Bayesian Calibration & Inverse Prediction

Mathematical Derivation and Assumptions

McClelland Lab, University College London

February 24, 2026

Contents

1 Introduction and Motivation

In analytical chemistry and bioassay work the standard calibration workflow is:

1. Prepare standards at known concentrations $x_1, \dots, x_n$.
2. Measure the instrument response y_i for each standard.
3. Fit a calibration curve $y = f(x;\, \boldsymbol{\theta})$.
4. For a new measurement y^*, *invert* the curve to estimate the unknown concentration x^*.

Step 4 is the **inverse prediction** (or *calibration*) problem. Classical frequentist approaches (e.g. Fieller's theorem, Wald intervals) provide approximate confidence intervals for x^*, but they rely on asymptotic normality, do not propagate the full parameter uncertainty, and become unreliable for nonlinear models.

The Bayesian framework solves this naturally: we obtain the full joint posterior distribution of the model parameters, then propagate every source of uncertainty—parameter uncertainty *and* measurement noise—through the inverse function to produce a posterior predictive distribution for x^*.

This document provides a self-contained mathematical derivation of the method implemented in the accompanying Streamlit application.

2 Notation

Symbol	Meaning
n	Number of calibration observations
$x_i \in \mathbb{R}$	Known standard value (concentration) for observation i
$y_i \in \mathbb{R}$	Measured instrument response for observation i
$\boldsymbol{x}^{\text{cal}} = (x_1, \dots, x_n)^\top$	Vector of calibration x-values
$\boldsymbol{y}^{\text{cal}} = (y_1, \dots, y_n)^\top$	Vector of calibration y-values
$f(x;\, \boldsymbol{\theta})$	Forward (calibration) function
$\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$	Model parameters
σ	Observation noise standard deviation (homoscedastic model)
σ_i	Observation-level noise s.d. (heteroscedastic model)
y^*	A new observed response for which we wish to find x^*
x^*	The unknown quantity to be estimated via inverse prediction

3 Forward Model

3.1 Homoscedastic Gaussian Likelihood

Assumption 1 (Functional form)**.** The calibration relationship is described by a known function $f : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$ that is continuous and differentiable almost everywhere with respect to both x and $\boldsymbol{\theta}$. The user specifies this function (e.g. $f(x;\, a, b) = a + bx$, or $f(x;\, a, b) = a\, e^{bx}$).

Assumption 2 (Additive Gaussian noise)**.** Observations are generated by

$$y_i = f(x_i;\, \boldsymbol{\theta}) + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0,\, \sigma^2), \qquad i = 1, \dots, n. \tag{1}$$

Assumption 3 (Independence)**.** The noise terms $\varepsilon_1, \dots, \varepsilon_n$ are mutually independent and independent of the true concentrations $x_1, \dots, x_n$.

Under these assumptions the likelihood function is

$$p(\boldsymbol{y}^{\text{cal}} \mid \boldsymbol{x}^{\text{cal}},\, \boldsymbol{\theta},\, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\Big[-\frac{(y_i - f(x_i;\, \boldsymbol{\theta}))^2}{2\sigma^2}\Big]. \tag{2}$$

Equivalently, in vector form with $\boldsymbol{\mu} = (f(x_1;\boldsymbol{\theta}), \ldots, f(x_n;\boldsymbol{\theta}))^\top$:

$$\boldsymbol{y}^{\text{cal}} \mid \boldsymbol{\theta}, \sigma \;\sim\; \mathcal{N}(\boldsymbol{\mu},\, \sigma^2 \mathbf{I}_n). \tag{3}$$

3.2 Heteroscedastic Gaussian Likelihood

Assumption 2 assumes constant variance. In many assays the noise variance grows with the signal level. We relax this with:

Assumption 4 (Non-constant variance)**.** The observation noise has standard deviation that depends on the predicted mean $\mu_i = f(x_i;\, \boldsymbol{\theta})$:

$$y_i = f(x_i;\, \boldsymbol{\theta}) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0,\, \sigma_i^2), \tag{4}$$

where $\sigma_i = g(\mu_i;\, \boldsymbol{\phi})$ for some variance function g with its own parameters $\boldsymbol{\phi}$.

The application offers two variance functions:

Linear variance model.

$$\sigma_i = \sigma_0 + \sigma_1\, |\mu_i|, \qquad \sigma_0 > 0,\ \sigma_1 \geq 0. \tag{5}$$

Here σ_0 captures a baseline noise floor and σ_1 captures the proportional increase in noise with signal magnitude.

Power-of-the-mean variance model.

$$\sigma_i = \sigma_0\, |\mu_i|^\delta, \qquad \sigma_0 > 0,\ \delta \geq 0. \tag{6}$$

When $\delta = 0$ this reduces to constant variance; when $\delta = 1$ it gives a coefficient-of-variation model (constant relative standard deviation).

The heteroscedastic likelihood is

$$p(\boldsymbol{y}^{\text{cal}} \mid \boldsymbol{x}^{\text{cal}},\, \boldsymbol{\theta},\, \boldsymbol{\phi}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\!\Big[-\frac{(y_i - \mu_i)^2}{2\sigma_i^2}\Big], \tag{7}$$

where each σ_i is given by (5) or (6).

4 Prior Distributions

Assumption 5 (Weakly informative priors)**.** We adopt the following independent prior distributions.

4.1 Homoscedastic Model

$$\theta_j \sim \mathcal{N}(0,\, 10^2), \qquad j = 1, \ldots, p, \tag{8}$$

$$\sigma \sim \mathcal{N}^+(0,\, 10^2), \tag{9}$$

where $\mathcal{N}^+(0,\, \tau^2)$ denotes the half-normal distribution (i.e. a normal distribution truncated to $(0, \infty)$).

4.2 Heteroscedastic Model

For the variance-function parameters:

$$\theta_j \sim \mathcal{N}(0,\, 10^2), \qquad j = 1, \dots, p, \tag{10}$$

$$\sigma_0 \sim \mathcal{N}^+(0,\, 10^2), \tag{11}$$

$$\sigma_1 \sim \mathcal{N}^+(0,\, 1^2), \qquad \text{(linear model)} \tag{12}$$

$$\delta \sim \mathcal{N}^+(0,\, 1^2). \qquad \text{(power model)} \tag{13}$$

Rationale. The normal priors on $\boldsymbol{\theta}$ are centred at zero with standard deviation 10, which is deliberately vague for standardised data. The half-normal priors on scale parameters ensure positivity while remaining weakly informative. These defaults work well across a wide range of calibration problems; users with strong domain knowledge may wish to substitute more informative priors in the source code.

Prior independence. All parameters are assumed *a priori* independent:

$$p(\boldsymbol{\theta}, \sigma) = \Big[\prod_{j=1}^{p} p(\theta_j)\Big]\, p(\sigma). \tag{14}$$

5 Posterior Distribution

By Bayes' theorem the joint posterior is

$$\boxed{p(\boldsymbol{\theta}, \sigma \mid \boldsymbol{y}^{\text{cal}}, \boldsymbol{x}^{\text{cal}}) = \frac{p(\boldsymbol{y}^{\text{cal}} \mid \boldsymbol{x}^{\text{cal}}, \boldsymbol{\theta}, \sigma)\; p(\boldsymbol{\theta}, \sigma)}{p(\boldsymbol{y}^{\text{cal}} \mid \boldsymbol{x}^{\text{cal}})},} \tag{15}$$

where the marginal likelihood (evidence) is

$$p(\boldsymbol{y}^{\text{cal}} \mid \boldsymbol{x}^{\text{cal}}) = \int p(\boldsymbol{y}^{\text{cal}} \mid \boldsymbol{x}^{\text{cal}}, \boldsymbol{\theta}, \sigma)\, p(\boldsymbol{\theta}, \sigma)\; \mathrm{d}\boldsymbol{\theta}\; \mathrm{d}\sigma. \tag{16}$$

For most nonlinear calibration functions f this integral is analytically intractable, so we resort to Markov chain Monte Carlo (MCMC) sampling.

6 MCMC Sampling via NUTS

6.1 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) augments the parameter space with auxiliary momentum variables $\boldsymbol{r} \in \mathbb{R}^d$ (where $d = p + 1$ for the homoscedastic model) and defines a joint density proportional to

$$p(\boldsymbol{\theta}, \sigma, \boldsymbol{r}) \propto \exp\big[-U(\boldsymbol{\theta}, \sigma) - \tfrac{1}{2}\boldsymbol{r}^\top \mathbf{M}^{-1} \boldsymbol{r}\big], \tag{17}$$

where the *potential energy* is

$$U(\boldsymbol{\theta}, \sigma) = -\log p(\boldsymbol{y}^{\text{cal}} \mid \boldsymbol{x}^{\text{cal}}, \boldsymbol{\theta}, \sigma) - \log p(\boldsymbol{\theta}, \sigma), \tag{18}$$

and $\mathbf{M}$ is a mass matrix (typically adapted during warm-up to approximate the posterior covariance).

HMC simulates Hamiltonian dynamics via the leapfrog integrator:

$$\begin{aligned} \boldsymbol{r}_{t+\epsilon/2} &= \boldsymbol{r}_t - \frac{\epsilon}{2}\,\nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}_t), && (19)\\ \boldsymbol{\theta}_{t+\epsilon} &= \boldsymbol{\theta}_t + \epsilon\,\mathbf{M}^{-1}\boldsymbol{r}_{t+\epsilon/2}, && (20)\\ \boldsymbol{r}_{t+\epsilon} &= \boldsymbol{r}_{t+\epsilon/2} - \frac{\epsilon}{2}\,\nabla_{\boldsymbol{\theta}} U(\boldsymbol{\theta}_{t+\epsilon}), && (21) \end{aligned}$$

for L leapfrog steps with step size ϵ. The proposal is accepted with probability $\min(1,\, \exp(-\Delta H))$ where ΔH is the change in Hamiltonian.

Assumption 6 (Differentiability for HMC)**.** The log-posterior $\log p(\boldsymbol{\theta}, \sigma \mid \boldsymbol{y}^{\text{cal}}, \boldsymbol{x}^{\text{cal}})$ is differentiable with respect to all continuous parameters. This is required for the gradient-based HMC sampler. SymPy-parsed expressions are analytically differentiable, and PyTensor provides automatic differentiation.

6.2 The No-U-Turn Sampler (NUTS)

NUTS [Hoffman and Gelman, 2014] eliminates the need to hand-tune L and ϵ. It builds a balanced binary tree of leapfrog steps, doubling the trajectory length until a "U-turn" criterion is met:

$$\boldsymbol{r} \cdot (\boldsymbol{\theta}^+ - \boldsymbol{\theta}^-) < 0 \quad \text{or} \quad \boldsymbol{r} \cdot (\boldsymbol{\theta}^- - \boldsymbol{\theta}^+) < 0, \tag{22}$$

where $\boldsymbol{\theta}^+$ and $\boldsymbol{\theta}^-$ are the forward and backward ends of the trajectory. A multinomial sampling scheme selects the next state from the trajectory, weighted by the un-normalised density.

The step size ϵ is adapted during the warm-up phase using dual averaging [Nesterov, 2009] to target an acceptance probability of approximately 0.8.

6.3 Warm-up and Adaptation

During the warm-up (tuning) phase:

1. The step size ϵ is adapted.
2. The mass matrix $\mathbf{M}$ is estimated from sample covariance.
3. Warm-up draws are *discarded* and not used for inference.

After warm-up, S draws $\{(\boldsymbol{\theta}^{(s)}, \sigma^{(s)})\}_{s=1}^{S}$ are collected from each of C independent chains, giving a total of $N = S \times C$ posterior draws.

6.4 Convergence Diagnostics

The application reports the ArviZ summary statistics including:

- $\hat{R}$ (Gelman–Rubin statistic): values $\lesssim 1.01$ indicate convergence.
- Effective sample size (n_{eff}): the number of effectively independent draws, accounting for autocorrelation.
- Monte Carlo standard error (MCSE): the standard error of the posterior mean estimate.

7 Inverse Prediction

7.1 Problem Statement

Given a new observed response y^*, we seek the posterior predictive distribution of the unknown x^* that produced it.

Assumption 7 (New observation model)**.** The new measurement follows the same generative process as the calibration data:

$$y^* = f(x^*; \boldsymbol{\theta}) + \varepsilon^*, \qquad \varepsilon^* \sim \mathcal{N}(0,\, \sigma^{*2}), \tag{23}$$

where $\sigma^* = \sigma$ in the homoscedastic case, or $\sigma^* = g(f(x^*; \boldsymbol{\theta});\, \boldsymbol{\phi})$ in the heteroscedastic case.

7.2 Posterior Predictive Distribution of x^*

The key quantity is

$$\boxed{p(x^* \mid y^*,\, \boldsymbol{y}^{\text{cal}},\, \boldsymbol{x}^{\text{cal}}) = \int p(x^* \mid y^*,\, \boldsymbol{\theta},\, \sigma)\; p(\boldsymbol{\theta}, \sigma \mid \boldsymbol{y}^{\text{cal}}, \boldsymbol{x}^{\text{cal}})\; \mathrm{d}\boldsymbol{\theta}\; \mathrm{d}\sigma.} \tag{24}$$

This integral marginalises over the full posterior uncertainty in $\boldsymbol{\theta}$ and σ. We approximate it by Monte Carlo:

Proposition 1 (Monte Carlo inverse prediction)**.** *For each posterior draw $(\boldsymbol{\theta}^{(s)}, \sigma^{(s)})$, $s = 1, \ldots, N$:*

1. *Sample a "noisy" response: $\tilde{y}^{(s)} = y^* + \epsilon^{(s)}$ where $\epsilon^{(s)} \sim \mathcal{N}(0,\, (\sigma^{*(s)})^2)$.*
   In the homoscedastic case $\sigma^{*(s)} = \sigma^{(s)}$.
   In the heteroscedastic case we approximate $\sigma^{*(s)} \approx g(y^*;\, \boldsymbol{\phi}^{(s)})$, using y^* as a proxy for $|f(x^*; \boldsymbol{\theta}^{(s)})|$.
2. *Invert the forward model: $x^{*(s)} = f^{-1}(\tilde{y}^{(s)};\, \boldsymbol{\theta}^{(s)})$.*

The collection $\{x^{*(s)}\}_{s=1}^{N}$ is a sample from the posterior predictive distribution $p(x^* \mid y^*, \boldsymbol{y}^{\text{cal}}, \boldsymbol{x}^{\text{cal}})$.

Proof. By the law of total probability:

$$p(x^* \mid y^*, \boldsymbol{y}^{\text{cal}}, \boldsymbol{x}^{\text{cal}}) = \int p(x^* \mid y^*, \boldsymbol{\theta}, \sigma)\, p(\boldsymbol{\theta}, \sigma \mid \boldsymbol{y}^{\text{cal}}, \boldsymbol{x}^{\text{cal}})\; \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\sigma. \tag{25}$$

For a given $(\boldsymbol{\theta}, \sigma)$, the forward model is deterministic, so $x^* = f^{-1}(y^* - \varepsilon^*;\, \boldsymbol{\theta})$ where $\varepsilon^* \sim \mathcal{N}(0, \sigma^2)$. Sampling $\tilde{y}^{(s)} = y^* + \epsilon^{(s)}$ with $\epsilon^{(s)} \sim \mathcal{N}(0, \sigma^{(s)2})$ is equivalent to sampling the "true" response $y^* - \varepsilon^*$ from the predictive distribution at that draw. Since $(\boldsymbol{\theta}^{(s)}, \sigma^{(s)})$ are draws from $p(\boldsymbol{\theta}, \sigma \mid \boldsymbol{y}^{\text{cal}}, \boldsymbol{x}^{\text{cal}})$, the composition yields draws from the marginal (24). □

7.3 Inversion Methods

Symbolic inverse. When f is algebraically invertible with respect to x, SymPy computes the closed-form solution $x = f^{-1}(y;\, \boldsymbol{\theta})$. If multiple real solutions exist (e.g. for a quadratic), the solution closest to the centroid of the calibration x-values is selected.

Numerical inverse. When no closed-form inverse exists, Brent's method is applied to find the root of

$$h(x) = f(x;\, \boldsymbol{\theta}^{(s)}) - \tilde{y}^{(s)} = 0 \tag{26}$$

over a search interval $[x_{\min} - 3\Delta x,\ x_{\max} + 3\Delta x]$ where $\Delta x = x_{\max} - x_{\min}$. Brent's method combines bisection, secant, and inverse quadratic interpolation, guaranteeing convergence for continuous functions with a sign change.

7.4 Credible Intervals

From the N draws $\{x^{*(s)}\}$, after removing any non-finite values (from failed inversions), the $100(1-\alpha)\%$ equal-tailed credible interval is

$$\text{CI}_{1-\alpha} = \big[Q_{\alpha/2},\ Q_{1-\alpha/2}\big], \tag{27}$$

where Q_q denotes the q-th quantile of the empirical distribution.

Point estimates reported:

$$\hat{x}_{\text{median}} = Q_{0.5}, \tag{28}$$

$$\hat{x}_{\text{mean}} = \frac{1}{N}\sum_{s=1}^{N} x^{*(s)}, \tag{29}$$

$$\hat{\sigma}_x = \sqrt{\frac{1}{N-1}\sum_{s=1}^{N}(x^{*(s)} - \hat{x}_{\text{mean}})^2}. \tag{30}$$

8 Residual Diagnostics

After fitting the homoscedastic model, the application performs residual diagnostics to assess model adequacy.

8.1 Residual Computation

The residuals are computed at the posterior median parameter values $\hat{\boldsymbol{\theta}} = \text{median}\{\boldsymbol{\theta}^{(s)}\}$:

$$e_i = y_i - f(x_i; \hat{\boldsymbol{\theta}}), \qquad i = 1, \ldots, n. \tag{31}$$

8.2 Breusch–Pagan Test for Heteroscedasticity

The Breusch–Pagan test [Breusch and Pagan, 1979] tests

$$H_0 : \ \text{Var}(\varepsilon_i) = \sigma^2 \ \ \forall\, i \qquad \text{vs.} \qquad H_1 : \ \text{Var}(\varepsilon_i) = h(\boldsymbol{z}_i^\top \boldsymbol{\gamma}), \tag{32}$$

where $\boldsymbol{z}_i$ is a vector of regressors (here, the fitted values $\hat{y}_i$).

The test procedure:

1. Compute squared residuals e_i^2.
2. Regress e_i^2 on $\boldsymbol{z}_i$ by OLS.
3. The LM statistic is $\text{LM} = \frac{1}{2}\,\text{ESS}$, where ESS is the explained sum of squares of this auxiliary regression.
4. Under H_0, $\text{LM} \sim \chi^2_k$ where k is the number of regressors (here $k = 1$).

A p-value below 0.05 indicates significant heteroscedasticity.

8.3 Wald–Wolfowitz Runs Test for Randomness

The runs test [Wald and Wolfowitz, 1940] assesses whether the sequence of residual signs (positive/negative) is random.

Let n_+ and n_- be the counts of positive and negative residuals, and R the number of runs (maximal consecutive subsequences of the same sign). Under H_0 (random ordering):

$$\mathbb{E}[R] = \frac{2n_+n_-}{n_+ + n_-} + 1, \tag{33}$$

$$\text{Var}(R) = \frac{2n_+n_-(2n_+n_- - n_+ - n_-)}{(n_+ + n_-)^2(n_+ + n_- - 1)}. \tag{34}$$

The standardised test statistic $Z = (R - \mathbb{E}[R])/\sqrt{\text{Var}(R)}$ is approximately standard normal for large n.

A p-value below 0.05 suggests systematic (non-random) structure in the residuals, indicating potential model misspecification.

9 Heteroscedastic Inverse Prediction

When the user elects to account for heteroscedasticity, the full parameter vector becomes $(\boldsymbol{\theta}, \boldsymbol{\phi})$ where $\boldsymbol{\phi} = (\sigma_0, \sigma_1)$ or $\boldsymbol{\phi} = (\sigma_0, \delta)$.

The inverse prediction procedure (Proposition 1) is modified as follows. For each posterior draw s:

1. Compute the approximate noise level at y^*:

$$\sigma^{*(s)} = g\big(|y^*|;\, \boldsymbol{\phi}^{(s)}\big), \tag{35}$$

   using $|y^*|$ as a proxy for $|\mu_i|$ (since $y^* \approx f(x^*; \boldsymbol{\theta})$ up to noise).
2. Sample $\tilde{y}^{(s)} = y^* + \epsilon^{(s)}$ with $\epsilon^{(s)} \sim \mathcal{N}(0, (\sigma^{*(s)})^2)$.
3. Invert: $x^{*(s)} = f^{-1}(\tilde{y}^{(s)};\, \boldsymbol{\theta}^{(s)})$.

This correctly propagates the heteroscedastic noise structure into the inverse-prediction uncertainty.

10 Summary of Assumptions

For clarity we collect all assumptions:

1 The user-specified functional form $f(x; \boldsymbol{\theta})$ is a correct (or adequate) description of the calibration relationship.

2 Observation noise is additive and Gaussian with constant variance (homoscedastic model) or signal-dependent variance (heteroscedastic model, Assumption 4).

3 Observations are mutually independent conditional on the parameters.

5 Prior distributions are weakly informative (zero-centred normals for location parameters, half-normals for scale parameters).

6 The log-posterior is differentiable w.r.t. all continuous parameters (required for NUTS).

7 New observations follow the same generative process as the calibration data (no distribution shift).

When assumptions may be violated.

- If the functional form is wrong, residual diagnostics (runs test) should flag systematic patterns.
- If noise is non-Gaussian (e.g. heavy-tailed), the posterior and credible intervals may undercover. Robust likelihood extensions (e.g. Student-t) can be implemented as future work.
- If observations are correlated (e.g. temporal drift), the independence assumption is violated and the credible intervals will be too narrow.

11 Software Implementation Notes

- **Equation parsing:** SymPy's parser with implicit multiplication and `convert_xor` transformations.
- **Automatic differentiation:** PyTensor (the tensor backend of PyMC) provides exact gradients for HMC/NUTS.
- **Sampling:** PyMC v5.10+ with the default NUTS sampler.
- **Diagnostics:** ArviZ for convergence statistics; statsmodels for the Breusch–Pagan and runs tests.
- **Numerical inversion:** SciPy's `brentq` with tolerance 10^{-10}.

References

T. S. Breusch and A. R. Pagan. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5):1287–1294, 1979.

M. D. Hoffman and A. Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1593–1623, 2014.

Y. Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120(1):221–259, 2009.

A. Wald and J. Wolfowitz. On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, 11(2):147–162, 1940.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis.* Chapman & Hall/CRC, 3rd edition, 2013.

J. Salvatier, T. V. Wiecki, and C. Fonnesbeck. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55, 2016.

R. M. Neal. MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, editors, *Handbook of Markov Chain Monte Carlo*, chapter 5. Chapman & Hall/CRC, 2011.

C. Eisenhart. The interpretation of certain regression methods and their use in biological and industrial research. *The Annals of Mathematical Statistics*, 10(2):162–186, 1939.

B. Hoadley. A Bayesian look at inverse linear regression. *Journal of the American Statistical Association*, 65(329):356–369, 1970.