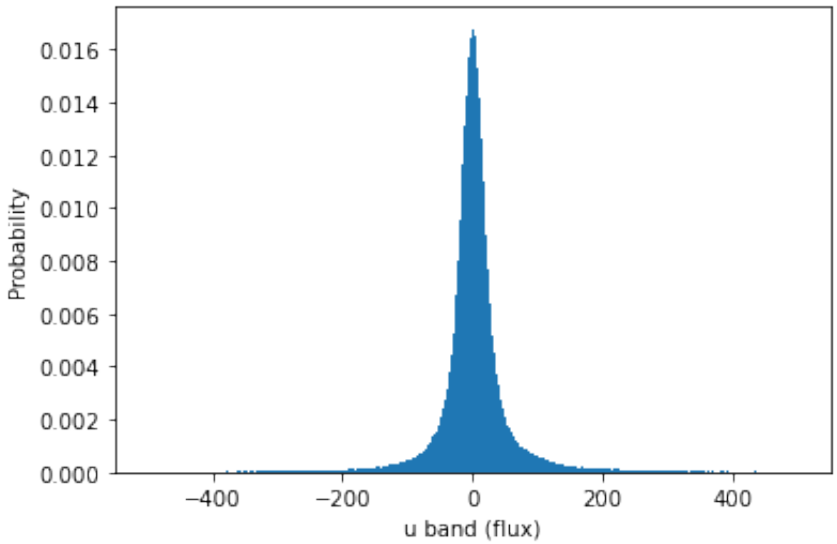
In [1]:
```
import numpy as np
import pandas as pd
import csv
import matplotlib.pyplot as plt
from matplotlib.pylab import *
from scipy.stats import norm
import matplotlib.mlab as mlab
```

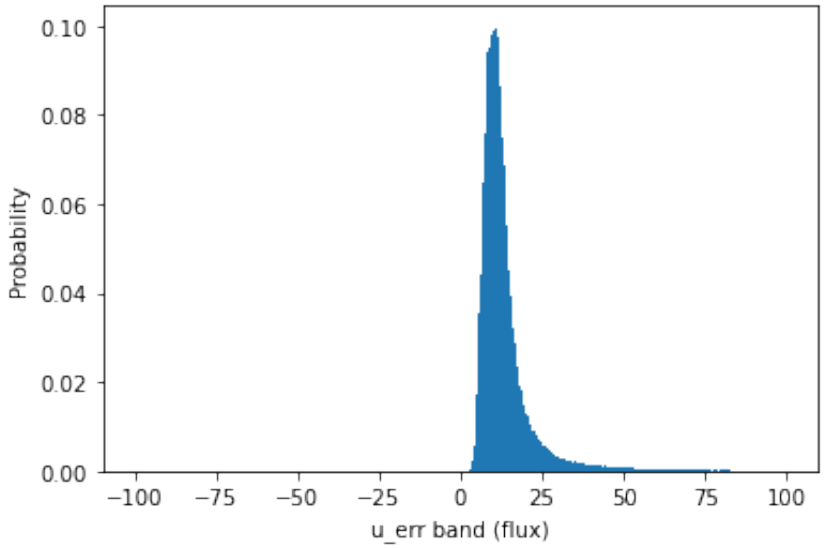
In [2]:
```
data1 = pd.read_csv('preprocessed_data/batch2_converted_AGN.csv')
data2 = pd.read_csv('preprocessed_data/batch3_converted_AGN.csv')
data3 = pd.read_csv('preprocessed_data/batch4_converted_AGN.csv')
data4 = pd.read_csv('preprocessed_data/batch5_converted_AGN.csv')
data5 = pd.read_csv('preprocessed_data/batch6_converted_AGN.csv')
data6 = pd.read_csv('preprocessed_data/batch7_converted_AGN.csv')
data7 = pd.read_csv('preprocessed_data/batch8_converted_AGN.csv')
data8 = pd.read_csv('preprocessed_data/batch9_converted_AGN.csv')
data9 = pd.read_csv('preprocessed_data/batch10_converted_AGN.csv')
data10 = pd.read_csv('preprocessed_data/batch11_converted_AGN.csv')
data11 = pd.read_csv('preprocessed_data/training_converted_AGN.csv'
)

frames = [data1, data2, data3, data4, data5, data6, data7, data8, d
ata9, data10, data11]
data = pd.concat(frames, ignore_index = True)
```

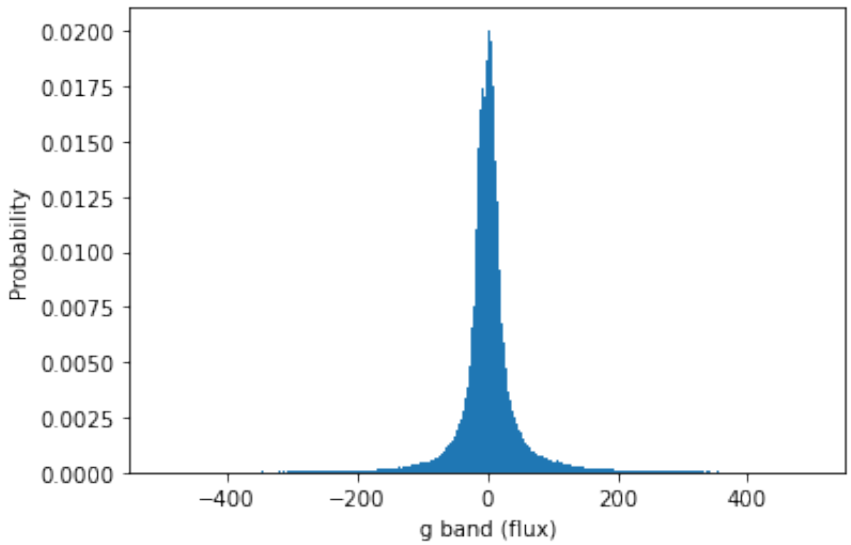
In [58]:
```
u_list = data['u'].dropna()
n, bins, patches = plt.hist(u_list,bins=1000, range=(-500,500), den
sity=1)
plt.ylabel('Probability')
plt.xlabel('u band (flux)')
plt.show()
```



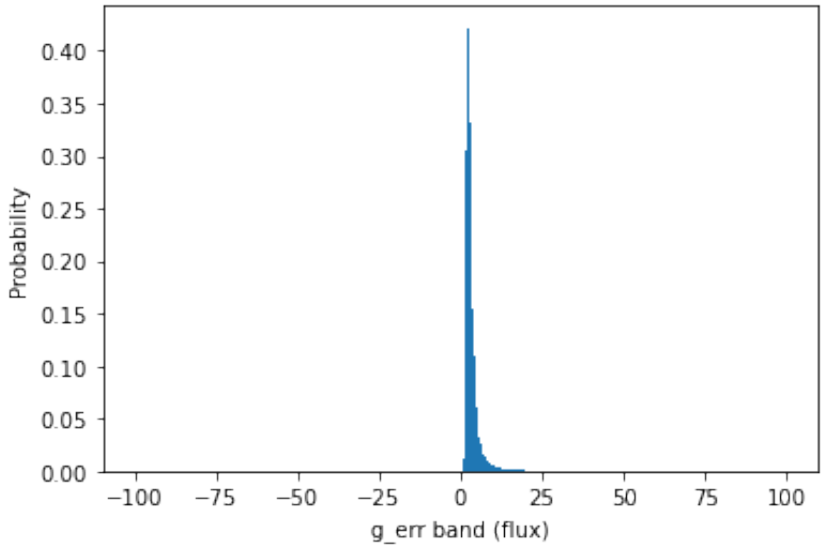
In [61]:
```
ue_list = data['u_err'].dropna()
plt.hist(ue_list,bins=500, range=(-100,100), density=1)
plt.xlabel('u_err band (flux)')
plt.ylabel('Probability')
plt.show()
```



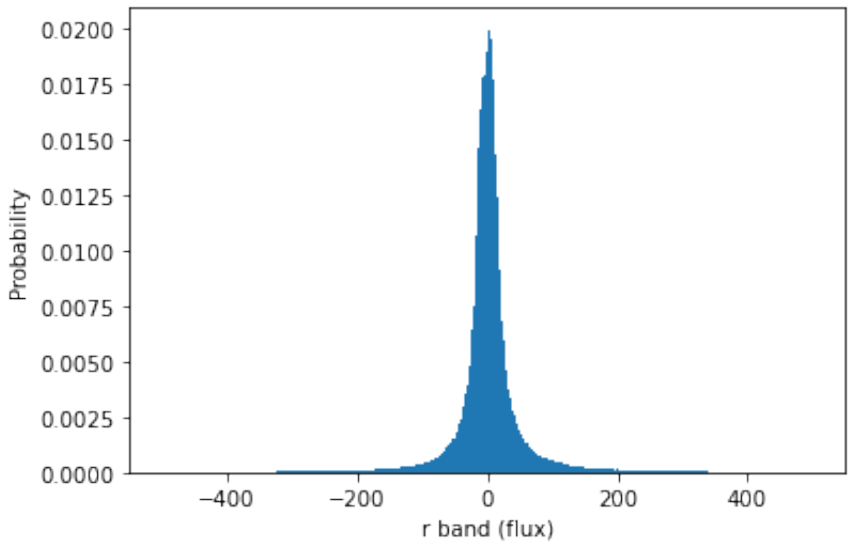
In [62]:
```
g_list = data['g'].dropna()
plt.hist(g_list,bins=500,range=(-500,500),density=1)
plt.xlabel('g band (flux)')
plt.ylabel('Probability')
plt.show()
```



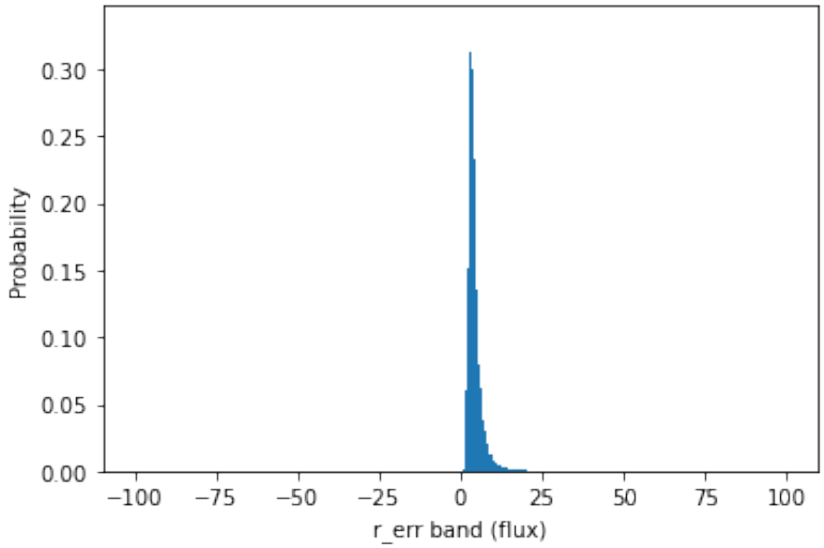
In [63]:
```
ge_list = data['g_err'].dropna()
plt.hist(ge_list,bins=500, range=(-100,100), density=1)
plt.xlabel('g_err band (flux)')
plt.ylabel('Probability')
plt.show()
```



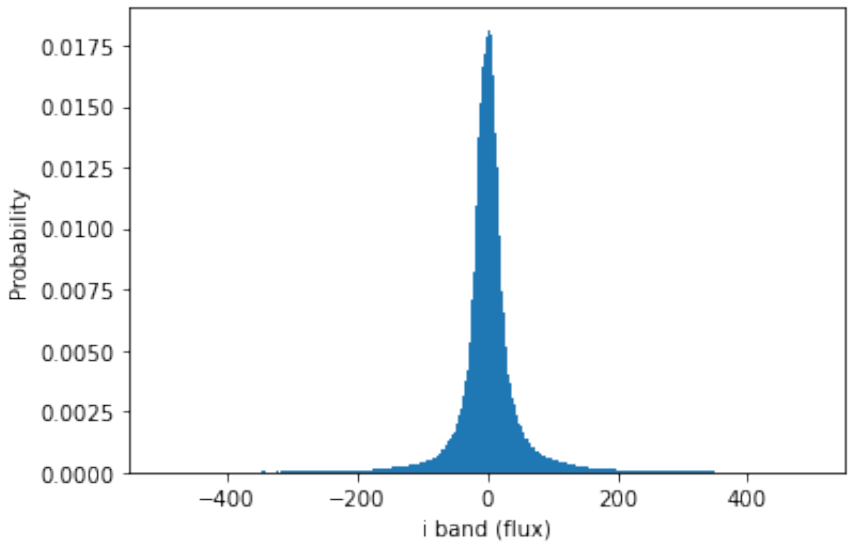
In [64]:
```
r_list = data['r'].dropna()
plt.hist(r_list,bins=500,range=(-500,500), density=1)
plt.xlabel('r band (flux)')
plt.ylabel('Probability')
plt.show()
```



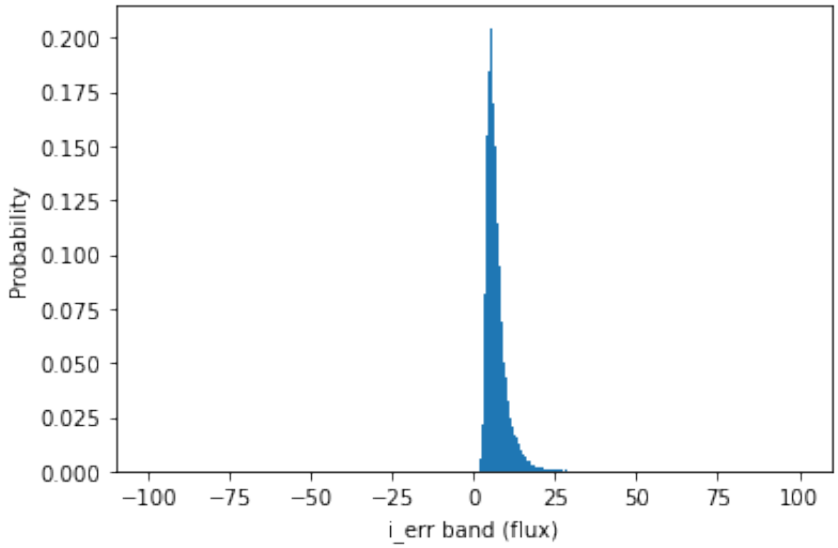
In [65]:
```
re_list = data['r_err'].dropna()
plt.hist(re_list,bins=500, range=(-100,100), density=1)
plt.xlabel('r_err band (flux)')
plt.ylabel('Probability')
plt.show()
```



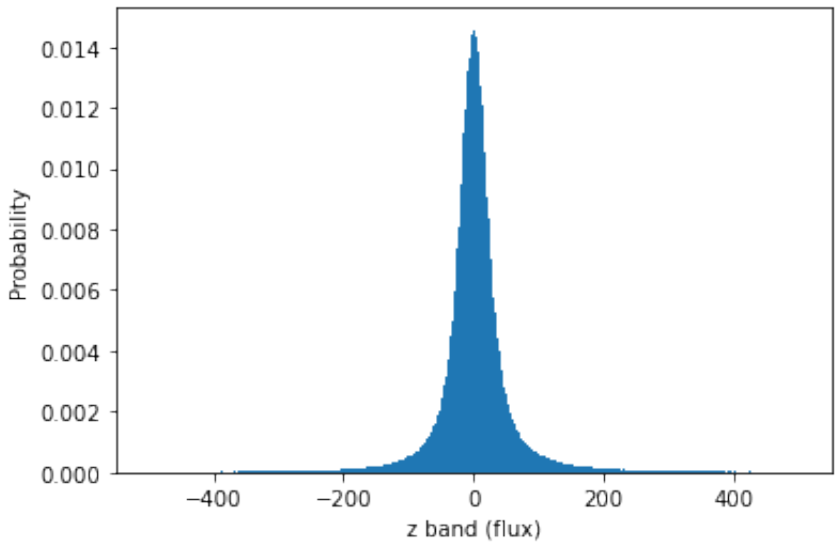
In [66]:
```
i_list = data['i'].dropna()
plt.hist(i_list,bins=500,range=(-500,500), density=1)
plt.xlabel('i band (flux)')
plt.ylabel('Probability')
plt.show()
```



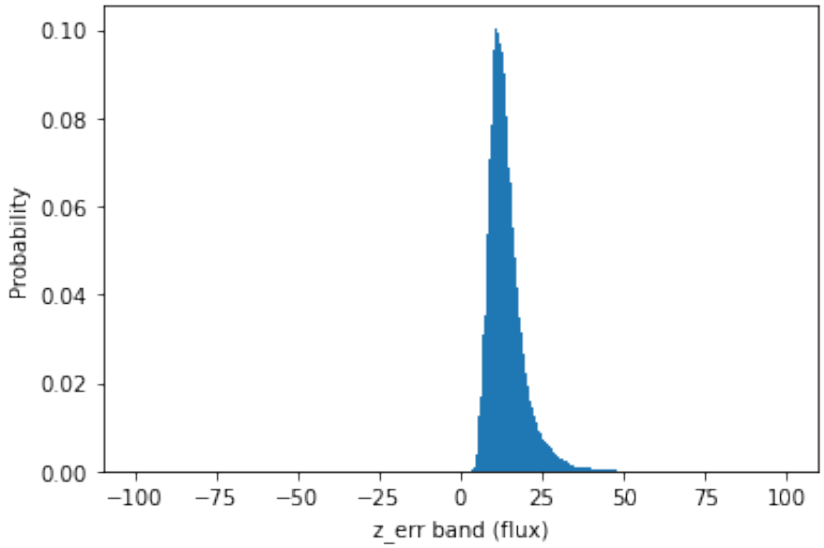
In [67]:
```
ie_list = data['i_err'].dropna()
plt.hist(ie_list,bins=500, range=(-100,100), density=1)
plt.xlabel('i_err band (flux)')
plt.ylabel('Probability')
plt.show()
```



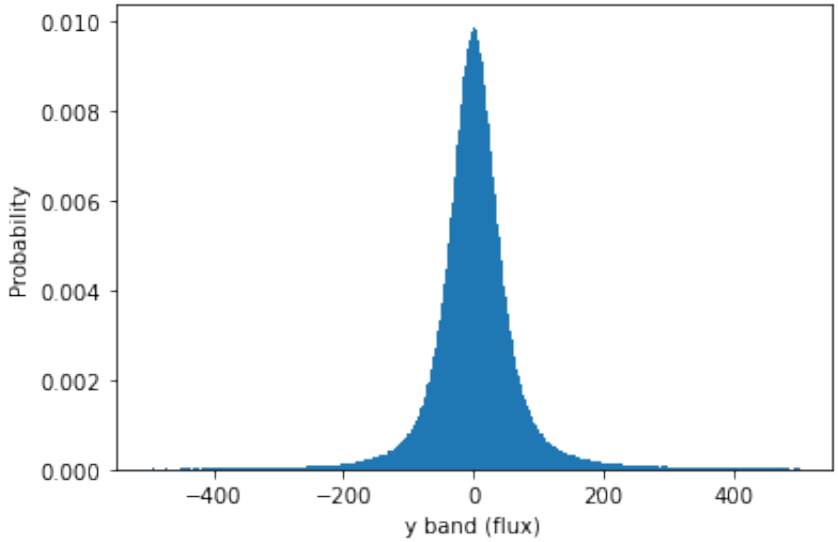
In [68]:
```
z_list = data['z'].dropna()
plt.hist(z_list,bins=500,range=(-500,500), density=1)
plt.xlabel('z band (flux)')
plt.ylabel('Probability')
plt.show()
```



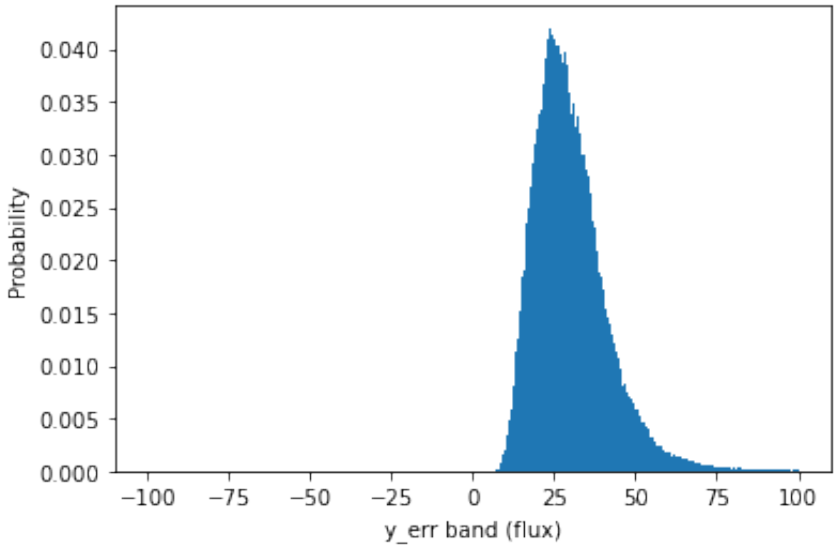
In [69]:
```
ze_list = data['z_err'].dropna()
plt.hist(ze_list,bins=500, range=(-100,100), density=1)
plt.xlabel('z_err band (flux)')
plt.ylabel('Probability')
plt.show()
```



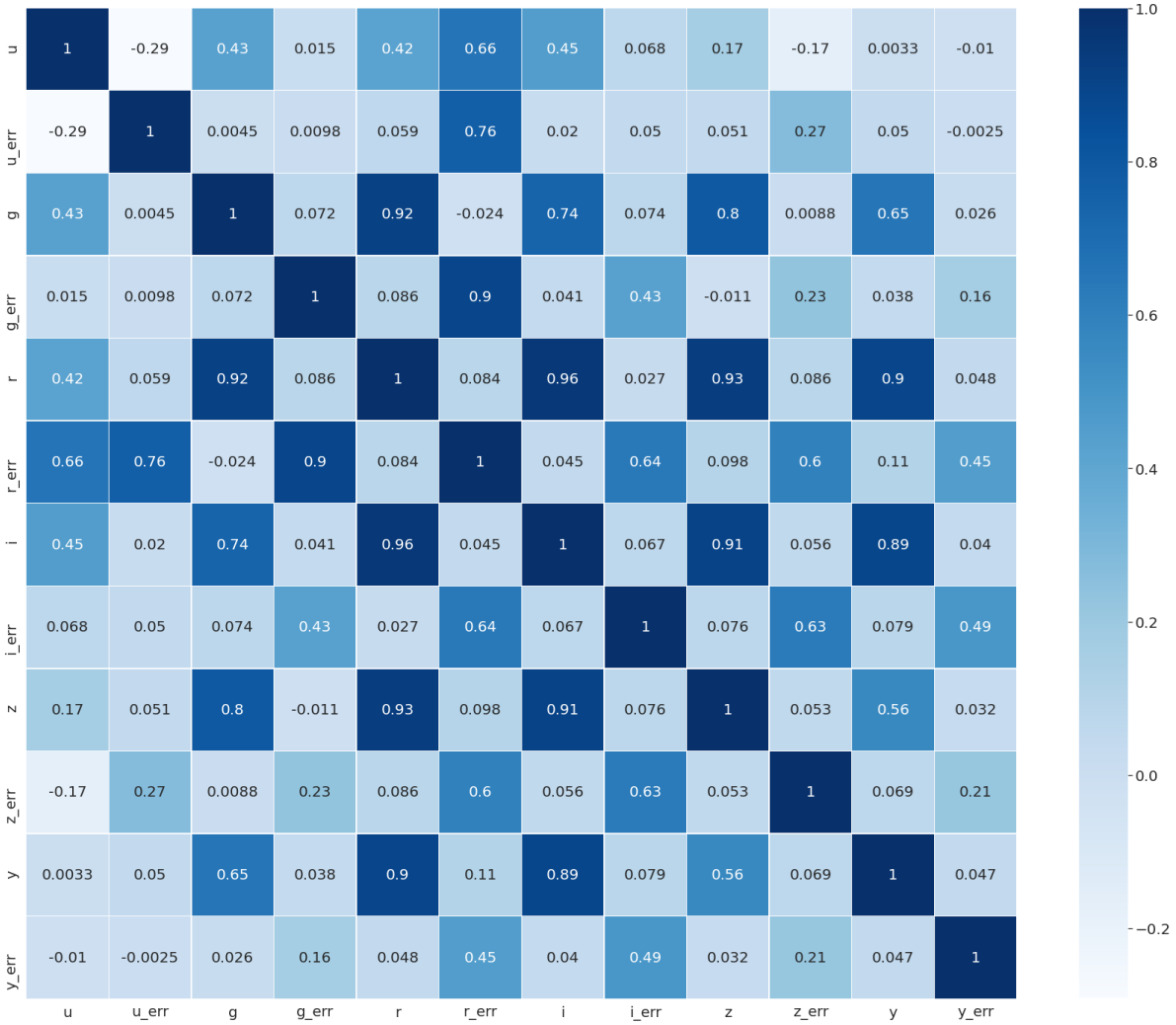
In [70]:
```
y_list = data['y'].dropna()
plt.hist(y_list,bins=500,range=(-500,500), density=1)
plt.xlabel('y band (flux)')
plt.ylabel('Probability')
plt.show()
```



In [3]:
```
ye_list = data['y_err'].dropna()
plt.hist(ye_list,bins=500, range=(-100,100), density=1)
plt.xlabel('y_err band (flux)')
plt.ylabel('Probability')
plt.show()
```



In [5]:
```
import seaborn as sns
internal_chars = ['u', 'u_err','g','g_err', 'r', 'r_err', 'i', 'i_e
rr','z','z_err','y','y_err']
# pre_data = data[internal_chars]
corrmat = data[internal_chars].corr()
f, ax = plt.subplots(figsize=(30, 30))
sns.set(font_scale=1.8)
sns.heatmap(corrmat, square=True, linewidths=.5, annot=True,annot_k
ws={"size": 20},cmap="Blues",cbar_kws={"shrink": 0.82})
plt.savefig('correlation_colour.png')
plt.show()
```


u 1 -0.29 0.43 0.015 0.42 0.66 0.45 0.068 0.17 -0.17 0.0033 -0.01
u_err -0.29 1 0.0045 0.0098 0.059 0.76 0.02 0.05 0.051 0.27 0.05 -0.0025
g 0.43 0.0045 1 0.072 0.92 -0.024 0.74 0.074 0.8 0.0088 0.65 0.026
g_err 0.015 0.0098 0.072 1 0.086 0.9 0.041 0.43 -0.011 0.23 0.038 0.16
r 0.42 0.059 0.92 0.086 1 0.084 0.96 0.027 0.93 0.086 0.9 0.048
r_err 0.66 0.76 -0.024 0.9 0.084 1 0.045 0.64 0.098 0.6 0.11 0.45
i 0.45 0.02 0.74 0.041 0.96 0.045 1 0.067 0.91 0.056 0.89 0.04
i_err 0.068 0.05 0.074 0.43 0.027 0.64 0.067 1 0.076 0.63 0.079 0.49
z 0.17 0.051 0.8 -0.011 0.93 0.098 0.91 0.076 1 0.053 0.56 0.032
z_err -0.17 0.27 0.0088 0.23 0.086 0.6 0.056 0.63 0.053 1 0.069 0.21
y 0.0033 0.05 0.65 0.038 0.9 0.11 0.89 0.079 0.56 0.069 1 0.047
y_err -0.01 -0.0025 0.026 0.16 0.048 0.45 0.04 0.49 0.032 0.21 0.047 1
u u_err g g_err r r_err i i_err z z_err y y_err
1.0 0.8 0.6 0.4 0.2 0.0 −0.2

In [ ]: