Probability Intro

GA DAT3

Agenda

Definition

Bayes’ Rule

Continuous Probability

Super important distributions

- **Sample space** Ω: The set of all the outcomes of a random experiment. Here, each outcome $\omega \in \Omega$ can be thought of as a complete description of the state of the real world at the end of the experiment.
- **Set of events** (or **event space**) $\mathcal{F}$: A set whose elements $A \in \mathcal{F}$ (called **events**) are subsets of Ω (i.e., $A \subseteq \Omega$ is a collection of possible outcomes of an experiment).[1].
- **Probability measure**: A function $P : \mathcal{F} \to \mathbb{R}$ that satisfies the following properties,
  - $P(A) \geq 0$, for all $A \in \mathcal{F}$
  - $P(\Omega) = 1$
  - If $A_1, A_2, \ldots$ are disjoint events (i.e., $A_i \cap A_j = \emptyset$ whenever $i \neq j$), then

$$P(\cup_i A_i) = \sum_i P(A_i)$$

e.g. die rolling

What are the **sample space**

and **event space?**

Sample vs event space with a k-sided die:

- sample: { 1…6 }
- event: “odd” {1, 3, 5 }
- Random variable e.g. “sum of the numbers”

Given two events, A and B, we define the probability of A or B as follows:

P(A or B) = P(A) + P(B) - P(A & B)

$$
\begin{aligned}
p(A \vee B) &= p(A) + p(B) - p(A \wedge B) \\
&= p(A) + p(B) \text{ if } A \text{ and } B \text{ are mutually exclusive}
\end{aligned}
$$

joint (= conditional)

$$p(A, B) = p(A \wedge B) = p(A|B)p(B)$$

marginal

$$p(A) = \sum_b p(A, B) = \sum_b p(A|B = b)p(B = b)$$

product rule (= joint)

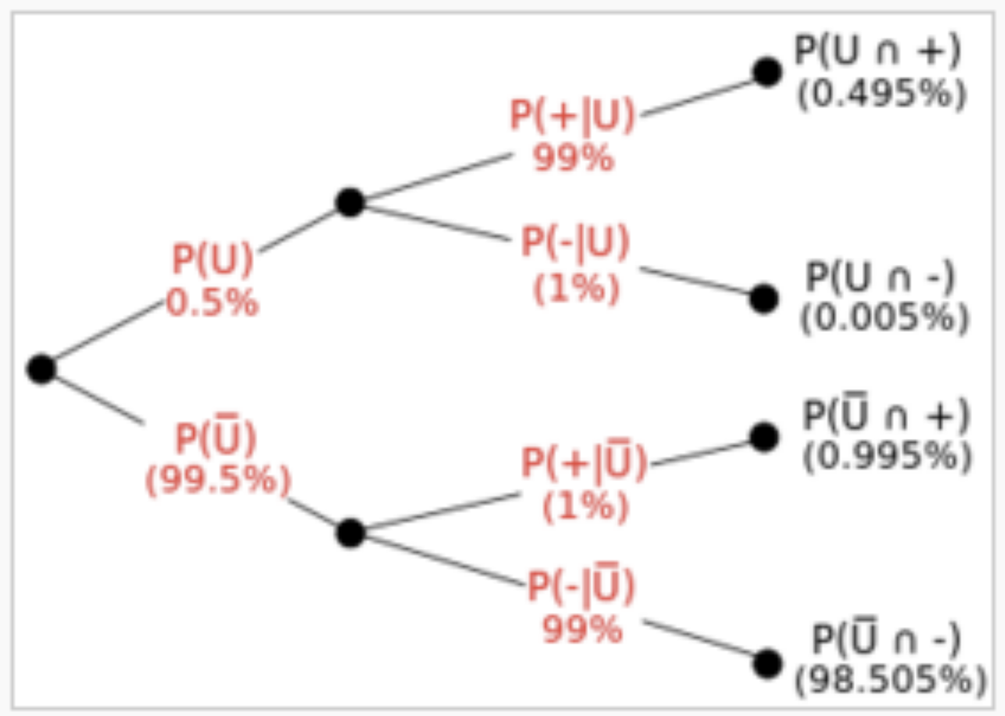
$$p(X_{1:D}) = p(X_1)p(X_2|X_1)p(X_3|X_2, X_1)p(X_4|X_1, X_2, X_3) \dots p(X_D|X_{1:D-1})$$

Bayes' Rule

Suppose a drug test is 99% sensitive and 99% specific. That is, the test will produce 99% true positive results for drug users and 99% true negative results for non-drug users. Suppose that 0.5% of people are users of the drug. If a randomly selected individual tests positive, what is the probability he or she is a user?

$$
\begin{aligned}
P(\text{User} \mid +) &= \frac{P(+ \mid \text{User}) P(\text{User})}{P(+ \mid \text{User}) P(\text{User}) + P(+ \mid \text{Non-user}) P(\text{Non-user})} \\
&= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \\
&\approx 33.2\%
\end{aligned}
$$

Baye’s Rule



(conditional) independence

$$X \perp Y|Z \iff p(X,Y|Z) = p(X|Z)p(Y|Z)$$

very fun (easy-ish) challenge - prove this:

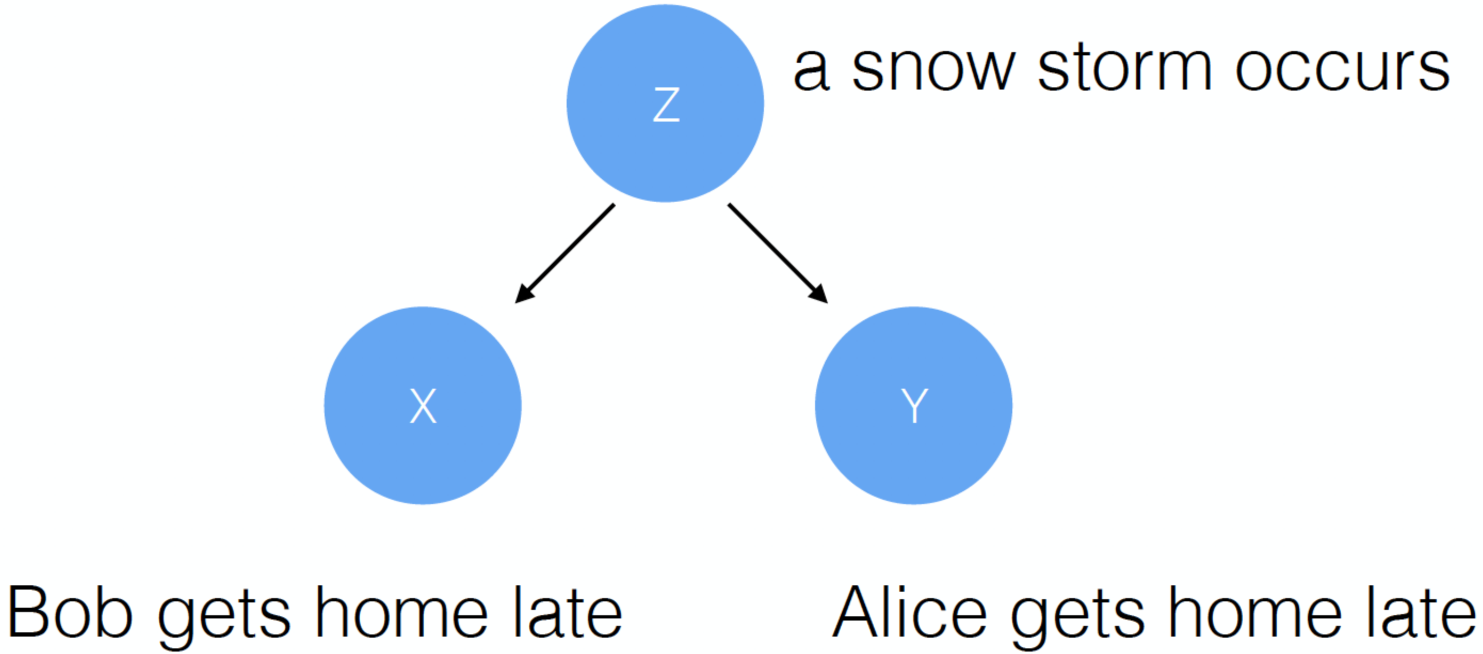
Theorem 2.2.1. *$X \perp Y|Z$ iff there exist function g and h such that*

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

for all x, y, z such that $p(z) > 0$.

conditional independence: example

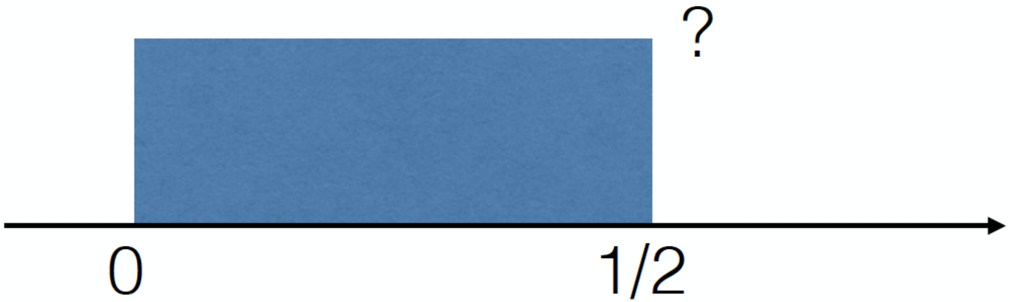
I like graphs



Continuous Probability

$$P(a < X \leq b) = \int_a^b f(x)dx$$

weird stuff - what goes on here?

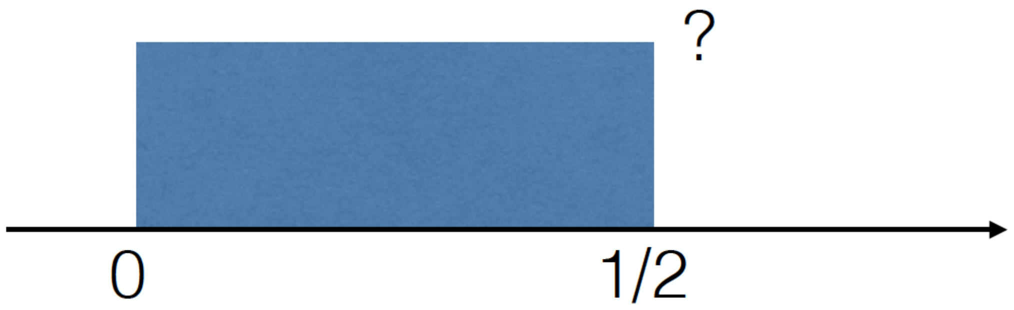


Continuous Probability

$$P(a < X \leq b) = \int_a^b f(x)dx$$

$$\text{Unif}(x|a,b) = \frac{1}{b-a}\mathbb{I}(a \leq x \leq b)$$

If we set $a = 0$ and $b = \frac{1}{2}$, we have $p(x) = 2$ for any $x \in [0, \frac{1}{2}]$



Mean, Variance, Covariance

$$E[X] = ?$$

$$\begin{aligned}\operatorname{var}[X] &\triangleq \mathbb{E}\left[(X-\mu)^2\right] = \int (x-\mu)^2 p(x)dx \\ &= \int x^2 p(x)dx + \mu^2 \int p(x)dx - 2\mu \int x p(x)dx = \mathbb{E}\left[X^2\right] - \mu^2\end{aligned}$$

$$\operatorname{cov}[X,Y] \triangleq \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

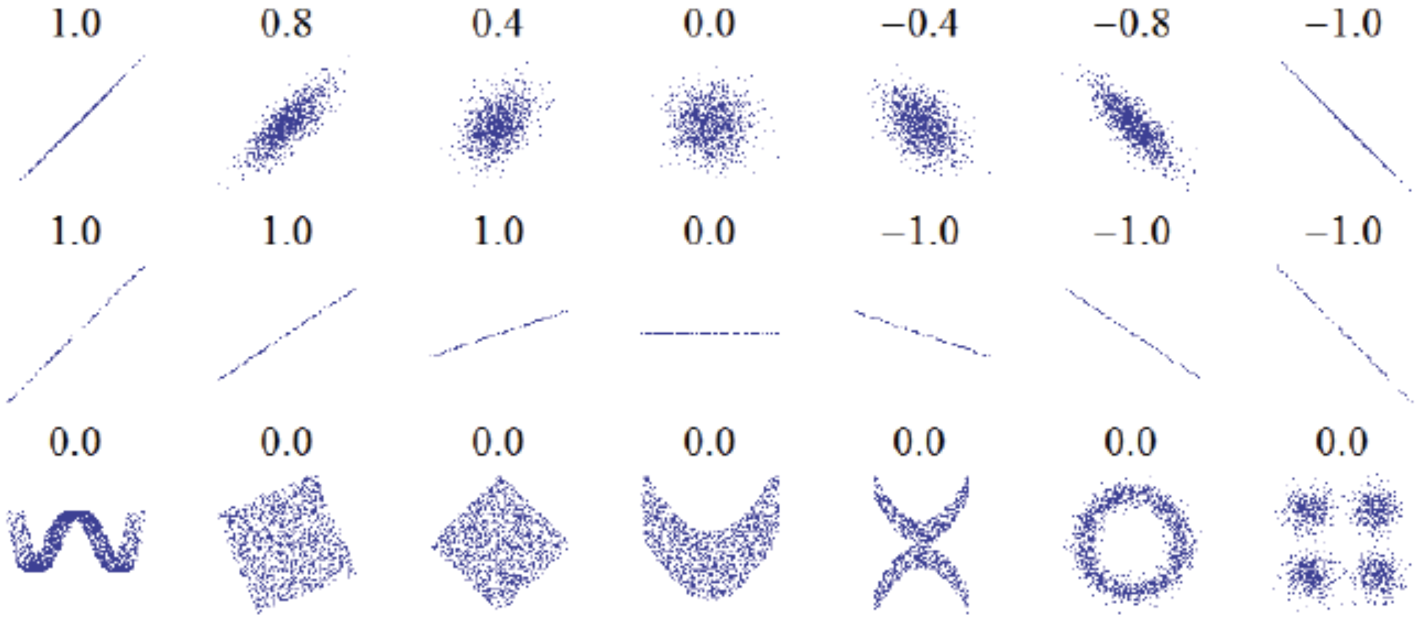


Figure 2.12 Several sets of (x, y) points, with the correlation coefficient of x and y for each set. Note that the correlation reflects the noisiness and direction of a linear relationship (top row), but not the slope of that relationship (middle), nor many aspects of nonlinear relationships (bottom). N.B.: the figure in the center has a slope of 0 but in that case the correlation coefficient is undefined because the variance of Y is zero. Source: `http://en.wikipedia.org/wiki/File:Correlation_examples.png`

Super important distributions - Bernoulli, Binomial

$$\mathrm{Ber}(x|\theta) = \theta^{\mathbb{I}(x=1)}(1-\theta)^{\mathbb{I}(x=0)}$$

In other words,

$$\mathrm{Ber}(x|\theta) = \begin{cases} \theta & \text{if } x = 1 \\ 1-\theta & \text{if } x = 0 \end{cases}$$

$$\mathrm{Bin}(k|n,\theta) \triangleq \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

where

$$\binom{n}{k} \triangleq \frac{n!}{(n-k)!k!}$$

why?

examples?

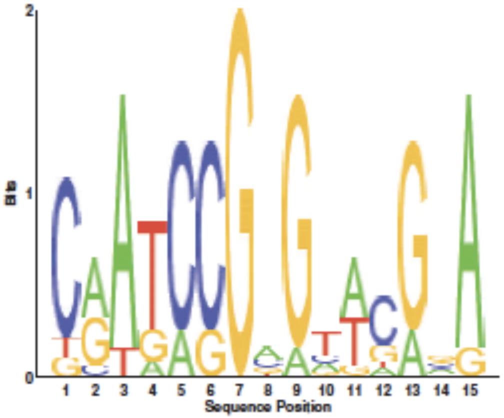
Super important distributions - Multinoulli

$$\mathrm{Mu}(\mathbf{x}|n,\boldsymbol{\theta}) \triangleq \binom{n}{x_1 \ldots x_K} \prod_{j=1}^{K} \theta_j^{x_j}$$

where θ_j is the probability that side j shows up, and

$$\binom{n}{x_1 \ldots x_K} \triangleq \frac{n!}{x_1! x_2! \cdots x_K!}$$

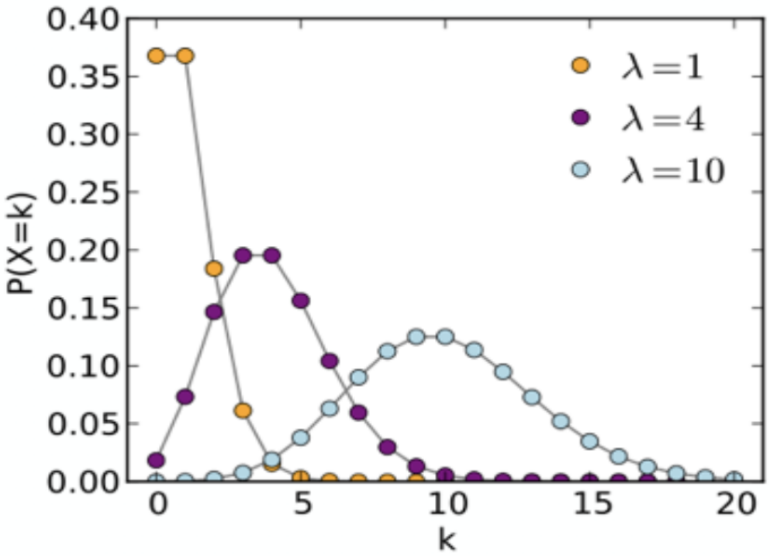
a	t	a	g	c	c	g	g	t	a	c	g	g	c	a
t	t	a	g	c	t	g	c	a	a	c	c	g	c	a
t	c	a	g	c	c	a	c	t	a	g	a	g	c	a
a	t	a	a	c	c	g	c	g	a	c	c	g	c	a
t	t	a	g	c	c	g	c	t	a	a	g	g	t	a
t	a	a	g	c	c	t	c	g	t	a	c	g	t	a
t	t	a	g	c	c	g	t	t	a	c	g	g	c	c
a	t	a	t	c	c	g	g	t	a	c	a	g	t	a
a	t	a	g	c	a	g	g	t	a	c	c	g	a	a
a	c	a	t	c	c	g	t	g	a	c	g	g	a	a

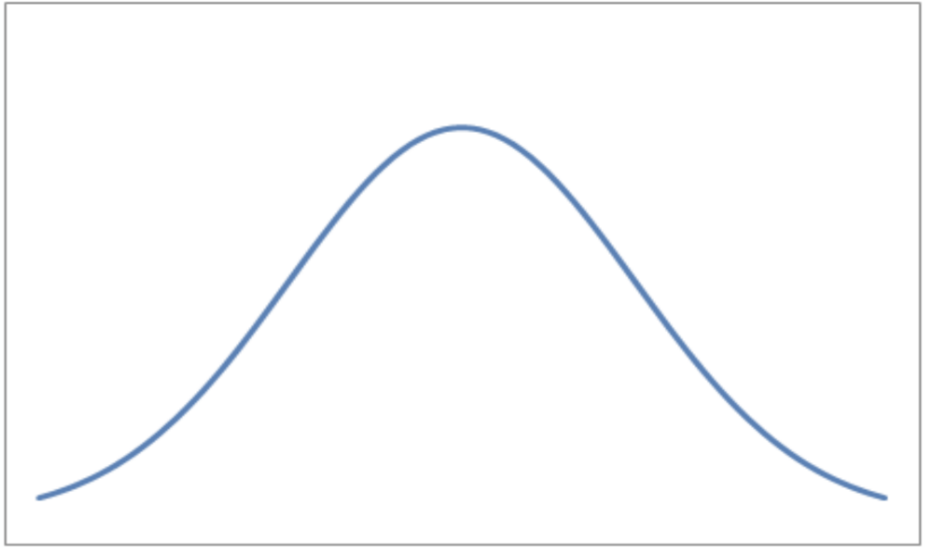


Super important distributions - Poisson, Gaussian

$$\mathrm{Poi}(x|\lambda) = e^{-\lambda}\,\frac{\lambda^x}{x!}$$

$$\mathcal{N}(x|\mu,\sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



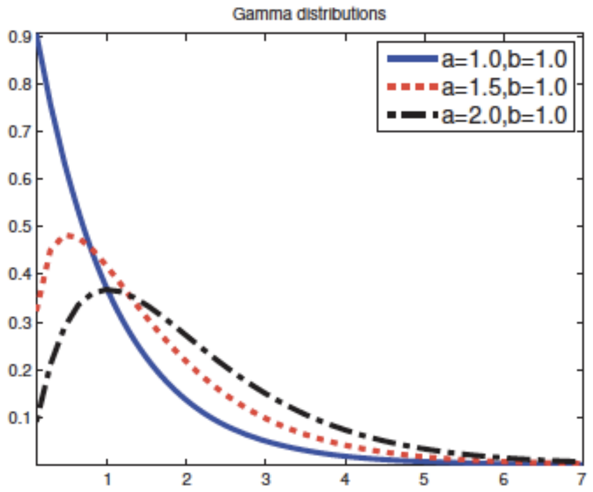


Super important distributions - Gamma, Beta

$$\mathrm{Ga}(T|\text{shape} = a, \text{rate} = b) \triangleq \frac{b^a}{\Gamma(a)} T^{a-1} e^{-Tb}$$

where $\Gamma(a)$ is the gamma function:

$$\Gamma(x) \triangleq \int_0^\infty u^{x-1} e^{-u} du$$



$$\mathrm{Beta}(x|a, b) = \frac{1}{B(a, b)} x^{a-1}(1 - x)^{b-1}$$

Here $B(p, q)$ is the beta function,

$$B(a, b) \triangleq \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$$