University of Bern
Daniil Krichevskiy (Master student in Theoretical Physics)
(daniil.krichevskiy@students.unibe.ch)

Advanced statistical methods for physicists

Solutions for Exercises Sheet 2

---

1. **Exercise 1. The Werewolves of Millers Hollow**

   Millers Hollow (pop. 842) is haunted by werewolves killing villagers at night. It is estimated that they currently represent 10% of the entire population. As an established statistician, the Council asks for your help in order to devise strategies to identify werewolves and bring Millers Hollow back to peace and tranquility. The Council worked relentlessly in the past year to build a werewolf detector, which samples a tiny bit of blood from each individual submitted to the test. They are very proud of it as this detector has an accuracy of 90% in detecting werewolves from gentle villagers. The Council claims that they would then be able to imprison 90% of the individuals who failed the test. Demonstrate that the Council's approach is too optimistic. If one villager fails the test, what is the probability that they are a werewolf?

   Let's define the following events: H_1 - the tested person is a villager, H_2 - the tested person is a werewolf. These are disjoint events ($H_i \cap H_j = \oslash,\ \forall\ i \neq j,\ i, j = 1, 2$) and form the full group of events Ω ($\sum_i H_i = \Omega$). Thus they are hypotheses. Let's denote the event "test is positive" by A.

   We need to calculate the conditional (posterior) probability of H_2 given that A is true $P(H_2 \mid A)$. For this we use the Bayes' theorem:

   $$P(H_2 \mid A) = \frac{P(A \mid H_2)P(H_2)}{P(A)} = \frac{P(A \mid H_2)P(H_2)}{P(A \mid H_2)P(H_2) + P(A \mid H_1)P(H_1)}, \tag{1}$$

   where $P(A \mid H_2) = 0.9$ - conditional probability of A given that H_2 is true, i.e., test accuracy, then $P(A \mid H_1) = 1 - P(A \mid H_2) = 0.1$, $P(H_1) = 0.9$ - share of (healthy) villagers in populations, $P(H_2) = 0.1$ - share of werewolfs in population of the village (prior probabilities). After substituting the numbers in the formula (1) we get

   $$P(H_2 \mid A) = 0.5. \tag{2}$$

   That means that the half of positively tested people get the wrong result thus the village Council can't imprison 90% of positively tested people (if they don't want to imprison many innocent people).

2. **Exercise 2. A new bag of marbles**

   Refer to the example given in the course if needed. It is not even Christmas yet but you were offered a shiny bag that contains a number of marbles. As in the lecture's example, the marbles can be either black or white. This bag is from another manufacturer than the one in the course so you have no prior knowledge about its contents. You shake the bag and draw the marbles 9 times with replacement. You obtain 3 white marbles and 6 black marbles.

   - What is the posterior probability that there are less than 20% of black marbles in the bag?
   - What is the posterior probability that there are more than 80% of black marbles in the bag?
   - What is the posterior probability of having between 20 and 80% of black marbles in the bag?

- What fraction of black marbles does the 25% of the posterior probability correspond to?
- What fraction of black marbles does the 75% of the posterior probability correspond to?

a Let's consider the number of black marbles drawn from the bag. It is a random variable which consists of the number of s successes in n Bernoulli trials with unknown probability of success q in $[0,1]$ (in our case $s=6$, $n=9$) with probability mass function

$$p(s) = \binom{n}{s} q^s (1-q)^{n-s}, \quad \binom{n}{s} = \frac{n!}{s!(n-s)!}. \tag{3}$$

In fact $q = N_b/N$, where N is a total number of marbles in the bag and N_b is a number of black marbles. We need to study the posterior probability $p(q = x \mid s, f)$. This is done using the conjugate prior distribution which is beta distribution with parameters α, β:

$$p(q) = \frac{q^{\alpha-1}(1-q)^{\beta-1}}{B(\alpha,\beta)}, \tag{4}$$

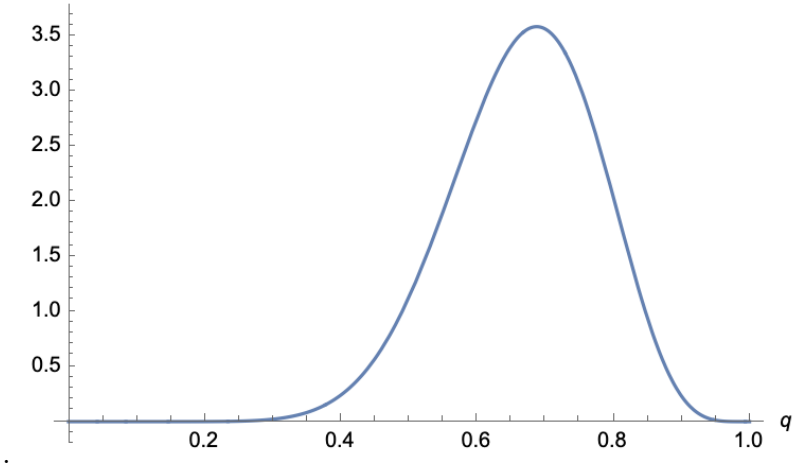
where $B(\alpha,\beta)$ - beta function ($\alpha = \beta = 1$ give a uniform distribution). Then the posterior probability $p(q = x \mid s, f)$ (f-number of successes) reads as [1]

$$\begin{aligned} p(q = x \mid s, f) &= \frac{p(s, f \mid x)p(x)}{\int dy\; p(s, f \mid y)p(y)} = \\ &= \frac{\binom{s+f}{s} x^{s+\alpha-1}(1-x)^{f+\beta-1}/B(\alpha,\beta)}{\int_{y=0}^{1} dy\; \binom{s+f}{s} y^{s+\alpha-1}(1-y)^{f+\beta-1}/B(\alpha,\beta)} = \\ &= \frac{x^{s+\alpha-1}(1-x)^{f+\beta-1}}{B(s+\alpha, f+\beta)}, \quad x \in [0,1]\,. \end{aligned} \tag{5}$$

The mean of beta distribution is $\frac{\alpha}{\alpha+\beta}$ which corresponds to α successes and β failures so we will set $\alpha = 6$ and $\beta = 3$ (a priori guess of the ratio of black and white marbles in the bag based on the draw result). Substituting our values we get

$$p(q = x \mid 6, 3) = 74256(1-x)^5 x^{11}. \tag{6}$$

Figure 1: Distribution function (6)



.

b To get the result for posterior probability of the black marbles' share being in some range we integrate (6) within the given boundaries:

$$\begin{aligned} P(q < 20\%) &= \int_0^{0.2} dx\ p(q = x \mid 6, 3) \approx 9.2 \times 10^{-6}, \\ P(q > 80\%) &= \int_{0.8}^{1} dx\ p(q = x \mid 6, 3) \approx 0.106, \\ P(20\% < q < 80\%) &= \int_{0.2}^{0.8} dx\ p(q = x \mid 6, 3) \approx 0.89. \end{aligned} \tag{7}$$

If we took a uniform distribution $\alpha = \beta = 1$ than we would get

$$\begin{aligned} P(q < 20\%) &\approx 0.00086, \\ P(q > 80\%) &\approx 0.121, \\ P(20\% < q < 80\%) &\approx 0.878. \end{aligned} \tag{8}$$

c To get the fraction of black marbles q_0 which corresponds to some given probability p_0 we integrate (6) to get the cumulative distribution function

$$P(q_0) = \int_0^{q_0} dx\ p(q = x \mid 6, 3) = x^1 2(6188 - 28560x + 53040x^2 - 49504x^3 + 23205x^4 - 4368x^5)\ |_0^{q_0} \tag{9}$$

and then numerically solve $P(q_0) = p_0$ (for $q_0 \in [0, 1]$). Then for $p_0 = 0.25$ we get $q_0 \approx 0.59$ and for $p_0 = 0.75$ we get $q_0 \approx 0.0.75$ Again, if we considered a uniform distribution $\alpha = \beta = 1$ than we would get for $p_0 = 0.25$ $q_0 \approx 0.54$ and for $p_0 = 0.75$ $q_0 \approx 0.0.74$

3. **Exercise 3. Visit to the marbles factory** Given your interest in bags of marbles, the factory invites you to spend a day at their premises. Upon greeting by the manager, they ask you to help them with a production issue. They manufactured several marbles for days until they noticed that the machine had a pigmentation issue, producing bags entirely composed of white marbles, which led to angry phone calls from unhappy customers. There is no way to discern the affected bags externally. To what extent would drawing a single marble from each bag improve the ability to identify affected bags? Assuming that P(normal bag — white) = 0.3, what can we say about the share of blue marbles in the unaffected bags?

a Let's note that if we draw from an affected (A) bag than the probability to get a white marble (W) is 1 $P(W, A) = 1$ (by definition of affected bag). The set of hypotheses is formed by $H_1 \equiv A$ (the bag is an affected one) and $H_2 \equiv N$ (the bag is a normal one). The by Bayes' theorem

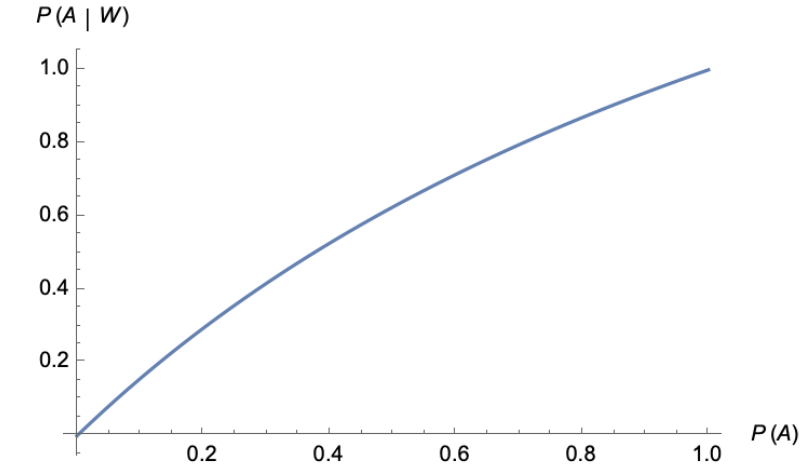
$$\begin{aligned} P(A \mid W) &= \frac{P(W \mid A)P(A)}{P(W \mid A)P(A) + P(W \mid N)P(N)} = \\ &= \frac{P(A)}{P(A) + P(W \mid N)(1 - P(A))}, \end{aligned} \tag{10}$$

where B means a blue marble. Let's suppose that there is the same number of white and blue marbles in a normal bag is approximately the same like in Exercise 2 (see Figure 1) : $P(W \mid N) \approx 0.4$. Than $P(A \mid W)$ has strong dependence on prior distribution of affected bags (see Figure 2).

b If we assume that $P(N|W) = 0.3$ than $P(B|W) = 1 - 0.3 = 0.7$. Thus from (**??**) we get the share of blue marbles in the normal bags

$$P(B|W) = 1 - P(W|N) = 1 - \frac{3}{7}\frac{P(A)}{1 - P(A)}. \tag{11}$$

Figure 2: Dependence of posterior probability $P(A \mid W)$ on $P(A)$



.

References

[1] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory.* Harvard Business School Publications. Division of Research, Graduate School of Business Adminitration, Harvard University, 1961.