STA 437/2005: Methods for Multivariate Data

Week 11: Conditional independence

Piotr Zwiernik

University of Toronto

Basic definitions

Random vector and independence

Let (X, Y) be a vector of two random variables.

Joint distribution

Density function $f_{XY}(x, y)$ if continuous.

Probability mass function $f_{XY}(x, y) = \mathbb{P}(X = x, Y = y)$ if discrete.

Marginal distribution

continuous: $f_X(x) = \int_{\mathbb{R}} f_{XY}(x, y)\mathrm{d}y$.

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

This can be generalized to random vectors.

Independence

If $f_{XY}(x, y)$ is the joint density (or PMF) of (X, Y) then X and Y are independent if and only if

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

We write $X \perp\!\!\!\perp Y$.

Recall:

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \quad \text{and} \quad \mathrm{var}(X) = \text{cov}(X, X).$$

The correlation $\rho_{X,Y}$ between X, Y is:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} \in [-1, 1].$$

If $X \perp\!\!\!\perp Y$ then $\rho_{X,Y} = 0$.

(but in general not the other way around, see slide 13)

Conditional distribution

Conditional distribution

In the discrete case the conditional probability mass function is defined as

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

for all x, y such that $\mathbb{P}(Y = y) > 0$ and so

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \text{ for all } x, y \text{ s.t. } f_Y(y) > 0.$$

In the continuous case we use the same definition.

Important reformulation of independence

$X \perp\!\!\!\perp Y$ if and only if $f_{X|Y}(x|y) = f_X(x)$.
(knowing Y brings no extra information about X)

A cautionary note

Note: $f_{X|Y}(x|y) \neq f_{Y|X}(y|x)$.

Example: A medical test for a disease D has outcomes $+$ and - with probabilities

	D	D^c
$+$	.009	.099
-	.001	.891

As needed $\mathbb{P}(+|D) = 0.9$ and $\mathbb{P}(-|D^c) = 0.9$. However, $\mathbb{P}(D|+) \approx 0.08$ (!)

Conditional independence

X, Y, Z random variables.

X is independent of Y given Z (write $X \perp\!\!\!\perp Y|Z$) if

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

Important reformulation of independence

$X \perp\!\!\!\perp Y|Z$ if and only if $f_{X|Y,Z}(x|y,z) = f_{X|Z}(x|z)$.
(if we observed Z, extra information about Y brings no extra information about X)

Testing independence

Recall: A statistical test

Given a statistical hypothesis $H_0 : \theta \in \Theta_0$, $H_1 : \theta \in \Theta_1$, a statistical test consists of a test statistics $T(X^{(1)}, \ldots, X^{(n)})$ and a rejection region, typically of the form

$$R = \{T(X^{(1)}, \ldots, X^{(n)}) > t\}.$$

If the null hypothesis is true T is unlikely to take large values.

Type I error: $\mathbb{P}(T \in R | H_0)$

although H_0 is true, it is rejected

Type II error: $\mathbb{P}(T \notin R | H_1)$

although H_0 is false, it is retained

A good test should minimize probabilities of both types of errors.

Testing independence

Data: $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{iid}{\sim} P_{X,Y}$.

Goal: Decide whether $X \perp\!\!\!\perp Y$.

Statistical test: $\qquad H_0 : X \perp\!\!\!\perp Y, \quad H_A : X \not\perp\!\!\!\perp Y$

There are many tests of independence.

We discuss some examples.

Test for vanishing correlation

Fisher's z-transform test for Gaussian data

Let r_n is the sample correlation coefficient from an *iid* sample $(X^{(i)}, Y^{(i)})$.

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 Z_n has asymptotically normal distribution with mean $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$ and variance $\frac{1}{n-3}$.

Fisher's z-transform test is implemented in R as `cor.test`.

Non-gaussianity may invalidate the test and affect its power.

Basic nonparametric test

Kendall's tau test for non-Gaussian data

Suppose a bivariate sample (x_i, y_i) for $i = 1, \ldots, n$ is given.

Pair (x_i, y_i) , (x_j, y_j) is concordant if $(x_i, y_i) < (x_j, y_j)$ or $(x_i, y_i) > (x_j, y_j)$. Otherwise discordant.

Define $\tau_{XY} = \frac{(\#\text{concordant}) - (\#\text{discordant})}{\binom{n}{2}} \in [-1, 1]$.

Test based on Kendell's τ statistic is implemented in R as `cor.test`.

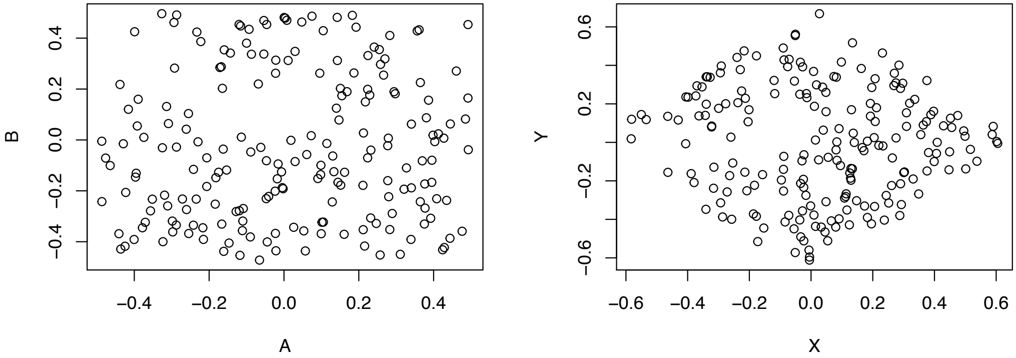
```
> set.seed(1); n <- 200; rho <- 0.2; Z <- runif(n);
> X <- runif(n)^2+sqrt(rho)*Z; Y <- runif(n)+sqrt(rho)*Z
> cor.test(X, Y, method = "pearson")$p.value
[1] 0.03417231
> cor.test(X, Y, method = "kendall")$p.value
[1] 0.01100592
```

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")
# Pearson's product-moment correlation
data:  X and Y
t = -0.84711, df = 198, p-value = 0.398
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1971897  0.0793095
sample estimates:
        cor
-0.06009275
```

X and Y are uncorrelated but dependent!



We see that X and Y are highly dependent.

Test based on distance correlation

Distance correlation $\mathcal{R}(X, Y)$ provides a test which applies when X, Y are two random **vectors** of any dimensions.

$\mathcal{R}(X, Y) = 0$ if and only if X and Y are independent.

The sample version of $\mathcal{R}(X, Y)$ gives a nonparametric test of independence.

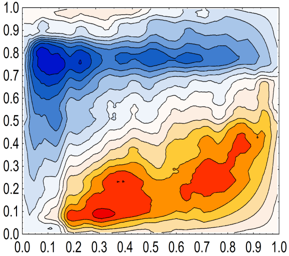
```
> library(energy); set.seed(1); n <- 200
> A <- runif(n)-1/2; B <- runif(n)-1/2
> X <- t(c(cos(pi/4),-sin(pi/4)) %*% rbind(A,B))
> Y <- t(c(sin(pi/4),cos(pi/4)) %*% rbind(A,B))
> dcor.test(X,Y,R=1000)

# dCor independence test (permutation test)
data:  index 1, replicates 1000
dCor = 0.21161, p-value = 0.004995
sample estimates:
      dCov       dCor    dVar(X)    dVar(Y)
0.03999654 0.21160982 0.17870935 0.19990601
```

Here $R = 1000$ is the number of the permutation bootstrap replications.

Another cautionary example

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




```
> library(sm); set.seed(1);
> X <- aircraft$Span
> Y <- aircraft$Speed
> cor.test(X,Y)$p.value
[1] 0.7816014
> dcor.test(X,Y,R=1000)$p.value
[1] 0.000999001
```

χ^2-test for discrete data

```
> 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))  # Prints test summary

Pearsons Chi-squared test

data:  M
X-squared = 30.07, df = 2, p-value = 2.954e-07

> Xsq$expected   # expected counts under the null
      party
gender Democrat Independent Republican
     F 703.6714    319.6453   533.6834
     M 542.3286    246.3547   411.3166
```

df = 2 is the difference between 5 (saturated model) and 3 (independence)

Testing conditional independence

Testing conditional independence is hard in general.

For discrete data we have the asymptotic χ^2-test.

Some parametric tests are implemented in the library `bnlearn`.
Many non-parametric methods have been implemented in `CondIndTest`

```
> 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")$pvalue
[1] 2.419926e-10
> bnlearn::ci.test(X,Y,Z)$p.value
[1] 1.15458e-25
```

See Section 3 in: C. Heinze-Deml, J. Peters, N. Meinshausen, Invariant Causal Prediction for Nonlinear Models, Journal of Causal Inference, 2018.

See: `http://www.bnlearn.com/documentation/man/conditional.independence.tests.html`

Simpson's paradox: UC Berkeley admissions example

The admission figures of the grad school at UC Berkeley in 1973: 8442 (44%) men, 4321 (35%) women admitted.

The same data conditioned on the department are:

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 et al, *Sex Bias in Graduate Admissions: Data From Berkeley*, Science, 1975)

In R:

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

Pearsons X^2
data:  Gender ~ Admit | Dept
x2 = 0, df = 6, p-value = 1
alternative hypothesis: true value is greater than 0

# gRim gives a slightly more refined output
> gRim::ciTest(as.data.frame(UCBAdmissions),set=~Gender+Admit+Dept)

set: [1] "Gender" "Admit"  "Dept"
Testing Gender _|_ Admit | Dept
Statistic (DEV):    0.000 df: 6 p-value: 1.0000 method: CHISQ

Slice information:
  statistic p.value df Dept
1         0       1  1    A
2         0       1  1    B
3         0       1  1    C
4         0       1  1    D
5         0       1  1    E
6         0       1  1    F
```

Conditional independence for Gaussian distributions

Recall: Marginal and conditional distributions

Split X into two blocks $X = (X_A, X_B)$. Denote

$$\mu = (\mu_A, \mu_B) \qquad \text{and} \qquad \Sigma = \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}.$$

Marginal distribution

$X_A \sim N_{|A|}(\mu_A, \Sigma_{AA})$

Conditional distribution

$X_A | X_B = x_B \sim N_{|A|} \left(\mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(x_B - \mu_B), \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA} \right)$

- Note that the conditional covariance is constant.

Conditional independence

Independence and conditional independence

$X_i \perp\!\!\!\perp X_j$ if and only if $\Sigma_{ij} = 0$.

$X_i \perp\!\!\!\perp X_j | X_C$ if and only if $\Sigma_{ij} - \Sigma_{i,C}\Sigma_{C,C}^{-1}\Sigma_{C,j} = 0$

Let $R = V \setminus \{i, j\}$. The following are equivalent:

- $X_i \perp\!\!\!\perp X_j | X_R$
- $\Sigma_{ij} - \Sigma_{i,R}\Sigma_{R,R}^{-1}\Sigma_{R,j} = 0$
- $(\Sigma^{-1})_{ij} = 0$

Useful: `https://en.wikipedia.org/wiki/Block_matrix#Block_matrix_inversion`