Variance, Covariance, correlation, dependence and causation

Independent Events

Definition:

$$P(A \cap B) = P(A)P(B)$$

What about these?

$$P(A \cap \bar{B}), P(\bar{A} \cap B), P(\bar{A} \cap \bar{B})$$

Implied by the definition

$$P(A \cap \bar{B}) = P(A - A \cap B) = P(A) - P(A \cap B) =$$

$$= P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\bar{B})$$

Conditional Probabilities

Definition:

$$P(A|B) \doteq \frac{P(A \cap B)}{P(B)}$$

Intuition:

Probability of A if we already know that sample is in B

If A and B are independent

$$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$$

Dependent Random Variables

Probability associated with each combination of values

marginal distributions

	X=0 0.4	X=1 0.3	X=2 0.3
Y=0 0.6	0.3	0.2	0.1
Y=1 0.4	0.1	0.1	0.2

Independent random variables

$$\forall x, y \quad P(X = x \wedge Y = y) = P(X = x)P(Y = y)$$

Independent random variables

$$\forall x, y \quad P(X = x \wedge Y = y) = P(X = x)P(Y = y)$$

marginal distributions

		X=0	X=1	X=2
		0.2	0.1	0.7
Y=0	0.4	0.08	0.04	0.28
Y=1	0.6	0.12	0.06	0.42

Expected Value

- Suppose X is a discrete random variable $P(X = a_i) = p_i$
  - The expected value of X is $E(X) = \sum_{i=1}^{n} p_i a_i$
- Suppose X is a continuous random variable with density f
  - The expected value of X is $E(X) = \int_{-\infty}^{+\infty} f(x) x dx$
- $E(X)$ is a property of the distribution, it is not a random variable.
- The average is a random variable:
  - $Average(x_1, x_2, \dots, x_n) \doteq \frac{1}{n} \sum_{i=1}^{n} x_i$
- When n is large, the average tends to be close to the mean.

Lets use $\mu \doteq E(X)$

We already know that $E(X-\mu)=0$

To find the width we could use $E\left(\left|X-\mu\right|\right)$

But it is much more convenient to use:

$$Var(X) \doteq E\left((X-\mu)^2\right)$$

Using the rules for expected value (remember that μ is a constant)

$$Var(X) \doteq E\left((X-\mu)^2\right) = E\left(X^2 - 2\mu X + \mu^2\right)$$

$$= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - E(X)^2$$

Properties of the variance

If a is a constant and X is a random variable then

$$
\begin{aligned}
Var(X+a) &= E\left[\left((X+a)-E[X+a]\right)^2\right] = \\
&= E\left[\left((X+a)-E[X]+a\right)^2\right] = \\
&= E\left[\left(X-E[X]\right)^2\right] = Var(X)
\end{aligned}
$$

$$
\begin{aligned}
Var(aX) &= E\left[\left(aX-E[aX]\right)^2\right] = \\
&= E\left[a^2\left(X-E[X]\right)^2\right] = a^2 Var(X)
\end{aligned}
$$

$$Var(X+Y) = E\left[\left((X+Y) - E[X+Y]\right)^2\right] =$$

$$= E\left[\left((X - E[X]) + (Y - E[Y])\right)^2\right] =$$

$$= E\left[(X - E[X])^2 + 2(X - E[X])(Y - E[Y]) + (Y - E[Y])^2\right]$$

$$= Var(X) + Var(Y) + 2Cov(X,Y)$$

$$Cov(X,Y) \doteq E\left[(X - E[X])(Y - E[Y])\right] = E[XY] - E[X]E[Y]$$

$$Cov(X,X) = Var(X)$$

If X,Y are independent then (assuming they are integer valued)

$$E[XY] = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} ij \Pr[X = i \wedge Y = j] = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} ij \Pr[X = i]\Pr[Y = j] =$$

$$= \sum_{i=-\infty}^{\infty} i \Pr[X = i] + \sum_{j=-\infty}^{\infty} j \Pr[Y = j] = E[X]E[Y] \Rightarrow Cov(X,Y) = 0$$

Getting the right dependence on units:

$Var(aX) = a^2 Var(X)$ -- does not represent the width of the distribution

$std(X) \doteq \sigma(X) \doteq \sqrt{Var(X)} \Rightarrow \sigma(aX) = a\sigma(X)$

Removing the effect of units:

$Cov(aX, bY) = abCov(X, Y)$ - Covariance depends on units

$$Corr(X,Y) \doteq \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} \Rightarrow \begin{array}{c} Corr(aX,bY) = Corr(X,Y) \\ \underline{\text{if } a,b>0} \end{array}$$

Unlike the Covariance, the Correlation Coefficient is unit-less,

Changing the units, or multiplying each random variable by some constant, does not change the correlation coefficient.

The correlation Coefficient is always in the range $[-1,+1]$

Expected value for a product of independent RVs

$$E(XY) = \sum_x \sum_y xyP(X = x \wedge Y = y) =$$

Expected value for a product of independent RVs

$$E(XY) = \sum_x \sum_y xyP(X = x \wedge Y = y) =$$

$$= \sum_x \sum_y xyP(X = x)P(Y = y) = \sum_x xP(X = x) \ \sum_y yP(Y = y) =$$

Expected value for a product of independent RVs

$$E(XY) = \sum_x \sum_y xyP(X = x \wedge Y = y) =$$

$$= \sum_x \sum_y xyP(X = x)P(Y = y) = \sum_x xP(X = x) \ \sum_y yP(Y = y) =$$

$$= E(X)E(Y)$$

Covariance

Recall $\mu_X \doteq E(X),\ \mu_Y \doteq E(Y)$

$$\mathrm{Cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y)) =$$

$$= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y = E(XY) - E(X)E(Y)$$

Recall $\mathrm{Var}(X) = E(X^2) - E(X)^2 = \mathrm{Cov}(X,X)$

Cov(X,Y)≠0 implies that X and Y are not independent
but Cov(X,Y)=0 **does not imply** that X and Y are independent

Go back to circle example

Correlation coefficient

$$\mathrm{Corr}(X,Y) \doteq \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

- Corr(aX+c,bY+d)=Corr(X,Y) if a,b>0
- Corr(X,Y) varies from -1 to +1
- Corr(X,Y)>0 ⇔ X and Y are "correlated"
- Corr(X,Y)<0 ⇔ X and Y are "anti-correlated"
- Corr(X,Y)=1 ⇔ X=aY, a>0
- Corr(X,Y)=-1 ⇔ X=aY a<0

Correlation coefficient

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

- The covariance depends on scaling and units, correlation coefficient does not

$$\forall a > 0, b > 0 \qquad Corr(aX, bY) = Corr(X,Y)$$

- The correlation coefficient varies between -1 and 1.

Examples

Correlated Variables

	X=1	X=2	X=3	X=4
Y=1	1/4	1/4	0	0
Y=2	0	0	0	0
Y=3	0	0	1/4	1/4

$$\mu(X) = 2.5,\ \mu(Y) = 2$$

$$\mathrm{cov}(X,Y) = \frac{1}{4}(-1.5 * -1) + \frac{1}{4}(-.5 * -1) + \frac{1}{4}(.5 * 1) + \frac{1}{4}(1.5 * 1) = 1$$

Anti Correlated Variables

	X=1	X=2	X=3	X=4
Y=1	0	0	0	1/4
Y=2	0	1/4	1/4	0
Y=3	1/4	0	0	0

$$\mu(X) = 2.5,\ \mu(Y) = 2$$

$$\mathrm{cov}(X,Y) = \frac{1}{4}(-1.5 * 1) + \frac{1}{4}(-.5 * 0) + \frac{1}{4}(.5 * 9) + \frac{1}{4}(1.5 * -1) = -\frac{3}{4}$$

Uncorrelated and independent

	X=1	X=2	X=3	X=4
Y=1	1/4	0	0	1/4
Y=2	0	0	0	0
Y=3	1/4	0	0	1/4

$$\mu(X) = 2.5,\ \mu(Y) = 2$$

$$\operatorname{cov}(X,Y) = \frac{1}{4}(-1.5*1) + \frac{1}{4}(-1.5*-1) + \frac{1}{4}(1.5*1) + \frac{1}{4}(1.5*-1) = 0$$

P(X=1)=P(X=4)=1/2, P(Y=1)=P(Y=3)=1/2

X and Y are independent because all of the joint probabilities are either 0 or 1/4

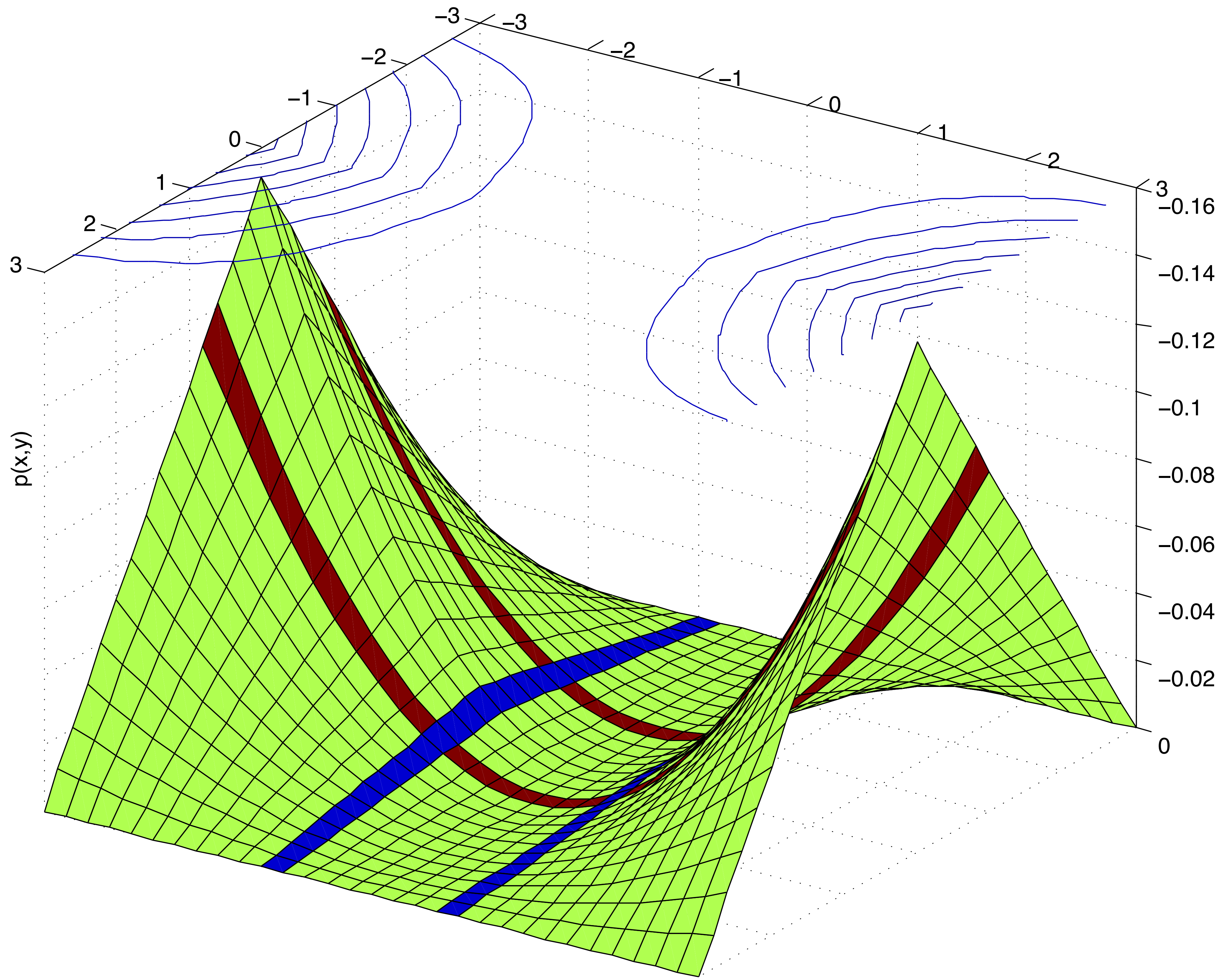
Uncorrelated but dependent

	X=1	X=2	X=3	X=4
Y=1	1/8	0	0	1/8
Y=2	0	1/4	1/4	0
Y=3	1/8	0	0	1/8

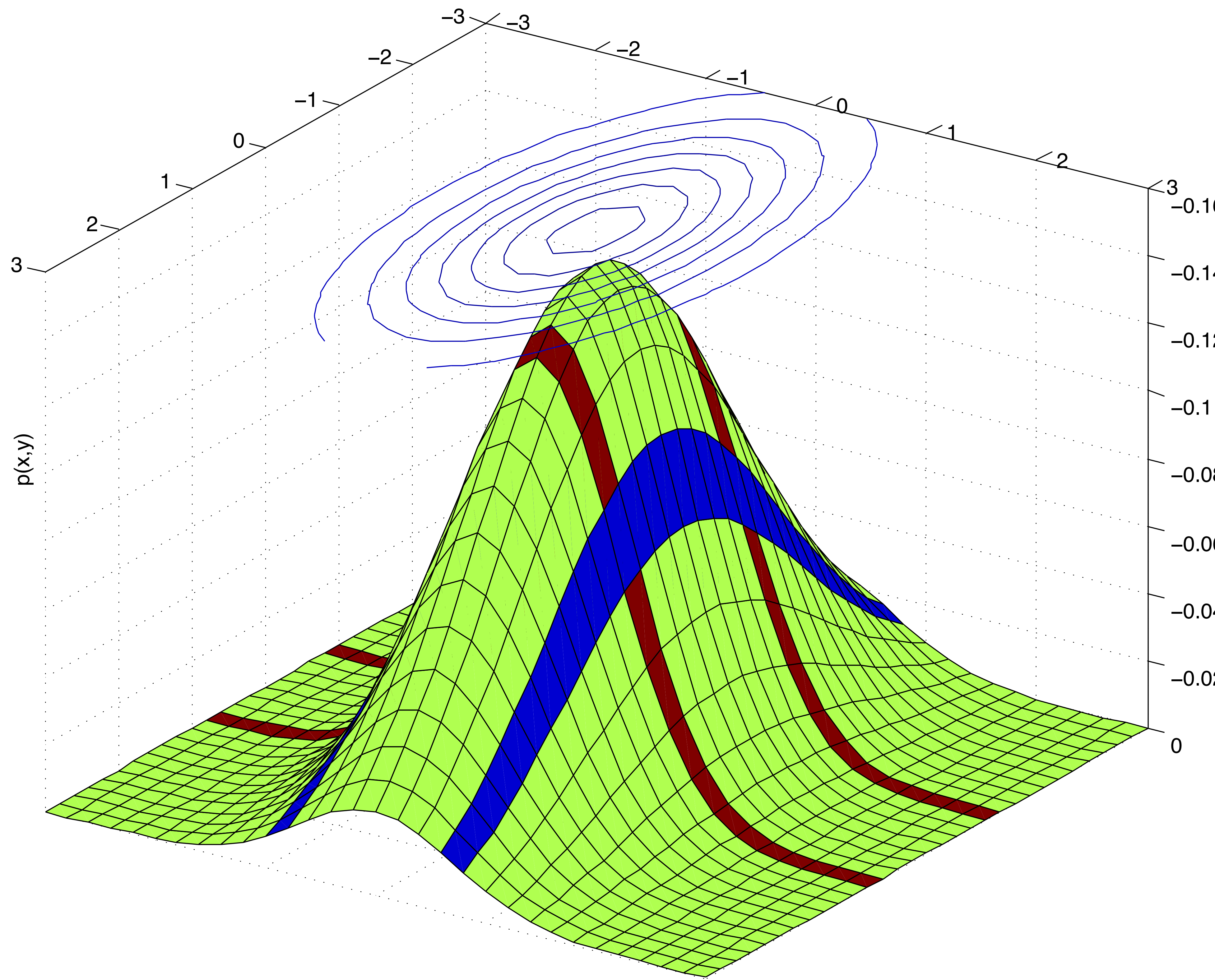
1. Cov(X,Y)=0
2. X and Y are independent

A. 1 and 2 B. 1 and not 2 C. not 1 and 2 D. not 1 and not 2

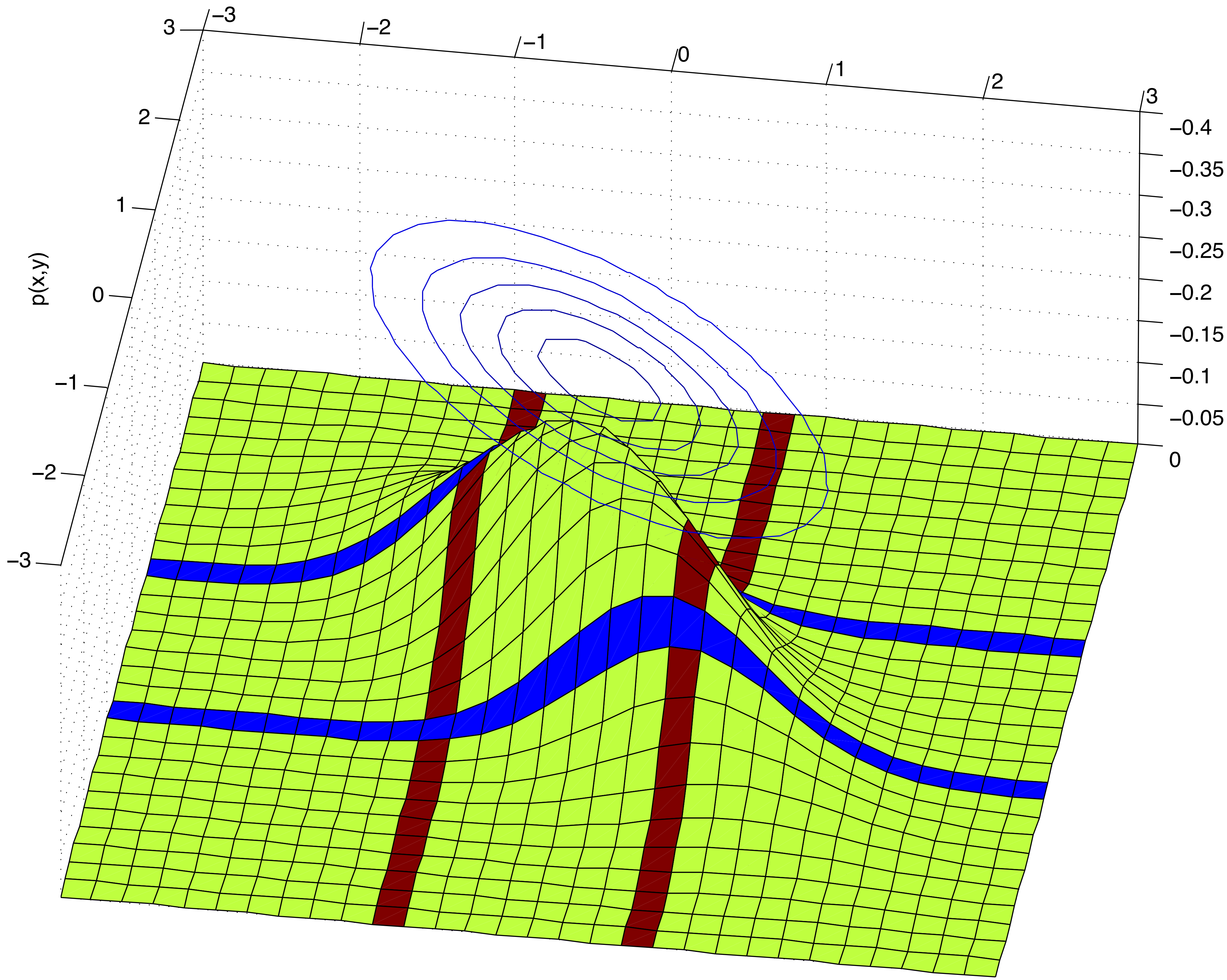
Example 1, independent RVs



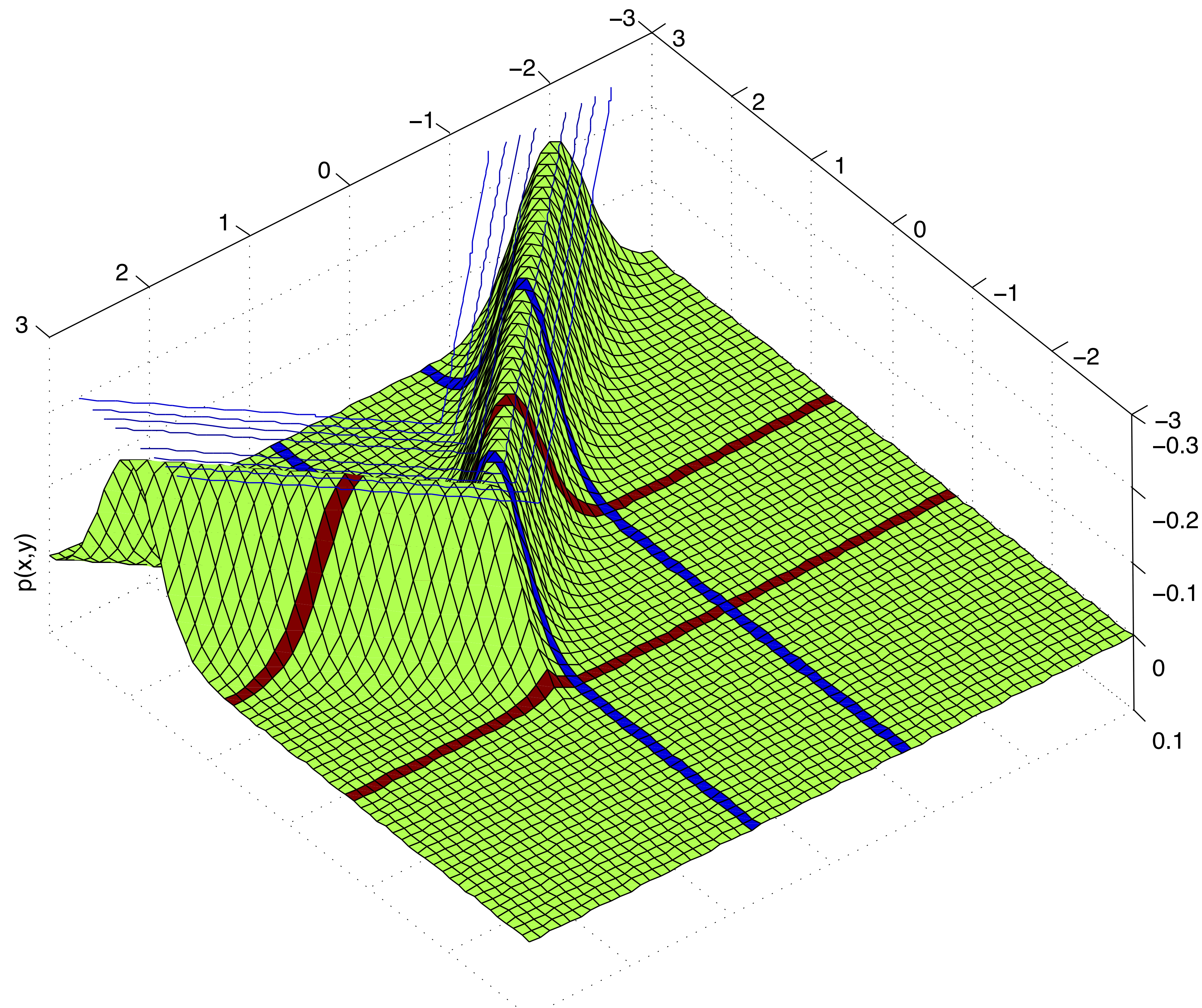
Example 2 independent RVs



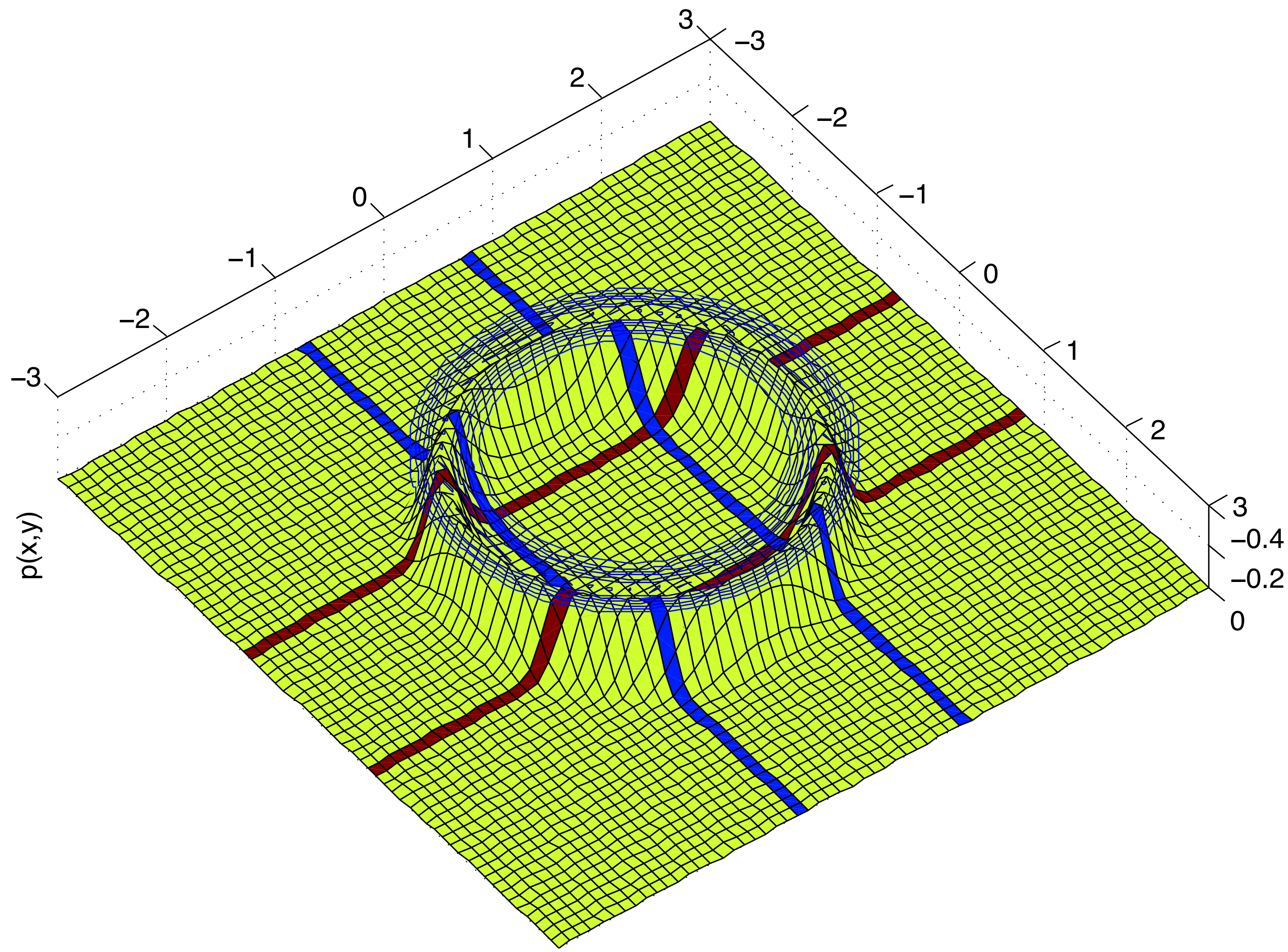
Example 3, Dependent RVs



Example 4, functional dependence



Example 5, Circle



Correlation vs. Dependence

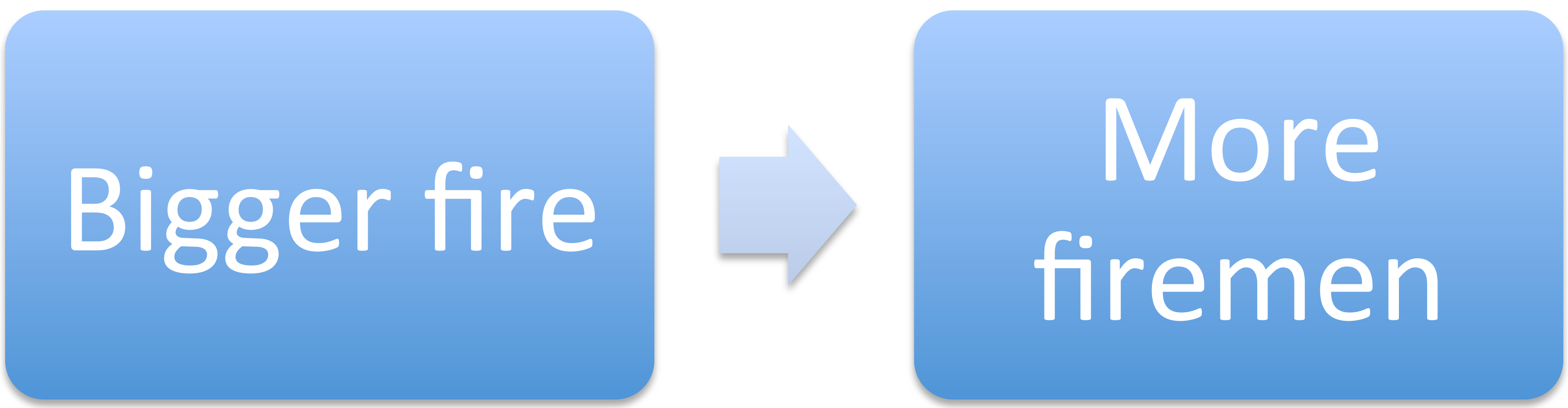
- Non-zero Correlation implies dependence
- Dependence does not imply correlation

Correlation vs Causation

- Using correlation because common, same can be said regarding Dependence vs. causation.
- The simple case is: the number of mosquitoes is correlated with the number of malaria cases. Therefor mosquitoes cause malaria. Which is true.
- However, one can deduce that malaria causes mosquitoes, which is false.

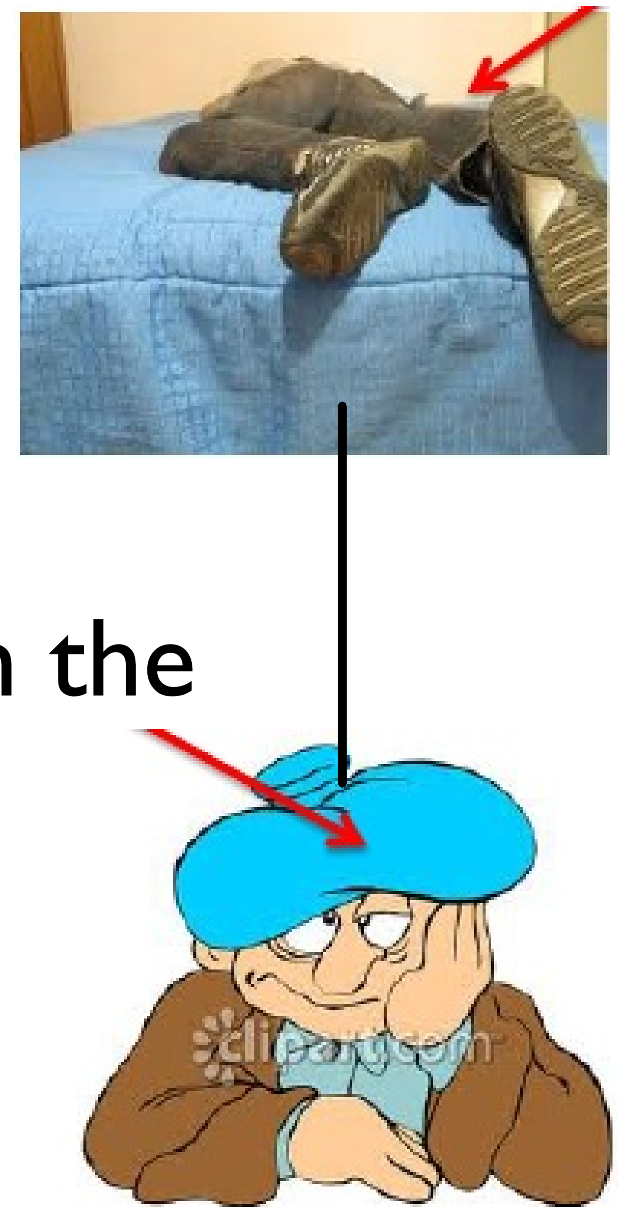
Correlation vs. causation 1

- The more firemen fighting a fire, the bigger the fire.
- Therefore firemen cause an increase in the size of a fire.



- Causation reversal. Correlation cannot distinguish between **A causes B** and **B causes A**

Dependence:
Sleeping with
shoes on is
correlated with
having a headache in the
morning

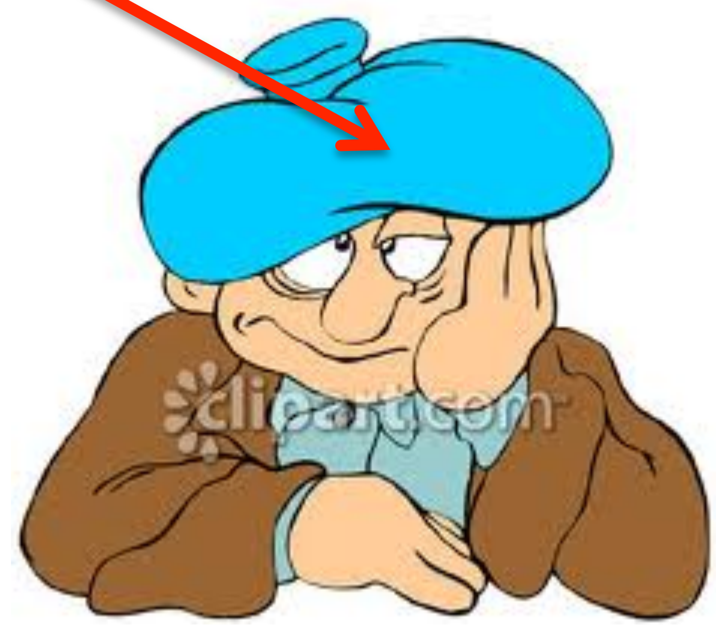


Correlation vs. causation 2

Excessive Drinking

Common Cause

Sleeping with shoes on



Morning Headache

Correlation vs. Causation 3

- For an ideal gas in a fixed volume, temperature is correlated with pressure.
- Gas, volume and temperature are related by the equation PV=nRT.
- Pressure and Temperature or co-dependent.
- Causation is bi-directional or not well defined.

Determining causation

- Can be very hard.
- Usually required intervention
- How can you do determine whether or not sleeping with shoes causes headaches?
  - A. Stop drinking.
  - B. Flip a coin to decide whether to wear shoes to bed.
  - C. Flip a coin to decide whether or not to drink.
  - D. Observe that every time you drank, you both slept with shoes and got up with a headache.

What is done in practice?

- Given random variables $X_1,\ldots,X_n$ and their joint distribution, we want to identify causal relationships. (For example, the causes for a particular disease).
- We perform a correlation analysis, computing the correlation for each pair Xi,Xj
- Sometimes, we know the causation direction, for example, a mutation in DNA causes a change in the protein and not vice versa.
- We pick the pairs with strongest correlations and use additional experiments to identify the causes.