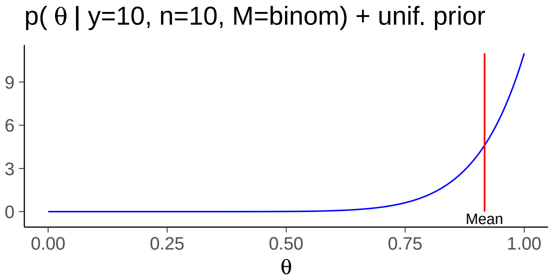
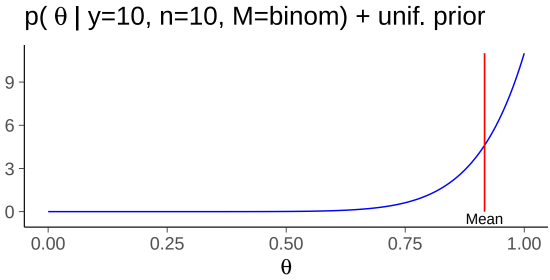
Benefits of integration and prior

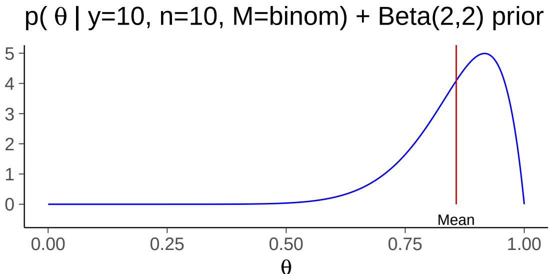
Example: $n = 10, y = 10$ - uniform vs Beta(2,2) prior



Benefits of integration and prior

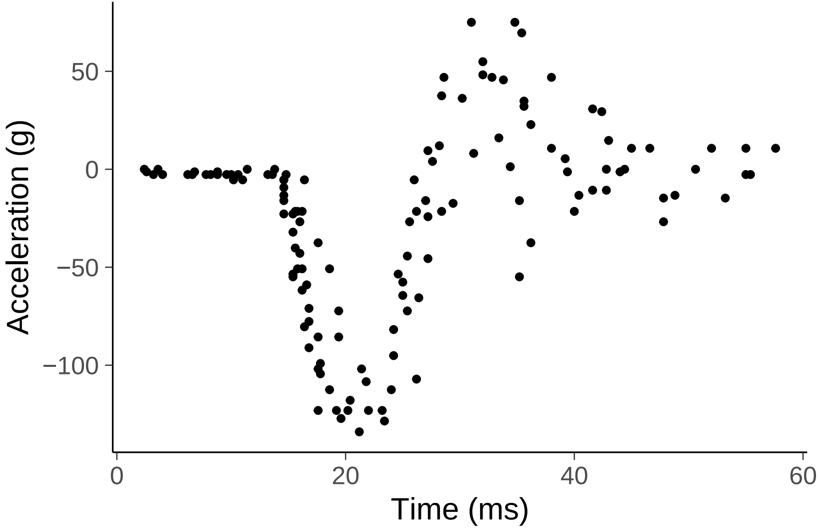
Example: $n = 10, y = 10$ - uniform vs Beta(2,2) prior





Benefits of integration: Gaussian process example

Simulated data of *g* forces in a motorcycle accident



Benefits of integration: Gaussian process example

- Gaussian process is a hierarchical normal model with multivariate normal population prior, where the off-diagonal covariances are determined by similarity measure (in the this example distance)

$$
\begin{aligned}
y &\sim \text{normal}(f(x), \sigma) \\
f &\sim GP(0, K_1) \\
\sigma &\sim \text{normal}^{+}(0, 1),
\end{aligned}
$$

Benefits of integration: Gaussian process example

- Gaussian process is a hierarchical normal model with multivariate normal population prior, where the off-diagonal covariances are determined by similarity measure (in the this example distance)

$$
\begin{aligned}
y &\sim \mathrm{normal}(f(x), \sigma) \\
f &\sim GP(0, K_1) \\
\sigma &\sim \mathrm{normal}^+(0, 1),
\end{aligned}
$$

- K_1 is a covariance matrix, defined by a covariance function, that has parameters lengthscale l and magnitude σ_f

Benefits of integration: Gaussian process example

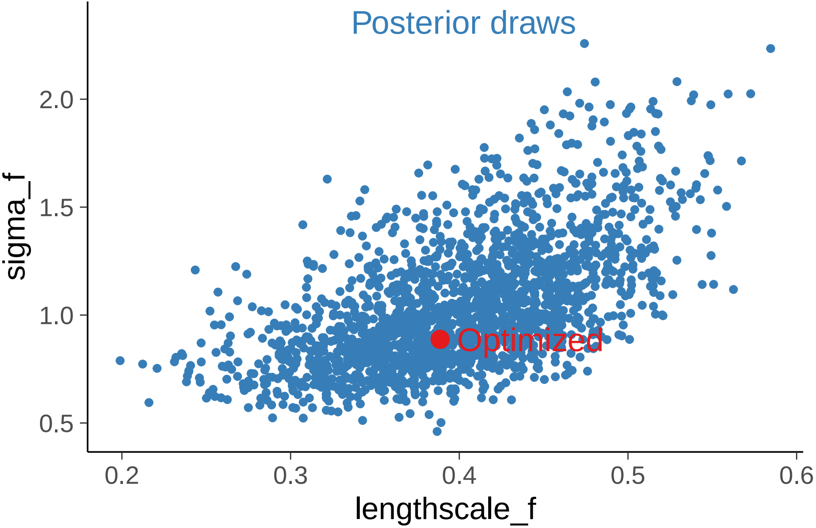
- Gaussian process is a hierarchical normal model with multivariate normal population prior, where the off-diagonal covariances are determined by similarity measure (in the this example distance)

$$
\begin{aligned}
y &\sim \mathrm{normal}(f(x), \sigma) \\
f &\sim GP(0, K_1) \\
\sigma &\sim \mathrm{normal}^{+}(0, 1),
\end{aligned}
$$

- K_1 is a covariance matrix, defined by a covariance function, that has parameters lengthscale l and magnitude σ_f
- Latent values f can be integrated out analytically, and the remaining marginal posterior is 3 dimensional

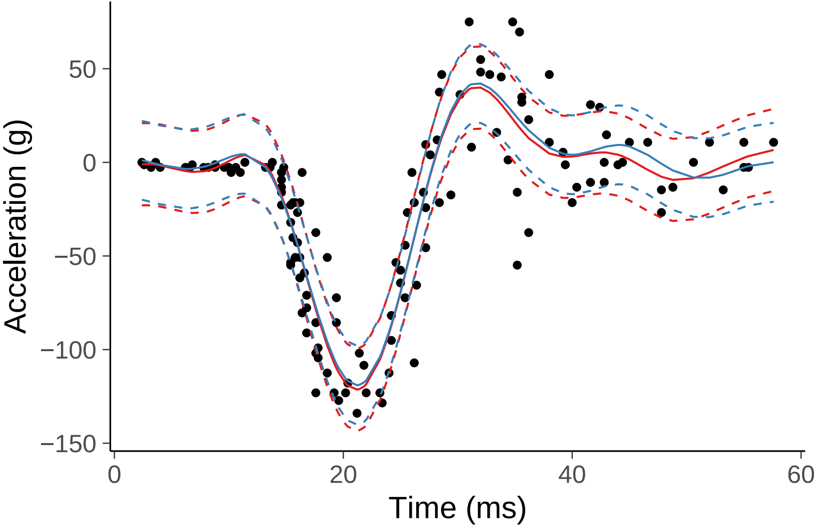
Benefits of integration: Gaussian process example

Plenty of data: the mode of th posterior can be representative



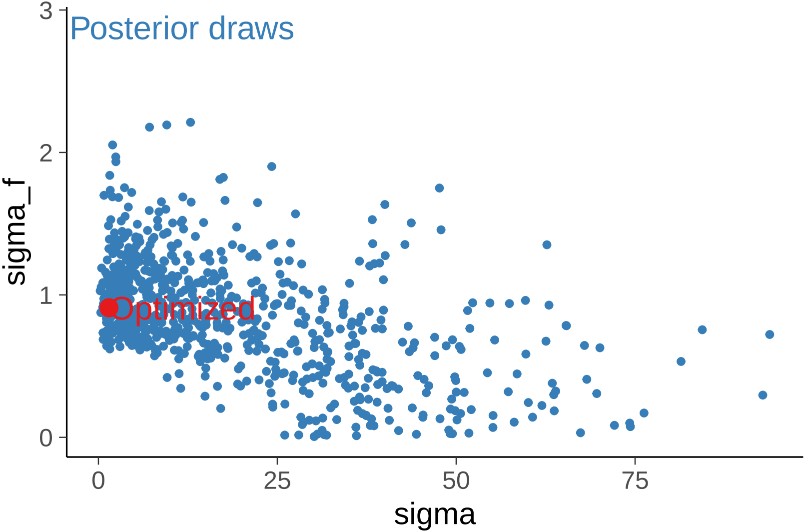
Benefits of integration: Gaussian process example

Plenty of data: the mode of th posterior can be representative



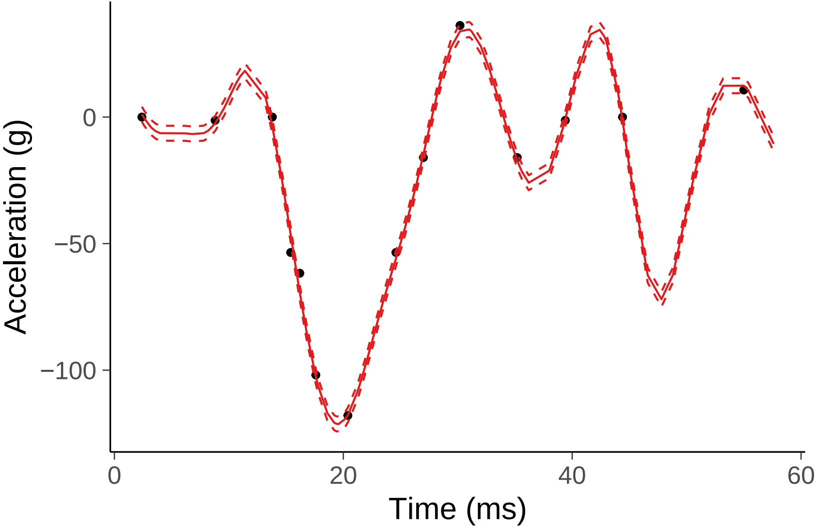
Benefits of integration: Gaussian process example

Small data: the mode of the posterior is not representative



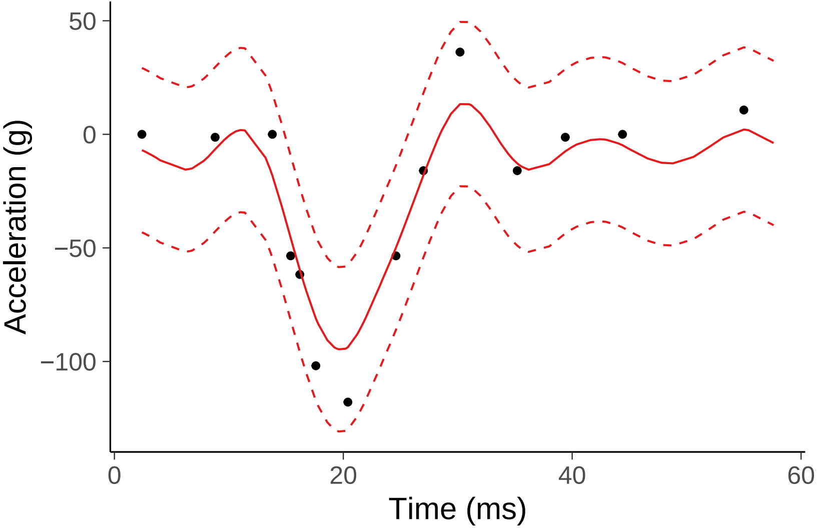
Benefits of integration: Gaussian process example

Small data: the mode of the posterior is not representative



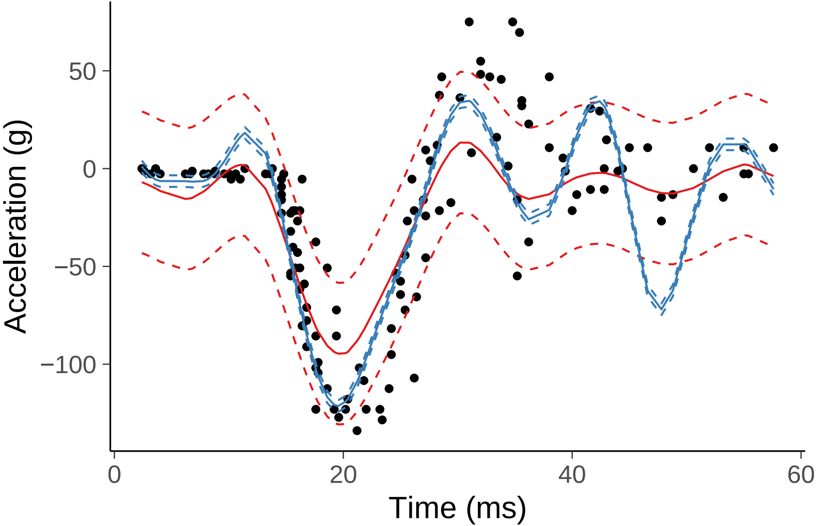
Benefits of integration: Gaussian process example

Small data: the mode of the posterior is not representative



Benefits of integration: Gaussian process example

Small data: the mode of the posterior is not representative



Benefits of integration: Gaussian process example

- More complex model, with time dependent residual variance $\exp(g(x))$

$$
\begin{aligned}
y &\sim \mathrm{normal}(f(x), \exp(g(x))) \\
f &\sim GP(0, K_1) \\
g &\sim GP(0, K_2).
\end{aligned}
$$

Benefits of integration: Gaussian process example

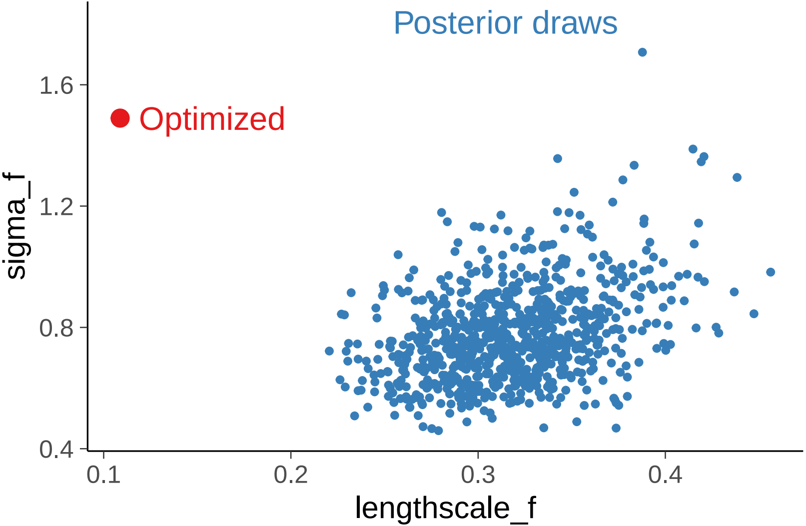
- More complex model, with time dependent residual variance $\exp(g(x))$

$$
\begin{aligned}
y &\sim \mathrm{normal}(f(x), \exp(g(x)) \\
f &\sim GP(0, K_1) \\
g &\sim GP(0, K_2).
\end{aligned}
$$

- Latent values can not be integrated out analytically, and the posterior is 270 dimensional

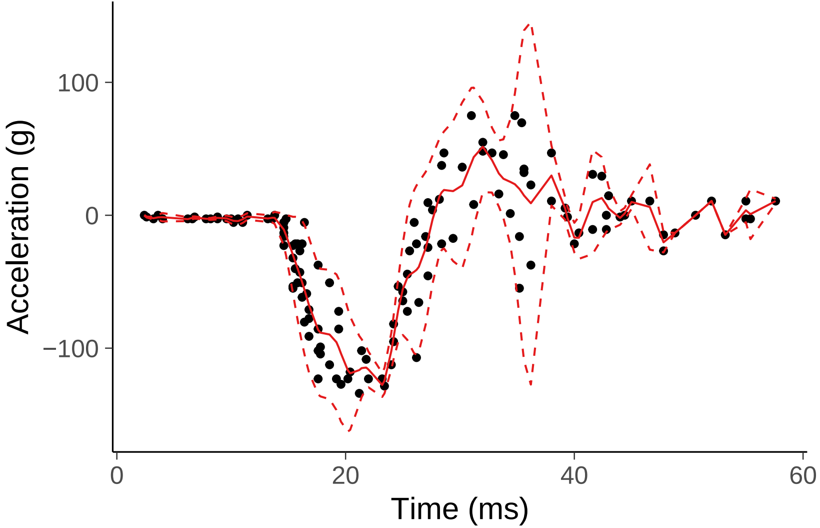
Benefits of integration: Gaussian process example

The mode of the posterior is not representative



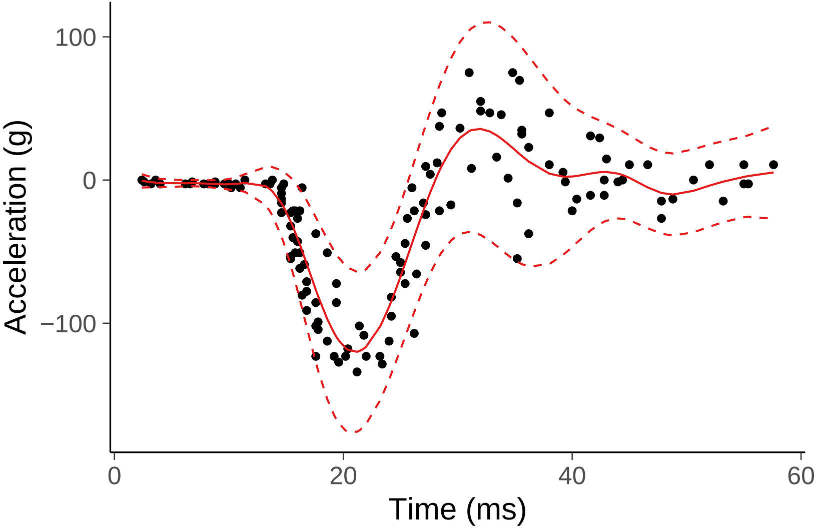
Benefits of integration: Gaussian process example

Optimization overfits



Benefits of integration: Gaussian process example

Integration works well



Benefits of better priors: logistic regression

$$y \sim \text{Bernoulli}\left(\text{logit}^{-1}(\alpha + \beta X)\right)$$

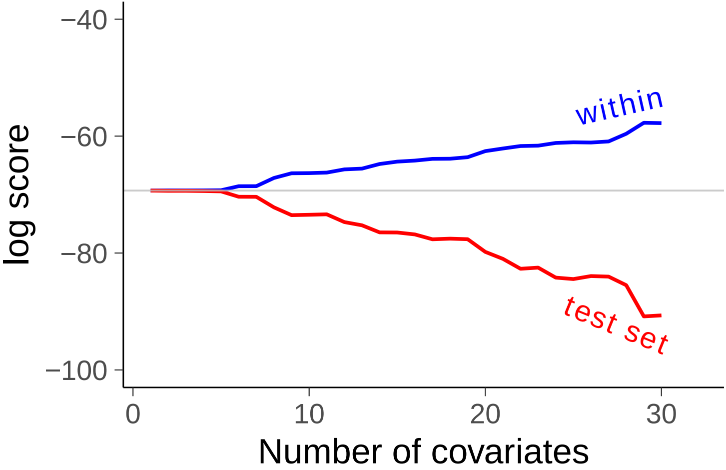
where α is a scalar intercept,
and β is a vector of coefficients

Model selection is needed to avoid overfitting?

logistic regression: 30 **completely irrelevant** variables, 100 observations

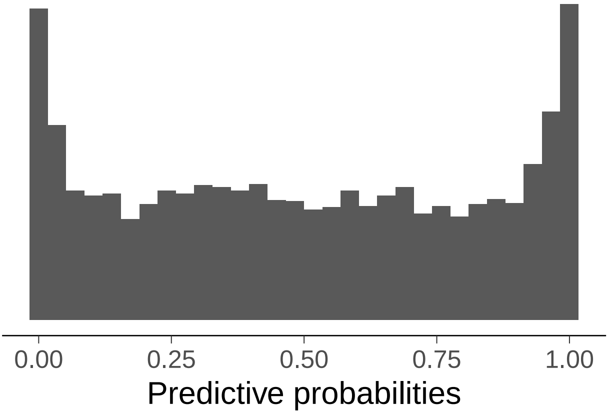
Model selection is needed to avoid overfitting?

logistic regression: 30 **completely irrelevant** variables, 100 observations



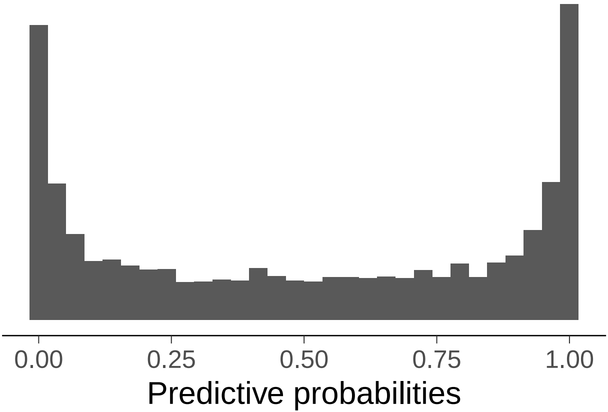
Prior on parameters vs predictions

N(0,3) prior on each coefficient
1 variable



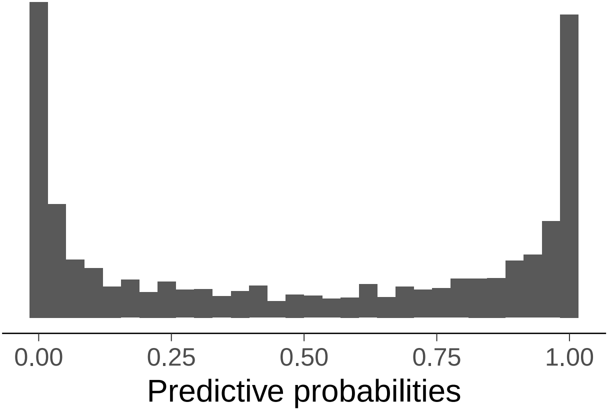
Prior on parameters vs predictions

N(0,3) prior on each coefficient
2 variables



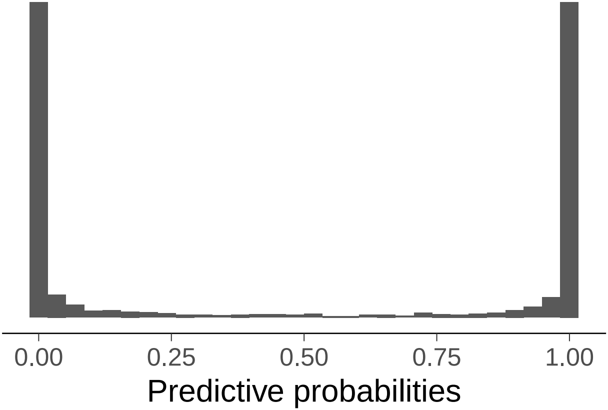
Prior on parameters vs predictions

N(0,3) prior on each coefficient
3 variables



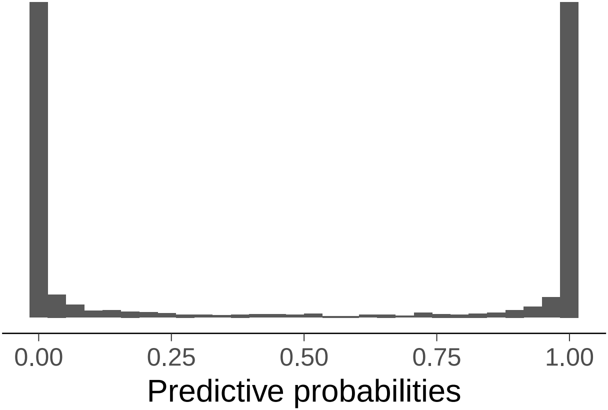
Prior on parameters vs predictions

N(0,3) prior on each coefficient
30 variables



Prior on parameters vs predictions

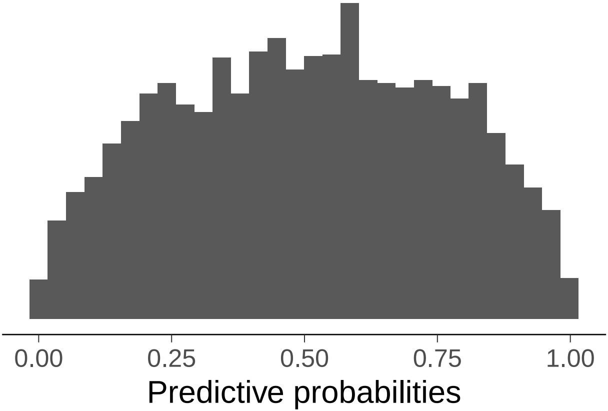
N(0,3) prior on each coefficient
30 variables



A weak prior on parameters can be a strong prior on predictions

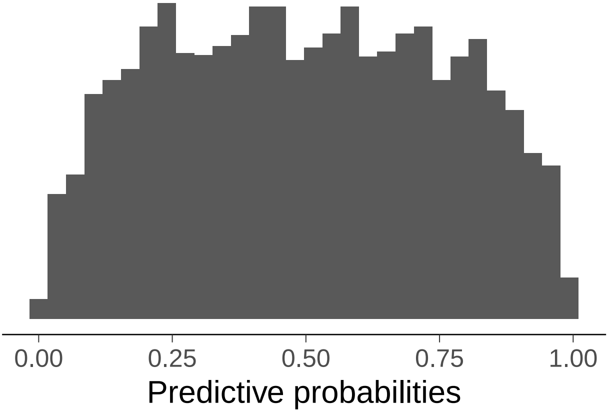
Better priors

$N(0, \frac{1}{\sqrt{p}})$ prior on each coefficient
1 variable



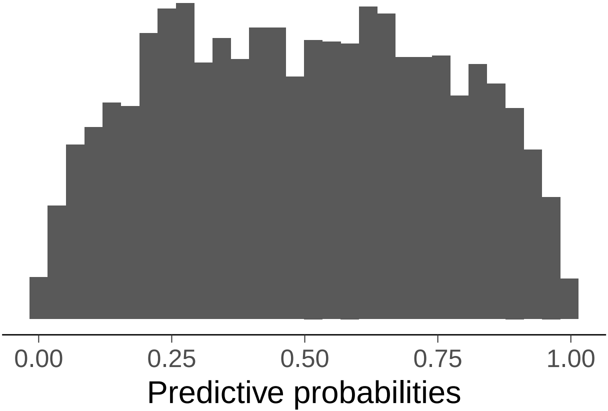
Better priors

$N(0, \frac{1}{\sqrt{p}})$ prior on each coefficient
2 variables



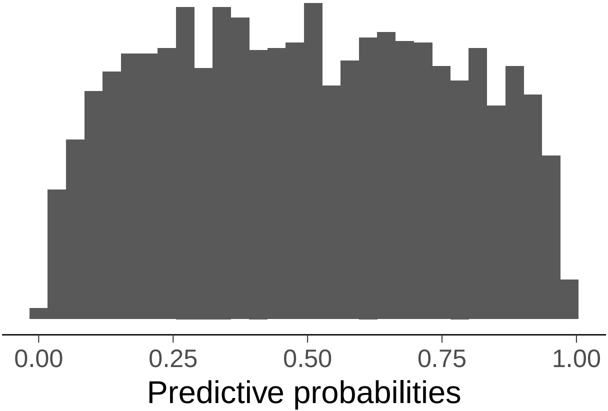
Better priors

$N(0, \frac{1}{\sqrt{p}})$ prior on each coefficient
3 variables



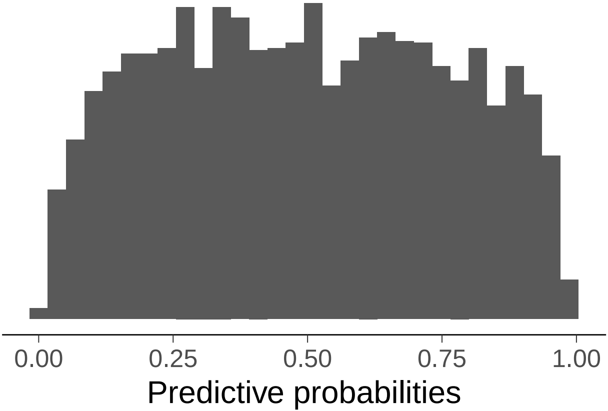
Better priors

$N(0, \frac{1}{\sqrt{p}})$ prior on each coefficient
30 variables



Better priors

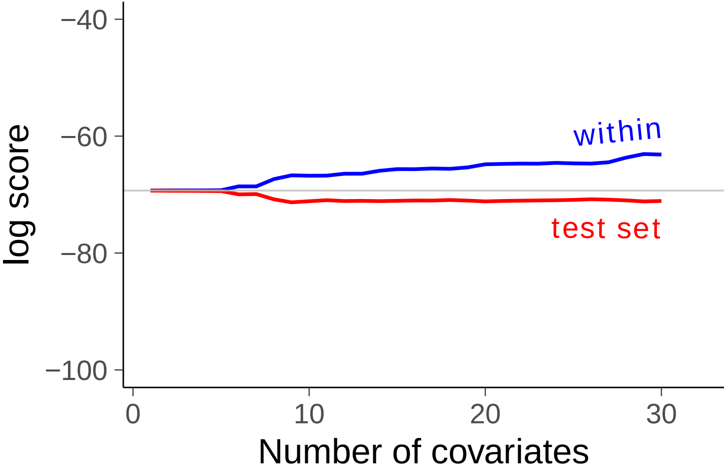
$N(0, \frac{1}{\sqrt{p}})$ prior on each coefficient
30 variables



Prior on predictions (almost) fixed when the model gets bigger

Better priors, no overfitting

logistic regression: 30 **completely irrelevant** variables, 100 observations

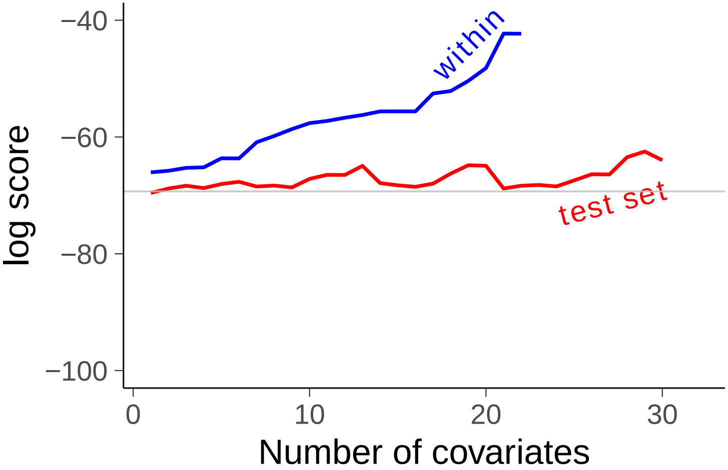


Many weak effects, wide prior on parameters

logistic regression: 30 **weakly relevant** variables,
100 observations, wide prior

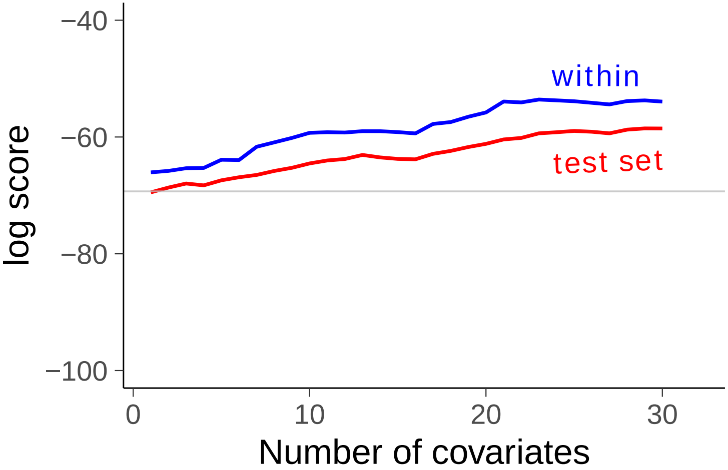
Many weak effects, wide prior on parameters

logistic regression: 30 **weakly relevant** variables,
100 observations, wide prior



Many weak effects, better prior

logistic regression: 30 **weakly relevant** variables,
100 observations, better prior

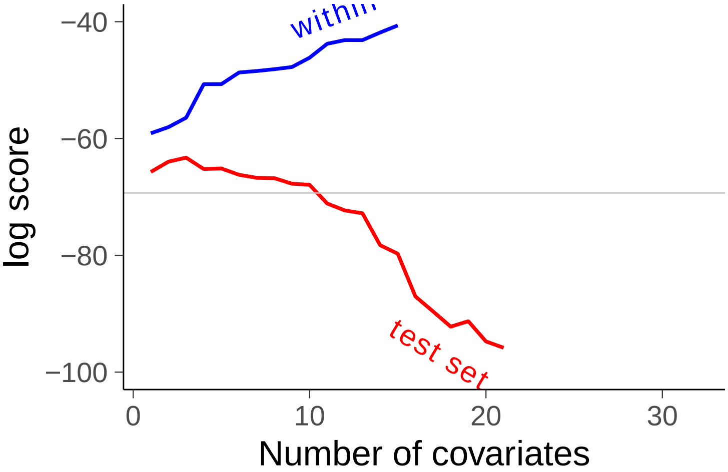


Correlating variables, wide prior on parameters

logistic regression: 30 **correlating relevant** variables,
100 observations

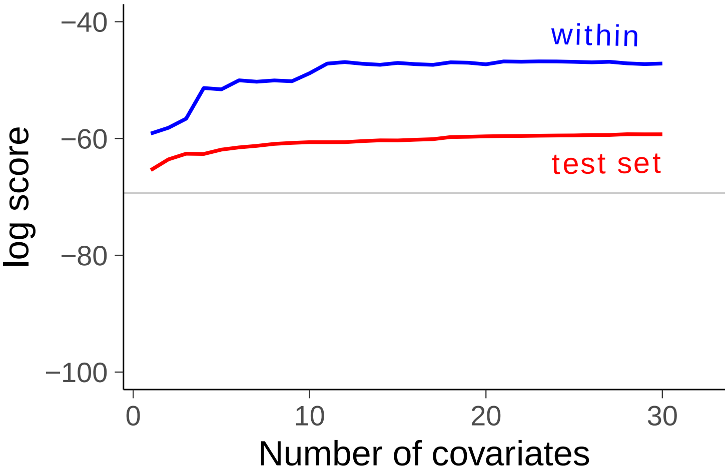
Correlating variables, wide prior on parameters

logistic regression: 30 **correlating relevant** variables, 100 observations



Correlating variables, better prior

logistic regression: 30 **correlating relevant** variables, 100 observations



Benefits of integration and prior

- Integration helps to avoid overfitting
- Integration is not able to counter bad priors