Bayesian Inference

Thursday 24th November, 2016

Order of the Day

- diving a bit deeper into statistical inference
- statistics - some religious aspects
- estimation - there's more to life than maximum likelihood estimation

"Probability Arithmetics"

Probability - simple arithmatics

Two $\boxed{fair}$ coins X and Y and one $\boxed{fair}$ dice Z

"Probability Arithmetics"

Probability - simple arithmatics

Two $\boxed{fair}$ coins X and Y and one $\boxed{fair}$ dice Z

- sum $\approx$ logical OR
  - What's the probability of the dice landing on 1 or 6

$$P(Z=1)+P(Z=6)=1/6+1/6$$

"Probability Arithmetics"

IsraelTech challenge

Probability - simple arithmatics

Two $\boxed{fair}$ coins X and Y and one $\boxed{fair}$ dice Z

- sum $\approx$ logical OR
  - What's the probability of the dice landing on 1 or 6

$$P(Z=1)+P(Z=6)=1/6+1/6$$

- product $\approx$ logical AND
  - What's the probability of both coins landing on heads?

$$P(X=H)\cdot P(Y=H)=0.5\cdot 0.5$$

Logical OR

$$P\left(A \text{ or } B\right) \stackrel{?}{=} P\left(A\right) + P\left(B\right)$$

Logical OR

$$P\left(A \text{ or } B\right) \stackrel{?}{=} P\left(A\right) + P\left(B\right)$$

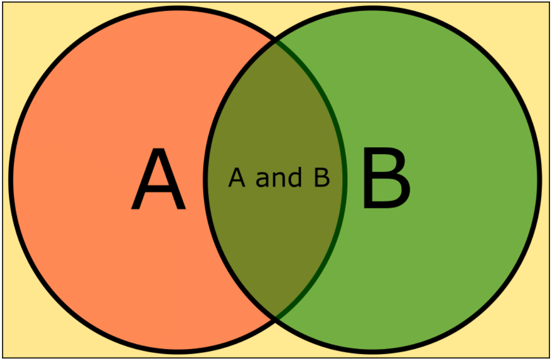
- When is this true?
  - A and B are disjoint events
- And if not?
  - X, Y are fair coins

$$\begin{aligned} P\left(or\left(X=1, Y=1\right)\right) &= 0.5 + 0.5 \\ &\stackrel{???}{=} 1 \end{aligned}$$

Logical OR

Israeltech challenge

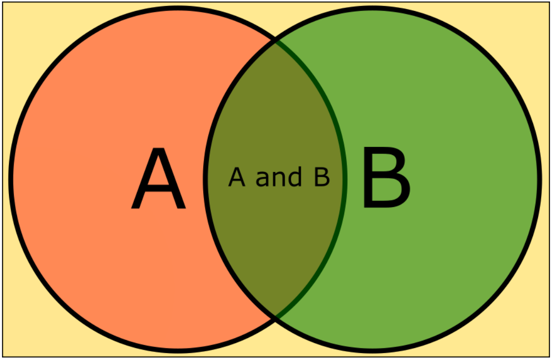
Darts



areas

Logical OR

Darts



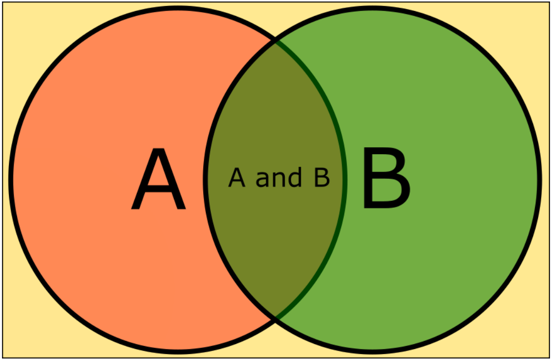
$$P(A) = P\left(and\left(A, \overline{B}\right)\right) + P\left(and\left(A, B\right)\right)$$

$$P(B) = P\left(and\left(\overline{A}, B\right)\right) + P\left(and\left(A, B\right)\right)$$

problem?

Logical OR

Darts



$$P\left(A\right) = P\left(and\left(A, \overline{B}\right)\right) + P\left(and\left(A, B\right)\right)$$

$$P\left(B\right) = P\left(and\left(\overline{A}, B\right)\right) + P\left(and\left(A, B\right)\right)$$

$$P\left(or\left(A, B\right)\right) = P\left(A\right) + P\left(B\right) - P\left(and\left(A, B\right)\right)$$

Joined Probability

$$P\left(X, Y\right)$$

- X someone falling asleep in the first row
- Y someone falling asleep in the last row

Joined Probability

$$P\left(X,Y\right)\stackrel{?}{=}P\left(X\right)\cdot P\left(Y\right)$$

Joined Probability

$$P(X,Y) \stackrel{?}{=} P(X) \cdot P(Y)$$

- X a fair coin
- Y the same coin

The probability $X = Y = 1$?

$$P(X=1) \cdot P(Y=1) \stackrel{???}{=} 0.25$$

Joined Probability

$$P\left(X,Y\right)\stackrel{?}{=}P\left(X\right)\cdot P\left(Y\right)$$

- When is this true?

Joined Probability

$$P\left(X,Y\right)\stackrel{?}{=}P\left(X\right)\cdot P\left(Y\right)$$

- When is this true?
  - X is independent of Y

Joined Probability

$$P(X,Y) \stackrel{?}{=} P(X) \cdot P(Y)$$

- When is this true?
  - X is independent of Y
- And if not?
  - $P(X,Y) = P(X \mid Y)\,P(Y)$

Conditional Probability

$P(X \mid Y)$ - probability of X conditional on Y

returning to the previous example:

$$P(X \mid Y) = \begin{cases} 1, & X = Y \\ 0, & X \neq Y \end{cases}$$

therefore

$$\begin{aligned} P(X = 1, Y = 1) &= P(X = 1 \mid Y = 1) \cdot P(Y = 1) \\ &= 0.5 \end{aligned}$$

Conditional Probability

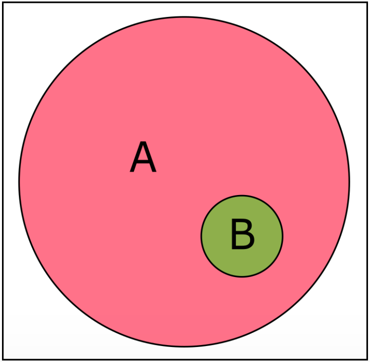
IsraelTech challenge

Another perspective

- A a Fellow sitting in class
- B the Fellow is listening

Calculate $P\left(B=1 \mid A=1\right)$

- start with $P\left(B=1, A=1\right)$



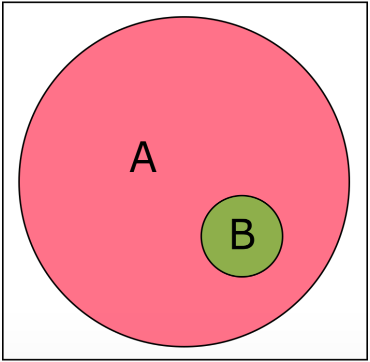
Conditional Probability

Another perspective

- A a Fellow sitting in class
- B the Fellow is listening

Calculate $P(B = 1 \mid A = 1)$

- start with $P(B = 1, A = 1)$
- divide by the probability of what's given $P(A = 1)$



The Law of Total Probability

description	probability
summery days	0.65
wintery days	0.35
rain on summery days	0.03
rain on wintery days	0.1

what's the probability of rain?

The Law of Total Probability

Israeltech challenge

description	probability
summery days	0.65
wintery days	0.35
rain on summery days	0.03
rain on wintery days	0.1

what's the probability of rain?

$$\begin{aligned} P(rain) &= P(rain \mid summery) \cdot P(summery) \\ &+ P(rain \mid wintery) \cdot P(wintery) \\ &= 0.03 \cdot 0.65 + 0.1 \cdot 0.35 \end{aligned}$$

The Law of Total Probability

$$\boxed{P\left(X\right)=\sum_{y}P\left(X\mid Y=y\right)P\left(Y=y\right)}$$

Bayes Theorem

- We know that

$$P(X \mid Y) = \frac{P(X, Y)}{P(Y)}$$

- Can we compute $P(Y \mid X)$ from

$$P(X), P(Y), P(X \mid Y)?$$

Bayes Theorem

$$P\left(Y \mid X\right) = \frac{P\left(X, Y\right)}{P\left(X\right)}$$

Bayes Theorem

$$
\begin{aligned}
P\left(Y \mid X\right) &= \frac{P\left(X, Y\right)}{P\left(X\right)} \\
&= \frac{P\left(X \mid Y\right) P\left(Y\right)}{P\left(X\right)}
\end{aligned}
$$

Bayes Theorem

$$\boxed{P\left(Y \mid X\right) = \frac{P\left(X \mid Y\right) P\left(Y\right)}{P\left(X\right)}}$$

Bayesian Vs. Frequentist Statistics

Question: What's the probability of rain today?

Bayesian Vs. Frequentist Statistics

Question: What's the probability of rain today?

- We can estimate
- Given N days
  - we can estimate what's the proportion of rainy days
  - rain on a given day is deterministic
  - **no probability**

What's your opinion?

Bayesian Vs. Frequentist Statistics

Israeltech challenge

Bayesian

Probability can model our belief or uncertainty regarding the world

Frequentist

Probability is a limit of frequency

Maximum Likelihood Revisited

Example: estimating coin parameter p

- coin tossed 10 times - 8 heads, 2 tails
- how would you estimate p?

Maximum Likelihood Revisited

Example: estimating coin parameter p

- coin tossed 10 times - 8 heads, 2 tails
- how would you estimate p?
- inside information - previous coins

$$p = \begin{cases} 0.3, & 9 \text{ out of } 10 \\ 0.8, & 1 \text{ out of } 10 \end{cases}$$

- can you use this information?

Maximum Likelihood Revisited

Example: estimating coin parameter p

- coin tossed 10 times - 8 heads, 2 tails
- how would you estimate p?
- inside information - previous coins

$$p = \begin{cases} 0.3, & 9 \text{ out of } 10 \\ 0.8, & 1 \text{ out of } 10 \end{cases}$$

- can you use this information?
- MLE - no
- if p is a constant - what's the point?

Maximum Likelihood Revisited

Example: estimating coin parameter p

- if you are willing to go Bayesian, there is a way
- compute probability (uncertainty) for p using Bayes Theorem

in this problem	name	formula
our inside information	**prior distribution**	$P(p)$
our uncertainty regarding p given the data	**posterior distribution**	$P(p \mid X)$

Maximum a Posteriori Estimation

- an alternative to MLE

$$\theta_{map} = \arg\max_{\theta} P\left(\theta \mid data\right)$$

Maximum a Posteriori Estimation

- an alternative to MLE

$$\theta_{map} = \arg\max_{\theta} P(\theta \mid data) = \arg\max_{\theta} \frac{P(data \mid \theta)\, P(\theta)}{P(data)}$$

Maximum a Posteriori Estimation

Israeltech challenge

- an alternative to MLE

$$\theta_{map} = \arg\max_{\theta} P\left(\theta \mid data\right)$$

- problem

$$\theta_{map} = \arg\max_{\theta} \frac{P\left(data \mid \theta\right) P\left(\theta\right)}{\color{red}{P\left(data\right)}}$$

Maximum a Posteriori Estimation

- an alternative to MLE

$$\theta_{map} = \arg\max_{\theta} P\left(\theta \mid data\right)$$

- problem

$$\theta_{map} = \arg\max_{\theta} \frac{P\left(data \mid \theta\right) P\left(\theta\right)}{P\left(data\right)}$$

- data is given - same for all θ

Maximum a Posteriori Estimation

$$\boxed{\theta_{map} = \arg\max_{\theta} P\left(data \mid \theta\right) P\left(\theta\right)}$$

MAP - examples

What season is this?

we observe a week with two days of rain

prior	likelihood
$P\left(summery\right) = 0.65$	$P\left(rain \mid summery\right) = 0.03$
$P\left(wintery\right) = 0.35$	$P\left(rain \mid wintery\right) = 0.1$

MAP - examples

What season is this?

we observe a week with two days of rain

prior	likelihood
$P\left(summery\right) = 0.65$	$P\left(rain \mid summery\right) = 0.03$
$P\left(wintery\right) = 0.35$	$P\left(rain \mid wintery\right) = 0.1$

$$\theta_{map} = \underset{wintery, summery}{\arg\max} \begin{cases} 0.65 \cdot 0.03^2 \cdot 0.97^5, & \theta = summery \\ 0.35 \cdot 0.1^2 \cdot 0.9^5, & \theta = wintery \end{cases}$$

MAP - examples

Israeltech challenge

What season is this?

we observe a week with two days of rain

prior	likelihood
$P\left(summery\right) = 0.65$	$P\left(rain \mid summery\right) = 0.03$
$P\left(wintery\right) = 0.35$	$P\left(rain \mid wintery\right) = 0.1$

$$\theta_{map} = \underset{wintery,summery}{\arg\max} \begin{cases} 0.0005, & \theta = summery \\ 0.002, & \theta = wintery \end{cases}$$

MAP - examples

Back to the coin

whiteboard

MAP - examples

Israeltech challenge

Back to the coin

whiteboard

$$p_{map} = \arg\max_{0.3,0.8} \begin{cases} 0.00067, & p = 0.8 \\ 0.000029, & p = 0.3 \end{cases}$$

MAP - examples

normal - normal

estimating mean height in Israel from samples
$D = X_1, \ldots, X_n$

- height - $H \sim \mathcal{N}\left(\theta, \sigma^2\right)$
- mean height across countries - $\theta \sim \mathcal{N}\left(\phi, \zeta^2\right)$

MAP - examples

normal - normal

estimating mean height in Israel from samples
$D = X_1, \dots, X_n$

- height - $H \sim \mathcal{N}\left(\theta, \sigma^2\right)$
- mean height across countries - $\theta \sim \mathcal{N}\left(\phi, \zeta^2\right)$

$$\theta_{map} = \arg\max_{\theta} P\left(D \mid \theta\right) P\left(\theta\right)$$

$$= \arg\max_{\theta} \left(\frac{1}{\sqrt{2\pi\zeta^2}} e^{-\frac{\left(\theta-\phi\right)^2}{2\zeta^2}}\right) \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(X_i-\theta\right)^2}{2\sigma^2}}\right)$$

MAP - examples

normal - normal

estimating mean height in Israel from samples
$D = X_1, \ldots, X_n$

- height - $H \sim \mathcal{N}\left(\theta, \sigma^2\right)$
- mean height across countries - $\theta \sim \mathcal{N}\left(\phi, \zeta^2\right)$

$$\theta_{map} = \arg\max_{\theta} P\left(D \mid \theta\right) P\left(\theta\right)$$

$$= \arg\max_{\theta} \left(\frac{1}{\sqrt{2\pi\zeta^2}\sqrt{2\pi\sigma^2}} e^{-\frac{\left(\theta-\phi\right)^2}{2\zeta^2} + \sum_{i=1}^{n} -\frac{\left(X_i-\theta\right)^2}{2\sigma^2}} \right)$$

MAP - examples

Israeltech challenge

normal - normal

estimating mean height in Israel from samples
$D = X_1, \ldots, X_n$

- height - $H \sim \mathcal{N}\left(\theta, \sigma^2\right)$
- mean height across countries - $\theta \sim \mathcal{N}\left(\phi, \zeta^2\right)$

$$\theta_{map} = \arg\max_{\theta} \left\{ -\frac{(\theta - \phi)^2}{2\zeta^2} + \sum_{i=1}^{n} -\frac{(X_i - \theta)^2}{2\sigma^2} \right\}$$

$$= \arg\min_{\theta} \left\{ \frac{\sigma^2}{\zeta^2} (\theta - \phi)^2 + \sum_{i=1}^{n} (X_i - \theta)^2 \right\}$$

MAP - examples

normal - normal

estimating mean height in Israel from samples
$D = X_1, \ldots, X_n$

- height - $H \sim \mathcal{N}\left(\theta, \sigma^2\right)$
- mean height across countries - $\theta \sim \mathcal{N}\left(\phi, \zeta^2\right)$

- if the problem is nice (this one is)
- find the unique minimizer

derivative

MAP - examples

normal - normal

estimating mean height in Israel from samples
$D = X_1, \ldots, X_n$

- height - $H \sim \mathcal{N}\left(\theta, \sigma^2\right)$
- mean height across countries - $\theta \sim \mathcal{N}\left(\phi, \zeta^2\right)$

$$\frac{d}{d\theta}\left\{\frac{\sigma^2}{\zeta^2}\left(\theta-\phi\right)^2+\sum_{i=1}^{n}\left(X_i-\theta\right)^2\right\} \overset{!}{=} 0$$

$$\frac{\sigma^2}{\zeta^2}\left(\theta-\phi\right) = -n\theta+\sum_{i=1}^{n}X_i$$

MAP - examples

normal - normal

estimating mean height in Israel from samples
$D = X_1, \ldots, X_n$

- height - $H \sim \mathcal{N}\left(\theta, \sigma^2\right)$
- mean height across countries - $\theta \sim \mathcal{N}\left(\phi, \zeta^2\right)$

$$\theta_{map} = \frac{\sum_{i=1}^{n} X_i + \phi \frac{\sigma^2}{\zeta^2}}{n + \frac{\sigma^2}{\zeta^2}}$$

$\zeta^2 \lesseqgtr \sigma^2$

Bayesian Inference

IsraelTech challenge

coin example

- prior - $p = 0.8$ only 10% of the time
- intuition - $p_{mle} = 0.8 \Rightarrow P\left(p = 0.8\right)$ will increase
- update - $P\left(p \mid data\right) \rightarrow P\left(p\right)$

Bayesian Inference

Bayesian Updating

we all do it, all the time

- bus to class $10min$
- walk to class $15min$
- a bus is quicker 95% of the time
- last few days - horrible traffic - bus time $25min$
- estimation of - best method of travel - will begin to change

Is Bayesian Inference Good?

Is the philosophy good?

- a matter of opinion
- probability is a modeling tool
- don't be an extremist

Is Bayesian Inference Good?

Is the practice good?

- are assumptions good?
- depends on the assumptions

Is Bayesian Inference Good?

Is the practice good?

- are assumptions good?
- depends on the assumptions
  - you lost something at your house
  - you think you remember where
  - start from there, expand search
    - you remember correctly - find item faster
    - you remember incorrectly - find item slower

Is Bayesian Inference Good?

Is the practice good?

- are assumptions good?
- depends on the assumptions
  - you see people going right or left to avoid an obstacle
  - assume best choice is normally distributed
  - choose the mean
  - bump into the obstacle

Is Bayesian Inference Good?

Is the practice good?

- are assumptions good?
- depends on the assumptions

discussion - assumptions

Conclusion

- joined and conditional probability
- Bayesian Vs. Frequntist statistics
- Bayesian Inference
- MAP estimator