

Implementation of Bayesian R-squared for GAMMs

Abdollah Jalilian, Aki Vehtari, Luigi Sedda

2024-10-17

A new function for Bayesian R^2

A new function for computing the Bayesian R^2 not only for the Gaussian and binomial families, but also for potentially any response family implemented in R and `stanreg` models. This includes GLMMs and GAMMs with group-specific random effect terms as implemented in the `stan_glmer` function of the `rstanarm` package.

```
bayes_R2_new <- function(fit)
{
  fam <- family(fit) # family dist. of the response
  eta <- posterior_linpred(fit) # linear predictor: eta
  mu <- fam$linkinv(eta) # conditional mean
  varfit <- apply(mu, 1, var)
  # conditional variance: adjust for dispersion parameter
  varres <- switch(fam$family, gaussian={
    as.matrix(fit, pars="sigma")^2
  }, binomial={
    v <- fam$variance(mu)
    apply(v, 1, mean)
  }, poisson={
    v <- fam$variance(mu)
    apply(v, 1, mean)
  }, Gamma={
    v <- fam$variance(mu)
    apply(v, 1, mean) / as.matrix(fit, pars="shape")
  }, beta={
    v <- fam$variance(mu)
    apply(v, 1, mean)
  }, neg_binomial_2={
    size <- as.matrix(fit, pars="reciprocal_dispersion")
    v <- fam$variance(mu, theta=c(size))
    apply(v, 1, mean)
  }, inverse.gaussian={
    v <- fam$variance(mu)
    apply(v, 1, mean) / as.matrix(fit, pars="lambda")
  }, stop("the specified family is not implemented"))
  R2 <- varfit / (varres + varfit) # Bayesian R-squared
  attributes(R2) <- list(varfit=varfit, varres=varres)
  return(R2)
}
```

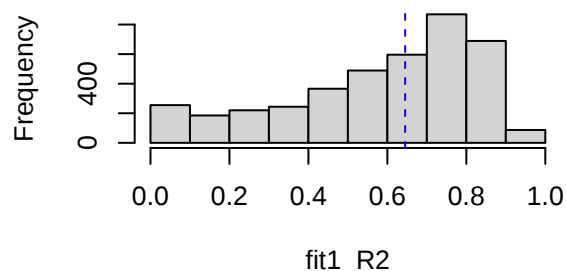
The new function produces identical results to the `bayes_R2` function in the `rstantools` package for Gaussian and binomial families.

Fitting a linear model with a Gaussian distribution for the response variable to the toy data with $n = 5$

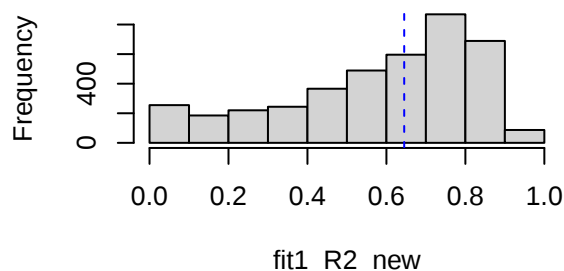
observation, as presented by Gelman et al. (2019), and a GLM with a binomial distribution and logistic link function for the binary response variable to a toy data with $n = 20$ observations as given in this webpage.

```
library("rstanarm")
xy <- data.frame(x=1:5 - 3, y=c(1.7, 2.6, 2.5, 4.4, 3.8) - 3)
fit1 <- stan_glm(y ~ x, family=gaussian(), data=xy, refresh=0)
set.seed(20)
data <- data.frame(income=1:20, rvote=rbinom(n=20, size=1, prob=(1:20 - 0.5) / 20))
fit2 <- stan_glm(rvote ~ income, family=binomial(link="logit"), data=data, refresh=0)
fit1_R2 <- bayes_R2(fit1); fit1_R2_new <- bayes_R2_new(fit1)
fit2_R2 <- bayes_R2(fit2); fit2_R2_new <- bayes_R2_new(fit2)
par(mfrow=c(2, 2))
hist(fit1_R2); abline(v=median(fit1_R2), col="blue", lty=2)
hist(fit1_R2_new); abline(v=median(fit1_R2_new), col="blue", lty=2)
hist(fit2_R2); abline(v=median(fit2_R2), col="blue", lty=2)
hist(fit2_R2_new); abline(v=median(fit2_R2_new), col="blue", lty=2)
```

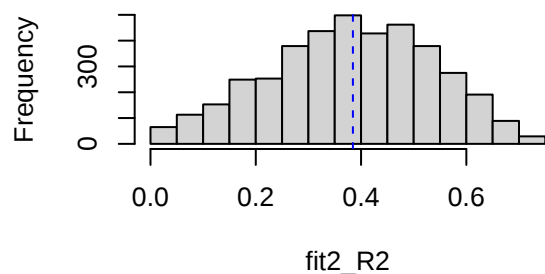
Histogram of fit1_R2



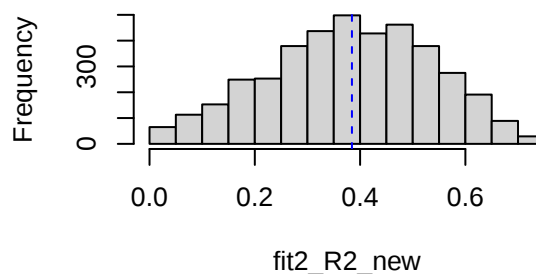
Histogram of fit1_R2_new



Histogram of fit2_R2



Histogram of fit2_R2_new



Partial Bayesian R^2

Partial R^2 can be defined by comparing a reduced model that includes either only fixed or only random effects with the full model that includes both. It represents how much of the total variation explained by the full model is due to inclusion of new terms (e.g., random effects) to the reduced model. This idea is implemented in the following function.

```
bayes_R2_partial <- function(fit, term=NULL, randall=FALSE)
{
  if (is.null(term))
  {
    term <- attr(terms(fit), "term.labels") # fixed effects
  }
}
```

```

fam <- family(fit)
eta <- posterior_linpred(fit)
mu <- fam$linkinv(eta)
R2 <- bayes_R2_new(fit)
varfit <- attr(R2, "varfit")
varres <- attr(R2, "varres")
theta <- as.matrix(fit) # extract the posterior sample
dat <- na.omit(fit$data)
eta0 <- matrix(NA, nrow=nrow(theta), ncol=nrow(dat)) # reduced model
for (i in 1:nrow(theta))
{
  eta0[i, ] <- theta[i, "(Intercept)"]
  for (xt in term)
    eta0[i, ] <- eta0[i, ] + theta[i, xt] * dat[, xt]
}
if (randall)
{
  eta0 <- eta - eta0 + theta[, "(Intercept)"]
}
mu0 <- fit$family$linkinv(eta0)
varterm <- apply(mu - mu0, 1, var)
R2_par <- varterm / (varres + varterm)
attributes(R2_par) <- list(varterm=varterm, varres=varres, varfit=varfit)
return(R2_par)
}

```

An example

To demonstrate the use of the Bayesian R^2 and its partial variants, we simulate a dataset with 200 observations from a negative binomial model with a size parameter of $\phi = 2$.

```

n <- 200
u1 <- runif(n, 0, 1)
f1 <- 2250 * (20 * u1^11 * (1 - u1)^6 + u1^3 * (1 - u1)^10) - 3/4
b1 <- rnorm(2, 0, 1)
dat <- data.frame(x1=runif(n, 0, 1),
  z1=sample(1:2, size=n, replace=TRUE),
  u1=u1, f1=f1)
eta <- 3 + 2 * dat$x1 + b1[dat$z1] + dat$f1
dat$y <- rnbino(n, mu=exp(eta), size=2)

```

We then fit three models to the data using the `rstanarm` package in R:

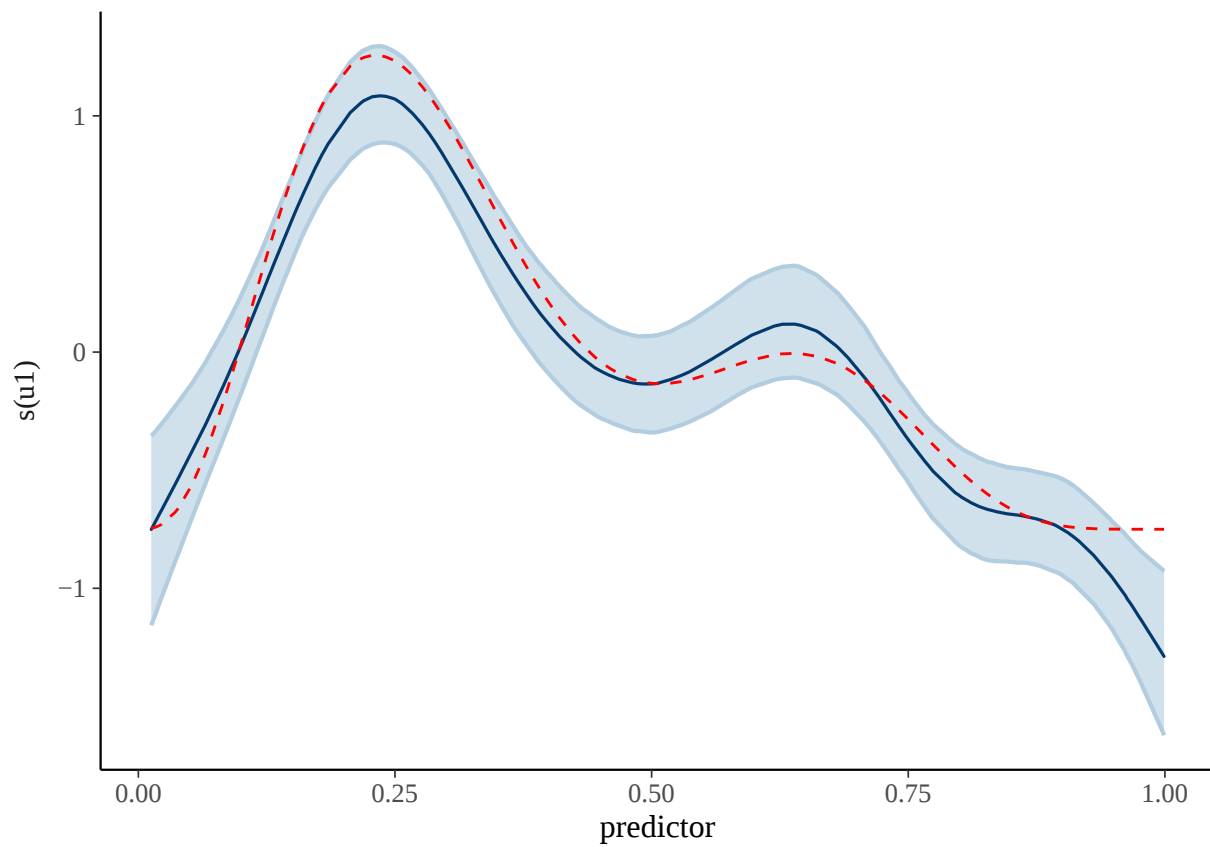
- A model with only fixed effects,
- A model with both fixed and random effects,
- A model that includes fixed effects, random effects, and a smooth term.

```

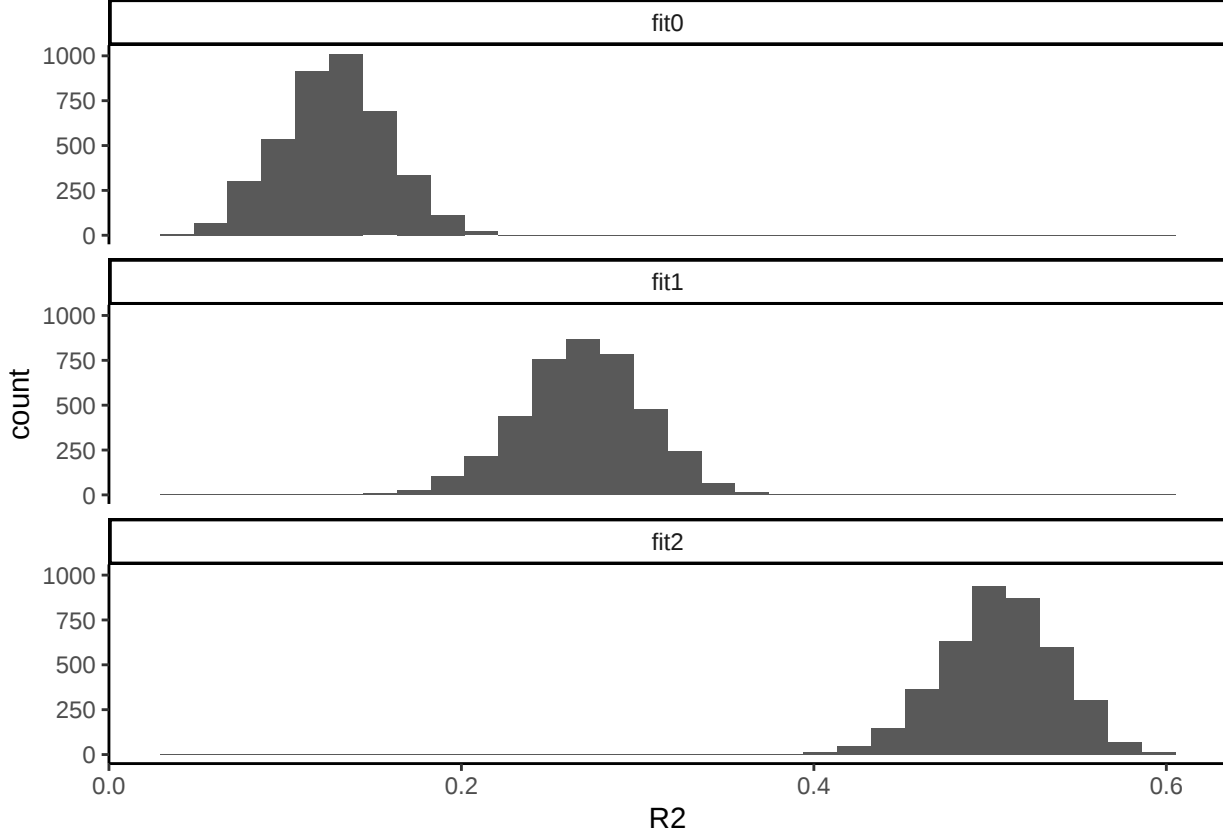
fit0 <- stan_glm(y ~ x1, data=dat, family=neg_binomial_2, refresh=0)
fit1 <- stan_glm(y ~ x1 + (1|z1), data=dat, family=neg_binomial_2, refresh=0)
fit2 <- stan_gamm4(y ~ x1 + s(u1), random= ~ (1|z1),
  data = dat, family=neg_binomial_2, refresh=0)

library("ggplot2")
plot_nonlinear(fit2) + geom_line(aes(x=u1, y=f1), linetype=2, col="red")

```



```
ggplot(data=data.frame(R2=c(bayes_R2_new(fit0), bayes_R2_new(fit1), bayes_R2_new(fit2)),
  y=rep(c("fit0", "fit1", "fit2"), each=4000)),
  aes(x=R2)) +
  geom_histogram() + facet_wrap(~y, nrow=3) + theme_classic()
```



Simulation study

For $n = 100$ observations, we consider the linear combination of two fixed effect covariates and two group-specific random effects; i.e.,

$$\eta_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + b_{j(i),1} + b_{l(i),2}, \quad i = 1, \dots, n.$$

In addition, we consider different scenarios for the response variable y_i , which may follow different distributions:

- Gaussian distribution with the identity link function:
 - $y_i \sim N(\mu_i, \sigma^2)$ with $\mu_i = \eta_i$ and $\sigma^2 > 0$
 - $\mathbb{E}[y_i] = \mu_i = \eta_i$ and $\text{Var}[y_i] = \sigma^2$
 - variance function $V(x) = 1$ and dispersion parameter $\phi = 1/\sigma^2$
- binomial distribution with the logistic link function:
 - $y_i \sim \text{Bin}(1, p_i)$ with $p_i = \exp(\eta_i)/(1 + \exp(\eta_i))$
 - $\mathbb{E}[y_i] = p_i = \exp(\eta_i)/(1 + \exp(\eta_i))$ and $\text{Var}[y_i] = p_i(1 - p_i) = \exp(\eta_i)/(1 + \exp(\eta_i))^2$
 - variance function $V(x) = x(1 - x)$ and no dispersion parameter
- Poisson distribution with the logarithmic link function:
 - $y_i \sim P(\lambda_i)$ with $\lambda_i = \exp(\eta_i)$
 - $\mathbb{E}[y_i] = \lambda_i = \exp(\eta_i)$ and $\text{Var}[y_i] = \lambda_i = \exp(\eta_i)$
 - variance function $V(x) = x$ and no dispersion parameter
- Gamma distribution with the inverse link function:
 - $y_i \sim G(\alpha, \beta_i)$ with $\alpha > 0$ and $\beta_i = \eta_i$
 - $\mathbb{E}[y_i] = \alpha/\beta_i = 1/\eta_i$ and $\text{Var}[y_i] = \alpha/\beta_i^2 = \alpha\eta_i^2$
 - variance function $V(x, \phi) = x^2/\phi$ and dispersion parameter $\phi = 1/\alpha$
- Beta distribution with the logistic link function:
 - $y_i \sim B(\alpha_i, \beta_i)$ with $\alpha_i = \phi \exp(\eta_i)/(1 + \exp(\eta_i))$ and $\beta_i = \phi/(1 + \exp(\eta_i))$
 - $\mathbb{E}[y_i] = \alpha_i/(\alpha_i + \beta_i) = \exp(\eta_i)/(1 + \exp(\eta_i))$ and $\text{Var}[y_i] = \frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)^2 (\alpha_i + \beta_i + 1)} = \frac{\exp(\eta_i)}{(1 + \exp(\eta_i))^2 (1 + \phi)}$

- variance function $V(x) = x(1-x)/(1+\phi)$ and dispersion parameter $\phi > 0$
- negative binomial distribution with the logarithmic link function:
 - $y_i \sim NBin(r, r/(r+\mu_i))$ with $\mu_i = \exp(\eta_i)$ and $r > 0$
 - $\mathbb{E}[y_i] = \mu_i = \exp(\eta_i)$ and $\text{Var}[y_i] = \mu_i + \mu_i^2/r = \exp(\eta_i) + \exp(2\eta_i)/r$
 - variance function $V(x) = x + x^2/\phi$ and dispersion parameter $\phi = r$
- inverse Gaussian distribution with the
 - $y_i \sim IG(\mu_i, \lambda)$ with $\mu_i = 1/\sqrt{\eta_i}$ and $\lambda > 0$
 - $\mathbb{E}[y_i] = \mu_i = 1/\sqrt{\eta_i}$ and $\text{Var}[y_i] = \mu_i^3/\lambda$
 - variance function $V(x) = x^3/\phi$ and dispersion parameter $\phi = \lambda$

```
n <- 200 # number of observations
x1 <- runif(n, -1, 1); x2 <- rnorm(n, 0, 1 / 3) # fixed effect covariates
beta <- c(0.6, -0.75) # intercept and fixed effect coefficients
b1 <- rnorm(n / 2, 0, 0.5); b2 <- rnorm(n / 10, 0, 1) # random effects
idx1 <- sample(rep(1:(n / 2), each=2)) # group indices for random effects
idx2 <- sample(rep(1:(n / 10), each=10)) # group indices for random effects
eta <- beta[1] * x1 + beta[2] * x2 + b1[idx1] + b2[idx2] # linear predictor
dat <- data.frame(x1=x1, x2=x2, idx1=idx1, idx2=idx2,
  y1=rnorm(n, 10 + eta, sd=1), # Gaussian response
  y2=rbinom(n, 1, exp(eta) / (1 + exp(eta))), # binomial response
  y3= rpois(n, 10 * exp(eta)), # Poisson response
  y4=rgamma(n, shape=2, rate=(10 + eta)), # Gamma response
  y5=rbeta(n, 2 * exp(eta) / (1 + exp(eta)),
    2 * 1 / (1 + exp(eta))), # beta response
  y6=rrnbinom(n, size=3, mu=exp(eta)), # negative binomial response
  y7=actuar::rinvgauss(n, 1 / sqrt(eta - min(eta) + 1e-06), shape=3))
```

Fitting the appropriate model (correct distribution family and link function) to each of the above response scenarios.

```
fit1 <- stan_glm(y1 ~ x1 + x2 + (1|idx1) + (1|idx2), data=dat,
  family=gaussian(), refresh=0)
fit2 <- stan_glm(y2 ~ x1 + x2 + (1|idx1) + (1|idx2), data=dat,
  family=binomial(link="logit"), refresh=0)
fit3 <- stan_glm(y3 ~ x1 + x2 + (1|idx1) + (1|idx2), data=dat,
  family=poisson(link="log"), refresh=0)
fit4 <- stan_glm(y4 ~ x1 + x2 + (1|idx1) + (1|idx2), data=dat,
  family=Gamma(link="inverse"), refresh=0)
fit5 <- stan_glm(y5 ~ x1 + x2 + (1|idx1) + (1|idx2), data=dat,
  family=mgcv::betar(link="logit"), refresh=0)
fit6 <- stan_glm(y6 ~ x1 + x2 + (1|idx1) + (1|idx2), data=dat,
  family=neg_binomial_2(link="log"), refresh=0)
fit7 <- stan_glm(y7 ~ x1 + x2 + (1|idx1) + (1|idx2), data=dat,
  family=inverse.gaussian(), refresh=0)
```

Computing the Bayesian R^2 for the full (including both fixed and random effects) model and partial Bayesian R^2 for two the reduced models:

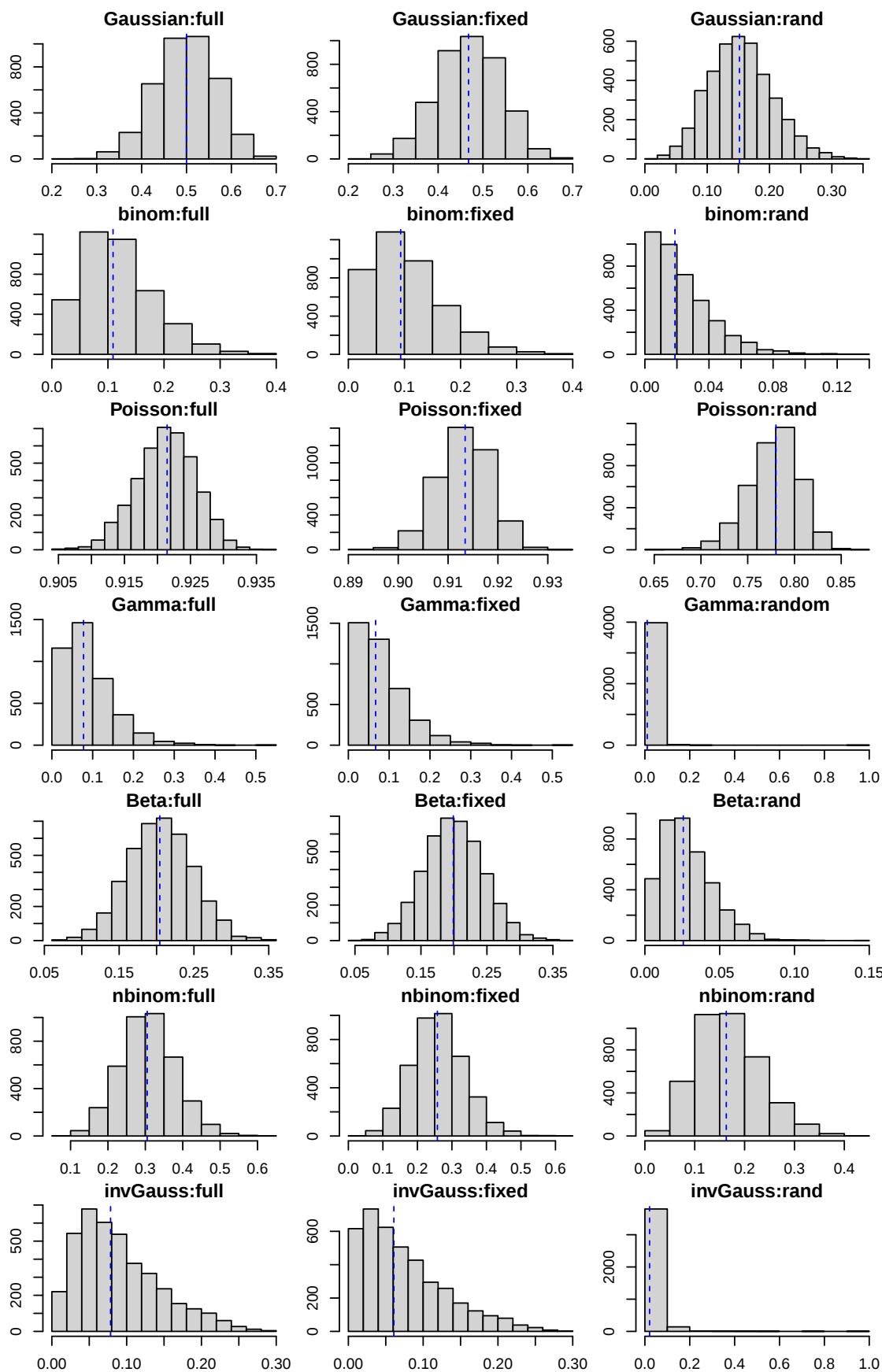
- containing only fixed effects
- containing only random effects (fixed intercept is included)

```
Rsqr1 <- bayes_R2_new(fit1); Rsqr2 <- bayes_R2_new(fit2); Rsqr3 <- bayes_R2_new(fit3)
Rsqr4 <- bayes_R2_new(fit4); Rsqr5 <- bayes_R2_new(fit5); Rsqr6 <- bayes_R2_new(fit6)
Rsqr7 <- bayes_R2_new(fit7)
Rsqrfix1 <- bayes_R2_partial(fit1); Rsqrfix2 <- bayes_R2_partial(fit2)
Rsqrfix3 <- bayes_R2_partial(fit3); Rsqrfix4 <- bayes_R2_partial(fit4)
```

```

Rsqfix5 <- bayes_R2_partial(fit5); Rsqfix6 <- bayes_R2_partial(fit6)
Rsqfix7 <- bayes_R2_partial(fit7)
Rsqrان1 <- bayes_R2_partial(fit1, randall=TRUE); Rsqrان2 <- bayes_R2_partial(fit2, randall=TRUE)
Rsqrان3 <- bayes_R2_partial(fit3, randall=TRUE); Rsqrان4 <- bayes_R2_partial(fit4, randall=TRUE)
Rsqrان5 <- bayes_R2_partial(fit5, randall=TRUE); Rsqrان6 <- bayes_R2_partial(fit6, randall=TRUE)
Rsqrان7 <- bayes_R2_partial(fit7, randall=TRUE)
par(mfrow=c(7, 3), mar=c(2, 2, 1.5, 1))
hist(Rsq1, main="Gaussian:full"); abline(v=median(Rsq1), col="blue", lty=2)
hist(Rsqfix1, main="Gaussian:fixed"); abline(v=median(Rsqfix1), col="blue", lty=2)
hist(Rsqrان1, main="Gaussian:rand"); abline(v=median(Rsqrان1), col="blue", lty=2)
hist(Rsq2, main="binom:full"); abline(v=median(Rsq2), col="blue", lty=2)
hist(Rsqfix2, main="binom:fixed"); abline(v=median(Rsqfix2), col="blue", lty=2)
hist(Rsqrان2, main="binom:rand"); abline(v=median(Rsqrان2), col="blue", lty=2)
hist(Rsq3, main="Poisson:full"); abline(v=median(Rsq3), col="blue", lty=2)
hist(Rsqfix3, main="Poisson:fixed"); abline(v=median(Rsqfix3), col="blue", lty=2)
hist(Rsqrان3, main="Poisson:rand"); abline(v=median(Rsqrان3), col="blue", lty=2)
hist(Rsq4, main="Gamma:full"); abline(v=median(Rsq4), col="blue", lty=2)
hist(Rsqfix4, main="Gamma:fixed"); abline(v=median(Rsqfix4), col="blue", lty=2)
hist(Rsqrان4, main="Gamma:random"); abline(v=median(Rsqrان4), col="blue", lty=2)
hist(Rsq5, main="Beta:full"); abline(v=median(Rsq5), col="blue", lty=2)
hist(Rsqfix5, main="Beta:fixed"); abline(v=median(Rsqfix5), col="blue", lty=2)
hist(Rsqrان5, main="Beta:rand"); abline(v=median(Rsqrان5), col="blue", lty=2)
hist(Rsq6, main="nbinom:full"); abline(v=median(Rsq6), col="blue", lty=2)
hist(Rsqfix6, main="nbinom:fixed"); abline(v=median(Rsqfix6), col="blue", lty=2)
hist(Rsqrان6, main="nbinom:rand"); abline(v=median(Rsqrان6), col="blue", lty=2)
hist(Rsq7, main="invGauss:full"); abline(v=median(Rsq7), col="blue", lty=2)
hist(Rsqfix7, main="invGauss:fixed"); abline(v=median(Rsqfix7, na.rm=TRUE), col="blue", lty=2)
hist(Rsqrان7, main="invGauss:rand"); abline(v=median(Rsqrان7, na.rm=TRUE), col="blue", lty=2)

```



The first row of plots in the above figure displays the Bayesian R^2 for the full model across various distribution scenarios for the response variable. In the second row, we observe the partial Bayesian R^2 for the reduced model containing only fixed effects compared to the full model. This metric represents the proportion of residual variation not explained by the reduced model, attributable to the inclusion of random effects in the reduced model of fixed effects. In the last row, we see the partial Bayesian R^2 for the reduced model containing only the intercept term and random effects versus the full model. Here, we measure the proportion of variation not explained by the reduced model due to the inclusion of fixed effects in the reduced model of random effects. It is important to note that because of potential interactions between random and fixed effects and also non linearity of the link function, there is no linear relation between marginal Bayesian R^2 for reduced models and the Bayesian R^2 for the full model.

For example, in the Gaussian model scenario, the full model explains approximately 60% of the variation in the future values of the response variable. When random effects are included in the reduced model, which initially contains only fixed effects, they account for around 40% of the variation that is not explained by the reduced model. However, about 50% of the variation not explained by the reduced model with random effects and intercept can be attributed to the inclusion of fixed effects. This indicates that fixed effects have a slightly higher contribution than random effect in explaining the variation in the future values of the response variable.