Probabilistic Graphical Models

Review of probability theory

Siamak Ravanbakhsh

Fall 2019

Learning objectives

- Probability distribution and density functions
- Random variable
- Bayes' rule
- Conditional independence
- Expectation and Variance

Sample space Ω

$\Omega = \{\omega\}$: the **set** of all possible **outcomes** (*a.k.a.* outcome space)

Example1: three tosses of a coin $\Omega = \{hhh, hht, hth, \ldots, ttt\}$

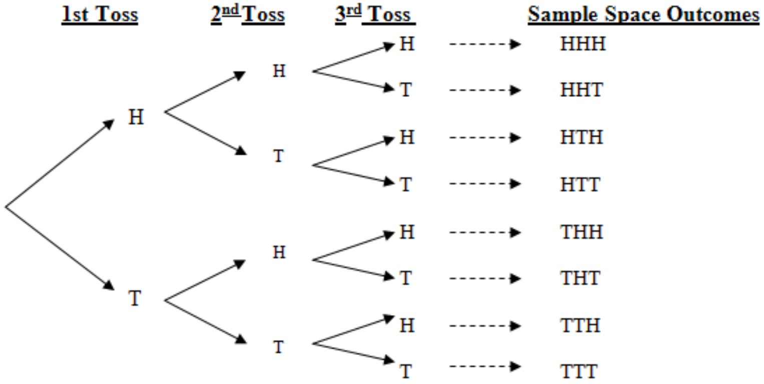


image: http://web.mnstate.edu/peil/MDEV102/U3/S25/Cartesian3.PNG

Sample space Ω

$\Omega = \{\omega\}$: the **set** of all possible **outcomes** (*a.k.a.* outcome space)

Example 2: two dice

$$\Omega = \{(1,1), \ldots, (6,6)\}$$

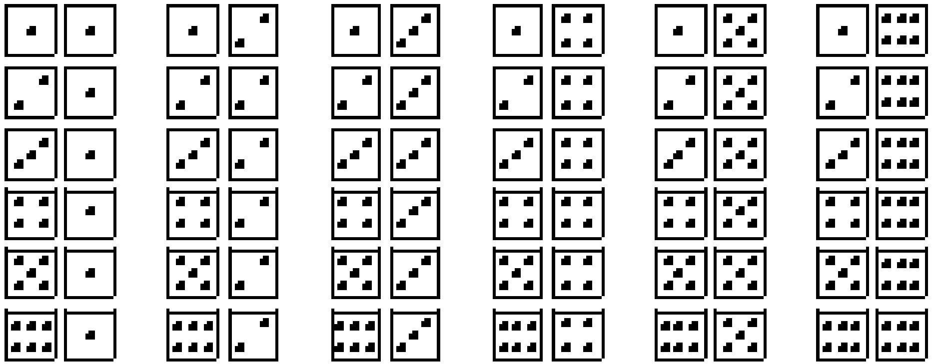


Image source: http://www.stat.ualberta.ca/people/schmu/preprints/article/Article.htm

Event space Σ

An event $E \subseteq \Omega$ is a set of outcomes

event space $\Sigma \subseteq 2^{\Omega}$ is a set of events

Event space Σ

An event $E \subseteq \Omega$ is a set of outcomes

event space $\Sigma \subseteq 2^{\Omega}$ is a set of events

Example:

Event: at least two heads $\Sigma = \{hht, thh, hth, hhh\}$

Event: draw a pair of aces from a deck

$|E| = 6$

Event space Σ

Requirements for event space $(\sigma - \text{algebra})$

- $\Omega \in \Sigma$
- *The complement of an event is also an event* $A \in \Sigma \rightarrow \Omega - A \in \Sigma$
- *(Countable) intersection of events is also an event*

$$A, B \in \Sigma \rightarrow A \cap B \in \Sigma$$

Example:

$$\text{at least one head, at least one tail} \in \Sigma \rightarrow \text{at least one head and one tail} \in \Sigma$$

$$\text{at least one head} \in \Sigma \rightarrow \text{no heads} \in \Sigma$$

Extends to uncountable sets (Real numbers)

Probability distribution

Assigns a real value to each event $P : \Sigma \to \mathbb{R}$

Probability axioms *(Kolmogorov axioms)*

other axiomatizations of probability?

- Probability is non-negative $P(A) \geq 0$
- The probability of disjoint events is (countably) additive

$$A \cap B = \emptyset \to P(A \cup B) = P(A) + P(B)$$

- $P(\Omega) = 1$

measure

The triple (Ω, Σ, P) is a probability space

Probability distribution

Probability axioms *(Kolmogorov axioms)*

- Probability is non-negative $P(A) \geq 0$
- disjoint events are additive: $A \cap B = \emptyset \rightarrow P(A \cup B) = P(A) + P(B)$
- $P(\Omega) = 1$

Derivatives:

- $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **union bound:** $P(A \cup B) \leq P(A) + P(B)$
- $P(\Omega \setminus A) = 1 - P(A)$
- $P(A \cap B) \leq \min\{P(A), P(B)\}$

Probability distribution: examples

$\Omega = \{1, 2, 3, 4, 5, 6\}$

$\Sigma = \{\emptyset, \Omega\}$ (a minimal choice of event space)

$P(\emptyset) = 0, P(\Omega) = 1$

Probability distribution: examples

$\Omega = \{1, 2, 3, 4, 5, 6\}$

$\Sigma = \{\emptyset, \Omega\}$ (a minimal choice of event space)

$P(\emptyset) = 0, P(\Omega) = 1$

$\Sigma = 2^\Omega$ (a maximal choice of event space)

$P(A) = \frac{|A|}{6}$ that is $P(\{1, 3\}) = \frac{2}{6}$ (any other *consistent* assignment is acceptable)


IS...IS THIS REALLY
NECESSARY?

Can't we always use 2^{Ω} even for uncountable outcome spaces?

Can't we always use 2^{Ω}
even for uncountable outcome spaces?

It turns out some events are not measurable

Banach-Tarski paradox

Can't we always use 2^Ω
even for uncountable outcome spaces?

It turns out some events are not measurable

Banach-Tarski paradox

Having a event space and probability measure avoids this

Conditional probability

Probability of an event A after observing the event B

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Conditional probability

Probability of an event A after observing the event B

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \longleftarrow P(B) > 0$$

Conditional probability

Probability of an event A after observing the event B

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \longleftarrow P(B) > 0$$

Example: three coin tosses

$$P(\text{at least one head} \mid \text{at least one tail}) = \frac{P(\text{at least one head and one tail})}{P(\text{at least one tail})}$$

Chain rule

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Chain rule

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$\downarrow$$

Chain rule: $$P(A \cap B) = P(B)P(A \mid B)$$

Chain rule

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$\downarrow$$

Chain rule: $P(A \cap B) = P(B)P(A \mid B)$ and $B = C \cap D$

Chain rule

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

↓

Chain rule: $P(A \cap B) = P(B)P(A \mid B)$ and $B = C \cap D$

↓

$$P(A \cap C \cap D) = P(C \cap D)P(A \mid C \cap D)$$

Chain rule

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

↓

Chain rule: $P(A \cap B) = P(B)P(A \mid B)$ and $B = C \cap D$

↓

$$P(A \cap C \cap D) = P(C \cap D)P(A \mid C \cap D)$$

↓

$$P(A \cap C \cap D) = P(D)P(C \mid D)P(A \mid C \cap D)$$

Chain rule

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

↓

Chain rule: $P(A \cap B) = P(B)P(A \mid B)$ and $B = C \cap D$

↓

$$P(A \cap C \cap D) = P(C \cap D)P(A \mid C \cap D)$$

↓

$$P(A \cap C \cap D) = P(D)P(C \mid D)P(A \mid C \cap D)$$

More generally: $P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2 \mid A_1)\ldots P(A_n \mid A_1 \cap \ldots \cap A_{n-1})$

Bayes' rule

Reasoning about event A:

likelihood of the event B if A were to happen

our **prior** belief about A

our **posterior** belief about A after observing B

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

Bayes' rule: example

- 1% of the population has cancer
- cancer test
  - False positive 10%
  - False negative 5%
- chance of **having cancer** given a **positive test** result?

posterior likelihood prior

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

Bayes' rule: example

- 1% of the population has cancer
- cancer test
  - False positive 10%
  - False negative 5%
- chance of **having cancer** given a **positive test** result?

posterior likelihood prior

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

- sample space? → {TP, TN, FP, FN}
- events A, B? → A = {TP, FN}, B = {TP, FP}
- prior? likelihood? → P(A) = .01, P(B|A) = .9
- P(B) is not trivial

Bayes' rule: example

- 1% of the population has cancer
- cancer test
  - False positive 10%
  - False negative 5%
- chance of **having cancer** given a **positive test** result?

posterior likelihood prior

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

- sample space? → {TP, TN, FP, FN}
- events A, B? → A = {TP, FN}, B = {TP, FP}
- prior? likelihood? → P(A) = .01, P(B|A) = .9
- P(B) is not trivial

$$P(cancer \mid +) \propto P(+ \mid cancer)P(cancer) = .009$$

$$P(\neg cancer \mid +) \propto P(+ \mid \neg cancer)P(\neg cancer) = .99 \times .1 = .099$$

→ $P(cancer \mid +) = \frac{.009}{.009+.099} \approx .08$

Independence

$$P \models (A \perp B)$$

$$\perp\!\!\!\perp$$

Events **A** and **B** are independent *iff*

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

Observing A does not change P(B)

Independence

$P \models (A \perp B)$ → $\perp\!\!\!\perp$

Events **A** and **B** are independent *iff*

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

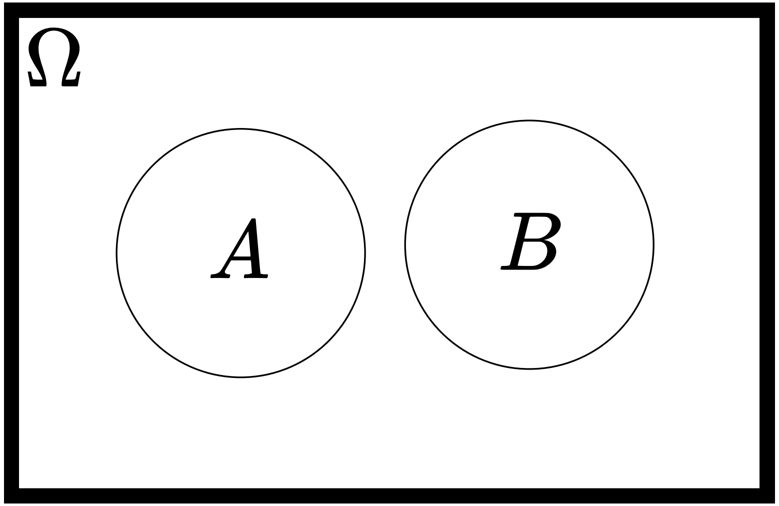
Observing A does not change P(B)

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

Equivalent definition: $P(B) = P(B \mid A)$ or $P(A) = 0$

Independence: example

Are A and B independent?



Independence: example

Example 1: $P(\mathrm{hhh}) = P(\mathrm{hht}) \ldots = P(\mathrm{ttt}) = \frac{1}{8}$

$$P(\mathrm{h} * * \mid * \mathrm{t} *) = P(\mathrm{h} * *) = \frac{1}{2}$$

equivalently: $P(\mathrm{h}\,\mathrm{t} *) = P(* \mathrm{t} *)P(\mathrm{h} * *) = \frac{1}{4}$

Independence: example

Example 1: $P(\mathrm{hhh}) = P(\mathrm{hht}) \ldots = P(\mathrm{ttt}) = \frac{1}{8}$

$$P(\mathrm{h} * * \mid * \mathrm{t} *) = P(\mathrm{h} * *) = \frac{1}{2}$$

equivalently: $P(\mathrm{h}\,\mathrm{t} *) = P(* \mathrm{t} *)P(\mathrm{h} * *) = \frac{1}{4}$

Example 2: are these two events independent?

$$P(\{ht, hh\}) = .3, P(\{th\}) = .1$$

Conditional independence $P \models (A \perp B \mid C)$

a more common phenomenon: $P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$

Conditional independence $P \models (A \perp B \mid C)$

a more common phenomenon: $P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$

using $P(A \cap B \mid C) = P(A \mid C)P(B \mid A \cap C)$

Conditional independence $P \models (A \perp B \mid C)$

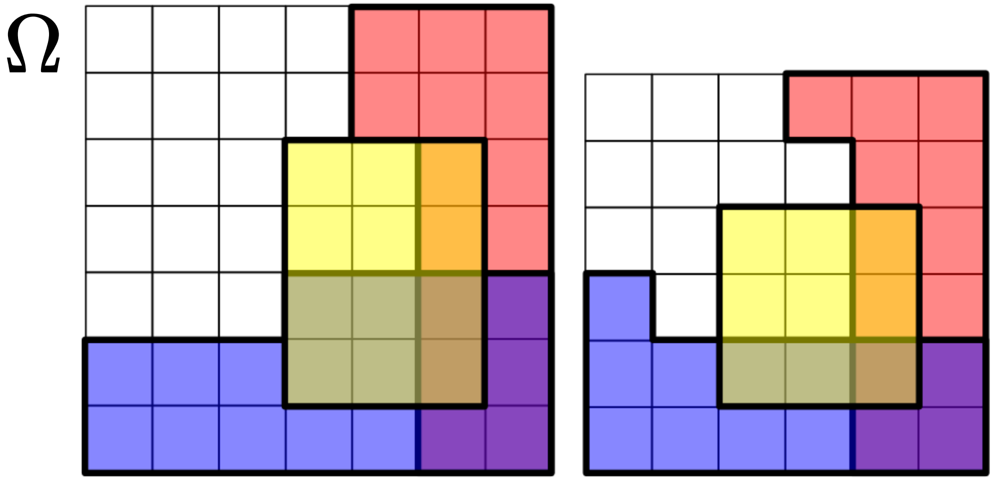
a more common phenomenon: $P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$

using $P(A \cap B \mid C) = P(A \mid C)P(B \mid A \cap C)$

Equivalent definition: $P(B \mid C) = P(B \mid A \cap C)$ or $P(A \cap C) = 0$

Conditional independence: example

Generalization of independence: $P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$



$P \models (R \perp B \mid Y)$

from: wikipedia

Summary

Basics of probability

- **Outcome space:** a set
- **Event:** a subset of outcomes
- **Event space:** a set of events
- **Probability dist.** is associated with events
- **Conditional probability:** based on intersection of events
- **Chain rule** follows from conditional probability
- **(Conditional) independence:** relevance of some events to others

Random Variable

is an **attribute** associated with each outcome $X : \Omega \rightarrow Val(X)$

- intensity of a pixel
- head/tail value of the first coin in multiple coin tosses
- first draw from a deck is larger than the second

a formalism to define **events** $P(X = x) \triangleq P(\{\omega \in \Omega \mid X(\omega) = x\})$

Random Variable

is an **attribute** associated with each outcome $X : \Omega \to Val(X)$

- intensity of a pixel
- head/tail value of the first coin in multiple coin tosses
- first draw from a deck is larger than the second

a formalism to define **events** $P(X = x) \triangleq P(\{\omega \in \Omega \mid X(\omega) = x\})$

Example: *three tosses of coin*

- *number of heads* $X_1 : \Omega \to \{0, 1, 2, 3\}$
- *number of heads in the first two trials* $X_2 : \Omega \to \{0, 1, 2\}$
- *at least one head* $X_3 : \Omega \to \{True, False\}$

Random Variable (RV)

is an **attribute** associated with each outcome $X : \Omega \rightarrow Val(X)$

a formalism to define **events** $P(X = x) \triangleq P(\{\omega \in \Omega \mid X(\omega) = x\})$

Multiple RVs: $X_1, \ldots, X_n$

- *outcomes that we care about:* $X_1 = x_1, \ldots, X_n = x_n$
- *cannonical outcome space:* $\Omega_c \triangleq Val(X_1) \times \ldots \times Val(X_n)$

Random Variable (RV)

is an **attribute** associated with each outcome $X : \Omega \rightarrow Val(X)$

a formalism to define **events** $P(X = x) \triangleq P(\{\omega \in \Omega \mid X(\omega) = x\})$

Multiple RVs: $X_1, \ldots, X_n$

- *outcomes that we care about:* $X_1 = x_1, \ldots, X_n = x_n$
- *cannonical outcome space:* $\Omega_c \triangleq Val(X_1) \times \ldots \times Val(X_n)$
- *joint probability:* $P(X_1 = x_1, \ldots, X_n = x_n) \triangleq P(X_1 = x_1 \cap \ldots \cap X_n = x_n)$

Random Variable (RV)

is an **attribute** associated with each outcome $X : \Omega \rightarrow Val(X)$

a formalism to define **events** $P(X = x) \triangleq P(\{\omega \in \Omega \mid X(\omega) = x\})$

Multiple RVs: $X_1, \ldots, X_n$

- *outcomes that we care about:* $X_1 = x_1, \ldots, X_n = x_n$
- *cannonical outcome space:* $\Omega_c \triangleq Val(X_1) \times \ldots \times Val(X_n)$
- *joint probability:* $P(X_1 = x_1, \ldots, X_n = x_n) \triangleq P(X_1 = x_1 \cap \ldots \cap X_n = x_n)$
- *marginal probability:* $P(X_1 = x_1) = \sum_{x_2, \ldots, x_n} P(X_1 = x_1, \ldots, X_n = x_n)$

Random Variable: example

three tosses of coin

number of heads $X_1 : \Omega \to \{0, 1, 2, 3\}$

first trial is a head $X_2 : \Omega \to \{True, False\}$

cannonical **outcome space**: $\Omega_c = \{(0, True), \ldots, (3, False)\}$

atomic outcome

a **joint** probability

marginal probability

	0	**1**	**2**	**3**	**P(X2)**
True	.1	.1	.4	.05	.65
False	.2	.01	.09	.05	.35
P(X1)	.3	.11	.49	.1	

Conditional independence *for RVs*

Given random variables X, Y, Z $P \models (X \perp Y \mid Z)$ *iff*

$$P \models (X = x \perp Y = y \mid Z = z) \quad \forall x, y, z$$

Therefore $P \models (X \perp Y \mid Z)$ *iff* $P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$

OR

$$P(X \mid Y, Z) = P(X \mid Z)$$

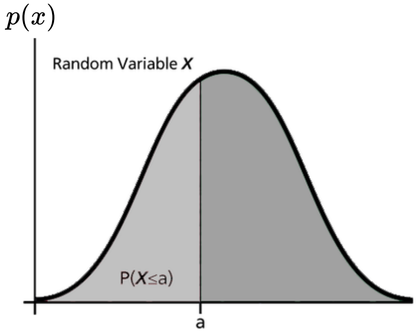
Marginal independence: $P \models (X \perp Y \mid \emptyset)$

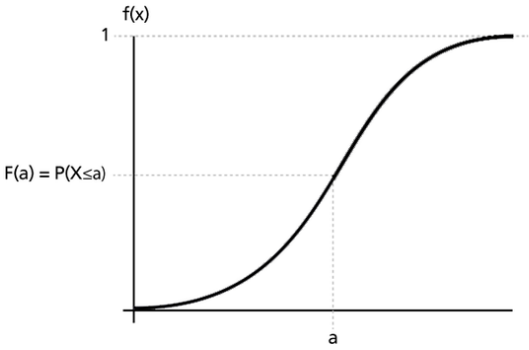
Continuous domain

probability **density** function (pdf) $p : Val(X) \to [0, +\infty) \quad s.t. \quad \int_{Val(X)} p(x)\mathrm{d}x = 1$

$$P(X \leq a) \triangleq \underline{\int_{-\infty}^{a} p(x)\mathrm{d}x}$$

$F(a)$: the cumulative distribution function (cdf)



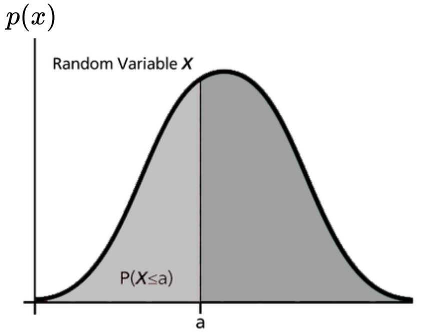


Continuous domain

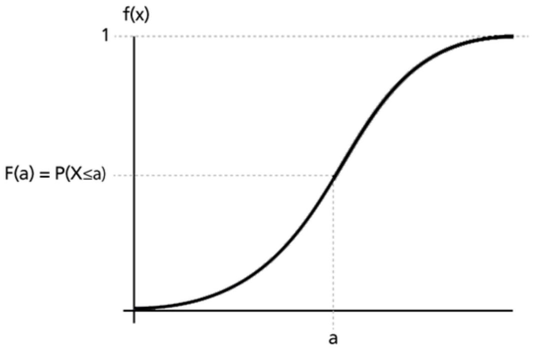
probability **density** function (pdf) $p : Val(X) \to [0, +\infty) \quad s.t. \quad \int_{Val(X)} p(x)\mathrm{d}x = 1$

$$P(X \leq a) \triangleq \underline{\int_{-\infty}^{a} p(x)\mathrm{d}x}$$

$F(a)$: the cumulative distribution function (cdf)



- note that often $P(X = x) = 0$
- $p(x)$ can be larger than 1
  - it is **not a probability distribution**
- $P(a \leq X \leq b) = F(b) - F(a)$
- may only consider *measurable* subsets A



Continuous domain

probability **density** function (pdf) $p : Val(X) \rightarrow [0, +\infty) \quad s.t. \quad \int_{Val(X)} p(x)\mathrm{d}x = 1$

for discrete domains:

probability **mass** function (pmf) $p(x) \triangleq P(X = x) \quad s.t. \quad \sum_{Val(X)} p(x) = 1$

Continuous domain: multivariate case

Joint density of multipe RVs: (same conditions)

$$P(X_1 \leq a_1, \ldots, X_n \leq a_n) \triangleq \underbrace{\int_{-\infty}^{a_1} \ldots \int_{-\infty}^{a_n} p(x_1, \ldots, x_n) \mathrm{d}x_n \ldots \mathrm{d}x_1}$$

$F(a_1, \ldots, a_n)$: joint CDF

Continuous domain: multivariate case

Joint density of multipe RVs: (same conditions)

$$P(X_1 \leq a_1, \dots, X_n \leq a_n) \triangleq \underbrace{\int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} p(x_1, \dots, x_n) \mathrm{d}x_n \dots \mathrm{d}x_1}$$

$F(a_1, \dots, a_n)$: joint CDF

Marginal density: $p(x_1) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p(x_1, \dots, x_n) \mathrm{d}x_n \dots \mathrm{d}x_2$

- marginal CDF $F(x_1) = \lim_{x_2, \dots, x_n \to \infty} F(x_1, \dots, x_n)$

Continuous domain: conditional density

Conditional distribution: $P(X \mid Y = y) = \frac{P(X, Y=y)}{P(Y=y)}$ zero measure!

Take the limit $\epsilon \to 0$ in: $P(X \leq a \mid y - \epsilon \leq Y \leq y + \epsilon) = \frac{\int_{-\infty}^{a} \int_{e=-\epsilon}^{\epsilon} p(x, y+e) \mathrm{d}e \mathrm{d}x}{\int_{e=-\epsilon}^{\epsilon} p(y+e) \mathrm{d}e}$

Continuous domain: conditional density

Conditional distribution: $P(X \mid Y = y) = \frac{P(X, Y=y)}{P(Y=y)}$ zero measure!

Take the limit $\epsilon \to 0$ in: $P(X \leq a \mid y - \epsilon \leq Y \leq y + \epsilon) = \frac{\int_{-\infty}^{a} \int_{e=-\epsilon}^{\epsilon} p(x, y+e) \mathrm{d}e \mathrm{d}x}{\int_{e=-\epsilon}^{\epsilon} p(y+e) \mathrm{d}e}$

using $\int_{e=-\epsilon}^{\epsilon} f(y+e)\mathrm{d}e = 2\epsilon f(y) + \mathcal{O}(\epsilon^2)$

↓

$$P(X \leq a \mid y - \epsilon \leq Y \leq y + \epsilon) \approx \frac{\int_{-\infty}^{a} p(x, y) \mathrm{d}x}{p(y)}$$

Continuous domain: conditional density

Conditional distribution: $P(X \mid Y = y) = \frac{P(X,Y=y)}{P(Y=y)}$ zero measure!

Take the limit $\epsilon \to 0$ in: $P(X \leq a \mid y - \epsilon \leq Y \leq y + \epsilon) = \frac{\int_{-\infty}^{a} \int_{e=-\epsilon}^{\epsilon} p(x,y+e)\mathrm{d}e\mathrm{d}x}{\int_{e=-\epsilon}^{\epsilon} p(y+e)\mathrm{d}e}$

using $\int_{e=-\epsilon}^{\epsilon} f(y+e)\mathrm{d}e = 2\epsilon f(y) + \mathcal{O}(\epsilon^2)$

↓

$$P(X \leq a \mid y - \epsilon \leq Y \leq y + \epsilon) \approx \frac{\int_{-\infty}^{a} p(x,y)\mathrm{d}x}{p(y)}$$

Conditional density of $P(X \mid Y = y)$ is $p(x \mid y) = \frac{p(x,y)}{p(y)}$

- *extends Bayes' rule and chain rule and conditional independence to* ***densities***

Functions of random variables

- RV is a function of the outcome $X : \Omega \rightarrow Val(X)$
- therefore $g(X) = g(X(\omega))$ is an RV itself
  - *E.g.,* $Y = X_1 + X_2$

Expectation & Variance

Expectation: $\mathbb{E}[X] \triangleq \sum_{x \in Val(X)} xp(x)$ OR $\mathbb{E}[X] \triangleq \int_{x \in Val(X)} xp(x)\mathrm{d}x$

- linearity: $\mathbb{E}[X + aY] = \mathbb{E}[X] + a\mathbb{E}[Y]$
  - X:# heads, Y:#heads in the first trial (X&Y are not independent)
- for independent X & Y

$$\mathbb{E}[XY] = \sum_{x,y \in Val(X) \times Val(Y)} p(x,y)(xy) = \sum_{x,y \in Val(X) \times Val(Y)} p(x)p(y)(xy)$$

$$= (\sum_{x \in Val(X)} xp(x))(\sum_{y \in Val(Y)} yp(y)) = \mathbb{E}[X]\mathbb{E}[Y]$$

Expectation & Variance

Variance: $Var[X] \triangleq \mathbb{E}[(X - \mathbb{E}[X])^2]$

$$= \mathbb{E}[X^2 + \mathbb{E}[X]^2 - 2X\mathbb{E}[X]] = \mathbb{E}[X^2] + \mathbb{E}[X]^2 - 2\mathbb{E}[X]\mathbb{E}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

- for independent X and Y $Var[X+Y] = Var[X] + Var[Y]$
  - if not independent $Var[X+Y] = Var[X] + Var[Y] + 2\,Cov[X,Y]$
- Covariance: $Cov[X,Y] \triangleq \mathbb{E}[X - \mathbb{E}[X]]\mathbb{E}[Y - \mathbb{E}[Y]] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
  - generalizes variance $Cov[X,X] = Var[X]$
  - symmetric & bilinear $Cov[aX,bY] = abCov[Y,X]$

Examples of probability dists.

Classical members of exponential family of distribution

more on this later

- Gaussian
- Bernoulli
- Binomial
- Multinomial
- Gamma
- Exponential
- Poisson
- Beta
- Dirichlet

Examples of probability dists.

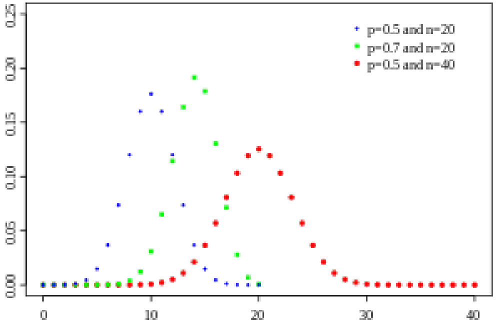
Bernoulli: $P(X = 1; \mu) = \mu \quad 0 \leq \mu \leq 1$ OR $p(x;\mu) = \mu^x (1-\mu)^{1-x}$

- discrete distribution with $Val(X) = \{0, 1\}$

Binomial: $P(X = k; \mu, n) = \binom{n}{k} \mu^k (1-\mu)^{n-k}$

- dist. over the number of ones in *n* independent Bernoulli trials
- *number of heads in n coin toss*

$$Val(X) = \{0, \ldots, n\}$$



Examples of probability dists.

Categorical (*aka.* multinulli) : $P(X = l; \mu) = \mu_l \quad \text{where} \quad \sum_l \mu_l = 1$

- *fully parameterized* discrete distribution with $Val(X) = \{0 \ldots, L\}$

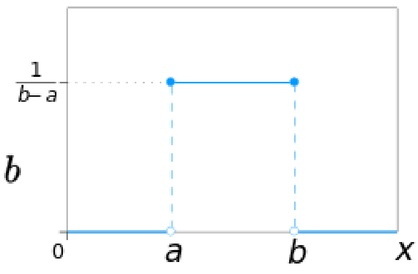
Multinomial distribution: $P(X_1 = x_1, \ldots, X_L = x_L; \mu, n) = \mathbb{I}(\sum_l x_l = n) \frac{n!}{\prod_l x_l!} \prod_l \mu_l^{x_l}$

- dist. over the number of different outcomes in *n* independent categorial trials

Examples of probability dists.

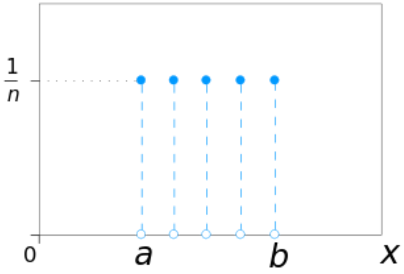
Uniform:

- CONTINUOUS $Val(X) = [a, b]$ $p(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$



- DISCRETE $Val(X) = \{a, a+1, \ldots, b\}$
  - max-entropy discrete distribution

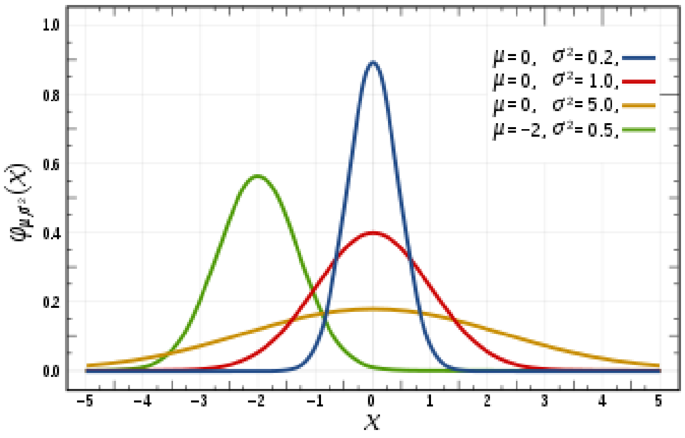
$$P(X = j) = \frac{1}{n}$$



Examples of probability dists.

Gaussian: $p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

- motivated by central limit theorem
- max-entropy dist. with a fixed variance



Summary

Adding random variables

- **Random variable:** assigns a value to each outcome
  - **Event (using RV):** set of outcomes with a particular attribute
  - **Prob. dist., cond. prob., chain rule, indep. …** are all extended to **RVs**
- **Continuous** domains**:** same definition of probability, event, RV etc.
  - **Specifying** the prob. dist. using **density function**

Summary

Notation

- random variable X, Y, Z $\mathbf{X} = [X_1, \ldots, X_n]$
- variable x, y, z
- PDF, PMF $p(x), p(\mathbf{x}), p(x, y)$
- probability distribution $P(X), P(x) \triangleq P(X = x)$

use interchangeably

- domain of an RV $Val(X), Val(X, Y, Z)$

bonus slides

Properties of conditional independence

- Symmetry: $(X \perp Y \mid Z) \Rightarrow (Y \perp X \mid Z)$
- Decomposition: $(X \perp Y, W \mid Z) \Rightarrow (X \perp Y \mid Z)$
- Weak union: $(X \perp Y, W \mid Z) \Rightarrow (X \perp Y \mid W, Z)$
- Contraction:

$$(X \perp Y \mid W, Z) \& (X \perp W \mid Y, Z) \Rightarrow (X \perp Y, W \mid Z)$$

- Intersection: *if P is positive*

$$(X \perp W \mid Y, Z) \& (X \perp Y \mid Z) \Rightarrow (X \perp Y, W \mid Z)$$

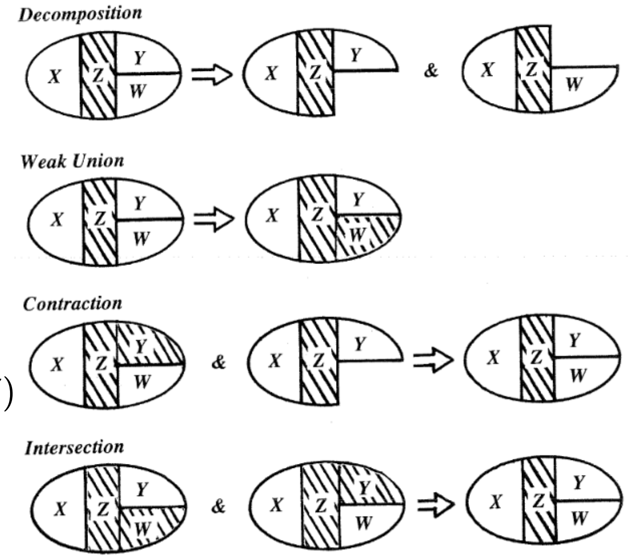


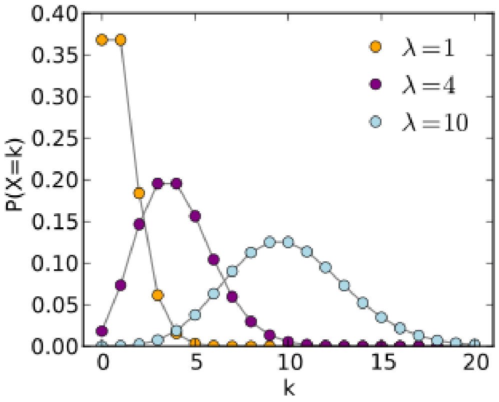
image: Pearl's book

Examples of probability dists.

Poisson: $p(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$ where $\lambda > 0$ is the *mean frequency* *(rate parameter)*

- frequency of rare events
- events are assumed independent
- similar to binomial with large number of trials $(\lambda \approx n\mu)$

$Val(X) = \mathbb{Z}^+$

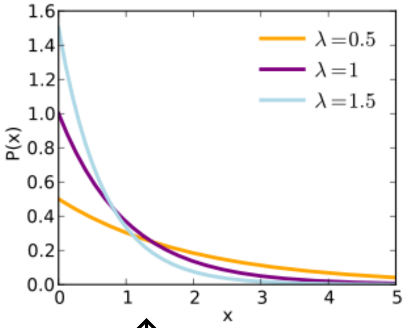


Examples of probability dists.

Exponential: $p(x;\lambda) = \lambda e^{-\lambda x} \quad \text{where} \quad \lambda > 0$

- time between events in Poisson dist.
- memoryless property

$Val(X) = \mathbb{R}^+$



$\updownarrow \quad (1-\mu) \equiv e^{-\lambda}$

Geometric: $p(x,k;\mu) = (1-\mu)^{k-1}\mu \quad \text{where} \quad 0 < \mu < 1$

- number of Bernoulli trials until success
- memoryless property

$Val(X) = \mathbb{N}$