Exploration of results from Bayesian modelling

[1]
```
import numpy as np
import pandas as pd
from matplotlib import cm
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.optimize import minimize
import pymc3 as pm
from tqdm import tqdm
```

```
Could not import matplotlib.animation 'ascii' codec can't decode byte
0xc2 in position 90: ordinal not in range(128)
```

Papermill - Parametrized

[2]
```
INPUT_FILE = 'n209d1000t500.pickle'
```

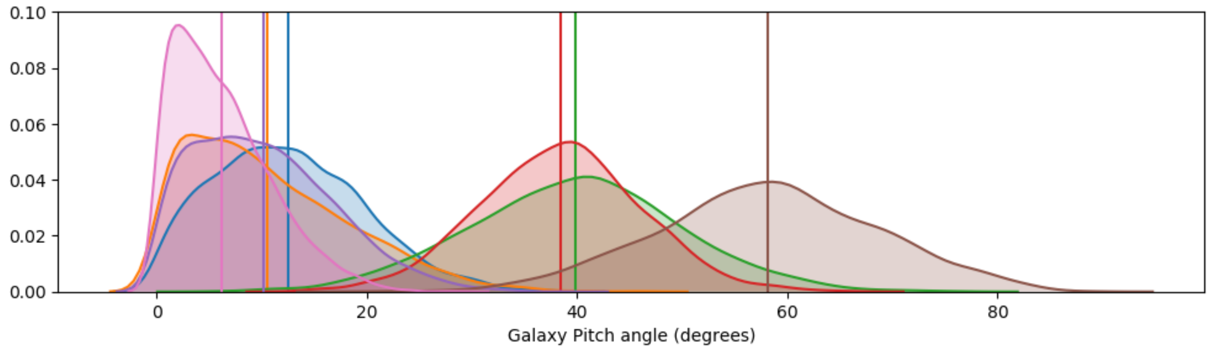
[3]
```
res = pd.read_pickle(INPUT_FILE)
bhsm = res['model']
trace = res['trace']
```

[4]
```
gal_pa_samples = pd.DataFrame(
    trace['phi_gal'].T,
    index=bhsm.galaxies.index
)
```

[5]
```
pa_expectation = gal_pa_samples.mean(axis=1)
pa_std = gal_pa_samples.std(axis=1)
```

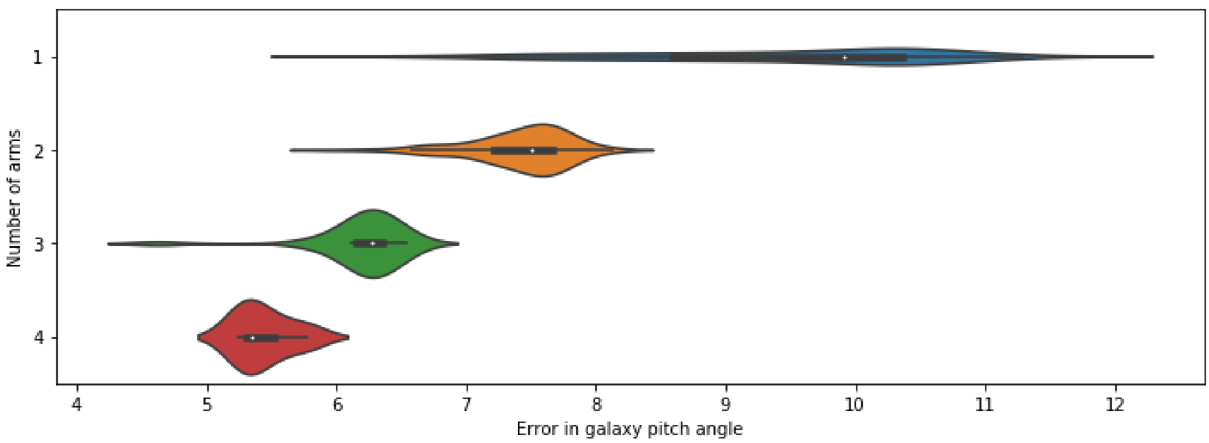
[6]
```
plt.figure(figsize=(12, 3), dpi=100)
sample = pa_expectation.sample(5)
for i, idx in enumerate(sample.index.values.tolist() +
[pa_expectation.idxmax(), pa_expectation.idxmin()]):
    sns.kdeplot((gal_pa_samples.loc[idx]), shade=True,
color=f'C{i}')
```

```
    plt.axvline(gal_pa_samples.loc[idx].mean(), color=f'C{i}')
# i = (gal_pa_samples.idxmax(), gal_pa_samples.idxmin())
plt.xlabel('Galaxy Pitch angle (degrees)')
plt.gca().get_legend().remove()
plt.savefig('plots/gal_pa_kde_sample.png', bbox_inches='tight')
```



[7]

```
plt.figure(figsize=(12, 4))
sns.violinplot(pa_std, bhsm.galaxies.apply(len),
orient='horizontal')
plt.xlabel('Error in galaxy pitch angle')
plt.ylabel('Number of arms')
plt.savefig('plots/error_vs_n_arms.png', bbox_inches='tight')
```



Comparison to length-weighted pitch angles

Let's get some values for the length-weighted pitch angle of our galaxies:

[8]

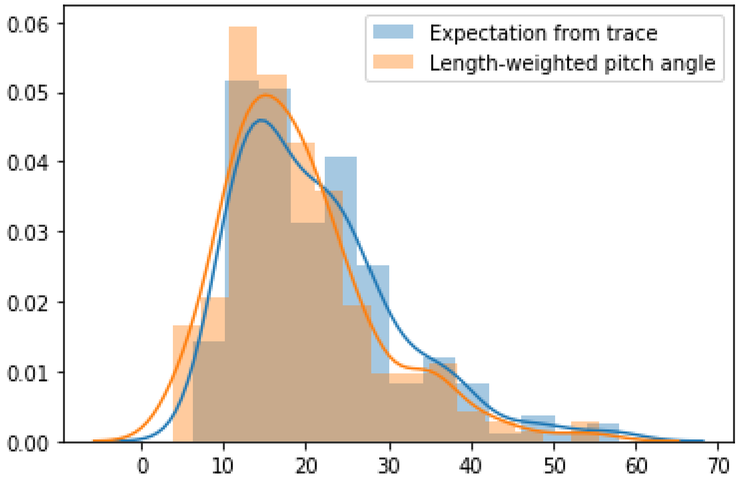
```
# sample extraction
galaxies_df = pd.read_pickle('lib/spiral_arms.pickle')
# keep only galaxies with one arm or more
galaxies_df = galaxies_df[galaxies_df.notna().any(axis=1)]
pa_lw_mean = galaxies_df.apply(
    lambda row:
row['pipeline'].get_pitch_angle(row.dropna().values[1:])[0],
```

```
    axis=1
).reindex_like(galaxies_df)
```

[9]

```
sns.distplot(pa_expectation, label='Expectation from trace')
sns.distplot(pa_lw_mean.dropna(), label='Length-weighted pitch
angle')
plt.legend()
```

```
<matplotlib.legend.Legend at 0x1c2ab315c0>
```



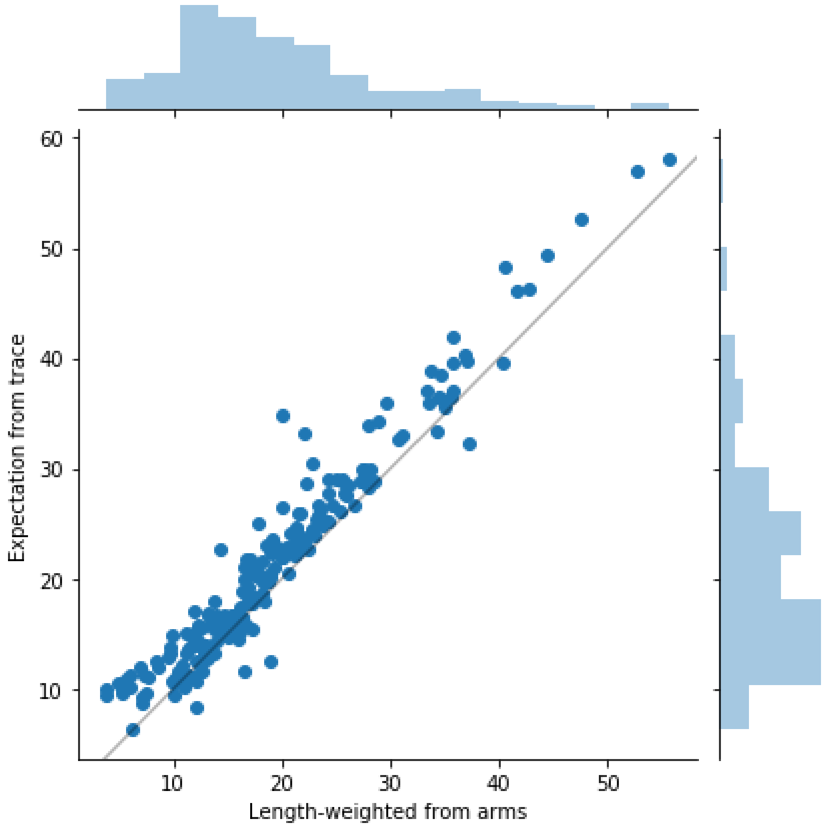
[36]

```
lw_df = pd.concat((
    pa_lw_mean.dropna().rename('lw'),
    pa_expectation.rename('E'),
    bhsm.galaxies.apply(len).rename('n'),
), axis=1).dropna()
```

[33]

```
g = sns.jointplot(
    lw_df['lw'], lw_df['E'],
).set_axis_labels('Length-weighted from arms', 'Expectation from
trace')
g.ax_joint.add_line(plt.Line2D(*[[0, 90]]*2, color='k',
alpha=0.3))
```

```
<matplotlib.lines.Line2D at 0x1c39579518>
```



The values are very consistent, as would be expected. However the measure of error generally used with length-weighted pitch angle measurement (sample error of arm segment pitch angle), does not provide as robust a measurement as the error in the posterior we obtain for ϕ_{gal}.

Morphology comparison

Does pitch angle vary with bulge or bar strength?

[12]

```
bar_fracs = pd.read_pickle('lib/bar_fractions.pkl')
bulge_fracs = pd.read_pickle('lib/bulge_fractions.pkl')
```

[13]

```
morphology_comparison = pd.concat((
    bar_fracs.rename(columns={'GZ2 bar fraction': 'bar
fraction'}).drop('GZB fraction', axis=1),
    bulge_fracs.rename(columns={'GZ2 no bulge': 'bulge
fraction'}).drop('GZB fraction', axis=1),
    pa_expectation.rename('phi'),
    pa_std.rename('sd'),
    bhsm.galaxies.apply(len).rename('n'),
), axis=1).dropna()
```

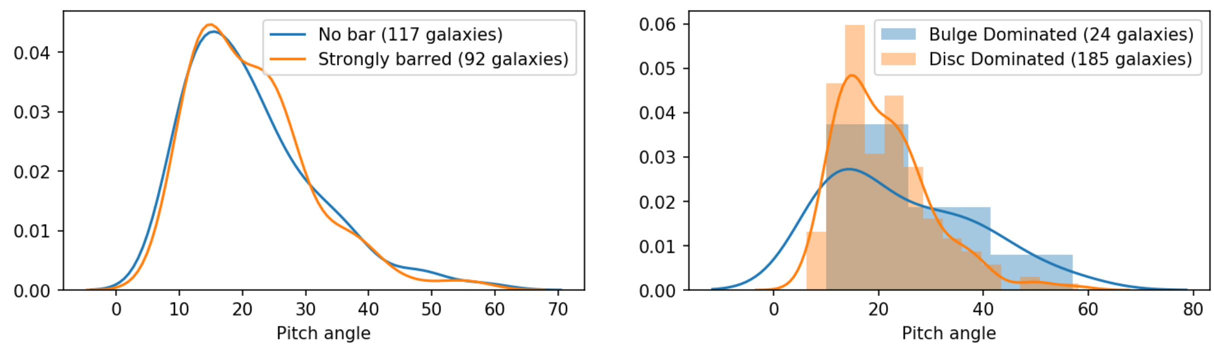
[14]

```python
n_bulge_dominated = morphology_comparison['GZ2 bulge dominated'].sum()
n_disc_dominated = np.logical_not(morphology_comparison['GZ2 bulge dominated']).sum()
n_bar = morphology_comparison['No bar'].sum()
n_no_bar = np.logical_not(morphology_comparison['No bar']).sum()
```

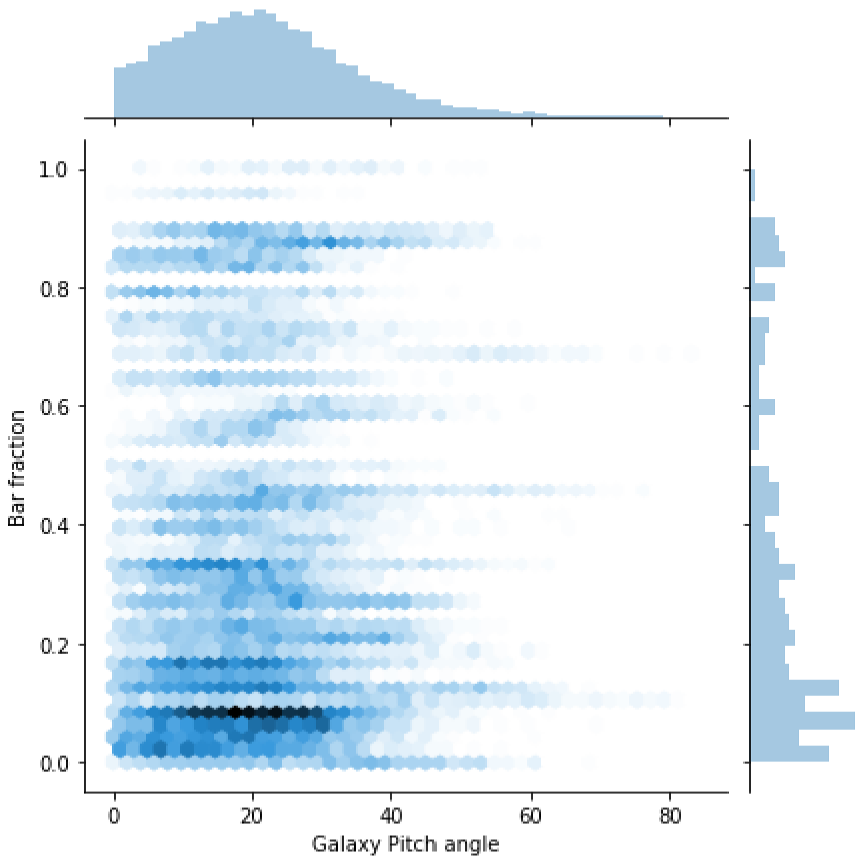
[15]

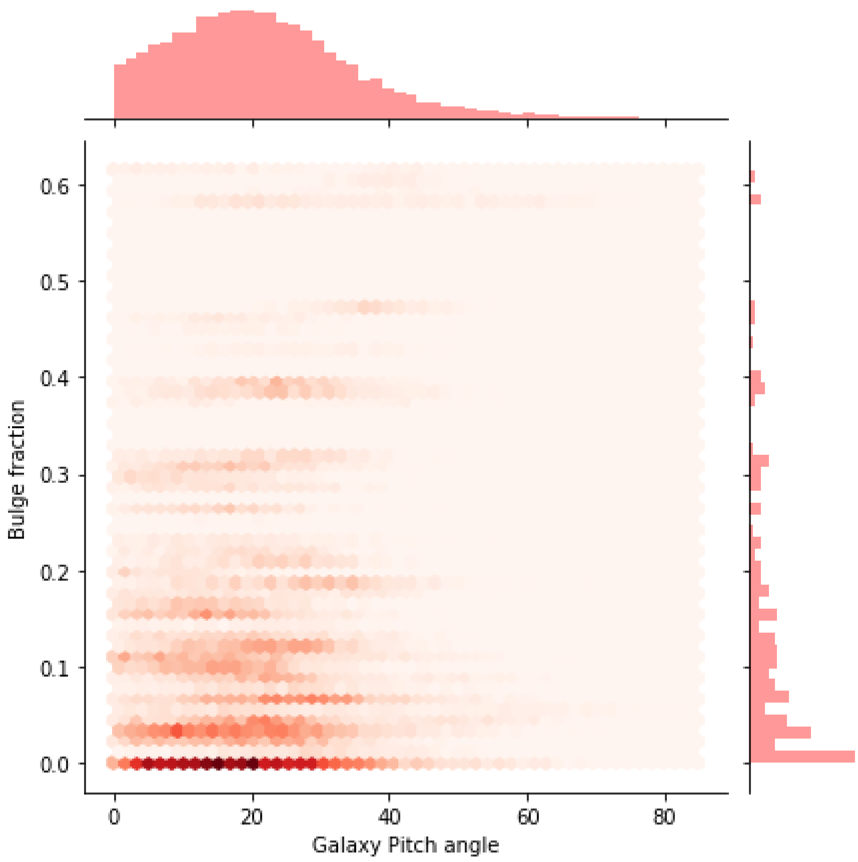
```python
plt.figure(figsize=(12, 3), dpi=150)
plt.subplot(121)
sns.kdeplot(
    morphology_comparison['phi'][morphology_comparison['No bar']].rename('No bar'),
    label=f'No bar ({n_no_bar} galaxies)'
)
sns.kdeplot(
    morphology_comparison['phi'][np.logical_not(morphology_comparison['No bar'])].rename('Bar'),
    label=f'Strongly barred ({n_bar} galaxies)'
)
plt.xlabel('Pitch angle')
plt.subplot(122)
sns.distplot(
    morphology_comparison['phi'][
        morphology_comparison['GZ2 bulge dominated']
    ],
    label=f'Bulge Dominated ({n_bulge_dominated} galaxies)'
)
sns.distplot(
    morphology_comparison['phi'][
        np.logical_not(morphology_comparison['GZ2 bulge dominated'])
    ],
    label=f'Disc Dominated ({n_disc_dominated} galaxies)'
)
plt.xlabel('Pitch angle')
plt.legend()
```

```
<matplotlib.legend.Legend at 0x1c36ba1940>
```



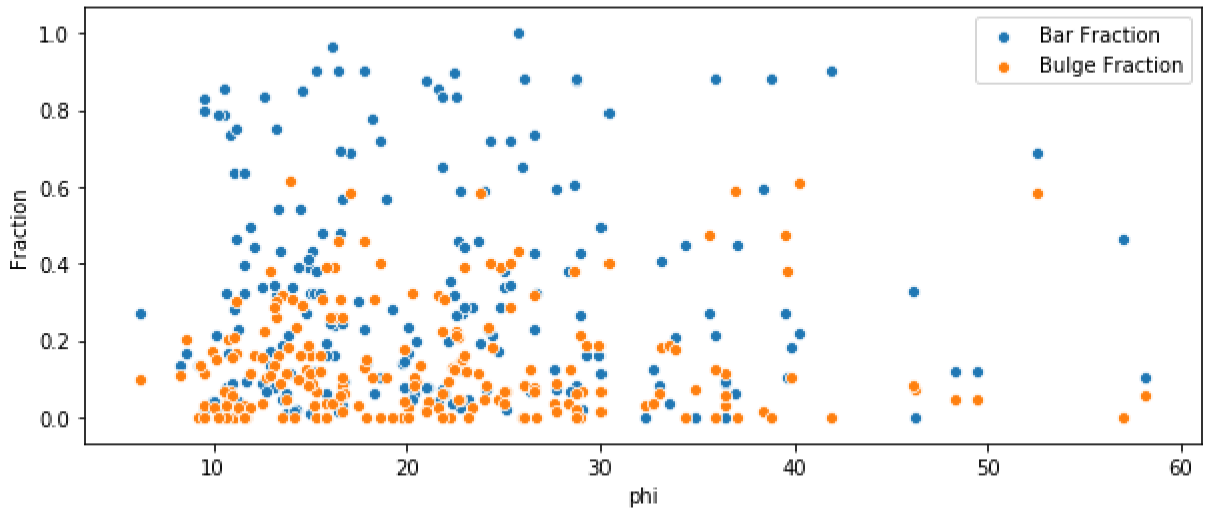
	bar fraction	bulge fraction	phi	sd	n
bar fraction	1	0.0725646	0.00778965	-0.0931531	0.170404
bulge fraction	0.0725646	1	0.0156806	-0.0155872	0.0366911
phi	0.00778965	0.0156806	1	0.139768	0.125944
sd	-0.0931531	-0.0155872	0.139768	1	-0.705629
n	0.170404	0.0366911	0.125944	-0.705629	1





[19]
```
plt.figure(figsize=(10, 4))
sns.scatterplot(x='phi', y='bar fraction',
data=morphology_comparison.dropna(), label='Bar Fraction')
sns.scatterplot(x='phi', y='bulge fraction',
data=morphology_comparison.dropna(), label='Bulge Fraction')
plt.ylabel('Fraction')
```

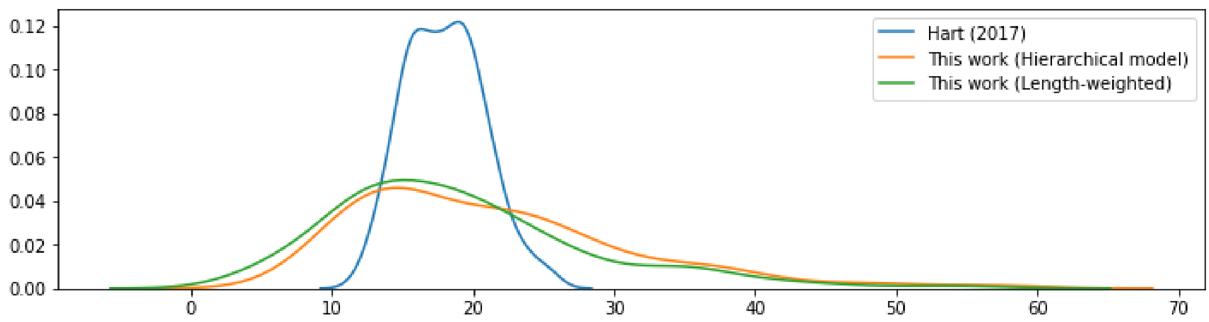
```
Text(0, 0.5, 'Fraction')
```



[20]
```
gz2_spiral_data = pd.read_csv('lib/gz2_spiral_data.csv',
index_col=0)
```

```
[21] plt.figure(figsize=(12, 3))
sns.kdeplot(gz2_spiral_data['hart_pa'].dropna(), label='Hart
(2017)')
sns.kdeplot(pa_expectation, label='This work (Hierarchical
model)')
sns.kdeplot(pa_lw_mean.dropna(), label='This work (Length-
weighted)')
```

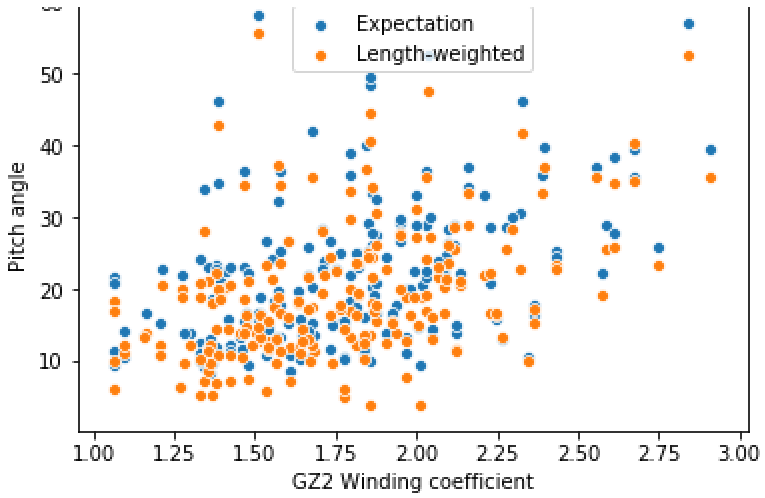
```
<matplotlib.axes._subplots.AxesSubplot at 0x1c36fceba8>
```




```
[22] sns.scatterplot(gz2_spiral_data['winding'], pa_expectation,
label='Expectation')
sns.scatterplot(gz2_spiral_data['winding'], pa_lw_mean,
label='Length-weighted')
plt.xlabel('GZ2 Winding coefficient')
plt.ylabel('Pitch angle')
pd.concat((
    gz2_spiral_data['winding'].rename('GZ2 Winding coefficient'),
    gz2_spiral_data['hart_pa'].rename('Hart (2017)'),
    pa_expectation.rename('Expectation'),
    pa_lw_mean.rename('Length-weighted'),
), axis=1).corr().style.apply(set_color_by_correlation)
```


	GZ2 Winding coefficient	Hart (2017)	Expectation	Length-weighted

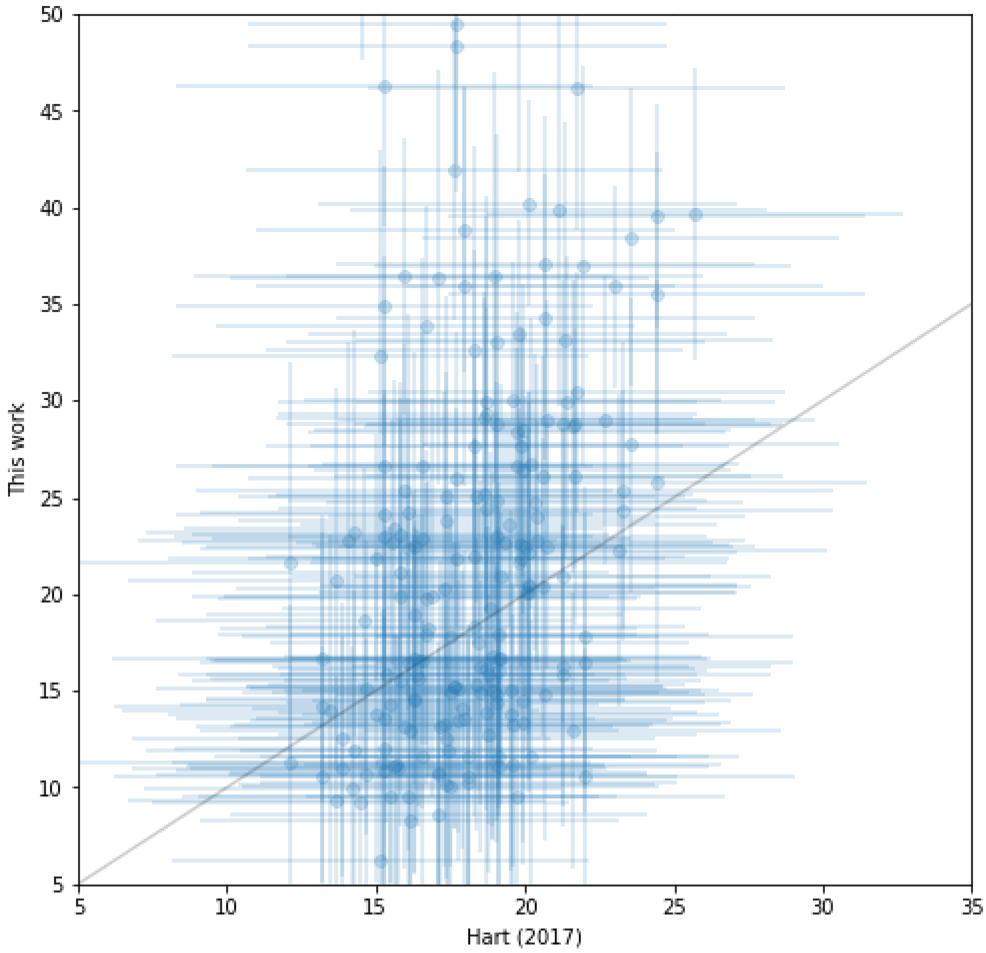
	GZ2 Winding coefficient	Hart (2017)	Expectation	Length-weighted
GZ2 Winding coefficient	1	0.957395	0.437746	0.437633
Hart (2017)	0.957395	1	0.376728	0.385381
Expectation	0.437746	0.376728	1	0.968659
Length-weighted	0.437633	0.385381	0.968659	1



[23]

```
plt.figure(figsize=(8, 8))
df = pd.concat((gz2_spiral_data['hart_pa'],
pa_expectation.rename('pa'), pa_std.rename('err')),
axis=1).dropna()
plt.errorbar(df['hart_pa'], df['pa'], xerr=7, yerr=df['err'],
fmt='o', alpha=0.2)
l = np.stack((plt.ylim(), plt.xlim()))
plt.plot((0, 90), (0, 90), c='k', alpha=0.2)
plt.xlim(5, 35)
plt.ylim(5, 50)
plt.xlabel('Hart (2017)')
plt.ylabel('This work')
```

```
Text(0, 0.5, 'This work')
```




```
[24]  import scipy.stats as st
      import warnings

      def cot(phi):
          return 1 / np.tan(np.radians(phi))

      def acot(a):
          return np.degrees(np.arctan(1 / a))

      warnings.simplefilter('ignore', UserWarning)
```

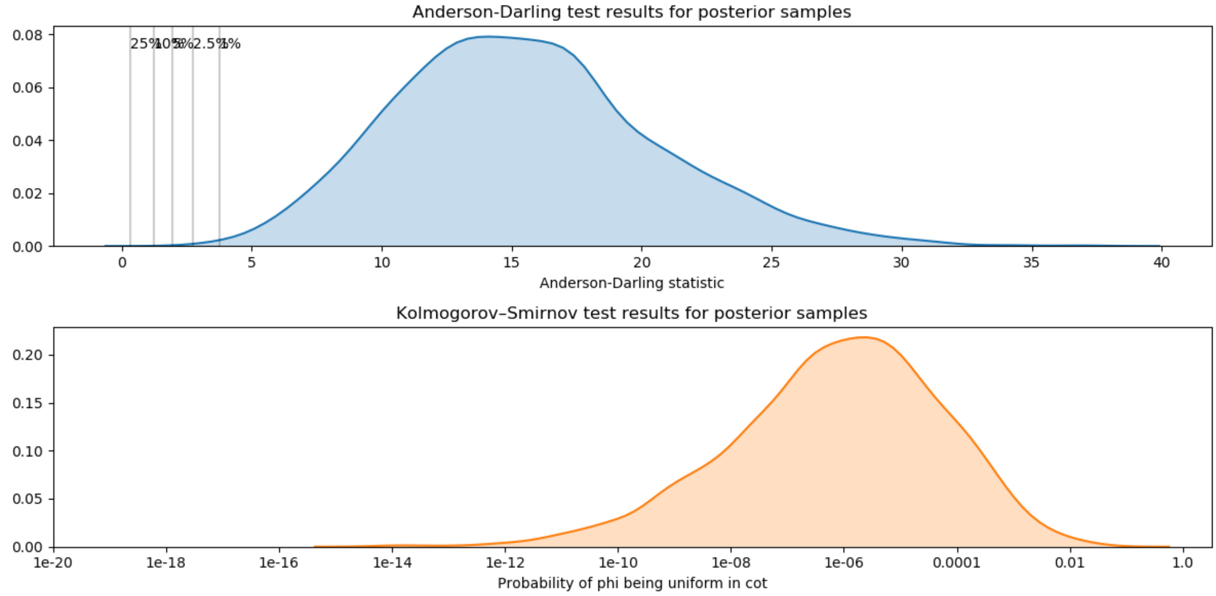
```
[25]  LOWER_COT_BOUND = 1.19
      UPPER_COT_BOUND = 4.75

      lower_phi_bound = acot(UPPER_COT_BOUND)
      upper_phi_bound = acot(LOWER_COT_BOUND)
```

```
100%|██████████| 4000/4000 [00:48<00:00, 82.74it/s]
```

```
Thresholds:
        25%: 0.33, reject 100% of the time
        10%: 1.23, reject 100% of the time
```

```
  5%: 1.96, reject 100% of the time
2.5%: 2.72, reject 100% of the time
  1%: 3.75, reject 100% of the time
```



Therefore we can reject the hypothesis that the galaxy pitch angles for our sample are uniformly distributed in $\cot$ between the limits present in Pringle & Dobbs (2019).

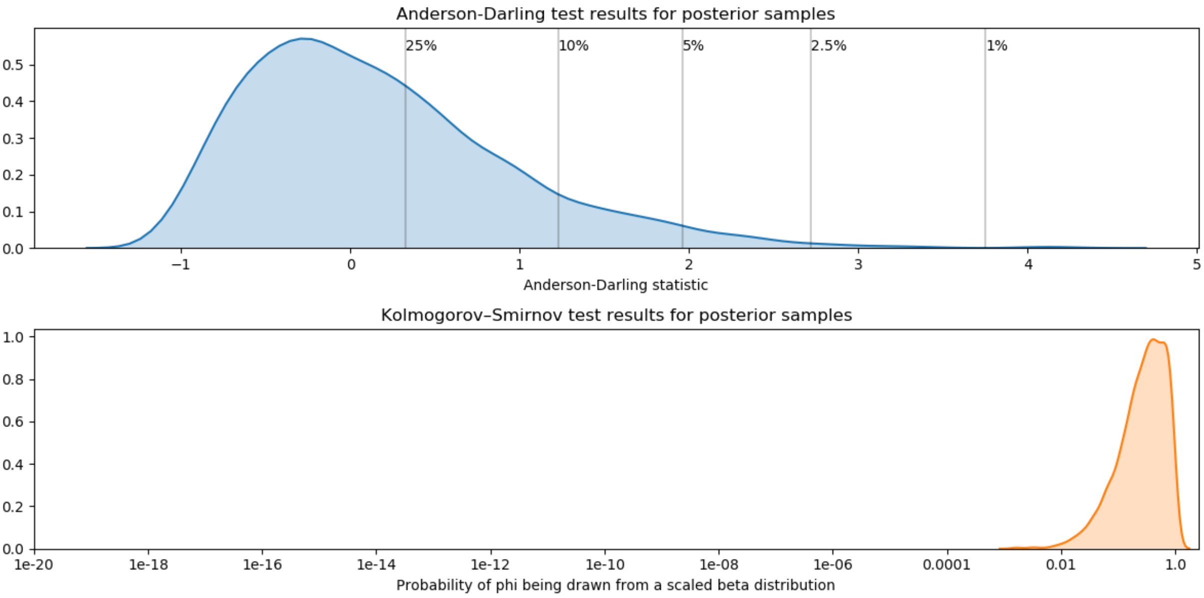
An alternate source distribution

Interestingly, the observed pitch angles **could** be drawn from a Beta distribution:

$$\phi_{\rm gal} \sim 90 \times \mathrm{Beta}(a, b)$$

```
100%|██████████| 4000/4000 [00:52<00:00, 75.65it/s]
```

```
Thresholds:
        25%: 0.33, reject 35% of the time
        10%: 1.23, reject 10% of the time
         5%: 1.96, reject 3% of the time
       2.5%: 2.72, reject 1% of the time
         1%: 3.75, reject 0% of the time
```


Anderson-Darling test results for posterior samples
25%
10%
5%
2.5%
1%
0.0
0.1
0.2
0.3
0.4
0.5
−1
0
1
2
3
4
5
Anderson-Darling statistic
Kolmogorov–Smirnov test results for posterior samples
0.0
0.2
0.4
0.6
0.8
1.0
1e-20
1e-18
1e-16
1e-14
1e-12
1e-10
1e-08
1e-06
0.0001
0.01
1.0
Probability of phi being drawn from a scaled beta distribution