Answers of BLG560E - Midterm 1

Question 1

(25 pts) Assume that you would like to learn about mean GPA of graduate students. You form a sample of 5 graduate students whose GPA's are 3.0, 2.7, 3.2, 3.7, and 2.9.

(a) Find the 95\% confidence interval for mean GPA of graduate students.

In [20]:
```python
import numpy as np
from scipy.stats import t

confidence = 0.95
sample = np.array([3, 2.7, 3.2, 3.7, 2.9])
N = sample.shape[0]

x_bar = sample.mean()
s = sample.std(ddof=1)

print(f"Sample size: {N} sample mean: {x_bar:.2f} sample std dev: {s:.2f}")

alpha = 1-confidence
val = t.ppf(1-alpha/2, N-1)*s/np.sqrt(N)
print(f"t_(N-1)_alpha/2: {t.ppf(1-alpha/2, N-1):.2f}  val: {val:.2f}")

L = x_bar - val
H = x_bar + val

print(f"{confidence*100}% confidence interval: [{L:.2f},{H:.2f}]")
```

```
Sample size: 5 sample mean: 3.10 sample std dev: 0.38
t_(N-1)_alpha/2: 2.78  val: 0.47
95.0% confidence interval: [2.63,3.57]
```

(b) Find the lower limit for mean GPA of graduate students with 90\% confidence.

In [21]:
```python
confidence = 0.9
alpha = 1-confidence

val = t.ppf(1-alpha, N-1)*s/np.sqrt(N)
print(f"t_(N-1)_alpha: {t.ppf(1-alpha, N-1):.2f}  val: {val:.2f}")

L = x_bar - val
print(f"{confidence*100}% confidence level lower limit: {L:.2f}")
```

```
t_(N-1)_alpha: 1.53  val: 0.26
90.0% confidence level lower limit: 2.84
```

Question 2

(25 pts) Assume that you trained a novel binary object classifier model. The model is tested on 64 images and 45 images are accurately classified.

(a) (5 pts) Find the maximum likelihood estimate of the classifier accuracy ratio.

In [22]:
```python
p_mle = 45/64

print(f"p_mle: {p_mle:.2f}")
```

```
p_mle: 0.70
```

(b) (10 pts) Find the 80% confidence interval of the classifier accuracy ratio.

As p_mle is not close to 0 or 1, and both Np_mle and $N(1-p_mle)$ are larger than 5, we can use normal approximation.

In [24]:
```python
from scipy.stats import norm

confidence = 0.8
alpha = 1 - confidence

z_half_alpha = norm.ppf(1-alpha*0.5)
val = z_half_alpha * np.sqrt(p_mle*(1-p_mle)/N)
print(f"z_half_alpha: {z_half_alpha:.2f}  val: {val:.2f}")

L = p_mle - val
H = p_mle + val

print(f"{confidence*100}% confidence interval: [{L:.2f},{H:.2f}]")
```

```
z_half_alpha: 1.28  val: 0.26
80.0% confidence interval: [0.44,0.96]
```

(c) (10 pts) Find the 70% confidence value of the upper limit for classifier accuracy ratio.

In [29]:
```python
confidence = 0.7
alpha = 1 - confidence

z_alpha = norm.ppf(1-alpha)
val = z_alpha * np.sqrt(p_mle*(1-p_mle)/N)

print(f"z_alpha: {z_alpha:.2f}  val: {val:.2f}")

H = p_mle + val

print(f"{confidence*100}% confidence level upper limit: {H:.2f}")
```

```
z_alpha: 0.52  val: 0.11
70.0% confidence level upper limit: 0.81
```

Question 3

(25pts) Consider a continuous uniform distribution whose probability distribution function is given below:

$$f(x;a) = \begin{cases} 0 & \text{if } x < 10-2a \\ \frac{1}{3a} & \text{if } 10-2a \leq x \leq 10+a \\ 0 & \text{if } x > 10+a \end{cases}$$

Let a be a positive real number.

Let X_i be independent and identically distributed (iid) random variables drawn from $f(x;a)$. For a sample of size 6, we observed $[6,4,12,10,8,14]$ for X_i.

(a) (10 pts) Write the likelihood function with respect to a

In [37]:
```
def likelihood(val, a):
    prob = 1
    for x in val:
        if 10-2*a <= x <= 10+a:
            prob *= 1/(3*a)
        else:
            prob *= 0
            return prob
    return prob

val = [6, 4, 12, 10,8, 14]
for a in range(10):
    likelihood_val = likelihood(val, a)
    print(f"a: {a} likelihood: {likelihood_val:}")
```

```
a: 0 likelihood: 0
a: 1 likelihood: 0
a: 2 likelihood: 0.0
a: 3 likelihood: 0.0
a: 4 likelihood: 3.348979766680384e-07
a: 5 likelihood: 8.779149519890261e-08
a: 6 likelihood: 2.9401194111858127e-08
a: 7 likelihood: 1.1659615572447303e-08
a: 8 likelihood: 5.232780885631e-09
a: 9 likelihood: 2.5811747917131966e-09
```

Hence;

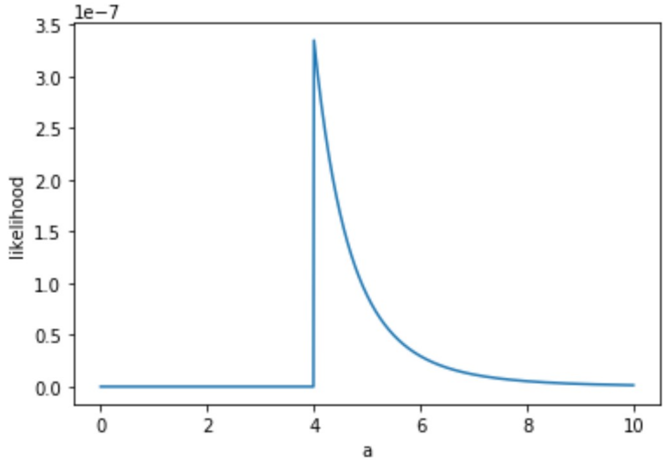
L(a) = 1/(3a)^6 if a>=4, 0 otherwise

(b) (10 pts) Draw the likelihood function with respect to a. Label horizontal and vertical axes and write important values on the related axes.

In [44]:
```python
import matplotlib.pyplot as plt

likelihood_list = []
delta = 0.01
a_range = np.arange(0, 10, delta)
for a in a_range:
    likelihood_val = likelihood(val, a)
    likelihood_list.append( likelihood_val)

plt.plot(a_range, likelihood_list)
plt.xlabel('a')
plt.ylabel('likelihood');
```



(c) (5 pts) Find the maximum likelihood estimation of parameter a.

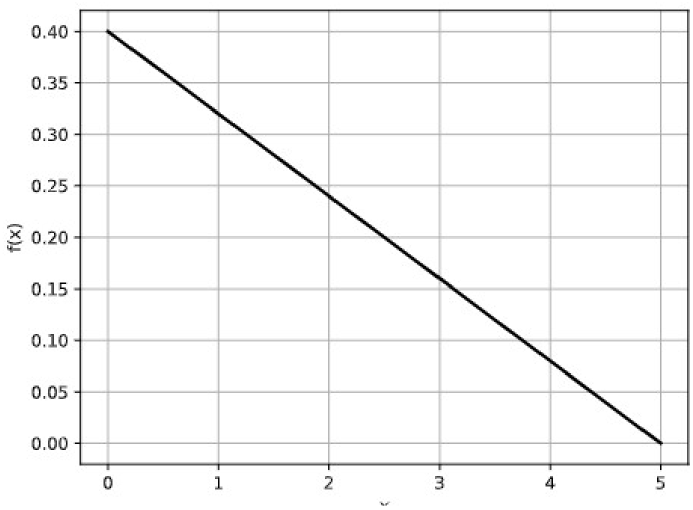
In [47]:
```python
a_mle = np.argmax(likelihood_list)*delta

print(f"MLE estimation of a is a_mle = {a_mle}")
```

```
MLE estimation of a is a_mle = 4.0
```

Question 4

[a] Consider a random variable X whose probability distribution function (pdf) is given below.



$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.4 - 0.08x & \text{if } 0 \leq x \leq 5 \\ 0 & \text{if } x > 5 \end{cases}$$

Assume following sample is formed from this distribution: $x = [1.2, 1.5, 2.8, 4.0]$. Transform these values to another random variable U that has uniform distribution in [0, 1].

(i) (5 pts) Write transformation formula

$$u = F_X(x) = 0.4x - 0.04x^2$$

(ii) (6 pts) Fill the following table.

In [50]:
```
val = [1.2, 1.5, 2.8, 4.0]

print(list(map(lambda x: 0.4*x-0.04*x**2, val)))
```

```
[0.4224, 0.5100000000000001, 0.8063999999999999, 0.9600000000000001]
```

(b) Consider random variables that comes from uniform distribution between [0, 1]: $u = [0.2, 0.5, 0.7]$. Transform these values into random variable Z that has standard normal distribution.

(i) (5 pts) Write transformation formula

$$z = \Phi^{-1}(u)$$

(ii) (9 pts) Fill the following table.

In [51]:
```
from scipy.stats import norm

u = [0.2, 0.5, 0.7]

print(list(map(lambda u: norm.ppf(u), u)))
```

```
[-0.8416212335729142, 0.0, 0.5244005127080407]
```