post_comment_time_rng

August 21, 2021

[1]:
```
import sys
import os

sys.path.insert(0, os.getcwd() + '/reddit_download')
```

[2]:
```
import sys

import matplotlib.pyplot as plt

sys.path.append('../..')
from plotting.matplotlib_setup import configure_latex, savefig,␣
 ↪set_size_decorator, savefig, thiner_border

tex_dir, images_dir = 'porocilo/main.tex', 'porocilo/images'

configure_latex(style=['science', 'notebook'], global_save_path=images_dir)

%config InlineBackend.figure_format = 'pdf'
```

[3]:
```
from reddit_download.RWV.pushshift.time_utils import timestamp_to_utc
from reddit_download.RWV.pushshift.utils import build_df,␣
 ↪apply_df_time_transforms

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

1 Load and preprocess

[4]:
```
# TODO: make this faster and more efficient
df_comments = build_df(content_type='comment', file_path=os.getcwd() + '/
 ↪reddit_download')
df_posts = build_df(content_type='post', file_path=os.getcwd() + '/
 ↪reddit_download')
```

[5]:
```
df_comments = apply_df_time_transforms(df_comments)
```

[6]:
```
ind = df_comments[df_comments['author'] == '[deleted]'].index
df_comments.drop(ind, inplace=True)

ind = df_comments[df_comments['author'] == 'AutoModerator'].index
df_comments.drop(ind, inplace=True)

ind = df_posts[df_posts['author'] == '[deleted]'].index
df_posts.drop(ind, inplace=True)

ind = df_posts[df_posts['author'] == 'AutoModerator'].index
df_posts.drop(ind, inplace=True)
```

[7]:
```
df_comments = df_comments.rename(columns={"link_id": "post_id"})

df_comments = df_comments.rename(columns={"created_utc": "timestamp"})
df_posts = df_posts.rename(columns={"created_utc": "timestamp"})
```

[8]:
```
df_comments['post_id'] = df_comments['post_id'].apply(lambda x: x.split('_')[1])
```

[9]:
```
# df_comments.sort_values(by='post_id', inplace=True)
# df_posts.sort_values(by='post_id', inplace=True)
```

[10]:
```
df_comments.head()
```

[10]:

	author	body	timestamp	post_id	parent_id	score	subreddit	permalink	datetime	time_in_day	weekday	date
1	STL	!removehelp	1622944798	nswqwe	t3_nswqwe	1	cpp	/r/cpp/comments/nswqwe/best_way_to_transfer_da…	2021-06-06 03:59:58	14398	6	2021-06-06
2	expekted	How does it compare to this\nhttps://github.co…	1622938839	nrngit	t3_nrngit	1	cpp	/r/cpp/comments/nrngit/tree_data_structure/h0q…	2021-06-06 02:20:39			
3	Xeverous	> I almost always end up writing the entire…	1622934345	ns4hl5	t1_h0qayak	3	cpp	/r/cpp/comments/ns4hl5/what_do_you_think_of_le…	2021-06-06 01:05:45			
4	igagis	> because most people using trees have very…	1622932430	nrngit	t1_h0pzo0k	1	cpp	/r/cpp/comments/nrngit/tree_data_structure/h0q…	2021-06-06 00:33:50			
5	AissySantos	I've just started building /w `make`. But I've…	1623018225	ntjgcp	t1_h0sstsl	2	cpp	/r/cpp/comments/ntjgcp/webkitgtk_desktop_app/h…	2021-06-07 00:23:45			

```
2         8439        6  2021-06-06
3         3945        6  2021-06-06
4         2030        6  2021-06-06
5         1425        0  2021-06-07
```

[11]:
```
df_posts.head()
```

[11]:
```
                author   timestamp post_id  num_comments  score subreddit  \
0     teahaikucollage  1623670812  nzkurx            58     22       tea
3        twbluenaxela  1623660782  nziece            11      1       tea
4   khalilurrehman777  1623656090  nzhcte             0      1       tea
5        IronLuncheon  1623653922  nzgvly            12      2       tea
6  PlaneElectronic8354  1623650662  nzg3x0             0      7       tea

                                          title  \
0                                 Huffing Jasmine
3           Different types of tea drunk?   vs
4                  Where Y'all Buy your tea from?
5              Question on using a ceramic teapot
6  Discovered at a recent sale. Could you Identif…

                                          selftext  \
0  Anyone else spend at least 5 minutes huffing t…
3  Hi guys I've just started getting into tea som…
4                                         [removed]
5  I own a cast iron tea pot already at home but …
6

                                          permalink
0          /r/tea/comments/nzkurx/huffing_jasmine/
3  /r/tea/comments/nziece/different_types_of_tea_…
4  /r/tea/comments/nzhcte/where_yall_buy_your_tea…
5  /r/tea/comments/nzgvly/question_on_using_a_cer…
6  /r/tea/comments/nzg3x0/discovered_at_a_recent_…
```

[12]:
```
# from pandarallel import pandarallel

# pandarallel.initialize(nb_workers=12, progress_bar=True, use_memory_fs=None)

# post_ids = df_posts['post_id'].unique()

# def check_post_id(x, post_ids):
#     if x in post_ids:
#         return x
#     else:
#         return 0
```

```
# df_comments['post_id'] = df_comments['post_id'].parallel_apply(check_post_id,␣
↪args=(post_ids, ))
```

[13]:
```
ind = df_comments[df_comments['post_id'] == 0].index
df_comments.drop(ind, inplace=True)
```

2 Link comments to posts

[14]:
```
k, v = df_posts['post_id'], df_posts['timestamp']
id_to_timestamp = dict(zip(k, v))

import swifter

def func(x, mapping):
    try:
        return mapping[x]
    except KeyError:
        return -1

df_comments['post_time'] = df_comments['post_id'].swifter.apply(func,␣
↪args=(id_to_timestamp, ))

ind = df_comments[df_comments['post_time'] == -1].index
df_comments.drop(ind, inplace=True)
```

```
Pandas Apply:   0%|          | 0/9931626 [00:00<?, ?it/s]
```

[15]:
```
df_comments.sort_values(by='score', inplace=True)
df_posts.sort_values(by='score', inplace=True)
```

3 Times from post to comment

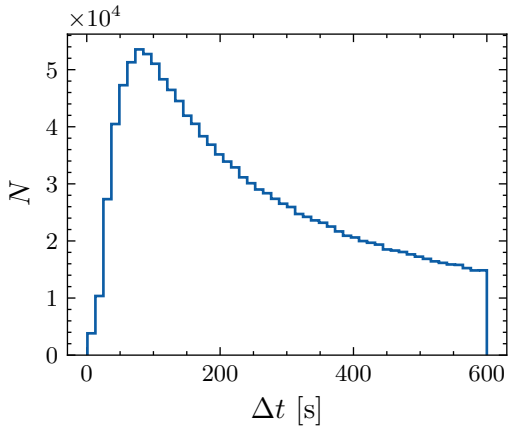
[17]:
```
from benford_helper_functions import get_first_digit, benfords_test,␣
↪construct_log_bins
from random_helper_functions import get_bitstring
from NIST_tests import RNG_test
```

[18]:
```
times = df_comments['timestamp'].values - df_comments['post_time'].values
times = times[times > 1]
```

[19]:
```
fig, ax = set_size_decorator(plt.subplots, fraction=0.5, ratio='4:3')(1, 1)

ax.hist(times, bins=50, range=[1, 60 * 10], histtype='step')
ax.ticklabel_format(style='sci', axis='y', scilimits=(0, 0))
ax.set_xlabel('$\Delta t$ [s]')
```

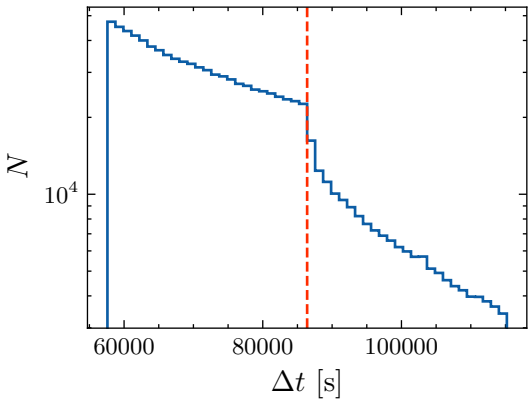
```
ax.set_ylabel('$N$')
# savefig('reddit_post_activity_10min', tight_layout=False)
```

[19]: Text(0, 0.5, 'N')



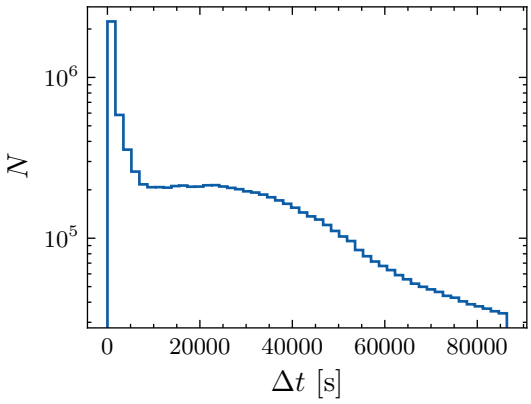
[20]:
```
fig, ax = set_size_decorator(plt.subplots, fraction=0.5, ratio='4:3')(1, 1)

plt.yscale('log')
ax.hist(times, bins=50, range=[60 * 60 * 16, 60 * 60 * 32], histtype='step')
ax.axvline(86400, lw=1, c='C3', ls='--')
ax.set_xlabel('$\Delta t$ [s]')
ax.set_ylabel('$N$')
# savefig('reddit_post_activity_16h_to32h', tight_layout=False)
```

[20]: Text(0, 0.5, 'N')



[21]:
```
fig, ax = set_size_decorator(plt.subplots, fraction=0.5, ratio='4:3')(1, 1)

ax.set_yscale('log')

ax.hist(times, bins=50, range=[1, 60 * 60 * 24 * 1], histtype='step')
ax.set_xlabel('$\Delta t$ [s]')
ax.set_ylabel('$N$')
# savefig('reddit_post_activity_1day', tight_layout=False)
```

[21]:
```
Text(0, 0.5, '$N$')
```



[22]:
```
from benford_helper_functions import do_full_rng_test
```

[23]:
```
def reshape_and_truncate(arr, shape):
    desired_size_factor = np.prod([n for n in shape if n != -1])
    if -1 in shape:  # implicit array size
        desired_size = arr.size // desired_size_factor * desired_size_factor
    else:
        desired_size = desired_size_factor
    return arr.flat[:desired_size].reshape(shape)
```

[24]:
```
split_times = np.array_split(times, 75)
```

[25]:
```
split_results = []

for s in split_times:
    f = 3
    i, j = len(s) // f, f

    a = reshape_and_truncate(s, (i, j))
    a = np.abs(a.astype(np.float64))
    b = np.prod(a, axis=1)
```

```
    f1s, fd, fracs, chi2_tests, ks_tests, df = do_full_rng_test(b,␣
↪rng_test=True, walk=False, end_bits=-1)

    split_results.append([f1s, fd, fracs, chi2_tests, ks_tests, df])
```

```
doing FT test
counting first digits
making fractions
doing chi2 test, 35.097120670142246
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.11it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.803928824360995
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.12it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.86223451648137
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.78256355433613
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.09it/s]
987597 total bits used
doing FT test
```

```
counting first digits
making fractions
doing chi2 test, 35.01113732293156
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.14it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.90263912152343
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.16it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.14391307102678
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.17773340957278
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.23it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.146124708607985
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.20it/s]
```

```
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.781381935582104
doing ks test
converting to bits

100%|
                   | 16/16 [00:07<00:00,  2.14it/s]

987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.47225111413591
doing ks test
converting to bits

100%|
                   | 16/16 [00:07<00:00,  2.19it/s]

987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.880385851812626
doing ks test
converting to bits

100%|
                   | 16/16 [00:07<00:00,  2.20it/s]

987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.253560004559624
doing ks test
converting to bits

100%|
                   | 16/16 [00:07<00:00,  2.23it/s]

987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.92116072453184
doing ks test
converting to bits
```

```
100%|
                   | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.02925726210607
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.18it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.7253283121357
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.66182351877545
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.20it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.2395721710069
doing ks test
converting to bits
100%|
                   | 16/16 [00:06<00:00,  2.29it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.52047987337511
```

```
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.27it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.896836423835744
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.22it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.001292002537106
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.23it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.01431736454608
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.25it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.15729786343361
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.16it/s]
987597 total bits used
doing FT test
counting first digits
```

```
making fractions
doing chi2 test, 34.82426772467499
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.97251006047109
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.097689100625054
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.05065946087655
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.14it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.96424551310864
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.19it/s]
```

```
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.98922423664851
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.16it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.57781117054355
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.14it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.01968788842652
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.07228158577543
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.13it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.00637097383776
doing ks test
converting to bits
```

```
100%|
                   | 16/16 [00:07<00:00,  2.18it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.79133420029255
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.14it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.39254163902836
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.34902572047839
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.18it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.717249903207296
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.11294485525274
```

```
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.16it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.17459414072059
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.919395624587104
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.88831176347634
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.07434631929861
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.18it/s]
987597 total bits used
doing FT test
counting first digits
```

```
making fractions
doing chi2 test, 34.77422007081492
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.18it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.8843391475565
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.22429146228455
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.223833980506654
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.14it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.688063771072926
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
```

```
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.99965646425794
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.11it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.14702927015544
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.894023861469826
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.855588956933396
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.10it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.68741230083426
doing ks test
converting to bits
```

```
100%|
                   | 16/16 [00:07<00:00,  2.16it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.256264359180996
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.72050590779144
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.13it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.98068861187728
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.115181147858586
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.15454660212894
```

```
doing ks test
converting to bits
100%|
                    | 16/16 [00:07<00:00,  2.26it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.7325164302473
doing ks test
converting to bits
100%|
                    | 16/16 [00:07<00:00,  2.16it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.669910586720476
doing ks test
converting to bits
100%|
                    | 16/16 [00:07<00:00,  2.17it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.87418554557812
doing ks test
converting to bits
100%|
                    | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.110739250756204
doing ks test
converting to bits
100%|
                    | 16/16 [00:07<00:00,  2.19it/s]
987597 total bits used
doing FT test
counting first digits
```

```
making fractions
doing chi2 test, 34.909955582923374
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.18it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.75593243577189
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.16it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.171904619080266
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.16it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.955658814196795
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.14it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.180465866658594
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.16it/s]
```

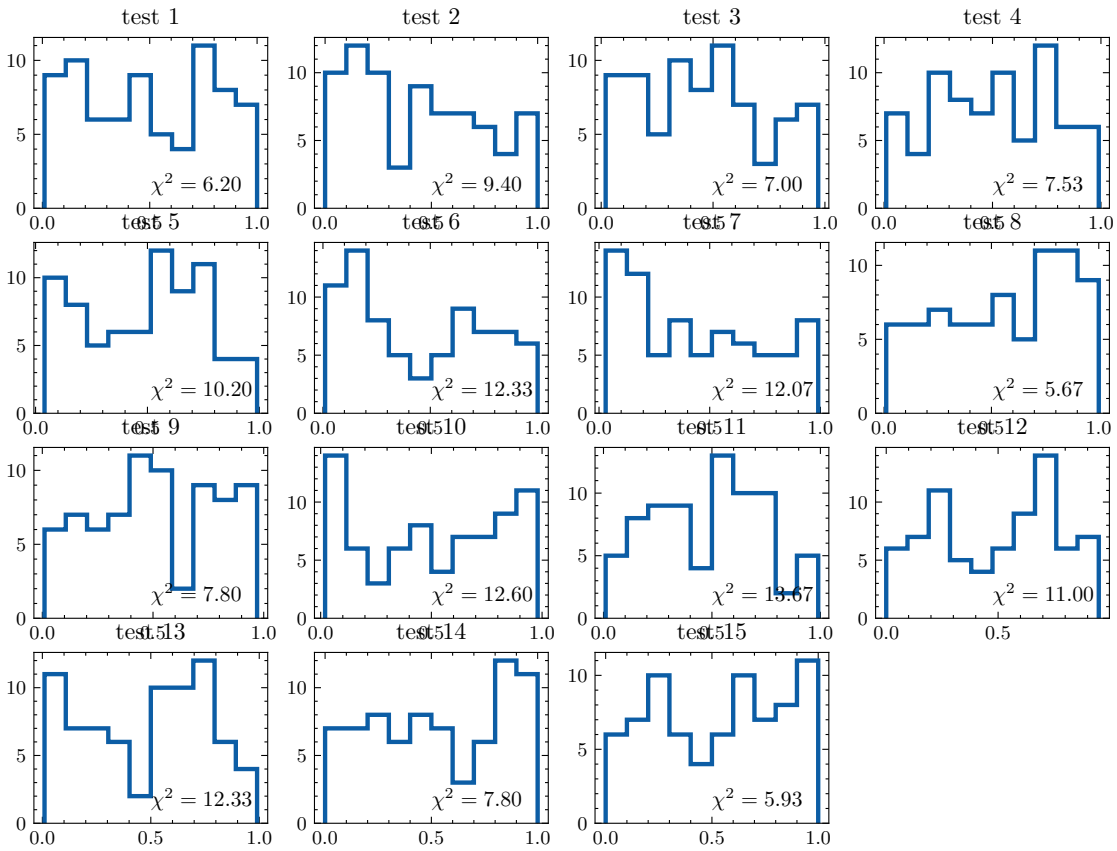
```
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.915448770997045
doing ks test
converting to bits

100%|
                   | 16/16 [00:07<00:00,  2.17it/s]

987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.280026356637656
doing ks test
converting to bits

100%|
                   | 16/16 [00:07<00:00,  2.17it/s]

987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.489541654727134
doing ks test
converting to bits

100%|
                   | 16/16 [00:07<00:00,  2.12it/s]

987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.14042231802518
doing ks test
converting to bits

100%|
                   | 16/16 [00:07<00:00,  2.19it/s]

987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.27152041123345
doing ks test
converting to bits
```

```
100%|
                   | 16/16 [00:07<00:00,  2.15it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 35.30746864578831
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.12it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.84901660349865
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.14it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.74202565448676
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.21it/s]
987597 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 34.93905732336465
doing ks test
converting to bits
100%|
                   | 16/16 [00:07<00:00,  2.23it/s]
987597 total bits used
```

[26]:
```
p_matrix = []

for r in split_results:
    p_matrix.append([float(i) for i in r[-1].iloc[0].values])

p_matrix = np.array(p_matrix)
```

[27]:
```
from stat_tests import chi2_test, ks_test

fig, ax = set_size_decorator(plt.subplots, fraction=1.5, ratio='4:3')(4, 4)
ax[-1, -1].set_visible(False)
axs = ax.flatten()

bins = 10
for i in range(p_matrix.shape[1]):
    m = p_matrix[:, i]

    t1 = chi2_test(m, n_bins=bins)
    t2 = ks_test(m)

    crit = t1[1]

    axs[i].hist(m, histtype='step', lw=2, bins=bins)
    axs[i].annotate(f'$\chi^2={t1[0][0][0]:.2f}$', xy=(0.5, 0.1),␣
 ↪xycoords='axes fraction', fontsize=10)
    axs[i].set_title(f'test {i+1}')

print(crit)
# savefig('p_test_dist')
```

```
23.209251158954356
```

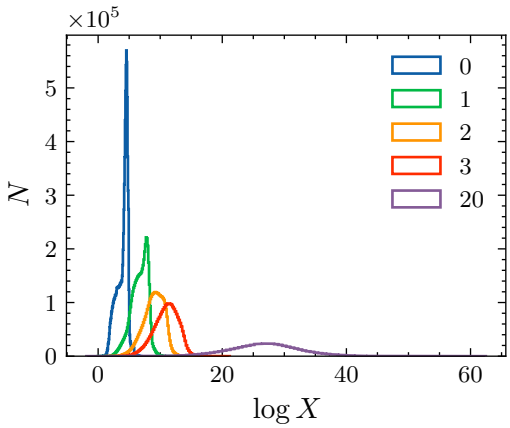


[29]:
```
fs = np.arange(1, 21, 1)

lognorms = []

for f in fs:
    i, j = len(times) // f, f

    a = reshape_and_truncate(times / f**2, (i, j))
    a = np.abs(a.astype(np.float64))
    b = np.prod(a, axis=1)

    lognorms.append(b)
```

[30]:
```
fig, ax = set_size_decorator(plt.subplots, fraction=0.5, ratio='4:3')(1, 1)


ax.hist(np.log10(lognorms[0]), bins=100, histtype='step')
ax.hist(np.log10(lognorms[1]), bins=100, histtype='step')
ax.hist(np.log10(lognorms[2]), bins=100, histtype='step')
ax.hist(np.log10(lognorms[3]), bins=100, histtype='step')
```

```
ax.hist(np.log10(lognorms[-1]), bins=100, histtype='step')

ax.ticklabel_format(style='sci', axis='y', scilimits=(0, 0))

ax.legend(['0', '1', '2', '3', '20'])

ax.set_xlabel(r'$\log X$')
ax.set_ylabel(r'$N$')

# savefig('reddit_lognorms', tight_layout=False)
```

[30]: Text(0, 0.5, 'N')



[31]:
```
results = []
for f in fs:
    i, j = len(times) // f, f

    a = reshape_and_truncate(times / f**2, (i, j))
    a = np.abs(a.astype(np.float64))
    b = np.prod(a, axis=1)

    f1s, fd, _, chi2_tests, ks_tests, df = do_full_rng_test(b, rng_test=True,␣
 ↪end_bits=10**5, walk=False)

    results.append([f1s, fd, chi2_tests, ks_tests, df])
```

```
doing FT test
counting first digits
making fractions
doing chi2 test, 255.1662469509177
doing ks test
converting to bits
```

```
100%|
                   | 16/16 [00:16<00:00,  1.04s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 165.67570178450194
doing ks test
converting to bits
100%|
                   | 16/16 [00:16<00:00,  1.06s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 147.29626893822973
doing ks test
converting to bits
100%|
                   | 16/16 [00:17<00:00,  1.06s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 134.24413547847297
doing ks test
converting to bits
100%|
                   | 16/16 [00:17<00:00,  1.06s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 124.54258289451053
doing ks test
converting to bits
100%|
                   | 16/16 [00:16<00:00,  1.04s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 117.20508181892075
```

```
doing ks test
converting to bits
100%|
                    | 16/16 [00:16<00:00,  1.05s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 111.29338781792413
doing ks test
converting to bits
100%|
                    | 16/16 [00:17<00:00,  1.07s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 106.44061224601488
doing ks test
converting to bits
100%|
                    | 16/16 [00:16<00:00,  1.06s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 102.41775030048731
doing ks test
converting to bits
100%|
                    | 16/16 [00:17<00:00,  1.07s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 98.69558630592743
doing ks test
converting to bits
100%|
                    | 16/16 [00:17<00:00,  1.06s/it]
2300000 total bits used
doing FT test
counting first digits
```

```
making fractions
doing chi2 test, 95.56086966607097
doing ks test
converting to bits
100%|
                   | 16/16 [00:16<00:00,  1.05s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 93.08079805484896
doing ks test
converting to bits
100%|
                   | 16/16 [00:17<00:00,  1.07s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 90.59561321221638
doing ks test
converting to bits
100%|
                   | 16/16 [00:16<00:00,  1.06s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 88.4981322932441
doing ks test
converting to bits
100%|
                   | 16/16 [00:17<00:00,  1.07s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 86.47846553410035
doing ks test
converting to bits
100%|
                   | 16/16 [00:17<00:00,  1.06s/it]
```

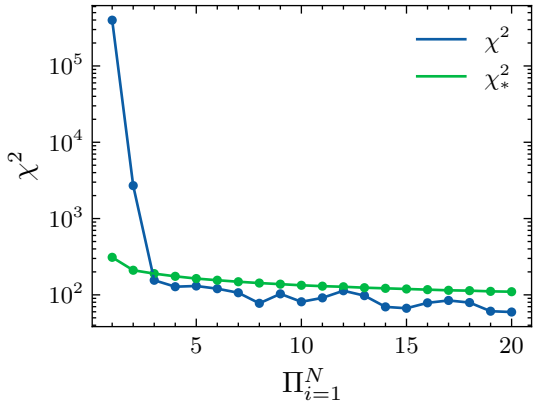
```
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 84.50297931060362
doing ks test
converting to bits
100%|
                   | 16/16 [00:17<00:00,  1.07s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 82.87426807026553
doing ks test
converting to bits
100%|
                   | 16/16 [00:16<00:00,  1.06s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 81.29066429548959
doing ks test
converting to bits
100%|
                   | 16/16 [00:16<00:00,  1.05s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 79.59314621853133
doing ks test
converting to bits
100%|
                   | 16/16 [00:16<00:00,  1.06s/it]
2300000 total bits used
doing FT test
counting first digits
making fractions
doing chi2 test, 78.72455733300659
doing ks test
converting to bits
```

```
100%|
                    | 16/16 [00:16<00:00,  1.05s/it]
2300000 total bits used
```

[32]:
```
chi2 = []
chi2crit = []
ks = []
kscrit = []
f1 = []
first = []
dfs = []

for r in results:
    f1s, fd, chi2_tests, ks_tests, df = r

    chi2.append(chi2_tests[0][0][0][0])
    chi2crit.append(chi2_tests[0][1])

    ks.append(ks_tests[0][0][0][0])
    kscrit.append(ks_tests[0][1][0])

    f1.append(f1s[0])

    first.append(fd[0][0])

    dfs.append(df)
```

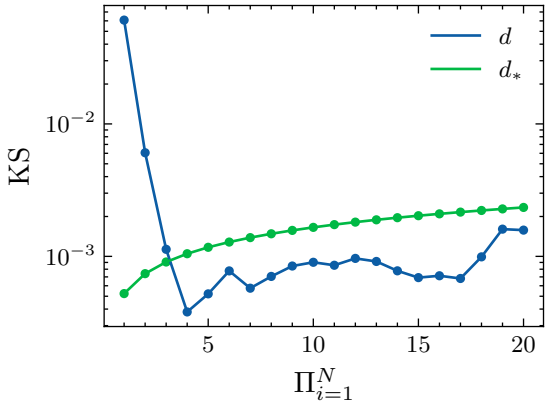
[33]:
```
fig, ax = set_size_decorator(plt.subplots, fraction=0.5, ratio='4:3')(1, 1)

ax.set_yscale('log')
ax.plot(fs, chi2, lw=1, label=r'$\chi^2$')
ax.scatter(fs, chi2, s=6)

ax.plot(fs, chi2crit, lw=1, label=r'$\chi^2_*$')
ax.scatter(fs, chi2crit, s=6)

ax.set_ylabel(r'$\chi^2$')
ax.set_xlabel(r'$\Pi_{i=1}^N$')

ax.legend()

# savefig('reddit_times_chi2')
```

[33]:
```
<matplotlib.legend.Legend at 0x7efce34eabb0>
```



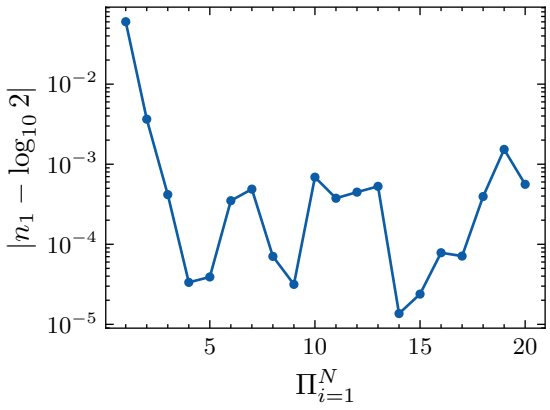
[34]:
```
fig, ax = set_size_decorator(plt.subplots, fraction=0.5, ratio='4:3')(1, 1)

ax.set_yscale('log')
ax.plot(fs, ks, lw=1, label=r'$d$')
ax.scatter(fs, ks, s=6)

ax.plot(fs, kscrit, lw=1, label=r'$d_*$')
ax.scatter(fs, kscrit, s=6)

ax.legend()

ax.set_ylabel(r'KS')
ax.set_xlabel(r'$\Pi_{i=1}^N$')

# savefig('reddit_times_ks')
```

[34]:
```
Text(0.5, 0, '$\\Pi_{i=1}^N$')
```



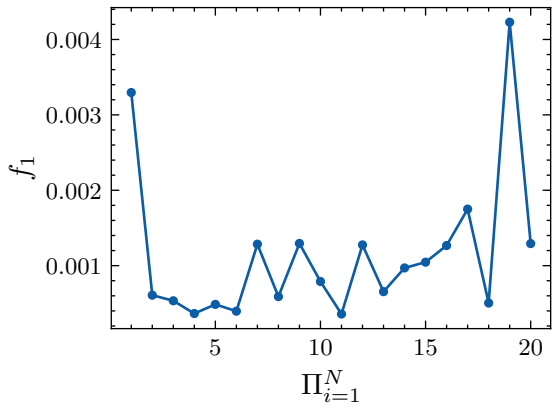
[35]:
```
fig, ax = set_size_decorator(plt.subplots, fraction=0.5, ratio='4:3')(1, 1)

ax.set_yscale('log')
ax.plot(fs, abs(first - np.log10(2)), lw=1)
ax.scatter(fs, abs(first - np.log10(2)), s=6)

ax.set_ylabel(r'$|n_1 - \log_{10}2|$')
ax.set_xlabel(r'$\Pi_{i=1}^N$')

# savefig('reddit_times_n1')
```

[35]:
```
Text(0.5, 0, '$\\Pi_{i=1}^N$')
```



[36]:
```
fig, ax = set_size_decorator(plt.subplots, fraction=0.5, ratio='4:3')(1, 1)

ax.plot(fs, f1, lw=1)
ax.scatter(fs, f1, s=6)

ax.set_ylabel(r'$f_1$')
ax.set_xlabel(r'$\Pi_{i=1}^N$')

# savefig('reddit_times_f1')
```

[36]:
```
Text(0.5, 0, '$\\Pi_{i=1}^N$')
```



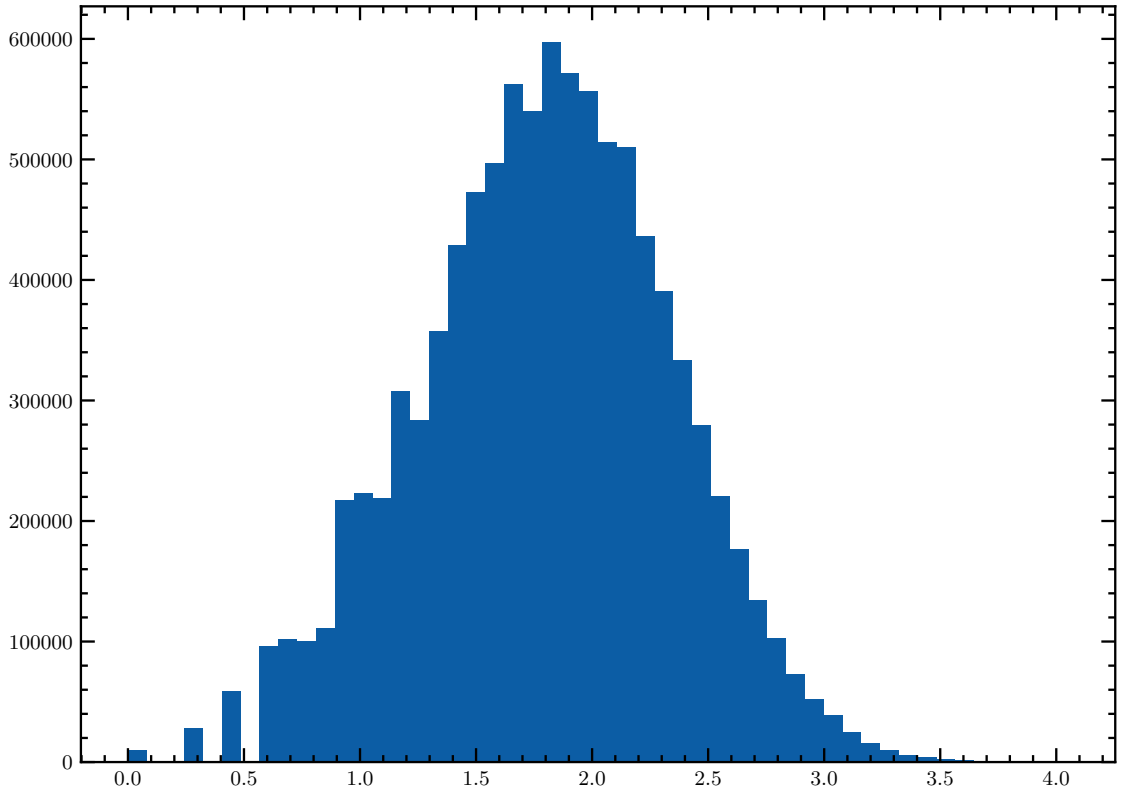
[37]:
```
# ne dela, ker ne poznamo tocne funkcijske odvisnosti g(x)
# from benford_helper_functions import normalize
# from numba import njit

# @njit
# def reject(us, g, bins, h2):
#     ys = []
#     for i, u in enumerate(us):
#         x = g[i]
#         fx = h2

#         ind = np.argmin(np.abs(bins - x))
#         gx = g[ind]

#         if u <= fx / gx:
#             ys.append(x)

#     return ys


# def uniform_from_any(g, us=None):
#     """g -> distribution used for making random numbers, u -> U(0, 1)␣
↪numbers"""
#     us = np.random.uniform(size=len(g))

#     g = g / np.max(g)

#     pdf, bins = np.histogram(g, int(np.sqrt(len(g))), density=True)
#     bins = bins[:-1]

#     s, pairs = [], []
#     for i in range(len(bins)):
```

```
#         h1, b1 = pdf[i], bins[i]
#         for j in range(len(bins)):
#             h2, b2 = pdf[j], bins[j]
#             S = b2 - b1 * h2
#             s.append(S)
#             pairs.append([h1, h2, b1, b2])

#     s = np.array(s)
#     ind = np.argsort(s)[::-1]
#     res = pairs[ind[0]]

#     h1, h2, b1, b2 = res

#     plt.plot(bins, pdf)
#     plt.scatter([b1, b2], [h2, h2])

#     ys = reject(us, g, bins, h2)

#     return ys

# y = uniform_from_any(times[times < 86400])
# plt.hist(y)
```

4 Length of comments

[38]:
```
df_comments['body_len'] = df_comments['body'].apply(len)

ind = df_comments[df_comments['body_len'] <= 0].index
df_comments.drop(ind, inplace=True)

df_comments.sort_values(by='score', inplace=True)

body_len = df_comments['body_len'].values
```

[39]:
```
plt.hist(np.log10(body_len), bins=50)
plt.show()
```



[40]:
```
f1s, first_digits, _, chi2_tests, ks_tests, df = do_full_rng_test(body_len,␣
↪rng_test=True, end_bits=10**6, walk=True)
```

```
doing FT test
counting first digits
making fractions
doing chi2 test, 207.71718961277233
doing ks test
converting to bits
```

```
100%|
                    | 16/16 [00:07<00:00,  2.18it/s]
```

```
1000000 total bits used
```

[41]:
```
ks_tests
```

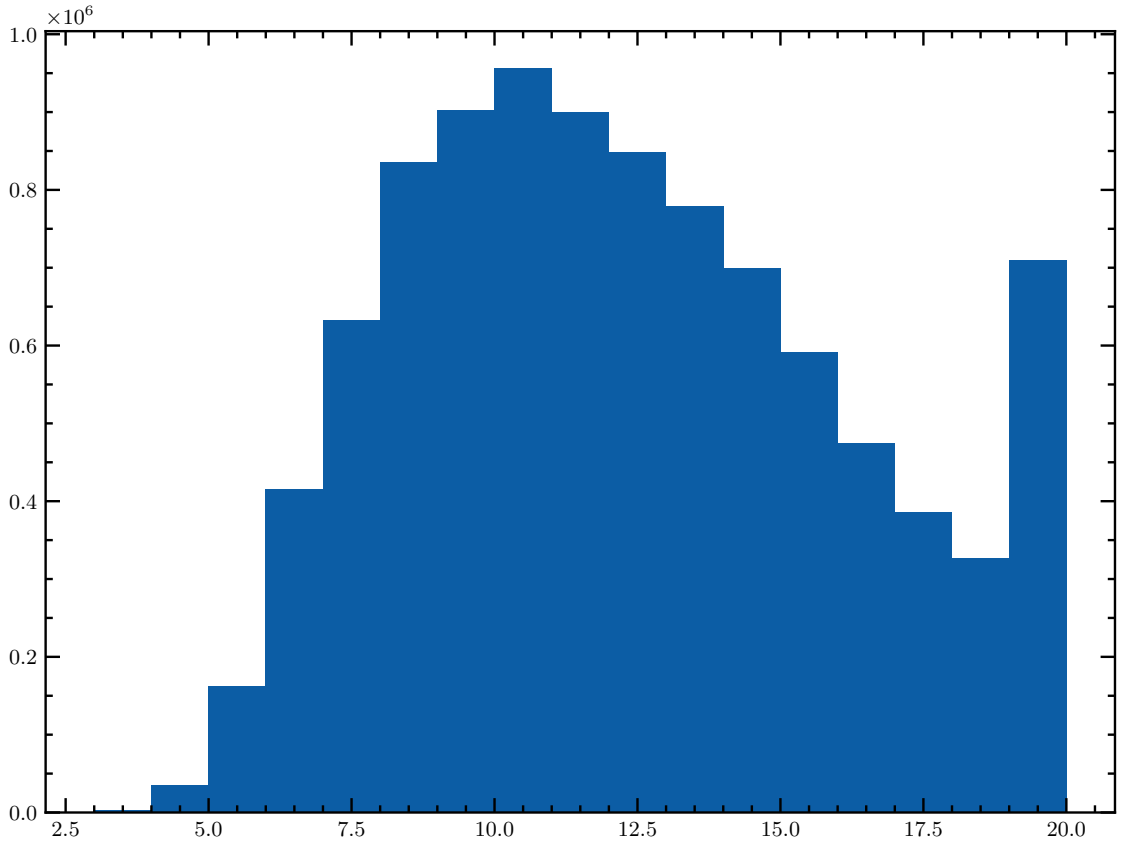
[41]:
```
[(array([[0.01881828, 0.        ]]), [0.0005236174429639164])]
```

5 Length of names

[42]:
```
df_comments['author_len'] = df_comments['author'].apply(len)
```

[43]:
```
author_len = df_comments['author_len'].values
```

[44]:
```
plt.hist(author_len, range=(3, 20), bins=17)
plt.show()
```



[ ]: