A program to refine C14 dates using a pair of relatives

Nick Patterson, Jakob Sedig, Iñigo Olalde, David Reich

April 10, 2021

1   Introduction

We describe and document usage of *refinedate*, a program that takes a prior distribution of date of death from two related individuals, and a prior distribution on the difference of the dates of death. It then constructs a joint posterior distribution. We document the process with an example of two individuals from Early Bronze Age England, I2457 (a male) and I2600 (a female). Genetics shows that the relationship is one of parent/child and I2457 is the father and I2600 his daughter, as the mitochondria belong to different haplogroups. We will show the prior distribution (smoothed) of date of death, as obtained from OxCal [Reimer et al., 2020, Bronk Ramsey, 2020].

2   Data Preparation

The calling sequence for *refinedate* requires 3 input files:

```
refinedate -a aname -b bname -d distancename
```

Here *aname, bname* are files giving prior information (such as from C14 dating) on the date at death of our 2 individuals. *distancename* gives a prior distribution on the difference of dates. Order, which matters for asymmetric relationships such as parent-child, is the date difference of the date of death of child minus date of death of parent. Our convention is that more remote dates are more negative. So for example 2100 *BCE* is given as -2100. We therefore expect that in the mean (date of child) - (date of parent) will be positive, though of course it may be negative for any particular pair of individuals. In our example we specify *aname* as the prior for child, *bname* as parent. We require all 3 distributions to be given at an interval of a year. Data (such as for example raw probability distributions) coming from *OxCal* may have a coarser 5 year quantization, and we provide a utility data smoother *gsmooth* to produce output files in the required format.

For our Early Bronze Age father-daughter pair, we use a prior distribution derived from [Kaplanis et al., 2018], which compiled genealogical information for millions of individuals from online historical and genealogical databases

We also provide sample distance distributions for some other common close relationships. These were created by collecting data for hundreds or thousands of individuals from online historical and genealogical databases (`genealogy.com,rootsweb.com,findagrave.com`) compiled in Supplement 2 of Sedig et al. 2021.
Note that the user is completely free to use other distributions that he or she supplies. For example in the case of a parent-child pair, it may be obvious that the child died in infancy, when our supplied distribution would be inappropriate.

3   Algorithm and Usage

We take from the input prior distributions $\{a(i), b(j)\}$ for the date of death in year i for sample A, year j for sample B. We take these priors to be independent. Then the joint distribution $z(i, j)$ is given by

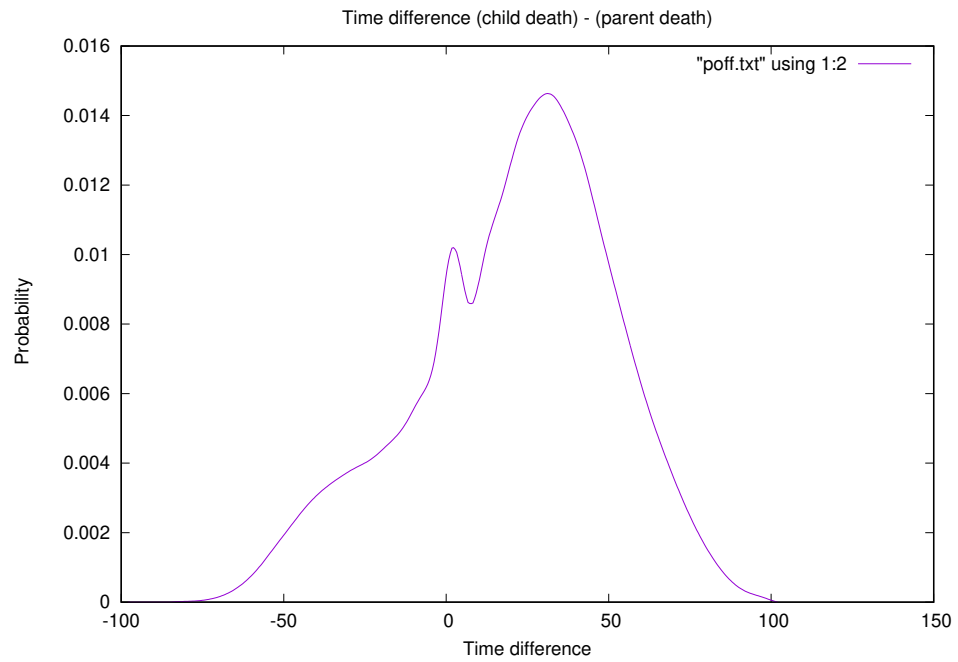$$z(i, j) \propto a(i)b(j)d(i - j)$$

where d is our difference distribution, and the constant of proportionality is chosen so that
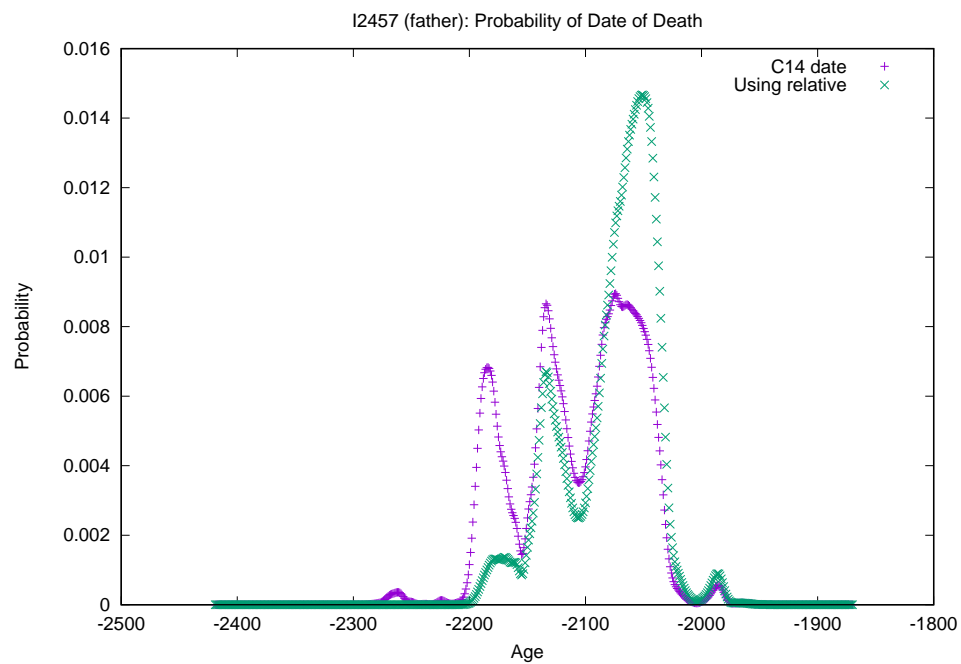$$\sum_{i,j} z(i, j) = 1$$

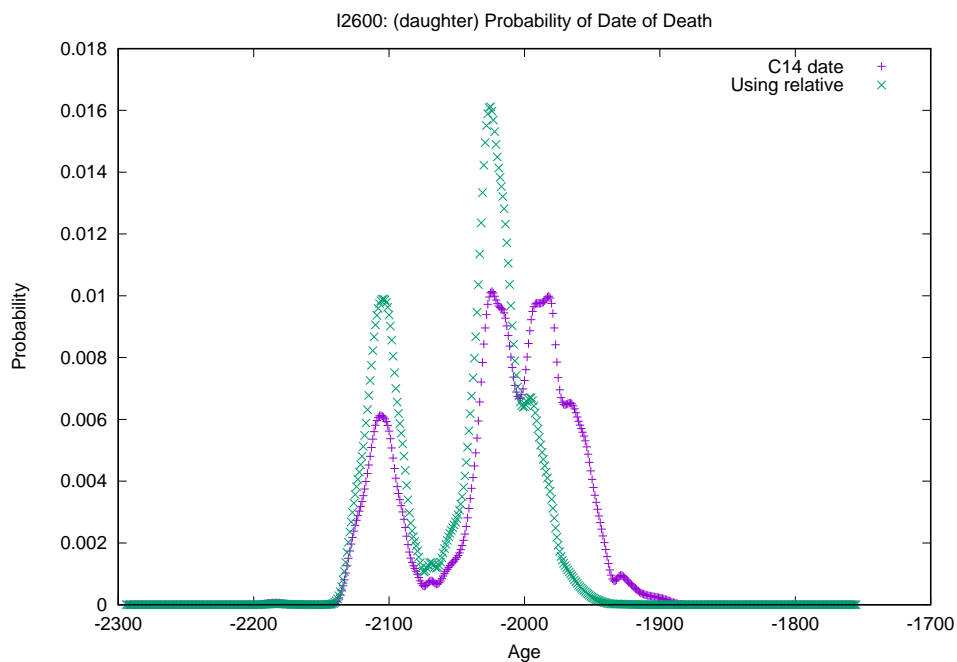We output marginal posterior estimates $a'(i), b'(j)$ and (optionally) $z(i, j)$.

4   An example

Here we give the prior distribution we use for a child-parent date difference.

We show prior and posterior distributions for the date of death for a father and daughter.

I2457 (father): Probability of Date of Death

I2600: (daughter) Probability of Date of Death

Note that both for father and daughter the prior distribution was multimodal with 3 peaks, while the posterior has just 2 peaks with one clearly dominant.

We also give prior and posterior means and standard deviations, though (of course) the distributions are far from normal.

Means + s.dev for Father and Daughter

	Prior		Posterior	
	Mean	s.dev	Mean	s.dev
Father	-2105.7	51.1	-2077.7	39.3
Daughter	-2018.0	52.8	-2045.4	45.5

Note that the mean difference on date of death has narrowed from an implausible 88 years to a much more likely 32 years.

To find and then download software and documentation,
Google 'github refinedate reich'. Detailed information about file formats is available on the *github* link.

References

[Bronk Ramsey, 2020] Bronk Ramsey, C. (2020). Oxcal 4.4.2. `https://c14.arch.ox.ac.uk/oxcal.html`.

[Kaplanis et al., 2018] Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., Gershovits, M., Markus, B., Sheikh, M., Gymrek, M., et al. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science*, 360(6385):171–175.

[Reimer et al., 2020] Reimer, P. J., Austin, W. E., Bard, E., Bayliss, A., Blackwell, P. G., Ramsey, C. B., Butzin, M., Cheng, H., Edwards, R. L., Friedrich, M., et al. (2020). The intcal20 northern hemisphere radiocarbon age calibration curve (0–55 cal kbp). *Radiocarbon*, 62(4):725–757.