Introduction to Probability

Lecture 8

Today

Poisson example

Sums of random variables

Covariance

Function of random variables (discrete)

Attendance: 61475838

Summary

Bernoulli

$$P(x|p) = p^x(1-p)^{1-x} \qquad x = 0,1$$

Binomial

$$P(k|N,p) = \binom{N}{k} p^k(1-p)^{N-k} \quad k = 0,1 \dots N$$

Poisson

$$P(k|\lambda) \equiv \frac{\lambda^k}{k!}e^{-\lambda} \quad k = 0,1 \dots$$

Multivariate Distributions

Multivariate Distributions

One variable

$$P(x)$$

$$\langle x \rangle \equiv \sum_{x} x\, P(x)$$

More than one variable

$$P(x, y)$$

$$\langle x \rangle \equiv \sum_{x} \sum_{y} x\, P(x, y) = \sum_{x} x\, P(x)$$

Example

What is $\langle x \rangle$ for the following?

$$P(x=0, y=0) = 0.1$$
$$P(x=1, y=0) = 0.1$$
$$P(x=0, y=1) = 0.4$$
$$P(x=1, y=1) = 0.4$$

$$\langle x \rangle \equiv \sum_x \sum_y x\, P(x,y) = \sum_x x\, P(x)$$

So

$$\langle x \rangle = 0 \times P(0,0) + 0 \times P(0,1) + 1 \times P(1,0) + 1 \times P(1,1)$$
$$= 1 \times 0.1 + 1 \times 0.4 = 0.5$$

Or

$$P(x=0) = P(0,0) + P(0,1) = 0.5$$
$$P(x=1) = P(1,0) + P(1,1) = 0.5$$

$$\rightarrow \langle x \rangle = 0 \times 0.5 + 1 \times 0.5 = 0.5$$

Sums of Random Variables

Notation

We say x was drawn according $\mathrm{P}(x)$ by writing

$$x \sim P(x)$$

This tells us the distribution that x follows.

Example: binomial, Poisson.

Sum of Random Variables

Consider N variables all drawn according to $P(x)$

$$x_1 \sim P(x); x_2 \sim P(x) \ldots$$

Define $t = x_1 + x_2 + \cdots x_N$

What is $P(t)$? Generally this is too difficult to calculate.

Can we calculate $\langle t \rangle$ or $\mathrm{var}(t)$?

Note:

$$\bar{x} = \frac{1}{N}(x_1 + x_2 + \cdots x_N) = \frac{t}{N}$$

Expectation Value

$$\langle t \rangle = \langle x_1 + x_2 + \cdots x_N \rangle$$

We already have linearity of expectation, so

$$\langle x_1 + x_2 + \cdots x_N \rangle = \langle x_1 \rangle + \langle x_2 \rangle + \cdots \langle x_N \rangle$$

Or we can do it explicitly. Consider $N = 2$

$$\langle x_1 + x_2 \rangle = \sum_{x_1, x_2} (x_1 + x_2) P(x_1, x_2)$$

$$= \sum_{x_1, x_2} (x_1) P(x_1, x_2) + \sum_{x_1, x_2} (x_2) P(x_1, x_2)$$

$$= \langle x_1 \rangle + \langle x_2 \rangle$$

In general

$$\left\langle \sum_n x_n \right\rangle = \sum_n \langle x_n \rangle$$

Variance

Let's use $N = 2$

$$t = x_1 + x_2; \;\; \mathrm{var}(t) = \langle t^2 \rangle - \langle t \rangle^2; \;\; \langle t \rangle = \langle x_1 \rangle + \langle x_2 \rangle$$

$$var(t) = \langle (x_1 + x_2)^2 \rangle - \langle (x_1 + x_2) \rangle^2$$

$$= \langle x_1^2 \rangle + \langle x_2^2 \rangle + 2\langle x_1 x_2 \rangle - \langle x_1 \rangle^2 - \langle x_2 \rangle^2 - 2\langle x_1 \rangle \langle x_2 \rangle$$

Re-ordering

$$= \langle x_1^2 \rangle - \langle x_1 \rangle^2 + \langle x_2^2 \rangle - \langle x_2 \rangle^2 + 2\langle x_1 x_2 \rangle - 2\langle x_1 \rangle \langle x_2 \rangle$$

$$= \mathrm{var}(x_1) + \mathrm{var}(x_2) + 2\mathrm{cov}(x_1, x_2)$$

Covariance

$$\mathrm{var}(x) \equiv \sum_x (x - \langle x \rangle)^2 \, P(x) = \langle x^2 \rangle - \langle x \rangle^2$$

$$\mathrm{cov}(x, y) \equiv \sum_{xy} (x - \langle x \rangle)(y - \langle y \rangle) P(x, y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle$$

$$= \langle xy \rangle - 2\langle x \rangle \langle y \rangle + \langle x \rangle \langle y \rangle$$
$$\langle xy \rangle - \langle x \rangle \langle y \rangle$$

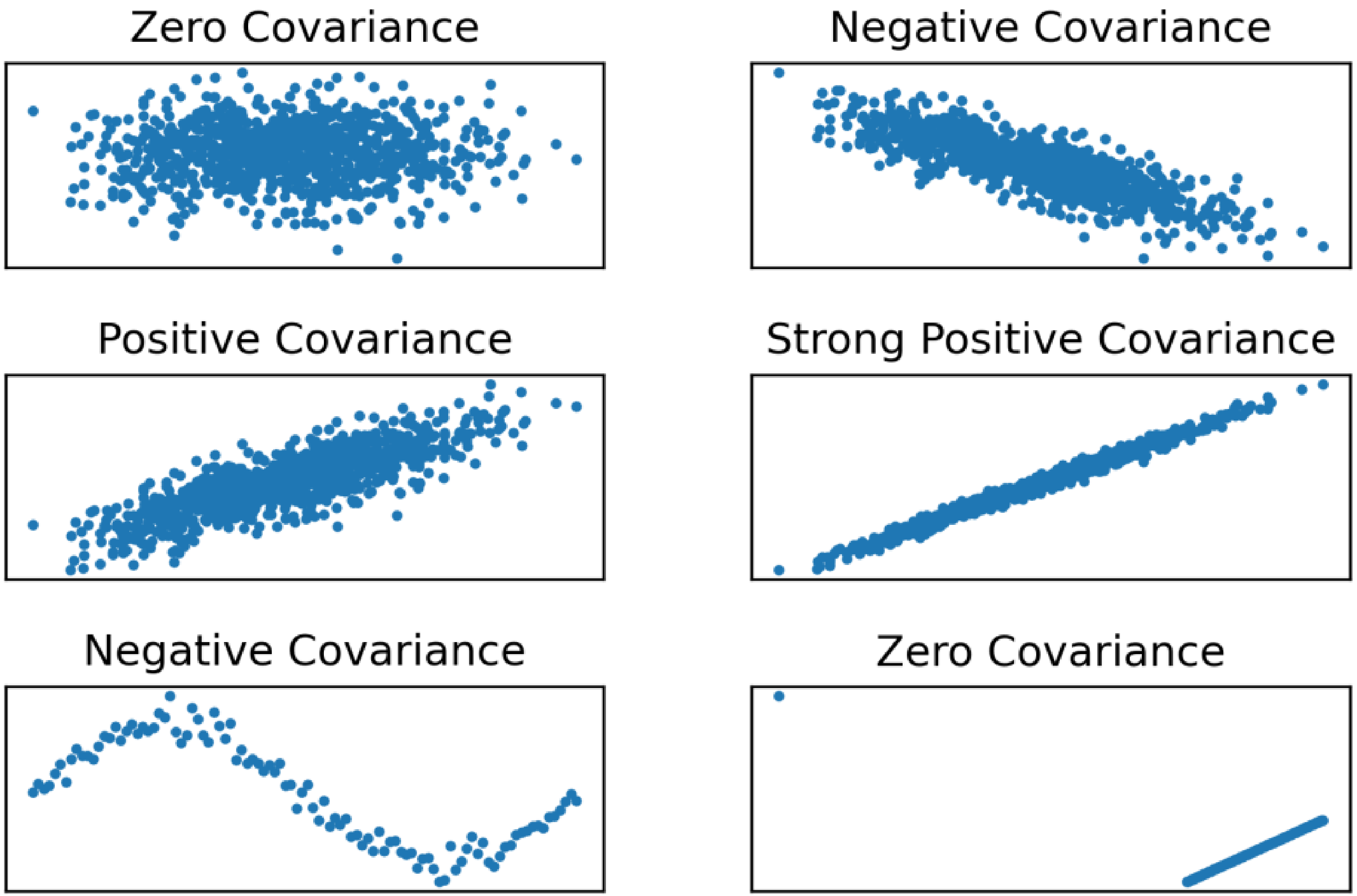
Note

$$\mathrm{var}(x) = \mathrm{cov}(x, x)$$

Covariance (2)

Covariance measures **linear association**

$$\mathrm{corr}(x,y) = \frac{\mathrm{cov}(x,y)}{\mathrm{std}(x)\mathrm{std}(y)}$$



Example

Calculate the covariance between x and y for the following distribution

$$P(x = 0, y = 0) = 0.2$$
$$P(x = 0, y = 1) = 0.2$$
$$P(x = 1, y = 0) = 0.2$$
$$P(x = 1, y = 1) = 0.4$$

$$\langle x \rangle = \sum_{xy} x\, P(x, y) = 0.6$$
$$\langle y \rangle = \sum_{xy} y\, P(x, y) = 0.6$$
$$\langle xy \rangle = \sum_{xy} xy\, P(x, y) = 0.4$$

$$\mathrm{cov}(x, y) = \langle xy \rangle - \langle x \rangle \langle y \rangle$$

$$= 0.4 - 0.6^2 = 0.04$$

Covariance and Independence

If x and y are **independent** then
$P(x,y) = P(x)P(y)$

$$\mathrm{cov}(x,y) = \langle xy\rangle - \langle x\rangle\langle y\rangle$$

$$\langle xy\rangle = \sum_{xy} xy\; P(x,y)$$
$$= \sum_{xy} xy\; P(x)P(y)$$

$$= \sum_{x} x\; P(x) \sum_{y} y\; P(y)$$
$$= \langle x\rangle\langle y\rangle$$

So

$$\mathrm{cov}(x,y) \to 0$$

Variance of Sum

If the variables are independent, then the following rule holds:

The variance of the sum is the sum of the variances

Otherwise we must include the covariance.

$$\text{var}(x_1 + x_2 + \cdots x_N) = \sum_n \text{var}(x_n) + 2 \sum_{m>n} \text{cov}(x_n, x_m)$$

Example

What is the expectation value and variance of the sum of N independent Bernoulli variables, each with parameter p.

For a Bernoulli

$$\langle x \rangle = p; \quad \mathrm{var}(x) = p(1-p)$$

Then

$$k = x_1 + x_2 + \cdots x_N$$

$$\rightarrow \langle k \rangle = \langle x_1 + x_2 + \cdots x_N \rangle = Np$$

$$\rightarrow \mathrm{var}(k) = Np(1-p)$$

Change of Variables

Change of Variables (Discrete)

Imagine we have some $P_x(x)$.

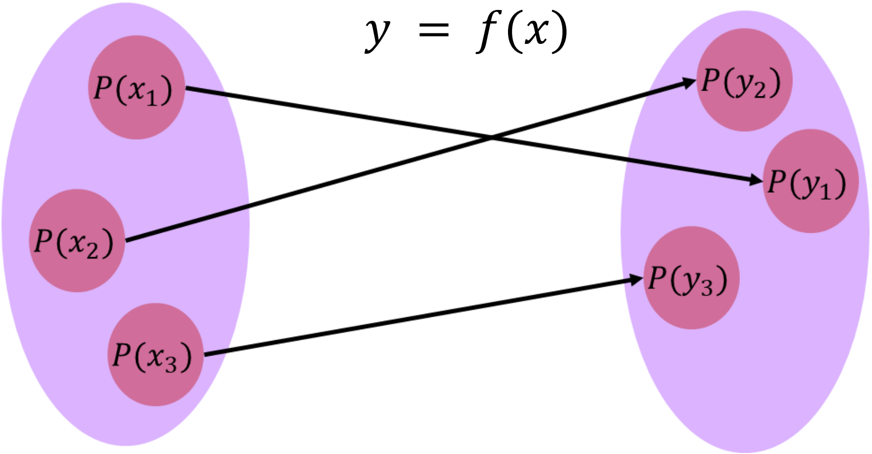
The sample space of x is Ω_x.

We make a function of x like $\mathrm{y} = f(x)$

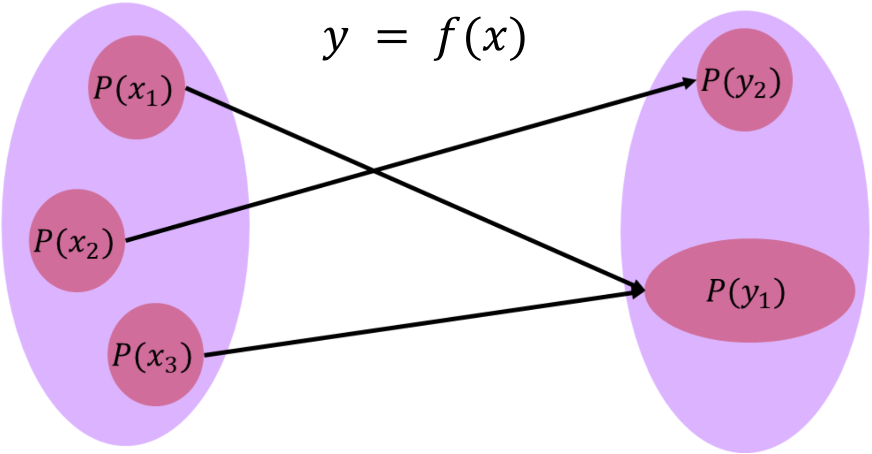
What is $P_y(y)$?

Or Ω_y?

Types of function



Paired (bijection)



Unpaired

Unique $x \rightarrow$ Unique y

Same x goes onto y

Example

Consider a fair six-sided die

$$P_x(x) = \frac{1}{6} \quad x = 1,2,3,4,5,6$$

1. If $y = x - 2$, what is Ω_y and $P_y(y)$

2. If $\mathrm{z} = |x - 2|$, what is Ω_z and $P_z(z)$

x	1	2	3	4	5	6
y	-1	0	1	2	3	4
z	1	0	1	2	3	4

$$\Omega_y = \{-1,0,1,2,3,4\}$$

So

$$P_y(y) = \frac{1}{6} \quad \text{for } y \in \Omega_y$$

$$\Omega_z = \{0,1,2,3,4\}$$

So

$$P_z(z) = \frac{1}{6} \quad z = \{0,2,3,4\}; P_z(z) = \frac{2}{6} \quad z = 1$$

General Formula

The general formula for $P_y(y)$ if we have $y = f(x)$

$$P_y(y) = \sum_{x:y=f(x)} P_x(x)$$

And Ω_y is the unique set of values that come $y = f(x)$ for all x in Ω_x.

Summary

Expectation value of sum

$$\langle x_1 + x_2 + \cdots x_N \rangle = \langle x_1 \rangle + \langle x_2 \rangle + \cdots \langle x_N \rangle$$

Variance of sum

$$\mathrm{var}(x_1 + x_2 + \cdots x_N) = \sum_n \mathrm{var}\,(x_n) + 2 \sum_{n>m} \mathrm{cov}\,(x_n, x_m)$$

Variance of sum (**independent)**

$$\mathrm{var}(x_1 + x_2 + \cdots x_N) = \sum_n \mathrm{var}\,(x_n)$$

Change of Variables

$$x \sim P(x), y = f(x) \rightarrow P(y) = \sum_{f(x)=y} P(x)$$

Examples

Class Example

Calculate the correlation between x and y for the following distribution

$$P(x=0,y=0)=0.2$$
$$P(x=0,y=1)=0.2$$
$$P(x=1,y=0)=0.2$$
$$P(x=1,y=1)=0.4$$

$$\mathrm{corr}(x,y)\equiv\frac{\mathrm{cov}(x,y)}{\mathrm{std}(x)\mathrm{std}(y)}$$

From earlier:

$$\langle x\rangle=0.4;\langle y\rangle=0.6;\ \mathrm{cov}(x,y)=0.04$$

$$\mathrm{var}(x)=0.4-0.4^2=0.24$$

$$\mathrm{var}(y)=0.6-0.6^2=0.24$$

$$\rightarrow \mathrm{corr}(x,y)=\frac{0.04}{0.24}=\frac{1}{7}$$

Class Example

Consider a fair six-sided die

$$P_x(x) = \frac{1}{6} \quad x = 1,2,3,4,5,6$$

If $y = |x - 3|$, what is Ω_y and $P_y(y)$ and $\langle y \rangle$?

x	1	2	3	4	5	6
y	2	1	0	1	2	3

$$\Omega_y = \{0,1,2,3\}$$

So

$$P_y(y) = \begin{cases} \frac{1}{6} & y = 0,3 \\ \frac{1}{3} & y = 1,2 \end{cases}$$

$$\langle y \rangle = \frac{1}{6} \times 0 + \frac{1}{3} \times 1 + \frac{1}{3} \times 2 + \frac{1}{6} \times 3 = \frac{3}{2}$$