Conditional false-alarm probabilities in Bayesian haemoglobin models of the Athlete Biological Passport: a normal-inverse-gamma alternative

Zuzana Kováčová and Richard Kollár
Department of Applied Mathematics and Statistics



Why was the ABP introduced?

- **Challenge:** Blood doping methods, such as autologous transfusions, are difficult to detect directly.
- **Approach:** The Athlete Biological Passport (ABP) monitors blood markers over time to detect unusual changes [2].

The ABP combines **population information** with an athlete's **previous measurements** $D_n = (X_1, \ldots, X_n)$ to set individual predictive limits.

False alarm: A measurement from a stable, non-doping athlete falls outside these limits.

What do we investigate?

1. Does population-level calibration guarantee the nominal false-alarm probability for every athlete?
2. Can alternative prior assumptions yield an analytically tractable model with comparable false-alarm behaviour?

Which Bayesian models do we compare?

All three models assume $X_j \mid \mu, S \overset{\text{iid}}{\sim} \mathcal{N}(\mu, S)$.

Normal–lognormal (NL), reference [1]

$$\mu \sim \mathcal{N}(m_0, \tau_0^2), \quad \sigma \sim \mathrm{LogN}(\phi, \omega^2), \quad \mu \perp \sigma.$$

Predictive: Normal mixture (numerical integration).

Fixed-variance (FV), benchmark

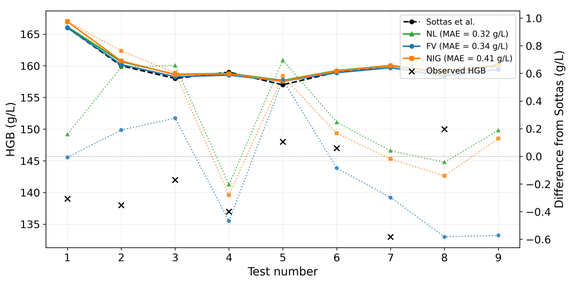
$$\mu \sim \mathcal{N}(m_0, \tau_0^2), \quad \sigma = s.$$

Predictive: $X_{n+1} \mid D_n \sim \mathcal{N}(m_n, s^2 + V_n)$.
Calibration: $s = \mathbb{E}_{\mathrm{NL}}[\sigma]$.

Normal–inverse-gamma (NIG), alternative

$$S \sim \mathrm{IG}(a_0, b_0), \quad \mu \mid S \sim \mathcal{N}(m_0, S/\kappa_0).$$

Predictive: $X_{n+1} \mid D_n \sim t_{2a_n}(m_n, q_n)$, $q_n^2 = \frac{b_n(\kappa_n+1)}{a_n\kappa_n}$.
Calibration: Match the NL prior variance of μ and the prior mean and variance of $S = \sigma^2$.



Reconstructed the one-sided 99.9% upper predictive limit [1].

How do we evaluate the models?

Assumption and conditional false-alarm probability:

$$X_j \overset{\text{iid}}{\sim} \mathcal{N}(\mu_*, \sigma_*^2), \quad r_n = \mathbb{P}_{\mu_*,\sigma_*}(X_{n+1} \notin I_n),$$

where I_n is the central 99% predictive interval based on n previous measurements, and r_n averages over possible measurement histories.

Prior parameters [1]: $\tau_0 = 7.6\,\mathrm{g/L}$, $\phi = 1.7$, $\omega = 0.10$. m_0 depends on six factors: sex, ethnic origin, altitude, age, sport discipline, and analyser type.

Settings:

$$d = \frac{\mu_* - m_0}{\tau_0}, \quad r = \frac{\sigma_*}{s}, \quad \alpha = 0.01.$$

Evaluation of r_n: FV analytical, NIG quadrature, NL Rao–Blackwellized MC.

What does theory tell us?

1. Population-level calibration
Under the model's joint prior π and its assumptions,

$$\boxed{\mathbb{E}_\pi[r_n(\mu, \sigma)] = \alpha} \qquad \text{for every } n.$$

This does not guarantee $r_n(\mu_*, \sigma_*) = \alpha$ for every athlete.

2. Asymptotic behaviour
Under stationary Gaussian observations,

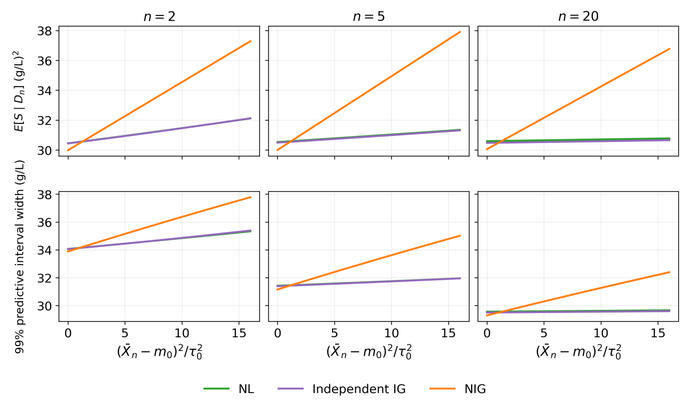
$$\lim_{n\to\infty} r_n^{\mathrm{FV}} = 2[1 - \Phi(z_{0.995}/r)], \qquad \lim_{n\to\infty} r_n^{\mathrm{NL}} = \lim_{n\to\infty} r_n^{\mathrm{NIG}} = \alpha.$$

Why do NL and NIG behave differently?

Prior differences:

- NL: Independent μ and S; marginals: $\mu \sim \mathcal{N}$; $\sigma, S \sim \mathrm{LogN}$.
- NIG: Dependent μ and S; marginals: $\mu \sim t_{2a_0}$, $S \sim \mathrm{IG}$.

Variance learning: For $t = (\bar{X}_n - m_0)^2$ and fixed Q_n, the sensitivity $\partial\mathbb{E}[S \mid D_n]/\partial t$ is asymptotically $O_p(n^{-2})$ in NL, versus $O(n^{-1})$ in NIG, under stationary Gaussian observations.



Posterior expected variance (top) and central 99% predictive interval width (bottom) versus $(\bar{X}_n - m_0)^2/\tau_0^2$, with $Q_n = (n-1)s^2$ fixed.

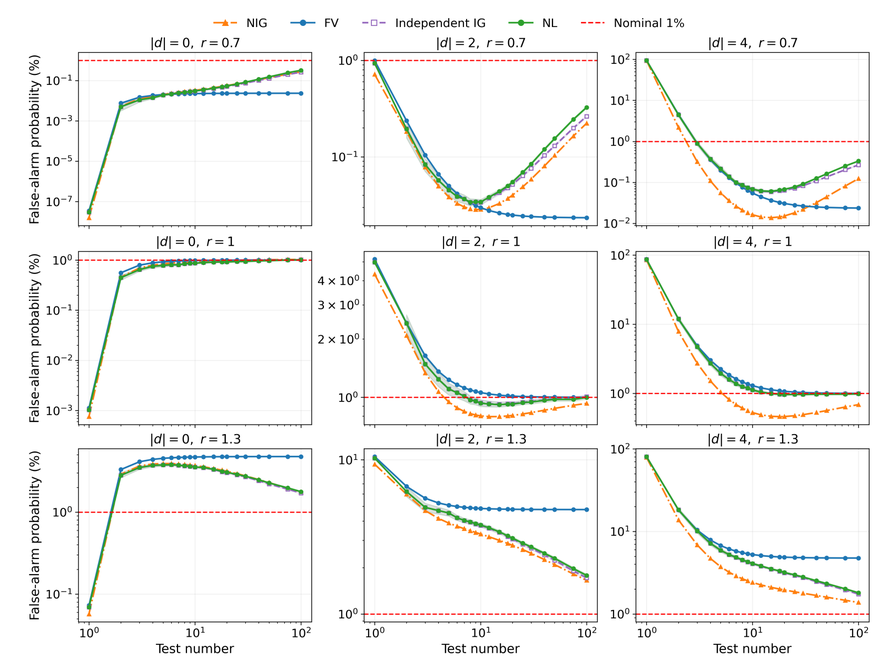
Effect of marginal priors:

- μ: Normal vs. Student-t ($2a_0 \approx 53$ df, close to normal).
- S: Lognormal vs. inverse-gamma (independent-IG and NL yield nearly identical false-alarm probabilities).

Selected references

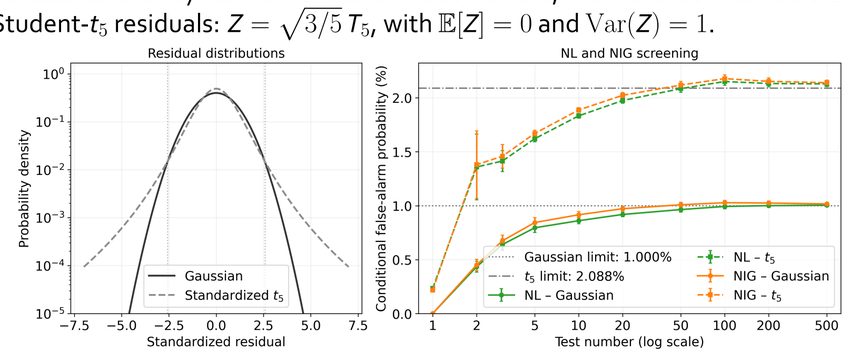
[1] Sottas, P.-E. et al. (2008). A forensic approach to the interpretation of blood doping markers. *Law, Probability and Risk*.
[2] WADA (2023). *Athlete Biological Passport Operating Guidelines*, v9.0.

How does individual false-alarm risk vary?



What if measurements are not Gaussian?

To test sensitivity to non-Gaussian observations, we consider standardized Student-t_5 residuals: $Z = \sqrt{3/5}\,T_5$, with $\mathbb{E}[Z] = 0$ and $\mathrm{Var}(Z) = 1$.



For NL and NIG (Gaussian likelihoods), $\lim_{n\to\infty} r_n = \mathbb{P}(|Z| > z_{0.995}) \approx \mathbf{2.09}\%$, versus the nominal $\alpha = 1\%$.

Heavy tails can cause persistent miscalibration even when the models learn the athlete's mean and variance.

Conclusions

Key findings

- **Individual risk:** False-alarm probabilities can differ substantially from the nominal level. Larger discrepancies from the population prior may require more measurements to reduce its influence.
- **Mean offset ($|d|$):** In FV, risk increases monotonically with $|d|$ for fixed n and r. In NL and NIG, mean deviations can also be interpreted as higher variability, widening predictive intervals.
- **Variability (r):** FV assumes a fixed variance and cannot adapt to individual variability. Its limiting false-alarm probability is below α for $r < 1$ and above α for $r > 1$. NL and NIG learn individual variability, and their false-alarm probabilities converge to α for every $r > 0$ under Gaussian sampling.
- **Model comparison:** Across the examined settings, neither NL nor NIG consistently yields false-alarm probabilities closer to α.

Limitations and outlook

- **NL vs NIG:** NIG provides an analytically tractable alternative to NL. Determining which prior specification is more appropriate requires real-world validation of both specificity and detection sensitivity.
- **Assumptions:** Athlete-specific mean and variance are assumed constant, with independent observations. Heavy-tailed observations can cause persistent miscalibration.
- **Scope:** Simulated single-marker screening, not the complete operational ABP or doping adjudication.

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
Univerzita Komenského v Bratislave

MATFYZ CONNECTIONS