Advanced techniques in applied economics

Lecture 1: Conditional independence as structure

Piotr Zwiernik

Universitat Pompeu Fabra Barcelona

Statistics, Probability and Machine Learning Research Group

Barcelona School of Economics

Spring 2026

About the course

Why this course?

- We live in a world of high-dimensional dependence.
- Standard tools often break down when too many variables interact.
- Economists increasingly work with data where structure matters: networks, latent factors, strategic interactions, and policy spillovers.

1. Efficiency

Conditional independence tells us what can be ignored. This can turn an impossible problem into a tractable one.

2. Interpretation

Graphical structure helps separate direct from indirect relations. This is crucial for causal and policy reasoning.

3. Expressivity

Latent variables capture hidden drivers such as confounders, common shocks, market factors, and unobserved heterogeneity.

Core message

Conditional independence + graphs + latent variables = scalable structure for modern empirical work.

Course structure and objectives

Format

- **8 content lectures + 2 project sessions**
- Mixture of theory, examples, and hands-on implementation

Tools

- Mathematical derivations and visual intuition
- `R` examples using packages such as `huge`, `glasso`, `pcalg`, `dagitty`

Applied economics angle

I will try to connect the theory to problems economists care about:

- causal identification and policy design,
- financial and macroeconomic dependence,
- latent heterogeneity and confounding,
- interference and network spillovers.

Assessment

The main assessment is a research project. You may choose:

- **Applied project:** empirical analysis of a structured model.
- **Methods project:** deeper study of one model class or algorithm.

Roadmap for today

1. Basic notions: joint, marginal, conditional distributions
2. Independence and conditional independence
3. Regression and the Gaussian bridge
4. Why conditioning can change conclusions
5. Testing dependence and conditional independence in real data

Conditioning is not a technicality — it is often the whole story.

For background reading, see Højsgaard, Edwards, and Lauritzen, *Graphical Models with R*, Chapter 1, and R. Alexander, *Telling Stories with Data*, Chapter 15.

Part 1: Conditional independence

Random vectors and joint distributions

Let (X, Y) be a random vector.

Joint distribution

It describes the full probabilistic behaviour of (X, Y).

- **Continuous case:** density $f_{XY}(x, y)$
- **Discrete case:** probability mass function

$$f_{XY}(x, y) = \mathbb{P}(X = x, Y = y)$$

A joint distribution describes how variables move *together*, for example:

- inflation and unemployment,
- returns on two assets,
- income and consumption,
- treatment assignment and outcomes.

Marginal distributions

Marginals are obtained by summing or integrating out the other variable from $f_{XY}(x, y)$.

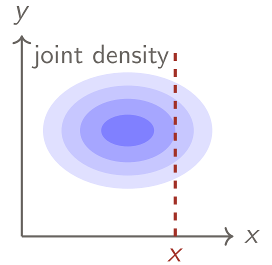
Marginal distribution of X

Continuous case:

$$f_X(x) = \int_{\mathbb{R}} f_{XY}(x, y)\, dy$$

Discrete case:

$$f_X(x) = \sum_y f_{XY}(x, y) = \mathbb{P}(X = x)$$



collect mass in the vertical slice

Conditional distributions

A conditional distribution tells us how one variable behaves once another variable is known.

Conditional distribution

For all x such that $f_X(x) > 0$,

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}.$$

In the discrete case this becomes

$$\mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)}.$$

Conditioning means that we restrict attention to a subpopulation:

- among employed workers,
- among treated units,
- among firms in distress,
- among borrowers with a given credit score.

Be careful interpreting conditional probabilities

Scenario: AI fraud detection

An algorithm flags fraudulent transactions (F). It catches 90% of fraud, and correctly labels 90% of clean transactions.

Joint probabilities

	Fraud (F)	Clean (F^c)
Flag (+)	0.009	0.099
No flag (-)	0.001	0.891

- **Sensitivity:** $\mathbb{P}(+ \mid F) = 0.9$
- **Specificity:** $\mathbb{P}(- \mid F^c) = 0.9$
- **Prevalence:** $\mathbb{P}(F) = 0.01$

Direction of conditioning matters

If a transaction is flagged, the probability that it is actually fraud equals

$$\mathbb{P}(F \mid +) = \frac{0.009}{0.009 + 0.099} \approx 0.083.$$

So even a strong test may generate many false alarms when the event is rare.

Independence

X and Y are **independent**, written $X \perp\!\!\!\perp Y$, if and only if

$$f_{XY}(x, y) = f_X(x) f_Y(y) \qquad \text{for all } x, y.$$

Equivalent reformulation

$X \perp\!\!\!\perp Y$ if and only if

$$f_{Y|X}(y|x) = f_Y(y) \qquad \text{for all } x \text{ such that } f_X(x) > 0.$$

So learning X does not change the distribution of Y.

Independence means that knowing X does not help to predict Y.

Conditional expectation and mean independence

The conditional expectation of Y given $X = x$ is

$$\mathbb{E}[Y \mid X = x] = \int y \, f_{Y|X}(y \mid x) \, dy.$$

$$\mathbb{E}[Y \mid X]$$

is the best predictor of Y from X under squared loss. It is the **regression function**.

Mean independence

We say that X and Y are mean independent if

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X] \qquad \text{and} \qquad \mathbb{E}[Y \mid X] = \mathbb{E}[Y].$$

Mean independence is strictly weaker than independence.

Conditional independence

X and Y are **independent given** Z, written $X \perp\!\!\!\perp Y \mid Z$, if

$$f_{XY|Z}(x, y|z) = f_{X|Z}(x|z) f_{Y|Z}(y|z) \qquad \text{for all } x, y, z.$$

Equivalently,

$$f_{Y|XZ}(y|x, z) = f_{Y|Z}(y|z).$$

Once Z is known, learning X gives no additional information about Y.

Conditional independence captures absence of a direct link.

Why economists should care

Conditional independence is everywhere in applied work.

Causal identification

Unconfoundedness is a conditional ind. statement: $(Y(1), Y(0)) \perp\!\!\!\perp D \mid X$.

Finance and macro

Sparse dependence means that many variables are unrelated once the right controls are included.

Selection problems

Apparent differences may disappear or reverse after conditioning on the right variables.

Latent structures

Many models are built by asserting that certain variables become independent once hidden factors are known.

In empirical work, assumptions are often best understood as conditional independence restrictions.

Part 2: Regression and the Gaussian bridge

CI and the regression function

Suppose

$$Y = \beta X + \gamma Z + \varepsilon, \qquad \varepsilon \perp\!\!\!\perp (X, Z).$$

If $\beta = 0$, then

$$Y \perp\!\!\!\perp X \mid Z.$$

So after controlling for Z, the variable X no longer matters.

Caution

If we only assume $\mathbb{E}[\varepsilon \mid X, Z] = 0$, then $\beta = 0$ implies only that the *conditional mean* does not depend on X. This does **not** imply full conditional independence in general.

The multivariate normal distribution

A random vector $X = (X_1, \ldots, X_m)$ is **Gaussian**, written

$$X \sim \mathcal{N}_m(\mu, \Sigma),$$

if the density is

$$f(x) = \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right\},$$

where $\mu \in \mathbb{R}^m$ is the mean vector and $\Sigma \in \mathbb{S}^m_+$ is the covariance matrix.

Why this matters

In Gaussian models, many conditional independence questions become algebraic. This is why Gaussian graphical models are so useful.

For a classic treatment of the multivariate normal distribution, see Mardia, Kent, and Bibby, *Multivariate Analysis*.

Gaussian marginals

Partition X into two subvectors:

$$X = \begin{pmatrix} X_A \\ X_B \end{pmatrix}, \qquad \mu = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix}.$$

For Gaussian vectors,

$$X_A \sim \mathcal{N}(\mu_A, \Sigma_{AA}).$$

Interpretation

To study X_A alone, we simply *drop* the other coordinates. No integration is needed.

Gaussian families are closed under marginalization.

Gaussian conditionals

Under the same partition, the conditional distribution of X_A given $X_B = x_B$ is Gaussian:

$$\mathbb{E}[X_A \mid X_B = x_B] = \mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(x_B - \mu_B),$$

$$\mathrm{var}(X_A \mid X_B = x_B) = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}.$$

- The conditional mean is linear in x_B.
- The conditional covariance matrix does not depend on x_B.

Under Gaussianity, conditioning preserves Gaussianity and linearity.

A special Gaussian feature

Recall: $X = (X_A, X_B)$.

For general distributions,

$$\operatorname{cov}(X_A, X_B) = 0$$

does **not** imply independence.

For jointly Gaussian vectors, however,

$$X_A \perp\!\!\!\perp X_B \qquad \iff \qquad \Sigma_{AB} = 0.$$

Many tools exploit Gaussian logic even when this is not made explicit. For example, some independence tests really test only zero covariance.

The precision matrix

The precision matrix is

$$K = \Sigma^{-1}.$$

Partition it in the same way:

$$K = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{BA} & K_{BB} \end{pmatrix}.$$

Using block matrix inversion and the Schur complement, one obtains

$$\operatorname{var}(X_A \mid X_B = x_B) = K_{AA}^{-1}.$$

Why this is important

- The covariance matrix describes marginal dependence.
- The precision matrix describes conditional dependence.

Precision matrix and conditional independence

Let $X = (X_1, \ldots, X_m)^\top \sim \mathcal{N}_m(\mu, \Sigma)$ and let $K = \Sigma^{-1}$.

For Gaussian data,

$$X_i \perp\!\!\!\perp X_j \mid X_{\{1,\ldots,m\}\setminus\{i,j\}} \qquad \Longleftrightarrow \qquad K_{ij} = 0.$$

A zero off-diagonal entry in the precision matrix means: once all other variables are controlled for, there is no remaining direct relation between X_i and X_j.

This is the bridge to graphs

Later we will encode these zeros by missing edges in an undirected graph.

Partial correlation

For three variables X, Y, Z, the partial correlation is

$$\rho_{X,Y\cdot Z} = \frac{\rho_{X,Y} - \rho_{X,Z}\rho_{Y,Z}}{\sqrt{(1-\rho_{X,Z}^2)(1-\rho_{Y,Z}^2)}}.$$

Measures the residual linear association between X and Y after removing the linear effect of Z.

Regression interpretation

$\rho_{X,Y\cdot Z} = 0$ if and only if in the linear regression of X on Y and Z, the coefficient of Y is zero.

In the Gaussian case, zero partial correlation is equivalent to conditional independence.

Economic interpretation of sparse precision matrices

Suppose X is a vector of many macro or financial variables.

Dense covariance

A dense covariance matrix means many variables move together marginally. This may simply reflect common shocks.

Sparse precision

A sparse precision matrix means that once the rest is controlled for, only a few pairs remain directly related.

Examples

- financial contagion networks,
- interbank or input-output dependence,
- psychometric and preference networks,
- sparse dependence in large macro panels.

Part 3: Why conditioning matters

A big lesson

Simpson’s paradox and Berkson’s paradox are not curiosities.

Statistical conclusions depend on what we condition on.

- The same data may suggest opposite conclusions under different conditioning sets.
- Conditioning can remove or add dependence.
- Positive association can become negative after conditioning and vice versa.

Conditioning can clarify a relationship, but it can also manufacture one.

An applied economics version of Simpson's paradox

Suppose we compare outcomes across two groups, say A and B.

Naive comparison: We observe

$$\mathbb{E}[Y \mid G = B] > \mathbb{E}[Y \mid G = A].$$

This might tempt us to conclude that group B does better.

What could go wrong?

The groups may differ in skill, education, sector, age, or risk exposure. So the marginal comparison may mix:

composition effects, treatment effects, and selection effects.

Before interpreting an association, ask: what should I condition on?

Example 1: Kidney stones

Ignoring confounding can lead to wrong causal conclusions.

- Study of 700 patients with kidney stones.
- Two treatments:
  - $T = a$: open surgery
  - $T = b$: percutaneous nephrolithotomy
- Overall recovery rates:
  - $\mathbb{P}(R = 1 \mid T = a) = 0.78$
  - $\mathbb{P}(R = 1 \mid T = b) = 0.83$

Based on overall rates, treatment b looks better.

This example comes from Charig et al., Comparison of treatment of kidney stones by open surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy, *British Medical Journal*, 1986.

Conditioning on stone size

Patients can be divided into two groups: small and large kidney stones.

	Overall	Small stones	Large stones
Treatment *a*	78%	93%	73%
Treatment *b*	83%	87%	69%

Treatment *a* is better in both subgroups.

Stone size acts as a **confounder**:

- large stones are harder to treat,
- more severe cases were more likely assigned to treatment *a*.

Example 2: UC Berkeley admissions

Admission figures for graduate school at UC Berkeley in 1973: 8442 men applied, with 44% admitted; 4321 women applied, with 35% admitted.

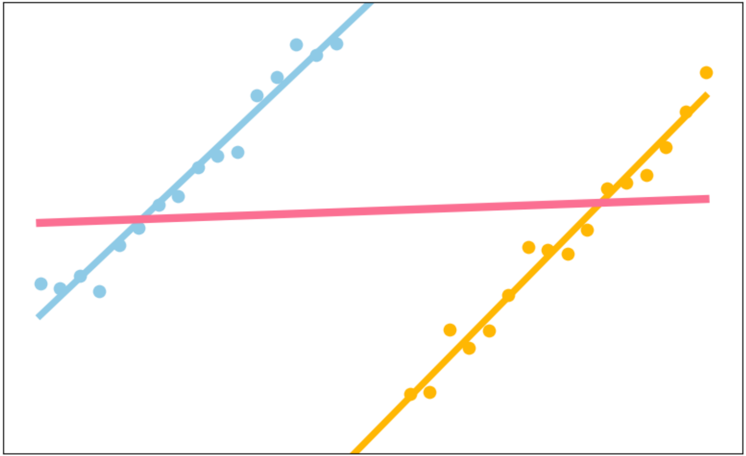
Conditioning on department gives:

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	825	62%	108	**82%**
B	560	63%	25	**68%**
C	325	**37%**	593	34%
D	417	33%	375	**35%**
E	191	**28%**	393	24%
F	373	6%	341	**7%**

Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation.

Bickel, Hammel, and OConnell, Sex Bias in Graduate Admissions: Data from Berkeley, *Science*, 1975.

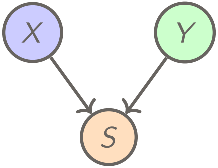
Simpson's paradox: Illustration



Berkson's paradox: conditioning on a collider

Suppose Looks (X) and Personality (Y) are independent in the population.

You only date people who are either attractive or nice. Then being selected into your dating pool depends on both variables.



Lesson

Even if $X \perp\!\!\!\perp Y$ in the population, conditioning on the common effect S may induce dependence:

$$X \not\!\perp\!\!\!\perp Y \mid S.$$

Economic versions of Berkson’s paradox

Selection can create misleading dependence in many economic settings.

Examples

- **Wage and job satisfaction** among workers who stay in a demanding sector.
  Workers with lower wages may stay only if job satisfaction is high, and vice versa.
- **Ability and family background** among students admitted to a selective program.
  Students with weaker background may enter only if ability is very high, and vice versa.

General lesson

These variables may be weakly related, or even unrelated, in the full population. But after conditioning on the selection event, they can become associated.

Conditioning on a consequence can create dependence that was not there before.

Exercise: mediation

Suppose

$$X,\ \varepsilon_Z,\ \varepsilon_Y \quad \text{are jointly independent,}$$

and

$$Z = aX + \varepsilon_Z, \qquad Y = bZ + \varepsilon_Y.$$

1. Is $X \perp\!\!\!\perp Y$?
2. Show that $Y \perp\!\!\!\perp X \mid Z$.
3. How would a regression of Y on X alone be interpreted?

X could be a policy variable, Z an intermediate behavioural response, and Y an outcome.

Exercise: confounding

Suppose

$$X = aU + \varepsilon_X, \qquad Y = bU + \varepsilon_Y,$$

with independent noise terms.

1. Is $X \perp\!\!\!\perp Y$?
2. Show that $X \perp\!\!\!\perp Y \mid U$.
3. Interpret U as an omitted variable.

U could be ability, firm quality, local demand, or risk preferences.

Exercise: selection bias

Suppose

$$X \perp\!\!\!\perp Y \qquad \text{and} \qquad S = X + Y + \varepsilon.$$

1. Are X and Y independent in the population?
2. What may happen to $\mathrm{cor}(X, Y \mid S > 0)$?
3. Give a real empirical example.

Hint

Selection is often an implicit conditioning step.

Part 4: Testing dependence in data

From theory to data

So far we discussed population properties such as

$$X \perp\!\!\!\perp Y \qquad \text{or} \qquad X \perp\!\!\!\perp Y \mid Z.$$

In data, we only observe a sample:

$$(X^{(1)}, Y^{(1)}), \ldots, (X^{(n)}, Y^{(n)}).$$

Goal: Use the sample to decide whether the dependence we see is plausibly just noise.

Practical question

Which notion of dependence are we testing?

- linear dependence,
- arbitrary dependence.

A first test: Pearson correlation

The most familiar test is based on the sample correlation coefficient r_n.

Consider the test

$$H_0 : \rho = 0 \qquad \text{vs.} \qquad H_A : \rho \neq 0.$$

When does this make sense?

Mainly when dependence is approximately linear and the Gaussian approximation is reasonable.

This is a test for **vanishing linear association**, not for full independence in general.

Fisher's z-transform test

Let r_n be the sample correlation from an *iid* sample $(X^{(1)}, Y^{(1)}), \ldots, (X^{(n)}, Y^{(n)})$. Define

$$Z_n = \frac{1}{2} \log\left(\frac{1+r_n}{1-r_n}\right).$$

If (X, Y) is bivariate normal with correlation ρ, then asymptotically

$$Z_n \approx \mathcal{N}\left(\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{n-3}\right).$$

```
# simulate sample correlations
library(MASS)
set.seed(1)
n <- 500
iter <- 1000
rho <- 0.2
srho <- rep(0, iter)
Sigma <- matrix(c(1, rho, rho, 1), 2, 2)
for (i in 1:iter) {
  x <- mvrnorm(n, mu = c(0, 0), Sigma = Sigma)
  srho[i] <- cor(x[,1], x[,2])
}
```

```
# Fisher transform and comparison
zrho <- 0.5 * log((1 + srho)/(1 - srho))

hist(zrho, prob = TRUE, breaks = 30,
     main = "", xlab = "z")

curve(dnorm(x,
      mean = 0.5*log((1+rho)/(1-rho)),
      sd = 1/sqrt(n-3)),
      add = TRUE, lwd = 2)
```

Pearson test in R

Fisher's test is implemented in R through `cor.test()`.

```
library(MASS)
set.seed(1)
rho <- 0.2
Sigma <- matrix(c(1, rho, rho, 1), 2, 2)
x <- mvrnorm(n, mu = c(0, 0), Sigma = Sigma)

cor.test(x[,1], x[,2], method = "pearson")
```

Try the same with $\rho = 0$.

What does the *p*-value mean?

It measures how surprising the observed sample correlation would be under the null hypothesis.

A non-significant result does **not** imply independence.

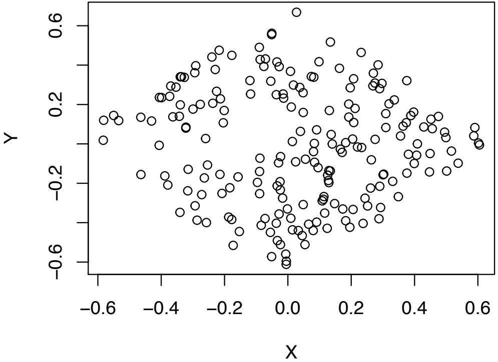
Limitation of Pearson correlation

A pair of variables can be strongly dependent and still have zero correlation.

Correlation measures only one aspect of dependence: the linear component.

Examples

- nonlinear deterministic relationships,
- symmetric patterns,
- dependence driven by tails or selection.



If economic data are heavy-tailed or nonlinear, Pearson correlation can miss important structure.

Non-Gaussianity issue

Vanishing covariance does not imply independence.

```
# generate sample from two uncorrelated but dependent random variables
set.seed(1); n <- 200
A <- runif(n) - 1/2
B <- runif(n) - 1/2
X <- t(c(cos(pi/4), -sin(pi/4)) %*% rbind(A, B))
Y <- t(c(sin(pi/4), cos(pi/4)) %*% rbind(A, B))
cor.test(X, Y, method = "pearson")
```

X and Y are uncorrelated but dependent.

Distance correlation

Distance correlation is a fully nonparametric measure of dependence.

It uses a test statistic $\mathcal{R}(X, Y)$ that satisfies

$$\mathcal{R}(X, Y) = 0 \qquad \Longleftrightarrow \qquad X \perp\!\!\!\perp Y.$$

```r
library(energy)
set.seed(1)
n <- 200
A <- runif(n) - 1/2
B <- runif(n) - 1/2
X <- A + B
Y <- A^2 - B^2
dcor.test(X, Y, R = 1000)
```

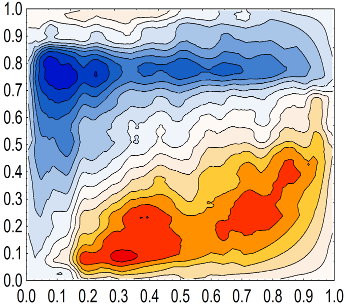
This gives a permutation test for general dependence, not just linear or monotone association.

Székely, Rizzo, and Bakirov, Measuring and testing dependence by correlation of distances, *Annals of Statistics*, 2007.

Cautionary example

Bowman and Azzalini (1997) analyze aircraft wing span and speed data.

```
library(sm); set.seed(1)
X <- aircraft$Span
Y <- aircraft$Speed
cor.test(X,Y)$p.value
dcor.test(X,Y,R=1000)$p.value
```



For discussion, see Ćmiel and Ledwina, Validation of Association, 2019.

χ^2-test for discrete data

For contingency tables, independence can be tested by comparing two models:

- the **full model**, which fits the cell probabilities without restrictions,
- the **independence model**, which assumes the row and column variables are independent.

Under the null hypothesis of independence, the likelihood-ratio statistic is asymptotically χ^2 with degrees of freedom equal to $(\#\text{rows} - 1)(\#\text{columns} - 1)$.

```
M <- as.table(rbind(c(762, 327, 468),
                    c(484, 239, 477)))
dimnames(M) <- list(gender = c("F", "M"),
                    party = c("Democrat","Independent","Republican"))

(Xsq <- chisq.test(M))
Xsq$expected
```

The output `Xsq$expected` gives the fitted counts under the independence model.

For contingency tables and asymptotic χ^2 tests, see Chapter 2 in Agresti, *Categorical Data Analysis*. For the graphical-model perspective, see Chapter 4 in Lauritzen, *Graphical Models*.

Part 5: Testing conditional independence

Testing conditional independence

Testing conditional independence is much harder than testing ordinary independence. Once we condition on Z, we must compare *conditional* laws rather than marginal ones.

- **Discrete data:** likelihood-ratio gives asymptotic χ^2 procedures.
- **Gaussian data:** conditional independence reduces to zero partial correlation.
- **Nonlinear / non-Gaussian data:** kernel and other nonparametric methods are available.

In practice, the difficulty of conditional independence testing depends strongly on the modelling assumptions.

```r
library(CondIndTests)
library(bnlearn)
set.seed(1)

n <- 100
Z <- rnorm(n)
X <- 4 + 2*Z + rnorm(n)
Y <- 3*X^2 + Z + rnorm(n)

CondIndTest(X, Y, Z,
          method = "KCI")$pvalue
bnlearn::ci.test(X, Y, Z)$p.value
```

Here X and Y are **not** conditionally independent given Z, so both procedures return very small p-values.

For a broader discussion of nonlinear CI testing, see Heinze-Deml, Peters, and Meinshausen, Invariant Causal Prediction for Nonlinear Models, 2018, and related literature on kernel CI tests.

Testing conditional independence in discrete data

For contingency tables, conditional independence can be tested by comparing:

- the **full model**, with no restrictions,
- the **conditional independence model**, which assumes $X \perp\!\!\!\perp Y \mid Z$.

The likelihood-ratio statistic compare the observed cell counts to the fitted counts under the conditional independence model. Under the null, these statistics are asymptotically χ^2.

```
data(UCBAdmissions)

bnlearn::ci.test(
  x = "Gender",
  y = "Admit",
  z = "Dept",
  test = "x2",
  data = as.data.frame(UCBAdmissions)
)
```

```
Pearson's X^2

data: Gender ~ Admit | Dept
x2 = 0, df = 6, p-value = 1
```

Here we test whether admission and gender are independent *within each department*. The null hypothesis is $\texttt{Gender} \perp\!\!\!\perp \texttt{Admit} \mid \texttt{Dept}$. The degrees of freedom are $(|\texttt{Gender}|-1)(|\texttt{Admit}|-1)|\texttt{Dept}| = (2-1)(2-1)\cdot 6 = 6$.

For log-linear models and conditional independence in contingency tables, see Chapter 7 in Agresti, *Categorical Data Analysis*. For the graphical-model perspective, see Chapter 4 in Lauritzen, *Graphical Models*.

Take-away

- Independence means factorization of the joint law.
- Conditional independence is a structural statement about what remains after conditioning.
- Mean independence is weaker than full independence.
- Conditioning can remove dependence, but it can also create it.
- In Gaussian models, conditional independence becomes algebraic: zeros in the precision matrix correspond to missing conditional associations.
- Testing dependence and conditional independence requires care: different tests target different notions of structure.

Main lesson of today: to understand dependence, always ask what is being conditioned on.

Looking ahead

Next lecture

We will encode conditional independence statements visually using **undirected graphs**.

What will come next?

- local, pairwise, and global Markov properties,
- Gaussian graphical models,
- sparse network estimation from data,
- applications to economic and financial dependence networks.

Missing edge $=$ conditional independence