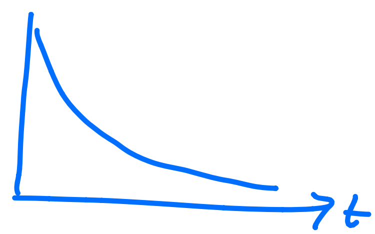
Probability and inference of parameters

Mackay chapter 3 and lecture 10
Sivia chapter 2
"What is Bayesian statistics" paper Eddy

a bacterial mutation wait times

- Measure times at which bacteria mutate
- mutation times t [0, 20] minutes

  it's a exponential decay



- all bacterium mutate independently

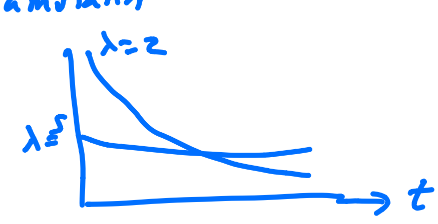
- Take measurements 20 minutes

  N = 6   1.2, 2.1, 3.4, 4.1, 7, 11 minutes

- we assume that the probability a bacterium mutates follows an exponential distribution

  T = time to wait to observe a mutation

  $P(T=t) \quad e^{-t/\lambda}$

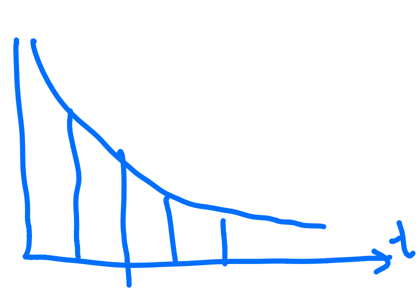


Q: Given the data ($N=6$), what can we say about the value of λ?

we want to infer λ

→ How? break rooms

* exponential $P(t|\lambda) = \frac{1}{Z(\lambda)} e^{-t/\lambda}$



* observations $N=6$

1.2, 2.1, 3.4, 4.1, 7, 11

$D_{KL}(E_\lambda \| O)$

calculate $D_{KL}(\lambda)$ for different values of λ

and optimize

$$\lambda_{mean} = \frac{1.2+2.1+3.4+4.1+7+11}{6} = 4.97.$$

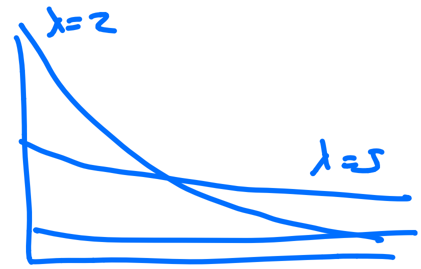
Exponential

$$P(t \mid \lambda, a, b) = \begin{cases} \frac{1}{Z(\lambda)} e^{-t/\lambda} & \text{if } a \le t \le b \\ 0 & \text{otherwise.} \end{cases}$$

$$Z(\lambda) = \int_a^b e^{-t/\lambda}\, dt$$ normalization constant.

$$= \lambda \left[e^{-a/\lambda} - e^{-b/\lambda} \right]$$

$$= \lambda \left[1 - e^{-20/\lambda} \right]$$



$$\boxed{P(\text{data} \mid \lambda)} = \prod_{i=1}^{f} P(t_i \mid \lambda)$$ (all are independent)

$$= \prod_{i=1}^{f} \frac{e^{-t_i/\lambda}}{Z(\lambda)} = \frac{e^{-\sum_i t_i/\lambda}}{Z^f(\lambda)}$$

$$= P(t_1 \ldots t_f \mid \text{exponential}, \lambda)$$

$$P(\text{data} \mid \lambda) \xrightarrow{\text{what can we say about?}} P(\lambda \mid \text{data})$$

forward probability

we can measure

describes outcome of random variable.

Inverse probability

quantity not directly measured

Bayes theorem

$$P(A,B) = P(A \mid B)\, P(B)$$
$$= P(B \mid A) \cdot P(A)$$

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

$$P(\lambda \mid \text{data}) = \frac{\overset{\text{likelihood}}{P(\text{data} \mid \lambda)} \cdot \overset{\text{prior}}{P(\lambda)}}{\underset{\text{evidence}}{P(\text{data})}}$$

posterior

$P(\text{data} \mid \lambda) \longrightarrow$ We know it

depends on data and hypothesis

"likelihood"

probability of data given the hypothesis and λ.

$P(\lambda) \longrightarrow$ prior probability of λ

"not an estimate"

MAXENT $P(\lambda) = \underline{1}$

$P(\text{data}) \longrightarrow$ evidence.

a constant that we can ignore for now.

$$P(\text{data}\mid\lambda) = \frac{e^{-\sum_i t_i/\lambda}}{Z^6(\lambda)}$$

$$P(\lambda\mid\text{data}) = \frac{P(\text{data}\mid\lambda)\cdot P(\lambda)}{P(\text{data})}$$

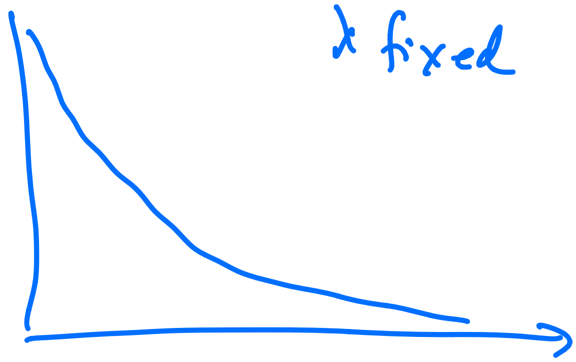
$$\propto \frac{e^{-\sum_i t_i/\lambda}}{Z^6(\lambda)}\cdot P(\lambda)$$

$$\propto \frac{e^{-\sum_i t_i/\lambda}}{Z^6(\lambda)}$$

$\boxed{\text{easy!}}$

$P(t|\lambda)$

λ fixed

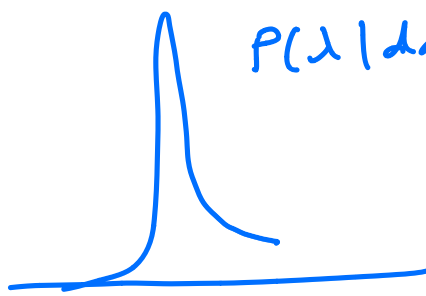


data fixed

$P(\lambda|\text{data}) \propto \prod_i P(t_i|\lambda)$

- best value so far
- variance

$P(\lambda|\text{data})$



as $N\uparrow$
your certainty
about λ increases

$P(\lambda)$ prior

Another example

the effectiveness of a vaccine

malaria vaccine

f = prob. of being malaria free after 1

N tested, n are malaria free after

1 year. each individual f, $(1-f)$

$$P(n \mid N f) = \binom{N}{n} f^n (1-f)^{N-n} = \frac{N!}{n!\,(N-n)!} f^n (1-f)^n$$

binomial distribution

$$P(f \mid n, N) = \frac{P(n \mid N f) \cdot P(f)}{P(n, N)}$$

bayes theorem.

$P(f) = 1$ by MAXENT principle.

$P(n, N)$ = prob that n out of N are malaria free no matter f.

$$P(n,N) = \int_f P(n,N)\, P(f)\, df$$

$$= \frac{N!}{n!\,(N-n)!} \underbrace{\int_0^1 f^n (1-f)^{N-n}\, df}_{\text{beta integral}}$$

$$= \frac{N!}{n!\,(N-n)!} \cdot \frac{n!\,(N-n)!}{(N+1)!} = \frac{1}{N+1}$$

$$P(f \mid n N) = \frac{N!}{n!\,(N-n)!} \cdot \frac{f^n (1-f)^{N-n}}{\frac{1}{N+1}} = \frac{(N+1)!}{n!\,(N-n)!} f^n (1-f)^{N-n}$$

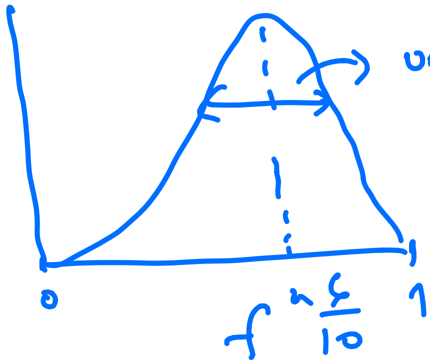
① $$P(n \mid N, f) = \frac{N!}{n!\,(N-n)!} f^n (1-f)^{N-n}$$ n is the variable f is fixed

② $$P(f \mid n N) = \frac{(N+1)!}{n!\,(N-n)!} f^n (1-f)^{N-n}$$ f is the variable N, n are fixed

<u>this is not a binomial distribution</u>

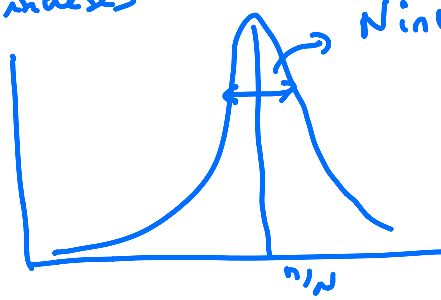
$N=10$

$n=6$



uncertainty due to $N=10$

$N \uparrow$ the evidence increases



N increases less uncertainty