Math 152

2-26-09 and March 3 3/3/09

Bayesian Parameter Estimation

Θ, the parameter in question

is treated as a random variable

with "prior" distribution $f_\theta(\theta)$.

We pick the prior $f_\theta(\theta)$ to represent

our prior opinion/belief/state of knowledge

about Θ before observing the data.

~~When~~ The distribution of X

is conditional on Θ.

i.e. given "$\Theta = \theta$"

the data have the distribution

$f_{X|\theta}(x|\theta)$.

and the joint distribution of

X and Θ is

$$f_{X,\theta}(x,\theta) = f_{X|\theta}(x|\theta)\, f_\Theta(\theta)$$

and the marginal of X

is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,\theta}(x \mid \theta)\, f_\theta(\theta)\, d\theta.$$

(~~these~~ might be a multiple integral)

We have then

$$f_{\theta|X}(\theta \mid X) = \frac{f_{X,\theta}(x,\theta)}{f_X(x)}$$

$$= \frac{f_{X|\theta}(x \mid \theta)\, f_\theta(\theta)}{\int f_{X,\theta}(x \mid \theta)\, f_\theta(\theta)\, d\theta}$$

and this is the "Posterior" distribution representing what is known about θ having observed X.

It is helpful to remeber.

$f_{\Theta|X}(\theta|x)$ is proportional to

$$f_{X|\Theta}(x|\theta) \cdot f_{\Theta}(\theta)$$

i.e (Posterior) is proportional to (Likelihood)$\cdot$(Prior.)

e.g. Poisson.

The unknown parameter λ has a prior $f_{\Lambda}(\lambda)$.

data:

$X_1, \ldots, X_n$ i.i.d.

Given $\Lambda = \lambda$

$$f_{X_i|\Lambda}(x_i|\lambda) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \qquad x_i = 0, 1, \ldots$$

and with $X = (X_1, \ldots, X_n)$

$$f_{X|\Lambda}(\vec{x}, \lambda) = \frac{\lambda^{\Sigma x_i} e^{-n\lambda}}{x_1! \cdots x_n!} \qquad \text{by indepen}$$

So

$$f_{\Lambda|X}(\lambda|\vec{x}) = \frac{\dfrac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} f_\Lambda(\lambda)}{\int \dfrac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod x_i!} f_\Lambda(\lambda)\, d\lambda}$$

$$= \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} f_\Lambda(\lambda)}{\int \lambda^{\sum x_i} e^{-n\lambda} f_\Lambda(\lambda)\, d\lambda}$$

e.g. Asbestos fiber counts.

$n=23$:

31	29	19	18	31	28
34	27	34	30	16	18
26	27	27	18	24	22
28	24	21	17	24	

$\cancel{\hat{\lambda} = 24.9}$ $\bar{X} = 24.9.$

① Before seeing the data, investigator believes: $\mu_1 = 15 \quad \sigma = 5$

And a Gamma distribution with this mean and s.d. has the following convenient mathematical property:

The posterior, in this (Poisson) situation will also be Gamma.

~~r.v. normal~~

We know from our earlier work that the 1st and second moments of a $X \sim \Gamma(\alpha, v)$ are related

$$f_X \sim \frac{v^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-vx}$$

to the parameters α, v by.

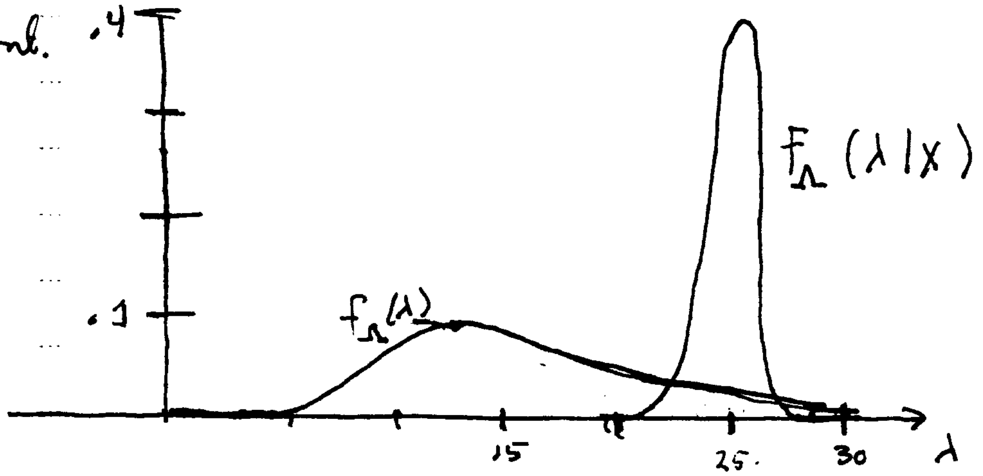
$$v = \frac{\mu_1}{\mu_2 - \mu_1^2}$$

$$\alpha = v\mu_1$$

and we have $\mu_2 = \mu_1^2 + \sigma^2 = 250$

so $v = 0.6$
$\alpha = 9.$

① cont.



$$f_\Lambda(\lambda) = \frac{v^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-v\lambda} \qquad \begin{array}{l} \alpha = 8 \\ v = .6. \end{array}$$

The investigator is in for a surprise when he sees the data.

$$f_{\Lambda|X}(\lambda|x) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \frac{v^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-v\lambda}}{\int_0^\infty \lambda^{\sum x_i} e^{-n\lambda} \frac{v^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-v\lambda} d\lambda}$$

$$= \frac{\lambda^{\sum x_i + \alpha - 1} e^{-(n+v)\lambda}}{\int_0^\infty \lambda^{\sum x_i + \alpha - 1} e^{-(n+v)\lambda} d\lambda}.$$

$$= C \cdot \lambda^{\alpha' - 1} \cdot e^{-v'\lambda}$$

where $\alpha' = \sum x_i + \alpha = 582$ and $v' = n + v = 23.6$.

We see that the posterior is also a Gamma so

$$C = \frac{(v')^{\alpha'}}{\Gamma(\alpha')}$$

and we don't need to do the integral on the bottom.

posterior mean $= \frac{\alpha'}{v'} = 24.7.$

posterior mode $= 24.6.$

$=$ max value of $\Gamma(\alpha', v') = \frac{\alpha' - 1}{v'}$

$$\sigma^2_{post} = \frac{\alpha'}{(v')^2} = 1.04.$$

$$\sigma_{post} = 1.02.$$

Bayesian 90% confidence interval

$\simeq [23.02, 26.34]$

5^{th} percentile to 95^{th} percentile of $\Gamma(\alpha', v')$.

alternatively: High posterior density interval.

②. Investigator believes $\lambda \le 100$
but ~~has no~~ expresses no opinion about a value between $[0, 100]$.

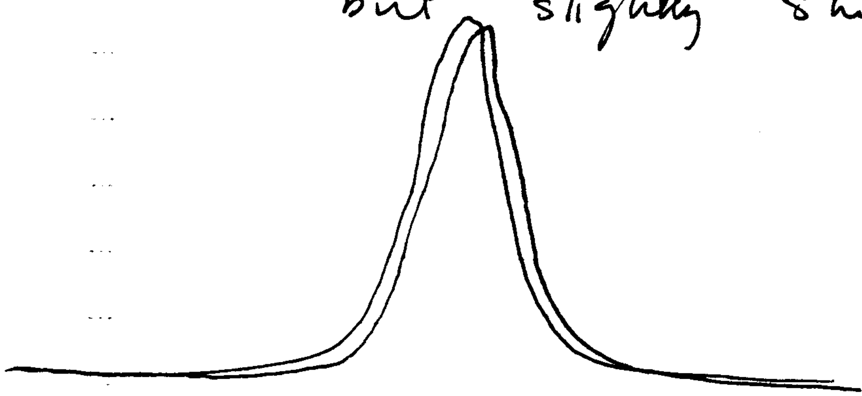
i.e.

$$f_\Lambda(\lambda) = \begin{cases} 1/100 & 0 \le \lambda \le 100 \\ 0 & \text{else} \end{cases}$$

and

$$f_{\Lambda|X}(\lambda|x) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \cdot \frac{1}{100}}{\frac{1}{100}\int_0^{100} \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}\, d\lambda}$$

The graph is similar to 1st investigator's posterior graph.
but slightly shifted to the right



posterior mean 25.
mode 24.9
s.d. 1.04
[5%, 95%] = [23.3, 26.7].

The maximum likelihood estimate was 24.9. and it is no coincidence that this is the same as ②'s posterior mode.

posterior $\propto$ (likelihood)(prior)

and the prior was constant on [0, 100].

so posterior and likelihood have the same maxima on [0, 100].

Comparing ① and ②, though the initial assumptions are different, the collected data forces similar conclusions.

~~remar~~

- The prior opinion of ① does have a slight effect, dragging the graph backwards a small amount.

- Since posterior is proportional to (likelihood)(prior).

note that ~~statistician~~ ①'s.

posterior is like a weighted average of likelihood with the weights a bit heavier on the left.

(looking at the interval near $\lambda = 25$ in which the likelihood is non-negligible).