DD2447 Statistical Methods in Applied Computer Science
Assignment 1

Leonardo De Clara, Riccardo Sena

November 20, 2023

Exercise 3

Exercise 3.1

If we use the exponential function as our bijective mapping we can define $y =$ $= f(x) = e^x$. We know that $p(x) = \mathcal{N}(x|\mu, \sigma^2)$ and we also know, using the transformation theorem, that $q(y) = p(f^{-1}(y))\left|\frac{df^{-1}(y)}{dy}\right|$.

We can compute the following:

- $f^{-1}(y) = log(y)$
- $\frac{df^{-1}(y)}{dy} = 1/y$

Substituting these two results in the previous expression we get:

$q(y) = \frac{1}{y\sqrt{2\pi\sigma^2}}e^{-\frac{log(y)-\mu^2}{2\sigma^2}}$

So $y \sim Lognormal(\mu_y, \sigma_y^2)$. This confirms the definition of the Lognormal distribution: the Lognormal distribution is a probability distribution of a random variable (in our case y) whose logarithm (in our case $x = log(y)$) is normally distributed.

Exercise 3.2

Since there is an earthquake every third year and we are interested in the probability of an earthquake happening in a single day:

$p(e = 1) = \epsilon = \frac{1}{3 \cdot 365}$

And since there is a burglary every second year:

$p(b=1) = \beta = \frac{1}{2 \cdot 365}$.

Our first objective is to compute the possible values for $p(a|b,e)$, which are 8 since the three variables are binary. Due to the fact that the events $E =$ $= (a=1|b=i,e=j)$ and $F = (a=0|b=i,e=j)$ are complementary we have that $p(a=1|b=i,e=j) = 1 - p(a=0|b=i,e=j)$. We can now list all the possible values for $p(a|b,e)$:

- $p(a=0|b=0,e=0) = 1-f = 0.999$
- $p(a=1|b=0,e=0) = f = 0.001$
- $p(a=0|b=1,e=0) = (1-f)(1-\alpha_b) = 0.00999$
- $p(a=1|b=1,e=0) = 1 - p(a=0|b=1,e=1) =$ $= 1-(1-f)(1-\alpha_b) = 0.99001$
- $p(a=0|b=0,e=1) = (1-f)(1-\alpha_e) = 0.98901$
- $p(a=1|b=0,e=1) = 1 - p(a=1|b=0,e=1) =$ $= (1-f)(1-\alpha_e) = 0.01099$
- $p(a=0|b=1,e=1) = (1-f)(1-\alpha_b)(1-\alpha_e) = 0.0098901$
- $p(a=1|b=1,e=1) = 1 - p(a=1|b=1,e=1) =$ $= 1-(1-f)(1-\alpha_b)(1-\alpha_e) = 0.9901099$

The first case is trivial since, if there isn't an earthquake and simultaneously a burglary is not happening then the alarm can only be triggered by some unknown reason. The next case is the probability that the alarm is not ringing given that there is a burglary but there isn't an earthquake. This event occurs if simultaneously neither the burglary nor the unknown reason activate the alarm. In the same way we can compute the probability of the alarm not ringing if there is an earthquake but there isn't a burglary. The only one that remains is the probability that the alarm does not ring if both an earthquake and a burglary are happening. Basically this takes place when not one event among burglary, earthquake and the unknown reason sets off the alarm.
The second task requires us to derive the posterior probability $p(b,e|a=1)$. The combination of the events b and e (which are independent) can have four different outcomes, since they are both binary variables: (0,0), (1,0), (0,1), (1,1). The expression of the posterior is therefore:

$$p(b=i,e=j|a=1) = \frac{p(a=1,b=i,e=j)}{p(a=1)} = \frac{p(a=1|b=i,e=j)p(b=i,e=j)}{p(a=1)} =$$

$$= \frac{(1-(1-f)(1-\alpha_b)^i(1-\alpha_e)^j)(\beta)^i(1-\beta)^{1-i}(\epsilon)^j(1-\epsilon)^{1-j}}{p(a=1)}, \text{ where i=0,1 and j=0,1.}$$

Here we can obtain $p(a=1)$ by marginalizing out the independent variables

b and e using the Law of Total Probability:
$p(a=1) = \sum_{i,j} p(a=1|b=i, e=j)p(b=i, e=j) =$
$= p(a=1|b=0, e=0)p(b=0)p(e=0) +$
$+ p(a=1|b=1, e=0)p(b=1)p(e=0) +$
$+ p(a=1|b=0, e=1)p(b=0)p(e=1) +$
$+ p(a=1|b=1, e=1)p(b=1)p(e=1) = 0.002364$

Finally we can compute $p(b=1|e=1, a=1)$. First, taking into account that e is independent of b, using the definition of conditional probability and applying Bayes' theorem we get:

$$p(b=1|e=1, a=1) = \frac{p(a=1,b=1,e=1)}{p(a=1,e=1)} = \frac{p(a=1|b=1,e=1)p(b=1,e=1)}{p(a=1|e=1)p(e=1)} = \frac{p(a=1|b=1,e=1)p(b=1)p(e=1)}{p(a=1|e=1)p(e=1)} =$$

$$= \frac{p(a=1|b=1,e=1)p(b=1)}{p(a=1|e=1)}$$

Next, using the Law of Total Probability we can compute the denominator by marginalizing out the variable b:

$$p(a=1|e=1) = \sum_{i \in \{0,1\}} p(a=1|e=1, b=i)p(b=i) =$$
$$= p(a=1|e=1, b=0)p(b=0) + p(a=1|e=1, b=1)p(b=1) =$$

$$= 0.01099 \cdot (1-\beta) + 0.9901099 \cdot \beta = 0.01233$$

Now we have everything to compute the probability that we are looking for:

$$p(b=1|e=1, a=1) = \frac{0.9901099 \cdot \beta}{0.01233} = 0.11.$$

Exercise 3.3

(i) The data is distributed as a multinomial with parameters $n_0 = 2000$, $k = 27$ and θ_k for all k=1,...,27. The prior distribution of our parameter vector θ is $p(\theta) \sim Dir(\alpha_1, ..., \alpha_{27})$. We know that the Dirichlet distribution is a conjugate prior to the multinomial distribution, which means that also the posterior probability $p(\theta|D)$ will be a Dirichlet distribution. First we need to compute the posterior distribution:

$$f(\theta|D) = \frac{f(D|\theta)f(\theta)}{f(D)} \propto f(D|\theta)f(\theta) \propto \frac{n_0!}{n_1!...n_k!} \prod_{i=1}^{K} (\theta_i)^{n_i} \prod_{i=1}^{K} (\theta_i)^{\alpha_i - 1} \propto \prod_{i=1}^{K} (\theta_i)^{\alpha_i + n_i - 1}$$

Where n_0 is the total number of samples, $n_i s$ are the counts of the old observations for each category i. If we set $\alpha'_i = \alpha_i + n_i$ then we have that $f(\theta|D) \sim Dir(\alpha'_1, ..., \alpha'_{27})$. Introducing the function $\mathrm{B}(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$, we can now compute the general predictive posterior distribution:

$f(X|D) = \int f(X|\theta)f(\theta|D)d\theta = \int \frac{x_0!}{x_1!...x_k!} \prod_{i=1}^{K} (\theta_i)^{x_i} \frac{1}{B(\alpha')} \prod_{i=1}^{K} (\theta_i)^{\alpha'_i - 1} d\theta =$

$= \frac{x_0!}{x_1!...x_k!} \frac{1}{B(\alpha')} \int \prod_{i=1}^{K} (\theta_i)^{x_i + n_i + \alpha_i - 1} d\theta$

We can notice that inside the integral we have a Dirichlet distribution without the normalizing constant. If we divide it by $B(x + \alpha + n)$ the integral will be exactly 1, since it is the integral of a distribution probability over its support. Therefore we get:

$$f(X|D) = \frac{x_0!}{x_1!...x_k!} \frac{1}{B(\alpha')} B(x + \alpha + n)$$

Where $x_i s$ are the counts of the new observation and x_0 is the total count of the new observation.

In our specific case we have $X_{2001} = e$, so we need to compute $f(X_{2001} = e|D)$:
$f(X_{2001} = e|D) = \int f(X_{2001} = e|\theta)f(\theta|D)d\theta = \int f(X_{2001} = e|\theta_e)f(\theta_e|D)d\theta_e$
Since the new observation depends only on θ_e we can avoid integrating for all values of θ. So the expression above becomes:

$f(X_{2001} = e|D) = \int \theta_e f(\theta_e|D)d\theta_e = \mathbb{E}[\theta_e|D] = \frac{\alpha_e + n_e}{\alpha_0 + n_0}$

Where we used the expected value of the Dirichlet distribution and where α_0 is the sum of all $\alpha_i s$.

So $p(X_{2001} = e|D) = \frac{\alpha_e + n_e}{\alpha_0 + n_0} = \frac{10+260}{270+2000} = 0.1189$

(ii) For this point we can directly use the form of the posterior predictive that we found above. For now we do not take into consideration the order of the new observations:

$p(X = (e, a)|D) = \frac{x_0!}{x_e! x_a!} \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i + n_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i + n_i)} \frac{\prod_{i=1}^{K} \Gamma(\alpha_i + n_i + x_i)}{\Gamma\left(\sum_{i=1}^{K} \alpha_i + n_i + x_i\right)}$

From the products the of the Gamma functions only the ones that have a different count after the new observation remain, hence the expression simplifies into:

$p(X = (e, a)|D) = \frac{x_0!}{x_e! x_a!} \frac{\Gamma(\alpha_0 + n_0)}{\Gamma(\alpha_0 + n_0 + x_0)} \frac{\Gamma(\alpha_e + n_e + x_e)}{\Gamma(\alpha_e + n_e)} \frac{\Gamma(\alpha_a + n_a + x_a)}{\Gamma(\alpha_a + n_a)}$

Thanks to the properties of the Gamma function and to the fact that both x_e and x_a are equal to 1 we get:

$p(X = (e, a)|D) = \frac{x_0!}{x_e! x_a!} \frac{(\alpha_e + n_e)(\alpha_a + n_a)}{(\alpha_0 + n_0 + 1)(\alpha_0 + n_0)} =$

$= \frac{2!}{1!1!} \frac{(10+260)(10+100)}{2271 \cdot 2270} = 0.01152.$

Since we care about the order of the new observations we can just divide by the multinomial coefficient, which in this case is just 2:

$p(X_{2001} = e, X_{2002}|D) = \frac{p(X=(e,a)|D)}{2} = 0.00576.$