3.26pt

An Introduction to Bayes' Theorem

Ruedi Port

Genentech Inc.
Early Development PKPD

October 21, 2008

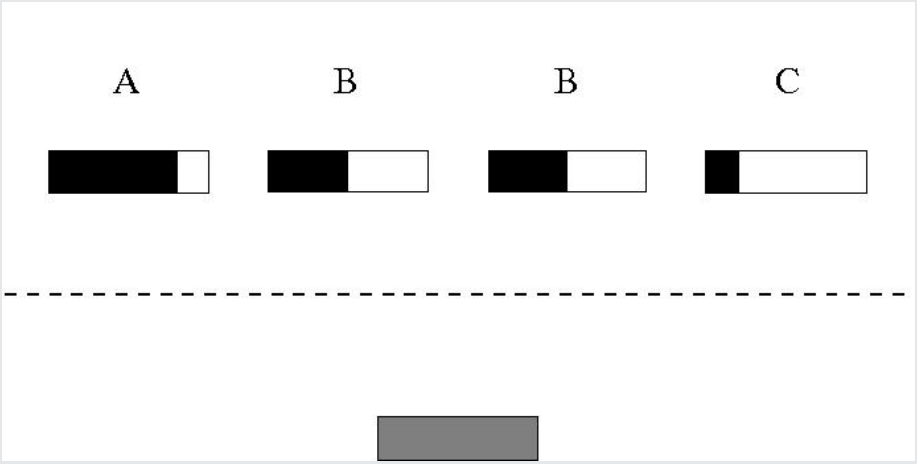
Introduction

The word “probability” can be used in two different ways. It may refer to the expected frequency of an event in a random experiment which is repeated many times, like the probability of getting face when tossing a coin one thousand times. On the other hand, in everyday life, it often describes our belief about something that is already a fact but which we don’t know for sure. For example, when the telephone is ringing, someone might say: “This is probably aunt Minnie who wants to ask...”. In this case, the word “probably” describes the degree to which we believe in a statement being right about something that is already a fact. The Bayesian approach offers a way to express the strengths of our beliefs in such statements in numbers. These numbers are usually derived by combining previous knowledge with recent observations. For example, the Bayesian probability of a medical diagnosis being right in a specific patient may be derived by combining information about the relative frequencies of several diseases which could all explain his/her symptoms with information from a recent lab test in this patient which, however, is also not completely conclusive because this lab test has a known large measurement error.

Thomas Bayes
1701? - 1761 (Tunbridge Wells, England)

Thomas Bayes was a presbyterian minister who was interested in mathematics. His famous theorem is mentioned as something quite obvious in the introduction to an assay that was found by a friend among his writings after his death. The concept can by explained by a hypothetical experiment as follows:

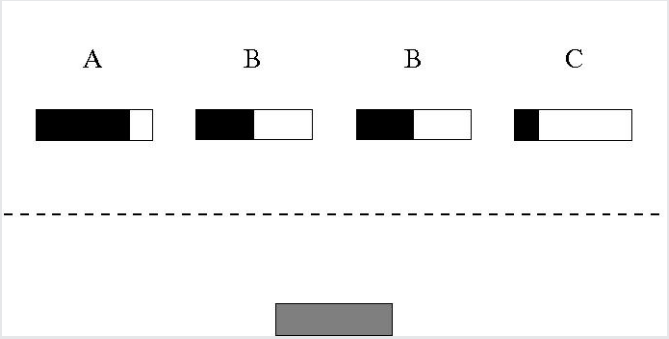
Urn experiment, prior probabilities



There are four urns looking exactly the same. One urn (“type A”) contains eight red and two white balls, two urns (“type B”) contain five red and and five white balls each, and one urn (“type C”) contains two red and eight white balls. We are given one of the four urns and are asked to guess what type it is: A, B, or C?

guess	probability of being right
“urn is type A”	0.25
“urn is type B”	0.50
“urn is type C”	0.25

Urn experiment, posterior probabilities after first observation



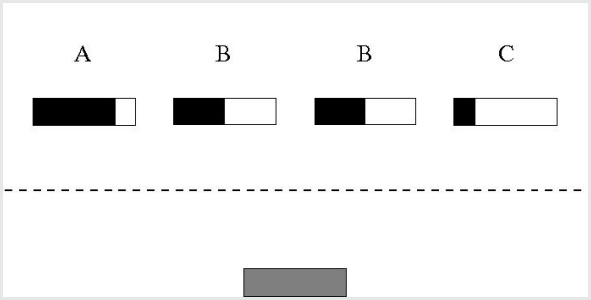
We make an observation by drawing a ball from the unknown urn; the ball is white. What is now the best guess?

		probability of occurring	probability of being right with this guess after a white ball has been drawn ("posterior probability")
urn is type A	**and** a white ball is drawn	$0.25 \cdot 0.20 = 0.05$	$0.05/0.50 = 0.10$
urn is type B	**and** a white ball is drawn	$0.50 \cdot 0.50 = 0.25$	$0.25/0.50 = 0.50$
urn is type C	**and** a white ball is drawn	$0.25 \cdot 0.80 = \underline{0.20}$	$0.20/0.50 = \underline{0.40}$
		0.50	1.00

The probabilities of occurring add up to 0.5 because there are three more possible cases where a red ball would have been drawn. The total probability of the "red-ball-cases" is also 0.5 because the total numbers of red and white balls, summed up over all urns, are the same.

The probabilities of being right by guessing a "white-ball-case" must add up to 1 because, after drawing a white ball, it is known that these three cases make up all possible cases.

Urn experiment, posterior probabilities after second observation



The ball drawn in the first observation is replaced and, after mixing, a ball is drawn for the second time; again, the ball is white. What is now the best guess?

case	probability of occurring	probability of being right with this guess after drawing a white ball two times (“posterior probability”)
“urn is type A **and** a white ball is drawn two times”	$0.25 \cdot 0.20 \cdot 0.20 = 0.010$	$0.010 / 0.295 = 0.034$
“urn is type B **and** a white ball is drawn two times”	$0.50 \cdot 0.50 \cdot 0.50 = 0.125$	$0.125 / 0.295 = 0.424$
“urn is type C **and** a white ball is drawn two times”	$0.25 \cdot 0.80 \cdot 0.80 = \underline{0.160}$	$0.160 / 0.295 = \underline{0.542}$
	0.295	1.000

Mammography for diagnosis of breast cancer

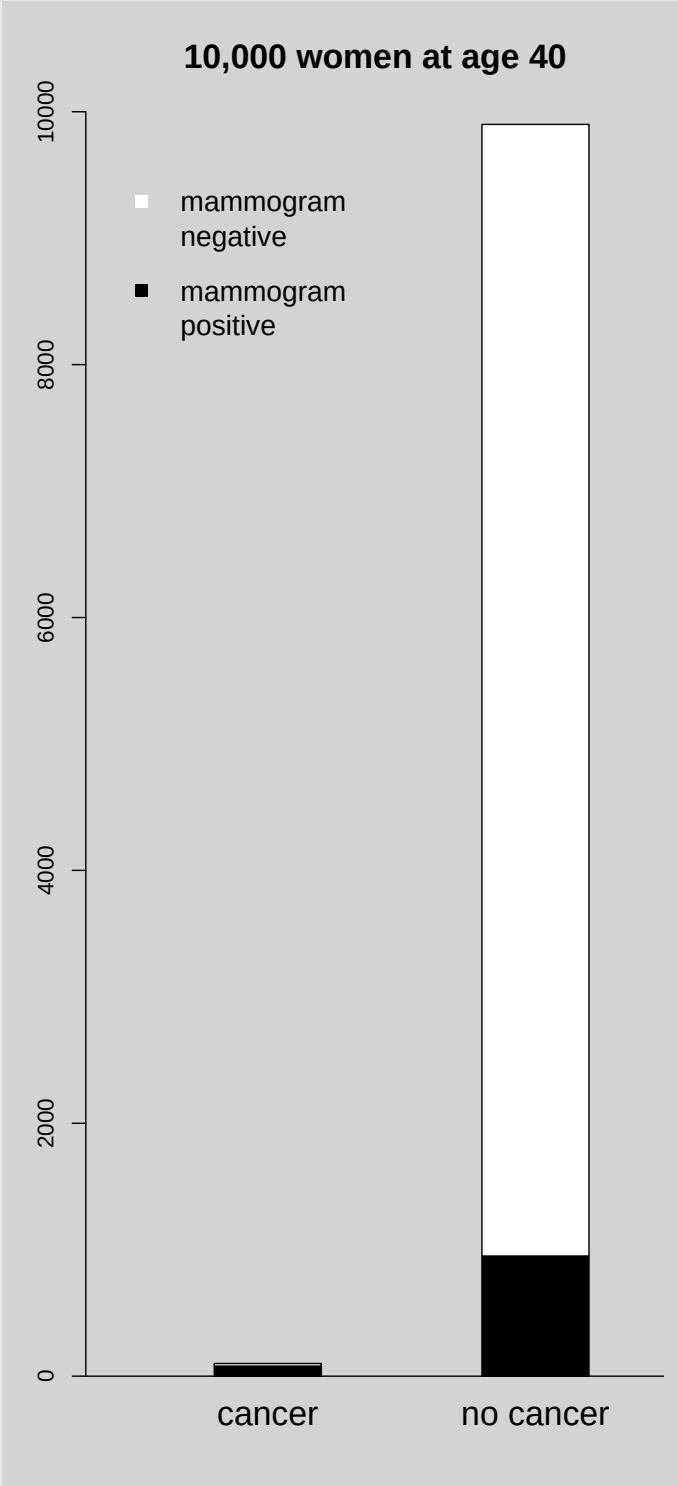
The prevalence of breast cancer in women at age 40 is 1%.

In women with breast cancer examined by mammography, the result is positive in 80% of cases.

In women without breast cancer, mammograms are positive in 9.6% of cases.

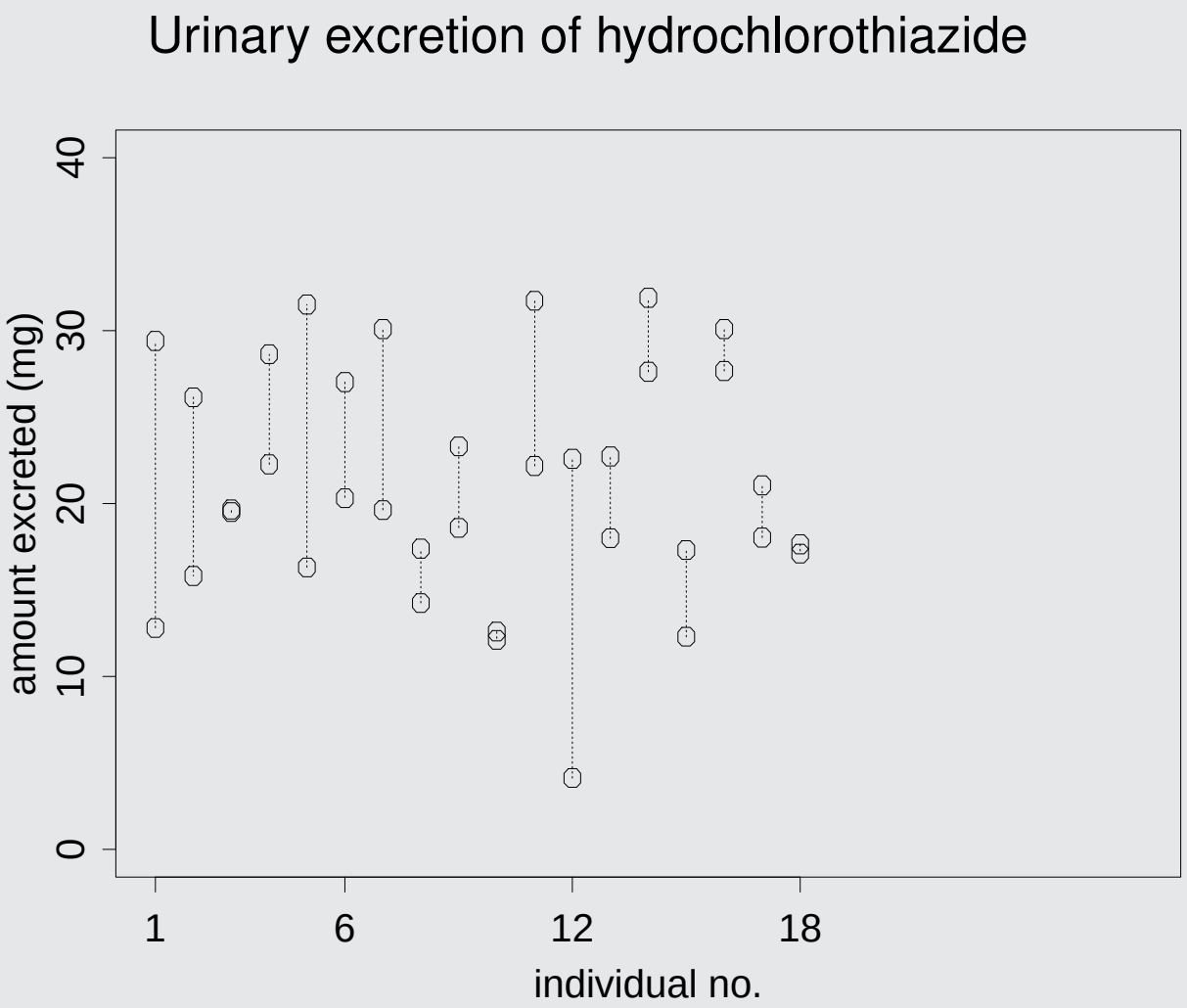
A 40 year old woman has a positive mammogram:

Is it cancer? Is it not?



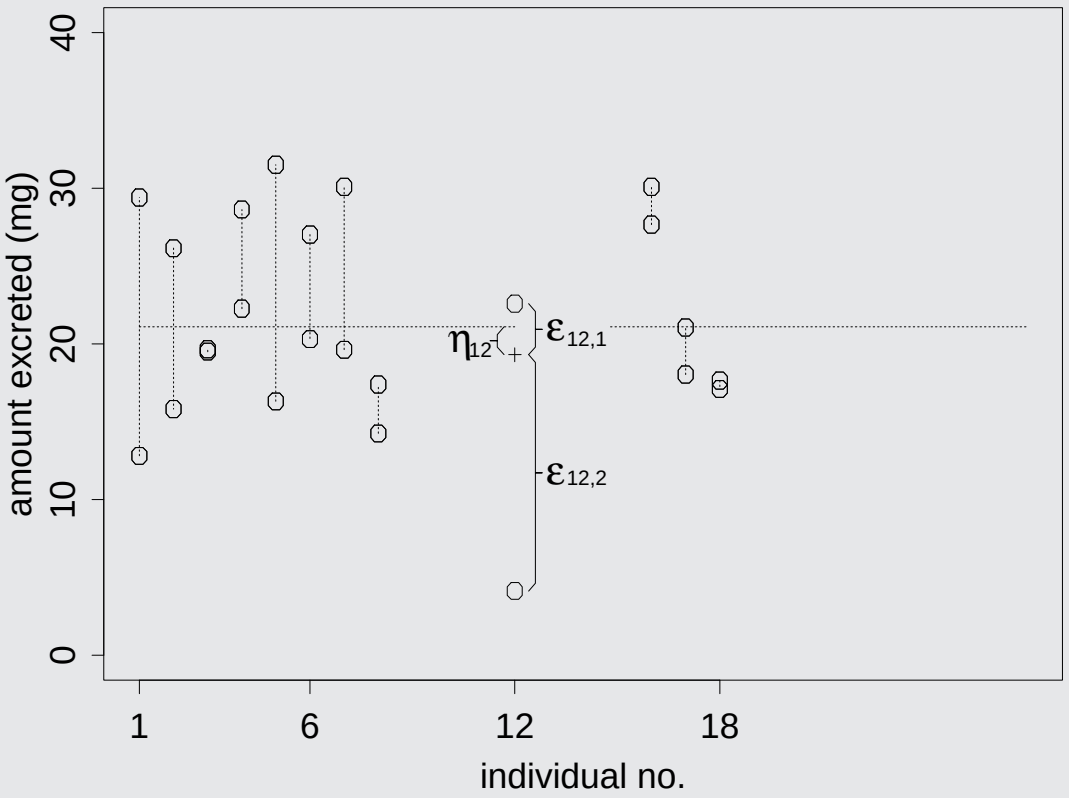
	Cancer	**No cancer**
Prior probabilities:		
	0.01	0.99
Probability of getting a positive mammogram:		
	$0.01 \cdot 0.8 = 0.008$	$0.99 \cdot 0.096 = 0.095$
Posterior probability when mammogram is positive:		
	$0.008/(0.008 + 0.095) = 0.078$	$0.095/(0.008 + 0.095) = 0.922$

Bayesian estimation of individual PK parameters



This drug was taken perorally by volunteers on two occasions, with a washout period in between, and the amount excreted in urine within 24 hours was determined. The dose was 50 mg. Each vertical line represents one individual; each dot represents one measurement.

The drug is not metabolized, so the amount excreted can be thought of as a measure of bioavailability. We may want to estimate each individual's expected mean amount excreted to assess individual bioavailability and decide about chronic dosing.



A parametric mixed-effects model ("population model"):

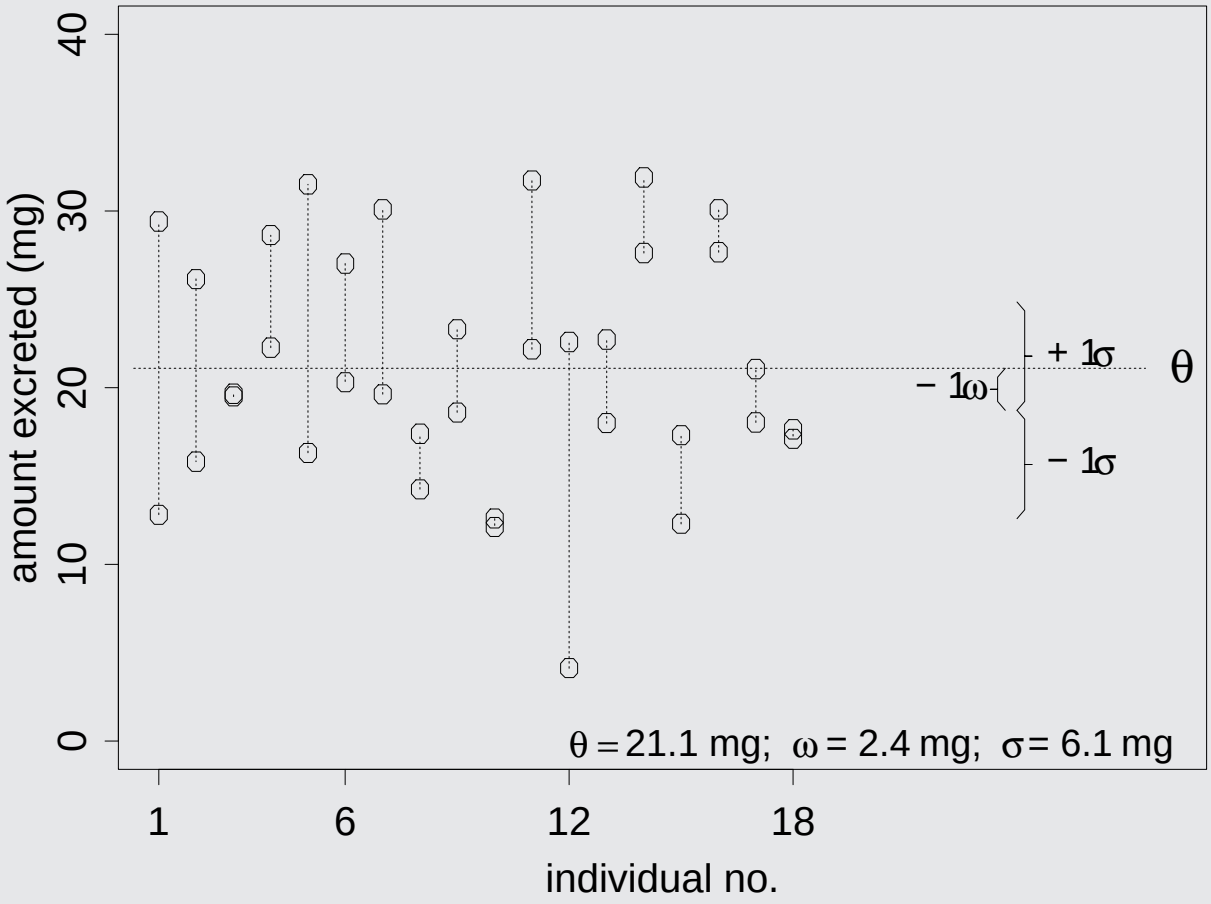
$$y_{ij} = \theta + \eta_i + \varepsilon_{ij}$$

fixed effect

random effects

The model describes each observation as the sum of a population mean, θ, a random deviation from the population mean, η_i, which is a characteristic of each individual, and an additional random deviation from the individual mean, ε_{ij}, which varies within each individual from occasion to occasion.

The population parameters to be estimated are:
the population mean, θ, the variance of η, ω^2, and the variance of ε, σ^2.

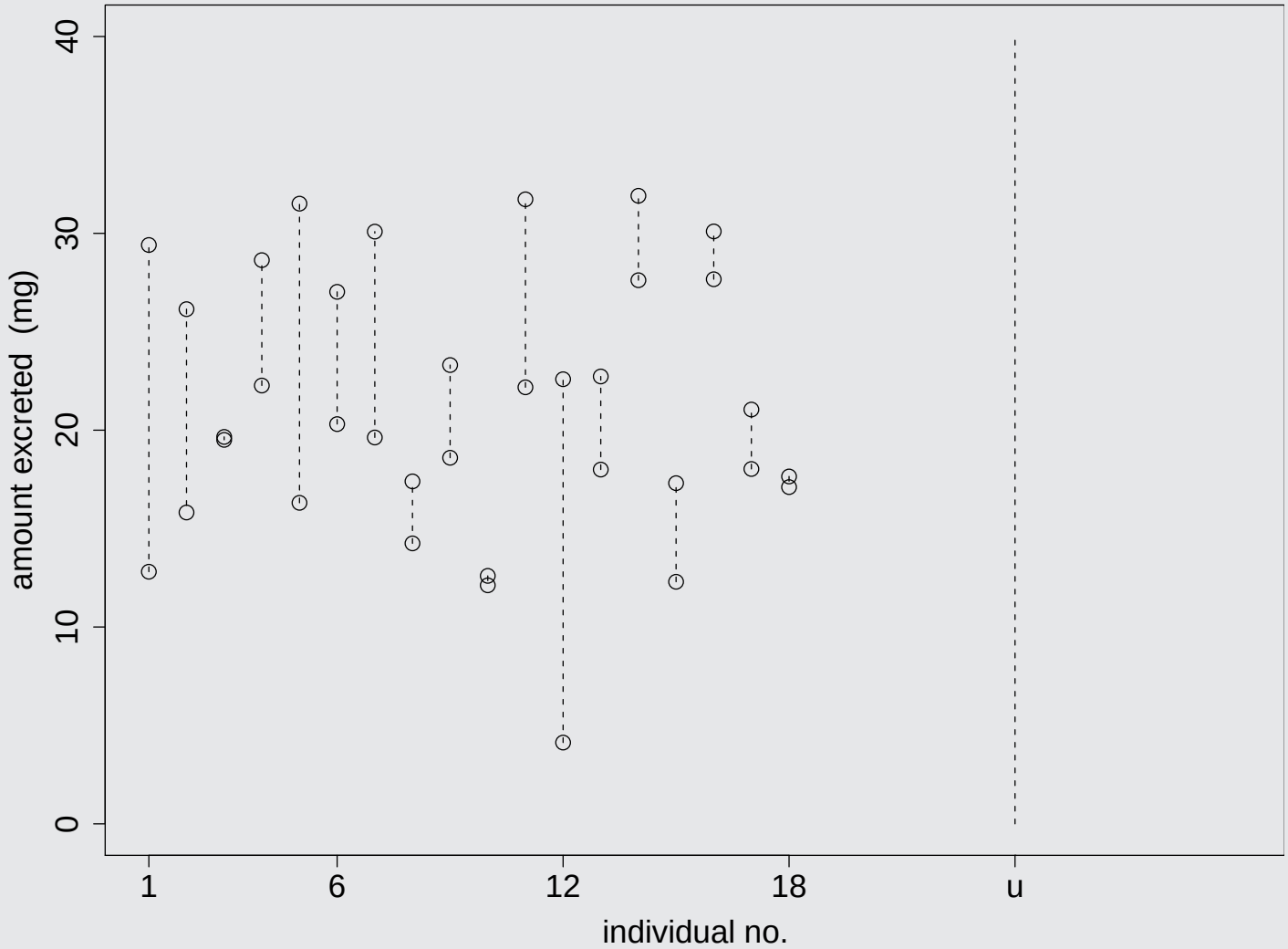


Result of fitting the mixed-effects model:

θ = 21.1 mg; ω^2 = 5.7 mg^2; σ^2 = 37.6 mg^2

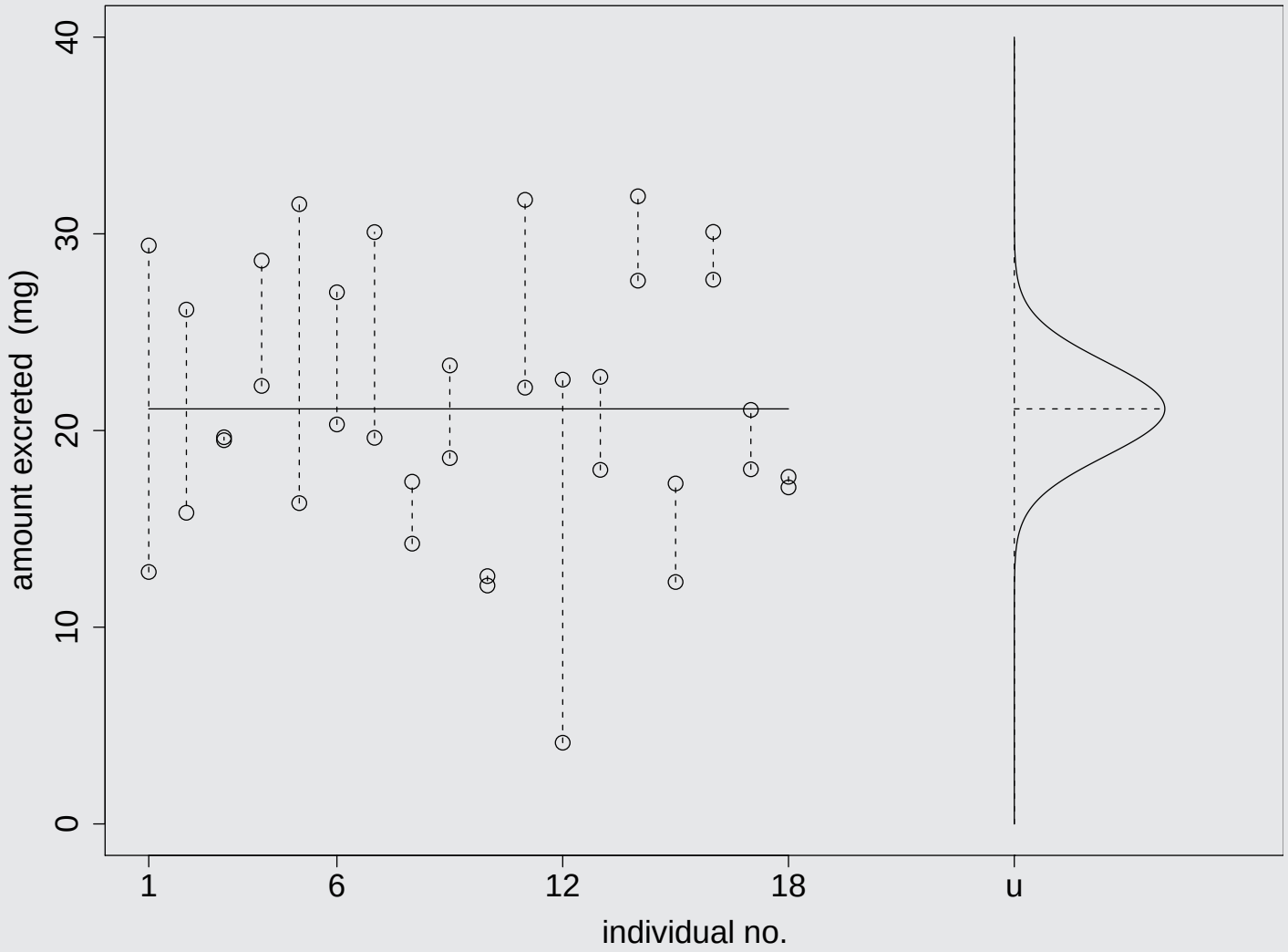
Most of the overall variability is attributed

to residual variability within individuals (σ^2).

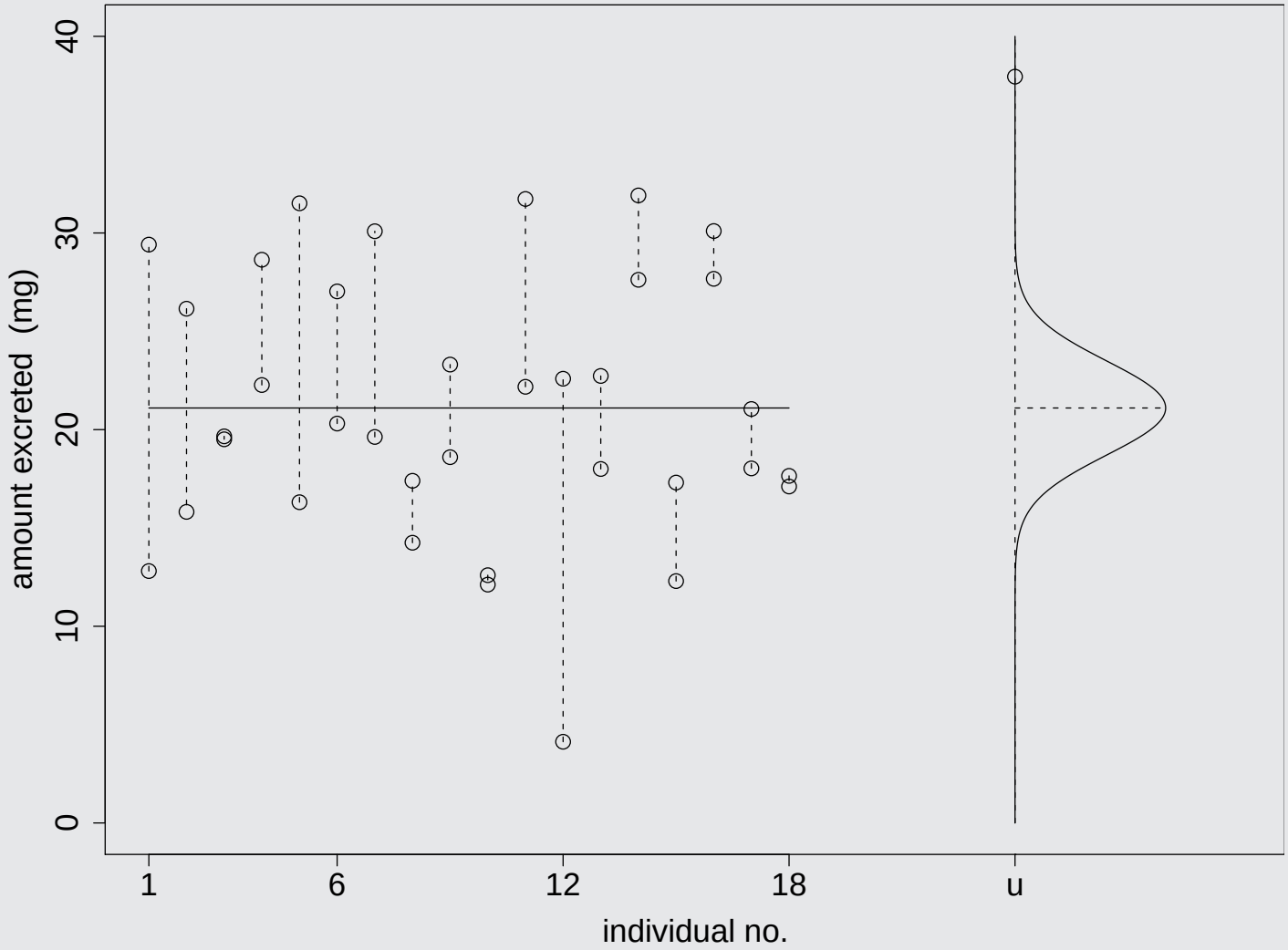


A new, unknown individual, “u”, comes in.
What can be expected regarding the individual

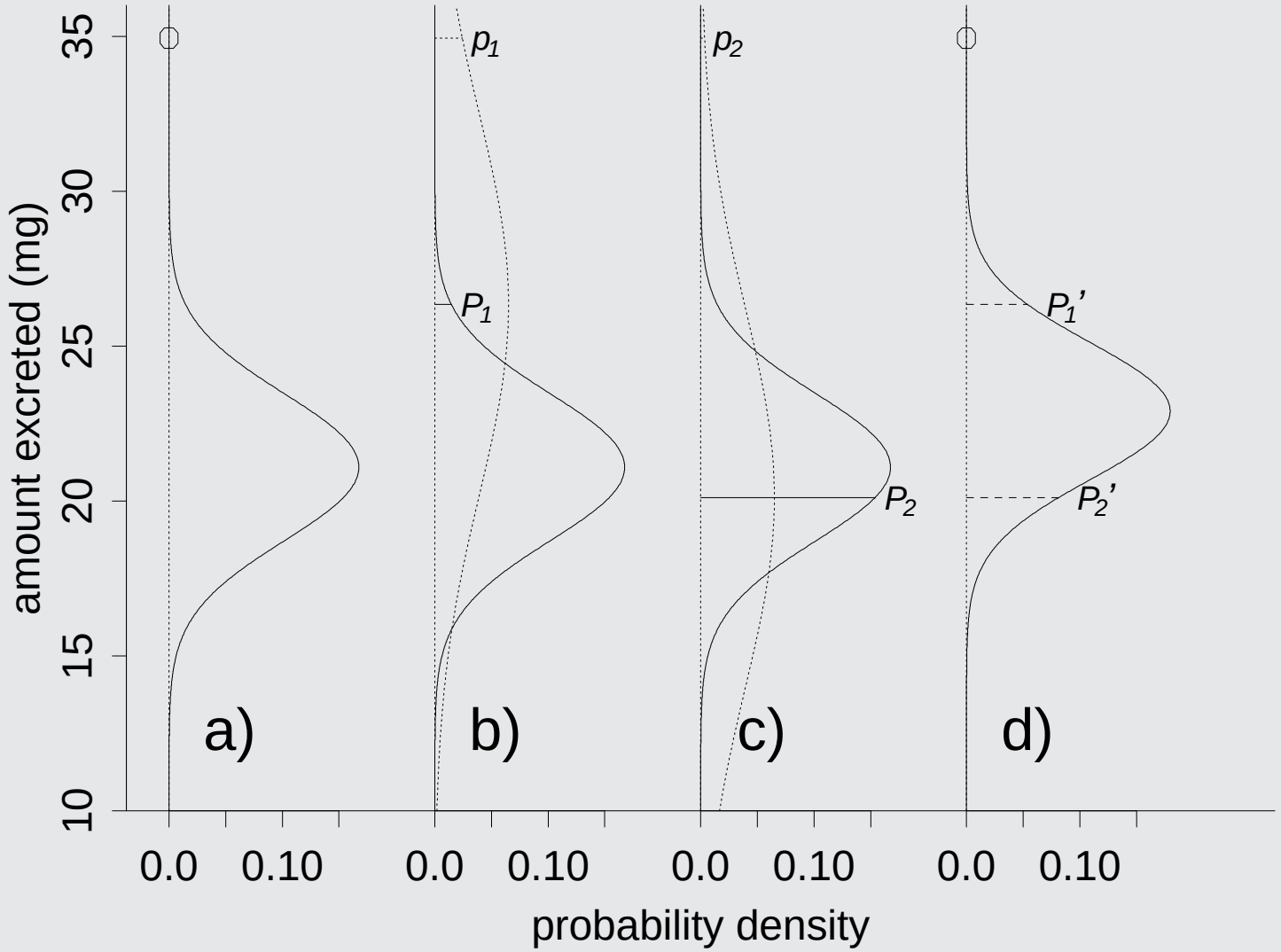
mean parameter in the absence of individual data?



The population parameters, as estimated from the data of individuals 1-18 (especially the small value of ω^2), imply that the individual means of any unknown individuals are to be expected within a narrow range around the population mean, as indicated by the bell-shaped curve on the right.



After obtaining one measurement in the new individual, it looks like the mean parameter of this individual will lie somewhere above the population mean. The information from this measurement can be combined with the information from individuals 1-18, as represented by the bell-shaped curve around the population mean, to obtain a new, updated probability distribution for the likely location of this individual's mean parameter:



probability of occurring = prior probability (P) · conditional probability (p)

(See next page for explanation.)

Graph a) shows the probability distribution for the likely location of the new individual's mean parameter in the absence of individual data ("prior distribution"). The dot indicates the first measurement in this individual. The prior distribution can now be updated.

As in the previous two examples, the updated probability distribution, "posterior distribution", is calculated by multiplying the prior probability for each possible case by the conditional probability that, if this was the true case, the given observation would be made. These probabilities are then scaled to sum up to 1 to obtain the posterior distribution.

In this example, the prior distribution is continuous so that it contains an infinite number of possible cases, as opposed to the previous examples where the number of possible cases was 3 for the urn experiment and was 2 for the mammography interpretation.

In this example, for each one of the infinite number of possible cases in the prior distribution, the same calculation is carried out as before for 3 or 2 possible cases: the prior probability of each case is multiplied by the probability of obtaining the current observation if this was the true case. The graphs b) and c) illustrate the calculation for two cases in the prior distribution, with prior probabilities P_1 and P_2. The ordinate under the solid bell-shaped curve at P_1 represents the prior probability (called "probability density" in a continuous distribution) of seeing an individual with mean excretion of 26.3 mg. If this was the true case, then observations would be expected in a wide range around this individual mean as indicated by the dotted bell-shaped curve centered at P_1; this curve represents a normal distribution with mean 26.3 mg and standard deviation equal to the residual standard deviation estimated from individuals 1-18, 6.1 mg.

The ordinate under the dotted curve at p_1 represents the conditional probability of obtaining the measurement indicated by the circle in graph a) if the individual mean were located at 26.3 mg. The product of the probabilities at P_1 and p_1, after upscaling of the entire posterior distribution to sum up to 1, becomes the posterior probability of 26.3 mg to be the mean parameter of this individual (graph d), ordinate at P_1'). The same calculation is illustrated for case no. 2 in graph c) where the prior probability, P_2, is much higher than that for case no. 1 in graph b) but is multiplied by a very low conditional probability, p_2. Carrying out this multiplication for all points of the prior probability distribution yields a continuous probability distribution which, by upscaling, becomes the posterior probability distribution shown in graph d). (An analytical solution exists for this example.)

In this case, the mean of the posterior distribution is still close to the population mean because of the large estimate of the residual variance, σ^2, indicating that single measurements have little information content.

Summary:

Bayesian calculations are useful when an assessment or a decision has to be based on measurements that are compromised by errors but when there is also previous knowledge about the probabilities of various scenarios or true values that could possibly be behind the current measurements. An example could be an attempt to estimate an individual patient's clearance of a drug based on one or two concentration measurements when there exists a database from a large number of patients to tell what the probability distribution of individual clearance values is and how large the concentration measurement errors are. Bayesian parameter estimation prevents from being overly confident in a few measurements at hand by balancing the current information against previous knowledge. The "posterior probability distributions" resulting from Bayesian parameter estimation do not describe expected frequencies of an event in a random experiment that may be repeated multiple times but describe the plausibilities of various possible explanations for observations that have already been made.