Homework 2

Solutions

1. (Geometric random variable) By definition of conditional probability, we have

$$\mathrm{P}(\tilde{a} = a \,|\, \tilde{a} > 5) = \frac{\mathrm{P}(\tilde{a} = a, \tilde{a} > a)}{\mathrm{P}(\tilde{a} > a)}. \tag{1}$$

The numerator is zero unless $a > 5$, in that case

$$\mathrm{P}(\tilde{a} = a \,|\, \tilde{a} > 5) = \frac{(1-\alpha)^{a-1}\alpha}{\sum_{b=6}^{\infty}(1-\alpha)^{b-1}\alpha}. \tag{2}$$

The denominator contains the geometric sum

$$\sum_{b=6}^{\infty}(1-\alpha)^{b-1} = \frac{(1-\alpha)^6}{\alpha}. \tag{3}$$

We conclude

$$\mathrm{P}(\tilde{a} = a \,|\, \tilde{a} > 5) = (1-\alpha)^{a-6}\alpha. \tag{4}$$

This is the same as the geometric pmf if we plug in $a - 5$. In terms of the coin example, if we have obtained 5 tails, now the probability that we have to flip $b = a - 5$ more times is distributed like a geometric random variable. This makes sense because the flips are independent, so there is no difference between this and just considering another sequence of flips starting from the beginning.

2. (Chess games)

   (a) Under the independence assumption, we have

$$\mathcal{L}_X(\theta) = \theta^4\alpha^2(1-\theta-\alpha)^4, \tag{5}$$

$$\log \mathcal{L}_X(\theta) = 4\log\theta + 2\log\alpha + 4\log(1-\theta-\alpha). \tag{6}$$

   The plot is shown in Figure 1.

   (b) From the plot we can see that the function has a single maximum. To find it, we set the partial derivatives to zero. We have

$$\frac{\mathrm{d}\log \mathcal{L}_X(\theta)}{\mathrm{d}\theta} = \frac{4}{\theta} - \frac{4}{1-\theta-\alpha}, \tag{7}$$

$$\frac{\mathrm{d}\log \mathcal{L}_X(\alpha)}{\mathrm{d}\theta} = \frac{2}{\alpha} - \frac{4}{1-\theta-\alpha}. \tag{8}$$

$$\tag{9}$$

   Setting the first expression equal to zero yields $\alpha_{\mathrm{ML}} = 1 - 2\theta_{\mathrm{ML}}$. Plugging into the second and solving the equation, we conclude $\theta_{\mathrm{ML}} = 0.4$ and $\alpha_{\mathrm{ML}} = 0.2$.

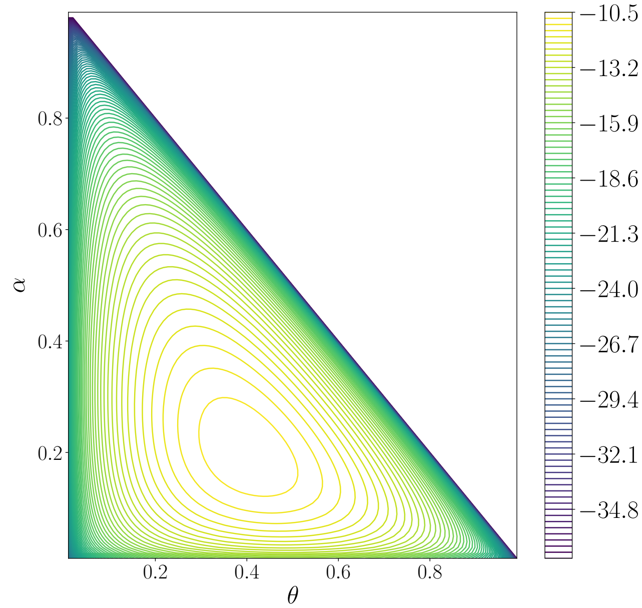


Figure 1: Log-likelihood for Problem 2.

(c) The empirical pmf would assign $4/10 = 0.4$ to the probability of Garry winning, $2/10 = 0.2$ to the probability of Anish winning, and $4/10 = 0.4$ to the probability of a draw. This is exactly equivalent to the parametric model.

3. (Darts) Let $\tilde{a}$ denote the random variable. Note that the last attempt must always be a success. We can therefore decompose the event *a attempts required* into the intersection of *$k-1$ successes over first $a-1$ attempts* and *ath attempt is a success*. Since the attempts are all independent, we have

$$p_{\tilde{a}}(a) = \mathrm{P}(k-1 \text{ successes over first } a-1 \text{ attempts})\mathrm{P}(a\text{th attempt is a success}). \tag{10}$$

By exactly the same reasoning we used to derive the binomial distribution, we have

$$\mathrm{P}(k-1 \text{ successes over first } a-1 \text{ attempts}) = \binom{a-1}{k-1}\theta^{k-1}(1-\theta)^{a-1-(k-1)}, \tag{11}$$

as long as $a \geq k$. We conclude

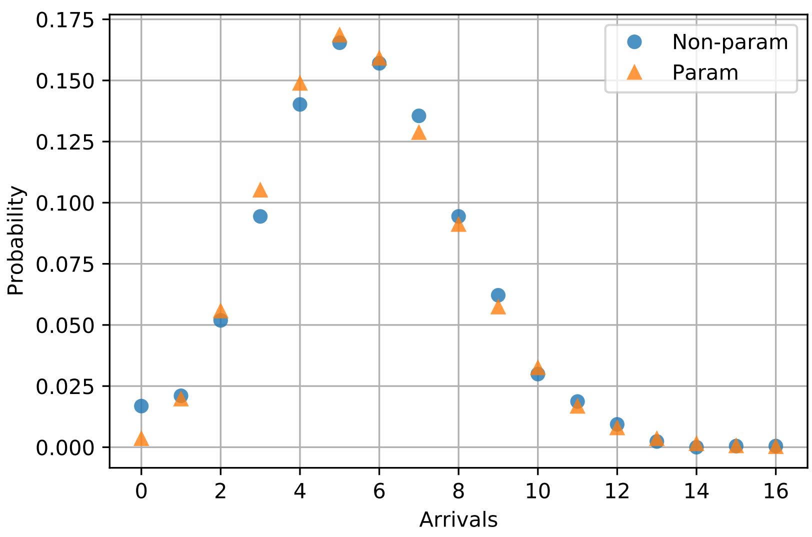
$$p_{\tilde{a}}(a) = \binom{a-1}{k-1}\theta^{k}(1-\theta)^{a-k}, \quad \text{for } a \geq k, \tag{12}$$

and zero otherwise.

4. (Air Traffic)

(a) We assume that flights arrive with a constant rate λ during the evening peak. We model the numbers of arrival with a Poisson distribution.

```
train_data = pd.read_csv("./train.csv", header=0, names=['time', 'arrivals']).set_
test_data = pd.read_csv("./test.csv", header=0, names=['time', 'arrivals']).set_in

x_range = np.arange(train_data.max()+1)

# non-parametric model
def non_param_model(data):
    pmf = np.zeros(data.max()+1)
    counts = data.value_counts()
    emprical_prob = counts/counts.sum()
    for i in emprical_prob.index:
        pmf[i] = emprical_prob[i]
    return pmf

pmf_nonparam = non_param_model(train_data)

# parametric model
pmf_param = poisson.pmf(x_range, train_data.mean())

plt.plot(x_range, pmf_nonparam, 'o', label="Non-param", alpha=0.8)
plt.plot(x_range, pmf_param, '^', label="Param", alpha=0.8)
plt.legend()
plt.ylabel('Probability')
plt.xlabel('Arrivals')
plt.grid("on")
```



(b) The test RMSE is 0.016 for non-parametric model and 0.021 for parametric model.