proof (i.e. answer to this question) to the partial correlation wikipedia page now! (Don't have enough reputation to comment/post my own answer so stuck with editing I'm afraid!)

Share Cite Improve this answer Follow

edited Jul 22, 2022 at 4:51
Community Bot
1

answered Oct 31, 2017 at 8:54
Po C.
351 3 5

If we let `i=j` , then `rho_ii V\{X_i, X_i} = -1` , How do we interpret those diagonal elements in the precision matrix? – Jason May 23, 2018 at 3:19

Good point. The formula should be only valid for i=/=j. From the proof, the minus sign comes from the 2-by-2 matrix inversion. It would not happen if i=j. – Po C. May 23, 2018 at 3:42

So the diagonal numbers can't be associated with partial correlation. What do they represent? They are not just inverses of the variances, are they? – Jason May 23, 2018 at 4:49

This formula is valid for i=/=j. It is meaningless for i=j. – Po C. May 23, 2018 at 8:21

1 Is this proof true only for the multi-variate normal case? – Maverick Meerkat Feb 27, 2021 at 16:51

6

Note that the sign of the answer actually depends on how you define partial correlation. There is a difference between regressing X_i and X_j on the other $n-1$ variables separately vs. regressing X_i and X_j on the other $n-2$ variables together. Under the second definition, let the correlation between residuals ϵ_i and ϵ_j be ρ. Then the partial correlation of the two (regressing ϵ_i on ϵ_j and vice versa) is $-\rho$.

This explains the confusion in the comments above, as well as on Wikipedia. The second definition is used universally from what I can tell, so there should be a negative sign.

I originally posted an edit to the other answer, but made a mistake - sorry about that!

Share Cite Improve this answer Follow

answered Oct 9, 2016 at 22:21
Johnny Ho
63 1 3

1

For another perspective, this will examine the left inverse of a finite data matrix A. We can consider the data to be a sample rather than a theoretical distribution. While any distribution -- even continuous -- will have a covariance matrix, you can't generally talk about a data matrix unless you get into infinite vectors and/or special inner products.

So we have a finite sample in an n-by-m data matrix A. Let each column be one random variable. Then it's n samples and m random variables. Let A's columns (the random variables) be linearly independent (this is independence in the linear algebra sense, *not* as in independent random variables).

Let A be mean-centered already. Then,

$$C = \frac{1}{n} A^T A$$

is our covariance matrix. It's invertible since A's columns are linearly independent.

And we'll use later that $C^{-1} = n(A^T A)^{-1}$

The left inverse of A is

$B = (A^T A)^{-1} A^T.$

And we have

$BA = I_{m-by-m}.$

What do we know about B?

1. It's m-by-n. There's a row of B corresponding to each column of A.
2. Because $BA = I$, we know the inner product of the ith row of B with the ith column in A equals 1 (diagonal of I).
3. An inner product of the ith row of B with a jth ($i \neq j$) column of A is 0 (off-diagonal of I).
4. The right-most term in the expression for B is A^T. Therefore B's rows are in the rowspace of A^T, the column space of A.
5. by (4) and the fact that A's columns are mean-centered, B's rows must also be mean-centered.

Let x_i be the ith column of A.

The only vectors that have a non-zero inner product with the x_i, zero inner product with all other x_j, and are linear combinations of the columns of A, are vectors parallel to the residual of x_i after projecting it into the space spanned by all the other x_j.

Call these residuals r_i. And call the projection (the linear regression result) p_i. So the ith row of B must be parallel to r_i (6).

Now we know its direction, but what about magnitude? Let b_i be the ith *row* of B.

$$\begin{aligned}
1 &= b_i \cdot x_i && \text{by (2)} \\
&= b_i \cdot (p_i + r_i) && x_i \text{ is the sum of its projection and residual} \\
&= (b_i \cdot p_i) + (b_i \cdot r_i) && \text{linearity of dot product} \\
&= 0 + (b_i \cdot r_i) && \text{by (3), and that } p_i \text{ is a linear combination of the } x_j\text{s } (j \neq i) \\
&= (c_i r_i) \cdot r_i && \text{for some constant } c_i\text{, by (6)}
\end{aligned}$$

Therefore, $c_i = \frac{1}{r_i \cdot r_i} = \frac{1}{\|r_i\|^2}$, so $b_i = \frac{r_i}{\|r_i\|^2}$.

We now know what each row of B looks like. Notice

$$BB^T = ((A^T A)^{-1} A^T)(A((A^T A)^{-1})^T) = (A^T A)^{-1} = \frac{1}{n} C^{-1}$$

We can look at any i, jth element

$$C^{-1}_{ij} = n(BB^T)_{ij} = n(b_i \cdot b_j) = n\frac{r_i \cdot r_j}{\|r_i\|^2 \|r_j\|^2}$$

The $(r_i \cdot r_j)$ part of that should tell you we're getting close to covariances and correlations of these residuals. Conveniently, the diagonal elements look like

$$C^{-1}_{ii} = n\frac{r_i \cdot r_i}{\|r_i\|^2 \|r_i\|^2} = n\frac{1}{\|r_i\|^2}.$$

This quantity is exactly 1 over the variance of the residual r_i, $\frac{\|r_i\|^2}{n}$ (the n makes it a variance instead of a squared vector magnitude).

Then to get partial correlations you just need to combine the elements of C^{-1} in the way others have shown.

- Gilbert Strang lecture on left inverses
- Gilbert Strang lecture on projection, residuals

Share Cite Improve this answer Follow

edited Apr 19, 2020 at 1:25

answered Mar 22, 2020 at 3:26

MathFoliage
96 1 1 4

Highly active question. Earn 10 reputation (not counting the association bonus) in order to answer this question. The reputation requirement helps protect this question from spam and non-answer activity.