# 732A66: Lab 1
# Decision Theory

Carles Sans Fuentes

October 20, 2017

---

## Assignment

Assume a person is visiting his General Practitioner (GP) for some health problem. The patient shows a symptom that from the GP:s point-of-view could be the consequence of each of three different diseases. This symptom appears with disease A1 in 25% of all cases, with disease A2 in 15% of all cases, and with disease A3 in 5% of all cases. One can further approximately assume that a person cannot have more than one of these diseases at the same time. The symptom can also appear for other reasons. When none of the three mentioned diseases are present the probability is approximately 0.5% that a person shows this symptom.

### a

Question: What diagnosis should the GP give if she uses the principles of inference to the best explanation?

Answer: Given that the best explanation is described by the likelihood (i.e.Using the likelihood as a measure of how likely is an event is a matter of inference to the best explanation), and given that:

$$P(symptoms|A_1) = 0.25$$
$$P(symptoms|A_2) = 0.15$$
$$P(symptoms|A_3) = 0.05$$
$$P(symptoms|\emptyset) = 0.005$$

The one with highest likelihood and (thus the best explanation) is $P(symptoms|A_1) = 0.25$

### b

Question: Assume now that the prevalence of the diseases A1, A2 and A3 are 0.1 %, 0.5 % and 1 % respectively. By prevalence is here meant the proportion of the relevant population (those people from which the current person belong) that has this disease at this specific point of time (point prevalence). What are the conditional probabilities of the person having respectively the diseases A1, A2 and A3? What is your opinion about diagnosis according to principles of inference to the best explanation in this case?

Answer: Following Bayes we get that the conditional distributions are:

$$P(A_1|symptoms) = \frac{(symptoms|A_1) * P(A_1)}{\int symptoms} =$$

$$\frac{(symptoms|A_1) * P(A_1)}{P(sym|A_1) * P(A_1) + P(sym|A_2) * P(A_2) + P(sym|A_3) * P(A_3) + P(sym|\emptyset) * P(\emptyset)} =$$

$$\frac{0.25 * 0.001}{(0.25 * 0.001 + 0.15 * 0.005 + 0.05 * 0.01 + (1 - 0.001 - 0.005 - 0.01) * 0.005} = 0.03894081 \approx 3.9\%$$

Analogously, we get

$$P(A_2|symptoms) = 0.1168224 \approx 11.7\%$$

$$P(A_3|symptoms) = 0.07788162 \approx 7.8\%$$

$$P(\emptyset|symptoms) = 0.7663551 \approx 76.7\%$$

Considering that, the best explanation is that the patient has nothing. Still, further evaluations should be done in order to assess the cost of doing further proofs to check if the patient in fact has some illness (e.g. cost of evaluations—cost not having any disease AND cost of evaluations— cost of having any disease) to be able to say if it is worth to give something to the patient and say it has illness $A_2$, which is the most probable illness.

## Question 2

Question: In sampling from a Bernoulli process, the posterior distribution is the same whether one samples with n fixed (binomial sampling) or with r fixed (Pascal sampling). Explain why this is true. Suppose that a statistician merely samples until he is tired and decides to go home. Would the posterior distribution still be the same (that is, is the stopping rule noninformative)?

Answer: The binomial sampling gets the number of successes (k) in a sequence of n trials (having n as fixed), whereas the Pascal sampling counts the number of events before a specified (non-random) number of failures (denoted r) occurs (being this r fixed). Both processes comes from Bernoulli trials, and for that, the posterior got is the same, being the relation between r and n the following: $n = r + k$.

Assuming for simplicity that our prior is beta ( whatever other prior will work, but to see it clear the choice of the beta given that it is its conjugate is optimal), we get that in the case of

$$PosteriorPA = p^{\alpha-1} * (1-p)^{\beta-1} p^r * (1-p)^k =$$

$$p^{\alpha-1+k}(1-p)^{\beta-1+r} \approx Beta(\alpha+k, \beta+r)$$

In the case of the binomial:

$$PosteriorBin = p^{\alpha-1} * (1-p)^{\beta-1} p^k * (1-p)^{n-k} =$$

$$p^{\alpha-1+k} * (1-p)^{\beta-1+n-k} \approx Beta(\alpha+k, \beta+n-k)$$

We can see that the distribution is the same, having only as an exchange r for n, and simulating for the same p being the probability of success and getting the same condition distribution probability of the result if the results are the same. That is why it can be said the posteriors are the same

If a statistician merely samples until he is tired and decides to go home the posterior distribution still will not be the same be the same at that point (concerning parameters) but it will converge to the same by when n or r is large. Also, the conditional distribution will be the same given the same p value in both cases for the same result.

## Question 3

Question:

A bank official is concerned about the rate at which the bank's tellers provide service for their customers. He feels that all of the tellers work at about the same speed, which is either 30, 40 or 50 customers per hour. Furthermore, 40 customers per hour is twice as likely as each of the two other values, which are assumed to be equally likely. In order to obtain more information, the official observes all five tellers for a two-hour period, noting that 380 customers are served during that period. Use this new information to revise the official's probability distribution of the rate at which the tellers provide service.

Answer: In order to answer the activity, the following calculus has been done:

We know first that our prior is the follwowing:

$$p(\lambda = 30) = 0.25$$

,

$$p(\lambda = 40) = 0.5$$

,

$$p(\lambda = 50) = 0.25$$

These ones are our prior parameters. Now we calculate the probability of the parameters given the data. This is from the formula:

$$P(\theta|Data) = \frac{p(Data|theta) * p(\theta)}{prob(Data)}$$

such that,

$$P(\lambda = 30|Data) = \frac{p(Data|\lambda = 30) * p(\lambda = 30)}{p(Data|\lambda = 30) * p(\lambda = 30) + p(Data|\lambda = 40) * p(\lambda = 40) + p(Data|\lambda = 50) * p(\lambda = 50)}$$

Using the formula from the Pois distribution

$$p(\lambda|r, t) = \frac{(\lambda * t)^r * \exp^{-\lambda * t}}{r!}$$

where r is the number of times and event occurs, t is time interval and $\lambda$ the average number of events in the interval. Thus, we have that it will be written as following as an example to calculate it manually ($\lambda = 30$, $t = 1$ and $r = 38$) :

$$p(\lambda = 30|Data) = \frac{30^38 * \exp^{-30}}{38!}$$

now I will be calling this for simplification on the steps dpois(r,$\lambda$) (as the function in R).

So, going back to the procedure, we get that the Posterior of our data is as following:

$$P(\lambda = 30|Data) = \frac{dpois(38, 30) * p(30)}{dpois(38, 30) * p(30) + dpois(38, 40) * p(40) + dpois(38, 50) * p(50)} =$$
$$\frac{0.0241 * 0.25}{0.0241 * 0.25 + 0.06137 * 0.5 + 0.0134 * 0.25} = 0.1507422$$

Analogously, I get that

$$P(\lambda = 40|Data) = \frac{dpois(38, 40) * p(40)}{dpois(38, 30) * p(30) + dpois(38, 40) * p(40) + dpois(38, 50) * p(50)} =$$
$$\frac{0.06137 * 0.5}{0.0241 * 0.25 + 0.06137 * 0.5 + 0.0134 * 0.25} = 0.7655824$$

$$P(\lambda = 50|Data) = \frac{dpois(38,50) * p(50)}{dpois(38,30) * p(30) + dpois(38,40) * p(40) + dpois(38,50) * p(50)} =$$

$$\frac{0.0134 * 0.25}{0.0241 * 0.25 + 0.06137 * 0.5 + 0.0134 * 0.25} = 0.7655824$$

## Question 4

Question: Show that the two-parameter beta distribution belongs to the exponential class of distribution, i.e. its probability density function can be written on the form of the exponential distribution:

Now, find the form of a conjugate prior distribution for the two parameters of a beta distribution. You need only to specify it up to a proportionality constant that needs not to be calculated. Using a sample of 5 observations having this beta distribution, how is the prior distribution updated to a posterior distribution?

Assume you assign a prior distribution that is as non-informative as possible, still being a proper distribution. How will the posterior distribution develop when you increase the sample size from 5 point to 100 points? From 100 points to 1000 points?

Answer:

In order to proof that the beta distribution, the following has been done:

$$Beta(x|\theta) = Beta(x|\alpha, \beta) = \frac{x^{\alpha-1} * (1-x)^{\beta-1}}{B(\alpha, \beta} =$$

$$\exp^{(\alpha-1)*ln(x)+(\beta-1)ln(1-x)-B(\alpha,\beta)}$$

, being

$$A_1 = \alpha - 1$$
$$A_2 = \beta - 1$$
$$B_1 = ln(x)$$
$$B_2 = ln(1-x)$$
$$C = 0$$
$$D = -ln(B(\alpha, \beta))$$

The prior distribution then is proportional to:

$$\exp^{\sum_{j=1}^{k} A_j(\theta)*\alpha_j + \alpha_{k+1}*D(\theta)+K(\alpha_1,\ldots,\alpha_k,\alpha_{k+1}}$$

that is as prior

$$\exp^{\alpha_1(\alpha-1)+\alpha_2(\beta-1)+\alpha_3*-log(B(\alpha,\beta))+k(\alpha_1,\alpha_2,\alpha_3)}$$

such that the posterior should be written as:

$$Posterior(\theta|Data) = \exp^{\alpha_1(\alpha-1)+\alpha_2(\beta-1)+\alpha_3*-log(B(\alpha,\beta))+k(\alpha_1,\alpha_2,\alpha_3)}$$
$$* \exp^{(\alpha-1)*\sum_{i=1}^{n} log(x_i)+(\beta-1)\sum_{i=1}^{n} log(1-x_i)-log(B(\alpha,\beta))}$$
$$\propto \exp^{(\alpha-1)(\alpha-1+\sum_{i=1}^{n} log(x_i))+(\beta-1)*(\alpha_2+\sum_{i=1}^{n} log(1-x_i))-\alpha_3*log(\alpha,\beta)}$$

being

$$\alpha posterior = \alpha * (\alpha_1 + \sum_{i=1}^{n} log(x_i))$$

$$\beta posterior = \beta * (\alpha_2 + \sum_{i=1}^{n}(1 - log(x_i))$$

The posterior distribution will be the same as the prior distribution. Thus, the only thing that will be necessary is to update the parameters in the posterior function putting there the values for $\alpha posterior$ $\beta posterior$ written above. It does not matter how many observations we have, that the distribution will still be the same. The only thing that will change will be the parameters.