Uncertainty in P-Values, Monty Hall Problem and, Bayesian and Neural Network Predictions

Priyantha Wijayatunga
Department of Statistics, Umeå University, Umeå, Sweden
Priyantha.wijayatunga@umu.se

Abstract

Probability is a better tool for handling uncertainty and reasoning. But it is often easy to go wrong with the task. Here we talk about some instances where the probability is misused and, then show how to use it correctly. Firstly, we talk about how to adjust p-values in statistical significance testing so that uncertainty in replication studies can be lowered, thus minimizing so-called replication crisis. Also, Monty Hall problem is discussed to show how puzzling the probability can be. The solution lies on differentiating the conditional probability from the marginal probability. And finally, a simple logical way of quantifying uncertainty of a probabilistic prediction is discussed. We use Bayesian network inference for the task. The method is extended to deep neural networks.

Introduction

Probability theory is regarded as a better and coherent framework for handling uncertainty in many natural phenomena and artificial systems. However, often it can be misunderstood easily. For example, in so-called statistical hypothesis tests, aka as significance testing, where the p-values are misused. Another case is probability puzzles. We show it using so-called Monty Hall problem. In fact, it is often referred as a puzzle since people do not analyze it properly. Solution to the puzzle is based on correct identification of conditional and marginal probabilities. Then, we show that it needs to inflate p-values that we calculate in order to reduce uncertainty in hypothesis test conclusions. Finally, we talk about how the probability theory can be used in uncertainty handling in predictions. In this case so-called probabilistic Bayesian networks are a better tool, therefore we show a simple way to express the uncertainty in their prediction which is otherwise being done in complicated ways or is ignored. This is done through defining a virtual parameter for the predictive probability distribution, which is a function of actual parameters of the network. It is shown how this method can be extended to deep neural networks.

Uncertainty in P-Values

An instance where insufficient handling of uncertainty is seen is in so-called replication studies in experimental sciences (Gibson 2021). There is a huge blame is put on so-called p-values in statistical null hypothesis testing. Weak uncertainty handling in the inference process causes the replication problems. We argue that the p-values that we obtained are often under some assumptions but there is some uncertainty in those assumptions, which we ignore. Therefore, our calculated p-value should be adjusted for compensating the uncertainty in our research design or in the modeling process. Suppose we are performing a one-sided T-test where we assume that we have a random sample of data of size n from the relevant population. Note that test statistics T has a t-distribution with $(n-1)$ degrees of freedom under the null hypothesis H_0 and then the p-value of the test is $p = P\{T \geq t_{ob} | H_0\}$ where t_{ob} is the observed value of T. However, practically we are not certain that our data are a random sample. Therefore, our observed p-value should be written as, $p_{ob} = P\{T \geq t_{ob} | H_0 \wedge R\}$ where R denotes the proposition that random sample assumption is true and $\wedge$ is the logical "AND" operator. We see that the true p-value must be at least the calculated value because,

$$p \geq P\{T \geq t_{ob} \wedge R | H_0\}$$
$$= \frac{P\{T \geq t_{ob} |\, H_0 \wedge R\}}{P\{R|H_0\}} = \frac{p_{ob}}{P\{R\}} \geq p_{ob}$$

The Monty Hall Puzzle

So-called Monty Hall problem or puzzle (Freedman 1998) is said to fool even trained human mind. Suppose you are on a game show. And you are given choices of three doors to select one, where behind one of the doors there is a car and behind the other two doors there are two goats, one for each. Trick is to select the door that has the car behind it. So, you select one of the three doors (say, Door-1). Then the game host reveals one non-selected door (say, Door-3) which does not have the car behind it. At this point, you can choose whether to stick with your original choice (i.e., Door-1), or switch to the remaining door (i.e. Door-2). What are the probabilities that you will win the car if you stick, versus if you switch?

Most people believe, upon first hearing of this problem, that the car is equally likely to be behind either of the two unopened doors, so the probability of winning is $1/2$ regardless of whether you stick or switch. However, in fact the probabilities of winning are $1/3$ if you stick, and $2/3$ if you switch. We can show that the two probabilities, $1/2$ and $2/3$ refer two different events. The confusion comes up when these two well-defined probabilities are interpreted as if they mean the same! The value $1/2$ corresponds to an unconditional event of winning the car on average (switching or sticking), and the value $2/3$ corresponds to a conditional event of winning the car if switch. Similarly, $1/3$ corresponds to a conditional event of winning the car if stick. So, Monty Hall problem is all about identifying what those probabilities mean.

Probabilistic Bayesian Network Prediction

In this section we talk about quantifying uncertainty of a probabilistic prediction. For simplicity we assume that all the random variables are discrete, therefore our prediction is categorical that means that it is a classification. A probabilistic Bayesian network (BN) (Cowell et al. 1999) that is used extensively for classification tasks represents a factorization of joint probability distribution of a given set of random variables based on conditional independences inherited in them. For a random variable vector of length n, say, $X_{[n]} = (X_1, \dots, X_n)$ we can write a BN representation as $p(x_{[n]}) = \prod_{i=1}^{n} p(x_i | x_{pa(i)})$ where $p(x_1 | x_{pa(1)}) = p(x_1)$ and $x_{pa(i)} \subseteq \{x_1, \dots, x_{i-1}\}$ for $i = 2, \dots, n$.

We assume multinomial distributions for X_1 and every conditional variable $X_i | X_{pa(i)}$ for $i = 2, \dots, n$ for parameterizing the BN. Uncertainty of every parameter can be model with a Dirichlet distribution whose parameters' sum corresponds to the "precision" of its component probabilities. When the BN used as a classifier, the classifying probability is just a function of BN parameters. Therefore, uncertainty in classifying probabilities can be model easily if we define a virtual parameter for the classifying probability distribution. When have done so, we can define the precision of it using, for example, the smallest precision in the answer to the respective probabilistic query. However, it will be a conservative uncertainty estimation for the probabilistic query. Our approach is easier to implement than those methods that are currently being used (Allen, et al. 2008).

Deep Neural Network Prediction

Often a deep neural network can have millions of parameters, therefore quantifying its prediction uncertainty through its parameters can be hard. See the paper Abdar et al. (2021) for a comprehensive review on the topic. However, to our surprise researchers often follow this line for doing it. But mathematically, we can define the feature space of the neural network classifier through output signal space of its last layer. Note that these signals are the input to the classification node of the network, that often uses so-called softmax activation function to generate the classifying probabilities. Since input to a deep neural network is high-dimensional (such as images, etc.) we can consider the whole network as a dimension reduction tool in this way. The reduced feature space is the output signal space of the last layer of the network.

Once we have derived the feature space we can define a reduced parameter set for the whole network, whose values ranges over the feature space. Note that these parameters are virtual, but their state spaces are not, i.e., they are observable. For a given training dataset, we should be able to count different configurations of these virtual parameters and output variable that has been predicted correctly or similar way. Thus, we can define precisions of softmax probabilities that can be transformed into a Dirichlet distribution. By this way, we have reduced a deep neural network classification task as multinomial and Dirichlet probability model. Uncertainty quantification is done with simple probabilistic calculations as in the case of Bayesian networks.

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavanmzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., and Nahavandi, S. 2021. A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges. *Information Fusion* 76: 243–297.

Allen, T. V., Singh, A., Greiner, R., and Hooper, P. 2008. Quantifying the Uncertainty of a Belief Net Response: Bayesian Errorbars for Belief Net Inference. *Artificial Intelligence* 172: 483–513.

Cowell, R.G., Dawid, P., Lauritzen, S. L. and Spiegelhalter, D. J. 1999. *Probabilistic Networks and Expert Systems.* New York: Springer-Verlag.

Freedman, D. 1998. Monty Hall's Three Doors: Construction and Deconstruction of Choice Anomaly. *The American Economic Review.* 88: 933–946.

Gibson, E. W. 2021. The Role of *p*-Values in Judging the Strength of Evidence and Realistic Expectations. *Statistics in Biopharmaceutical Research.* 13(1): 6–18.