APPLICATIONS OF BAYES' THEOREM

C&PE 940, 21 September 2005

Geoff Bohling
Assistant Scientist
Kansas Geological Survey
geoff@kgs.ku.edu
864-2093

Notes, overheads, Excel example file available at

http://people.ku.edu/~gbohling/cpe940

Development of Bayes' Theorem

Terminology:

$P(A)$: Probability of occurrence of event A (*marginal*)
$P(B)$: Probability of occurrence of event B (*marginal*)
$P(A,B)$: Probability of simultaneous occurrence of events A and B (*joint*)
$P(A|B)$: Probability of occurrence of A *given* that B has occurred (*conditional*)
$P(B|A)$: Probability of occurrence of B *given* that A has occurred (*conditional*)

Relationship of joint probability to conditional and marginal probabilities:

$$P(A,B) = P(A|B)P(B) \quad \text{or} \quad P(A,B) = P(B|A)P(A)$$

So . . .

$$P(A|B)P(B) = P(B|A)P(A)$$

Rearranging gives simplest statement of Bayes' theorem:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Often, *B* represents an underlying *model* or *hypothesis* and *A* represents observable *consequences* or *data*, so Bayes' theorem can be written schematically as

$$P(model|data) \propto P(data|model)P(model)$$

This lets us turn a statement about the forward problem:

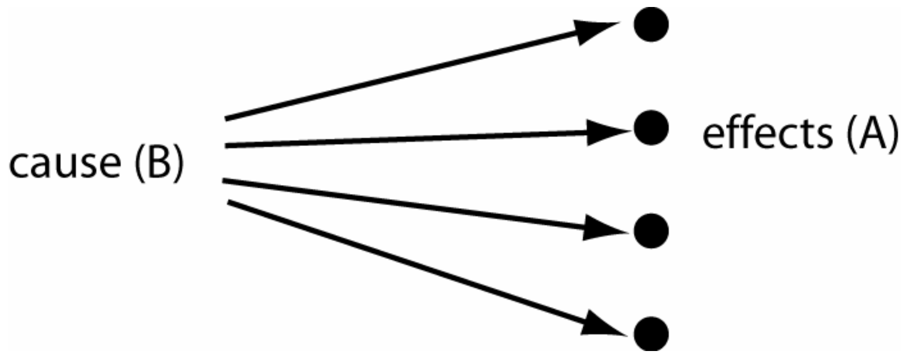
$P(data|model)$: probability of obtaining observed data given certain model

into statements about the corresponding inverse problem:

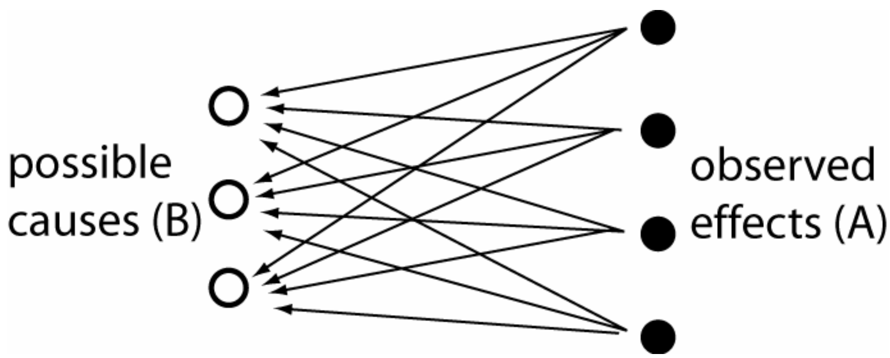
$P(model|data)$: probability that certain model gave rise to observed data

as long as we are willing to make some guesses about the probability of occurrence of that model, $P(model)$, prior to taking the data into account.

Or graphically, Bayes' theorem lets us turn information about the probability of different effects from each possible cause:



into information about the probable cause given the observed effects:



(Illustration styled after Sivia, 1996, Figure 1.1)

Assume that B_i represents one of n possible mutually exclusive events and that the conditional probability for the occurrence of A given that B_i has occurred is $P(A|B_i)$. In this case, the total probability for the occurrence of A is

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

and the conditional probability that event B_i has occurred given that event A has been observed to occur is given by

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)} = \frac{P(A|B_i)P(B_i)}{P(A)}.$$

That is, if we assume that event A arises with probability $P(A|B_i)$, from each of the underlying "states" B_i, $i=1,\ldots,n$, we can use our observation of the occurrence of A to update our *a priori* assessment of the probability of occurrence of each state, $P(B_i)$, to an improved *a posteriori* estimate, $P(B_i|A)$.

Discrete-Probability Example: Dolomite/Shale Discrimination Using Gamma Ray Log Threshold

Reservoir with dolomite "pay" zones and shale "non-pay" zones.

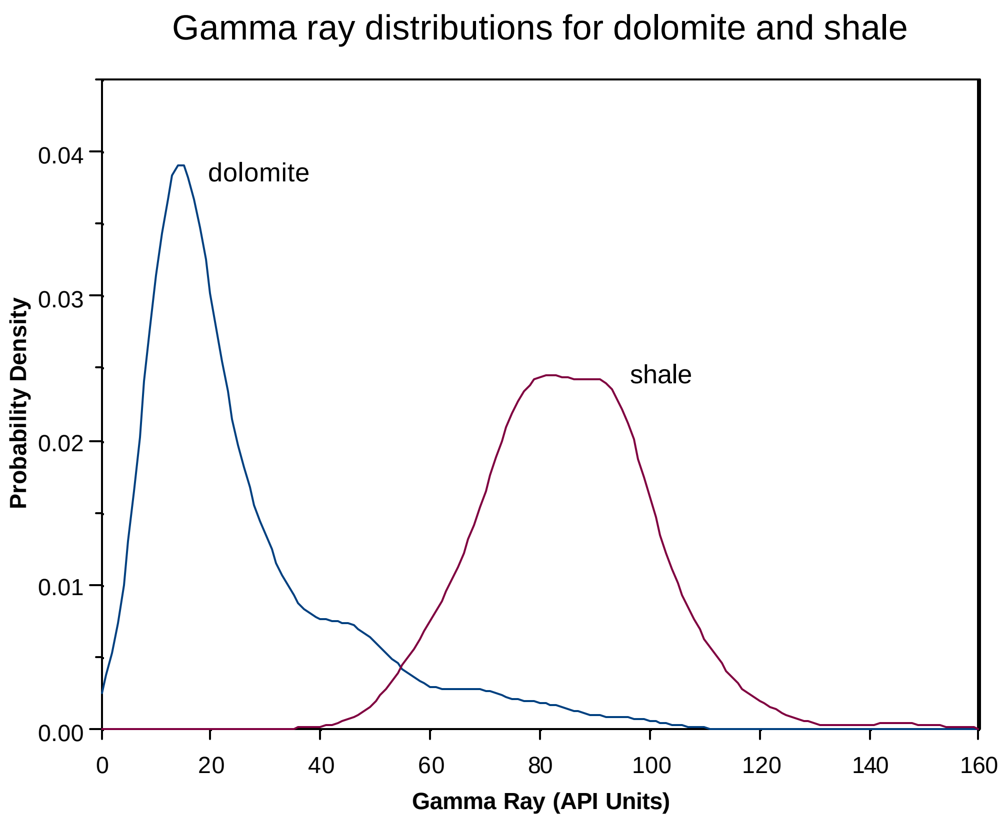
Gamma ray log: Measures natural radioactivity of rock; measured in API units

Shales: Typically high gamma ray (~110 API units) due to abundance of radioactive isotopes in clay minerals; somewhat lower in this reservoir (~80 API units) due to high silt content

Dolomite: Typically low gamma ray (~10-15 API units), but some "hot" intervals due to uranium

Can characterize gamma ray distribution for each lithology based on core samples from wells in field:

	Dolomite	Shale
Mean	25.8	85.2
Std. Dev.	18.6	14.9
Count	476	295



Will use these distributions to predict lithology from gamma ray in uncored wells, first using a simple rule:

- if *GammaRay* > 60, call the logged interval a shale
- if *GammaRay* < 60, call it a dolomite

Using Bayes' rule we can determine the posterior probability of occurrence of dolomite and shale given that we have actually observed a gamma ray value greater than 60. Let's define events & probabilities as follows:

A: $GammaRay > 60$
B_1: occurrence of dolomite
B_2: occurrence of shale
$P(B_1)$: prior probability for dolomite based on overall prevalence $\cong$ 60% (476 of 771 core samples)
$P(B_2)$: prior probability for shale based on overall prevalence $\cong$ 40% (295 of 771 core samples)
$P(A|B_1)$: probability of $GammaRay > 60$ in a dolomite = 7% (34 of 476 dolomite samples)
$P(A|B_2)$: probability of $GammaRay > 60$ in a shale = 95% (280 of 295 shale samples)

Then the denominator in Bayes' theorem, the *total probability* of A, is given by

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$$
$$= 0.07 * 0.60 + 0.95 * 0.40 = 0.422$$

If we measure a gamma ray value greater than 60 at a certain depth in a well, then the probability that we are logging a dolomite interval is

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{0.07*0.60}{0.422} = 0.10$$

and the probability that we are logging a shale interval is

$$P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A)} = \frac{0.95*0.40}{0.422} = 0.90.$$

Thus, our observation of a high gamma ray value has changed our assessment of the probabilities of occurrence of dolomite and shale from 60% and 40%, based on our prior estimates of overall prevalence, to 10% and 90%.

We can do simple sensitivity analysis with respect to prior probabilities. For example, if we take prior probability for shale to be 20% (meaning prior for dolomite is 80%), then get posterior probability of 77% for shale (23% for dolomite) if the gamma ray value is greater than 60 API units.

Continuous-Probability Example: Dolomite/Shale Discrimination Using Gamma Ray Density Functions

It is also possible to formulate Bayes' theorem using probability density functions in place of the discrete probabilities $P(A|B_i)$. We could represent the probability density function that a continuous variable, X, follows in each case as $f(x|B_i)$ or, more compactly, $f_i(x)$. Then

$$P(B_i|x) = \frac{f_i(x)P(B_i)}{\sum_{j=1}^{n} f_j(x)P(B_j)}.$$

That is, if we can characterize the distribution of X for each category, B_i, we can use the above equation to compute the probability that event B_i has occurred given that the observed value of X is x. For example, based on the observed distribution of gamma ray values for dolomites and shales, a gamma ray measurement of 110 API units almost certainly arises from a shale interval, because the probability density function for gamma ray in dolomites evaluated at 110 API units, $f_1(x=110)$, is essentially 0.

This form of Bayes' theorem lets us develop a continuous mapping from gamma ray value to posterior probability.

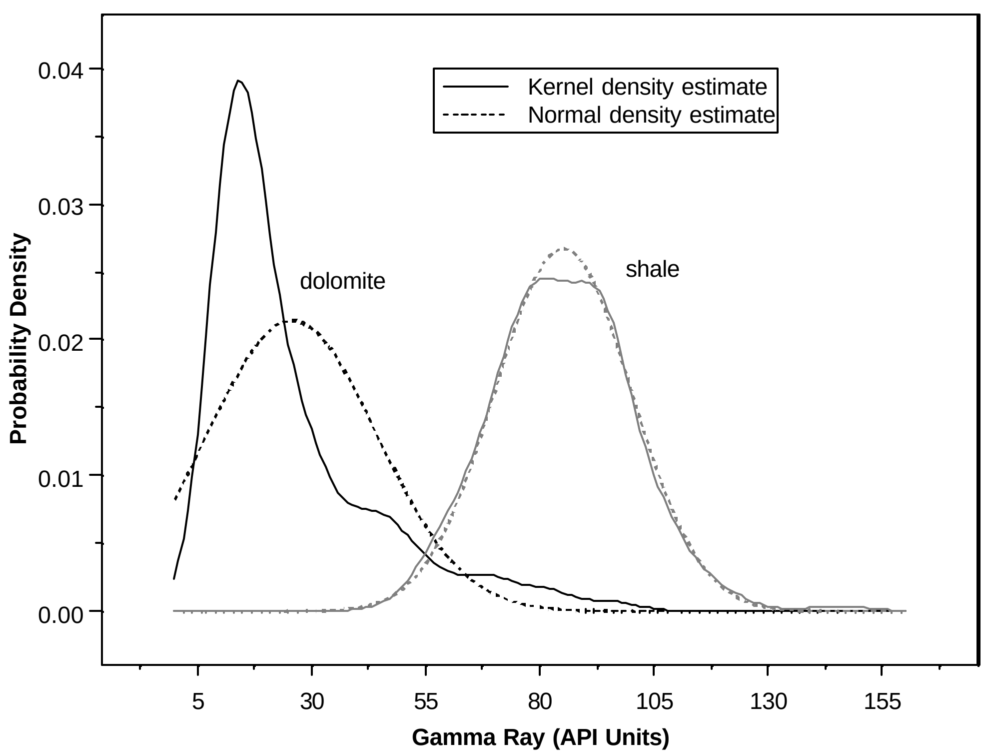
Shale/Dolomite Discrimination Using Normal Density Functions

	Dolomite (1)	Shale (2)
Mean ($\bar{x}$)	25.8	85.2
Std. Dev. (s)	18.6	14.9
Count	476	295

$$f_1(x) = \frac{1}{s_1\sqrt{2p}} \exp\left[-(x - \bar{x}_1)^2 / 2s_1^2\right]$$

$$f_2(x) = \frac{1}{s_2\sqrt{2p}} \exp\left[-(x - \bar{x}_2)^2 / 2s_2^2\right]$$

Normal Approximations for Gamma Ray Distributions



Let

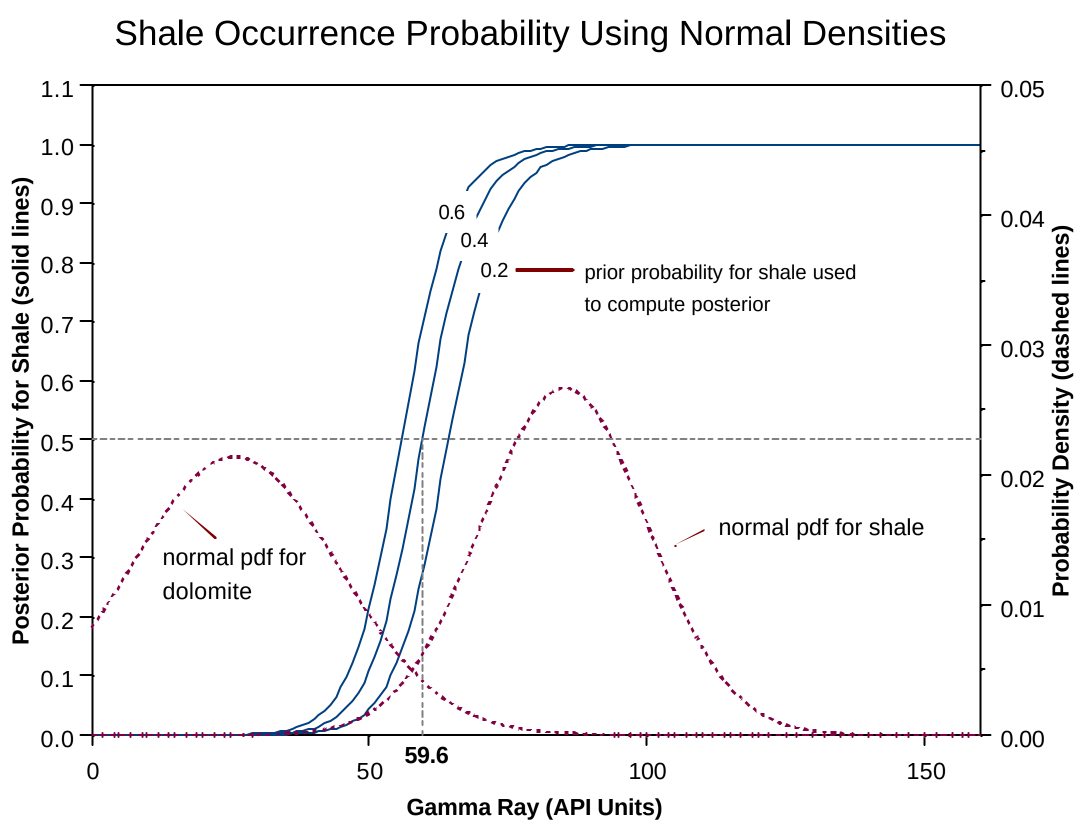
$q_2 = P(B_2)$ represent prior probability for shale
prior for dolomite is then $P(B_1) = 1 - q_2$

Let

$p_2(x) = P(B_2|x)$ represent posterior probability for shale
posterior for dolomite is then $P(B_1|x) = 1 - p_2(x)$

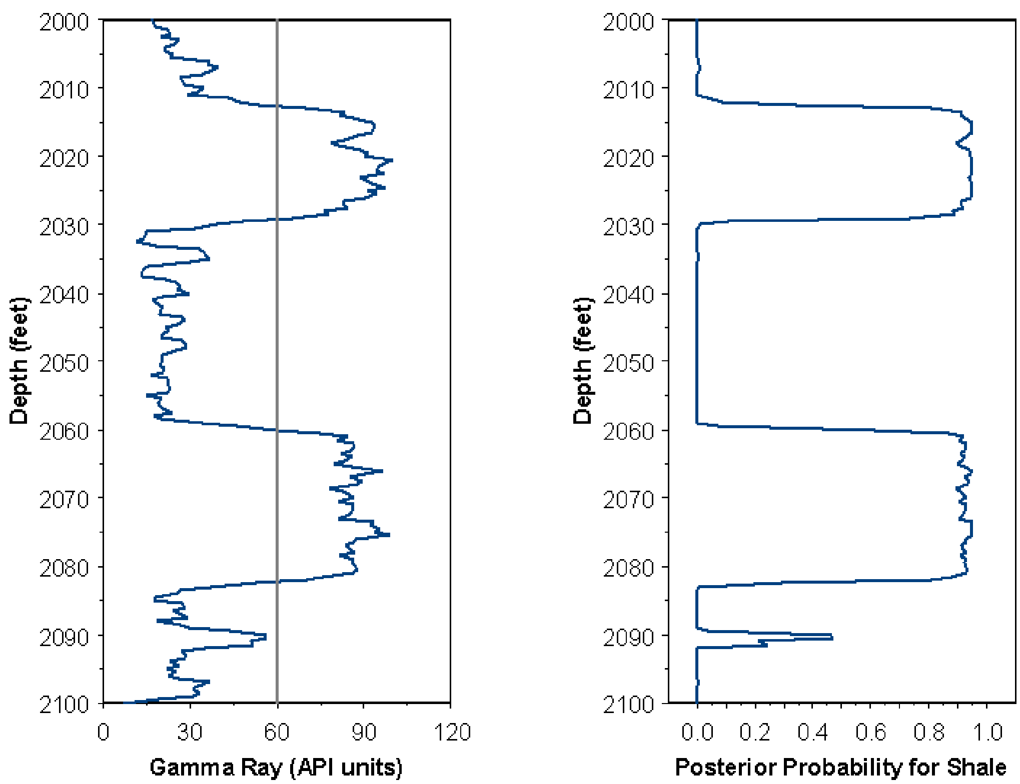
So, posterior probability for shale given that the observed gamma ray value = x is

$$p_2(x) = \frac{q_2 f_2(x)}{(1-q_2) f_1(x) + q_2 f_2(x)}$$



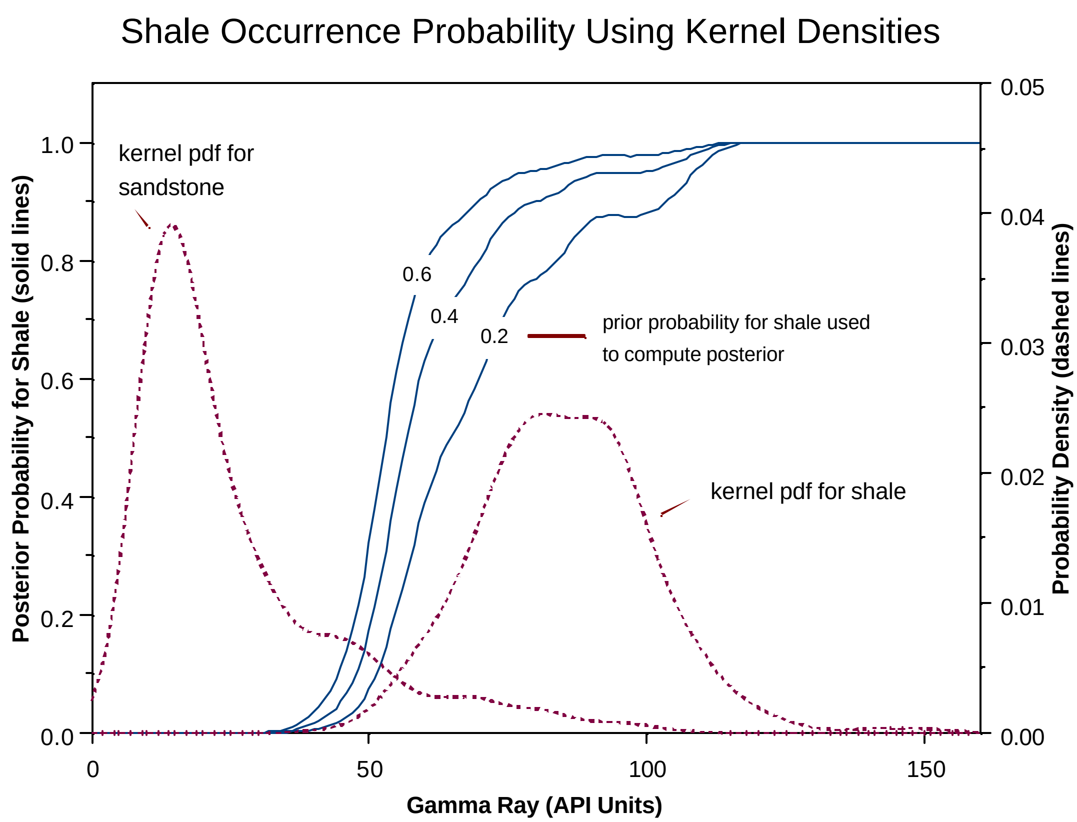
Bayes' rule allocation: Assign observation to class with highest posterior probability.

For "base case" prior of 40% for shale, 50% posterior probability point occurs at gamma ray of 59.6 – so Bayes' rule allocation leads to basically same results as thresholding at 60 API units. But now have means for converting gamma ray to continuous "shale probability" log.



Shale/Dolomite Discrimination Using Kernel Density Estimates

No need to restrict approach to just normal densities. Could use any other form of probability density function for each category, including the kernel density estimates shown initially:



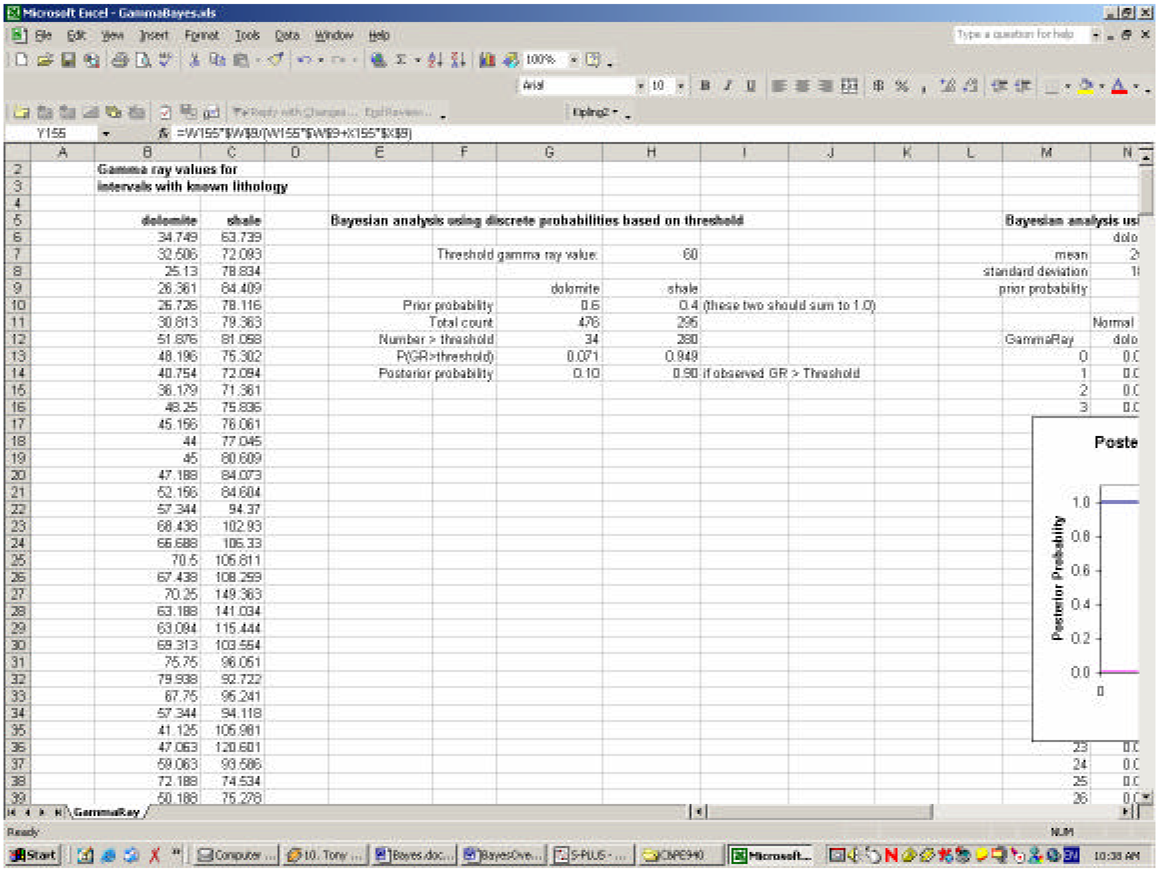
Relationship to Discriminant Analysis

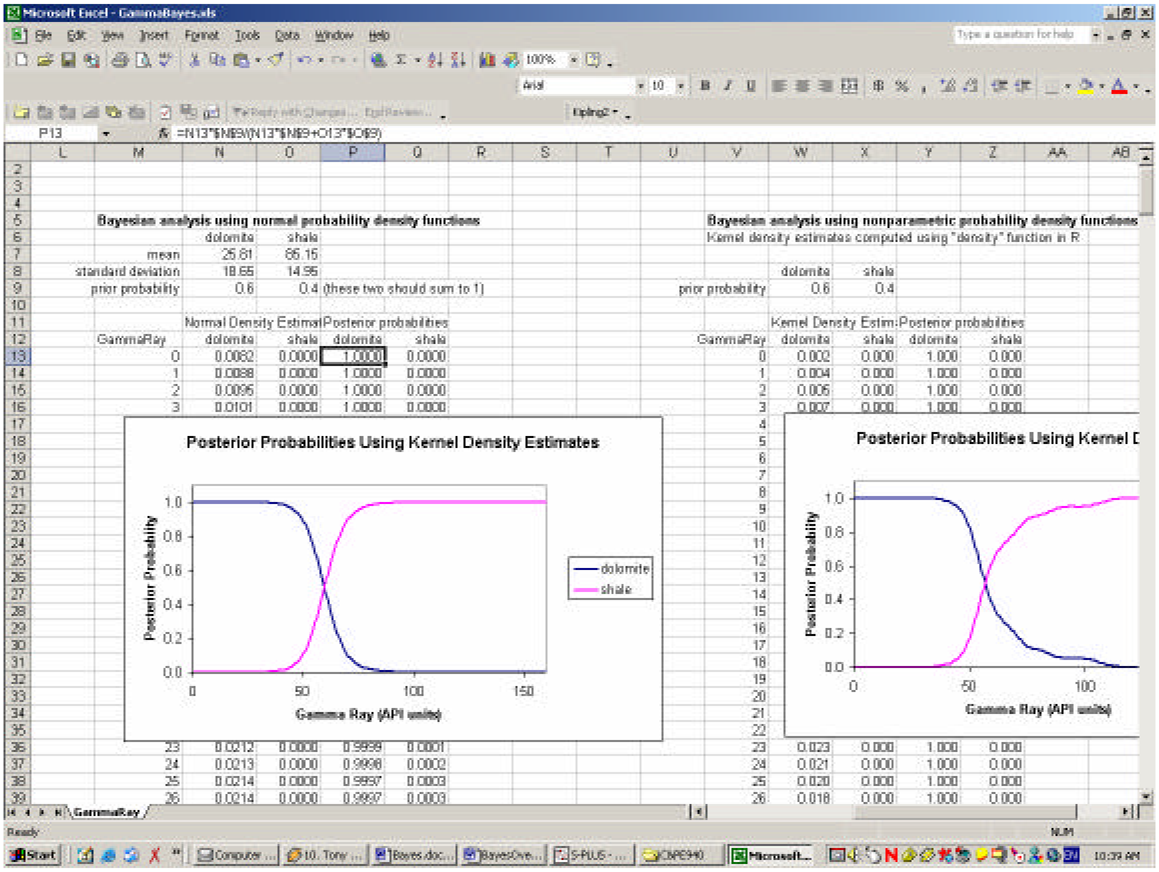
Could just as easily use multivariate density functions in Bayes' theorem. For example, could be discriminating facies based on a vector of log measurements, $\boldsymbol{x}$, rather than a single log.

If use multivariate normal density functions for each class, Bayes' rule allocation leads to classical *discriminant analysis.*

Assuming covariance matrices all equal for different classes leads to *linear discriminant analysis*: Bayes' rule allocation draws linear boundaries between classes in x space.

Assuming unequal covariance matrices leads to *quadratic discriminant analysis*: Bayes' rule allocation draws quadratic boundaries between classes.


Microsoft Excel - GammaBayes.xls
Y155
=W155*$W$9/(W155*$W$9+X155*$X$9)
Gamma ray values for
intervals with known lithology
dolomite shale
34.749 63.739
32.506 72.093
25.13 78.834
26.361 84.409
26.726 78.116
30.813 79.363
51.876 81.068
48.196 75.302
40.754 72.094
36.179 71.361
48.25 75.836
45.156 76.061
44 77.046
45 80.609
47.188 84.073
52.156 84.604
57.344 94.37
68.438 102.93
66.688 106.33
70.5 106.811
67.438 108.299
70.25 149.363
63.188 141.034
63.094 115.444
69.313 103.554
75.75 96.061
79.938 92.722
67.75 95.241
57.344 94.118
41.125 106.981
47.063 120.601
59.063 93.586
72.188 74.534
50.188 75.278
Bayesian analysis using discrete probabilities based on threshold
Threshold gamma ray value: 60
dolomite shale
Prior probability 0.6 0.4 (these two should sum to 1.0)
Total count 476 296
Number > threshold 34 280
P(GR>threshold) 0.071 0.949
Posterior probability 0.10 0.90 if observed GR > Threshold
mean
standard deviation
prior probability
GammaRay
Posterior Probability
GammaRay



Microsoft Excel - GammaBayes.xls
=N13*$N$9/(N13*$N$9+O13*$O$9)
Bayesian analysis using normal probability density functions
Bayesian analysis using nonparametric probability density functions
Kernel density estimates computed using "density" function in R
mean
standard deviation
prior probability
(these two should sum to 1)
Normal Density Estimat
Posterior probabilities
Kernel Density Estim
GammaRay
dolomite
shale
Posterior Probabilities Using Kernel Density Estimates
Posterior Probability
Gamma Ray (API units)
GammaRay