



RANDOM VARIABLES

CS6140

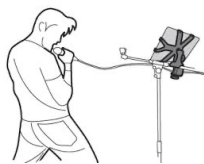
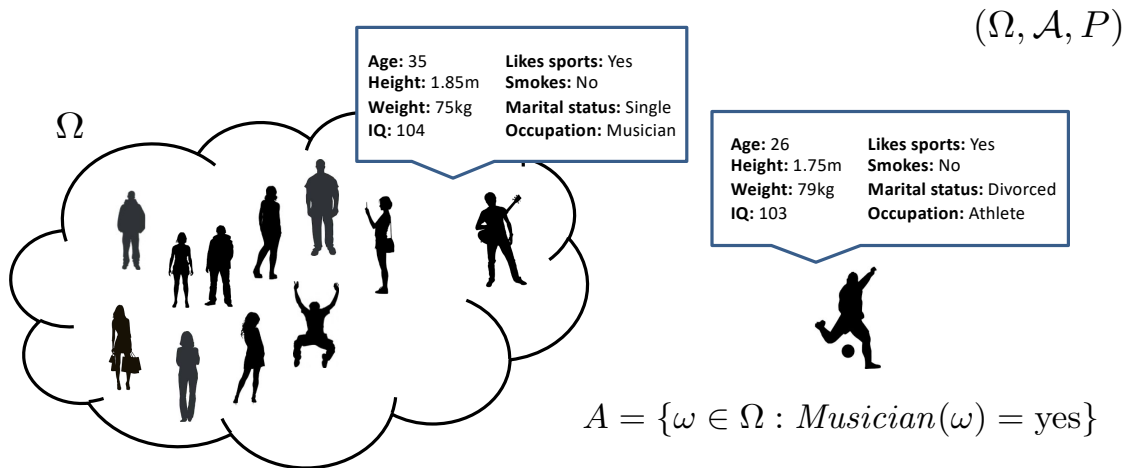
Predrag Radivojac

KHOURY COLLEGE OF COMPUTER SCIENCES

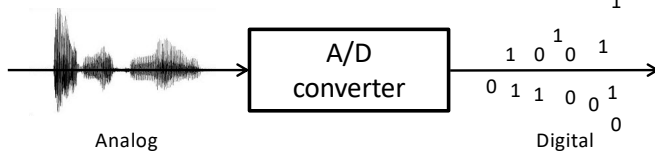
NORTHEASTERN UNIVERSITY

Fall 2026

RANDOM VARIABLES



$\Omega = \text{voltage at any time } t$



EXAMPLE: RANDOM VARIABLES

Experiment: three consecutive (fair) coin tosses

X = the number of heads in the first toss

Y = the number of heads in all three tosses

Find the probability spaces after the transformations.



What is the probability space $(\Omega, \mathcal{A}, P)$?

Where does the randomness come from?

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

$$\mathcal{A} = \mathcal{P}(\Omega)$$

$$P = ?$$

$$P(\Omega) = 1$$

$$P(\{HHH, TTT\}) = \frac{2}{8}$$

$$\vdots$$

EXAMPLE: RANDOM VARIABLES

$$X : \Omega \rightarrow \{0, 1\}$$

$$Y : \Omega \rightarrow \{0, 1, 2, 3\}$$



ω	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(\omega)$	1	1	1	1	0	0	0	0
$Y(\omega)$	3	2	2	1	2	1	1	0

What are the probability spaces $(\Omega_X, \mathcal{A}_X, P_X)$ and $(\Omega_Y, \mathcal{A}_Y, P_Y)$?

Where does the randomness come from?

RANDOM VARIABLE: FORMAL DEFINITION

$(\Omega, \mathcal{A}, P)$ = a probability space

Random variable:

1. $X : \Omega \rightarrow \Omega_X$
2. $\forall A \in \mathcal{B}(\Omega_X)$ it holds that $\{\omega : X(\omega) \in A\} \in \mathcal{A}$

It follows that:

$$P_X(A) = P(\{\omega : X(\omega) \in A\})$$

DISCRETE RANDOM VARIABLE

$(\Omega, \mathcal{A}, P)$ = a discrete probability space

Probability mass function (pmf):

$$\begin{aligned} p_X(x) &= P_X(\{x\}) \\ &= P(\{\omega : X(\omega) = x\}) \\ &= P(X = x) \end{aligned} \quad \forall x \in \Omega_X$$

The probability of an event A :

$$P_X(A) = \sum_{x \in A} p_X(x)$$

$$P(\{\omega : X(\omega) \in A\})$$
$$\forall A \in \mathcal{B}(\Omega_X)$$

EXERCISE: QUANTIZATION

$(\Omega, \mathcal{A}, P)$ = probability space

$\Omega = [-1, 1]$, $\mathcal{A} = \mathcal{B}(\Omega)$, P induced by uniform density

$X : \Omega \rightarrow \{0, 1\}$ such that

$$X(\omega) = \begin{cases} 0 & \omega^2 < 0.25 \\ 1 & \omega^2 \geq 0.25 \end{cases}$$

Question: Find $(\Omega_X, \mathcal{A}_X, P_X)$

CONTINUOUS RANDOM VARIABLE

Cumulative distribution function (cdf):

$$\begin{aligned}F_X(t) &= P_X(\{x : x \leq t\}) \\&= P_X((-\infty, t]) \\&= P(\{\omega : X(\omega) \leq t\}) \\&= P(X \leq t)\end{aligned}$$

Probability density function (pdf), if it exists:

$$p_X(x) = \left. \frac{dF_X(t)}{dt} \right|_{t=x}$$

CONTINUOUS RANDOM VARIABLE

If the probability density function (pdf) exists:

$$F_X(t) = \int_{-\infty}^t p_X(x) dx$$

The probability of an event $A = (a, b]$:

$$\begin{aligned} P_X((a, b]) &= \int_a^b p_X(x) dx \\ &= F_X(b) - F_X(a) \end{aligned}$$

$$P(a < X \leq b)$$



JOINT AND MARGINAL DISTRIBUTIONS

$(\Omega, \mathcal{A}, P)$ = a discrete probability space

Joint probability distribution:

$$\begin{aligned} p_{XY}(x, y) &= P(\{\omega : X(\omega) = x\} \cap \{\omega : Y(\omega) = y\}) \\ &= P(X = x, Y = y) \end{aligned}$$

Extend to d -D vector $\mathbf{X} = (X_1, X_2, \dots, X_d)$

Marginal probability distribution:

$$p_{X_i}(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_d} p_{\mathbf{X}}(x_1, \dots, x_d)$$

EXERCISE: JOINT DISTRIBUTIONS

Example: X = the number of heads in the first toss
 Y = the number of heads in all three tosses

		Y			
		0	1	2	3
X	0				
	1				



JOINT AND MARGINAL DISTRIBUTIONS

$(\Omega, \mathcal{A}, P) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R})^d, P_{\mathbf{X}})$ = a continuous probability space

Joint probability distribution:

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{t}) &= P_{\mathbf{X}}(\{\mathbf{x} : x_i \leq t_i, i = 1 \dots d\}) \\ &= P(X_1 \leq t_1, X_2 \leq t_2 \dots) \end{aligned}$$

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^d}{\partial t_1 \dots \partial t_d} F_{\mathbf{X}}(t_1, \dots, t_d) \Big|_{\mathbf{t}=\mathbf{x}} \quad (\text{if it exists})$$

Marginal probability distribution:

$$p_{X_i}(x_i) = \int_{x_1} \dots \int_{x_{i-1}} \int_{x_{i+1}} \dots \int_{x_d} p_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_d$$

CONDITIONAL DISTRIBUTIONS

Conditional probability distribution:

$$p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}$$

The probability of an event A , given that $X = x$, is:

$$P_{Y|X}(Y \in A|X = x) = \begin{cases} \sum_{y \in A} p_{Y|X}(y|x) & Y : \text{discrete} \\ \int_{y \in A} p_{Y|X}(y|x) dy & Y : \text{continuous} \end{cases}$$

CHAIN RULE

Conditional probability distribution:

$$p(x_d|x_1, \dots, x_{d-1}) = \frac{p(x_1, \dots, x_d)}{p(x_1, \dots, x_{d-1})}$$

This leads to:

$$p(x_1, \dots, x_d) = p(x_1) \prod_{l=2}^d p(x_l|x_1, \dots, x_{l-1})$$

INDEPENDENCE OF RANDOM VARIABLES

X and Y are **independent** if:

$$p_{XY}(x, y) = p_X(x) \cdot p_Y(y)$$

X and Y are **conditionally independent** given Z if:

$$p_{XY|Z}(x, y|z) = p_{X|Z}(x|z) \cdot p_{Y|Z}(y|z)$$

What if we had d random variables?

EXPECTATIONS

$(\Omega_X, \mathcal{B}(\Omega_X), P_X)$ = a probability space

Consider a function $f : \Omega_X \rightarrow \mathbb{C}$

$$\mathbb{E}[f(X)] = \begin{cases} \sum_{x \in \Omega_X} f(x)p_X(x) & X : \text{discrete} \\ \int_{\Omega_X} f(x)p_X(x)dx & X : \text{continuous} \end{cases}$$

WELL-KNOWN EXPECTATIONS

$f(x)$	Symbol	Name
x	$\mathbb{E}[X]$	Mean
$(x - \mathbb{E}[X])^2$	$V[X]$	Variance
x^k	$\mathbb{E}[X^k]$	k-th moment; $k \in \mathbb{N}$
$(x - \mathbb{E}[X])^k$	$\mathbb{E}[(x - \mathbb{E}[X])^k]$	k-th central moment; $k \in \mathbb{N}$
e^{tx}	$M_X(t)$	Moment generating function
e^{itx}	$\varphi_X(t)$	Characteristic function
$\log \frac{1}{p_X(x)}$	$H(X)$	(Differential) entropy
$\log \frac{p_X(x)}{q(x)}$	$D(p_X q)$	Kullback-Leibler divergence
$\left(\frac{\partial}{\partial \theta} \log p_X(x \theta)\right)^2$	$\mathcal{I}(\theta)$	Fisher information

CONDITIONAL EXPECTATIONS

Consider a function $f : \Omega_Y \rightarrow \mathbb{C}$

$$\mathbb{E}[f(Y)|x] = \begin{cases} \sum_{y \in \Omega_Y} f(y)p_{Y|X}(y|x) & Y : \text{discrete} \\ \int_{\Omega_Y} f(y)p_{Y|X}(y|x)dy & Y : \text{continuous} \end{cases}$$

$$\mathbb{E}[Y|x] = \sum yp_{Y|X}(y|x)$$

$$\mathbb{E}[Y|x] = \int yp_{Y|X}(y|x)dy$$

 Regression function!

EXERCISE: EXPECTATIONS

Example: X = the number of heads in the first toss
 Y = the number of heads in all three tosses



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

$$\mathbb{E}[X] =$$

$$\mathbb{E}[Y|X = 0] =$$

EXPECTATIONS FOR TWO VARIABLES

Consider a function $f : \mathbb{R}^2 \rightarrow \mathbb{C}$

$$\mathbb{E}[f(X, Y)] = \begin{cases} \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} f(x, y) p_{XY}(x, y) & X, Y : \text{discrete} \\ \int_{\Omega_X} \int_{\Omega_Y} f(x, y) p_{XY}(x, y) dx dy & X, Y : \text{continuous} \end{cases}$$

WELL-KNOWN EXPECTATIONS

$f(x, y)$	Symbol	Name
$(x - \mathbb{E}[X])(y - \mathbb{E}[Y])$	$\text{Cov}[X, Y]$	Covariance
$\frac{(x - \mathbb{E}[X])(y - \mathbb{E}[Y])}{\sqrt{V[X]V[Y]}}$	$\text{Corr}[X, Y]$	Correlation
$\log \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)}$	$I(X; Y)$	Mutual information
$\log \frac{1}{p_{XY}(x, y)}$	$H(X, Y)$	Joint entropy
$\log \frac{1}{p_{X Y}(x y)}$	$H(X Y)$	Conditional entropy

MIXTURES OF DISTRIBUTIONS

Mixture model:

A set of m probability distributions, $\{p_i(x)\}_{i=1}^m$

$$p(x) = \sum_{i=1}^m w_i p_i(x)$$

where $\mathbf{w} = (w_1, w_2, \dots, w_m)$ are non-negative and

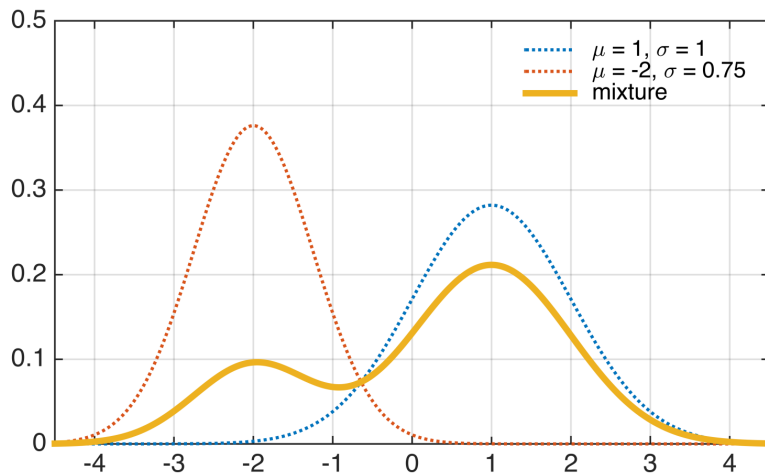
$$\sum_{i=1}^m w_i = 1$$

MIXTURES OF GAUSSIANS

Mixture of $m = 2$ Gaussian distributions:

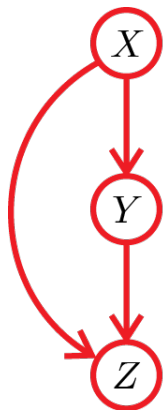
$$w_1 = 0.75, w_2 = 0.25$$

$$p(x) = \sum_{i=1}^m w_i p_i(x)$$



GRAPHICAL REPRESENTATIONS

Bayesian Network:
$$p(\mathbf{x}) = \prod_{i=1}^d p(x_i | \mathbf{x}_{\text{Parents}(X_i)})$$



$P(X = 1)$
0.3

X	$P(Y = 1 X)$
0	0.5
1	0.9

X	Y	$P(Z = 1 X, Y)$
0	0	0.3
0	1	0.1
1	0	0.7
1	1	0.4

Factorization:

$$p(x, y, z) = p(x)p(y|x)p(z|x, y)$$

GRAPHICAL REPRESENTATIONS

Bayesian Network:
$$p(\mathbf{x}) = \prod_{i=1}^d p(x_i | \mathbf{x}_{\text{Parents}(X_i)})$$



$P(X = 1)$
0.3

Y	$P(Z = 1 Y)$
0	0.2
1	0.7

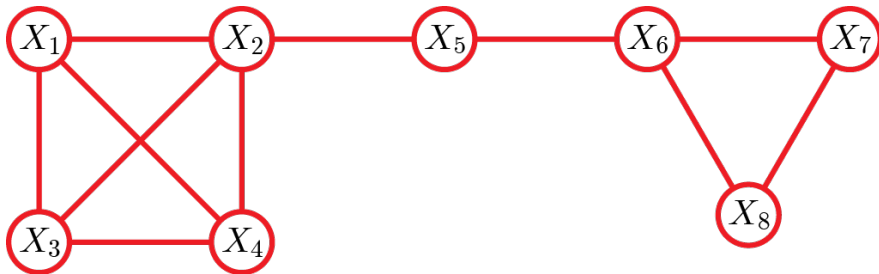
X	$P(Y = 1 X)$
0	0.5
1	0.9

Factorization:

$$p(x, y, z) = p(x)p(y|x)p(z|y)$$

GRAPHICAL REPRESENTATIONS

Markov Network: $p(x_i | \mathbf{x}_{-i}) = p(x_i | \mathbf{x}_{N(X_i)})$



Factorization:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(\mathbf{x}_C)$$