In [1]:
```
import biom
import pandas as pd

from birdman import NegativeBinomial

table = biom.load_table("../data/tablebiomcorrected.biom")
metadata = pd.read_csv(
    "../data/metadatacorrected.txt",
    sep="\t",
    index_col=0
)
```

In [2]:
```
prevalence = table.to_dataframe().clip(upper=1).sum(axis=1)
features_to_keep = prevalence[prevalence >= 5].index.tolist()
table_filt = table.filter(features_to_keep, axis="observation")
table_filt.head()
```

Out[2]:
```
5 x 5 <class 'biom.table.Table'> with 20 nonzero entries (80% dense)
```

In [3]:
```
import pandas as pd

metadata = pd.read_csv(
    "../data/metadatacorrected.txt",
    sep="\t",
    index_col=0
)
metadata.index = metadata.index.astype(str)
metadata.head()
```

Out[3]:

sample-id	Temperature	Nitrate	Treatn	Treatment	Duplicate_n	Duplicate	Aquariu
18-c-1a	18C	control	1	18C_control	0	one	18C_contrc
18-c-1a-rep	18C	control	1	18C_control	1	two	18C_contrc
18-c-1b	18C	control	1	18C_control	0	one	18C_contrc
18-c-1b-rep	18C	control	1	18C_control	1	two	18C_contrc
18-c-1c	18C	control	1	18C_control	0	one	18C_contrc

In [4]:
```
metadata.groupby(["Temperature", "Nitrate"]).size()
```

Out[4]:
```
Temperature  Nitrate
18C          N100       18
             N50        18
             control    18
22C          N100       18
             N50        18
             control    18
26C          N100       18
             N50        18
             control    18
28C          N100       18
             N50        18
             control    18
Fluctuating  N100       18
             N50        18
             control    18
dtype: int64
```

In [5]:
```
from birdman import NegativeBinomial

nb = NegativeBinomial(
    table=table_filt,
    formula="C(Temperature, Diff, levels=['18C', '22C', '26C', '28C', 'Fluct
    metadata=metadata,
    num_iter=1000,
    chains=4,
    beta_prior=3.0,
    inv_disp_sd=0.5
)
```

In [6]:
```
nb.compile_model()
```

In [7]:
```
%%time

import logging

cmdstanpy_logger = logging.getLogger("cmdstanpy")
cmdstanpy_logger.disabled = True

nb.fit_model()
```

```
chain 1 |          | 00:00 Status
chain 2 |          | 00:00 Status
chain 3 |          | 00:00 Status
chain 4 |          | 00:00 Status

CPU times: user 2min 4s, sys: 26.6 s, total: 2min 31s
Wall time: 20h 17min 30s
```

In [8]:
```
inference = nb.to_inference()
```

In [9]:
```
import birdman.diagnostics as diag

diag.loo(inference, pointwise=True)
```

```
/usr/local/lib/python3.11/site-packages/arviz/stats/stats.py:803: UserWarnin
g: Estimated shape parameter of Pareto distribution is greater than 0.7 for
one or more samples. You should consider using a more robust model, this is
because importance sampling is less likely to work well if the marginal post
erior and LOO posterior are very different. This is more likely to happen wi
th a non-robust model and highly influential observations.
  warnings.warn(
```

Out[9]:

```
Computed from 4000 posterior samples and 194940 observations log-likelihood
matrix.

         Estimate       SE
elpd_loo -169592.24   821.56
p_loo     7503.98        -

There has been a warning during the calculation. Please check the results.
------

Pareto k diagnostic values:
                         Count   Pct.
(-Inf, 0.5]   (good)     191741   98.4%
 (0.5, 0.7]   (ok)         1737    0.9%
   (0.7, 1]   (bad)        1138    0.6%
   (1, Inf)   (very bad)    324    0.2%
```

In [10]:

```
diag.rhat(inference).max()
```

Out[10]: xarray.Dataset

▶ Dimensions:

▶ Coordinates: (0)

▼ Data variables:

inv_disp	()	float64	1.006
beta_var	()	float64	1.009

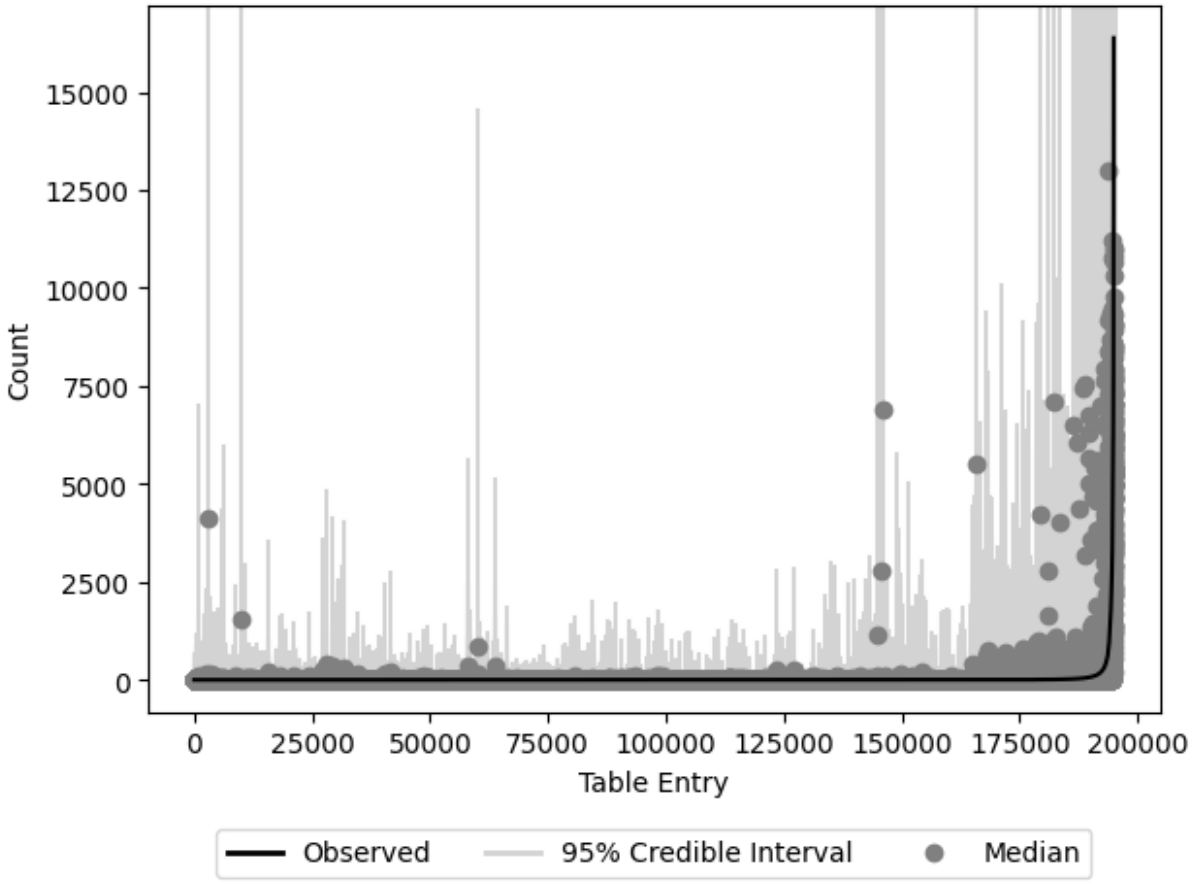
▶ Indexes: (0)

▶ Attributes: (0)

In [11]:

```
import birdman.visualization as viz
import matplotlib.pyplot as plt

viz.plot_posterior_predictive_checks(inference);
```



In [12]:
```
%%capture

from birdman import NegativeBinomial

nb_null = NegativeBinomial(
    table=table_filt,
    formula="1",
    metadata=metadata,
    num_iter=1000,
    chains=4
)
nb_null.compile_model()
nb_null.fit_model()
```

In [13]:
```
inference_null = nb_null.to_inference()
```

In [14]:
```
diag.loo(inference_null, pointwise=True)
```

```
/usr/local/lib/python3.11/site-packages/arviz/stats/stats.py:803: UserWarnin
g: Estimated shape parameter of Pareto distribution is greater than 0.7 for
one or more samples. You should consider using a more robust model, this is
because importance sampling is less likely to work well if the marginal post
erior and LOO posterior are very different. This is more likely to happen wi
th a non-robust model and highly influential observations.
  warnings.warn(
```

Out[14]:
```
Computed from 4000 posterior samples and 194940 observations log-likelihood
matrix.

         Estimate       SE
elpd_loo -187662.19   889.26
p_loo     1862.51        -

There has been a warning during the calculation. Please check the results.
------

Pareto k diagnostic values:
                         Count   Pct.
(-Inf, 0.5]   (good)     194704   99.9%
 (0.5, 0.7]   (ok)          146    0.1%
   (0.7, 1]   (bad)          70    0.0%
   (1, Inf)   (very bad)     20    0.0%
```

In [ ]:
```
import arviz as az

az.compare({"null": inference_null, "model": inference})
```

In [ ]:
```

```