A New Robust Statistical Model for Radiocarbon Data

J. Andrés Christen   Sergio Pérez-Elizalde

Centro de Investigación en Matemáticas
Guanajuato, México.

Paleochronology Building Workshop, San Miguel Allende, Guanajuato, México. August 2010.

Motivation

- Radiocarbon dating is a method to approximate the age of organic samples and after a complex and costly process a “radiocarbon date” and a standard error is the output of the dating process, $y \pm \sigma$ (eg. 4500 ± 30).
- The general method currently used to analyze radiocarbon data (y) is conditional on the standard deviation (σ). Nevertheless, σ is assumed as known in the usual statistical model for radiocarbon data.
- We want to propose a robust analysis in the presence of atypical data.
- and understand and explain the scatter in radiocarbon data seen in interlaboratory studies.

Motivation

- Radiocarbon dating is a method to approximate the age of organic samples and after a complex and costly process a "radiocarbon date" and a standard error is the output of the dating process, $y \pm \sigma$ (eg. 4500 ± 30).
- The general method currently used to analyze radiocarbon data (y) is conditional on the standard deviation (σ). Nevertheless, σ is assumed as known in the usual statistical model for radiocarbon data.
- We want to propose a robust analysis in the presence of atypical data.
- and understand and explain the scatter in radiocarbon data seen in interlaboratory studies.

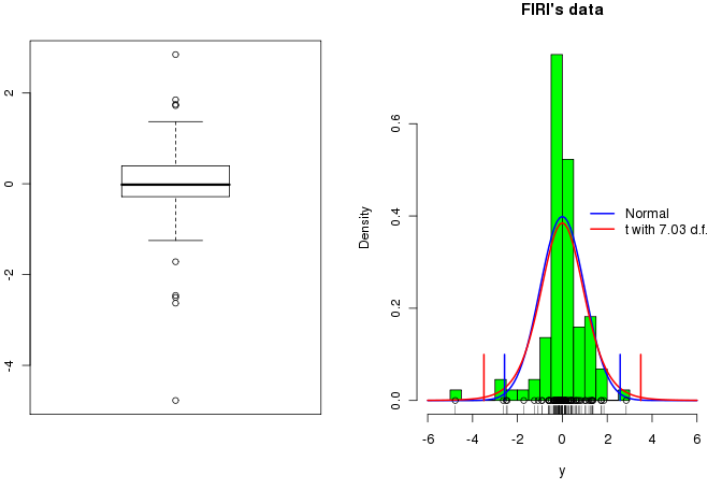
Motivation

- Radiocarbon dating is a method to approximate the age of organic samples and after a complex and costly process a "radiocarbon date" and a standard error is the output of the dating process, $y \pm \sigma$ (eg. 4500 ± 30).
- The general method currently used to analyze radiocarbon data (y) is conditional on the standard deviation (σ). Nevertheless, σ is assumed as known in the usual statistical model for radiocarbon data.
- We want to propose a robust analysis in the presence of atypical data.
- and understand and explain the scatter in radiocarbon data seen in interlaboratory studies.

Motivation

- Radiocarbon dating is a method to approximate the age of organic samples and after a complex and costly process a “radiocarbon date” and a standard error is the output of the dating process, $y \pm \sigma$ (eg. 4500 ± 30).
- The general method currently used to analyze radiocarbon data (y) is conditional on the standard deviation (σ). Nevertheless, σ is assumed as known in the usual statistical model for radiocarbon data.
- We want to propose a robust analysis in the presence of atypical data.
- and understand and explain the scatter in radiocarbon data seen in interlaboratory studies.

Distribution of offsets relative to the dendro-dated samples



Traditional model

The traditional statistical model for a ^{14}C determination y_j is given by

$$y_j \sim \mathrm{N}\left(\mu(\theta), \sigma_j^2\right), \quad j = 1, 2, \ldots, m \tag{1}$$

- $\mu(\cdot)$ is the calibration curve
- θ is the associated calendar year
- σ_j is the reported standard deviation for y_j
- For a given θ we use an estimate of both $\mu(\theta)$ and its standard deviation $\sigma(\theta)$ (for example INTCAL04). Model (1) becomes

$$y_j \sim \mathrm{N}\left(\mu(\theta), \sigma_j^2 + \sigma^2(\theta)\right), \tag{2}$$

where σ_j is assumed as known; this is the basic statistical model currently used for the statistical analysis of ^{14}C data

Traditional model

The traditional statistical model for a ^{14}C determination y_j is given by

$$y_j \sim \mathrm{N}\left(\mu(\theta), \sigma_j^2\right), \quad j = 1, 2, \ldots, m \tag{1}$$

- $\mu(\cdot)$ is the calibration curve
- θ is the associated calendar year
- σ_j is the reported standard deviation for y_j
- For a given θ we use an estimate of both $\mu(\theta)$ and its standard deviation $\sigma(\theta)$ (for example INTCAL04). Model (1) becomes

$$y_j \sim \mathrm{N}\left(\mu(\theta), \sigma_j^2 + \sigma^2(\theta)\right), \tag{2}$$

where σ_j is assumed as known; this is the basic statistical model currently used for the statistical analysis of ^{14}C data

Traditional model

The traditional statistical model for a ^{14}C determination y_j is given by

$$y_j \sim \mathrm{N}\left(\mu(\theta), \sigma_j^2\right), \quad j = 1, 2, \ldots, m \tag{1}$$

- $\mu(\cdot)$ is the calibration curve
- θ is the associated calendar year
- σ_j is the reported standard deviation for y_j
- For a given θ we use an estimate of both $\mu(\theta)$ and its standard deviation $\sigma(\theta)$ (for example INTCAL04). Model (1) becomes

$$y_j \sim \mathrm{N}\left(\mu(\theta), \sigma_j^2 + \sigma^2(\theta)\right), \tag{2}$$

where σ_j is assumed as known; this is the basic statistical model currently used for the statistical analysis of ^{14}C data

Traditional model

The traditional statistical model for a ^{14}C determination y_j is given by

$$y_j \sim \mathrm{N}\left(\mu(\theta), \sigma_j^2\right), \quad j = 1, 2, \ldots, m \tag{1}$$

- $\mu(\cdot)$ is the calibration curve
- θ is the associated calendar year
- σ_j is the reported standard deviation for y_j
- For a given θ we use an estimate of both $\mu(\theta)$ and its standard deviation $\sigma(\theta)$ (for example INTCAL04). Model (1) becomes

$$y_j \sim \mathrm{N}\left(\mu(\theta), \sigma_j^2 + \sigma^2(\theta)\right), \tag{2}$$

where σ_j is assumed as known; this is the basic statistical model currently used for the statistical analysis of ^{14}C data

The traditional likelihood function

- The likelihood function for θ given a random sample $\mathbf{y} = (y_1, \ldots, y_m)$ of m ^{14}C determinations is

$$L_N(\theta \mid \mathbf{y}) \propto \prod_{j=1}^{m} \frac{1}{\omega_j(\theta)} \exp\left\{ -\frac{1}{2\omega_j^2(\theta)} \left(y_j - \mu(\theta)\right)^2 \right\}, \qquad (3)$$

where $\omega_j^2(\theta) = \sigma^2(\theta) + \sigma_j^2$.

- We derive the posterior distribution of θ by formal use of the Bayes' rule; that is,

$$\pi(\theta \mid \mathbf{y}) \propto L(\theta \mid \mathbf{y})\pi(\theta). \qquad (4)$$

The traditional likelihood function

- The likelihood function for θ given a random sample $\mathbf{y} = (y_1, \ldots, y_m)$ of m ^{14}C determinations is

$$L_N(\theta \mid \mathbf{y}) \propto \prod_{j=1}^{m} \frac{1}{\omega_j(\theta)} \exp\left\{-\frac{1}{2\omega_j^2(\theta)} \left(y_j - \mu(\theta)\right)^2\right\}, \tag{3}$$

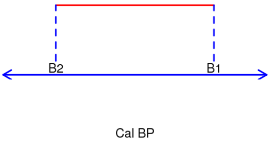
where $\omega_j^2(\theta) = \sigma^2(\theta) + \sigma_j^2$.

- We derive the posterior distribution of θ by formal use of the Bayes' rule; that is,

$$\pi(\theta \mid \mathbf{y}) \propto L(\theta \mid \mathbf{y})\pi(\theta). \tag{4}$$

The traditional posterior function

- For the prior of θ we will use a uniform distribution on the interval (B_2, B_1), $B_1 < B_2$.

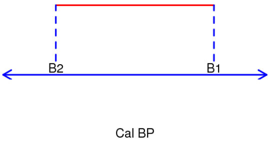


- Further prior information about θ may be included through any other prior distribution.
- With the conventional normal model the posterior is proportional to the likelihood L_N, i.e.,

$$\pi_N(\theta \mid y) \propto L_N(\theta \mid y)$$

The traditional posterior function

- For the prior of θ we will use a uniform distribution on the interval $(B_2, B_1)\,, \quad B_1 < B_2$.

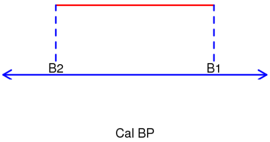


- Further prior information about θ may be included through any other prior distribution.
- With the conventional normal model the posterior is proportional to the likelihood L_N, i.e.,

$$\pi_N(\theta \mid y) \propto L_N(\theta \mid y)$$

The traditional posterior function

- For the prior of θ we will use a uniform distribution on the interval $(B_2, B_1)\,, \quad B_1 < B_2$.

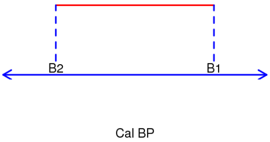


- Further prior information about θ may be included through any other prior distribution.
- With the conventional normal model the posterior is proportional to the likelihood L_N, i.e.,

$$\pi_N(\theta \mid y) \propto L_N(\theta \mid y)$$

The traditional posterior function

- For the prior of θ we will use a uniform distribution on the interval $(B_2, B_1)\,, \quad B_1 < B_2$.

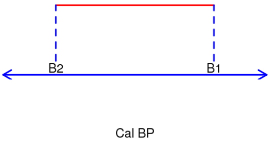


- Further prior information about θ may be included through any other prior distribution.
- With the conventional normal model the posterior is proportional to the likelihood L_N; i.e.,

$$\pi_N(\theta \mid \mathbf{y}) \propto L_N(\theta \mid \mathbf{y}).$$

The traditional posterior function

- For the prior of θ we will use a uniform distribution on the interval $(B_2, B_1)\,, \quad B_1 < B_2$.

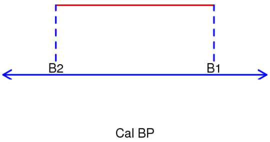


- Further prior information about θ may be included through any other prior distribution.
- With the conventional normal model the posterior is proportional to the likelihood L_N; i.e.,

$$\pi_N(\theta \mid \mathbf{y}) \propto L_N(\theta \mid \mathbf{y}).$$

The traditional posterior function

- For the prior of θ we will use a uniform distribution on the interval $(B_2, B_1)\,, \quad B_1 < B_2$.

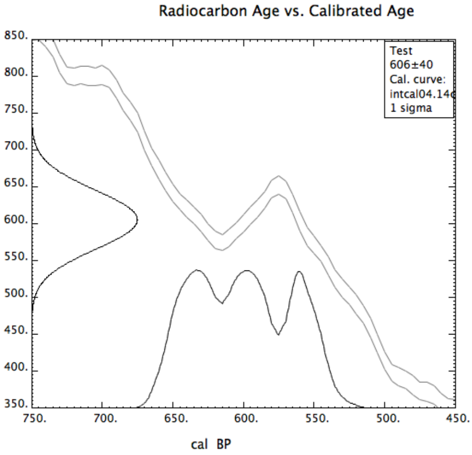


- Further prior information about θ may be included through any other prior distribution.
- With the conventional normal model the posterior is proportional to the likelihood L_N; i.e.,

$$\pi_N(\theta \mid \mathbf{y}) \propto L_N(\theta \mid \mathbf{y}).$$

Classic calibration (using software CALIB)

Calibration of a ^{14}C determination 606 ± 40 using CALIB.



Traditional model: disadvantages

- This traditional model assumes that σ_j is known exactly. However, σ_j is calculated at each laboratory and strictly speaking is not known precisely.
- Also, the presence of outliers is a constant factor in the analysis of ^{14}C data, which may influence notably the inference results given the small sample sizes common in practice (Blaauw et al., 2005).
- Even for the simplest of cases Christen (1994) approach to *detect* outliers requires the use of complex numerical techniques (eg. MCMC).
- International interlaboratory studies show "unexplained" scatter in ^{14}C data. An unexplored alternative would be to change the model to a heavier tailed distribution than the Normal

Traditional model: disadvantages

- This traditional model assumes that σ_j is known exactly. However, σ_j is calculated at each laboratory and strictly speaking is not known precisely.
- Also, the presence of outliers is a constant factor in the analysis of ^{14}C data, which may influence notably the inference results given the small sample sizes common in practice (Blaauw et al., 2005).
- Even for the simplest of cases Christen (1994) approach to *detect* outliers requires the use of complex numerical techniques (eg. MCMC).
- International interlaboratory studies show “unexplained” scatter in ^{14}C data. An unexplored alternative would be to change the model to a heavier tailed distribution than the Normal

Traditional model: disadvantages

- This traditional model assumes that σ_j is known exactly. However, σ_j is calculated at each laboratory and strictly speaking is not known precisely.
- Also, the presence of outliers is a constant factor in the analysis of ^{14}C data, which may influence notably the inference results given the small sample sizes common in practice (Blaauw et al., 2005).
- Even for the simplest of cases Christen (1994) approach to *detect* outliers requires the use of complex numerical techniques (eg. MCMC).
- International interlaboratory studies show "unexplained" scatter in ^{14}C data. An unexplored alternative would be to change the model to a heavier tailed distribution than the Normal

Traditional model: disadvantages

- This traditional model assumes that σ_j is known exactly. However, σ_j is calculated at each laboratory and strictly speaking is not known precisely.
- Also, the presence of outliers is a constant factor in the analysis of ^{14}C data, which may influence notably the inference results given the small sample sizes common in practice (Blaauw et al., 2005).
- Even for the simplest of cases Christen (1994) approach to *detect* outliers requires the use of complex numerical techniques (eg. MCMC).
- International interlaboratory studies show "unexplained" scatter in ^{14}C data. An unexplored alternative would be to change the model to a heavier tailed distribution than the Normal

Christen 1994 Outlier identification

Christen (1994) approach to *detect* outliers considers that *each* radiocarbon determination might need a shift in the radiocarbon scale δ_j, in order to be be properly explained in terms of the rest of dates and the contextual information used.

That is:

$$y_j \sim \mathrm{N}\left(\mu(\theta) + \phi_j\delta_j, \sigma_j^2\right), \quad j = 1, 2, \ldots, m,$$

where $\phi_j = 1, 0$ depending on whether determination j does require or does not requiere a shift (δ_j) to be properly explained (is or is not and outlier).

The posterior probability $P[\phi_j = 1|\text{All data and prior info}]$ is calculated. This main be interpreted as our (posterior) probability that determination j is an outlier.

Christen 1994 Outlier identification

Christen (1994) approach to *detect* outliers considers that *each* radiocarbon determination might need a shift in the radiocarbon scale δ_j, in order to be be properly explained in terms of the rest of dates and the contextual information used.

That is:

$$y_j \sim \mathrm{N}\left(\mu(\theta) + \phi_j\delta_j, \sigma_j^2\right), \quad j = 1, 2, \ldots, m,$$

where $\phi_j = 1, 0$ depending on whether determination j does require or does not requiere a shift (δ_j) to be properly explained (is or is not and outlier).

The posterior probability $P[\phi_j = 1|\text{All data and prior info}]$ is calculated. This main be interpreted as our (posterior) probability that determination j is an outlier.

Christen 1994 Outlier identification

Christen (1994) approach to *detect* outliers considers that *each* radiocarbon determination might need a shift in the radiocarbon scale δ_j, in order to be be properly explained in terms of the rest of dates and the contextual information used.

That is:

$$y_j \sim \mathrm{N}\left(\mu(\theta) + \phi_j\delta_j, \sigma_j^2\right), \quad j = 1, 2, \ldots, m,$$

where $\phi_j = 1, 0$ depending on whether determination j does require or does not requiere a shift (δ_j) to be properly explained (is or is not and outlier).

The posterior probability $P[\phi_j = 1|\text{All data and prior info}]$ is calculated. This main be interpreted as our (posterior) probability that determination j is an outlier.

Christen 1994 Outlier identification

Christen (1994) approach to *detect* outliers considers that *each* radiocarbon determination might need a shift in the radiocarbon scale δ_j, in order to be be properly explained in terms of the rest of dates and the contextual information used.

That is:

$$y_j \sim \mathrm{N}\left(\mu(\theta) + \phi_j \delta_j, \sigma_j^2\right), \quad j = 1, 2, \ldots, m,$$

where $\phi_j = 1, 0$ depending on whether determination j does require or does not requiere a shift (δ_j) to be properly explained (is or is not and outlier).

The posterior probability $P[\phi_j = 1 | \text{All data and prior info}]$ is calculated. This main be interpreted as our (posterior) probability that determination j is an outlier.

Normal model with a variance multiplier

- We know that σ_j varies jointly with y_j.
- The uncertainty about the variance of y_j in model (1) may be introduced by considering the product $\alpha\sigma_j^2$, where $\alpha > 0$. The new model is

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha\sigma_j^2\right)$$

- α is an unknown "variance multiplier" to the laboratory reported variance σ_j^2.
- If we also consider a model which uses the variance $\sigma^2(\theta)$ in the calibration curve,

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha\left(\sigma^2(\theta) + \sigma_j^2\right)\right)$$

Normal model with a variance multiplier

- We know that σ_j varies jointly with y_j.
- The uncertainty about the variance of y_j in model (1) may be introduced by considering the product $\alpha\sigma_j^2$, where $\alpha > 0$. The new model is

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha\sigma_j^2\right)$$

- α is an unknown "variance multiplier" to the laboratory reported variance σ_j^2.
- If we also consider a model which uses the variance $\sigma^2(\theta)$ in the calibration curve,

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha\left(\sigma^2(\theta) + \sigma_j^2\right)\right)$$

The new model

Normal model with a variance multiplier

- We know that σ_j varies jointly with y_j.
- The uncertainty about the variance of y_j in model (1) may be introduced by considering the product $\alpha\sigma_j^2$, where $\alpha > 0$. The new model is

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha\sigma_j^2\right)$$

- α is an unknown "variance multiplier" to the laboratory reported variance σ_j^2.
- If we also consider a model which uses the variance $\sigma^2(\theta)$ in the calibration curve,

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha\left(\sigma^2(\theta) + \sigma_j^2\right)\right)$$

Normal model with a variance multiplier

- We know that σ_j varies jointly with y_j.
- The uncertainty about the variance of y_j in model (1) may be introduced by considering the product $\alpha\sigma_j^2$, where $\alpha > 0$. The new model is

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha\sigma_j^2\right)$$

- α is an unknown "variance multiplier" to the laboratory reported variance σ_j^2.
- If we also consider a model which uses the variance $\sigma^2(\theta)$ in the calibration curve,

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha\left(\sigma^2(\theta) + \sigma_j^2\right)\right)$$

The new model

- Calibration curve based on high quality data, so might need more optimistic α for $\sigma(\theta)$. But, then MCMC needed to infer model parameters.
- We assume multiplier α also affects $\sigma(\theta)$, ensures mathematical tractability and analytically feasible representation of θ posterior distribution.

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha_1\sigma^2(\theta) + \alpha_2\sigma_j^2\right). \tag{5}$$

- Typically $\sigma(\theta)$ small compared to σ, model well behaved approximation to the more realistic model.

The new model

- Calibration curve based on high quality data, so might need more optimistic α for $\sigma(\theta)$. But, then MCMC needed to infer model parameters.
- We assume multiplier α also affects $\sigma(\theta)$, ensures mathematical tractability and analytically feasible representation of θ posterior distribution.

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha_1 \sigma^2(\theta) + \alpha_2 \sigma_j^2\right). \tag{5}$$

- Typically $\sigma(\theta)$ small compared to σ, model well behaved approximation to the more realistic model.

The new model

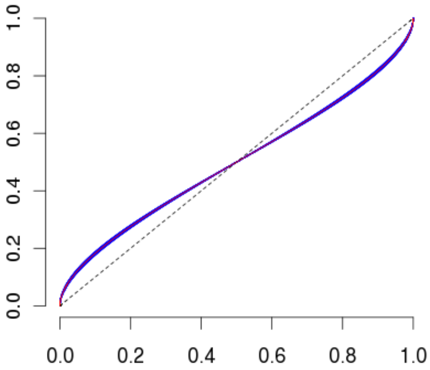
- Calibration curve based on high quality data, so might need more optimistic α for $\sigma(\theta)$. But, then MCMC needed to infer model parameters.
- We assume multiplier α also affects $\sigma(\theta)$, ensures mathematical tractability and analytically feasible representation of θ posterior distribution.

$$y_j \sim \mathrm{N}\left(\mu(\theta), \alpha_1 \sigma^2(\theta) + \alpha_2 \sigma_j^2\right). \tag{5}$$

- Typically $\sigma(\theta)$ small compared to σ, model well behaved approximation to the more realistic model.

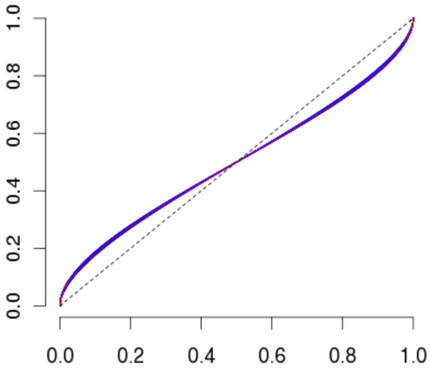
Double mulplier model vs. single multiplier model

Probability plot for the double multiplier normal model (vertical axis, blue region), single multiplier normal model (vertical axis, red line) vs. the traditional normal model (horizontal axis). $\theta = 500$, $\sigma = 50$.



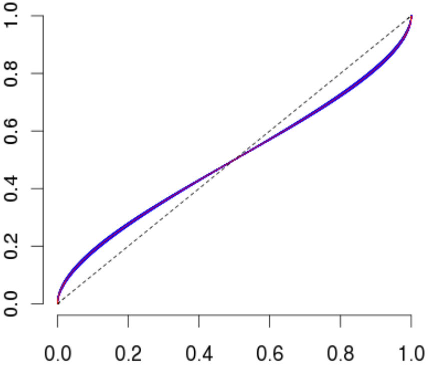
Double mulplier model vs. single multiplier model

Probability plot for the double multiplier normal model (vertical axis, blue region), single multiplier normal model (vertical axis, red line) vs. the traditional normal model (horizontal axis). $\theta = 500$, $\sigma = 50$.



Double mulplier model vs. single multiplier model

Probability plot for the double multiplier normal model (vertical axis, blue region), single multiplier normal model (vertical axis, red line) vs. the traditional normal model (horizontal axis). $\theta = 500$, $\sigma = 50$.



The prior for α

We assume that the prior distribution for α is an inverted gamma with parameters *a* and *b*

$$\pi(\alpha) = \mathrm{InvGa}(\alpha \mid a, b). \tag{6}$$

Then, given θ, the prior distribution of $\alpha\omega_j^2$ is the inverted gamma

$$\alpha\omega_j^2 \mid \theta \sim \mathrm{InvGa}\left(a, b\left(\sigma_j^2 + \sigma^2(\theta)\right)\right), \tag{7}$$

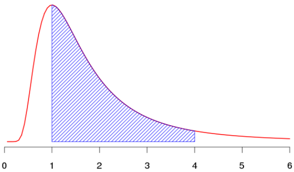
such that

$$\mathrm{E}\left(\alpha\omega_j^2 \mid \theta\right) = \frac{b}{a-1}\left(\sigma_j^2 + \sigma^2(\theta)\right)$$

is the prior expected variance of y_j.

It is clear that for particular applications, $\pi(\alpha)$ should be set according to *a priori* considerations about possible error multipliers for the sample at hand.

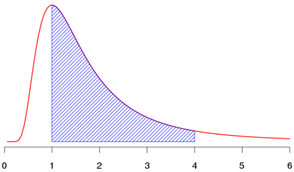
A proposed prior for α



Prior density for the variance multiplier α with expected value $\mathrm{E}(\alpha) = b/(a-1) = 2$, mode $\mathrm{Mo}(\alpha) = b/(a+1) = 1$ and median $\mathrm{Me}(\alpha) \approx 1.5$, $\Pr(\alpha \leq 1) \approx 0.24$, $P(\alpha \geq 4) \approx 0.08$, $a = 3, b = 4$.

- $\mathrm{InvGa}(\alpha \mid a = 3, b = 4)$ represents $\Pr(\sqrt{\alpha} > 2) \approx 0.08$, $\Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $\Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.
- The most likely scenario is that the error term was correctly reported.
- The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.

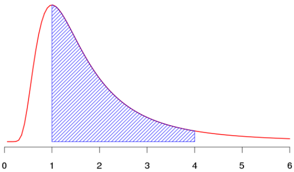
A proposed prior for α



Prior density for the variance multiplier α with expected value $\mathrm{E}(\alpha) = b/(a-1) = 2$, mode $\mathrm{Mo}(\alpha) = b/(a+1) = 1$ and median $\mathrm{Me}(\alpha) \approx 1.5$, $\Pr(\alpha \leq 1) \approx 0.24$, $P(\alpha \geq 4) \approx 0.08$, $a = 3, b = 4$.

- $\mathrm{InvGa}(\alpha \mid a = 3, b = 4)$ represents $\Pr(\sqrt{\alpha} > 2) \approx 0.08$, $\Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $\Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.
- The most likely scenario is that the error term was correctly reported.
- The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.

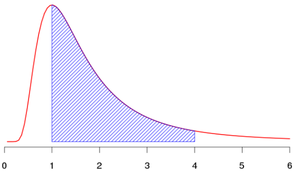
A proposed prior for α



Prior density for the variance multiplier α with expected value $\mathrm{E}(\alpha) = b/(a-1) = 2$, mode $\mathrm{Mo}(\alpha) = b/(a+1) = 1$ and median $\mathrm{Me}(\alpha) \approx 1.5$, $\Pr(\alpha \leq 1) \approx 0.24$, $P(\alpha \geq 4) \approx 0.08$, $a = 3, b = 4$.

- $\mathrm{InvGa}(\alpha \mid a = 3, b = 4)$ represents $\Pr(\sqrt{\alpha} > 2) \approx 0.08$, $\Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $\Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.
- The most likely scenario is that the error term was correctly reported.
- The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.

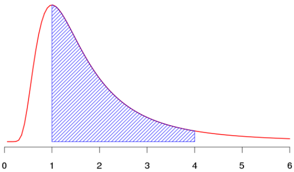
A proposed prior for α



Prior density for the variance multiplier α with expected value $\mathrm{E}(\alpha) = b/(a-1) = 2$, mode $\mathrm{Mo}(\alpha) = b/(a+1) = 1$ and median $\mathrm{Me}(\alpha) \approx 1.5$, $\Pr(\alpha \leq 1) \approx 0.24$, $P(\alpha \geq 4) \approx 0.08$, $a = 3, b = 4$.

- $\mathrm{InvGa}(\alpha \mid a = 3, b = 4)$ represents $\Pr(\sqrt{\alpha} > 2) \approx 0.08$, $\Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $\Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.
- The most likely scenario is that the error term was correctly reported.
- The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.

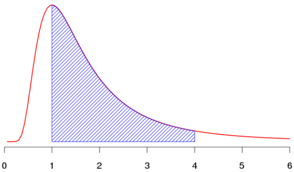
A proposed prior for α



Prior density for the variance multiplier α with expected value $\mathrm{E}(\alpha) = b/(a-1) = 2$, mode $\mathrm{Mo}(\alpha) = b/(a+1) = 1$ and median $\mathrm{Me}(\alpha) \approx 1.5$, $\Pr(\alpha \leq 1) \approx 0.24$, $P(\alpha \geq 4) \approx 0.08$, $a = 3, b = 4$.

- $\mathrm{InvGa}(\alpha \mid a = 3, b = 4)$ represents $\Pr(\sqrt{\alpha} > 2) \approx 0.08$, $\Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $\Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.
- The most likely scenario is that the error term was correctly reported.
- The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.

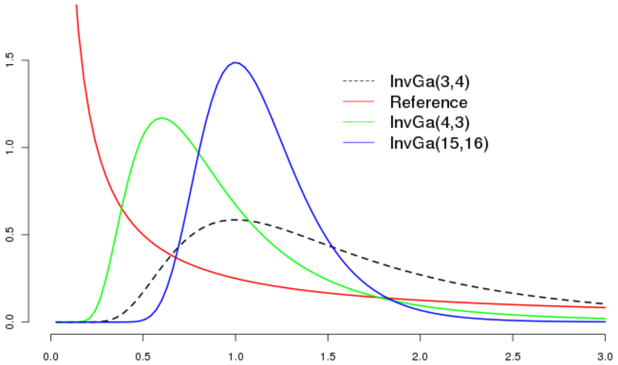
A proposed prior for α



Prior density for the variance multiplier α with expected value $\mathrm{E}(\alpha) = b/(a-1) = 2$, mode $\mathrm{Mo}(\alpha) = b/(a+1) = 1$ and median $\mathrm{Me}(\alpha) \approx 1.5$, $\Pr(\alpha \leq 1) \approx 0.24$, $P(\alpha \geq 4) \approx 0.08$, $a = 3, b = 4$.

- $\mathrm{InvGa}(\alpha \mid a = 3, b = 4)$ represents $\Pr(\sqrt{\alpha} > 2) \approx 0.08$, $\Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $\Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.
- The most likely scenario is that the error term was correctly reported.
- The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.

Examples for the prior for α



The new (integrated) likelihood

- The parameter of interest is the true calendar age θ, being α a nuisance parameter.
- In a Bayesian setting nuisance parameters are naturally eliminated by integrating out them from either the posterior distribution or the likelihood function. Here we derive the posterior distribution for θ using the integrated likelihood:

$$L_{a,b}(\theta \mid \mathbf{y}) = \int_0^\infty \prod_{j=1}^{m} p(y_j \mid \theta, \alpha) \pi_{a,b}(\alpha) \, d\alpha. \quad (8)$$

- Note that we are assuming prior independence of the parameters (θ, α).

The new (integrated) likelihood

- The parameter of interest is the true calendar age θ, being α a nuisance parameter.
- In a Bayesian setting nuisance parameters are naturally eliminated by integrating out them from either the posterior distribution or the likelihood function. Here we derive the posterior distribution for θ using the integrated likelihood:

$$L_{a,b}(\theta \mid \mathbf{y}) = \int_0^\infty \prod_{j=1}^m p(y_j \mid \theta, \alpha)\pi_{a,b}(\alpha)\, \mathrm{d}\alpha. \tag{8}$$

- Note that we are assuming prior independence of the parameters (θ, α).

The new (integrated) likelihood

- The parameter of interest is the true calendar age θ, being α a nuisance parameter.
- In a Bayesian setting nuisance parameters are naturally eliminated by integrating out them from either the posterior distribution or the likelihood function. Here we derive the posterior distribution for θ using the integrated likelihood:

$$L_{a,b}(\theta \mid \mathbf{y}) = \int_0^\infty \prod_{j=1}^{m} p(y_j \mid \theta, \alpha)\pi_{a,b}(\alpha)\,\mathrm{d}\alpha. \tag{8}$$

- Note that we are assuming prior independence of the parameters (θ, α).

The new (integrated) likelihood

- Therefore, under the prior distribution (6) the integrated likelihood, given $\mathbf{y} = (y_1, \ldots, y_m)$, is

$$L_{a,b}(\theta \mid \mathbf{y}) = \int_0^\infty \prod_{j=1}^m \mathrm{N}\left(y_j \mid \mu(\theta), \alpha\omega_j(\theta)\right) \mathrm{InvGa}\left(\alpha \mid a, b\right) \mathrm{d}\alpha$$

$$\propto \left[1 + a^{-1} \sum_{j=1}^m \frac{\left(y_j - \mu(\theta)\right)^2}{\omega_j(\theta) b/a}\right]^{-\frac{2a+m}{2}}$$

$$\propto \mathrm{t}\left(\mathbf{y} \middle| \mu(\theta)\mathbf{1}_m, \Sigma(\theta) b/a, 2a\right), \quad (9)$$

where $\mathbf{1}_m = (1, \ldots, 1)^t$ and $\Sigma(\theta) = \mathrm{diag}\left(\omega_1(\theta), \ldots, \omega_m(\theta)\right)$.

- The integrated likelihood for θ given $\mathbf{y}$ is proportional to a t distribution with location parameter $\mu(\theta)\mathbf{1}_m$, covariance matrix $\Sigma(\theta)b/(a-1)$ and $2a$ d.f..

The new (integrated) likelihood

- Therefore, under the prior distribution (6) the integrated likelihood, given $\mathbf{y} = (y_1, \ldots, y_m)$, is

$$L_{a,b}(\theta \mid \mathbf{y}) = \int_0^\infty \prod_{j=1}^m \mathrm{N}\left(y_j \mid \mu(\theta), \alpha\omega_j(\theta)\right) \mathrm{InvGa}\left(\alpha \mid a, b\right) \mathrm{d}\alpha$$

$$\propto \left[1 + a^{-1} \sum_{j=1}^m \frac{\left(y_j - \mu(\theta)\right)^2}{\omega_j(\theta) b/a}\right]^{-\frac{2a+m}{2}}$$

$$\propto \mathrm{t}\left(\mathbf{y} \middle| \mu(\theta)\mathbf{1}_m, \Sigma(\theta) b/a, 2a\right), \quad (9)$$

where $\mathbf{1}_m = (1, \ldots, 1)^t$ and $\Sigma(\theta) = \mathrm{diag}\left(\omega_1(\theta), \ldots, \omega_m(\theta)\right)$.

- The integrated likelihood for θ given $\mathbf{y}$ is proportional to a t distribution with location parameter $\mu(\theta)\mathbf{1}_m$, covariance matrix $\Sigma(\theta)b/(a-1)$ and $2a$ d.f..

The new (integrated) likelihood

- Note that with the sequence of parameters values $a = i + 1, b = i, i = 1, 2, \ldots,$ we obtain a sequence of prior distributions which converges to the degenerate Dirac distribution at $\alpha = 1$, leading to a sequence of integrated heavy tail models, with covariance matrix $\Sigma(\theta)$, which converges to the traditional normal model.
- As expected, our new model has as limiting case the standard normal model when *a priori* $\Pr(\alpha = 1) = 1$; that is, σ_j is known exactly.

The new (integrated) likelihood

- Note that with the sequence of parameters values $a = i + 1, b = i, i = 1, 2, \ldots$, we obtain a sequence of prior distributions which converges to the degenerate Dirac distribution at $\alpha = 1$, leading to a sequence of integrated heavy tail models, with covariance matrix $\Sigma(\theta)$, which converges to the traditional normal model.
- As expected, our new model has as limiting case the standard normal model when *a priori* $\Pr(\alpha = 1) = 1$; that is, σ_j is known exactly.

Posterior distribution

- Now we derive the posterior distribution of θ by formal use of the Bayes' rule; that is,

$$\pi\left(\theta \mid \mathbf{y}\right) \propto L(\theta \mid \mathbf{y})\pi(\theta). \tag{10}$$

- As a prior of θ we will use a uniform distribution on the interval $\left(B_2, B_1\right), \quad B_1 < B_2$. Of course, if the researcher has further prior information about θ they may properly include it through any other prior distribution, exactly the same as in the traditional normal case.
- The posterior distribution for θ is

$$\pi_{a,b}\left(\theta \mid \mathbf{y}\right) \propto L_{a,b}(\theta \mid \mathbf{y})\pi(\theta) \propto L_{a,b}(\theta \mid \mathbf{y}), \tag{11}$$

Posterior distribution

- Now we derive the posterior distribution of θ by formal use of the Bayes' rule; that is,

$$\pi\left(\theta \mid \mathbf{y}\right) \propto L(\theta \mid \mathbf{y})\pi(\theta). \tag{10}$$

- As a prior of θ we will use a uniform distribution on the interval $\left(B_2, B_1\right), \quad B_1 < B_2$. Of course, if the researcher has further prior information about θ they may properly include it through any other prior distribution, exactly the same as in the traditional normal case.
- The posterior distribution for θ is

$$\pi_{a,b}\left(\theta \mid \mathbf{y}\right) \propto L_{a,b}(\theta \mid \mathbf{y})\pi(\theta) \propto L_{a,b}(\theta \mid \mathbf{y}), \tag{11}$$

Posterior distribution

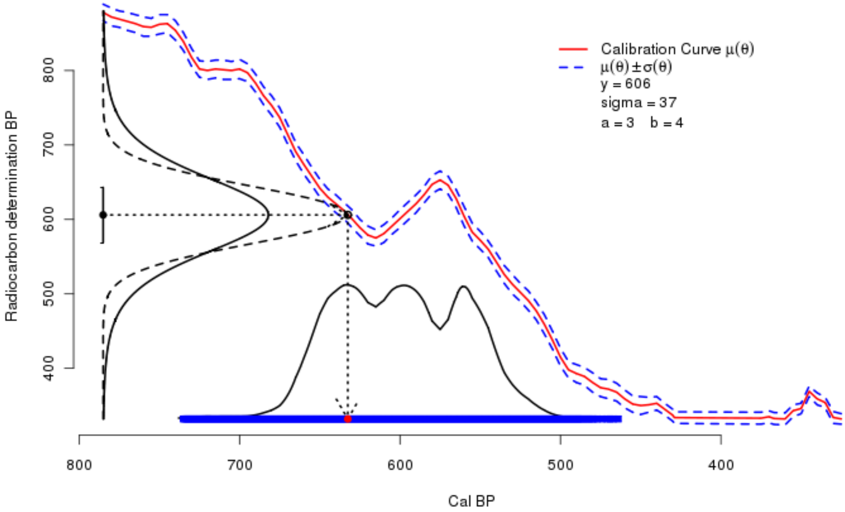
- Now we derive the posterior distribution of θ by formal use of the Bayes' rule; that is,

$$\pi\,(\theta \mid \mathbf{y}) \propto L(\theta \mid \mathbf{y})\pi(\theta). \tag{10}$$

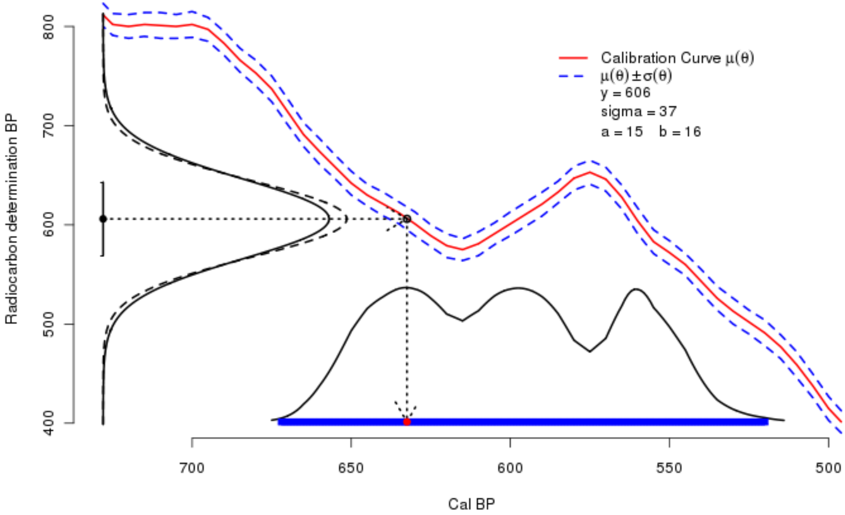
- As a prior of θ we will use a uniform distribution on the interval $(B_2, B_1)\,, \quad B_1 < B_2$. Of course, if the researcher has further prior information about θ they may properly include it through any other prior distribution, exactly the same as in the traditional normal case.
- The posterior distribution for θ is

$$\pi_{a,b}\,(\theta \mid \mathbf{y}) \propto L_{a,b}(\theta \mid \mathbf{y})\pi(\theta) \propto L_{a,b}(\theta \mid \mathbf{y}), \tag{11}$$

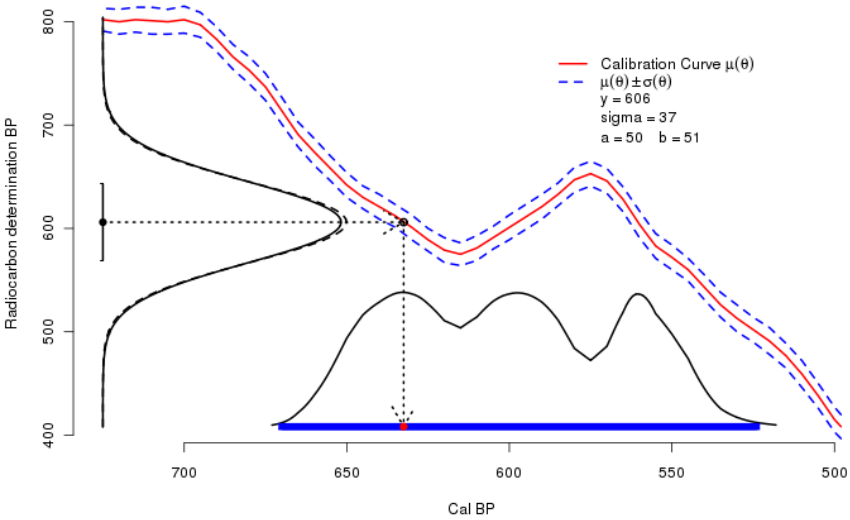
Single ^{14}C calibration



Single ^{14}C calibration



Single ^{14}C calibration



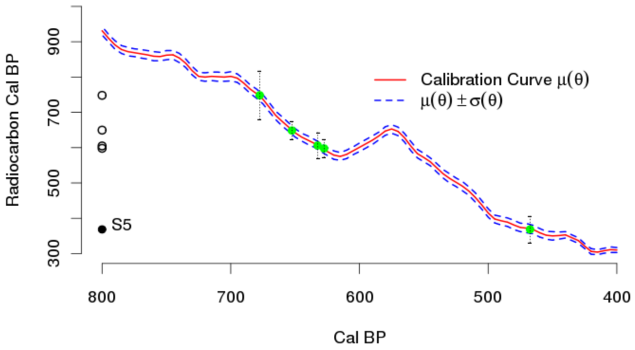
Simulated Example

We analyze a set of simulated of $m = 5$ radiocarbon observations. The parameter values are $\theta = 650$, $\sigma(\theta) = 12$, and $m = 5$.

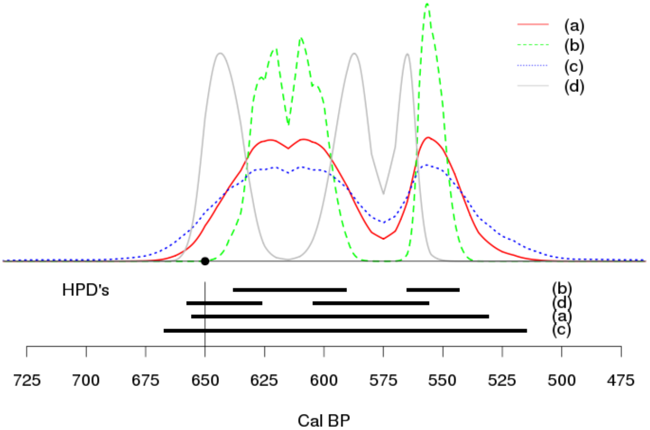
id	Determination
S1	649 ± 25
S2	598 ± 25
S3	748 ± 69
S4	606 ± 37
S5	368 ± 37

Simulated Example

The figure exhibits the simulated radiocarbon determinations plotted over the calibration curve. Note that there is an atypical observation (S5).



Simulated Example



Posterior densities for θ, and their corresponding %95 HPD credible sets, for (a) $\pi_{3,4}$, (b) π_N, (c) π_R and (d) π_N^* (normal model not including observation S5).

Simulated Example

- Note that π_N looks rougher, reproducing the wiggles in the calibration curve, while $\pi_{3,4}$ is smoother, and concentrated over the most likely region given the data.
- The effect of the outlying observation S5 causes the normal likelihood to shrink and shift to the right, leaving the true value for θ out of the 95% HPD region for π_N

Simulated Example

- Note that π_N looks rougher, reproducing the wiggles in the calibration curve, while $\pi_{3,4}$ is smoother, and concentrated over the most likely region given the data.
- The effect of the outlying observation S5 causes the normal likelihood to shrink and shift to the right, leaving the true value for θ out of the 95% HPD region for π_N

Simulated Example

- If we drop from the data S5 the resulting posterior arising from π^*_N is not more informative than π_N.
- $\pi_{3,4}$ is based on all the data and the heavy tails of the underlaying model ensure that we are properly including the information provided by possible extreme values.
- Our new approach is more cautious and results in wider smoother distributions.
- Shorter intervals may be obtained by dropping outlier determinations, but the gain in precision, given the amount of atypical information, is an illusion

Simulated Example

- If we drop from the data S5 the resulting posterior arising from π_N^* is not more informative than π_N.
- $\pi_{3,4}$ is based on all the data and the heavy tails of the underlaying model ensure that we are properly including the information provided by possible extreme values.
- Our new approach is more cautious and results in wider smoother distributions.
- Shorter intervals may be obtained by dropping outlier determinations, but the gain in precision, given the amount of atypical information, is an illusion

Simulated Example

- If we drop from the data S5 the resulting posterior arising from π_N^* is not more informative than π_N.
- $\pi_{3,4}$ is based on all the data and the heavy tails of the underlaying model ensure that we are properly including the information provided by possible extreme values.
- Our new approach is more cautious and results in wider smoother distributions.
- Shorter intervals may be obtained by dropping outlier determinations, but the gain in precision, given the amount of atypical information, is an illusion

Simulated Example

- If we drop from the data S5 the resulting posterior arising from π_N^* is not more informative than π_N.
- $\pi_{3,4}$ is based on all the data and the heavy tails of the underlaying model ensure that we are properly including the information provided by possible extreme values.
- Our new approach is more cautious and results in wider smoother distributions.
- Shorter intervals may be obtained by dropping outlier determinations, but the gain in precision, given the amount of atypical information, is an illusion

The Shroud of Turin

Table: Radiocarbon determinations for the 'Shroud of Turin'.

Laboratory	id	Determination
Arizona	A1	591 ± 30
	A2	690 ± 35
	A3	606 ± 41
	A4	701 ± 33
Oxford	O1	795 ± 65
	O2	730 ± 45
	O3	745 ± 55
Zurich	Z1	733 ± 61
	Z2	722 ± 56
	Z3	635 ± 57
	Z4	639 ± 45
	Z5	679 ± 51

The Shroud of Turin

Differences in the determination process suggest the use a different α for each laboratory. The likelihood is

$$L(\theta, \boldsymbol{\alpha} \mid \mathbf{y}) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} \left(2\pi\alpha_i\omega_{ij}^2(\theta)\right)^{-m_i/2} \exp\left\{-\frac{1}{2\alpha_i\omega_{ij}^2(\theta)}\left(y_{ij} - \mu(\theta)\right)^2\right\} \quad (12)$$

Where $n = 3$, $m_1 = 4$, $m_2 = 3$ and $m_3 = 5$. Integrating the likelihood function w.r.t. an InvGa prior for each α_i, the integrated likelihood is

$$L_{a,b}(\theta \mid \mathbf{y}) \quad \propto \quad \prod_{i=1}^{3} \mathrm{t}\left(\mathbf{y}_i \middle| \mu(\theta)\mathbf{1}_{m_i}, \Sigma_i(\theta)b/a, 2a\right), \quad (13)$$

where $\Sigma_i(\theta) = \mathrm{diag}\left(\omega_{i1}(\theta), \ldots, \omega_{m_i1}(\theta)\right)$. Thus, $L_{a,b}$ is the product of three multivariate **t** densities.

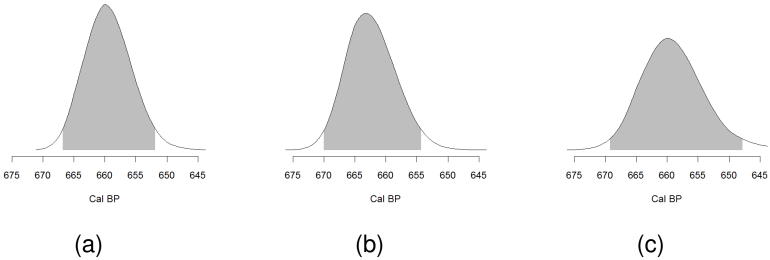


Figure: Posterior densities and 95% HPD regions (under shaded area) of θ for the Shroud of Turin data. (a) π_N, (b) π_N^* and (c) $\pi_{3,4}$.

Operating Characteristics

In order to analyze the performance of our proposed model we estimate with Monte Carlo simulation the "coverage probability" of 95% HPD sets

Table: Estimated coverage probability of the 95% HPD sets for different values of p.

Posterior Distribution	p (outlier probability)			
	0.0	0.01	0.05	0.1
$\pi_{a=3,b=4}$	0.9806	0.9700	0.9564	0.9332
$\pi_{a=15,b=16}$	0.9650	0.9372	0.8714	0.7936
π_N	0.9556	0.9198	0.8194	0.7014
π_R	0.9558	0.9550	0.9582	0.9692

Operating Characteristics

Since multimodal posteriors lead commonly to unconnected HPD regions, at each iteration the size of each HPD region (in cal. years) was counted and the average used as an indicator of the precision

Table: Average rounded count of the 95% HPD sets for different values of p.

Posterior Distribution	p (outlier probability)			
	0.0	0.01	0.05	0.1
$\pi_{a=3,b=4}$	72	74	82	92
$\pi_{a=15,b=16}$	65	65	66	68
π_N	61	61	58	56
π_R	73	80	102	130

More complex dating problems

- We may consider the above formulation as the hierarchical model

$$\mathbf{y} \longleftarrow \theta \longleftarrow (a, b)$$

  hyperparameters are introduced to model specific features of the data
- The general dating model is of the form

$$\mathbf{y} \longleftarrow \theta \longleftarrow \psi,$$

  were $\mathbf{y}$ is a generic representation of data obtained under diverse sampling schemes, θ is a vector of several calendar BP years and ψ contains (a, b) and quantities related to the phenomena being dated

More complex dating problems

- We may consider the above formulation as the hierarchical model

$$\mathbf{y} \longleftarrow \theta \longleftarrow (a, b)$$

hyperparameters are introduced to model specific features of the data

- The general dating model is of the form

$$\mathbf{y} \longleftarrow \boldsymbol{\theta} \longleftarrow \psi,$$

were $\mathbf{y}$ is a generic representation of data obtained under diverse sampling schemes, $\boldsymbol{\theta}$ is a vector of several calendar BP years and ψ contains (a, b) and quantities related to the phenomena being dated

More complex dating problems

- This representation is given by Christen (1994) and implemented in BCal, OxCal, Bpeat, etc.
- That is, for any dating problem we obtain a more robust analysis of radiocarbon data by substituting the normal model (2) by the t model in (9).
- Setting $a = i + 1, b = i$ for large i the normal model is recovered. Our working recommendation, both from conceptual and analytical perspectives, is $a = 3$ and $b = 4$.

More complex dating problems

- This representation is given by Christen (1994) and implemented in BCal, OxCal, Bpeat, etc.
- That is, for any dating problem we obtain a more robust analysis of radiocarbon data by substituting the normal model (2) by the t model in (9).
- Setting $a = i + 1, b = i$ for large i the normal model is recovered. Our working recommendation, both from conceptual and analytical perspectives, is $a = 3$ and $b = 4$.

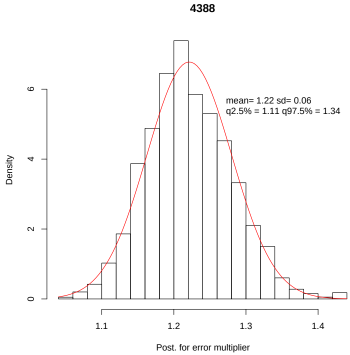
More complex dating problems

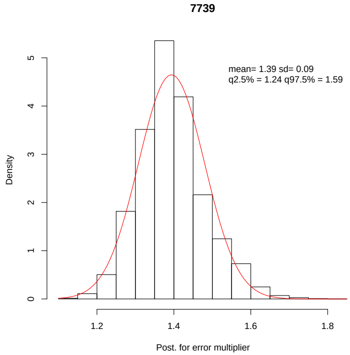
- This representation is given by Christen (1994) and implemented in BCal, OxCal, Bpeat, etc.
- That is, for any dating problem we obtain a more robust analysis of radiocarbon data by substituting the normal model (2) by the t model in (9).
- Setting $a = i + 1, b = i$ for large i the normal model is recovered. Our working recommendation, both from conceptual and analytical perspectives, is $a = 3$ and $b = 4$.

Posterior of $\sqrt{\alpha}$

Using MCMC we estimate

$$\pi\left(\alpha \mid \mathbf{y}\right) \propto \int_{\theta \in \Theta} p(\mathbf{y} \mid \theta, \alpha)\pi\left(\theta, \alpha\right) \mathrm{d}\theta.$$





Discussion

- The effect of outlier observations is reduced without additional parameters nor removing determinations.
- The posterior for θ has a smoother shape and the coverage of HPD regions is closer to the posterior $1 - \alpha$ probability.
- By plugging in the new model into the general statistical framework proposed by Christen (1994) and Buck et al. (2003) we obtain a method robust to outlier observations and other causes of overdispersed data, with far fewer parameters

Discussion

- The effect of outlier observations is reduced without additional parameters nor removing determinations.
- The posterior for θ has a smoother shape and the coverage of HPD regions is closer to the posterior $1 - \alpha$ probability.
- By plugging in the new model into the general statistical framework proposed by Christen (1994) and Buck et al. (2003) we obtain a method robust to outlier observations and other causes of overdispersed data, with far fewer parameters

Discussion

- The effect of outlier observations is reduced without additional parameters nor removing determinations.
- The posterior for θ has a smoother shape and the coverage of HPD regions is closer to the posterior $1 - \alpha$ probability.
- By plugging in the new model into the general statistical framework proposed by Christen (1994) and Buck et al. (2003) we obtain a method robust to outlier observations and other causes of overdispersed data, with far fewer parameters

Discussion

- As we increase either *a* or sample size, the proposed method will produce inferential results similar to those obtained under the normal model.
- Further research is needed regarding the sensitivity of posterior summaries to the choice of the prior α and the general applicability of our choice ($a = 3$ and $b = 4$).

Discussion

- As we increase either a or sample size, the proposed method will produce inferential results similar to those obtained under the normal model.
- Further research is needed regarding the sensitivity of posterior summaries to the choice of the prior α and the general applicability of our choice ($a = 3$ and $b = 4$).