

Normalized Prediction Distribution Errors

Mutaz M. Jaber

September 10, 2021

Normalized prediction distribution errors (npde) is an evaluation metric used in nonlinear mixed-effect approach (NLME). This metric is a nonparametric version of population weighted residual calculation.

0.1 General notation

Let i denote the i^{th} individual and j is the j^{th} measurement point in an individual. n_{tot} is the total observations in dataset.

$$n_{tot} = \sum_1^i n_i$$

Let Y_i be the observation vector for subject i , and $f(\cdot)$ is the PK model, then for the observation y_{ij} :

$$y_{ij} = f(t_{ij}, \theta_i) + \epsilon_{ij}$$

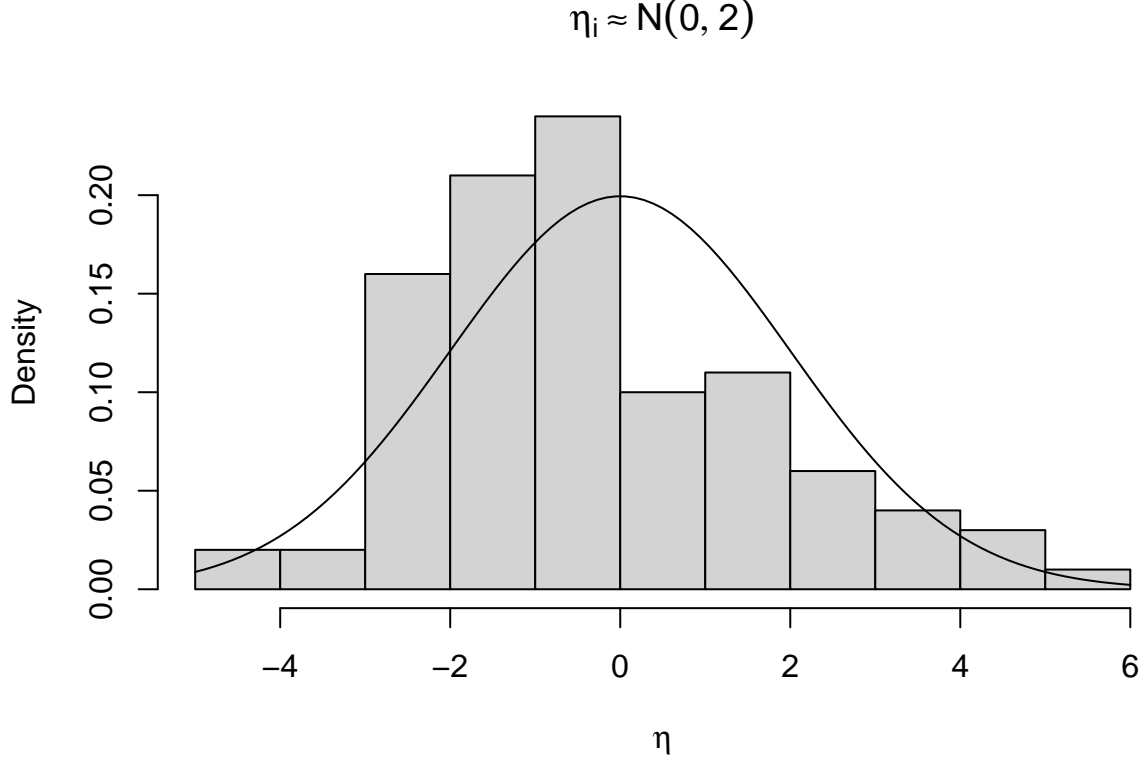
Where θ_i is vector of individual PK parameters, that we assumed it is log-normally distributed

$$\theta_i = h(\mu)e^{\eta_i}$$

Where μ is the population value, and η_i is the vector of random effects, and usually we assume that η_i are normally distributed $\eta_i \sim \mathcal{N}(0, \omega^2)$

Reminder:

```
set.seed(1234)
x <- rnorm(100, 0, 2)
hist(x, freq=F, xlab=expression(eta), main=expression(eta[i] %~~% N(0, 2)))
curve(dnorm(x, 0, 2), ylab='PDF', add = TRUE)
```



0.2 Predictive Distribution

Taking into consideration that population model f and population parameters $\{\mu, \text{diag}(\Omega), \sigma\}$ are given and they define the null model (H_0). We can define the predictive distribution (pd_i) for the i^{th} subject as:

$$p_i = \int p(y|\theta_i)p(\theta_i)d\theta_i$$

Where $p(y|\theta_i)$ is normally distributed with mean $f(t_{ij}, \theta_i)$ and variance of $\Sigma_i(\theta_i) = \text{diag}(\sigma^2 h(t_{ij}, \theta_i))$. Since we almost deal with nonlinear systems, there is no analytical expression for $p_i(y)$. Thus, we approximate the distribution using discrete distribution using stochastic simulation MCMC.

$$p_i(y) = \frac{1}{K} \sum_{k=1}^K p(y|\theta_k) = \frac{1}{K} \sum_{k=1}^K \phi(y; f(t_{ij}, \theta_k), \Sigma_i(\theta_k))$$

Where ϕ is the multivariate normal PDF with mean of μ and variance of Ω evaluated at population model value. Prediction distribution can compare the observed data with the theoretical distribution of model prediction based on MCMC simulation.

0.3 Prediction Discrepancy

Prediction discrepancy is a method uses stochastic MCMC simulation to generate prediction distribution for each observation. If the model and population parameters are correct, its expected that prediction discrepancy follows a uniform distribution: $pd_i \sim U(0, 1)$.

Let F_{ij} be the prediction distribution function of the observation y_{ij} for the i^{th} under the estimated model M . We can calculate prediction discrepancy pd_{ij} as the value of F_{ij} at observation y_{ij} .

We can simulate K datasets and let $Y_i^{sim(k)}$ denotes the vector of simulated observations for the i^{th} subject in the k^{th} dataset. pd_i is computed as the percentile of y_{ij} based on the theoritical distribution of $y_{ij}^{sim(k)}$

$$pd_{ij} = F_{ij}(y_{ij}) \approx \frac{1}{K} \sum \delta_{ijk}$$

Where δ_{ijk} is an indicator function

$$\delta_{ijk} = \begin{cases} 1 & y_{ij}^{sim(k)} < y_{ij} \\ 0 & o.w. \end{cases}$$

However, as we mentioned the distribution of pd is uniform but that the case of one observation per individual without taken into consideration with-subject correlation of multiple observation.

To overcome this issue, a decorrelation step has been proposed

$$pde_{ij} = F_{ij}^*(y_{ij}^*) \approx \frac{1}{K} \sum \delta_{ijk}^*$$

0.4 Computation of NPDE

Applying the inverse normal distribution function to pde_{ij} we can obtain normalized prediction distribution error ($npde_{ij}$):

$$npde_{ij} = \phi^{-1}(pde_{ij})$$

With that, if H_0 is true, them NPDE will follow a standard normal distribution $npde \sim \mathcal{N}(0, 1)$

0.5 NONMEM example

Lets go back to PK hello world;IV-bolus one-compartment model with linear elimination. The implementation is simple in NONMEM, few parameters are added to **\$TABLE** block for postprocessing

```
$TABLE ... NPDE NPD ESAMPLE=n
```

The full model:

```
$PROB base
; Drug X; one compartment
; estimat
$INPUT ID TIME AMT DROP DV MDV NUM
$DATA ../../../../data/dat1.csv ignore=@
```

```
$SUBR ADVAN1 TRANS2
$EST MET=1 PRINT=5 NOABORT INTER
```

```
$PK
TVCL = THETA(1)
TVV = THETA(2)

CL = TVCL * EXP(ETA(1))
V = TVV * EXP(ETA(2))

S1 = V
```

```
$ERROR
```

```
IPRED = F
```

```
Y = IPRED * (1 + ERR(1))
```

```
$THETA
```

```
(0, 5) ; CL
```

```
(0, 20) ; V
```

```
$OMEGA
```

```
0.1
```

```
0.1
```

```
$$SIGMA 0.01
```

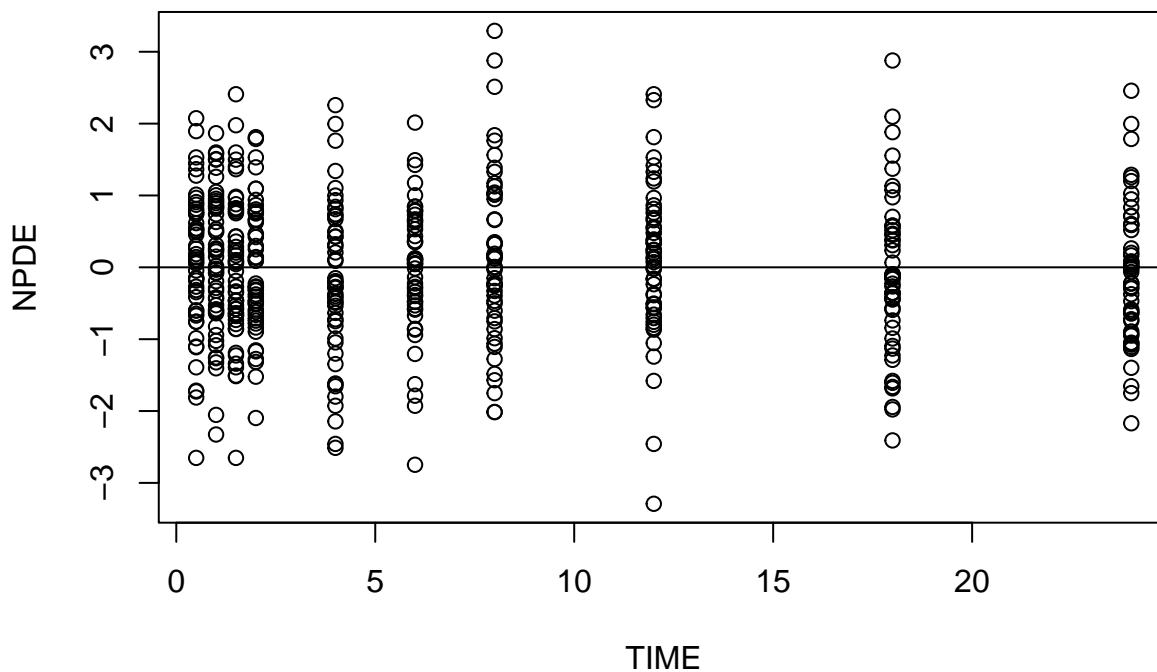
```
$COVR
```

```
$TAB ID NUM TIME AMT DV MDV IPRED PRED CWRES NPDE NPD ESAMPLE=10000 NOPRINT NOAPPEND ONEHEADER FILE=r2.
```

Where ESAMPLE is the K simulated datasets using MCMC. The default value is 300.

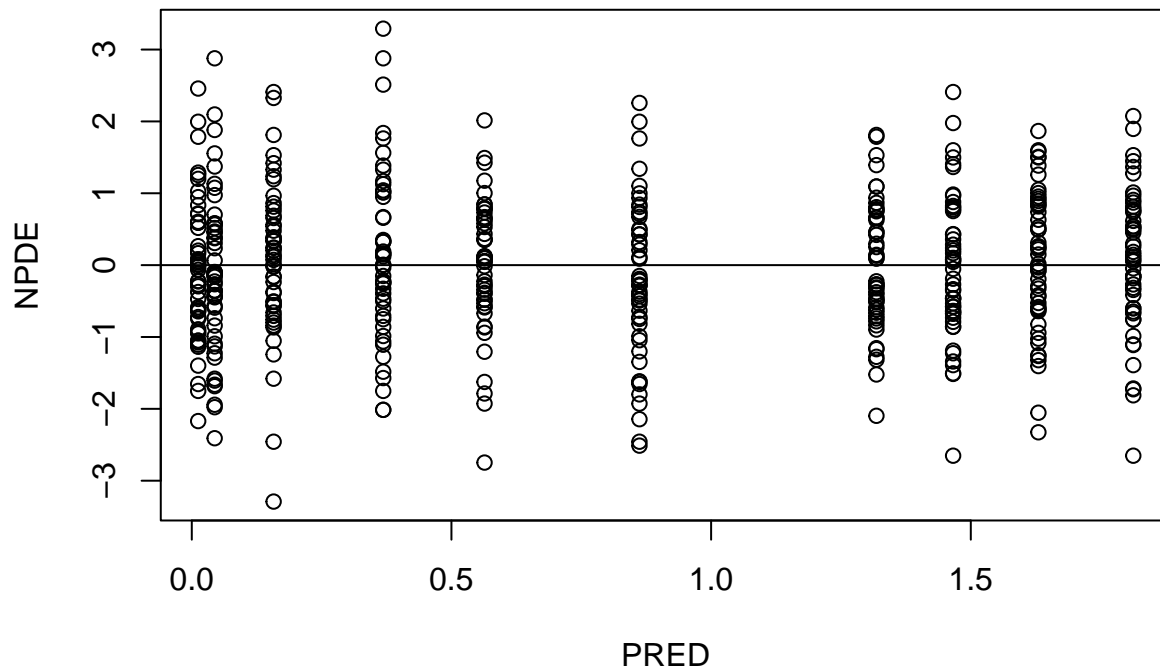
We can plot the data for correlated NPDE and decorrelated one

```
dat <- read.table('src/r1.tab', skip=1, header=TRUE)
dat <- dat[dat$CWRES!=0, ]
with(dat, plot(TIME, NPDE))
abline(h=0)
```



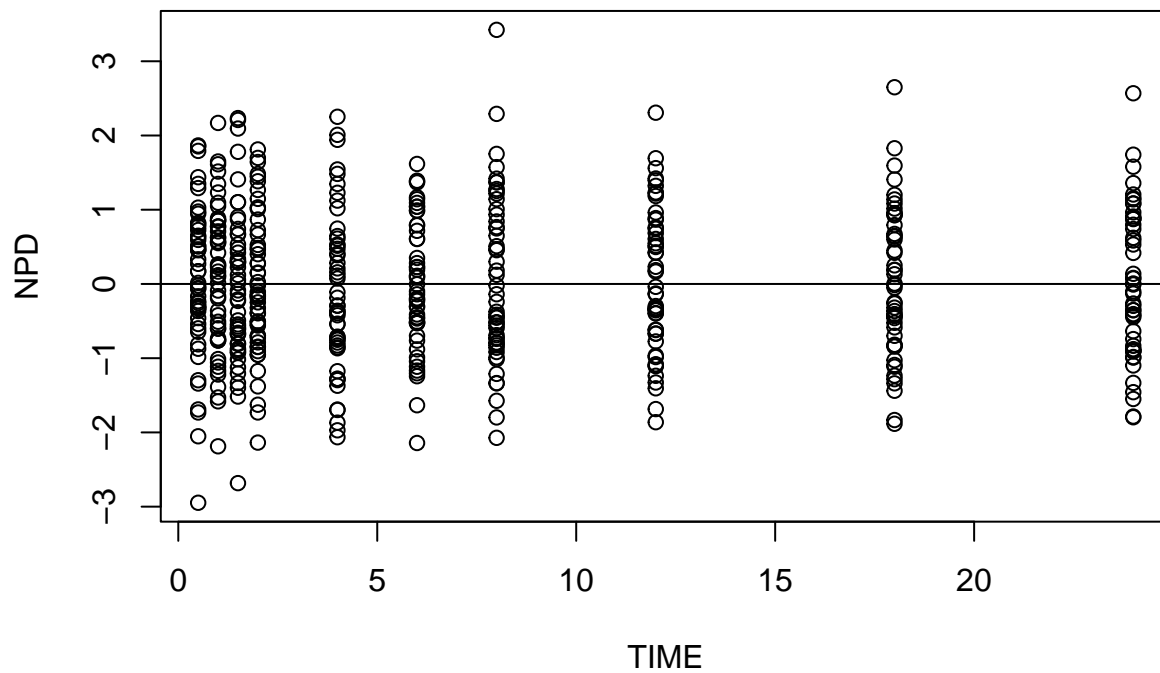
Now with population prediction

```
with(dat, plot(PRED, NPDE))
abline(h=0)
```



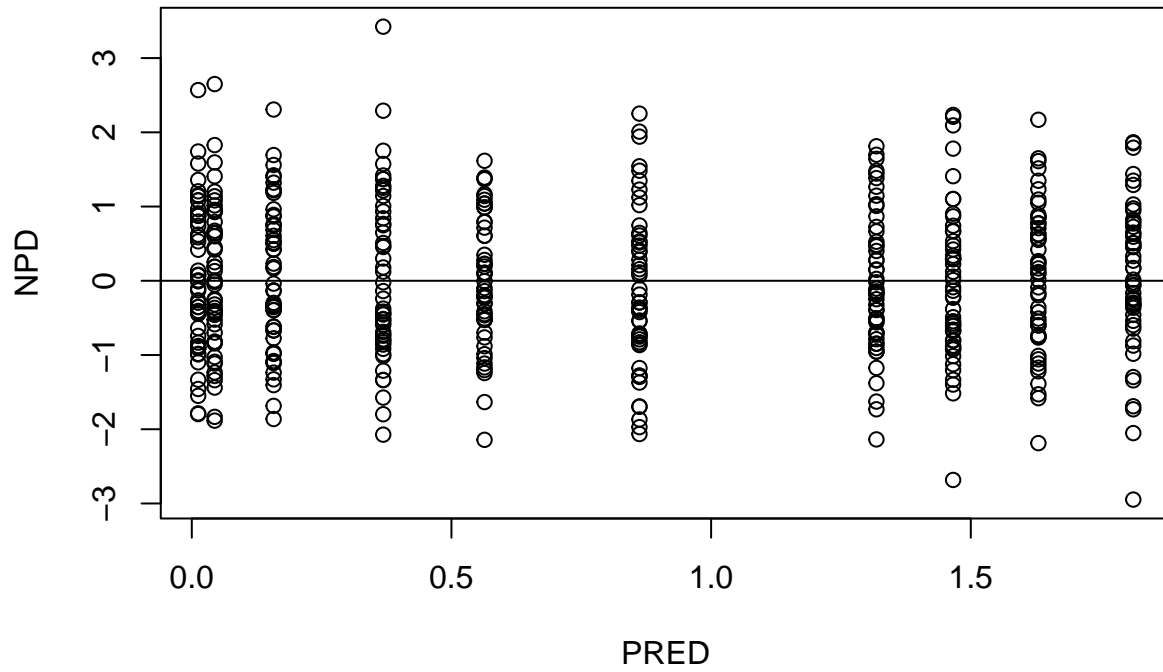
Same with correlated normalized prediction

```
with(dat, plot(TIME, NPD))
abline(h=0)
```



And again with population prediction

```
with(dat, plot(PRED, NPD))
abline(h=0)
```



References:

- Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: The npde add-on package for R [<https://pubmed.ncbi.nlm.nih.gov/18215437/>]
- NONMEM VIII User guide