Random Variables

Short Review

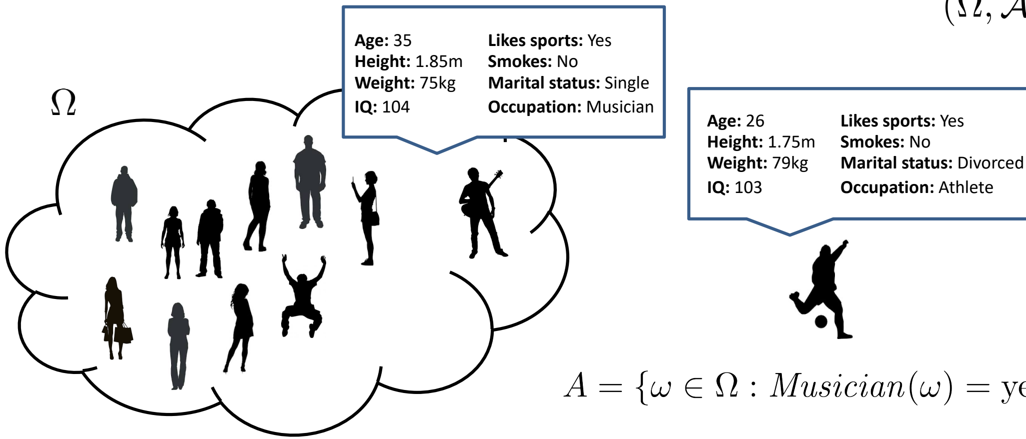
CS6140

Predrag Radivojac
Khoury College of Computer Sciences
Northeastern University

Spring 2021

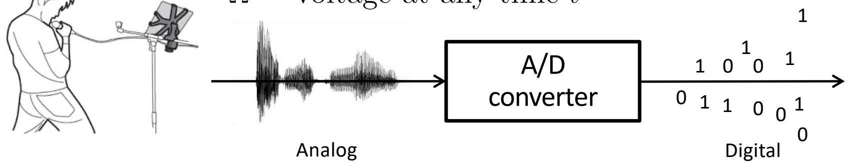
RANDOM VARIABLES

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



$$A = \{\omega \in \Omega : Musician(\omega) = \text{yes}\}$$

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



Pictures from the Internet.

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?

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

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

$$P = \,?$$

$$P(\Omega) = 1$$
$$P(\{\text{HHH}, \text{TTT}\}) = \tfrac{2}{8}$$
$$\vdots$$

Picture from the Internet.

Example: Random Variables

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

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?

Picture from the Internet.

Random Variable: Formal Definition

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

Random variable:

1. $X : \Omega \to \Omega_X$
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):

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

The probability of an event A:

$$P_X(A) = \sum_{x \in A} p_X(x) \qquad \forall A \in \mathcal{B}(\Omega_X)$$

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

Exercise: Quantization

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

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

$X : \Omega \to \{0, 1\}$ 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):

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

Probability density function (pdf), if it exists:

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

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 $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$

Marginal probability distribution:

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

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				

Picture from the Internet.

Joint and Marginal Distributions

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

Joint probability distribution:

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

$$p_{\boldsymbol{X}}\left(\boldsymbol{x}\right) = \left.\frac{\partial^d}{\partial t_1 \cdots \partial t_d} F_{\boldsymbol{X}}\left(t_1, \ldots t_d\right)\right|_{\boldsymbol{t}=\boldsymbol{x}} \qquad \text{(if it exists)}$$

Marginal probability distribution:

$$p_{X_i}\left(x_i\right) = \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_d} p_{\boldsymbol{X}}\left(\boldsymbol{x}\right) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d$$

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:

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

This leads to:

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

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

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

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

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

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 $f : \Omega_Y \to \mathbb{C}$

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

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

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

$\Leftarrow$ Regression function!

Exercise: Expectations

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

$$
\begin{array}{cc|c|c|c|c|}
 & & \multicolumn{4}{c}{Y} \\
 & & 0 & 1 & 2 & 3 \\
\hline
X & 0 & 1/8 & 1/4 & 1/8 & 0 \\
\hline
 & 1 & 0 & 1/8 & 1/4 & 1/8 \\
\hline
\end{array}
$$

$\mathbb{E}[X] =$

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

Picture from the Internet.

Expectations for Two Variables

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

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

WELL-KNOWN EXPECTATIONS

$f(x,y)$	Symbol	Name
$(x-\mathbb{E}[X])(y-\mathbb{E}[Y])$	$\mathrm{Cov}[X,Y]$	Covariance
$\frac{(x-\mathbb{E}[X])(y-\mathbb{E}[Y])}{\sqrt{V[X]V[Y]}}$	$\mathrm{Corr}[X,Y]$	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
$\log\frac{1}{p_{X\mid Y}(x\mid y)}$	$H(X\mid Y)$	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 $\boldsymbol{w} = (w_1, w_2, \ldots, w_m)$ are non-negative and

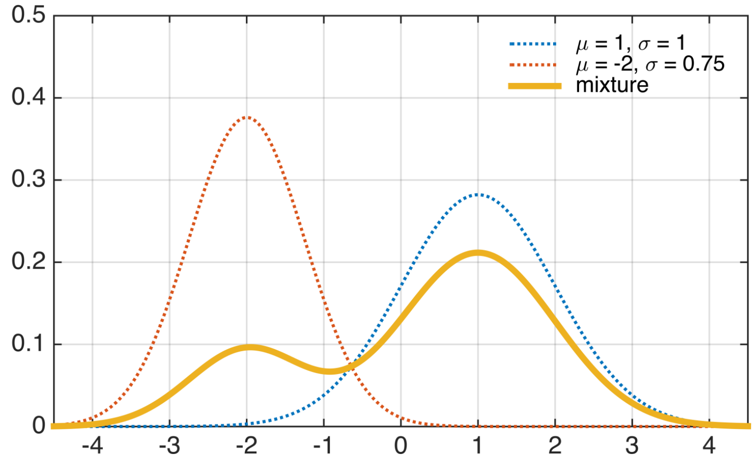
$$\sum_{i=1}^{m} w_i = 1$$

Mixtures of Gaussians

Mixture of $m = 2$ Gaussian distributions:

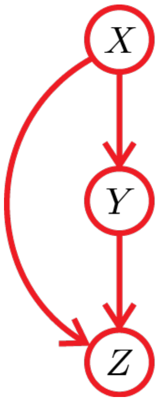
$w_1 = 0.75,\ w_2 = 0.25$

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



Graphical Representations

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



$P(X=1)$
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: $$p(\boldsymbol{x}) = \prod_{i=1}^{d} p\left(x_i | \boldsymbol{x}_{\text{Parents}(X_i)}\right)$$

$P(X=1)$
0.3

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

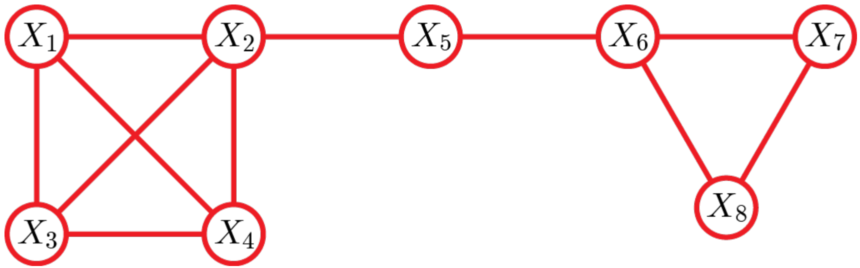
X	$P(Y=1 \mid X)$
0	0.5
1	0.9

Factorization:

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

Graphical Representations

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



Factorization:

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