Bayesian Theory

— Exercises with Solutions —

Laurenz Wiskott
Institut für Neuroinformatik
Ruhr-Universität Bochum, Germany, EU

4 February 2017

Contents

These exercises complement my corresponding lecture notes, and there is a version with and one without solutions. The table of contents of the lecture notes is reproduced here to give an orientation when the exercises can be reasonably solved. For best learning effect I recommend to first seriously try to solve the exercises yourself before looking into the solutions.
More teaching material is available at https://www.ini.rub.de/PEOPLE/wiskott/Teaching/Material/.

1 Bayesian inference

1.1 Discrete random variables and basic Bayesian formalism

Joint probability

1.1.1 Exercise: Heads-tails-tails-heads

1. With four tosses of a fair coin, what is the probability to get exactly heads-tails-tails-heads, in this order?

   Solution: Each toss is independent of the others and the probability for each toss to get the desired result is $\frac{1}{2}$. Thus, the probability to get exactly heads-tails-tails-heads is $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$. This, by the way, holds for any concrete combination of length four.

2. With four tosses of a fair coin, what is the probability to get each heads and tails twice, regardless of the order?

   Solution: The probability for any particular combination of four times heads or tails is $\frac{1}{16}$, see above. Since there are six different ways to get heads and tails twice (namely tthh, thth, thht, htth, htht, hhtt), the probability to get any of these is $\frac{6}{16} = \frac{3}{8}$.

Total probability

1.1.2 Exercise: Election and Bridge

Three candidates run for an election as a major in a city. According to a public opinion poll their chances to win are 0.25, 0.35 und 0.40. The chances that they build a bridge after they have been elected are 0.60, 0.90 und 0.80. What is the probability that the bridge will be build after the election.

Solution: Let $C, c \in \{1, 2, 3\}$, be the random variable indicating the winning candidate and $B, b \in \{t, f\}$, the random variable indicating whether the bridge will be built. Then the total probability that the bridge will be built is

$$P(B = t) = \sum_{c=1}^{3} P(B = t|c)P(c) = 0.60 \times 0.25 + 0.90 \times 0.35 + 0.80 \times 0.40 = 0.785\,.$$

Bayes formula

1.1.3 Exercise: Bayes theorem in four variables

Consider four random variables A, B, C, and D. Given are the (marginal) joint probabilities for each pair of variables, i.e. probabilities of the form $P(A, B), P(A, C)$ etc., and the conditional probability $P(A, B|C, D)$. Calculate $P(A, C|B, D)$.

Solution:

$$\begin{aligned} P(A, C|B, D) &= \frac{P(A, B, C, D)}{P(B, D)} && (1) \\ &= \frac{P(A, B|C, D)P(C, D)}{P(B, D)}\,. && (2) \end{aligned}$$

1.1.4 Exercise: Airport security

On an airport all passengers are checked carefully. Let T with $t \in \{0, 1\}$ be the random variable indicating whether somebody is a terrorist ($t = 1$) or not ($t = 0$) and A with $a \in \{0, 1\}$ be the variable indicating arrest. A terrorist shall be arrested with probability $P(A = 1|T = 1) = 0.98$, a non-terrorist with probability $P(A = 1|T = 0) = 0.001$. One in hundredthousand passengers is a terrorist, $P(T = 1) = 0.00001$. What is the probability that an arrested person actually is a terrorist?

Solution: This can be solved directly with the Bayesian theorem.

$$
\begin{aligned}
P(T = 1|A = 1) &= \frac{P(A = 1|T = 1)P(T = 1)}{P(A = 1)} && (1)\\
&= \frac{P(A = 1|T = 1)P(T = 1)}{P(A = 1|T = 1)P(T = 1) + P(A = 1|T = 0)P(T = 0)} && (2)\\
&= \frac{0.98 \times 0.00001}{0.98 \times 0.00001 + 0.001 \times (1 - 0.00001)} = 0.0097 && (3)\\
&\approx \frac{0.00001}{0.001} = 0.01 && (4)
\end{aligned}
$$

It is interesting that even though for any passenger it can be decided with high reliability (98% and 99.9%) whether (s)he is a terrorist or not, if somebody gets arrested as a terrorist, (s)he is still most likely not a terrorist (with a probability of 99%).

1.1.5 Exercise: Drug Test

A drug test (random variable T) has 1% false positives (i.e., 1% of those not taking drugs show positive in the test), and 5% false negatives (i.e., 5% of those taking drugs test negative). Suppose that 2% of those tested are taking drugs. Determine the probability that somebody who tests positive is actually taking drugs (random variable D).

Solution:

$T = p$ means Test positive,
$T = n$ means Test negative,
$D = p$ means person takes drug,
$D = n$ means person does not take drugs

We know:

$$
\begin{aligned}
P(T = p|D = n) &= 0.01 \quad \text{(false positives)} && (1)\\
\text{(false negatives)} \quad P(T = n|D = p) &= 0.05 \Longrightarrow P(T = p|D = p) = 0.95 \quad \text{(true positives)} && (2)\\
P(D = p) &= 0.02 \Longrightarrow P(D = n) = 0.98 && (3)\\
& && (4)
\end{aligned}
$$

We want to know the probability that somebody who tests positive is actually taking drugs:

$$
P(D = p|T = p) = \frac{P(T = p|D = p)P(D = p)}{P(T = p)} \text{(Bayes theorem)} \qquad (5)
$$

We do not know $P(T = p)$:

$$
P(T = p) = P(T = p|D = p)P(D = p) + P(T = p|D = n)P(D = n) \qquad (6)
$$

We get:

$$
\begin{aligned}
P(D=p|T=p) &= \frac{P(T=p|D=p)P(D=p)}{P(T=p)} && (7) \\
&= \frac{P(T=p|D=p)P(D=p)}{P(T=p|D=p)P(D=p)+P(T=p|D=n)P(D=n)} && (8) \\
&= \frac{0.95 \cdot 0.02}{0.95 \cdot 0.02 + 0.01 \cdot 0.98} && (9) \\
&= 0.019/0.0288 \approx 0.66 && (10)
\end{aligned}
$$

There is a chance of only two thirds that someone with a positive test is actually taking drugs.

An alternative way to solve this exercise is using decision trees. Let's assume there are 1000 people tested. What would the result look like?

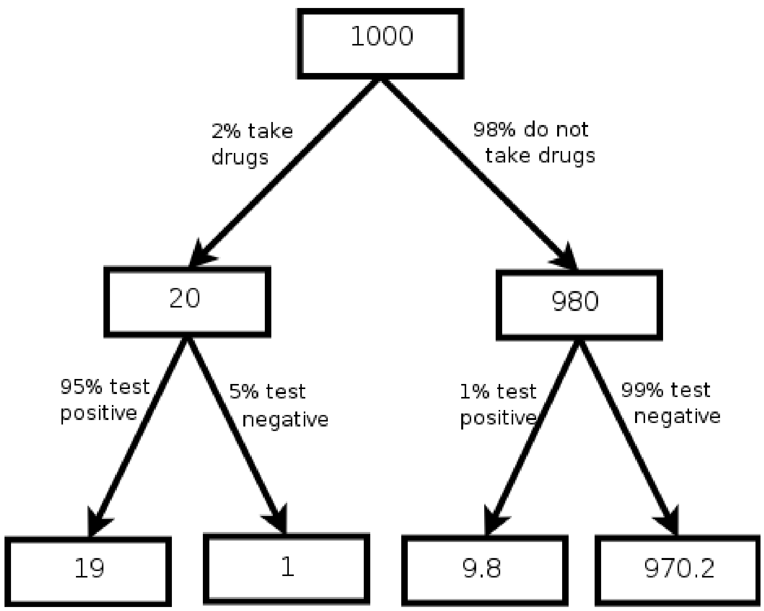


Figure: (Uknown)

Now we can put this together in a contingency table:

	$D=p$	$D=n$	sum
$T=p$	19	9.8	28.8
$T=n$	1	970.2	971.2
sum	20	980	1000

To determine the probability that somebody who tests positive is actually taking drugs we have to calculate:

$$\frac{\text{taking drugs and positive test}}{\text{all positive test}} = \frac{19}{28.8} \approx 0.66 \quad (11)$$

1.1.6 Exercise: Oral Exam

In an oral exam you have to solve exactly one problem, which might be one of three types, A, B, or C, which will come up with probabilities 30%, 20%, and 50%, respectively. During your preparation you have solved 9 of 10 problems of type A, 2 of 10 problems of type B, and 6 of 10 problems of type C.

(a) What is the probability that you will solve the problem of the exam?

Solution: The probability to solve the problem of the exam is the probability of getting a problem of a certain type times the probability of solving such a problem, summed over all types. This is known as the total probability.

$$\begin{aligned} P(\text{solved}) &= P(\text{solved}|\text{A})P(\text{A}) + P(\text{solved}|\text{B})P(\text{B}) + P(\text{solved}|\text{C})P(\text{C}) && (1) \\ &= 9/10 \cdot 30\% + 2/10 \cdot 20\% + 6/10 \cdot 50\% && (2) \\ &= 27/100 + 4/100 + 30/100 = 61/100 = 0.61\,. && (3) \end{aligned}$$

(b) Given you have solved the problem, what is the probability that it was of type A?

Solution: For this to answer we need Bayes theorem.

$$\begin{aligned} P(\text{A}|\text{solved}) &= \frac{P(\text{solved}|\text{A})P(\text{A})}{P(\text{solved})} && (4) \\ &= \frac{9/10 \cdot 30\%}{61/100} = \frac{27/100}{61/100} = \frac{27}{61} = 0.442...\,. && (5) \\ & && (6) \end{aligned}$$

So we see that given you have solved the problem, the *a posteriori* probability that the problem was of type A is greater than its *a priori* probability of 30%, because problems of type A are relatively easy to solve.

1.1.7 Exercise: Radar station

Consider a radar station monitoring air traffic. For simplicity we chunk time into periods of five minutes and assume that they are independent of each other. Within each five minute period, there may be an airplane flying over the radar station with probability 5%, or there is no airplane (we exclude the possibility that there are several airplanes). If there is an airplane, it will be detected by the radar with a probability of 99%. If there is no airplane, the radar will give a false alarm and detect a non-existent airplane with a probability of 10%.

1. How many airplanes fly over the radar station on average per day (24 hours)?

   Solution: There are $24 \times 12 = 288$ five-minute periods per day. In each period there is a probability of 5% for an airplane being present. Thus the average number of airplanes is $288 \times 5\% = 288 \times 0.05 = 14.4$.

2. How many false alarms (there is an alarm even though there is no airplane) and how many false no-alarms (there is no alarm even though there is an airplane) are there on average per day.

   Solution: On average there is no airplane in $288 - 14.4$ of the five-minute periods. This times the probability of 10% per period for a false alarm yields $(288 - 14.4) \times 10\% = 273.6 \times 0.1 = 27.36$ false alarms.

   On average there are 14.4 airplanes, each of which has a probability of 1% of getting missed. Thus the number of false no-alarms is $14.4 \times 1\% = 14.4 \times 0.01 = 0.144$.

3. If there is an alarm, what is the probability that there is indeed an airplane?

   Solution: For this question we need Bayes theorem.

$$
\begin{aligned}
& P(\text{airplane}|\text{alarm}) && (1)\\
=\; & \frac{P(\text{alarm}|\text{airplane})P(\text{airplane})}{P(\text{alarm})} && (2)\\
=\; & \frac{P(\text{alarm}|\text{airplane})P(\text{airplane})}{P(\text{alarm}|\text{airplane})P(\text{airplane}) + P(\text{alarm}|\text{no airplane})P(\text{no airplane})} && (3)\\
=\; & \frac{0.99 \cdot 0.05}{0.99 \cdot 0.05 + 0.1 \cdot (1 - 0.05)} = 0.342... && (4)\\
\approx\; & \frac{0.05}{0.05 + 0.1} = 0.333.... && (5)
\end{aligned}
$$

   It might be somewhat surprising that the probability of an airplane being present given an alarm is only 34% even though the detection of an airplane is so reliable (99%). The reason is that airplanes are not so frequent (only 5%) and the probability for an alarm given no airplane is relatively high (10%).

Miscellaneous

1.1.8 Exercise: Gambling machine

Imagine a simple gambling machine. It has two display fields that can light up in red or green. The first one lights up first with green being twice as frequent as red. The color of the second field depends on the first one. If the first color is red, green appears five times as often as red in the second field. If the first color is green, the two colors are equally likely.

A game costs 8€ and goes as follows. The player can tip right in the beginning on both colors, or he can tip the second color after he sees the first color, or he can tip not at all. He is allowed to decide on when he tips during the game. The payout for the three tip options is different of course, highest for tipping two colors and lowest for no tip at all.

1. To get a feeling for the question, first assume for simplicity that each color is equally likely and the second color is independent of the first one. How high must the payout for each of the three tip options be, if the tip is correct, to make the game just worth playing?

   Solution: If all colors are equally likely, then one would tip a two-color combination correctly with probability 1/4, the second color alone with 1/2, and no color with certainty. Thus the payout if the tip is correct must be a bit more than 32€, 16€, and 8€, respectively, to make the game worth playing.

2. Do the chances to win get better or worse if the colors are not equally likely anymore but have different probabilities and you know the probabilities? Does it matter whether the two fields are statistically independent or not?

   Solution: If the probabilities different, then some combinations are more frequent than others. If one systematically tips these more frequent combinations, the mean payout is increased. Thus, chances get better.

3. Given the payouts for the three tip options are 20€, 12€, and 7€. What is the optimal tip strategy and what is the mean payout?

   Solution: The solution to this question can be put in a table.

<table>
<tr><td>cost of one game</td><td colspan="4">-8€</td></tr>
<tr><td>best tip now</td><td colspan="4">green-red or green-green</td></tr>
<tr><td>mean payout for best tip now</td><td colspan="4">$+6/18 \cdot 20€ = +6\frac{2}{3}€$</td></tr>
<tr><td>mean payout for best tip later</td><td colspan="4">$+1/3 \cdot 10€ + 2/3 \cdot 7€ = +24/3€ = +8€$</td></tr>
<tr><td>prob. of first color</td><td colspan="2">1/3 red</td><td colspan="2">2/3 green</td></tr>
<tr><td>best tip now</td><td colspan="2">green</td><td colspan="2">red or green</td></tr>
<tr><td>mean payout for best tip now</td><td colspan="2">$+5/6 \cdot 12€ = +10€$</td><td colspan="2">$+3/6 \cdot 12€ = +6€$</td></tr>
<tr><td>mean payout for no tip now</td><td colspan="2">+7€</td><td colspan="2">+7€</td></tr>
<tr><td>prob. of second color given first</td><td>1/6 red</td><td>5/6 green</td><td>3/6 red</td><td>3/6 green</td></tr>
<tr><td>prob. of color combination</td><td>1/18 red-red</td><td>5/18 red-green</td><td>6/18 green-red</td><td>6/18 green-green</td></tr>
<tr><td>payout for no tip</td><td>+7€</td><td>+7€</td><td>+7€</td><td>+7€</td></tr>
</table>

Thus the best strategy is not to tip initially and then tip green as the second color if red comes up as the first color. If green comes up as the first color, don't tip at all. The mean payout of this optimal strategy is 8€, which just cancels the costs of the game.

1.1.9 Exercise: Probability theory

1. A friend offers you a chance to win some money by betting on the pattern of heads and tails shown on two coins that he tosses hidden from view. Sometimes he is allowed to give you a hint as to the result, sometimes not. Calculate the following probabilities:

   (a) If your friend stays silent, the probability that the coins show TT.

   (b) If your friend stays silent, the probability that the coins show HT in any order.

   (c) If you friend tells you that at least one of the coins is an H, the probability that the coins show HH.

2. Your friend now invents a second game. This time he tosses a biased coin which can produce three different results. The coin shows an H with probability 0.375 and a T with probability 0.45. The rest of the time the coin shows neither an H nor a T but lands on its side. A round of the game consists of repeatedly tossing the coin until either an H or a T comes up, whereupon the round ends.

   (a) Calculate the probability the coin needs to be tossed more than three times before either an H or a T comes up and the round ends.

   (b) Your friend proposes that if a T comes up you have to pay him 8 Euros, while if an H comes up he has to pay you 10 Euros. Calculate the expectation value of your winnings per round when playing this game. Would you agree to play using these rules?

   (c) Assume you agree with your friend that if a T comes up you have to pay him 10 Euros. What is the minimum amount you should receive if an H comes up in order to give a positive expectation value for your winnings per round?

   (d) Your friend now produces a coin which is always either an H or a T. In other words, it cannot land on its side. He claims that using this new coin eliminates the need to re-toss the coin without changing the statistics of the game in any other way. Assuming this is true, what is the probability of getting an H and a T on this new coin?

Solution: Not available!

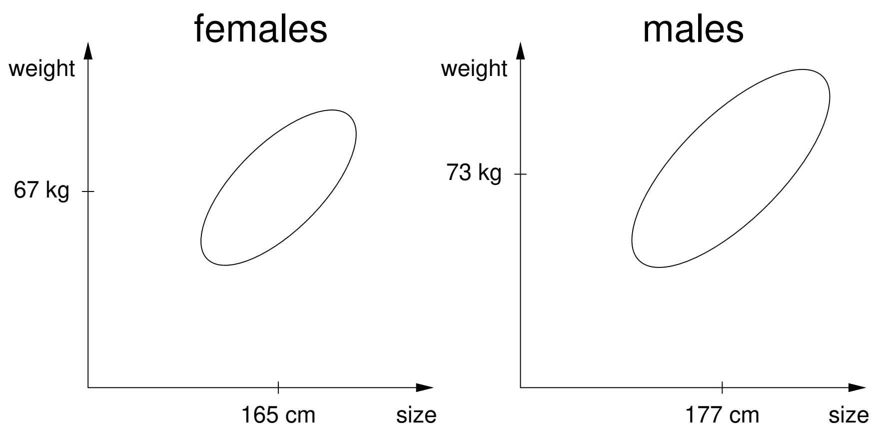
1.2 Partial evidence

1.3 Expectation values

1.4 Continuous random variables

1.4.1 Exercise: Probability densities

Let w, s, and G be random variables indicating body weight, size, and gender of a person. Let $p(w, s|G = f)$ and $p(w, s|G = m)$ be the conditional probability densities over weight and size for females (f) and males (m), respectively, in the shape of Gaussians tilted by 45°, see figure.



1. What is the probability $P(w = 50\text{kg}, s = 156\text{cm}|G = f)$?

   Solution: The probability that the weight is exactly 50kg and the size exactly 156cm is zero.

2. What is the probability $P(w \in [49\text{kg}, 51\text{kg}], s \in [154\text{cm}, 158\text{cm}]|G = f)$?

   Hint: You don't need to calculate a value here. Give an equation.

   Solution: This probability can be calculated by integrating the probability densities over the respective intervals.

$$P(w \in [49\text{kg}, 51\text{kg}], s \in [154\text{cm}, 158\text{cm}]|G = f) = \int_{49}^{51} \int_{154}^{158} p(w, s|G = f)\, \mathrm{d}s\, \mathrm{d}w\,. \tag{1}$$

3. Are weight and size statistically independent? Explain your statement.

   Solution: No, tall persons are typically heavier than short persons.

4. Can you derive variables that are statistically independent?

   Solution: Yes, for instance the weighted sum of weight and size in the principal direction of the Gaussians is statistically independent of the weighted sum orthogonal to that direction.

1.4.2 Exercise: Maximum likelihood estimate

Given N independent measurements $x_1, \ldots, x_N$. As a model of this data we assume the Gaußian distribution

$$p(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \, \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

1. Determine the probability density function $p(x_1, \ldots, x_N)$ of the set of data points $x_1, \ldots, x_N$ given the two parameters μ, σ^2.

   Solution: The probability density function is simply the product of the probabilities of the individual data points, which is given by the Gaußian.

   $$\begin{aligned} p(x_1, \ldots, x_N) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \, \mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}} & (1) \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^N} \, \mathrm{e}^{-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}} \, . & (2) \end{aligned}$$

2. Determine the natural logarithm (i.e. ln) of the probability density function of the data given the parameters.

   Solution:

   $$\begin{aligned} \ln(p(x_1, \ldots, x_N)) &= \ln\left(\frac{1}{(\sqrt{2\pi\sigma^2})^N} \, \mathrm{e}^{-\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2}} \right) & (3) \\ &= -\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^2} - \frac{N}{2} \ln\left(2\pi\sigma^2\right) \, . & (4) \end{aligned}$$

3. Determine the optimal parameters of the model, i.e. the parameters that would maximize the probability density determined above. It is equivalent to maximizing the logarithm of the pdf (since it is a strictly monotonically increasing function).

   Hint: Calculate the derivative wrt the parameters (i.e. $\frac{\partial}{\partial\mu}$ and $\frac{\partial}{\partial(\sigma^2)}$).

   Solution: Taking the derivatives and setting them to zero yields

   $$\begin{aligned} 0 \overset{!}{=} \frac{\partial \ln(p(x_1, \ldots, x_N))}{\partial\mu} &\overset{(4)}{=} \frac{\sum_{i=1}^{N}(x_i-\mu)}{\sigma^2} & (5) \\ \iff \quad \mu &= \frac{1}{N} \sum_{i=1}^{N} x_i \, , & (6) \\ 0 \overset{!}{=} \frac{\partial \ln(p(x_1, \ldots, x_N))}{\partial(\sigma^2)} &\overset{(4)}{=} \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{2\sigma^4} - \frac{N}{2}\frac{2\pi}{2\pi\sigma^2} & (7) \\ \iff \quad \sigma^2 &= \frac{1}{N} \sum_{i=1}^{N} (x_i-\mu)^2 \, . & (8) \end{aligned}$$

   Interestingly, μ becomes simply the mean and σ^2 the variance of the data.

   Extra question: Why is often a factor of $1/(N-1)$ used instead of $1/N$ in the estimate of the variance?

   Extra question: How does $p(x_1, ..., x_N)$ depend on the number of data points?

   Extra question: Can you make fits for different distributions comparable?

1.4.3 Exercise: Medical diagnosis

Imagine you go to a doctor for a check up to determine your health status H, i.e. whether you are sick ($H = \text{sick}$) or well ($H = \text{well}$). The doctor takes a blood sample and measures a critical continuous variable B. The probability distribution of the variable depends on your health status and is denoted by $p(B|H)$, concrete values are denoted by b. The *a priori* probability for you to be sick (or well) is indicated by $P(H)$.

In the following always arrive at equations written entirely in terms of $p(B|H)$ and $P(H)$. B and H may, of course, be replaced by concrete values.

1. What is the probability that you are sick before you go to the doctor?

   Solution: If you do not know anything about B, the probability that you are sick is obviously the *a priori* probability $P(H = \text{sick})$.

2. If the doctor has determined the value of B, i.e. $B = b$, what is the probability that you are sick? In other words, determine $P(H = \text{sick}|B = b)$.

   Solution: We use Bayesian theory.

$$P(H = \text{sick}|B = b) = \frac{p(b|\text{sick})P(\text{sick})}{p(b)} \tag{1}$$

$$= \frac{p(b|\text{sick})P(\text{sick})}{p(b|\text{sick})P(\text{sick}) + p(b|\text{well})P(\text{well})} . \tag{2}$$

3. Assume the doctor diagnoses that you are sick if $P(H = \text{sick}|B = b) > 0.5$ and let D be the variable indicating the diagnosis. Given b, what is the probability that you are being diagnosed as being sick. In other words, determine $P(D = \text{sick}|B = b)$.

   Solution: For a given b one can calculate a concrete probability for being sick, i.e. $P(H = \text{sick}|B = b)$. Now since the diagnosis is deterministic we have

$$P(D = \text{sick}|B = b) = \Theta(P(H = \text{sick}|B = b) - 0.5) , \tag{3}$$

   with the Heaviside step function defined as $\Theta(x) = 0$ if $x < 0$, $\Theta(x) = 0.5$ if $x = 0$, and $\Theta(x) = 1$ otherwise.

4. Before having any concrete value b, e.g. before you go to the doctor, what is the probability that you will be diagnosed as being sick even though you are well? In other words, determine $P(D = \text{sick}|H = \text{well})$.

   Solution: We use the rule for the total probability.

$$P(D = \text{sick}|H = \text{well})$$

$$= \int P(D = \text{sick}|b)\, p(b|H = \text{well})\, \mathrm{d}b \tag{4}$$

$$\stackrel{(3)}{=} \int \Theta(P(H = \text{sick}|b) - 0.5)\, p(b|H = \text{well})\, \mathrm{d}b \tag{5}$$

$$\stackrel{(2)}{=} \int \Theta\left(\frac{p(b|\text{sick})P(\text{sick})}{p(b|\text{sick})P(\text{sick}) + p(b|\text{well})P(\text{well})} - 0.5\right) p(b|H = \text{well})\, \mathrm{d}b . \tag{6}$$

   Extra question: Here we have seen how to turn a continuous random variable into a descrete random variable. How do you describe the distribution of a continuous random variable that deterministically assumes a certain value?

1.4.4 Exercise: Bayesian analysis of a face recognition system

You run a company that sells a face recognition system called 'faceIt'. It consists of a camera and a software system that controls an entry door. If a person wants to identify himself to the system he gets a picture taken with the camera, the probe picture, and that picture is compared to a gallery of stored pictures, the gallery pictures. FaceIt gives a scalar score value between 0 and 1 for each comparison. If the probe picture and the gallery picture show the same person, the score is distributed like

$$p(s|\text{same}) = \alpha_s \exp(\lambda_s s)\,, \tag{1}$$

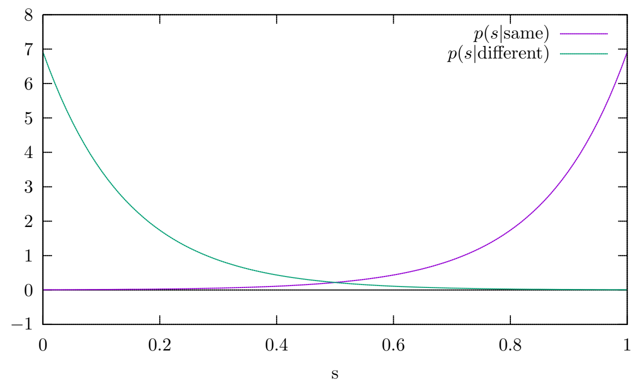
if they show different persons, the score is distributed like

$$p(s|\text{different}) = \alpha_d \exp(-\lambda_d s) \tag{2}$$

with some positive decay constants $\lambda_{\{s,d\}}$ and suitable normalization constants $\alpha_{\{s,d\}}$. All comparisons shall be independent of each other and the score depends only on whether the probe picture and the gallery picture show the same person or not.

1. Draw the score distributions and provide an intuition for why these pdfs might be reasonable.

   Solution:



   The pdfs are reasonable because if the persons are identical you get high probability densities for high scores and if the persons are identical you get high probability densities for low scores.

2. Determine the normalization constants $\alpha_{\{s,d\}}$ for given decay constants $\lambda_{\{s,d\}}$. First give general formulas and then calculate concrete values for $\lambda_s = \lambda_d = \ln(1000)$

   Solution: The normalization constant can be derived from the condition that the probability density function integrated over the whole range should be 1.

$$\begin{aligned}
1 &\stackrel{!}{=} \int_0^1 p(s|\text{same})\,\mathrm{d}s && (3)\\
&= \int_0^1 \alpha_s \exp(\lambda_s s)\,\mathrm{d}s = \alpha_s \left[\frac{1}{\lambda_s}\exp(\lambda_s s)\right]_0^1 = \frac{\alpha_s}{\lambda_s}(\exp(\lambda_s) - 1) && (4)\\
\iff \quad \alpha_s &= \frac{\lambda_s}{\exp(\lambda_s) - 1} = \frac{\ln(1000)}{1000 - 1} \approx 0.00691\,, && (5)\\
\text{and similarly} \quad \alpha_d &= \frac{-\lambda_d}{\exp(-\lambda_d) - 1} = \frac{\ln(0.001)}{0.001 - 1} \approx 6.91\,. && (6)
\end{aligned}$$

   With these constants we find that the two pdfs are actually mirrored versions of each other, since the exponents are the negative of each other and the normalization scales them to equal amplitude.

3. What is the probability that the score s is less (or greater) than a threshold θ if the probe picture and the gallery picture show the same person, and what if they show different persons. First give general formulas and then calculate concrete values for $\theta = 0.5$.

   Solution: This is straight forward. One simply has to integrate from the threshold to the upper or lower limit of the probability distribution.

$$\begin{aligned}
P(s < \theta|\text{same}) &= \int_0^\theta p(s|\text{same})\,\mathrm{d}s = \alpha_s \left[\frac{1}{\lambda_s}\exp(\lambda_s s)\right]_0^\theta && (7)\\
&= \frac{\alpha_s}{\lambda_s}(\exp(\lambda_s\theta) - 1) = \frac{\exp(\lambda_s\theta) - 1}{\exp(\lambda_s) - 1} \approx 0.031\,, && (8)\\
P(s > \theta|\text{same}) &= 1 - P(s < \theta|\text{same}) = \frac{(\exp(\lambda_s) - 1) - (\exp(\lambda_s\theta) - 1)}{\exp(\lambda_s) - 1} && (9)\\
&= \frac{\exp(\lambda_s) - \exp(\lambda_s\theta)}{\exp(\lambda_s) - 1} \approx 0.969\,, && (10)\\
P(s < \theta|\text{different}) &= \frac{\alpha_d}{-\lambda_d}(\exp(-\lambda_d\theta) - 1) = \frac{\exp(-\lambda_d\theta) - 1}{\exp(-\lambda_d) - 1} \approx 0.969\,, && (11)\\
P(s > \theta|\text{different}) &= 1 - P(s < \theta|\text{different}) = \frac{\exp(-\lambda_d) - \exp(-\lambda_d\theta)}{\exp(-\lambda_d) - 1} \approx 0.031\,. && (12)
\end{aligned}$$

   Notice that due to the finite probability densities it does not make any difference whether we write $<$ and $>$ or $<$ and $\geq$ or $\leq$ and $>$ in the probabilites on the lhs (left hand side).

4. Assume the gallery contains N pictures of N different persons (one picture per person). If N concrete score values $s_i, i = 1, ..., N$, are given and sorted to be in increasing order. What is the probability that gallery picture j shows the correct person? Assume that the probe person is actually in the gallery and that the *a priori* probability for all persons is the same. Give a general formula and calculate a concrete value for $N = 2$ and $s_1 = 0.3$ and $s_2 = 0.8$, and for $s_1 = 0.8$ and $s_2 = 0.9$ if $j = 2$.

Solution: This can be solved directly with Bayes' theorem.

$$
\begin{aligned}
& P(\mathrm{same}_j, \mathrm{different}_{i\neq j}|s_1, ..., s_N) && (13)\\
= & \frac{p(s_1, ..., s_N|\mathrm{same}_j, \mathrm{different}_{i\neq j})P(\mathrm{same}_j, \mathrm{different}_{i\neq j})}{p(s_1, ..., s_N)} && (14)\\
= & \frac{p(s_j|\mathrm{same})\left(\prod_{i\neq j} p(s_i|\mathrm{different})\right)(1/N)}{p(s_1, ..., s_N)} && (15)\\
& \quad \text{(since the scores are independent of each other and} \\
& \quad \text{the } \textit{a priory} \text{ probability is the same for all gallery images)} \\
\overset{(1,2)}{=} & \frac{\alpha_s \exp(\lambda_s s_j)\left(\prod_{i\neq j} \alpha_d \exp(-\lambda_d s_i)\right)(1/N)}{p(s_1, ..., s_N)} && (16)\\
= & \frac{\alpha_s \alpha_d^{N-1} \exp(\lambda_s s_j)\exp(\lambda_d(s_j - S))(1/N)}{p(s_1, ..., s_N)} && (17)\\
& \left(\text{with } S := \sum_i s_i\right) && (18)\\
= & \frac{\alpha_s \alpha_d^{N-1} \exp((\lambda_s + \lambda_d)s_j)\exp(-\lambda_d S)(1/N)}{p(s_1, ..., s_N)} && (19)\\
= & \frac{\alpha_s \alpha_d^{N-1} \exp((\lambda_s + \lambda_d)s_j)\exp(-\lambda_d S)(1/N)}{\sum_{j'} \alpha_s \alpha_d^{N-1} \exp((\lambda_s + \lambda_d)s_{j'})\exp(-\lambda_d S)(1/N)} && (20)\\
& \quad \text{(since } P \text{ must be normalized to 1)} \\
= & \frac{\exp((\lambda_s + \lambda_d)s_j)}{\sum_{j'} \exp((\lambda_s + \lambda_d)s_{j'})}\,. && (21)
\end{aligned}
$$

This is a neat formula. It is instructive to calculate the ratio between the probabilty that j is the correct gallery image and that k is the correct gallery image, which is

$$
\begin{aligned}
\frac{P(\mathrm{same}_j, \mathrm{different}_{i\neq j}|s_1, ..., s_N)}{P(\mathrm{same}_k, \mathrm{different}_{i\neq k}|s_1, ..., s_N)} &\overset{(21)}{=} \frac{\exp((\lambda_s + \lambda_d)s_j)}{\exp((\lambda_s + \lambda_d)s_k)} && (22)\\
&= \exp((\lambda_s + \lambda_d)(s_j - s_k))\,. && (23)
\end{aligned}
$$

We see that the ratio only depends on the difference between the score values but not on the values themselves. Thus, for the two examples given above it is clear that gallery image 2 is more likely the correct one in the first example even though its absolute score value is greater in the second example. We can verify this by calculating the actual probabilities.

$$
\begin{aligned}
P(\mathrm{different}_1, \mathrm{same}_2|s_1 = 0.3, s_2 = 0.8) &\overset{(21)}{=} \frac{\exp(2\ln(1000)0.8)}{\exp(2\ln(1000)0.3) + \exp(2\ln(1000)0.8)} && (24)\\
&\approx \frac{63096}{63159} \approx 0.9990\,, && (25)\\
P(\mathrm{different}_1, \mathrm{same}_2|s_1 = 0.8, s_2 = 0.9) &\overset{(21)}{=} \frac{\exp(2\ln(1000)0.9)}{\exp(2\ln(1000)0.8) + \exp(2\ln(1000)0.9)} && (26)\\
&\approx \frac{251189}{314284} \approx 0.7992\,. && (27)
\end{aligned}
$$

5. Without any given concrete score values, what is the probability that a probe picture of one of the persons in the gallery is recognized correctly if one simply picks the gallery picture with the highest score as the best guess for the person to be recognized. Give a general formula.

Solution: The probability of correct recognition is the probability density that the correct gallery picture gets a certain score s', i.e. $p(s'|\text{same})$, times the probability that all the other gallery pictures get score below s', i.e. $P(s < s'|\text{different})^{(N-1)}$, integrated over all possible scores s'. Note that the integration turns the probability density $p(s'|\text{same})$ into a proper probability.

$$
\begin{aligned}
& P(\text{correct recognition}) && (28) \\
= \; & \int_0^1 p(s'|\text{same}) P(s < s'|\text{different})^{(N-1)} \, \mathrm{d}s' && (29) \\
= \; & \int_0^1 \alpha_s \exp(\lambda_s s') \left(\frac{\exp(-\lambda_d s') - 1}{\exp(-\lambda_d) - 1} \right)^{(N-1)} \mathrm{d}s' && (30) \\
= \; & \int_0^1 \frac{\lambda_s \exp(\lambda_s s')}{\exp(\lambda_s) - 1} \left(\frac{\exp(-\lambda_d s') - 1}{\exp(-\lambda_d) - 1} \right)^{(N-1)} \mathrm{d}s' \quad \text{(this is ok as a solution)} && (31) \\
= \; & \underbrace{\frac{\lambda_s}{(\exp(\lambda_s) - 1)(\exp(-\lambda_d) - 1)^{(N-1)}}}_{=:A} \int_0^1 \exp(\lambda_s s')(\exp(-\lambda_d s') - 1)^{(N-1)} \, \mathrm{d}s' && (32) \\
= \; & \ldots \quad \text{(one could simplify even further)} && (33)
\end{aligned}
$$

1.5 A joint as a product of conditionals

1.6 Marginalization

2 Application: Visual attention +