Probability Rules

- Seen how to work with samples from prob functions
- Seen how to summarize them.
- Need to formally manipulate probability distros

When we discuss a generic probabily distro,

write $P(X)$

samples $x_i \sim P(X)$

Sometimes $prob(X)$

~~Some distros~~

If ~~Some~~ X is discrete, for exampe True/False; Integers,

$P(X)$ has no units, $0 \leq P(X) \leq 1$.

~~If continuous $[P(X)] = \frac{1}{[X]}$~~.

"Or" $P(X) + P(\neg X) = 1$

normalized over all possibilities.

Continuous:

$$[P(X)] = \frac{1}{[X]}$$

Normalization: $\int_{\forall X} P(X)\,dX = 1$

Note: $P(X)$ can be > 1 in narrow regions

- $P(X)$ is a probability density. or pdf

High prob density $\not\Rightarrow$ High probability.

Joint Probability

Can talk about more than 1 var:

$$P(X,Y) \quad \text{or} \quad P(\vec{X}) = P(x_1, x_2, x_3, x_4)$$

discrete: X "AND" Y at same time

continuous: multi-dimensional PDF

Conditional Probability

$P(X|Y)$ — probability X is true given Y is true

given

or

probability X given Y is fixed to some value

Product Rule

$$P(X,Y) = P(X|Y)P(Y)$$

works the other way too

$$P(X,Y) = P(Y|X)P(X)$$

probability of X irregardless of Y

Also works ~~if~~ for Conditional Probs:

$$P(X,Y|Z) = P(X|Y,Z)P(Y|Z)$$

Bayes Rule - Put product rule together

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Independence

$$P(X,Y) = \cancel{P(X|Y)}P(Y) = P(X)P(Y)$$

$$\rightarrow P(X)$$

Marginalization - When we don't know a variable's value, or don't care

$$P(X) = \int_{\forall Y} P(X,Y)\,dY = \int_{\forall Y} P(X|Y)P(Y)\,dY$$

discrete: $P(X) = \sum_{Y} P(X|Y)P(Y)$

[Supernovae Example]

[Prisoner Example]

Tail: $P(>X) = \int_{X}^{\infty} P(x)\,dX$

Changing Variables

$P(X)dx = P(Y)dy$

$P(X) = P(Y)\left|\frac{dy}{dx}\right|$

Supernova

$P(I_a) = 0.3$ probability of type Ia

$P(I_b) = 0.2$ probability of type Ib

$P(II) = 0.5$

Note: - Discrete variable w/ 3 options

- Normalization satisfied

$P(R) = 0.3$ Probability of radio emission detected from SN, regardless of type

- 2 option variable

$P(II|R) = 0.75$ - Probability that SN is type II GIVEN Radio detected

$P(Ib|R) = 0.25$ - Probability that SN is type Ib

What is $P(Ia|R) = ?$ 0. → Normalization rule!

3 Questions, with neighbors

1.) $P(R \mid I_b)$ probability that radio is detected GIVEN that it is I_b

2.) $P(R \mid I_a \text{ or } I_b)$ probability of Radio Given type I

3.) $P(I_a \mid \neg R)$ probability of I_a given no radio

Answers:

1.) $$P(R \mid I_b) = \frac{P(I_b \mid R)P(R)}{P(I_b)} = \frac{(0.25)(0.3)}{0.2} = 0.375$$

2.) $$P(R \mid I_a \text{ or } I_b) = \frac{P(I_a \text{ or } I_b \mid R)P(R)}{P(I_a \text{ or } I_b)}$$

$$= \overset{0}{\cancel{\frac{P(I_a \mid R)P(R)}{P(I_a \text{ or } I_b)}}} + \frac{P(I_b \mid R)P(R)}{P(I_a \text{ or } I_b)} = \frac{(0.25)(0.3)}{0.5}$$

$$= 0.15$$

Note: $P(R \mid I_a \text{ or } I_b) \neq P(R \mid I_a) + P(R \mid I_b)$

~~3.) P~~

3.) $$P(I_a \mid \neg R) = \frac{P(\neg R \mid I_a)P(I_a)}{1 - P(R)}, \quad P(\neg R \mid I_a) = 1 - \cancelto{0}{P(R \mid I_a)}$$

$$= \frac{1(0.3)}{0.7} = 0.43$$

[Notebook Demo of Answers]

3 Prisoners ~~Dilem~~ Problem [For Piazza]

3 prisoners in jail: A, B, C

One has randomly been selected for release.

The warden knows, but cannot say.

A asks the warden, "~~Tell me if B or C will be released~~

" If B is to be released, tell me not C;

~~If~~ C, then not B;

if me, then flip a coin and tell me which one is not released."

Warden says "Not B", ~~after some thought i a coin fli~~

A thinks his chances are now 50/50 for release, between A, C.

~~B~~ A tells C. C smirks thinks he now has 2/3 chance.

WHo is right?

A, B, C $\equiv$ released $\qquad P(A) = P(B) = P(C) = 1/3$

b $\equiv$ told not B, c $\equiv$ told not C

$$P(A|b) = \frac{P(b|A)\,P(A)}{P(b)} =$$

$$P(b|A) = 1/2, \quad P(b|B) = 0, \quad P(b|C) = 1$$

~~$\Rightarrow P(A|b) = P(b|A$~~

$$P(b) = P(b|A)P(A) + P(b|B)P(B) + P(b|C)P(C)$$

$$\Rightarrow P(A|b) = \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 0 + 1/3} = \frac{1/6}{3/6} = 1/3 \quad !$$

A gets no information about his fate, but C does!

Maximum Likelihood

or "Maximum a Posteriori Solution"

~~$P(\vec{X}) \Rightarrow \frac{dP}{dX} = 0$ is a local max or minimum.~~

Imagine we have a model $f(\vec{x}; \vec{\alpha})$

independent variable — parameters.

We have measured data $\vec{y}$ at $\vec{x}$.

We want to know

$$P(\vec{\alpha}|\vec{y}) = P(\vec{y}|\vec{\alpha})\frac{P(\vec{\alpha})}{P(\vec{y})}$$

probability of parameters after seeing data

Bayes Rule

$P(\vec{y}|\vec{\alpha})$ usually far easier to write down "Likelihood"

$P(\vec{\alpha})$ prior - what we think params are w/o seeing any data
or → seeing all possible datasets

$P(\vec{y})$ evidence - normalization factor
↳ important, hard to calculate.

↳ come back to this.

Let's say all the data is independent.

$$P(\vec{\alpha}|\vec{y}) = \frac{P(\vec{\alpha})}{P(\vec{y})} \prod_i P(y_i|\vec{\alpha}).$$

Lets also say $P(y_i|\vec{\alpha})$ is Gaussian, and we know the noise $\vec{\sigma}$

$$P(y_i|\vec{\alpha}) = \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\frac{1}{2}\left(\frac{y_i - f(x_i;\vec{\alpha})}{\sigma_i}\right)^2}$$

What is $P(\vec{\alpha})$? Might be bad, but $P(\vec{\alpha}) = U(\vec{\alpha}) = 1/V$

constant

$P(\vec{\alpha}|\vec{y})$ getting messy, hard to calc. We can always take its Log:

$$\ln P(\vec{\alpha}|\vec{y}) = \sum_i \ln(P(y_i|\vec{\alpha})) + K$$

$$= \sum_i \left[-\frac{1}{2}\left(\frac{y_i - f(x_i;\vec{\alpha})}{\sigma_i}\right)^2 - \ln(\sqrt{2\pi}\,\sigma_i)\right] + K$$

That looks familiar!

$$\chi^2 \equiv \sum_i \left(\frac{y_i - f(x_i;\vec{\alpha})}{\sigma_i}\right)^2$$

$$\Rightarrow \ln P(\vec{\alpha}|\vec{y}) \propto -\frac{1}{2}\chi^2 + K$$

~~when~~ wrt $\vec{\alpha}$.

So maximizing $\ln P$ is same as minimizing χ^2.

what we do to find the mode of the distro

But what assumptions?

- i.i.d. data
- Gaussian Noise
- $P(\vec{\alpha}) = $ Uniform.

Often not the case!

What if we had 1D Gaussian data? Want μ?

$$\Rightarrow \ln P(\mu|\vec{x}) \propto -\frac{1}{2}\sum_i \left(\frac{x_i - \mu}{\sigma}\right)^2$$

num datapoints

$$\frac{d\ln P}{d\mu} = +\sum_i x_i - \mu = \sum_i x_i - N\mu = 0$$

$$\Rightarrow \mu = \frac{1}{N}\sum_i x_i$$

Max like estimate of μ ~~mean~~ is sample mean!