

Making predictions and decisions with medical trial data

We have data for 6 studies that have been done on the effect of *specific allergen immunotherapy* (SIT) on eczema and the following success rates have been observed. In each of the trials, the investigator rated whether each patient's condition improved or not.

This data set is from the [Cochrane Database of Systematic Reviews](#) article cited below, available for free [here](#). The Cochrane Database is a great resource for high quality data on all sorts of medical trials.

Tam H., Calderon M.A., Manikam L., Nankervis H., García Núñez I., Williams H.C., Durham S., Boyle R.J. (2016). Specific allergen immunotherapy for the treatment of atopic eczema. *Cochrane Database of Systematic Reviews, Issue 2*. Art. No.: CD008774. DOI: 10.1002/14651858.CD008774.pub2.

Study	Improved	Not improved
Di Rienzo 2014	20	3
Galli 1994	10	6
Kaufman 1974	13	3
Qin 2014	35	10
Sanchez 2012	22	9
Silny 2006	7	3
Totals	107	34

Task 1: Modeling

Build a statistical model for this data set. As your data, use the total number of patients improved (107) and not improved (34).

Steps in the modeling process:

1. Choose and motivate for an appropriate likelihood function for this type of data.
 1. Make sure the support of your likelihood function matches the type of data we are working with.
 - Binomial
2. Describe the unobserved parameter(s) of your model, and describe how it/they can be interpreted.

- Success rate

2. Use a conjugate prior distribution for your likelihood function.

1. Select appropriate values for the prior hyperparameters and motivate your choice.

- Beta ($a=1$, $b=1$) - uniform distribution because we don't know anything about the treatment

2. Visualize your prior distribution.

3. Compute and visualize the posterior distribution over the unobserved parameter(s) of your model. Also describe what the posterior tells you about the parameter(s).

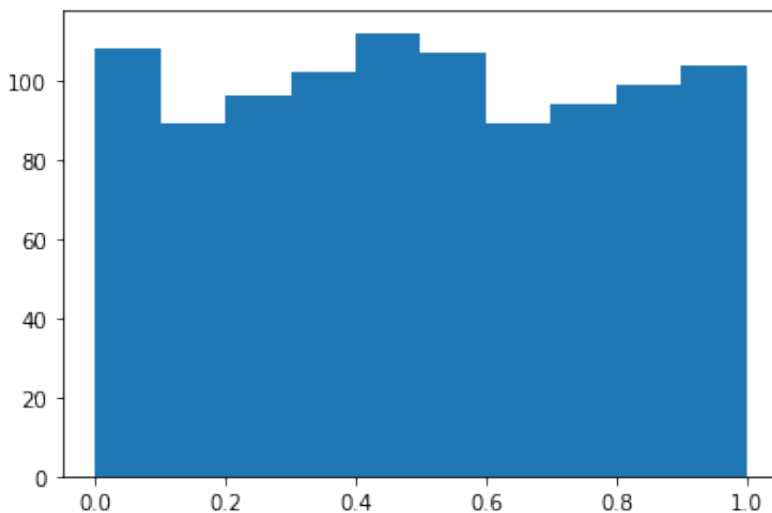
Feel free to discuss this task with other students in the course, or attempt it on your own if you prefer.

Discussing your modeling ideas with other students is a useful way to get feedback on your assumptions.

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats as sts
```

```
alpha_0 = 1
beta_0 = 1
prior_distr = sts.beta(a=alpha_0, b=beta_0)
```

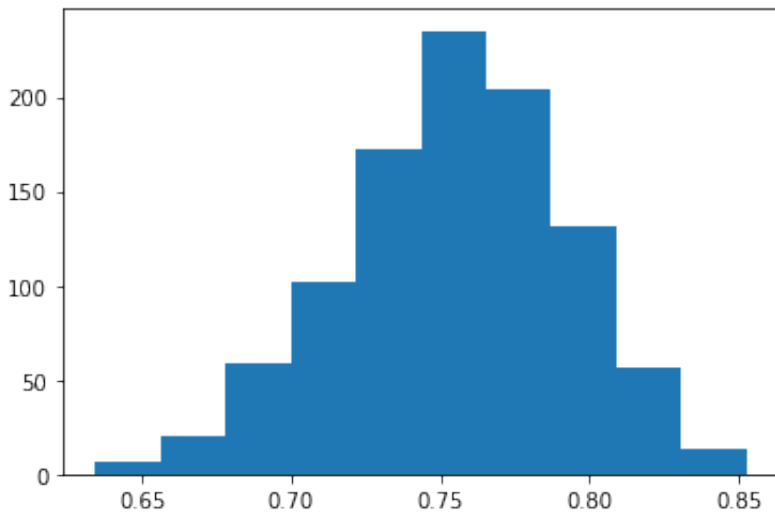
```
plt.hist(prior_distr.rvs(size=1000))
plt.show()
```



```
improved = [20, 10, 13, 35, 22, 7]
patients = [23, 16, 16, 45, 31, 10]
```

```
alpha_1 = alpha_0 + sum(improved)
beta_1 = beta_0 + sum(patients) - sum(improved)
post_distr = sts.beta(a=alpha_1, b=beta_1)
```

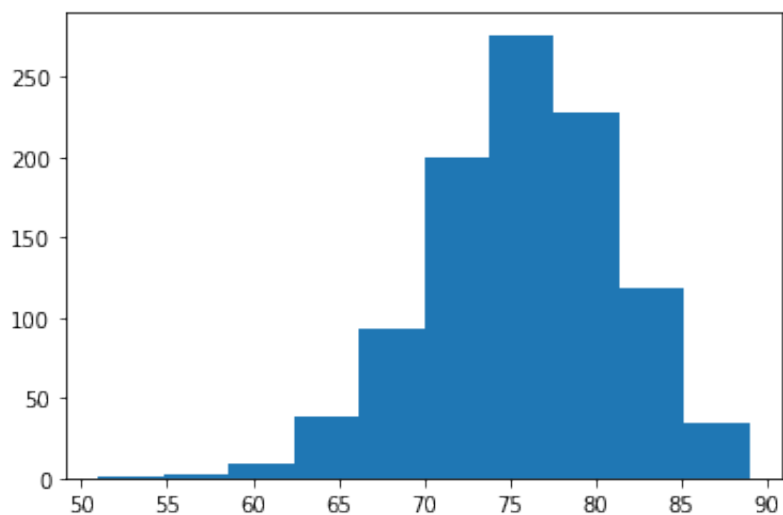
```
plt.hist(post_distr.rvs(size=1000))  
plt.show()
```



Task 2: Questions, predictions, and decisions

1. Posterior predictive distribution: Use your model to predict how many patients would improve if we treated 100 new eczema patients using SIT. Express your answer as a probability distribution over the number of patients improved.
2. Use your posterior to answer these questions:
 1. What is the probability that at least two thirds of eczema patients get better with SIT?
 2. What is the probability that at least 75% of eczema patients get better with SIT?
3. Use your model to decide whether the treatment works or not. Motivate your answer.

```
num_trials = 100  
p = post_distr.rvs(size=1000)  
binom_distr = sts.binom(p=p, n=num_trials)  
  
post_pred_samples = binom_distr.rvs(1000)  
plt.hist(post_pred_samples)  
plt.show()
```



```
# probability that >2/3 patients get better  
num_success_a = 2/3 * num_trials  
print(sum(post_pred_samples > num_success_a) / len(post_pred_samples))
```

```
0.95
```

```
# probability that >75% of patients get better  
num_success_b = 0.75 * num_trials  
print(sum(post_pred_samples > num_success_b) / len(post_pred_samples))
```

```
0.511
```

To decide whether the treatment works, we would need to compare this to the control group's results.
Maybe eczema can cure itself over time.