Stochastic space-time models, non-trivial observation mechanisms, and practical inference

Finn Lindgren (finn.lindgren@ed.ac.uk)

THE UNIVERSITY *of* EDINBURGH

Edinburgh 2017-11-07

EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

- GRF+Spectral+Properties+GMRF
- Bayes+Linear obs+Gaussian mixture posterior (ILA)
- Non-Gaussian observations; transformations, Point process, INLA
- Non-linear predictors; location+group size+detection probability

EUSTACE

Covariance functions and SPDEs

The Matérn covariance family on $\mathbb{R}^d$

$$\mathrm{Cov}(x(\mathbf{0}), x(\boldsymbol{s})) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \|\boldsymbol{s}\|)^\nu K_\nu(\kappa \|\boldsymbol{s}\|)$$

Scale $\kappa > 0$, smoothness $\nu > 0$, variance $\sigma^2 > 0$

Whittle (1954, 1963): Matérn as SPDE solution

Matérn fields are the stationary solutions to the SPDE

$$\left(\kappa^2 - \nabla \cdot \nabla\right)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \alpha = \nu + d/2$$

$\mathcal{W}(\cdot)$ white noise, $\nabla \cdot \nabla = \sum_{i=1}^d \frac{\partial^2}{\partial s_i^2}$, $\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)\kappa^{2\nu}(4\pi)^{d/2}}$

White noise has $K(\mathbf{s}, \mathbf{s}') = \delta(\mathbf{s} - \mathbf{s}')$. Do not confuse with independent noise, $K(\mathbf{s}, \mathbf{s}') = \mathbb{I}(\mathbf{s} = \mathbf{s}')$, which has non-integrable realisations.

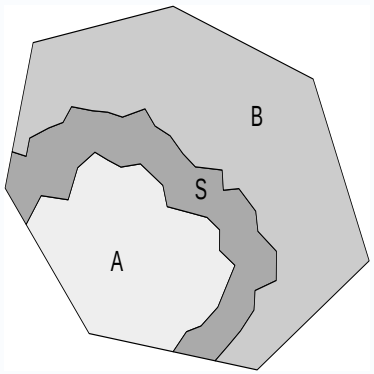
GMRFs: Gaussian Markov random fields

Continuous domain GMRFs

If $x(\boldsymbol{s})$ is a (stationary) Gaussian random field on Ω with covariance kernel $K(\boldsymbol{s}, \boldsymbol{s}')$, it fulfills the *global Markov property*

$$\{x(\mathcal{A}) \perp x(\mathcal{B}) | x(\mathcal{S}), \text{ for all } \mathcal{AB}\text{-separating sets } \mathcal{S} \subset \Omega\}$$

if the power spectrum can be written as $1/S_x(\boldsymbol{\omega}) =$ polynomial in $\boldsymbol{\omega}$, for some polynomial order p. (Rozanov, 1977)



Generally: Markov iff the precision operator $\mathcal{Q} = \mathcal{R}^{-1}$ is local.

Discrete domain GMRFs

$\boldsymbol{x} = (x_1, \ldots, x_n) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ is Markov with respect to a neighbourhood structure $\{\mathcal{N}_i, i = 1, \ldots, n\}$ if $Q_{ij} = 0$ whenever $j \neq \mathcal{N}_i \cup i$.

- Continuous domain basis representation with Markov weights: $x(\boldsymbol{s}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{s}) x_k$
- Many stochastic PDE solutions are Markov in continuous space, and can be approximated by Markov weights on local basis functions.

Matérn driven heat equation variations

- The iterated heat equation is a simple non-separable space-time SPDE family:

$$(\kappa^2 - \Delta)^{\gamma/2} \left[\phi \frac{\partial}{\partial t} + (\kappa^2 - \Delta)^{\alpha/2}\right]^{\beta} x(\mathbf{s}, t) = \mathcal{W}(\mathbf{s}, t)/\tau$$

- Fourier spectra are based on eigenfunctions $e_{\boldsymbol{\omega}}(\mathbf{s})$ of $-\Delta$.
  On $\mathbb{R}^2$, $-\Delta e_{\boldsymbol{\omega}}(\mathbf{s}) = \|\boldsymbol{\omega}\|^2 e_{\boldsymbol{\omega}}(\mathbf{s})$, and $e_{\boldsymbol{\omega}}$ are harmonic functions.
  On $\mathbb{S}^2$, $-\Delta e_k(\mathbf{s}) = \lambda_k e_k(\mathbf{s}) = k(k+1) e_k(\mathbf{s})$, and e_k are spherical harmonics.
- The power spectrum on $\mathbb{R}^2 \times \mathbb{R}$ is

$$S_x(\boldsymbol{\omega}_s, \omega_t) = \frac{1}{\tau^2 (2\pi)^3 (\kappa^2 + \|\boldsymbol{\omega}_s\|^2)^{\gamma} \left[\phi^2 \omega_t^2 + (\kappa^2 + \|\boldsymbol{\omega}_s\|^2)^{\alpha}\right]^{\beta}}$$

  which leads to Matérn covariances marginally in space, and in time for each spatial frequency.
- The finite element approximation has precision matrix structure

$$\boldsymbol{Q} = \sum_{i=0}^{\alpha+\beta+\gamma} \boldsymbol{M}_i^{[t]} \otimes \boldsymbol{M}_i^{[\mathbf{s}]}$$

  even, e.g., if κ is spatially varying.

EUSTACE

Linear models

Statistical linear models can be formulated as Bayesian hierarchical models, with a simple network of conditional prior densities:

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta}) \quad \text{(variance parameters)}$$
$$\boldsymbol{x}|\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x(\boldsymbol{\theta})^{-1}) \quad \text{(latent Gaussian variables)}$$
$$\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{Q}_{y|x}(\boldsymbol{\theta})^{-1}) \quad \text{(observed linear combinations)}$$

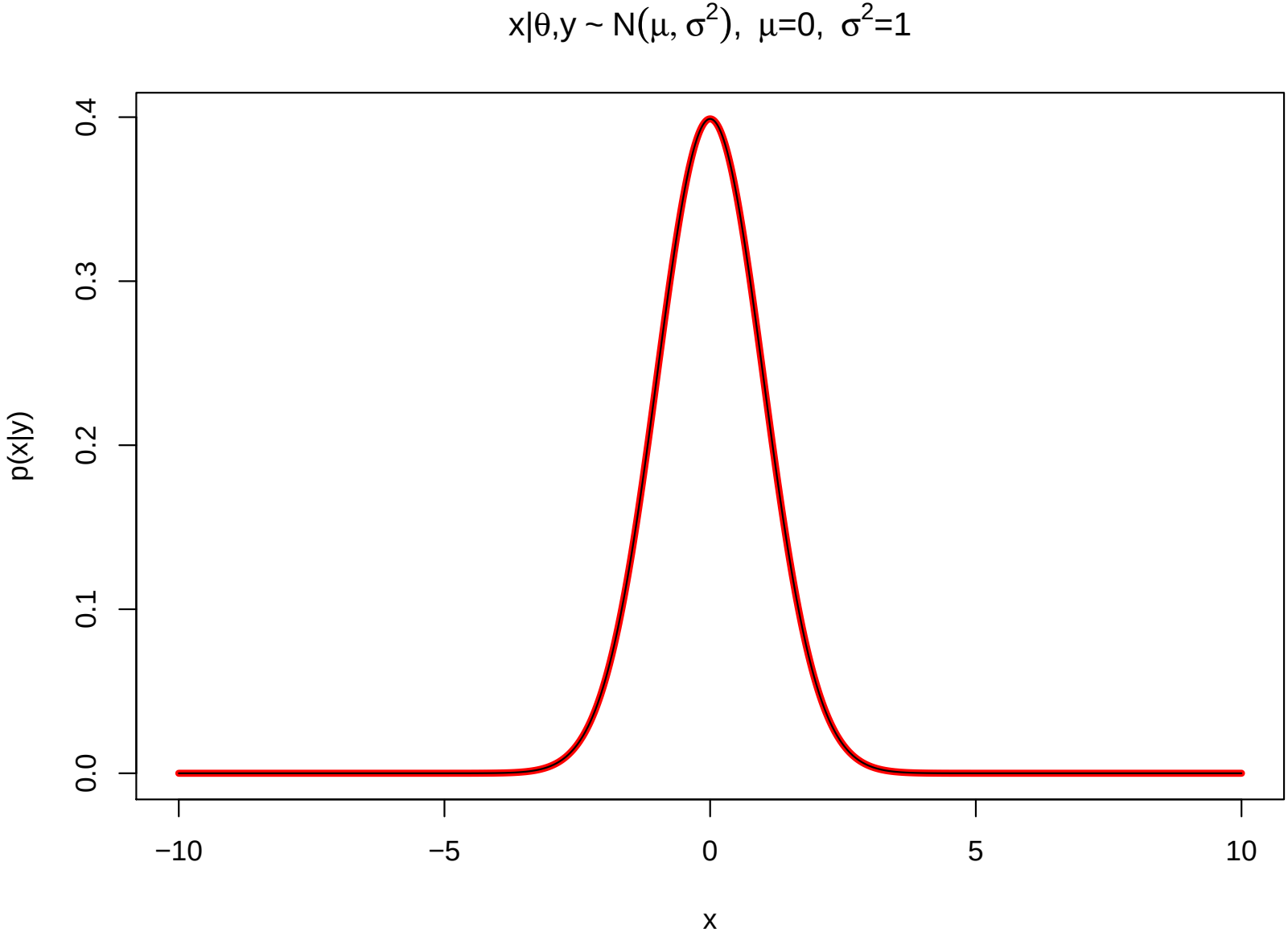
Inference about $\boldsymbol{\theta}$ and $\boldsymbol{x}$ is based on the posterior densities

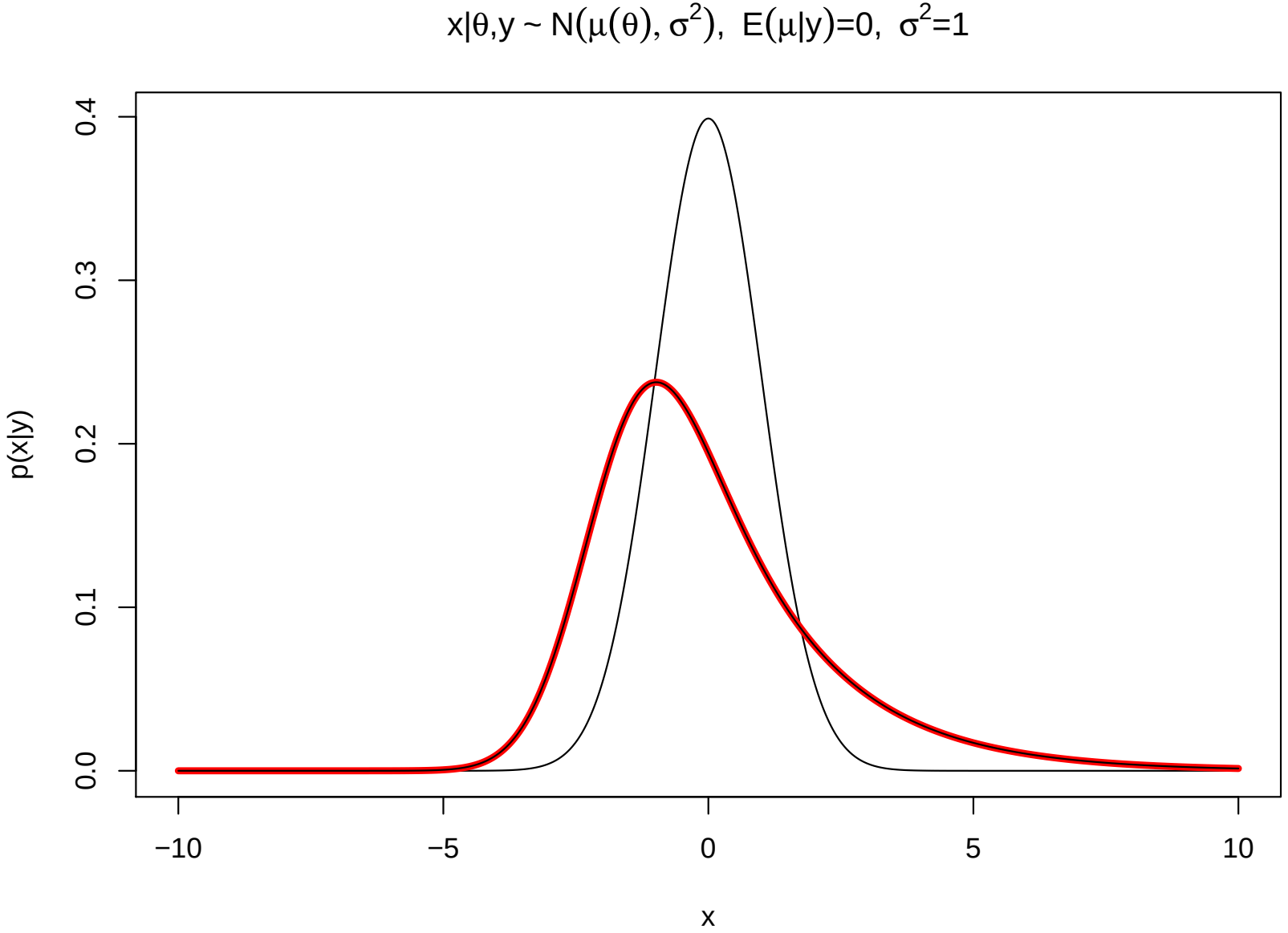
$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \int p(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})\, \mathrm{d}\boldsymbol{x} \quad \text{(Soon: How can we compute this?)}$$
$$p(\boldsymbol{x}|\boldsymbol{y}) = \int p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{y})\, \mathrm{d}\boldsymbol{\theta} \quad \text{(continuous mixture distribution)}$$

where $\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{x|y}, \boldsymbol{Q}_{x|y}^{-1})$, with

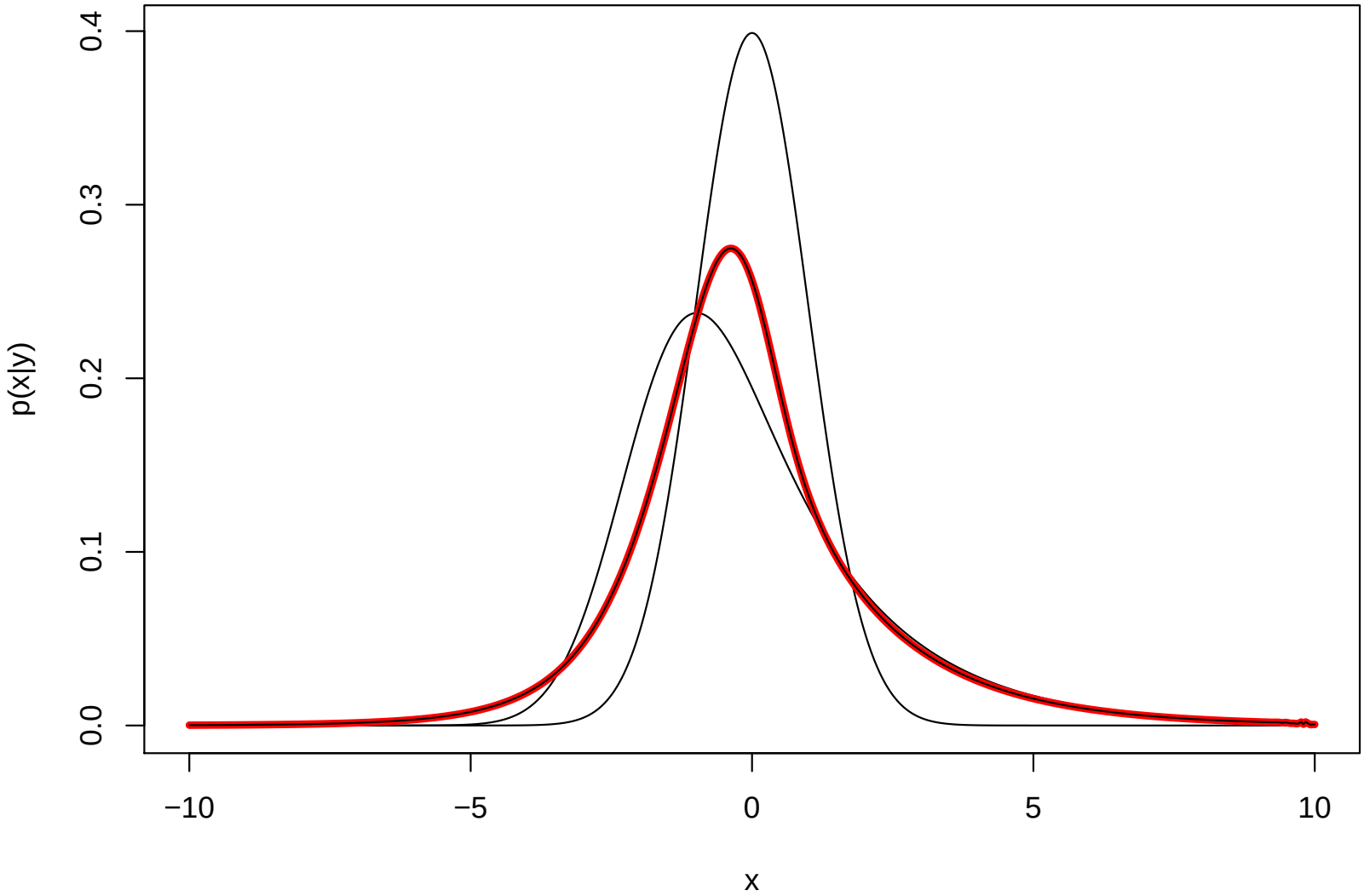
$$\boldsymbol{Q}_{x|y} = \boldsymbol{Q}_x + \boldsymbol{A}^\top \boldsymbol{Q}_{y|x} \boldsymbol{A}, \quad \boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \boldsymbol{Q}_{x|y}^{-1} \boldsymbol{A}^\top \boldsymbol{Q}_{y|x} (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}_x).$$

EUSTACE





$x|\theta,y \sim N(\mu(\theta), \sigma(\theta)^2),\ E(\mu|y)=0,\ E(\sigma^2|y)=1$



Integration techniques for linear models

For $p(\boldsymbol{\theta}|\boldsymbol{y})$, we don't need to actually integrate. For *any* $\boldsymbol{x}^*$,

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{y}) &= \frac{p(\boldsymbol{\theta},\boldsymbol{y})}{p(\boldsymbol{y})} \left.\frac{p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}{p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*} \\ &= \frac{p(\boldsymbol{\theta})p(\boldsymbol{x}^*|\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})}{p(\boldsymbol{y})p(\boldsymbol{x}^*|\boldsymbol{\theta},\boldsymbol{y})} \end{aligned}$$

All components are known except for the normalisation factor $p(\boldsymbol{y})$.
Use numerical optimisation to find the mode w.r.t. $\boldsymbol{\theta}$ and place integration points around it. Then

$$p(\boldsymbol{x}|\boldsymbol{y}) \approx \sum_k p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\boldsymbol{y})w_k$$

where $w_k^{-1} \equiv \sum_k p(\boldsymbol{\theta}_k|\boldsymbol{y})$ for a regular integration grid.

EUSTACE

Integration techniques for generalised linear models

If $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x})$ is non-Gaussian or the expectation if non-linear in $\boldsymbol{x}$, $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ is also non-Gaussian.

Find the mode of $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ w.r.t. $\boldsymbol{x}$, and evaluate the expression for $p(\boldsymbol{\theta}|\boldsymbol{y})$ there. This is a *Laplace approximation* $p_{\text{LA}}(\boldsymbol{\theta}|\boldsymbol{y})$, using Gaussian approximations $p_{\text{G}}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$.

The resulting mixture posterior density for $\boldsymbol{x}|\boldsymbol{y}$ is an *integrated Laplace approximation*.

Going one step further, evaluating

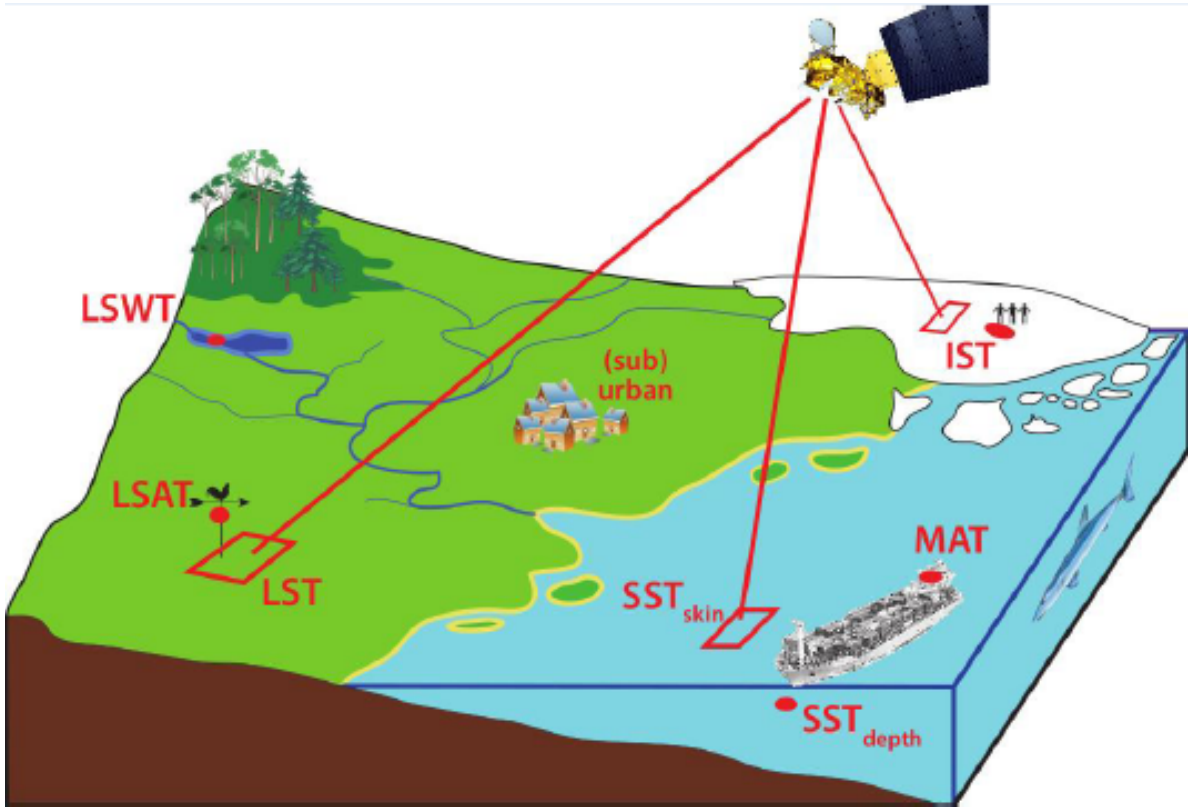
$$p(x_i|\boldsymbol{y}) \approx \sum_k p_{\text{LA}}(x_i|\boldsymbol{y}, \boldsymbol{\theta}_k) p_{\text{LA}}(\boldsymbol{\theta}_k|\boldsymbol{y}) w_k$$

where $p_{\text{LA}}(x_i|\boldsymbol{y}, \boldsymbol{\theta}_k)$ are Laplace approximations of the componentwise marginal densities, we've found the *integrated nested Laplace approximation* (INLA) method.

EUSTACE

EU Surface Temperatures for All Corners of Earth

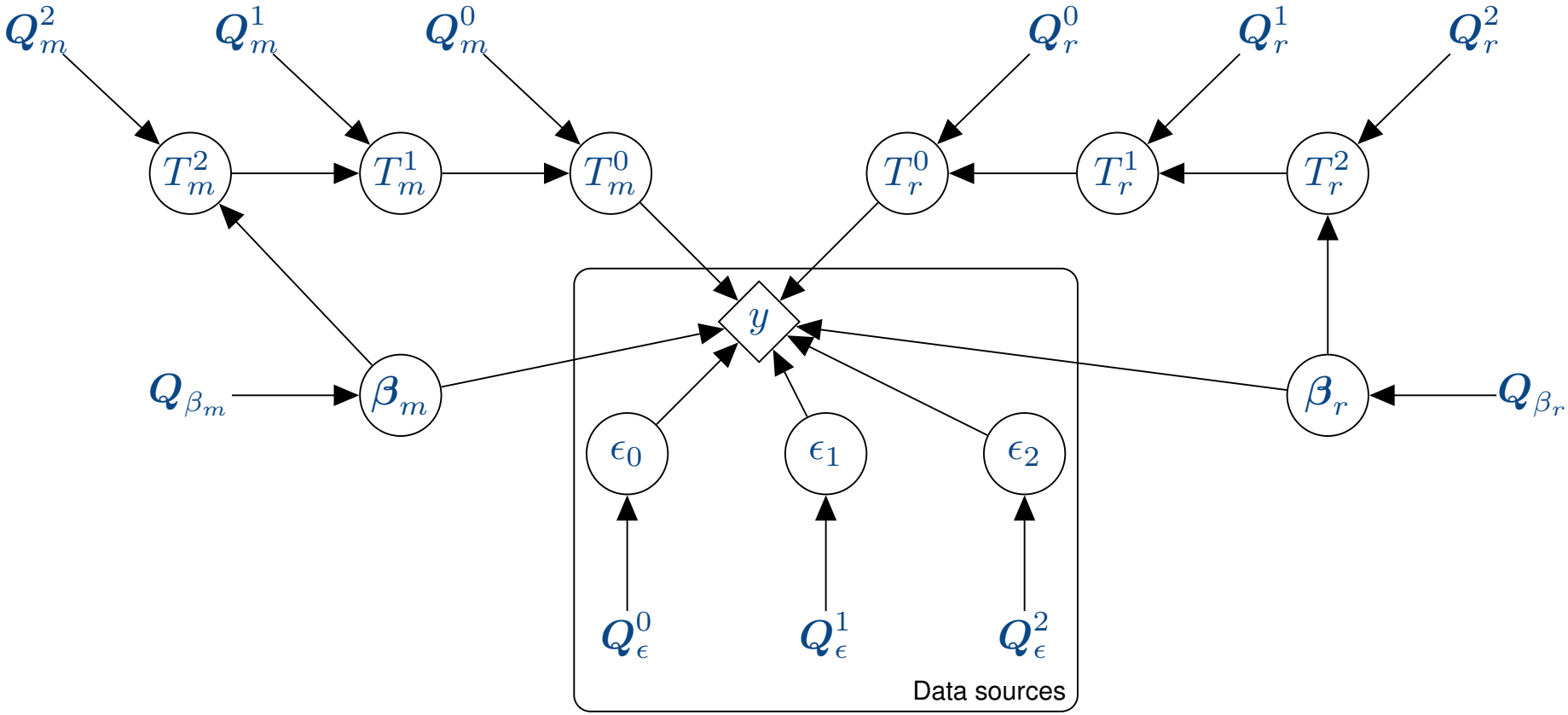
EUSTACE will give publicly available daily estimates of surface air temperature since 1850 across the globe for the first time by combining surface and satellite data using novel statistical techniques.



EUSTACE

Partial hierarchical representation

Observations of *mean*, *max*, *min*. Model *mean* and *range*.

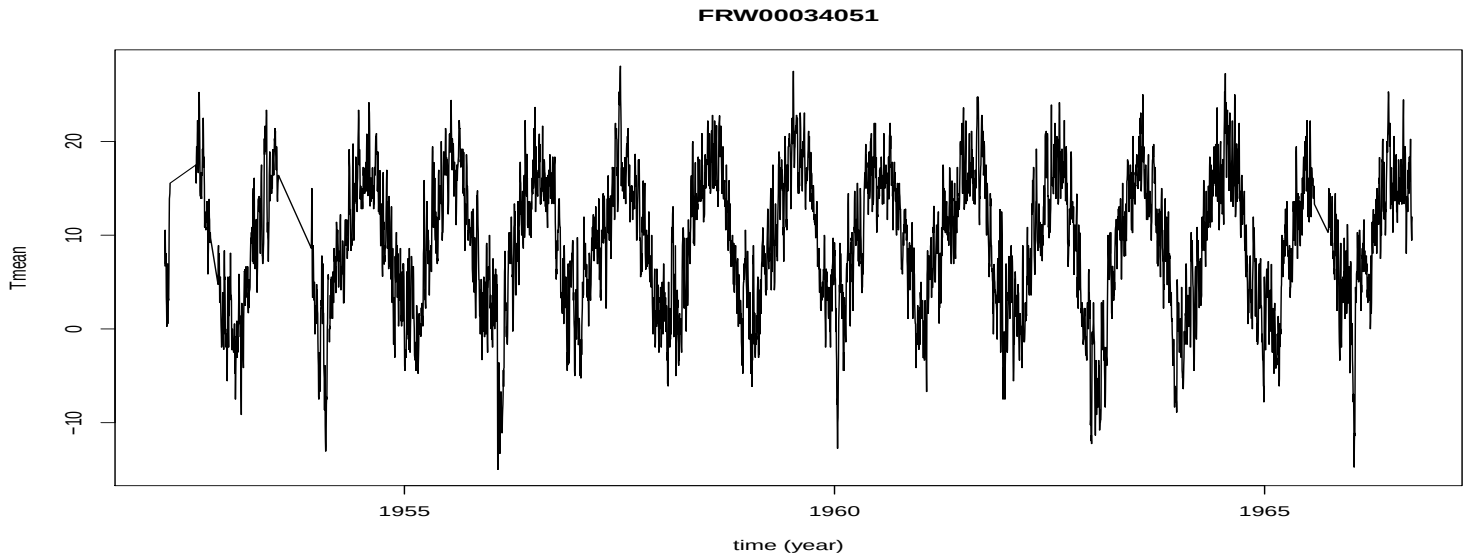


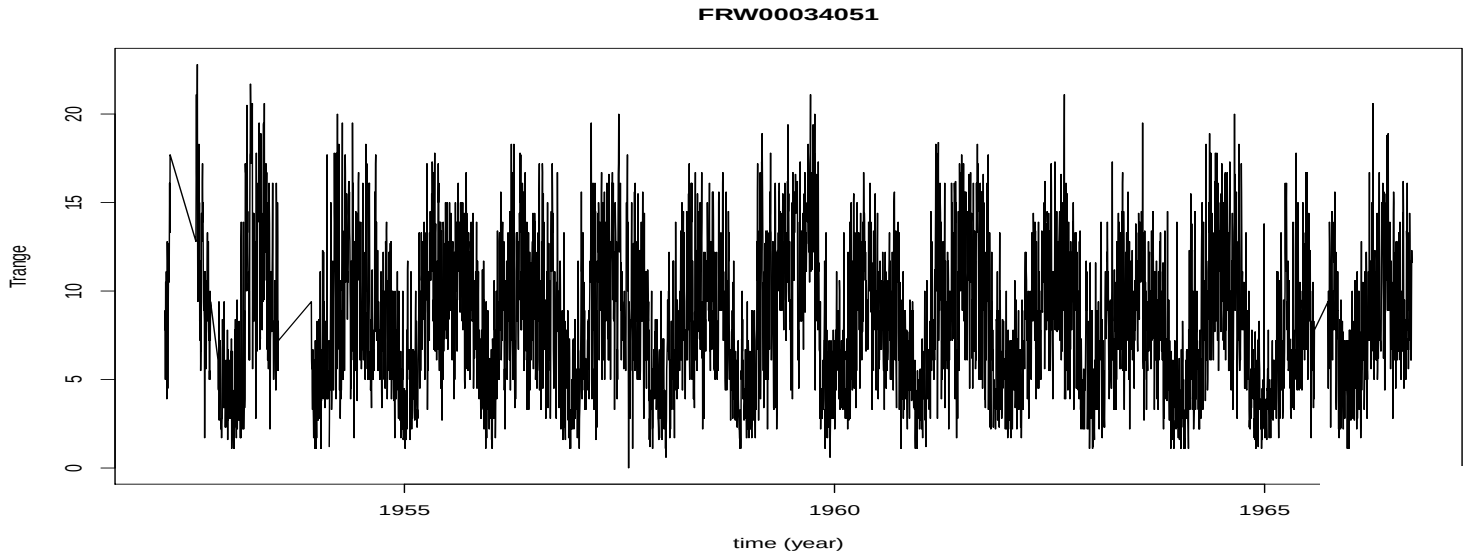
Conditional specifications, e.g.

$$(T_m^0 | T_m^1, Q_m^0) \sim \mathcal{N}\left(T_m^1, {Q_m^0}^{-1}\right)$$

Observed data

Observed daily T_{mean} and T_{range} for station FRW00034051





Power tail quantile (POQ) model

The quantile function (inverse cumulative distribution function) $F_{\boldsymbol{\theta}}^{-1}(p)$, $p \in [0,1]$, is defined as a quantile blend of left and right tailed generalized Pareto distributions,

$$f_\theta^-(p) = \begin{cases} \frac{1-(2p)^{-\theta}}{2\theta}, & \theta \neq 0, \\ \frac{1}{2}\log(2p), & \theta = 0, \end{cases}$$

$$f_\theta^+(p) = -f_\theta^-(1-p) = \begin{cases} \frac{(2(1-p))^{-\theta}-1}{2\theta}, & \theta \neq 0, \\ -\frac{1}{2}\log(2(1-p)), & \theta = 0. \end{cases}$$

$$F_{\boldsymbol{\theta}}^{-1}(p) = \theta_0 + \frac{\tau}{2}\left[(1-\gamma) f_{\theta_3}^-(p) + (1+\gamma) f_{\theta_4}^+(p)\right],$$

The parameters $\boldsymbol{\theta} = (\theta_0, \theta_1 = \log\tau, \theta_2 = \text{logit}[(\gamma+1)/2], \theta_3, \theta_4)$ control the median, spread/scale, skewness, and the left and right tail shape.
This model is also known as the *five parameter lambda model*.

A spatio-temporally dependent Gaussian field $u(\mathbf{s}, t)$ with expectation 0 and variance 1 can be transformed into a POQ field by

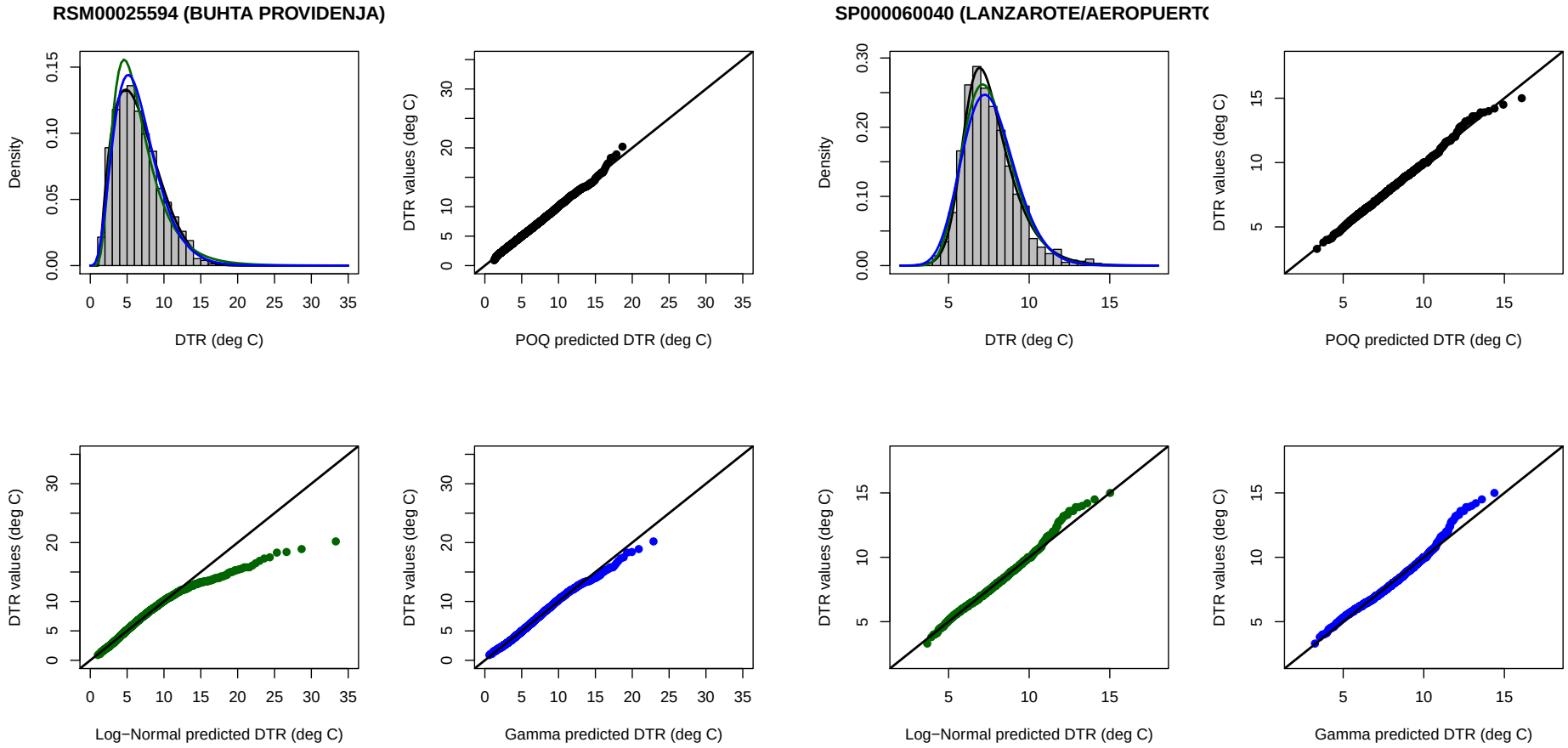
$$\widetilde{u}(\mathbf{s}, t) = F_{\boldsymbol{\theta}(\mathbf{s},t)}^{-1}(\Phi(u(\mathbf{s}, t)),$$

where the parameters can vary with space and time.

EUSTACE

Diurnal range distributions

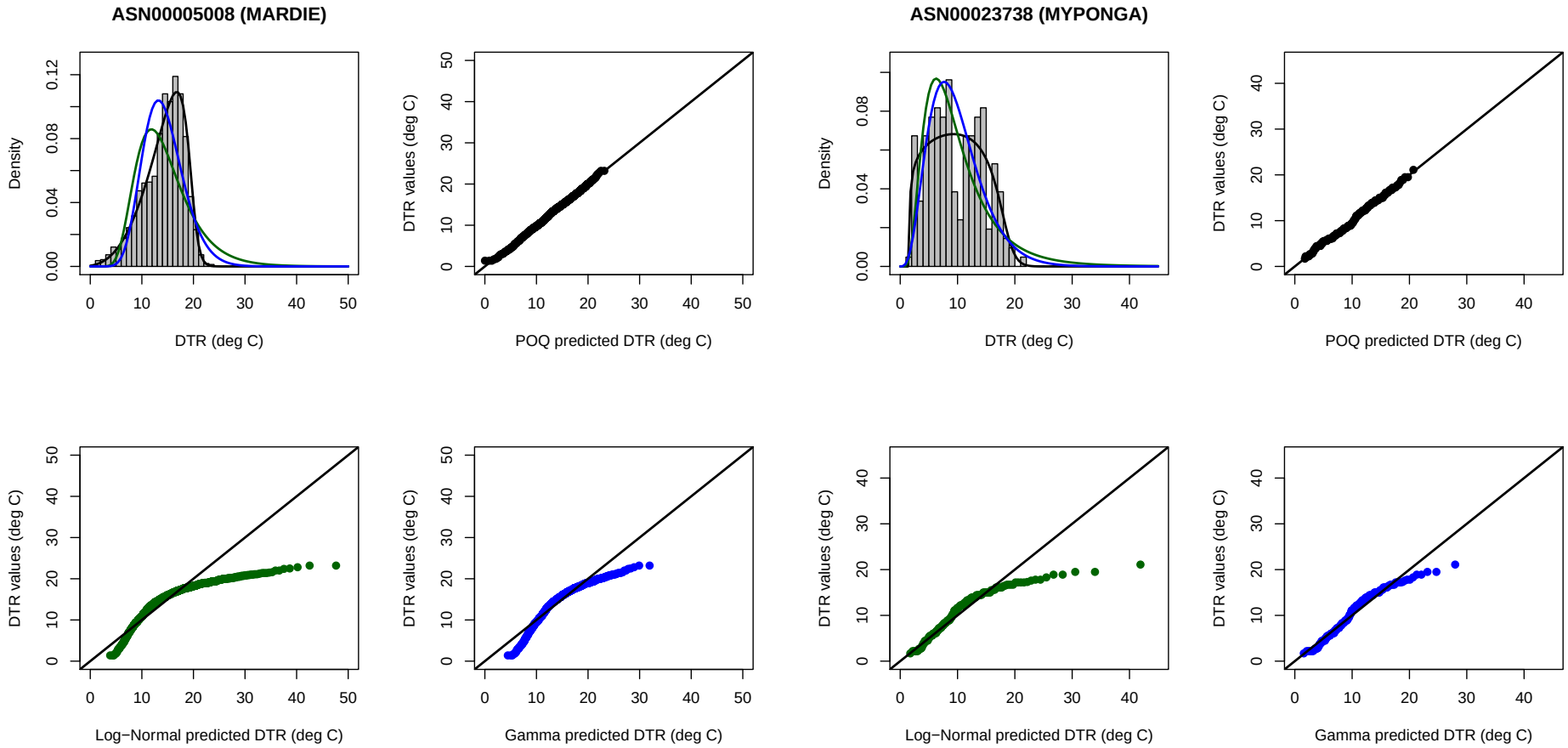
After seasonal compensation:



For these stations, POQ does a slightly better job than a Gamma distribution.

Diurnal range distributions; quantile model

After seasonal compensation:



For these stations only POQ comes close to representing the distributions.

Note: Some of the mixture-like distribution shapes may be an effect of unmodeled station inhomogeneities as well as temporal shift effects.

Transformed expectations

The transformation model essentially leads to a hierarchical model of the form

$$\begin{aligned}
\boldsymbol{\theta} &\sim p(\boldsymbol{\theta}) \quad \text{(variance parameters)} \\
\boldsymbol{x}|\boldsymbol{\theta} &\sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x(\boldsymbol{\theta})^{-1}) \quad \text{(latent Gaussian random fields)} \\
\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x} &\sim \mathcal{N}(h(\boldsymbol{A}\boldsymbol{x}), \boldsymbol{Q}_{y|x}(\boldsymbol{\theta})^{-1}) \quad \text{(observations)}
\end{aligned}$$

for a nonlinear function $h(\boldsymbol{\eta}) = [h_1(\eta_1), \ldots, h_n(\eta_n)]^\top$.
Constructing the Laplace approximations involves finding the Gaussian approximation $\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}^*_{x|y}, [\boldsymbol{Q}^*_{x|y}]^{-1})$ by quasi-Newton iteration:

$$\begin{aligned}
f(\boldsymbol{x}) &= \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_x)^\top \boldsymbol{Q}_x(\boldsymbol{x} - \boldsymbol{\mu}_x) + \frac{1}{2}(\boldsymbol{y} - h(\boldsymbol{A}\boldsymbol{x}))^\top \boldsymbol{Q}_{y|x}(\boldsymbol{y} - h(\boldsymbol{A}\boldsymbol{x})), \\
\boldsymbol{Q}^{\text{new}}_{x|y} &= \boldsymbol{Q}_x + \boldsymbol{A}^\top \boldsymbol{J}_*^\top \boldsymbol{Q}_{y|x} \boldsymbol{J}_* \boldsymbol{A}, \quad \text{(one h.o.t. eliminated)} \\
\boldsymbol{\mu}^{\text{new}}_{x|y} &= \boldsymbol{\mu}^*_{x|y} - a[\boldsymbol{Q}^*_{x|y}]^{-1} \left[\boldsymbol{Q}_x(\boldsymbol{\mu}^*_{x|y} - \boldsymbol{\mu}_x) - \boldsymbol{A}^\top \boldsymbol{J}_*^\top \boldsymbol{Q}_{y|x}(\boldsymbol{y} - h(\boldsymbol{A}\boldsymbol{\mu}^*_{x|y})) \right]
\end{aligned}$$

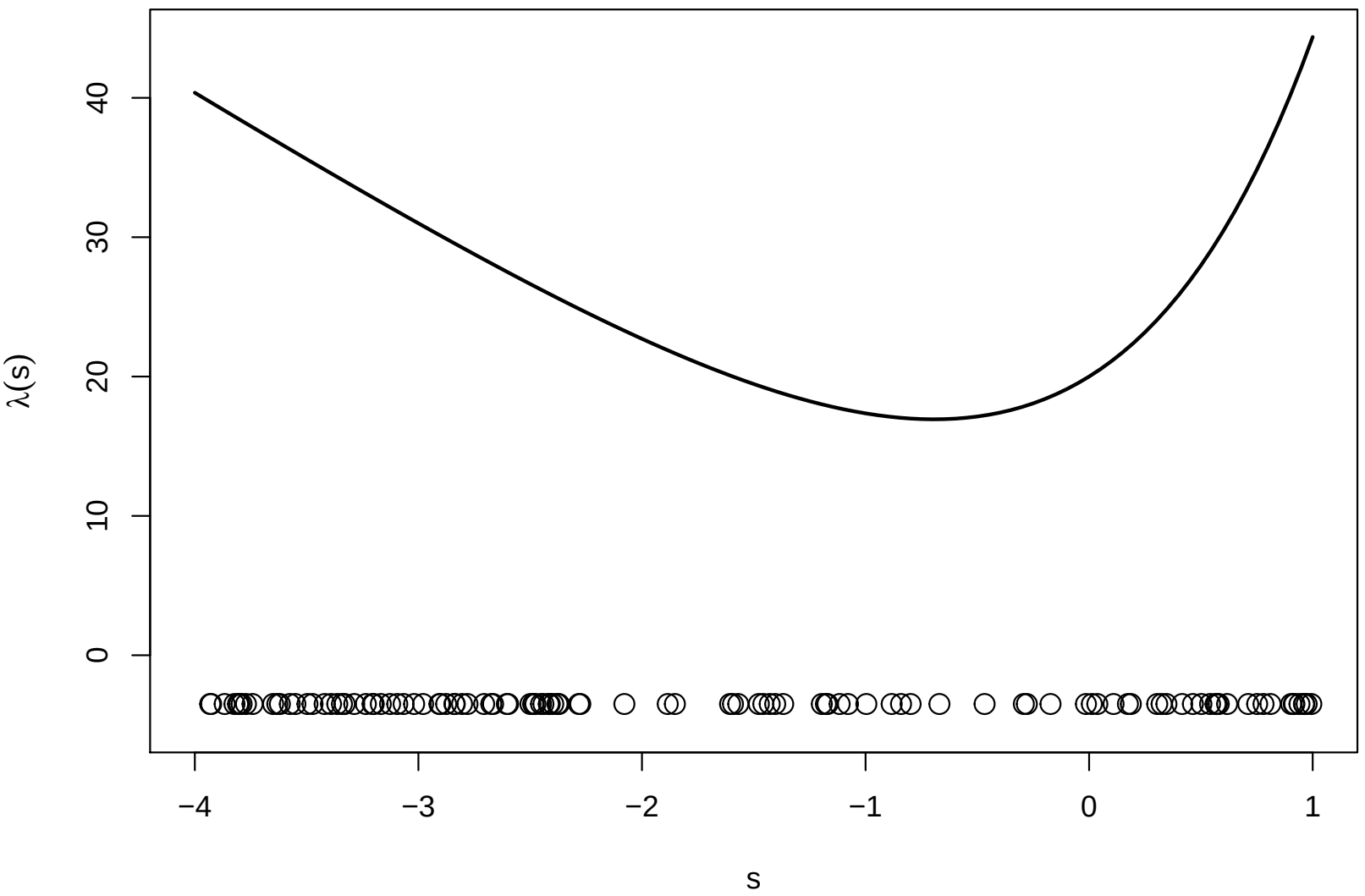
EUSTACE

Finding animals: Poisson point processes

Animals detected at sea can be modelled by inhomogeneous point processes. Given an *intensity function* $\lambda(\mathbf{s})$, the observation log-likelihood is

$$\log p(\boldsymbol{y}|\lambda) = -\int \lambda(\mathbf{s})\,\mathrm{d}\mathbf{s} + \sum_{i=1}^{n} \log(\lambda(y_i))$$

Usually, λ is modelled as a log-linear function of Gaussian latent variables and random fields.



EUSTACE

Finding animals: Poisson point processes

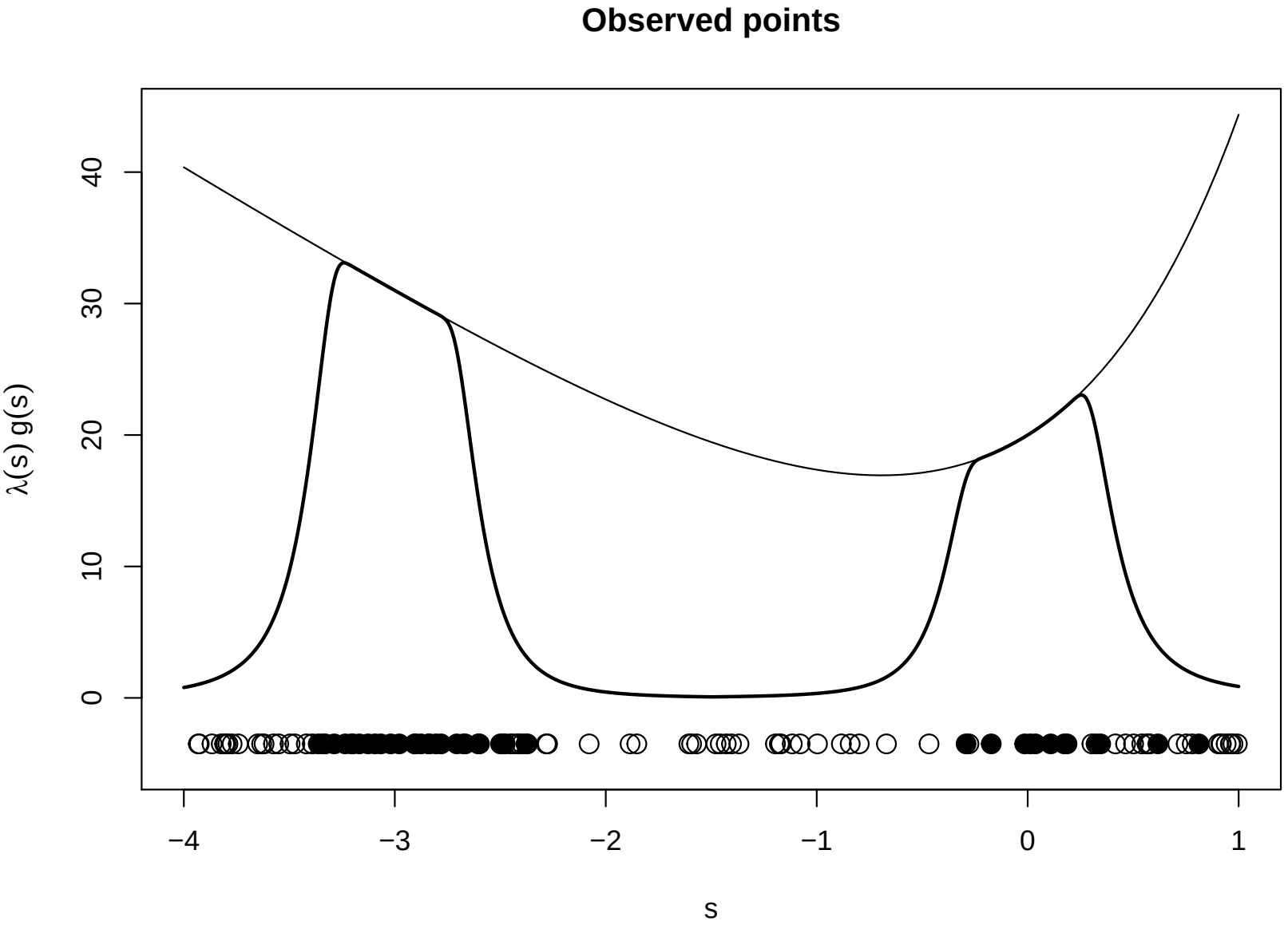
Animals detected at sea can be modelled by inhomogeneous point processes. Given an *intensity function* $\lambda(\mathbf{s})$, the observation log-likelihood is

$$\log p(\boldsymbol{y}|\lambda) = -\int \lambda(\mathbf{s})\,\mathrm{d}\mathbf{s} + \sum_{i=1}^{n} \log(\lambda(y_i))$$

Usually, λ is modelled as a log-linear function of Gaussian latent variables and random fields.

Not all animals are detected, since they are far away from the observer. This changes the intensity function by the *probability of detection*, $g(\mathbf{s})$, so that the density of the observation model is $\lambda(\mathbf{s})p(\mathbf{s})$

EUSTACE



Finding animals: Poisson point processes

Animals detected at sea can be modelled by inhomogeneous point processes. Given an *intensity function* $\lambda(\mathbf{s})$, the observation log-likelihood is

$$\log p(\boldsymbol{y}|\lambda) = -\int \lambda(\mathbf{s})\,\mathrm{d}\mathbf{s} + \sum_{i=1}^{n} \log(\lambda(y_i))$$

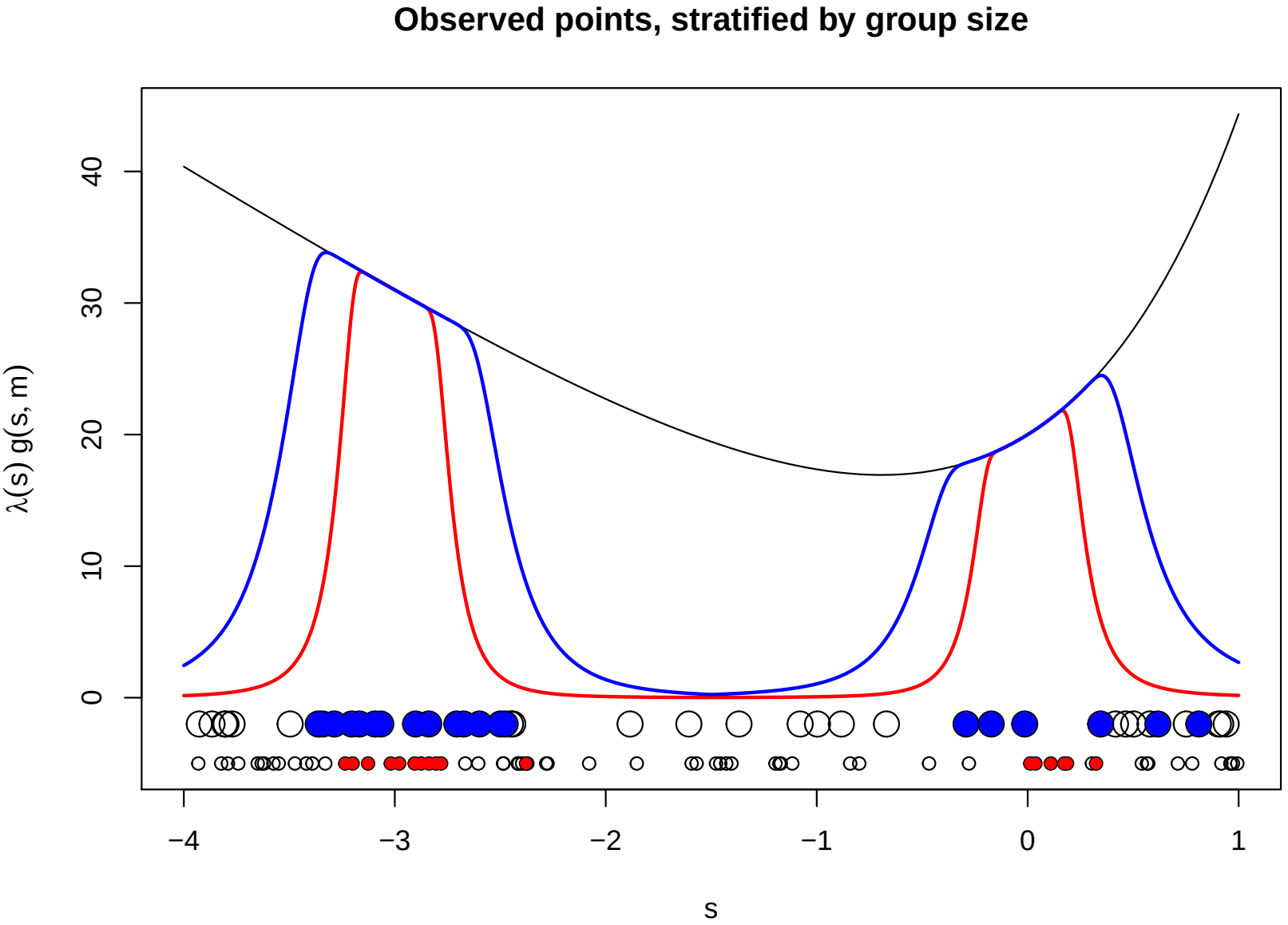
Usually, λ is modelled as a log-linear function of Gaussian latent variables and random fields.

Not all animals are detected, since they are far away from the observer. This changes the intensity function by the *probability of detection*, $g(\mathbf{s})$, so that the density of the observation model is $\lambda(\mathbf{s})g(\mathbf{s})$.

Larger groups are easier to detect than others. We therefore need to model $g(\mathbf{s}, m)$ and $p(m|\mathbf{s})$, the distribution of group sizes as a function of space!

The combined locations and *marks*, (y_i, m_i) follow a joint point process wth intensity $\lambda(\mathbf{s})p(m|\mathbf{s})g(\mathbf{s}, m)$.

EUSTACE


Observed points, stratified by group size
λ(s) g(s, m)
0
10
20
30
40
−4
−3
−2
−1
0
1
s

The likelihood

$$p(m|\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}\exp(x_2)} \exp\left[-\frac{(m-x_1)^2}{2\exp(2x_2)}\right]$$

has Poisson-process log-likelihood version

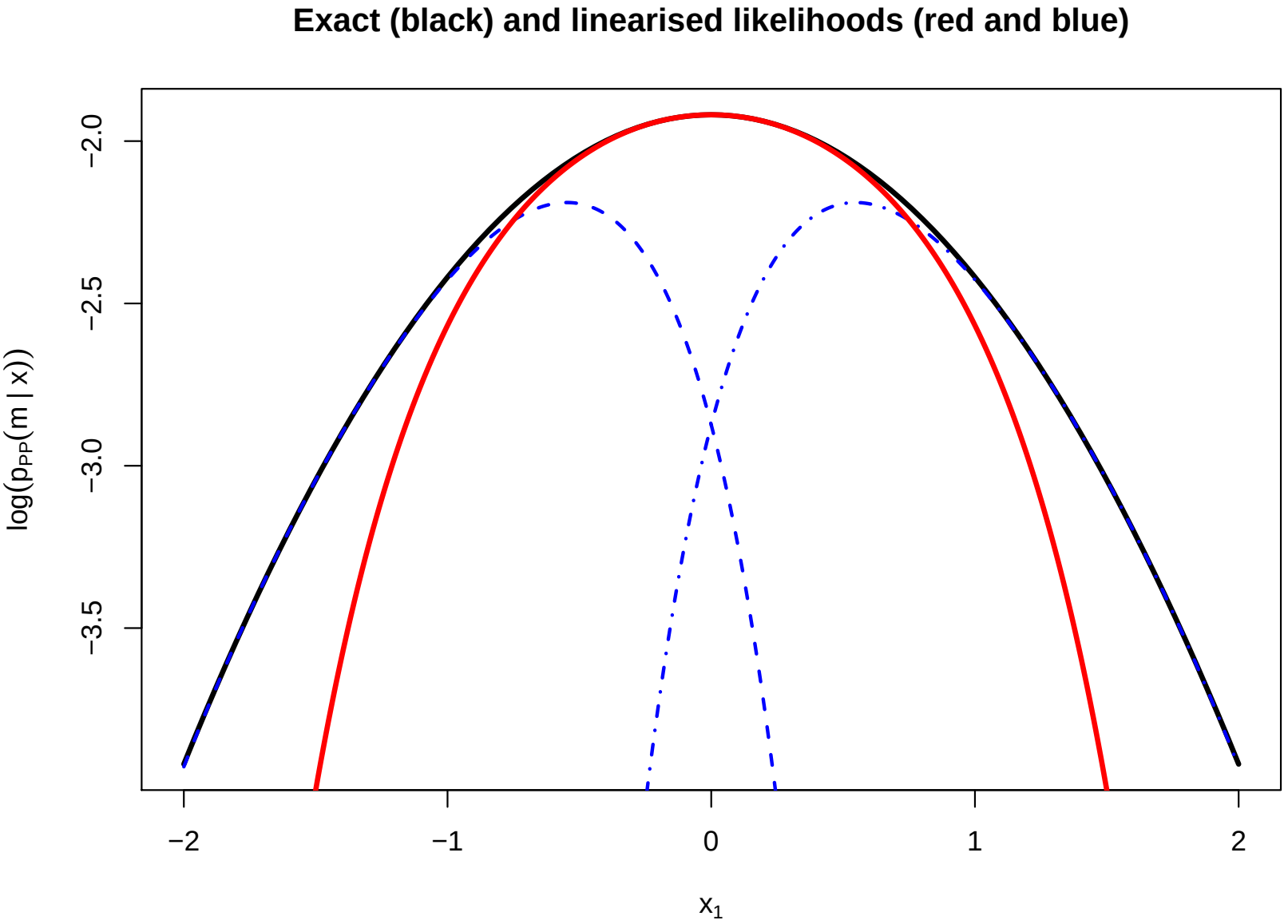
$$\log p_{\mathsf{PP}}(m|\boldsymbol{x}) = -\int p(m'|\boldsymbol{x})\,\mathrm{d}m' + \log p(m|\boldsymbol{x}) = -1 + \log p(m|\boldsymbol{x}).$$

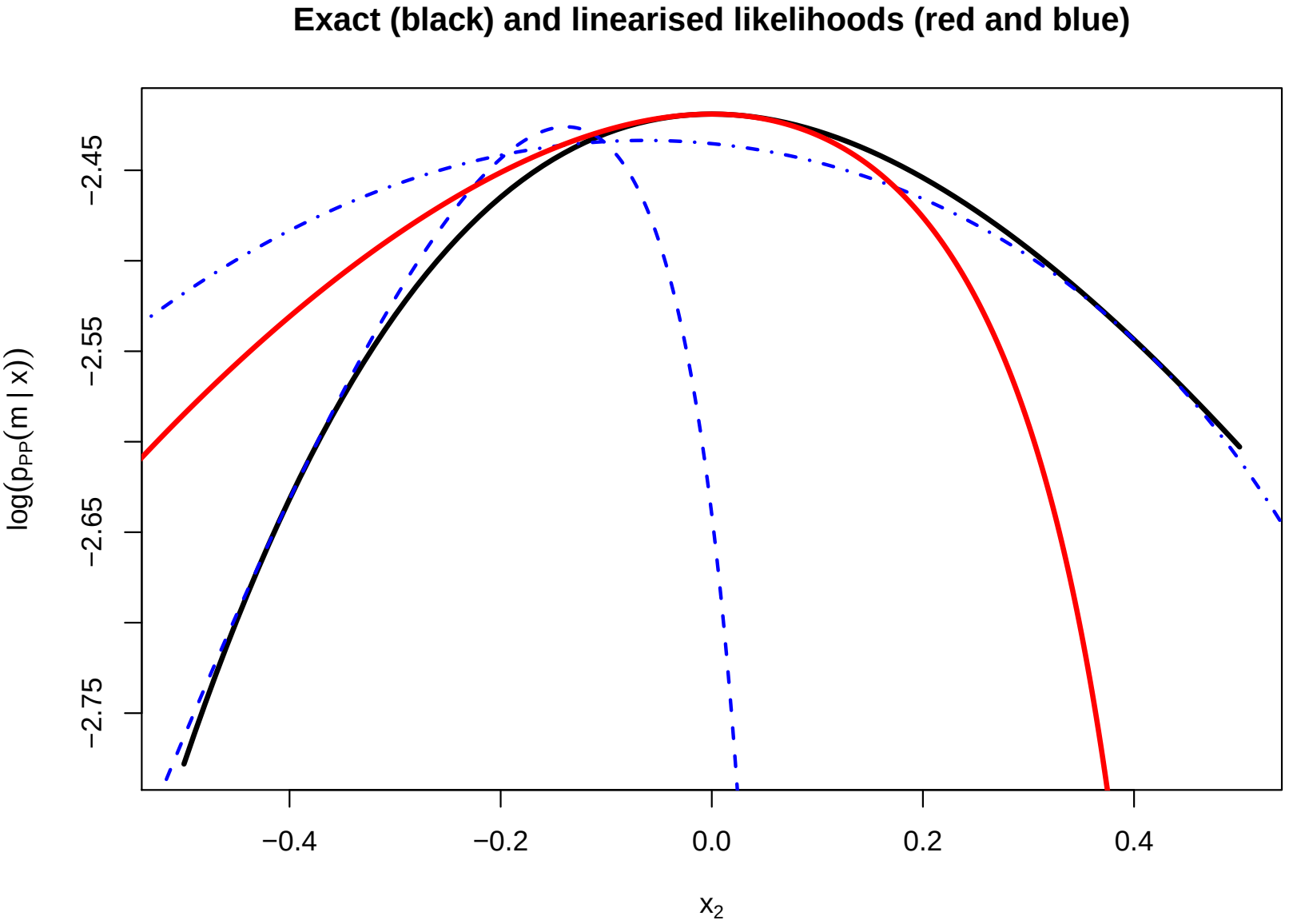
Linearising $\log p(m|\boldsymbol{x})$ at some $\boldsymbol{x}^*$,

$$\log p^*(m|\boldsymbol{x}) = \log p(m|\boldsymbol{x}^*) + \sum_{k=1}^{2} \left.\frac{\partial \log p(m|\boldsymbol{x})}{\partial x_k}\right|_{\boldsymbol{x}=\boldsymbol{x}^*} (x_k - x_k^*),$$

which leads to the Poisson-process likelihood

$$\log p^*_{\mathsf{PP}}(m|\boldsymbol{x}) = -\int p^*(m'|\boldsymbol{x})\,\mathrm{d}m' + \log p^*(m|\boldsymbol{x}).$$


Exact (black) and linearised likelihoods (red and blue)
log(p_PP(m | x))
−2.0
−2.5
−3.0
−3.5
−2
−1
0
1
2
x_1



Summary

- Gaussian SPDEs provide a multitude of practically useful spatial statistics models with Markov properties
- INLA methods provide fast approximate Bayesian inference for a large class of latent Gaussian models
- Complex observation mechanisms can be incorporated

EUSTACE

Addendum (LaTeXing notes that, for obvious reasons, did not make it into the talk)

In the Poisson process version of the linearised likelihood for multiple (n) observations,

$$\log p^*_{\mathsf{PP}}(\boldsymbol{m}|\boldsymbol{x}) = -n \int p^*(m'|\boldsymbol{x})\,\mathrm{d}m' + \sum_k \log p^*(m_k|\boldsymbol{x}).$$

the second term is linear in $\boldsymbol{x}$ and the integral is not 1 except for at least $\boldsymbol{x} = \boldsymbol{x}^*$.

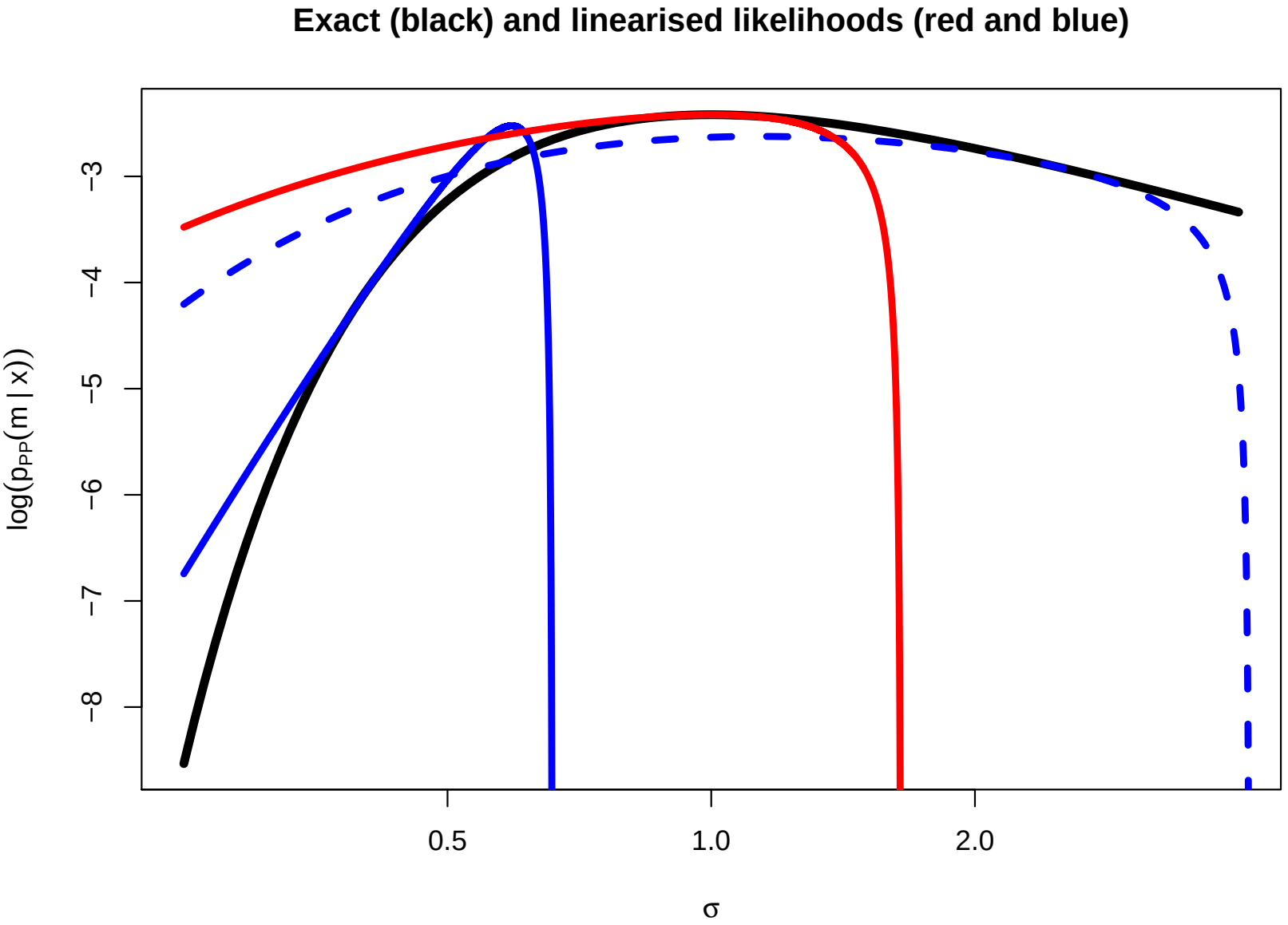
Spelling out the details:

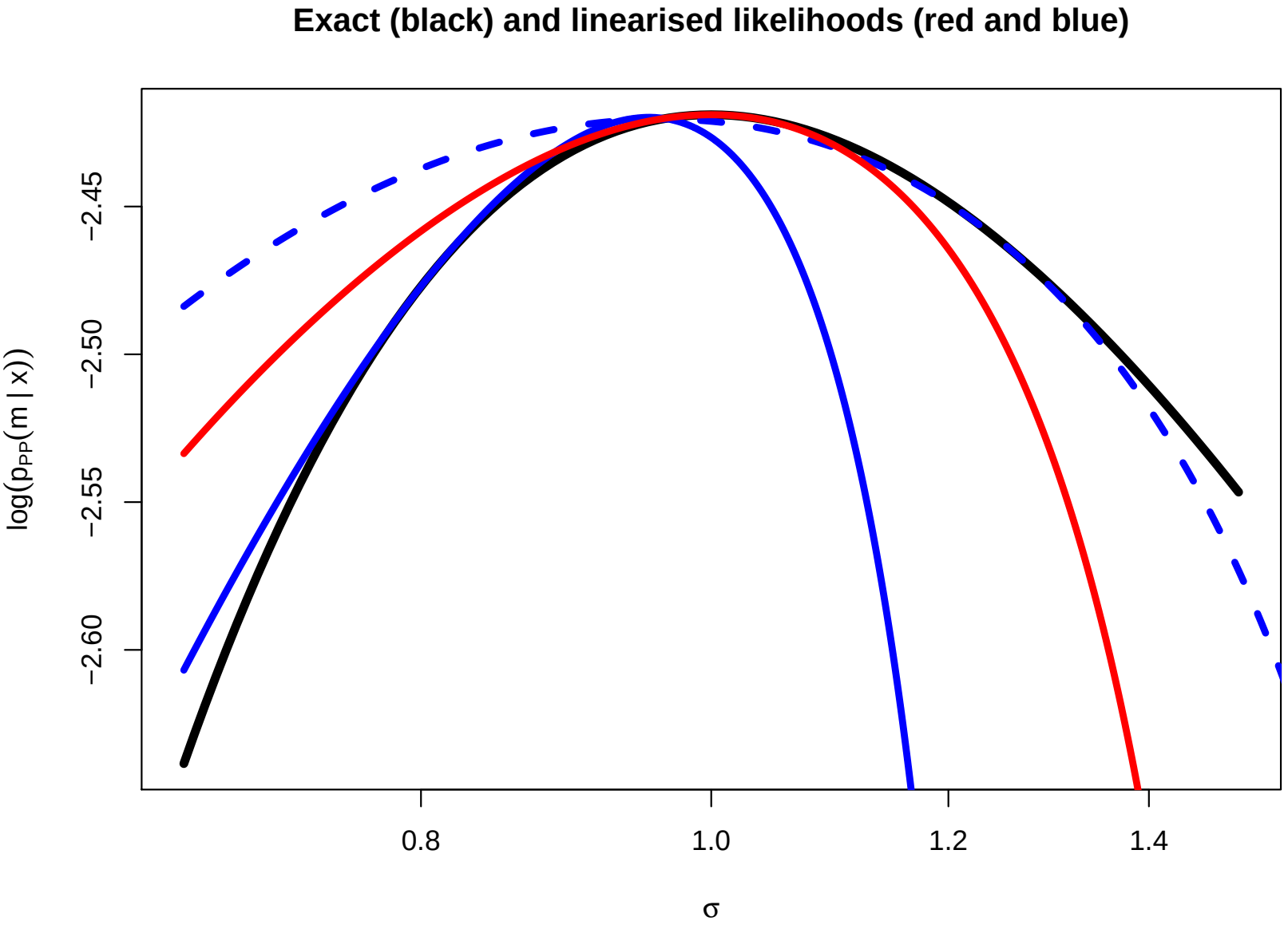
$$\begin{aligned}
\log p^*(m|\boldsymbol{x}) &= -\log\sqrt{2\pi} - x_2^* - \frac{(m-x_1^*)^2}{2\mathrm{e}^{2x_2^*}} \\
&\quad + \frac{m-x_1^*}{\mathrm{e}^{2x_2^*}}(x_1-x_1^*) + \left(-1+\frac{(m-x_1^*)^2}{\mathrm{e}^{2x_2}}\right)(x_2-x_2^*) \\
&= -\log\sqrt{2\pi} - x_2 - \frac{1}{2\mathrm{e}^{2x_2^*}}\,[\\
&\quad (m-x_1^*)^2(1-2(x_2-x_2^*)) - 2(m-x_1^*)(x_1-x_1^*)] \\
&= -\log\sqrt{2\pi} - x_2 \\
&\quad - \frac{1-2(x_2-x_2^*)}{2\mathrm{e}^{2x_2^*}}\left[\, m - x_1^* - \frac{x_1-x_1^*}{1-2(x_2-x_2^*)}\right]^2 \\
&\quad + \frac{1-2(x_2-x_2^*)}{2\mathrm{e}^{2x_2^*}}\left[\frac{x_1-x_1^*}{1-2(x_2-x_2^*)}\right]^2
\end{aligned}$$

If $1-2(x_2-x_2^*) > 0$,

$$\int_{\mathbb{R}} p^*(m|\boldsymbol{x})\,\mathrm{d}m = \frac{\mathrm{e}^{x_2^*}}{\sqrt{1-2(x_2-x_2^*)}}\exp\left\{-x_2 + \frac{(x_1-x_1^*)^2}{2\mathrm{e}^{2x_2^*}[1-2(x_2-x_2^*)]}\right\}$$

EUSTACE




Exact (black) and linearised likelihoods (red and blue)
log(p_PP(m | x))
−2.45
−2.50
−2.55
−2.60
0.8
1.0
1.2
1.4
σ

General behaviour (written for a 1-dimensional x for simplicity). Derivative of $f(m,x) = \log p(m|x)$ w.r.t. denoted $f_x(m,x)$.

$$
\begin{aligned}
f(m,x) &= \log p(m|x) \\
f_P(m,x) &= -\int \mathrm{e}^f(m,x)\,\mathrm{d}m + f(m,x) = -1 + f(m,x) \\
f^*(m,x) &= f(m,x^*) + f_x(m,x^*)(x-x^*) \\
f_P^*(m,x) &= -\int \mathrm{e}^{f(m,x^*)+f_x(m,x^*)(x-x^*)}\,\mathrm{d}m + f(m,x^*) + f_x(m,x^*)(x-x^*) \\
\frac{\partial f_P^*(m,x)}{\partial x} &= -\int f_x(m,x^*)\mathrm{e}^{f(m,x^*)+f_x(m,x^*)(x-x^*)}\,\mathrm{d}m + f_x(m,x^*) \\
&= [\text{at } x = x^*] = -\int f_x(m,x^*)\mathrm{e}^{f(m,x^*)}\,\mathrm{d}m + f_x(m,x^*) \\
&= -\int \mathrm{e}^{f(m,x^*)}\,\frac{\partial \mathrm{e}^{f(m,x)}/\partial x}{\mathrm{e}^{f(m,x)}}\bigg|_{x=x^*}\mathrm{d}m + f_x(m,x^*) \\
&= -\frac{\partial}{\partial x}\int \mathrm{e}^{f(m,x)}\,\mathrm{d}m\bigg|_{x=x^*} + f_x(m,x^*) \\
&= f_x(m,x^*)
\end{aligned}
$$

EUSTACE

(cont.)

$$
\begin{aligned}
\frac{\partial^2 f_P^*(m,x)}{\partial x^2} &= -\int f_x(m,x^*)^2 \mathrm{e}^{f(m,x^*)+f_x(m,x^*)(x-x^*)}\,\mathrm{d}m \\
&= [\text{at } x = x^*] = -\int f_x(m,x^*)^2 \mathrm{e}^{f(m,x^*)}\,\mathrm{d}m \\
&= -\mathsf{E}\left[f_x(m,x^*)^2 | m \sim p(m|x^*)\right] \\
&= [\text{(from likelihood theory)}] \\
&= \mathsf{E}\left[-\left[\frac{p_x(m|x^*)}{p(m|x^*)}\right]^2 + \frac{p_{xx}(m|x^*)}{p(m|x^*)}\middle| m \sim p(m|x^*)\right] \\
&= \mathsf{E}\left[\frac{\partial}{\partial x}\frac{p_x(m|x)}{p(m|x)}\middle|_{x=x^*}\middle| m \sim p(m|x^*)\right] \\
&= \mathsf{E}\left[\frac{\partial^2}{\partial x^2}\log p(m|x)\middle|_{x=x^*}\middle| m \sim p(m|x^*)\right] \\
&= \mathsf{E}\left[f_{xx}(m,x^*) | m \sim p(m|x^*)\right]
\end{aligned}
$$

i.e. the second order derivative at the linearisation point is the expected Fisher information.

EUSTACE